Nomenclature Issues and the Y-Chromosome

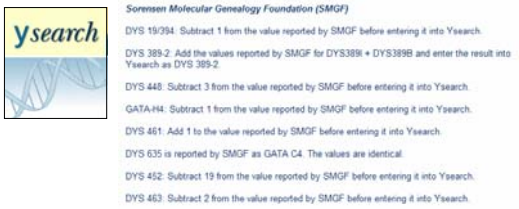
John M. Butler, Ph.D.
National Institute of Standards and Technology

4th International Conference on Genetic Genealogy
Houston, TX
October 20, 2007

The Problem

- **STR marker and allele nomenclatures differ between DNA testing laboratories** that make data comparisons frustrating at best

http://www.ysearch.org/conversion_page.asp



The Solution

- **Standardize the nomenclature** and encourage (or require) all testing laboratories to calibrate their results to this standard nomenclature

Questions to Address Regarding STR Marker and Allele Nomenclature

- Where do STR marker names come from?
- Who decides on STR allele nomenclature?
- How is calibration of STR allele calls performed in the forensic DNA community?
- What are some potential solutions to aid the genetic genealogy community in marker standardization and STR allele calibration?

Presentation Outline

- Introduction to myself and NIST
- Overview of DNA Typing Process
- STR Allele Repeat Nomenclature
- NIST Standard Reference Materials

NIST and NIJ Disclaimer

Funding: Interagency Agreement 2003-IJ-R-029 between the National Institute of Justice and NIST Office of Law Enforcement Standards

Points of view are mine and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

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.

Our Team Mission Statement

- The NIST Human Identity Project Team is trying **to lead the way in forensic DNA**… through research that helps bring traceability and technology to the scales of justice.

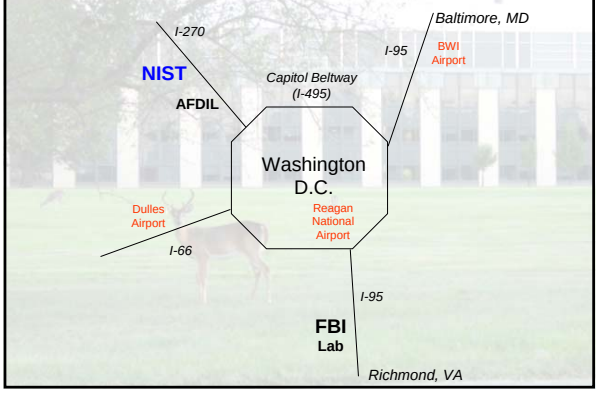
NIST History and Mission

- 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.**

$573 for 3 jars


DNA typing standard

Location of NIST (near Washington, DC)



NIST Gaithersburg Campus


Administration (Building 101)

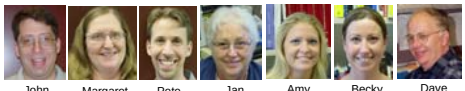
Located in Gaithersburg, Maryland, on approximately 234 hectares (578 acres) just off Interstate 270 about **25 miles northwest of Washington, D.C.**

http://www.nist.gov

~2,500 staff

Advanced Chemical Sciences Laboratory (Building 227)

NIST Human Identity Project Team


John Butler, Margaret Kline, Pete Vallone, Jan Redman, Amy Decker, Becky Hill, Dave Duewer

All NIST publications and presentations available on STRBase: http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

- **22 publications** since 2002 on Y-chromosome work
- **>40 presentations** and **10 training workshops** to the community in the past year

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
- **Training Materials**
  - Review articles and workshops on STRs, CE, validation
  - PowerPoint and pdf files available for download

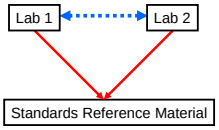
Standard Reference Materials

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



Calibration with SRMs enables confidence in comparisons of results between laboratories

Information Resources

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

Includes information on:
- Core STR loci
- Validation
- STR reference list
- NIST publications
- miniSTRs
- Forensic SNPs
- Variant STR alleles
- Population data resources
- Addresses of scientists

Provides up-to-date information and has been used in court cases to support application of DNA technology

Overview of DNA Typing Process

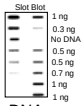
Steps in DNA Analysis

Usually 1-2 day process (a minimum of ~5 hours)

Steps Involved
- Collection
- Specimen Storage
- Extraction
- Quantitation
- Multiplex PCR
- STR Typing
- Interpretation of Results
- Database (Storage & Searching)
- Calculation of Match Probability

Blood Stain Buccal swab
Sample Collection & Storage

DNA Extraction



DNA Quantitation

Biology

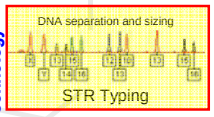
Multiplex PCR Amplification

Genetics

If a match occurs, comparison of DNA profile to population allele frequencies to generate a case report with probability of a random match to an unrelated individual

DNA Database Search

Technology

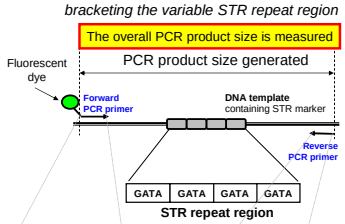


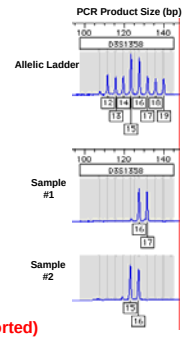
Male: 13,14-15,16-12,13-10,13-15,16

Interpretation of Results

Short Tandem Repeat (STR) Markers

PCR primers anneal to unique sequences bracketing the variable STR repeat region



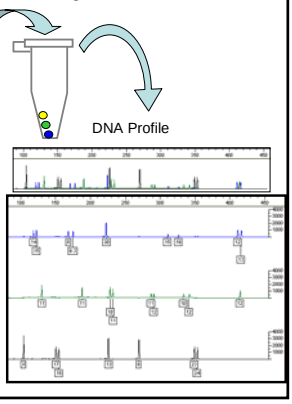


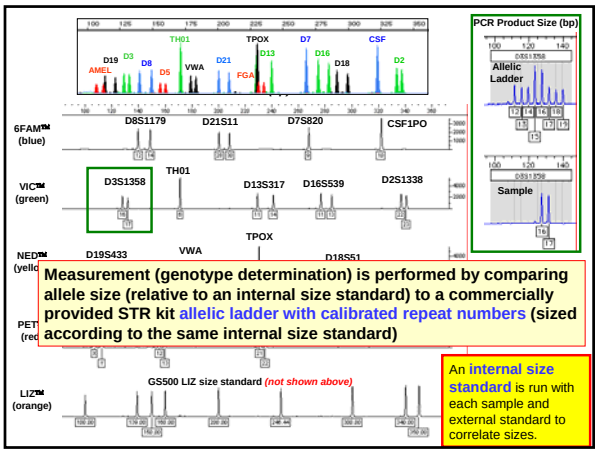
TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGA
AGACAGGTG**GATAGATAGATAGATAGATAGATAGATA**
GATAGATAGATAGATATCATTGAAAGACAAAACAGAGA
TGGATGATAGATACATGCTTACAGATGCACAC

= 11 GATA repeats ("11" is all that is reported)

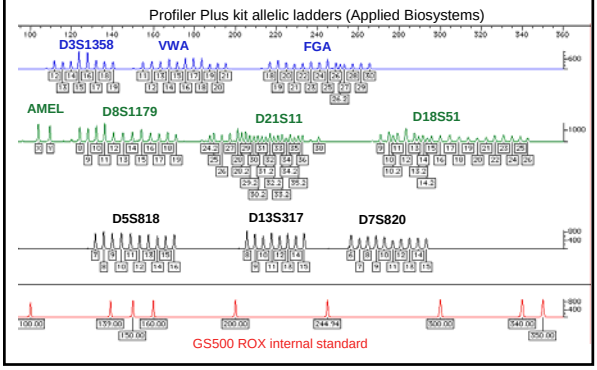
PCR Primers in an STR Kit

Locus	Dye	Promega PP16 Primer Sequences
D3S1358-F		ACTGCAGTCCAATCTGGGT
D3S1358-R	FL	ATGAAATCAACAGAGGCTTGC
TH01-F	FL	GTGATTCCCATTGGCCTGTTC
TH01-R		ATTCCTGTGGGCTGAAAAGCTC
D21S11-F		ATATGTGAGTCAATTCCCCAAG
D21S11-R	FL	TGTATTAGTCAATGTTCTCCAGAGAC
D18S51-F	FL	TTCTTGAGCCCAGAAGGTTA
D18S51-R		ATTCTACCAGCAACAACACAAATAAAC
PentaE-F		ATTACCAACATGAAAGGGTACCAATA
PentaE-R	FL	TGGGTTATTAATTGAGAAAACTCCTTACAATTT
D5S818-F		GGTGATTTTCCTCTTTGGTATCC
D5S818-R	JOE	AGCCACAGTTTACAACATTTGTATCT
D13S317-F		ATTACAGAAGTCTGGGATGTGGAGGA
D13S317-R	JOE	GGCAGCCCAAAAAGACAGA
D7S820-F	JOE	ATGTTGGTCAGGCTGACTATG
D7S820-R		GATTCCACATTTATCCTCATTGAC
D16S539-F		GGGGGTCTAAGAGCTTGTAAAAAG
D16S539-R	JOE	GTTTGTGTGTGCATCTGTAAGCATGTATC
CSF1PO-F	JOE	CCGGAGGTAAAGGTGTCTTAAAGT
CSF1PO-R		ATTTCCTGTGTCAGACCCTGTT
PentaD-F	JOE	GAAGGTCGAAGCTGAAGTG
PentaD-R		ATTAGAATTCTTTAATCTGGACACAAG
AMEL-F	TMR	CCCTGGGCTCTGTAAAGAA
AMEL-R		ATCAGAGCTTAAACTGGGAAGCTG
vWA-F		GCCCTAGTGGATGATAAGAATAATCAGTATGTG
vWA-R	TMR	GGACAGATGATAAATACATAGGATGGATGG
D8S1179-F	TMR	ATTGCAACTTATATGTATTTTTGTATTTCATG
D8S1179-R		ACCAAATTGTGTTCATGAGTATAGTTTC
TPOX-F		GCACAGAACAGGCACTTAGG
TPOX-R	TMR	CGCTCAAACGTGAGGTTG
FGA-F	TMR	GGCTGCAGGGCATAACATTA
FGA-R		ATTCTATGACTTTGCGCTTCAGGA

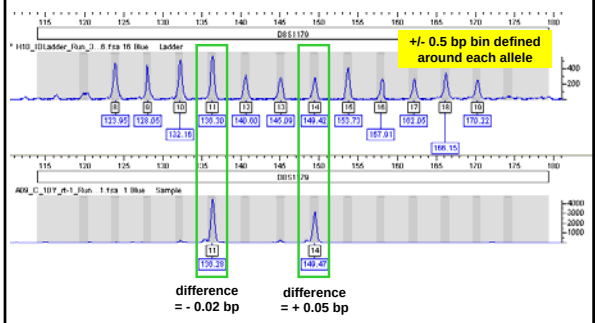




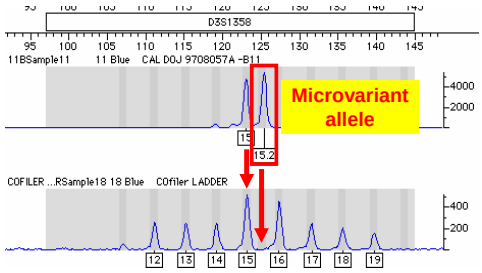
Companies Supply **Allelic Ladders** in STR Kits to **Aid Interlaboratory Consistency**



Comparison of Allelic Ladder to Samples to Convert Size into Allele Repeat Number

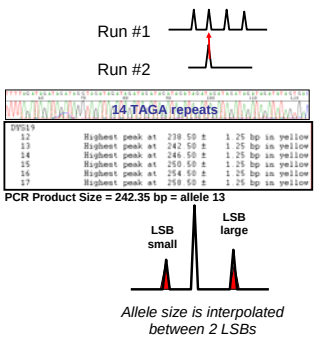


STR genotyping is performed by comparison of sample data to allelic ladders

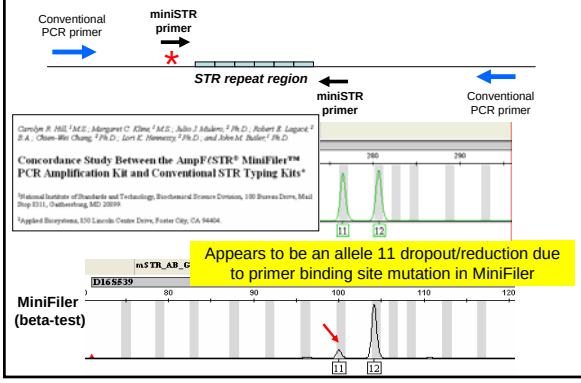


Methods for Typing STR Samples

- Allelic Ladders
  - Available in commercial kits
- Precise Sizing (and a sequenced sample)
- Locus-specific brackets (LSBs)



Concordance Studies Reveal Potential Primer Binding Site Mutations with Different Primer Sets



How Forensic DNA Typing Differs from Genetic Genealogy

- **Common set of core STR loci** are used to enable DNA database compatibility
- **Commercially available STR kits are used with allelic ladders** to ensure consistency between laboratories
- Required **calibration of STR allele calls to available NIST SRMs** to ensure accuracy

STR Allele Designation

Types of STR Repeat Units

Requires size based DNA separation to resolve different alleles from one another

- **Di**nucleotide (CA)(CA)(CA)(CA)
- **Tri**nucleotide (GCC)(GCC)(GCC)
- **Tetra**nucleotide (AATG)(AATG)(AATG)
- **Penta**nucleotide (AGAAA)(AGAAA)
- **Hexa**nucleotide (AGTACA)(AGTACA)

Short tandem repeat (STR) = microsatellite = simple sequence repeat (SSR)

Equivalent Tetranucleotide STR Motifs

Top Strand	Complementary (Bottom Strand)
• AGAT	• TCTA
• GATA	• CTAT
• ATAG	• TATC
• TAGA	• ATCT

Same DNA sequence with different starting points for naming the repeat)

…AGATAGATAGATATAGATAGATAG…

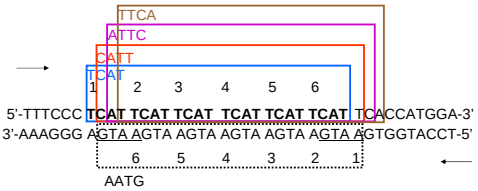
…AGATAGATAGATATAGATAGATAG…

…AGATAGATAGATATAGATAGATAG…

…AGATAGATAGATATAGATAGATAG…

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… **The NIST SRMs are viewed positively for aiding nomenclature disputes.**

STR Repeat Motifs

From D.N.A. Box 5.1 – J.M. Butler (2005) *Forensic DNA Typing*, 2nd Ed.

- Theoretically, there are 256 possible motifs for tetranucleotide repeats (e.g., GATA). However, because STRs are tandemly repeated, some motifs are actually equivalent to others.
- Two rules can be used to identify whether motif A is equivalent to motif B. Motif A is considered equivalent to motif B when (1) motif A is inversely complementary to motif B or (2) **motif A is different from motif B or the inversely complementary sequence of motif B by frameshift.**

Jin L, Zhong Y, Chakraborty R (1994) The exact numbers of possible microsatellite motifs [letter]. *Am. J. Hum. Genet.* 55:582-583

Categories for STR Markers

Category	Example Repeat Structure	13 CODIS Loci
Simple repeats – contain units of identical length and sequence	(GATA)(GATA)(GATA)	TPOX, CSF1PO, D5S818, D13S317, D16S539
Simple repeats with non-consensus alleles (e.g., TH01 9.3)	(GATA)(GAT-)(GATA)	TH01, D18S51, D7S820
Compound repeats – comprise two or more adjacent simple repeats	(GATA)(GATA)(GACA)	VWA, FGA, D3S1358, D8S1179
Complex repeats – contain several repeat blocks of variable unit length	(GATA)(GACA)(CA)(CATA)	D21S11

These categories were first described by **Urquhart *et al.* (1994) *Int. J. Legal Med.* 107:13-20**

STR Typing Measurement Issues

- STR genotypes are generated using PCR amplification and electrophoretic sizing that involves an internal size standard with each sample.
- The forensic DNA community almost exclusively uses STR typing kits to obtain results (there are different kits available that examine the same common markers).
- PCR amplification is expected to generate consistent genotypes as long as primer positions are not changed between kits. Primer changes can result in allele dropout due to primer site mutations.
- Occasionally new commercial kits are created with additional loci.
- General STR repeat nomenclature rules have been established but do have some subjectivity in them permitting possible differences in how STR alleles are named.

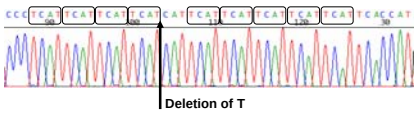
STR Repeat Nomenclature Rules

DNA Commission of the 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 **(as far as possible on the 5'end)**
- If two sequences are repeated (i.e., compound repeats), 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*

Microvariant Alleles

- Defined as alleles that are not exact multiples of the basic repeat motif or sequence variants of the repeat motif or both
- Alleles with partial repeat units are designated by the number of full repeats and then a decimal point followed by the number of bases in the partial repeat
- Example: **TH01 9.3 allele**: $[TCAT]_4$ -CAT $[TCAT]_5$



Variation in the Flanking Region Can Cause Variant Alleles

D7S820 Example

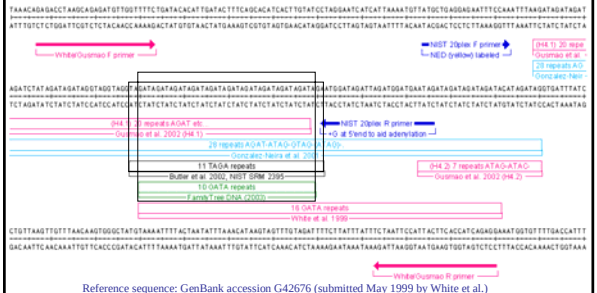
GATAGAACACTTGTCATAGTTTAGAACGAACTAACGATAGATAGATAGATAGATAGATAG
CTATCTTGTGAACAGTATCAAATCTTGCTTGATTGCTATCTATCTATCTATCTATCTATC

13 repeat units = $(GATA)_{13}$

TTTTTTTT x.3
TTTTTTTTT x.1

ATAGATAGATAGATAGATAGATAGATAGACAGATTGATAGTTTTTTTTTAATCTCACTAAA
TATCTATCTATCTATCTATCTATCTATCTGTCTAACTATCAAAAAAAAATTAGAGTGATTT

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

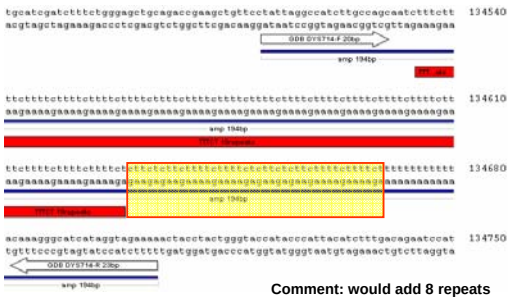


Reference sequence: GenBank accession G42676 (submitted May 1999 by White et al.)

This simple difference has impacted the genetic genealogy community

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

DYS714 Reference Sequence



Comment: would add 8 repeats

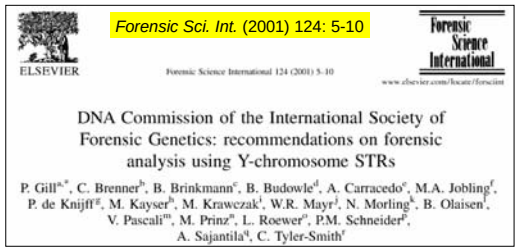
Alternate: $[TTTCT]_n[CTTCT]_2[TTTCT]_2[CTTCT]_2[TTTCT]_2$

19 + 8 = 27

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

Nomenclature Issues for Y STR Alleles

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002



Forensic Sci. Int. (2001) 124: 5-10

ELSEVIER

Forensic Science International

DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs

P. Gill[a,*], C. Brenner[b], B. Brinkmann[c], B. Budowle[d], A. Carracedo[e], M.A. Jobling[f], P. de Knijff[g], M. Kayser[h], M. Krawczak[i], W.R. Mayr[j], N. Morling[k], B. Olaisen[l], V. Pascali[m], M. Prinz[n], L. Roewer[o], P.M. Schneider[p], A. Sajantila[q], C. Tyler-Smith[r]

ISFG Guidelines for Y STRs

- Locus nomenclature should be DYS number if possible
- Allelic ladders should be used
- Allele nomenclature discussed…

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

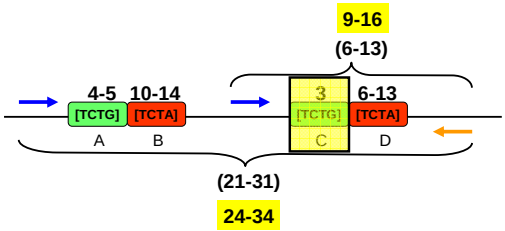
ISFG Guidelines for Y STR Allele Nomenclature

Gill et al. (2001) *Forensic Sci. Int*. 124: 5-10

- Number of complete repeats
- A partial repeat (variant allele) is designated by number of complete repeats separated by a dot followed by the number of bases in the incomplete repeat (e.g., 17.3)
- Some locus nomenclatures take into account the total number of repetitive units (non-variant plus variant) while others have taken into account only the variable repetitive stretches
  - "If a nomenclature is already in use, it is recommended that it should be continued. However, to encourage consistency for newly reported STRs, it is recommended that alleles should be named according to the total number of repeat units of the DNA that comprises both variant and non-variant repeats"
- Duplicated systems such as DYS385 have to be treated as genotypes and alleles should be separated by a hyphen (e.g., "11-14")

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

Nomenclature Issues with DYS389 I/II



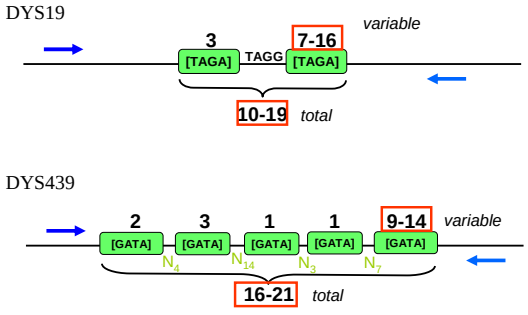
DYS389I = 3+11 = **14**; $[TCTG]_3[TCTA]_{11}$

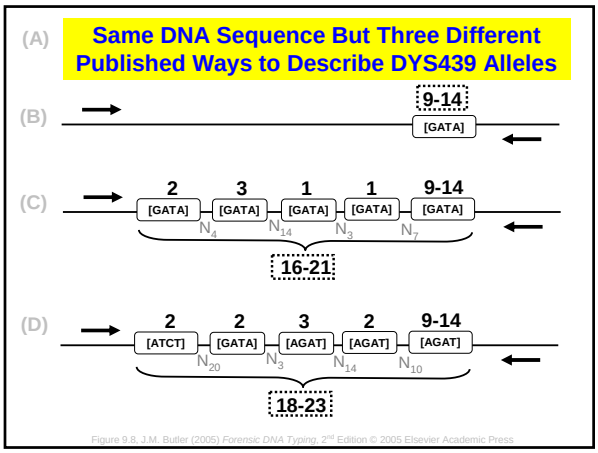
DYS389II = 5+12+3+11 = **31**; $[TCTG]_5[TCTA]_{12}[TCTG]_3[TCTA]_{11}$

Original paper (Kayser et al. (1997) Int. J. Legal Med. 110:141-149) defines allele nomenclature without repeat segment "C"; it has now been added in more recent nomenclatures thus making alleles +3 repeats larger

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

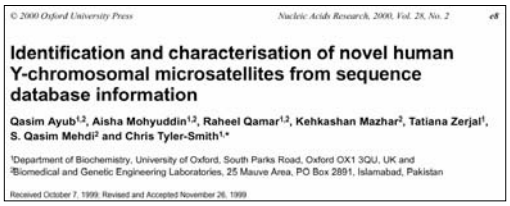
Y STR Allele Nomenclatures





From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

Issues with DYS439 Nomenclature



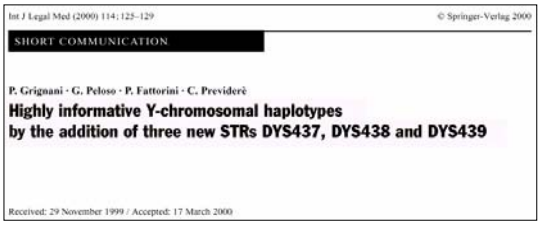
Original description of DYS439 (only variable repeat used)

$[GATA]_2N_4[GATA]_3N_{14}[GATA]_1N_3[GATA]_1N_7$ **$[GATA]_{9-14}$**

Alleles 9-14

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

Issues with DYS439 Nomenclature



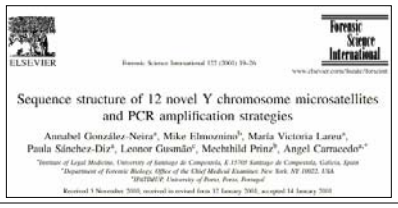
$[GATA]_2N_4[GATA]_3N_{14}[GATA]_1N_3[GATA]_1N_7[GATA]_{9-14}$

+7 repeats

Alleles 16-21

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

Issues with DYS439 Nomenclature



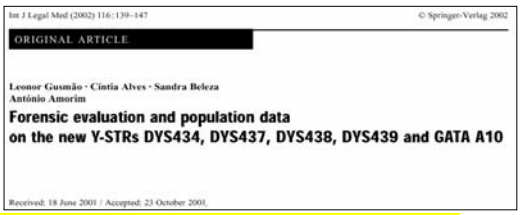
+9 repeats

Alleles 18-23

Repeat designation changed to AGAT from previous GATA

From a presentation by John Butler at 3rd Forensic Y User Group (Porto, Portugal) Nov 2002

Issues with DYS439 Nomenclature



Follow Ayub et al. designation (only variable repeat used)

$[GATA]_2N_4[GATA]_3N_{14}[GATA]_1N_3[GATA]_1N_7$ **$[GATA]_{9-14}$**

Alleles 9-14

Genetic Genealogy Companies

FamilyTreeDNA
http://www.familytreedna.com
http://www.dna-fingerprint.com

Sorenson Genomics
http://www.sorensongenomics.com

Relative Genetics
GeneTree
http://www.relativegenetics.com

Ethnoancestry
http://www.ethnoancestry.com

oxford ancestors
http://www.oxfordancestors.com

DNA Heritage
http://www.dnaheritage.com

GeoGene
http://www.geogene.com

http://www.isogg.org

Each company is using different markers and often different Y-STR allele nomenclature

A Nomenclature Comparison Chart

Commercial testing companies follow different standards when determining marker values. The SMGF Search Page automatically converts marker values when you select a lab. However, we provide the following chart for comparisons between testing companies.

Marker	DNA-Fingerprint	DNA Heritage	Family Tree DNA	Genographic	Oxford Ancestors	Relative Genetics
DYS385	=	=	=	=	?	=
DYS388	=	=	=	=	=	=
DYS389I	=	=	=	=	+ 3	=
DYS389II	**	**	**	**	**	**
DYS390	=	=	=	=	=	=
DYS391	=	=	=	=	=	=
DYS392	=	=	=	=	=	=
DYS393	=	=	=	=	=	=
DYS394	+ 1	+ 1	+ 1	+ 1	+ 1	+ 1
DYS426	=	=	=	=	=	=
DYS437	=	=	=	?	?	=

http://www.smgf.org/ychromosome/marker_standards.jspx

Note: Differences in standards are due to different methods of counting STR values.

Attempts to Resolve Nomenclature Issues

- Mulero, J.J., Budowle, B., **Butler, J.M.**, Gusmão, L. (2006) Letter to the Editor--Nomenclature and allele repeat structure update for the Y-STR locus GATA H4. *J. Forensic Sci.* 51(3): 694.
- Gusmão, L., **Butler, J.M.**, et al. (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci. Int.* 157:187-197.

Yfiler Kit H4 Nomenclature Resolution

- Establishing a consensus nomenclature can facilitate data comparison for proficiency testing, quality assurance, and casework results. **Efforts into nomenclature standardization should be supported and lauded**…
- There are differences in allele designations at the GATA H4 marker between those recommended in the Applied Biosystems AmpFlSTRs Yfiler™ polymerase chain reaction amplification kit (Applied Biosystems, Foster City, CA) and the ISFG recommendations. **The nomenclature for the GATA H4 marker in the Yfiler kit is based on the allele repeat structure defined by the National Institute of Standards and Technology Standard reference material (SRM) 2395** and the work of Butler et al.
- Those who choose to follow the allele nomenclature recommendations of the ISFG Commission **should add a correction factor of nine** to the Yfiler allele number, and they should refer to this marker as GATA H4.1.

Mulero, J.J., Budowle, B., **Butler, J.M.**, Gusmão, L. (2006) Letter to the Editor--Nomenclature and allele repeat structure update for the Y-STR locus GATA H4. *J. Forensic Sci.* 51(3): 694.

ISFG DNA Commission

International Society of Forensic Genetics

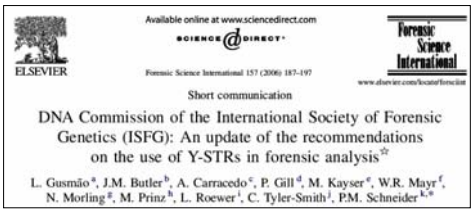
http://www.isfg.org/

- An international organization responsible for the promotion of scientific knowledge in the field of genetic markers analyzed with forensic purposes.
- Founded in 1968 and represents more than 800 members from over 50 countries.
- **A DNA Commission regularly offers recommendations on forensic genetic analysis.**

DNA Commission of the ISFG

- DNA polymorphisms (1989)
- PCR based polymorphisms (1992)
- Naming variant alleles (1994)
- **Repeat nomenclature (1997)**
- Mitochondrial DNA (2000)
- **Y-STR use in forensic analysis (2001)**
- **Additional Y-STRs - nomenclature (2006)**
- Mixture Interpretation (2006)
- Disaster Victim Identification (2007)

ISFG DNA Commission on Y-STRs



Available online at www.sciencedirect.com
SCIENCE DIRECT
Forensic Science International 157 (2006) 187–197
Forensic Science International
www.elsevier.com/locate/forsciint
ELSEVIER

Short communication

DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis☆

L. Gusmão [a], J.M. Butler [b], A. Carracedo [c], P. Gill [d], M. Kayser [e], W.R. Mayr [f], N. Morling [g], M. Prinz [h], L. Roewer [i], C. Tyler-Smith [j], P.M. Schneider [k,*]

Subject Matter Experts
Leonor Gusmão (Portugal)
John Butler (USA)
Peter Gill (UK)
Manfred Kayser (Netherlands)
Lutz Roewer (Germany)
Chris Tyler-Smith (UK)

ISFG Board Members
Angel Carracedo (Spain)
Wolfgang Mayr (Austria)
Niels Morling (Denmark)
Mecki Prinz (USA)
Peter Schneider (Germany)

Nomenclature Covered by ISFG (2006)

- Nomenclature of **11 core Y-STRs** (Table 1)
  - Some nomenclature is historical and not optimal
- "To avoid further confusion due to nomenclature changes, the nomenclature of widely used Y-STRs should not be altered, even if the present guidelines are not followed…"
- Coverage of **63 additional loci** **that were known at the time** (Table 2)

Genetic genealogy has gone beyond these defined loci…

NIST Standard Reference Materials (SRMs)

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 Marligen kit

$316

Certified for all loci in commercial Y-STR kits:
Y-PLEX 6
Y-PLEX 5
Y-PLEX 12
PowerPlex Y

SWGDAM recommended loci:
DYS19, DYS385 a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439

Y-filer - adds DYS635 (C4); now sequenced

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

Y-SNP Results on SRM 2395

from Marligen Signet™ Multiplexes (Luminex bead assay)

42 Y-SNPs measured across all samples

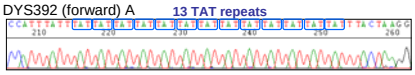
SRM 2395	AMEL	M207	M45	M89	DYS391	M2	M170	M172	M201
		(A/G)	(A/G)	(C/T)	(C/G)	(A/G)	(A/C)	(G/T)	(G/T)
Component A	XY	**G**	**A**	T	C	A	A	T	G
Component B	XY	A	G	T	C	A	A	**G**	G
Component C	XY	A	G	**C**	**G**	**G**	A	T	G
Component D	XY	A	G	T	C	A	A	T	**T**
Component E	XY	A	G	T	C	A	**C**	T	G
Component F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

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

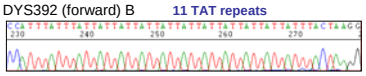
5 male components in SRM 2395 have 5 different Y-SNP backgrounds: **R1b, J2, E3a, G, and I**

Sequence Determination of Y STR Repeat Region for Each Component

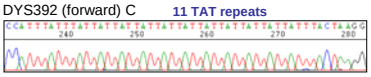
DYS392 (forward) A **13 TAT repeats**



DYS392 (forward) B **11 TAT repeats**

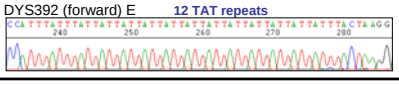


DYS392 (forward) C **11 TAT repeats**



DYS392 (forward) D **11 TAT repeats**

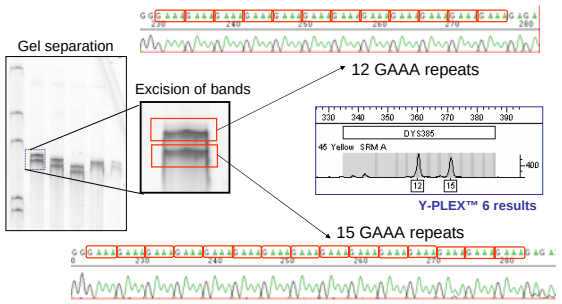
DYS392 (forward) E **12 TAT repeats**



Sequencing Performed

DYS19	DYS388
DYS385 a/b	DYS426
DYS389 I/II	
DYS390	DYS435
DYS391	DYS436
DYS392	DYS437
DYS393	DYS438
	DYS439
DYS447	DYS460 (A7.1)
DYS448	Y-GATA-H4
DYS464 a/b/c/d	

Sequencing Individual DYS385 Alleles



Sequence Summaries for SRM 2395

DYS19	A	14	[TAGA]$_3$ tagg[TAGA]$_{11}$
	B	14	[TAGA]$_3$ tagg[TAGA]$_{11}$
	C	16	[TAGA]$_3$ tagg[TAGA]$_{13}$
	D	15	[TAGA]$_3$ tagg[TAGA]$_{12}$
	E	17	[TAGA]$_3$ tagg[TAGA]$_{14}$

DYS438	A	12	[TTTTC]$_{12}$
	B	9	[TTTTC]$_9$
	C	11	[TTTTC]$_{11}$
	D	11	[TTTTC]$_{11}$
	E	10	[TTTTC]$_{10}$

DYS390	A	25	[TCTG]$_8$ [TCTA]$_{12}$ [TCTG]$_3$ [TCTA]$_4$
	B	23	[TCTG]$_8$ [TCTA]$_{10}$ [TCTG]$_3$ [TCTA]$_4$
	C	21	[TCTG]$_8$ [TCTA]$_8$ ACTA [TCTA]$_3$ [TCTG]$_3$ [TCTA]$_4$
	D	22	[TCTG]$_8$ [TCTA]$_9$ [TCTG]$_3$ [TCTA]$_4$
	E	24	[TCTG]$_8$ [TCTA]$_{11}$ [TCTG]$_3$ [TCTA]$_4$

We will continue to add information on new Y-STR loci as they are adopted by the community and put into commercial kits

Y-STR Kits Used in Forensic Testing

Allele size range and locus dye colors

PowerPlex® Y
Released by Promega Corporation in Sept 2003

100 bp 200 bp 300 bp 400 bp

FL: DYS391 DYS389I DYS439 DYS389II
JOE: DYS438 DYS437 DYS19 DYS392
TMR: DYS393 DYS390 DYS385a/b

3 dye colors
12-plex PCR

AmpFℓSTR® Yfiler™
Released by Applied Biosystems in Dec 2004

100 bp 200 bp 300 bp 400 bp

6-FAM: DYS456 DYS389I DYS390 DYS389II
VIC: DYS458 DYS19 DYS385a/b
NED: DYS393 DYS391 DYS439 DYS635 DYS392
PET: H4 DYS437 DYS438 DYS448

4 dye colors
17-plex PCR

Y-STR Types Being Added to NIST SRM 2395 Certificate

- DYS635 (to complete Yfiler loci)
- **DYS449**
- **DYS481**
- **DYS570**
- **DYS576**

These Y-STR loci have been sequenced across all 5 male components in SRM 2395

- DYS492, DYS522, DYS532, DYS534, DYS572, DYS607, DYS652, DYS709, DYS712, DYS715, DYS717
- **Also examining nomenclature on additional Y-STR loci used by genetic genealogy companies**

http://www.cstl.nist.gov/biotech/strbase/srm2395.htm

Modes Used at NIST for Value-Assignment of Reference Materials for Chemical Measurements to Obtain a NIST Certified Value

- Certification at NIST Using a Single Primary Method with Confirmation by Other Method(s)
- Certification at NIST Using Two Independent Critically-Evaluated Methods
- Certification/Value-Assignment Using One Method at NIST and Different Methods by Outside Collaborating Laboratories

NIST Special Publication 260-136 "Definitions of Terms and Modes Used at NIST for Value-Assignment of Reference Materials for Chemical Measurements, Table 1 (p.2)

Y-STR Loci in Use for Genetic Genealogy

- **Family Tree DNA** (12, 37, or 67 loci) – DYS19, DYS385 a/b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS439, DYS437, DYS447, DYS448, DYS449, DYS454, DYS455, DYS458, DYS459 a/b, DYS464 a/b/c/d, DYS438, DYS442, DYS460, GATA-H4, YCA II a/b, DYS456, DYS570, DYS576, DYS607, DYS724 a/b (CDY a/b), DYF395S1a/b, DYF406S1, DYS413 a/b, DYS425, DYS436, DYS444, DYS446, DYS450, DYS472, DYS481, DYS487, DYS490, DYS492, DYS511, DYS520, DYS531, DYS534, DYS537, DYS557, DYS565, DYS568, DYS572, DYS578, DYS590, DYS594, DYS617, DYS640, DYS641
- **Relative Genetics** (18, 26, or 43 loci) – DYS19, DYS385 a/b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS426, DYS438, DYS439, DYS447, DYS448, DYS460, YCAII a/b, GATA-H4, DYS393, DYS437, DYS454, DYS455, DYS461, DYS462, GGAAT1B07, GATA-A10, DYS635, DYS441, DYS442, DYS444, DYS445, DYS446, DYS449, DYS452, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d
- **Ethnoancestry** (18 loci) –DYS481, DYS487, DYS490, DYS494, DYS505, DYS522, DYS531, DYS533, DYS549, DYS556, DYS575, DYS578, DYS589, DYS594, DYS636, DYS638, DYS641, DYF406S1 + *Y-SNPs*
- **Oxford Ancestors** (10 loci) – DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS425, DYS426
- **GeoGene** (6 loci) – DYS19, DYS388, DYS390, DYS391, DYS392, DYS393 + *Y-SNPs*

67 Y-STR Loci in Use by FamilyTree DNA

Family Tree DNA (12, 37, or 67 loci) – **DYS19**, **DYS385 a/b**, **DYS388**, **DYS389I**, **DYS389II**, **DYS390**, **DYS391**, **DYS392**, **DYS393**, **DYS426**, **DYS439**, **DYS437**, **DYS447**, **DYS448**, DYS449, DYS454, DYS455, **DYS458**, DYS459 a/b, **DYS464 a/b/c/d**, **DYS438**, DYS442, **DYS460**, **GATA-H4**, **YCA II a/b**, **DYS456**, DYS570, DYS576, DYS607, DYS724 a/b (CDY a/b), DYF395S1a/b, DYF406S1, DYS413 a/b, DYS425, **DYS436**, DYS444, DYS446, **DYS450**, DYS472, DYS481, DYS487, DYS490, DYS492, DYS511, DYS520, DYS531, DYS534, DYS537, DYS557, DYS565, DYS568, DYS572, DYS578, DYS590, DYS594, DYS617, DYS640, DYS641

Loci in bold and underlined have certified values in current NIST SRM 2395

43 Y-STR Loci in Use by Relative Genetics

Relative Genetics (18, 26, or 43 loci) – **DYS19**, **DYS385 a/b**, **DYS388**, **DYS389I**, **DYS389II**, **DYS390**, **DYS391**, **DYS392**, **DYS426**, **DYS438**, **DYS439**, **DYS447**, **DYS448**, **DYS460**, **YCAII a/b**, **GATA-H4**, **DYS393**, **DYS437**, DYS454, DYS455, **DYS461**, **DYS462**, GGAAT1B07, GATA-A10, DYS635, DYS441, DYS442, DYS444, DYS445, DYS446, DYS449, DYS452, **DYS456**, **DYS458**, DYS459 a/b, DYS463, **DYS464 a/b/c/d**

Loci in bold and underlined have certified values in current NIST SRM 2395

New Y-STR Loci Offer a Great View…

but without the underlying support of standardization which takes time

http://www.photos.org.au/userimages/user756_1174912615a.jpg

Points to Keep in Mind…

- Initial selection of material (SRM components) was for a specific purpose usually and may not address every need in the future (a new locus may not exhibit a diverse set of alleles)
- The forensic community uses commercial STR typing kits – and only wants a confirmation of the allele calls against an allelic ladder
- Some duplicated Y-STR loci (e.g., DYS464) will not be able to have every allele sequenced
- **There are lots of loci that could be "certified" – how do we decide which ones to include in future certificate updates?**

STRBase Resources for Genetic Genealogists

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

- PowerPoint slides providing background information on Y-chromosome loci and testing
  - http://www.cstl.nist.gov/biotech/strbase/YmtDNAworkshop.htm
- Information on commonly used Y-STR loci
  - http://www.cstl.nist.gov/biotech/strbase/ystr_fact.htm
- Information on Y-STR nomenclature
  - http://www.cstl.nist.gov/biotech/strbase/YSTRs/H4_nomenclature.htm
- Links to Y-STR haplotype databases
  - http://www.cstl.nist.gov/biotech/strbase/y_strs.htm
- Links to genetic genealogy websites
  - http://www.cstl.nist.gov/biotech/strbase/weblink.htm#Genetic_Genealogy