Advanced Topics in STR DNA Analysis

Y-STRs and mtDNA

john.butler@nist.gov
NIST

AAFS 2006 Workshop #6
Seattle, WA
February 20, 2006

Dr. John M. Butler
Dr. Bruce R. McCord

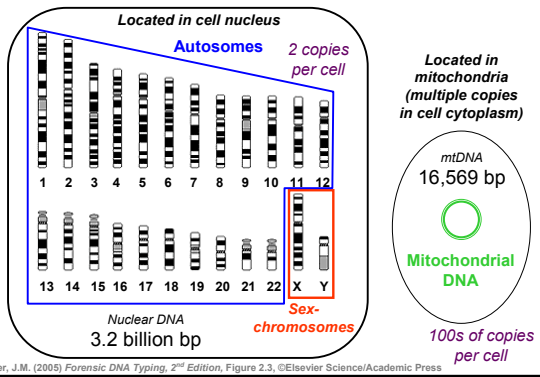
mccordb@fiu.edu
FLORIDA INTERNATIONAL UNIVERSITY

Outline for This Section

- Role of Y-STRs and mtDNA compared to autosomal STRs
- Advantages and disadvantages of lineage markers
- Y-STR core loci and available kits
- Y-STR haplotype databases and statistics
- mtDNA characteristics
- Efforts to resolve common types
- Hair shaft analysis with mtDNA and STRs

Human Genome

23 Pairs of Chromosomes + mtDNA



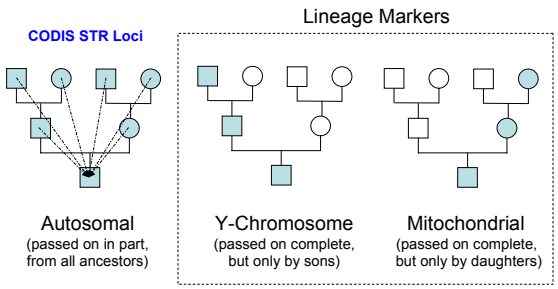
Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 2.3, ©Elsevier Science/Academic Press

Role of Y-STRs and mtDNA

Compared to Autosomal STRs

- **Autosomal STRs provide a higher power of discrimination and <u>are the preferred method</u> whenever possible**
- **Due to capabilities for male-specific amplification**, Y-chromosome STRs (**Y-STRs**) can be useful in extreme female-male mixtures (e.g., when differential extraction is not possible such as fingernail scrapings)
- **Due to high copy number**, mitochondrial DNA (**mtDNA**) may be the only source of surviving DNA in highly degraded specimens or low quantity samples such as hair shafts

Different Inheritance Patterns



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 9.1, ©Elsevier Science/Academic Press

Lineage Markers: Y-STRs and mtDNA

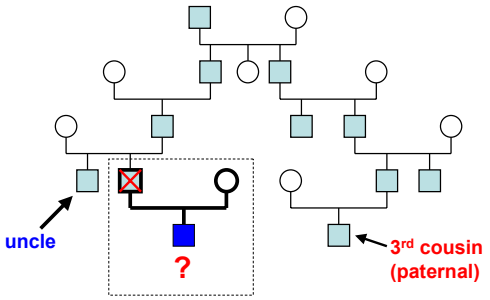
<u>Advantages</u>

- Extend possible reference samples beyond a single generation (benefits missing persons cases and genetic genealogy)
- Family members have indistinguishable haplotypes unless mutations have occurred

<u>Disadvantages</u>

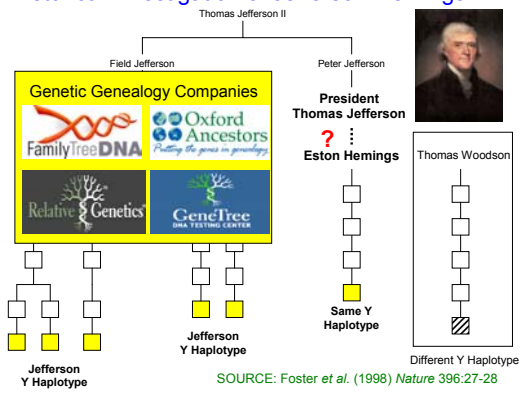
- Lower power of discrimination due to no genetic shuffling with recombination
- Family members have indistinguishable haplotypes unless mutations have occurred

Y-STRs permit extension of possible reference samples in missing persons cases



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 9.3, ©Elsevier Science/Academic Press

Historical Investigation of Jefferson-Hemings DNA



SOURCE: Foster *et al.* (1998) *Nature* 396:27-28

Figure 9.10, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Value of Y-Chromosome Markers

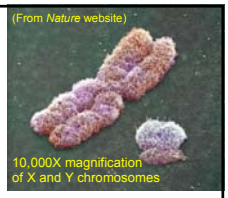
J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition; Table 9.1

Application	Advantage
Forensic casework on sexual assault evidence	**Male-specific amplification** (can avoid differential extraction to separate sperm and epithelial cells)
Paternity testing	Male children can be tied to fathers in motherless paternity cases
Missing persons investigations	**Patrilineal male relatives may be used for reference samples**
Human migration and evolutionary studies	Lack of recombination enables comparison of male individuals separated by large periods of time
Historical and genealogical research	Surnames usually retained by males; can make links where paper trail is limited

THE HUMAN Y CHROMOSOME: AN EVOLUTIONARY MARKER COMES OF AGE

Mark A. Jobling & Chris Tyler-Smith

***Nature Reviews Genetics* (2003) 4, 598-612**

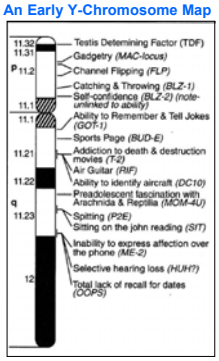


Abstract

- **Until recently, the Y chromosome seemed to fulfill the role of juvenile delinquent among human chromosomes — rich in junk, poor in useful attributes, reluctant to socialize with its neighbors and with an inescapable tendency to degenerate. The availability of the near-complete chromosome sequence, plus many new polymorphisms, a highly resolved phylogeny and insights into its mutation processes, now provide new avenues for investigating human evolution. Y-chromosome research is growing up.**

Traits found on the Y - Chromosome

An Early Y-Chromosome Map



Science (1993) 261:679

- spitting
- incessant use of TV remote buttons
- if lost, cannot stop and ask for directions
- ability to recall facts about baseball/basketball/hockey/golf/etc.
- male pattern baldness
- congregates with other Y-chromosome bearers to do "guy things"
- Source of "Testosterone poisoning"

What has happened in the past few years

- "Full" Y-chromosome sequence became available in June 2003; over 200 Y-STR loci identified (only ~20 in 2000)
- Selection of core Y-STR loci (SWGDAM Jan 2003)
- Multiple commercial Y-STR kits released
  - Y-PLEX 6,5,12 (2001-03), PowerPlex Y (9/03), Yfiler (12/04)
- Many population studies performed and databases generated with thousands of Y-STR haplotypes
- Forensic casework demonstration of value of Y-STR testing along with court acceptance

Disadvantages of the Y-Chromosome

- Loci are not independent of one another and therefore rare random match probabilities cannot be generated with the product rule; must use haplotypes (combination of alleles observed at all tested loci)
- Paternal lineages possess the same Y-STR haplotype (barring mutation) and thus fathers, sons, brothers, uncles, and paternal cousins cannot be distinguished from one another
- Not as informative as autosomal STR results
  - **More like addition (10 + 10 + 10 = 30) than multiplication (10 x 10 x 10 = 1,000)**

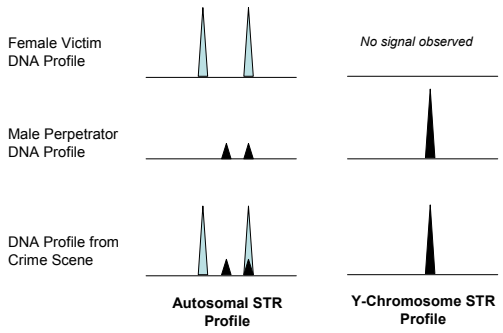
Forensic Advantages of Y-STRs

- **Male-specific amplification** extends range of cases accessible to obtaining probative DNA results (e.g., fingernail scrapings, sexual assault without sperm)
- **Technical simplicity due to single allele profile**; can potentially recover results with lower levels of male perpetrator DNA because there is not a concern about heterozygote allele loss via stochastic PCR amplification; number of male contributors can be determined
- **Courts have already widely accepted STR typing**, instrumentation, and software for analysis (Y-STR markers just have different PCR primers)
- **Acceptance of statistical reports using the counting method** due to previous experience with mtDNA

Scenarios Where Y-STRs Can Aid Forensic Casework

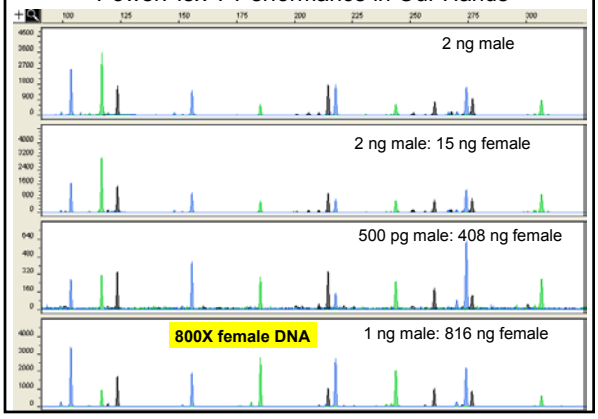
- Sexual assaults by vasectomized or azoospermic males (no sperm left behind for differential extraction)
- Extending length of time after assault for recovery of perpetrator's DNA profile (greater than 48 hours)
- Fingernail scrapings from sexual assault victims
- Male-male mixtures
- Other bodily fluid mixtures (blood-blood, skin-saliva)
- Gang rape situation to include or exclude potential contributors

Y-STRs can permit simplification of male DNA identification in sexual assault cases

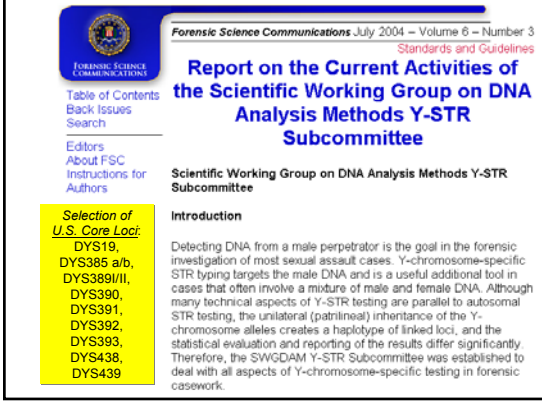


Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.2, ©Elsevier Science/Academic Press

PowerPlex Y Performance in Our Hands



Selection of Core Y-STR Loci



Forensic Science Communications July 2004 – Volume 6 – Number 3
Standards and Guidelines

Report on the Current Activities of the Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee

Table of Contents
Back Issues
Search
Editors
About FSC
Instructions for Authors

Selection of U.S. Core Loci:
DYS19,
DYS385 a/b,
DYS389I/II,
DYS390,
DYS391,
DYS392,
DYS393,
DYS438,
DYS439

Scientific Working Group on DNA Analysis Methods Y-STR Subcommittee

Introduction

Detecting DNA from a male perpetrator is the goal in the forensic investigation of most sexual assault cases. Y-chromosome-specific STR typing targets the male DNA and is a useful additional tool in cases that often involve a mixture of male and female DNA. Although many technical aspects of Y-STR testing are parallel to autosomal STR testing, the unilateral (patrilineal) inheritance of the Y-chromosome alleles creates a haplotype of linked loci, and the statistical evaluation and reporting of the results differ significantly. Therefore, the SWGDAM Y-STR Subcommittee was established to deal with all aspects of Y-chromosome-specific testing in forensic casework.

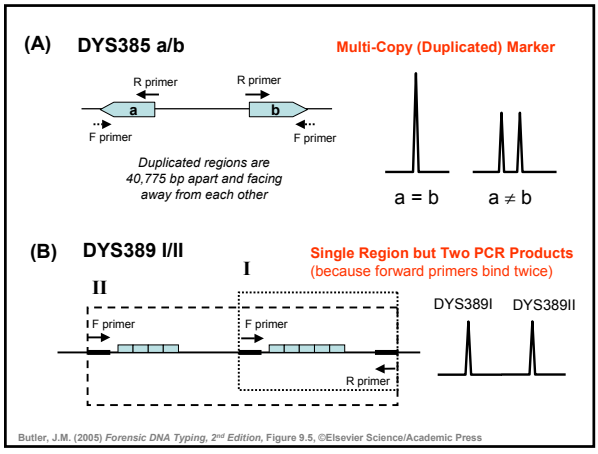
11 PCR products
9 primer sets

Core Y-STR Characteristics

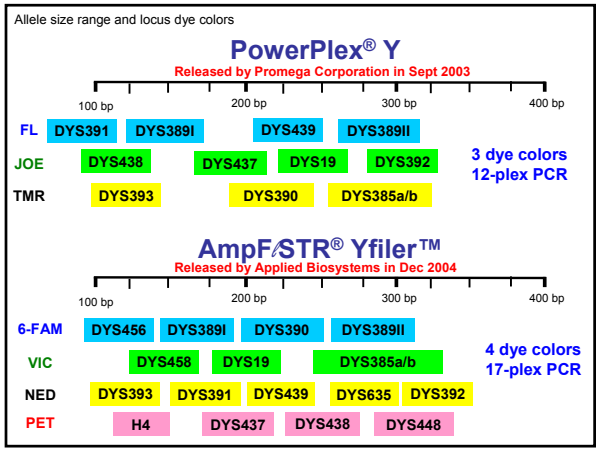
STR Marker	Position (Mb)	Repeat Motif	Allele Range	Mutation Rate
DYS393	3.17	AGAT	8-17	0.05%
DYS19	10.12	TAGA	10-19	0.20%
DYS391	12.54	TCTA	6-14	0.40%
DYS439	12.95	AGAT	8-15	0.38%
DYS389 I/II	13.05	[TCTG] [TCTA]	9-17 / 24-34	0.20%, 0.31%
DYS438	13.38	TTTTC	6-14	0.09%
DYS390	15.71	[TCTA] [TCTG]	17-28	0.32%
DYS385 a/b	19.19, 19.23	GAAA	7-28	0.23%
DYS392	20.97	TAT	6-20	0.05%

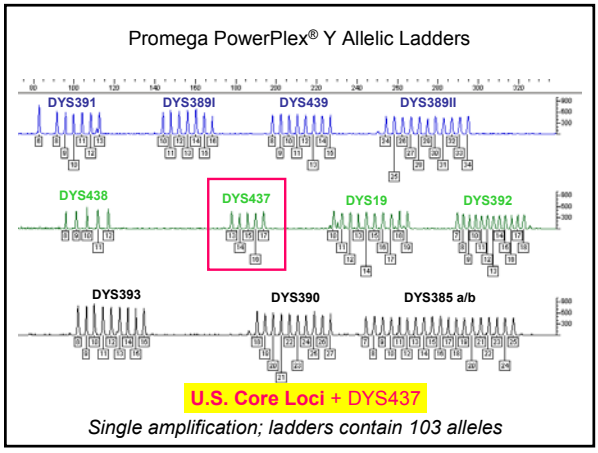
Positions in megabases (Mb) along the Y-chromosome were determined with NCBI build 35 (May 2004) using BLAT. Allele ranges represent the full range of alleles reported in the literature. Mutation rates summarized from YHRD (http://www.yhrd.org; accessed 6 Apr 2005).

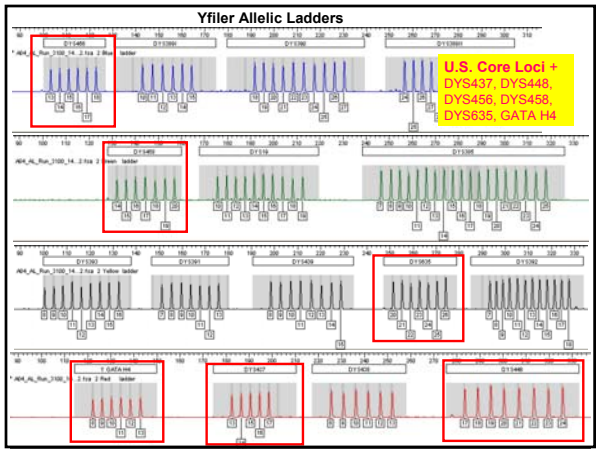
Butler, J.M. (2006) Genetics and genomics of core STR loci used in human identity testing. *J. Forensic Sci., in press.*

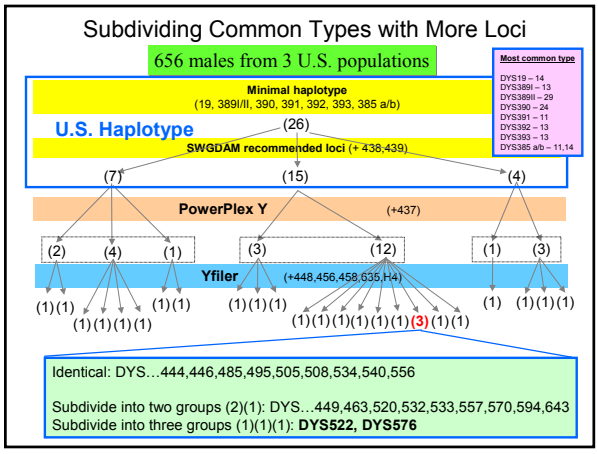


Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.5, ©Elsevier Science/Academic Press





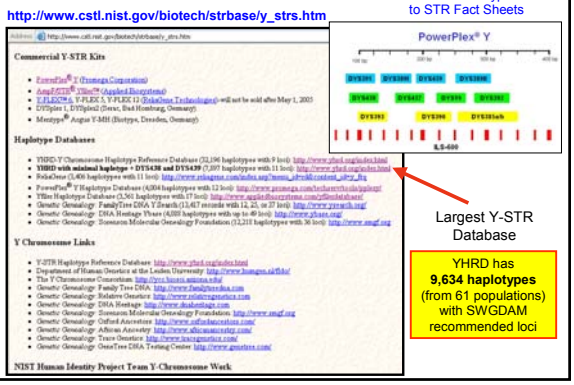




New Y-Chromosome Information Resources on STRBase

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

Locus boxes are hyperlinked to STR Fact Sheets



Largest Y-STR Database

YHRD has **9,634 haplotypes** (from 61 populations) with SWGDAM recommended loci

Y-Chromosome Haplotype Reference Database (YHRD)

http://www.yhrd.org

As of 12/5/05: **34,558 haplotypes**

9,634 haplotypes with all US required loci

Commercial Y-STR kits exist to amplify all of the core loci in a single reaction (plus a few additional markers)

Run only with minimal haplotype

DYS19
DYS389I/II
DYS390
DYS391
DYS392
DYS393
DYS385 a/b

US haplotype requires 2 additional loci:
DYS438
DYS439

Haplotype Databases for Y-STR Kits

http://www.promega.com/techserv/tools/pplexy/
http://www.appliedbiosystems.com/yfilerdatabase/

PowerPlex Y

1311 Caucasians
325 Asians
894 Hispanics
1108 African Americans
366 Native Americans

4,004 total
(as of March 2005)

Yfiler

1276 Caucasians
330 Asians
597 Hispanics
985 African Americans
106 Native Americans
105 Filipino
59 Sub-Saharan Africans
103 Vietnamese

3,561 total
(as of December 2004)

Statistics with Y-STR Haplotypes

Most labs will probably go with the **counting method** (number of times a haplotype is observed in a database) as is typically done with mtDNA results

Example Y-STR Haplotype

Core US Haplotype

- DYS19 – 14
- DYS389I – 13
- DYS389II – 29
- DYS390 – 24
- DYS391 – 11
- DYS392 – 14
- DYS393 – 13
- DYS385 a/b – 11,15
- DYS438 – 12
- DYS439 – 13

Matches by Databases

- YHRD (9 loci)
  - 7 matches in 27,773
- YHRD (11 loci)
  - 0 matches in 6,281
- ReliaGene (11 loci)
  - 0 matches in 3,403
- PowerPlex Y (12 loci)
  - 0 matches in 4,004
- Yfiler (17 loci)
  - 0 matches in 3,561

Y-Chromosome Haplotype Reference Database

www.YHRD.org

Release "15" from 2004-12-17 16:11:24

7 matches in 27,773 individuals from 236 worldwide populations

Population	#	Metapopulation
Bogotá, Colombia [European]	1 / 147	Eurasian MP / European MP
Central Portugal	1 / 230	Eurasian MP / European MP
Cologne, Germany	1 / 135	Eurasian MP / European MP
Leipzig, Germany	1 / 661	Eurasian MP / European MP
Liguria, Italy	1 / 81	Eurasian MP / European MP
London, UK	1 / 285	Eurasian MP / European MP
Lyon, France	1 / 125	Eurasian MP / European MP

Minimal Haplotype Result

DYS19 – 14
DYS389I – 13
DYS389II – 29
DYS390 – 24
DYS391 – 11
DYS392 – 14
DYS393 – 13
DYS385 a/b – 11,15

Frequency Estimate Calculations
Using the Counting Method

In cases where a Y-STR profile is observed a particular number of times (X) in a database containing N profiles, its frequency (p) can be calculated as follows:

$p = X/N$

7 matches in 27,773

p = 7/27,773 = 0.000252 = **0.025%**

An upper bound confidence interval can be placed on the profile's frequency using:

$$p + 1.96\sqrt{\frac{(p)(1-p)}{N}}$$

$$0.000252 + 1.96\sqrt{\frac{(0.000252)(1-0.000252)}{27,773}}$$

= 0.000252 + 0.000187 = 0.000439
= **0.044% (~1 in 2270)**

When there is no match with the counting method…

In cases where the profile has not been observed in a database, the upper bound on the confidence interval is

$1-\alpha^{1/N}$

0 matches in 4,004

where α is the confidence coefficient (0.05 for a 95% confidence interval) and N is the number of individuals in the database.

$1-\alpha^{1/N} = 1-(0.05)^{[1/4,004]} = 0.000748$
= 0.075% (~1 in 1340)

If using database of 2,443, then the best you can do is 1 in 816

The Meaning of a Y-Chromosome Match

Conservative statement for a match report:

The Y-STR profile of the crime sample matches the Y-STR profile of the suspect (at *xxx* number of loci examined). Therefore, ***we cannot exclude the suspect*** as being the donor of the crime sample. In addition, we cannot exclude all patrilineal related male relatives and an unknown number of unrelated males as being the donor of the crime sample.

Difficult Questions…

- **Which database(s) should be used for Y-STR profile frequency estimate determination?**
- Are any of the current forensic Y-STR databases truly adequate for reliable estimations of Y-STR haplotype frequencies?
  - Some individuals share identical Y-STR haplotypes due to recurrent mutations, not relatedness…
  - Is the database a random collection reflecting Y-STR haplotype frequencies of the population?
  - Is the Y-STR haplotype frequency relevant for the population of the suspect?

Issues raised by Peter de Knijff at his Promega meeting presentation (Oct 2004)

Conclusions from Peter de Knijff

From his presentation at the Promega meeting (Oct 2004)

A haplotype frequency taken from any Y-STR database should not be reported or seen as a random match probability

- Because all male relatives have the same haplotype
- Males can share haplotypes without being related

Database estimates are at most qualitative…

What Peter de Knijff Reports with a Y-STR Match

From his presentation at the Promega meeting (Oct 2004)

- The Y-STR profile of the stain matches with the suspect.
- Therefore, the suspect cannot be excluded as the donor of the stain.
- On the basis of this DNA evidence, I **can also not exclude all paternally related male relatives of the suspect** as possible donors of this stain.
- In addition, **an unknown number of males from the same region cannot be excluded**. A more accurate answer can only be obtained if (1) we have detailed knowledge of the population structure of the region of interest, (2) the Y-STR frequencies therein are known, and (3) we have knowledge about the family structure of the suspect.

Can Y-STR results be combined with autosomal STR information?

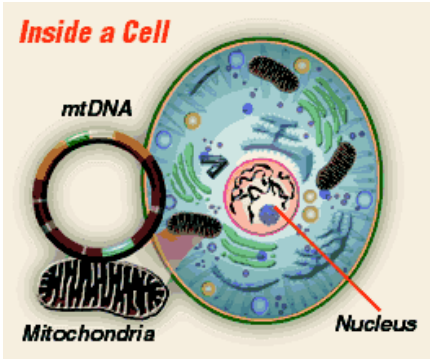
- Still subject to some debate among experts (most say "yes")
- Problem of different inheritance modes
- Multiply random match probability from the autosomal STR profile obtained with the upper bound confidence limit from the Y-STR haplotype frequency estimate

International Forensic Y-User Workshops

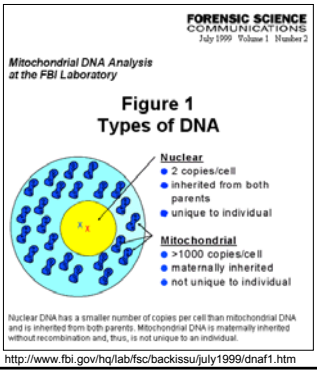
- Next meeting (5th): Sept 26-30, 2006 (Innsbruck, Austria) – **will also cover mtDNA**
- 1st – Berlin, Germany June 1996
- 2nd – Berlin, Germany June 2000
- 3rd – Porto, Portugal Nov 2002
- 4th – Berlin, Germany Nov 2004

For more information, see: http://www.yhrd.org/index.html

Mitochondrial DNA (mtDNA)



Comparison of Nuclear and Mitochondrial DNA



http://www.fbi.gov/hq/lab/fsc/backissu/july1999/dnaf1.htm

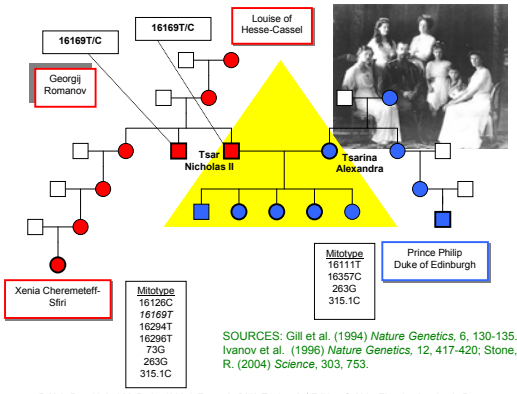
Advantages of mtDNA testing:

Higher copy number per cell
Results with highly degraded DNA
Results with limited sample (hair shaft)

Disadvantages of mtDNA testing:

Low power of discrimination
Labor intensive
Expensive

Identifying the Romanov Remains (the Last Russian Czar)



SOURCES: Gill et al. (1994) *Nature Genetics*, 6, 130-135.; Ivanov et al. (1996) *Nature Genetics*, 12, 417-420; Stone, R. (2004) *Science*, 303, 753.

D.N.A. Box 10.2, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

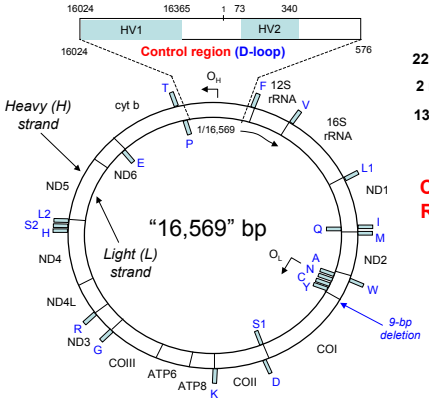


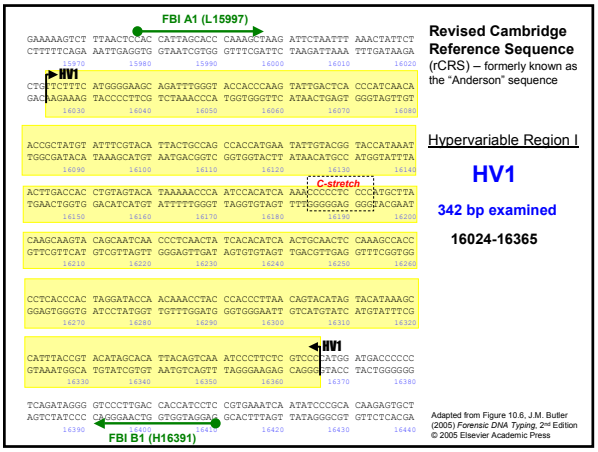
Figure 10.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

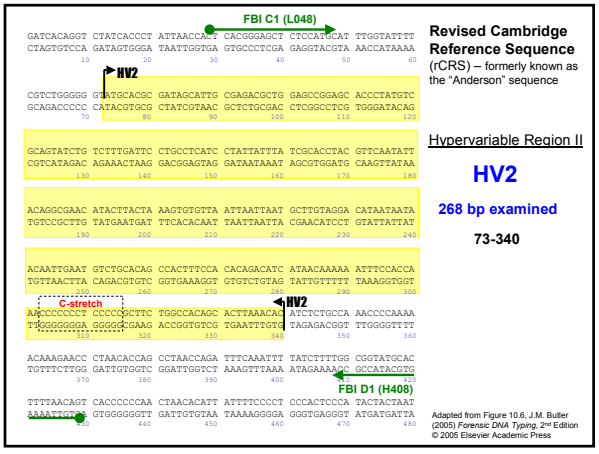
Control Region (16024-576)

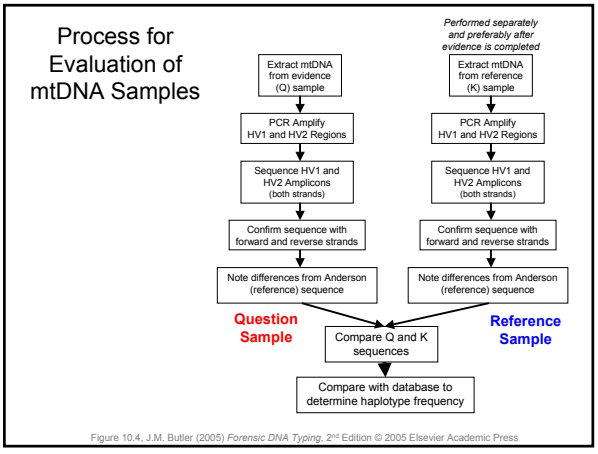
- 1,122 nucleotide positions
- Typically only **610 bases examined**
  - (HVI: 16024-16365; HVII: 73-340)

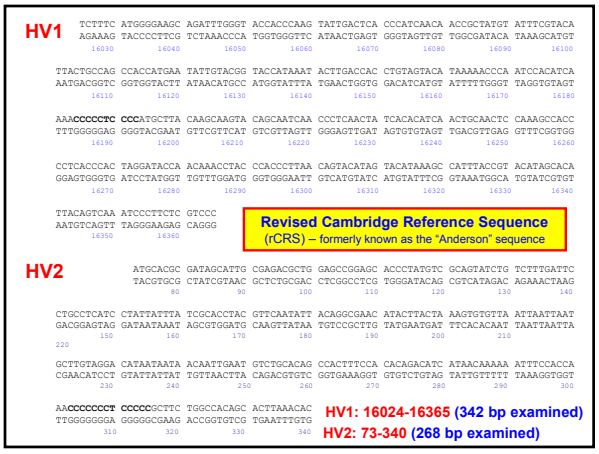
Coding Region (577-16023)

- 15,446 nucleotide positions
- Challenges with typing widely spaced SNPs
  - Multiplex PCR required
- Polymorphisms may have medical significance









Differences from Reference Sequence

mtDNA sequences from tested samples are aligned with the reference rCRS sequence (e.g., positions 16071-16140)

```
          16090     16100     16110     16120     16130     16140
rCRS ACCGCTATGT ATTCCGTACA TTACTGCCAG CCACCATGAA TATTGTACAG TACCATAAAT
Q    ACCGCTATGT ATTCCGTACA TTACTGCCAG CCACCATGAA TATTGTACAG TACCATAAAT
K    ACCGCTATGT ATTCCGTACA TTACTGCCAG CCACCATGAA TATTGTACAG TACCATAAAT
               16093                                 16129
```

Differences are reported by the position and the nucleotide change (compared to the rCRS)

Sample Q	Sample K
16093C	16093C
16129A	16129A

Adapted from Figure 10.8, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Challenges with mtDNA

- Data Interpretation
  - Heteroplasmy
  - Sample mixtures (currently not possible)
- DNA Database Sizes
  - Similar issues to Y-STRs but takes longer to generate mtDNA data than Y-STR haplotypes
- DNA Database Quality

Sequence Heteroplasmy at Position 16093

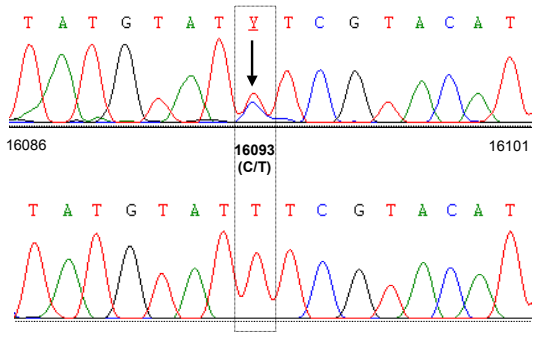
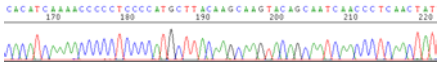


Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Disadvantages to Sequencing

- Expensive
  - Primarily due to intensive labor in data analysis
- Error possibilities with more data to review
- Most information is not used



Review forward and reverse sequences across 610 bases only to report…

263G, 315.1C **Most common type: found in ~7% of Caucasians…**

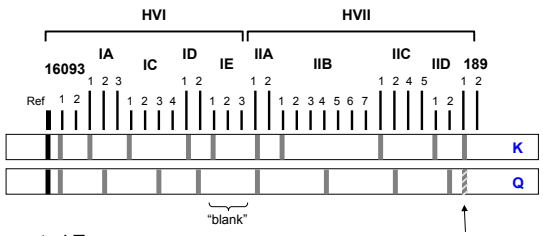
Advantages to Screening Methods

- Rapid results
- Aids in exclusion of non-matching samples
- Less labor intensive
- Usually less expensive
- Permits more labs to get involved in mtDNA

Screening assays are essentially a presumptive test prior to final confirmatory DNA sequencing.

Sequencing is necessary to certify that every position matches between a question and a known sample.

LINEAR ARRAY mtDNA Typing Strips: New Screening Method



Reported Types

K: 1-1-1-1-1-1-1-1-1-1

Q: 1-2-3-2-0-1-4-2-2-*w*1

If known (K) and question (Q) samples do not match, there is no need to involve the expense of mtDNA sequencing

Figure 10.10, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

A Common Use of mtDNA is for **Hair Shaft Analysis**

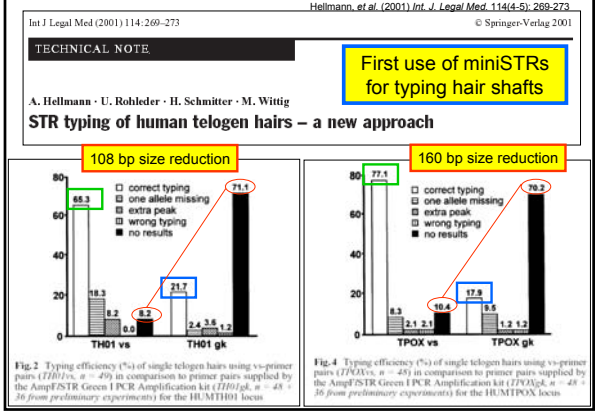
- Human hair shafts contain very little DNA but because mtDNA is in higher copy number it can often be recovered and successfully analyzed
- Melanin found in hair is a PCR inhibitor

Important Publications:

- **Wilson, M.R., *et al.* (1995) Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. Biotechniques 18(4): 662-669.**
  - **Tissue grinding method described by FBI Lab**
- **Melton *et al.* (2005) Forensic mitochondrial DNA analysis of 691 casework hairs. J. Forensic Sci. 50(1): 73-80.**
  - **Obtained a full or partial mtDNA profile for >92% of hairs tested**

PCR Product Size Reduction Improves Recovery of STR Information from Telogen Hairs

Hellmann, *et al.* (2001) *Int. J. Legal Med.* 114(4-5): 269-273



mtDNA and miniSTRs

- Due to the higher copy number, mtDNA will still have a role in many highly degraded DNA scenarios or where limited biological material is present, such as hair shafts.
- However, miniSTRs will most likely extend the range of cases where highly informative STR data can be obtained

THANK YOU FOR YOUR ATTENTION…

- Thank you for attending and participating in this Advanced Topics in STR DNA Analysis Workshop
- Feel free to contact us if you have further questions:

John Butler (NIST): john.butler@nist.gov

Bruce McCord (FIU): mccordb@fiu.edu