National Forensic Science Technology Center

Mitochondrial DNA Workshop

Michael D. Coble, PhD

March 13-15, 2006

Goals and Objectives

- Overview and theory behind mtDNA analysis
- The science behind mtDNA sequencing
- Forensic casework applications of mtDNA (validation and examples)
- Tools for mtDNA screening – Linear Arrays
- Emerging mtDNA technologies – mtDNA genome sequencing, species identification, dHPLC for resolving mixtures.
- Summary and **Questions**

June 26, 2000

"A day for the ages"

Associated Press

April 09, 1981

Mitochondrial DNA – The OTHER Human Genome

Fred Sanger standing in front of the Sanger Institute

http://www.wellcome.ac.uk/

A (Very) Brief DNA Review

Watson and Crick - 1953

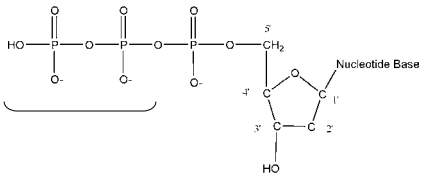
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

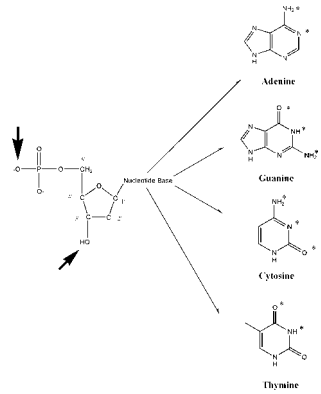
A Structure for Deoxyribose Nucleic Acid

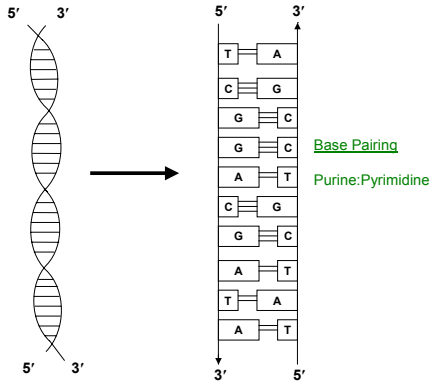
WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

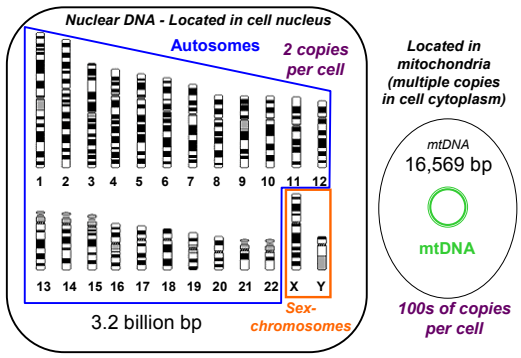
dNTP - Monomer







The Human Genome



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

You Say Tomato…

- Cowdry (1918) review of what microscopists called "mitochondria"

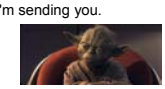
mitos = thread
chondros = granule

Blepharoblasts	Fadenkorper
Chondriokonts	Mitogel
Chondriomites	Parabasal bodies
Chondrioplasts	Plasmabioblasts
Chondriosomes	Plastochondria
Chondriospheres	Plastosomes
Filia	Vermicules
Fuchsinophilic	Sarcosomes
Granules	Interstitial bodies
Korner	Bioblasts

The Secret of the Force?

Star Wars Episode I: The Phantom Menace

QUI-GON : Make an analysis of this blood sample I'm sending you.
QUI-GON : I need a **midi-chlorian** count.



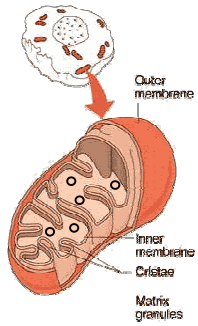
OBI-WAN : All right. I've got it.
QUI-GON : What are your readings?

OBI-WAN : Strange… the reading's off the chart...over twenty thousand.
OBI-WAN : Even Master Yoda doesn't have a **midi-chlorian** count that high!

QUI-GON : No Jedi has.
OBI-WAN : What does it mean?
QUI-GON : I'm not sure.

The Cell

Mitochondrial Morphology



Cytoplasmic organelle

Double membrane

Outer membrane – porin proteins for the transportation of materials.

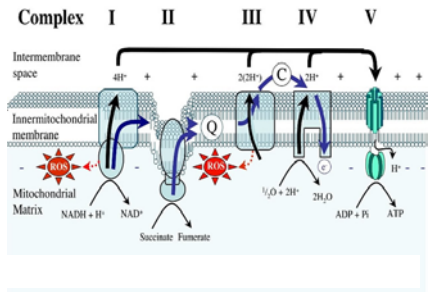
Inner membrane – highly folded (increased surface area) and highly impermeable.

Inner Matrix – several copies of mtDNA

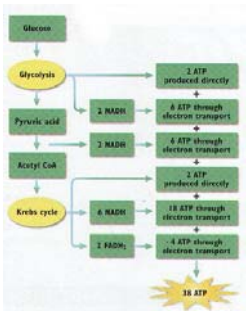
Mitochondrial Functions

Cellular Respiration – ATP production via oxidative-phosphorylation (OX-PHOS).

Apoptosis – programmed cell death
Steroid synthesis
Elongation of fatty acids
Oxidation of epinephrine (adrenaline)
Degradation of tryptophan
Heme synthesis
Heat production



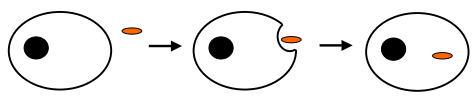
http://myweb.uiowa.edu/bballard/Research%20Figures/OXPHOS.jpg



http://sps.k12.ar.us/massengale/cell_respiration_bi.htm

Mitochondrial Evolution

- Endosymbiotic Theory – Ivan Wallin (1920s) and Lynn Margulis (1981).
- Proto-Eukaryotic cell incorporated a proto-bacterial cell and formed a symbiotic relationship.



Support for the Endosymbiotic Theory

- Mitochondria have double membranes – and the inner membrane is rich in cardiolipin.
- Mitochondria have their own genome, which is circular like bacteria (no histones), and use a genetic code for amino acids different that the nuclear DNA.
- New mitochondria are formed by a process similar to binary fission.
- Mitochondrial ribosomes are very similar to bacterial ribosomes (affected by antibiotics such as linezolid).

Lucky Guess or Clairvoyant?

- 1890 – R. Altman writes that "bioplasts" (mitochondria) are, "autonomous, elemental living units, forming bacteria-like colonies in the cytoplasm of the host cell."

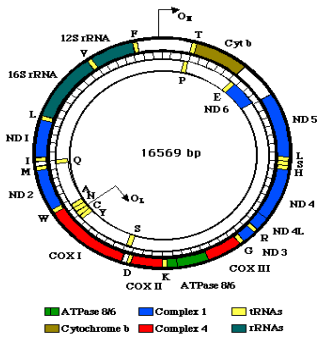
Immo Scheffler, Mitochondria (1999)

Mitochondrial Evolution

Complex	I	II	III	IV	V
Enzyme	NADH-CoQ Reductase	Succinate-CoQ Reductase	CoQ-Cytochrome C Reductase	Cytochrome C Oxidase	ATP Synthase
Inhibitor	Rotenone Amytal	TTFA malonate	Antimycin A	Cyanide Carbon Monoxide Azide	Oligomycin
Nuclear DNA Subunits	~43	4	10	10	~14
mtDNA Subunits	7 ND1-6, ND4L	0	1 Cytochrome b	3 COX I, II, III	2 ATPase 6 ATPase 8

~81 subunits encoded by the nuclear genome

mtDNA Genome



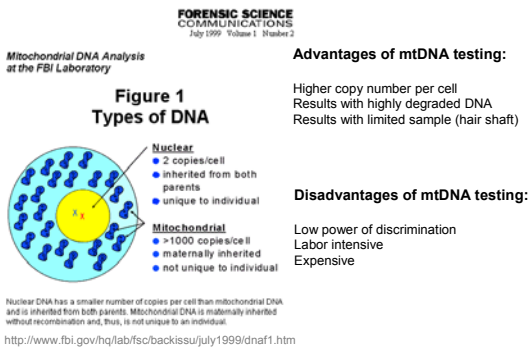
Coding Region
13 polypeptides
2 rRNAs
13 tRNAs

All necessary for OXPHOS

Highly compact (few intergenic spaces)

http://www.mitomap.org/

mtDNA as a Forensic Marker



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:

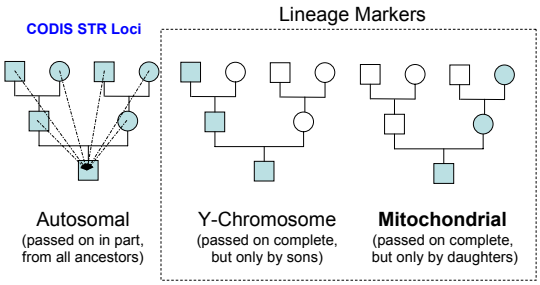
Low power of discrimination
Labor intensive
Expensive

Role of mtDNA

Compared to Autosomal STRs

- **Autosomal STRs provide a higher power of discrimination and <u>are the preferred method</u> whenever possible**
- **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
- A mtDNA result is better than no result at all…

Different Inheritance Patterns



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

Comparison of Human nucDNA and mtDNA

Characteristics	Nuclear DNA (nucDNA)	Mitochondrial DNA (mtDNA)
Size of genome	~3.2 billion bp	~16 569 bp
Copies per cell	2 (1 allele from each parent)	Can be > 1000
Percent of total DNA content per cell	99.75%	0.25%
Structure	Linear; packaged in chromosomes	Circular
Inherited from	Father and Mother	Mother
Chromosomal pairing	Diploid	Haploid
Generational recombination	Yes	No
Replication repair	Yes	No
Unique	Unique to individual (except identical twins)	Not unique to individual (same as maternal relatives)
Mutation rate	Low	At least 5–10 times nucDNA
Reference sequence	Described in 2001 by the Human Genome Project	Described in 1981 by Anderson and co-workers

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

Lineage Markers: Y-STRs and mtDNA

Advantages

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

Disadvantages

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

Location and Copy Number of mtDNA

- Found within the mitochondria in the cellular cytoplasm.
- On average 4-5 copies of mtDNA molecules per mitochondria (range of 1-15 mtDNA copies).
- Number of mitochondria vary by cell type (e.g., muscles have more…).
- Generally, hundreds of mitochondria per cell.

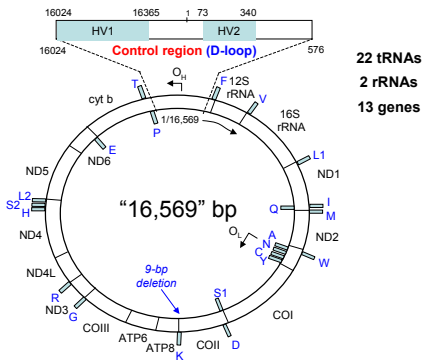


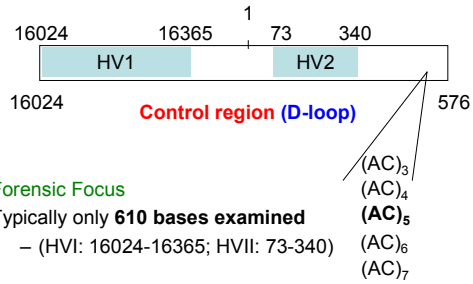
Figure 10.1, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

mtDNA Is Not Always 16,569 bp …

- Dinucleotide repeat at positions 514-524 (near end of control region)
  - Usually ACACACACAC or $(AC)_5$ in most individuals
  - Can vary from $(AC)_3$ to $(AC)_7$
- Other insertions and deletions may occur
  - 9 bp deletion (positions 8277 to 8285) in some individuals from Asia and Pacific Islands (haplogroup B) and Africans (haplogroup L).

Control Region (16024-576)

1,122 nucleotide positions



Forensic Focus

Typically only **610 bases examined**

– (HVI: 16024-16365; HVII: 73-340)

"Heavy" vs. "Light" Strand

- The two strands ("inner" and "outer" loops) of mtDNA can be separated with an alkaline CsCl gradient.
- Heavy or H-strand contains a greater number of guanine nucleotides (largest molecular weight of the four nucleotides) – purine rich.
- Light or L-strand contains more C and T nucleotides and is thus physically lighter (pyrimidine rich).
- H-strand codes for 28 gene products while the L-strand is used to transcribe 8 tRNAs and the ND6 enzyme.

Original Reference Sequence

- Human mtDNA was first sequenced in 1981 in Frederick Sanger's lab located in Cambridge, England.
- Authors for this paper (Nature 1981, 290:457-465) were listed in alphabetical order so Stan Anderson was the first author.
- This sequence has come to be referred to as the **"Anderson" sequence** (GenBank accession: M63933).
- This first sequence is sometimes called the **Cambridge Reference Sequence (CRS).**

Re-Sequencing of CRS

- The 1981 sequence was derived primarily from a placenta of an individual with European ancestry; however, some HeLa and bovine sequence was used to fill in gaps due to early sequencing procedures performed.
- Re-analysis of original placental material by Andrews et al. (1999) found 11 nucleotides that differed from Anderson et al. (1981) sequence.
- This **revised Cambridge Reference Sequence (rCRS)** is now the accepted standard for comparison.

Evaluation of Sequence Differences

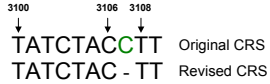
Between CRS (Anderson et al. 1981) and rCRS (Andrews et al. 1999)

Nucleotide Position	Region of mtGenome	Original CRS	Revised CRS	Remarks
3106–3107	16S rRNA	CC	C	Error
3423	ND1	G	T	Error
4985	ND2	G	A	Error
9559	COII	G	C	Error
11335	ND4	T	C	Error
13702	ND5	G	C	Error
14199	ND6	G	T	Error
14272	ND6	G	C	Error (bovine sequence inserted)
14365	ND6	G	C	Error (bovine sequence inserted)
14368	ND6	G	C	Error
14766	cyt b	T	C	Error (HeLa sequence inserted)

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Table 10.3, ©Elsevier Science/Academic Press

Further Comparison of CRS and rCRS

- No differences seen between CRS and rCRS within the mtDNA control region.
- **The original CRS contained a "CC" at positions 3106-3107 but rCRS was found to possess only a single "C"**



- Thus, rCRS is only 16,568 bp!

Maternal Inheritance of mtDNA

- Fertilizing sperm contributes only nuclear DNA.
- Cellular components including the mitochondria in the cytoplasm come from the mother's ovum.
- Any sperm mitochondria that may enter a fertilized egg are selectively destroyed due to a ubiquitin tag added during spermatogenesis.
- Barring mutation, a mother passes her mtDNA type on to her children.

Maternal Inheritance of mtDNA

Note that mtDNA is not unique to an individual

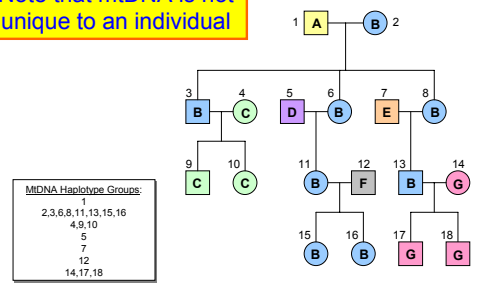


Figure 10.2, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Summary – mtDNA Characteristics

- High copy number of mtDNA.
- Maternal inheritance of mtDNA.
- Lack of recombination.
- High mutation rate compared to single copy nucDNA.

Methods for Measuring mtDNA Variation

- Low-resolution RFLP (1980s)
- High-resolution RFLP (1990s)
- Sequence analysis of HV1 and HV2 within control region (1991-present)
- Sequence analysis of complete mtDNA genome (2000-present)

Mitochondrial DNA Sequencing in Forensic Casework

Issues and Examples

Candidates for mtDNA Testing

- Shed hairs lacking root bulb or attached tissue
- Fragments of hair shafts.
- Aged bones or teeth that have been subjected to long periods of exposure.
- Crime scene stains or swabs that were unsuccessful for nuclear DNA testing.
- Tissues (muscle, organ, skin) that were unsuccessful for nuclear DNA testing.

Terry Melton – International Symposium on the Application of DNA Technologies in Analytical Sciences

mtDNA Testing on Hairs

- 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**

The Mitotyping Experience

Terry Melton,[1] Ph.D.; Gloria Dimick,[1] M.S.; Bonnie Higgins,[1] M.S.; Lynn Lindstrom,[1,2] B.S.; and Kimberlyn Nelson,[1] Ph.D.

Forensic Mitochondrial DNA Analysis of 691 Casework Hairs*

Journal of Forensic Science (2005) **50(1):** 73-80.

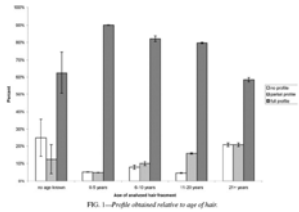


FIG. 1—*Profile obtained relative to age of hair.*

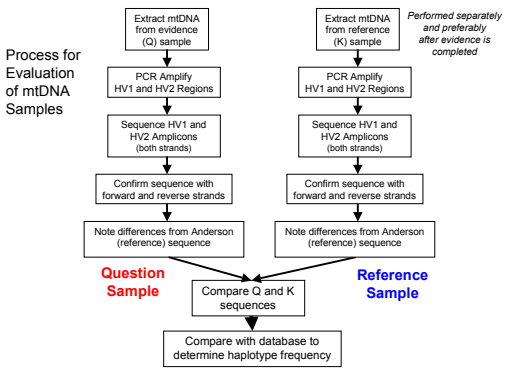
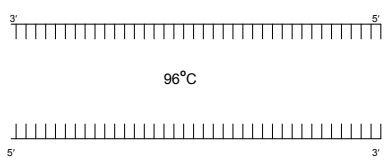


Figure 10.4, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

PCR Amplification of mtDNA

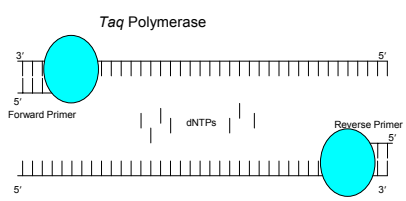
- Usually performed with 34-38 cycles
- Some protocols may go to 42 cycles for highly degraded specimens

PCR Amplification of mtDNA



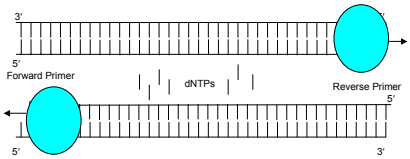
Denaturation

PCR Amplification of mtDNA



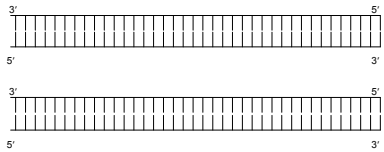
Primer Annealing

PCR Amplification of mtDNA



Primer Extension

PCR Amplification of mtDNA



Two Exact Copies

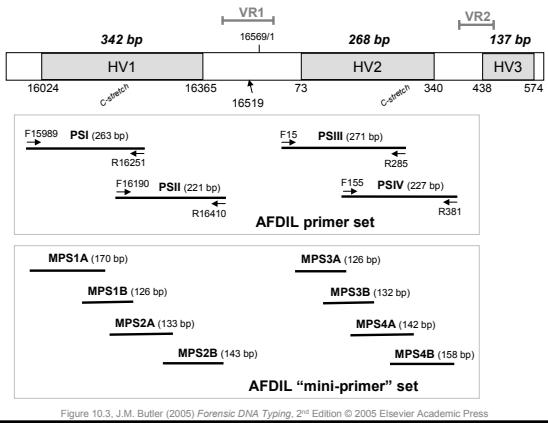
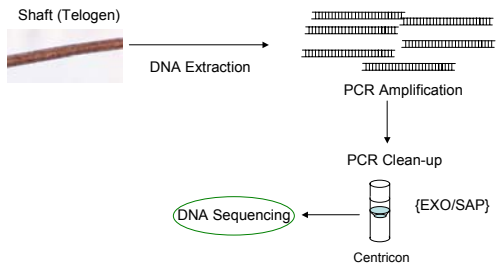


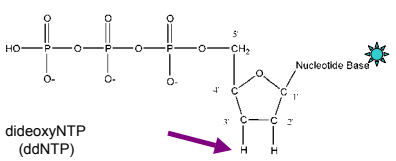
Figure 10.3, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Process for Evaluation of mtDNA Samples

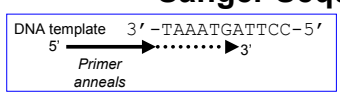


Science of DNA Sequencing

- Fred Sanger – developed the dideoxy method of sequencing in the 1970s… still used today.



Sanger Sequencing



Extension produces a series of ddNTP terminated products each one base different in length

Each ddNTP is labeled with a different color fluorescent dye

Sequence is read by noting peak color in electropherogram (possessing single base resolution)

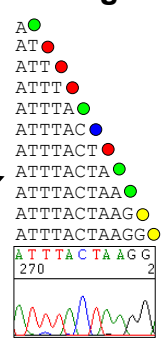
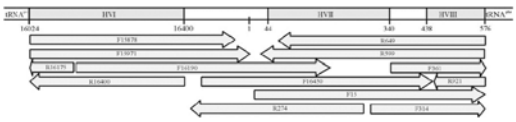
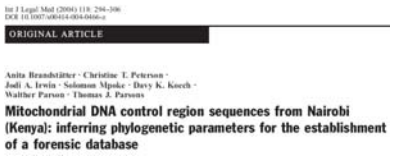


Figure 10.5, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Primers Used for Control Region Amplification and Sequencing

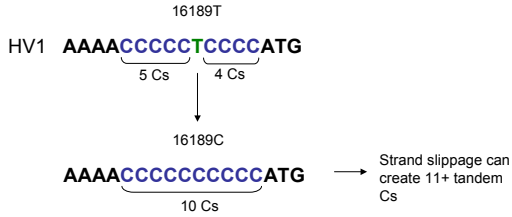


13 primers – 11 for routine sequencing



Why the Redundancy?

- Homopolymeric stretches of Cytosines (C-stretches).



Challenges with Sequencing Beyond the Polymeric C-Stretches in HV1 and HV2

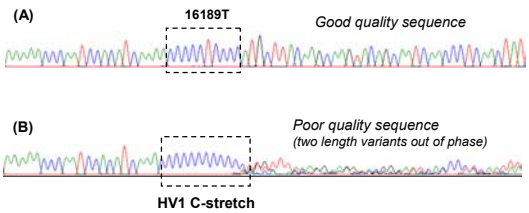


Figure 10.7, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Challenges with Sequencing Beyond the Polymeric C-Stretches in HV1 and HV2

Primer strategies typically used with C-stretch containing samples

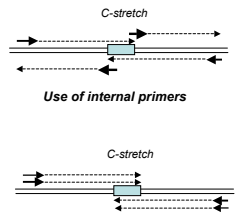
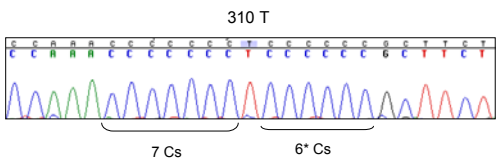


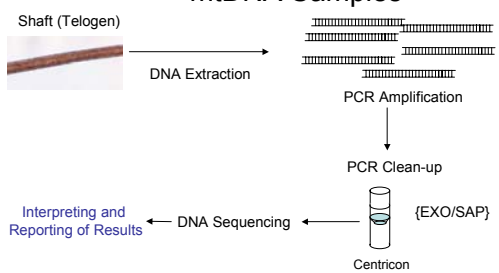
Figure 10.7, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

HV2 C-Stretch



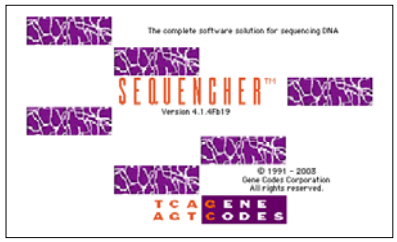
*The rCRS has 5 Cs in this region

Process for Evaluation of mtDNA Samples

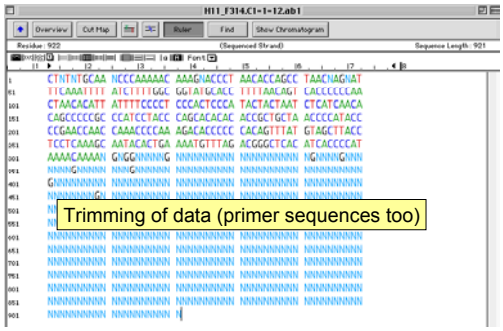


Interpreting and Reporting mtDNA Results

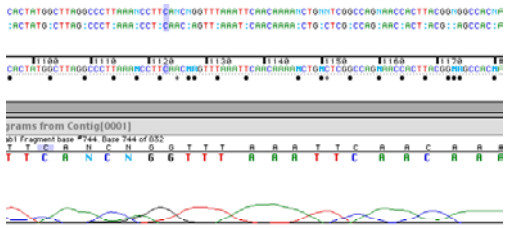
Data Review and Editing



Data Review and Editing

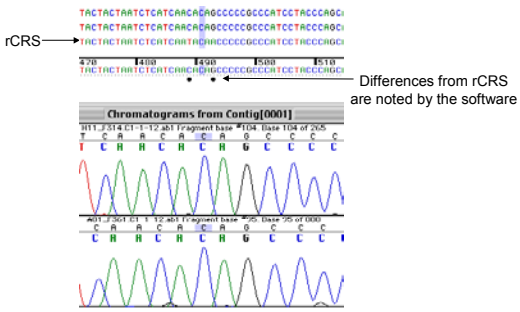


Data Review and Editing

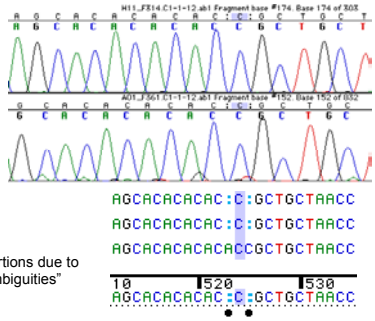


Poor resolution at the end of long fragments

Data Review and Editing

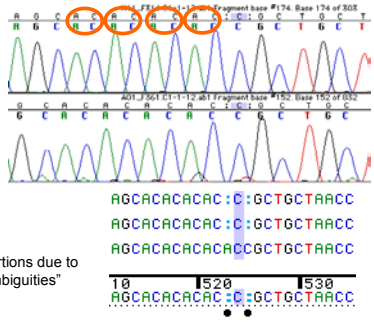


Data Review and Editing



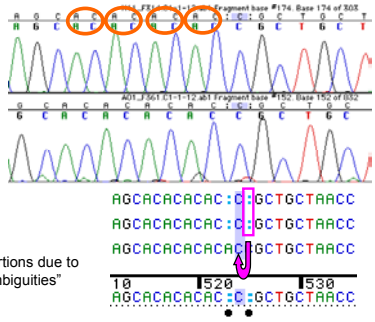
"Insertions due to ambiguities"

Data Review and Editing



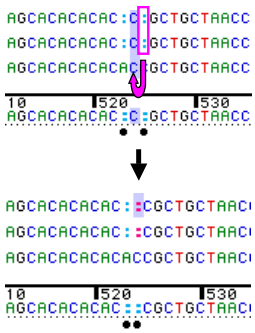
"Insertions due to ambiguities"

Data Review and Editing

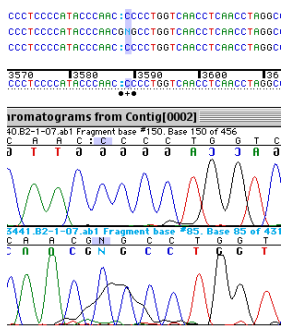


"Insertions due to ambiguities"

Data Review and Editing



Data Review and Editing

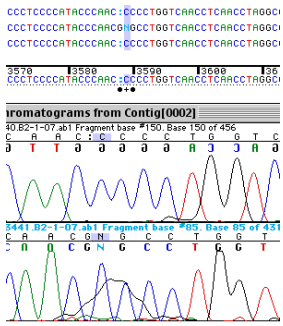


R primer – great data!

Small dye blob ~80bp into F primer creates ambiguities

Data Review and Editing

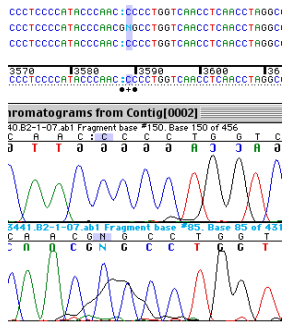
R AAC : CCCCT
F AACGNGCCT
rCRS AAC : CCCCT



Data Review and Editing

R AAC : CCCCT
F AACGNGCCT
rCRS AAC : CCCCT

Delete without affecting the sequence

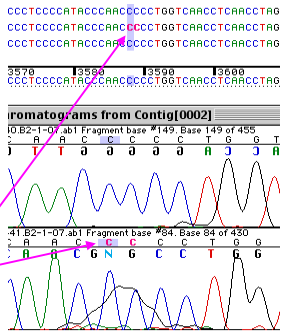


Data Review and Editing

R AACCCCCT
F AACNGCCT
rCRS AACCCCCT

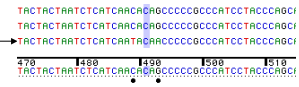
Change to "C"

Changes recorded in pink

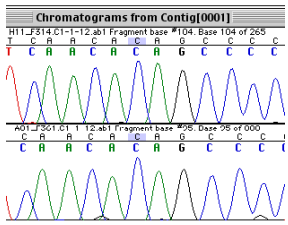


Reporting Differences from rCRS

rCRS



Point Mutations are listed as differences from the rCRS

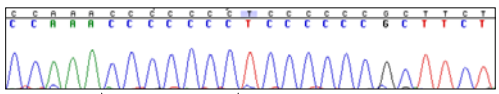


489 T-C
493 A-G

Reporting Differences from rCRS

- Insertions and Deletions (Indels)

310 T

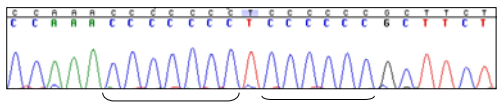


7 Cs 6* Cs

*The rCRS has 5 Cs in this region

Reporting Differences from rCRS

310 T



7 Cs 6* Cs

Sequence AAACCCCCCCTCCCCCCGCTT
rCRS AAACCCCCCCTCCCCCGCTTC
303 310 315

Reporting Differences from rCRS

Sequence AAACCCCCCCTCCCCCCGCTT
rCRS AAACCCCCCCTCCCCCGCTTC
303 310 315

Sequence AAACCCCCCCTCCCCCCGCTT
rCRS AAACCCCCCCTCCCCC:GCTTC
303 310 315

315.1 C

Reporting Differences from rCRS

- What if….

Sequence AAACCCCCCCTCCCCCCCGCTT
rCRS AAACCCCCCCTCCCCCGCTTC
303 310 315

2 additional Cs after 315?

Reporting Differences from rCRS

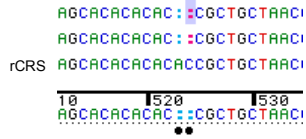
- What if….

Sequence AAACCCCCCCTCCCCCCCGCTT
rCRS AAACCCCCCCTCCCCC : : GCTTC
303 310 315

315.1 C
315.2 C

Deletions

- Deletions – report the position and bases deleted…



523 A-del
524 C-del

Nomenclature Issues

Wilson, MR et al. (2002) "Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region." ***Forensic Science International*** **129(1): 35-42.**

1. Use the least number of differences
2. Prioritization of indels > transitions > transversions

Transitions	Transversions
A-G	A-T
C-T	A-C
	G-T
	G-C
(Purine-Purine) (Pyrimidine-Pyrimidine)	(Purine-Pyrimidine) (Pyrimidine-Purine)

3. Indels placed at the 3' end with respect to the light strand

Nomenclature Issues

Wilson, MR et al. (2002) "Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region." ***Forensic Science International*** **129(1): 35-42.**

1. Use the least number of differences
2. Prioritization of indels > transitions > transversions

Sample TTTA : CCCAT 4 T-A
rCRS TTTTGCCCAT 5 G-del

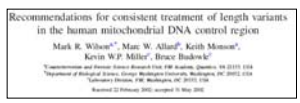
Sample TTTACCCAT
rCRS TTTTGCCCAT
1 10

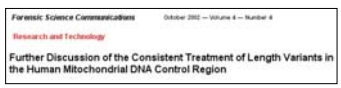
Sample TTT : ACCCAT 4 T-del
rCRS TTTTGCCCAT 5 A-G

3. Indels placed at the 3' end with respect to the light strand

Nomenclature Issues

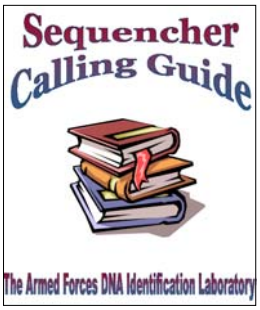
Wilson, MR ***et al.*** **(2002) "Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region."** ***Forensic Science International*** **129(1): 35-42.**


Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region


Forensic Science Communications
Research and Technology
Further Discussion of the Consistent Treatment of Length Variants in the Human Mitochondrial DNA Control Region

Wilson, MR ***et al.*** **(2002)** ***Forensic Science Communications***

Nomenclature Issues



Initiated by Jennifer O'Callaghan, AFDIL. Presented at the Promega HID Symposium, 2004.

Gives examples along with EPGs

POP QUIZ!!!

ATACAACCCCACCCAT
rCRS ATACAACCCCCACCCAT
488 504

ATTGATGTC
rCRS ATTGAATGTC
244 253

ATTGAAATGTC
rCRS ATTGAATGTC
244 253

CATAACAAAATTT
rCRS CATAACAAAAAATTT
280 294

POP QUIZ!!!

TGGCACTTTTCGTCT
rCRS TGGTATTTTCGTCT
52 65

TATCTTTCGT
rCRS TATTTTCGT
55 63

TATTTTTCGTCT
rCRS TATTTTCGTCT
55 65

AAACCCCCCCCCCTCCCCCCCGCT
rCRS AAACCCCCCCCTCCCCCGCT
300 318

Nomenclature Issues

- Consistency is needed – especially for database searches.

513
Lab 01 Sample GCACACACACACACCGCT
rCRS GCACACACACACCGCT

Lab 02 Sample GCACACACACACACCGCT
rCRS GCACACACACACCGCT

Lab 03 Sample GCACACACACACACCGCT
rCRS GCACACACACACCGCT

Nomenclature Issues

- Consistency is needed – especially for database searches.

513
Lab 01 Sample GCACACACACACACCGCT 524.1 A
rCRS GCACACACACAC::CGCT 524.2 C
1 2 3 4 5

Lab 02 Sample GCACACACACACACCGCT 523.1 C
rCRS GCACACACACA::CCGCT 523.2 A
1 2 3 4 5

Lab 03 Sample GCACACACACACACCGCT 514.1 A
rCRS GC::ACACACACACCGCT 514.2 C
1 2 3 4 5

Nomenclature Issues

Lab 01		Lab 02		Lab 03
16519 T-C		16519 T-C		16519 T-C
263 A-G		263 A-G		263 A-G
315.1 C	≠	315.1 C	≠	315.1 C
524.1 A		**523.1 C**		**514.1 A**
524.2 C		**523.2 A**		**514.2 C**

Each lab submits 20 sequences above into the population DB (N=1000)

Lab 04 → Will match the 20 samples submitted by Lab 01
16519 T-C
263 A-G
315.1 C
524.1 A
524.2 C

Apparent Frequency = 20/1000 (0.02)
True Frequency = 60/1000 (0.06)

Underestimation of the true frequency

Interpretational Issues - Heteroplasmy

- Heteroplasmy – the presence of more than one mtDNA type in an individual (Melton 2004).
- Once thought to be rare, heteroplasmy exists (at some level) in all tissues (Melton 2004).
- Especially important in hair analysis (semi-clonal).

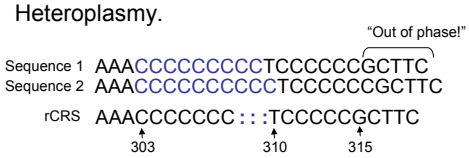
Heteroplasmy

- Some interesting papers (forensic focus)…

- Melton, T. (2004) Mitochondrial DNA heteroplasmy. *Forensic Science Reviews* **16:**1-20.
- Calloway *et al.* (2000) The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet.* **66(4):**1384-1397.
- Stewart *et al.* (2001) Length variation in HV2 of the human mitochondrial DNA control region. *Journal of Forensic Science* **46(4):**862-870.
- Sekiguchi *et al.* (2003) Inter- and intragenerational transmission of a human mitochondrial DNA heteroplasmy among 13 maternally-related individuals and differences between and within tissues in two family members. *Mitochondrion* **2(6):**401-414.
- Salas *et al.* (2001) Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. *Int J Legal Med.* **114(3):**186-190.
- Tully, L *et al.* (2000) A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *Am J Hum Genet.* **67(2):**432-443.

Interpretational Issues - Heteroplasmy

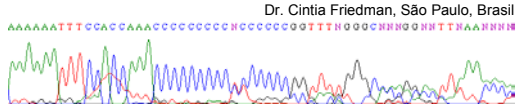
- Two types: Length (most common) and Point Heteroplasmy.



Sequence 1 has 9 Cs before 310T
Sequence 2 has 10 Cs before 310T

HV2 Length Heteroplasmy

Dr. Cintia Friedman, São Paulo, Brasil



Double coverage is important to determine sequences surrounding HV1, HV2, HV3 C-stretches.

Point Heteroplasmy

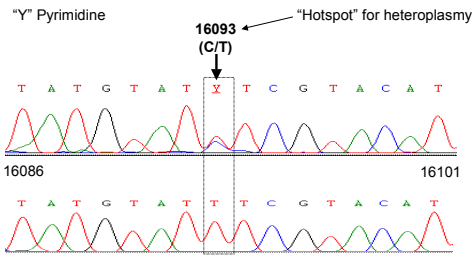


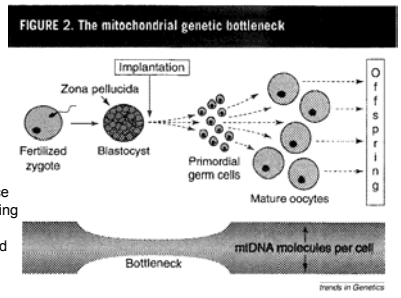
Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Origination of Heteroplasmy

Ovum – 100K mitochondria

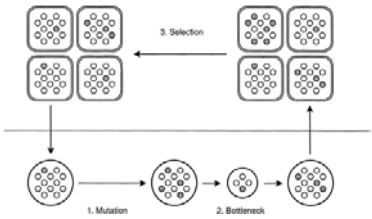
Very little mito growth until implantation

Females – produce ~7 million ova during fetal development only a few hundred become mature oocytes



Chinnery *et al.* (2000) Trends in Genetics

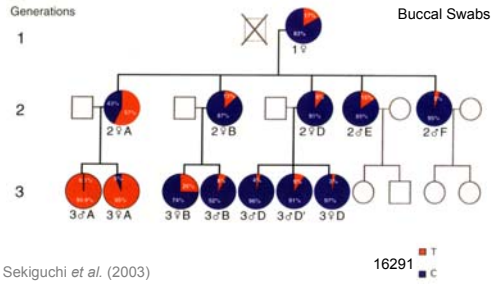
Origination of Heteroplasmy



Bergstrom and Pritchard (1998) *Genetics*

Heteroplasmic Variation

Heteroplasmic variation among family members



Sekiguchi *et al.* (2003)

Heteroplasmy in Hairs

Individual	Sample	Cytosines Before Position 310 T: Major Component	Minor Component	Type
Z	Blood sample			C_8TC_6
	1 hair	8 Cs	9 Cs	C_9TC_6
	1 hair	8 and 9 Cs	10 Cs	C_8TC_6 C_9TC_6 $C_{10}TC_6$
	1 hair	7 and 8 Cs		C_7TC_6 C_8TC_6
	1 hair	7 Cs		C_7TC_6

Stewart *et al.* (2001)

Blood + 1 hair	C_8TC_6 C_9TC_6
1 hair	C_7TC_6
1 hair	C_8TC_6 C_9TC_6 $C_{10}TC_6$

Length variants in HV2 alone should not be used to support an interpretation of exclusion.

Heteroplasmy in Hairs

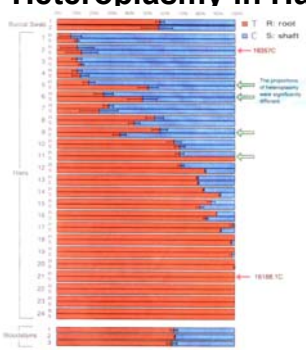
TABLE 3—*Seventy-eight independent observations of sequence heteroplasmy in 691 hairs.*

Heteroplasmy Position*	Other Matched K or Q Sample in Case for Comparison?	Nucleotide Present in Comparison Sample
70[†]	7 Q hairs	CRS in all 7
94	3 Q and 1 K hairs	CRS in all 4
120[†]	No	...
150	No	...
152	No	...
152	2 Q hairs and 1 K hair	All 3 hairs have T/C heteroplasmy
185	1 Q hair and 1 K blood	Both have A (substitution from CRS)

Melton *et al.* (2005)

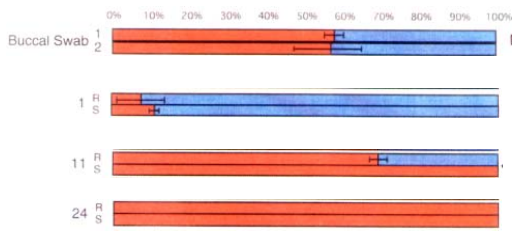
78 observations of point heteroplasmy in 691 hairs (11.4%)

Heteroplasmy in Hairs



Sekiguchi *et al.*

Heteroplasmy in Hairs

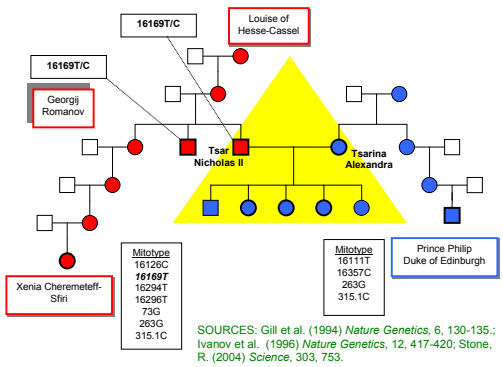


Heteroplasmy Detection

- Detection of heteroplasmy – sequencing can detect only to ~10% level.
- Other methods (e.g. DGGE) are much more sensitive.

Famous Case Involving Heteroplasmy

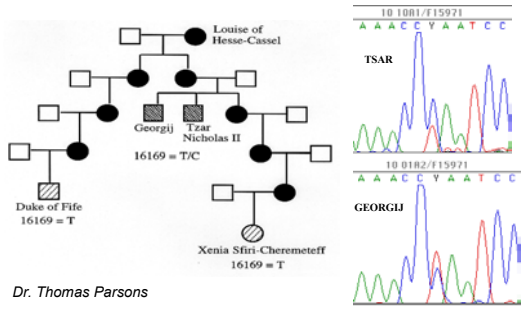
Identification of the Romanov Remains
(the Last Russian Czar)



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

AFDIL – Confirmation of FSS



Dr. Thomas Parsons

Interpretation of mtDNA Results

- Once the sequence has been generated (Q and K), and the differences from the rCRS are noted, what next?

SWGDAM Guidelines for Mitochondrial DNA (mtDNA) Nucleotide Sequence Interpretation

(1) Exclusion

(2) Inconclusive

(3) Cannot Exclude (Failure to Exclude)

Interpretation of mtDNA Results

- Exclusion – if there are **two or more** nucleotide differences between the questioned and known samples, the samples can be excluded as originating from the same person or maternal lineage.

Q TATTG***C***AC***A***G
K TATTGTACGG

Sample Q	Sample K
6 T-C	263 A-G
9 G-A	315.1 C
263 A-G	
315.1 C	

Exclusion

Interpretation of mtDNA Results

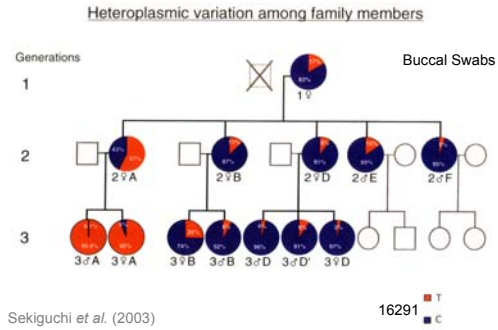
- Inconclusive – if there is **one** nucleotide difference between the questioned and the known samples, the result will be inconclusive.

Q TATTG*C*ACGG
K TATTG*T*ACGG

Sample Q	Sample K
6 T-C	263 A-G
263 A-G	315.1 C
315.1 C	

Inconclusive

Heteroplasmic Variation



Sekiguchi *et al.* (2003)

Interpretation of mtDNA Results

- Cannot Exclude – if the sequences from questioned and known samples under comparison have a common base at each position or a common length variant in the HV2 C-stretch, the samples cannot be excluded as originating from the same person or the same maternal lineage.

Q TATTGTACGG
K TATTGTACGG

Sample Q	Sample K
152 T-C	152 T-C
263 A-G	263 A-G
315.1 C	315.1 C

Cannot Exclude

POP QUIZ!!!

- How would you interpret these results?

Q TATTGTAC*A/G*G
K TATTGTAC *G* G

Sample Q	Sample K
9 G-R	263 A-G
263 A-G	315.1 C
315.1 C	

Q TATTGTAC*A/G*G
K TATTGTAC*G/A*G

Sample Q	Sample K
9 G-R	9 G-R
152 T-C	152 T-C
263 A-G	263 A-G
315.1 C	315.1 C

POP QUIZ!!!

- How would you interpret these results?

Q TA*T*TGTAC*A*G
K TA*C*TGTAC*G*G

Sample Q	Sample K
16519 T-C	16519 T-C
9 G-A	3 T-C
263 A-G	263 A-G
315.1 C	315.1 C

Q TATTGTACGG
K TATTGTACGG

Sample Q	Sample K
16519 T-C	16519 T-C
152 T-C	152 T-C
263 A-G	263 A-G
309.1 C	315.1 C
315.1 C	

POP QUIZ!!!

Q TATTGTACGG
K TATTG*C*ACGG

Sample Q	Sample K
16519 T-C	16519 T-C
263 A-G	152 T-C
309.1 C	263 A-G
315.1 C	315.1 C

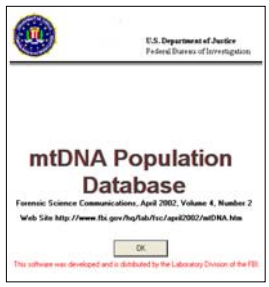
Q TATTGT*T*ACGG
K TAT**-**GT**-**ACGG

Sample Q	Sample K
16519 T-C	16519 T-C
6.1 T	4 T-del
263 A-G	263 A-G
315.1 C	315.1 C

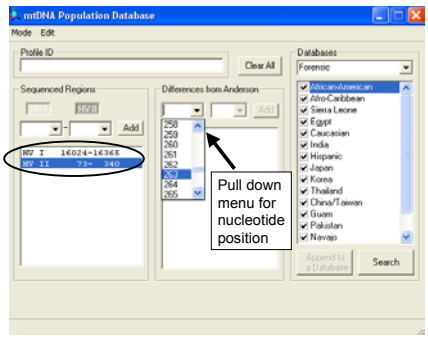
Reporting Statistics

- When “cannot exclude” is the interpretation, then a statistical estimate is needed in order to weigh the significance of the observed match
- Counting method is most common approach used and involves counting the number of times that a particular mtDNA haplotype (sequence) is seen in a database
- The larger the number of unrelated individuals in the database, the better the statistics will be for a random match frequency estimate.

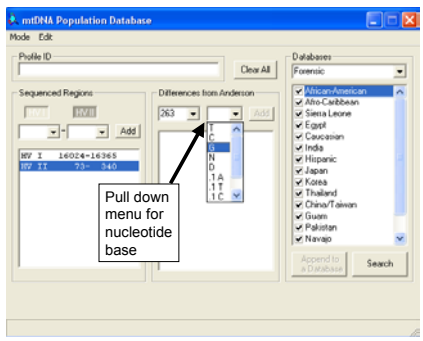
FBI mtDNA Population DB



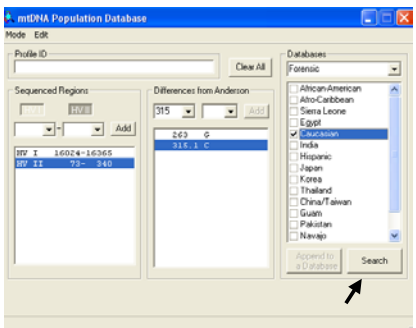
FBI mtDNA Population DB



FBI mtDNA Population DB



FBI mtDNA Population DB



FBI mtDNA Population DB

Summary of Databases

Database	Classification	# Profiles
Caucasian	Caucasian origin	1655
Total		1655

Overall Match Results

Number of Differences From Search Profile	Number	Freq.	Cum. Num.	Cum. Freq.
0	70	0.0423	70	0.0423
1	171	0.1033	241	0.1456
2	218	0.1317	459	0.2773
3	173	0.1045	632	0.3819
4	195	0.1178	827	0.4997
5	198	0.1196	1025	0.6193
>5	630	0.3807	1655	1.0000

Average Number of Differences = 4.845

FBI mtDNA Population DB

Summary of Databases

Database	Classification	# Profiles
Caucasian	Caucasian origin	1655
Total		1655

~10% of the database is 1 mutation away from the searched haplotype

Number of Differences From Search Profile	Number	Freq.	Cum. Num.	Cum. Freq.
0	70	0.0423	70	0.0423
1	171	0.1033	241	0.1456
2	218	0.1317	459	0.2773
3	173	0.1045	632	0.3819
4	195	0.1178	827	0.4997
5	198	0.1196	1025	0.6193
>5	630	0.3807	1655	1.0000

Average Number of Differences = 4.845

Example Calculation of mtDNA Profile Frequency Estimate

The frequency (p) of observing a mtDNA profile (X) times in a database having a size of (N) is…

$p = X/N$

$p = 70/1665 = 0.042$

A 95% confidence interval can be determined by using a normal approximation of the binomial

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

$$p \pm 1.96 \sqrt{\frac{(0.042)\,(0.958)}{1665}} = p \pm 1.96\,(0.0049) = 0.032 \text{ and } 0.052$$

Holland and Parsons (1999) *Forensic Sci. Rev.*

Issues Impacting mtDNA Interpretation

Challenges with mtDNA

- Data Interpretation
  - Heteroplasmy
  - Sample mixtures (Dr. Danielson)
  - Other Issues (Pseudogenes, etc…)
- DNA Database Sizes
  - Similar issues to Y-STRs but takes longer to generate mtDNA data than Y-STR haplotypes
- DNA Database Quality

Nuclear Pseudogenes

- Thoughout history – movement of mtDNA genes into the nucleus.
- Nuclear Pseudogenes (nuclear-mitochondrial like sequence ***numts***) – could potentially be amplified, confounding interpretation. "Molecular Fossils"

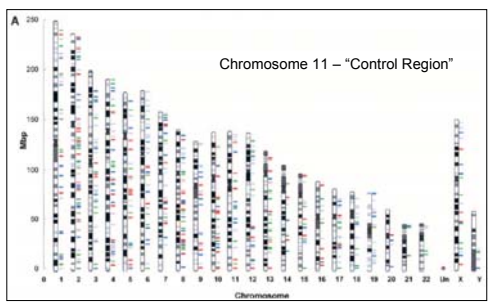
Article — Genome Research (2002)

Pattern of Organization of Human Mitochondrial Pseudogenes in the Nuclear Genome

Markus Woischnik and Carlos T. Moraes[1]

Department of Neurology, University of Miami-School of Medicine, Miami, Florida 33136, USA

Nuclear Pseudogenes



Woischnik and Moraes (2002)

Nuclear Pseudogenes

- Typically – numts are not a problem for forensics – mtDNA high copy number
- "Mitochondrial DNA pseudogenes in the nuclear genome as possible sources of contamination" - *Goios A, Amorim A, Pereira L.* ISFG meeting in the Azores, 2005.
- Extraordinary measures to observe a numt (Possibly seen by Grzbowski 2000 – nested PCR ~ 60 cycles).

mtDNA Recombination

- Adam Eyre-Walker and colleagues - proposed that paternal contribution of mtDNA has caused recombination.
- Some of their assumptions along with the data that was analyzed have been wrong (more tomorrow).

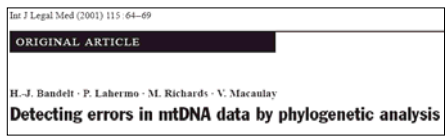
mtDNA Population Database: Size and Quality of Information

- Population databases are critical for estimating expected frequencies. The more, the better.

Database	# Profiles
African-American	1148
Afro-Caribbean	0
Sierra Leone	109
Caucasian	1655
Hispanic	686
Japan	163
Korea	182
Thailand	52
Navajo	146
Apache	180
Egypt	48
China/Taiwan	356
Guam	87
India	19 ←
Pakistan	8 ←
Total	4839

mtDNA Population Database: Size and Quality of Information

- Recently – mtDNA database quality has become an issue…


Int J Legal Med (2001) 115 : 64–69

ORIGINAL ARTICLE

H.-J. Bandelt · P. Lahermo · M. Richards · V. Macaulay

Detecting errors in mtDNA data by phylogenetic analysis

Artificial recombinations and phantom mutations plague the quality of mtDNA data in population genetics, forensics, and clinical studies

mtDNA Population Database: Size and Quality of Information

- Bandelt et al. (2001)

"In order to meet high-quality standards in forensics, sequencing should be performed in both directions (Bär et al. 2000). It is then important to read the two series of outputsseparately (against the CRS) and to transform either series into a data table independently, preferably of different formats (motif vs dot table); finally, the two tables should be compared by computer."

mtDNA Population Database: Size and Quality of Information

Problems in FBI mtDNA Database

Bandelt, Salas, and Bravi (2004) *Science*

Found 5 examples of artificial recombination among the 1148 African Americans in the database

mtDNA Population Database: Size and Quality of Information

SWGDAM mtDNA database - USA.AFR.000942

HV1	HV2
16126-16187-16189-16223-16264 16270-16278-16293-16311-16519	73-249d-263-290d-291d 309.1C-315.1C-489

mtDNA Population Database: Size and Quality of Information

SWGDAM mtDNA database - USA.AFR.000942

HV1	HV2
16126-16187-16189-16223-16264 16270-16278-16293-16311-16519	73-249d-263-290d-291d 309.1C-315.1C-489
African haplogroup L1b	Asian haplogroup C1

mtDNA Population Database: Size and Quality of Information

- Phantom mutations – Bandelt et al. (2002); Brandstatter et al. (2005).
- Phantom mutations are systematic artifacts generated during cycle sequencing. These can be created by either the sequencing chemistry, the automated sequencer, or lab procedures.
- Single-strand sequencing (e.g. F only) is highly susceptible to generating phantom mutations.

mtDNA Population Database: Size and Quality of Information

Table 1. Combinations of sequencing instruments, PCR-purification procedures, and sequencing chemistries in the experimental study

Sequencer	Sequencing chemistries[a]	Post-PCR treatment	Number of sequences	Artificial deletions	Artificial insertions	Artificial substitutions	Artifacts per sequence	Sample score (average QV)	Source[b]
ABI310	BD1	Enzymatic	23	10	2	13	1.1	30.6	1
ABI310	BD1	*Via* columns	20	5	1	10	0.8	28.9	1
ABI310	BD2	*Via* columns	30	0	1	6	0.2	28.4	2
ABI310	DR	Enzymatic	48	14	4	57	1.6	29.7	1
ABI310	DR	*Via* columns	45	16	3	51	1.6	29.0	1
ABI377	BD2	*Via* columns	45	0	6	35	0.9	31.6	2
ABI377	BD3	*Via* columns	18	0	7	48	3.1	26.6	2
ABI377	DP	*Via* columns	34	5	2	56	1.9	28.1	2
ABI377	DR	*Via* columns	31	0	1	88	2.9	24.9	2
ABI3100	BD1	Enzymatic	50	20	2	65	1.7	31.6	1
ABI3100	BD2	*Via* columns	46	3	5	49	1.2	28.2	2
ABI3100	BD3	*Via* columns	38	2	5	52	1.6	31.6	1
ABI3100	DR	Enzymatic	21	29	3	181	10.1	27.7	1
ABI3100	DR	*Via* columns	16	23	0	169	12.0	28.9	1
ABI3700	BD2	Enzymatic	72	0	12	43	0.8	31.3	2

dRhodamine Terminator Cycle Sequencing Kit

Brandstatter *et al.* (2005).

mtDNA Population Database: Size and Quality of Information

Table 5. Haplotypes truncated to positions 317, 320, 330, 343, and 345 as observed in the SWGDAM database and in complete mtDNA sequences

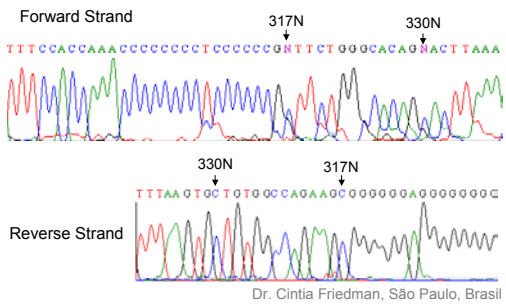
317	320	330	343	345	SWGDAM[a]	Complete mtDNA
C	C	C	C	C	*N* = 4839	*N* > 1700
A	.	.	.	.	1	1[b]
G	.	.	.	.	0	2[b,d]
T	.	.	.	.	1	0
N	.	.	.	.	14	0
del	.	.	.	.	1	0
.	T	.	.	.	0	2[c]
.	N	.	.	.	29	0
.	.	G	.	.	0	1[b]
.	.	N	.	.	10	0
.	.	.	N	.	3	0
.	.	.	.	N	4	0
.	.	.	.	T	1	1[e]
N	N	.	.	.	25	0
N	.	N	.	.	1	0
N	.	N	N	.	1	0
.	N	N	.	.	2	0
.	N	.	.	N	3	0
.	N	.	N	N	1	0
.	.	N	.	N	1	0
.	.	.	N	N	1	0
					99	7

HV2 phantom mutations

Post C-stretch ambiguities

Brandstatter *et al.* (2005).

mtDNA Population Database: Size and Quality of Information



Dr. Cintia Friedman, São Paulo, Brasil

mtDNA Population Database: Size and Quality of Information

- Recent efforts to increase DB sizes and quality have been undertaken by the NIJ (Grant to AFDIL Research Section) for entire control region sequences.
- EDNAP Mitochondrial Population Database (EMPOP) – developing QC tools to check sequences, including the ability to see electropherograms of all polymorphisms.

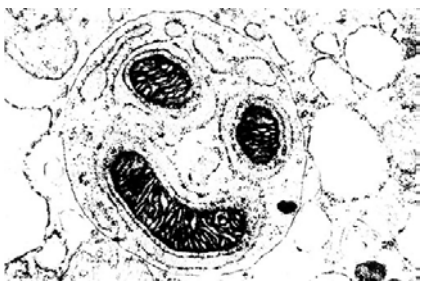
Acknowledgments

- Dr. John Butler (NIST)
- Dr. Tom Parsons (ICMP, formally AFDIL)
- Dr. Cintia Friedman (Brasil)
- Jodi Irwin, Rebecca Just, and the AFDIL Research Section.

Disclaimer

This project was funded by the National Institute of Justice through interagency agreement 2003-IJ-R-029 to the NIST Office of Law Enforcement Standards. 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.

QUESTIONS?



http://www-medlib.med.utah.edu/WebPath/jpeg2/EM003.jpg