Y-Chromosome and Mitochondrial DNA Analysis

mitochondrial DNA

NEAFS 2006 Workshop
Rye Brook, NY
November 1, 2006

Northeastern Association of Forensic Scientists

Dr. John M. Butler
Dr. Michael D. Coble

john.butler@nist.gov
Michael.Coble@afip.osd.mil

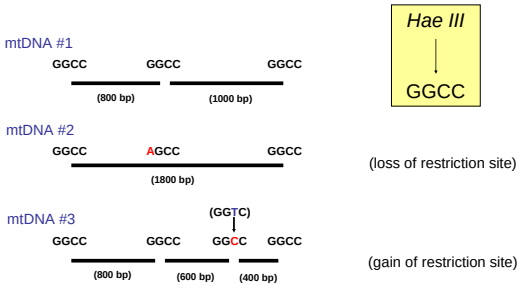
A Brief Sidestep…

- mtDNA as a genetic tool… "mitogenomics"
- The lack of apparent recombination, and high mutation rate make mtDNA an excellent tool for studying human evolution.
- Some of these insights have also been useful for the mtDNA forensic scientist.

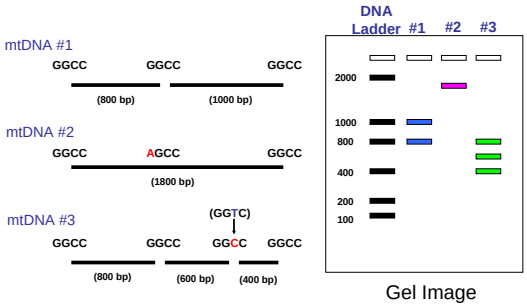
Methods for Measuring mtDNA Variation

- 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)

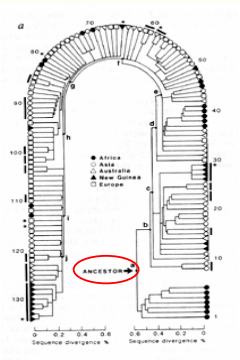
RFLP Analysis



RFLP Analysis



mtDNA as a Genetic Marker

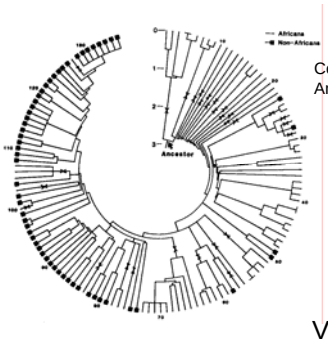


Cann *et al.* 1987

High res. RFLP analysis of 134 mtDNA types from 148 individuals ~ 370 different restriction sites per individual.

"Out of Africa"

mtDNA as a Genetic Marker



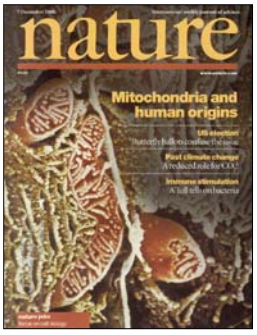
Control Region Sequence Analysis of 189 individuals

Vigilant *et al.* 1991

mtDNA as a Genetic Marker

- Templeton (1992) *Science* – Found phylogenetic trees that were more parsimonious than Vigilant *et al.* **AND** these trees did not suggest an "Out of African" origin.
- More sequence data and better tree-building methods confirmed the OOA hypothesis (Penny *et al.* 1995; Watson *et al.* 1997)

mtDNA as a Genetic Marker



Ingman *et al.* (2000)

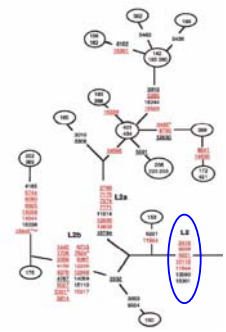
53 entire genome sequences from diverse global populations.

Confirmation for OOA.

mtDNA as a Genetic Marker

- RFLP variation has revealed continent-specific polymorphisms for classifying mtDNAs.
- Haplotype – the mtDNA sequence variations within an individual (e.g. your HV1/HV2 type).
- Haplogroup – a group of related haplotypes. These form monophyletic clades on a phylogenetic tree.

mtDNA Haplogroups

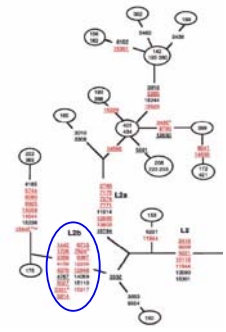


Each haplogroup cluster is defined by a set of ***specific, shared*** polymorphisms.

In this cluster, all individuals belonging to the African L2 haplogroup share a set of 7 SNPs in the coding region.

Herrnstadt et al. AJHG (2002)

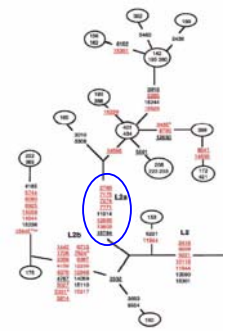
mtDNA Haplogroups



In this cluster, all individuals belonging to the African L2b sub-haplogroup share a set of 17 SNPs in the coding region.

Herrnstadt et al. AJHG (2002)

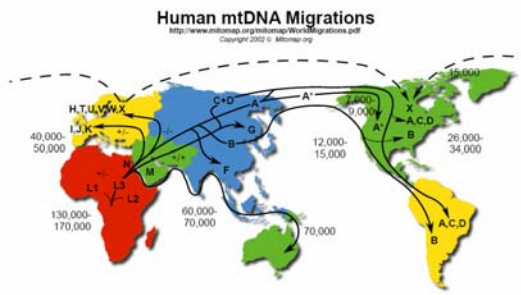
mtDNA Haplogroups



In this cluster, all individuals belonging to the African L2a sub-haplogroup share a set of 8 SNPs in the coding region.

Herrnstadt et al. AJHG (2002)

Human Migration Model



mtDNA Haplogroups (HV1/HV2)

- J - 16069 C-T 16126 T-C 73 A-G 295 C-T
- T - 16126 T-C 16294 C-T 73 A-G
- V - 16298 T-C 72 T-C
- L3e3 - 16223 C-T 16265 A-C 73 A-G 150 C-T 195 T-C

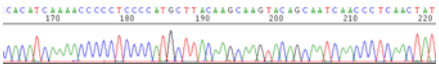
Generally, very good concordance between CR and coding haplogroups

Macaulay *et al.* (1999) *AJHG* **64:** 232-249
Allard *et al.* (2002) *JFS* **47:** 1215-1223
Brandstatter *et al.* (2004) *IJLM* **118:** 294-306

Tools for mtDNA Screening

Disadvantages to Sequencing

- Expensive
  - Primarily due to intensive labor in data analysis
- Error potential 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

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

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

Methodologies for SNP Typing

High-tech
- **SNaPshot (minisequencing)**
- Luminex 100 allele-specific hybridization
- Pyrosequencing
- TaqMan
- Primer extension with time-of-flight mass spectrometry
- TagArray (SNPstream UHT)
- Affymetrix hybridization chip

Low tech
- **Reverse dot blot (LINEAR ARRAYs)**
- PCR-RFLP
- Allele-specific PCR

See Budowle *et al.* (2004) *Forensic Sci. Rev.* 16:21-36 for a review of some SNP typing technologies

Mito "Strips"

- Roche Applied Science (Indianapolis, IN) has released a mtDNA typing kit.
- LINEAR ARRAY Mitochondrial DNA HVI/HVII Region-Sequence Typing Kit
- Cat. No. 03 527 867 001
- Cost $1500 for 50 reactions
- NIST was involved in beta-testing and performed a population study with these LINEAR ARRAYs

Previous Publications on mtDNA Typing Assays with SSO Probes (dot blot, reverse dot blot, linear arrays)

- Stoneking *et al.* (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet*. 48:370-382
- Skowasch, K., *et al.* (1994) Development of PCR-based reverse dot-blot typing system for the control region of mtDNA. *Proceedings of the Fifth International Symposium on Human Identification*, Madison, WI: Promega, p. 127.
- Comas, D., *et al.* (1999) Eur. J. Hum. Genet. 7:459-468
- Calloway, C.D., *et al.* (2000) *Am. J. Hum. Genet.* 66:1384-1397
- Reynolds, R., *et al.* (2000) *J. Forensic Sci.* 45(6):1210-1231
- Gabriel, M.N., *et al.* (2001) *Croatian Medical Journal* 42(3):328-335
- Gabriel, M.N., et al. (2003) *Croatian Medical Journal* 44(3):293-298
- Calloway, C., et al. (2003) Validation of the LINEAR ARRAY Mitochondrial DNA HVI/HVII Region-Sequence Typing kit. *Proceedings of the 14th International Symposium on Human Identification*.
- Calloway, C., et al. (2003) Applications of the LINEAR ARRAY Mitochondrial DNA HVI/HVII Region-Sequence Typing kit. *Proceedings of the 14th International Symposium on Human Identification*.
- Kline, M.C., et al. (2005) *J. Forensic Sci.* 50(2):377-385

Terry Melton population studies…

Am. J. Hum. Genet. 48:370-382, 1991

Population Variation of Human mtDNA Control Region Sequences Detected by Enzymatic Amplification and Sequence-specific Oligonucleotide Probes

Mark Stoneking, Dennis Hedgecock, Russell G. Higuchi, Linda Vigilant, and Henry A. Erlich

Department of Human Genetics, Cetus Corporation, Emeryville, CA; Bodega Marine Laboratory, University of California, Bodega Bay, CA; and Division of Biochemistry and Molecular Biology, University of California, Berkeley

Dot blot assay

Control Region

HVI **HVII**

A B C D — A B C D E

IA1 IA2 IA3 | IB1 IB2 IB3 | IC1 IC2 IC3 | ID1 ID2

IIA1 IIA2 | IIB1 IIB2 IIB3 | IIC1 IIC2 IIC3 | IID1 IID2 | IIE1 IIE2

23 probes across 9 regions

14 SNPs

16126 16129 | 16217 16223 | 16304 16311 | 16362

73 | 146 152 | 195 199 | 247 | 309.1

7 SNP sites **7 SNP sites**

European Journal of Human Genetics (1999) 7, 459–468

http://www.stockton-press.co.uk/ejhg

ARTICLE

Analysis of mtDNA HVRII in several human populations using an immobilised SSO probe hybridisation assay

David Comas[1], Rebecca Reynolds[2] and Antti Sajantila[1]

[1]*Department of Forensic Medicine, University of Helsinki, Finland*
[2]*Roche Molecular Systems, Alameda, CA, USA*

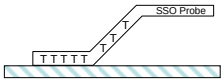
HVII

A B C D E

IIA1 IIA2 | IIB1 IIB2 IIB3 | IIC1 IIC2 IIC4 IIC5 | IID1 IID2 | IIE2

73 | 146 150 152 | 189 195 198 200 | 247 | 309.1

10 SNP sites



Detection Chemistry

Saiki, R.K., *et al.* (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc. Natl. Acad. Sci. USA* 86:6230-6234

16026C

3,3',5,5'-Tetramethyl-benzidine (TMB)

Colorless substrate → Colored Precipitate

Streptavidin-horseradish peroxidase enzyme conjugate

HRP

Streptavidin

Biotin

probe

3 3′-CTTATAACATGTCATGGTA-5′

2 3′-CTTATAACGTGCCATGGTA-5′ 2

1 3′-CTTATAACATGCCATGGTA-5′

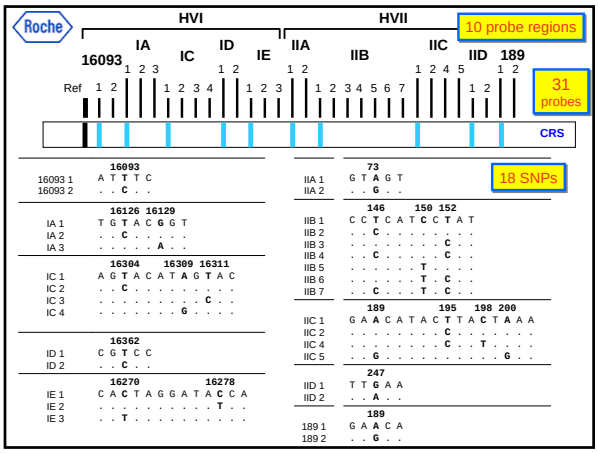
16026 16029

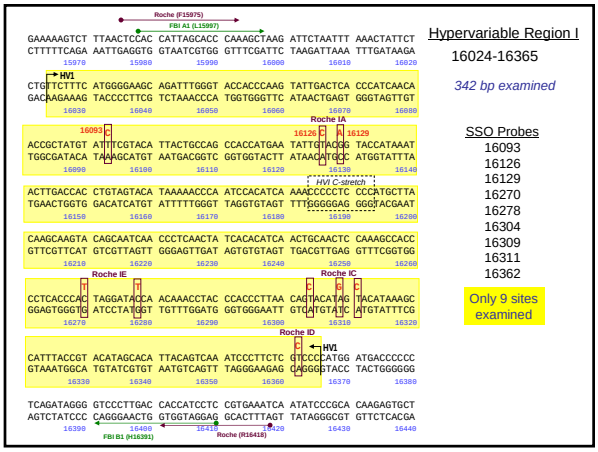
Biotin-labeled PCR product

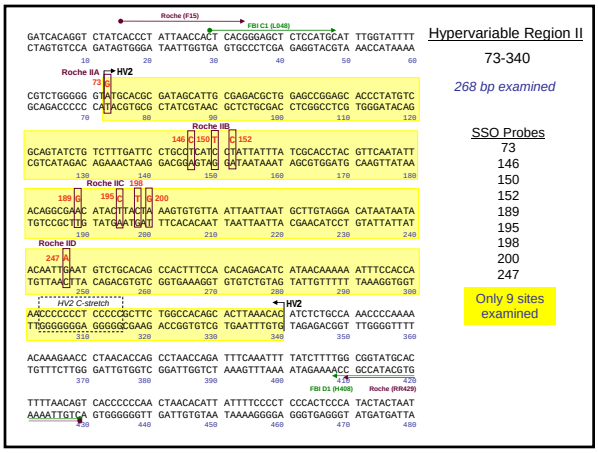
Biotin……GAATATTGTACGGTACCAT……3'

CTTATAACATGCCATGGTA-5' **Probe**

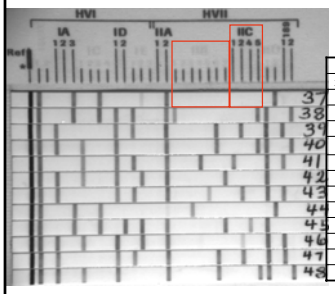
Nylon membrane


Roche
HVI
HVII
10 probe regions
IA
IC
ID
IE
IIA
IIB
IIC
IID
16093
189
Ref
31 probes
CRS
18 SNPs


Hypervariable Region I
16024-16365
342 bp examined
SSO Probes
16093
16126
16129
16270
16278
16304
16309
16311
16362
Only 9 sites examined


Hypervariable Region II
73-340
268 bp examined
SSO Probes
73
146
150
152
189
195
198
200
247
Only 9 sites examined

Data Interpretation for LINEAR ARRAYs



Analysis of probe results is still manual!

	HVI					HVII			
1609 2	A 3	C 3	D 0	E w2	A 2	B 0	C 0	D 2	18 0
1	3	3	1	1	2	1	5	1	2
1	1	0	1	1	2	6	2	1	1
1	1	1	1	0	2	6	5	1	2
1	0	1	1	1	2	4	1	1	1
1	1	1	1	2	2	7	0	1	0
1	1	0	1	1	2	5	2	1	1
1	3	3	0	0	w2	3	0	2	0
1	3	3	2	2	2	7	4	1	0
1	1	1	1	2	2	0	w2	1	1
1	1	0	1	1	2	6	2	1	1
1	1	1	1	0	2	6	5	1	2

Typing results from 50 ng of each PCR product

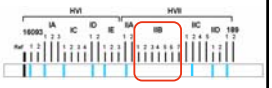
Margaret Kline, NIST

"Blank" Calls on LINEAR ARRAYs

NIST observed 640 "blanks" (**9.6% of calls**) on 346 different individuals (**52% of samples typed**).

*Different individuals typing as a blank for the same probe region could have different substitutions but for the purposes of data analysis the blanks are considered to represent the same variant (see Melton *et al.* (2001) *J. Forensic Sci.* 46(1):46-52)

SSO Probe Region	Number Observed	Frequency	Budowle et al. (1999) Cau, AA, His
16093	23	3.5%	
HVIA	33	5.0%	3, 9, 3%
HVIC	76	11.4%	3, 20, 10%
HVID	33	5.0%	7, 17, 4%
HVIE	60	9.0%	
HVIIA	3	0.5%	0, 0, 0%
HVIIB	96	14.4%	16, 70, 55%
HVIIC	122	18.3%	11, 47, 13%
HVIID	42	6.3%	5, 5, 18%
189	152	22.8%	



PCR product fails to hybridize to any probe in region due to additional polymorphisms in the probe region that prevent hybridization

Probe Region HVIIB

	146							150	152		
IIB 1	C	C	T	C	A	T	C	C	T	A	T
IIB 2	.	.	C	.	.	.	.	.	.	.	.
IIB 3	.	.	.	.	.	.	.	C	.	.	.
IIB 4	.	.	C	.	.	.	.	C	.	.	.
IIB 5	.	.	.	.	.	.	T	.	.	.	.
IIB 6	.	.	.	.	.	.	T	C	.	.	.
IIB 7	.	.	C	.	.	.	T	C	.	.	.

Blanks expected based on full sequence analysis of 1393 individuals

Nucleotide positions 151 and 153 are common variants in African Americans

Kline *et al.* (2005) *JFS*

Typing frequencies for 666 NIST population samples

#*	Freq	% Types	% People
1	185	65.6	27.8
2	46	16.3	13.8
3	18	6.4	8.1
4	4	1.4	2.4
5	3	1.1	2.3
6	4	1.4	3.6
7	1	0.4	1.1
8	9	3.2	10.8
9	2	0.7	2.7
10	4	1.4	6.0
11	1	0.4	1.7
12	1	0.4	1.8
18	1	0.4	2.7
23	1	0.4	3.5
28	1	0.4	4.2
51	1	0.4	7.7

Summary of NIST Population Typing with Roche mtDNA LINEAR ARRAYS

- 282 different types
- 185 were unique (occurred only once)
- 51 samples had "Most Common Type"

"Most Common Type" evaluated further with mtDNA coding region SNP assay (later)

Comparison of Other U.S. Population Data with SSO Probes

	Population	N	#types	diversity	Most Common Type	MCT frequency
11	Caucasian	922	226	0.964	11111111	15.4%
10	African Am	805	251	0.983	12112021	6.8%
7	Hispanic	555	170	0.963	12122011	11.7%
	Total	2282	502	0.998	11111111	**7.2%**

8 regions, 21 probes, 13 SNPs

Melton *et al.* (2001) *J. Forensic Sci.* 46(1): 46-52

Population	N	#types	diversity	Most Common Type	MCT frequency
Caucasian	286	116	0.960	1111111111	16.4%
African Am	252	129	0.977	1141224211	10.7%
Hispanic	128	74	0.954	1102120111	16.4%
Total	666	282	0.985	1111111111	**7.7%**

10 regions, 31 probes, 18 SNPs

Kline *et al.* (2003) NIST population study

HV1 Null Alleles

A

HVI Array Locus (rCRS position)	African American	Hispanic	Caucasian	Haplogroup Associated Polymorphisms	Percentage of Null Alleles from Haplogroup Polymorphisms
16093	8	8	7	--	--
HVIA (16126; 16124)	24	2	7	16124C - L3b and L3d	27/33 (82%)
HVIC (16304; 16309; 16311)	28	38	10	16320T - L3e2 16290T and 16319A - A (Asian)	69/76 (91%)
HVID (16362)	29	3	2	16360T - L1c	29/34 (85%)
HVIE (16270; 16278)	40	11	8	16270T and 16278T - L1b 16264T - L3e4 16265T - L3e3 16265C - L1c2	33/59 (56%)
	129	62	34		158/225 (70%)
			225 total		

Coble et al. (in review)

HV2 Null Alleles

B

HVII Array Locus (rCRS position)	African American	Hispanic	Caucasian	Haplogroup Associated Polymorphisms	Percentage of Null Alleles from Haplogroup Polymorphisms
HVIIA (73)	1	0	2	72T-C - preV	1/3 (33%)
HVIIB (146; 150; 152)	46	36	15	151C-T - L1c 143G-A - L2a* 153A-G - A2; X*	88/97 (91%)
HVIIC (189; 195; 198; 200)	66	19	37	186 C-A - L1c 185G-T - L1b 189A-C - L2b/c 185G-A; 189A-G; and 200A-G - L3e* 194C-T - D*/D4b2 199T-C and 204T-C - I	78/122 (64%)
HVIID (247)	5	27	7	249A-del - CZ 242C-T - J1* 250T-C - I2*	32/39 (82%)
HVIIE (189)	103	18	32	182C-T and 195T-C L1*/L2* 185G-T - L1b 195T-C and 198C-T - L2a* 189A-C - L2b/c 185G-A and 200A-G - L3e1a 185G-A - J1* 194C-T - D*/D4b2	123/153 (80%)
	221	100	93		322/414 (78%)
			414 total		

Coble et al. (in review)

Sequencing – Increased Discrimination

# times haplotype observed	LINEAR ARRAY	HV1	HV1+HV2	control region	
1	**185**	**334**	**454**	**502**	
2	45	38	36	37	
3	18	11	11	10	
4	4	7	4	1	
5	4	4	6	5	
6	3	5	2	3	
7	1	7	3	·	
8	9	·	1	·	
9	2	1	·	·	
10	4	1	·	**1**	
11	1	2	·	·	
12	1	1	·	·	
17	·	·	**1**	·	
18	1	·	·	·	
23	1	·	·	·	
28	1	·	·	·	
40	·	**1**	·	·	
51	**1**	·	·	·	
HD	0.9869	0.9936	0.9982	0.9990	
% DC	42.19%	61.86%	77.78%	83.93%	total
# HT	**281**	**412**	**518**	**559**	**666**

Coble et al. (in review)

SNP Typing Instrumentation

Multi-Color Capillary Electrophoresis (ABI 310 or 3100)

PCR & primer extension

TaqMan

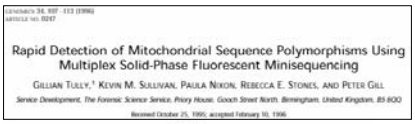
ABI 7000 SDS

SNP Extension Primer Design

- Must anneal to DNA template with 3'end of primer next to SNP site
- Can anneal to either top strand or bottom strand
- Should have uniform annealing temperature (by lengthening 5'end of SNP primer)
- Should not form significant hairpins or dimers with other SNP or PCR primers

```
          ddC
-TCTCATAATA(G/A)GATAAAACAC-
-AGAGTATTAT(C/T)CTATTTTGTG-
          ddG
```

Early Multiplex SNP Detection Work



Rapid Detection of Mitochondrial Sequence Polymorphisms Using Multiplex Solid-Phase Fluorescent Minisequencing

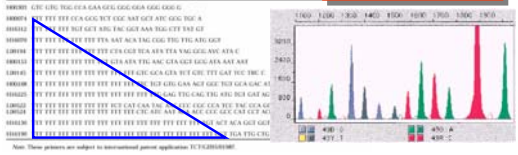
Gillian Tully,† Kevin M. Sullivan, Paula Nixon, Rebecca E. Stones, and Peter Gill

Service Development, The Forensic Science Service, Priory House, Gooch Street North, Birmingham, United Kingdom, B5 6QQ

16069 16129
16189 16224
16311
73 146
152 195
247 309.1
523 525

Variable poly(T) 5'tails to spread products

Control region SNPs

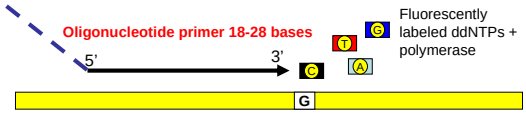


Allele-Specific Primer Extension

SNP Primer is extended by one base unit

ABI PRISM® SNaPshot™ Multiplex System

"tail" used to vary electrophoretic mobility



Fluorescently labeled ddNTPs + polymerase

PCR Amplified DNA Template

Products can be electrophoretically separated on an ABI 310, 3100

Allele-Specific Primer Extension

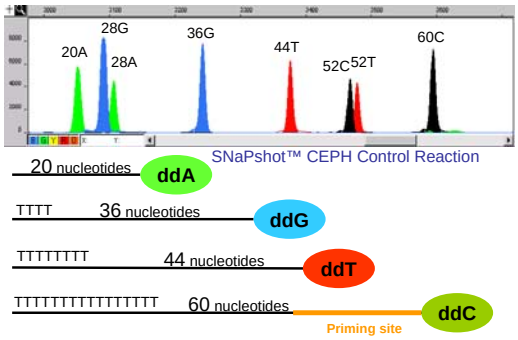
The use of "tailed" SNP primers allows for multiplexing in the SNaPshot assay

Sequences for 11 SNP primers

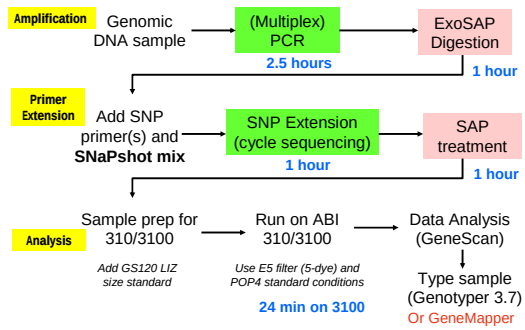
TGTTGGATCAGGACATCCC	19/19
TTTTTCAGAAGTGAAAGGGGGC	18/22
TTTTTTTTTTACTAAGAAGAATTTTATGGA	20/30
TTTTTTTTTTTTTTTAGACCCAGCTACGCAAAATC	20/34
TTTTTTTTTTTTTTTTTTTTGACACGTACTACGTTGTAGC	20/38
TTTTTTTTTTTTTTTTTTTTTTTTCCACAACACTTTCTCGGCCT	20/42
TTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGGGCTATTTAGGCTTTATG	22/46
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAGCCATTCAAGCAATCCTATA	23/50
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTTAGAACTGGAATAAAAGCTAG	25/54
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCCTCCCACTCCCATACTAC	20/58
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGAATGATGGTTGTCTTTGG	21/62

Template binding sequence – black
Tailed sequence for fragment separation - red

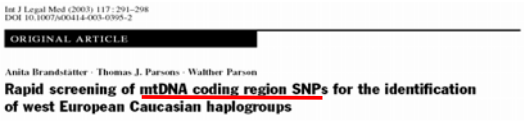
Allele-Specific Primer Extension



Protocol with SNaPshot™ "Kit"



Use of Haplogroup Defining mtSNPs


Int J Legal Med (2003) 117 : 291–298
DOI 10.1007/s00414-003-0395-2

ORIGINAL ARTICLE

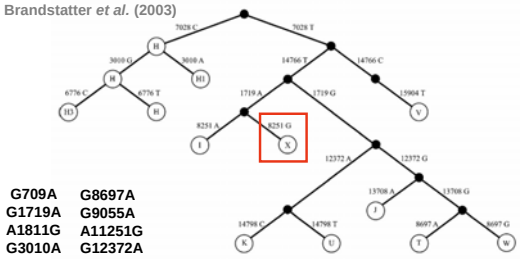
Anita Brandstätter · Thomas J. Parsons · Walther Parson

Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups

Identified coding region SNP to classify the 9 major Western European haplogroups (plus 2 sub-haplogroups).

Use of Haplogroup Defining mtSNPs

Brandstatter *et al.* (2003)



G709A	G8697A
G1719A	G9055A
A1811G	A11251G
G3010A	G12372A
T6365C	G13708A
T6776C	C14766T
C7028T	T14798C
G8251A	C15904T

16 mtSNPs run in two SNaPshot 8plex reactions

Use of Haplogroup Defining mtSNPs

Haplogroups Defined by Control Region mtSNPs in SWGDAM Caucasian Samples (n = 1771)

Allard *et al.* (2002) *J. Forensic Sci.* 47(6):1215-1223

- **H** (46%): 73A
- **U** (15.6%): 16270T
- **T** (10.5%): 16126C, 16294T
- **J** (10%): 16069T, 16126C, 295T
- **K** (8.9%): 16224C, 16311C
- **I** (2%): 16223T, 199C, 204C, 250C
- **V** (1.9%): 16298C, 72C
- **W** (1.9%): 16223T, 189G, 195C, 204C, 207A
- **X** (1.6%): 16189C, 16223T, 16278T, 195C
- **M** (1.9%): 16223T, 16298C

If a **G is observed at 8251**, then the sample can be classified as a member of **haplogroup X** so the following control region SNPs should be expected: 16189C, 16223T, 16278T, 195C

Use of Haplogroup Defining mtSNPs

Advantages

Sensitive – 1 pg genomic DNA
Short amplicons – degraded DNA
Multiplexed PCR – conserves template
POD – 88.6% among 277 unrelated Austrian Caucasians

Disadvantages

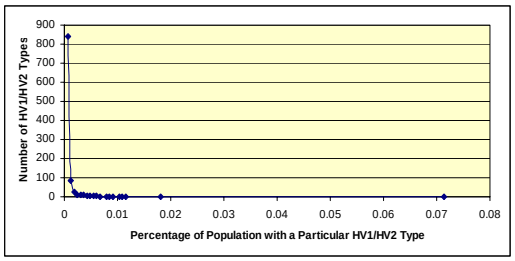
No "kit" – need to order, validate primers.
Probability of random match among Caucasians is ~11.4%

Emerging mtDNA technologies

mtDNA genome sequencing for increased discrimination

mtDNA Population Distribution

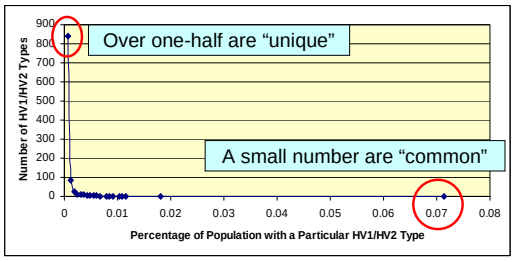
Caucasians (n=1665)



Coble *et al.* (2004)

mtDNA Population Distribution

Caucasians (n=1665)



Coble *et al.* (2004)

Framing the Problem

- The greatest limitation for mtDNA testing lies with the small number of common types for which the power of discrimination is low.
- ~20% of the time, the Forensic Scientist encounters a HV1/HV2 type that occurs at greater than ~0.5% of the population.
- In database or mass fatality comparisons: multiple hits will occur for these common types.

A Case Example

- September 15, 1943 - B17F Bomber returning from a mission to Port Moresby, New Guinea

A Case Example

- The plane crashes in the Owen Stanley Mountain range due to “adverse weather.”
- Subsequent searches proved negative.
- 11 crewmen declared non-recoverable on July 22, 1949.

A Case Example

- October 9, 1992 - A private company helicopter discovers crash site.
- mtDNA testing reveals that 3/11 crewmen share the same HV type (263 A-G, 315.1 C).
- Further VR testing could distinguish 1 of the 3 crewmen (16519 T-C). However, 2 crewmen still matched.

A Case Example

- Partial dental records were used to associate 3 teeth among the 2 crewmen matching in the CR.
- One L femur could not be associated with either crewmen, and was buried in a grave containing group remains

Strategy for SNP Identification

- Sequence the entire genome of unrelated individuals sharing common HV1/HV2 types in the Caucasian population (focus on 18 of 22 common types that occur at a frequency of 0.5% or greater).

Ethical Considerations

- ~265 characterized diseases associated with mtDNA mutations in the coding region (Mitomap – www.mitomap.org)
- To avoid having forensic testing from evolving into genetic counseling, we decided to focus on neutral SNPs in the mtGenome.

SNPs for Discrimination

- Non-coding sites in the control region (outside of HV1/HV2).
- Non-coding “spacer” regions throughout the mtGenome.
- Silent mutations in protein coding genes.

SNPs for Discrimination

- Practical application – A set of SNP sites that can be rapidly assayed to provide maximal discrimination.
- Avoids further sequencing.
- Allele Specific Primer Extension – small amplicons, multiplexed - can conserve template, run on standard instrumentation.

Common mtDNA Haplogroups

Com	Haplo	Seq (+ CRS)
31	H1	CRS
25	H2	152 C
11	H3	16129 A
8	H4	16263 C
12	H5	16304 C
11	H6	73 G
7	H7	16162 G 16209 C 73 G

Length Variation in HV2 C-stretch – ignored (Stewart *et al.* (2001)

Common mtDNA Haplogroups

Com	Haplo	Seq (+ CRS)
15	J1	16069 T 16126 C 73 G 185 A 228 A 295 T
6	J2	16069 T 16126 C 73 G 228 A 295 T
12	J3	16069 T 16126 C 73 G 185 A 188G 228 A 295 T
3	J4	16069 T 16126 C 16145 A 16172 C 16222 T 16261 T 73 G 242 T 295 T
20	T1	16126 C 16294 T 16296 T 16304 C 73 G
10	T2	16126 C 16163 G 16186 T 16189 C 16294 T 73 G 152 C 195 C
8	T3	16126 C 16294 T 16296 T 73 G
25	V1	16298 C
14	K1	16224 C 16311 C 73 G 146 C 152 C
8	K2	16093 C 16224 C 16311 C 73 G
7	K3	16224 C 16311 C 73 G

241 total genomes from 18 common HV1/HV2 types
(~14% of the total database)

Strategies for Whole mtGenome Analysis

58 PCR rxn
116 seq rxn
Levin *et al.* (1999)

24 PCR rxn
48 seq rxn
Rieder *et al.* (1998)
Ingman *et al.* (2000)

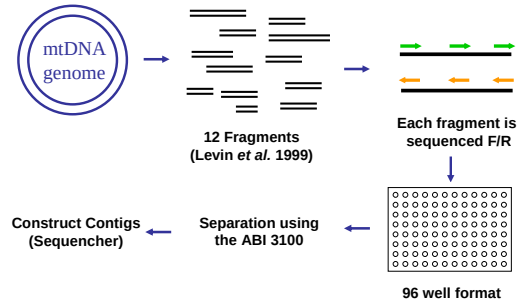
18 PCR rxn
36 seq rxn
Aldridge *et al.* (2003)

12 PCR rxn
95 seq rxn
Coble *et al.* (2004)

32 PCR rxn
64 seq rxn
Maca-Meyer *et al.* (2001)

15 PCR rxn
47 seq rxn
Kong *et al.* (2003)

Sequencing Strategy



The Nature of the SNPs

- Would the SNPs that resolve one group be useful for resolving other closely related groups?

Com	Haplo	Seq (+ CRS)		
31	H1	CRS		
25	H2	152 C		
11	H3	16129 A		"Hot Spots"
8	H4	16263 C		
12	H5	16304 C		
11	H6	73 G		
7	H7	16162 G	16209 C	73 G

The Nature of the SNPs

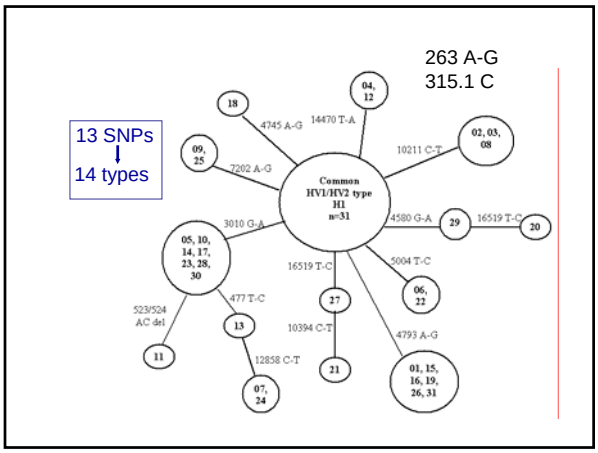
- Are resolving SNPs **slow and rare?** Did these SNPs arise once during the evolution of a haplogroup?

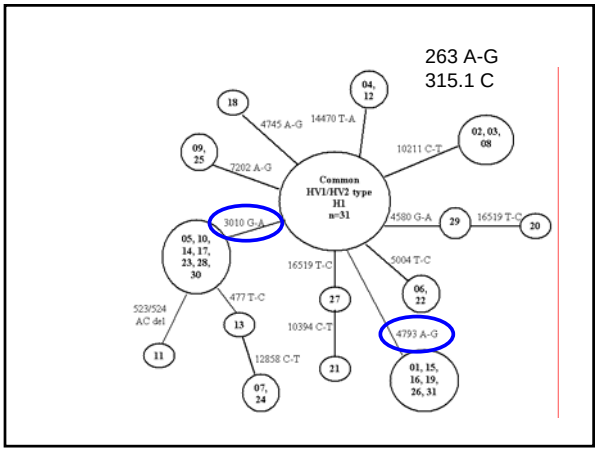
OR…

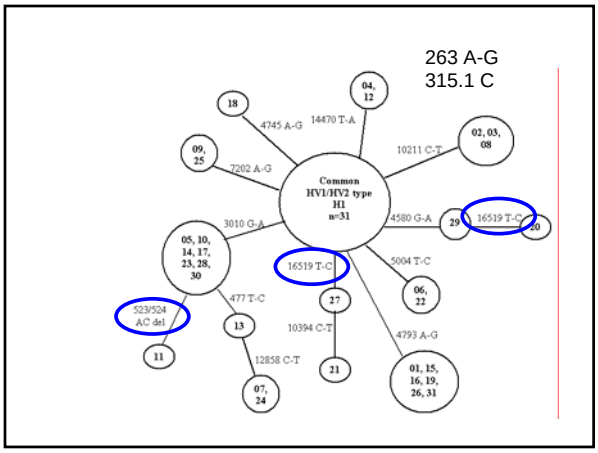
- Are resolving SNPs "universally" **fast hot spots**, useful for all haplogroups (L, M, N)?

OR….

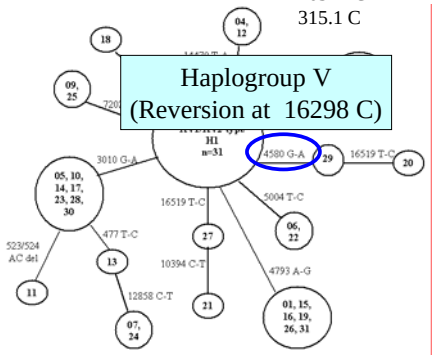
- Are resolving SNPs a combination of the two?


263 A-G
315.1 C
13 SNPs
14 types
Common HV1/HV2 type H1 n=31
04, 12
18
4745 A-G
14470 T-A
02, 03, 08
10211 C-T
09, 25
7202 A-G
4580 G-A
29
16519 T-C
20
3010 G-A
05, 10, 14, 17, 23, 28, 30
16519 T-C
5004 T-C
06, 22
27
523/524 AC del
477 T-C
13
10394 C-T
4793 A-G
11
12858 C-T
21
01, 15, 16, 19, 26, 31
07, 24

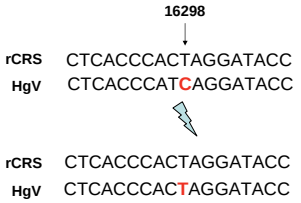

263 A-G
315.1 C
Common HV1/HV2 type H1 n=31
04, 12
18
4745 A-G
14470 T-A
02, 03, 08
10211 C-T
09, 25
7202 A-G
4580 G-A
29
16519 T-C
20
3010 G-A
05, 10, 14, 17, 23, 28, 30
16519 T-C
5004 T-C
06, 22
27
523/524 AC del
477 T-C
13
10394 C-T
4793 A-G
11
12858 C-T
21
01, 15, 16, 19, 26, 31
07, 24


263 A-G
315.1 C
Common HV1/HV2 type H1 n=31
04, 12
18
4745 A-G
14470 T-A
02, 03, 08
10211 C-T
09, 25
7202 A-G
4580 G-A
29
16519 T-C
20
3010 G-A
05, 10, 14, 17, 23, 28, 30
16519 T-C
5004 T-C
06, 22
27
523/524 AC del
477 T-C
13
10394 C-T
4793 A-G
11
12858 C-T
21
01, 15, 16, 19, 26, 31
07, 24

263 A-G
315.1 C

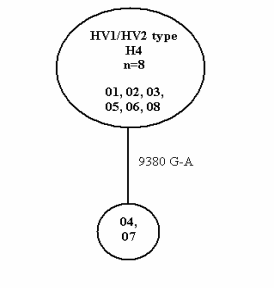


Reversion

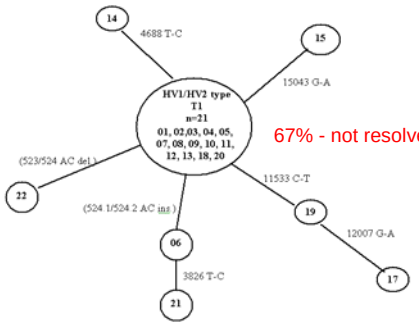


Sequence with HgV background…appears as a HgH (H1)

H4 - rCRS + 16263 T-C



T1 – Low Resolution

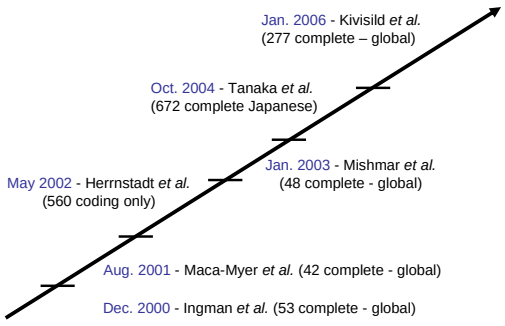


67% - not resolved

Brute-Force Sequencing

- Why not used information from the literature??
- Prior to 1999, only a handful of whole genome sequences in GenBank. Most of the mtDNA coding region data was from RFLP studies (assays ~ 20% of the genome)

mtGenomics



mtDB - Human Mitochondrial Genome Database

- Max Ingman (Uppsala University, Sweden)
- 1711 complete sequences and 839 coding region sequences.
- 2550 coding region sequences.

mtDB - Human Mitochondrial Genome Database

CRS		2550 Sequences									
Posn	Base	A	G	C	T	Gap	Location	Codon	Position	Amino Change	Syn?
9380	G	13	2537				COIII	58	3	Trp -> Trp	Yes

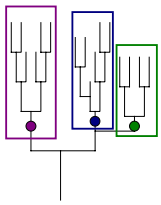
9380 G-A has only been observed in 11/2296 (0.48%) coding regions… would not be a good candidate if one was "trolling" the database for discriminating SNPs

Problem – very few common types in global DB

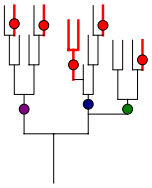
Summary

- 241 mtGenomes – 420 polymorphic sites in the coding region.
- 32/241 (13%) – matched one or more individuals over the entire mtGenome (0/12 H5 individuals matched; 4/8 H7 individuals matched).
- Homoplasies – common in HV1/HV2.

Homoplasy – Parallel Substitutions



3 basal polymorphisms that define3 clusters



“Red” mutation has occurred multiple times on the tree

Summary

- Percentage of sites that varied ranged from 1.0% (16S rRNA) to 6.6% (non-coding regions outside of the control region).
- ATP Synthase 8 (4.8%) and ATP Synthase 6 (3.7%) showed the greatest non-synonymous variation in the protein coding genes.

Gene	Length	Synonymous	Nonsynonymous	Total	% NonSyn
ND1	956	14	8	22	36.4%
ND2	1,042	25	11	36	30.6%
CO1	1,542	29	9	38	23.7%
CO2	684	14	4	18	22.2%
ATP8	207	3	5	8	62.5%
ATP6	681	7	20	27	74.1%
CO3	784	14	4	18	22.2%
ND3	346	5	2	7	28.6%
ND4L	297	5	1	6	16.7%
ND4	1,378	30	7	37	18.9%
ND5	1,812	39	15	54	27.8%
ND6	525	8	7	15	46.7%
CYB	1,141	23	15	38	39.5%
Total	11,341	216	108	324	33.1%