


SNPs and Strips:

Approaches to Rapid Screening of mtDNA types

John Butler

AAFS Workshop:
Forensic Human Mitochondrial
DNA Analysis
February 16, 2004



Presentation Outline

Advantages to Screening Methods

SNPs

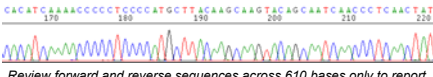
- Uses for mtSNPs
- Methodologies for SNP Typing
- SNP Assay Design for SNaPshot

Strips

- SSO Probe History and Chemistry
- Results with Roche LINEAR ARRAYS

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

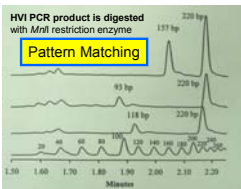
- Rapid results
- Aids in exclusion of non-matching samples
- Less labor intensive
- Usually less expensive
- Permits more labs to get involved in mtDNA

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.

Two Examples of mtDNA Screening Methods

Butler et al. (1998) *Electrophoresis* 19:119-124
mtDNA typing with PCR-RFLP and CE-LIF



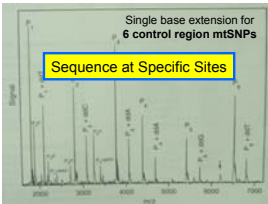
Pattern Matching

3 different mtDNA sequences can be distinguished from one another

DNA sizes (bp) after *MnI* digestion

220-35-157-28-3
220-35-39-93-3-22-28-3
220-35-39-118-28-3

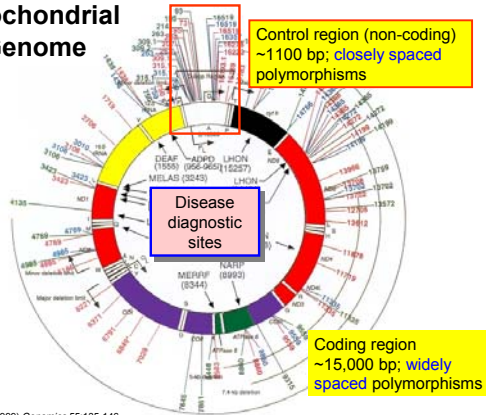
Li, Butler, et al. (1999) *Electrophoresis* 20:1258-1265
mtDNA SNP typing with TOF mass spectrometry



Sequence at Specific Sites

HV1	HV2
H16069	L00146
H16311	H00152
	L00195
	H00247

Mitochondrial Genome



Control region (non-coding)
~1100 bp; closely spaced polymorphisms

Disease diagnostic sites

Coding region
~15,000 bp; widely spaced polymorphisms

Levin et al (1999) *Genomics* 55:135-146

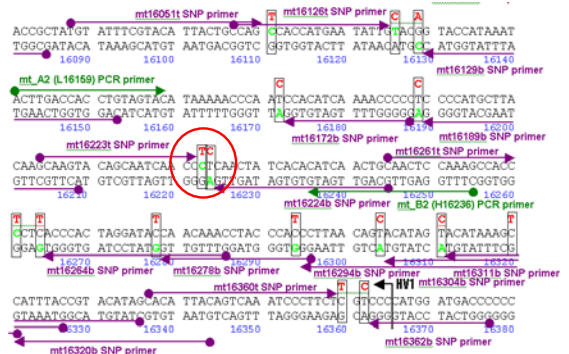
Control Region (16024-576) SNPs

- 1,122 nucleotide positions; typically only 610 bases analyzed (HVI: 16024-16365; HVII: 73-340)
- Challenges with typing closely spaced SNPs
 - Probes are disrupted by neighboring polymorphism(s)

Coding Region (577-16023) SNPs

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

Closely Spaced Polymorphisms Make Assay Design Challenging for Control Region Mitochondrial SNPs



SNPs

Use of mtSNPs

- Rapid screen with informative control region SNPs
 - Minisequencing
 - Tully *et al.* (1996) *Genomics* 34: 107-113
 - Budowle *et al.* (2003) *Annu. Rev. Genomics Hum. Genet.* 4:119-141
- Quality control for control region sequence data with informative coding region SNPs
 - Coding region SNPs used to classify haplogroup
 - Brandstätter *et al.* (2003) *Int. J. Legal Med.* 117: 291-298
 - Allard *et al.* (2002) *J. Forensic Sci.* 47(6): 1215-1223
- Aid resolution of most common HV1/HV2 sequence types with coding region SNPs to improve forensic discrimination
 - Full mtGenome sequencing to find optimal SNP markers
 - Parsons and Coble (2001) *Croat. Med. J.* 42(3): 304-309
 - Coble *et al.* (2004) *Int. J. Legal Med.*, in press
 - Vallone *et al.* (2004) *Int. J. Legal Med.*, in press

Early Multiplex SNP Detection Work

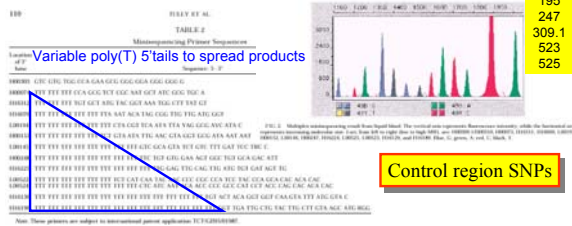
GENOMICS 34, 107-113 (1996)
ARTICLE NO. 0287

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

GILLIAN TULLY,¹ 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 40Q

Received October 25, 1995; accepted February 20, 1996.

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



Evaluation of Useful SNP Sites in mtDNA Control Region Across African American and Caucasian Samples

CRS site	AI Am	Cauc	CRS site	AI Am	Cauc	CRS site	AI Am	Cauc	CRS site	AI Am	Cauc
16048	G	A	16215	A	G	16327	C	T	199	A	C
16051	A	G	16222	C	T	16343	A	G	194	C	T
16089	C	T	16223	C	T	16355	C	T	195	T	C
16096	T	C	16230	T	C	16356	T	C	196	C	T
16093	T	C	16230	A	G	16360	C	T	199	T	C
16114	C	A	16231	T	C	16362	T	C	200	A	G
16124	T	C	16256	C	T	16390	G	A	204	T	C
16126	T	C	16261	C	T	16399	A	G	207	G	A
16129	G	A	16263	T	C	16519	T	C	215	A	G
16145	G	A	16264	C	T	16527	C	T	217	T	C
16148	C	T	16265	A	C	84	C	T	225	G	A
16153	G	T	16270	C	T	72	T	C	226	T	C
16162	A	G	16271	T	C	73	A	G	228	G	A
16163	A	G	16278	C	T	93	A	G	236	T	C
16166	C	T	16285	A	G	95	A	G	239	T	C
16172	T	C	16291	C	T	119	T	C	242	C	T
16183	A	C	16292	C	T	143	G	A	247	G	A
16185	C	T	16293	A	G	148	T	C	250	T	C
16186	C	T	16294	C	T	150	C	T	263	A	G
16187	C	T	16295	C	T	151	T	T	295	C	T
16188	C	T									
16189	T	C									
16192	C	T									
16193	C	T									
16207	A	G									
16209	T	C									
16213	G	A									

FORENSICS AND MITOCHONDRIAL DNA: Applications, Debates, and Foundations*

Bruce Budowle,¹ Marc W. Allard,² Mark R. Wilson,³ and Ranajit Chakraborty¹

Annual Reviews in Genomics and Human Genetics 2003, 4:119-141

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

Use of Haplogroup Defining mtSNPs

ORIGINAL ARTICLE

Anita Brandstätter · Thomas J. Parsons · Walthar Parson

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

Defines 11 different haplogroups

16 mtSNPs run in two SNaPshot 8plex reactions

16 mtSNPs run in two SNaPshot 8plex reactions

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

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

Efforts with Coding Region Sequencing applied to human identity testing

- Tzen et al. (2001) *Forensic Sci. Int.* 120:204-209
 - Nucleotide positions 8389 to 8865 (portions of ATP8, ATP6)
 - 119 Chinese individuals
- Andreasson et al. (2002) *Biotechniques* 32:124-133
 - Pyrosequencing for 11 coding region reactions
 - 190 Swedish individuals
- Lee et al. (2002) *Int. J. Legal Med.* 116:74-78
 - mtCyt B
 - 98 Korean individuals
- Lutz-Bonengel et al. (2003) *Int. J. Legal Med.* 117:133-142
 - mtATP6, mtATP8, mtND4
 - 109 German individuals
- Poetsch et al. (2003) *Mitochondrion* 3:133-137
 - Nucleotide positions 8306 to 9021 (portions of tRNA K, ATP8, ATP6)
 - 180 German individuals
- Coble et al. (2004) *Int. J. Legal Med.*, in press
 - 241 complete mtGenomes from 18 most common Caucasian HV1/HV2 types

How many SNPs found in each study?

Recent Efforts with Whole mtGenome Analysis

- Ingman et al. (2000) *Nature* 408:708-713
 - 53 mtGenomes from diverse worldwide origin
 - GenBank AF348963-AF347015
- Maca-Meyer et al. (2001) *BMC Genetics* 2:13
 - 33 mtGenomes from diverse worldwide origin
 - GenBank AF381981-AF382013
- Herrnstadt et al. (2002) *Am. J. Hum. Genet.* 70:1152-1171
 - 560 mtGenomes (coding region only) from major African, Asian, and European origins
 - See MitoKor website for sequences: <http://www.mitokor.com>
- Kong et al. (2003) *Am. J. Hum. Genet.* 73:671-676
 - 48 mtGenomes from East Asian lineages
 - GenBank AY255133-AY255180
- Ingman et al. (2003) *Genome Res.* 13:1600-1606
 - 52 mtGenomes from Australian and New Guinean Aborigines and Polynesians
 - GenBank AY289051-AY289102
- Coble et al. (2004) *Int. J. Legal Med.*, in press
 - 241 mtGenomes from most common Caucasian types
 - GenBank AY495090-AY495330

Approximately 1,000 full mtGenomes have been sequenced

68 PCR rxn
136 seq rxn
Herrnstadt et al. (2002)

Strategies for Whole mtGenome Analysis

- Levin et al. (1999): 58 PCR rxn, 116 seq rxn
- Rieder et al. (1998) / Ingman et al. (2000): 24 PCR rxn, 48 seq rxn
- Aldridge et al. (2003): 18 PCR rxn, 36 seq rxn
- Coble et al. (2004): 12 PCR rxn, 95 seq rxn
- Maca-Meyer et al. (2001): 32 PCR rxn, 64 seq rxn
- Kong et al. (2003): 15 PCR rxn, 47 seq rxn

Publication of 560 Complete mtDNA Coding Region Sequences

Address: <http://www.genpat.uu.se/mitDE/polysties.html>

Am. J. Hum. Genet. 70:1152-1171, 2002

Anderson		746 Sequences					
Posn.	Base	A	G	C	T	Gap	Ins
3010	G	134	612				
3027	T		3	743			
3028	A	745				1	
3083	T		1	745			
3105	A	745	1				
3107	C		4		742		
3116	C		744	2			
3197	T		26	720			

497 polymorphisms identified outside the control region

coding region sequences from unrelated individuals to develop a more complete understanding of sequence diversity both within and between populations. We report here the presence or absence of a relatively small set of highly conserved sequences of the D-loop region. We report here the presence or absence of a relatively small set of highly conserved sequences of the D-loop region. We report here the presence or absence of a relatively small set of highly conserved sequences of the D-loop region.

Blazej, et al. (2003) *Genome Res.* 13:287-293

Polymorphism Ratio Sequencing: A New Approach for Single Nucleotide Polymorphism Discovery and Genotyping

Robert G. Blazej,¹ Brian M. Paegel,² and Richard A. Mathies^{1,2,3}

¹University of California, Berkeley/University of California, San Francisco Joint Bioengineering Graduate Group, Berkeley, California 94720, USA; ²Department of Chemistry, University of California, Berkeley, California 94720, USA;

30 minutes to sequence an entire mtGenome compared against another sample

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

SNP Typing Instrumentation at NIST

PCR & primer extension

Luminex Beads hybridization

Primer Extension

TaqMan

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-
    ← ddC
    
```

Detection of SNPs with ABI 310/3100

SNaPshot™ CEPH Control Reaction

- 20 nucleotides ddA
- 36 nucleotides ddG
- 44 nucleotides ddT
- 60 nucleotides ddC

Poly(T) tail or non-nucleotide linker to aid separation

Priming site

Multiplexing possible by use of different length primers

Protocol with SNaPshot™ "Kit"

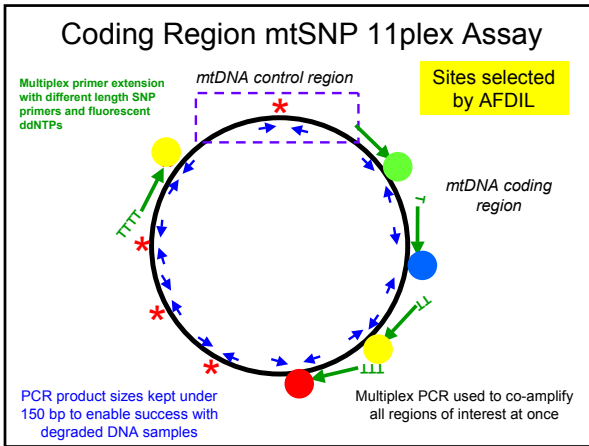
```

    Genomic DNA sample → (Multiplex) PCR → ExoSAP Digestion
                                2.5 hours                1 hour

    Add SNP primer(s) and SNaPshot mix → SNP Extension (cycle sequencing) → SAP treatment
                                1 hour                1 hour

    Sample prep for 310/3100 → Run on ABI 310/3100 → Data Analysis (GeneScan)
                                Add GS120 LIZ size standard   Use E5 filter (5-dye) and POP4 standard conditions
                                24 min on 3100
    
```

Type sample (Genotyper 3.7) Or GeneMapper

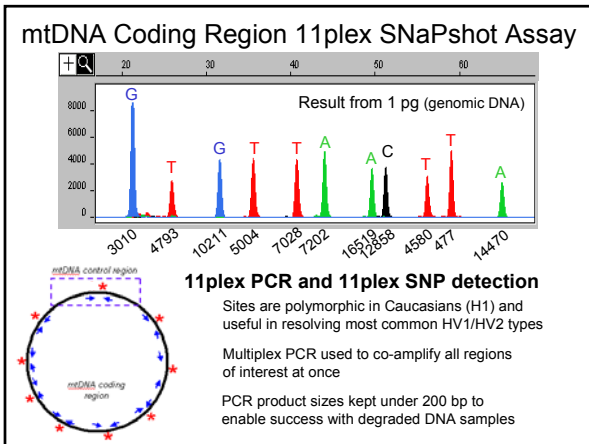


The use of "tailed" SNP primers allows for multiplexing in the SNaPshot assay

Sequences for 11 SNP primers

TGTTGGATCAGGACATCCC	19/19
TTTTTCAGAAAGTAAAGGGGGC	18/22
TTTTTTTTTTACTAAGAAGAATTTTATGGA	20/30
TTTTTTTTTTAGACCCAGCTACGCAAATC	20/34
TTTTTTTTTTGACACGTACTACGTTGTAGC	20/38
TTTTTTTTTTCCACAACACTTTCTCGGCCT	20/42
TTTTTTTTTTTGTGGGTATTTAGGCTTATG	22/46
TTTTTTTTTTGCAGCCATTCAAGCAATCCTATA	23/50
TTTTTTTTTTTGGTTAGAAGTGAATAAAAGCTAG	25/54
TTTTTTTTTTCCCTCCCACTCCCACTACTAC	20/58
TTTTTTTTTTGGGAATGATGGTTGTCTTTGG	21/62

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



Strips

Roche LINEAR ARRAYS
(a.k.a. reverse dot blots)

Commercial kit so SNP selection and assay design is already done for you...

Mito "Strips"

- Roche Applied Science (Indianapolis, IN) recently released a mtDNA typing kit
- LINEAR ARRAY Mitochondrial DNA HV1/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 HV1/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 HV1/HVII Region-Sequence Typing kit. *Proceedings of the 14th International Symposium on Human Identification.*
- Kline, M.C., *et al.* (2003) Semi-automation of mtDNA arrays: results from 666 population samples and comparisons. *Proceedings of the 14th International Symposium on Human Identification.*

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

A B C D

IA1 IA2 IA3 IB1 IB2 IB3 IC1 IC2 IC3 ID1 ID2

16126 16217 16304 16362
16129 16223 16311

7 SNP sites

HVII

A B C D E

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

73 146 195 247 309.1
152 199

7 SNP sites

23 probes across 9 regions 14 SNPs

European Journal of Human Genetics (1999) 7, 459-468
© 1999 Stocker from All rights reserved 1098-8139/99 \$12.00
http://www.stockerpress.co.uk/ejhg

ARTICLE

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

David Comas¹, Rebecca Reynolds² and Antti Sajantila¹

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

HVII

A B C D E

IA1 IA2 IA3 IB1 IB2 IB3 IC1 IC2 IC3 IC4 IC5 IID1 IID2 IIE2

73 146 150 152 189 195 198 200 247 309.1

10 SNP sites

SSO Probe

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

3,3',5,5'-Tetramethylbenzidine (TMB)

Colorless substrate → Colored Precipitate

Streptavidin-horseradish peroxidase enzyme conjugate

HRP

Streptavidin

Biotin

Biotin-labeled PCR product

Probe

Nylon membrane

16026C

2

3' -CTTATAACATGTCATGGTA-5'

3' -CTTATAACGTGCCATGGTA-5'

3' -CTTATAACATGCCATGGTA-5'

16026 16029

Roche

HVI HVII

16093 IA IC ID IE IIA IIB IIC IID 189

Ref 1 2 3 1 2 3 4 1 2 3 1 2 3 4 5 6 7 1 2 4 5 1 2

10 probe regions

31 probes

CRS

Roche

HVI HVII

16093 IA IC ID IE IIA IIB IIC IID 189

Ref 1 2 3 1 2 3 4 1 2 3 1 2 3 4 5 6 7 1 2 4 5 1 2

40 probe regions

31 probes

CRS

16093 16092

IA 1 IA 2 IA 3

IC 1 IC 2 IC 3 IC w2/w3

ID 1 ID 2

IE 1 IE 2 IE 3

IIA 1 IIA 2

IIB 1 IIB 2 IIB 3 IIB 4 IIB 5 IIB 6 IIB 7

IIC 1 IIC 2 IIC 3 IIC 4

IID 1 IID 2

189 189.1 189.2

18 SNPs

16126 16129 16304 16309 16311 16362 16270 16278

73 146 150 152 189 195 198 200 247

Lessons Learned with LINEAR ARRAYS

- pH is important to wash solutions
 - If above protocol's pH 7.4, then the blue color will not develop correctly and signal will be lost
- Quantification value obtained on Agilent is not equivalent to yield gel information provided by Roche
 - May need to make adjustments—do a sensitivity titration in your lab

NIST U.S. Population Samples

As of 06/2003 **666 males** (anonymous; self-identified ethnicities)

286 Caucasians
252 African Americans
128 Hispanics

Whole blood received from Interstate Blood Bank (Memphis, TN)

Working tubes/plates 1 ng/uL

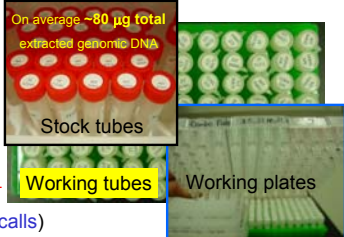
To date: (~50,000 allele calls)

Identifier (15 autosomal markers + Amelogenin) (10,608)

Roche Linear Arrays (HV1/HV2 10 regions) (6,630)

Y STRs 22 loci—27 amplicons (17,388)

Y SNPs 50 markers on sub-set of samples (11,498)



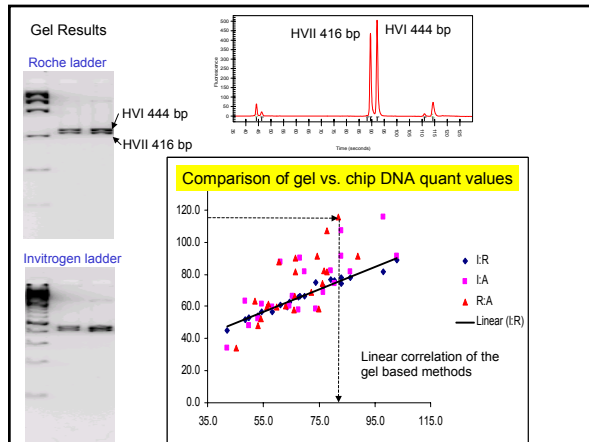
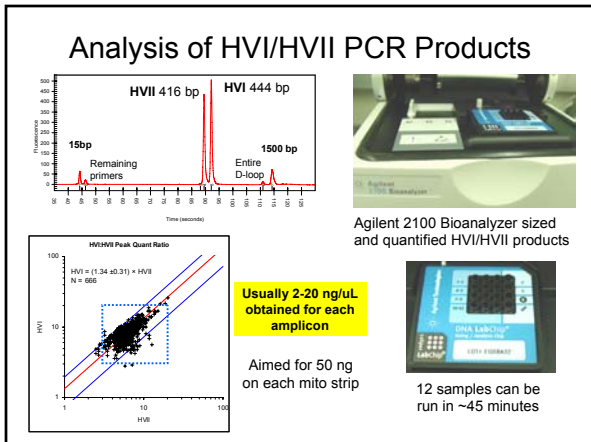
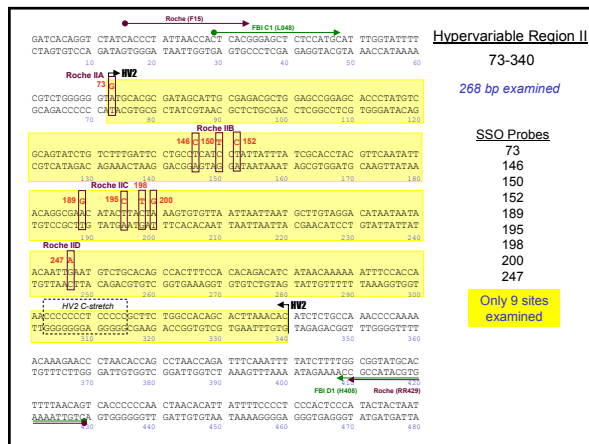
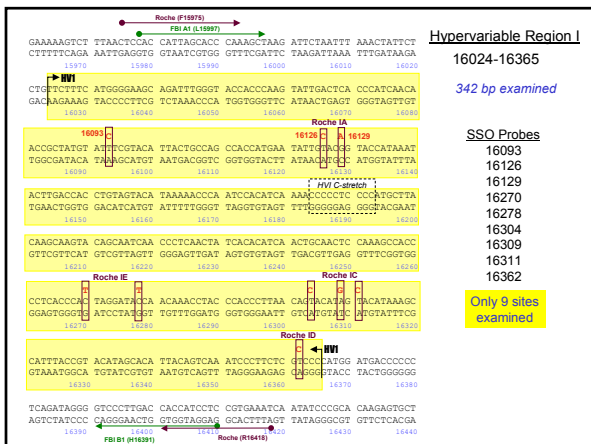
Samples supplied to OhioU for miniSTR typing and AFDIL for whole mtGenome sequencing

PCR Amplification

- Protocol calls for 5 pg input DNA (based on nuclear DNA measurement)
- We used 1 ng DNA instead and reduced PCR cycle number from 34 (protocol) to 28
- Thermal cycling on GeneAmp 9700:
 - 94 °C for 14 minutes
 - 28 cycles: 92 °C for 15s, 59 °C for 30s, and 72 °C for 30s
 - 72 °C for 10 minutes
 - hold at 10 °C
- 50 uL PCR volume (protocol) with duplex amplification

HVI primers amplify a 444 bp PCR product
Forward (L15975-15993) 5'-biotin-CTCCACCATTAGCACCCAA-3'
Reverse (H16418-16401) 5'-biotin-ATTTCACGGAGGATGGT-3'

HVII primers amplify a 416 bp PCR product
Forward (L15-34) 5'-biotin-CACCCATTATAACCACTCACG-3'
Reverse (H429-410) 5'-biotin-CTGTAAAGTCATACC-3'



Analysis of HVI/HVII PCR Products

16189T Good quality sequence
 CACATCAAAGCCCTCCCTATGCTTACAGCAAGCAATCAACCTCAACTAT
 170 180 190 200 210 220

The Agilent electropherogram also gives an indication of the HVI "C-stretch" by the presence of extra peaks (see Butler et al. (1998) *Electrophoresis* 19, 119-124).

16184 **16193** Poor quality sequence following C-stretch (out of phase)
 CACATCAAAGCCCTCCCTATGCTTACAGCAAGCAATCAACCTCAACTAT
 170 180 190 200 210

HVI C-stretch

Automated Washing and Color Development

Tecan Proflibot – processes sample through wash steps

24 strips per run
 ~2 hours per run
 2 or 3 runs easily performed per day

Step	File	Time	Temp	Solution
1	Temp		55 °C	
2	Disp		55 °C	Wash
3	Pause		55 °C	
4	Inc	15 min	55 °C	
5	Asp		55 °C	
6	Disp		55 °C	Wash
7	Asp		55 °C	
8	Disp		55 °C	SA-HRP Conjugate
9	Inc	5 min	55 °C	
10	Asp		55 °C	
11	Disp		55 °C	Wash
12	Asp		55 °C	
13	Disp		55 °C	Wash
14	Inc	12 min	55 °C	
15	Asp		55 °C	
16	Cool		25 °C	
17	Disp		25 °C	Wash
18	Asp		25 °C	
19	Disp		25 °C	Citrate
20	Inc	5 min	25 °C	
21	Asp		25 °C	
22	Disp		25 °C	Color Dev
23	Inc	15 min	25 °C	
24	Asp		25 °C	
25	Disp		25 °C	DI Water
26	Asp		25 °C	
27	Disp		25 °C	DI Water
28	Inc	5 min	25 °C	
29	Asp		25 °C	
30	Disp		25 °C	DI Water
31	End		25 °C	

Digital Recording of LINEAR ARRAYS

UV epi-fluorescence

GeneGnome (Syngene)

Yellow light used with filters so blue color lines could be archived as black and white images

Data Interpretation for LINEAR ARRAYS

Analysis of probe results is still manual!

16093	HVI				HVII			
	A	C	D	E	A	B	C	D
1	3	3	0	w2	2	0	0	2
1	3	3	1	1	2	1	5	1
1	1	0	1	1	2	6	2	1
1	1	1	1	0	2	6	5	1
1	0	1	1	1	2	4	1	1
1	1	1	1	2	2	7	0	1
1	1	0	1	1	2	5	2	1
1	3	3	0	0	w2	3	0	2
1	3	3	2	2	2	7	4	1
1	1	1	1	2	2	0	w2	1
1	1	0	1	1	2	6	2	1
1	1	1	1	0	2	6	5	1

Typing results from 50 ng of each PCR product

"Blank" Calls on LINEAR ARRAYS

We 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%	
HVIAA	3	0.5%	0, 0, 0%
HVIIB	96	14.4%	16, 70, 85%
HVIIC	122	18.3%	11, 47, 13%
HVIID	42	6.3%	5, 5, 18%
189	152	22.8%	

Blanks expected based on full sequence analysis of 1393 individuals

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
11B 1	C	C	C
11B 2	C	C	C
11B 3	C	C	C
11B 4	C	C	C
11B 5	C	C	C
11B 6	C	C	C
11B 7	C	C	C

Nucleotide positions 151 and 153 are common variants in African Americans

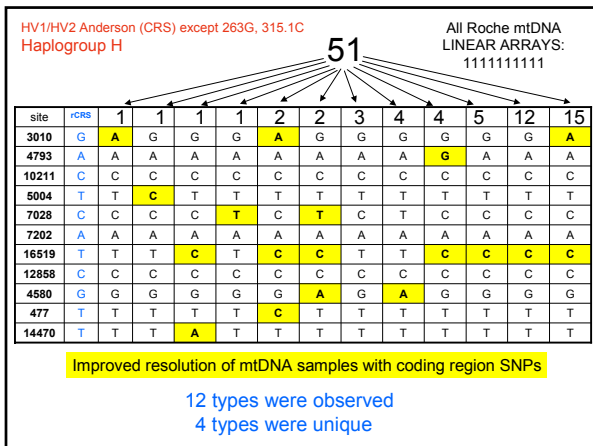
Summary of Our Population Typing with Roche mtDNA LINEAR ARRAYS

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

•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



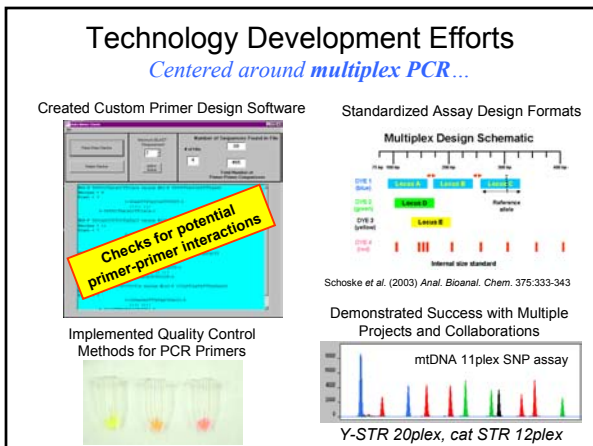
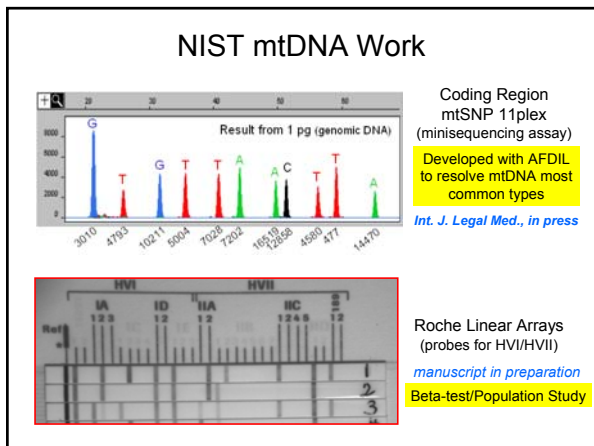
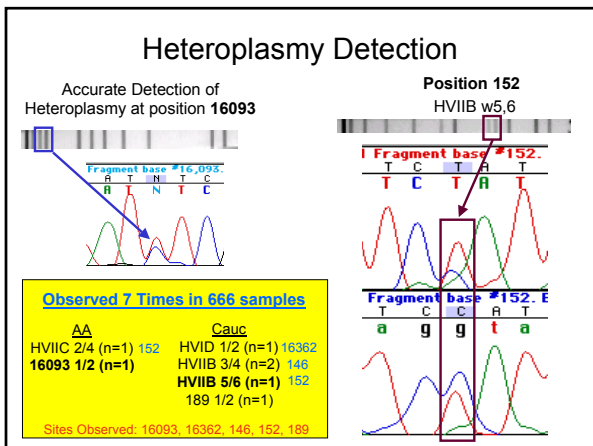
Comparison of Other U.S. Population Data with SSO Probes

Population	N	#types	diversity	Most Common Type	MCT frequency
Caucasian	922	226	0.964	111111111	15.4%
African Am	805	251	0.983	12112021	6.8%
Hispanic	555	170	0.963	12122011	11.7%
Total	2282	502	0.998	111111111	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	11111111111	16.4%
African Am	252	129	0.977	1141224211	10.7%
Hispanic	128	74	0.954	1102120111	16.4%
Total	666	282	0.985	11111111111	7.7%

10 regions, 31 probes, 18 SNPs Kline et al. (2003) NIST population study



Acknowledgments

Funding:
Interagency Agreement between National Institute of Justice and NIST Office of Law Enforcement Standards

NIST Project Team:
John Butler (project leader)
Pete Vallone
Margaret Kline
Jan Redman
Dave Duewer

Collaborators:
Tom Parsons, Mike Coble, and Rebecca Just (AFDIL)
Sandy Calloway (Roche Molecular Systems) for letting us be part of the LINEAR ARRAY beta-testing

Lois Tully (NIJ) for encouragement and support

This presentation available as pdf file from
<http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm>