Analyzing Single Nucleotide Polymorphisms

Peter M. Vallone
NIST
Biotechnology Division
February 11, 2002

National Institute of Standards and Technology

Forensic Mitochondrial DNA Analysis: A Community Forum
2002 AAFS Annual Meeting Atlanta, GA

Overview

- SNPs
  - Definition
  - mtSNP systems of interest
- Instrumentation for SNP detection
  - Capillary Electrophoresis
  - TaqMan
  - Luminex
  - MALDI-TOF MS
- Assay Design
  - SNaPshot primer extension
  - Multiplex PCR
  - SNP Primer Design
- Typing Results
  - Coding and Control Region SNPs
  - MALDI

What is a Single Nucleotide Polymorphism?

- A single nucleotide polymorphism (SNP) is a single base variation in an otherwise conserved region of DNA
- SNPs are the most common type of DNA sequence variation and occurs in ~ 1 of every 1000 bases in the human genome
- An SNP can be an insertion, deletion, or sequence variation

```
–TCTCATAATAGGATAAAACAC–
–AGAGTATTATCCTATTTTGTG–

–TCTCATAATACGATAAAACAC–
–AGAGTATTATGCTATTTTGTG–
```

A G/C transversion highlighted in red

Mitochondrial SNPs

Human identification

Control Region/D-loop highly polymorphic

Use of coding region mtSNPs

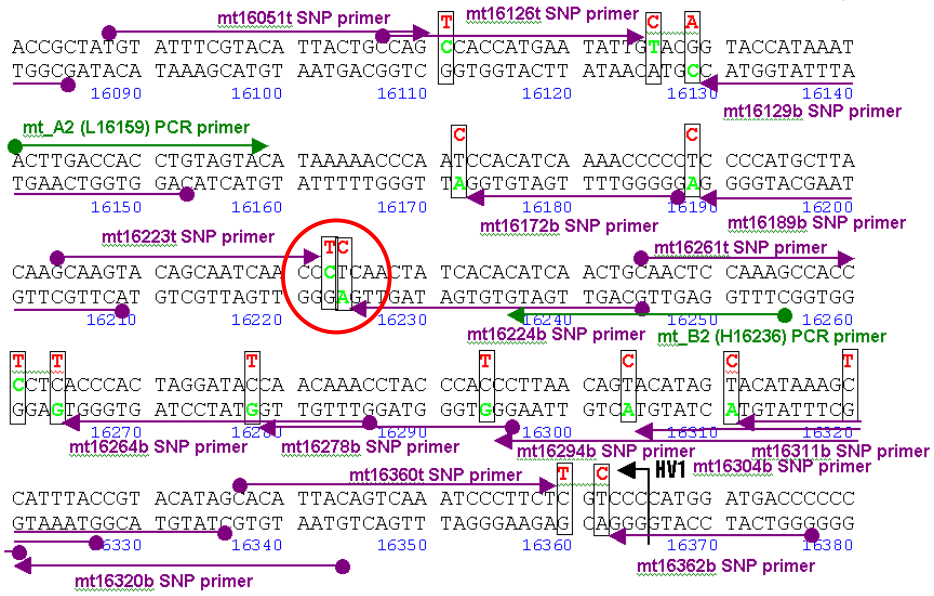
Collaboration with the FBI and AFDIL to design assays for candidate mtSNP markers

Assay design challenges: high GC content, insertions/deletions, closely spaced SNP sites, multiplexing

Recent Efforts with mtSNPs

- Mark Wilson (FBI Laboratory) provided us with list of informative SNPs across entire control region
- SNP extension primers designed to target most informative sites in HV1
- Multiplexed probing of 10 sites in the control region
- We are testing the limitations of probing closely spaced SNP sites with primer extension assays

Mitochondrial SNPs



Recent Efforts with mtSNPs

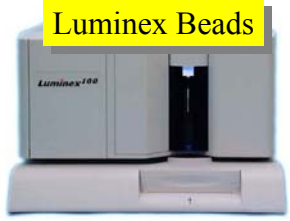
- Tom Parsons (AFDIL) provided us with 29 sequence variation sites located in the coding region
- PCR and SNP extension primers were designed to target these 29 sites
- SNaPshot assay for multiplex amplification of 10 sites is being developed
- Typing is being performed by MS and CE methods

Instrumentation for SNP Assays

SNaPshot

Multi-Color Capillary Electrophoresis (ABI 310 or 3100)

Luminex Beads



Luminex 100 Flow Cytometer

Primer Extension

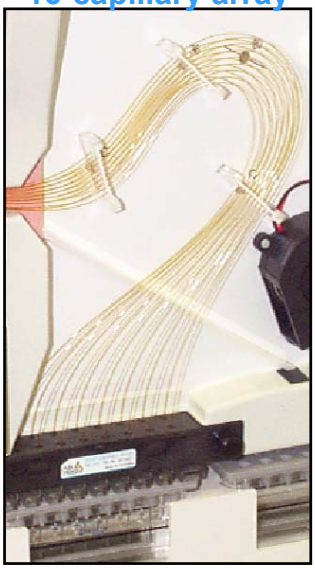
Time-of-Flight Mass Spectrometer

TaqMan

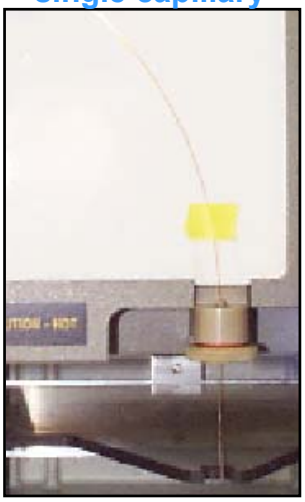
Roche LightCycler

Capillary Electrophoresis Instrumentation

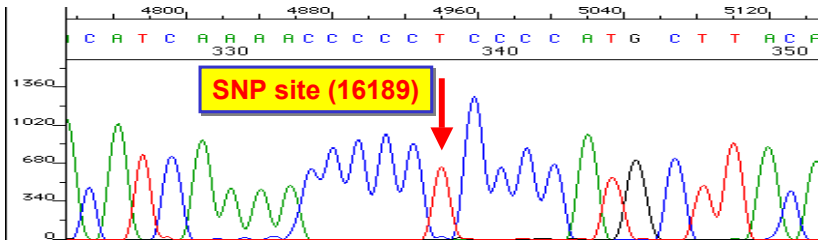
ABI 3100
16-capillary array



ABI 310
single capillary



mtDNA Sequence Analysis



Positions 16176-16201 (from HV1)

Generate mtDNA sequence across HV1 and HV2 (~600 bases)

Rapid screening method could help exclude samples that do not match and enable more samples to be examined

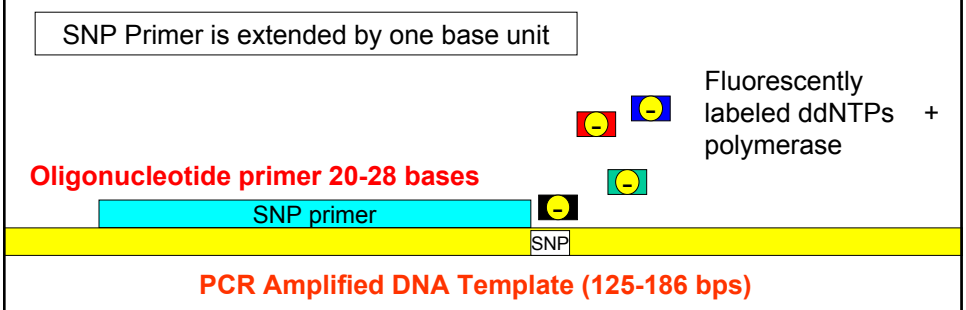
We must identify and target most informative sites

ABI PRISM® SNaPshot™ Multiplex System

- Primer extension assay that utilizes fluorescently labeled ddNTPs
- Analysis of fragment size and fluorescent label identity by capillary electrophoresis allow genotyping of multiple SNP sites
- Multiplexed amplicons or pooled singleplex PCR amplicons can be used as templates
- Kit contains polymerase, Fl-ddNTPs, buffer

-you provide the sites and primers (design/QC)

Genotyping SNPs with SNaPshot™



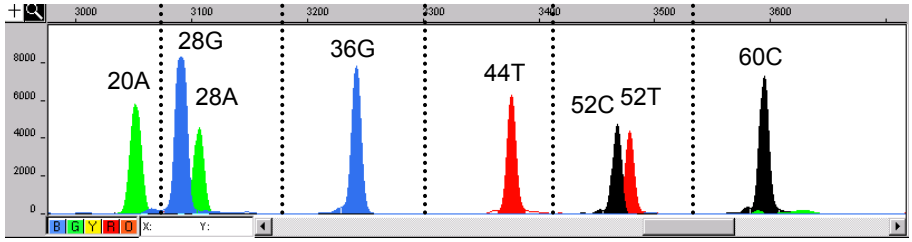
ddNTP	Dye label	Color
A	dR6G	Green
C	dTAMRA	Black
G	dR110	Blue
T	dROX	Red

25 Cycles
96°C 10s
50°C 5s
60°C 30s

ABI SNaPshot Assay

- Typical reaction protocol
  - 3uL of PCR amplicon(s) (ExoSAP treated)
  - 5uL of SNaPshot reaction mix
  - 1uL of SNaPshot primer(s) (0.5 – 2.0 μM)
  - 1uL water
- Thermal cycling (~1 hour)
- SAP treat sample for 1 hour
- Separate and detect on CE system 310/3100/3700
  - LIZ-120 sizing standard
  - 5 dye chemistry, POP4, ~24 min per run
- GeneScan – Genotyper - GeneMapper

Detection of SNPs with ABI 310/3100



SNaPshot™ CEPH Control Reaction

20 nucleotides — ddA

36 nucleotides — ddG

44 nucleotides — ddT

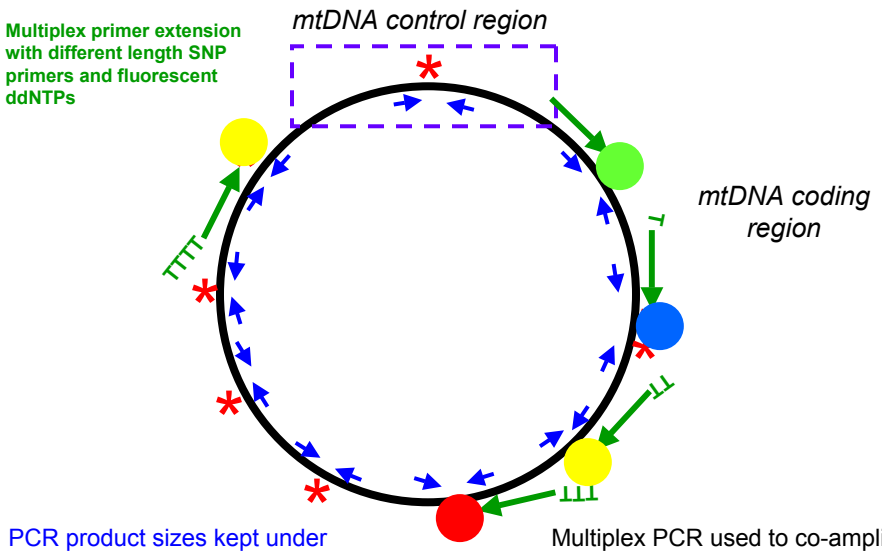
60 nucleotides — ddC

Multiplexing possible by use of different length primers

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

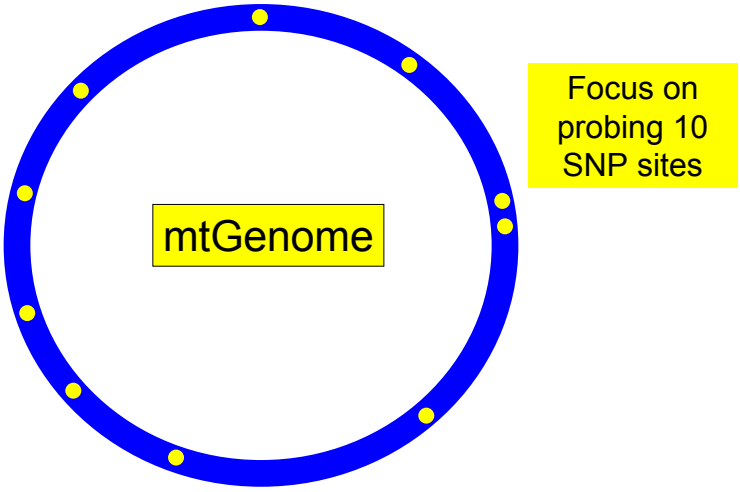
Priming site

Coding Region mtSNP 10plex Assay



Multiplex PCR Design

Identify SNP markers of interest
(collaborations, literature, research)

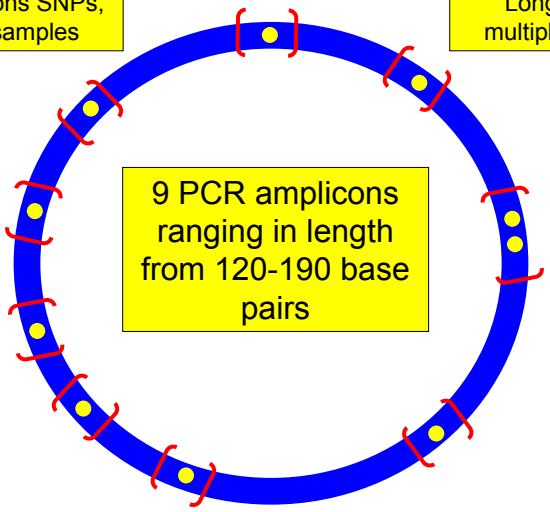


Multiplex PCR Design

Estimate necessary amplicon size

Short amplicons SNPs, degraded samples

Longer amplicons multiplex STR assays



Multiplex PCR Design

Select singleplex PCR primers for each of the 9 amplicons using Primer 3 software

```
OLIGO            start  len      tm      gc%    any    3'  seq
LEFT PRIMER         27   20   60.06    50.00   4.00   2.00 GGGATAACAGCGCAATCCTA
RIGHT PRIMER       174   22   60.31    50.00   8.00   3.00 CGGTCTGAACTCAGATCACGTA
SEQUENCE SIZE: 205
INCLUDED REGION SIZE: 205

PRODUCT SIZE: 148, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00
EXCLUDED REGIONS (start, len)*: 70,65

    1 CTTGACCAACGGAACAAGTTACCCTAGGGATAACAGCGCAATCCTATTCTAGAGTCCATA
                                >>>>>>>>>>>>>>>>>>>>

   61 TCAACAATAGGGTTTACGACCTCGATGTTGGATCAGGACATCCCRATGGTGCAGCCGCTA
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

  121 TTAAAGGTTCGTTTGTTCAACGATTAAAGTCCTACGTGATCTGAGTTCAGACCGGAGTAA
      XXXXXXXXXXXXXX                  <<<<<<<<<<<<<<<<<<<<<<<<

  181 TCCAGGTCGGTTTCTATCTACCTTC
```

Multiplex PCR Design

Checking for significant primer dimer interactions between ALL the primers in the multiplex is VERY IMPORTANT

Screening for potential intermolecular primer-dimer formation

Multiplex PCR Design

- Obtain primer pairs and test singleplex PCR
  QC all primers with MS/CE/HPLC
- Begin initial testing of multiplex PCR
  Start with a PCR mix containing
  1.0 μM of each primer pair
- Evaluate amplicon yields, presence and balance
- Vary primer pair concentrations, increase amount of polymerase, number of cycles
- Redesign and retest failing loci

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

Sequences for 10 SNP primers

TCAGAAGTGAAAGGGGGC	18/na
TTTTTTTGTTGGATCAGGACATCCC	19/26
TTTTTTTTTTACTAAGAAGAATTTTATGGA	20/30
TTTTTTTTTTTTTTAGACCCAGCTACGCAAAATC	20/34
TTTTTTTTTTTTTTTTTTGACACGTACTACGTTGTAGC	20/38
TTTTTTTTTTTTTTTTTTTTTTCCACAACACTTTCTCGGCCT	20/42
TTTTTTTTTTTTTTTTTTTTTTTTTGTGGGCTATTTAGGCTTTATG	22/46
TTTTTTTTTTTTTTTTTTTTTTTTTTGCAGCCATTCAAGCAATCCTATA	23/50
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTTAGAACTGGAATAAAAGCTAG	25/54
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAACCATAACCAATACTACCAATCA	25/58

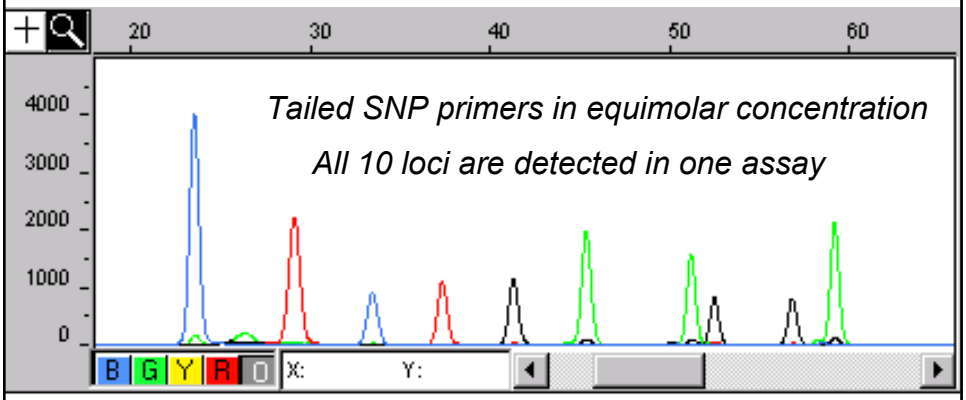
Template binding sequence – black
Tailed sequence for fragment separation - red

SNP Primer Program Output

Label	Length	Sequence	Position	Tm
Forward Primers Salt = 0.3Ct = 10				
M42 340 bp (A/T 297 W) AC010889	18	ATTTAGGACACAAAAGCW	280	60.65398
M42 340 bp (A/T 297 W) AC010889	19	GATTTAGGACACAAAAGCW	279	61.96716
M42 340 bp (A/T 297 W) AC010889	20	AGATTTAGGACACAAAAGCW	278	63.67808
Reverse Primers				
M42 340 bp (A/T 297 W) AC010889	23	GCTCTCTTTTTCATTATGTAGTW	319	63.5462
M42 340 bp (A/T 297 W) AC010889	21	TCTCTTTTTCATTATGTAGTW	317	59.28964
M42 340 bp (A/T 297 W) AC010889	20	CTCTTTTTCATTATGTAGTW	316	57.50257

Hairpin	Dimer	Template	Mass	Rank	Mutation	+ddC	+ddT	+ddA	+ddG
4	8	10	5273.48	2.133333	W	N/A	5561.67998	5570.68998	N/A
5	10	10	5602.69	2	W	N/A	5890.889941	5899.899941	N/A
5	10	11	5915.9	2	W	N/A	6204.099902	6213.109902	N/A
4	8	22	6734.42	2.133333	W	N/A	7022.619922	7031.629922	N/A
4	8	20	6116.02	2.133333	W	N/A	6404.22002	6413.23002	N/A
4	8	19	5811.82	2.133333	W	N/A	6100.019824	6109.029824	N/A

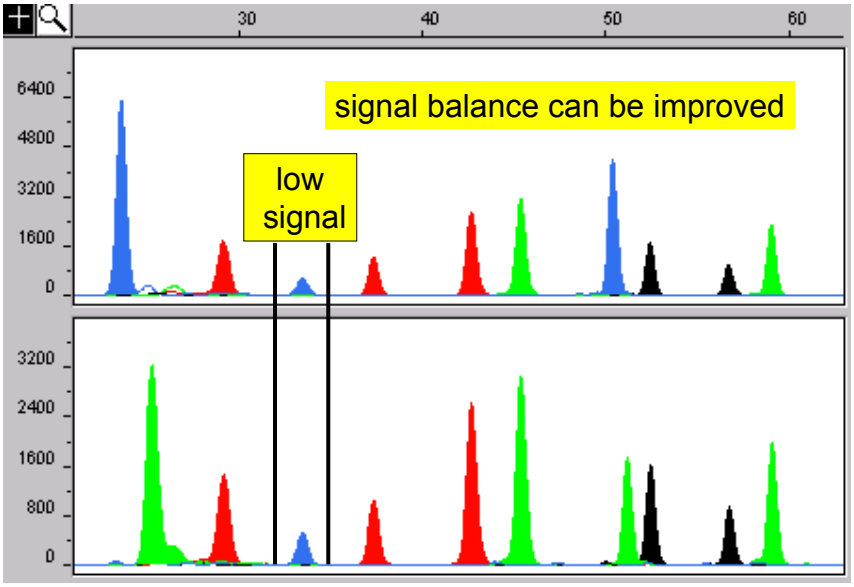
SNaPshot Results Using *pooled PCR products*



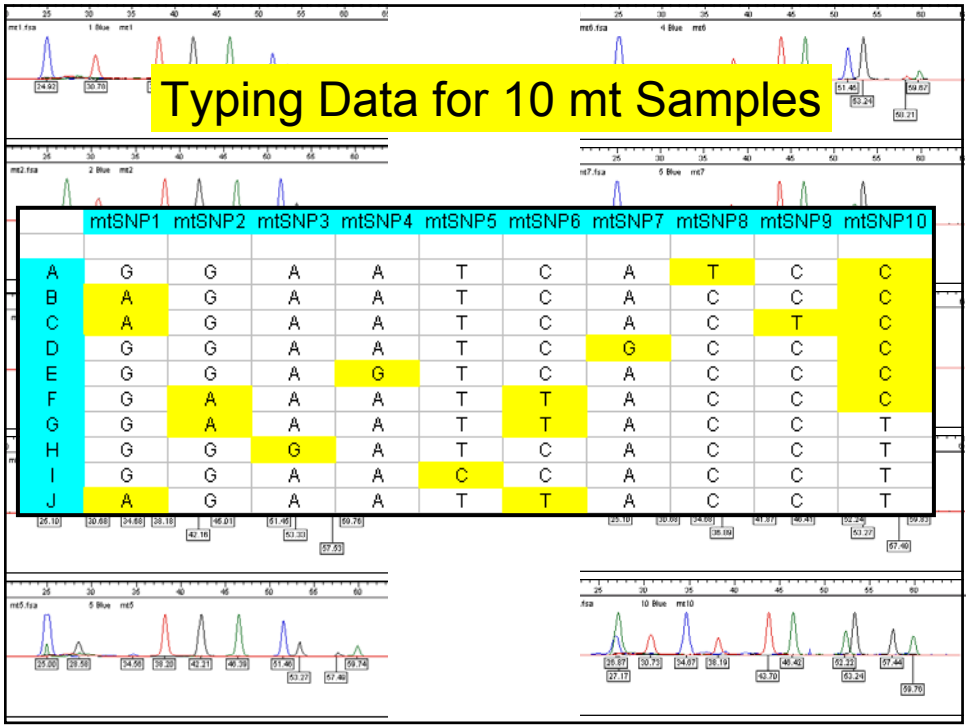
9 PCR amplicons generated in singleplex were combined and used as templates in a SNaPshot mtSNP 10plex reaction

Coding region mtSNPs

SNaPshot results using *multiplex PCR products*



Coding region mtSNPs



	mtSNP1	mtSNP2	mtSNP3	mtSNP4	mtSNP5	mtSNP6	mtSNP7	mtSNP8	mtSNP9	mtSNP10
A	G	G	A	A	T	C	A	T	C	C
B	A	G	A	A	T	C	A	C	C	C
C	A	G	A	A	T	C	A	C	T	C
D	G	G	A	A	T	C	G	C	C	C
E	G	G	A	G	T	C	A	C	C	C
F	G	A	A	A	T	T	A	C	C	C
G	G	A	A	A	T	T	A	C	C	T
H	G	G	G	A	T	C	A	C	C	T
I	G	G	A	A	C	C	A	C	C	T
J	A	G	A	A	T	T	A	C	C	T

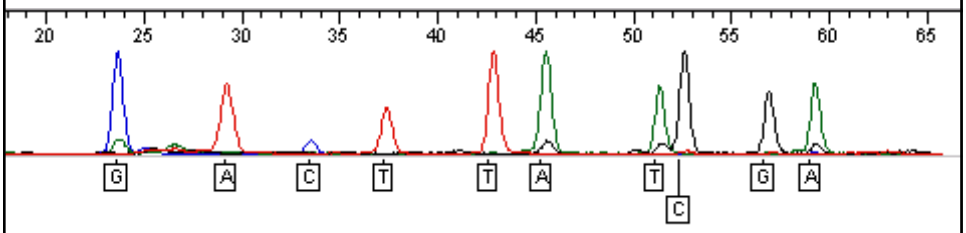
Sizing of Fragments in mtSNP 10plex

Actual versus observed

Actual length (bases)	alelle 1	allele 2	Δ allele1	Δ allele 2
18	25.0	27.1	-7.0	-9.1
26	28.6	30.7	-2.6	-4.7
30	34.7	35.6	-4.7	-5.6
34	36.9	38.2	-2.9	-4.2
38	42.2	43.7	-4.2	-5.7
42	45.0	46.4	-3.0	-4.4
46	51.4	52.2	-5.4	-6.2
50	53.3	54.2	-3.3	-4.2
54	57.5	58.3	-3.5	-4.3
58	59.2	59.7	-1.2	-1.7

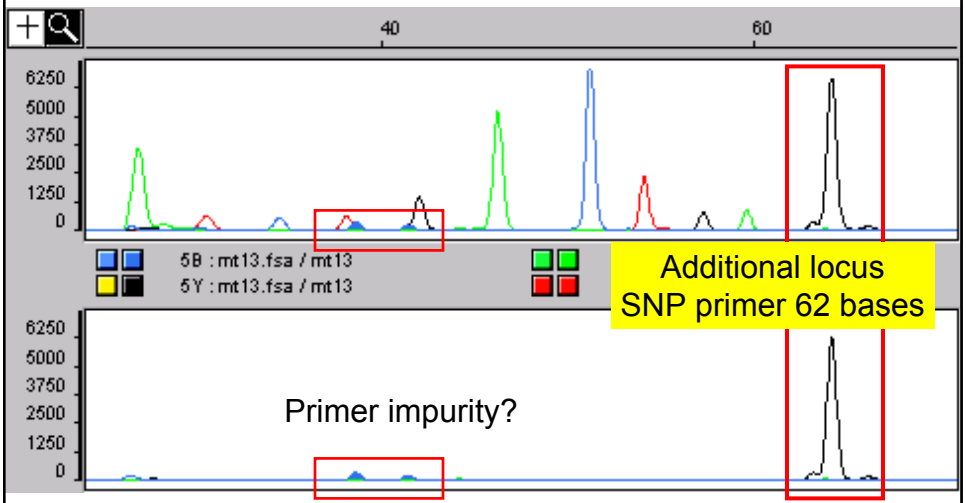
Sizing differences vary with sequence, length and fluorescent dye attachment

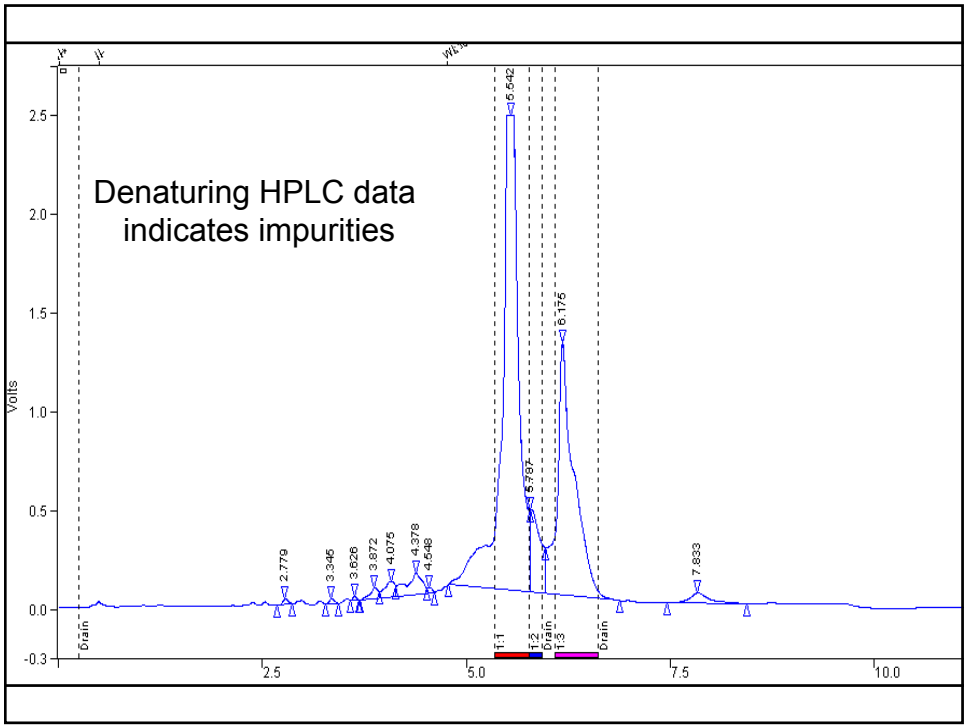
Genotyper Macro for SNaPshot



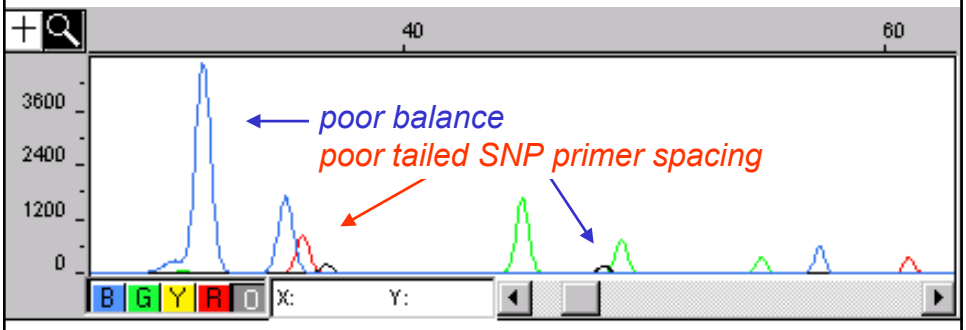
Beginning work on developing macros for automated genotyping of mtSNP assays

Additional Loci to the 10plex





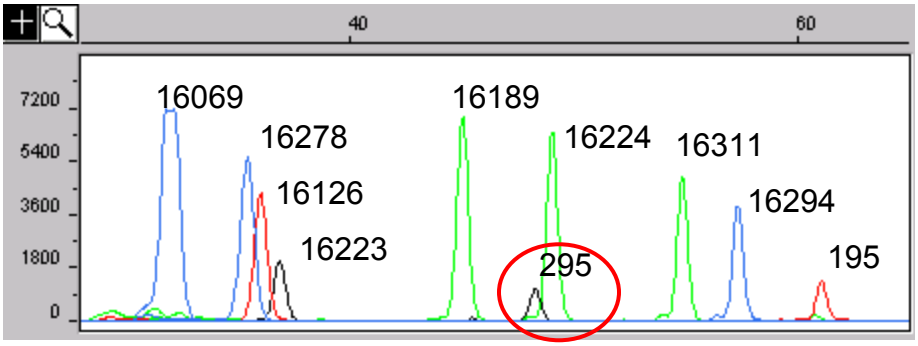
Preliminary SNaPshot Results for Multiplex Typing of Control Region mtSNPs



A single PCR amplicon was used as a template in a SNaPshot Control Region mtSNP 10plex reaction

Control region PCR amplicon 1304bp

Control region mtSNP singleplex overlay



mt16069	22	**mt16224**	**42**
mt16278	26	**mt295**	**46**
mt16126	30	mt16311	50
mt16223	34	mt16294	54
mt16189	38	mt195	58

Control region PCR amplicon 1304bp

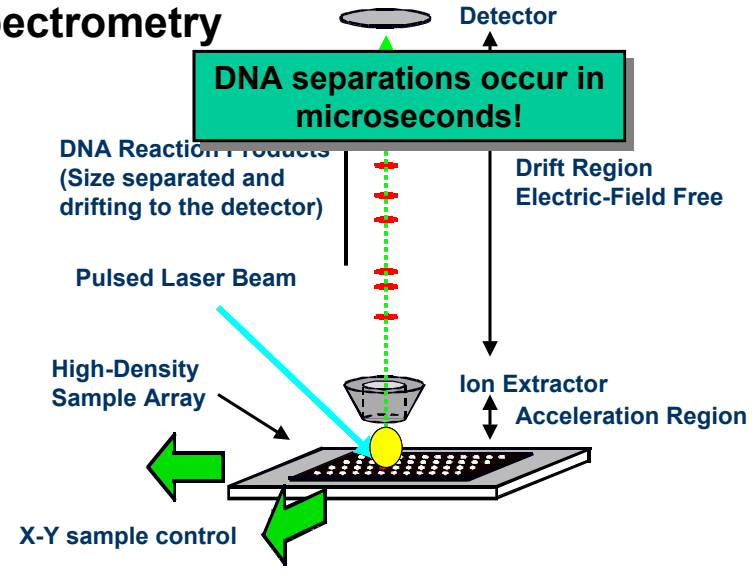
Future Directions with SNaPshot Assay

- Improve signal balance for coding region 11plex
- Adjust primer lengths for control region 10plex
- Develop macro for typing mtSNP assays
- Gain an understanding of how sequence, tails, and fluorescent dye labels effect electrophoretic mobility
- Sensitivity of assay with degraded/low template copies

The use of Mass Spectrometry for SNP Genotyping

- The speed of the MALDI TOF MS technique makes it a good candidate for quickly genotyping a large number of samples for a few (less than 10) SNP markers
- Sample preparation, data collection, and data analysis are amenable to automation

Time-of-Flight Mass Spectrometry

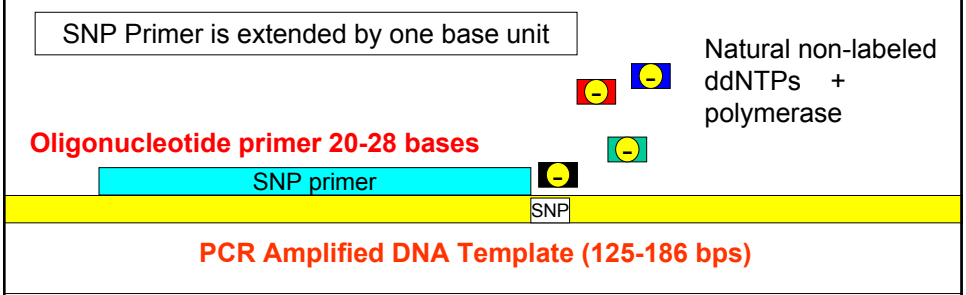


The Principle of MALDI TOF MS

Matrix Assisted Laser Desorption Ionization

- Sample is combined with a matrix and allowed to dry
- Crystalline sample is irradiated by a short pulse laser
- The beam volatilizes the sample, producing molecular ions
- These ions are accelerated by a strong electric field and directed toward the detector
- Ions of different m/z are separated and their flight times are converted to mass
- The resulting mass spectrum yields useful structural and chemical information

Genotyping SNPs with Primer Extension for MALDI Analysis



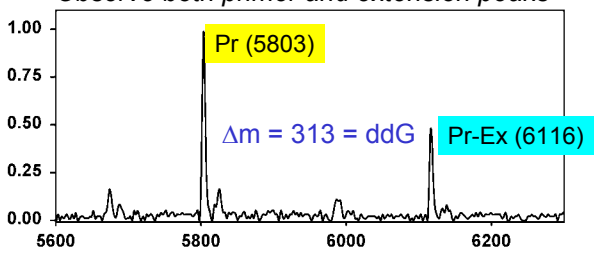
ddNTP	Mass (Da)
A	297
C	273
G	313
T	288

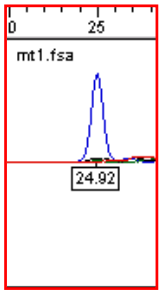
45 Cycles
96°C 10s
50°C 20s
72°C 30s

Mass difference between SNP primer and single base extension product provides genotype

Typing mtSNP1 by Mass Spectrometry

Observe both primer and extension peaks



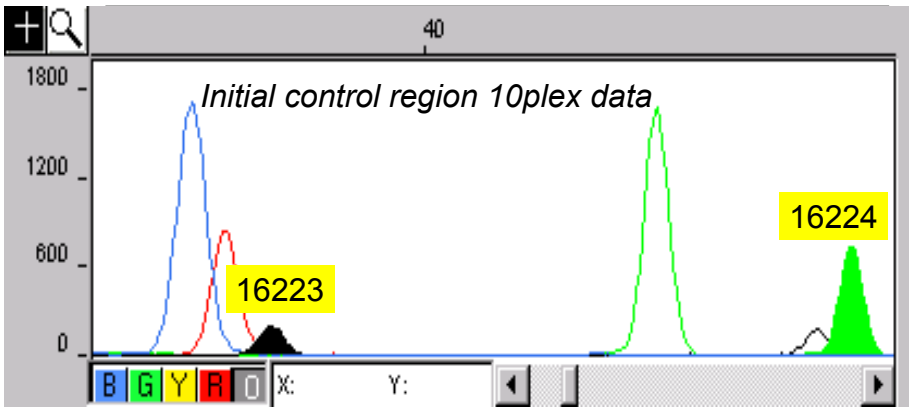


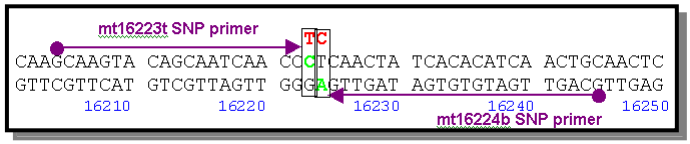
	mtSNP1 SNaPshot	mtSNP1 MALDI
A	G	G
B	A	A
C	A	A
D	G	G
E	G	G
F	G	G
G	G	G
H	G	G
I	G	G
J	A	A

Genotypes determined by SNaPshot and Mass Spectrometry agree

Approximately 20 sec per SNP by MALDI

Two Adjacent Mitochondrial SNPs 16223 (C/T) and 16224 (A/G)





Future Directions

- Continue evaluation of assays and techniques for mtSNP and Y chromosome SNP analysis
- Continue developing quality control methods for synthetic oligonucleotides
- Further development of software for designing multiplex assays – increasing number of loci 15plex, 20plex, etc
- Collaborations

Acknowledgments

Funding:

National Institute of Justice Grant #97-LB-VX-0003

John Butler (NIST)
Thomas Parsons (AFDIL)
Mike Coble (AFDIL)
Mark Wilson (FBI)

petev@nist.gov