DNA Mass Spectrometry at NIST

Peter M. Vallone
Biotechnology Division, NIST
Bruker Daltonics Annual User Meeting
May 27, 2001

NIST
National Institute of Standards and Technology

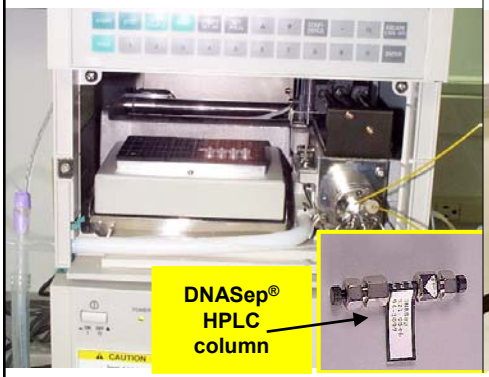
National Institute of Standards and Technology (NIST)

- Biotechnology Division: 5 groups
- DNA Technologies Group
  - Standard Reference Materials for Forensic DNA Labs
  - Quality Control Testing/Interlaboratory Tests
  - Multiplex PCR for Human Identity Testing
  - Development/Evaluation of New Technologies

Presentation Overview

- Methodologies for quality control testing DNA oligomers
  - Multiplex PCR primer mixtures
  - Sequencing short oligonucleotides
- Genotyping single nucleotide polymorphisms (SNPs)
  - Y chromosome SNPs
  - mitochondrial SNPs
- Evaluating genotyping chemistries
  - Primer extension
  - GOOD assay
  - Cleavable primer

Transgenomic SystemWAVE® DNA Fragment Analysis System

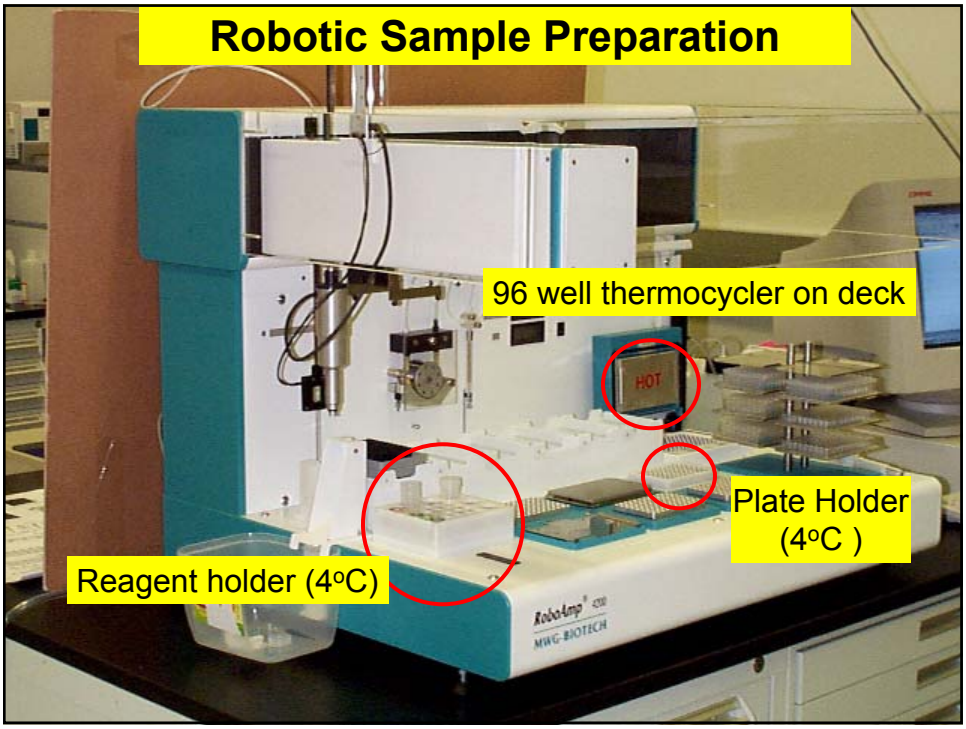


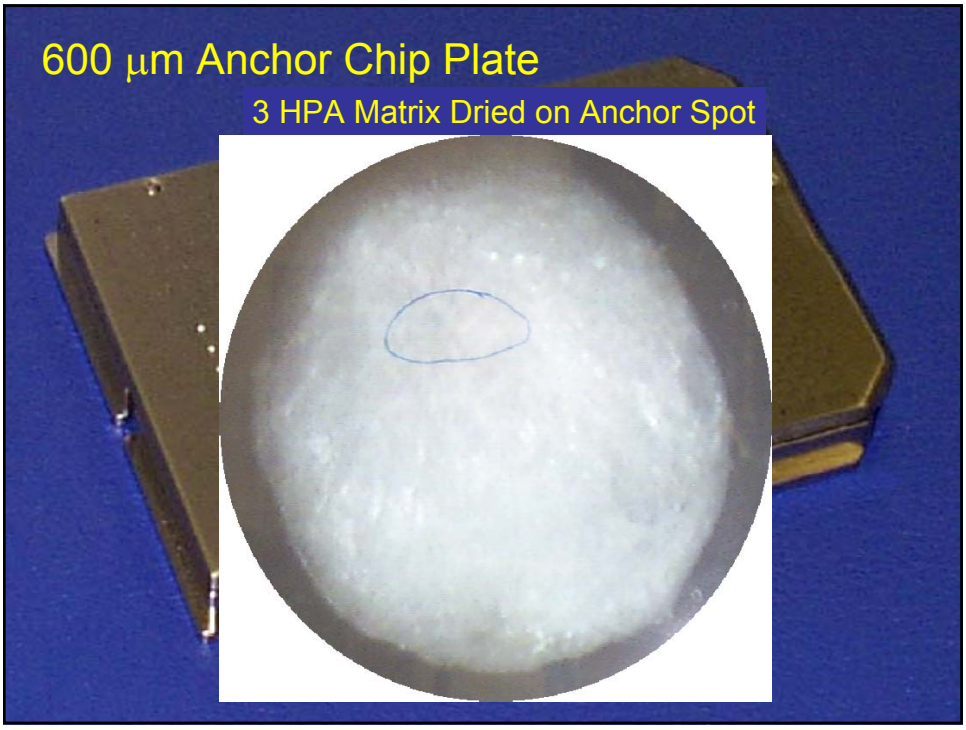
Bruker BIFLEX III MALDI-TOF MS

Instrumentation

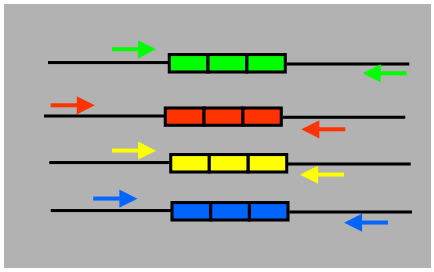
•Bruker BIFLEX III MALDI-TOF mass spectrometer (Bremen, Germany) capable of operation in both linear and reflector mode.

•Reverse phase-ion pairing HPLC (Transgenomic, San Jose, CA) was employed for the separation, concentration, and desalting of mixtures of nucleic acid oligomers.

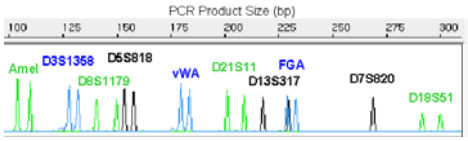

Robotic Sample Preparation
96 well thermocycler on deck
HOT
Plate Holder
(4°C)
Reagent holder (4°C)


600 μm Anchor Chip Plate
3 HPA Matrix Dried on Anchor Spot

Multiplex PCR (Parallel Sample Processing)



- **Successful multiplexing depends on high quality oligonucleotides or "primers"**
- **Complex mixtures of primers need to be quality control tested**

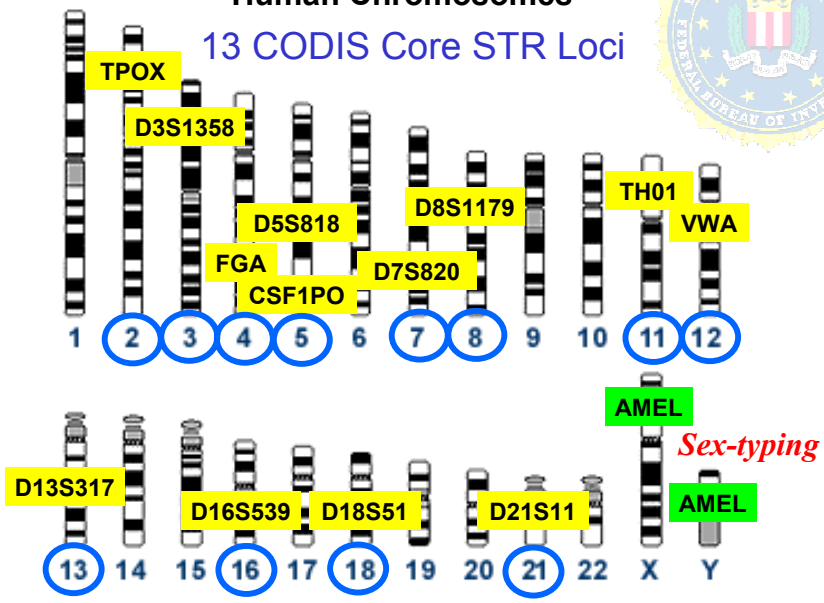


Advantages of Multiplex PCR

–Increases information obtained per unit time (increases power of discrimination)
–Reduces labor to obtain results
–Reduces template required (smaller sample consumed)

Position of Forensic STR Markers on Human Chromosomes

13 CODIS Core STR Loci



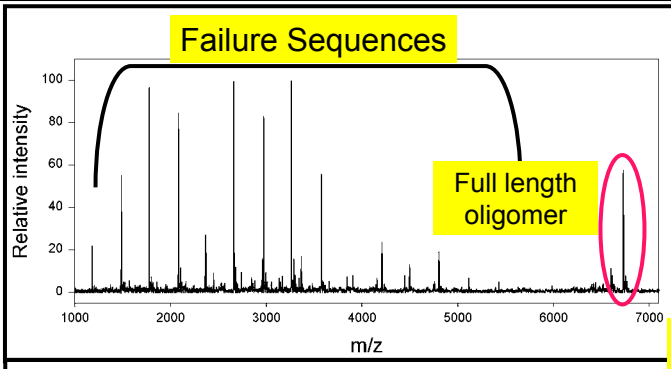
Quality Control Testing of Multiplex PCR Primer Kits

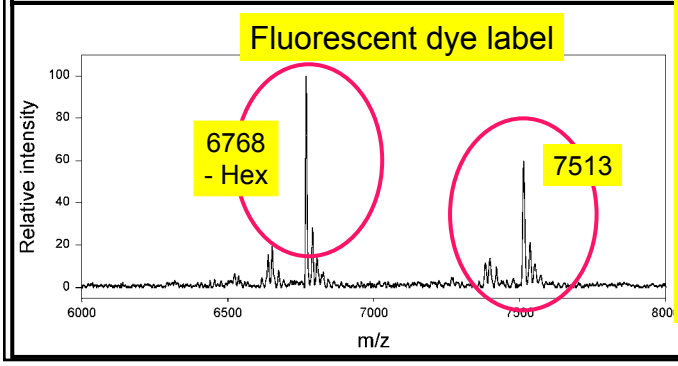
•Reliable amplification of short tandem repeat (STR) DNA markers with the polymerase chain reaction (PCR) is dependent on high quality PCR primers

•Particular primer combinations and concentrations are especially important with multiplex amplification reactions where multiple STR loci are simultaneously copied

•Commercially available kits are now widely used for STR amplification and subsequent DNA typing

•Use of HPLC and MALDI-TOF MS methods for characterization of commercially available STR kits



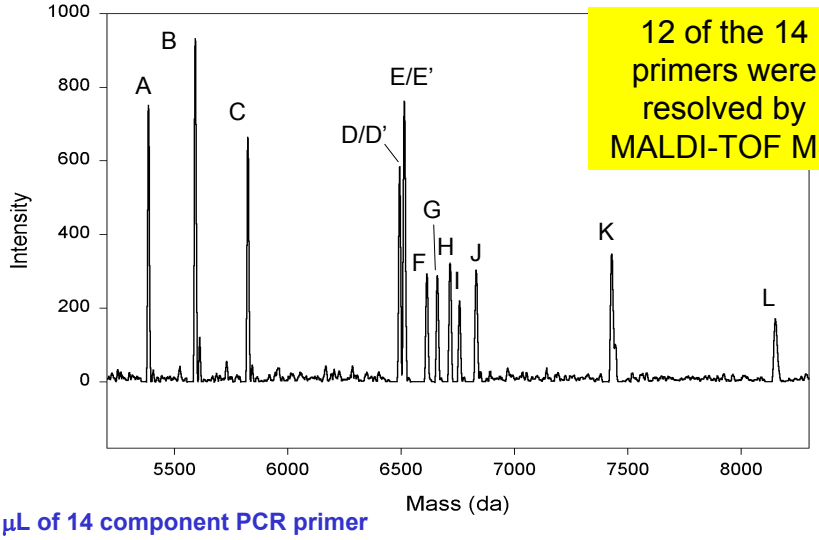


Examples of automated synthesis failure and/or failed purification

•Mass Spectrometry provides an independent means to probe oligonucleotide purity

•Impure Hex labeled primer provides the mass of the Hex dye 7513 – 6768 = 745 Da

Mass Spectrum of Multiplex Set of PCR Primers

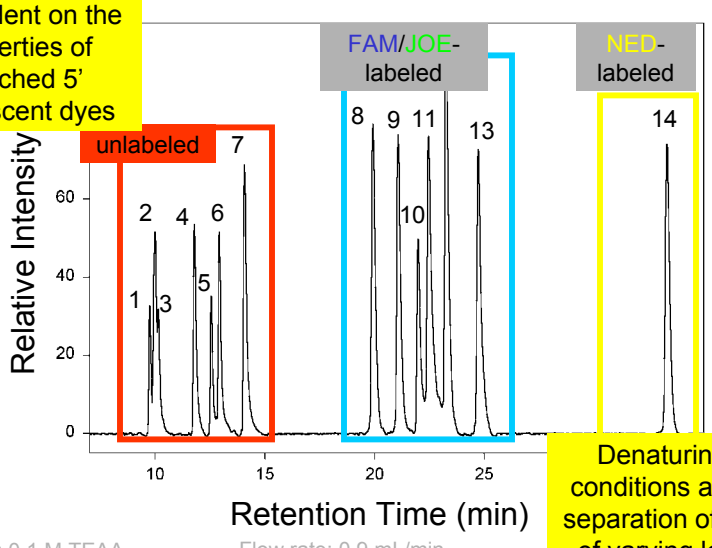


12 of the 14 primers were resolved by MALDI-TOF MS

1 μL of 14 component PCR primer kit spotted in 1 μL 3HPA matrix

HPLC Separation of a Multiplex Set of 14 PCR Primers

Separation dependent on the properties of attached 5' fluorescent dyes



2 μL injected onto DNASep® column run at 70 °C

Solvent A: 0.1 M TEAA

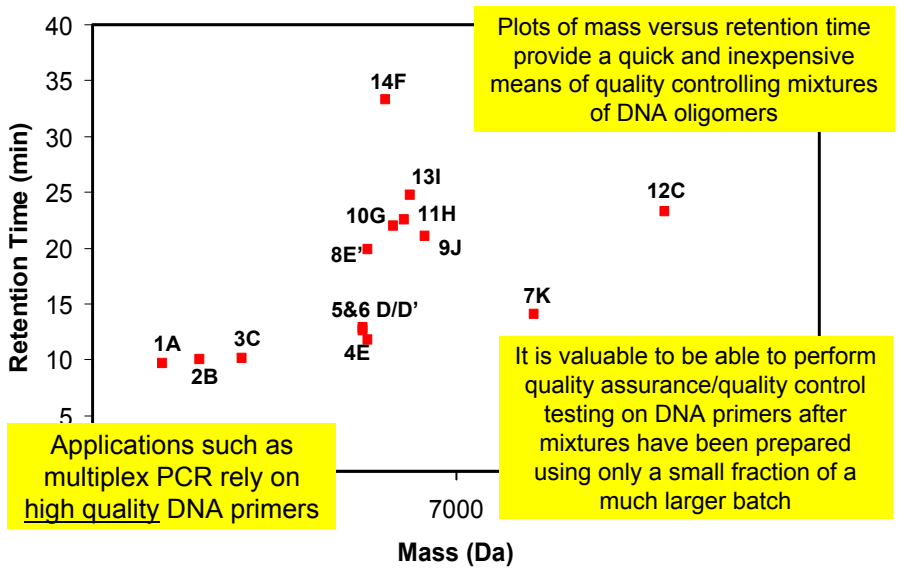
Solvent B: 0.1 M TEAA-25% ACN

Flow rate: 0.9 mL/min

Gradient of 18%-50% B -40 min

Denaturing HPLC conditions allowed the separation of oligomers of varying length and sequence content

Plot of Mass versus HPLC Retention Time



Plots of mass versus retention time provide a quick and inexpensive means of quality controlling mixtures of DNA oligomers

It is valuable to be able to perform quality assurance/quality control testing on DNA primers after mixtures have been prepared using only a small fraction of a much larger batch

Applications such as multiplex PCR rely on high quality DNA primers

General Strategy for Sequencing Short DNA Oligonucleotide Mixtures

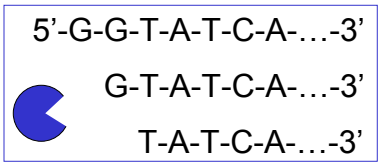
- Oligonucleotides are isolated and collected by HPLC
- The mass of each fraction is measured by MALDI-TOF MS
- Each fraction is subjected to exonuclease degradation and the products are analyzed by MALDI-TOF MS
- Resulting sequence information is tabulated and compared to a reference sequence
- The confirmation of sequence content/context is valuable for quality controlling large oligonucleotide mixtures

Exonuclease Digestions and Mass Analysis

•HPLC isolated DNA oligomers are subject to 5' and 3' acting exonucleases

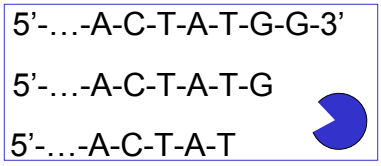
•Digestion reactions are analyzed by MALDI-TOF and the oligomer sequence is determined

•The use of denaturing HPLC allows the sequencing of a relatively complex (up to ~32 components) mixture of DNA oligonucleotides



CSP

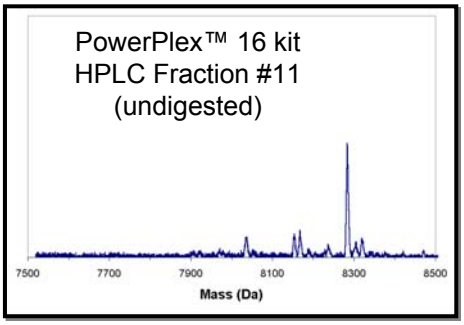
5'-exonuclease **(calf spleen phosphodiesterase)**



SVP

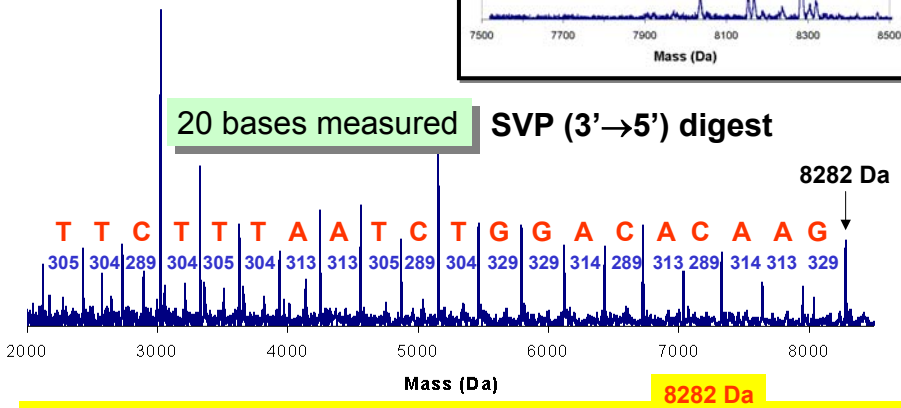
3'-exonuclease **(snake venom phosphodiesterase)**

Results of Sequencing Solving Strategy



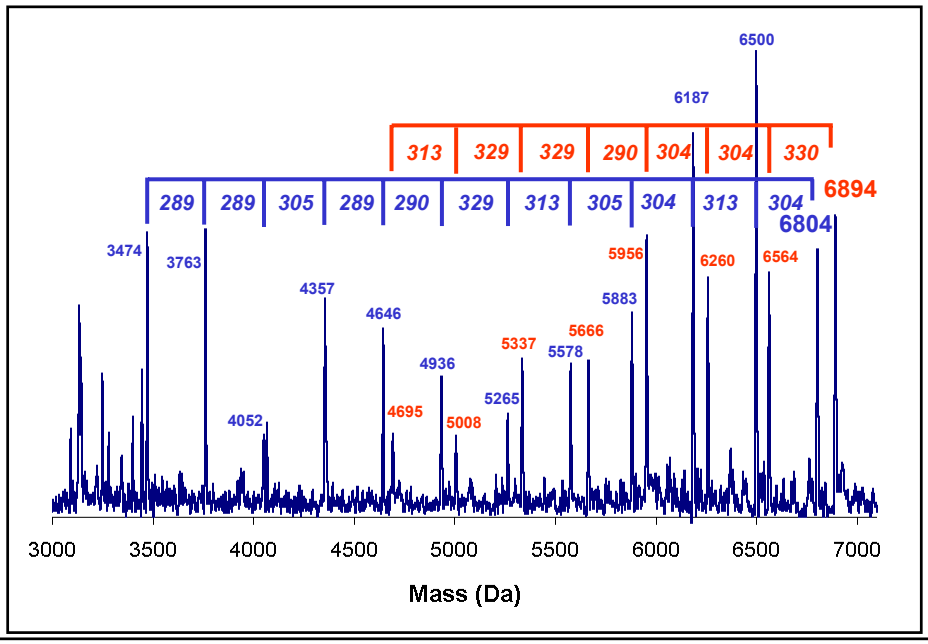
20 bases measured

SVP (3'→5') digest

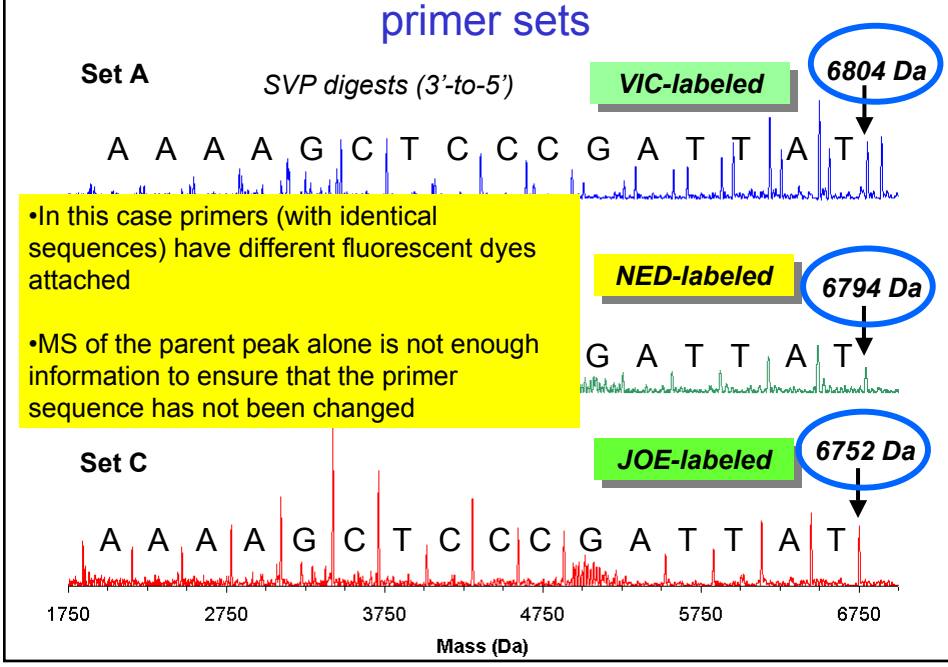


***Sequence received from Promega*: ATTAGAATTCTTTAATCTGGACACAAG**

Simultaneous Digest of 2 Oligonucleotides



Primers for the same locus but from different primer sets



- A MS profile of a multiplex PCR kit allows the unique mass of each primer to be measured
- This is a rapid and inexpensive method of quality controlling the kit by confirming that primers for a specific loci are present
- Sequencing of short DNA oligomers allows the user/community to confirm that primers remain consistent over time
  —Incorporation of a different fluorescent dye
  —Addition of mobility modifying linkers
  —Altering primer binding site

SNP Research Focus

•Evaluate SNP detection assays that can be analyzed by MALDI-TOF MS methods

•Genotype known SNP sites (we are not scanning or validating SNPs)

•Improve multiplex capabilities for MS based SNP detection

•Develop tools/protocols for high throughput genotyping

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 for the M96 marker located on the Y-chromosome

The Significance of SNPs

- A key aspect of research in genetics is associating sequence variations with heritable phenotypes

- Because SNPs are expected to facilitate large-scale association genetics studies, there has recently been great interest in SNP discovery and detection

  - Disease association
  - Genetic mapping
  - Pharmacogenetics
  - Evolutionary studies
  - Human identification
    - –Paternity testing
    - –Forensic testing

3 Steps to Characterizing SNPs

- Discovery
- Validation
- **Genotyping (scoring) samples**

Technologies Used

•Discovery – sequence alignments or DHPLC

•Validation – resequencing or database searches

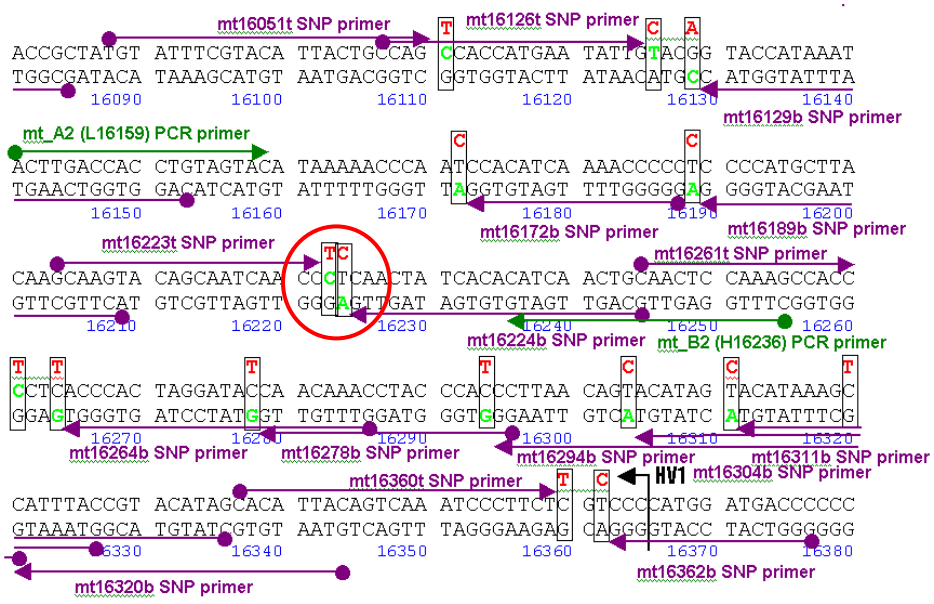
•Genotyping (scoring) samples – variety of technologies (single base extension or hybridization); MALDI-TOF mass spectrometry; capillary electrophoresis

Advantages of Mass Spectrometry for Genotyping SNPs

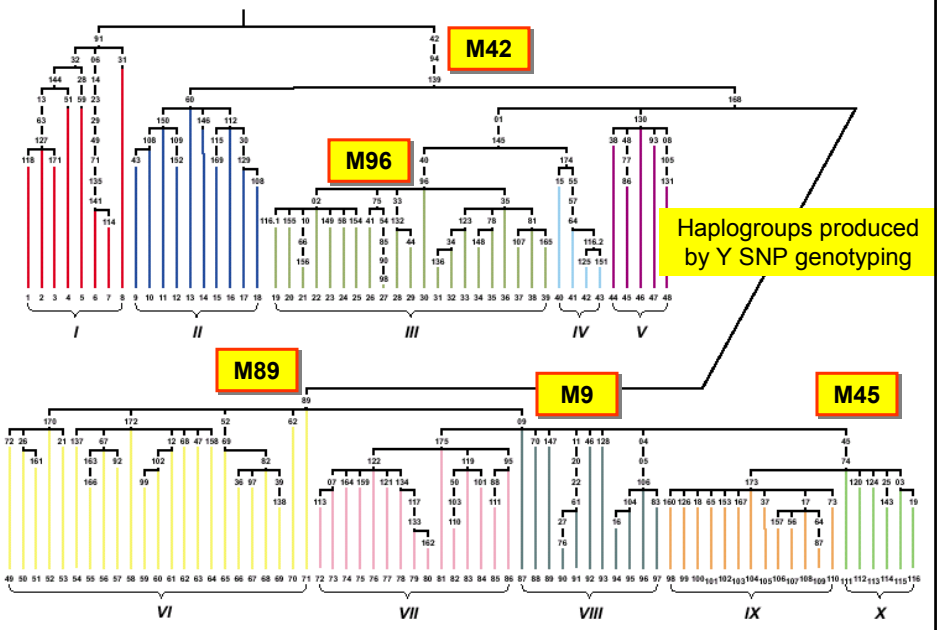
- Accuracy
  - internal standards are not required to correct mobility differences between runs as in electrophoretic systems
- Speed
  - separations may be performed hundreds of times faster than with electrophoretic systems because ions are in the gas phase

Mitochondrial SNPs	Y Chromosome SNPs
•Human identification	•Human identification
•Control Region/D-loop highly polymorphic	•Defines genetic affinities between contemporary global populations
•10plex already demonstrated	•Over 200 SNPs have been discovered
•Collaboration with the FBI to find candidate mtSNP markers	•Initial research already performed for multiplexing 5 Y SNP markers M9, M42, M45, M89, M96
•Assay design challenges: high GC content, insertions/deletions, closely spaced SNP sites	

Mitochondrial SNPs



Y Chromosome SNPs



Underhill *et al.* (2000) *Nature Genetics* 26: 358-361

Primer Extension Assay

- PinPoint Assay
  Haff and Smirnov (1997) *Genome Res* 7:378-388

- A single primer 16-30 (bases) is designed to bind upstream/downstream from a known SNP site

- Mixture of ddNTPs and polymerase extend the primer by a single base

- The sequence identity of the extended base is the complement of the base at the SNP site

Steps in Primer Extension Assay

Select SNP site of interest and obtain DNA sequence

Design PCR primers and SNP extension primer

Perform PCR

Purification of PCR Template

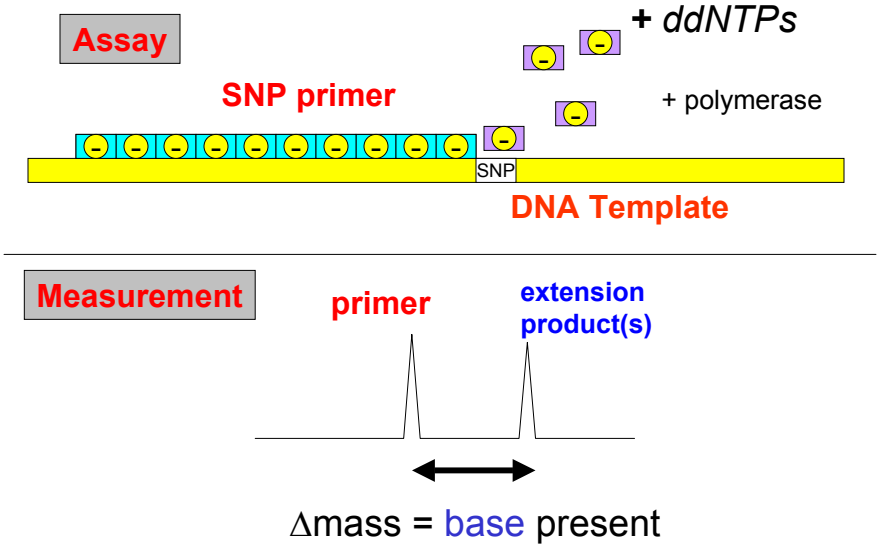
Add SNP reagents and perform single base extension

Purification

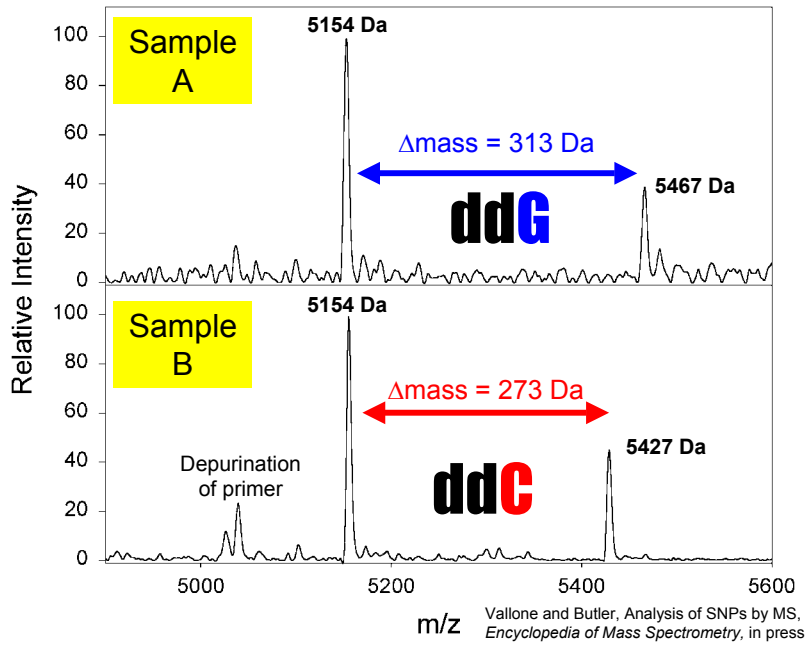
Spot sample and perform mass spec analysis

Genotype based on mass difference measurement

Primer Extension Assay

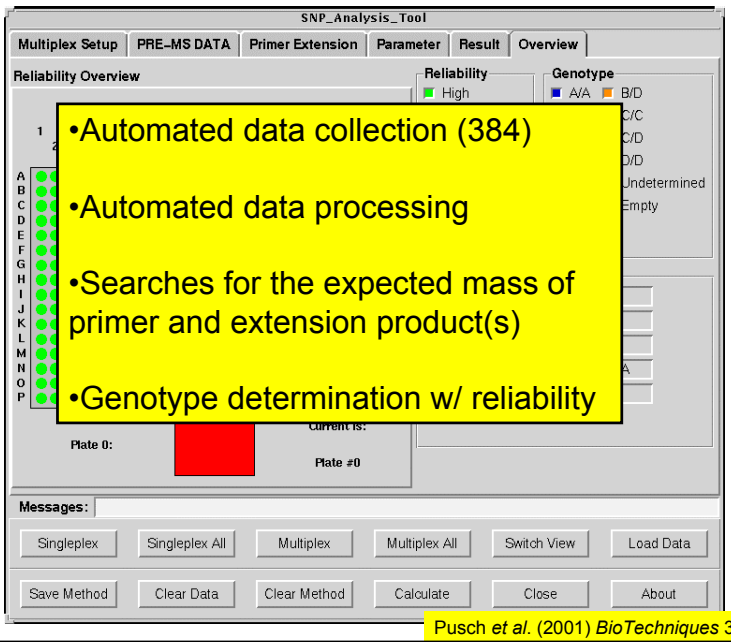


MS Data from Y SNP Marker M96



Vallone and Butler, Analysis of SNPs by MS, *Encyclopedia of Mass Spectrometry*, in press

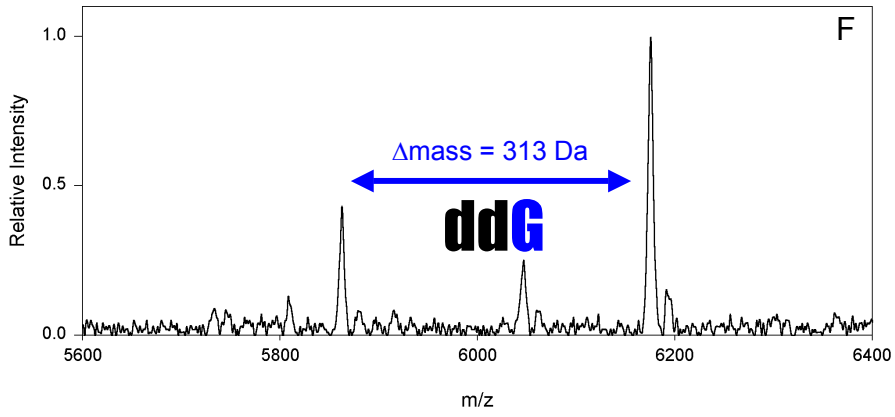
Bruker SNP Manager Genotyping Software



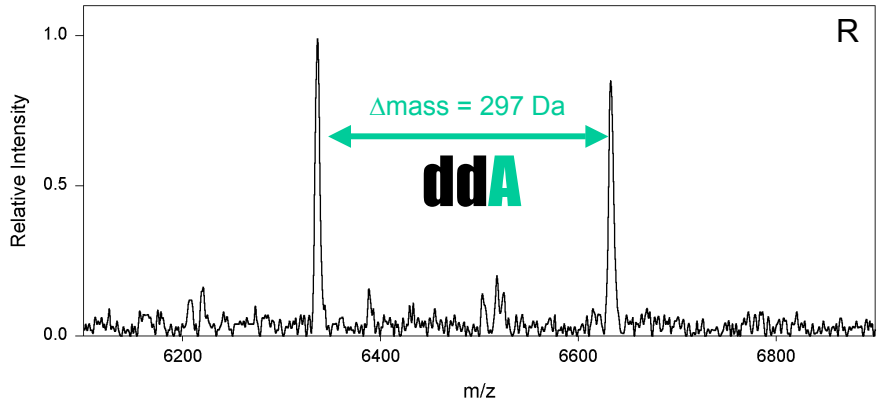
- •Automated data collection (384)
- •Automated data processing
- •Searches for the expected mass of primer and extension product(s)
- •Genotype determination w/ reliability

Pusch *et al.* (2001) *BioTechniques* 30: 210-215

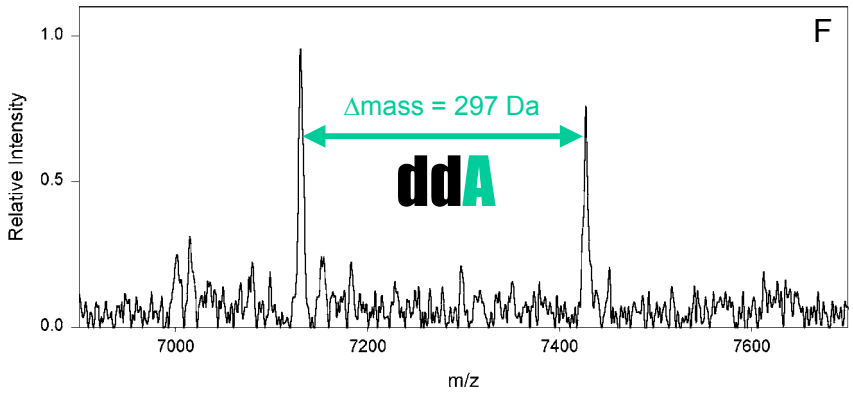
M9 (G/C) Y Chromosome Marker



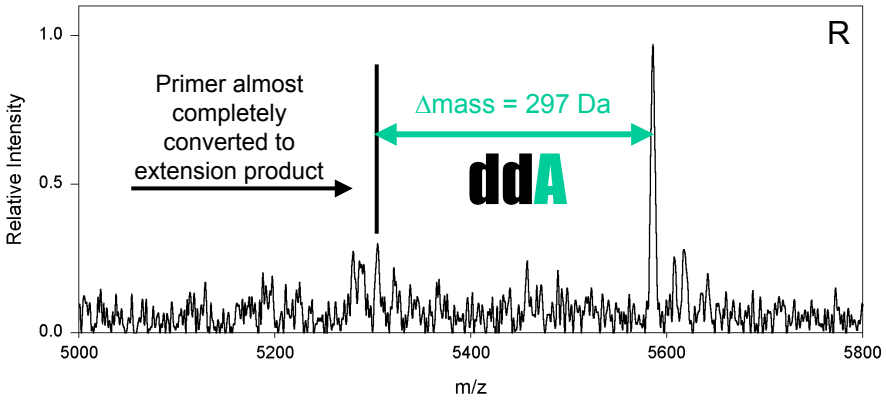
M42 (A/T) Y Chromosome Marker



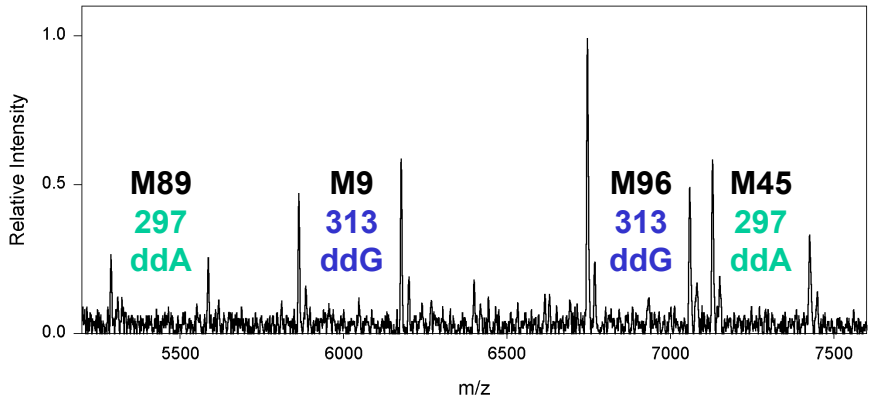
M45 (G/A) Y Chromosome Marker



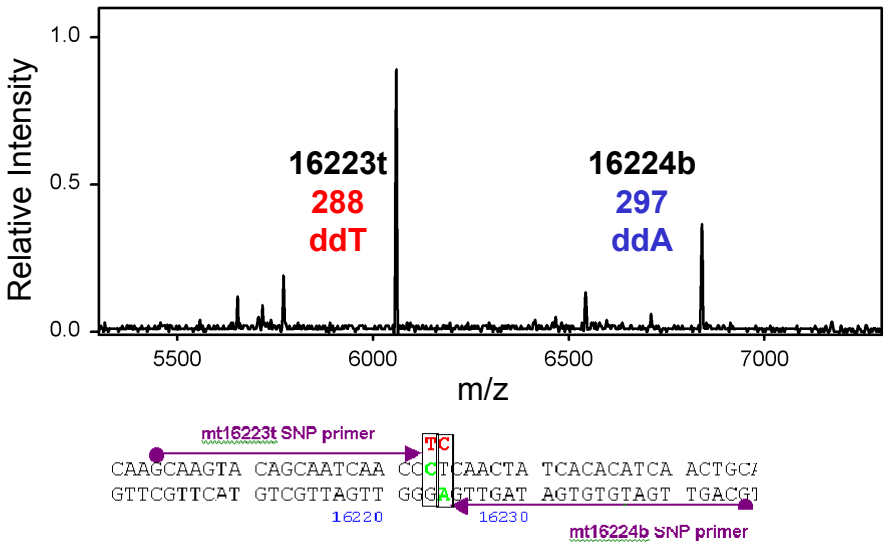
M89 (C/T) Y Chromosome Marker



M9, M45, M89, and M96 Y Chromosome Markers Multiplexed



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



Challenges for Analysis of DNA by MALDI-TOF MS

- Non-uniform matrix crystallization leads to non-homogeneous sample spots; MS signal can thus vary across the sample ("sweet spots" exist). This can make automated data collection difficult.
- Sample salts affect resolution and sensitivity; a "clean up" step is required prior to MALDI analysis.
- Limited mass range with high accuracy and resolution (best below 10 kDa but works up to ~35 kDa). The limited mass range restricts multiplex designs.
- Salt adducts Na^+(+22) can interfere with accurate genotyping of heterozygous samples.

Advantages/Disadvantages of the Basic Primer Extension Assay

•Uses readily available reagents
  –Synthetic primers (no modifications)
  – ddNTPs
•Automation of liquid handling can be performed with robotics

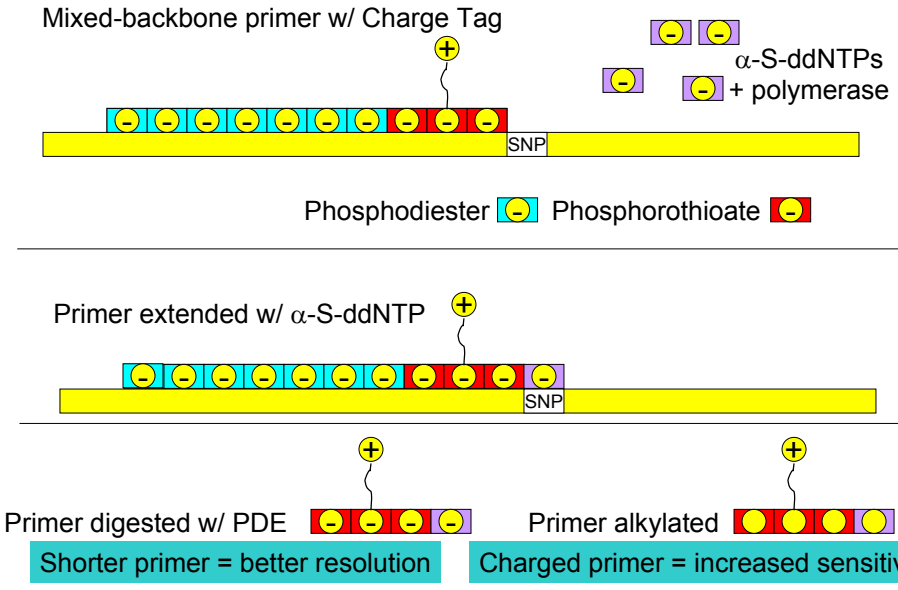
---

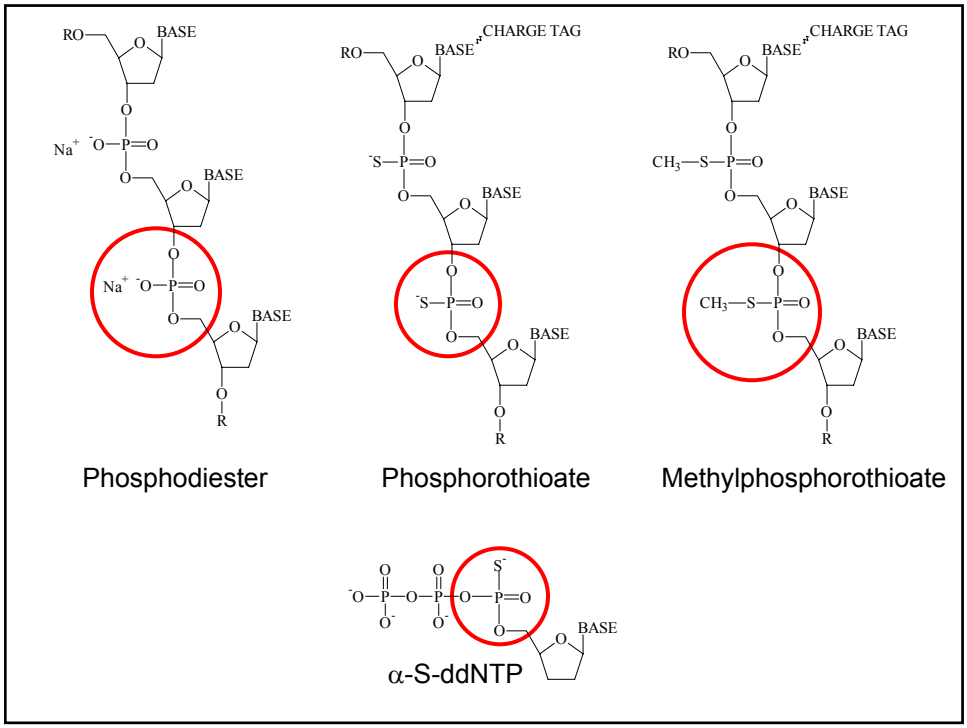
•Limited multiplexing capabilities (~ 5plex)
•As mass range increases, resolution decreases
  – heterozygous samples difficult to resolve
•Salt adducts may interfere with data interpretation
  – products must be purified
•3HPA matrix
  – non-homogeneous crystal formation

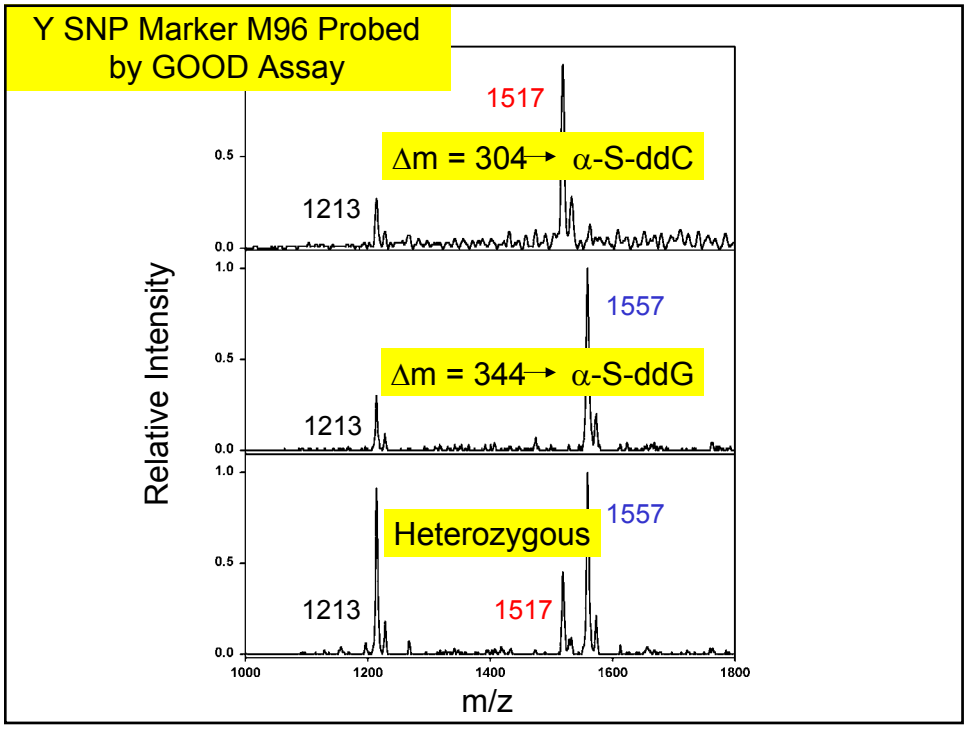
"GOOD" Assay

- Sauer *et al*., (2000) *Nucleic Acids Res* 28:E13
- Variation on Primer Extension Assay
- Uses a chemically modified primer
  - "Charge tag"
  - Phosphorothioate linkage
- Chromatography is not required for purification
- Alkylated/Charged tagged primer increases sensitivity in MALDI TOF MS by 100 fold

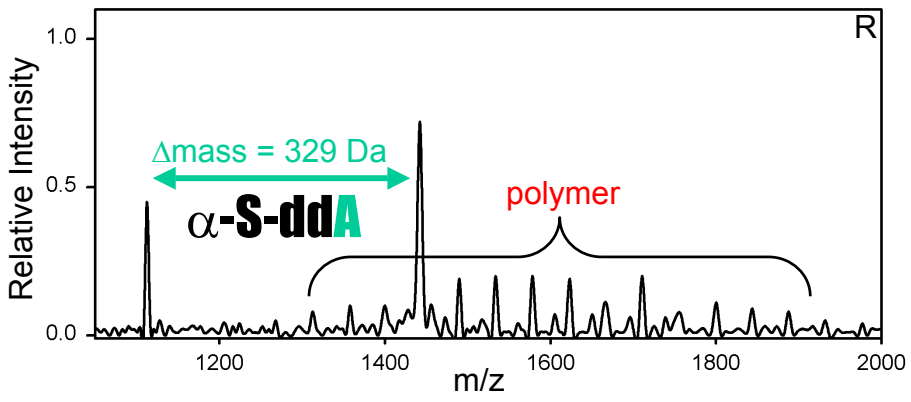
SNP Detection by "Good" Assay




RO
BASE
CHARGE TAG
Na+ ⁻O—P=O
⁻S—P=O
CH3—S—P=O
R
Phosphodiester
Phosphorothioate
Methylphosphorothioate
α-S-ddNTP


Y SNP Marker M96 Probed by GOOD Assay
Relative Intensity
m/z
1517
Δm = 304 → α-S-ddC
1213
1557
Δm = 344 → α-S-ddG
Heterozygous
0.0
0.5
1.0
1000
1200
1400
1600
1800

M42 (A/T) Y Chromosome Marker GOOD Assay



Advantages/Disadvantages of the "GOOD" Assay

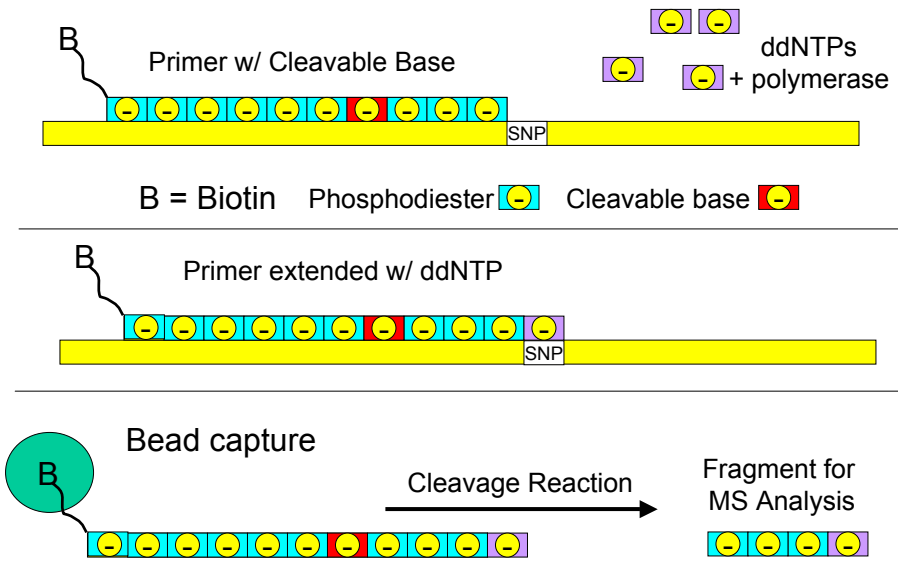
- 5plex has been shown
- Higher sensitivity facilitates data collection in automated mode
- Control of fragment mass with "Mass Tags"
- Lower mass primers fragment easier to ionize
- No salt adducts
- No chromatography required

- Requires specially modified synthetic primers
- Requires modified α-S-ddNTPs
- Performing assay requires some skill/steps
- Automated in development lab

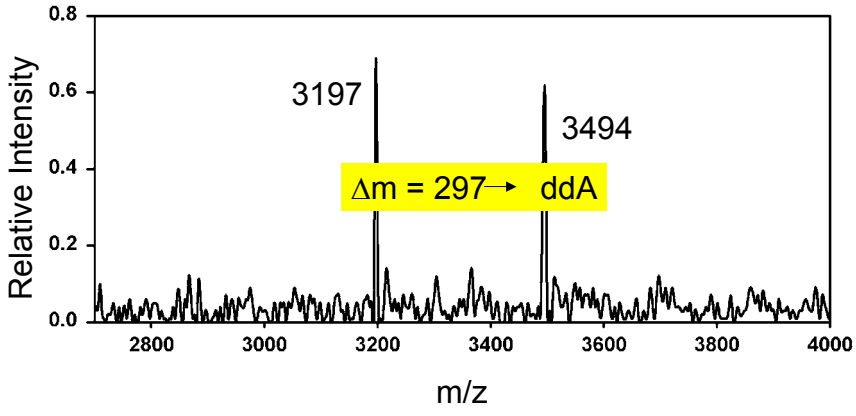
Primer Extension with Cleavable Primer

- GeneTrace Systems Inc.; US Patent 5,700,642
- Variation on the Primer Extension Assay
- A 5' biotin labeled primer containing a cleavable base is used
- Streptavidin coated magnetic beads are employed for purification

Use of Cleavable Primer

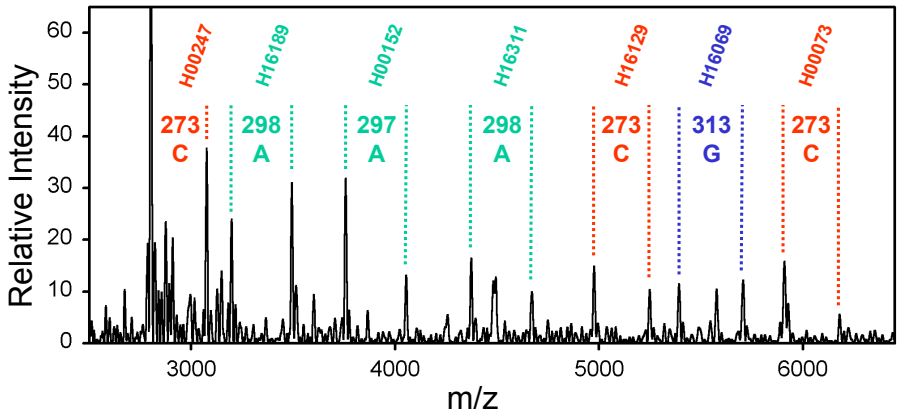


mtSNP H16189 Probed Using a Cleavable Primer



DNA oligomer used in primer extension assay
Biotin – 5' GGTTGATTGCTGTACTTGCTTGTAAGCATGGGG 3'

mtSNP 7plex



Seven SNP sites located in the control region of mtDNA were probed by the extension assay and analyzed by MALDI-TOF MS

Advantages/Disadvantages of Using a Cleavable Primer

- Multiplexing (10plex has been shown)
- Control of fragment mass with cleavable base
- Reduced mass primer fragments are easier to ionize
- Better resolution (due to lower mass range)
- Only bead chromatography required

- Uses a modified primer
- Time required for bead washing steps
- Multiple steps for automation

Future Directions

- Optimize automation of assays/data analysis for high throughput genotyping
- Perform comparisons with other technologies for SNP detection
- Increase multiplexing capabilities of SNP detection by MALDI-TOF MS

Acknowledgments

Funding:

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

John Butler (NIST)

Jay Stoerker (Bruker Daltonics)
Thomas Froehlich (Bruker Daltonics)
Michael Marino (Transgenomic)
Joseph Devaney (Transgenomic)