Multiplex SNP Assays for the Evaluation of Forensic Markers

Dr. Peter M. Vallone
National Institute of Standards and Technology (U.S.)

Department of Biotechnology, Royal Institute of Technology, Stockholm, Sweden November 18, 2003

NIST is located 30 miles (~50 km) north of Washington D.C.

Founded in 1901, NIST is a non-regulatory federal agency within the U.S. Commerce Department's Technology Administration.

NIST's mission is to develop and promote measurement, standards, and technology to enhance productivity, and facilitate trade.

NIST is comprised of 8 laboratories
Chemical Science and Technologies Laboratories
Biotechnology Division
DNA Technologies Group
Human Identity Project

National Institute of Standards and Technology

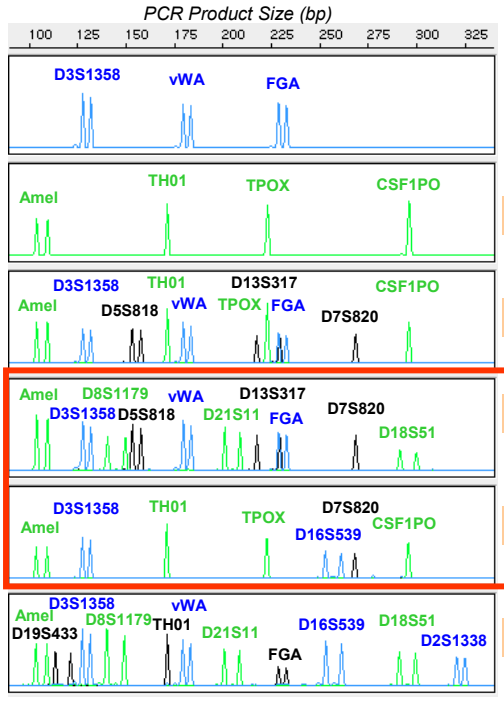
NIST

...working with industry to develop and apply technology, measurements and standards

Human Identity Project

Project leader Dr. John Butler

- Working with the forensic community and industry
- Provide Standard Reference Materials
- Develop multiplex PCR primer sets for new markers
- Evaluate newly discovered forensic markers in U.S. populations
- Coordinate Inter-laboratory Studies
- Maintain STRbase (http://www.cstl.nist.gov/biotech/strbase/)

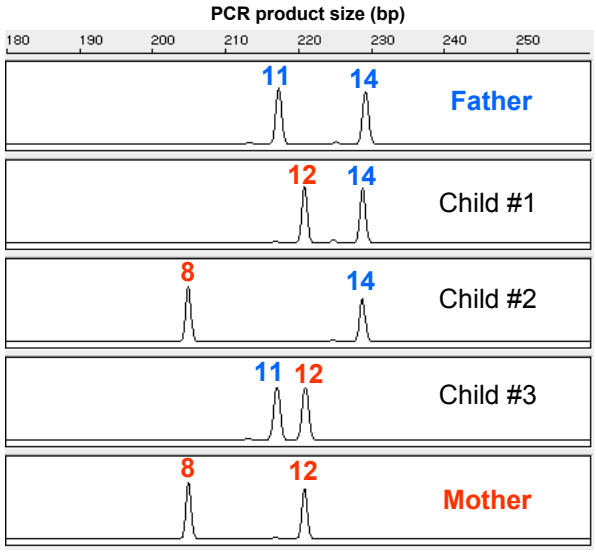


Same DNA Sample Run with Each of the STR Kits

	Power of Discrimination
Blue	1:5000
Green I	1:410
Profiler™	$1:3.6 \times 10^{9}$
Profiler Plus™	$1:9.6 \times 10^{10}$
COfiler™	$1:8.4 \times 10^{5}$
SGM Plus™	$1:3.3 \times 10^{12}$

PATERNITY TESTING

Family Inheritance of STR Alleles (D13S317)



Past Projects

- SRM 2391b PCR-based DNA Typing Standard
- Y-Chromosome Standard SRM 2395
- DNA Stability Studies from Aged Blood Stains
- Interlaboratory Studies on Analysis of Sample Mixtures

- STR and SNP Typing with MALDI-TOF MS
- Y-STR Multiplex Development: 20-plex, 10-plex, 11-plex
- NIST Population Sample Collection and Initial Typing
- Evaluation of Optimal Y-STRs in U.S. Populations

Areas of Research for 2004

- Projects to Aid Degraded DNA Analysis
  - miniSTR with CODIS loci
  - Develop new miniSTR loci
  - Autosomal SNP typing
  - Mitochondrial coding region SNP assays (collaboration with Tom Parsons)
  - Evaluation of mtDNA LINEAR ARRAYS (collaboration with Roche)
- Projects to Aid Y-Chromosome Studies
  - Y-SNP markers and assays (collaboration with Mike Hammer/Alan Redd)
  - Y-STR markers and multiplex assays
- DNA Quantitation
  - NIST Interlaboratory Comparison Quantitation Study A
  - Develop SRM 2372-Human DNA Quantitation Standard

Overview

SNPs

Assay Platforms and Instrumentation

Multiplexing

U.S. Population Samples

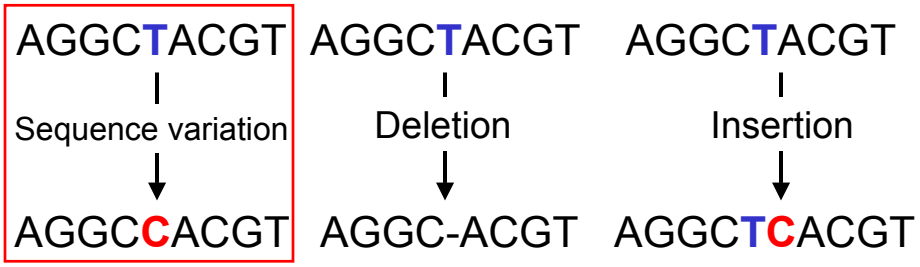
Y Chromosome and Mitochondrial Markers

Results

mtSNP 11 plex

Y-SNP multiplexes

SNP
Single Nucleotide Polymorphism



Low mutation rate 10^{-8}
Typically Bi-allelic

SNP Facts

Most common type of variation in the human genome (90%)

Estimated to occur every 100-300 bases

For a SNP to be defined it must occur in at least 1 % of the population

2 out of every 3 SNPs is a C-T transition

Occur in coding and non coding regions of the genome

http://www.ornl.gov/TechResources/Human_Genome/faq/snps.html

Biomedical Importance of SNPs

Variations in DNA sequence can have a major impact on how humans respond to disease, bacteria, viruses, toxins, chemicals and drugs.

This makes SNPs of great value for biomedical research and for developing pharmaceutical products or medical diagnostics.

SNPs are also evolutionarily stable --not changing much from generation to generation --making them easier to follow in population studies

Forensic Utility of SNPs

Human identification purposes (criminal, paternity, evolutionary, population studies, predicting ethnicity)

The short PCR amplicons required for typing SNPs may result in success with degraded samples and possibly higher sensitivity

Simplicity in testing – typically bi-allelic markers (versus length polymorphisms)

Improve assay development (both multiplex PCR and SNP detection)

For serious forensic usage parallel high-throughput methods will be required for typing

Forensic Utility of SNPs

Short tandem repeat (STR)
CTGATGCTA**(GATA)$_n$**GACTACTTA
n = 5 to 15 = 66 possible allelic combinations

Single Nucleotide Polymorphism (SNP)
CTGATGCTA**(G/A)**GACTACTTA
3 possible allelic combinations

For human ID purposes more SNPs would be needed than STRs
Multiplexing is essential

Overview

SNPs

Assay Platforms and Instrumentation

Multiplexing

U.S. Population Samples

Y Chromosome and Mitochondrial Markers

Results

- mtSNP 11 plex
- Y-SNP multiplexes

Instrumentation

PCR & primer extension

Multi-Color Capillary Electrophoresis
(ABI 310 or 3100)

Luminex Beads hybridization

Luminex 100 Flow Cytometer

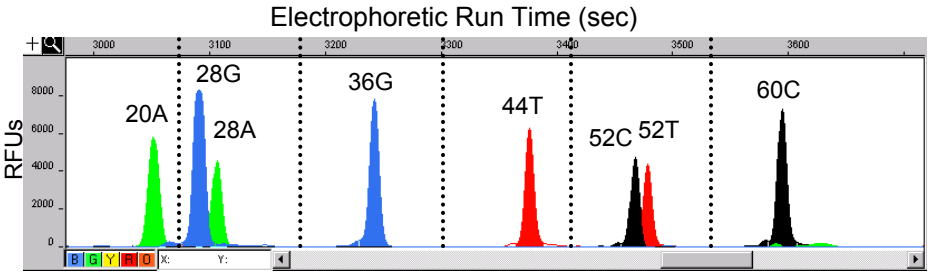
Primer Extension

Time-of-Flight Mass Spectrometer

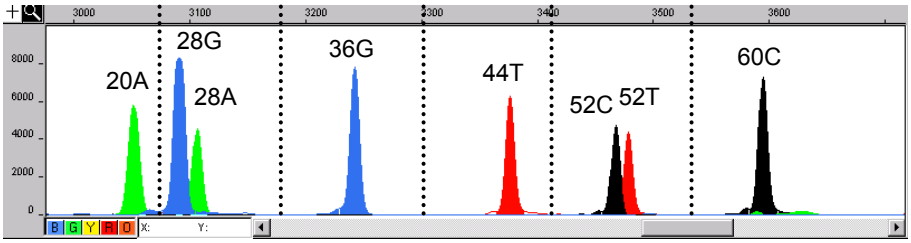
TaqMan

ABI 7000 SDS

Allele-Specific Primer Extension (ASPE)



Detection of SNPs with ABI 310/3100



SNaPshot™ CEPH Control Reaction

20 nucleotides — ddA

36 nucleotides — ddG

44 nucleotides — ddT

60 nucleotides — ddC

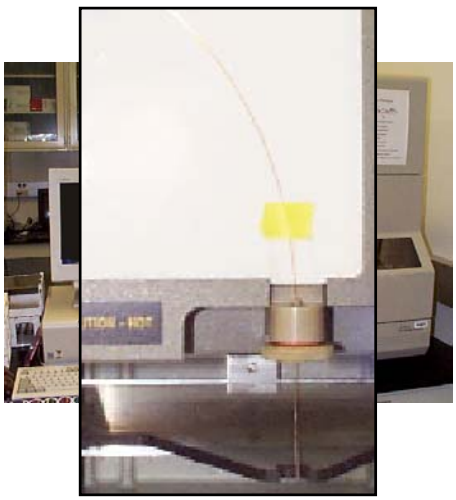
Multiplexing possible by use of different length primers

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

Priming site

Capillary Electrophoresis Instrumentation

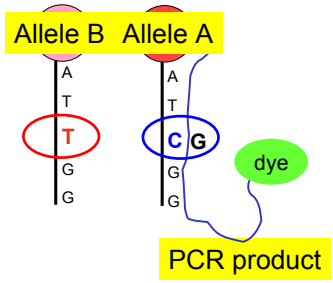
ABI 310
single capillary

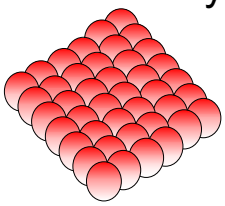


ABI 3100
16-capillary array

SNP Detection by Hybridization

Luminex Bead Array Assay



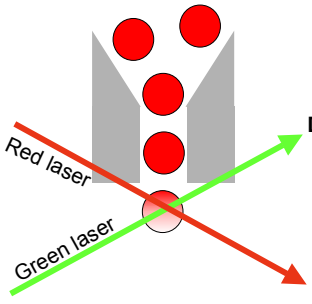


100 different colored beads are possible (potential for multiplexing 50 SNP markers)

Luminex 100 Flow Cytometer

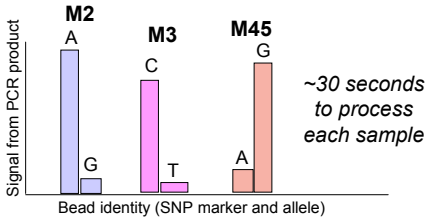
MARLIGEN BIOSCIENCE

Signet™ Y-SNP Typing System (42 Y-SNPs + AMEL)



Detects labeled PCR product

Identity of bead (probe)



~30 seconds to process each sample

ASPE combined with MALDI-TOF-MS Analysis

Primer is extended by one base unit

Natural non-labeled ddNTPs + polymerase

Oligonucleotide primer 18-28 bases

5’ 3’

SNP

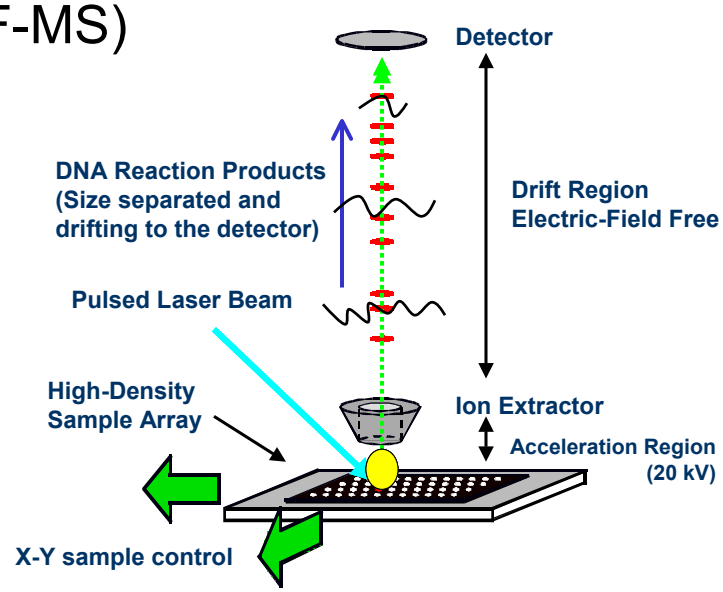
PCR Amplified DNA Template

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

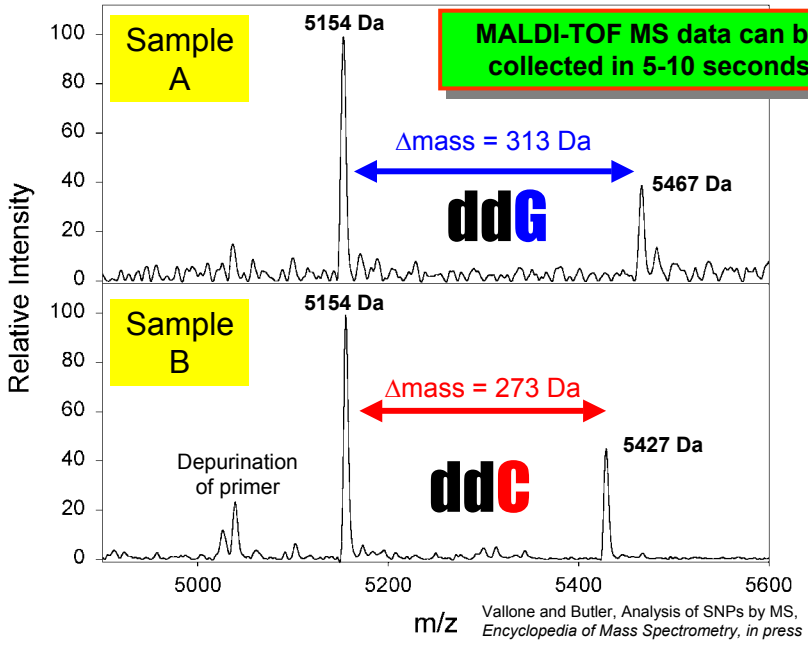
40 Cycles
96°C 10s
50°C 20s
72°C 30s

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

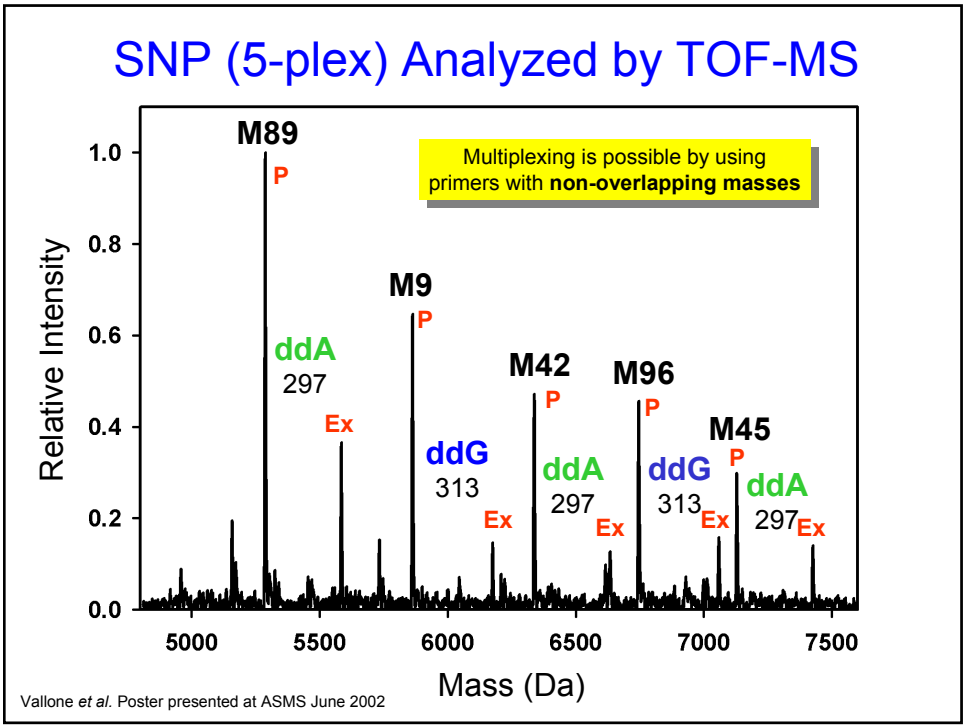
Time-of-Flight Mass Spectrometry (TOF-MS)

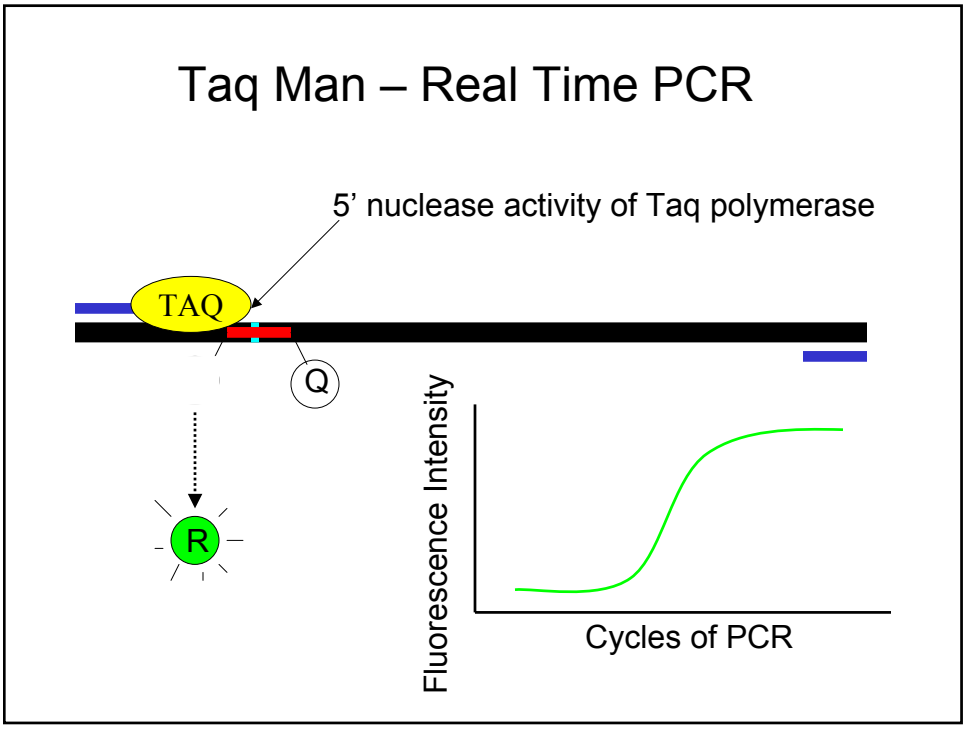


MS Data from Y SNP Marker M96



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


SNP (5-plex) Analyzed by TOF-MS
Multiplexing is possible by using primers with non-overlapping masses
M89
P
ddA
297
Ex
M9
P
ddG
313
Ex
M42
P
ddA
297
Ex
M96
P
ddG
313
Ex
M45
P
ddA
297
Ex
Relative Intensity
1.0
0.8
0.6
0.4
0.2
0.0
5000
5500
6000
6500
7000
7500
Mass (Da)
Vallone et al. Poster presented at ASMS June 2002


Taq Man – Real Time PCR
5’ nuclease activity of Taq polymerase
TAQ
Q
R
Fluorescence Intensity
Cycles of PCR

Allelic Discrimination Assay using TaqMan

Design 2 Taqman probes for each expected SNP
Each probe will have a different reporter dye (FAM/VIC)
Possible assays outcomes
SNP A = FAM
SNP B = VIC
SNP A/B = FAM/VIC

Typically TaqMan genotyping assays are singleplex!

Summary of SNP Assays

	Advantages	Disadvantages
ASPE-CE	Moderate degree of	Development of multiplex
ASPE-M		ultiplexing nt than CE
Microbe (Lumine		custom
TaqMan	Rapid - one step Good for one marker on 1000's of sample	No multiplexing Costly for typing many SNPs & few samples

Other Technologies
Pyrosequencing
Chip Based – Affymetrix - Agilent
Allele specific PCR
Invader-mismatch cleavage
Orchid SNPstream
Illumina Bead Arrays

Overview

SNPs

Assay Platforms and Instrumentation

Multiplexing

U.S. Population Samples

Y Chromosome and Mitochondrial Markers

Results

- mtSNP 11 plex
- Y-SNP multiplexes

Advantages of Multiplexing

Obtain more information per unit time

Reduce the amount of limited forensic sample used

Save on reagents; enzyme, buffers, DNA oligomers

Reduces labor

Streamlines data analysis

For certain markers it is essential (SNPs, YSTRs)

Coincides with high capacity instrumentation and new SNP typing technologies

Goals for Multiplex Assay Development

Working with collaborators who have markers of forensic interest

Evaluate the forensic utility of newly discovered markers (medium sized multiplexes 5 – 10 loci)

Further the understanding of developing multiplex assays (primer design, QC)

Publish assay details for others to evaluate (commercial and research)

Multiplex PCR Primer Selection

Identify markers of interest (collaborations, literature, research)

Organize sequences with ~250 bases of sequence up- and downstream of the SNP

PCR product size
- Short amplicons for degraded samples, SNPs
- Longer amplicons for STRs

Use software for selecting singleplex primer pairs

Primer3
www-genome.wi.mit.edu/genome_software/other/primer3.html

Steve Rozen and Helen J. Skaletsky (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology.* Humana Press, Totowa, NJ, pp 365-386

Multiplex PCR Design

Select singleplex PCR primers for each amplicon 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
```

Running Primer3 Locally

Sending multiple sequences over the web for primer selection can be tedious

The Primer3 web output is acceptable for the screen viewing or printing but not for organizing in spreadsheets

Primer3 is publicly available and can be run on a Unix, PC (Linux), or Mac (OSX) computer

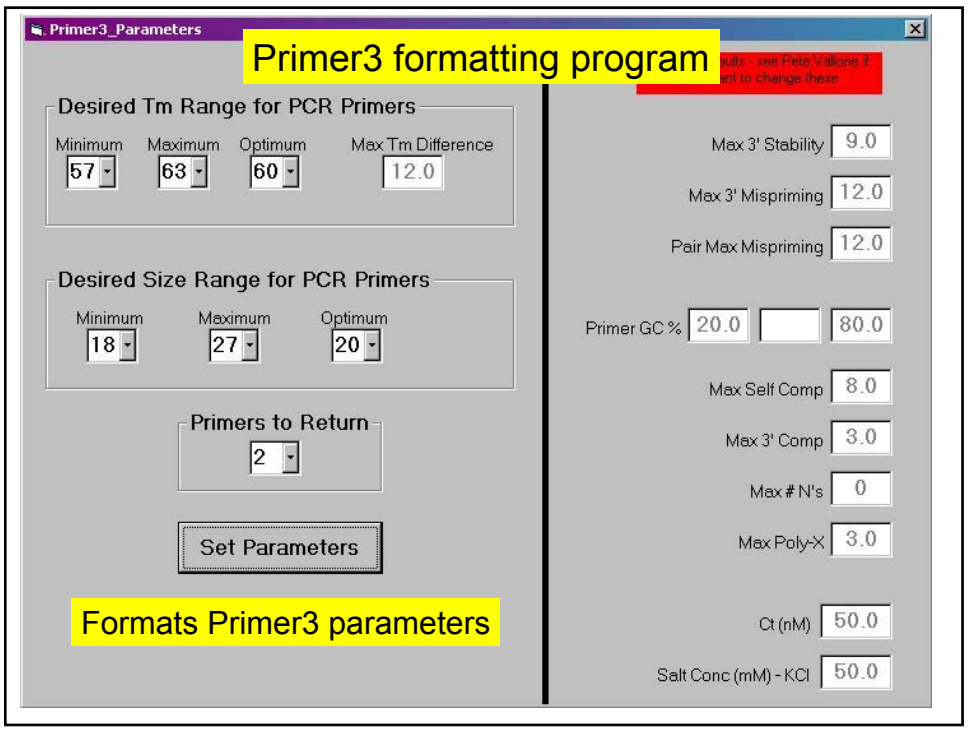
Developed a program that formats files for Primer3 input

Reference sequences that are stored in Excel can be quickly formatted for Primer3

Format of Template Sequences

Locus	Total Length	Minimum	Maximum	Optimal	Excluded Region	SNP site
M3	255	105	150	125	174,60	204
TGATTATTTAGAAACAAAACAATAAACAATAACAAAACAATGGTTCCCTGTTAAAATGTG						
M9	255	105	150	125	237,60	267
CCTGTGCACGCCAAAGCGGAAGCTGAAGTGCGGCGTCTTTGATCTCTCAATCCTGGAG/						

Sequences stored in excel
Will be adapted for FASTA format & comma delimited



Example input format for Primer3

```
PRIMER_SEQUENCE_ID=M9
SEQUENCE=GCAGCATATAAAACTTTCAGGACCCTGAAATACAGAACTG
CAAAGAAACGGCCTAAGATGGTTGAATNCTCTTTATTTTTCTTTAATTTAG
ACATGTTCAAACGTTCAATGTCTTACATACTTAGTTATGTAAGTAAGGTAG
CGCTTACTTCATTATGCATTTCAATACTCAAAAAAAATTCCTTTGTGAAAT
GTTGAAATATTTTTCTAATCTGTTTCACGAGCTTCAAAAATGAGGAAAAAA
GATTCAGTTTACATTTCAGCAAAATGCCTCTTTTTAATCGGATTTATGTTT
ACTTAACATTTACAGTACATTTACGCTTGAGCAAAGTTAGGTTTT
PRIMER_COMMENT=(340 bp); G to C at position 68
PRIMER_MISPRIMING_LIBRARY=/Users/vallone/Desktop/primer3/misprM9
PRIMER_MAX_MISPRIMING=8
PRIMER_PAIR_MAX_MISPRIMING=20
EXCLUDED_REGION=38,60
PRIMER_PRODUCT_SIZE_RANGE=90-150
PRIMER_PRODUCT_OPT_SIZE=105
PRIMER_MIN_SIZE=18
PRIMER_MAX_SIZE=27
PRIMER_OPT_SIZE=20
PRIMER_OPT_TM=60
PRIMER_MIN_TM=57
PRIMER_MAX_TM=63
PRIMER_NUM_RETURN=1
PRIMER_EXPLAIN_FLAG=1
PRIMER_LIBERAL_BASE=1
=
PRIMER_SEQUENCE_ID=M42
SEQUENCE=AAAGCGAGAGATTCAATCCAGGATGACAGAATGCGTTCAC
CTTTAAAGGGATTAAAAGAAGTATAATACAGTCTGTATTATTAGATCACCC
AGAGACACACAAAACAAGAACCGTGAATTGAATTAGTGGTATACTAATAG
```

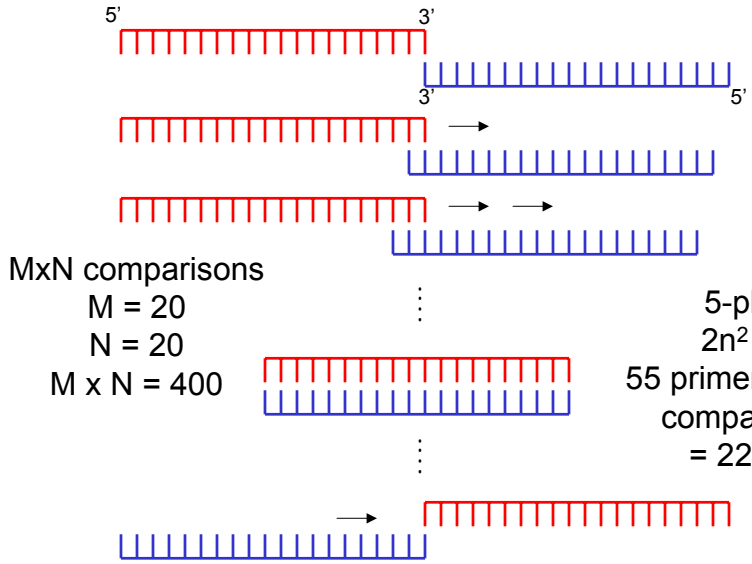
Non-Specific Interactions

Primers that interact with non-specific (undesired) regions of a genome OR with each other can degrade PCR performance

Screening for alternate genomic binding regions can be accomplished using **BLAST** *http://www.ncbi.nlm.nih.gov*

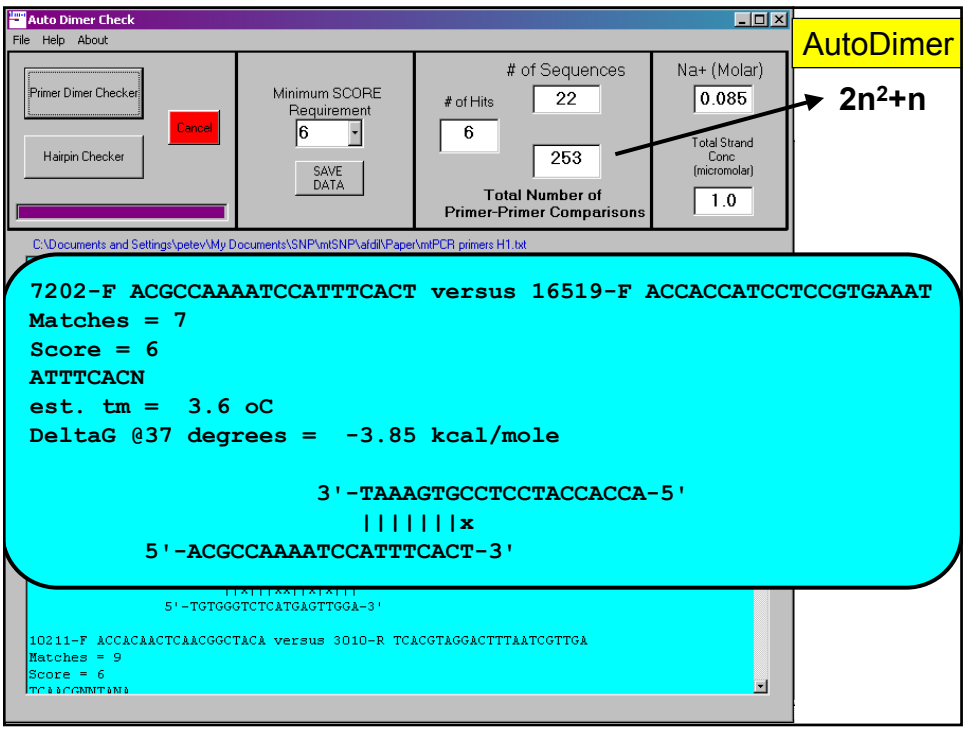
Screening for potential primer-dimer interactions is accomplished using in house software - **AutoDimer**

Basic Sliding Algorithm for Complementarity Check



MxN comparisons
M = 20
N = 20
M x N = 400

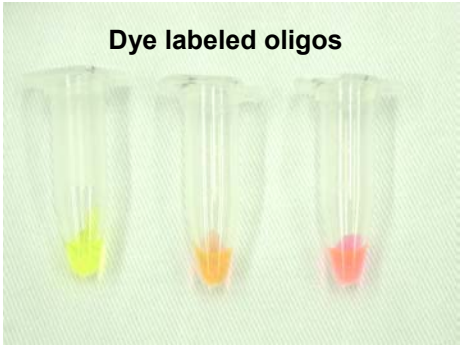
5-plex
$2n^2 + n$
55 primer–primer comparisons
= 22,000



AutoDimer

$2n^2+n$

PCR Primer Quality Control



6FAM (yellow), **VIC** (orange), **NED** (red)

- UV Spec to determine concentration
- HPLC to evaluate purity
- TOF-MS to confirm correct sequence

Butler *et al.* (2001) *Forensic Sci. Int.* 119: 87-96

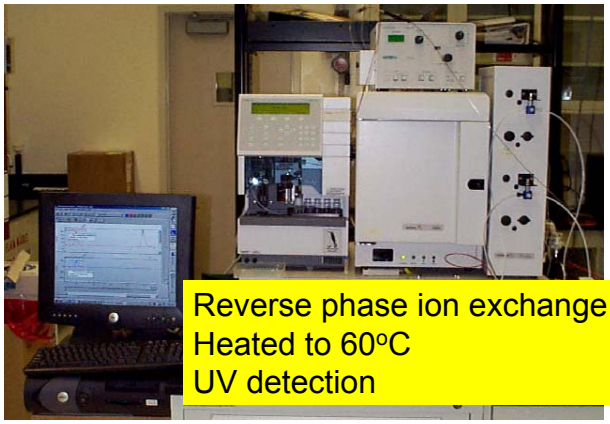
Determination of DNA Oligomer Concentrations

Expected 100 μM

Seq	μM	% deviation
1	173.3	42.3
2	164.8	39.3
3	155.0	35.5
4	124.1	19.4
5	116.4	14.1
6	98.5	-1.5
7	108.6	7.9
8	103.1	3.0
9	120.8	17.2
10	79.6	-25.7
11	83.0	-20.5

Concentrations were estimated by UV Spec readings @260 using extinction coefficients determined from nearest-neighbor values

Varian Helix DHPLC System

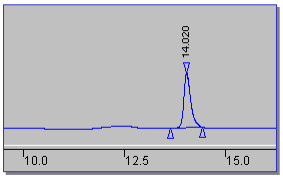


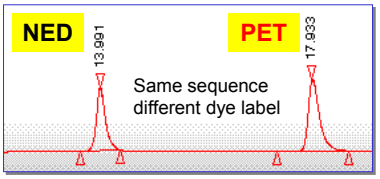
Oligo QC

Oligo Purification

Fluorescent dye studies (excess dye removal)

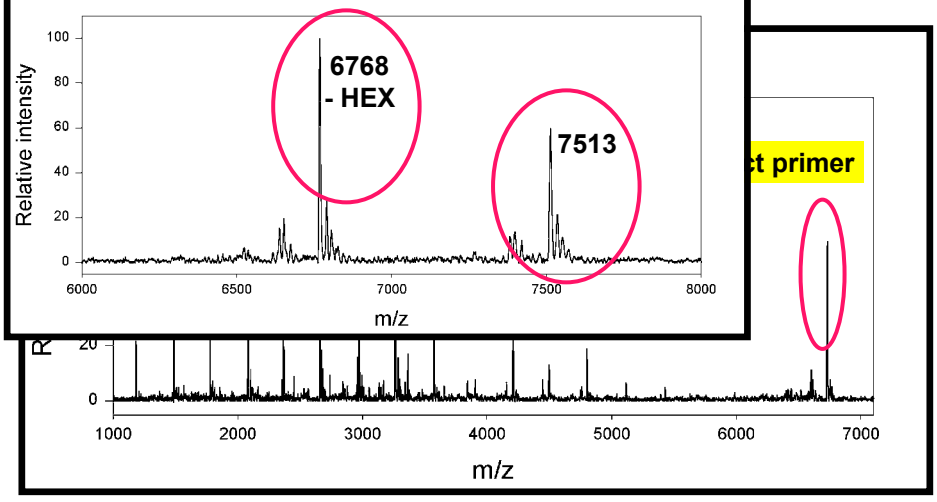
Reverse phase ion exchange column
Heated to 60°C
UV detection



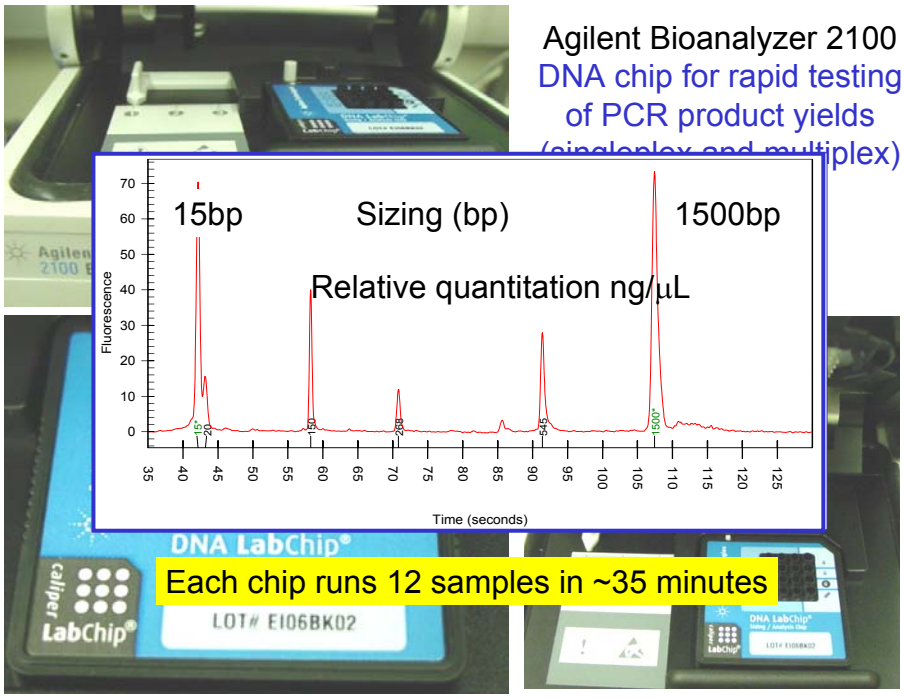


MALDI QC of Commercial Oligos

Loss of Fluorescent dye



Vallone and Butler (Oct 2000) *International Symposium on Human Identification* (Biloxi, MS)



Publications Describing Multiplex Assay Design

Schoske, R., Vallone, P.M., Ruitberg, C.M., Butler, J.M. (2003) Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Anal. Bioanal. Chem.*, 375: 333-343.

Butler, J.M., Schoske, R., Vallone, P.M. Highly multiplexed assays for measuring polymorphisms on the Y-chromosome. (2003) *Progress in Forensic Genetics 9* (Brinkmann, B. and Carracedo, A., eds.), Elsevier Science: Amsterdam, The Netherlands, International Congress Series 1239, pp. 301-305.

Schoske, R., Vallone, P.M., Kline, M.C., Redman, J.W., Butler, J.M. (2003) High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays, *Forensic Sci. Int., in press*

Butler, J.M. (2003) Constructing STR multiplex assays. *Methods in Molecular Biology: Forensic DNA Typing Protocols* (Carracedo, A., ed.), Humana Press: Totowa, New Jersey, *in press*.

Butler, J.M., Schoske, R., Vallone, P.M., Kline, M.C., Redd, A.J., Hammer, M.F. (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Int*. 129: 10-24.

Butler, J.M., David, V.A., O'Brien, S.J., Menotti-Raymond, M. (2002) The MeowPlex: a new DNA test using tetranucleotide STR markers for the domestic cat. *Profiles in DNA*, Promega Corporation, Volume 5, No. 2, pp. 7–10. http://www.promega.com/profiles/502/ProfilesInDNA_502_07.pdf

Butler, J.M., Devaney, J.M., Marino, M.A., Vallone, P.M. (2001) Quality control of PCR primers used in multiplex STR amplifications. *Forensic Sci. Int*., 119: 87-96.

Butler, J.M., C.M. Ruitberg, Vallone, P.M. (2001) Capillary electrophoresis as a tool for optimization of multiplex PCR reactions, *Fresenius J. Anal. Chem.* 369: 200-205.

Overview

SNPs

Assay Platforms and Instrumentation

Multiplexing

U.S. Population Samples

Y Chromosome and Mitochondrial Markers

Results

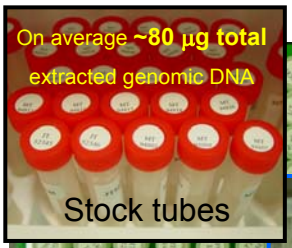
- mtSNP 11 plex
- Y-SNP multiplexes

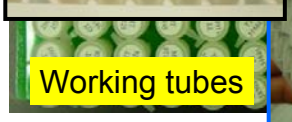
NIST U.S. Population Samples

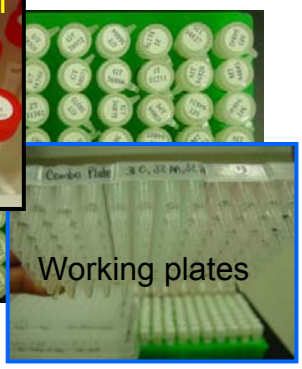
As of 09/2003 666 males (anonymous; self-identified ethnicities)

260 Caucasians
260 African Americans
143 Hispanics
3 Asians

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







Working tubes/plates 1 ng/uL

To date: (~50,000 allele calls)

Identifiler (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

Overview

SNPs

Assay Platforms and Instrumentation

Multiplexing

U.S. Population Samples

Y Chromosome and Mitochondrial Markers

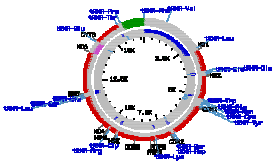
Results

mtSNP 11 plex

Y-SNP multiplexes

Markers of Interest

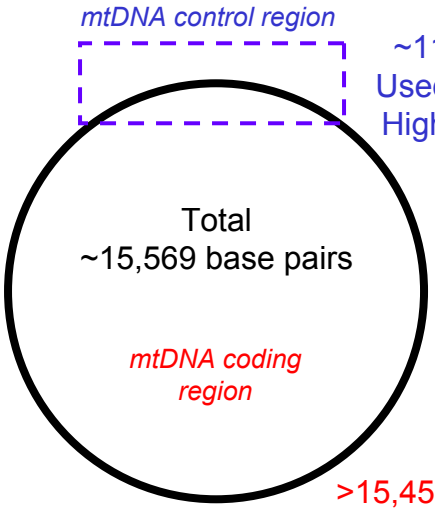
- Mitochondrial DNA (mtDNA)
  - maternally inherited
  - polymorphic control region (D-loop)
  - ~500-2000 copies per cell
  - coding region



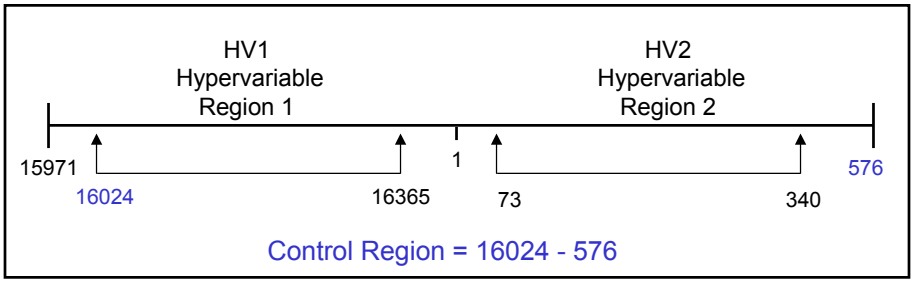
- Y chromosome
  - paternally inherited
  - variety of Y-STR and Y-SNP markers
  - ***haplotype rather than genotype***

Require large databases because recombination does not occur

Mitochondrial Genome (mt Genome)



The Current mtDNA Amplification & Sequencing Strategy Focuses on the Hypervariable Regions of the mitochondrial genome HV1 and HV2



In Caucasians, approximately 7% of HV1 and HV2 sequences are identical

The Use of Full mtGenome Polymorphisms

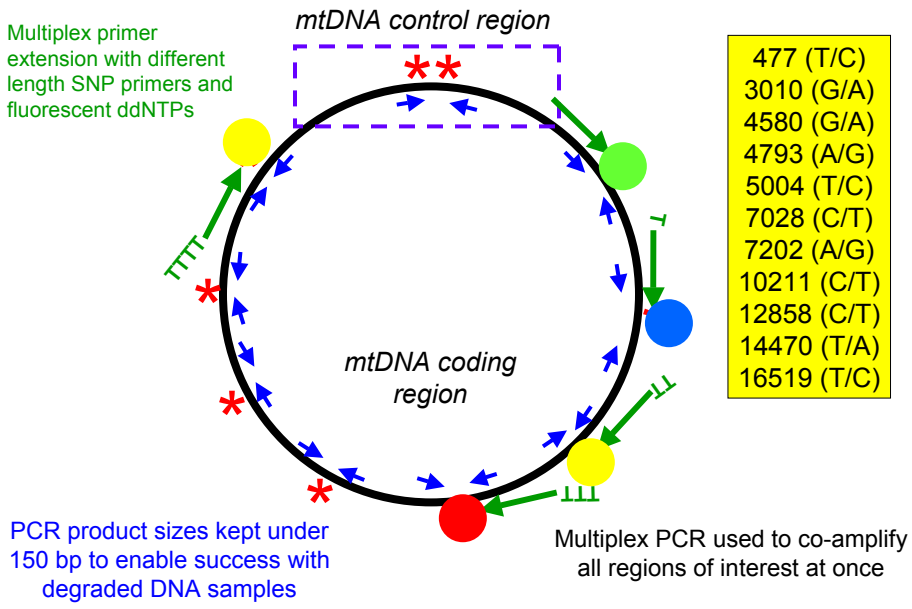
mtGenome sequencing data (241 mt genomes) reveals numerous SNPs that can help distinguish Caucasians sharing common HV types (Tom Parsons and Mike Coble AFDIL)

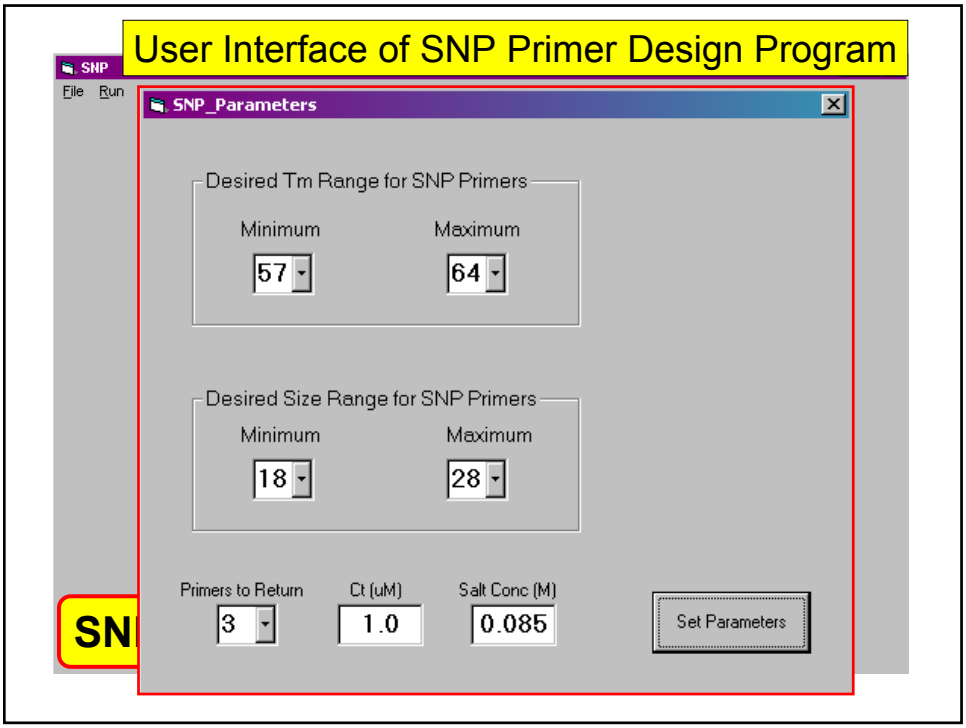
11 SNP sites were selected to help resolve Caucasian individuals having the most common HV1/HV2 type

mtSNPs: Silent and at third codon positions or fall in the short non-coding regions between genes in the coding region

Detect in a multiplex assay run on a common forensic instrumental platform

mtSNP 11-plex Assay





Tailed SNP primers allows for multiplexing in the SNaPshot assay

Sequences for 11 extension primers

3010-F	TGTTGGATCAGGACATCCC	19 19
4793-R	$(T)_4$ – TCAGAAGTGAAAGGGGGC	18 22
10211-R	$(T)_{10}$ – ACTAAGAAGAATTTTATGGA	20 30
5004-F	$(T)_{14}$ – AGACCCAGCTACGCAAAATC	20 34
7028-F	$(T)_{18}$ –GACACGTACTACGTTGTAGC	20 38
7202-F	$(T)_{22}$ –CCACAACACTTTCTCGGCCT	20 42
16519-R	$(T)_{24}$ –TGTGGGCTATTTAGGCTTTATG	22 46
12858-F	$(T)_{27}$ –GCAGCCATTCAAGCAATCCTATA	23 50
4580-R	$(T)_{29}$ –TGGTTAGAACTGGAATAAAAGCTAG	25 54
477-F	$(T)_{38}$ –CCCTCCCACTCCCATACTAC	20 58
14470-R	$(T)_{41}$ –GGGAATGATGGTTGTCTTTGG	21 62

Program Output

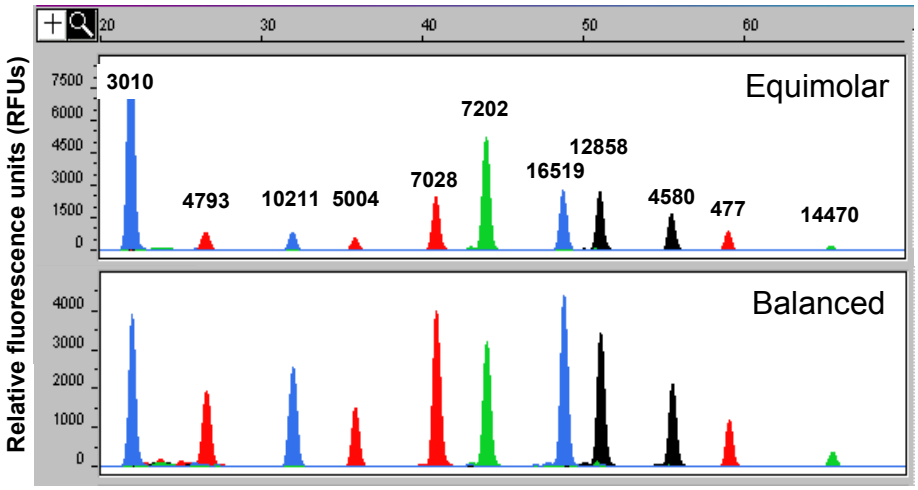
Label	Length	Sequence	Position	Tm
Forward Primers Salt = 0.3Ct = 10				
M42 340 bp (A/T 297 W) AC010889	18	ATTTAGGACACAAAAGCW	280	60.65398
M42 340 bp (A/T 297 W) AC010889	19	GATTTAGGACACAAAAGCW	279	61.96716
M42 340 bp (A/T 297 W) AC010889	20	AGATTTAGGACACAAAAGCW	278	63.67808
Reverse Primers				
M42 340 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

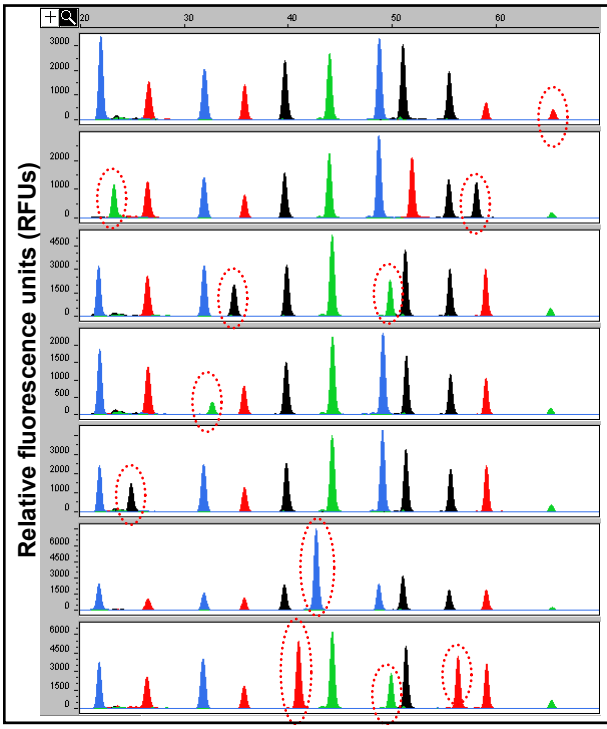
Hairpin	Dimer	Template	Mass	Rank	Mutation	+ddC	+ddT	+ddA	+ddG
4	8	10	5273.48	2.133333	W	N/A	5561.67998	5570.68998	N/A
5	10	10	5602.69	2	W	N/A	5890.889941	5899.899941	N/A
5	10	11	5915.9	2	W	N/A	6204.099902	6213.109902	N/A
4	8	22	6734.42	2.133333	W	N/A	7022.619922	7031.629922	N/A
4	8	20	6116.02	2.133333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

mtSNP 11-plex run on ABI 3100

Multiplex PCR and Multiplex SNP Detection

Measured size (nt) (relative to GSLIZ-120 size standard)



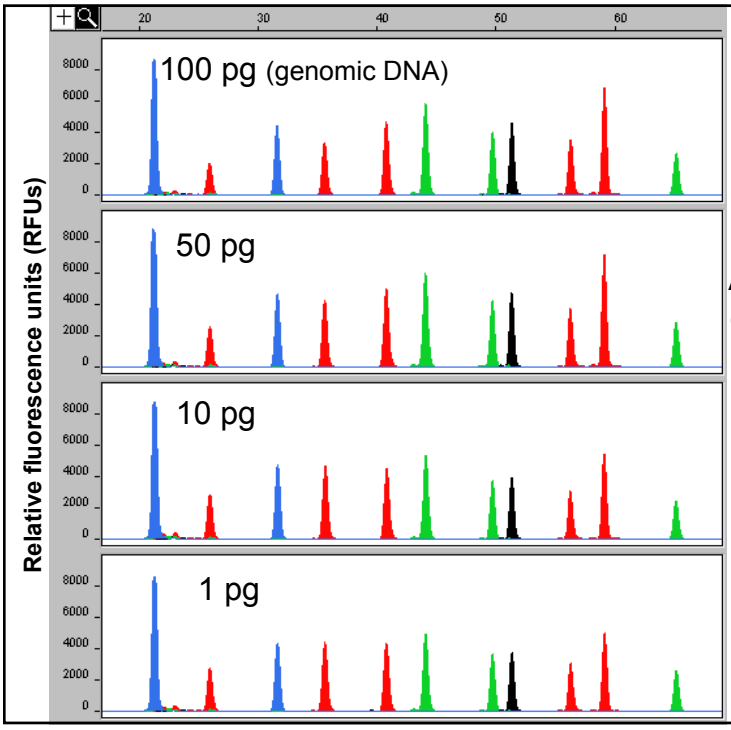


11 plex run on 7 unique samples

All allele variations are represented in these 7 samples

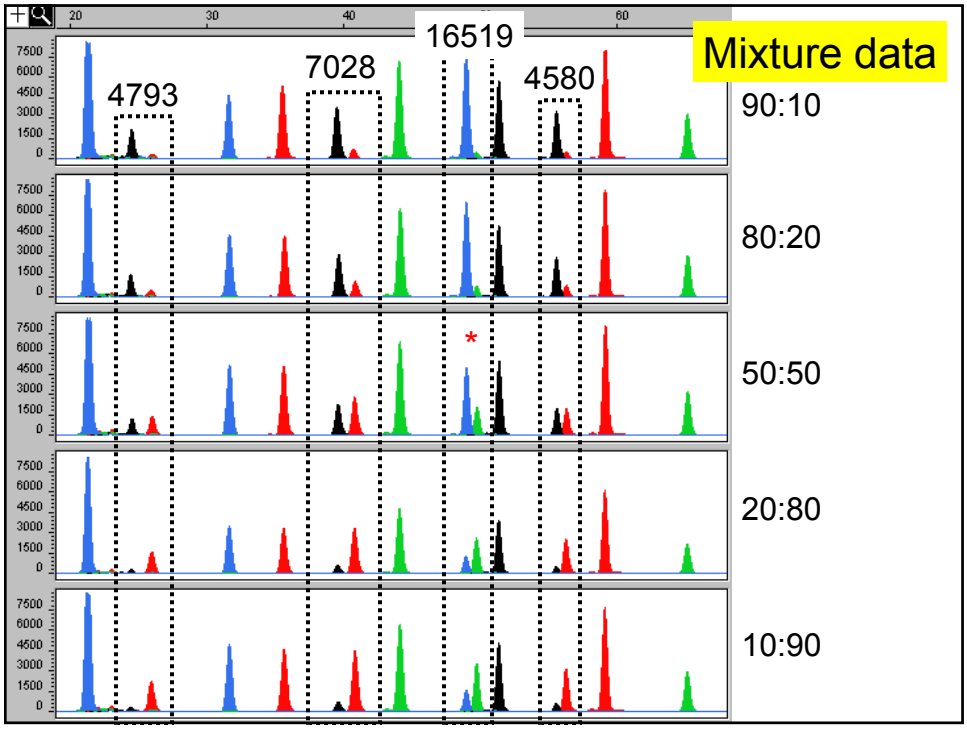
The assay accurately detects each variant

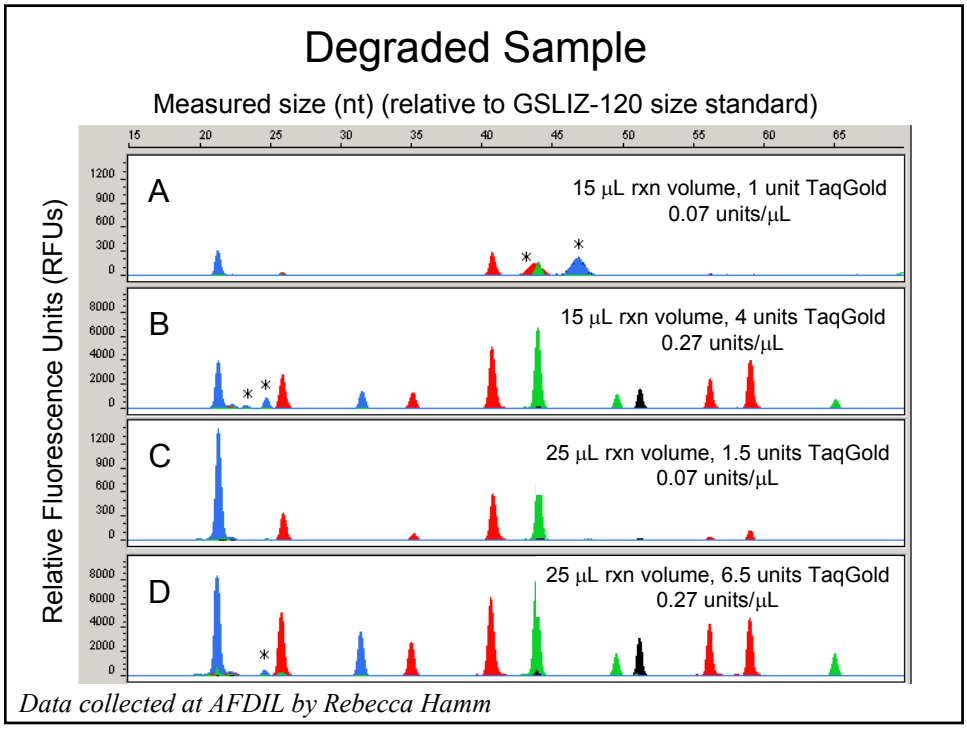
Sizing can be used to develop a macro for automated typing



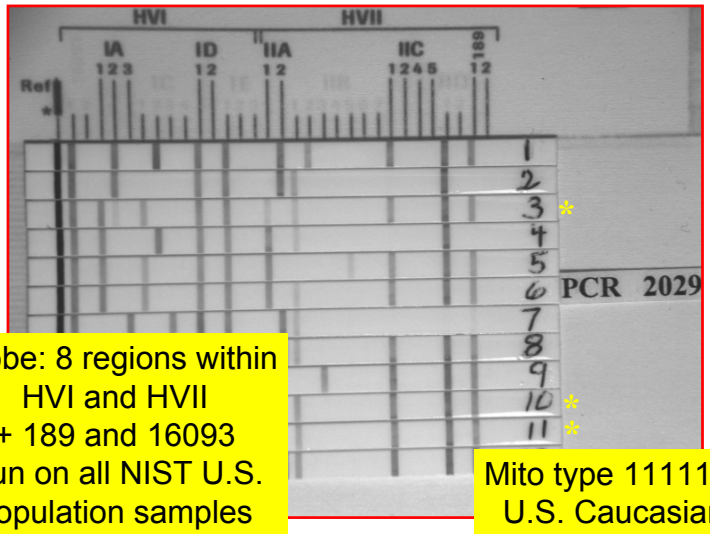
Sensitivity Study

Assay performs down to 1 pg of genomic DNA


Mixture data
4793
7028
16519
4580
90:10
80:20
50:50
20:80
10:90


Degraded Sample
Measured size (nt) (relative to GSLIZ-120 size standard)
Relative Fluorescence Units (RFUs)
A
15 µL rxn volume, 1 unit TaqGold
0.07 units/µL
B
15 µL rxn volume, 4 units TaqGold
0.27 units/µL
C
25 µL rxn volume, 1.5 units TaqGold
0.07 units/µL
D
25 µL rxn volume, 6.5 units TaqGold
0.27 units/µL
Data collected at AFDIL by Rebecca Hamm

Roche Linear Arrays



Probe: 8 regions within HVI and HVII + 189 and 16093 Run on all NIST U.S. population samples

Mito type 11111111AT U.S. Caucasian pop 47 / 286 = 16.4%

Data collected by Margaret Kline and Jan Redman

Typing 51 samples with mt 11 plex assay

51 (47 cauc/4 hisp) samples were identical by Roche linear array assay (most common Haplogroup observed in NIST U.S. Caucasian population samples)

3010	G	A	G	G	G	A	G	G	G	G	G	G	A
4793	A	A	A	A	A	A	A	A	A	G	A	A	A
10211	C	C	C	C	C	C	C	C	C	C	C	C	C
5004	T	T	C	T	T	T	T	T	T	T	T	T	T
7028	C	C	C	C	T	C	T	C	T	C	C	C	C
7202	A	A	A	A	A	A	A	A	A	A	A	A	A
16519	T	C	T	C	T	C	C	T	T	C	C	C	C
12858	C	T	C	C	C	C	C	C	C	C	C	C	C
4580	G	G	G	G	G	G	A	G	A	G	G	G	G
477	T	C	T	T	T	C	T	T	T	T	T	T	T
14470	T	T	T	A	T	T	T	T	T	T	T	T	T
	rCRS	1	1	1	1	2	2	3	4	4	5	12	15

12 haplogroups were observed
4 haplogroups were unique
2 of 11 sites did not vary

11-plex mtSNP assay

Assay is capable of accurately detecting 11 mtSNP in a single assay

The 11-plex assay is currently being validated for case work samples at AFDIL

Manuscript has been submitted

Additional multiplex mtSNP assays are being developed for other common HV1/HV2 types in collaboration with AFDIL

Overview

SNPs

Assay Platforms and Instrumentation

Multiplexing

U.S. Population Samples

Y Chromosome and Mitochondrial Markers

Results

mtSNP 11 plex

Y-SNP multiplexes

The Y Chromosome

60,000kb total size

The non-recombining region (NRY) consists of 95% of the Y chromosome

NRY is passed on as a block of information

Sequence variations in the NRY are due to mutation

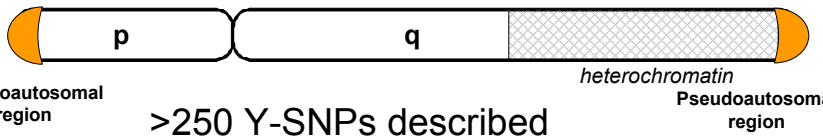
Potential for predicting geographical origin?

Forensic Utility of Y Chromosome SNPs

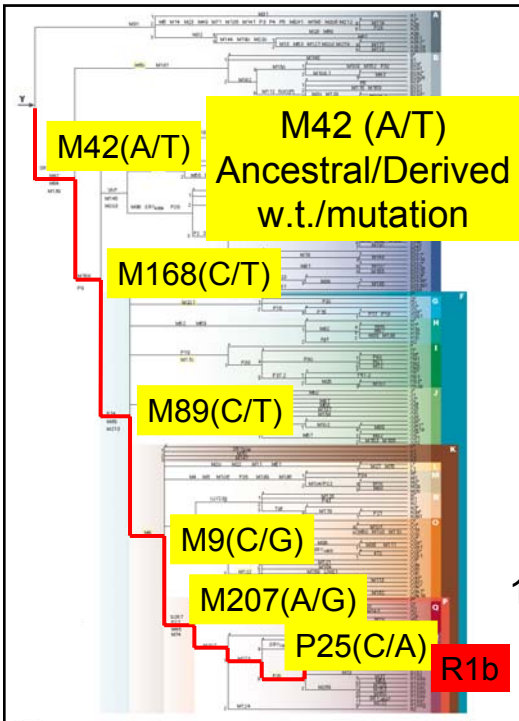
Y chromosome markers are useful in mixed male - female samples

Haplogroups are non-randomly distributed among populations therefore potential exists for predicting population of origin

Low mutation rate of SNPs $2e^{-8}$ per base per generation



>250 Y-SNPs described



The Y Chromosome Consortium Map (2003) Nat Rev Genet. 4 :598-612

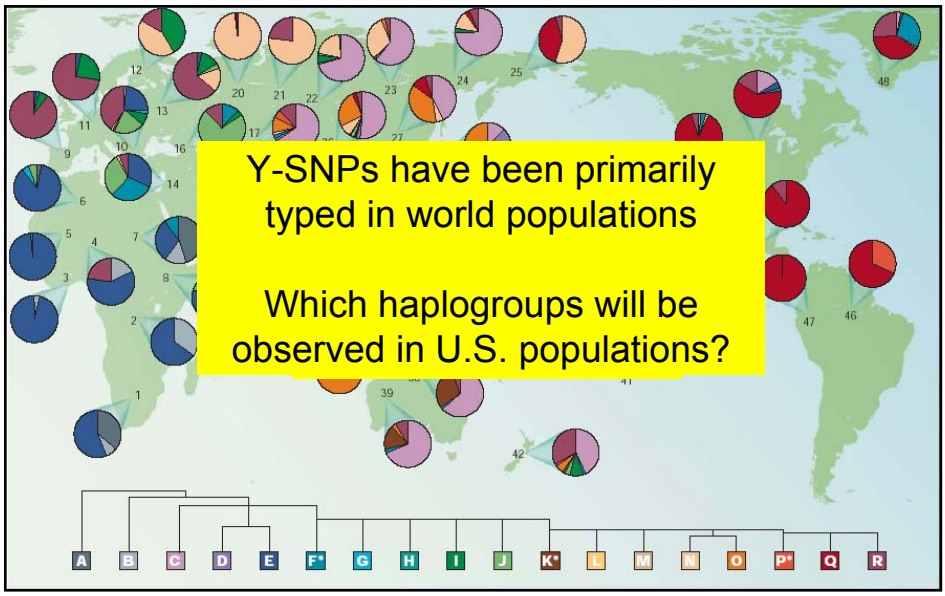
Tree contains over 250 Y-SNPs

Samples were typed for 48 world populations

18 main groups A-R

159 haplogroups defined

Global Distribution of Y Haplogroups



Y Chromosome Consortium (2003) Nat Rev Genet. 4 :598-612

Y-SNPs in U.S. populations

Which haplogroups will be observed?

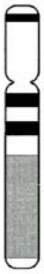
How specific will certain Y-SNPs be for a U.S. population group?

Forensic utility in comparison/addition to Y-STRs

Commercial kit (Marligen) 42 Y-SNPs

Medium sized multiplexes developed in-house (CE or MS)

Y-SNPs Typed at NIST



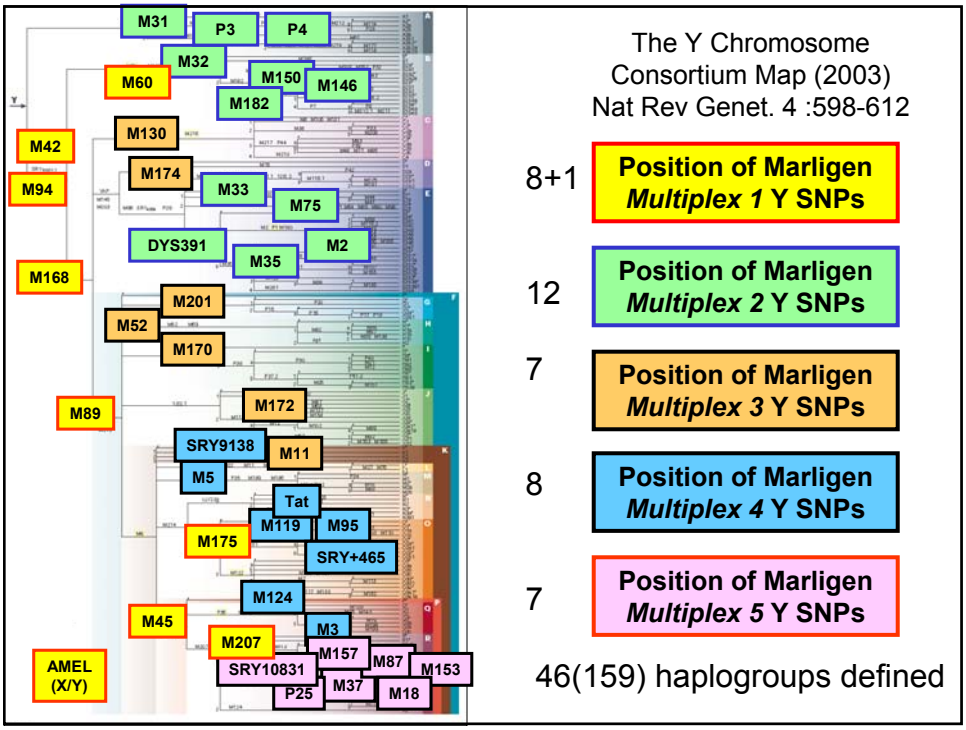
42 SNPs + Amelogenin present in 5 multiplexes
(commercially available kit from Marligen)

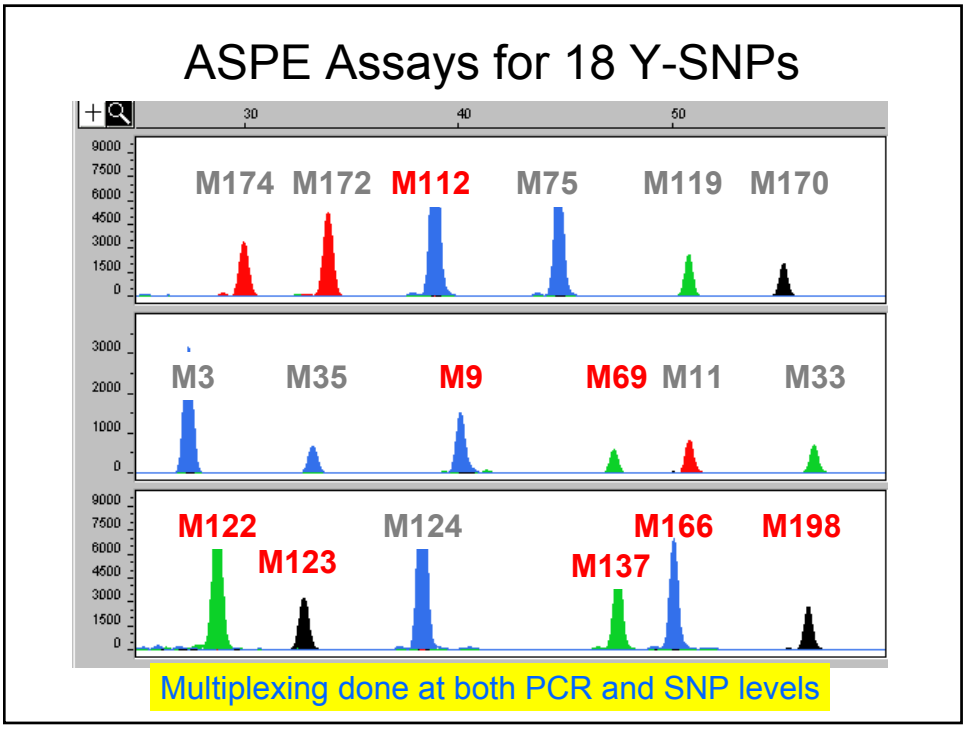
18 SNPs in 3 NIST-designed 6plexes (8 unique)
10 SNPs in 2 NIST-designed 5plexes (1 unique)

19 of the SNP sites overlapped…

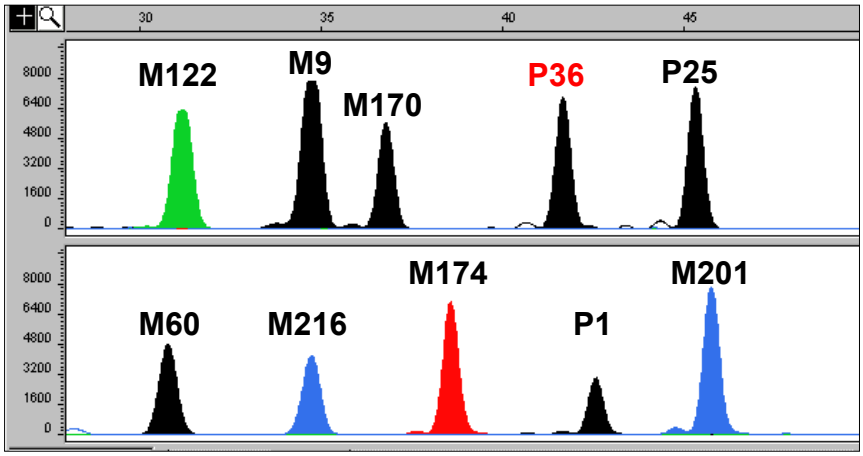
Resulting in a total of 51 Y-SNPs

115 African Americans
114 Caucasians
95 Hispanics (presently typed for 10 Y-SNPs)


M31
P3
P4
M32
M60
M150
M146
M182
M130
M42
M174
M94
M33
M75
DYS391
M35
M2
M168
M201
M52
M170
M89
M172
SRY9138
M11
M5
Tat
M119
M95
M175
SRY+465
M124
M45
M3
M207
M157
M87
M153
AMEL (X/Y)
SRY10831
M37
M18
P25
The Y Chromosome Consortium Map (2003) Nat Rev Genet. 4 :598-612
8+1
Position of Marligen Multiplex 1 Y SNPs
12
Position of Marligen Multiplex 2 Y SNPs
7
Position of Marligen Multiplex 3 Y SNPs
8
Position of Marligen Multiplex 4 Y SNPs
7
Position of Marligen Multiplex 5 Y SNPs
46(159) haplogroups defined


ASPE Assays for 18 Y-SNPs
M174 M172 M112 M75 M119 M170
M3 M35 M9 M69 M11 M33
M122 M123 M124 M137 M166 M198
Multiplexing done at both PCR and SNP levels

ASPE Assays for 10 Y-SNPs



Equimolar PCR primer concentration (5plex)
Empirical balancing of extension primers

Summary of Y-SNP Data

(115 African Americans and 114 Caucasians)

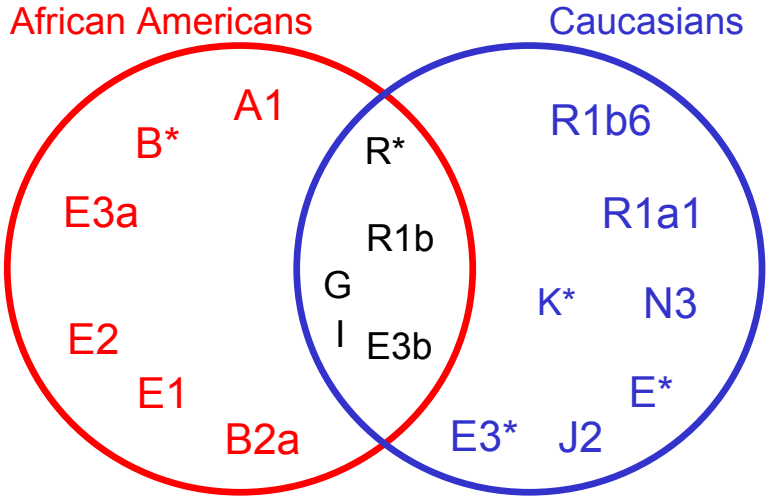
A total of 20 ng of genomic DNA was consumed for the 10 multiplexes

18 out of 46 haplogroups observed

Over 99 % success rate for allele calls (both methods)
Variation was observed in 24 of the 51 Y-SNPs

100% concordance for the 19 overlapping markers (>3,800 allele calls)

Observed Haplogroups in Two U.S. Populations

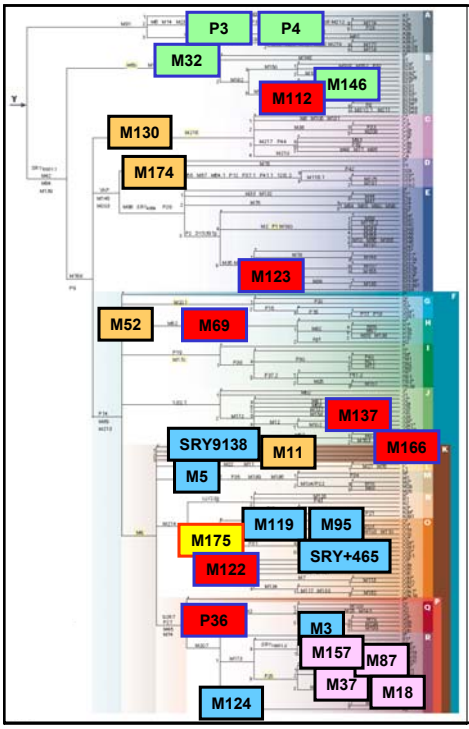


18 total Hgs; 5 shared

Variation was not observed for 27 Y-SNPs

(in AA and CAUC populations)

M175 +/-	M119 A/C	M37 C/T
M146 A/C	M124 C/T	M87 T/C
M32 T/C	M3 C/T	M69 T/C
P3 (C/T)	M5 C/T	M112 G/A
P4 (G/A)	M95 C/T	M122 T/C
M11 A/G	SRY465 C/T	M123 G/A
M130 C/T	SRY9138 C/T	M137 T/C
M174 T/C	M157 A/C	M166 G/A
M52 A/C	M18 -/+	P36



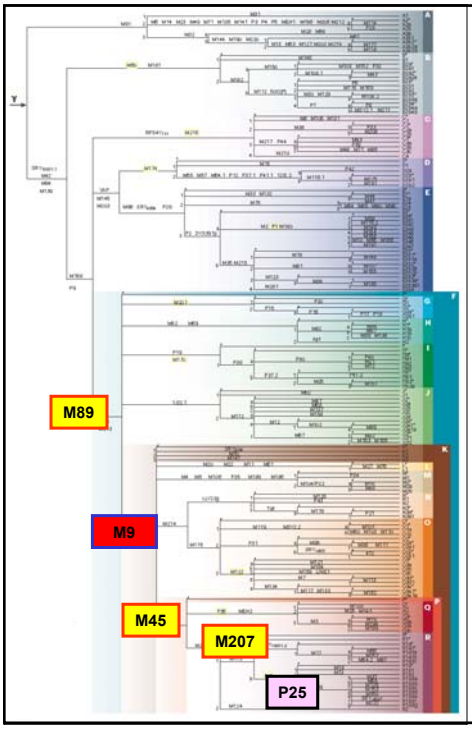
Variation was not observed for 27 Y-SNPs

(in AA and CAUC populations)

P3 (A2)
P4 (A2)
M32 (A3a)
M146 (B1)
M112 (B2b)
M130 (C)
M174 (D)
M123 (E3b)
M52 (H)
M69 (H)
M137 (J2c)
M166 (J2f2)
SRY9138 (K1)
M11 (L)
M5 (M)
M175 (O)
M119 (O1)
M95 (O2a)
SRY+465 (O2b)
M122 (O3)
P36 (Q)
M3 (Q3)
M124 (R2)
M157 (R1a1b)
M87 (R1a1c)
M37 (R1b3a)
M18 (R1b1)

Y-SNPs derived at greater than 25 % in more than one population

Locus	All	AA	Cauc	Hisp	Hap
M207 A/G	0.46	0.27	0.65	na	R
M45 G/A	0.46	0.27	0.64	na	P-R
M89 C/T	0.64	0.32	0.96	na	F-R
P25 C/A	0.47	0.30	0.57	0.53	R1b
M9 C/G	0.53	0.31	0.65	0.64	K-R



Y-SNPs observed at greater than 25 % in more than one population

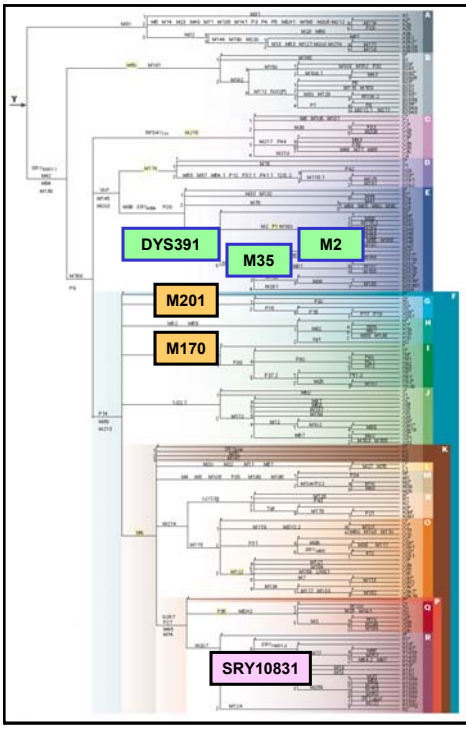
M45 (P-R)
M89 (F-R)
M9 (K-R)

M207 (R)
P25 (R1b)

Derived in more than one population

Locus	All	AA	Cauc	Hisp	Hap
M2 A/G	0.23	0.58	not obs	0.08	E3a
DYS391 C/G	0.31	0.60	0.04	na	E3
M170 A/C	0.10	0.04	0.21	0.04	I
M35 G/C	0.02	0.02	0.03	na	E3b
M201 G/T	0.03	0.01	0.03	0.05	G
SRY10831 A/G	0.03	0.01	0.05	na	R1a

M2 is not derived in the U.S. Caucasian population

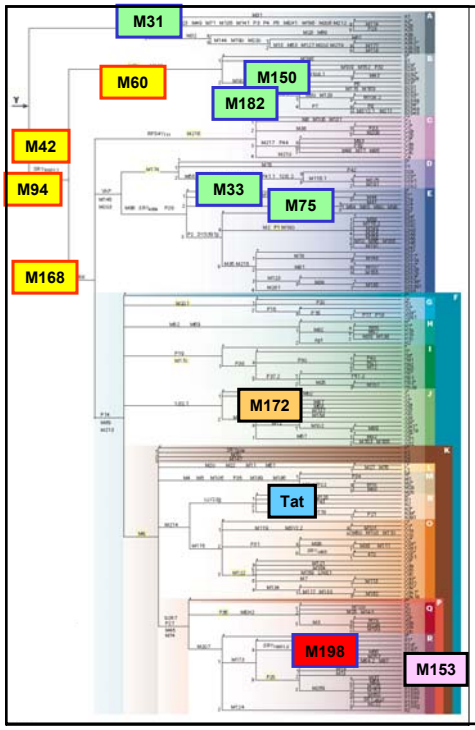


Derived in more than one population

M2 (E3a)
DYS391 (E3)
M170 (I)
M35 (E3b)
M201 (G)
SRY10831 (R1a)

Low frequency and population specific

Locus	All	AA	Cauc	Hisp
M168 C/T	0.01	0.03	not obs	na
M42 A/T	0.04	0.01	not obs	na
M60 -/+	0.01	0.02	not obs	not obs
M94 C/A	0.01	0.01	not obs	na
M150 C/T	0.01	0.01	not obs	na
M182 C/T	0.01	0.01	not obs	na
M31 G/C	0.01	0.01	not obs	na
M33 A/C	0.01	0.03	not obs	na
M75 G/A	0.01	0.03	not obs	na
M172 T/G	0.03	not obs	0.05	na
M198 C/T	0.03	not obs	0.05	na
Tat T/C	0.01	not obs	0.01	na
M153 T/A	0.01	not obs	0.01	na

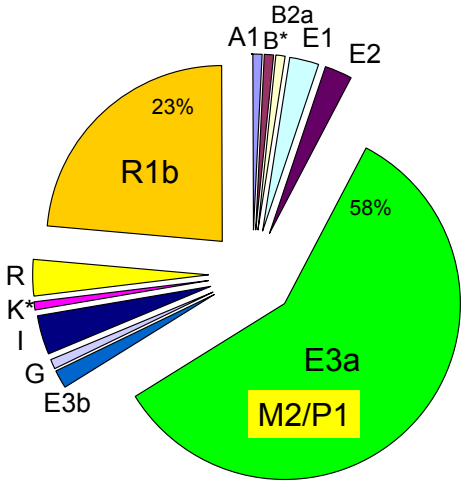


Low frequency and population specific

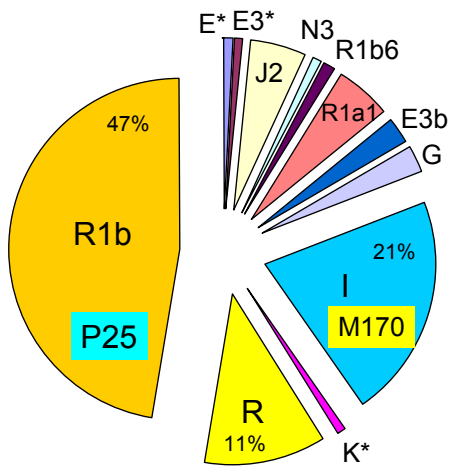
AA	Cauc
M168 (C-E)	M172 (J2)
M42 (B)	M198 (R1a1)
M60 (B)	TaT (N3)
M94 (B)	M153 (R1b3d)
M150 (B2a)	
M182 (B2a)	
M31(A1)	
M33 (E1)	
M75 (E2a)	

Derived less than 5 %

Y-SNP haplogroups for 115 African Americans

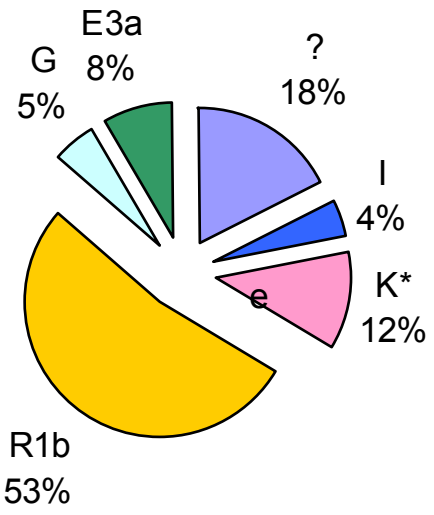


Y-SNP haplogroups for 114 Caucasians



18 different haplogroups observed in 229 males

Y-SNP haplogroups
for 95 Hispanics



Only 10 SNPs were typed for these 95 samples

Forensic Utility
51 Y-SNPs versus 1 Y-STR

For N = 211 male samples

	51Y-SNPs	Y-STR DYS464
Amount of sample consumed	10ng	1ng
Number for types observed	**18**	**62**
Analysis	Multiple	1 reaction
Degraded samples	+	?

As a stand alone forensic assay
1 Y-STR is better than 51 Y-SNPs

Conclusions

- Full concordance was observed between hybridization and primer extension technologies on 18 different Y-SNPs (>3,800 allele calls)
- Caucasian admixture was observed with our African American population (Hg R and R1b in ~30%)—agrees with Kayser *et al.* (2003) *Genome Res.* 13:624-634 done with 9 Y-STRs
- Y-SNPs may have limited value for ethnic differentiation in U.S. populations
  - One exception: M2 not in Caucasians

Issues with Y-SNP P25

Initially when typing P25 with the Marligen kit the derived allele (A) was not observed

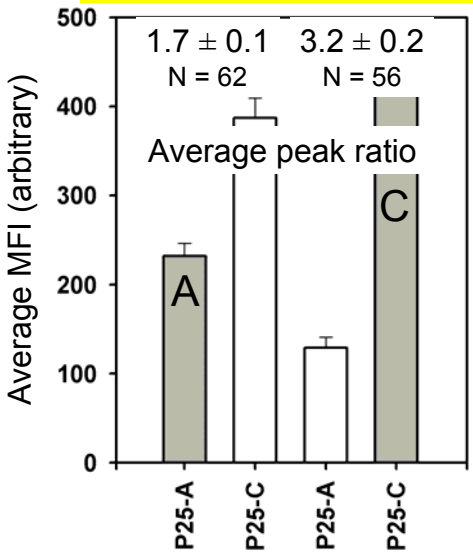
Alan Redd (Univ of AZ) informed us that P25 is a multi copy locus

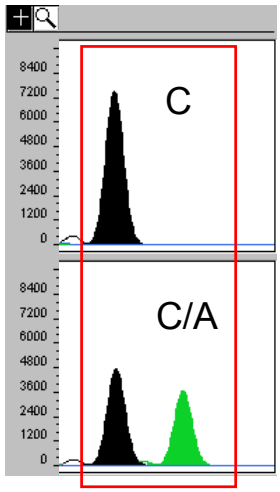
After further review of our data we were able to make correct allele call for the P25 marker based on signal intensity ratio

BLAST results indicate that the region surrounding P25 is present 3 times on the Y chromosome

Typing P25 by ASH or ASPE

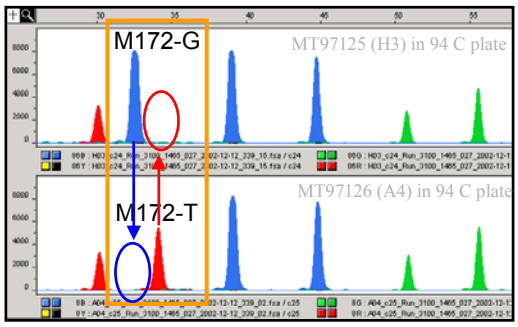
Solid bar = correct allele call





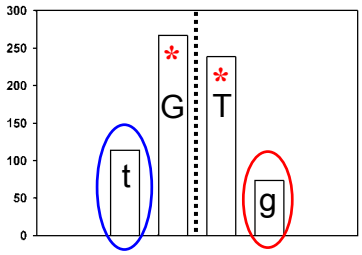
Signal to noise Ratio

ASPE



Alternate allele is not observed

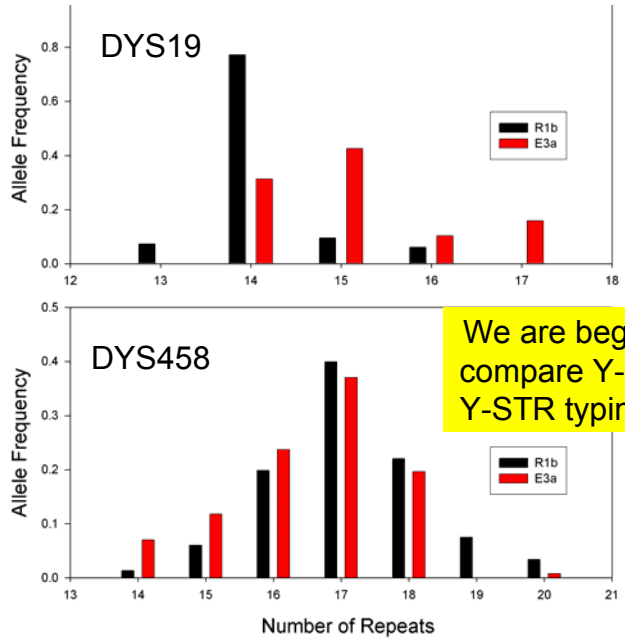
Hybridization



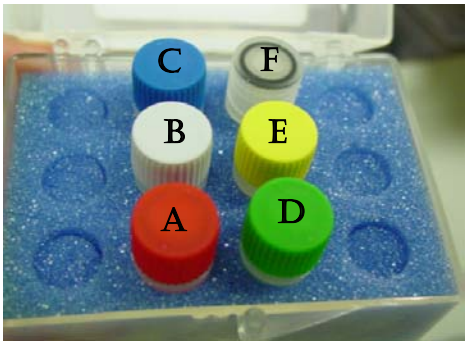
267	73.5
114	238.5

S/N = 2.3 and 3.2

Y-STR Alleles Observed with Various Y-SNP Haplogroups



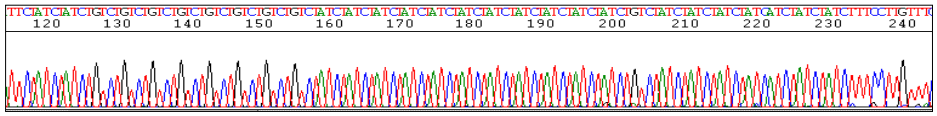
We are beginning to compare Y-SNP and Y-STR typing results



Sequencing Results for 23 Y STR Loci

50 Y SNP Loci Typed

DYS390 (forward) E



Y SNP Results on SRM 2395

SRM 2395	AMEL	M207	M45	M89	DYS391	M2	M170	M172	M201
		(A/G)	(A/G)	(C/T)	(C/G)	(A/G)	(A/C)	(G/T)	(G/T)
Component A	XY	**G**	**A**	T	C	A	A	T	G
Component B	XY	A	G	T	C	A	A	**G**	G
Component C	XY	A	G	**C**	**G**	**G**	A	T	G
Component D	XY	A	G	T	C	A	A	T	**T**
Component E	XY	A	G	T	C	A	**C**	T	G
Component F	XX								

SRM components are all distinguishable from one another with these Y SNPs

50 Y SNPs measured across all samples

Acknowledgments

Funding:

U.S. National Institute of Justice

Interagency Agreement between NIJ and NIST Office of Law Enforcement Standards

John Butler

Collaborators
Thomas Parsons, Rebecca Hamm and Mike Coble (AFDIL)
David Carlson (Marligen)
Mike Hammer and Alan Redd (U of AZ)

Jan Redman

Margaret Kline