Different Classes of SNPs for Human Identification

Kristen Lewis O'Connor, Ph.D.

Applied Genetics Group
National Institute of Standards and Technology
Gaithersburg, Maryland

HITA/AABB SNP Workshop
San Antonio, Texas
October 10, 2010

NIST
National Institute of Standards and Technology

Types of Genetic Variation

Length Variation

- short tandem repeats (STRs)

  CTAGTCGT**(GATA)(GATA)(GATA)**GCGATCGT

Sequence Variation

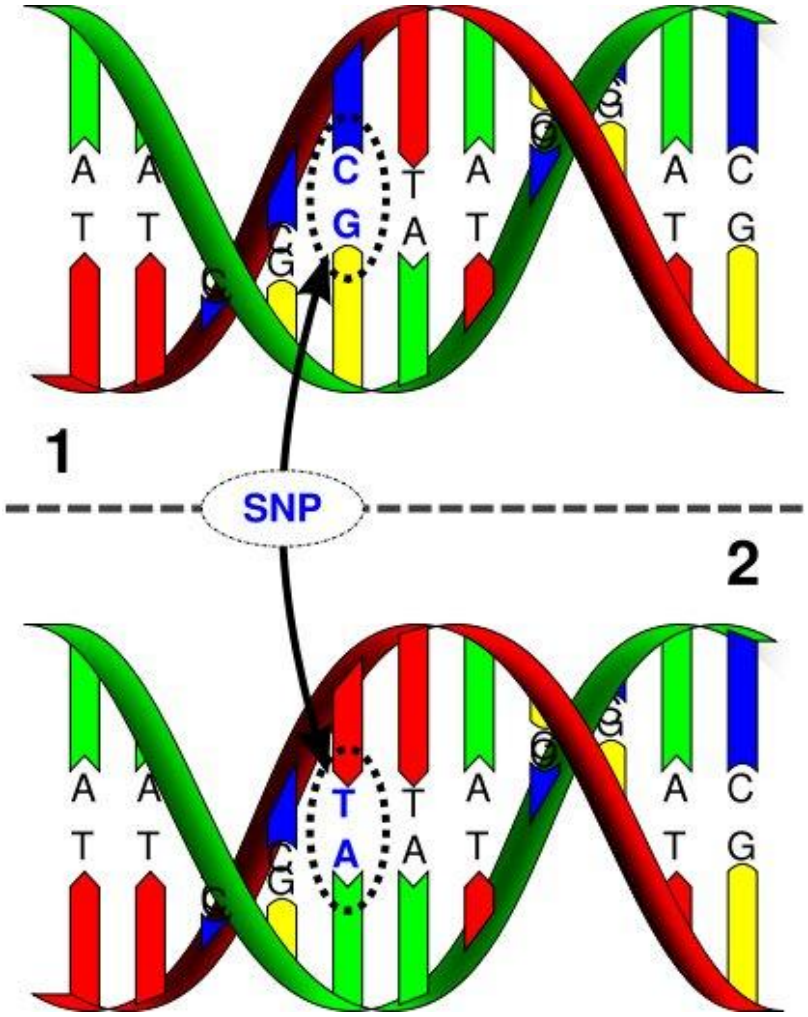
- insertions/deletions
- **single nucleotide polymorphisms** (SNPs)

  GCTAGTCGATGCTC**(G/A)**GCGTATGCTGTAGC

Structural Variation

- duplications/deletions (copy number variants, CNVs)
- genomic rearrangements (translocation, inversion, transposition)

SNP



Typically biallelic

http://www.science.marshall.edu/murraye/341/snps/Human%20Genetics%20MTHFR%20SNP%20Page.html

Advantages/Benefits

- Small amplicon size better for analyzing degraded samples
- Abundant in the human genome (avg. 1 SNP/1000 bases)
- Lower mutation rate compared with STRs (10^{-8} vs. 10^{-3})
- Can provide specific information
  - Identity, lineage, ancestry, or phenotype
- Multiple typing platforms provide assay design flexibility and throughput

Limitations/Challenges

- SNPs are not currently represented in national DNA databases
- No widely established core loci in U.S.
- Mixture resolution issues/interpretation
- Multiple typing platforms make validation and concordance difficult
- Lack of available commercial kits
- Statistical issues with using genome-wide SNPs in linkage disequilibrium
- Sensitivity
- Cost

SNP Typing Platforms

- Serial vs parallel throughput?
- What are the DNA template requirements?
- What is cost per SNP?
- Is the instrumentation/technique common to a forensic laboratory?

SNP Typing Platforms

- RT-PCR (TaqMan, Light Cycler, Molecular Beacon)
- ASPE (SNaPshot, MALDI, Fluorescence Polarization)
- Mass Spectrometry (Electrospray, IBIS)
- Sequencing
- Flow Cytometry (Luminex)
- Pyrosequencing
- Ligation (SNPlex/GenPlex, Illumina)
- Invader assay
- ARMS assay (Forensic Science Service - UK)
- RFLP

Sensitivity, multiplexing, accurate typing

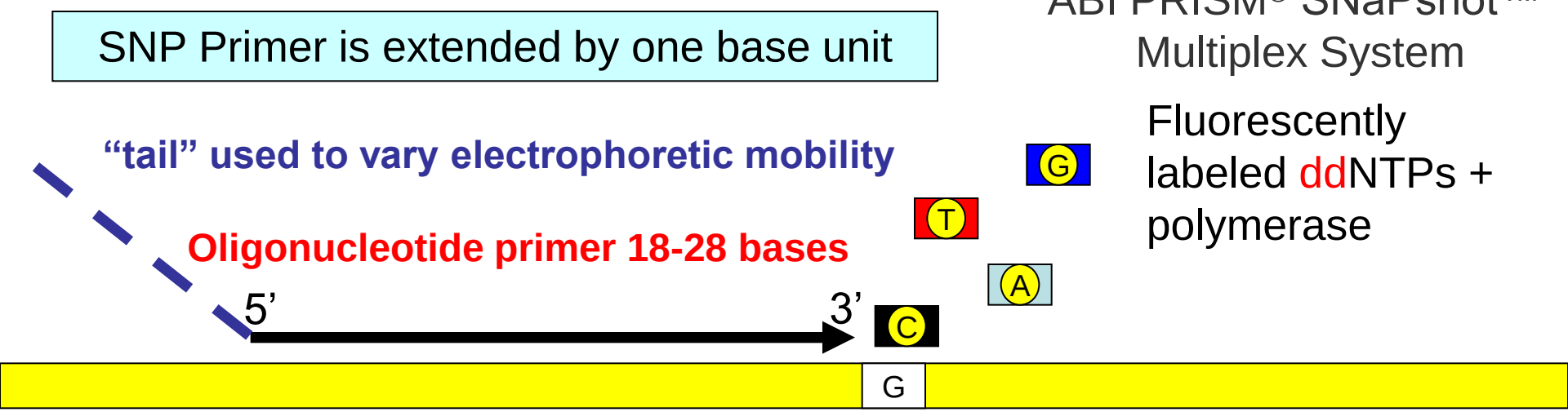
Budowle, 2004 FSI 139-142; Sobrino et al., 2005 FSI (epub); Dixon et al., 2005 FSI (epub)

Allele-Specific Primer Extension

SNP Primer is extended by one base unit

ABI PRISM® SNaPshot™ Multiplex System

Fluorescently labeled ddNTPs + polymerase

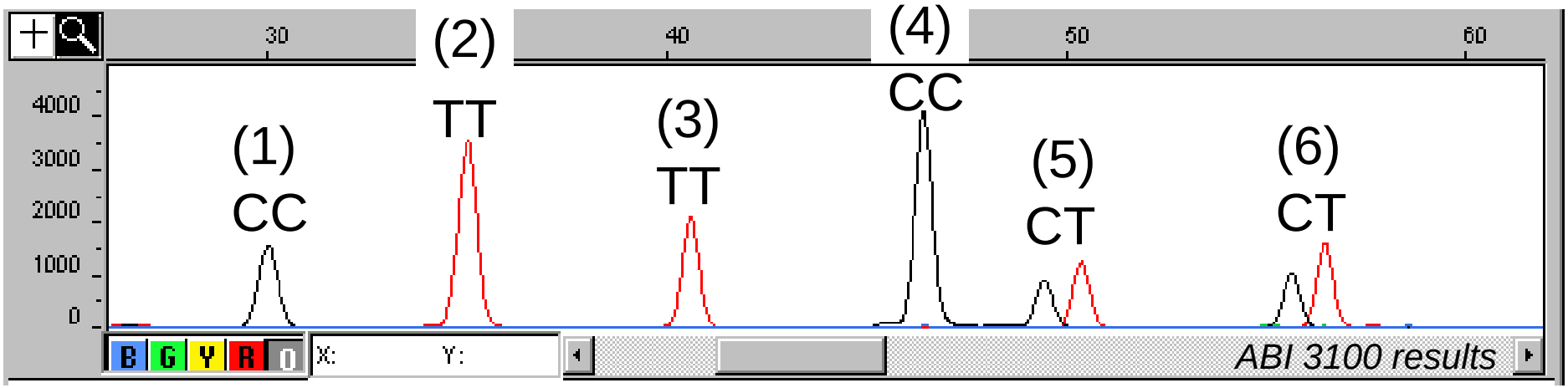


PCR Amplified DNA Template

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

6-plex SNP Assay



CHR: 2 13 6 6 15 17

Extension primers for 6-plex

(1) TTTTTAGCTCCTAATTTCTTGATGGG
(2) TTTTTTTTTTTCATCTGATGCCATGAGAAAGC
(3) TTTTTTTTTTTTTTTGTTCTGCTTTAATACAAAACCAG
(4) TTTTTTTTTTTTTTTTTTTTTTTTATAAAGGGCAGAATGAGGATTA
(5) TTTTTTTTTTTTTTTTTTTTTTTTTTTAGAAAGTATCTTGCAAAAGGTCCA
(6) TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATAATCACAGCTTTTTTCTCCCAA

SNaPshot assays designed by Peter Vallone, NIST

SNP Categories

- Individual Identification SNPs (IISNPs): SNPs that collectively give very low probabilities of two individuals having the same multi-locus genotype
- Ancestry Informative SNPs (AISNPs): SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world
- Phenotype Informative SNPs (PISNPs): SNPs that provide a high probability that the individual has particular phenotypes, such as a particular skin color, hair color, eye color, etc.
- Lineage Informative SNPs (LISNPs): Sets of tightly linked SNPs that function as multi-allelic markers that can serve to identify relatives with higher probabilities than simple bi-allelic SNPs

Budowle, B. and van, Daal. A. (2008) Forensically relevant SNP classes, *Biotechniques 44*, 603-8, 610.
Butler, J.M., Budowle, B., Gill, P., Kidd, K.K., Phillips, C., Schneider, P.M., Vallone, P.M., Morling, N. (2008) Report on ISFG SNP panel discussion. *Forensic Science International: Genetics Supplement Series (Progress in Forensic Genetics 12*) 1: 471-472.

Individual Identification SNPs

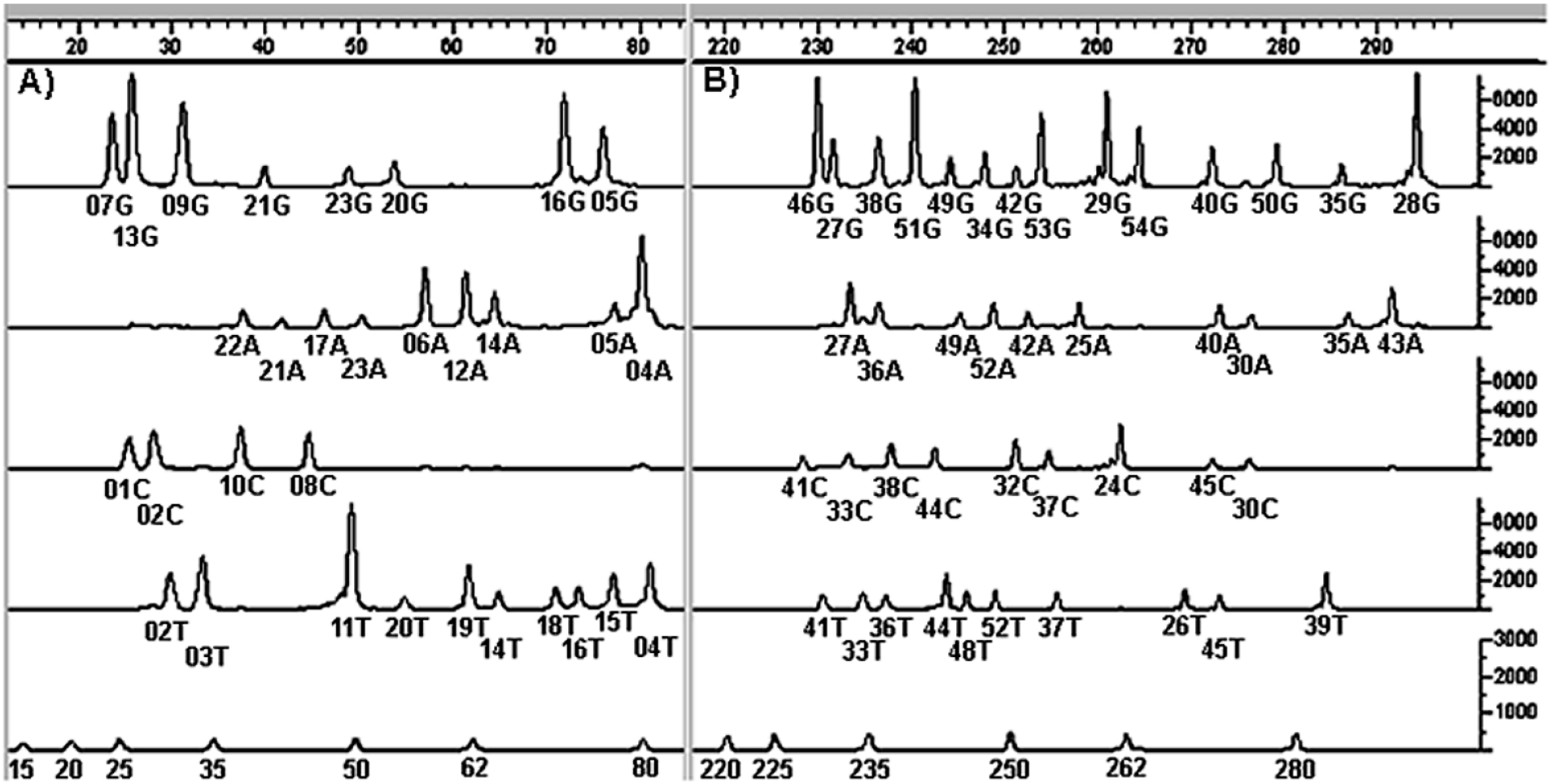
- Use for individual identification of a sample
- Power of Discrimination – how many SNPs are needed to match STRs?
- Can the assay amplify > 30 loci using a small amount of template DNA?
- Use on a degraded sample
- Issues with a mixture

General Criteria

- Low F_{ST} (not population specific)
- No linkage disequilibrium between SNPs or CODIS loci
- Amplicon size < 120 bp
- Minimum 30% heterozygosity
- Minimum distance of 100 kb between SNPs and neighboring genes

SNP*for*ID 52 plex

Sanchez, J. J., Phillips, C., Borsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, Schneider, P. M., Carracedo, A., and Morling, N. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis 27*, 1713-1724.



59 to 115 bp
500 pg
2 ASPE reactions

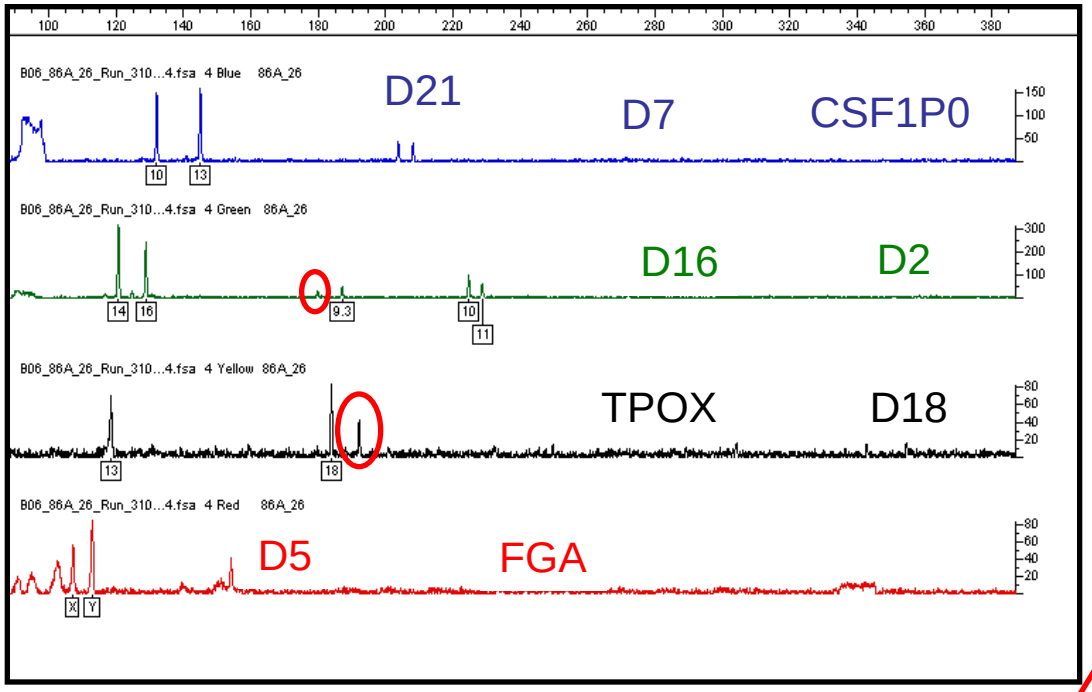
Figure 1. Analysis of the 52 SNP-plex assay. (A) Results of the first injection (23 SNPs). (B) Results of the second injection (29 SNPs).

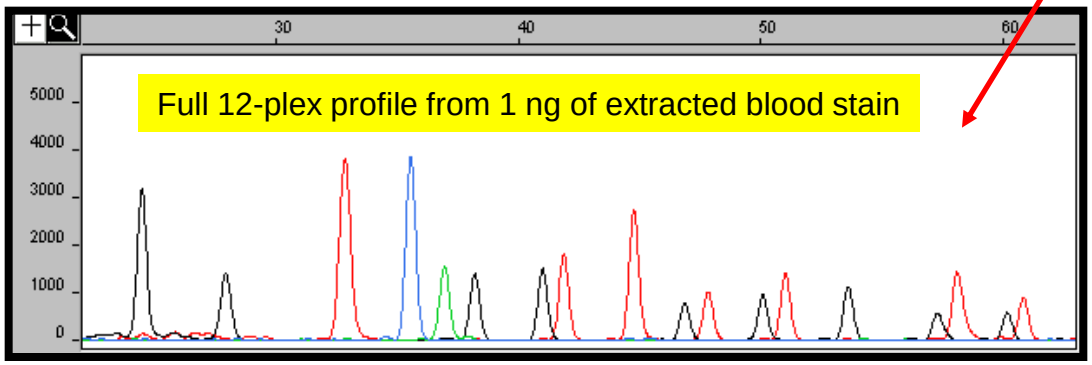
Probability of Identity

STRs vs. SNPs

Set of Loci	# Loci	Probability of Identity	
CODIS 13	13	4.58×10^{-16}	STRs (combined U.S. population)
PowerPlex16	15	2.87×10^{-19}	
Identifiler	15	5.46×10^{-19}	
ESI 16/NGM	15	2.58×10^{-20}	
ESI 17/NGM SElect	16	1.61×10^{-22}	
48plex assay	48	9.6×10^{-18}	SNPs (combined European population)
52plex assay	52	5.06×10^{-21}	

Typing an Aged Blood Stain



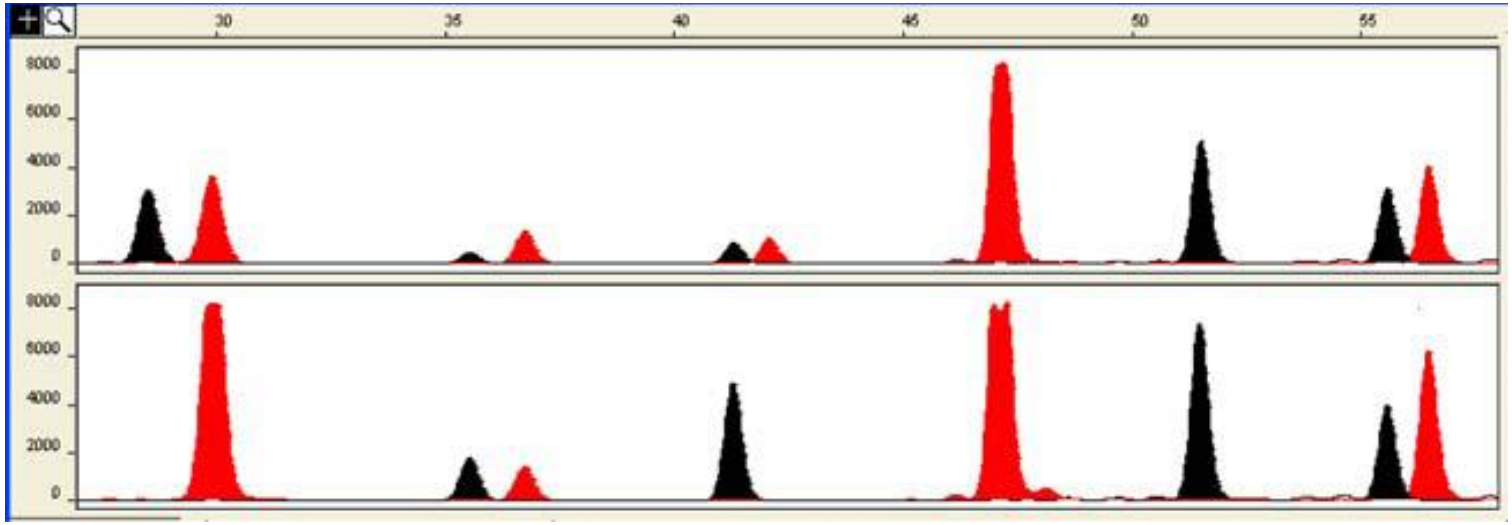


Identifiler genotyping result from a blood stain aged 15 years stored at room temperature. (stored on 903 paper, Chelex extracted)

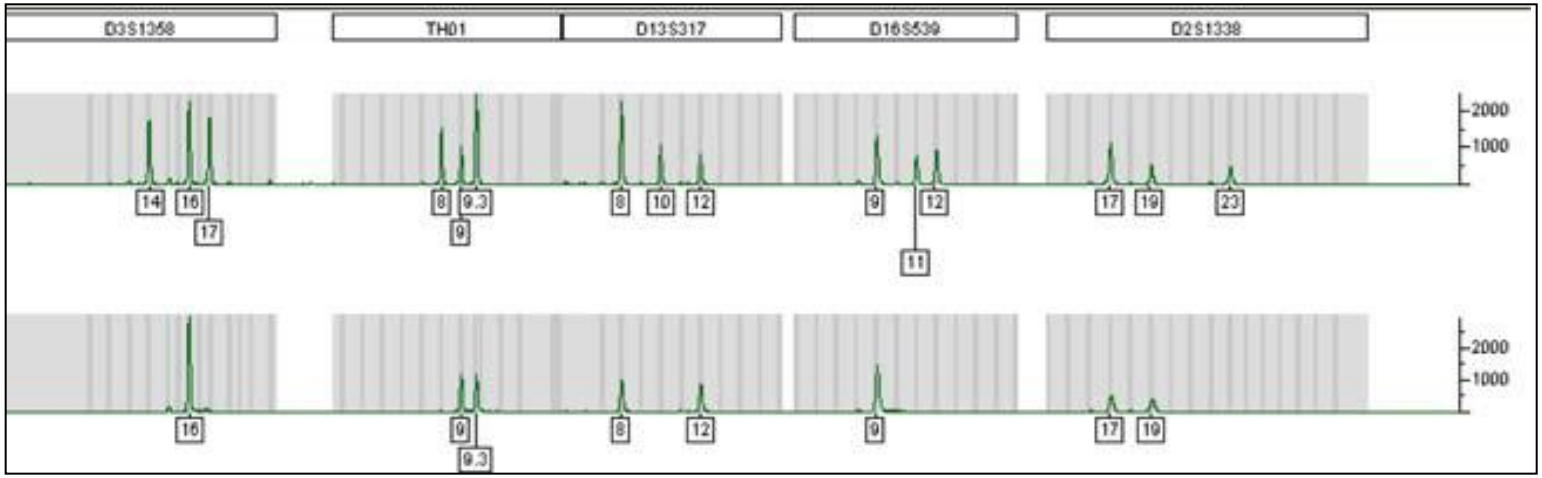
The same sample extract as above typed by the 12-plex SNP assay.

11 different samples that gave partial profiles with Identifiler gave full profiles typed with the 12-plex assay.

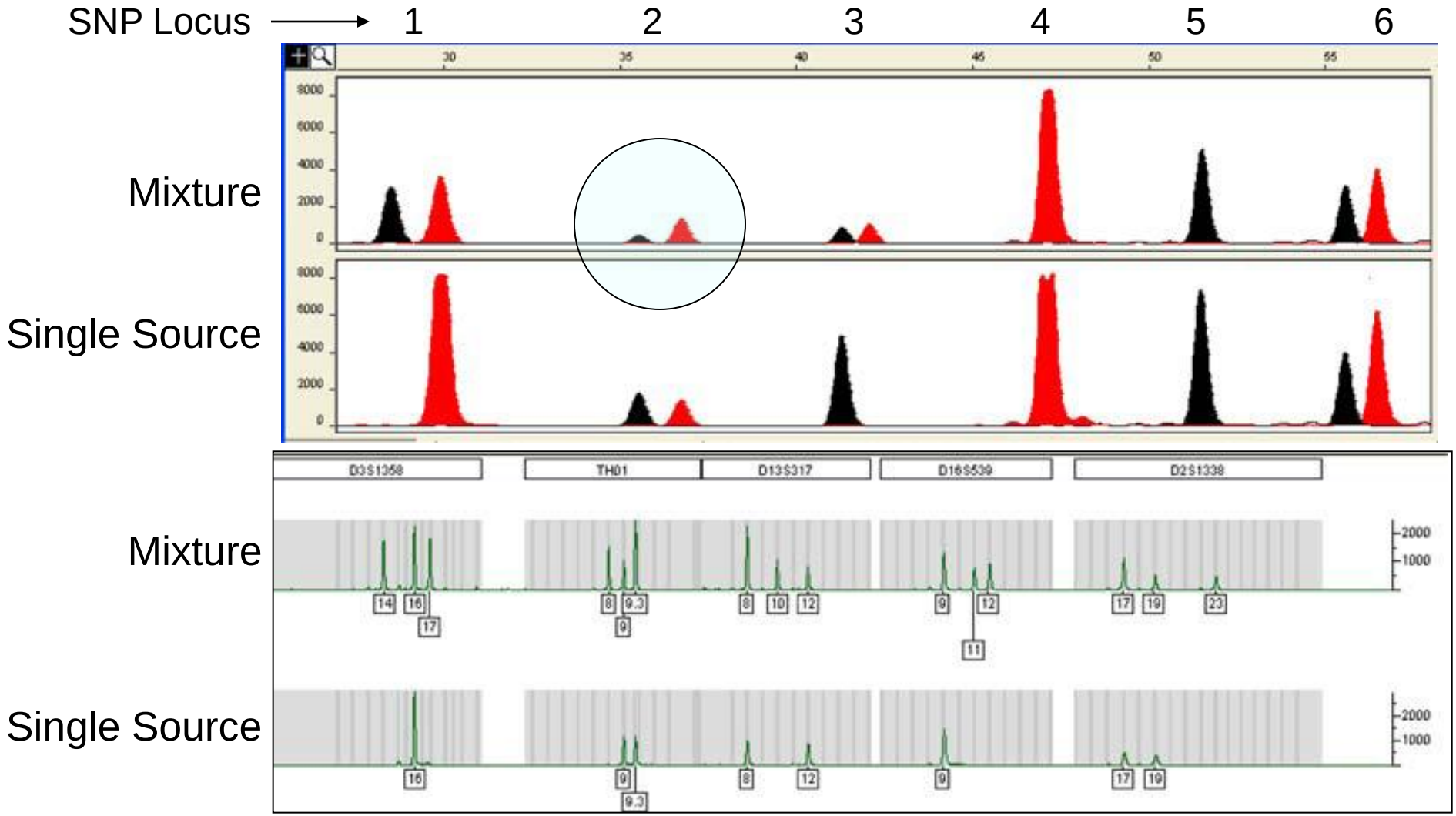
SNP Locus → 1 2 3 4 5 6



Mixture?



Butler, J.M., Coble, M.D., Vallone, P.M. (2007) STRs vs SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci. Med. Pathol.* 3: 200-205.



- Only SNP locus 2 (top panel) has an allele imbalance suggesting that a possible second contributor is present
- 5 STR loci shown from the green dye channel of the Identifiler kit contain three alleles making mixture detection straightforward

Butler, J.M., Coble, M.D., Vallone, P.M. (2007) STRs vs SNPs: thoughts on the future of forensic DNA testing. *Forensic Sci. Med. Pathol.* 3: 200-205.

Challenges of Using IISNPs for Forensic Testing

- U.S. and international databases consist of STR profiles
  - Is there a benefit to changing the DNA typing technology for databanking and routine casework?
- Mixture analysis using genome-wide arrays
  - Detection is possible but interpretation is not here yet
- SNPs in linkage disequilibrium: match probability calculations?
- Sensitivity
  - Genome-wide arrays require >500 ng DNA
  - Non-biased whole genome amplification is necessary but not here yet
- Cost
  - Approximately $500 per sample for genome-wide arrays
  - Arrays cost the same amount for typing 1,000 SNPs or 1 million SNPs

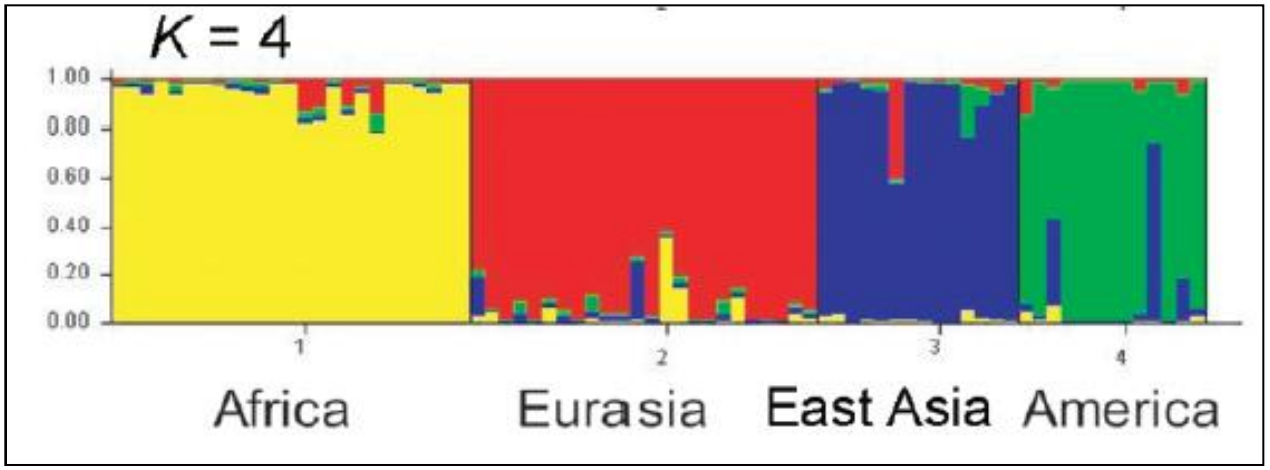
Ancestry Informative SNPs

- SNPs that are population specific
  - Opposite of Individual ID SNPs…
- Estimate genetic ancestry
  - Biogeographical population structure
- Also called ASMs, AIMs
  - Ancestry sensitive makers
  - Ancestry informative markers
  - Can also include Y and mito SNPs

Source of AISNPs

- Genotype world population samples on a high throughput array (e.g. Affymetrix 500k SNP)
  - Use bioinformatics determine a minimal set of SNPs for prediction purposes
- Scanning SNPs from the literature & databases
- SNP with higher F_{ST} values
  - F_{ST} is a statistic that describes the proportion of variance within a species that is due to population subdivision

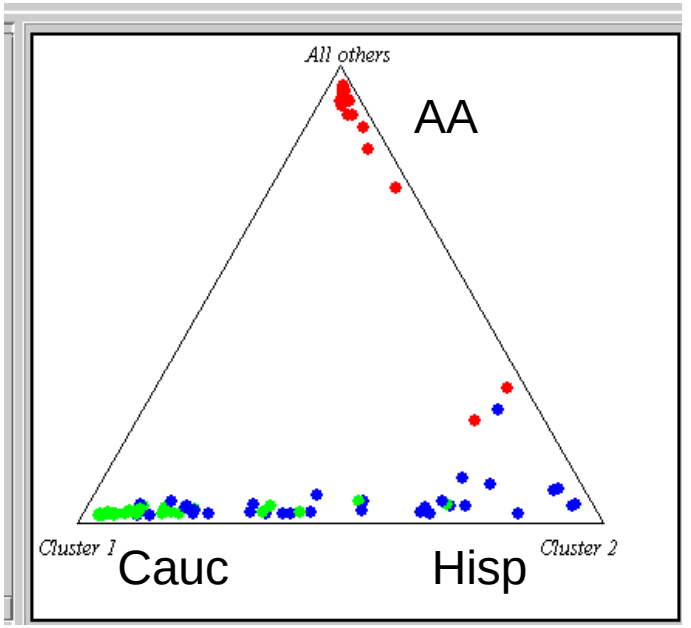
Ancestry Informative SNPs



Individuals of varying ancestry exhibit similar genetic profiles

With an appropriate set of SNPs, genetic ancestry can be estimated

Using Structure software or
Principle Component Analysis (PCA)
http://pritch.bsd.uchicago.edu/structure.html



Data from Lao et al. AJHG 2006 78: 680-690

Phenotype Informative SNPs

- Eye Color

Wide range of pigmentation in humans

- Hair Color
- Multiple genes involved (~100), complex phenotype
- Some key pigmentation genes have been characterized
- Gene discovery and characterization is ongoing

Tully G. Genotype versus phenotype: human pigmentation. Forensic Sci Int Genet. 2007 1: 105-10.

Utility of PISNPs for Forensics

- Speed of test is important
- Use minimal sample
- Predict an observable trait – definitively
- Should not be predictive of disease

Recent Work on PISNPs

- Kayser M., Schneider P.M. (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci. Int. Genet. 3(3):154-61

- Liu F., et al. (2009). Eye color and the prediction of complex phenotypes from genotypes, Curr. Biol. 19:R192–R193

- Walsh S., et al. IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci. Int. Genet. (in press)

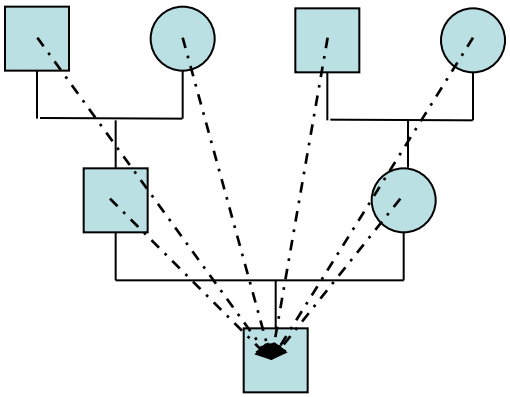
- Mengel-From J., et al. Human eye colour and HERC2, OCA2 and MATP. Forensic Sci. Int. Genet. (in press)

Lineage Informative SNPs

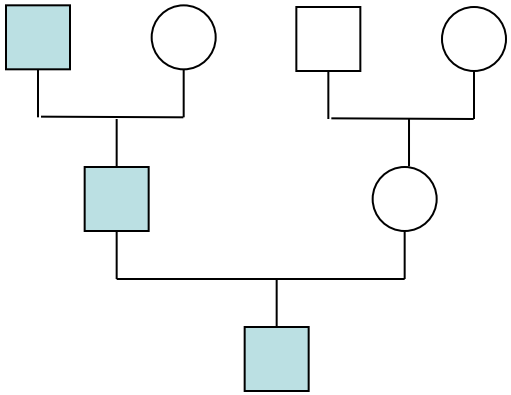
- Much of the LISNP literature focused on Y-chromosome and mitochondrial DNA SNPs
- With genome-wide arrays, autosomal SNP typing for lineage analysis is possible
- Lack of recombination
- Useful in evolutionary studies and kinship analysis

Different Inheritance Patterns

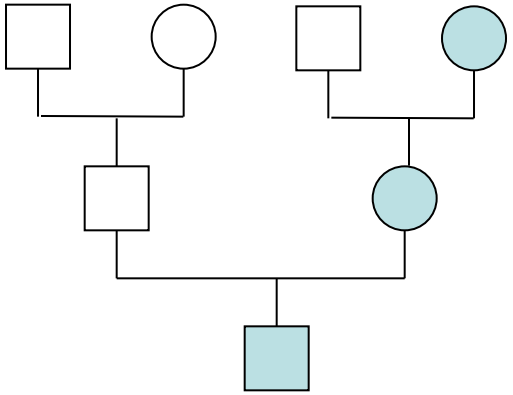
Lineage Markers



Autosomal
(passed on in part, from all ancestors)



Y-Chromosome
(passed on complete, but only by sons)



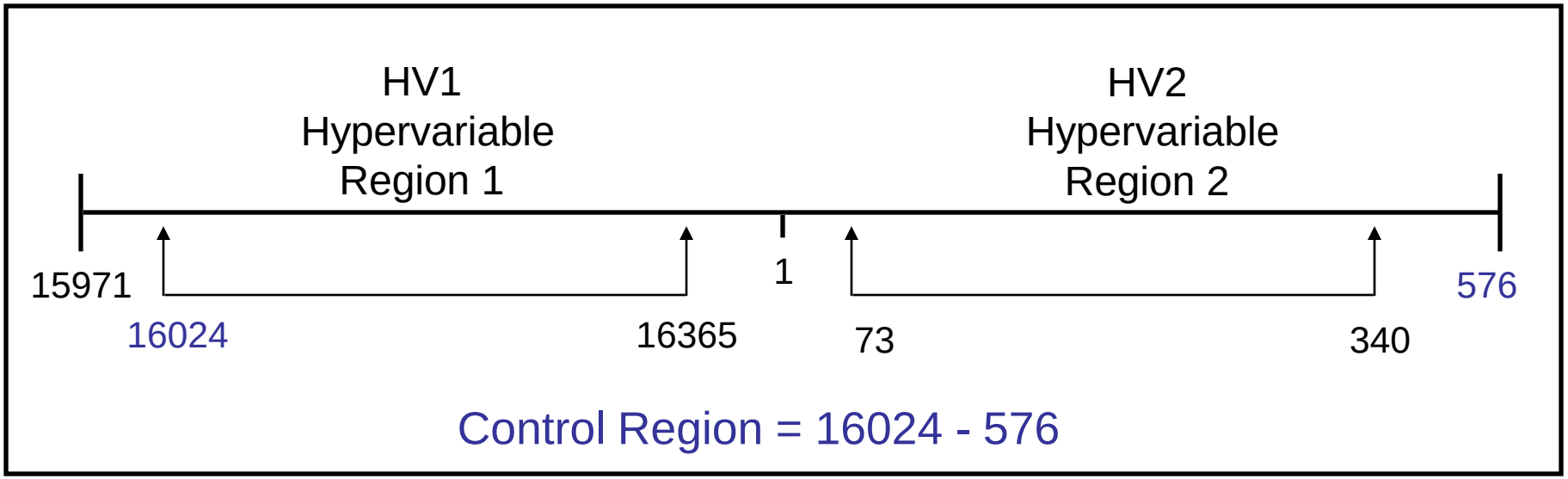
Mitochondrial
(passed on complete, but only by daughters)

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.1, ©Elsevier Science/Academic Press

Mitochondrial SNPs

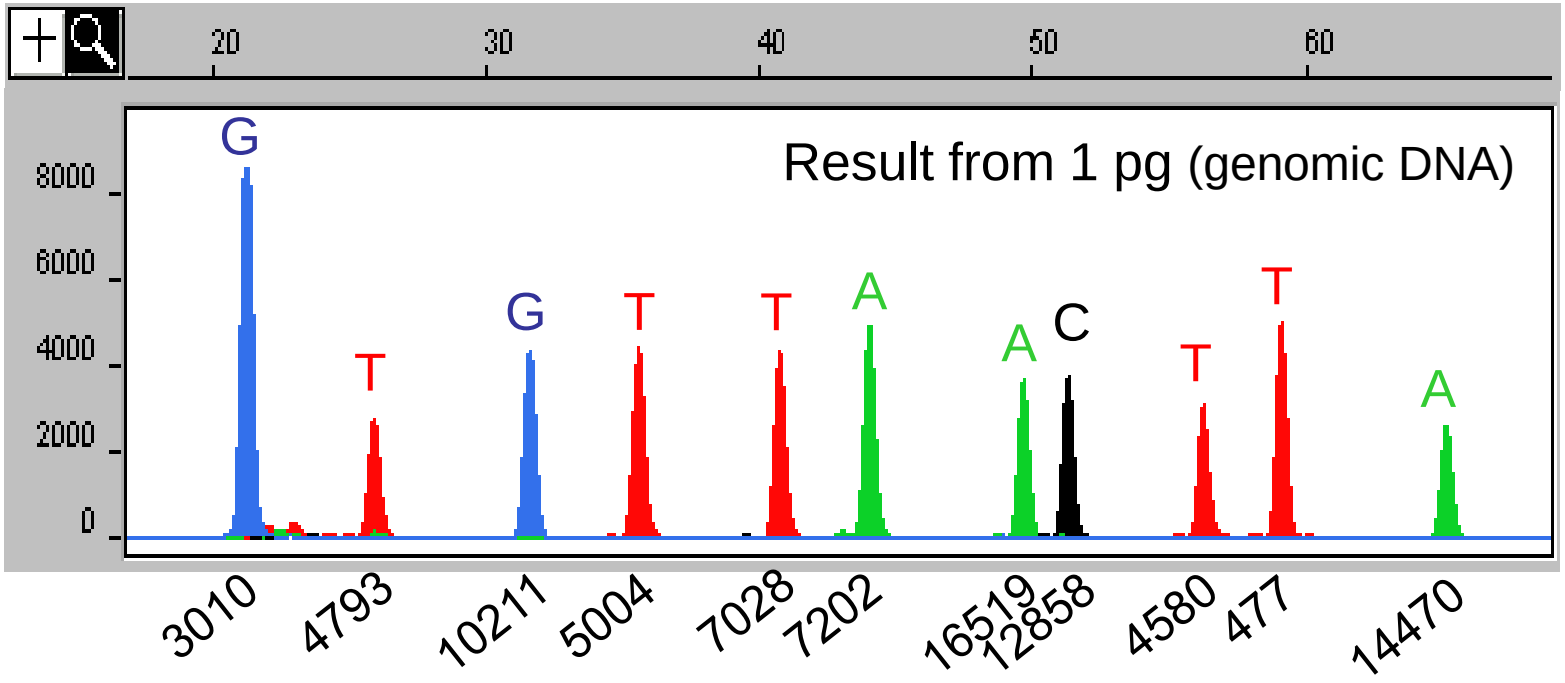
- HV1/HV2 sequencing
- LINEAR ARRAY mtDNA HVI/HVII Region-Sequencing Typing Kit (Roche Applied Science)
- For resolving a common mitochondrial HV1/HV2 haplogroup
- Mitochondrial haplogroup designation

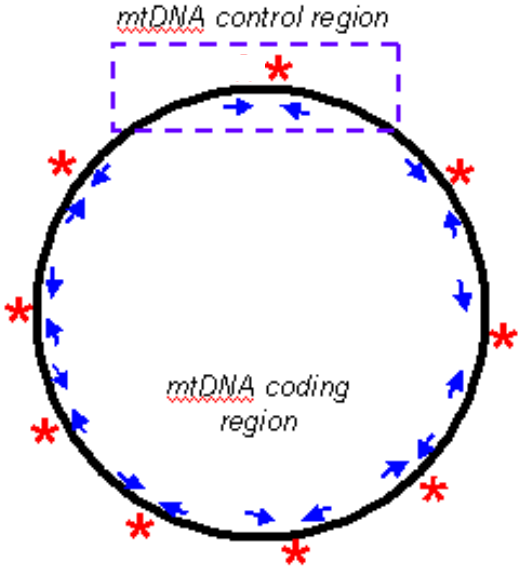
The Current mtDNA Amplification & Sequencing Strategy Focuses on the Hypervariable Regions of the mitochondrial genome HV1 and HV2



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

mtDNA Coding Region 11-plex ASPE Assay





11-plex PCR and 11-plex SNP detection

Sites are polymorphic in Caucasians (H1) and useful in resolving most common HV1/HV2 types

Multiplex PCR used to co-amplify all regions of interest at once

PCR product sizes kept under 200 bp to enable success with degraded DNA samples

Vallone et al., IJLM 2004 118: 147-57

Typing 51 samples with mt 11-plex assay

51 samples with similar HV1/HV2 sequence regions based on linear array results

3010	G	A	G	G	G	G	A	G	G	G	G	A	G
4793	A	A	A	A	A	A	A	A	A	A	G	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	T	C	T	C	T	C	C	C
7202	A	A	A	A	A	A	A	A	A	A	A	A	A
16519	T	C	T	C	C	T	C	C	T	T	C	C	C
12858	C	T	C	C	C	C	C	C	C	C	C	C	C
4580	G	G	G	G	G	G	G	A	G	A	G	G	G
477	T	C	T	T	T	T	C	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	1	2	2	3	4	4	15	16

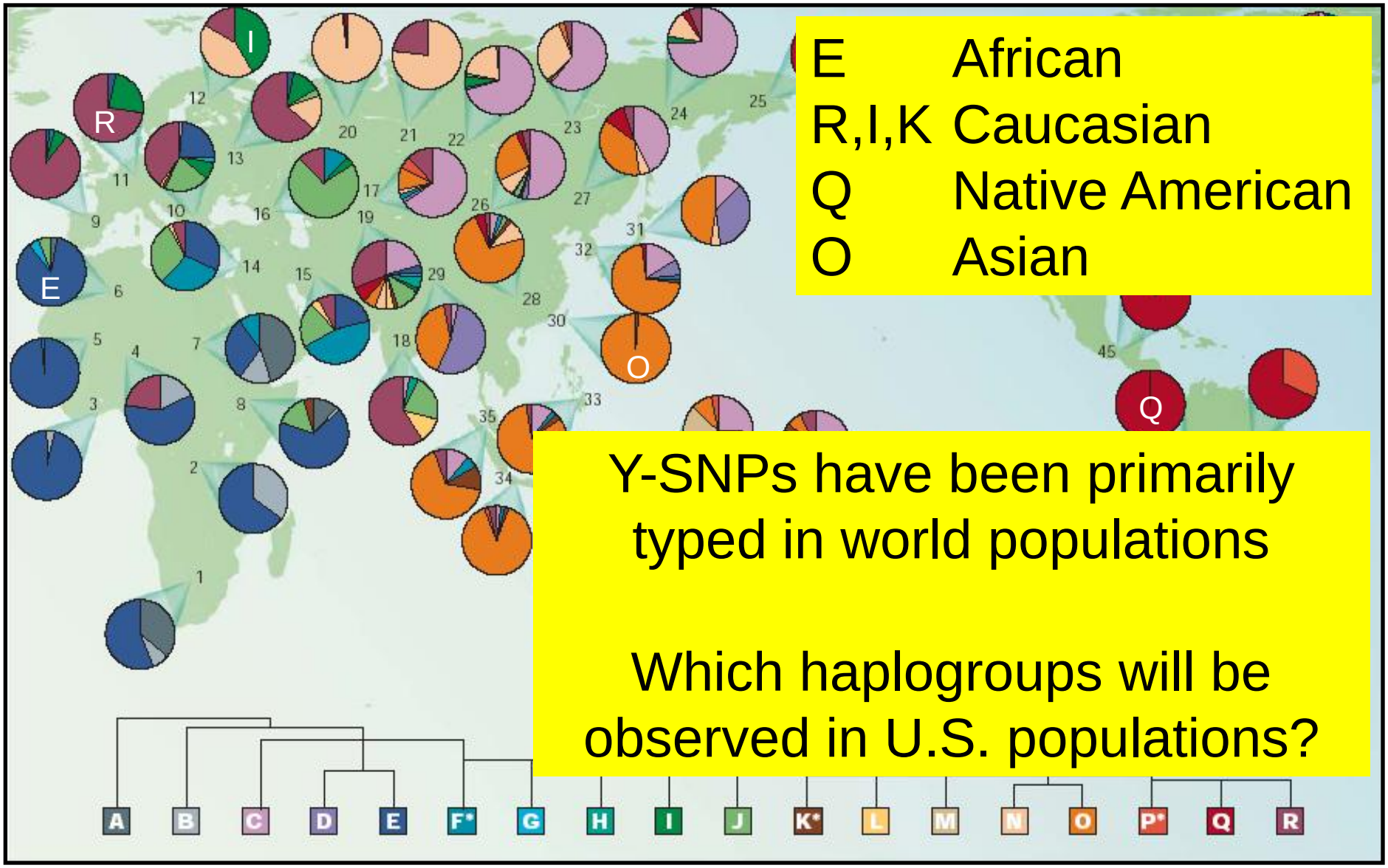
12 haplogroups were observed (5 unique)
2 of 11 sites did not vary
The 11-plex assay is currently in use at AFDIL

Kline, M.C., Vallone, P.M., Redman, J.W., Duewer, D.L., Calloway, C.D., Butler, J.M. (2005) Mitochondrial DNA typing screens with control region and coding region SNPs. J. Forensic Sci. 50: 377-385.

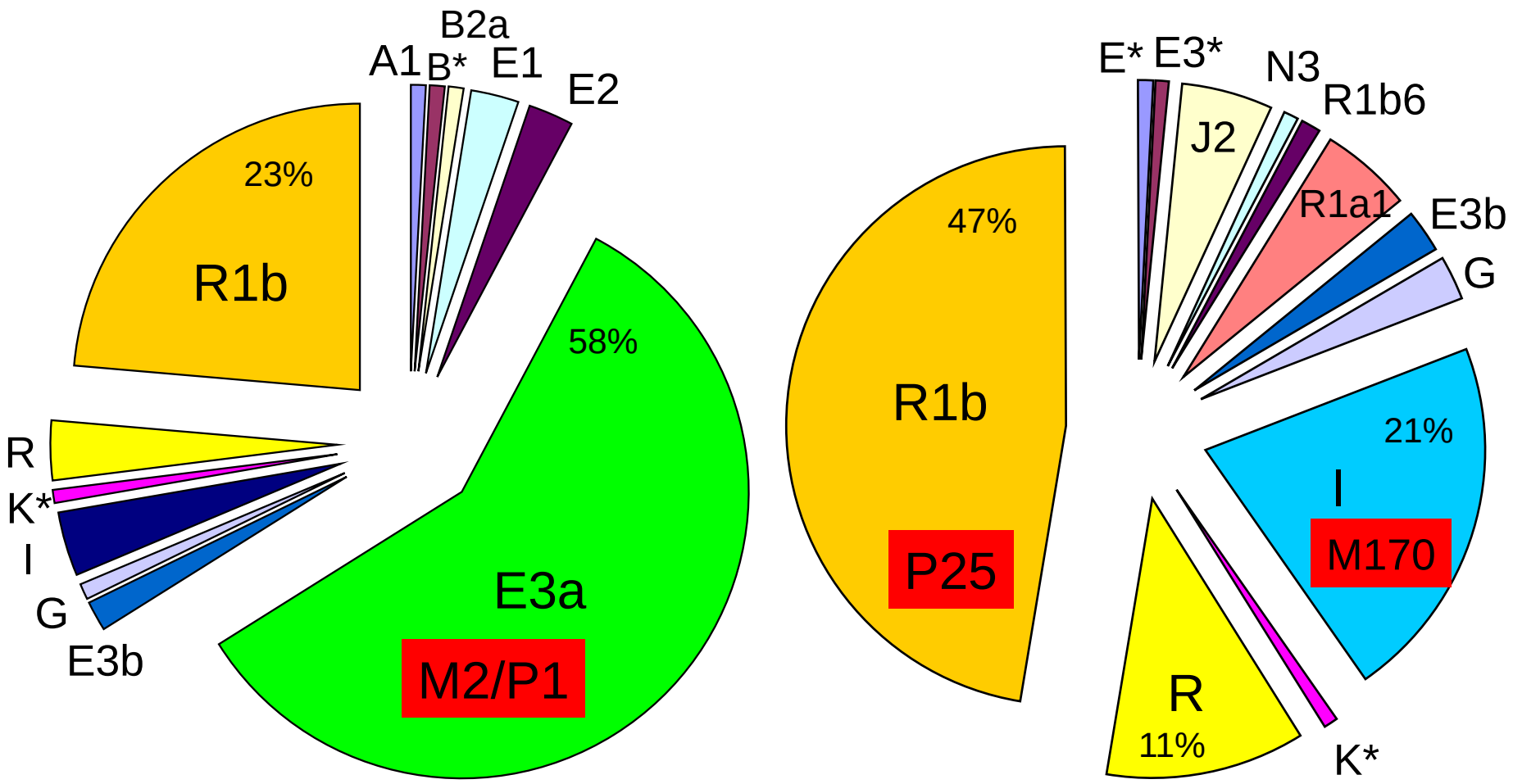
Y-SNPs

- Biogeographical information
- Evolutionary studies (human migration)
- Still discovering Y-SNPs and refining the Y haplogroup tree
- Use for identification? (similar to Y-STRs)
- Degraded samples?

Global Distribution of Y Haplogroups



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



18 of 46 different haplogroups observed in 229 males

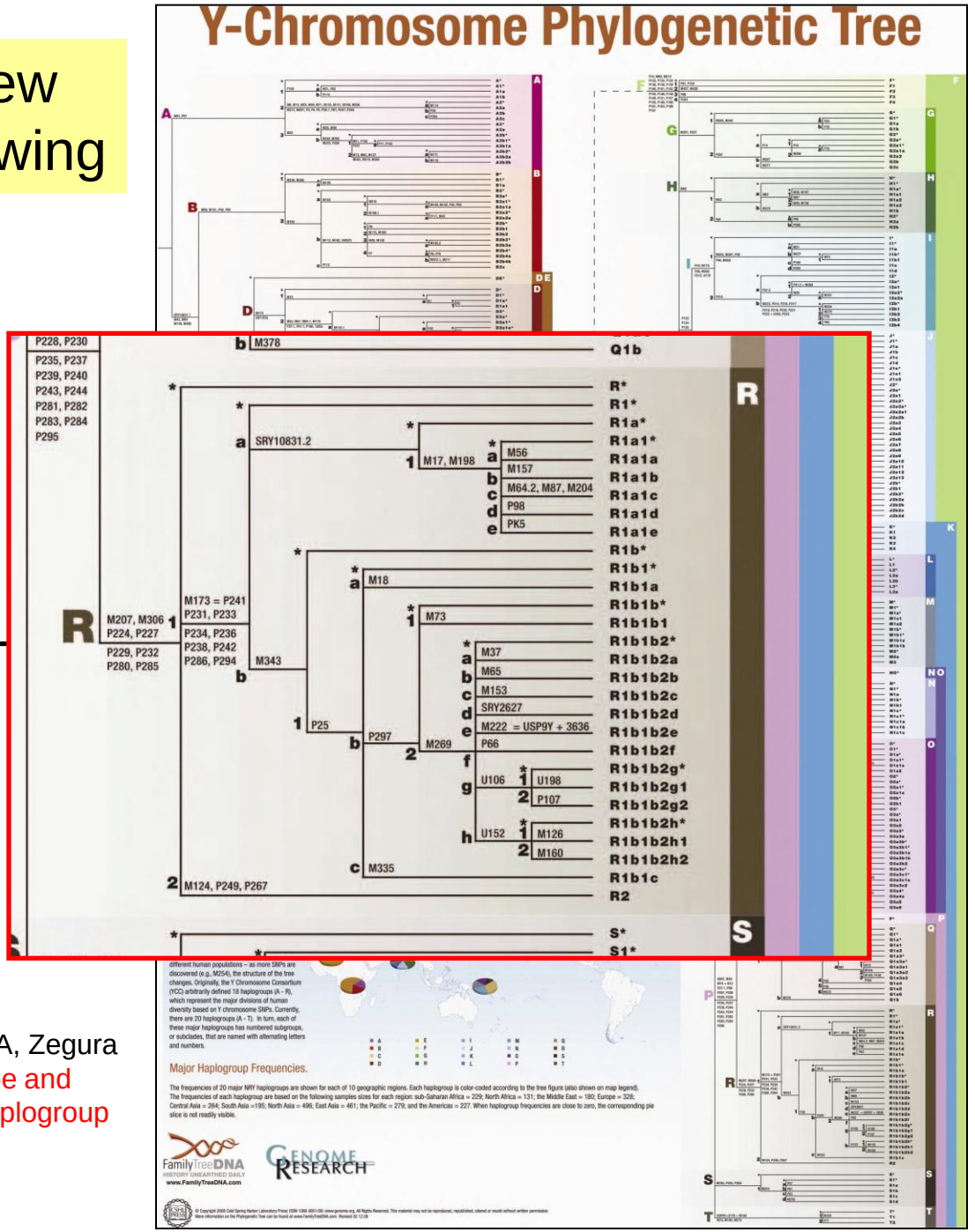
Vallone and Butler, J Forensic Sci. 2004 49: 723-32

Variation was observed in
24 of the 51 Y-SNPs

Research to discover new Y-SNPs is continually growing

In 2002
243 Y-SNPs
Defining 153 Y haplogroups

In 2008
~600 Y-SNPs
Defining 311 Y haplogroups



Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. 2008 18: 830-8.

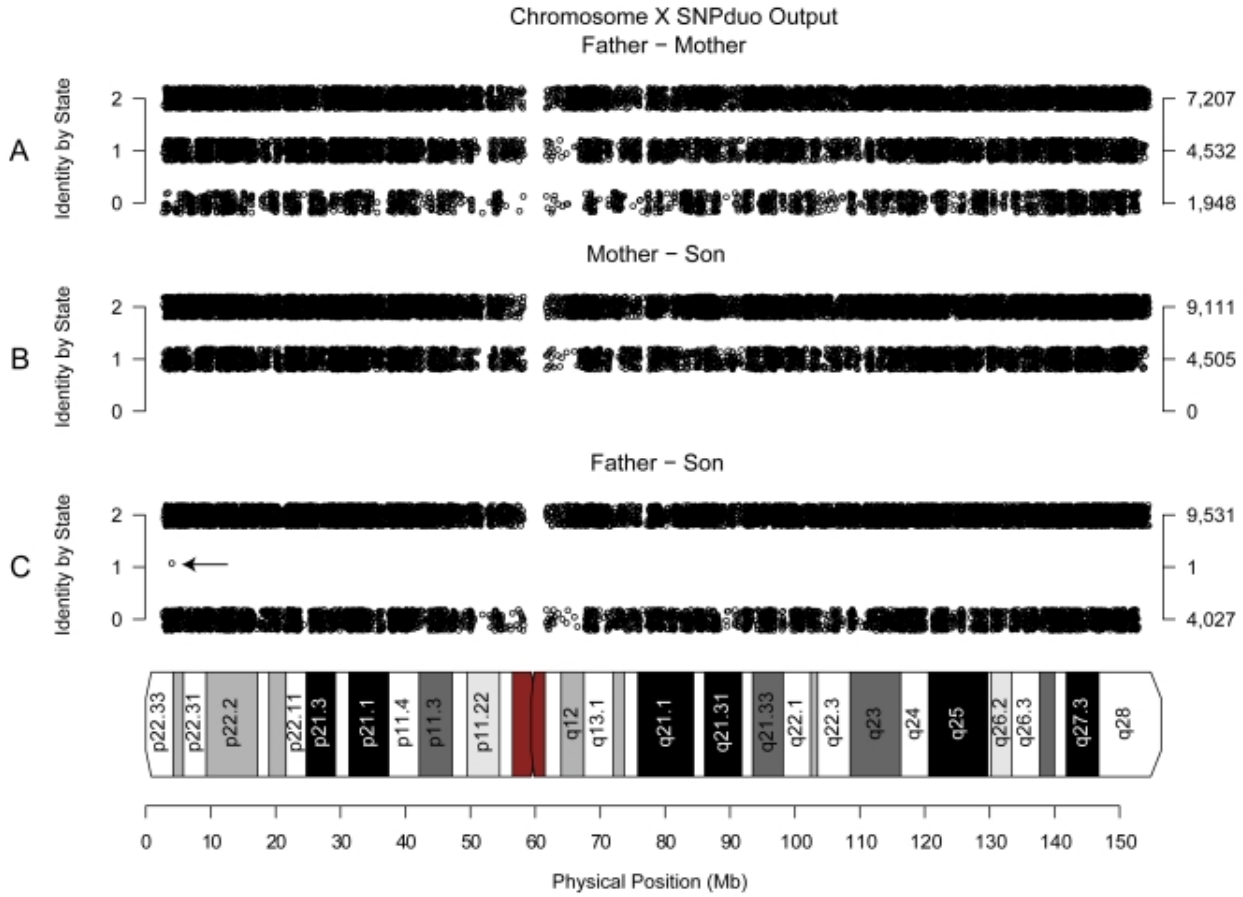
Kinship Analysis with Autosomal SNPs

- Supplement autosomal STR testing for paternity and complex kinship analysis

  C. Borsting, N. Morling, Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers, Forensic Sci. Int. Genet. (in press)

- SNP arrays
  - Identity by state (AA, AB, or BB)
    - No allele frequencies required
  - Haplogroup blocks
  - Clustering different relationship types

Identity by State Patterns for Father, Mother, and Son on the X-chromosome



Unrelated parents, many instances of no shared alleles (e.g. AA vs. BB)

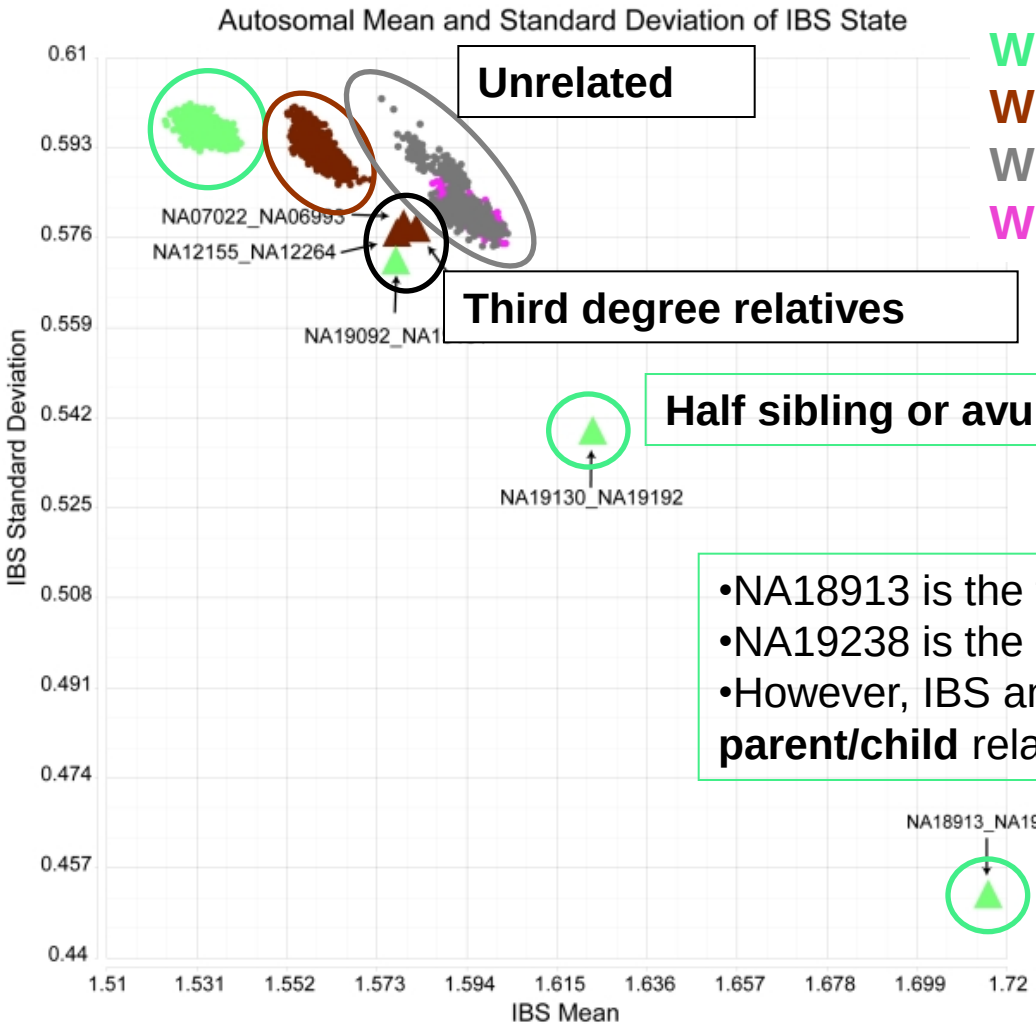
Mother-son, no IBS-0 SNPs because the son inherited a copy of the maternal X

Father/son comparison, no IBS-1 SNPs since the X chromosomes were non-identical.

SNPduo analysis with Illumina HumanHap 550K data

E. Roberson, J. Pevsner. Visualization of shared genomic regions and meiotic recombination in high-density SNP data, PLoS One. 2009; 4(8): e6711.

IBS Analysis in “Unrelated” HapMap Data



Within African
Within Caucasian
Within Japanese
Within Chinese

- NA18913 is the father in a trio
- NA19238 is the mother in a separate trio
- However, IBS analysis suggested **parent/child** relationship

Identical twins
Self vs self

IBS mean ~2

E. Roberson, J. Pevsner. Visualization of shared genomic regions and meiotic recombination in high-density SNP data, PLoS One. 2009; 4(8): e6711.

Summary

- Individual Identification SNPs
  - No core loci or commercial kits, issues with mixtures
  - 52-plex used in European community, degraded samples, supplement kinship testing
- Ancestry Informative SNPs
  - Panels are being selected and tested
  - Could be of practical use?
  - U.S. admixture is an issue
- Phenotype Informative SNPs
  - Complex traits, area still growing
  - Rapid screen could be useful as an investigative lead
- Lineage Informative SNPs
  - Mito SNP assays are useful for resolving common HVI/II haplogroup
  - Y-SNPs have limited forensic utility (especially compared to Y-STRs)
  - High-density SNP arrays may be powerful for detecting distant relationships

Thank you for your attention!

Kristen O'Connor

kristen.oconnor@nist.gov

301-975-5205

PDF of talk

http://www.cstl.nist.gov/strbase/NISTpub.htm

Forensic SNP Information

http://www.cstl.nist.gov/strbase/SNP.htm