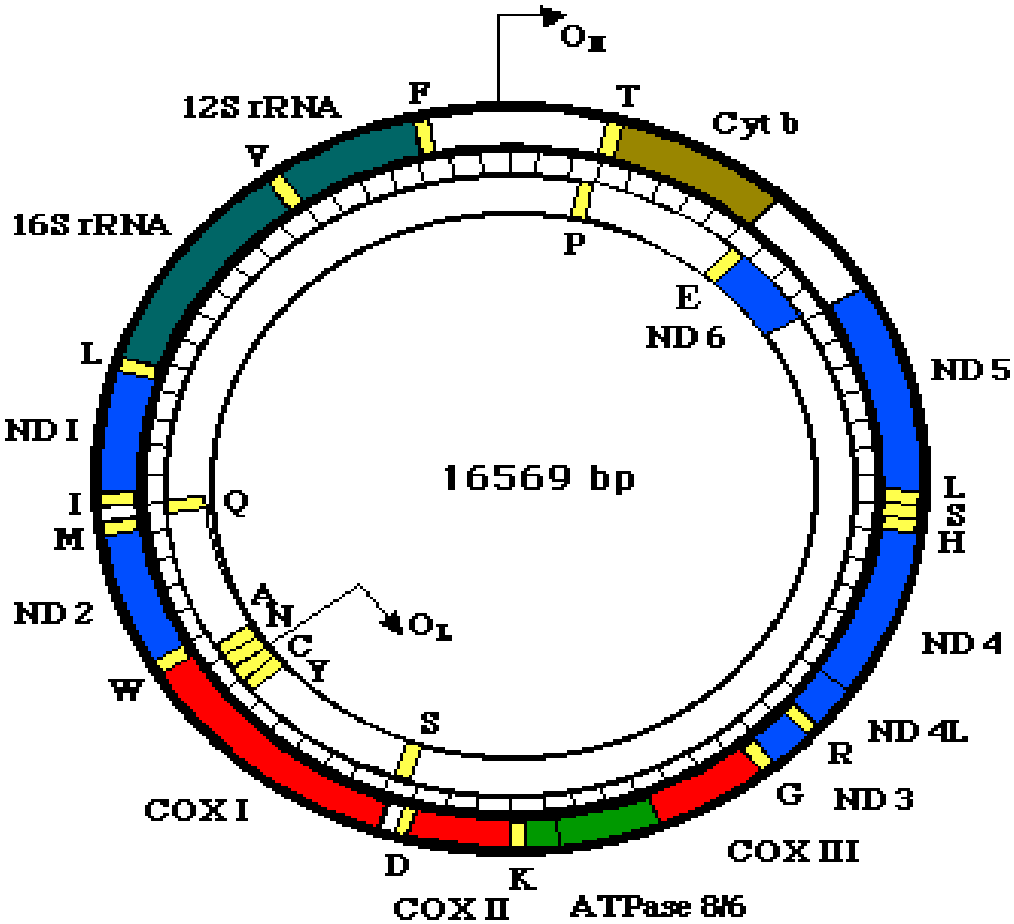
# Identification of SNPs in the Mitochondrial Genome to Resolve Common HV1/HV2 Types in Caucasian Populations.

Michael D. Coble[1], Rebecca S. Just[2], Jessica L. Saunier[2]; Jennifer E. O'Callaghan[2]; Ilona H. Letmanyi[2], Christine T. Peterson[2], Jodi A. Irwin[2]; Peter M. Vallone[1], John M. Butler[1], and Thomas J. Parsons[2].

[1]National Institute of Standards and Technology, Gaithersburg, MD, USA

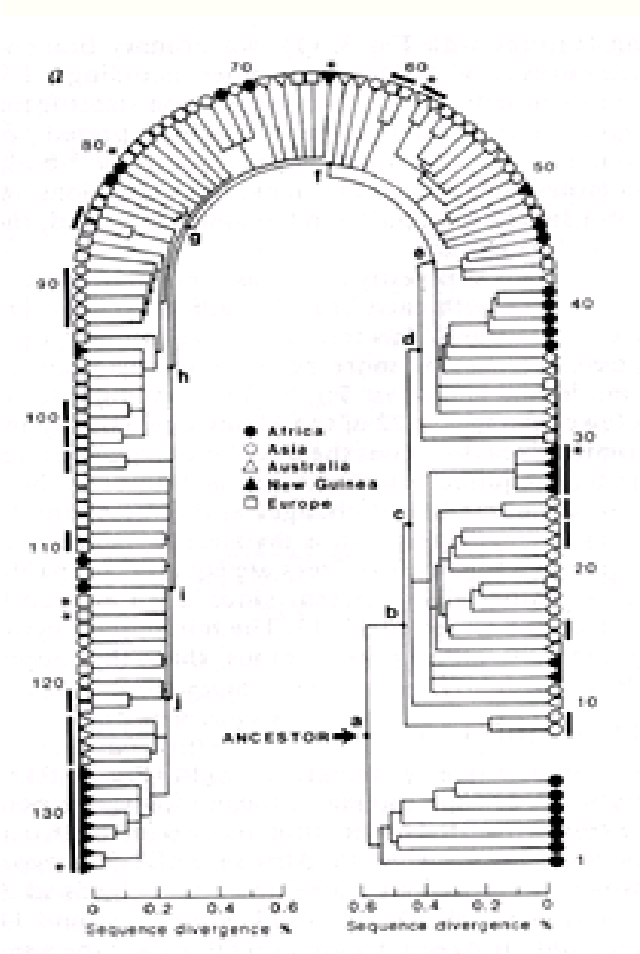
[2]The Armed Forces DNA Identification Laboratory, Rockville, MD, USA

# The Mitochondrial Genome




http://www.mitomap.org/

# mtDNA as a Genetic Marker

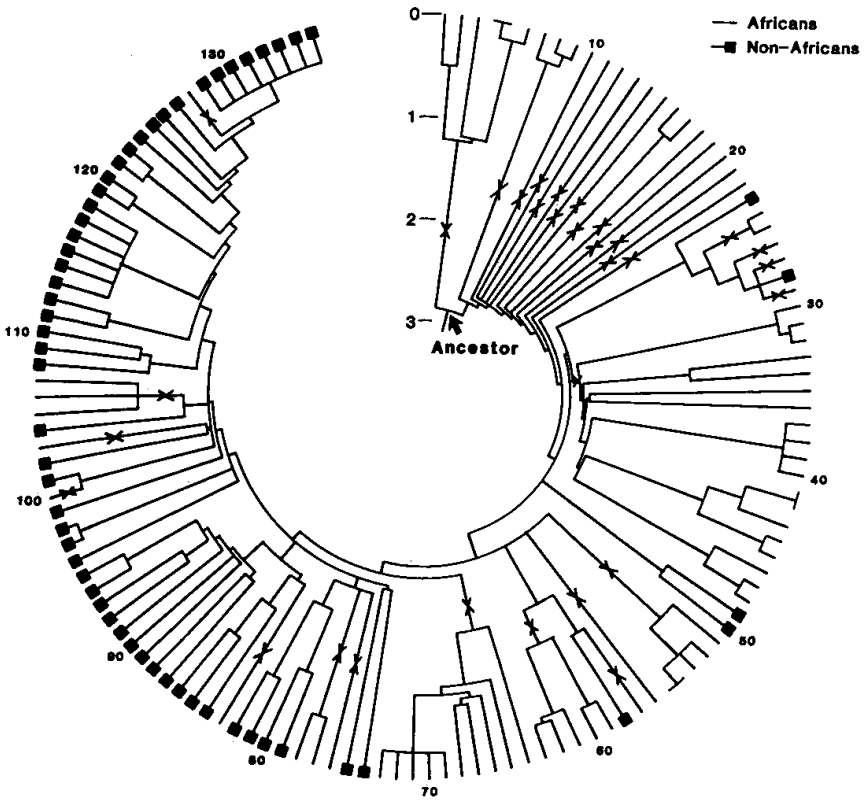



RFLP analysis of 134 individuals

Cann et al. 1987

“Out of Africa”

# mtDNA as a Genetic Marker

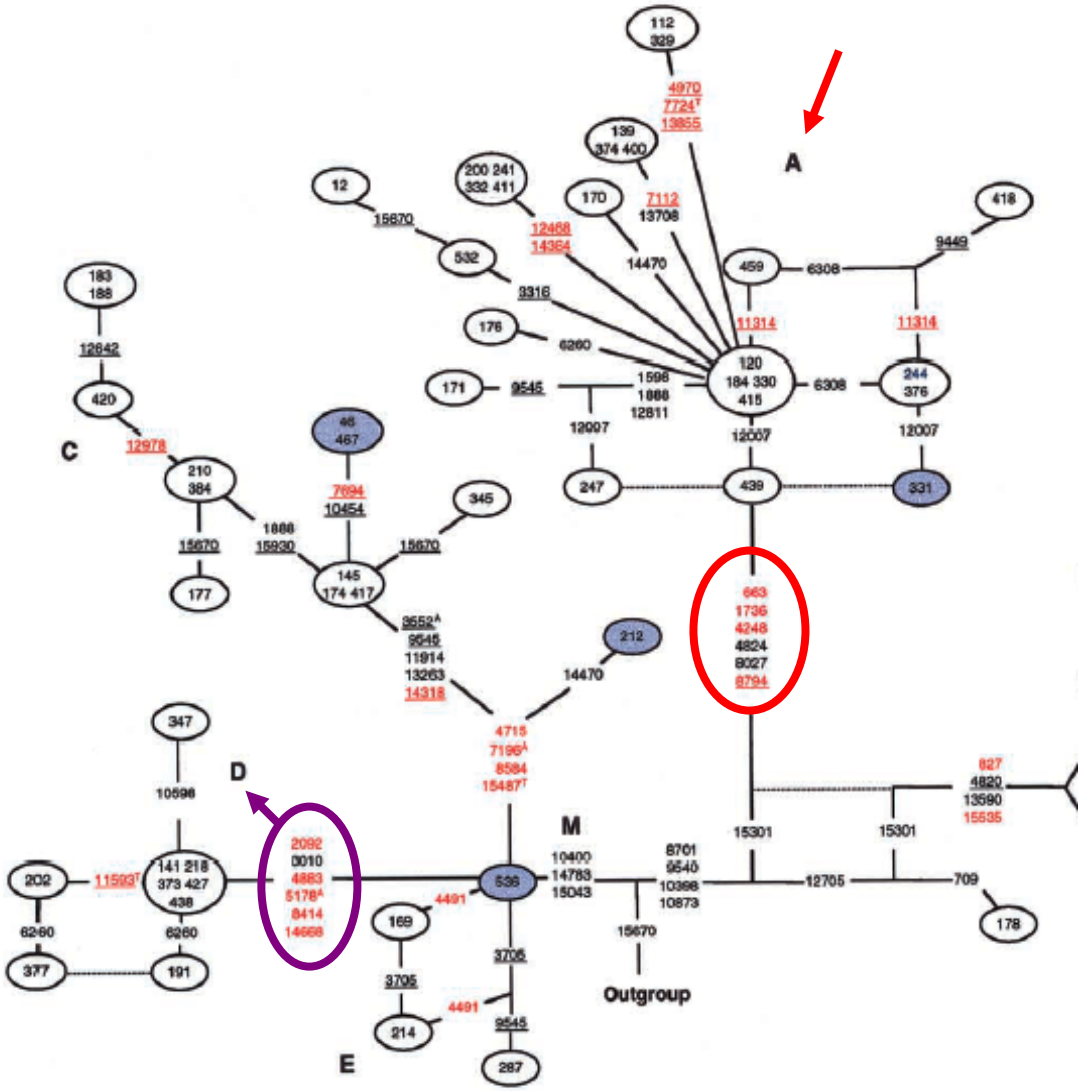



Control Region Sequence Analysis of 189 individuals

Vigilant et al. 1991

# Mitochondrial Haplogroups

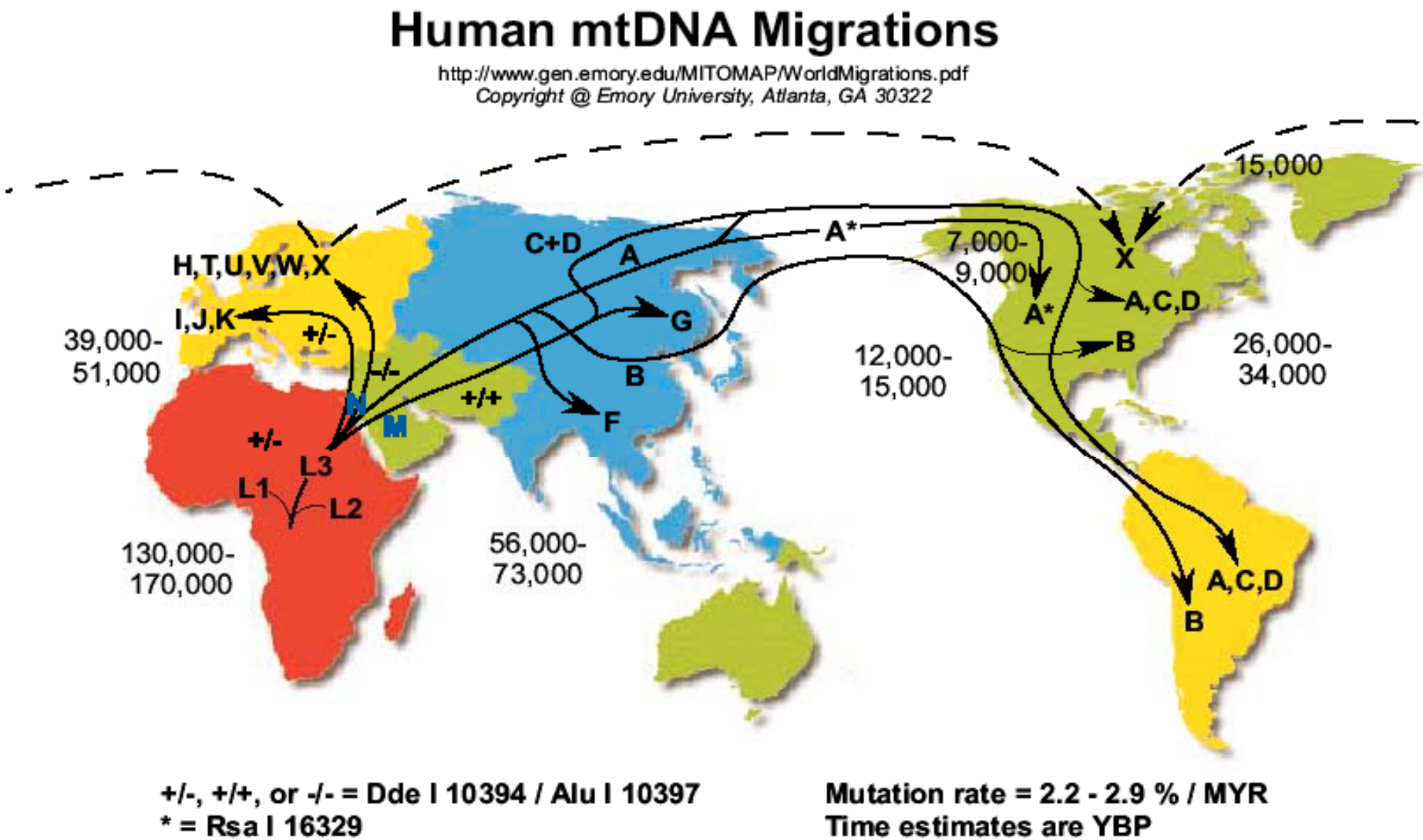
Herrnstadt et al.: Network Analysis of mtDNA Haplogroups




Haplogroup - A group of related haplotypes.

Each haplogroup cluster is defined by a set of specific, shared polymorphisms.

# mtDNA Haplogroups (RFLP)



# Caucasian mtDNA Haplogroups (HV1/HV2)

- H - CRS +/- variants
- J - 16069 C-T 16126 T-C 73 A-G 295 C-T
- T - 16126 T-C 16294 C-T 73 A-G
- V - 16298 T-C 72 T-C

Macaulay et al. (1999) *AJHG* **64:** 232-249.
Allard et al. (2002) *JFS* **47:** 1215-1223.

# mtDNA as a Forensic Tool

## Advantages of Using mtDNA

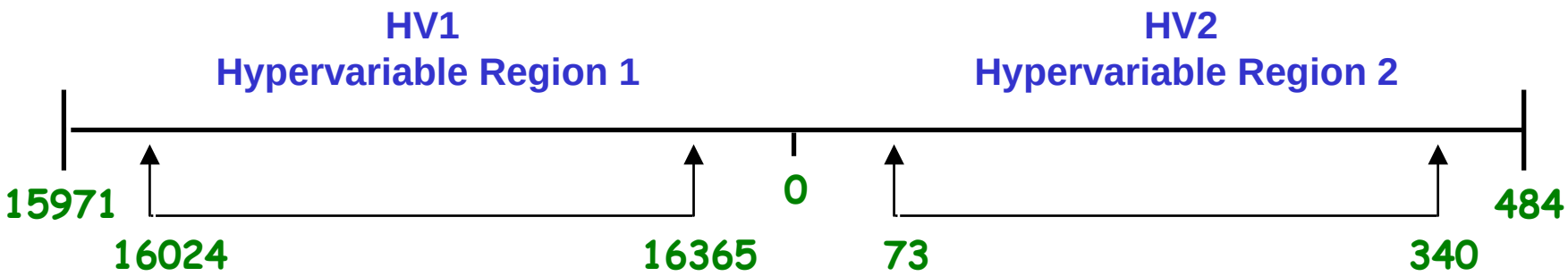
- Maternal Inheritance
- Lack of Recombination
- High Copy Number
- Cases where:
  - DNA is degraded
  - Only maternal references are available
  - Samples with little or no Nuclear DNA
    - Shed Hairs
    - Fingernails

# mtDNA as a Forensic Tool

## Disadvantages of Using mtDNA

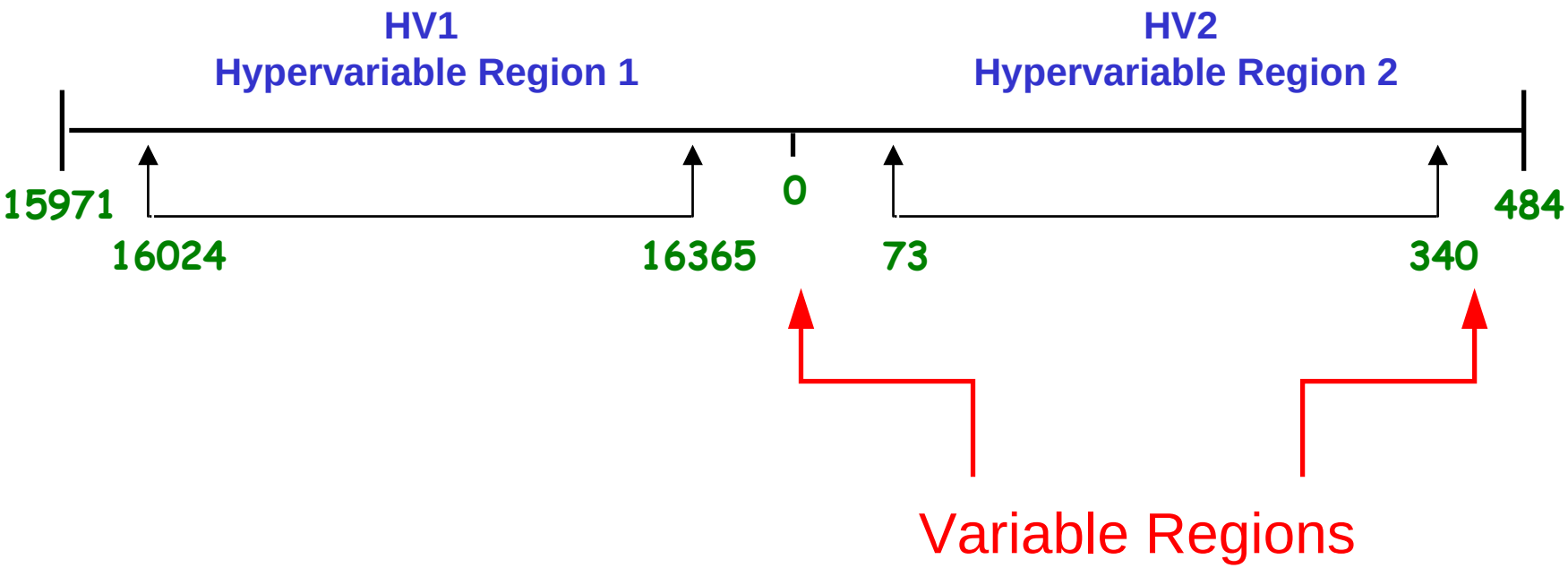
- Maternal Inheritance – You have many!
- Not a unique identifier
- Some mtDNA types are common in the population

# Current mtDNA Amplification & Sequencing Strategy

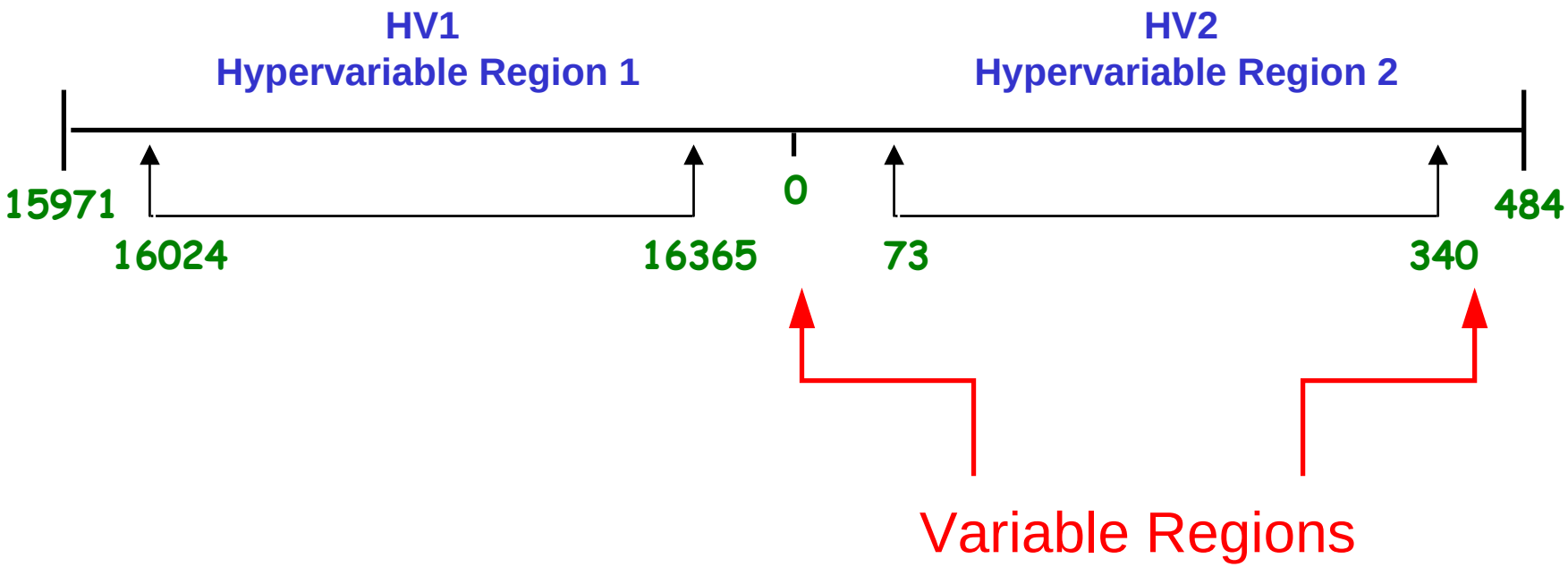



HV1 + HV2 = 610 bp

# Current mtDNA Amplification & Sequencing Strategy

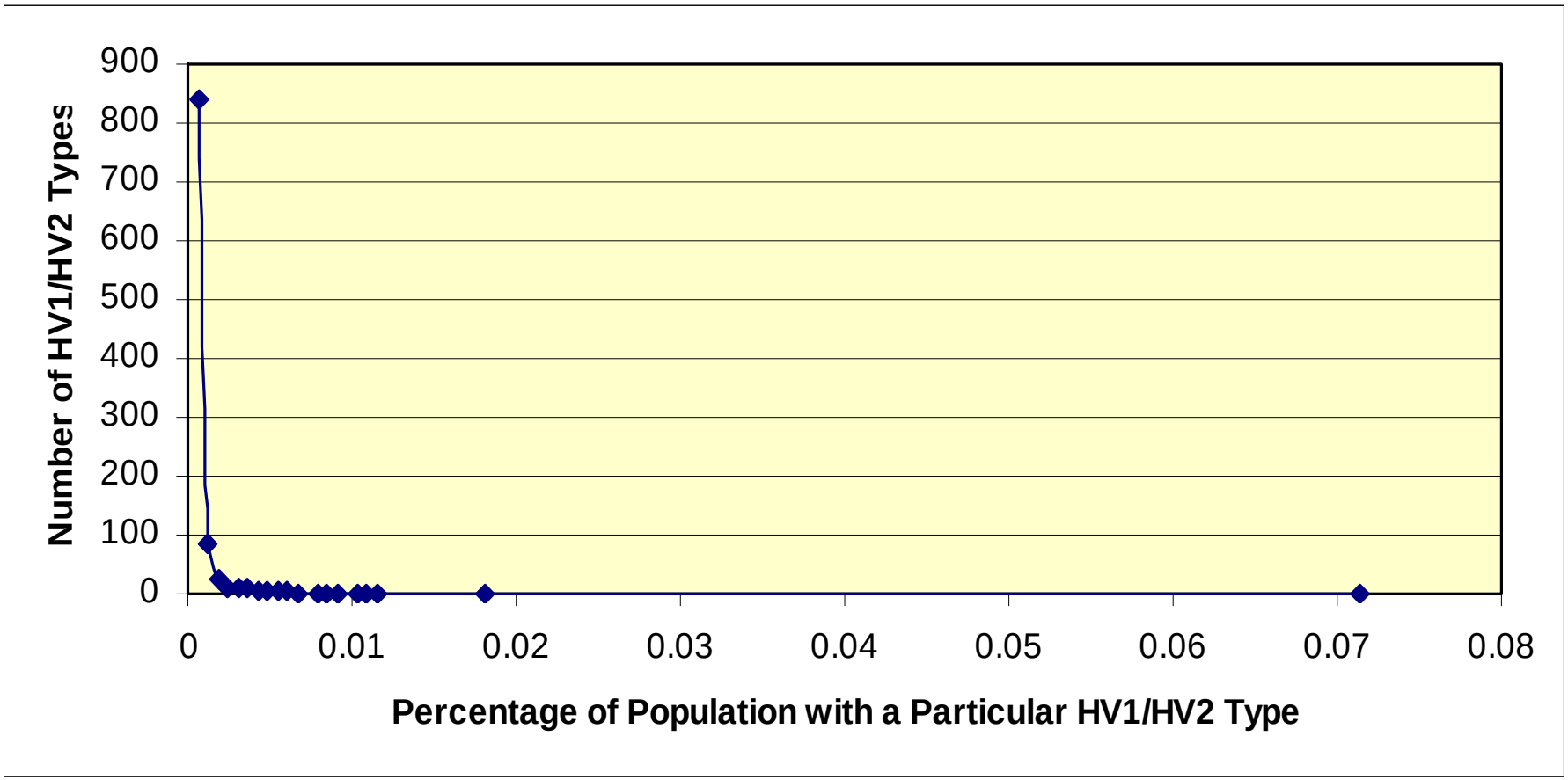


# Current mtDNA Amplification & Sequencing Strategy

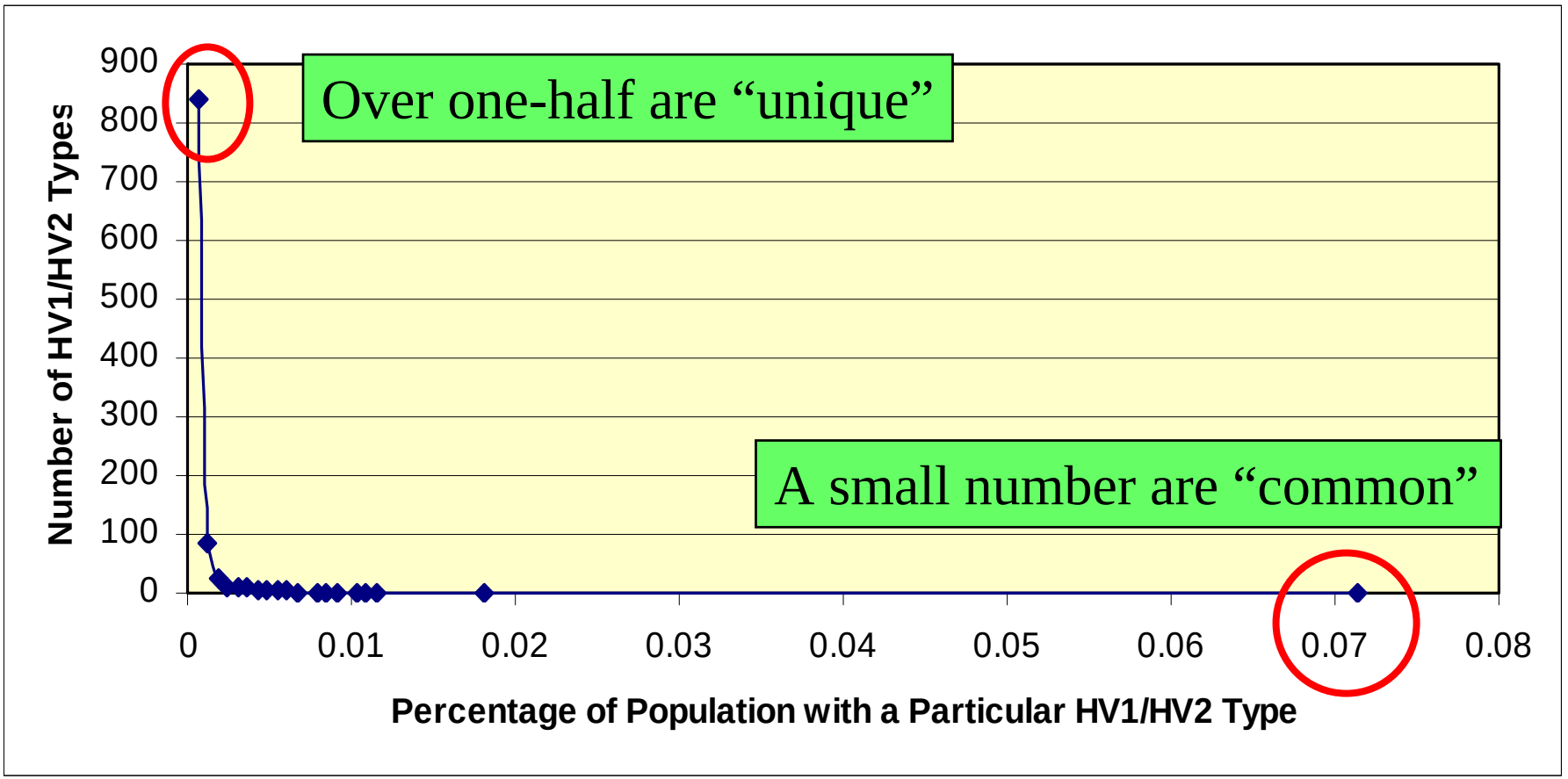



More Sequence Information

# mtDNA Population Distribution Caucasians (n=1665)



# mtDNA Population Distribution Caucasians (n=1665)



# Framing the Problem

The greatest limitation for mtDNA testing lies with the small number of common types for which the power of discrimination is low.

~20% of the time, the Forensic Scientist encounters a HV1/HV2 type that occurs at greater than ~0.5% of the population

In database or mass fatality comparisons: multiple hits will occur for these common types.

# A Case Example

- September 15, 1943 - B17F Bomber returning from a mission to Port Moresby, New Guinea

# A Case Example

- The plane crashes in the Owen Stanley Mountain range due to “adverse weather.”
- Subsequent searches proved negative.
- 11 crewmen declared non-recoverable on July 22, 1949.

# A Case Example

- October 9, 1992 - A private company helicopter discovers crash site.
- mtDNA testing reveals that 3/11 crewmen share the same HV type (263 A-G, 315.1 C).
- Further VR testing could distinguish 1 of the 3 crewmen (16519 T-C). However, 2 crewmen still matched.

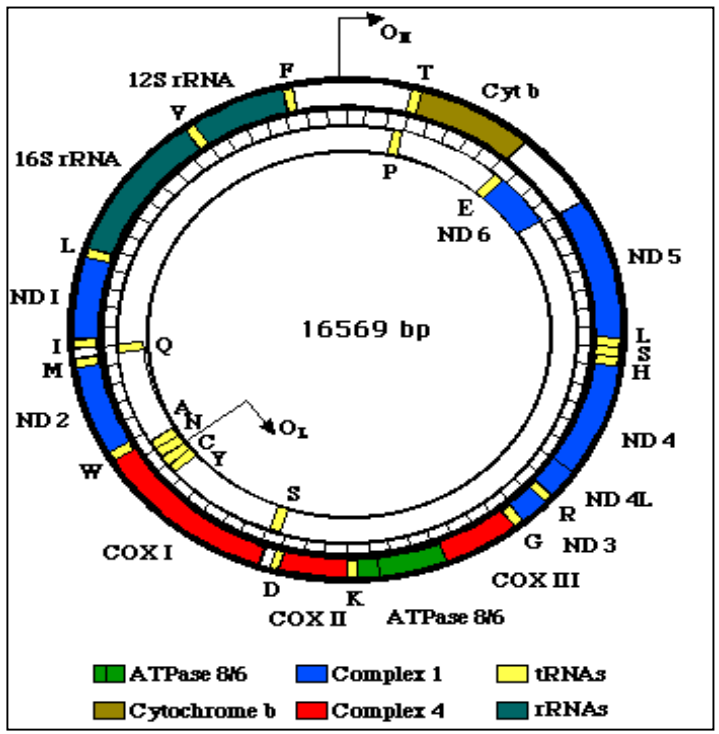
# A Case Example

- Partial dental records were used to associate 3 teeth among the 2 crewmen matching in the CR.

- One L femur could not be associated with either crewmen, and was buried in a grave containing group remains

# Central Effort of the Project

- Sequence variation outside of HV1/HV2 can be used to distinguish Caucasian individuals sharing common types.




Coding Region – evolutionary rate is 4-fold less than the control region.

However…15X Amount of DNA

# Ethical Considerations

- More than 100 characterized diseases associated with mtDNA mutations (Mitomap – www.mitomap.org)

- To avoid having forensic testing from evolving into genetic counseling, we focused on neutral SNPs in the mtGenome.

# SNPs for Discrimination

- Non-coding sites in the control region (outside of HV1/HV2).
- Non-coding “spacer” regions throughout the mtGenome.
- Silent mutations in protein coding genes.

# SNPs for Discrimination

- Practical application – A set of SNP sites that can be rapidly assayed to provide maximal discrimination.
- Avoids further sequencing.
- SNaPShot™ (ABI) – small amplicons, multiplexed - can conserve template.

# Strategy for SNP Identification

- Sequence the entire genome of unrelated individuals sharing common HV1/HV2 types in the Caucasian population (focus on 18 of 22 common types that occur at a frequency of 0.5% or greater).

# Common mtDNA Haplogroups

| Com | Haplo | Seq (+ CRS) |
|---|---|---|
| 31 | **H1** | CRS |
| 25 | **H2** | 152 C |
| 11 | **H3** | 16129 A |
| 8 | **H4** | 16263 C |
| 12 | **H5** | 16304 C |
| 11 | **H6** | 73 G |
| 7 | **H7** | 16162 G 16209 C 73 G |

Length Variation in HV2 C-stretch – ignored (see Stewart et al., 2001)

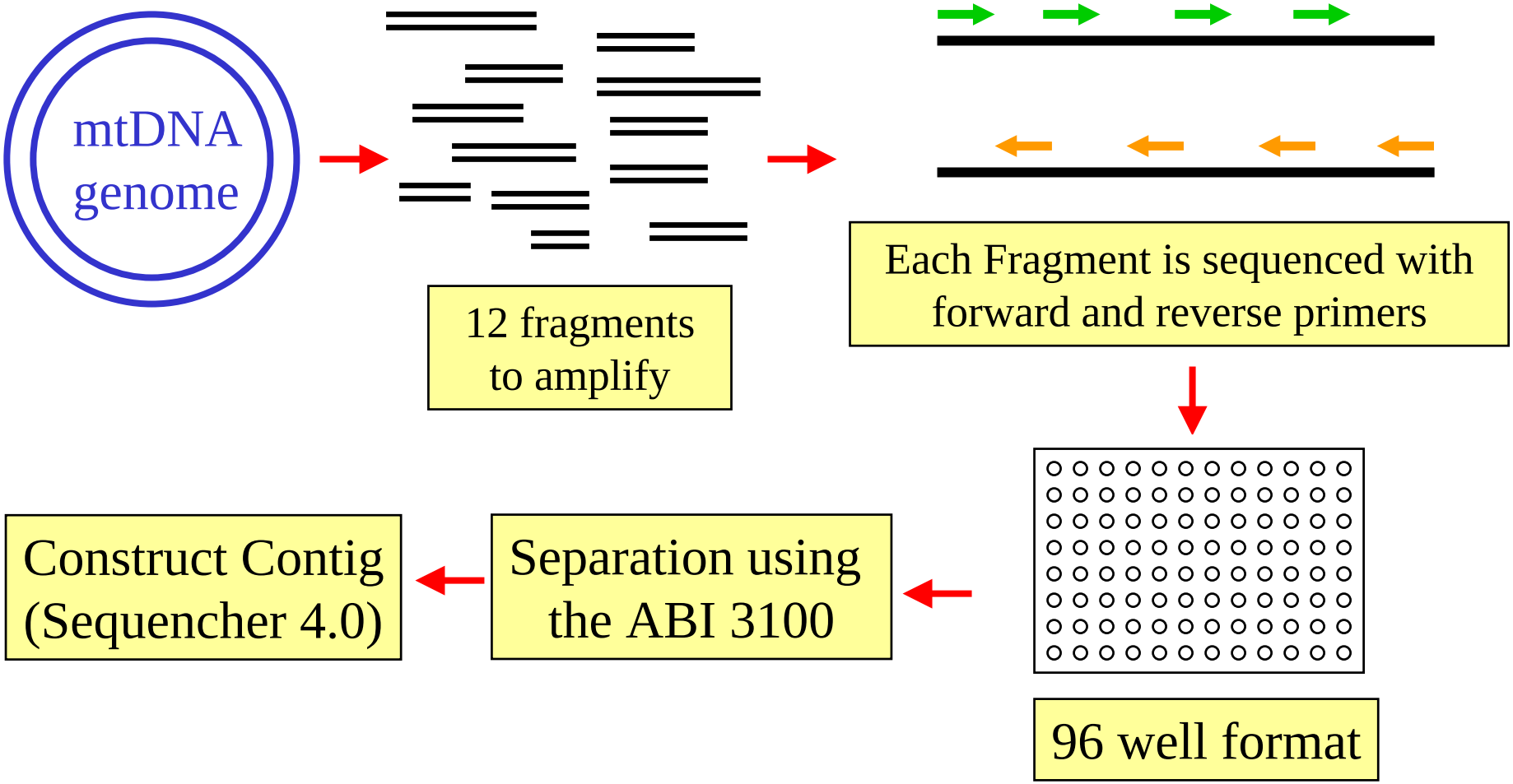
# Common mtDNA Haplogroups

| Com | Haplo | Seq (+ CRS) |
|---|---|---|
| 15 | J1 | 16069 T 16126 C 73 G 185 A 228 A 295 T |
| 6 | J2 | 16069 T 16126 C 73 G 228 A 295 T |
| 12 | J3 | 16069 T 16126 C 73 G 185 A 188G 228 A 295 T |
| 3 | J4 | 16069 T 16126 C 16145 A 16172 C 16222 T 16261 T 73 G 242 T 295 T |
| | | |
| 20 | T1 | 16126 C 16294 T 16296 T 16304 C 73 G |
| 10 | T2 | 16126 C 16163 G 16186 T 16189 C 16294 T 73 G 152 C 195 C |
| 8 | T3 | 16126 C 16294 T 16296 T 73 G |
| | | |
| 25 | V1 | 16298 C |
| | | |
| 14 | K1 | 16224 C 16311 C 73 G 146 C 152 C |
| 8 | K2 | 16093 C 16224 C 16311 C 73 G |
| 7 | K3 | 16224 C 16311 C 73 G |

241 total genomes from 18 common HV1/HV2 types
(~14% of the total database)

# Whole Genome Sequencing Strategy

- Human mtDNA standard reference material (Levin et al., 1999)



# High Throughput Sequencing
# MWG RoboAmp 4200

# High Throughput Sequencing from NIJ Support

Jennifer O'Callaghan Rebecca Just Jessica Saunier

Former Team Members – Christine Peterson and Ilona Letmanyi

# Criteria for SNP Selection

- Neutral.
- Should be shared (within or among individuals sharing the common types).
- Non-redundant

# The Nature of the SNPs

- Would the SNPs that resolve one group be useful for resolving other closely related groups?

| Com | Haplo | Seq (+ CRS) | |
|---|---|---|---|
| 31 | H1 | CRS | |
| 25 | H2 | 152 C | |
| 11 | H3 | 16129 A | "Hot Spots" |
| 8 | H4 | 16263 C | |
| 12 | H5 | 16304 C | |
| 11 | H6 | 73 G | |
| 7 | H7 | 16162 G 16209 C 73 G | |

# The Nature of the SNPs

- Are resolving SNPs **slow, rare polymorphisms** that occurred once during the evolution of a haplogroup?

OR….

# The Nature of the SNPs

- Are resolving SNPs **slow, rare polymorphisms** that occurred once during the evolution of a haplogroup?

OR….

- Are resolving SNPs “universally” **fast hot spots**, useful for all haplogroups (L, M, N)?

OR….

# The Nature of the SNPs

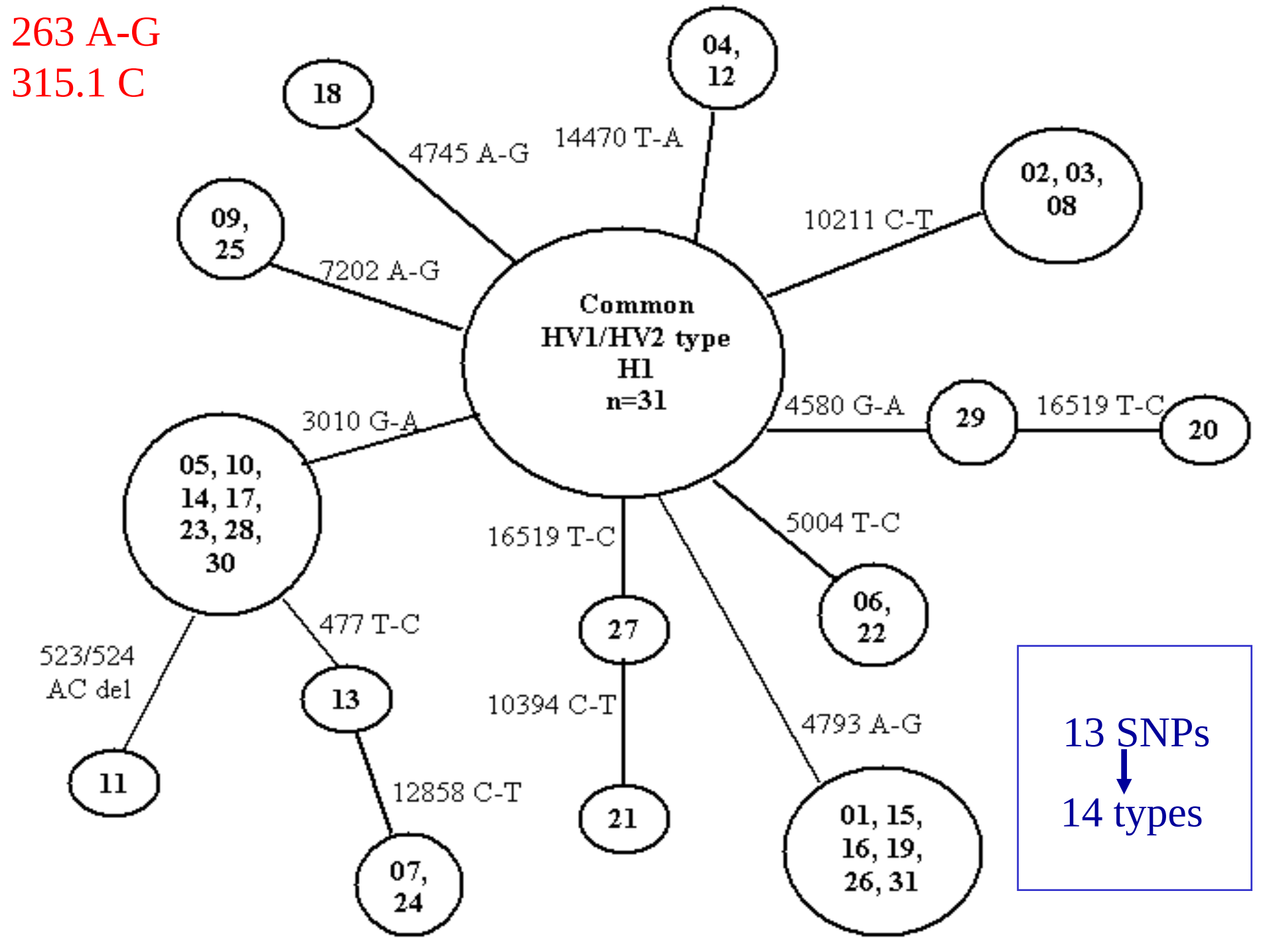
- Are resolving SNPs **slow, rare polymorphisms** that occurred once during the evolution of a haplogroup?

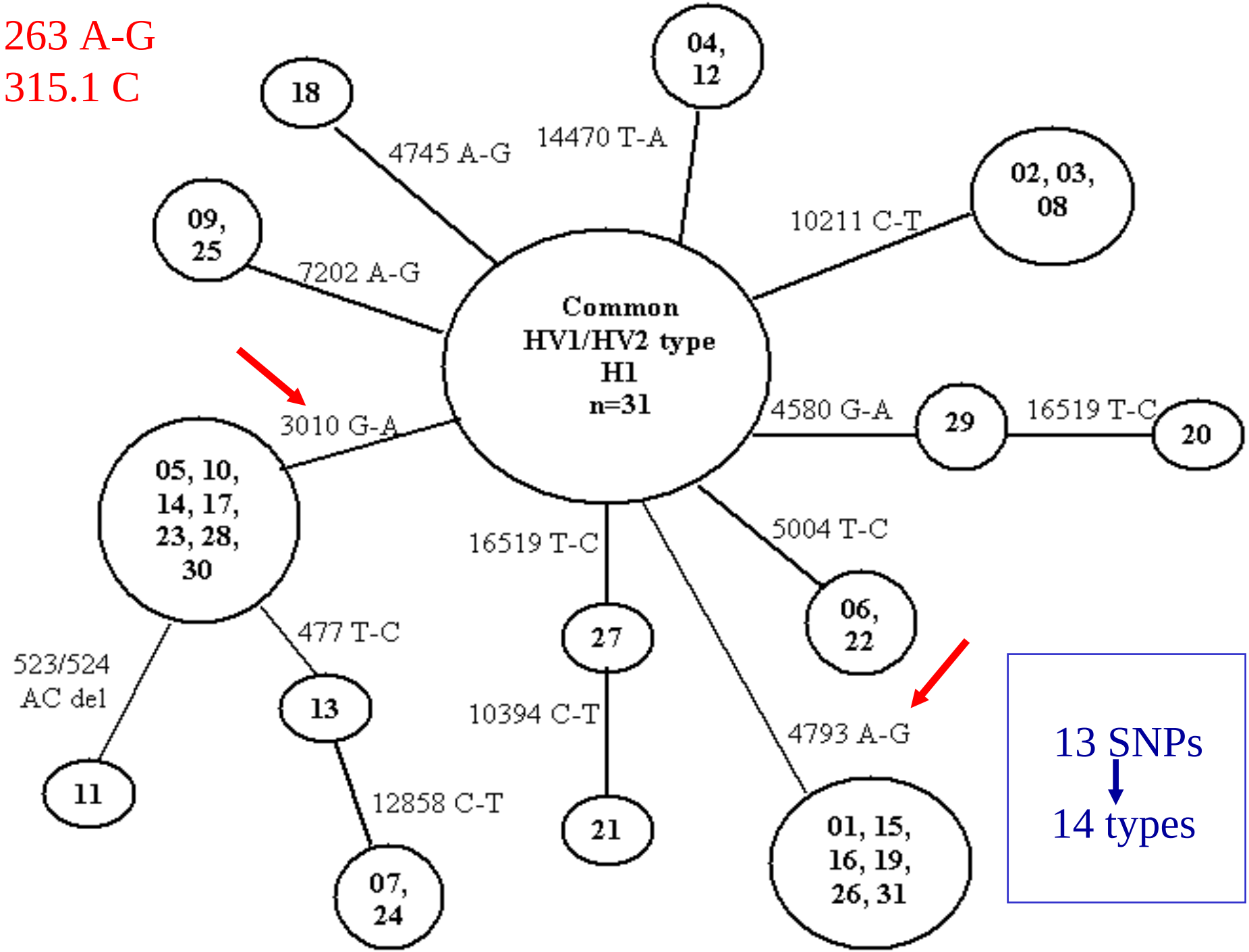
OR….

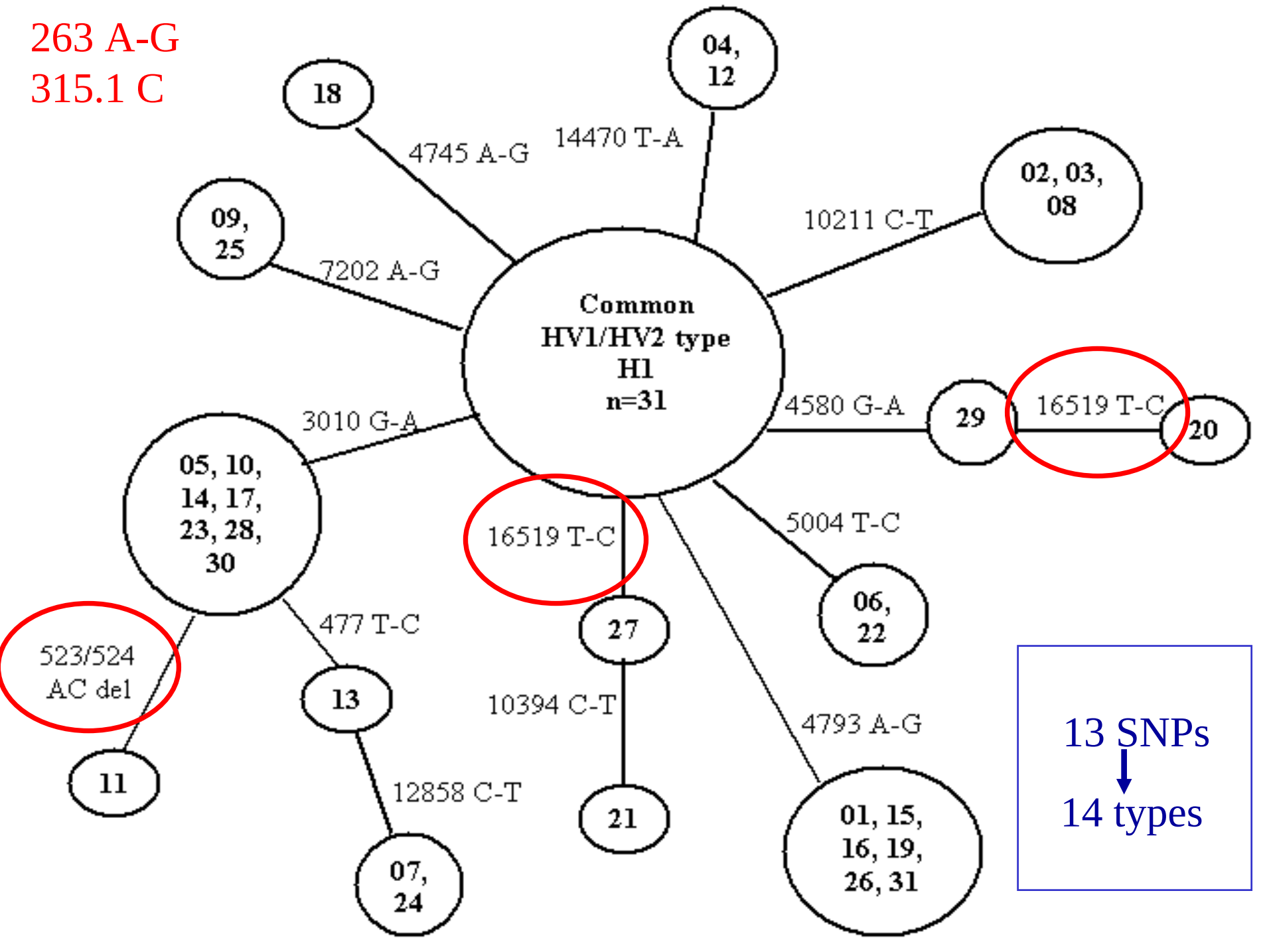
- Are resolving SNPs “universally” **fast hot spots**, useful for all haplogroups (L, M, N)?

OR….

- Are resolving SNPs a combination of the two?

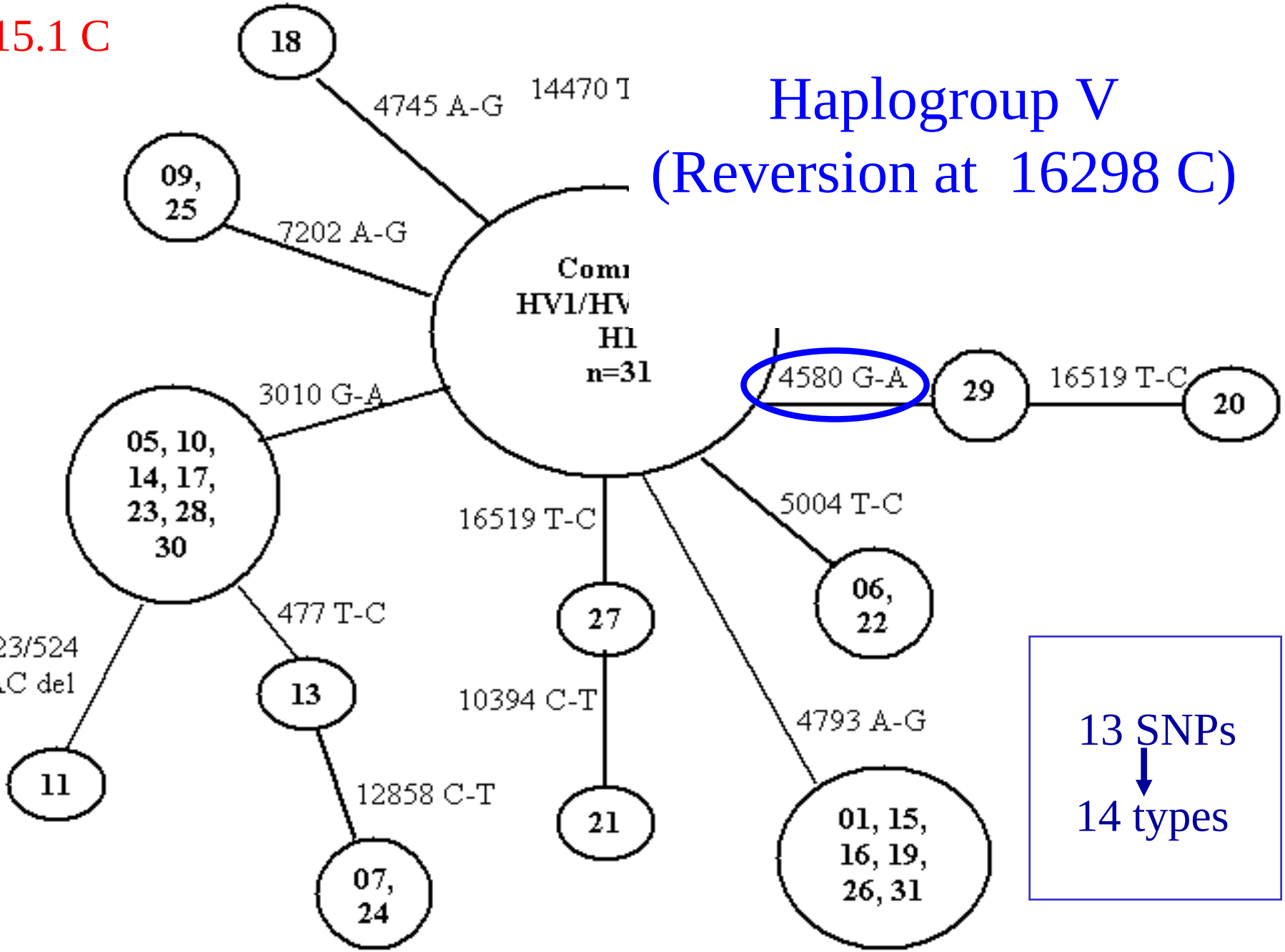

263 A-G
315.1 C
18
04, 12
4745 A-G
14470 T-A
02, 03, 08
09, 25
10211 C-T
7202 A-G
Common HV1/HV2 type H1 n=31
4580 G-A
29
16519 T-C
20
3010 G-A
05, 10, 14, 17, 23, 28, 30
5004 T-C
16519 T-C
06, 22
27
477 T-C
523/524 AC del
13
10394 C-T
4793 A-G
11
12858 C-T
21
01, 15, 16, 19, 26, 31
07, 24
13 SNPs
14 types


263 A-G
315.1 C
18
04, 12
4745 A-G
14470 T-A
09, 25
02, 03, 08
10211 C-T
7202 A-G
Common HV1/HV2 type H1 n=31
3010 G-A
4580 G-A
29
16519 T-C
20
05, 10, 14, 17, 23, 28, 30
16519 T-C
5004 T-C
06, 22
27
477 T-C
523/524 AC del
13
10394 C-T
4793 A-G
11
12858 C-T
21
01, 15, 16, 19, 26, 31
07, 24
13 SNPs
14 types


263 A-G
315.1 C
18
04, 12
4745 A-G
14470 T-A
02, 03, 08
09, 25
10211 C-T
7202 A-G
Common HV1/HV2 type H1 n=31
4580 G-A
29
16519 T-C
20
3010 G-A
05, 10, 14, 17, 23, 28, 30
16519 T-C
5004 T-C
06, 22
523/524 AC del
477 T-C
27
13
10394 C-T
4793 A-G
11
12858 C-T
21
01, 15, 16, 19, 26, 31
07, 24
13 SNPs
14 types

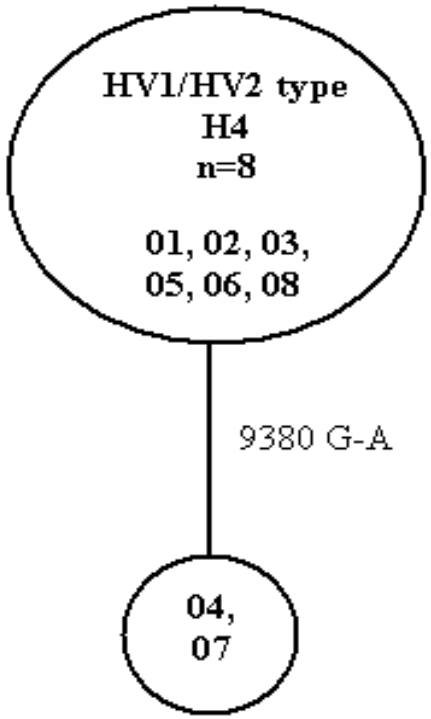
263 A-G
315.1 C

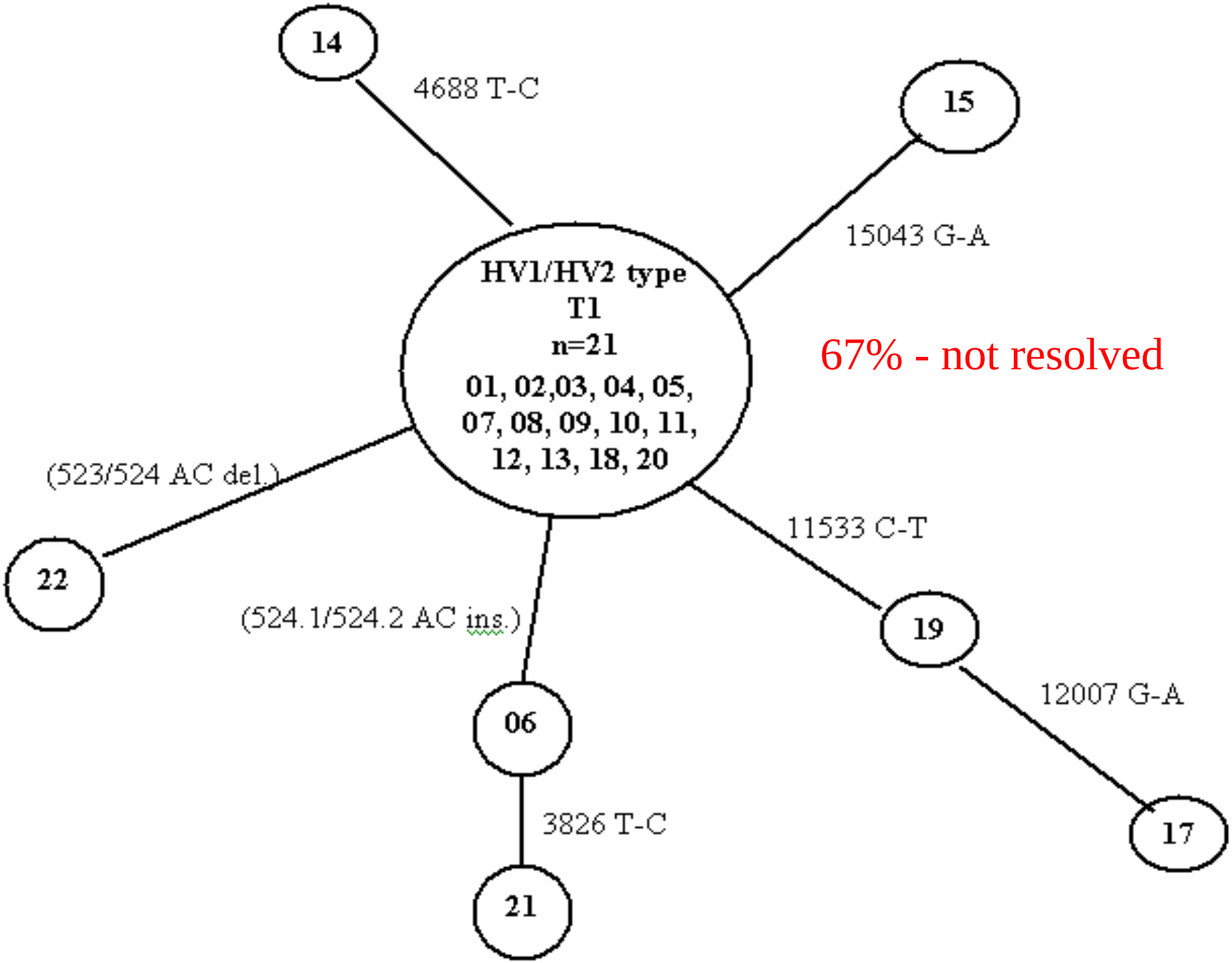
# Haplogroup V
# (Reversion at 16298 C)




13 SNPs
↓
14 types

# H4 - CRS + 16263 T-C




14
4688 T-C
15
15043 G-A
HV1/HV2 type
T1
n=21
01, 02,03, 04, 05,
07, 08, 09, 10, 11,
12, 13, 18, 20
67% - not resolved
(523/524 AC del.)
22
(524.1/524.2 AC ins.)
11533 C-T
19
12007 G-A
17
06
3826 T-C
21

# Summary

- 241 mtGenomes – 420 polymorphic sites in the coding region.
- 32/241 – matched one or more individuals over the entire mtGenome (0/12 H5 individuals matched; 4/8 H7 individuals matched).
- Homoplasies – common in HV1/HV2.

# Summary

- Percentage of sites that varied ranged from 1.0% (16S rRNA) to 6.6% (non-coding regions outside of the control region).

- ATP Synthase 8 (4.8%) and ATP Synthase 6 (3.7%) showed the greatest variation in the protein coding genes.

# Synonymous and Non-synonymous mutations, by Gene

| Gene | Length | Synonymous | Nonsynonymous | Total | % NonSyn. |
|---|---|---|---|---|---|
| ND1 | 956 | 14 | 8 | 22 | 36.4% |
| ND2 | 1,042 | 25 | 11 | 36 | 30.6% |
| CO1 | 1,542 | 29 | 9 | 38 | 23.7% |
| CO2 | 684 | 14 | 4 | 18 | 22.2% |
| ATP8 | 207 | 3 | 5 | 8 | 62.5% |
| ATP6 | 681 | 7 | 20 | 27 | 74.1% |
| CO3 | 784 | 14 | 4 | 18 | 22.2% |
| ND3 | 346 | 5 | 2 | 7 | 28.6% |
| ND4L | 297 | 5 | 1 | 6 | 16.7% |
| ND4 | 1,378 | 30 | 7 | 37 | 18.9% |
| ND5 | 1,812 | 39 | 15 | 54 | 27.8% |
| ND6 | 525 | 8 | 7 | 15 | 46.7% |
| CYB | 1,141 | 23 | 15 | 38 | 39.5% |
| | | | | | |
| Total | 11,341 | 216 | 108 | 324 | 33.1% |

c.f. Mishmar et al. (2003) PNAS

# SNPs for Forensic Discrimination

- 59 SNPs – that met our criteria (neutral, shared, non-redundant).
  - 49 – Protein coding (silent)
  - 8 – Control Region (outside HV1/2)
  - 1 – Non-coding spacer region
  - 1 – 16S rRNA*

* 3010 G-A

# SNPs for Forensic Discrimination

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 477 | 477 | 72 | 482 | 4808 | 64 | 3826 | 64 |
| 3010 | 3010 | 513 | 5198 | 5147 | 4745 | 3834 | 4688 |
| 4580 | 3915 | 4580 | 6260 | 9380 | 10211 | 4688 | 11377 |
| 4793 | 5004 | 5250 | 9548 | 9899 | 10394 | 6293 | 12795 |
| 5004 | 6776 | 11719 | 9635 | 11914 | 10685 | 7891 | 13293 |
| 7028 | 8592 | 12438 | 11485 | 15067 | 11377 | 11533 | 14305 |
| 7202 | 10394 | 12810 | 11914 | 16519 | 14470 | 12007 | 16519 |
| 10211 | 10754 | 14770 | 15355 | | 14560 | 12795 | |
| 12858 | 11864 | 15833 | 15884 | | 16390 | 15043 | |
| 14470 | 15340 | 15884 | 16368 | | 14869 | 16390 | |
| 16519 | 16519 | 16519 | | | | 16519 | |
| | | | | | | | |
| H1 | H2 H3 H6 | V1 H5 | J1 J2 K2 K3 | J4 T2 T3 H4 | V1 H1 H2 H3 | J1 J3 T1 | K1 |

# SNPs for Forensic Discrimination

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 477 | 477 | 72 | 482 | 4808 | 64 | 3826 | 64 |
| 3010 | 3010 | 513 | 5198 | 5147 | 4745 | 3834 | 4688 |
| 4580 | 3915 | 4580 | 6260 | 9380 | 10211 | 4688 | 11377 |
| 4793 | 5004 | 5250 | 9548 | 9899 | 10394 | 6293 | 12795 |
| 5004 | 6776 | 11719 | 9635 | 11914 | 10685 | 7891 | 13293 |
| 7028 | 8592 | 12438 | 11485 | 15067 | 11377 | 11533 | 14305 |
| 7202 | 10394 | 12810 | 11914 | 16519 | 14470 | 12007 | 16519 |
| 10211 | 10754 | 14770 | 15355 | | 14560 | 12795 | |
| 12858 | 11864 | 15833 | 15884 | | 16390 | 15043 | |
| 14470 | 15340 | 15884 | 16368 | | 14869 | 16390 | |
| 16519 | 16519 | 16519 | | | | 16519 | |
| | | | | | | | |
| H1 | H2 H3 H6 | V1 H5 | J1 J2 K2 K3 | J4 T2 T3 H4 | V1 H1 H2 H3 | J1 J3 T1 | K1 |

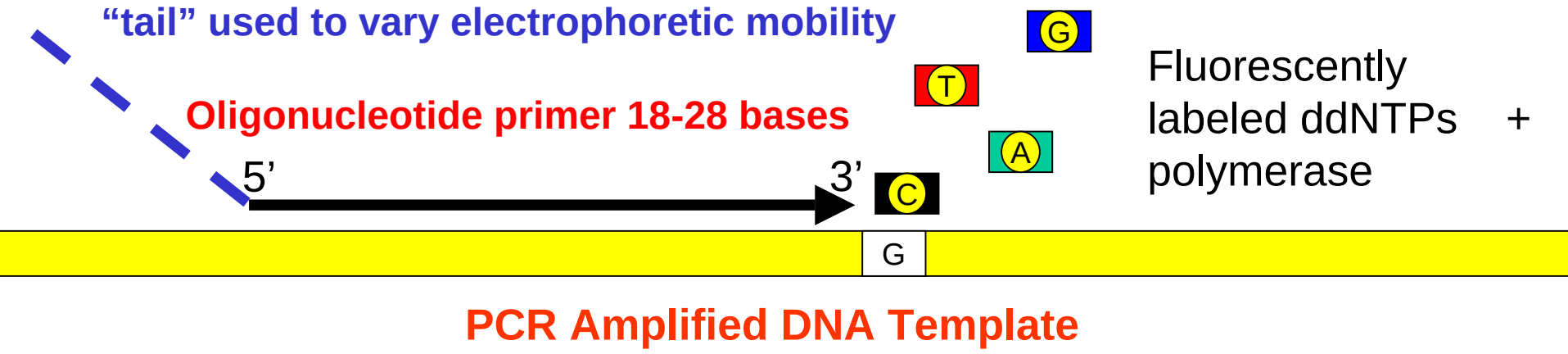
# SNPs for Forensic Discrimination

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 477 | 477 | 72 | 482 | 4808 | 64 | 3826 | 64 |
| 3010 | 3010 | 513 | 5198 | 5147 | 4745 | 3834 | 4688 |
| 4580 | 3915 | 4580 | 6260 | 9380 | 10211 | 4688 | 11377 |
| 4793 | 5004 | 5250 | 9548 | 9899 | 10394 | 6293 | 12795 |
| 5004 | 6776 | 11719 | 9635 | 11914 | 10685 | 7891 | 13293 |
| 7028 | 8592 | 12438 | 11485 | 15067 | 11377 | 11533 | 14305 |
| 7202 | 10394 | 12810 | 11914 | 16519 | 14470 | 12007 | 16519 |
| 10211 | 10754 | 14770 | 15355 | | 14560 | 12795 | |
| 12858 | 11864 | 15833 | 15884 | | 16390 | 15043 | |
| 14470 | 15340 | 15884 | 16368 | | 14869 | 16390 | |
| 16519 | 16519 | 16519 | | | | 16519 | |
| | | | | | | | |
| H1 | H2 H3 H6 | V1 H5 | J1 J2 K2 K3 | J4 T2 T3 H4 | V1 H1 H2 H3 | J1 J3 T1 | K1 |

# Allele-Specific Primer Extension

ABI PRISM® SNaPshot™ Multiplex System

SNP Primer is extended by one base unit

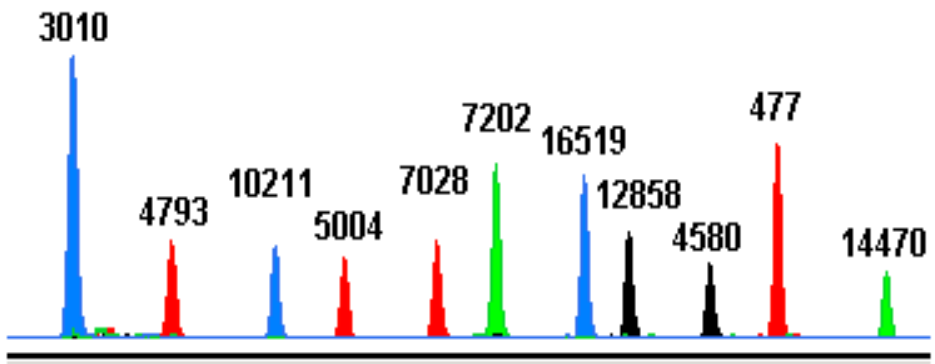



Products can be electrophoretically separated on an ABI 310, 3100

# The SNaPShot™ Platform

| Locus | SNP Primer Sequence | Length |
|---|---|---|
| 3010-F | TCAGAAGTGAAAGGGGGC | 18/na |
| 4763-R | TTTTTTTGTTGGATCAGGACATCCC | 19/26 |
| 10211-R | TTTTTTTTTACTAAGAAGAATTTTATGGA | 20/30 |
| 5004-F | TTTTTTTTTTTTTAGACCCAGCTACGCAAAATC | 20/34 |
| 7028-F | TTTTTTTTTTTTTTTTTGACACGTACTACGTTGTAGC | 20/38 |
| 7202-F | TTTTTTTTTTTTTTTTTTTTCCACAACACTTTCTCGGCCT | 20/42 |
| 16519-R | TTTTTTTTTTTTTTTTTTTTTTGTGGGCTATTTAGGCTTTATG | 22/46 |
| 12858-F | TTTTTTTTTTTTTTTTTTTTTTTTTGCAGCCATTCAAGCAATCCTATA | 23/50 |
| 4580-R | TTTTTTTTTTTTTTTTTTTTTTTTTTTGGTTAGAACTGGAATAAAAGCTAG | 25/54 |
| 477-F | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCTCCCACTCCCATACTAC | 20/58 |
| 14470-R | TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGAATGATGGTTGTCTTTGG | 21/62 |

Rebecca Hamm
Dr. Peter Vallone




Vallone et al. *IJLM* (2004) **118:** 147- 157.

# SNPs for Forensic Discrimination

18 common HV1/HV2 types, 241 individuals

# SNPs for Forensic Discrimination

18 common HV1/HV2 types, 241 individuals

+8 Multiplexes (59 SNPs)

105 types (55 “unique”)

# SNPs for Forensic Discrimination

18 common HV1/HV2 types, 241 individuals

| +8 Multiplexes (59 SNPs) | +8 Multiplexes (with AC indel) |
| --- | --- |
| 105 types (55 “unique”) | 112 types (64 “unique”) |

# SNPs for Forensic Discrimination

18 common HV1/HV2 types, 241 individuals

| +8 Multiplexes (59 SNPs) | +8 Multiplexes (with AC indel) |
| --- | --- |
| 105 types (55 “unique”) | 112 types (64 “unique”) |

6-fold improvement!

# The Nature of the SNPs

| A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|
| 477 | 477 | 72 | 482 | 4808 | 64 | 3826 | 64 |
| 3010 | 3010 | 513 | 5198 | 5147 | 4745 | 3834 | 4688 |
| 4580 | 3915 | 4580 | 6260 | 9380 | 10211 | 4688 | 11377 |
| 4793 | 5004 | 5250 | 9548 | 9899 | 10394 | 6293 | 12795 |
| 5004 | 6776 | 11719 | 9635 | 11914 | 10685 | 7891 | 13293 |
| 7028 | 8592 | 12438 | 11485 | 15067 | 11377 | 11533 | 14305 |
| 7202 | 10394 | 12810 | 11914 | 16519 | 14470 | 12007 | 16519 |
| 10211 | 10754 | 14770 | 15355 | | 14560 | 12795 | |
| 12858 | 11864 | 15833 | 15884 | | 16390 | 15043 | |
| 14470 | 15340 | 15884 | 16368 | | 14869 | 16390 | |
| 16519 | 16519 | 16519 | | | | 16519 | |
| | | | | | | | |
| H1 | H2 H3 H6 | V1 H5 | J1 J2 K2 K3 | J4 T2 T3 H4 | V1 H1 H2 H3 | J1 J3 T1 | K1 |

# The Nature of the SNPs

- Are the SNPs useful for discrimination mostly slow, rare types restricted to a particular HV1/HV2 type

(OR)

- Do the SNPs have a general utility across many different haplotypes?

- How should one proceed to identify SNPs to resolve common HV1/HV2 types in other forensically relevant populations (e.g. African American)?

# Why not survey the literature for Polymorphisms?

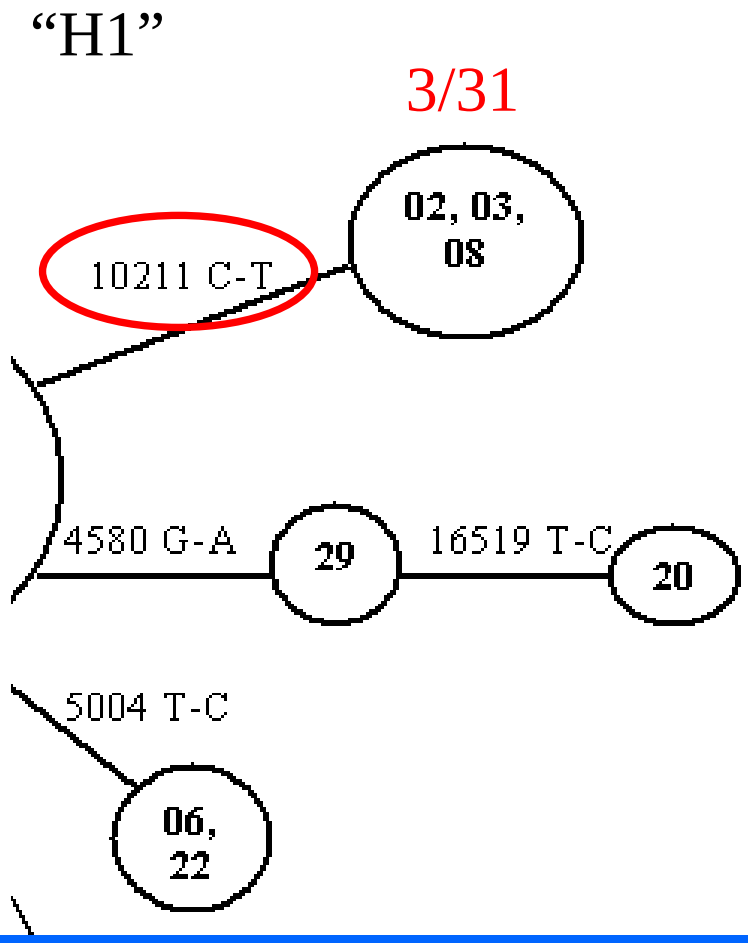
- Prior to Dec. 2000 - handful of complete human genomes (mostly RFLP data ~20% of the genome)
- Dec. 2000 - Ingman et al. (53 complete)
- June 2001 - Finnila et al. (192 genomes - CSGE)
- August 2001 - Maca-Myer et al. (42 complete)
- May 2002 - Herrnstadt et al. (560 coding only)

# Why not survey the literature for Polymorphisms?

- Prior to Dec. 2000 - handful of complete human genomes (mostly RFLP data ~20% of the genome)
- Dec. 2000 - Ingman et al. (53 complete)
- June 2001 - Finnila et al. (192 genomes - CSGE)
- August 2001 - Maca-Myer et al. (42 complete)
- May 2002 - Herrnstadt et al. (560 coding only)

**Problem - Very Few *Common* Types**

| Anderson | | 746 Sequences | | | | | |
|---|---|---|---|---|---|---|---|
| Posn. | Base | A | G | C | T | Gap | Ins |
| 10192 | C | 1 | | 743 | 2 | | |
| 10197 | G | 2 | 744 | | | | |
| 10199 | C | | | 745 | 1 | | |
| 10211 | C | | | 745 | 1 | | |
| 10238 | T | | | 31 | 715 | | |
| 10245 | T | | | 1 | 745 | | |




- mtDB - Human Mitochondrial Genome Database
- http://www.genpat.uu.se/mtDB/

# Recent recommendations to increase forensic discrimination

- Andreasson et al. (2002) – Sequenced short fragments of the mtGenome that are most informative
- Lee et al. (2002) – Sequenced the CytB gene for Koreans
- Lutz-Bonengel et al. (2003) – Sequenced the ATPase and ND4 genes (highly variable genes)
- Poetsch et al. (2003) – Sequenced the ATPase genes

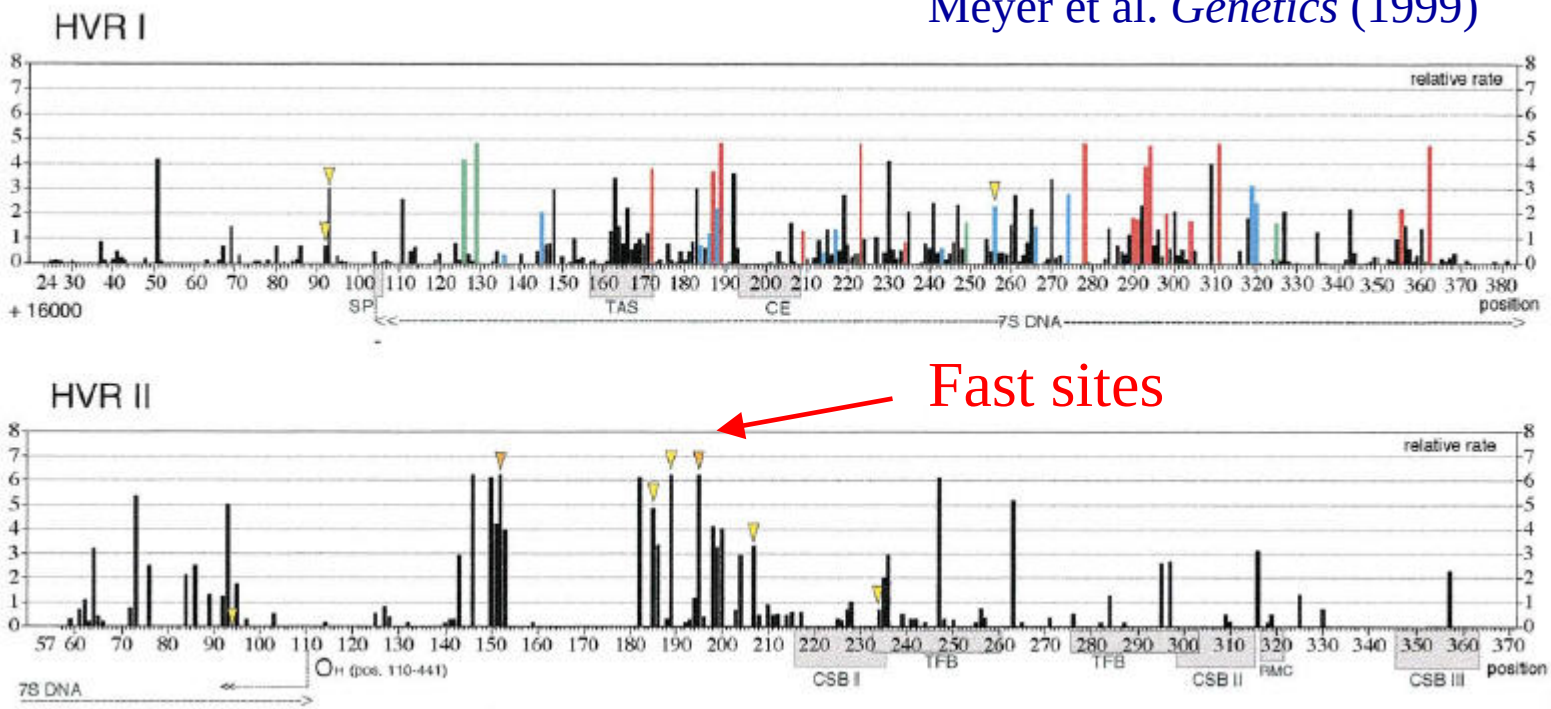
# Flaws with this approach

- Variation in one gene is not guaranteed (or likely) to resolve *common* types.
- Focus on one segment could miss SNPs scattered throughout the mtGenome.
- Unintended effect of revealing medically significant information.

# Mutation Rate Analysis in the mtDNA Coding Region

Mutation rate heterogeneity – the variation of mutation rates among sites.

Meyer et al. *Genetics* (1999)



# Mutation Rate Analysis in the mtDNA Control Region

Mutation rate heterogeneity – has been well characterized in the control region using a variety of methods for analysis (Parsimony, Maximum Likelihood, Pairwise Distance methods).

# Mutation Rate Analysis in the mtDNA Coding Region – Previous Assumptions (I)

- Eyre-Walker et al. *Proc. R. Soc. Lond B* 1999. Using partial DNA sequences of the human mtDNA genome (filled with errors), this group observed a significant amount of recurrent mutations (homoplasy) in their data.
- Conclusion – **Recombination!** (between paternal and maternal mtDNA)

# Mutation Rate Analysis in the mtDNA Coding Region – Previous Assumptions (I)

- Eyre-Walker et al. assume mutation rate ***Homo***geneity…

- "There is no evidence of variation in the mutation rate."

- (Mostly discredited for their poor data choice and method of calculating LD)

# Mutation Rate Analysis in the mtDNA Coding Region – Previous Assumptions (II)

- Herrnstadt et al. (2002) *AJHG* – 560 coding region sequences.

- "One important result to emerge from these studies is the ***relatively large number of sites*** at which ***homoplasic events*** have occurred."

(Referring to their Table 2)

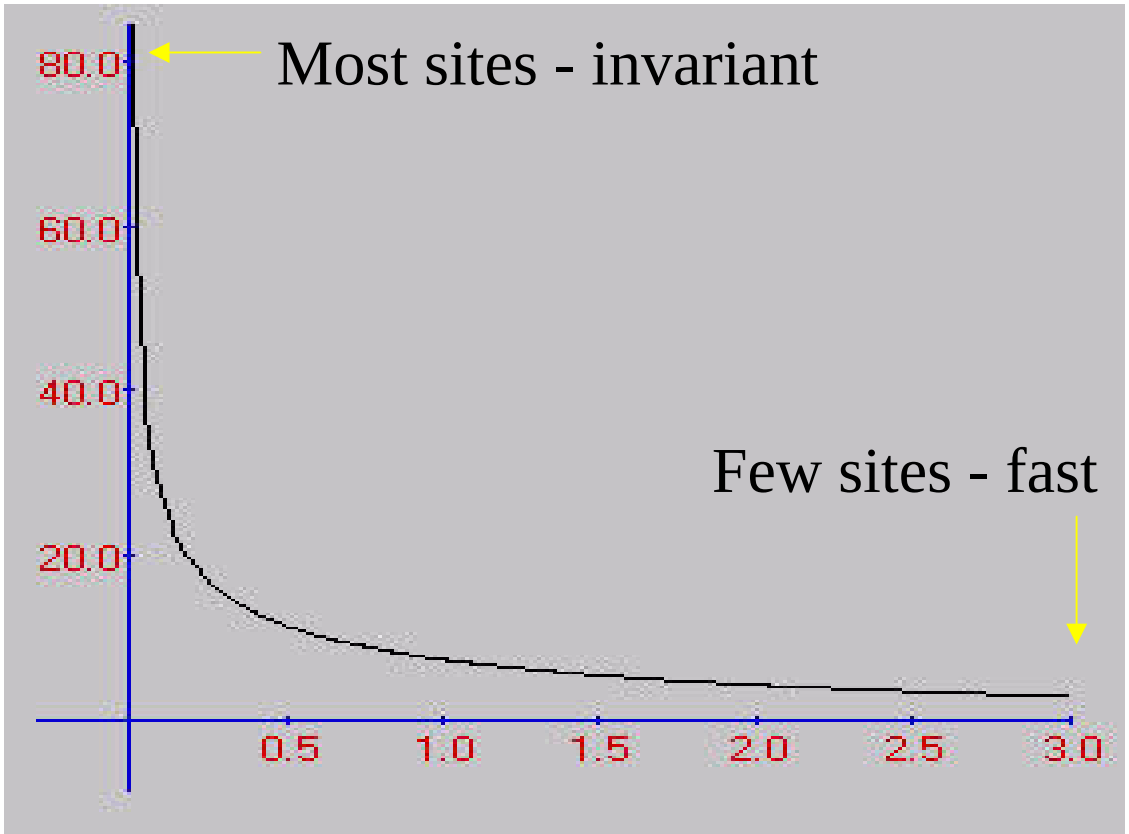
# Mutation Rate Analysis in the mtDNA Coding Region – Previous Assumptions (II)

- Yao et al. (2003) *AJHG* – in response to an Amerindian paper filled with sequence errors.

- "***Homoplasy*** in the coding region is ***much less*** than in the control region and may have ***only a few*** hot spots (see, e.g., table 2 of Herrnstadt et al. [2002])"

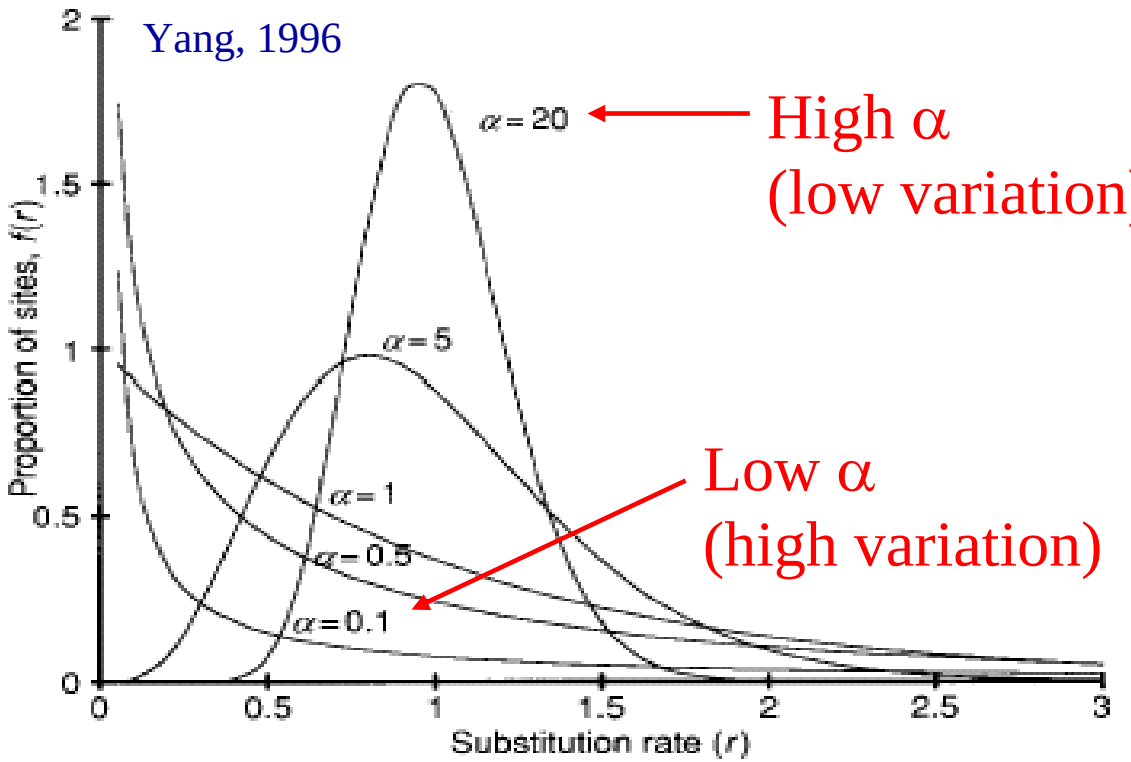
# How is Mutation Rate Variation Measured?

- Control region rates follow a negative binomial distribution (gamma distribution).



# How is Mutation Rate Variation Measured?

- The SHAPE of the curve (α) is inversely related to the amount of heterogeneity



# Current Literature

- Only one study has examined the mutation rate heterogeneity in the coding region.

- Meyer and von Haeseler (2003) *Mol. Biol Evol.* Analyzed the 53 mtGenomes from Ingman et al. (2000).

# Methods

- Parsimony analysis of phylogenetic trees (646 coding region sequences).
- Count the number of character changes mapped upon the MPT to determine the relative mutation rate.
- Calculate the $\alpha$ parameter using the method of Yang and Kumar (1996).

# Results

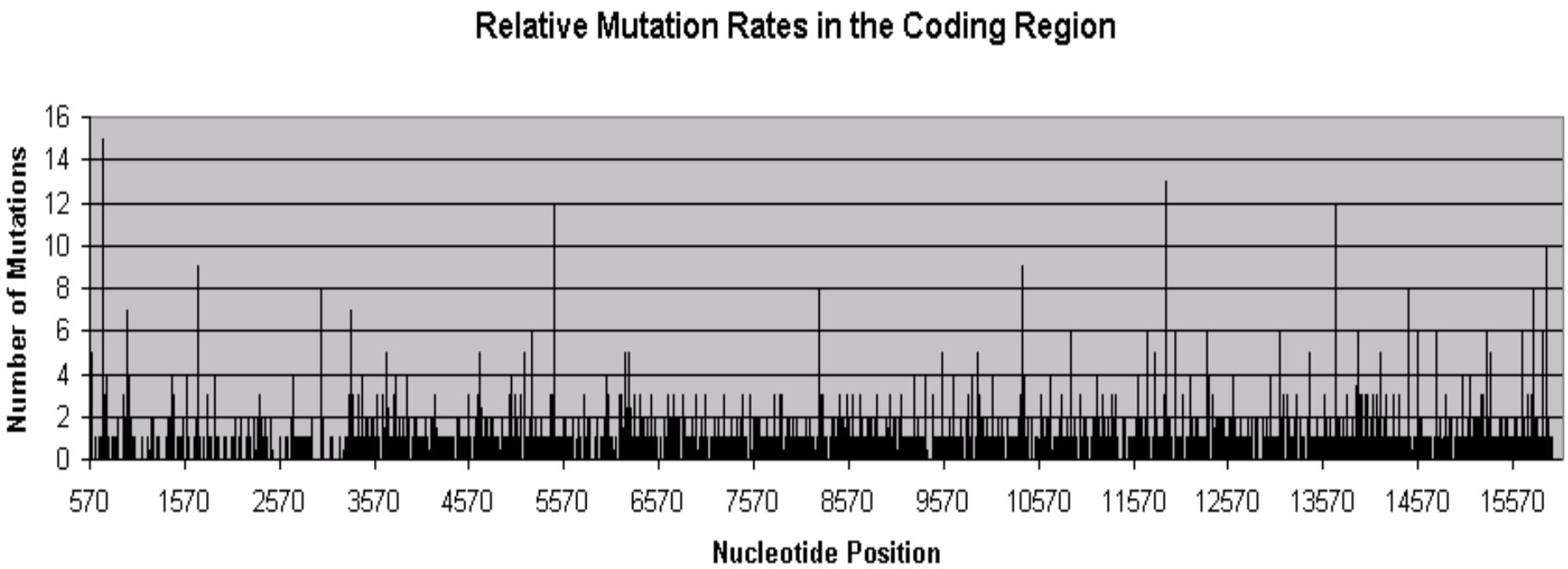
- Analysis of 646 coding region genomes.

| | Parsimony | | NJ | |
|---|---|---|---|---|
| **Data Set (# genomes)** | **Tree Length** | **α estimation** | **Tree Length** | **α estimation** |
| Ingman HV1 (53) | 144 | 0.2091 | 144 | 0.2081 |
| Ingman Control Region (53) | 273 | 0.0038 | 281 | 0.0036 |
| Ingman Coding Region (53) | 588 | 0.0075 | 588 | 0.0074 |
| Ingman Full Data (53) | 873 | 0.0050 | 876 | 0.0067 |
| Total Coding Data (646) | 2352 | 0.0086 | 2353 | 0.0083 |

Meyer and von Haeseler – α estimation = 0.002 (full data)

**Extreme** rate variation exists in the coding region

# Relative Mutation Rates



# The Mutation Rate Spectrum

| Length | Character | Gene | codon |
|---|---|---|---|
| | | | |
| 15 | 709 | 12S | * |
| | | | |
| 13 | 11914 | ND4 | 3 |
| | | | |
| 12 | 5460 | ND2 | 1 |
| 12 | 13708 | ND5 | 1 |
| | | | |
| 10 | 15924 | tRNA(thr) | * |
| | | | |
| 9 | 1719 | 16S | * |
| 9 | 10398 | ND3 | 1 |
| | | | |
| 8 | 3010 | 16S | * |
| 8 | 8251 | COII | 3 |
| 8 | 14470 | ND6 | 3 |
| 8 | 15784 | CYTB | 3 |
| | | | |
| 7 | 961 | 12S | * |
| 7 | 3316 | ND1 | 1 |
| | | | |
| 6 | 5237 | ND2 | 3 |
| 6 | 10915 | ND4 | 3 |
| 6 | 11719 | ND4 | 3 |
| 6 | 12007 | ND4 | 3 |
| 6 | 12346 | ND5 | 1 |
| 6 | 13105 | ND5 | 1 |
| 6 | 13928 | ND5 | 2 |
| 6 | 14569 | ND6 | 3 |
| 6 | 14766 | CYTB | 2 |
| 6 | 15301 | CYTB | 3 |
| 6 | 15670 | CYTB | 3 |
| 6 | 15884 | NC | - |

19/25 – Protein Coding

Synonymous sites = 11
Non-synonymous = 8

# The Mutation Rate Spectrum

- How does our rate spectrum compare to the rate spectrum of sites determined by the method of Meyer and von Haeseler (2003)?

# The Mutation Rate Spectrum

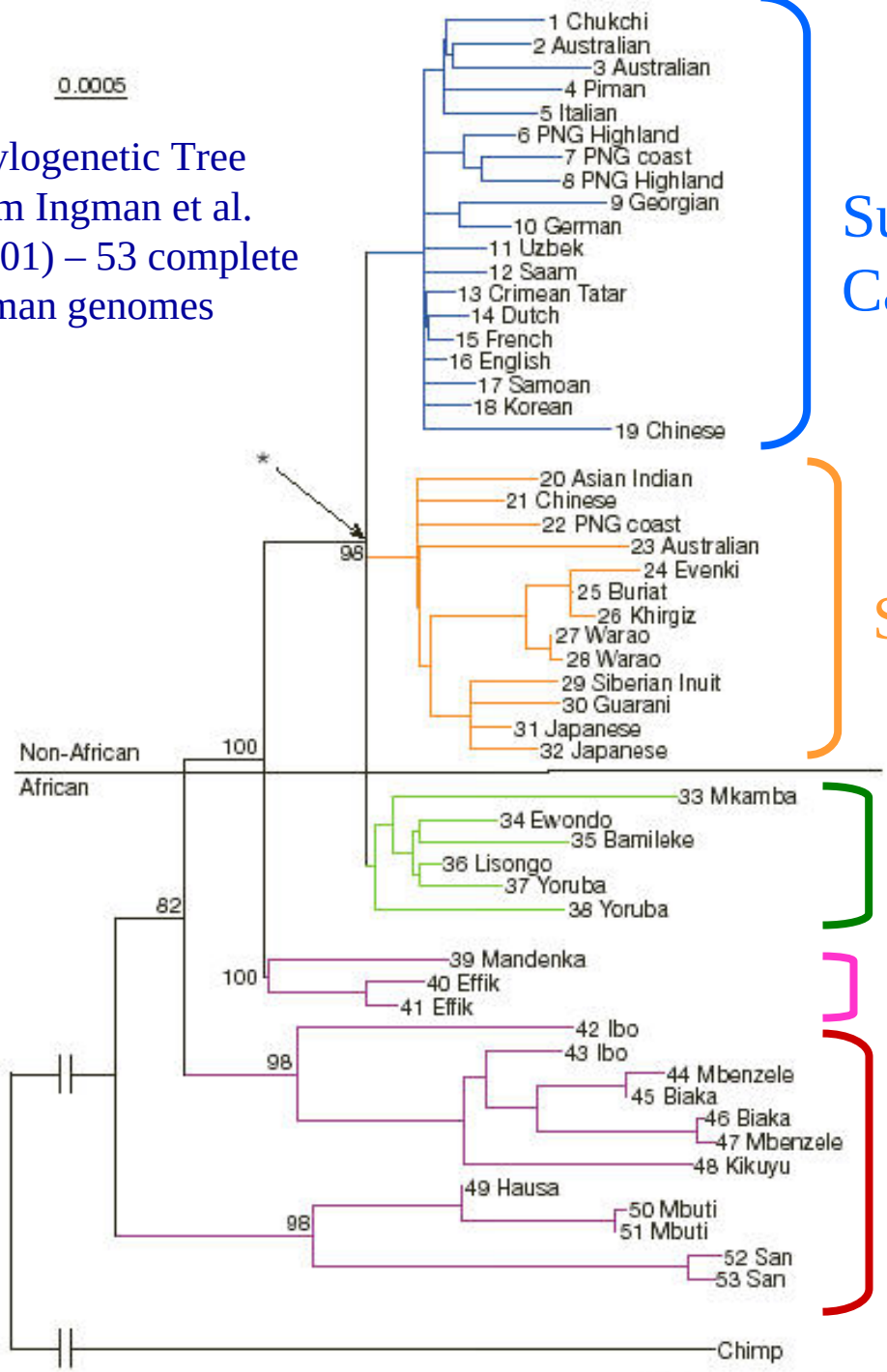
| Rate Score | Character | Length |
|---|---|---|
| | | |
| 175.21 | 15301 | 6 |
| | | |
| 162.82 | 10398 | 9 |
| | | |
| 155.20 | 8701 | 2 |
| 155.20 | 9540 | 1 |
| 155.20 | 10873 | 1 |
| | | |
| 129.16 | 12705 | 2 |
| | | |
| 119.30 | 7521 | 3 |
| | | |
| 112.03 | 769 | 1 |
| 112.03 | 1018 | 1 |
| 112.03 | 3594 | 1 |
| 112.03 | 4104 | 2 |
| 112.03 | 7256 | 3 |
| 112.03 | 13650 | 1 |
| | | |
| 105.84 | 11914 | 13 |
| | | |
| 100.77 | 10400 | 1 |
| 100.77 | 14783 | 1 |
| 100.77 | 15043 | 4 |
| | | |
| 96.96 | 10688 | 2 |
| 96.89 | 13105 | 7 |
| | | |
| 89.38 | 825 | 1 |
| 89.38 | 2758 | 1 |
| 89.38 | 2885 | 1 |
| 89.38 | 8468 | 1 |
| 89.38 | 8655 | 1 |
| 89.38 | 10810 | 2 |
| 89.38 | 13506 | 1 |

Only 4 sites shared among the top 26 fastest sites as determined by Meyer and von Haeseler (2003)

Most of the “fastest” sites change once on our MPT

?????

Phylogenetic Tree from Ingman et al. (2001) – 53 complete human genomes




Super-haplogroup N (Asian and Caucasian)

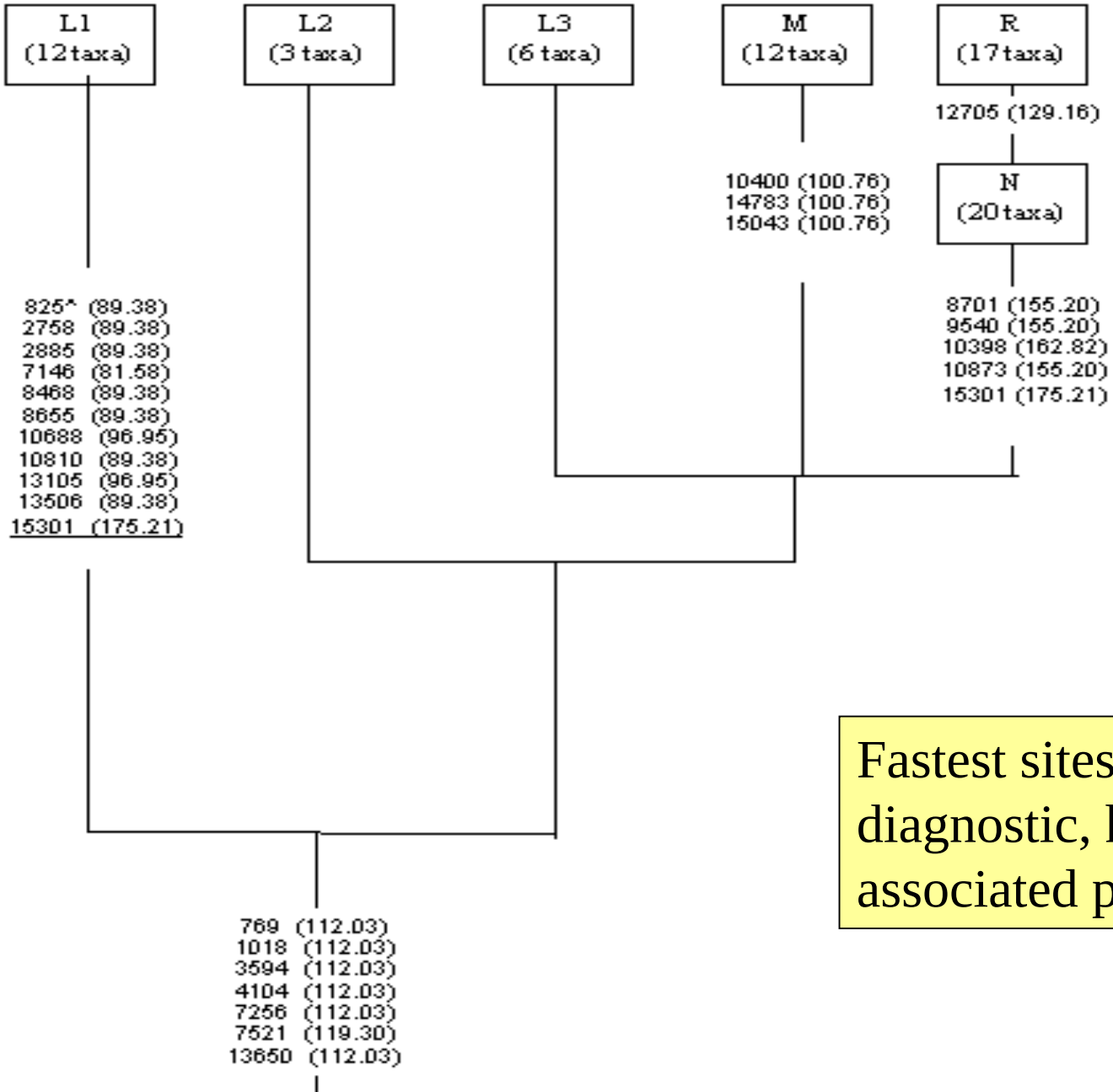
Super-haplogroup M (Asian)

Haplogroup L3 (African)

Haplogroup L2 (African)

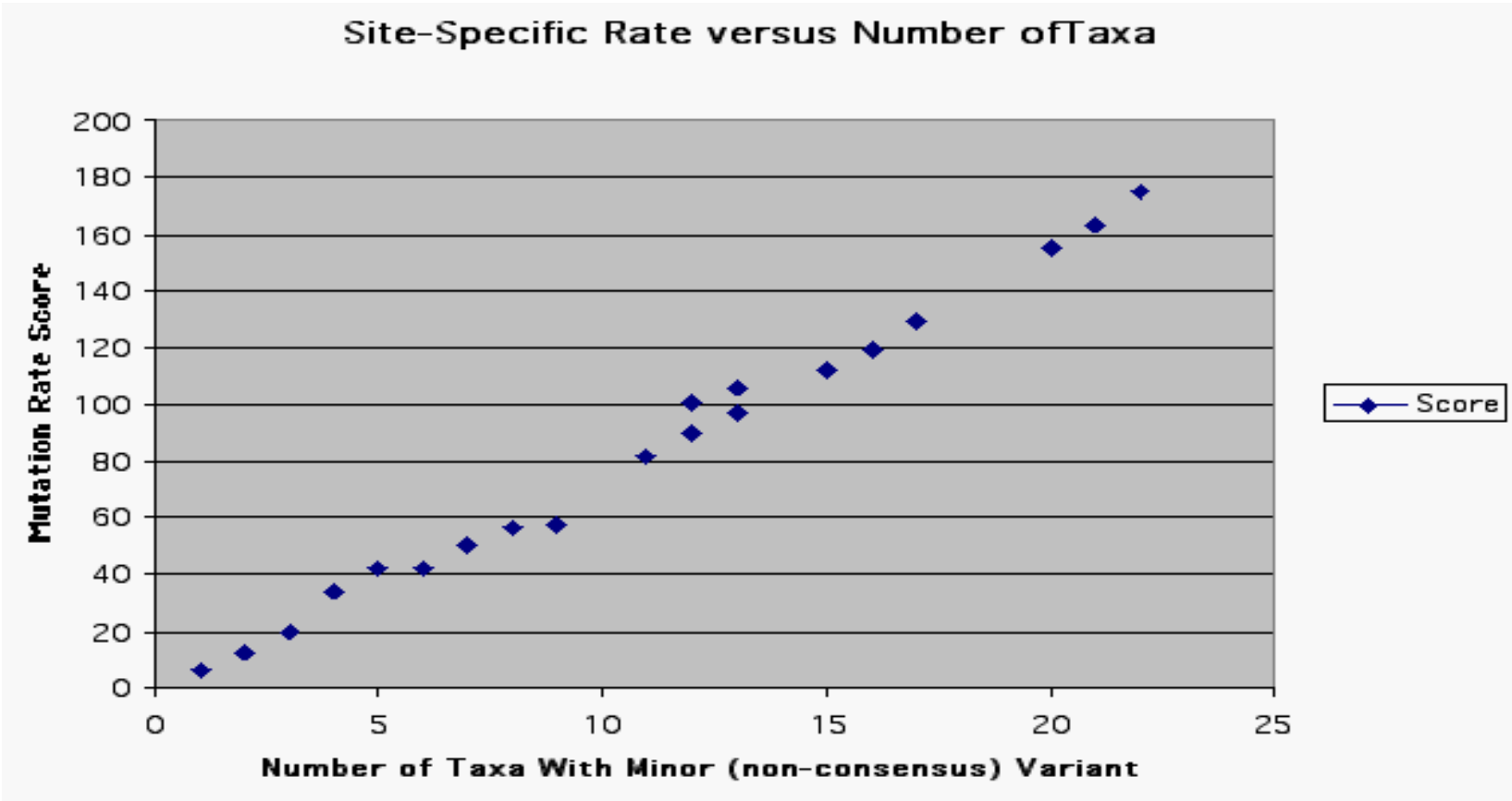
Haplogroup L1 (African)

# Skeleton Tree based on Human mtDNA Phylogeny

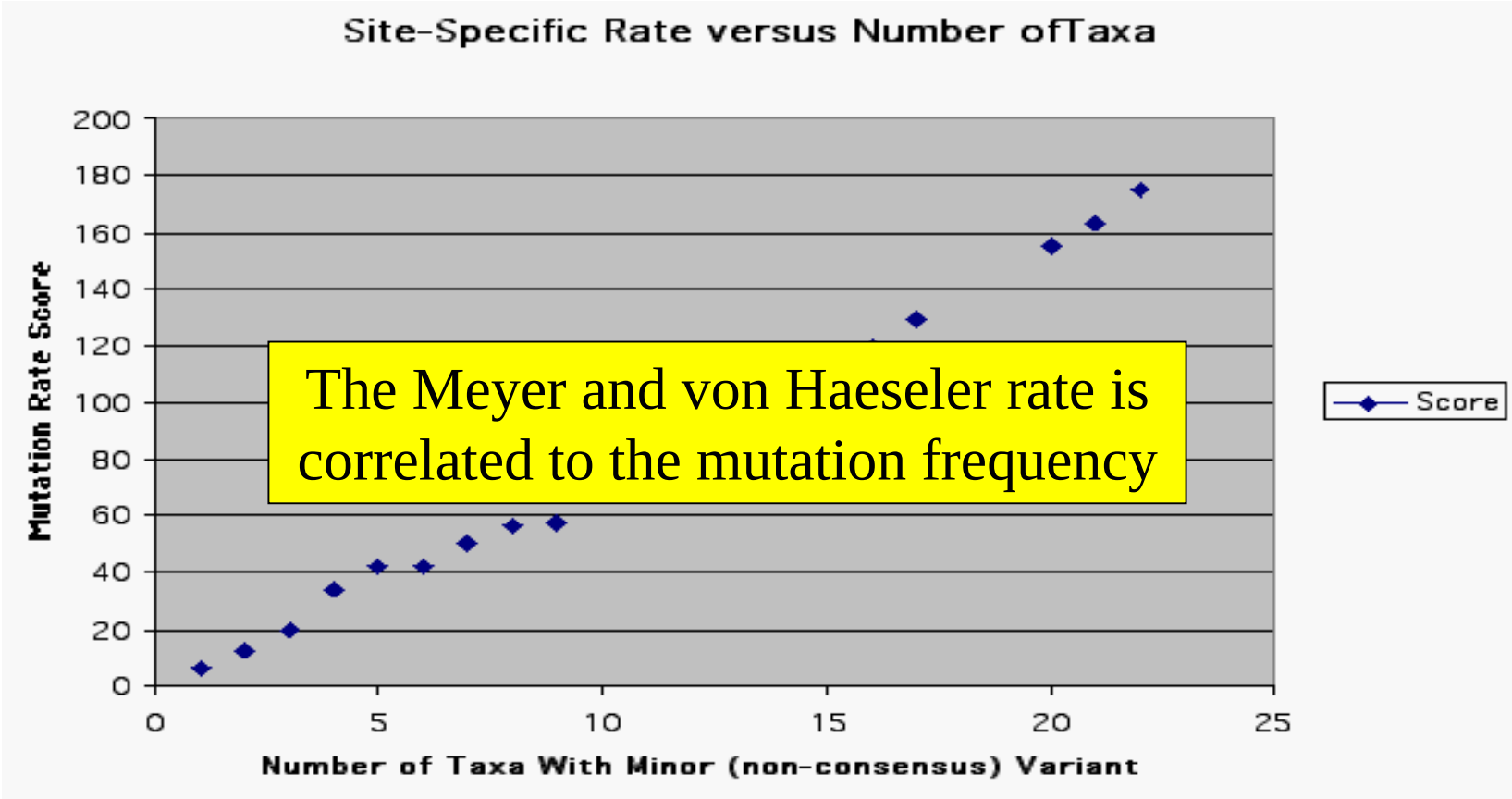



Fastest sites are actually diagnostic, haplogroup-associated polymorphisms!

# Pairwise Genetic Distances to Estimate Mutation Rates

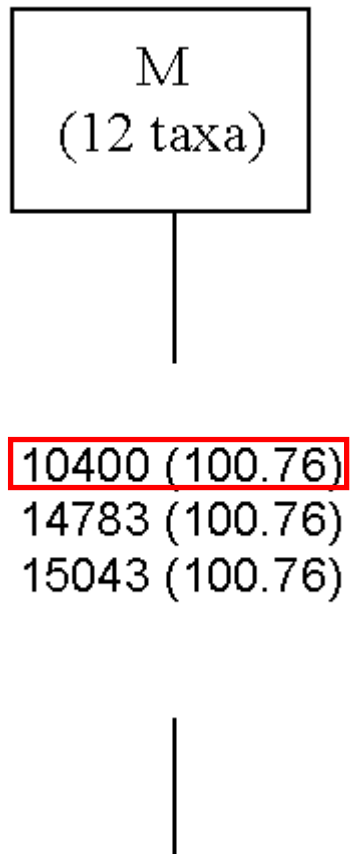


# Pairwise Genetic Distances to Estimate Mutation Rates



# The Mutation Rate Spectrum

"It is hard to believe that 10400 has actually mutated … because ***no single homoplasious change*** at this site has been observed in >900 coding-region sequences or fragments that cover site 10400…" (Yao et al. *AJHG* 2003 – in response to Silva et al. 2002).



# Mutation Rate Analysis and the 8 Multiplex SNP Panels

| Length | Character | Gene | codon | 241 Caucasians |
|---|---|---|---|---|
| | | | | |
| 15 | 709 | 12S | * | Yes |
| | | | | |
| 13 | 11914 | ND4 | 3 | Yes-SNP |
| | | | | |
| 12 | 5460 | ND2 | 1 | Yes |
| 12 | 13708 | ND5 | 1 | Yes |
| | | | | |
| 10 | 15924 | tRNA(thr) | * | Yes |
| | | | | |
| 9 | 1719 | 16S | * | Yes |
| 9 | 10398 | ND3 | 1 | Yes |
| | | | | |
| 8 | 3010 | 16S | * | Yes-SNP |
| 8 | 8251 | COII | 3 | |
| 8 | 14470 | ND6 | 3 | Yes-SNP |
| 8 | 15784 | CYTB | 3 | |
| | | | | |
| 7 | 961 | 12S | * | |
| 7 | 3316 | ND1 | 1 | |
| | | | | |
| 6 | 5237 | ND2 | 3 | Yes |
| 6 | 10915 | ND4 | 3 | Yes |
| 6 | 11719 | ND4 | 3 | Yes-SNP |
| 6 | 12007 | ND4 | 3 | Yes-SNP |
| 6 | 12346 | ND5 | 1 | |
| 6 | 13105 | ND5 | 1 | Yes |
| 6 | 13928 | ND5 | 2 | |
| 6 | 14569 | ND6 | 3 | |
| 6 | 14766 | CYTB | 2 | |
| 6 | 15301 | CYTB | 3 | |
| 6 | 15670 | CYTB | 3 | |
| 6 | 15884 | CYTB | nc | Yes-SNP |

Only 6 of the 59 SNPs are among the “fastest” sites

# Mutation Rate Analysis and the 8 Multiplex SNP Panels

| Length | Character | Gene | codon | 241 Caucasians |
|---|---|---|---|---|
| | | | | |
| 15 | 709 | 12S | * | Yes |
| | | | | |
| 13 | 11914 | ND4 | 3 | Yes-SNP |
| | | | | |
| 12 | 5460 | ND2 | 1 | Yes |
| 12 | 13708 | ND5 | 1 | Yes |
| | | | | |
| 10 | 15924 | tRNA(thr) | * | Yes |
| | | | | |
| 9 | 1719 | 16S | * | Yes |
| 9 | 10398 | ND3 | 1 | Yes |
| | | | | |
| 8 | 3010 | 16S | * | Yes-SNP |
| 8 | 8251 | COII | 3 | |
| 8 | 14470 | ND6 | 3 | Yes-SNP |
| 8 | 15784 | CYTB | 3 | |
| | | | | |
| 7 | 961 | 12S | * | |
| 7 | 3316 | ND1 | 1 | |
| | | | | |
| 6 | 5237 | ND2 | 3 | Yes |
| 6 | 10915 | ND4 | 3 | Yes |
| 6 | 11719 | ND4 | 3 | Yes-SNP |
| 6 | 12007 | ND4 | 3 | Yes-SNP |
| 6 | 12346 | ND5 | 1 | |
| 6 | 13105 | ND5 | 1 | Yes |
| 6 | 13928 | ND5 | 2 | |
| 6 | 14569 | ND6 | 3 | |
| 6 | 14766 | CYTB | 2 | |
| 6 | 15301 | CYTB | 3 | |
| 6 | 15670 | CYTB | 3 | |
| 6 | 15884 | CYTB | nc | Yes-SNP |

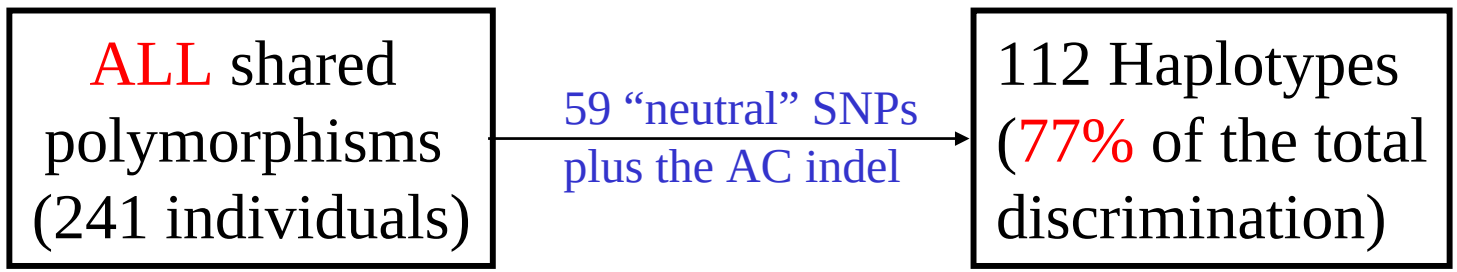
What about These highly polymorphic mutations?

# Mutation Rate Analysis and the 8 Multiplex SNP Panels

- How much information is lost by focusing only on mutations not associated with a potential for changing the phenotype?

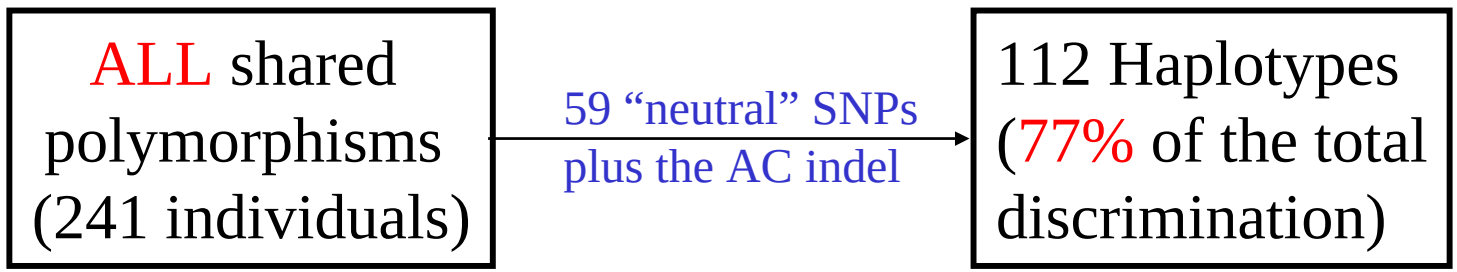
# Mutation Rate Analysis and the 8 Multiplex SNP Panels

- How much information is lost by focusing only on mutations not associated with a potential for changing the phenotype?



# Mutation Rate Analysis and the 8 Multiplex SNP Panels

- How much information is lost by focusing only on mutations not associated with a potential for changing the phenotype?




Additional SNP panels with fast, non-synonymous sites that vary widely in the population have been developed.

# A Case Example

Skeletal remains - “H1” in the HV1/HV2 region.

Thought to belong to one of two individuals (Smith or Jones)

Family references for Smith and Jones were obtained.

| Smith Family | Jones Family |
| --- | --- |
| 263 A-G | 263 A-G |
| 315.1 C | 315.1 C |

# A Case Example

Skeletal remains - “H1” in the HV1/HV2 region.

Thought to belong to one of two individuals (Smith or Jones)

Family references for Smith and Jones were obtained.

| Smith Family | Jones Family |
|---|---|
| 263 A-G | 263 A-G |
| 315.1 C | 315.1 C |
| 477 T-C | 16519 T-C |
| 16519 T-C | |

Remains tested for VR region: 477 T-C and 16519 T-C

# A Case Example

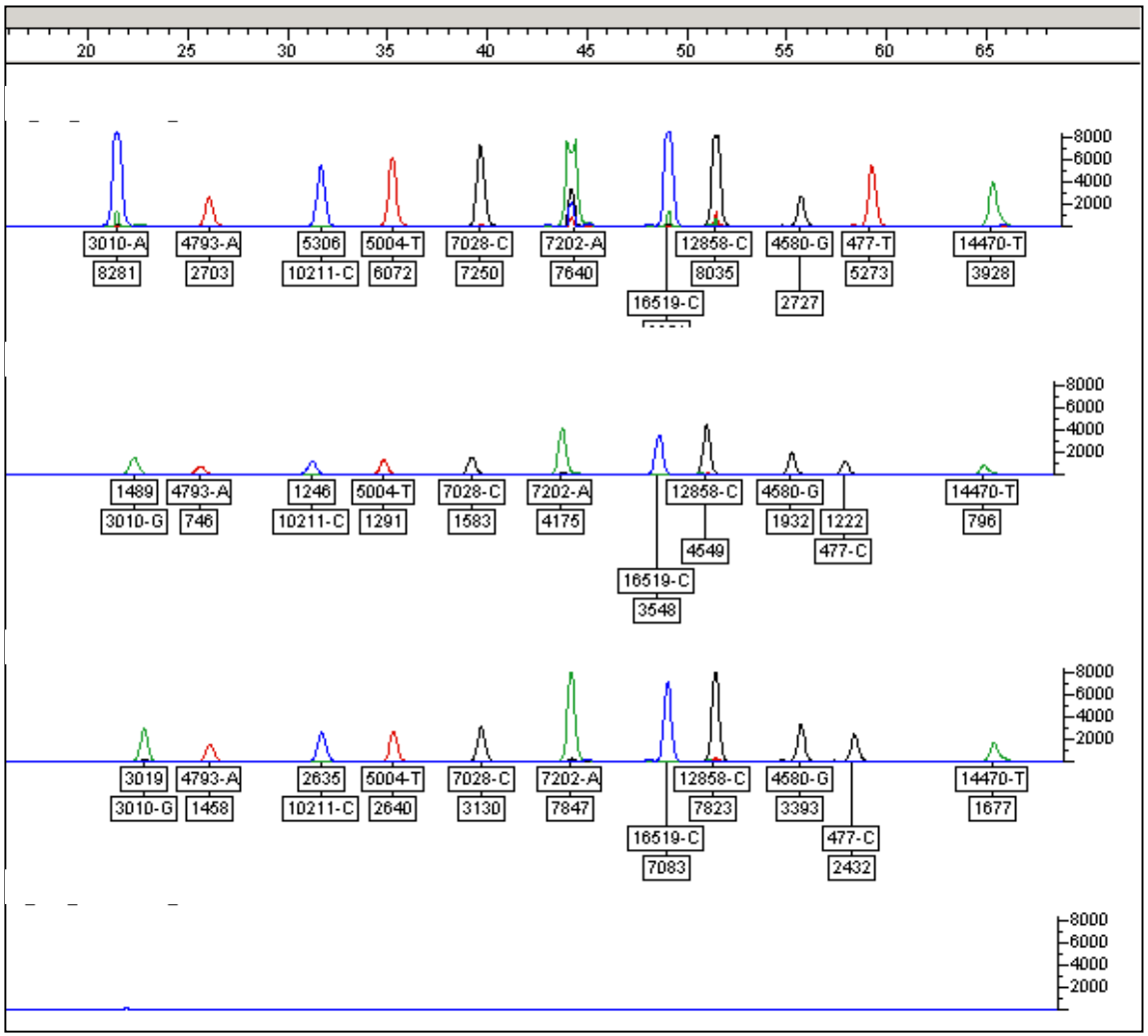
| Smith Family | Jones Family |
| --- | --- |
| 263 A-G | 263 A-G |
| 315.1 C | 315.1 C |
| 477 T-C | 16519 T-C |
| 16519 T-C | |

Remains tested for VR region: 477 T-C and 16519 T-C

Since there was a single difference between the remains and the Jones family, AFDIL could not make an exclusion

# A Case Example

The remains and the family references were typed with multiplex A




Jones Reference

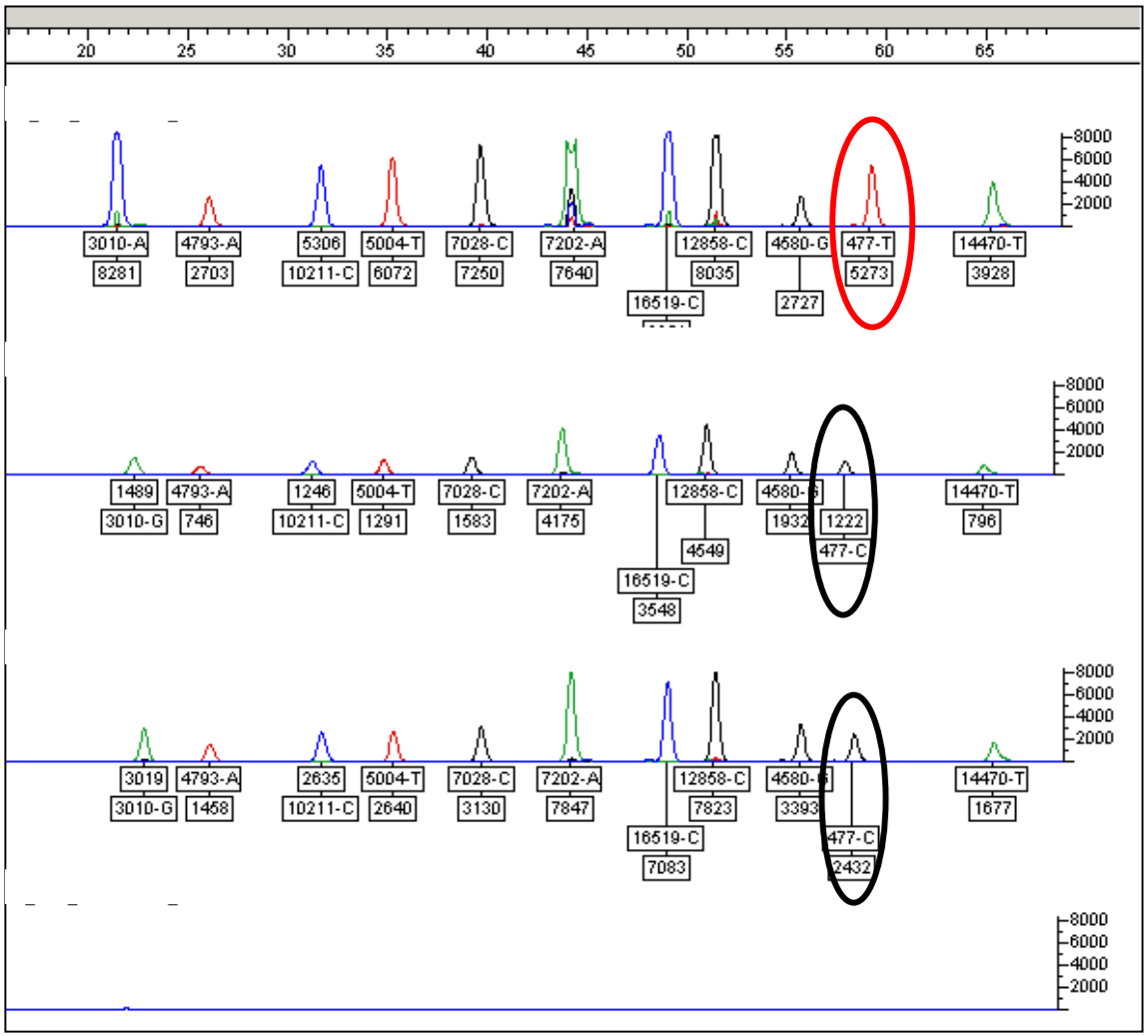
Smith Reference #1

Smith Reference #2

Negative Control

# A Case Example

Reference extracts confirmed the polymorphism at 477.




Jones Reference

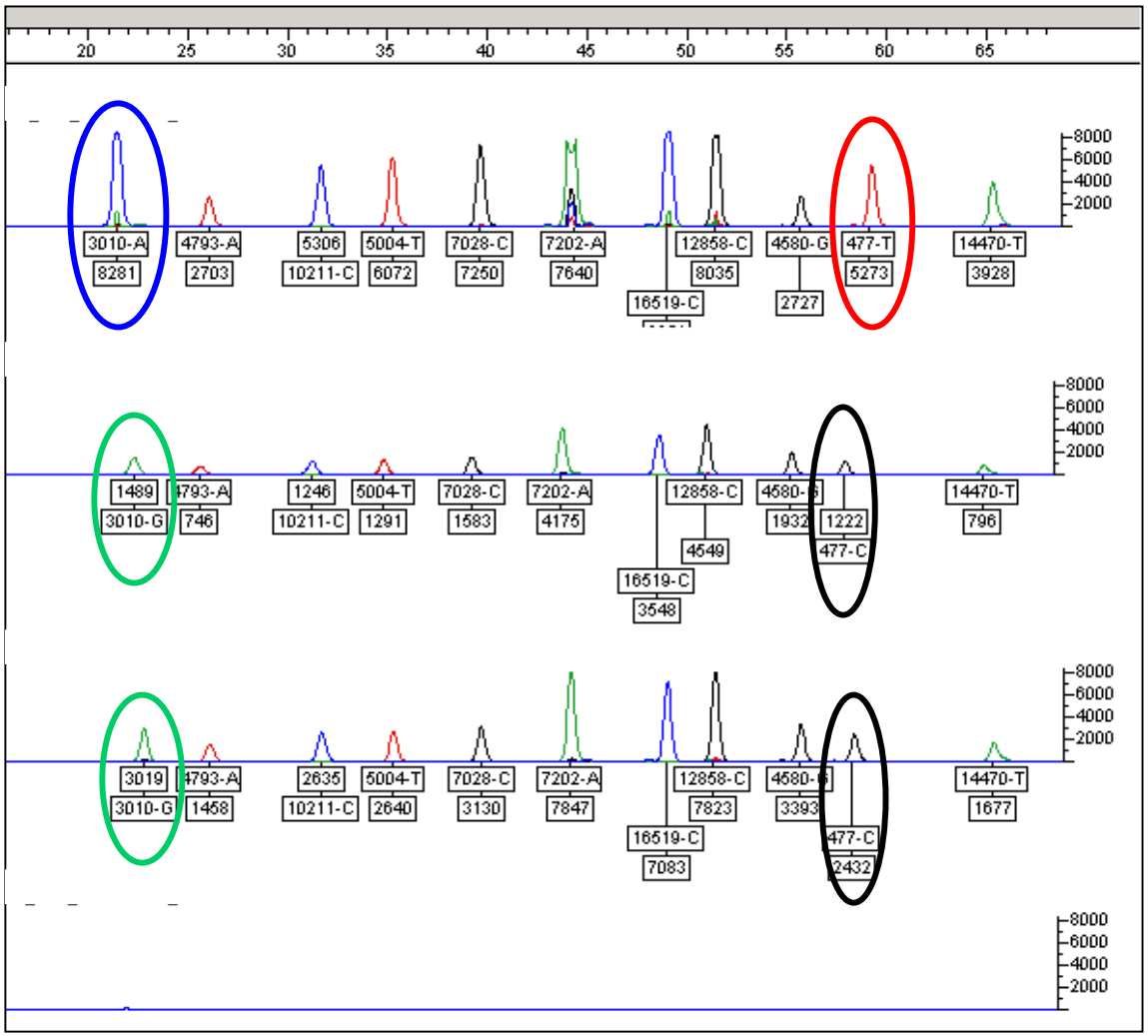
Smith Reference #1

Smith Reference #2

Negative Control

# A Case Example

An additional difference was observed at position 3010.



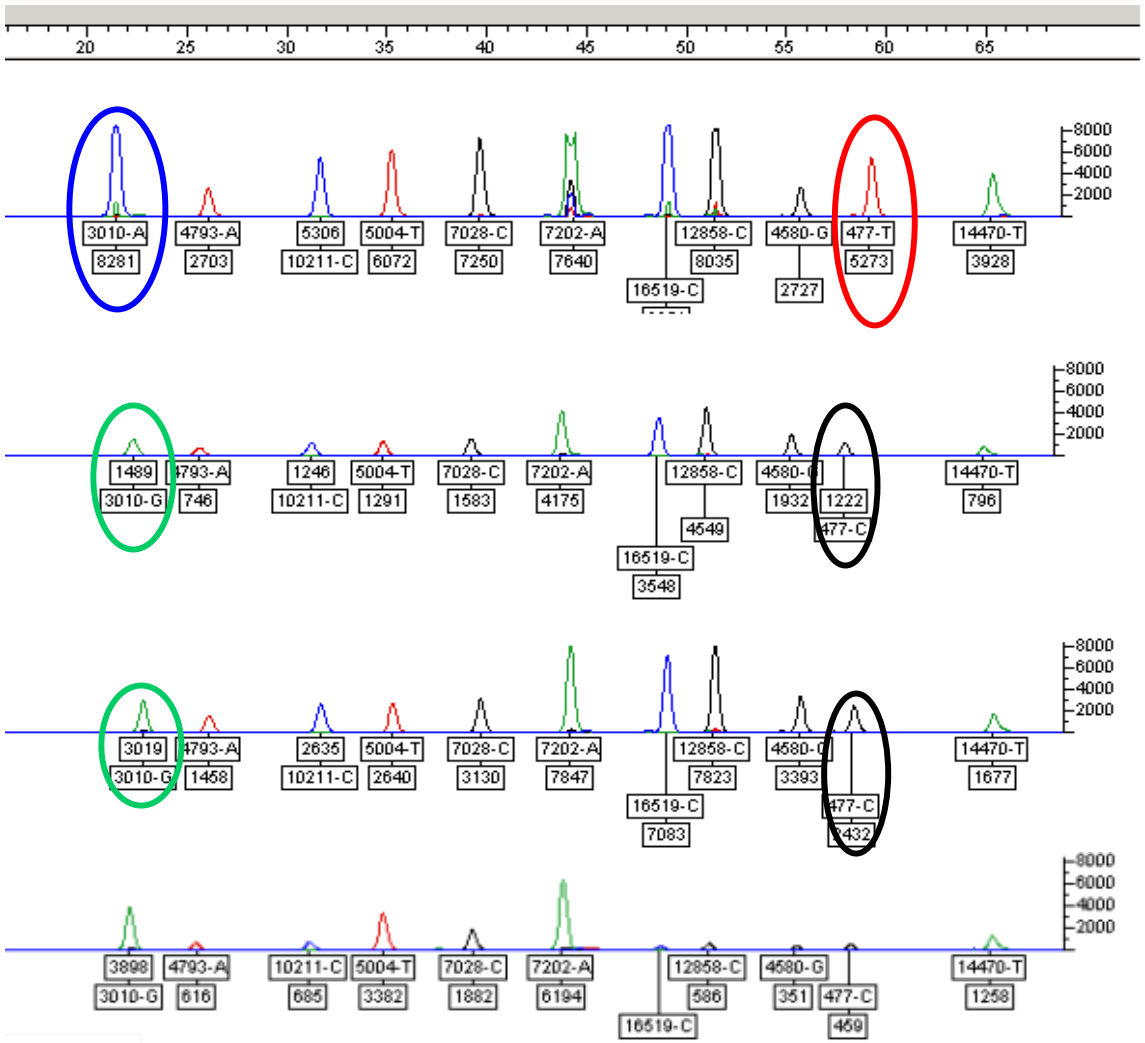

Jones Reference

Smith Reference #1

Smith Reference #2

Negative Control

# A Case Example




Jones Reference

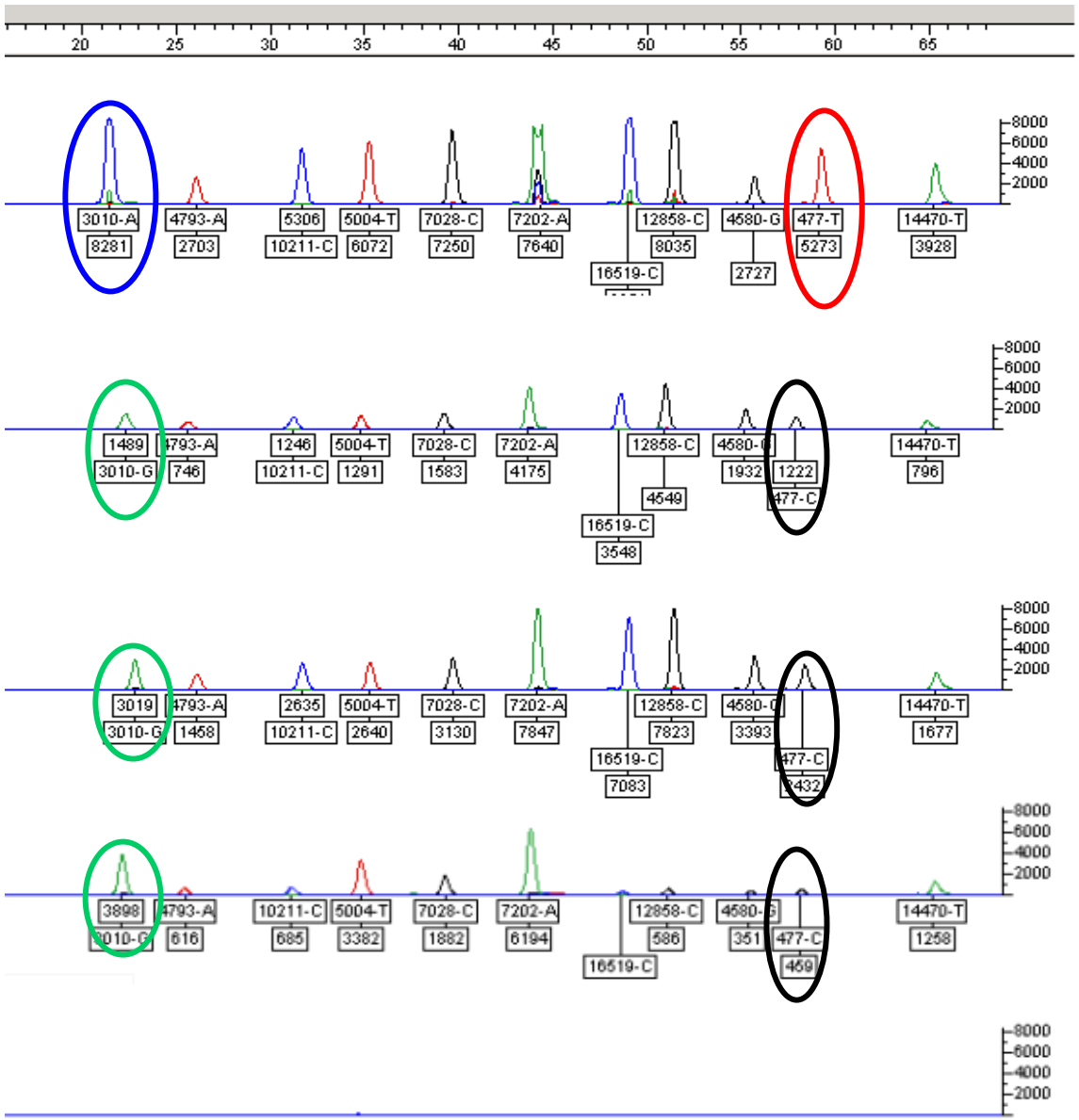
Smith Reference #1

Smith Reference #2

Bone Extract

15uL Reaction; 0.07Units/uL Taq; 31 cycles --- 100 RFU cutoff

# A Case Example




Jones Reference

Smith Reference #1

Smith Reference #2

Bone Extract

Negative Control

# A Case Example

| Smith Family | Skeletal Remains | Jones Family |
|---|---|---|
| 263 A-G | 263 A-G | 263 A-G |
| 315.1 C | 315.1 C | 315.1 C |
| 477 T-C | 477 T-C | 16519 T-C |
| 3010 A-G | 3010 A-G | |
| 16519 T-C | 16519 T-C | |

Remains – match exactly the Smith family, now 2 differences from the Jones family – can be excluded.

# Summary

- Purpose – Maximize Discrimination.
- A supplement to current HV1/HV2 testing.
- When the Forensic Scientist encounters a common type, select the most discriminating SNP panel.

# Summary

- We – focused on sites that are not associated with the potential for phenotypic change.

- Most of the informative sites are rare, slow polymorphisms that are useful for discrimination in a particular common type.

- A few SNP sites may be useful for resolving common HV1/HV2 types from various backgrounds.

# Summary

- Mutation rate analysis of the coding region using parsimony-evaluated phylogenetic trees revealed extreme rate variation using a relatively large data set.

- Parsimony distinguished fast sites from slow, haplogroup-associated polymorphisms (compared to Meyer and von Haeseler, 2003).

# Summary/Future Goals

- Future efforts to identify discriminatory SNPs to resolve common types in other populations – will require whole genome sequencing.

- Evaluation of  non-synonymous sites that are not associated with diseases and are useful for forensic discrimination.

# Publications

Michael D. Coble · Rebecca S. Just
Jennifer E. O'Callaghan · Ilona H. Letmanyi
Christine T. Peterson · Jodi A. Irwin · Thomas J. Parsons

**Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians**

*IJLM* (2004) **118:** 137-146.

Peter M. Vallone · Rebecca S. Just · Michael D. Coble
John M. Butler · Thomas J. Parsons

**A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome**

*IJLM* (2004) **118:** 147- 157.

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

# Acknowledgements

National Institutes of Justice
2000-1J-CX-K010 and Dr. Lois Tully

**The Armed Forces DNA Identification Laboratory**

Co-Authors
(AFDIL)
Dr. Thomas Parsons
Rebecca Just
Jessica Saunier
Jennifer O'Callaghan
Christine Peterson
Ilona Letmanyi
Jodi Irwin

(NIST)
Dr. Peter Vallone
Dr. John Butler

Research Interns
Rachel Barry
Trina Bersola
Serena Filosa
Victoria Glynn
Carrie Guyan
William Ivory
Devon Pierce

Administration (AFDIL)
Col. Brion Smith
James Canik
Scott Carroll

Casework Section (AFDIL)
Suzie Barritt

ILM – Innsbruck, Austria
Dr. Walther Parson
Harold Niederstaetter

http://www.cstl.nist.gov/biotech/strbase/Coble.htm
michael.coble@nist.gov