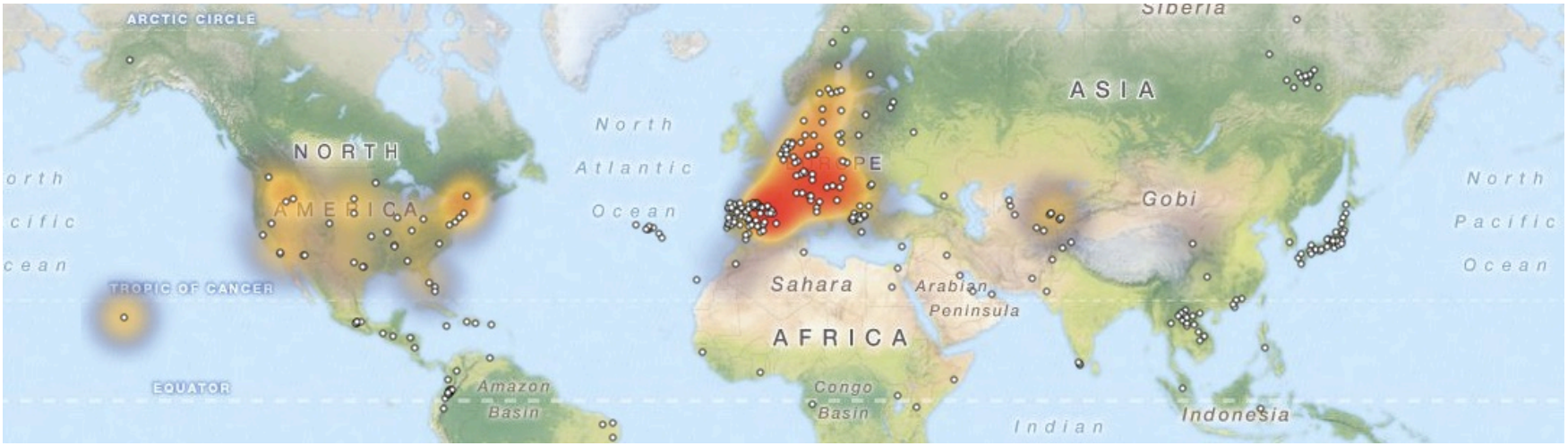
The use of mtDNA in forensics - Lausanne, Switzerland, Nov 24 2015

EMPOP mtDNA database



Introduction to mtDNA and its use in a forensic laboratory

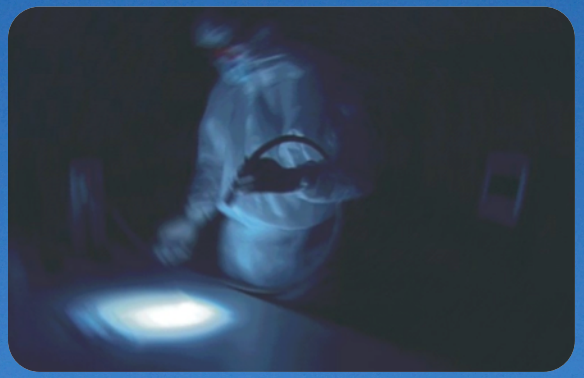
Dr. Walther Parson
assoc. Professor, Institute of Legal Medicine, Innsbruck
adj. Professor, Penn State University, PA, USA
walther.parson@gmail.com

CSI laboratory (ISO17025)

Austrian central DNA laboratory

EDNAP, ENFSI, Interpol

Int. reference lab for forensic mtDNA typing



DVI laboratory (ISO17025)

Tsunami (Sri Lanka), Chile, Former Yugoslavia,…

Forensic molecular research laboratory

Mitochondrial DNA databasing (EMPOP)

STR databasing (ENFSI STRbase)

Population genetics (mito, nDNA)

Predictive Forensic Markers (AIMs, phenotypic)

New technologies (MS, RAPID, MPS)

What to expect today

Introduction to mtDNA

Lab protocols

(a posteriori) quality control

mtDNA alignment

mtDNA haplogroups

Exclusion criteria

Updated ISFG guidelines on mtDNA typing

practical EMPOP work

ISFG mtDNA recommendations

Forensic Science International: Genetics 13 (2014) 134–142

Contents lists available at ScienceDirect

Forensic Science International: Genetics

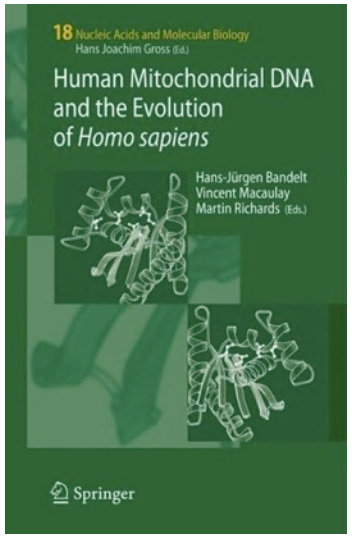
journal homepage: www.elsevier.com/locate/fsig

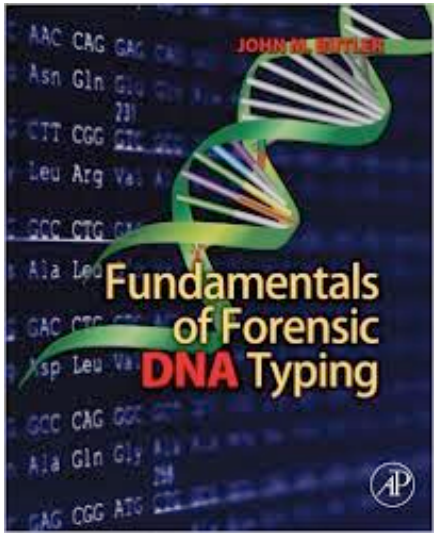
DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing

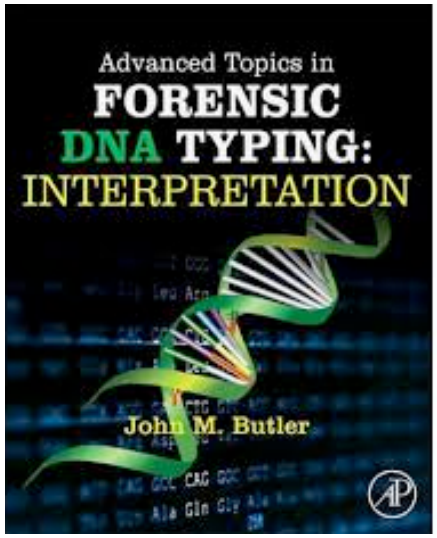
W. Parson [a,b,*], L. Gusmão [c,d], D.R. Hares [e], J.A. Irwin [e], W.R. Mayr [f], N. Morling [g], E. Pokorak [e], M. Prinz [h], A. Salas [i], P.M. Schneider [j], T.J. Parsons [k]

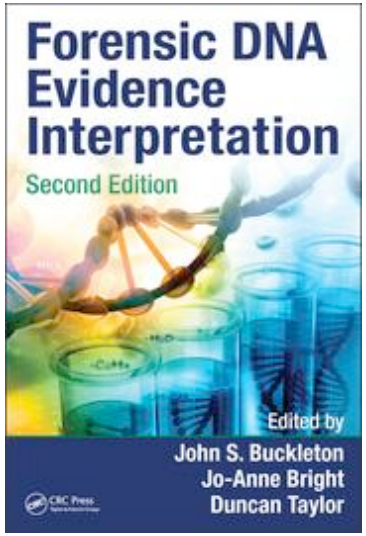
[a] Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[b] Penn State Eberly College of Science, University Park, PA, USA
[c] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Brazil
[d] IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, Portugal
[e] FBI Laboratory, Quantico, VA, USA
[f] Division of Blood Group Serology, Medical University of Vienna, Austria
[g] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[h] Department of Sciences, John Jay College for Criminal Justice, New York, NY, USA
[i] Unidade de Xenética, Departamento de AnatomíaPatolóxica e CienciasForenses, and Instituto de CienciasForenses, Grupo de MedicinaXenómica (GMX), Facultade de Medicina, Universidade de Santiago de Compostela, 15872 Galicia, Spain
[j] Institute of Legal Medicine, Medical Faculty, University of Cologne, Cologne, Germany
[k] International Commission on Missing Persons, Alipasina 45a, 71000 Sarajevo, Bosnia and Herzegovina

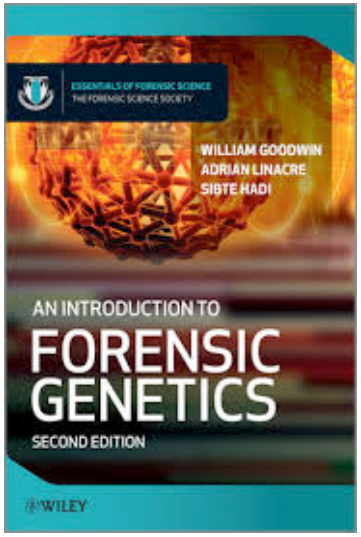
Books on mitochondria / mtDNA

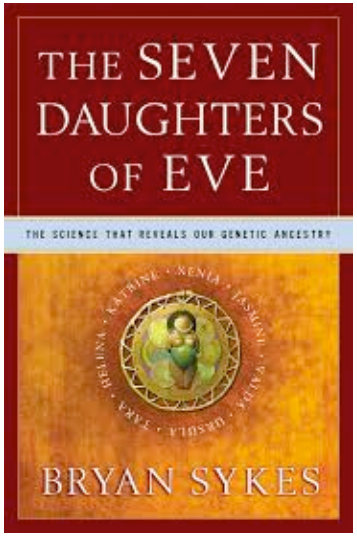


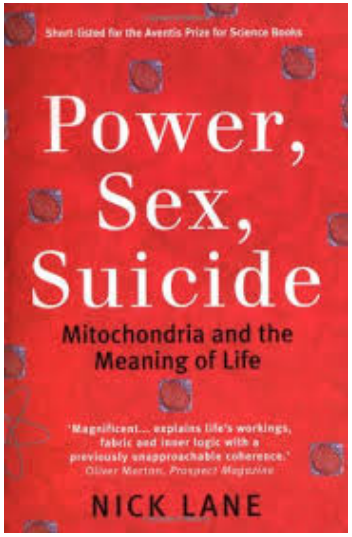






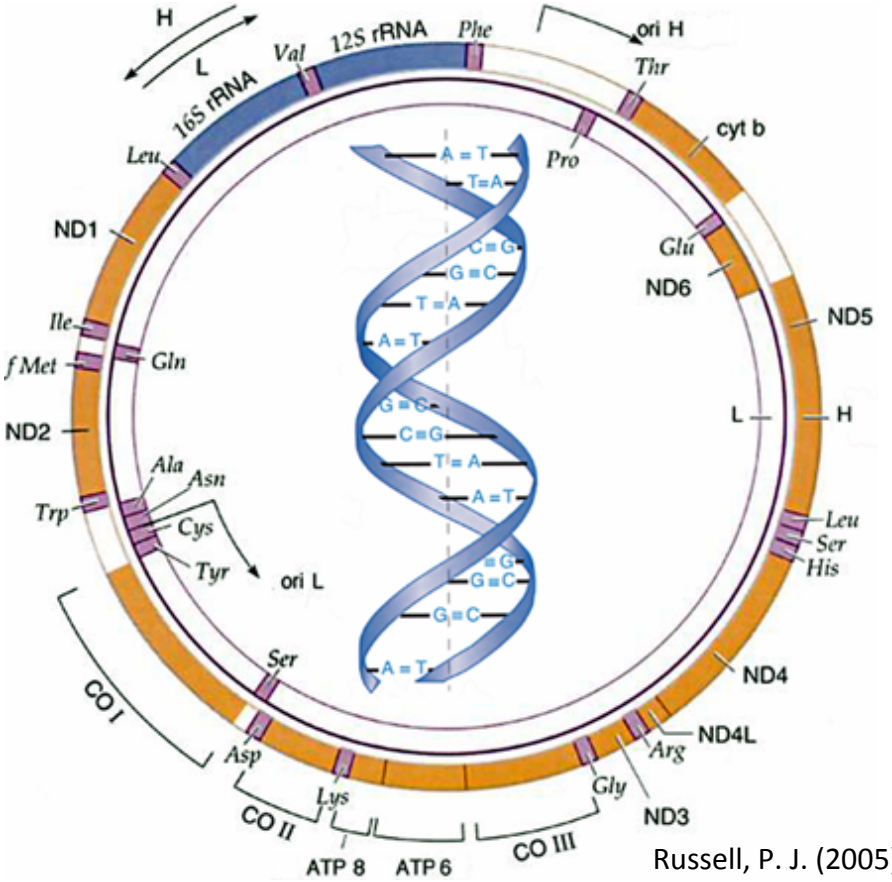






Human mitochondrial DNA

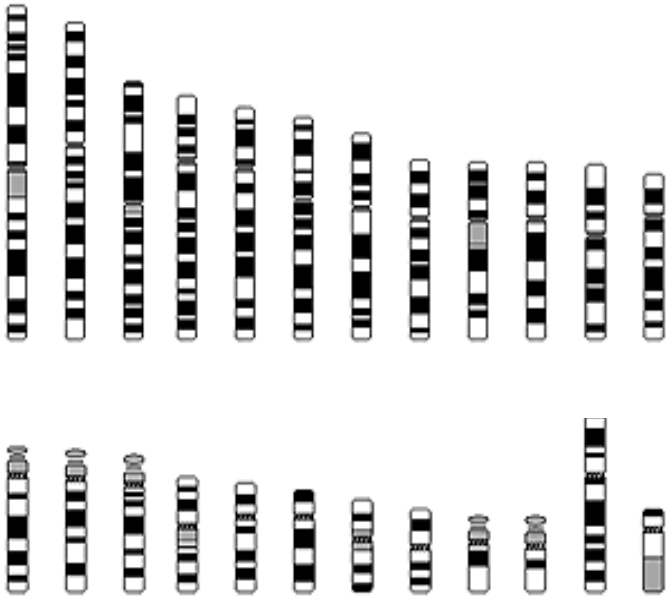
circular double-stranded molecule
16.5 kbp in size
compact and reduced
coding region (15 kb) 37 genes
  13 OSPHOX proteins
  22 tRNAs
    2 rRNAs
control region (1.1 kb) d-loop
  non-coding, regulatory
evolutionary rate ~10x of nDNA



Russell, P. J. (2005)

Human mitochondrial DNA

Nuclear DNA (nDNA)



46 chromosomes, 3.2 x 10^9 bp
diploid

mitochondrial DNA (mtDNA)

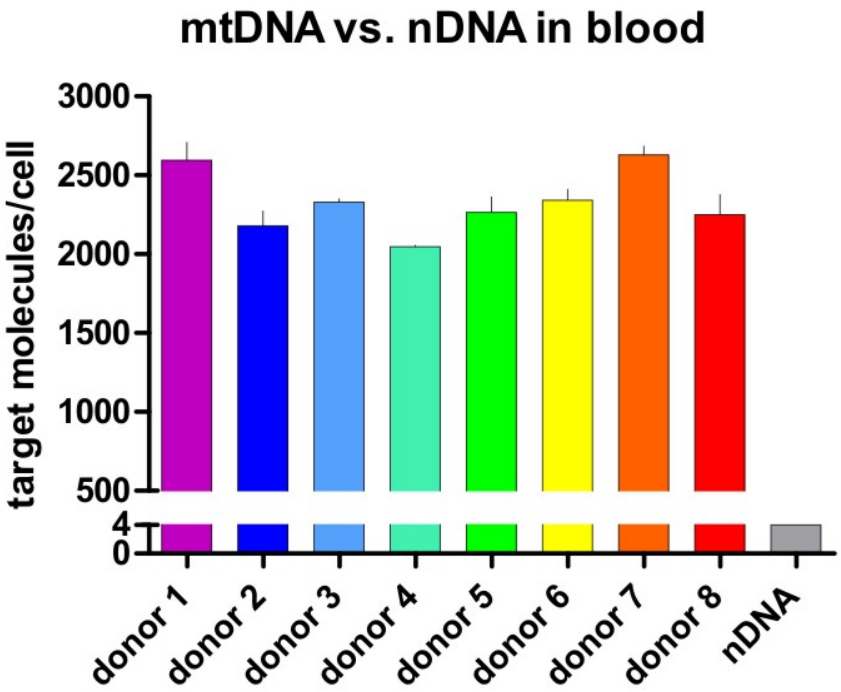
100(0)s per cell, 16.6 kbp
haploid

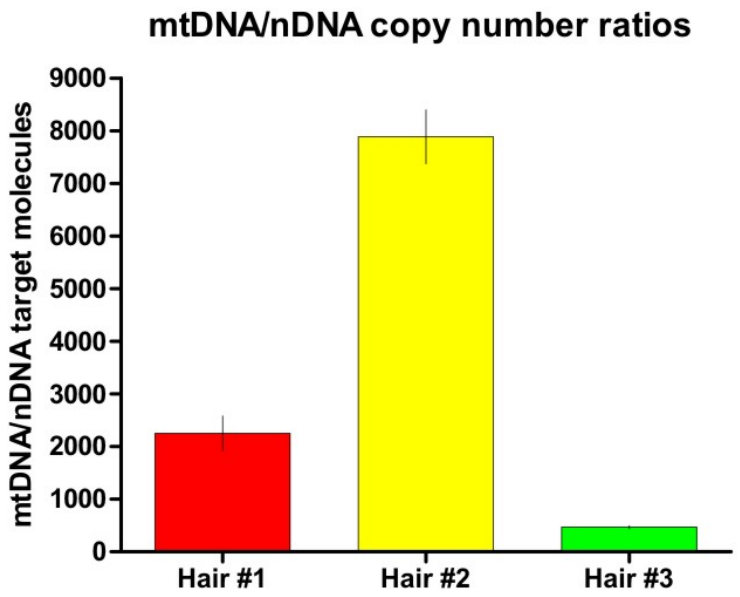
High mtDNA copy number

Higher copy number than nDNA

4-5 mtDNA (avg) molecules/mitochondrion (Satoh and Kuroiwa, 1991)

up to 1,000 mitochondria/cell (Robin and Wong, 1988)

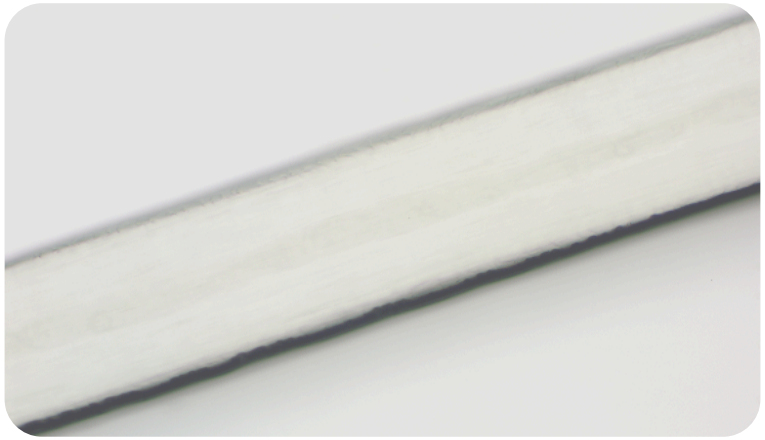


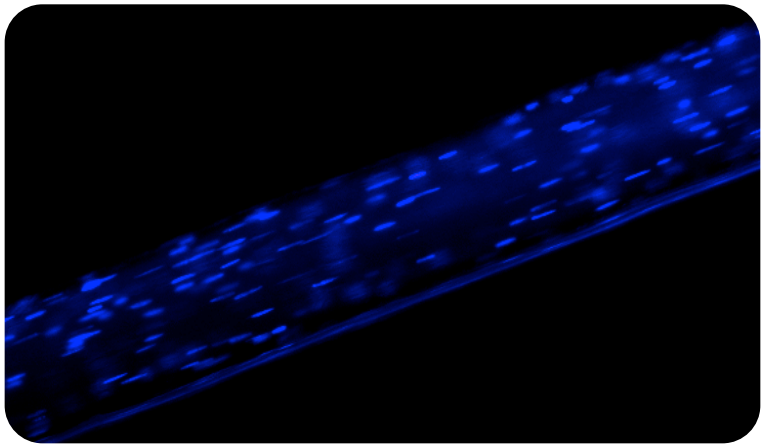


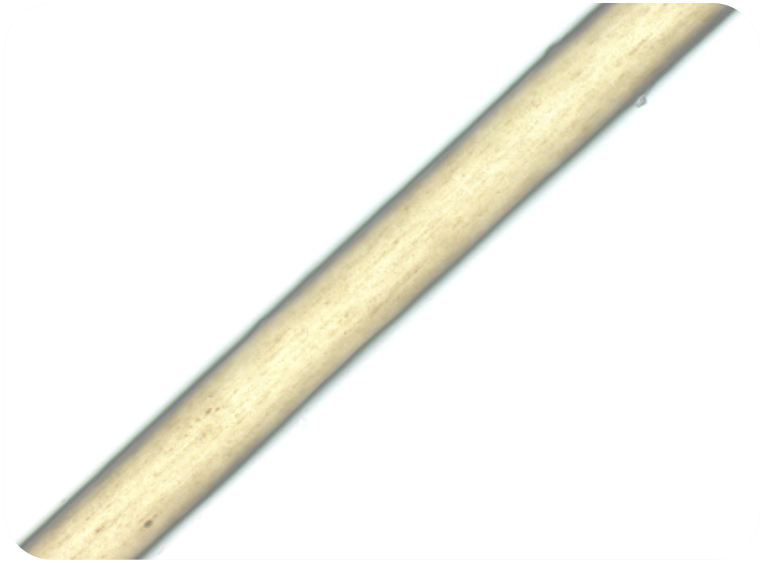
Blood: 500 – 650x

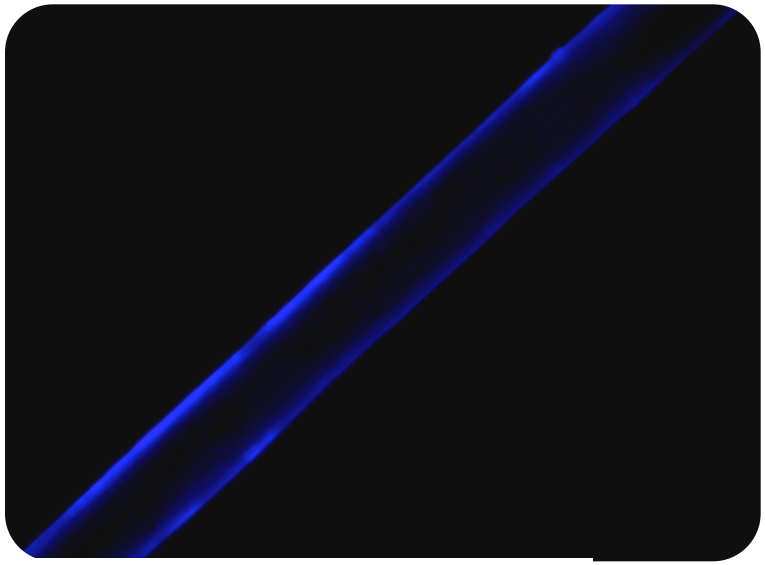
Hair: much higher

Application to challenging crime stains (e.g. hair)









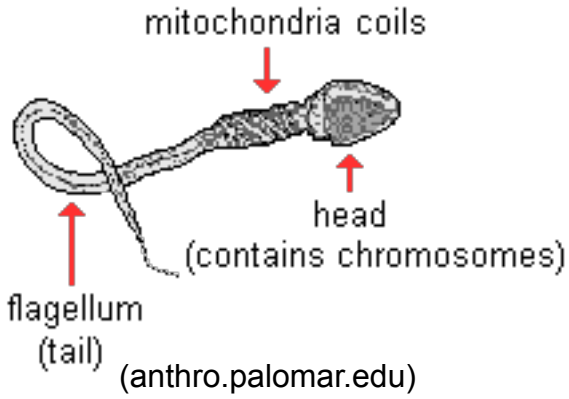
Szabo et al 2012

MtDNA is maternally inherited

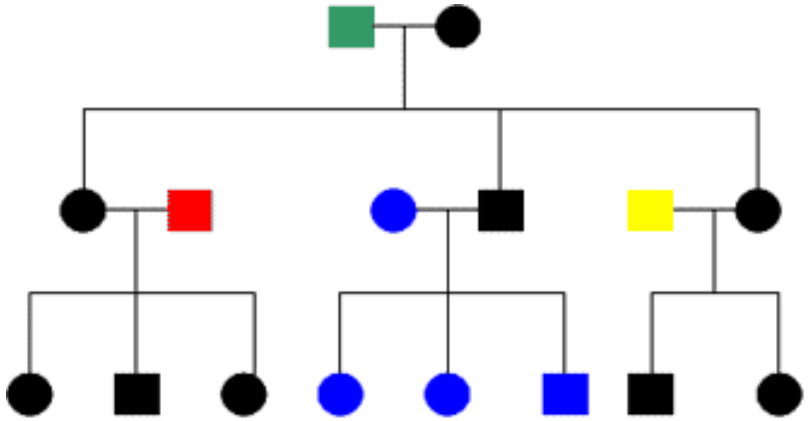
Mitochondria derive from the fertilized egg

(100.000s versus few in sperm neck)

Ubiquitin tagging of paternal mtDNA



(anthro.palomar.edu)



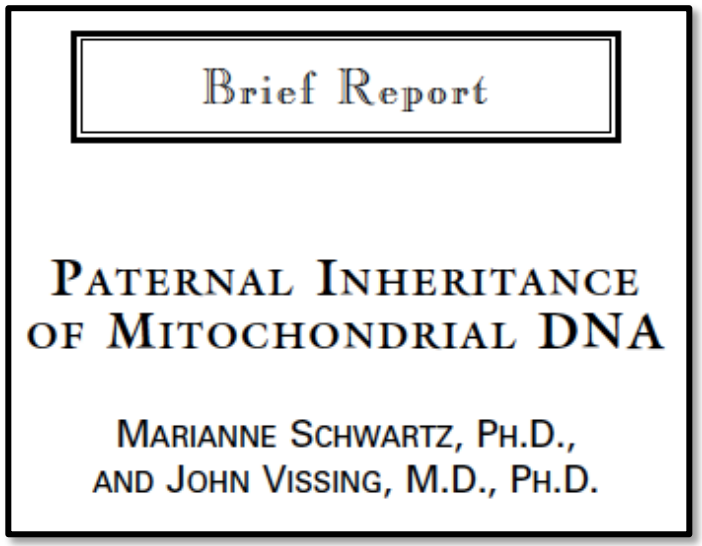
*Colors reflect inheritance of the same mitochondrial genome

Identification of maternal lineages (not individuals)

“Paternal inheritance and recombination”

Schwarz and Vissing (2002) found paternal mtDNA (90%) in skeletal muscle of 28 years old mitochondrial myopathy patient including a 2bp deletion in ND2 gene.

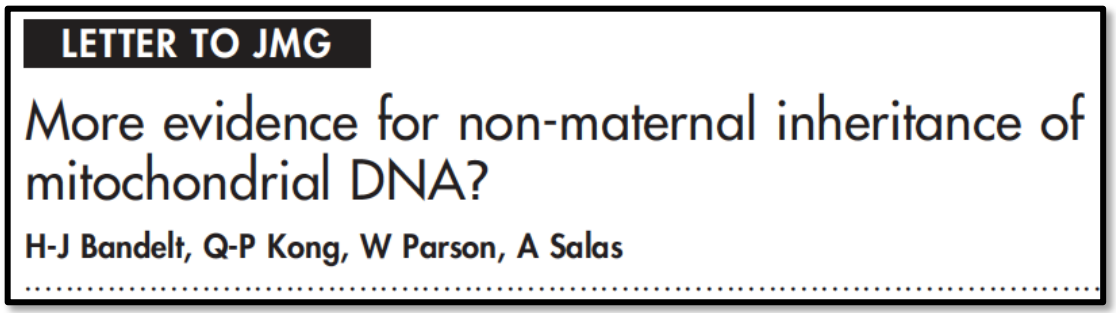
No further proofs in similar cases.



Brief Report

PATERNAL INHERITANCE OF MITOCHONDRIAL DNA

MARIANNE SCHWARTZ, PH.D., AND JOHN VISSING, M.D., PH.D.



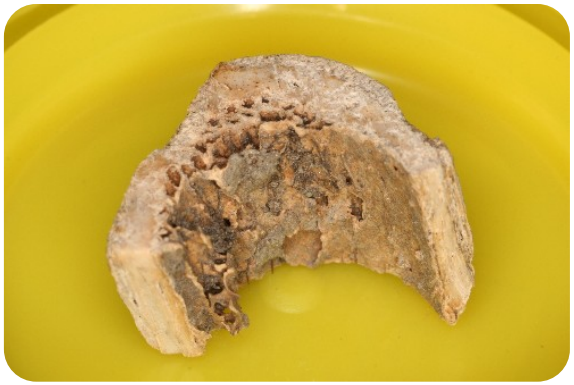
LETTER TO JMG

More evidence for non-maternal inheritance of mitochondrial DNA?

H-J Bandelt, Q-P Kong, W Parson, A Salas

>20 studies with mixed or mosaic mtDNA in tissues of the inestigated individuals phenomena can be ascribed to contamination and sample mix-up

Application to historical cases



Wolfgang A. Mozart (+1791)
Parson 2006

Romanov family (+1918)
Coble et al 2009

King Richard III (+1485)
King et al 2014

Statistics

EMPOP mtDNA database

Lab protocols

mtDNA extraction

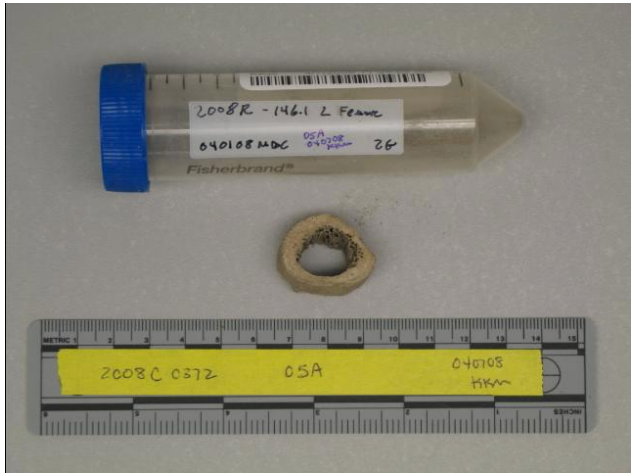
Identification of the two missing Romanov children (Coble et al 2009)

100-400 mg bone powder in 3 ml extraction buffer

EDTA 0.5 M, 0.5% lauryl-sarcosinate, 100 ml pro K

Purification with Centricon 30 (Millipore)

QIAquick PCR purification (Qiagen)

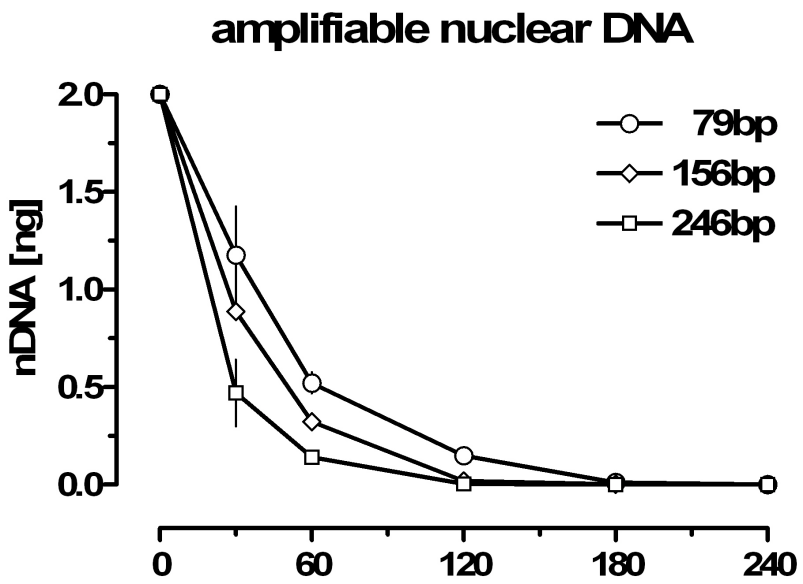


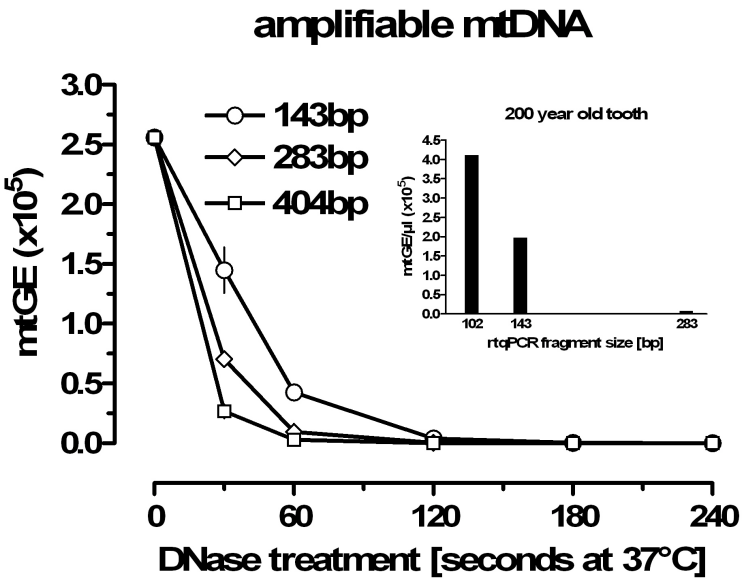
Update in Bauer et al 2013

mtDNA quantitation

Modular system (Niederstätter et al 2007)

- parallel determination of nDNA (Alu Yb8) and mtDNA
- or nDNA/mtDNA quantitation with internal PCR control (IPC)
- inhibition and degradation sensitive

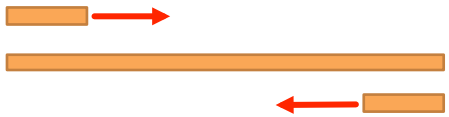




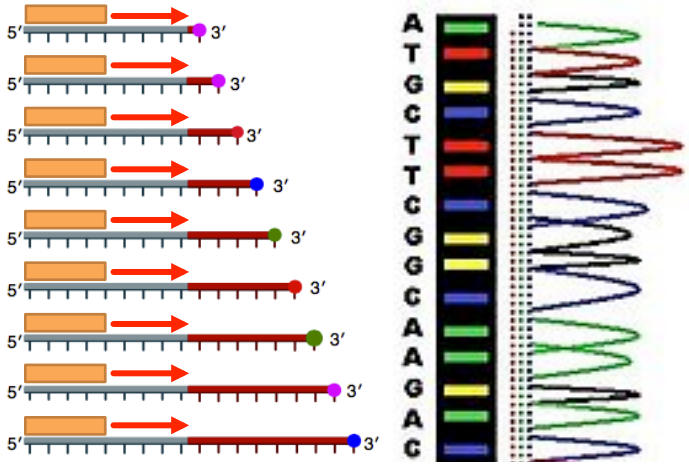
Update in Bauer et al 2013

mtDNA sequencing - Sanger

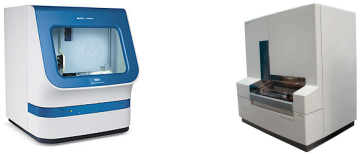
Amplification



Sanger-type
Dye Terminator
Sequencing



CE



Forensic mtDNA Sanger sequencing protocols

Brandstätter et al (2007) Generating population data for the **EMPOP database** - an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. Forensic Sci Int 166: 164-175
in response to high error rate due to phantom mutations

Parson W, Bandelt HJ (2007) **Extended** guidelines for mtDNA typing of population data in forensic science. Forensic Sci Int Genet 1: 13-19

Eichmann C, Parson W (2008) 'Mitominis': multiplex PCR analysis of **reduced size** amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. Int J Legal Med 122: 385-388
for severely degraded DNA

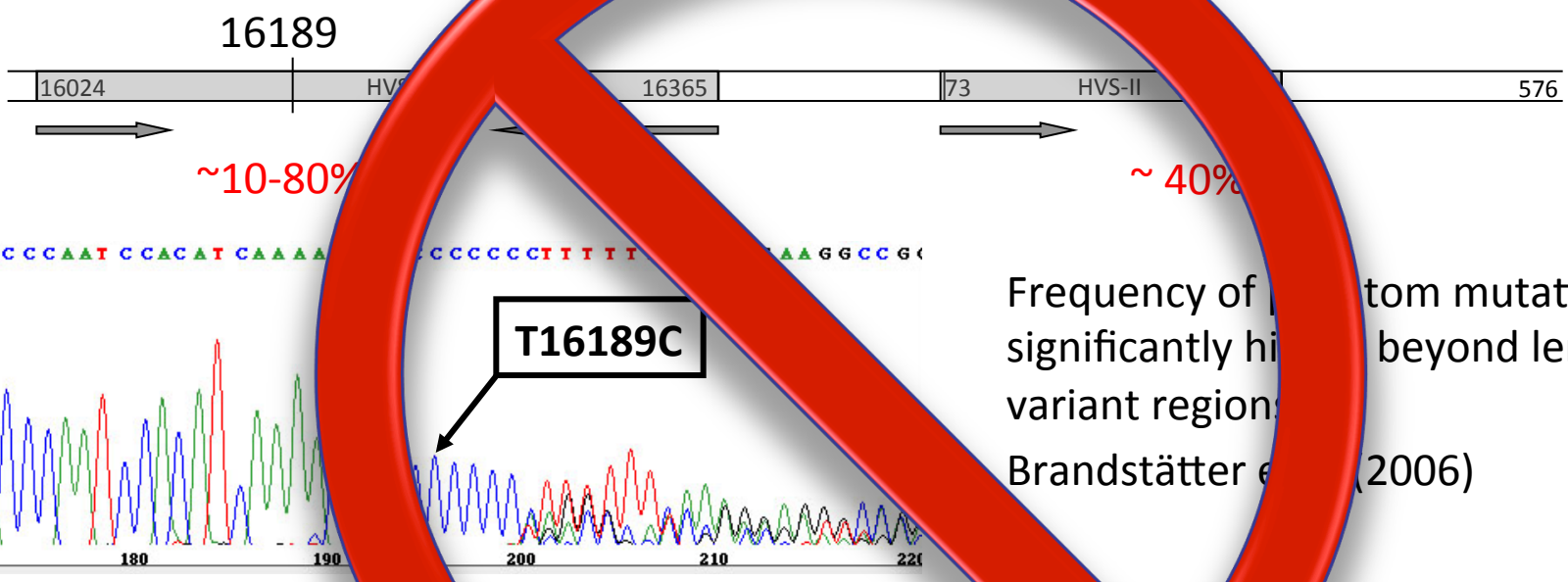
Berger C, Parson W (2009) **Mini-midi-mito**: Adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. Forensic Sci Int Genet 3: 149-153
for typical casework samples

Fendt L, Zimmermann B, Daniaux M, Parson W (2009) Sequencing strategy for the **whole mitochondrial genome** resulting in high quality sequences. BMC Genomics 10: 139
for high quality DNA

Outdated PCR and sequencing strategies

Separate HVS-I/HVS-II analysis

"double-stranded" consensus by [illegible] uencing



Frequency of [illegible] tom mutations significantly hi[illegible] beyond length variant region[illegible]

Brandstätter e[illegible] (2006)

> [illegible] length heteroplasmy, blurre[illegible] reads

Outdated PCR and sequencing strategies

Separate HVS-I/HVS-II analysis

Risk of artificial recombination (r gions between different samples)

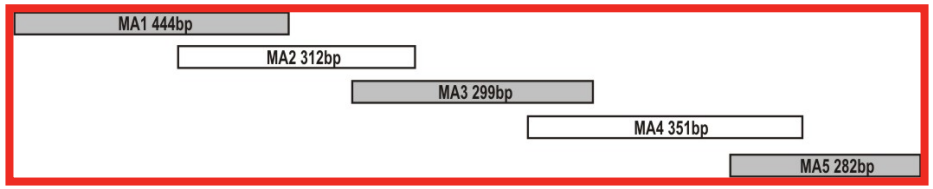
Bandelt et al (2004), Sal (2005)

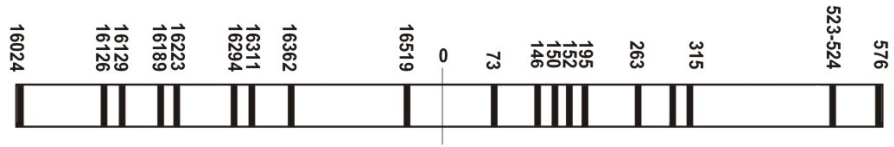
Example South Asi p. 09) – two consecutive sa

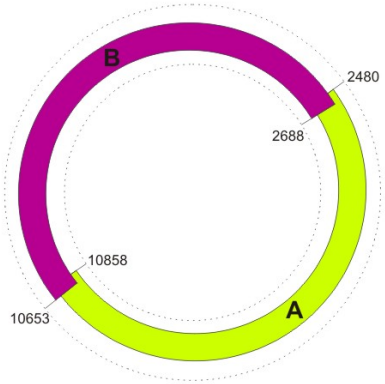
H 263G 309. 15.1C 489C

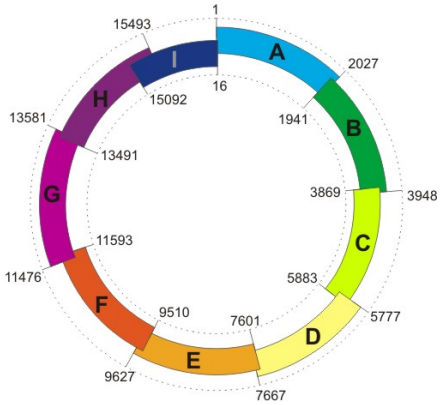
M 16223T 162 311C 16362C 73G 200G 263 489C is missing

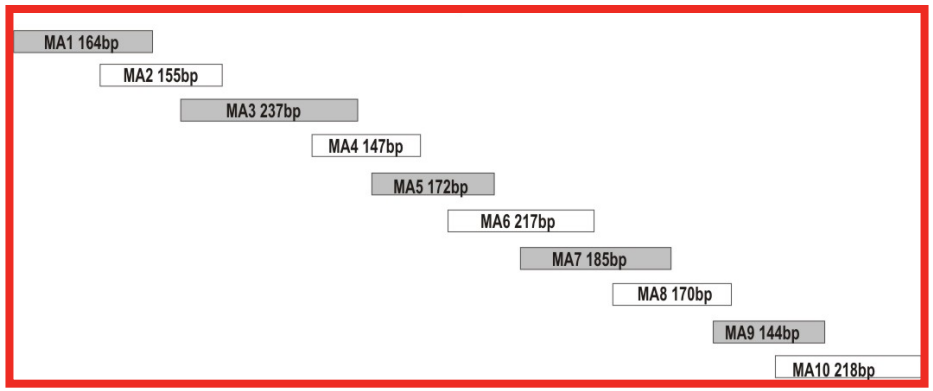
Forensic mtDNA Sanger sequencing protocols

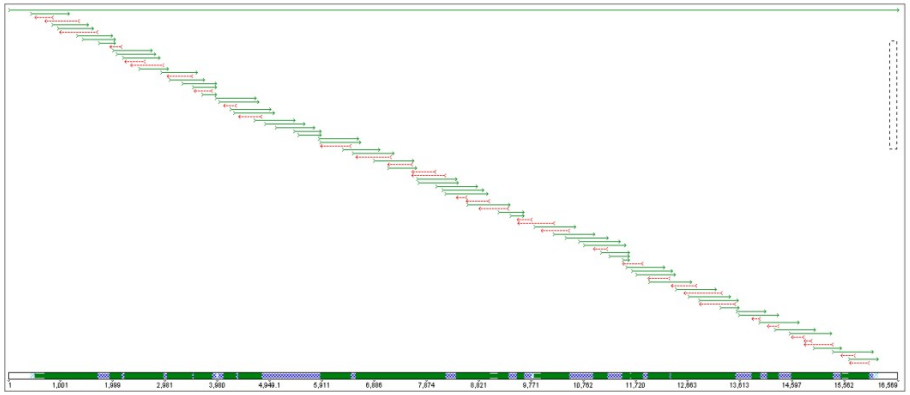










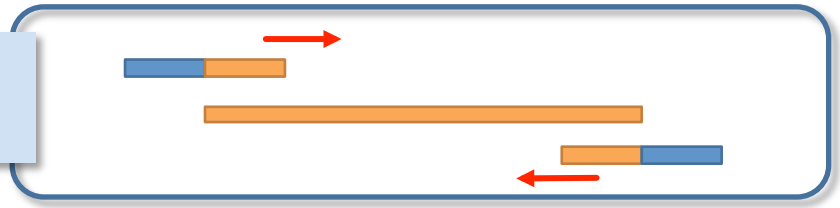


Berger and Parson 2009

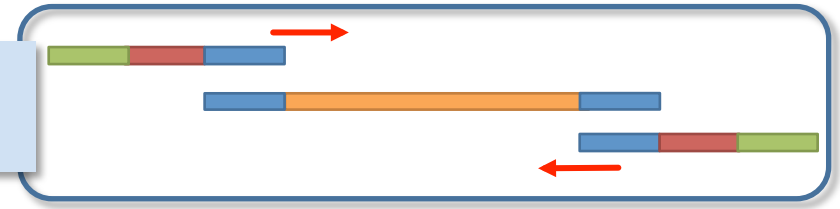
Fendt et al 2009

PCR-based mtDNA MPS

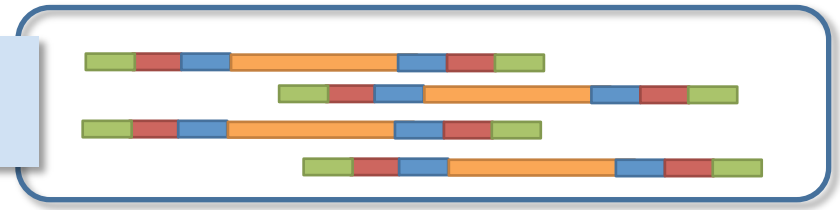
Amplification



Enrichment

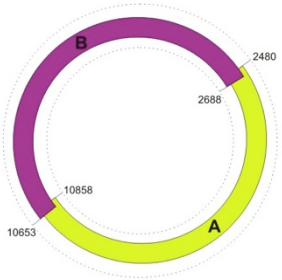


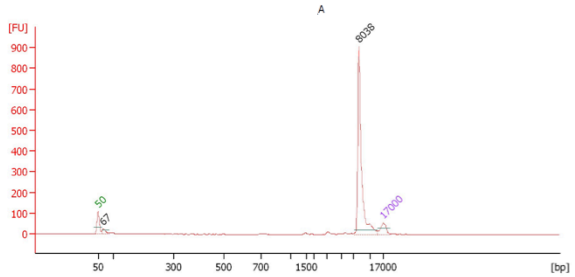
Final Library



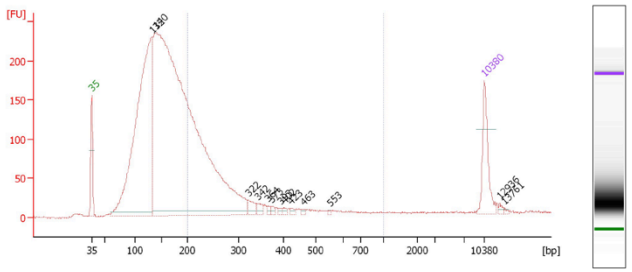
MPS

mtGenomes MPS of high quality DNA

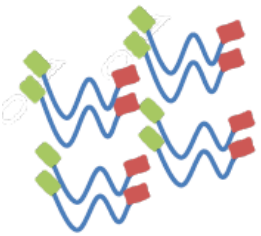




PCR

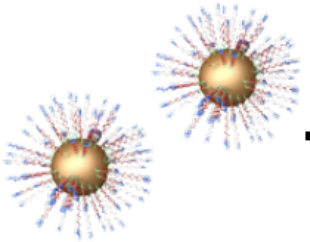


e-shearing

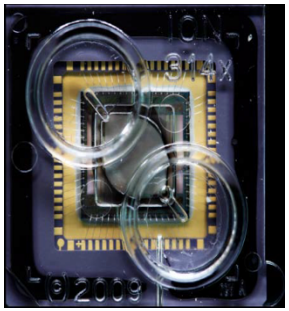


Prepare Library

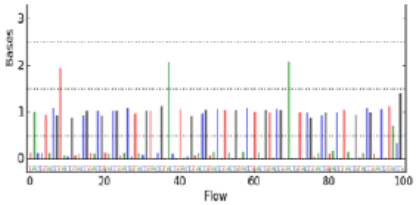
Clonal Amplification



Isolate Positive Ion Sphere™ Particles



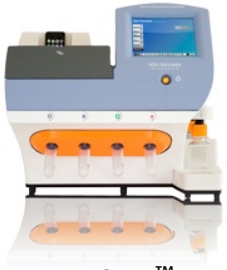
200 bp chemistry
316 chips



Data Analysis

OneTouch 2™

OneTouch™ES



PGM™

Torrent Server &
Torrent Browser

Parson et al 2012

courtesy Applied Biosystems by Life Technologies

Resolving most common West Eurasian CR type

Forensic Science International: Genetics 15 (2015) 21–26

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Helena, the hidden beauty: Resolving the most common West Eurasian mtDNA control region haplotype by massively parallel sequencing an Italian population sample

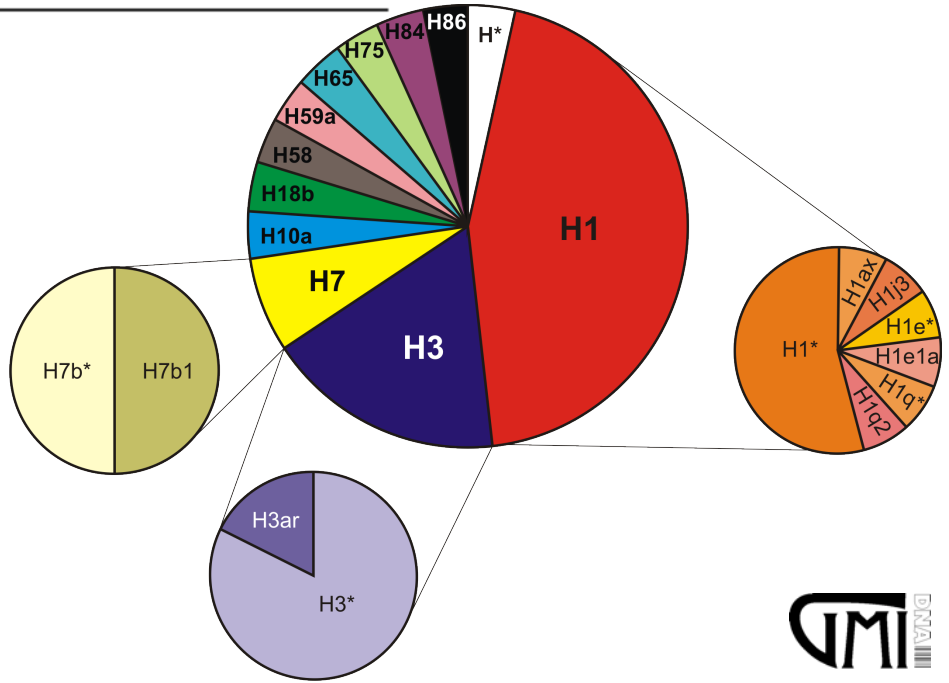
Martin Bodner [a], Alessandra Iuvaro [a,b], Christina Strobl [a], Simone Nagl [a], Gabriela Huber [a], Susi Pelotti [b], Davide Pettener [c], Donata Luiselli [c,*], Walther Parson [a,d,**]

[a] Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[b] Department of Medical and Surgical Sciences, Institute of Legal Medicine, University of Bologna, Bologna, Italy
[c] Department of Biological, Geological and Environmental Science, Laboratory of Molecular Anthropology, University of Bologna, Bologna, Italy
[d] Penn State Eberly College of Science, University Park, PA, USA

Resolving most common West Eurasian CR type

Table 1
Comparison of the diversity parameters in the 29 Italian samples using different sequence ranges.

	mtDNA range		
	CR	CR+39 codR SNPs[a]	Complete mtGenome
Haplotypes	1	6	28
Unique haplotypes	0	2	27
Haplogroups[b]	1	6	20
Unique haplogroups[b]	0	2	18
RMP[c]	1.000	0.296	0.037
Haplotype diversity	0.0%	72.9%	99.8%

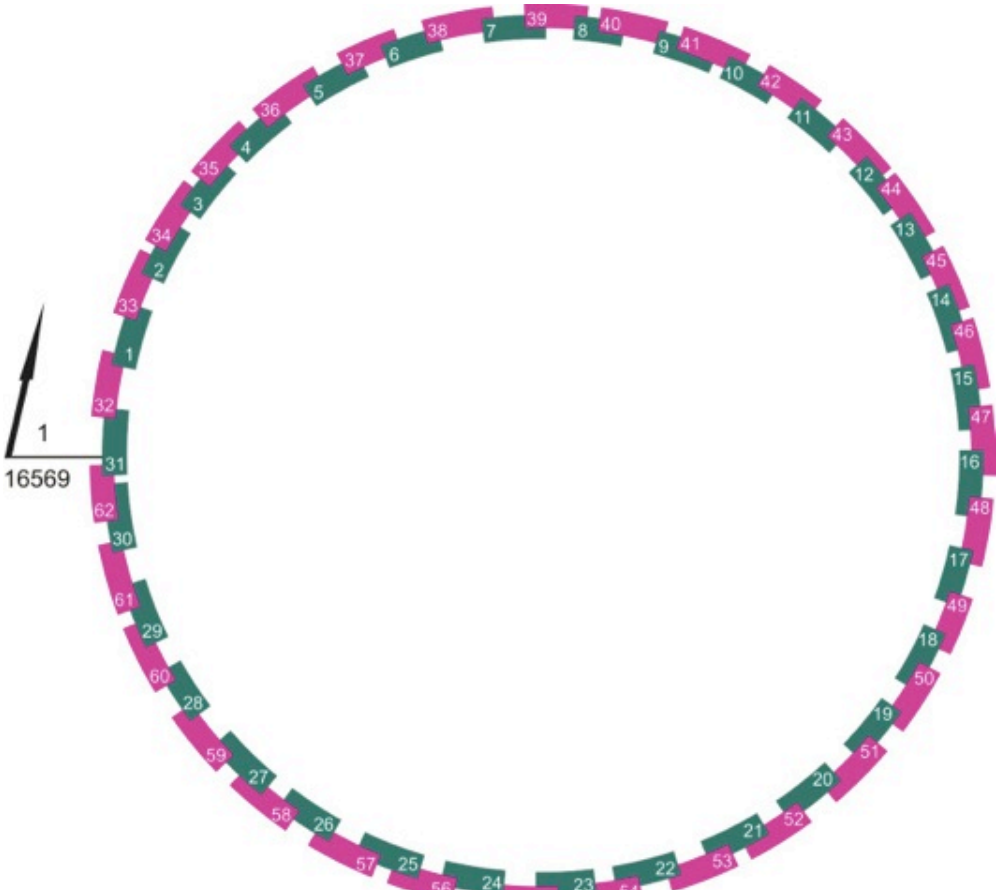


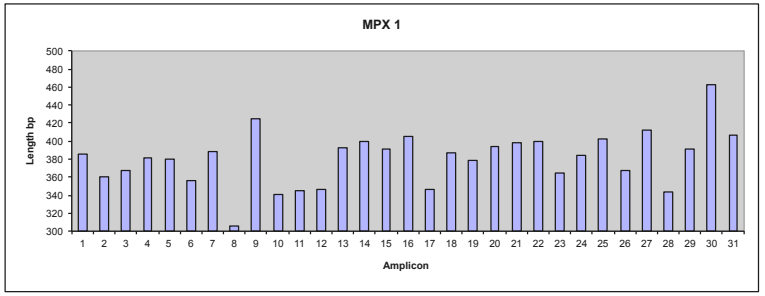
mtGenome MPS of degraded DNA

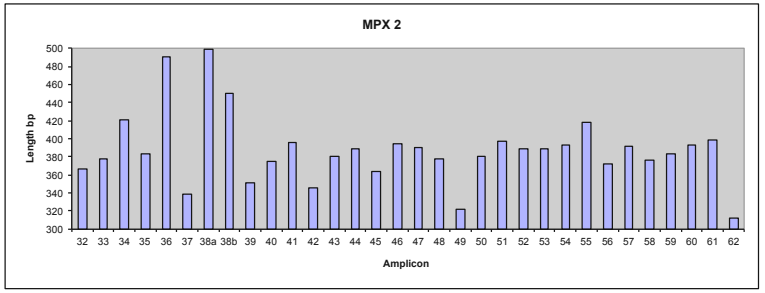
Original Research Paper

Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples

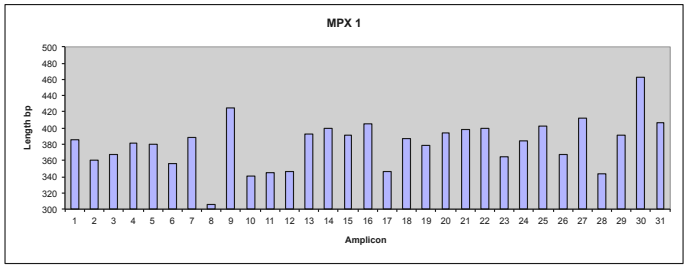
Walther Parson [a,b,*], Gabriela Huber [a], Lilliana Moreno [c], Maria-Bernadette Madel [a], Michael D. Brandhagen [c], Simone Nagl [a], Catarina Xavier [a], Mayra Eduardoff [a], Thomas C. Callaghan [c], Jodi A. Irwin [c,*]



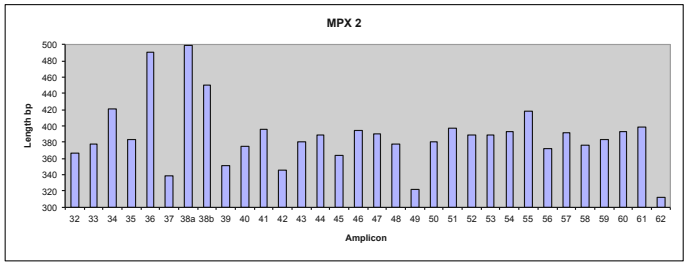




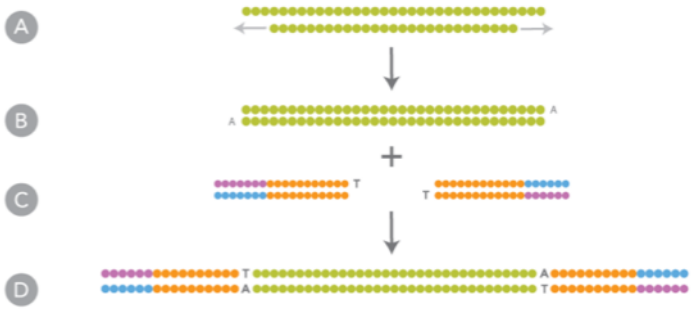
mtGenome MPS of degraded DNA



mtG-Midi-plex 1

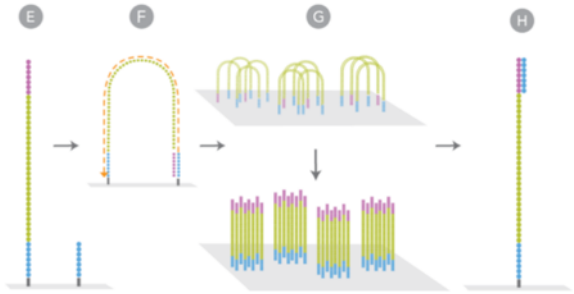


mtG-Midi-plex 2

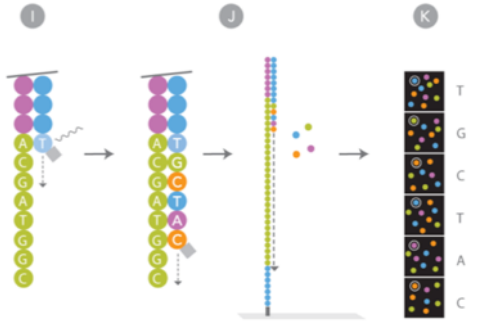


Nextera XT

Flowcell



Cluster Generation



& Sequencing by Synthesis

Illumina MiSeq

Parson et al 2015

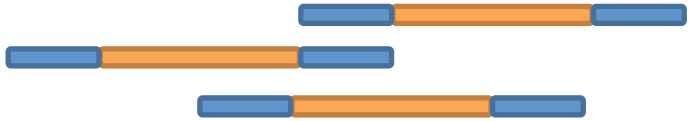
modified by courtesy of Illumina

mtGenome MPS using mini-amplicons

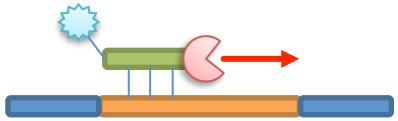
Strobl et al in prep

PEC mtDNA MPS

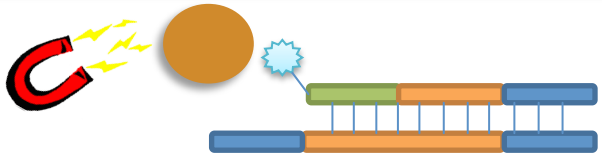
Ligation



Primer **E**xtension



Capture



MPS

PEC mtDNA MPS

Eduardoff et al in prep

The New York Times

AMERICAS

Remains of Student in Mexico Identified

By RANDAL C. ARCHIBOLD and PAULINA VILLEGAS DEC. 6, 2014

Ezequiel Mora, father of Alexander Mora, whose remains were said to have been identified, mourns at his home in El Pericon, Mexico, on Sunday. Jorge Dan Lopez/Reuters

Nuclear elements of mtDNA - numts

Segregation of mtDNA segments into nDNA

Databases for numts (e.g. mitomap)

numts can produce a **mixed sequence** with autochthonous mtDNA

numts may even mask autochthonous mtDNA, e.g. when

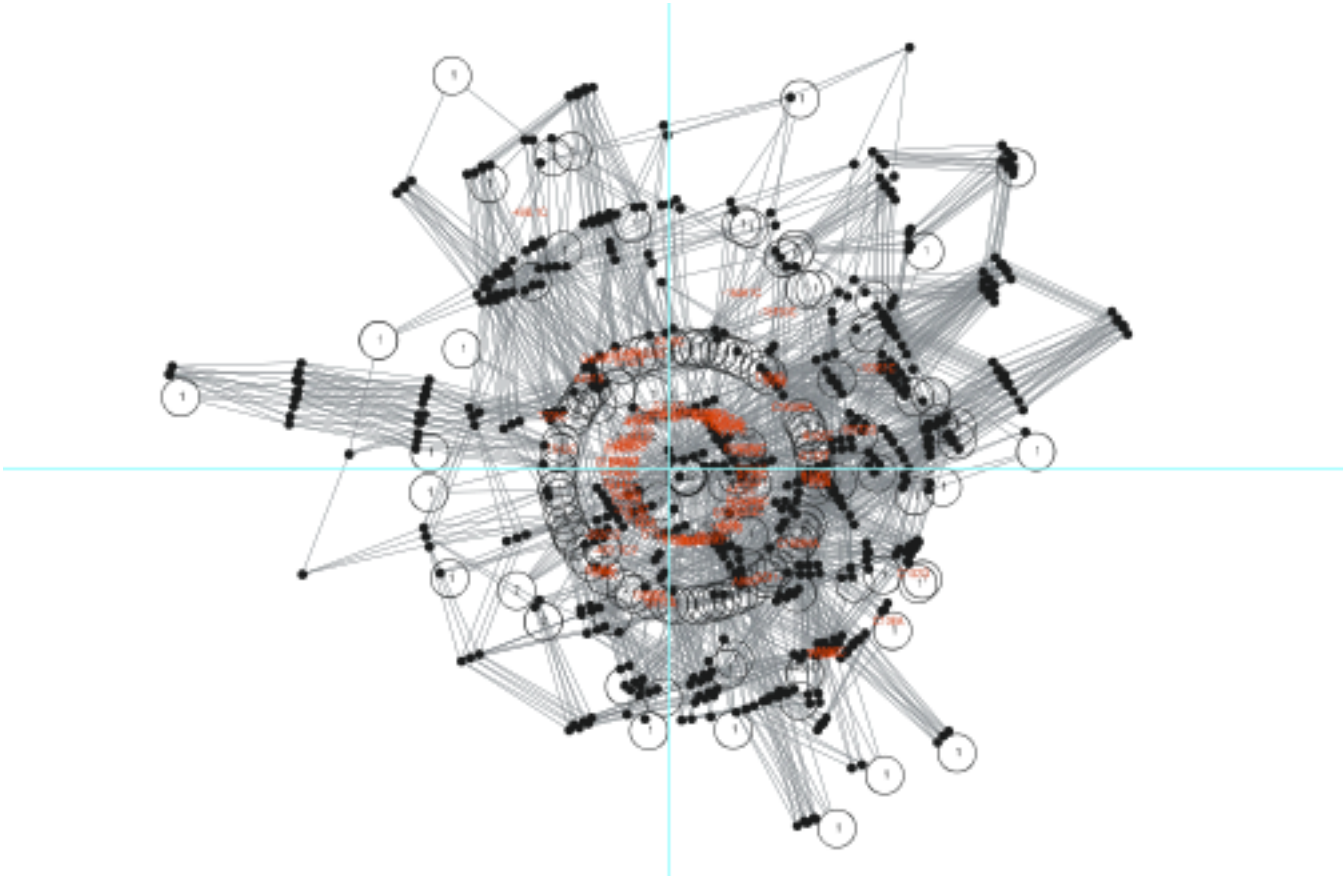
primer binding sites favor the amplification of the numt

Forensic relevance:

numts may produce mixtures or even mask autochthonous mtDNA

careful primer design and use of appropriate amounts of (mt)DNA (quantitation)

mtDNA quality control



Error reports

Annals of human genetics

Commentary

To Err is Human

P. Forster

Survey on mtDNA population studies in 37 journals (1981-2002)

137 publications, 80 were erroneous – error rate 58%
Forensic journals performed similar to others

"Phylogenetic analyses make it clear that error free publications on mtDNA sequences are extremely rare (Bandelt *et al.* 2001; 2002)"

Collaborative EDNAP study

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Forensic Science International 139 (2004) 215–226

Forensic Science International

www.elsevier.com/locate/forsciint

The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives

Walther Parson[a,*], Anita Brandstätter[a], Antonio Alonso[b], Nathalie Brandt[c], Bernd Brinkmann[d], Angel Carracedo[e], Daniel Corach[f], Olivier Froment[g], Ivana Furac[h], Tomasz Grzybowski[i], Karin Hedberg[j], Christine Keyser-Tracqui[k], Tomasz Kupiec[l], Sabine Lutz-Bonengel[m], Bente Mevag[n], Rafal Ploski[o], Hermann Schmitter[p], Peter Schneider[q], Denise Syndercombe-Court[r], Eric Sørensen[s], Heather Thew[t], Gillian Tully[u], Richard Scheithauer[a]

ıses of sample mix-up and 1 case of point heteroplasmic mixture, where the 2 sequencing reactions b
ılls. This corresponds to an error rate of 10.7% in a virtual mtDNA database consisting of the collat
owever, this estimate is still conservative compared to conclusions drawn by authors of meanwhile

71% clerical error

Lab-specific mutation processes

Bandelt H-J, Kivisild T, Parik J, Villems R, Bravi C, Yao Y-G, Brandstätter A, Parson W

In: ***Human mitochondrial DNA and the evolution of Homo sapiens***

Springer-Verlag eds. Hans-Jürgen Bandelt, Vincent Macaulay, Martin Richards (2006)

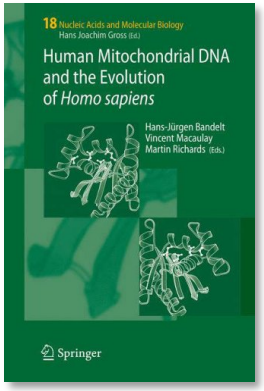
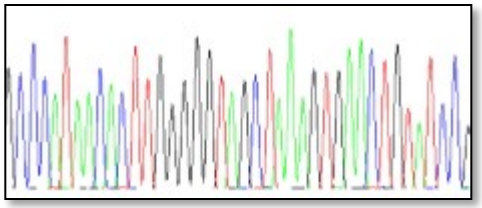
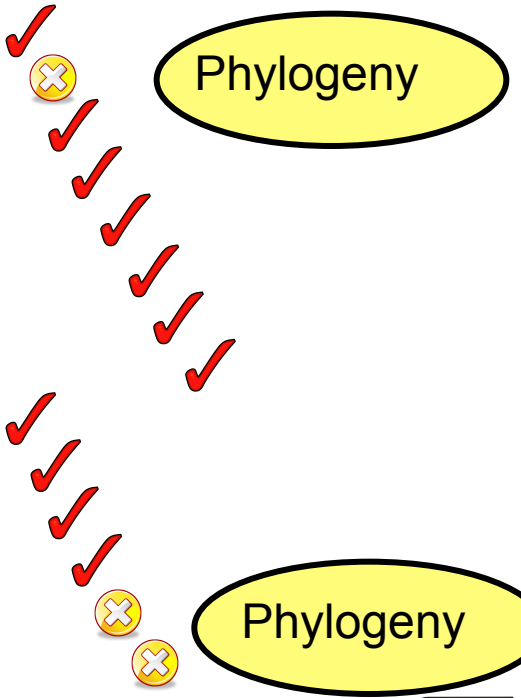


Table 1 Error classification "Stage– Cause– Phenotype"

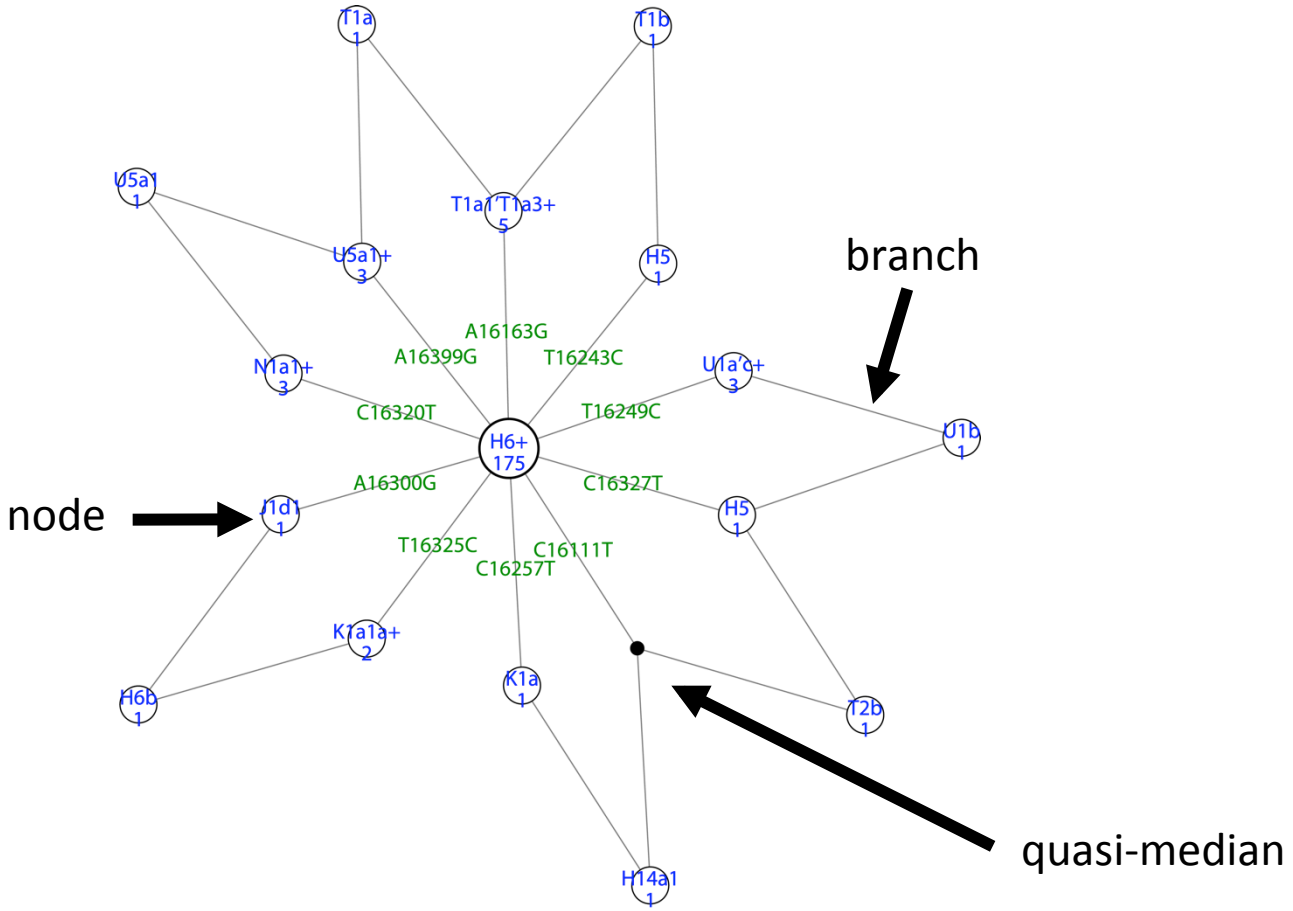
Stage	Cause	Phenotype	Process
1			DNA extraction
2			PCR
3			Sequencing
4			Electrophoresis
5			Interpretation and documentation
	C		Contamination
	M		Sample mix-up
	A		Sequencing artefact
	S		Sample manipulation and bias
	L		Misalignment or incorrect reference sequence
	B		Basecall misinterpretation
	N		Nomenclature violation
	T		Transcription error
		I	Base shift
		II	Reference bias
		III	Phantom mutation
		IV	Base misreporting
		V	Artificial recombination
		VI	Skewed variation





Quasi-Median Networks (QMN)

Network torso HVS-I (16024-16569)
EMPOPSpeedyWE
N=202

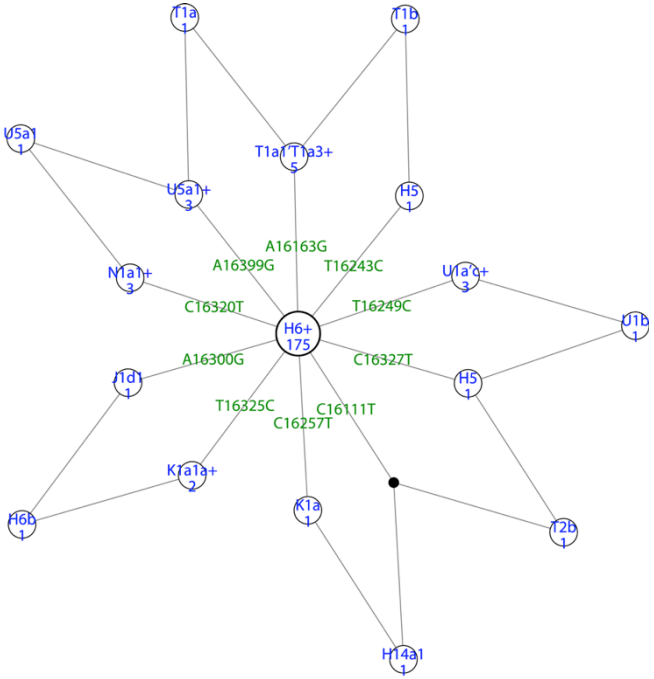


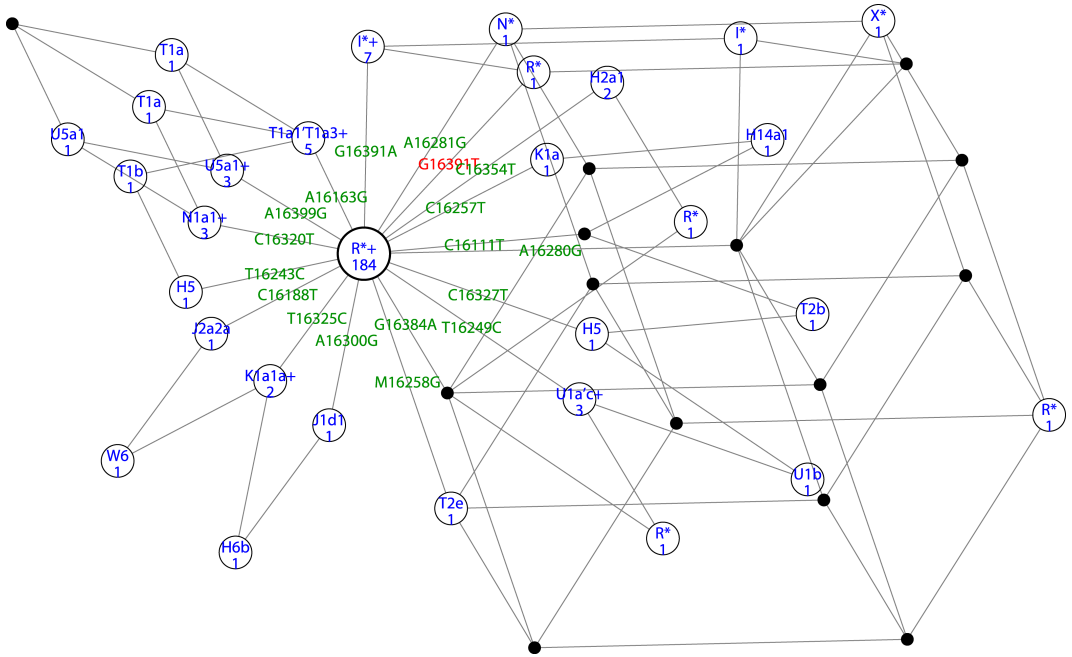
Zimmermann et al (2011)

Quasi-Median Networks (QMN)

Network torso HVS-I (16024-16569)
EMPOPSpeedyWE
N=WE-Etalon (202)

+29 (Caucasus, 2001)





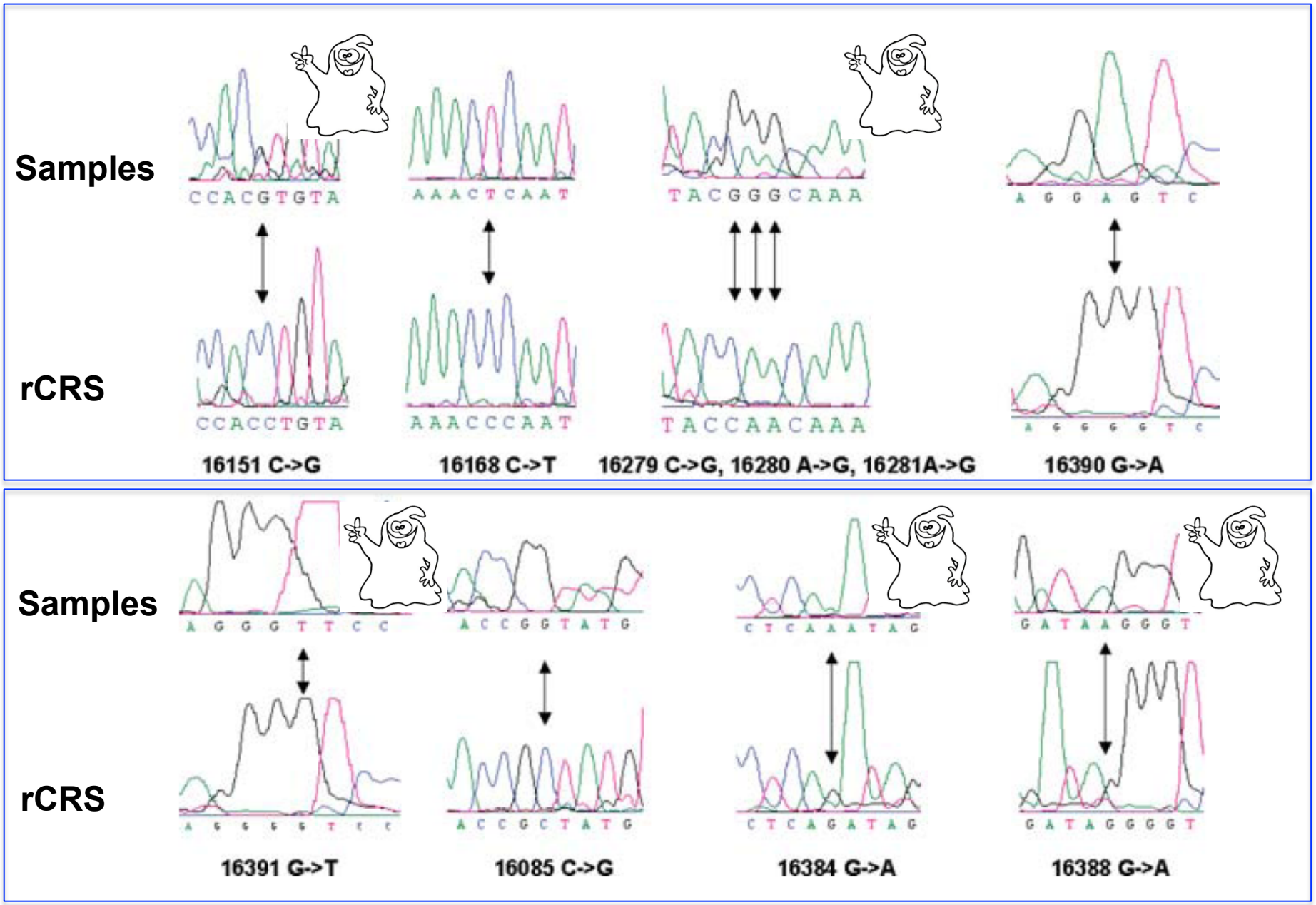
Zimmermann et al (2011)

13 new quasi-medians

29 haplotypes (published data)

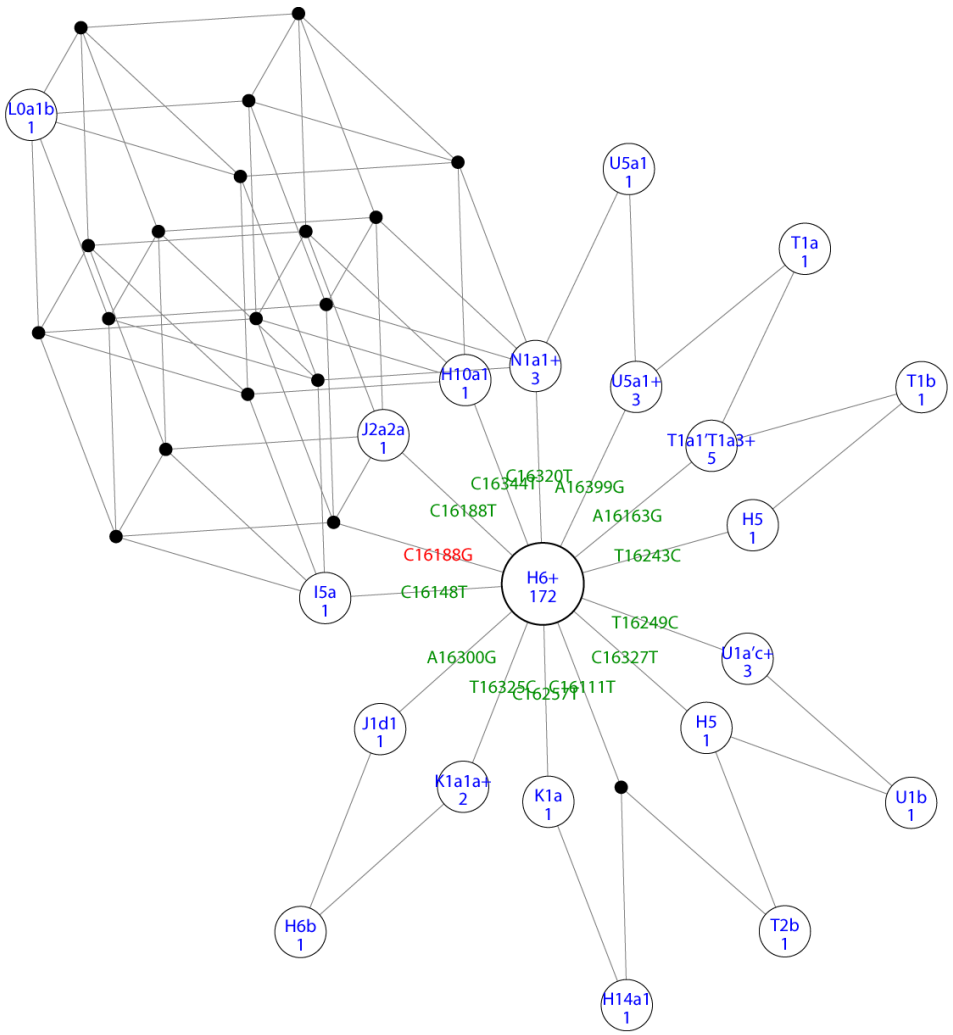
3	# 29 samples											
4	#! 16024-16400											
5	DAR04_Darginian	R0	1	16093C								
6	DAR012_Darginian	R0	1	16189C								
7	DAR16_Darginian	R0	1	16189C								
8	DAR49_Darginian	R0	1	16093C								
9	DAR014_Darginian	J*	1	16069T	16126C	16145A	16261T					
10	DAR42_Darginian	J*	1	16126C	16192T	16294T	16296T					
11	DAR06_Darginian	J*	1	16126C	16213A	16269G	16294T					
12	DAR31_Darginian	K*	1	16093C	16224C	16278T	16291T	16311C				
13	DAR03_Darginian	T1a	1	16126C	16163G	16186T	16189C	16294T	16320T			
14	DAR11_Darginian	U2	1	16051G	16148T	16184T	16234T	16294T	16342C	16357C		
15	DAR013_Darginian	U4	1	16261T	16356C							
16	DAR02_Darginian	U5	1	16189C	16192T	16256T	16270T					
17	DAR3_Darginian	U5	1	16256T	16270T							
18	DAR011_Darginian	U5	1	16256T	16270T							
19	DAR37_Darginian	U5	1	16256T	16270T							
20	DAR18_Darginian	U8	1	16051G	16093C	16148T	16184T	16210G	16233G	16234T	16266T	16294T
21	**DAR05_Darginian**	N*	1	16129A	16223T	**16281G**						
22	**DAR015_Darginian**	R*	1	16234T	**16388A**							
23	**DAR016_Darginian**	R*	1	16150T	16354T	**16384A**						
24	**DAR07_Darginian**	R*	1	16189C	16249C	**16373A**	**16381C**	**16384A**				
25	**DAR35_Darginian**	R*	1	16294T	**16391T**	**16392C**						
26	**DAR36_Darginian**	R*	1	16258G	**16280G**	**16281G**	**16384A**					
27	DAR01_Darginian	R*	1	16292T								
28	DAR08_Darginian	R*	1	16223T								
29	DAR010_Darginian	R*	1	16292T								
30	DAR51_Darginian	R*	1	16234T								
31	**DAR14_Darginian**	I*	1	16114T	16129A	16223T	**16280G**	16391A				
32	DAR09_Darginian	I*	1	16129A	16223T	16391A						
33	DAR018_Darginian	I*	1	16129A	16223T	16391A						
34	**DAR23_Darginian**	X*	1	16129A	16223T	16278T	**16280G**	**16281G**				
35	DAR46_Darginian	W6	1	16188T	16192T	16223T	16292T	16325C				

red mutations not previously observed



Data from Caucasus (2001) Proc R Soc Lond B Biol Sci 268, 1197–1206

QMN - distant phylogenies



EMPOP as mtDNA QC tool

Forensic Science International: Genetics 4 (2010) 145–147

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Editorial

Publication of population data for forensic purposes

Int J Legal Med (2010) 124:505–509
DOI 10.1007/s00414-010-0492-y

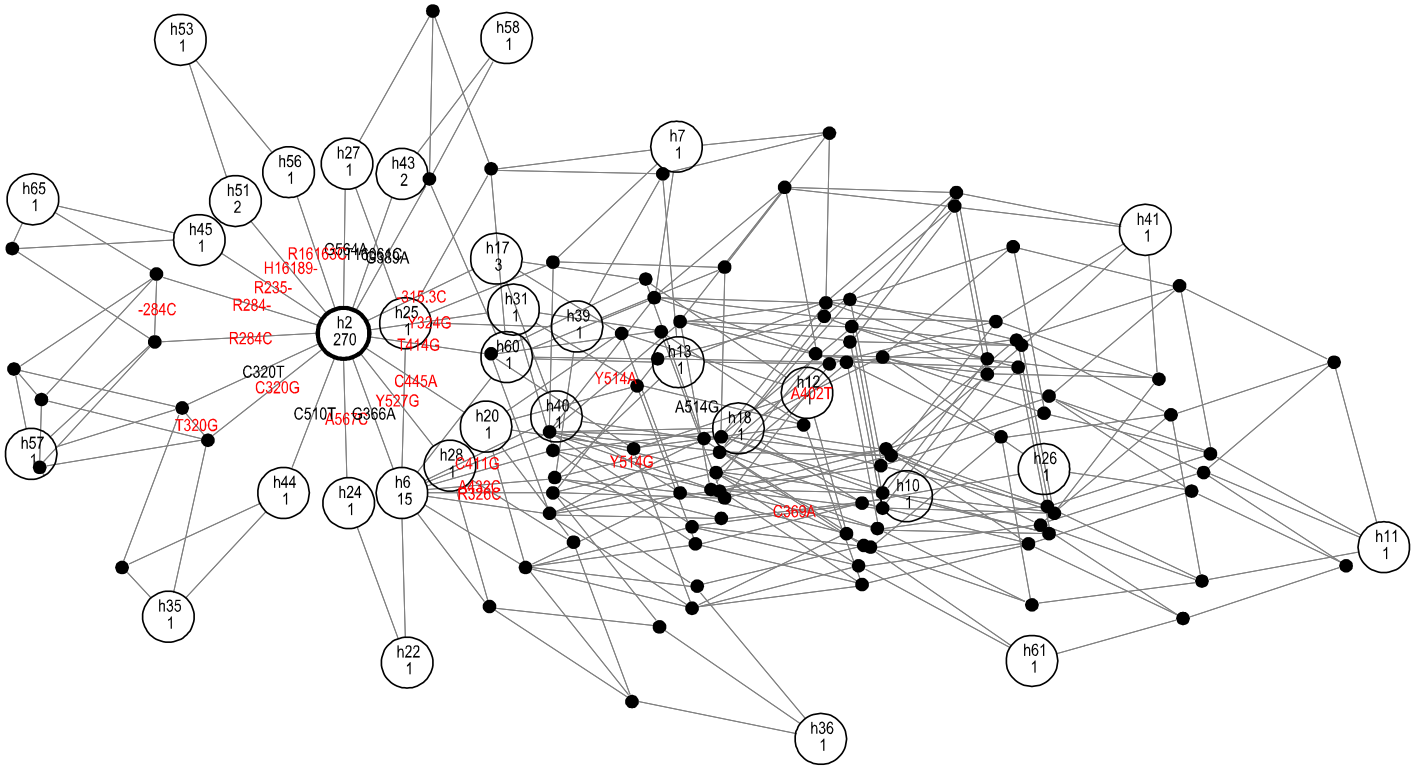
SHORT COMMUNICATION

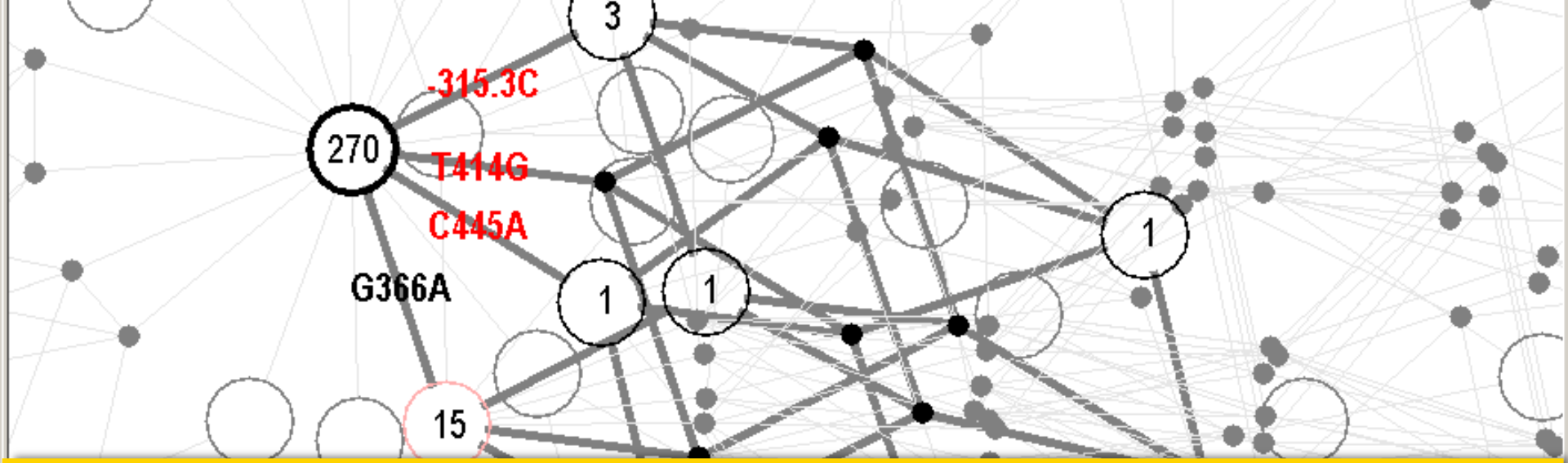
Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine

Walther Parson · Lutz Roewer

Dataset A: 320 West Eurasian haplotypes

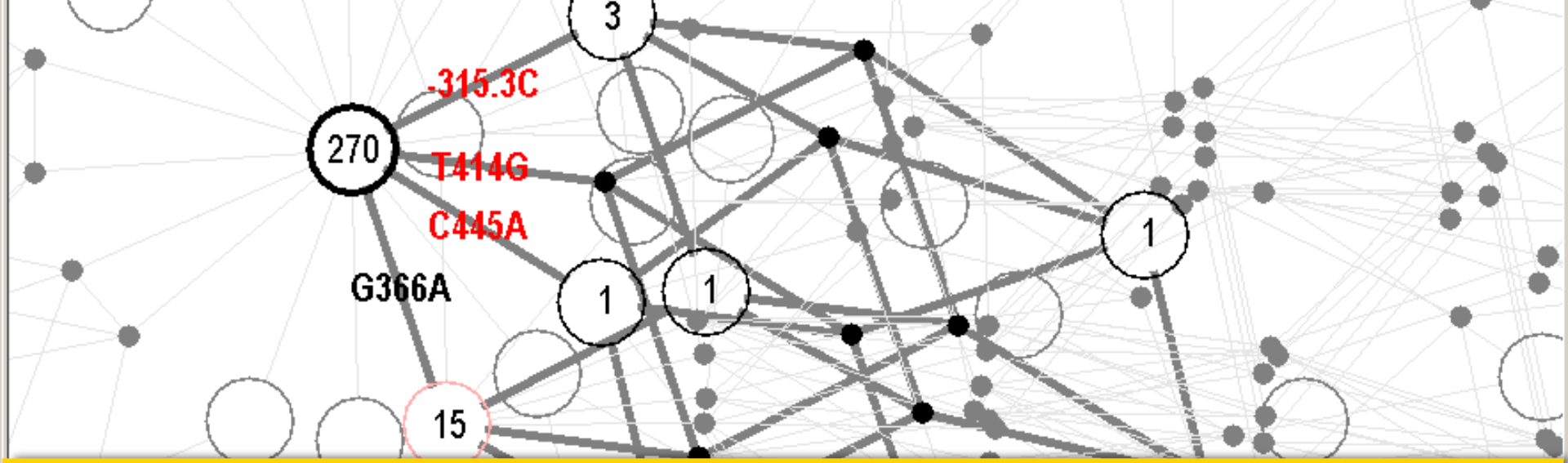
Type: Torso
Name: WE_320.dnw
Filter: Hotspots_EMPOPall_R7
Range: 16024-576






```
Detailed information for h6:
WE230 RO 1 16129A 16519C 263G 309.1C 366A
WE499 RO 1 195C 263G 309.1C 366A 523DEL 524DEL
WE256 H15 1 16184T 44.1C 55C 57C 146C 263G 309.1C 366A
WE443 RO 1 16259T 16519C 263G 309.1C 366A
WE263 RO 1 16519C 263G 309.1C 366A
WE275 RO 1 16519C 263G 309.1C 366A
WE244 RO 1 16140C 16519C 71.1G 249DEL 263G 309.1C 366A
WE336 RO 1 16519C 263G 309.1C 366A
WE278 H2a2b 1 16235G 16291T 93G 263G 309.1C 366A
WE321 H6 1 16519C 239C 263G 309.1C 366A
WE226 H6 1 16519C 152C 239C 263G 309.1C 366A
WE258 H6a1a1a 1 16311C 16362C 16482G 71.1G 73G 239C 263G 309.1C 366A
WE567 HV0 1 16075C 16240G 16298C 72C 263G 309.1C 366A
WE327 J1c2 1 16069T 16126C 16278T 16519C 73G 185A 188G 228A 263G 295T 309.1C 366A 462T 489C 523DEL 524DEL
WE334 K1a 1 16224C 16311C 16519C 73G 152C 263G 309.1C 366A 497T
```

Connection: A366G



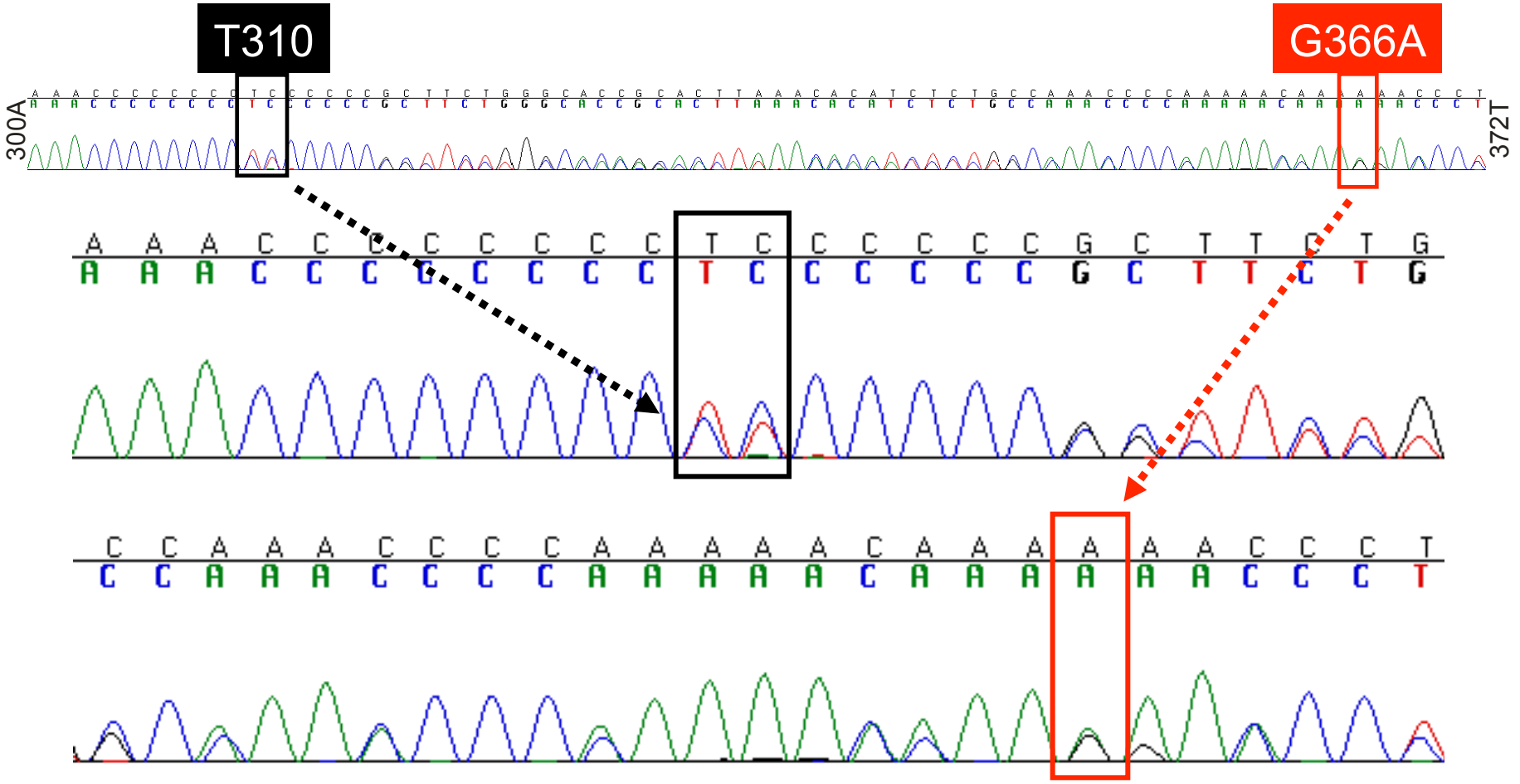

```
Detailed information for h6:
WE230 RO 1 16129A 16519C 263G 309.1C 366A
WE499 RO 1 195C 263G 309.1C 366A 523DEL 524DEL
WE256 H15 1 16184T 44.1C 55C 57C 146C 263G 3
WE443 RO 1 16259T 16519C 263G 309.1C 366A
WE263 RO 1 16519C 263G 309.1C 366A
WE275 RO 1 16519C 263G 309.1C 366A
WE244 RO 1 16140C 16519C 71.1G 249DEL 263G 3
WE336 RO 1 16519C 263G 309.1C 366A
WE278 H2a2b 1 16235G 16291T 93G 263G 309.1C
WE321 H6 1 16519C 239C 263G 309.1C 366A
WE226 H6 1 16519C 152C 239C 263G 309.1C 366A
WE258 H6a1a1a 1 16311C 16362C 16482G 71.1G 73G 239C 263G 309.1C 366A
WE567 HV0 1 16075C 16240G 16298C 72C 263G 309.1C 366A
WE327 J1c2 1 16069T 16126C 16278T 16519C 73G 185A 188G 228A 263G 295T 309.1C 366A 462T 489C 523DEL 524DEL
WE334 K1a 1 16224C 16311C 16519C 73G 152C 263G 309.1C 366A 497T
```

G366A found in hgs

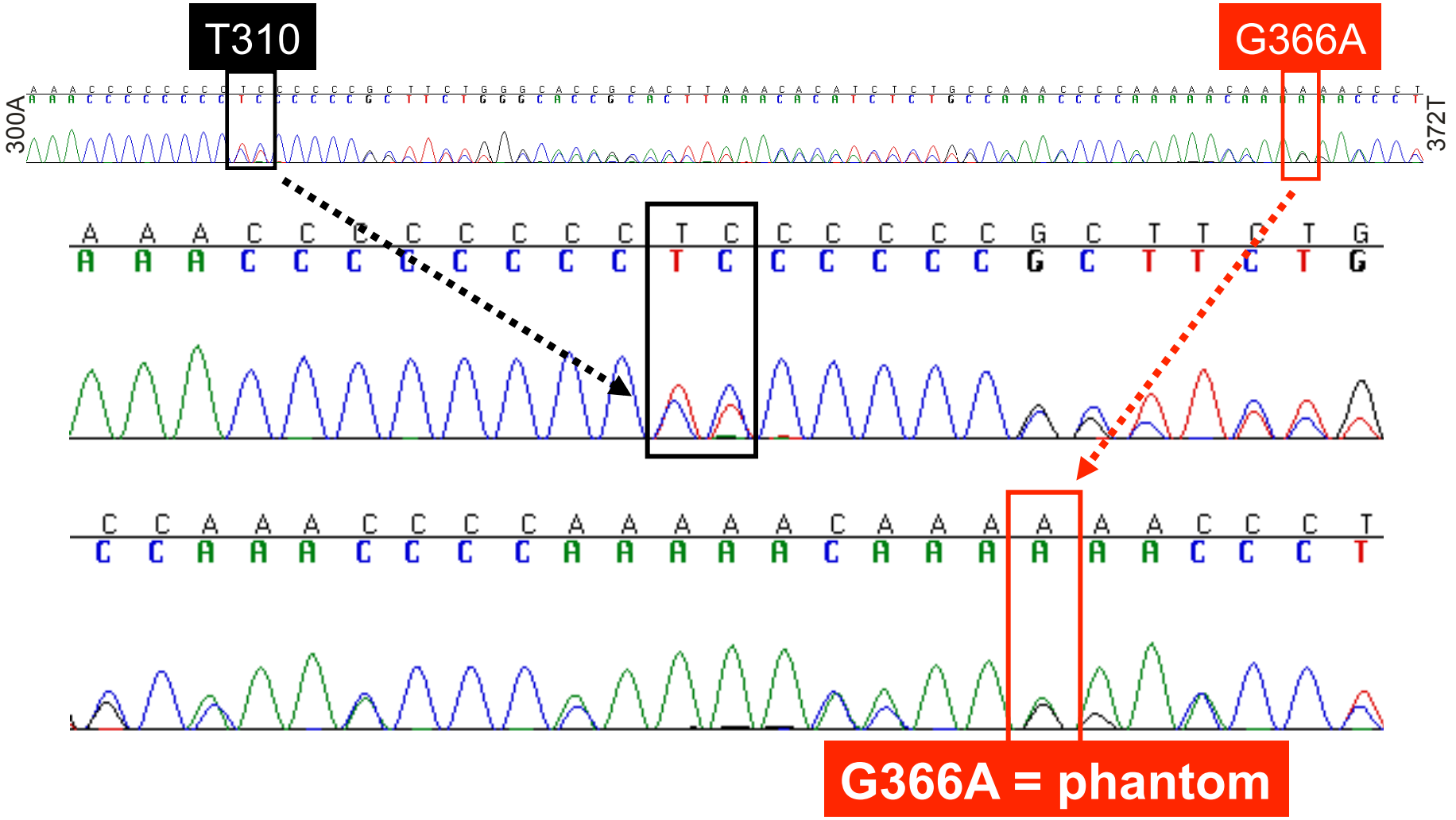
R0, H2, H6, HV0, J, K

Connection: A366G

Raw data confirm that G366A is in fact a phantom mutation caused by length variation around 310

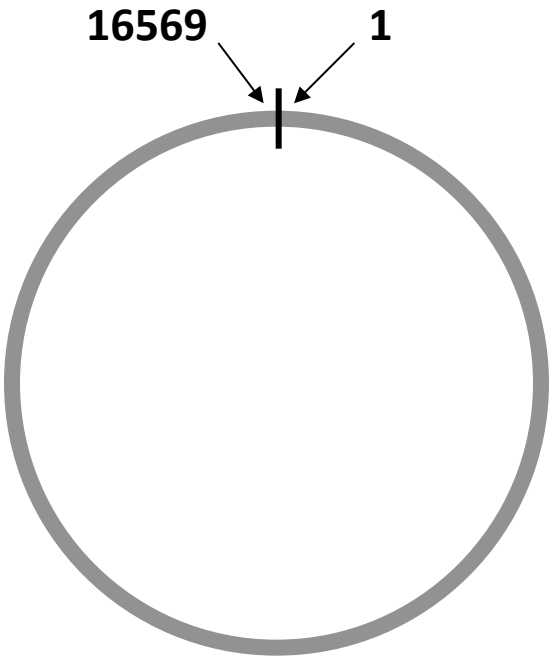


Raw data confirm that G366A is in fact a phantom mutation caused by length variation around 310



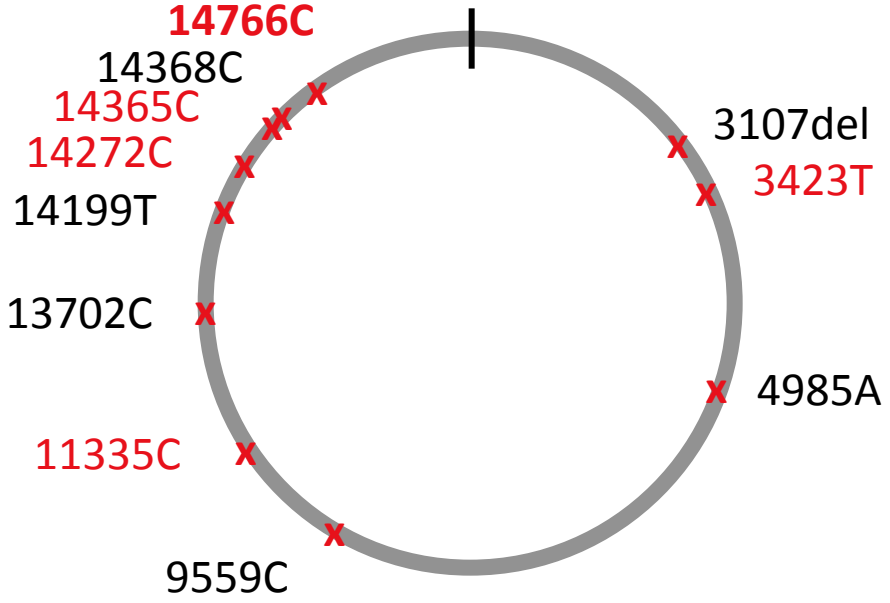
mtDNA alignment

MtDNA reference sequence



CRS

Cambridge Reference Sequence (CRS)
GenBank: M63933
Anderson et al (1981)
“Anderson-Sequence”



rCRS

revised Cambridge Reference Sequence (rCRS)
GenBank: **NC012920** (mitomap)
Andrews et al (1999)
“Andrews-Sequence”

Multiple “Anderson”-sequences!!

Alignment of mtDNA sequences

Control Region sequence WP (16024-576; 1224 bp)

TTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAA
ATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCCATGCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCA
AAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCCTCTCGCCCCCA
TGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTA
AAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAAC
CACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCTATGTCGCAGTATCTGTCTTTGATTC
CTGCCTCATCCTATTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATRATAATAACAATTGAATGTCTGCAC
AGCCGCTTTCCACACAGACATCATAACAAAAAATTTCCACCAAACCCCCCCTCCCCCGCTTCTGGCCACAGCACTTAAACACATCTCTGCCAAACCCCAAAAACAAAGAACCCT
AACACCAGCCTAACCAGATTTCAAATTTTATCTTTTGGCGGTATGCACTTTTAACAGTCACCCCCCAACTAACACATTATTTTCCCCTCCCACTCCCATACTACTAATCTCATCAATA
CAACCCCCGCCCATCCTACCCAGCACACACACA--CGCTGCTAACCCCATACCCCGAACCAACCAAACCCCAAAGACACCCCCCCCACA

16024-576
T16189C -16193.1C T16356C T16362C T16519C A234R A263G -315.1C A523del C524del -573.1C -573.2C

Difference-coded alignment is easy as long as there are no indels, because they can be aligned in multiple ways and we have no experimental approach to determine which is the correct one

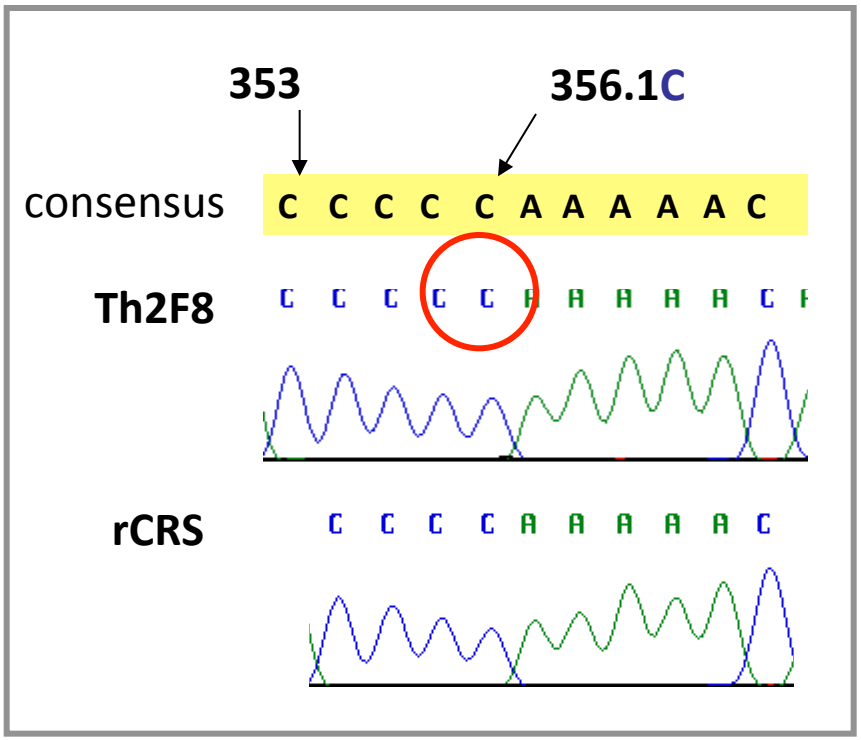
Problem
Alignment of indels

Alignment rules (length variants)

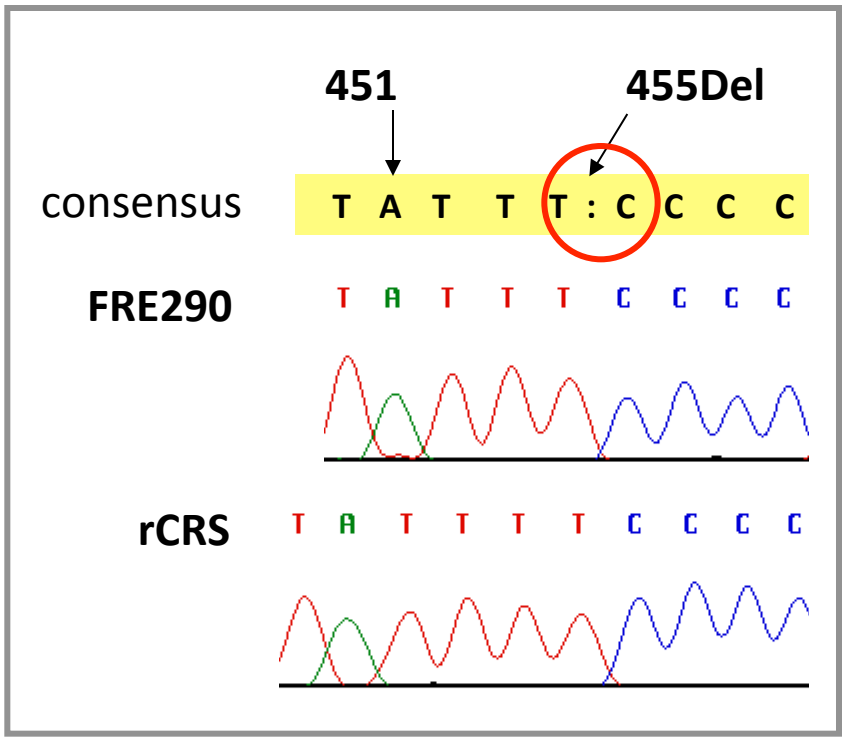
ISFG guidelines (Bär et al 2000, Carracedo et al 2000)

3’ convention for indels in length variants

365.1C (Th2F8, Thailand, EMPOP)



455Del (FRE390, Germany, EMPOP)



Alignment rules (length variants)

ISFG guidelines (Bär et al. 2000, Carracedo et al. 2000)

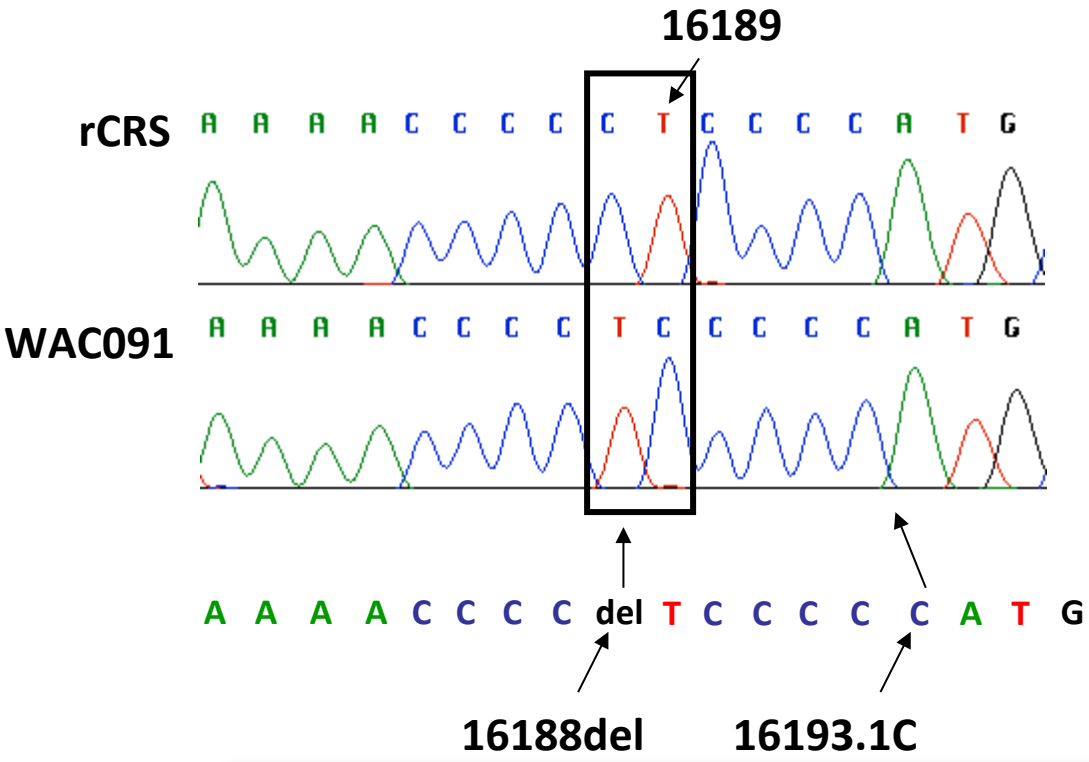
- 3’ convention for indels in length variants

Hierarchical Wilson rules (Wilson et al. 2002)

1) minimal number of differences (unweighted parsimony)
2) indels > transitions > transversions
3) 3’ convention for indels

Multiple possible alignments

16188T 16189C 16356C 16519C 263G 315.1C (WAC091)



16188T 16189C

16188Del 16193.1C

Wilson et al (2002)

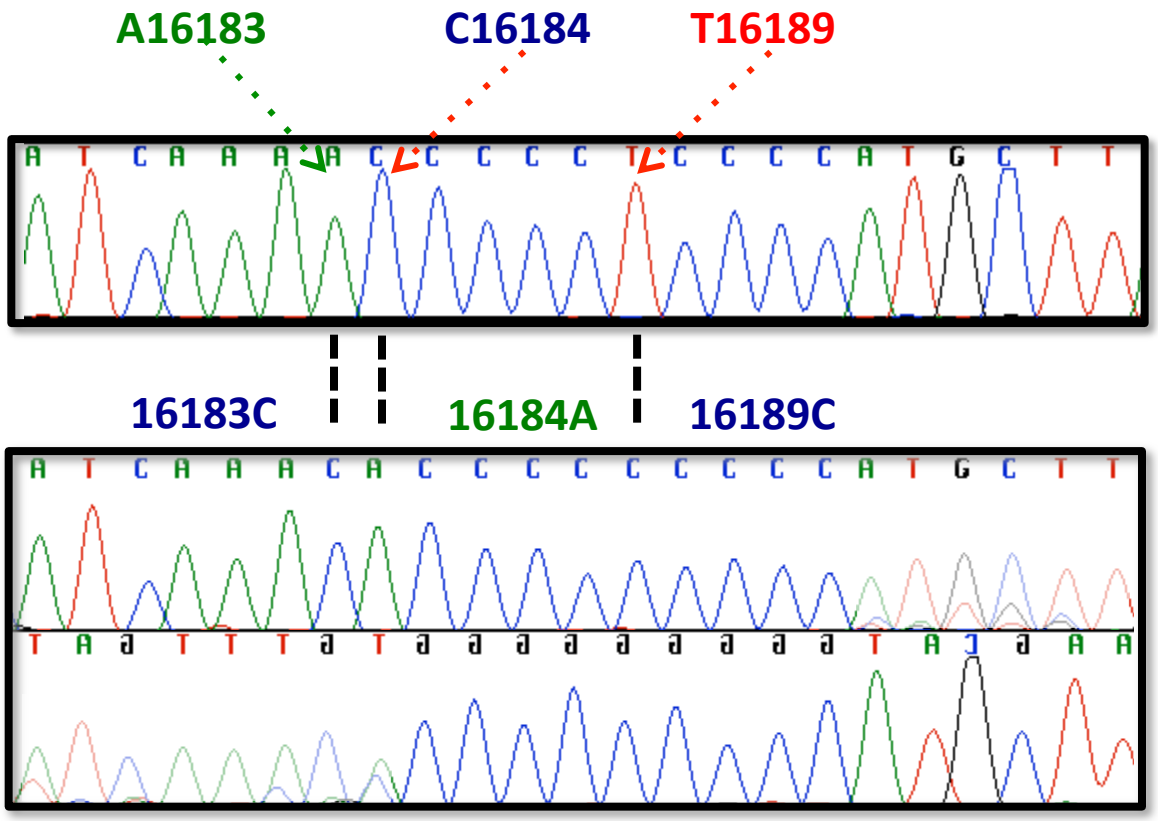
Rule 1. Parsimony rule – search for least differences to rCRS
Rule 2. Favor indels over transitions over transversions
Rule 3. 3‘ alignment

Effect on database searches

Diff (16184-16194)	16188T 16189C	16188del 16193.1C
0	17	0
1	3,524	21

(EMPOP dev 2008-05; N=26.930)

Unweighted Parsimony and alignment



MAS019

16147T
16183C
16184A
16189C
16217C
16235G
16294T
16519C
73G
263G
309.1C
315.1C
523DEL
524DEL

16183C 16184A 16189C — **phylogenetic alignment**

16182.1C 16189DEL — **parsimonious alignment**

most parsimonious type sometimes difficult to determine

Unweighted parsimony and alignment

Parsimony

It seems to be a straight forward approach to determine the most parsimonious alignment because it uses the fewest number of differences to the rCRS, but…

CHN.ASN.000206 (24 differences to rCRS)

16181C 16182C 16183C 16189C 16213A 16217C 16242T 16261T 16292T 16301T 16519C

61A 62A 73G 183G 263G 309.1C 309.2C 309.3C 315.1C 323N 324N 523Del 524Del

How many possible alignments tolerating 24+1 mutations?

953,110

It is practically **impossible** to manually determine all alignments (and therefore the most parsimonious alignment)

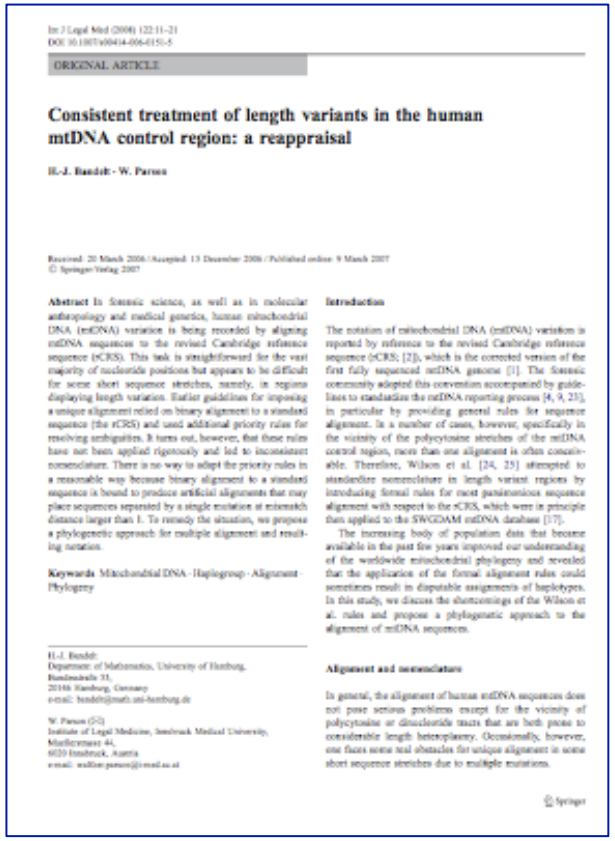
Phylogenetic alignment

Bandelt and Parson (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal, Int J Legal Med 122:1-21

Rule 1. Phylogenetic rule

Rule 2. Anchor 16189 and 310

Rule 3. 3’ alignment



Int J Legal Med (2008) 122:11–21
DOI 10.1007/s00414-006-0151-5

ORIGINAL ARTICLE

Consistent treatment of length variants in the human mtDNA control region: a reappraisal

H.-J. Bandelt · W. Parson

Received: 20 March 2006 / Accepted: 13 December 2006 / Published online: 9 March 2007
© Springer-Verlag 2007

Abstract In forensic science, as well as in molecular anthropology and medical genetics, human mitochondrial DNA (mtDNA) variation is being recorded by aligning mtDNA sequences to the revised Cambridge reference sequence (rCRS). This task is straightforward for the vast majority of nucleotide positions but appears to be difficult for some short sequence stretches, namely, in regions displaying length variation. Earlier guidelines for imposing a unique alignment relied on binary alignment to a standard sequence (the rCRS) and used additional priority rules for resolving ambiguities. It turns out, however, that these rules have not been applied rigorously and led to inconsistent nomenclature. There is no way to adapt the priority rules in a reasonable way because binary alignment to a standard sequence is bound to produce artificial alignments that may place sequences separated by a single mutation at mismatch distance larger than 1. To remedy the situation, we propose a phylogenetic approach for multiple alignment and resulting notation.

Keywords Mitochondrial DNA · Haplogroup · Alignment · Phylogeny

Introduction

The notation of mitochondrial DNA (mtDNA) variation is reported by reference to the revised Cambridge reference sequence (rCRS; [2]), which is the corrected version of the first fully sequenced mtDNA genome [1]. The forensic community adopted this convention accompanied by guidelines to standardize the mtDNA reporting process [4, 9, 23], in particular by providing general rules for sequence alignment. In a number of cases, however, specifically in the vicinity of the polycytosine stretches of the mtDNA control region, more than one alignment is often conceivable. Therefore, Wilson et al. [24, 25] attempted to standardize nomenclature in length variant regions by introducing formal rules for most parsimonious sequence alignment with respect to the rCRS, which were in principle then applied to the SWGDAM mtDNA database [17].

The increasing body of population data that became available in the past few years improved our understanding of the worldwide mitochondrial phylogeny and revealed that the application of the formal alignment rules could sometimes result in disputable assignments of haplotypes. In this study, we discuss the shortcomings of the Wilson et al. rules and propose a phylogenetic approach to the alignment of mtDNA sequences.

Alignment and nomenclature

In general, the alignment of human mtDNA sequences does not pose serious problems except for the vicinity of polycytosine or dinucleotide tracts that are both prone to considerable length heteroplasmy. Occasionally, however, one faces some real obstacles for unique alignment in some short sequence stretches due to multiple mutations.

H.-J. Bandelt
Department of Mathematics, University of Hamburg,
Bundesstraße 55,
20146 Hamburg, Germany
e-mail: bandelt@math.uni-hamburg.de

W. Parson (✉)
Institute of Legal Medicine, Innsbruck Medical University,
Müllerstrasse 44,
6020 Innsbruck, Austria
e-mail: walther.parson@i-med.ac.at

Springer

Alignment rules

ISFG guidelines (Bär et al. 2000, Carracedo et al 2000)

- 3’ convention for indels in length variants

Hierarchical Wilson rules (Wilson et al. 2002)

1) minimal number of differences (unweighted parsimony)
2) indels > transitions > transversions
3) 3’ convention for indels

Phylogenetic alignment (Bandelt and Parson, 2008)

1) alignment according to mtDNA phylogeny
2) C-tract conventions
3) 3’ convention for indels except phylogeny suggests otherwise

“MitoTyper” rules (Budowle et al. 2010)

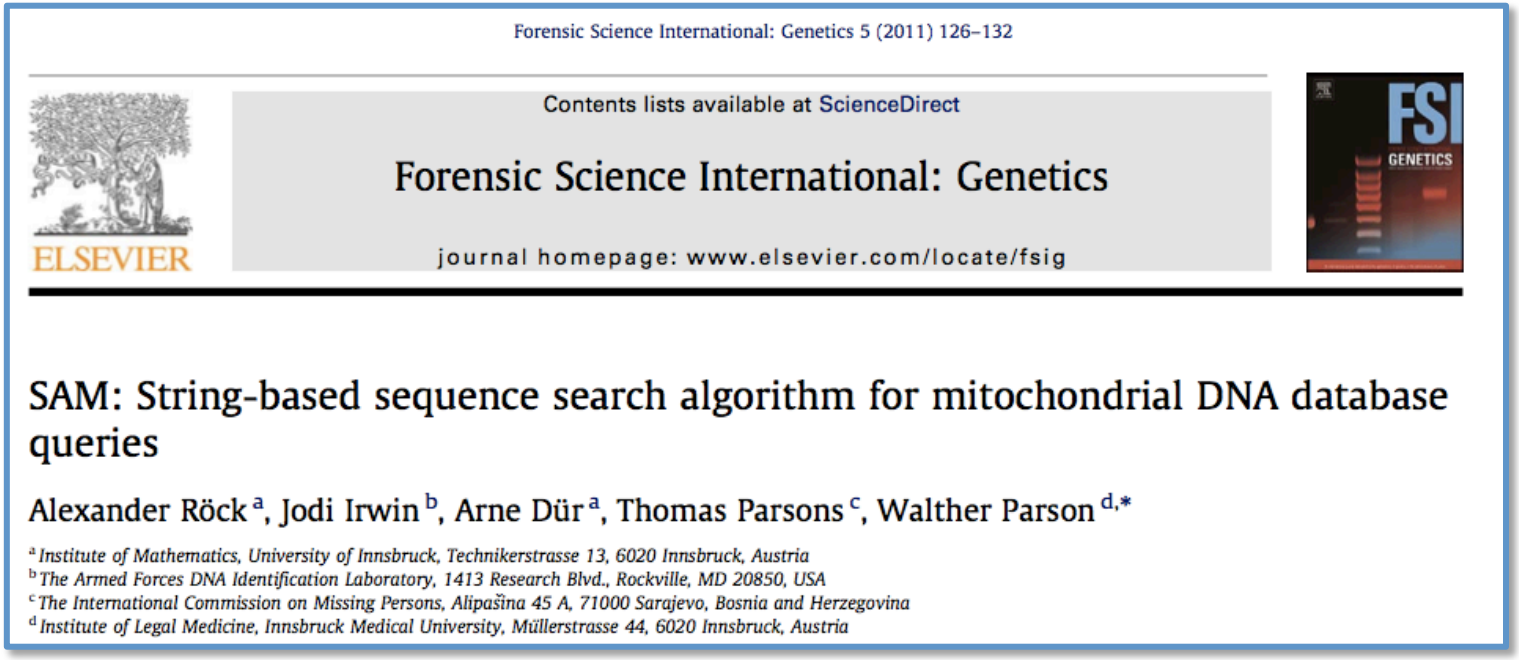
- Corrected and extended Wilson rules based on unweighted parsimony

New SWGDAM mtDNA guidelines use phylogenetic-based alignment (July 2013)

Solution to the alignment dilemma

How should a practitioner know about phylogeny?

EMPOP translates all difference-coded haplotypes into **alignment-free** strings of nucleotides. Database search becomes **independent** from alignment



Forensic Science International: Genetics 5 (2011) 126–132

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

SAM: String-based sequence search algorithm for mitochondrial DNA database queries

Alexander Röck [a], Jodi Irwin [b], Arne Dür [a], Thomas Parsons [c], Walther Parson [d,*]

[a] Institute of Mathematics, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria
[b] The Armed Forces DNA Identification Laboratory, 1413 Research Blvd., Rockville, MD 20850, USA
[c] The International Commission on Missing Persons, Alipašina 45 A, 71000 Sarajevo, Bosnia and Herzegovina
[d] Institute of Legal Medicine, Innsbruck Medical University, Müllerstrasse 44, 6020 Innsbruck, Austria

Haplotype not missed in a database search

EMPOP uses SAM since Vs. 2 (April 2010)

mtDNA haplogrouping

mtDNA haplogroups - Phylotree

Control Region sequence WP (16024-576; 1124 bp)

TTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAA
ATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCCATGCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCA
AAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCCTCTCCGCCCCCA
TGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTA
AAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAAC
CACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCTATGTCGCAGTATCTGTCTTTGATTC
CTGCCTCATCCTATTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATRATAATAACAATTGAATGTCTGCAC
AGCCGCTTTCCACACAGACATCATAACAAAAAATTTCCACCAAACCCCCCCTCCCCGCCGCTTCTGGCCACAGCACTTAAACACATCTCTGCCAAACCCCAAAAACAAAGAACCCT
AACACCAGCCTAACCAGATTTCAAATTTTATCTTTTGGCGGTATGCACTTTTAACAGTCACCCCCCAACTAACACATTATTTTCCCCTCCCACTCCCATACTACTAATCTCATCAATA
CAACCCCCGCCCATCCTACCCAGCACACACACC--CGCTGCTAACCCCATACCCCGAACCAACCAAACCCCAAAGACACCCCCCCCACA

rCRS-coded

16024-576
16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523del 524del 573.1C 573.2C

Haplotype

genetic markers that are inherited together and do not undergo recombination (e.g. SNPs).

Haplogroup

Clusters of (similar) haplotypes that share a common ancestor. MtDNA haplogroups are named with alternating letters and numbers (e.g. H1b1).

mtDNA haplogroups - Phylotree

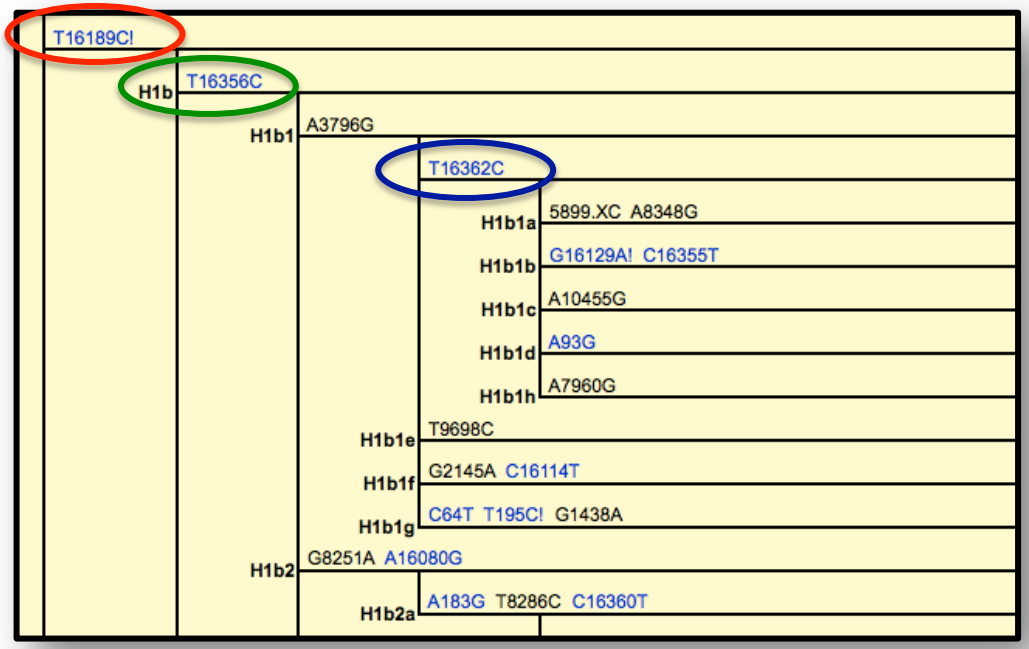
Control Region sequence WP (16024-576; 1124 bp)

TTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACTCACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAA
ATACTTGACCACCTGTAGTACATAAAAACCCAATCCACATCAAAACCCCCCCCCCCATGCTTACAAGCAAGTACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCA
AAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGTACATAGTACATAAAGCCATTTACCGTACATAGCACATTACAGTCAAATCCCCCTCTCCCCCCA
TGGATGACCCCCCTCAGATAGGGGTCCCTTGACCACCATCCTCCGTGAAATCAATATCCCGCACAAGAGTGCTACTCTCCTCGCTCCGGGCCCATAACACTTGGGGGTAGCTA
AAGTGAACTGTATCCGACATCTGGTTCCTACTTCAGGGCCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAAC
CACTCACGGGAGCTCTCCATGCATTTGGTATTTTCGTCTGGGGGGTATGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCTATGTCGCAGTATCTGTCTTTGATTC
CTGCCTCATCCTATTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACTTACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATRATAATAACAATTGAATGTCTGCAC
AGCCGCTTTCCACACAGACATCATAACAAAAAATTTCCACCAAACCCCCCCTCCCCCCGCTTCTGGCCACAGCACTTAAACACATCTCTGCCAAACCCCAAAAACAAAGAACCCT
AACACCAGCCTAACCAGATTTCAAATTTTATCTTTTGGCGGTATGCACTTTTAACAGTCACCCCCCAACTAACACATTATTTTCCCCTCCCACTCCCATACTACTAATCTCATCAATA
CAACCCCCGCCCATCCTACCCAGCACACACACC--CGCTGCTAACCCCATACCCCGAACCAACCAAACCCCAAAGACACCCCCCCCACA

rCRS-coded

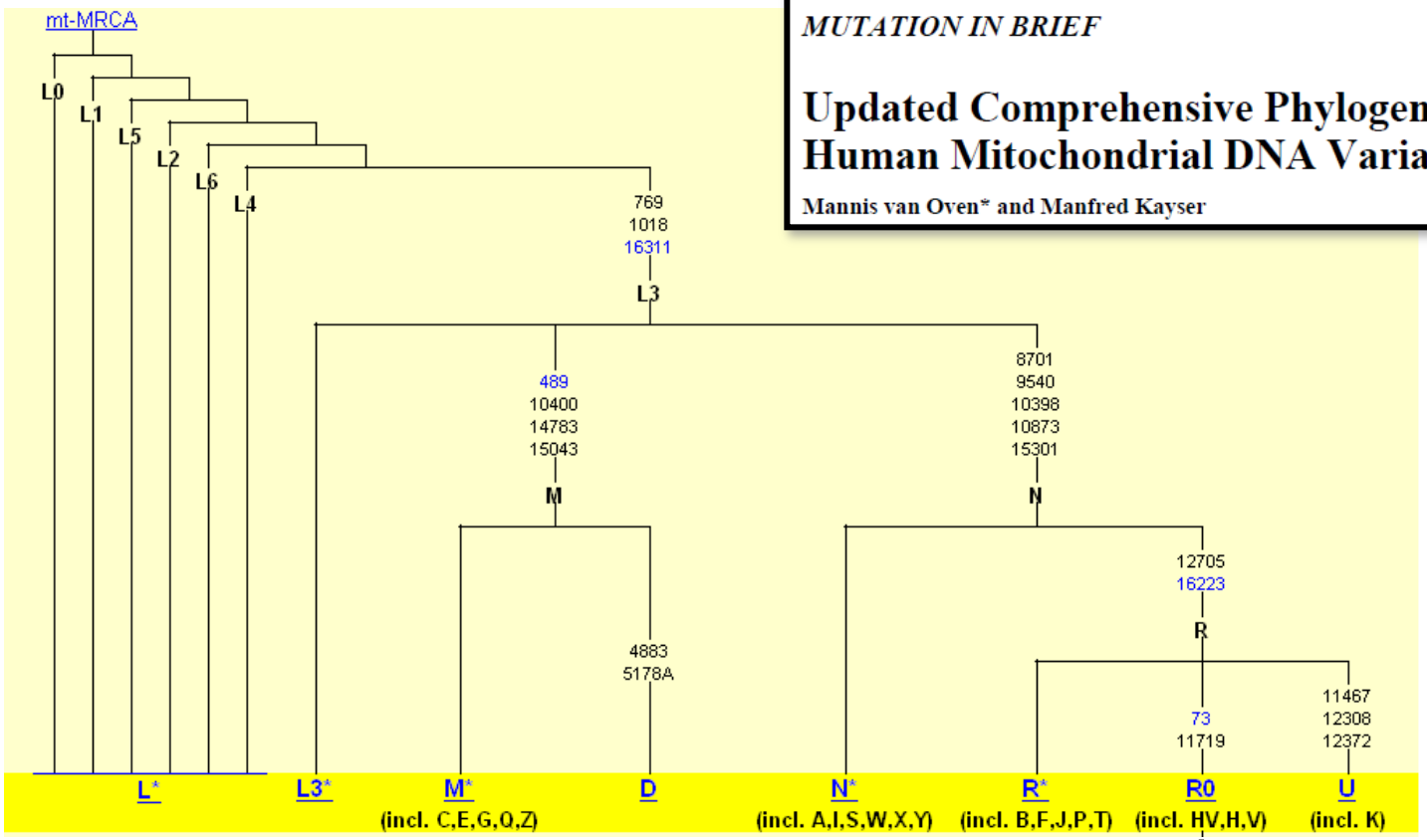
16024-576

16189C 16193.1C **16356C** **16362C** 16519C 234R 263G 315.1C 523del 524del 573.1C 573.2C - **hg H1b1**



vanOven and Kayser, 2009; www.phylotree.org

Where are the haplogroup names defined?



HUMAN MUTATION Mutation in Brief #1039, 30:E386-E394, (2008) Online

MUTATION IN BRIEF

Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation

Mannis van Oven* and Manfred Kayser

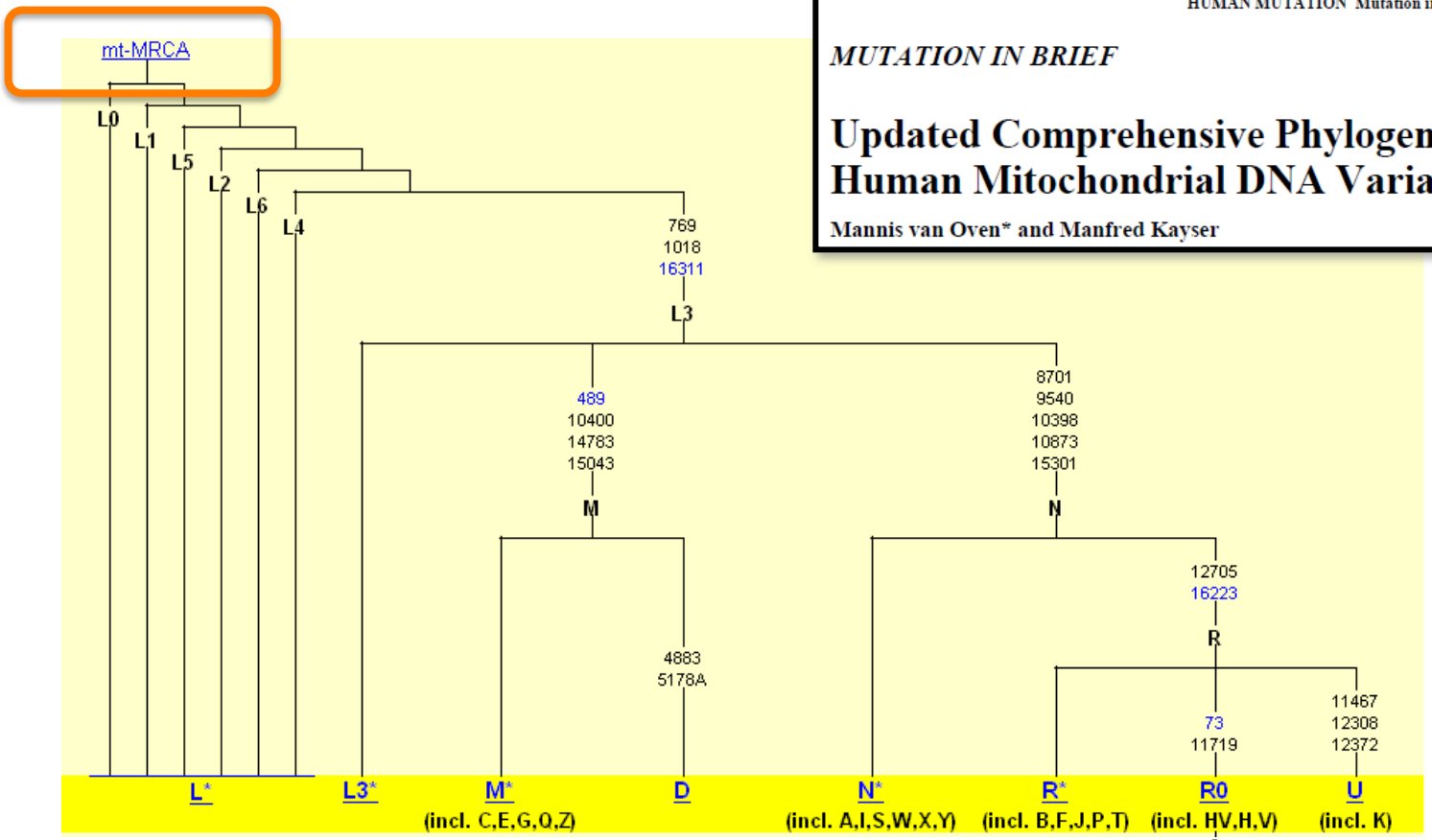
rCRS

Haplogroup L
Africa
(America)

Haplogroup M
Eastern Eurasia
Native America

Haplogroup N
Eurasia
(Native)America

Where are the haplogroup names defined?



HUMAN MUTATION Mutation in Brief #1039, 30:E386-E394, (2008) Online

MUTATION IN BRIEF

Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation

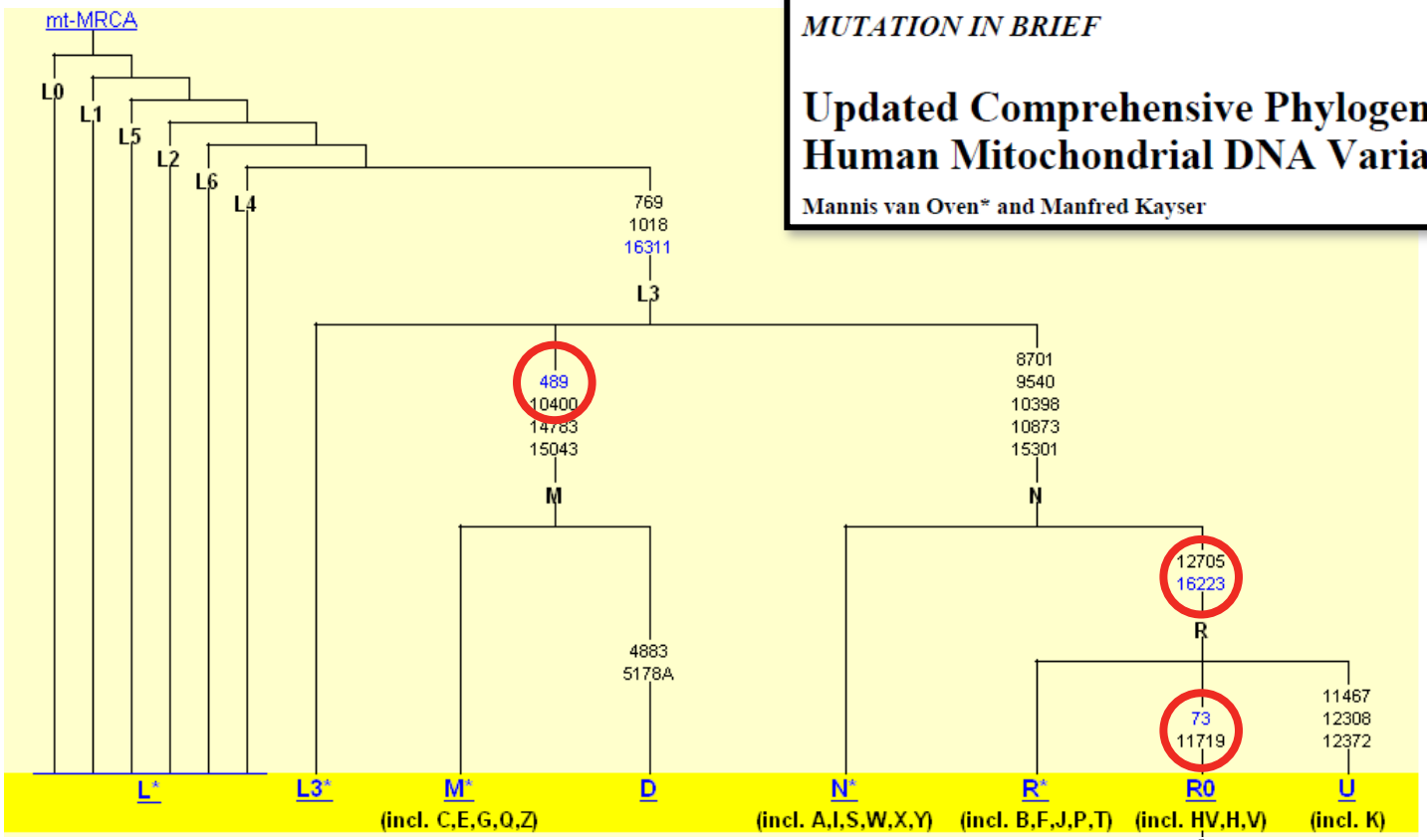
Mannis van Oven* and Manfred Kayser

rCRS

Note: Phylotree is **rooted** in the **Most Recent Common Ancestor** of all human mtDNA lineages known today.

But: mtDNA haplotypes are **reported** to the **rCRS** (RSRS not recommended in forensics)

Where are the haplogroup names defined?



HUMAN MUTATION Mutation in Brief #1039, 30:E386-E394, (2008) Online

MUTATION IN BRIEF

Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation

Mannis van Oven* and Manfred Kayser

489

16223

73

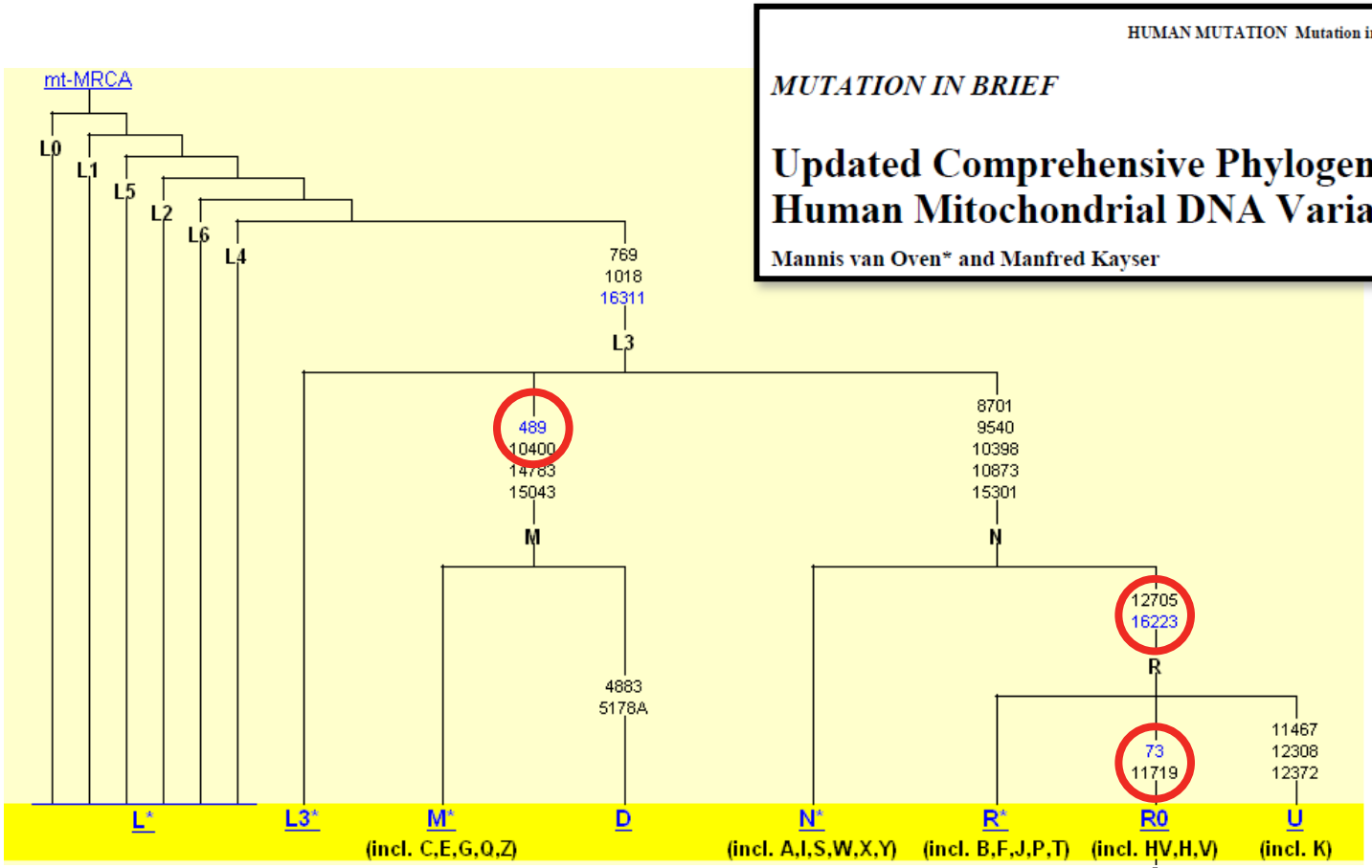
rCRS

Haplogroup L
Africa
(America)

Haplogroup M
Eastern Eurasia
Native America

Haplogroup N
Eurasia
(Native)America

Phylotree



HUMAN MUTATION Mutation in Brief #1039, 30:E386-E394, (2008) Online

MUTATION IN BRIEF

Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation

Mannis van Oven* and Manfred Kayser

489

16223

73

rCRS

16024-576

H1c3	195C	257G	263G	315.1C	477C	16519C			
M2	73G	199C	263G	315.1C	447G	489C	16223T	16258-	16274A
K1a	73G	146C	263G	315.1C	497T	16224C	16311C	16519C	

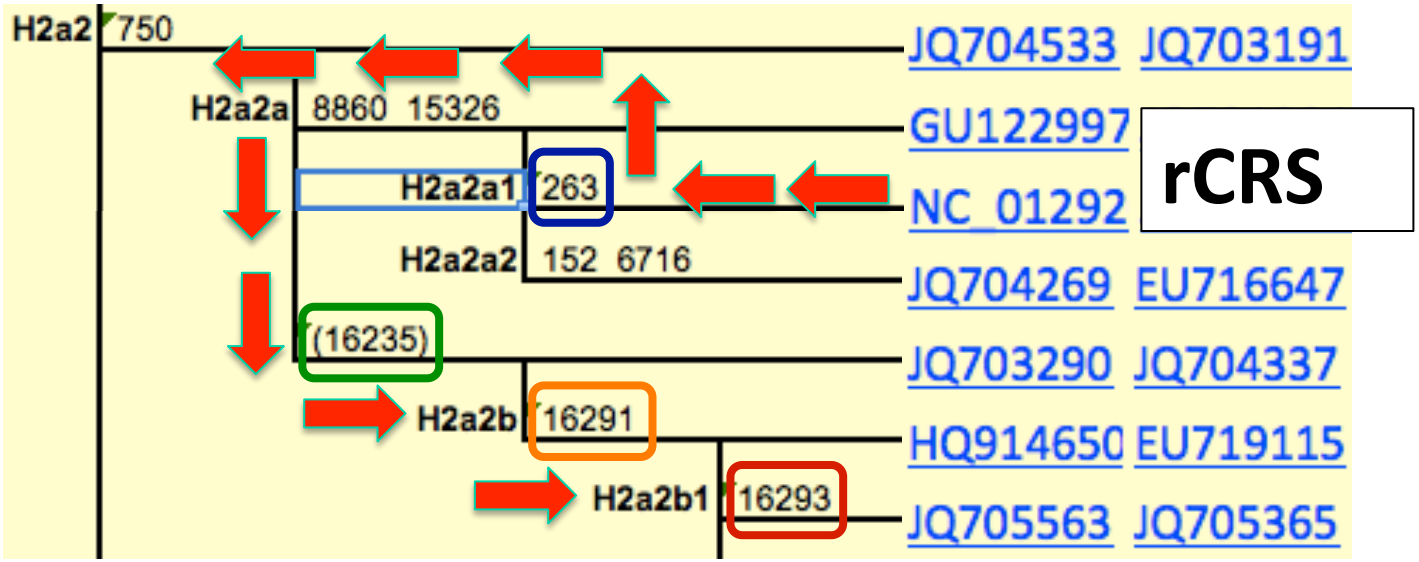
Manual haplogrouping using Phylotree

EMP00539 (Arizona)

16024-576

16235G 16291T 16293G 263G 315.1C

16235G 16291T 16293G 263G n.a. - hg H2a2b1



modified from Phylotree B.16

Limitations of the tree

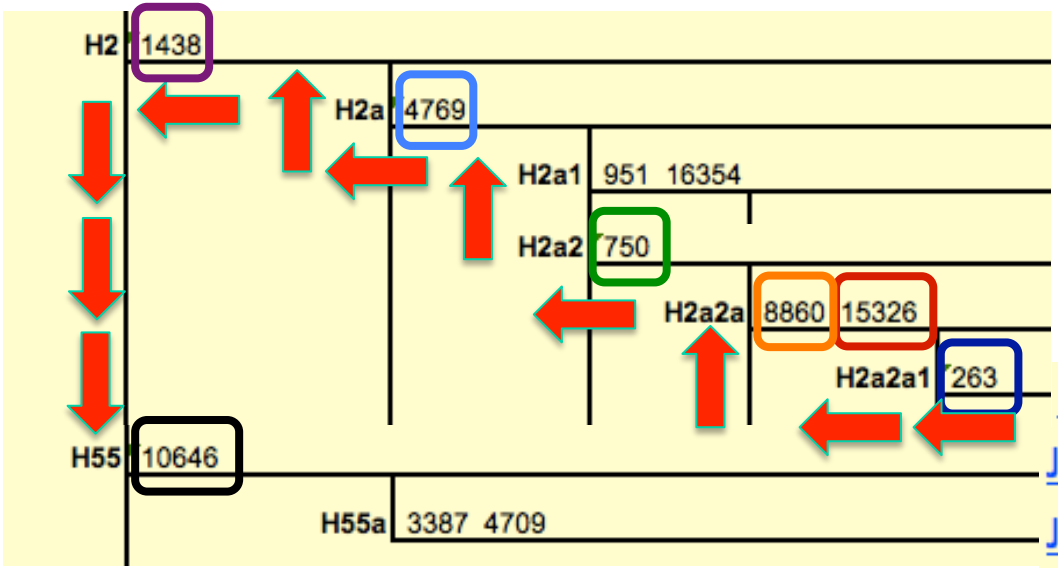
ABS133 (Argentina) 16024-576 10646

16189C 16292T 16519C 71A 153G 204C 207A 263G 315.1C 373G 10646A - hg H55

mtGenome - hg H55

16189C 16519C 153G 204C 263G 315.1C 750G 1438G 4769G 8860G 10646A 15326G

263G n.a. 750G 1438G 4769G 8860G 10646A 15326G



rCRS

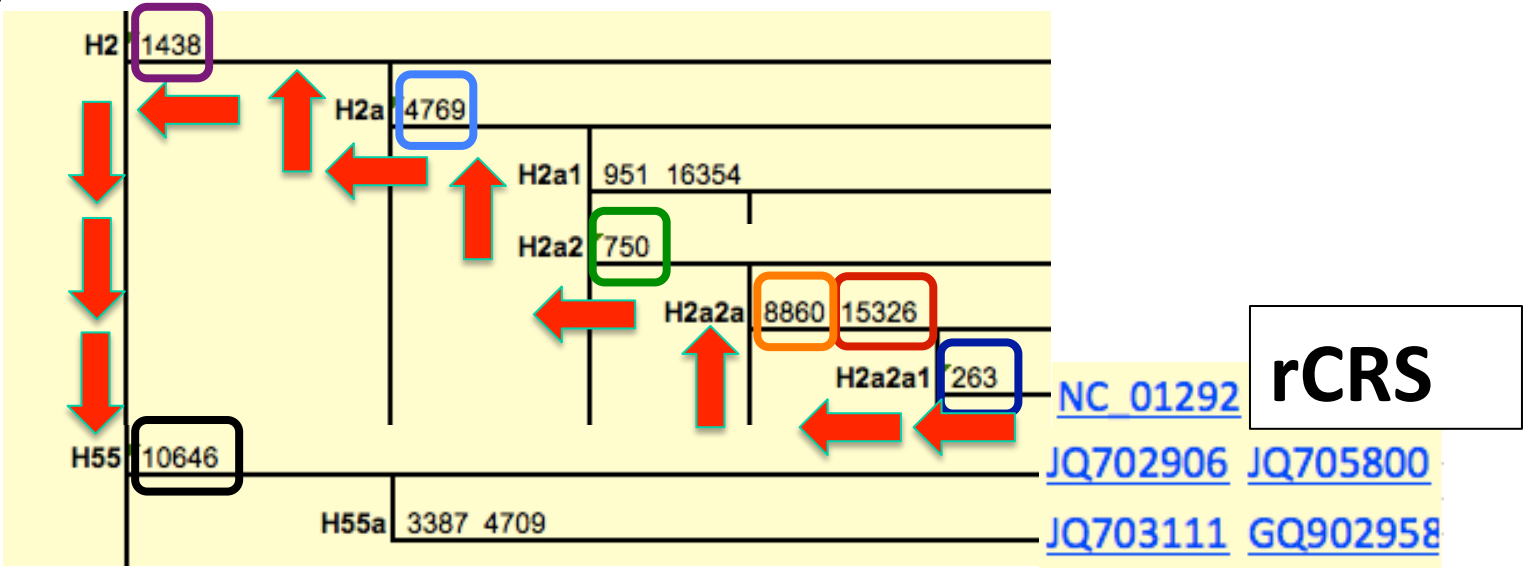
modified from Phylotree B.16

Limitations of the tree

Phylotree branches only list selected mutations required to describe the haplogroup (private mutations are missing)

“virtual haplotypes”

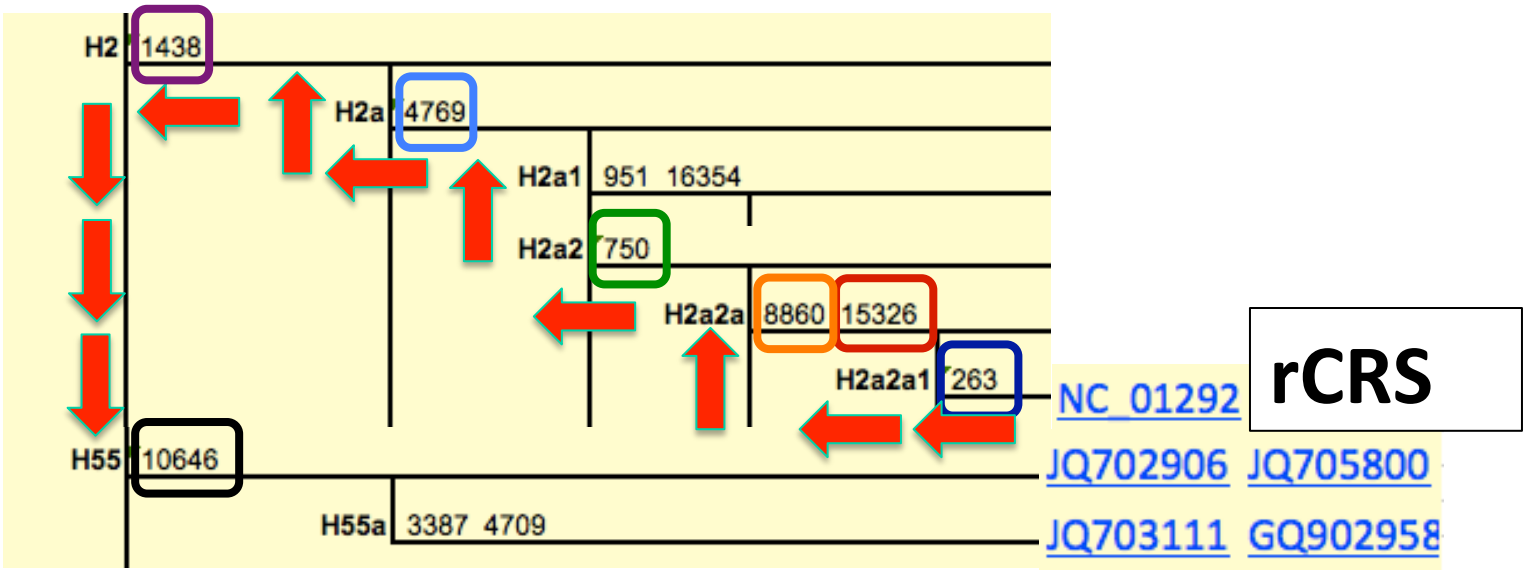
Matches may be missed due to absence of private mutations in branches



modified from Phylotree B.16

Limitations of the tree

To increase chances of finding the correct haplogroups, haplogrouping should be performed on real haplotypes, preferable full mtGenomes, not only virtual haplotypes



modified from Phylotree B.16

EMMA

Forensic Science International: Genetics 7 (2013) 601–609

Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA)☆

Alexander W. Röck[a], Arne Dür[b], Mannis van Oven[c], Walther Parson[a,d,*]

[a] Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
[b] Institute of Mathematics, University of Innsbruck, Innsbruck, Austria
[c] Department of Forensic Molecular Biology, Erasmus MC, University Medical Center Rotterdam, The Netherlands
[d] Penn State Eberly College of Science, University Park, PA, USA

EMMA uses

- phylotree haplogroup nomenclature
- virtual phylotree haplotypes (4,806; phylotree build 16)
- curated database of full mtGenomes used for phylotree (19,299)
- curated database of additional full mtGenomes (ca. 5,000)

Mutation rates differ across mtGenome and within hgs

	worldwide	hg X1'3	hg K	hg T
T16519C	in 18,362 of 40,246	in 8 of 8	in 893 of 931	in 1,124 of 1,171
%	45.6	100	95.5	95.9

Skewed variation between haplogroups

need to determine rate within haplogroups

606 discernible CR-HGs (relevant to forensics, phylotree, b. 16)

Manual haplogrouping of 19,171 EMPOP CR haplotypes according to phylotree build 12-15 (Nov 2011 - Sep 2012)

Requirements:

high quality sequences (EMPOP QC process)

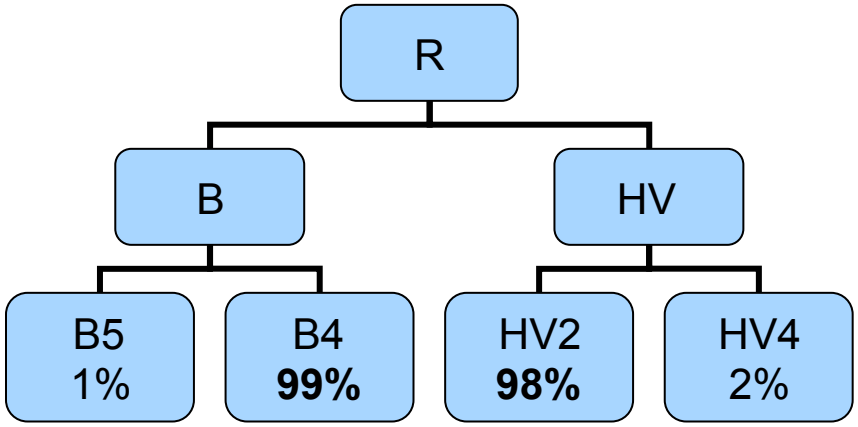
consistent phylogenetic alignment (Bandelt and Parson, 2008)

409 CR-HGs >4 haplotypes in EMPOP (R9)

Fluctuation rates

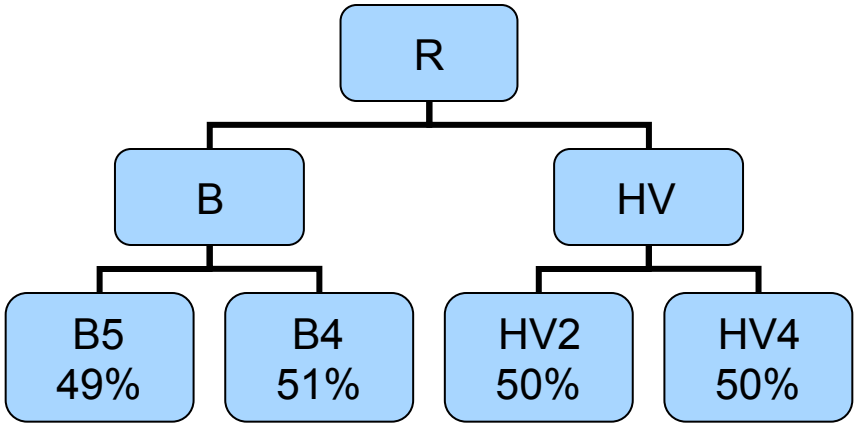
The fluctuation rate (r) is a measure for the stability of a “mutation” within a given haplogroup

Observations for T16217C



T16217C is a stable marker in hgs B4 and HV2 and therefore a strong signature for hg-estimation

Observations for T152C



T152C is strongly fluctuating in all 4 hgs and therefore of little relevance for hg-estimation

Fluctuation rates

The fluctuation rate (r) is a measure for the stability of a “mutation” within a given haplogroup

$$r_{\alpha\beta} = \frac{\sum_{\gamma} \min(n(\alpha, \gamma), n(\beta, \gamma))}{\sum_{\gamma} n(\gamma)}$$

α, β … A, C, G or T; α unequal β
γ … runs over all CR-hgs
n(x, γ) … denotes number of samples in CR-hg γ with symbol x
n(γ) … denotes total number of samples in CR-hg γ

Control region fluctuation rates

CR-HG	N	T16217C	Difference to majority	T152C	Difference to majority
A2	30	0	0	16	14
B2	40	39	1	19	19
C1	50	0	0	27	23
D1	40	2	2	12	12

$$r_{(T16217C)} = (0+1+0+2) / (30+40+50+40) = 3 / 160 = 0.01875$$

$$r_{(T152C)} = (14+19+23+12) / (30+40+50+40) = 68 / 160 = 0.425$$

$$0 \leq r \leq 0.5$$

Algorithm - maximum likelihood

Concept

Compare test haplotype to all database haplotypes by striving for maximum likelihood

$$L_t(b) = \prod_i r(b_i \rightarrow t_i)$$

b … database haplotype, t … test haplotype, i … positions

Calculation of the product is computationally intensive, therefore minimal costs are computed instead

Algorithm - minimal cost function

$$C_t(b) = \lg(\prod_i r(t_i \rightarrow t_i)/L_t(b)) \quad \text{where} \quad \lg(x) = \log 10(x)/3$$

For short motif lists, such as differences to rCRS between database and test haplotypes, the cost function can be efficiently evaluated by

$$C_t(b) = \sum_i c(b_i, t_i)$$

and

$$c(b_i,t_i) = \lg(r(t_i \rightarrow t_i)/r(b_i \rightarrow t_i))$$

are real numbers termed positional costs for the change from the base profile symbol to the test profile symbol

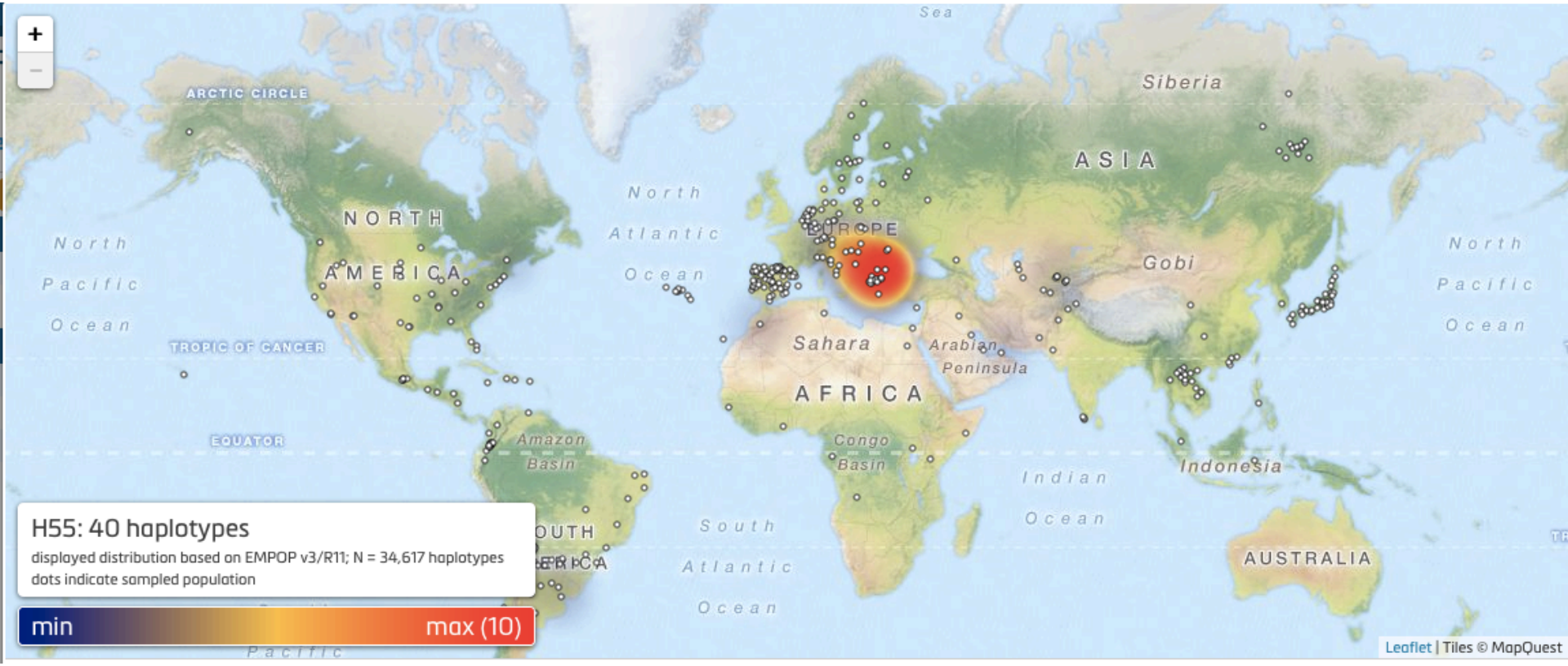
Average "mutations" yield value of approx. 1.0, unobserved transitions 2.0 and unobserved transversions 3.0

Ranking of haplotypes by total costs equals ranking by max. likelihood

H55 Example by EMMA

ABS133 (Argentina) 16024-576

16189C 16292T 16519C 71A 153G 204C 207A 263G 315.1C ~~373G~~



Exclusion criteria

Guidelines (2000-2003) define

Exclusion

when "[] ...there are more than 2 differences between 2 sequences" or
"[]... unequivocally different... []... including tissue specificity and mutation rates"
Which mutation rates? Evolutionary? Haplogroup-specific? Pedigree? Somatic?
Do we have data? Tissues?

Cannot Exclude

when sequences are identical (except for PHP and LHP in HVS-II C-tract)
Seems straight forward. Extend to all length variant regions around 16193, 455 and 573.

Inconclusive

for one difference between two sequences
Depends on position. Maybe hg-specific difference

ISFG (Bär et al 2000), EDNAP (Tully et al 2001), SWGDAM (Budowle et al 2003)

2015 ISFG mtDNA recommendations

Recommendation #1

Good laboratory practice and specific protocols for work with mtDNA must be followed in accordance with previous guidelines.

Recommendation #2

Negative and positive controls as well as extraction reagent blanks must be carried through the entire laboratory process.

Recommendation #3

Reported consensus sequences must be based on redundant sequence information, using forward and reverse sequencing reactions whenever practical.

2015 ISFG mtDNA recommendations

Recommendation #4

Manual transcription of data should be avoided and independent confirmation of consensus haplotypes by two scientists must be performed.

Recommendation #5

Laboratories using mtDNA typing in forensic casework shall participate regularly in suitable proficiency testing programs.

Recommendation #6

In population genetic studies for forensic databasing purposes, the **entire mitochondrial DNA control region** should be sequenced.

2015 ISFG mtDNA recommendations

Recommendation #7

MtDNA sequences should be aligned and reported relative to the revised Cambridge Reference Sequence (rCRS, ~~NC001807)~~, and should include the interpretation range (excluding primer sequence information). (rCRS, **NC012920**).

Recommendation #8

IUPAC conventions using **capital letters** shall be used to describe differences to the rCRS and (point heteroplasmic) mixtures. **Lower case letters** should be used to indicate mixtures **between deleted and non-deleted** (inserted and non-inserted) bases. N-designations should only be used when all four bases are observed at a single position (or if no base call can be made at a given position). For the representation of deletions, “DEL”, “del” or “-“ shall be used.

2015 ISFG mtDNA recommendations

Recommendation #9

The alignment and notation of mtDNA sequences should be performed in agreement with the mitochondrial phylogeny (established patterns of mutations). Tools to assist with the notation of mtDNA sequences are available at http://empop.org/.

Recommendation #10

In forensic casework, laboratories must establish their own interpretation and reporting guidelines for observed **length and point heteroplasmy**. The evaluation of heteroplasmy depends on the limitations of the technology and the quality of the sequencing reactions as well as the experience of the laboratory. Differences in both PHP and LHP do not constitute evidence for excluding two otherwise identical haplotypes as deriving from the same source or same maternal lineage.

2015 ISFG mtDNA recommendations

Recommendation #11

For population database samples, **length heteroplasmy** in homopolymeric sequence stretches should be interpreted by calling the **dominant variant**, which can be determined by identifying the position with the highest representation of a non-repetitive peak downstream of the affected stretch.

Recommendation #12

MtDNA population data should be subjected to analytical software tools that facilitate phylogenetic checks for data quality control. A comprehensive suite of QC tools is provided by EMPOP.

Recommendation #13

The **entire database** of available sequences should be searched with respect to the sequencing (interpretation) range to avoid biased query results.

2015 ISFG mtDNA recommendations

Recommendation #14

Laboratories must be able to justify the choice of database(s) and statistical approach used in reporting.

Recommendation #15

Laboratories must establish statistical guidelines for use in reporting an mtDNA match between two samples.

Recommendation #16

Highly variable positions such as length variants in homopolymeric stretches should be disregarded from searches for determining frequency estimates. Heteroplasmic calls should be queried in a manner that does not exclude any of the heteroplasmic variants.

Acknowledgements

EUROFORGEN Network of Excellence

FP7-SEC-2011-285487

FWF Der Wissenschaftsfonds.

Translational Research project L397
"EMPOP–an innovative human mtDNA database"

FWF Der Wissenschaftsfonds.

Research project P22880-B12
"Genetic discovery of an early medieval Alpine population "

NIJ National Institute of Justice 2011-MU-MU-K402

Maximizing mtDNA Testing Potential with the Generation of High-Quality mtGenome Reference Data

applied biosystems by life technologies

Robert Lagacé
Sharon Wootton
Reina Langit
Lisa Calandro

illumina

Nicola Oldroyd
Cydne Holt
Oliver Goldenberg
Silvio Scheel

EMPOP GMI

Mayra Eduardoff
Catarina Xavier
Christina Strobl
Gabriela Huber
Harald Niederstätter
Bettina Zimmermann
Martin Bodner
Alexander Röck

Collaborators
Arne Dür, EMPOP contributors, EMPOP IT