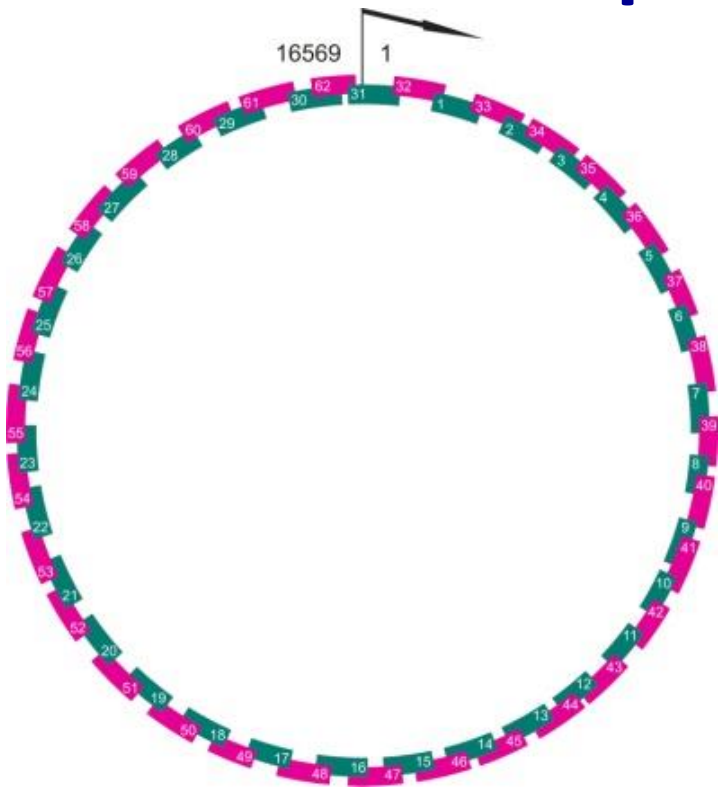
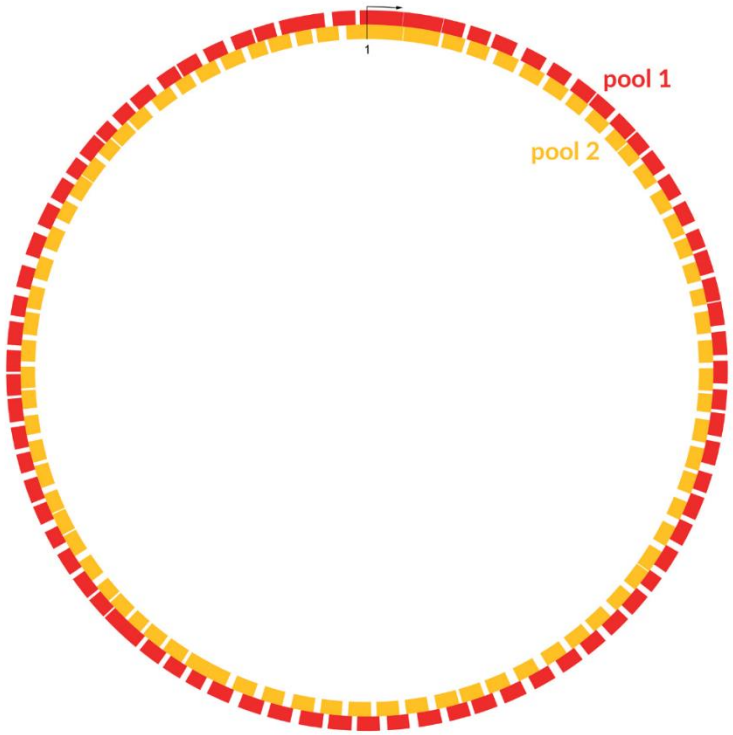
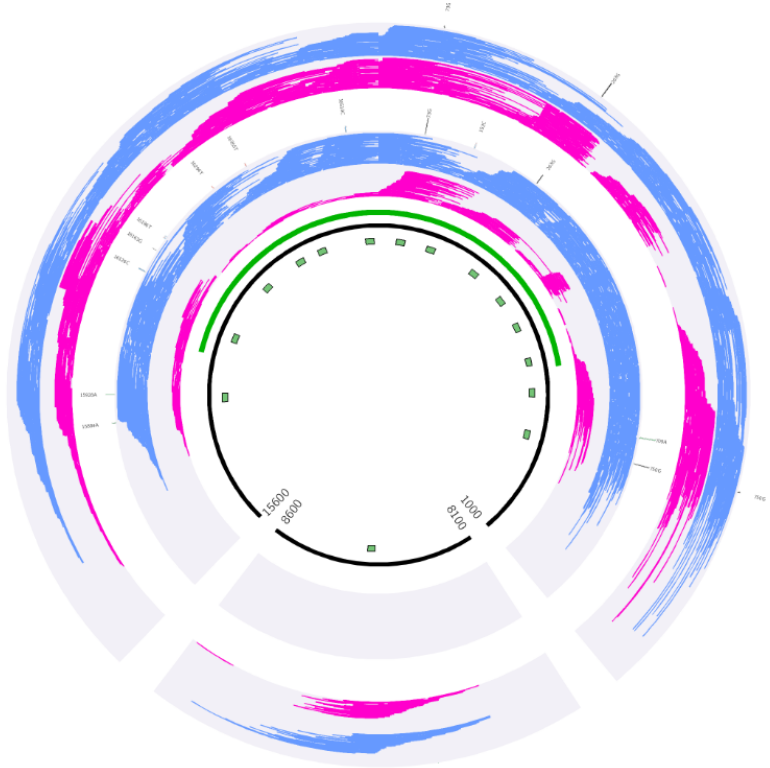
Moving Implementation Mountains: Experiencing the Forensic Laboratory NGS and Bioinformatics workflow through Simulation Minneapolis MN, Sep 26 2016

Implementing Mitochondrial DNA Massively Parallel Sequencing into Forensic Casework







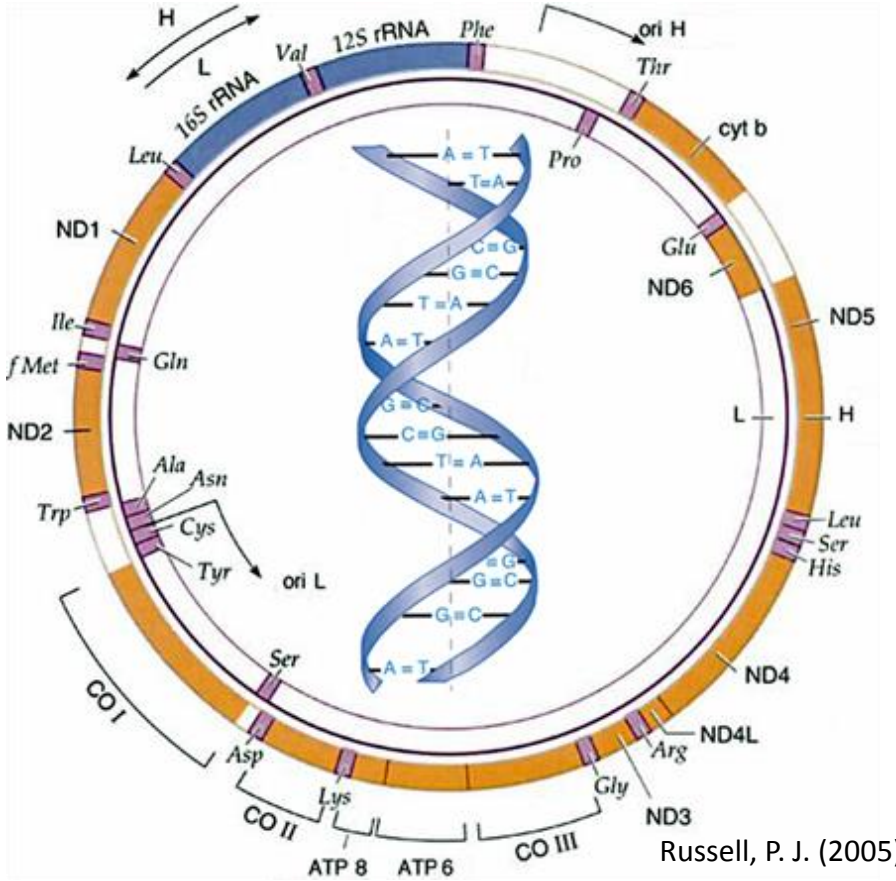
Dr. Walther Parson
assoc. Prof. Institute of Legal Medicine, Innsbruck, Austria
adj. Prof. Penn State University, PA, USA
walther.parson@gmail.com

Table of Contents

Mitochondria / mitochondrial DNA

Introduction

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)

Mitochondrial DNA copy number

Nuclear DNA (nDNA)

46 chromosomes, 3.2 x 10^9 bp
diploid

mitochondrial DNA (mtDNA)

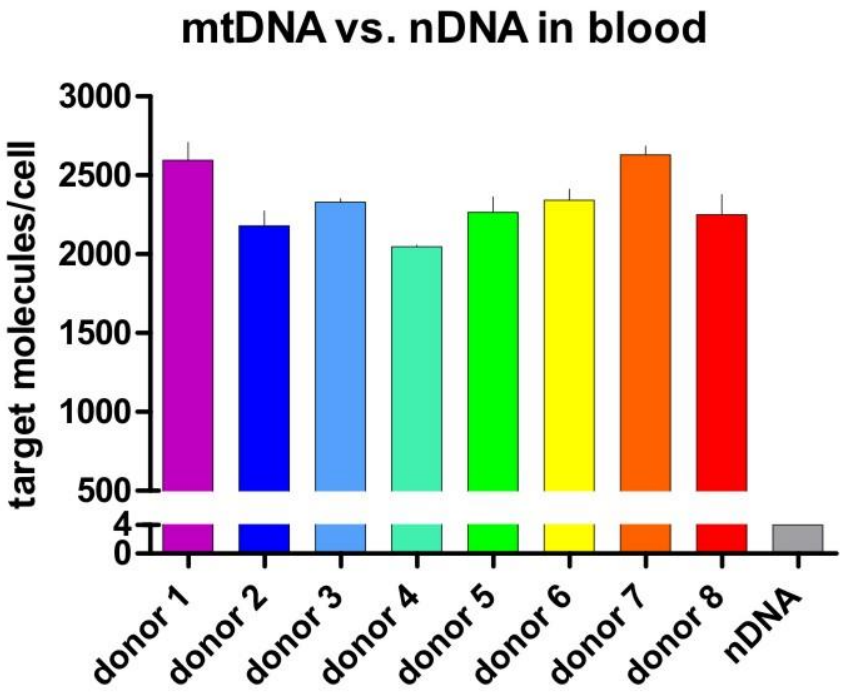
100(0)s per cell, 16.6 kbp
haploid

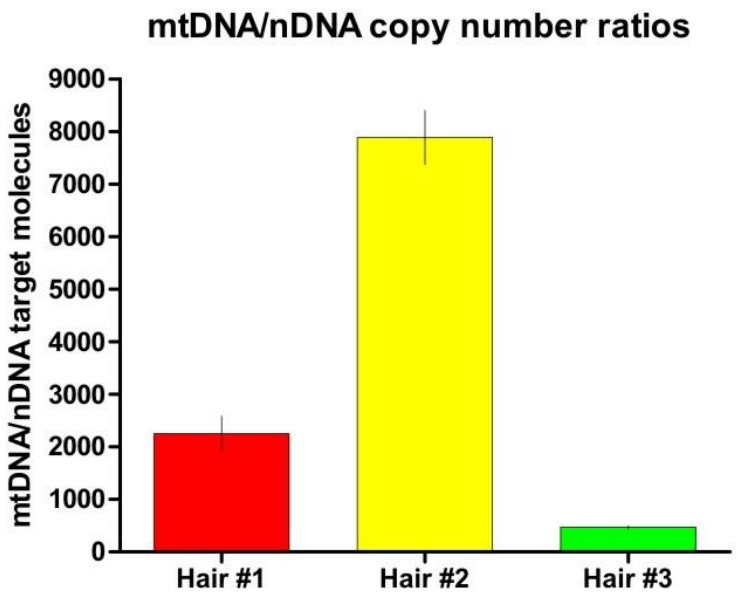
Mitochondrial DNA - copy number

Higher copy number than nDNA

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

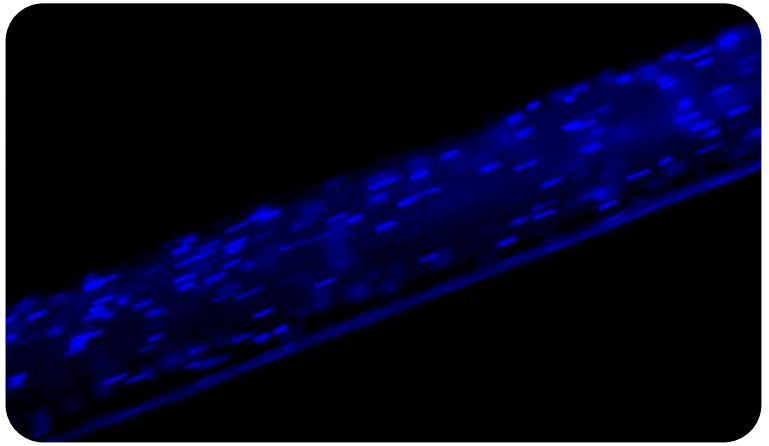
up to 1,000 mitochondria/cell (Robin and Wong, 1988)





Blood: 500 – 650x

Hair: much higher



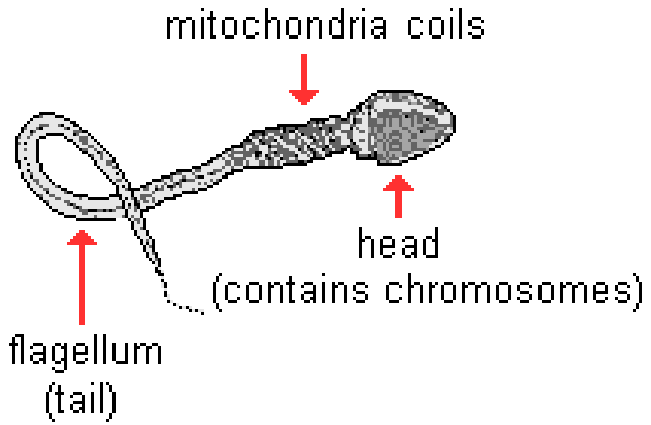
Szabo et al 2012

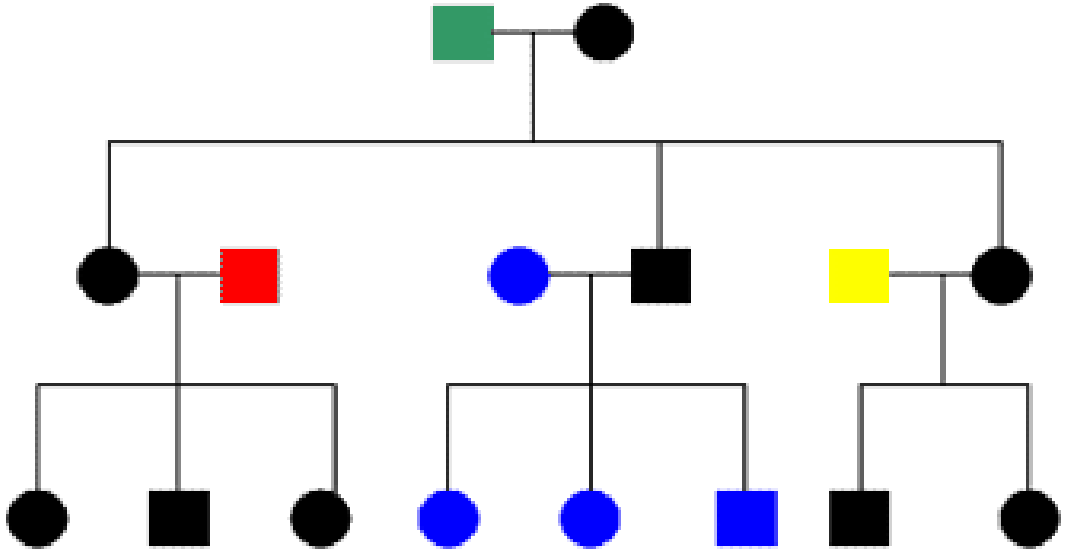
Mitochondrial DNA - Maternal Inheritance

Mitochondria derive from the fertilized egg

(100.000s versus few in sperm neck)

Ubiquitin tagging of paternal mtDNA

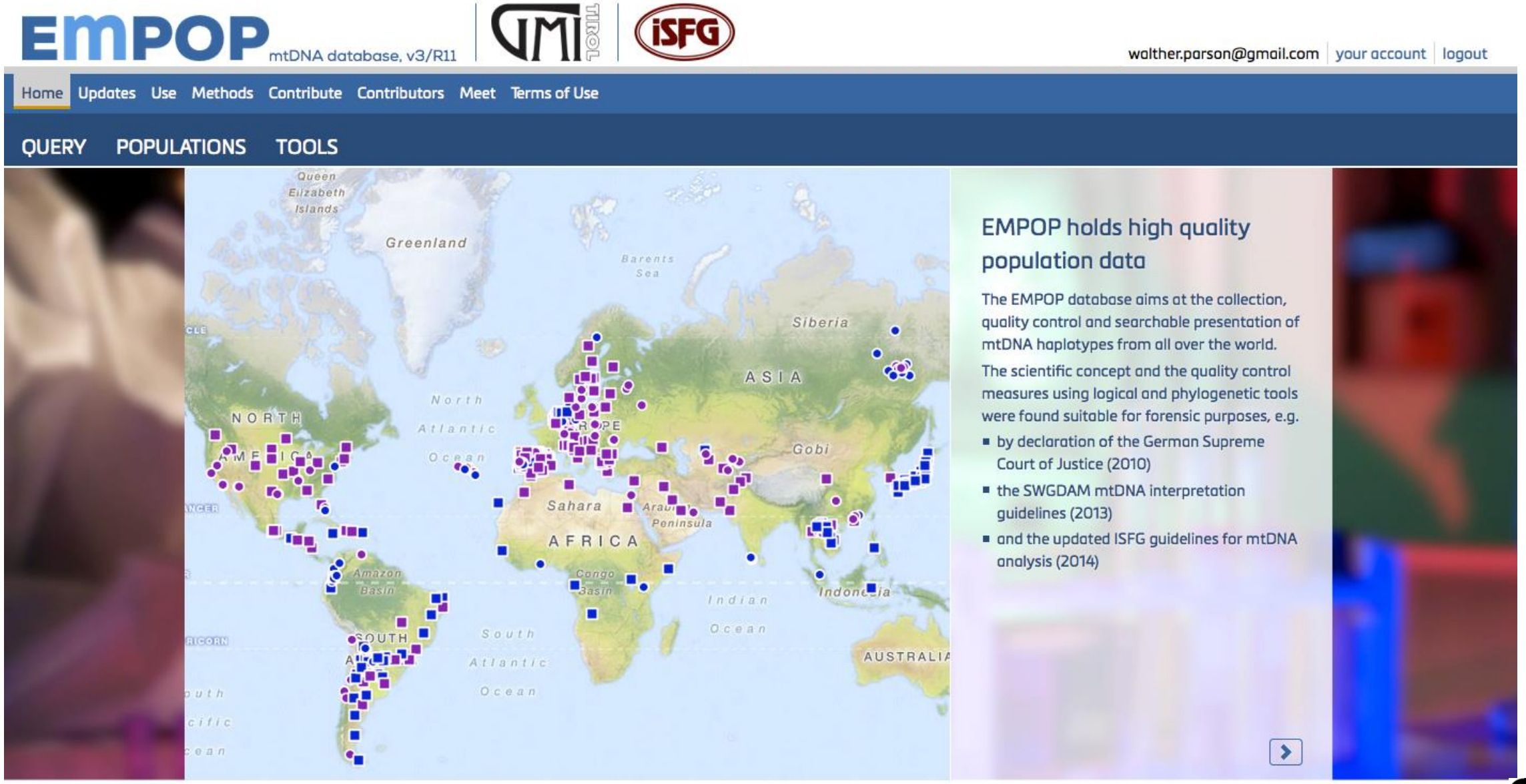




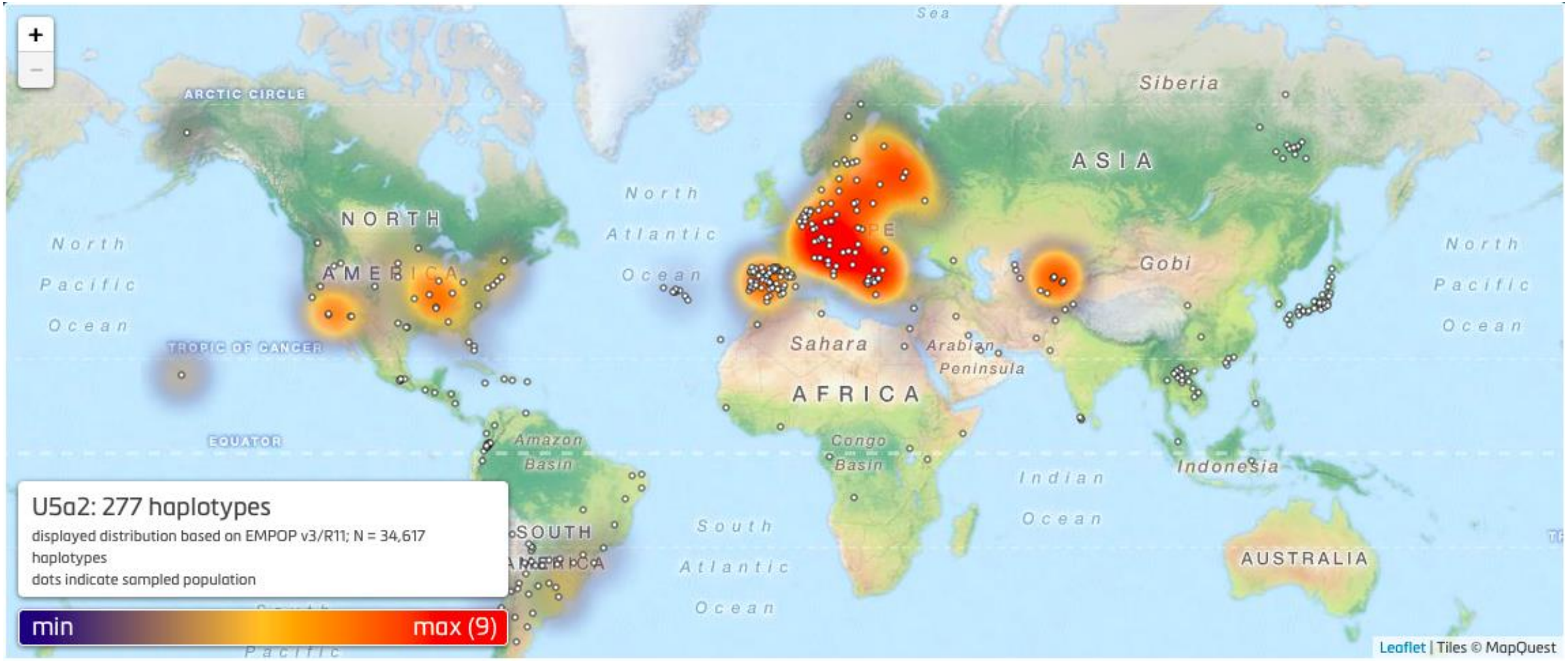
*Colors reflect inheritance of the same mitochondrial genome

Identification of maternal lineages (not individuals)

Mitochondrial DNA - Maternal Inheritance - Statistical Evaluation

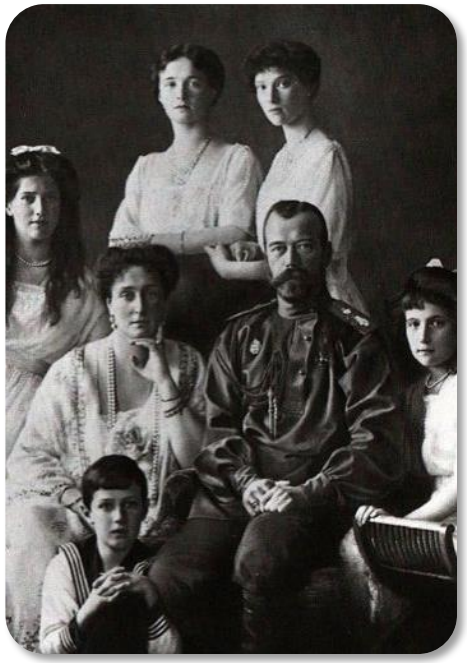


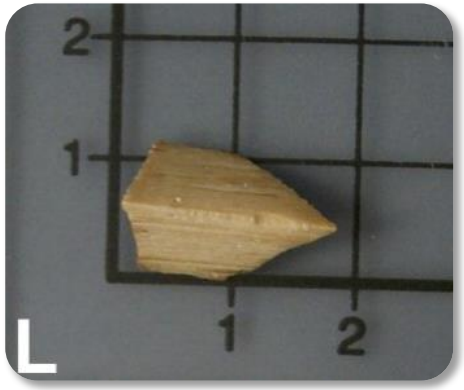
Mitochondrial DNA - Maternal Inheritance - Haplogroup Information



EMPOP

Mitochondrial DNA - Maternal Inheritance





Leopold III (+1136)
Bauer et al. 2013 ***FSIG***

King Richard III (+1485)
King et al 2014 ***Nature Com***

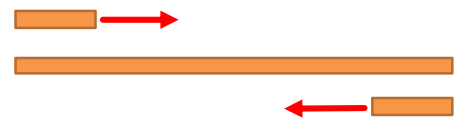
Wolfgang A. Mozart (+1791)
Parson 2006 (pers. comm)

Friedrich v. Schiller (+1805)
Parson 2008 (pers. comm.)

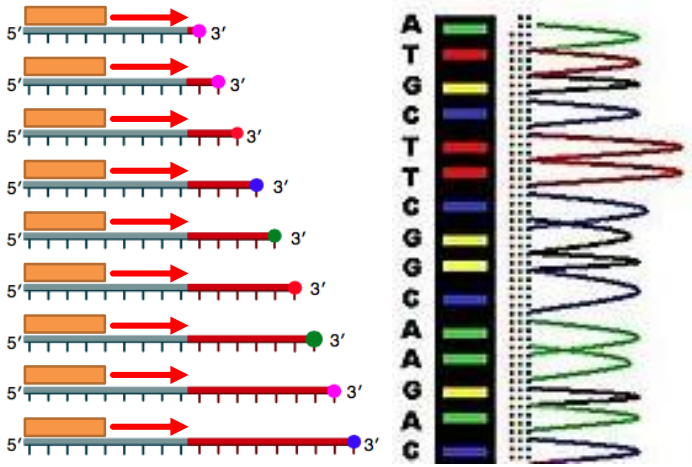
Romanov family (+1918)
Coble et al 2009 ***PLoS ONE***

PCR-based Sanger-type Dye Terminator Sequencing

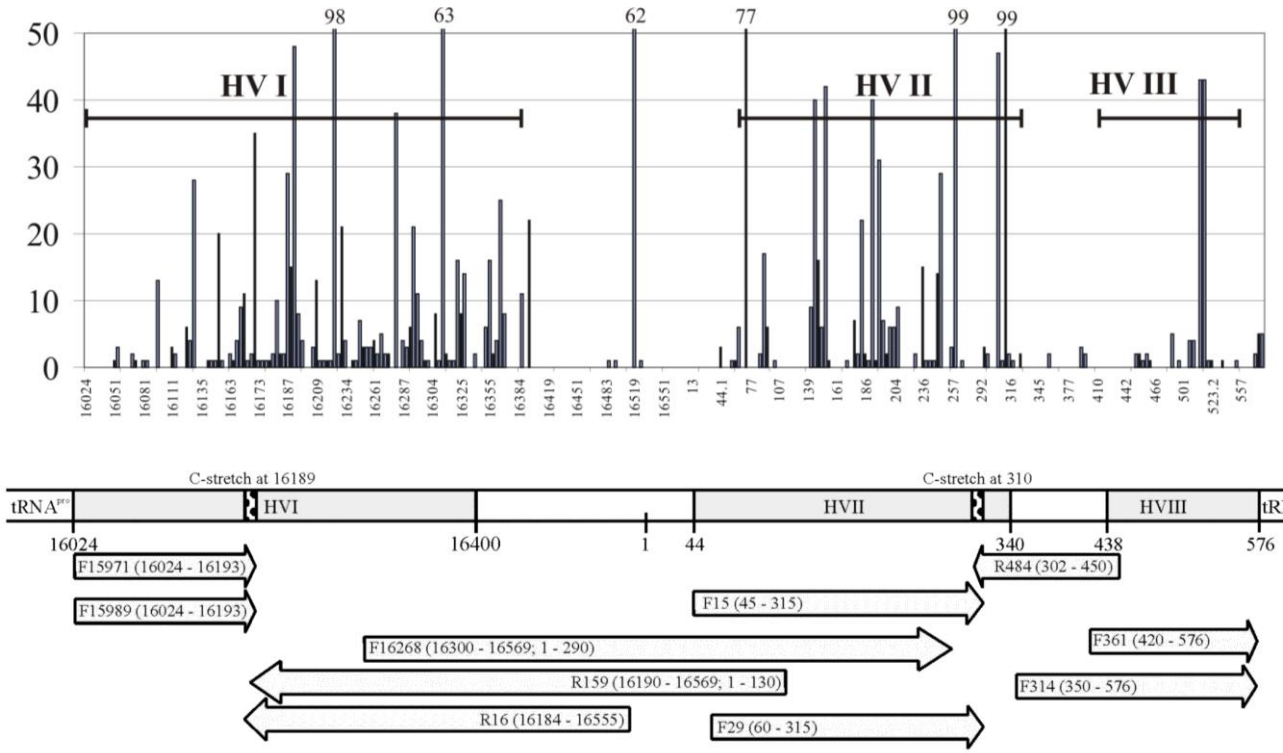
PCR



Dye Terminator Sequencing



CE



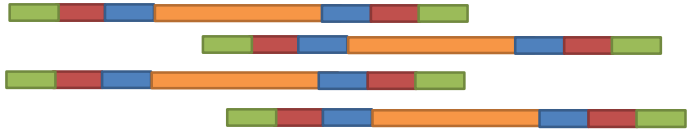
Brandstätter et al (2004) ***IJLM***

MtDNA Analysis by Next Generation (Massively Parallel) Sequencing

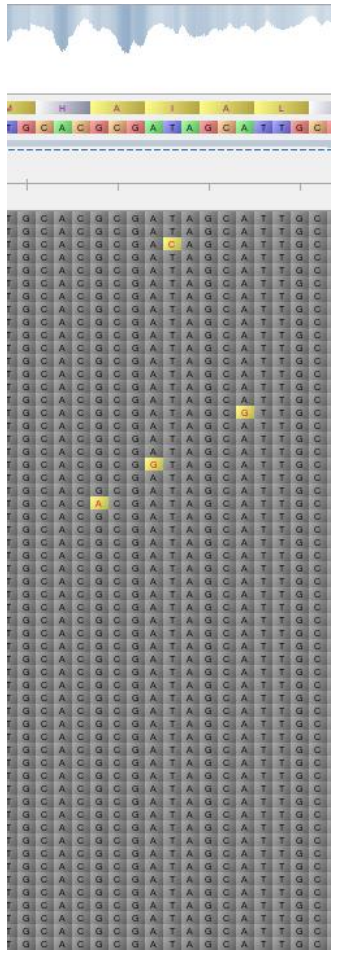
PCR

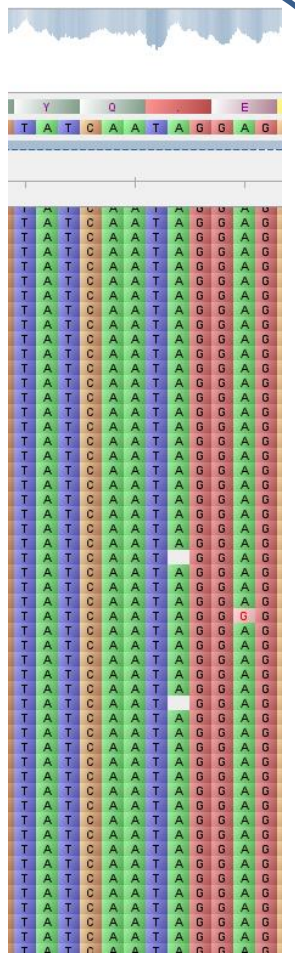
Enrich

Library



MPS





Sequence

MtDNA MPS: New Avenues

Capture **more sequence information with MPS** from a sample per assay/run than CE-based methods

- increased sequence depth
- analyze larger regions up to full mitogenomes
- higher resolution of mixtures (heteroplasmy)

Direct determination of the sequence

- actual counts of reads (not peak heights)

Amenable to **alternative library generation** methods

- e.g. Capture hybridisation, Primer Capture Extension, Shotgun sequencing

Early Research on MtDNA MPS

Forensic Science International: Genetics 7 (2013) 543–549

Contents lists available at SciVerse ScienceDirect

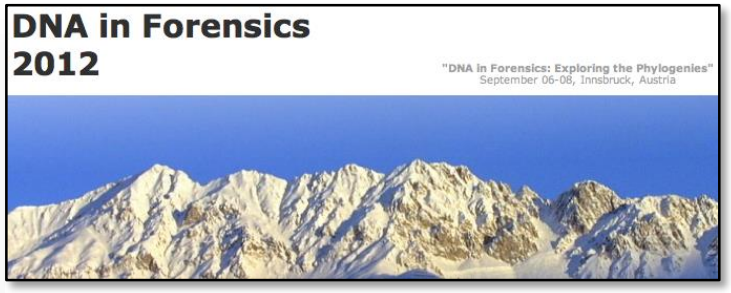
Forensic Science International: Genetics

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

Forensic Population Genetics—Original Research

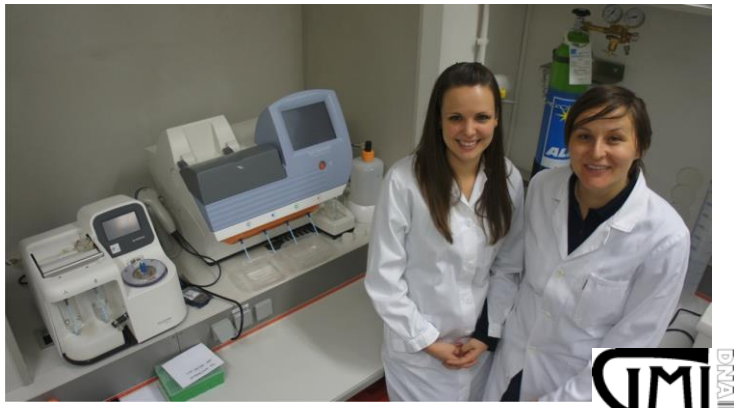
Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM)☆

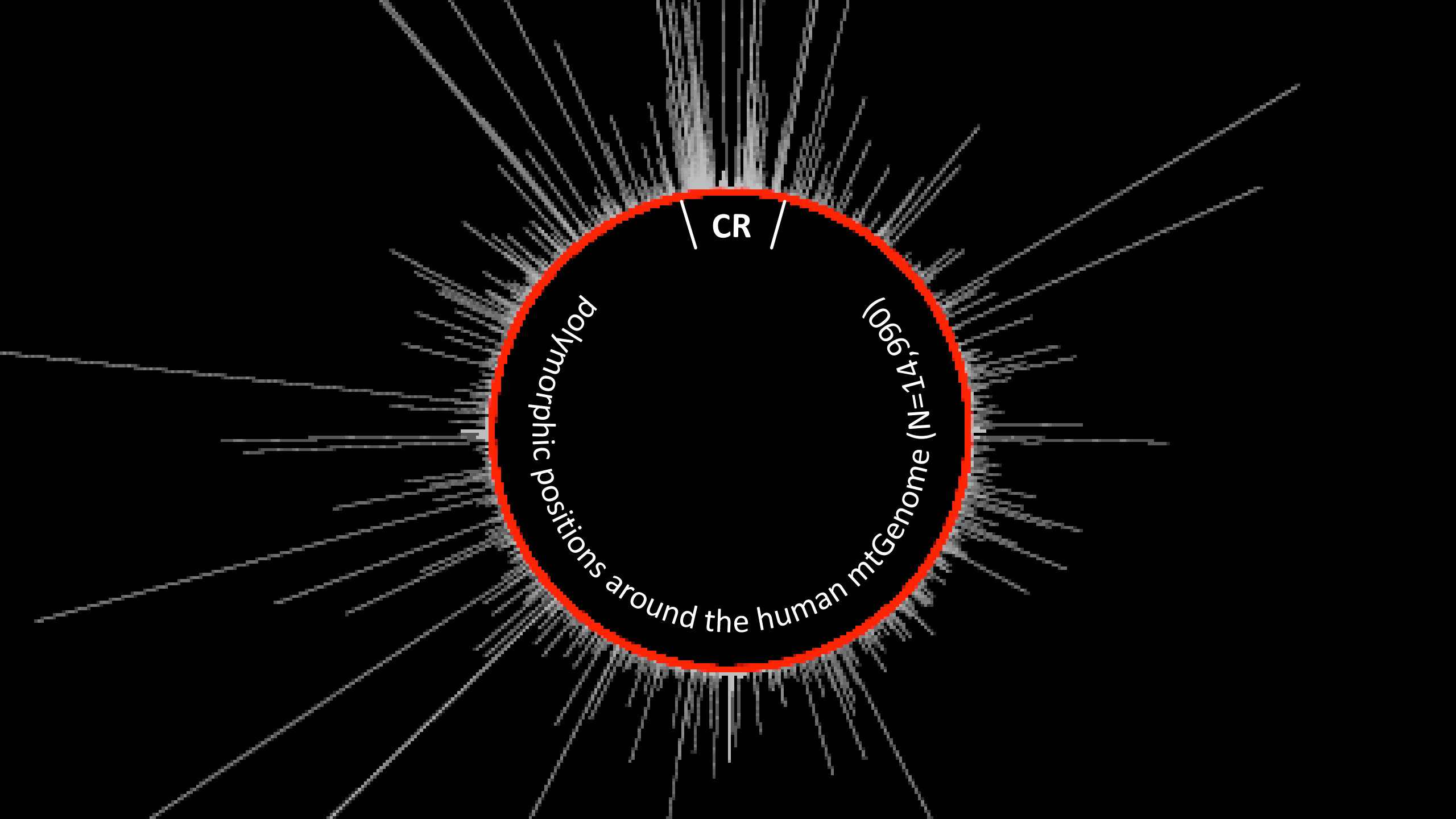
Walther Parson [a,b,*], Christina Strobl [a], Gabriela Huber [a], Bettina Zimmermann [a], Sibylle M. Gomes [c], Luis Souto [c], Liane Fendt [a,d], Rhena Delport [e], Reina Langit [f], Sharon Wootton [f], Robert Lagacé [f], Jodi Irwin [g]



42 mtGenomes STS & PGM
high concordance
except C-stretches
Software improvements

Origin	#	Source	Reference
Sub-Saharan (Angola)	5	blood	Fendt et al 2012
Southeast Asian (East Timor)	8	buccal	Parson et al 2013
Westeurasian (Austria)	6	paraffin-embedded tissue	Fendt et al 2011
Westeurasian (Austria)	23	buccal	Parson et al 2013




CR
polymorphic positions around the human mtGenome (N=14,990)

Evaluating variation in the mtDNA coding region

mitogenomes significantly increase PD in random samples

Three major U.S. populations (n=588)

African American (n=170)	HV1	HV1/HV2	CR	mtG
# Haplotypes	124	140	148	169
# Unique Haplotypes	106	120	130	168
Power of Discrimination	99,20%	99,67%	99,81%	99,99%
U.S. Caucasian (n=263)	HV1	HV1/HV2	CR	mtG
# Haplotypes	151	200	229	259
# Unique Haplotypes	122	170	211	255
Power of Discrimination	97,62%	99,42%	99,78%	99,99%
U.S. Hispanic (n=155)	HV1	HV1/HV2	CR	mtG
# Haplotypes	119	134	141	147
# Unique Haplotypes	102	121	130	140
Power of Discrimination	99,37%	99,74%	99,86%	99,92%

Just et al ***FSIG*** (2014)

Three major U.S. populations (n=283)

	HV1/HV2			mtG		
	AFA	CAU	HIS	AFA	CAU	HIS
# Individuals	87	83	113	87	83	113
# Unique haplotypes	76	77	96	85	83	111

Populations	*n*	HVI/HVII RMP	HVI/HVII GD	mtGenome RMP	mtGenome GD
AFA	*87*	2.42%	98.72%	1.31%	99.84%
CAU	*83*	3.12%	98.06%	1.20%	100.00%
HIS	*113*	3.33%	97.53%	0.98%	99.91%
Mean ±SD		2.96 ±0.48%	98.10 ±0.59%	1.16[c] ±0.17%	99.91[d] ±0.08%

King et al ***FSIG*** (2014)

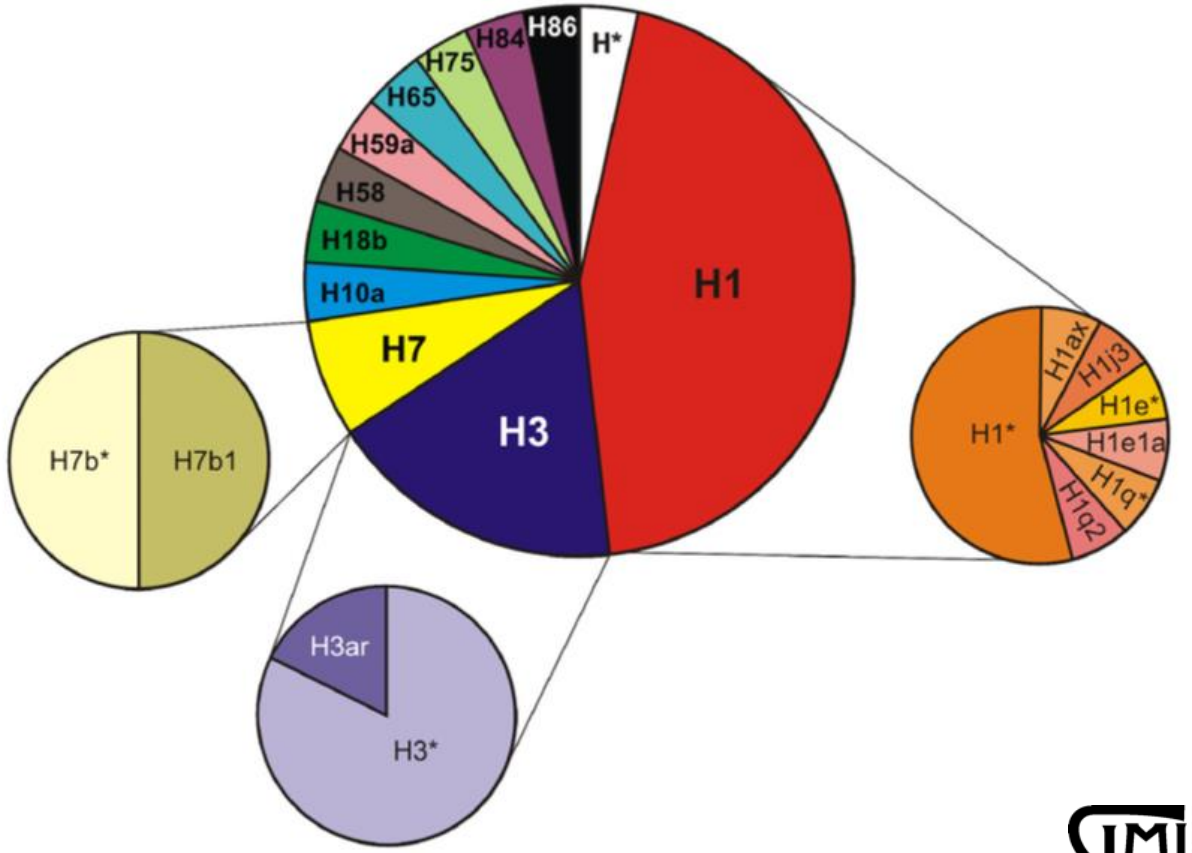
Evaluating variation in the mtDNA coding region

mitogenomes significantly increase PD in common types

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%

[a] Thereof 17 specific for haplogroup H clades [44].
[b] According to Ref. [15], build 16. H* is considered a haplogroup.
[c] Random match probability.



Bodner et al ***FSIG*** (2015)

Full mitogenomes from hair shafts

Forensic Science International: Genetics 15 (2015) 8–15

Contents lists available at ScienceDirect

Forensic Science International: Genetics

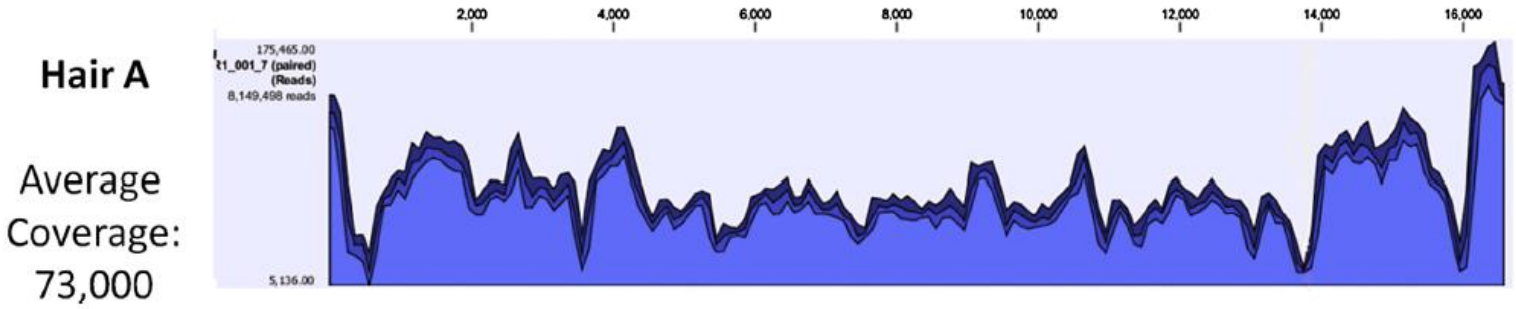
journal homepage: www.elsevier.com/locate/fsig

Original Research Paper

Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples

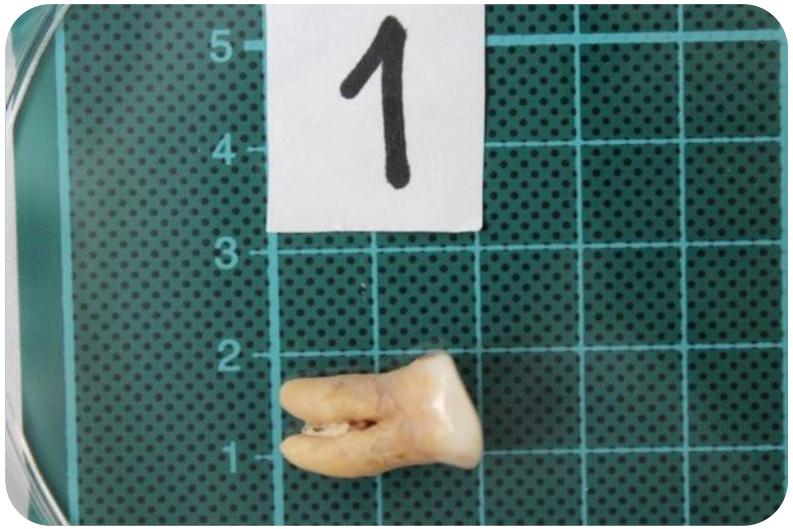
Walther Parson [a,b,*], Gabriela Huber [a], Lilliana Moreno [c], Maria-Bernadette Madel [a], Michael D. Brandhagen [c], Simone Nagl [a], Catarina Xavier [a], Mayra Eduardoff [a], Thomas C. Callaghan [c], Jodi A. Irwin [c,*]

300-500 bp ***midi***-amplicon assay



Full mitogenomes from aDNA

~2000 year old Colombian tooth sample
Muisca burial place around Sun Temple (Sogamoso, Boyaca, Colombia)



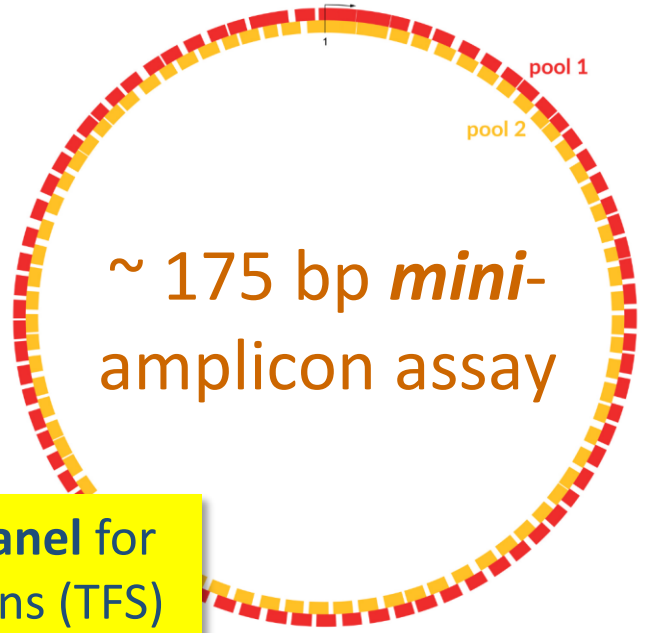
DNA extraction: full demin protocol (Bauer et al 2013 ***FSIG***)

rt-PCR mtDNA quant: 2,668 mtGE/µl (Niederstätter et al 2007 ***FSIG***)

full mitogenome – hg B2d

73G 263G 309.1C 315.1C 498del 499A 750G 827G 1438G 2706G 3547G 4122G 4123G 4769G 4820A 4977C 6473T 7028T 8281-8289del 8860G 8875C 9682C 9950C 11177T 11719A 13590A 14766T 15326G 15535T 16093C 16183C 16189C 16217C

mean cov mtG: 306
mean cov diff to rCRS: 362



Early access mtDNA Ion AmpliSeq™ panel for mitogenome sequencing - 162 amplicons (TFS)


personal genome machine
We
Mito
Tiling
Early access mtDNA Ion AmpliSeq™ panel for mtGenome sequencing
162 amplicons – about 175 bp in size

Full mitogenomes from aDNA

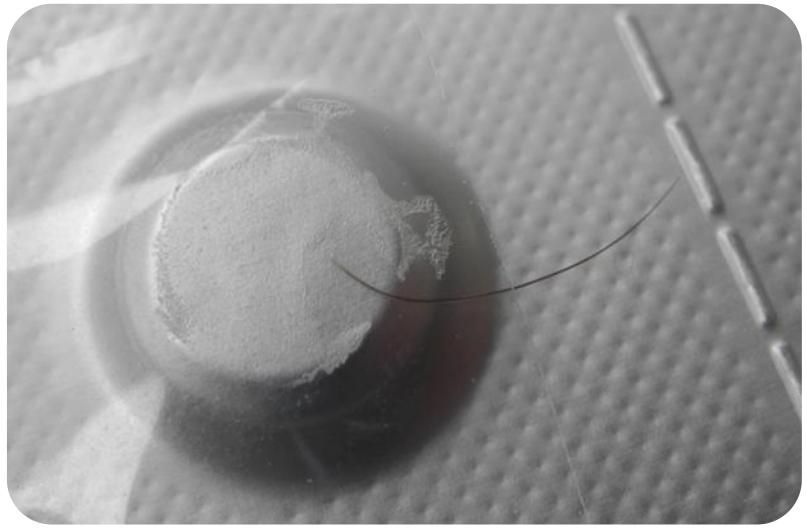
Remains from an Austrian medieval cemetery (5th/6th and 12th/13th centuries)

	tissue	mtGE/µL 143bp	mean cov mtG	mean cov diff to rCRS	full/partial
No.1	T	55,871	2,502	490	full
No.2	T	3,536	50	30	full
No.3	T	4,881	44	22	full
No.4	T	655	18	24	partial
No.5	T	3,432	26	17	full
No.6	T	60,364	38	26	full
No.7	T	4,516	47	23	full
No.8	B	729	3,494	3,326	partial
No.9	B	0	20	10	partial
No.10	T	8,847	25	14	full

(Bauer et al 2013 ***FSIG***)

MitoTiling, HID-Ion_AmpliSeq_Mito_Library_Prep_2-to-1, 200bp_Hi-Q, One Touch, 318v2PGM (TFS)

Mitogenomes from hair



Hair sample (1.2 cm): “root” and shaft

EZ-1 extraction

hair “root”: 94 mtGE/µl

hair shaft: 220 mtGE/µl

CE CR mini, “root” and shaft, HV0-ht

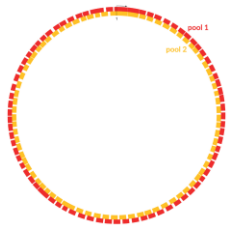
MPS mito tiling, “root” and shaft, V1a1-ht

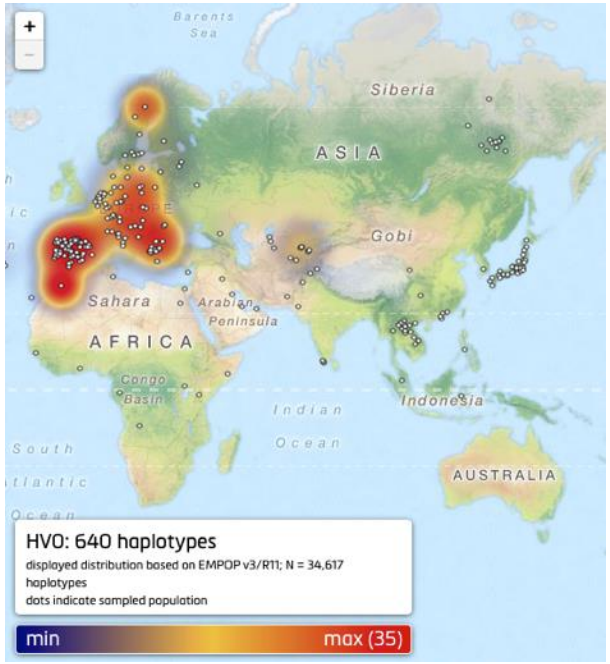
16184T
16298C
16519C
73G
263G
309.1C
315.1C
466C

CR - HV0

750G
1438G
2706G
4580A
4639C
4769G
5263T
7028T
8860G
8869G
15326G
15904T

codR - V1a1





Highly degraded DNA

SEPTEMBER 26, 2014, IGUALA, MEXICO

43 male students from the Ayotzinapa Rural Teachers College went missing

We received **17 severely burnt** samples

One resulted in CE-STR profile matching 1 family

Remaining samples gave **no detectable DNA** (mito)

The New York Times

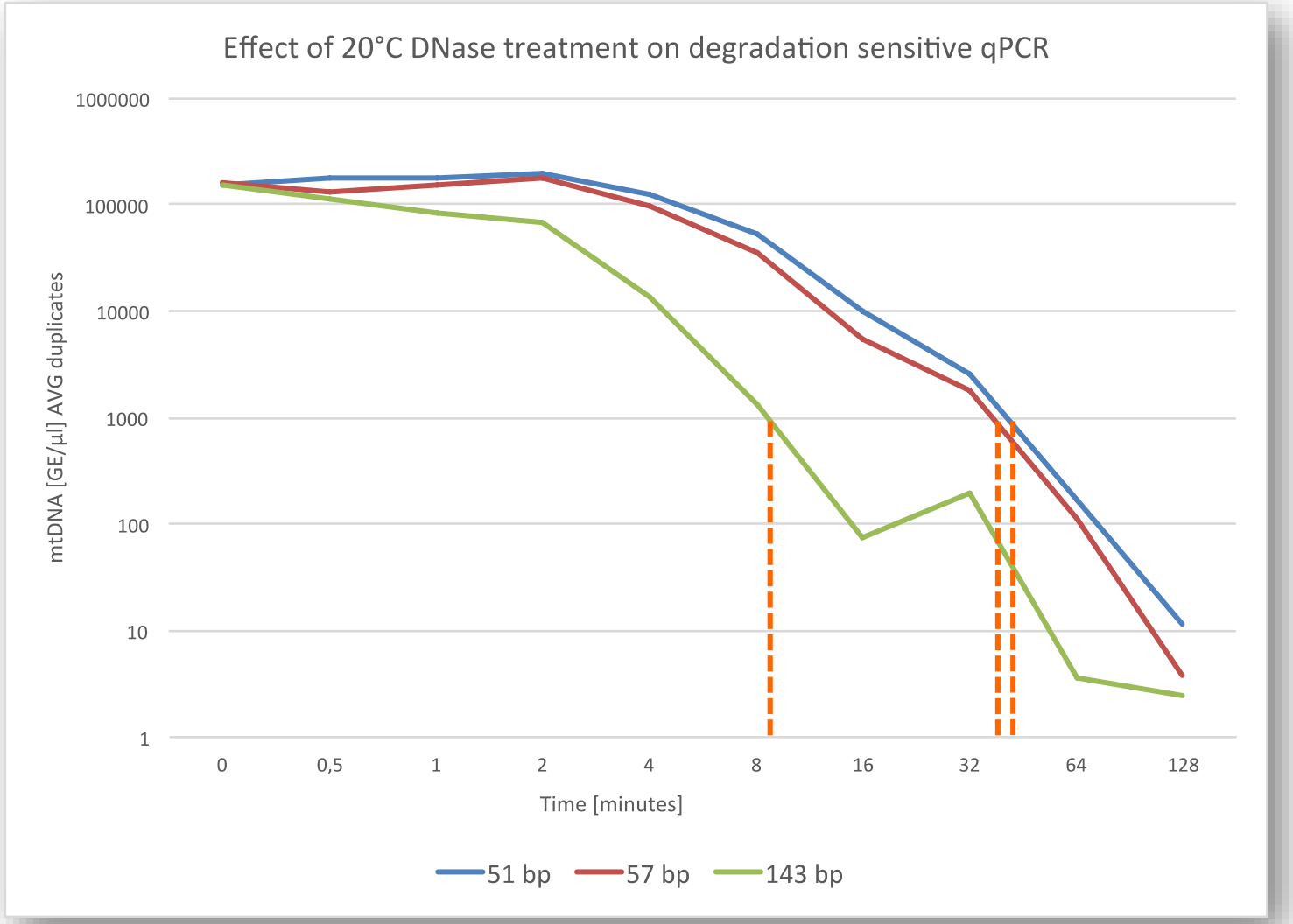
AMERICAS

Remains of Student in Mexico Identified

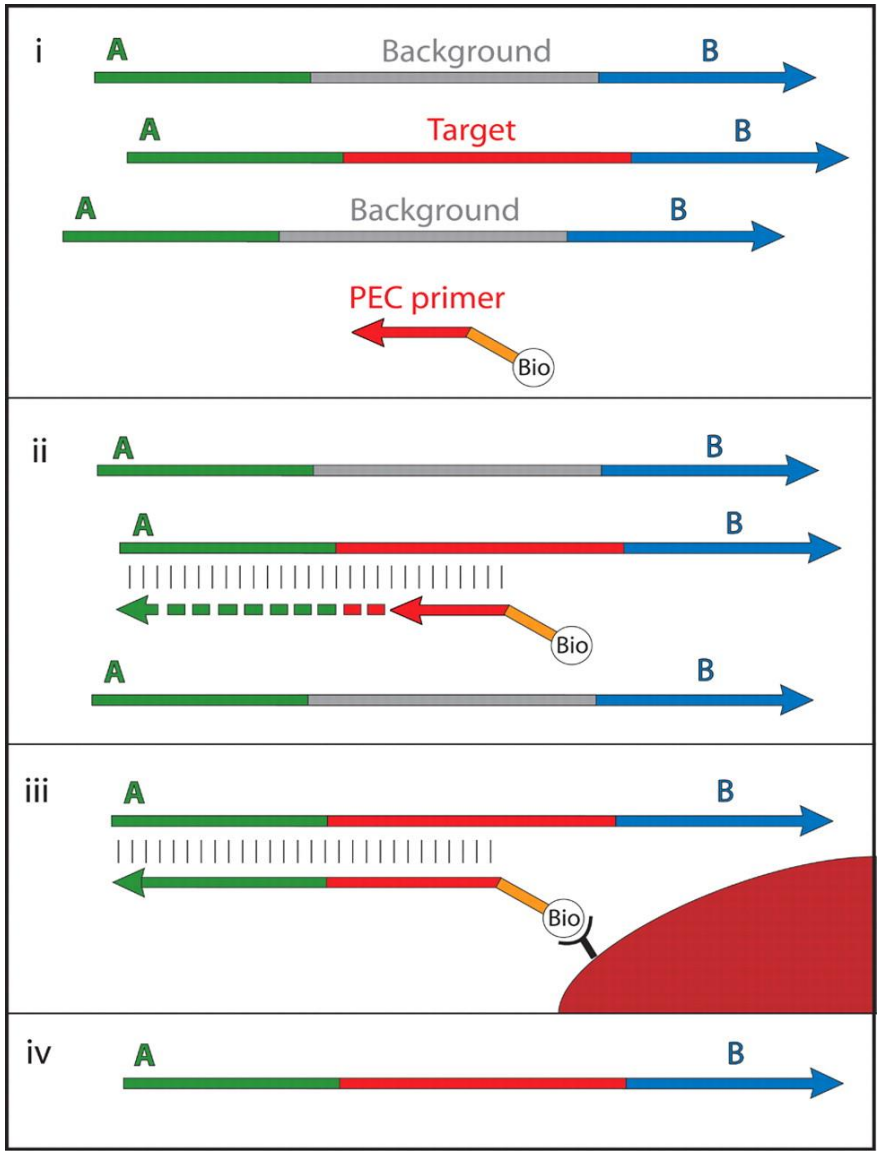
By RANDAL C. ARCHIBOLD and PAULINA VILLEGAS DEC. 6, 2014

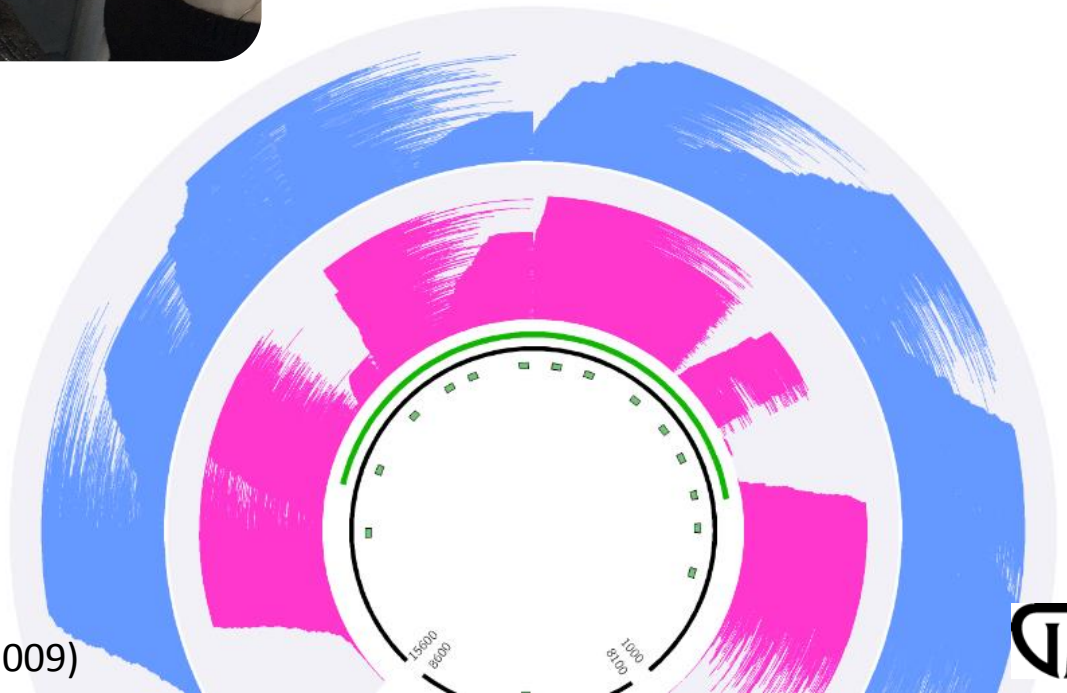
Ezequiel Mora, father of Alexander Mora, whose remains were said to have been identified, mourns at his home in El Pericon, Mexico, on Sunday. Jorge Dan Lopez/Reuters

qPCR values of DNase-degraded mtDNA



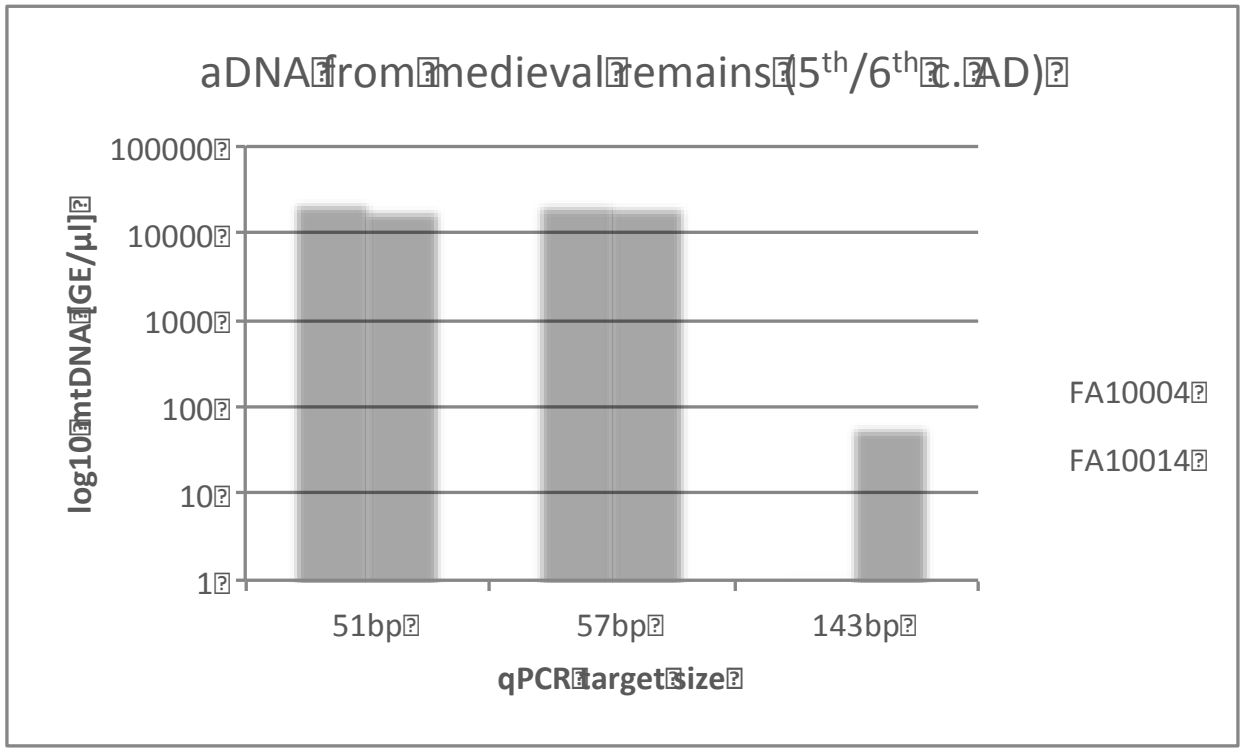
Primer Extension Capture mtDNA MPS

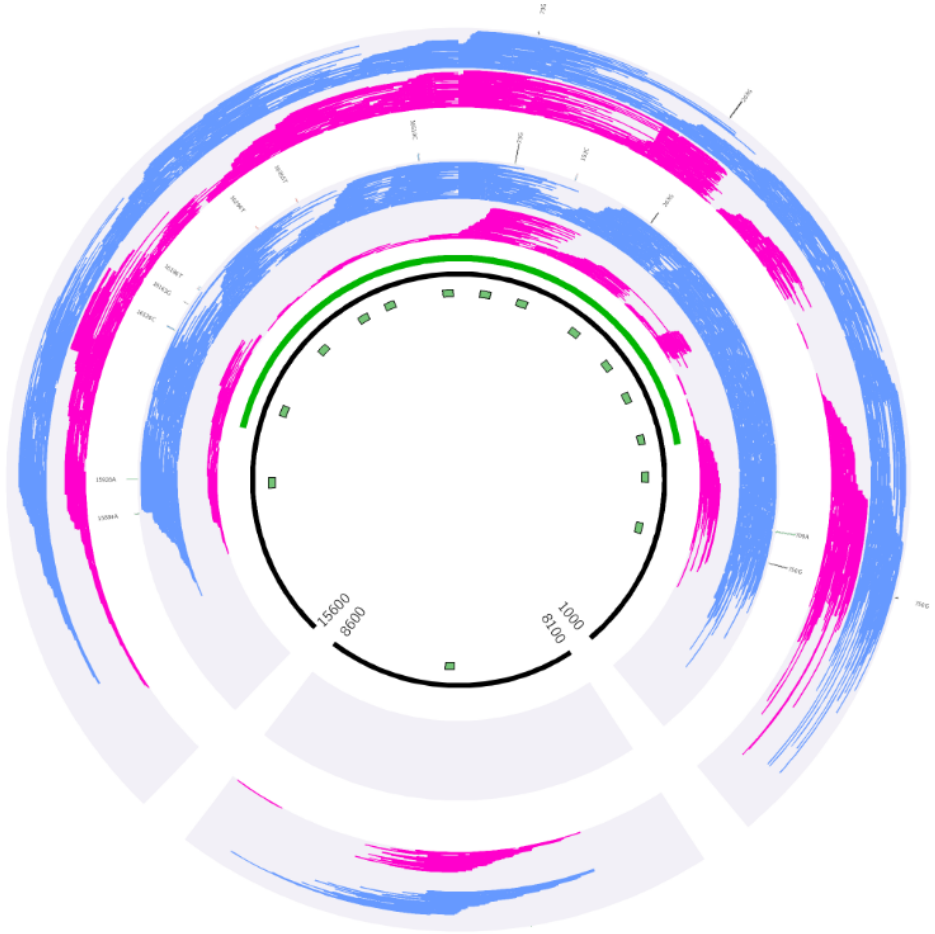




Briggs et al ***Science*** (2009)

Primer Extension Capture mtDNA MPS





FA10014

Sanger (2012) *60 mtGE/µl [143bp]*
16126C 16163G 16186T 16189C 16294T 16355T 16519C 73G 152C
16126C 16163G 16186T 16189C 16294T 16355T 16519C 73G 152C 263G 315.1C 709A 750G 15884A 15928A
PEC (2015) *20.000 mtGE/µl [57bp]*

hg T1a12

PEC applied to Mexican samples

NEWS

Home | Video | World | UK | Business | Tech | Science | Magazine | Entertainment & Arts

World | Africa | Asia | Australia | Europe | Latin America | Middle East | US & Canada

Latin America & Caribbean

Remains of second Mexican student identified

17 September 2015 | Latin America & Caribbean

The relatives of the 43 missing students have held regular peaceful protests and marches for the past year

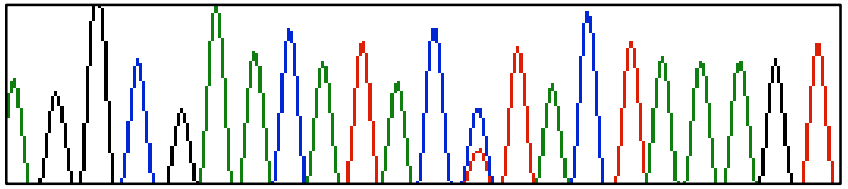
Of 17 shipped samples

one gave CE-STRs matching one set of family references (LR>)

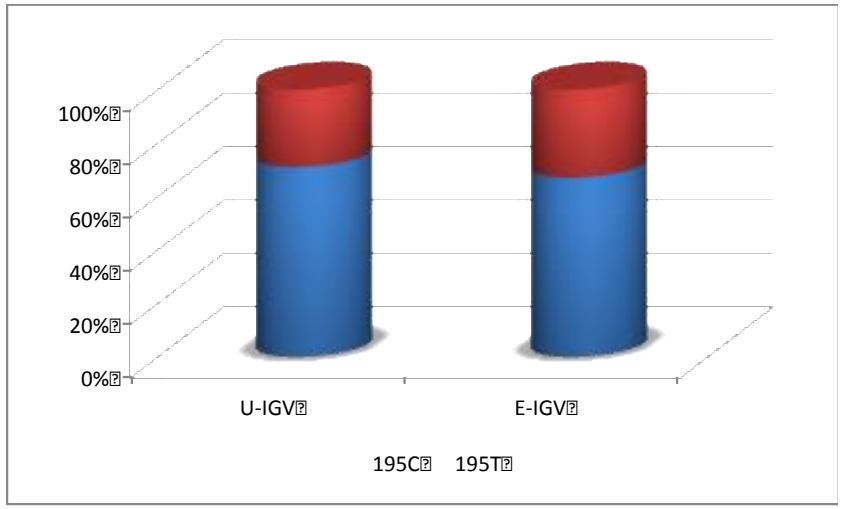
Nine brought sequences using PEC MPS, of which **two** were human specific. One matched the **earlier identified student**, the other matched **a new set** of family references

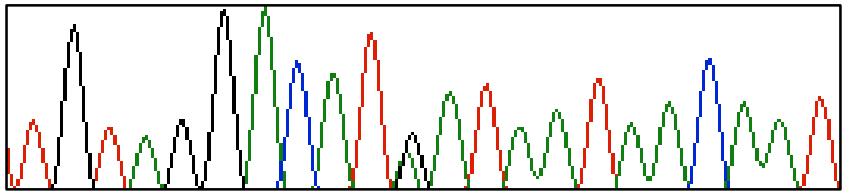
Seven remaining samples brought non-human results

Mixtures (point heteroplasmy) with MPS

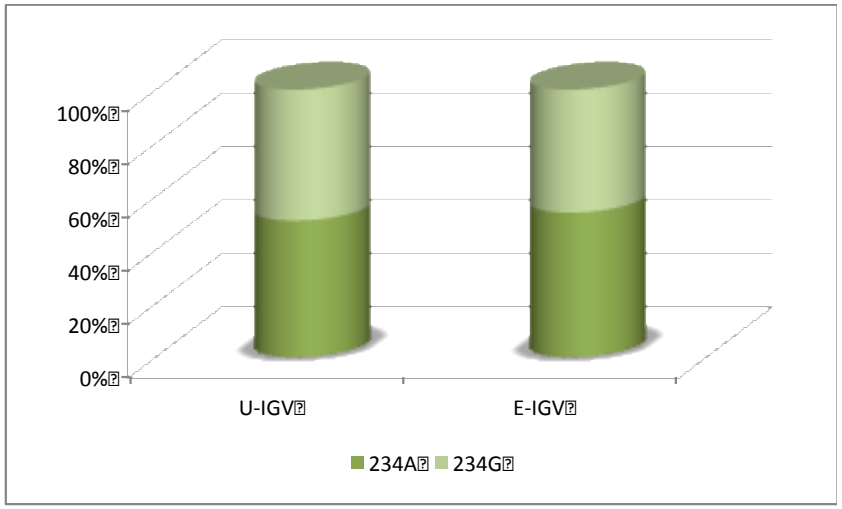


PGM	195C	195T	CV
U-IGV	0.74	0.26	2129
E-IGV	0.67	0.33	4953
MiSeq	**195C**	**195T**	**CV**
U-IGV	0.71	0.29	4150





PGM	234A	234G	CV
U-IGV	0.53	0.47	1417
E-IGV	0.54	0.46	3841
MiSeq	**234A**	**234G**	**CV**
U-IGV	0.51	0.49	2375



With Sanger we established …

… that in the Control Region, point heteroplasmy

occurs in about 6% in saliva and blood (Irwin et al 2009), **max 3/Ind**

is **consistent with evolutionary hotspot mutations** (few exceptions, e.g. 214, 215)

... that in the entire mitogenome, point heteroplasmy

occurs in about 24% in serum (Just et al 2015), **max 3/Ind**

occurs randomly (not associated with evoltuionary hotspots nor signature mutations)

Is Heteroplasmy Detection More Sensitive with MPS?

Generally yes, but interpretation is dependent on coverage, strand bias, background, and other factors. Great care is indicated with superficial literature data!

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Next-Generation Sequencing of Human Mitochondrial Reference Genomes Uncovers High Heteroplasmy Frequency

Bullet points:

HapMap samples: 20 European ancestry (CEU), 20 African ancestry (YRI)

454 GS FLX pyrosequencing platform

Sequencing error rate < 5.63×10^{-4}

Higher rate of heteroplasmy (10-50%)

"NGS technologies allow interrogation of the mitochondrial genome in greater depth than previously possible which may be of value in biology and medicine"

NA12891	14872	H13
NA12891	2259	H13a
NA12891	4745	H13a1
NA12891	7337	H13a1a1
NA12891	6755	H13a1a1a
NA12891	2706	outside H
NA12891	7028	outside H
NA12891	152	recurrent
NA12891	6266	recurrent

9 "heteroplasmies" - contamination!!!

apparent mixture of H13 and a sample outside hg H

numerous other examples…..

Is Heteroplasmy Detection More Sensitive with MPS?

Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals

Bullet points:

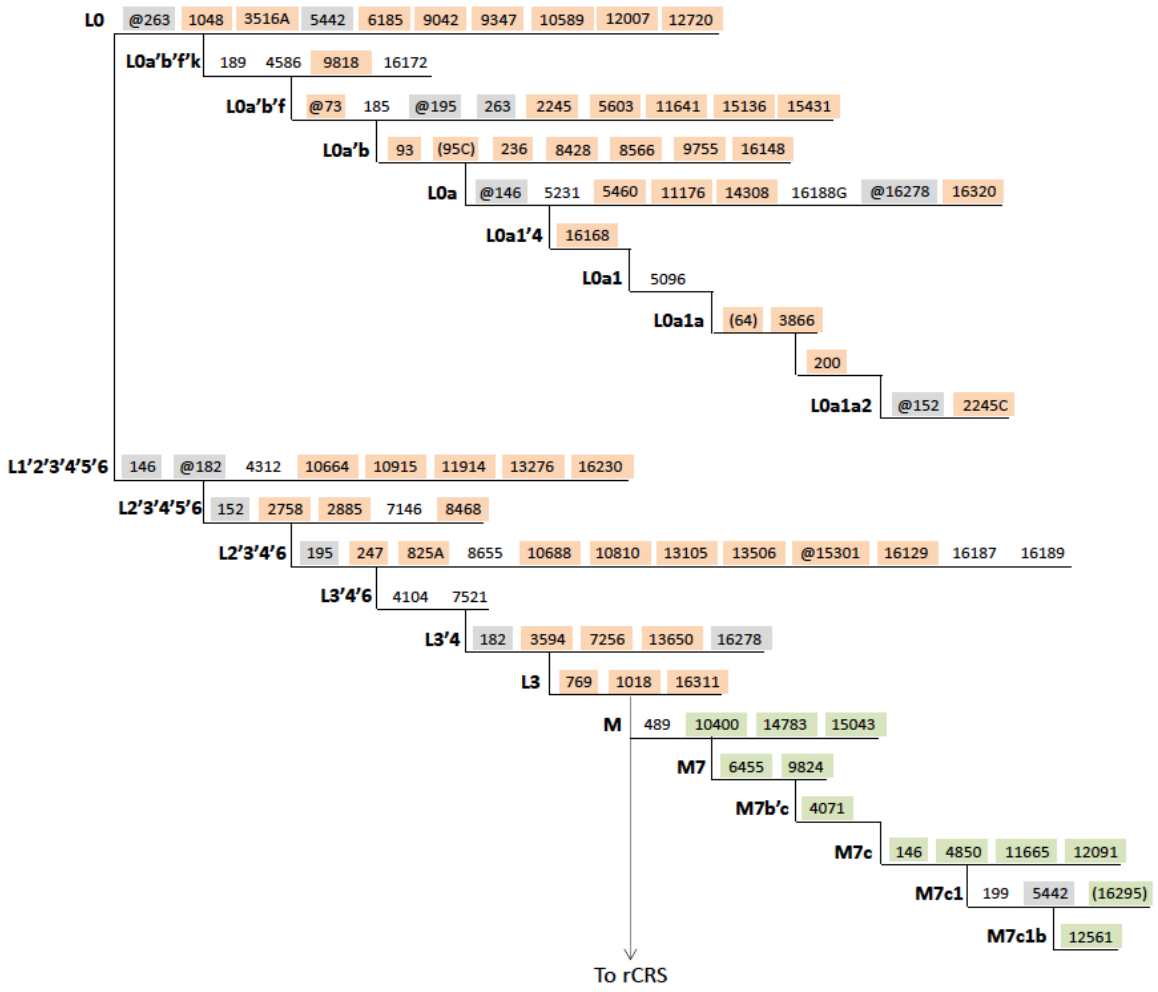
1000 Genomes Project

Mean coverage of ~2,000x

Use a combination of stringent thresholds and a maximum-likelihood method to define heteroplasmy

~90% of the individuals carry at least one heteroplasmy (1% minor allele frequency (MAF) threshold)

Positive correlation between substitution rates and heteroplasmy rates (not found in Sanger)

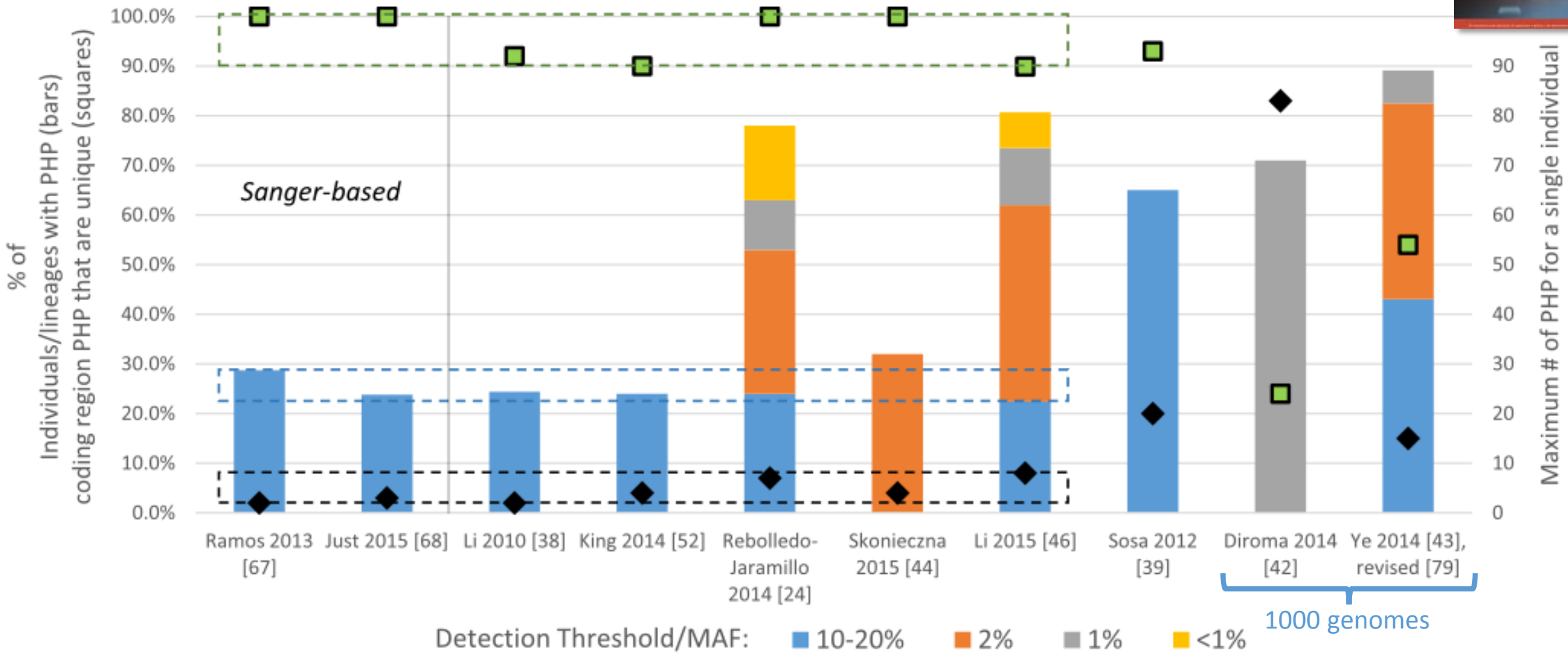


>70 "heteroplasmies"??? - contamination!!!

mixture of M7c and L0a

Heteroplasmy Detection Is More Sensitive with MPS, but ….

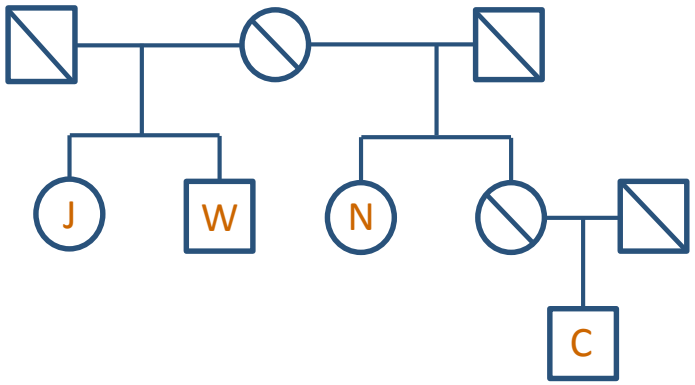
… its authenticity depends on various factors including rate of contamination, presence of numts, total sequencing coverage, and sequence background.



Just et al (2015)

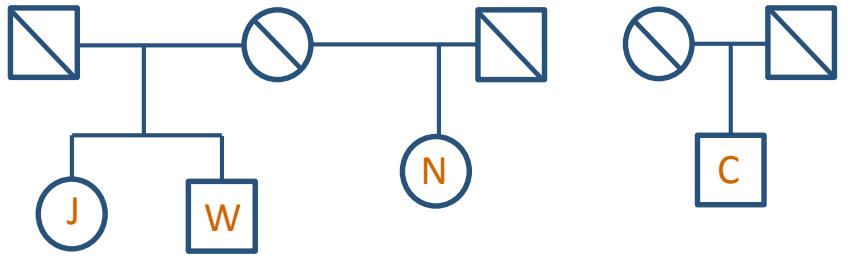
Immigration case - aSTRs & X-STRs (CE)

H1



W_1 = 99.9966 %

H2

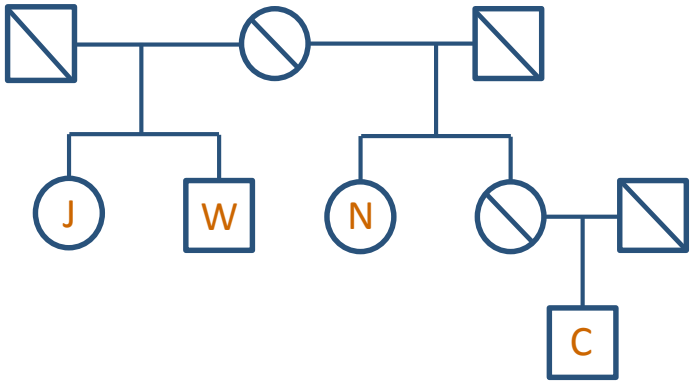


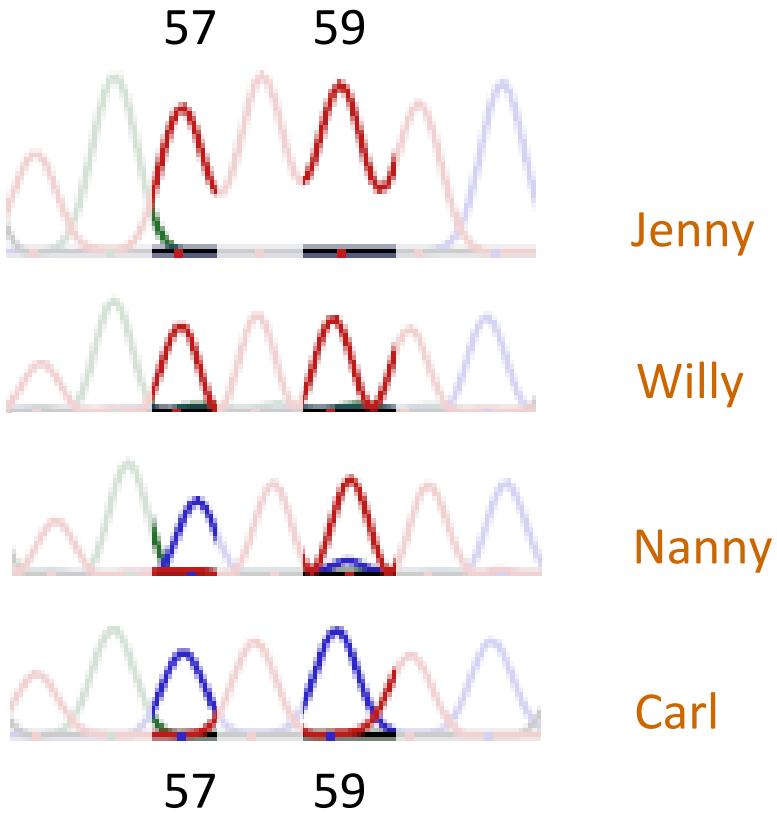
W_2 = 0.0034 %

eurofins B. Rolf, K. Koop, Eurofins Medigenomix Forensik GmbH, Munich, Germany

Immigration case - mtDNA Sanger

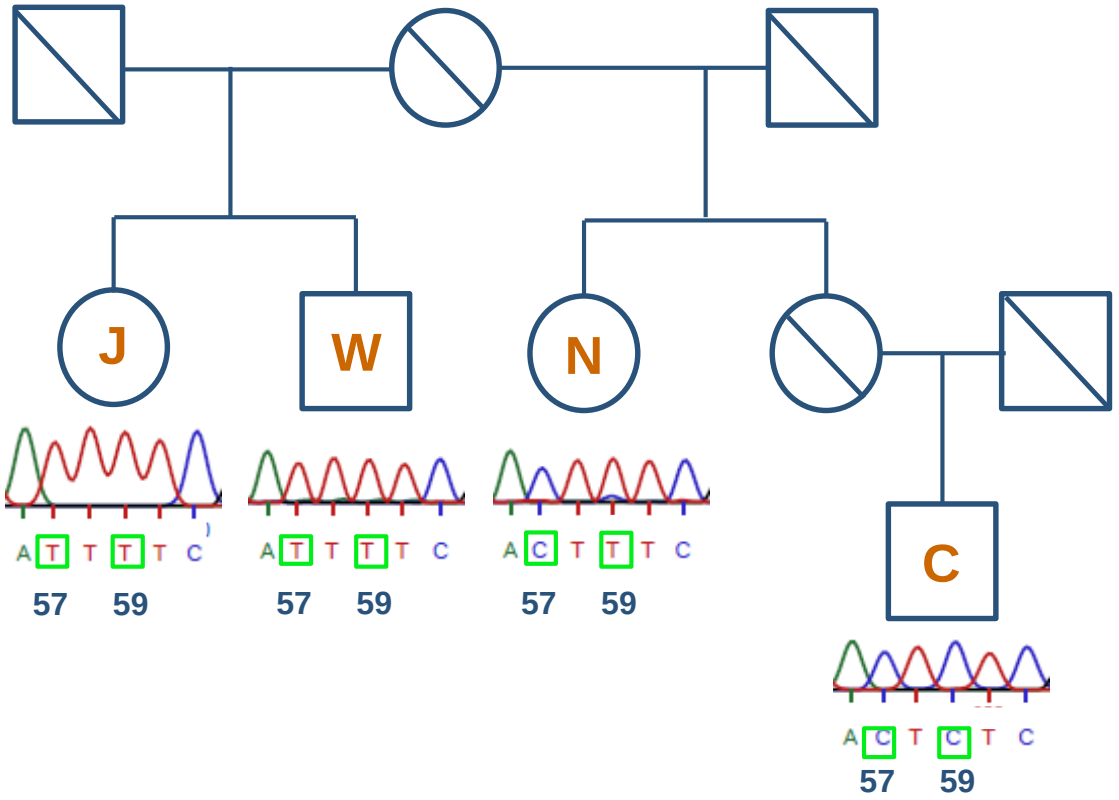
rCRS	Position	Jenny	Carl
G	16129	A	A
C	16148	T	T
A	16166	G	G
A	16183	C	C
C	16186	T	T
T	16189	C	C
C	16223	T	T
C	16278	T	T
T	16311	C	C
C	16355	T	T
T	16362	C	C
T	**57**	**T**	**C**
T	**59**	**T**	**C**
A	73	G	G
T	152	C	C
C	182	T	T
T	195	C	C
G	247	A	A
A	263	G	G
-	315.1	C	C





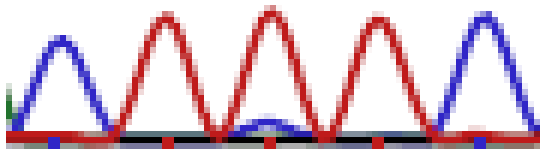
eurofins

Immigration case - apparently 2 differences between J and C



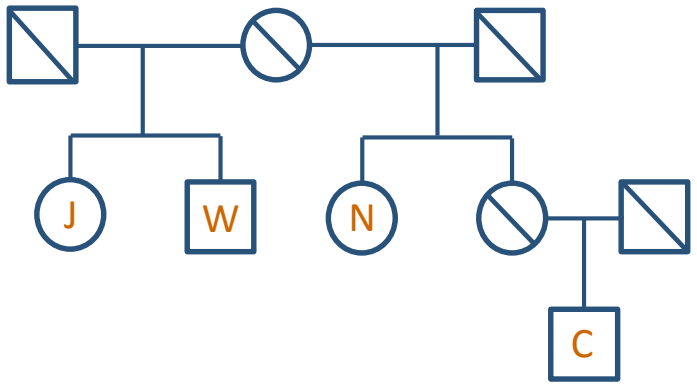
Low level point heteroplasmy?

approx. 10% C point heteroplasmy at T59 in N



57 and 59 not known as hotspots for heteroplasmy (Irwin et al 2009)

Immigration case - mtDNA MPS



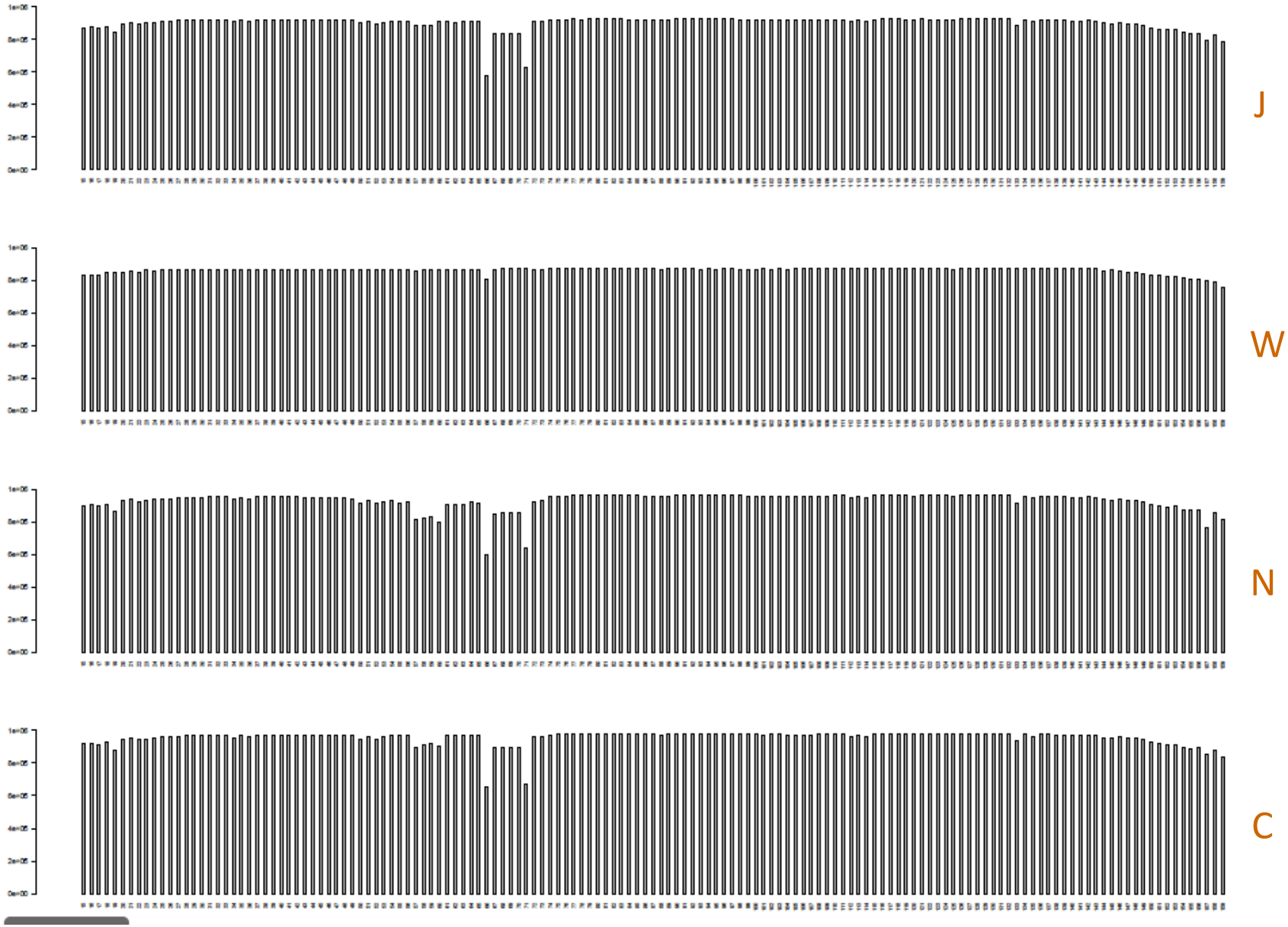
PCR amplicon 16-158 (HVS-II)

$1*10^6$ Coverage

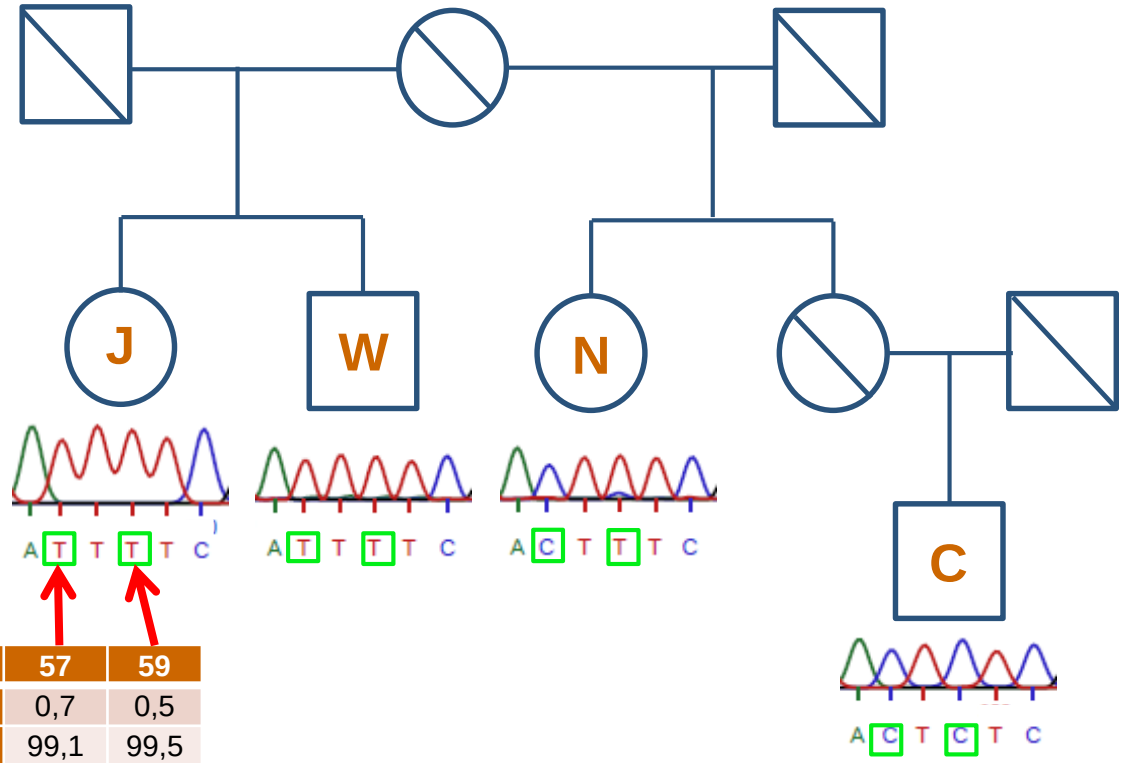
Hi-Q Enzyme

200 bp Chemistry

318 Chip, PGM

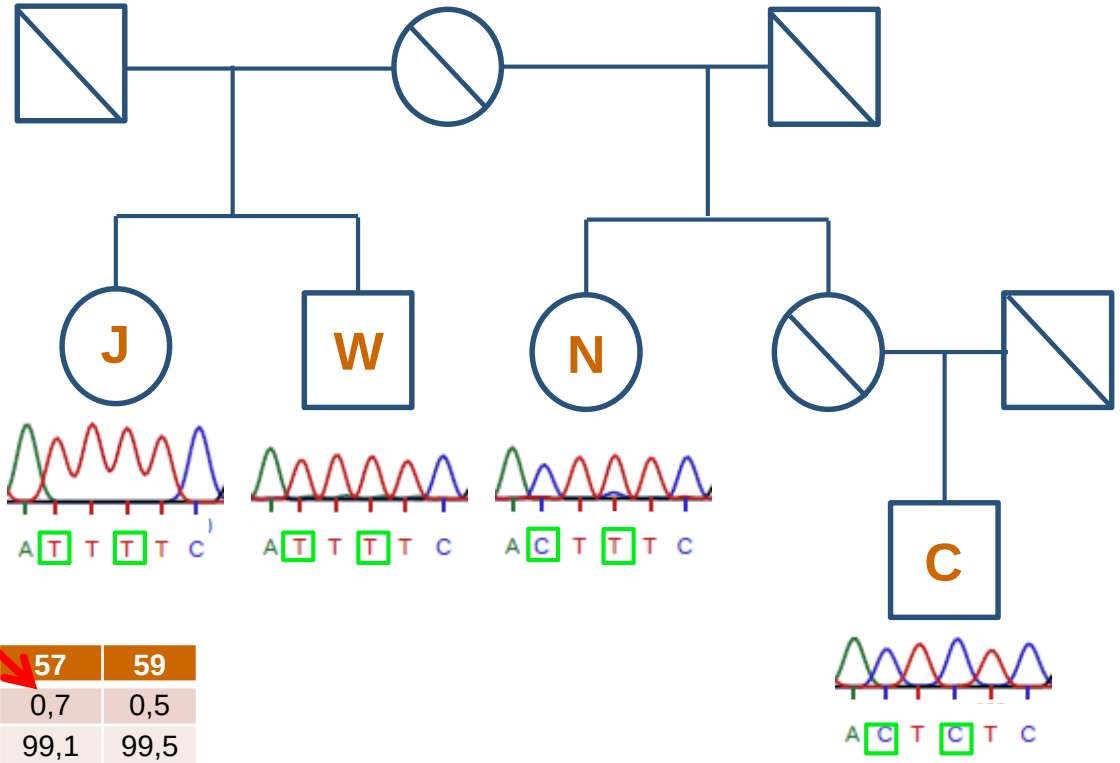


Immigration case - low level contribution detected by MPS



Position	57	59
C	0,7	0,5
T	99,1	99,5

Immigration case - low level contribution detected by MPS

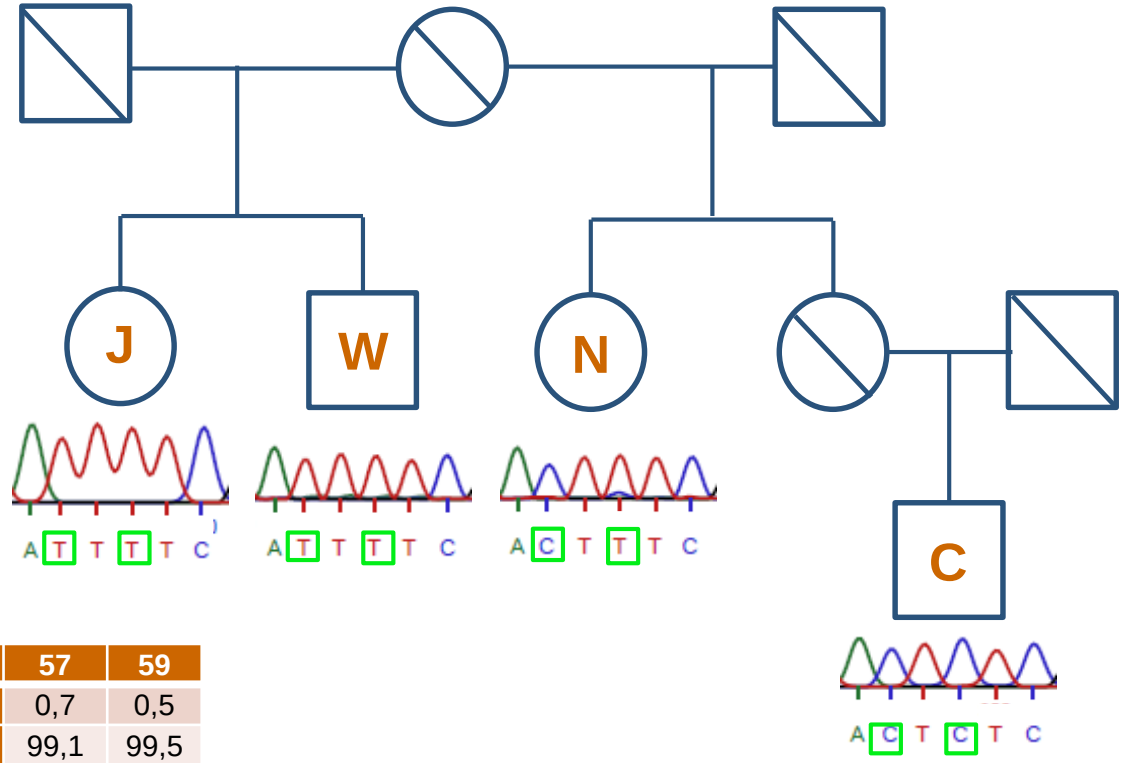


6.186 PGM-Reads

Position	57	59
C	0,7	0,5
T	99,1	99,5

841.309 PGM-Reads

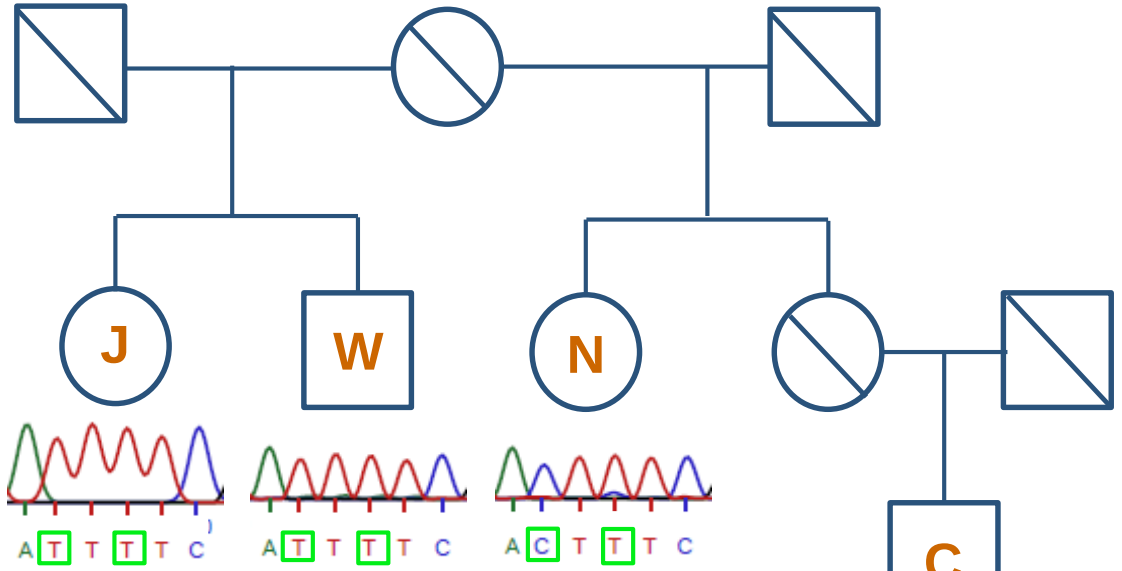
Immigration case - low level contribution detected by MPS



Position	57	59
C	0,7	0,5
T	99,1	99,5

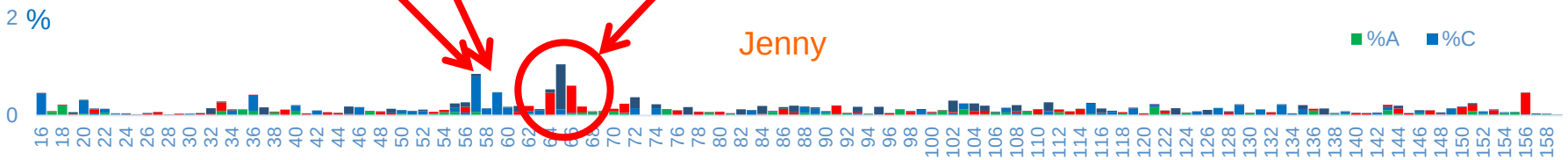
relative contribution of non-dominant bases (background + PHP)

Immigration case - low level contribution detected by MPS



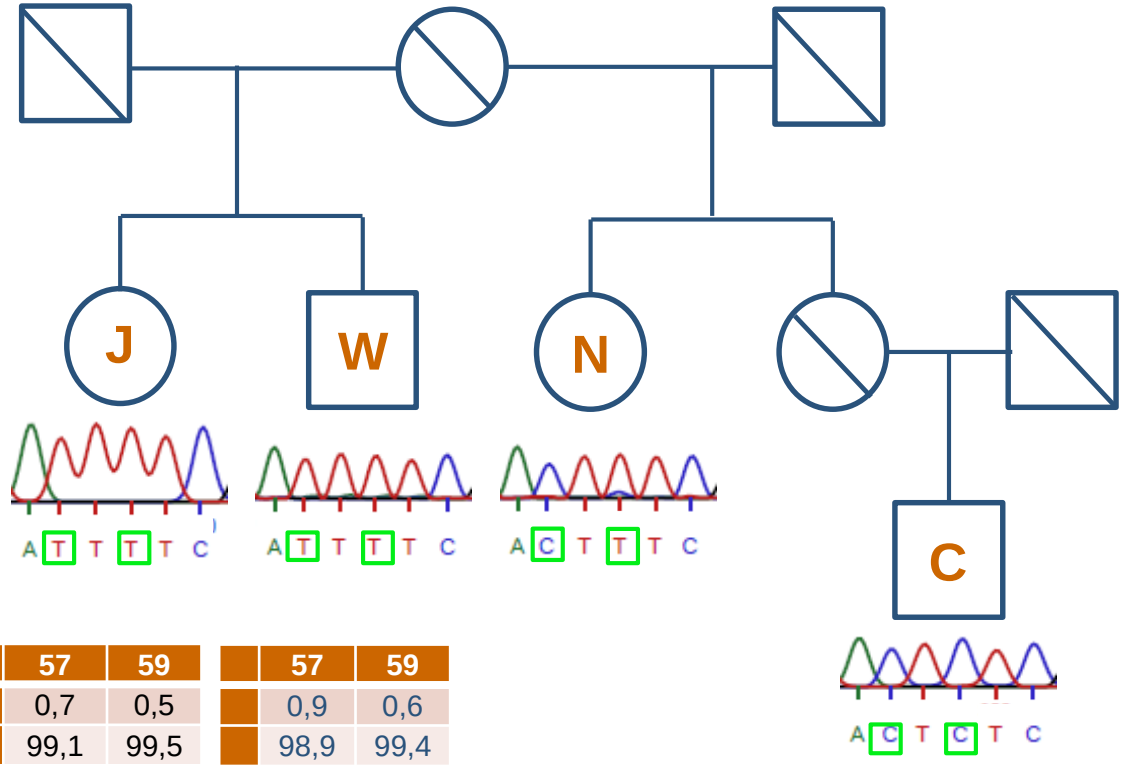
Position	57	59
C	0,7	0,5
T	99,1	99,5

Alignment artefact due to G-tract 65 TGGGGGGTA 73



relative contribution of non-dominant bases (background + PHP)

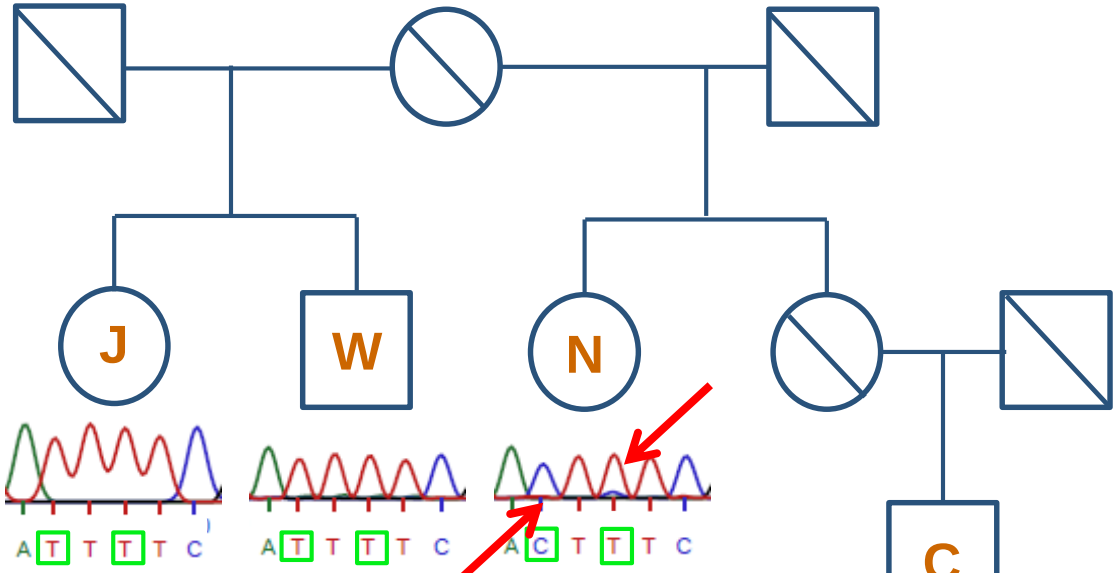
Immigration case - low level contribution detected by MPS



Position	57	59
C	0,7	0,5
T	99,1	99,5

	57	59
	0,9	0,6
	98,9	99,4

Immigration case - low level contribution detected by MPS



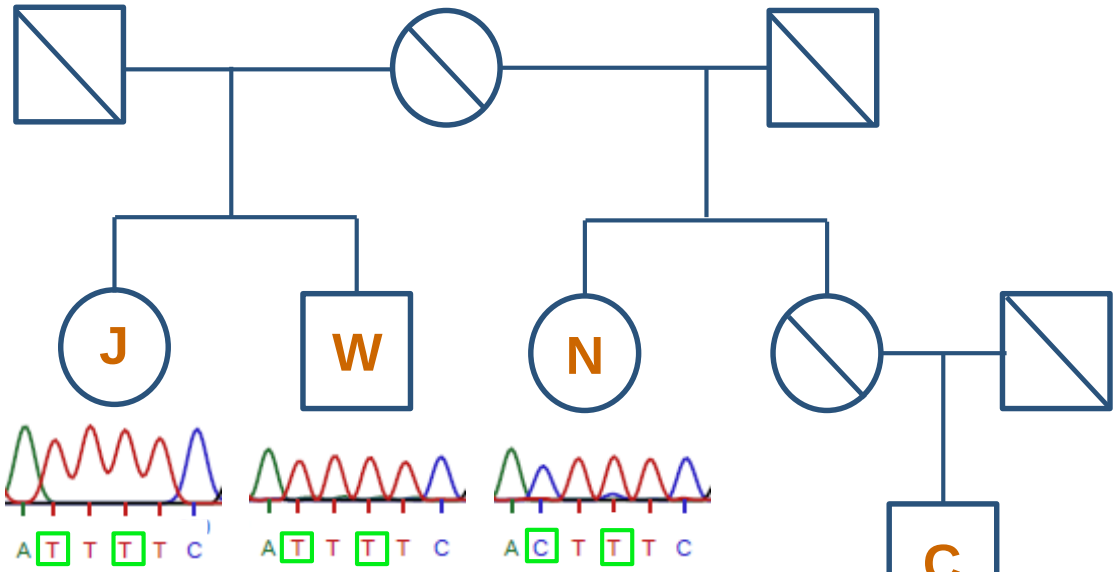
Position	57	59
C	0,7	0,5
T	99,1	99,5

	57	59
	0,9	0,6
	98,9	99,4

	57	59
	88,7	9,7
	10,6	90,2

10% 57T point heteroplasmy not detected by Sanger

Immigration case - low level contribution detected by MPS



Position	57	59
C	0,7	0,5
T	99,1	99,5

	57	59
	0,9	0,6
	98,9	99,4

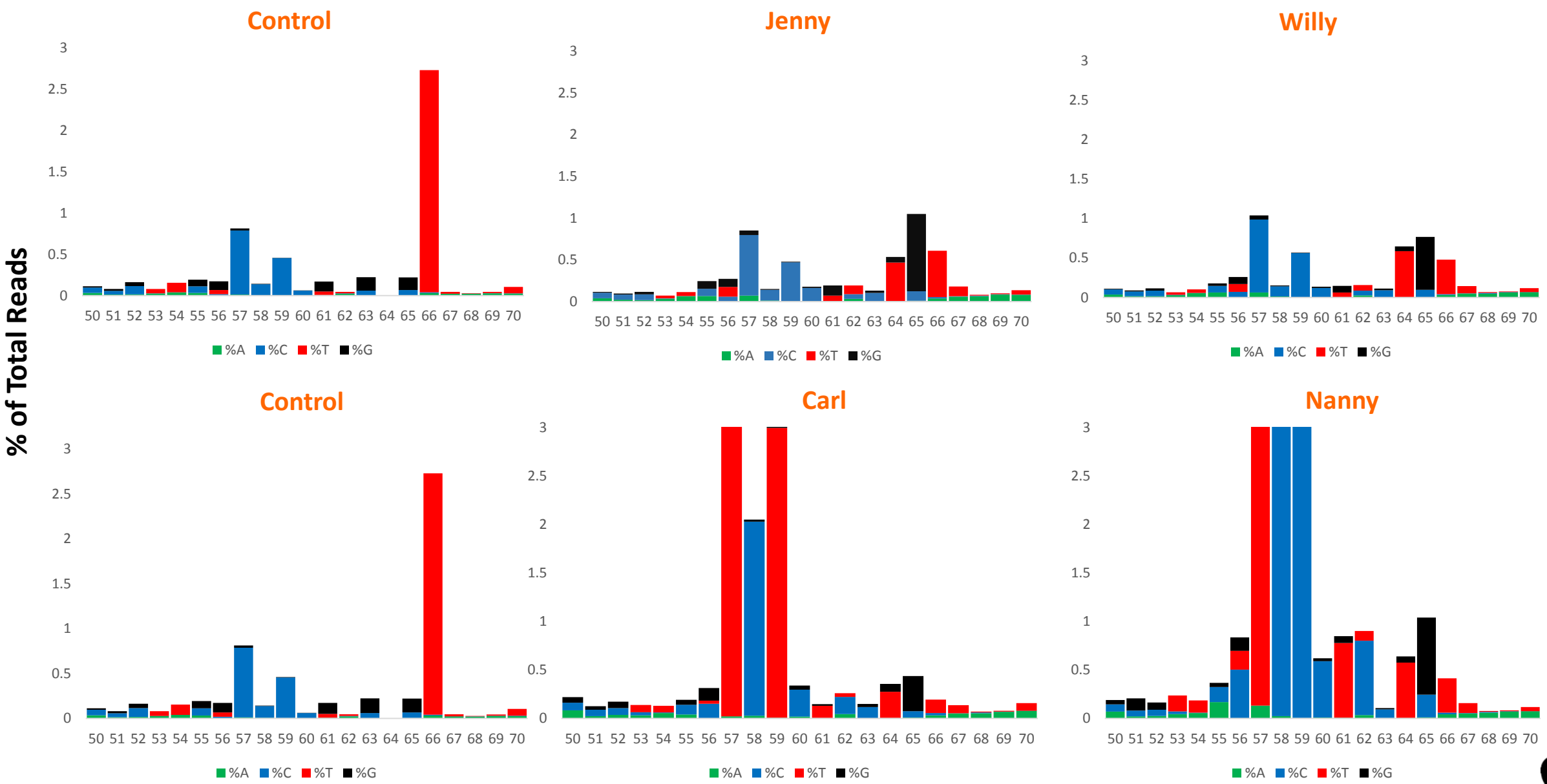
	57	59
	88,7	9,7
	10,6	90,2

	57	59
	96,2	97,0
	3,8	3,0

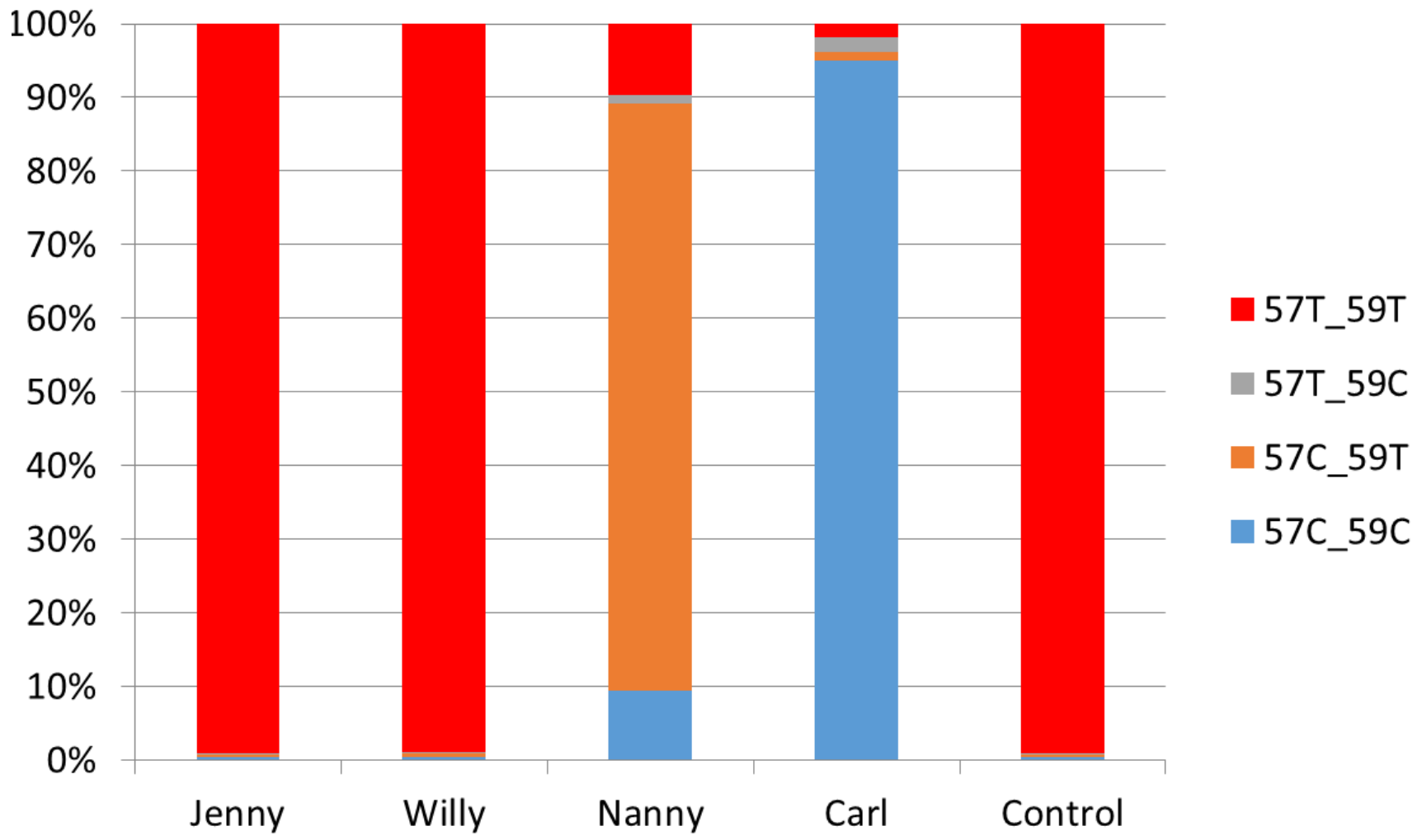
4% T57 point heteroplasmy not detected by Sanger

3% T59 point heteroplasmy not detected by Sanger

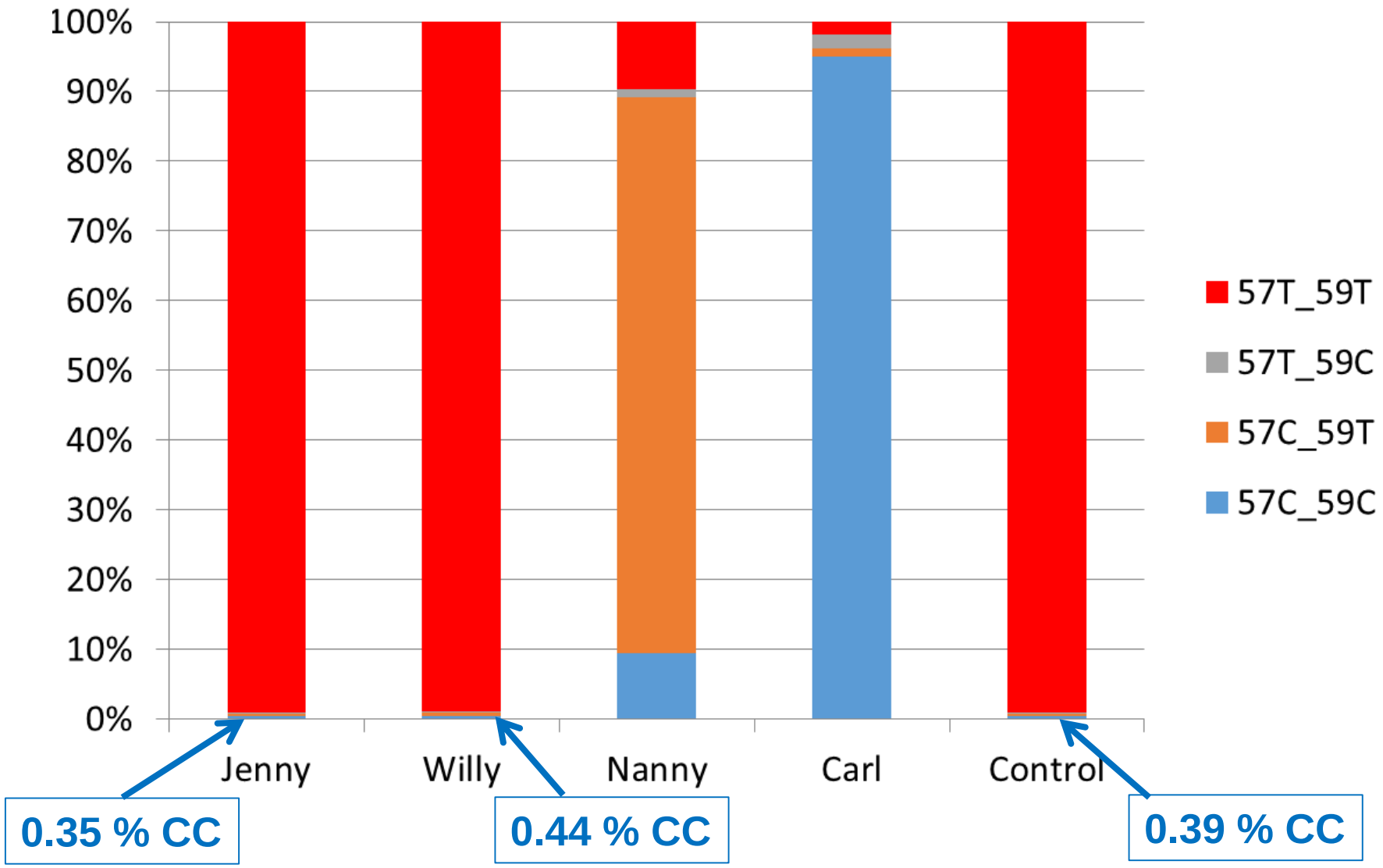
Immigration case - non-dominant signal detected by MPS



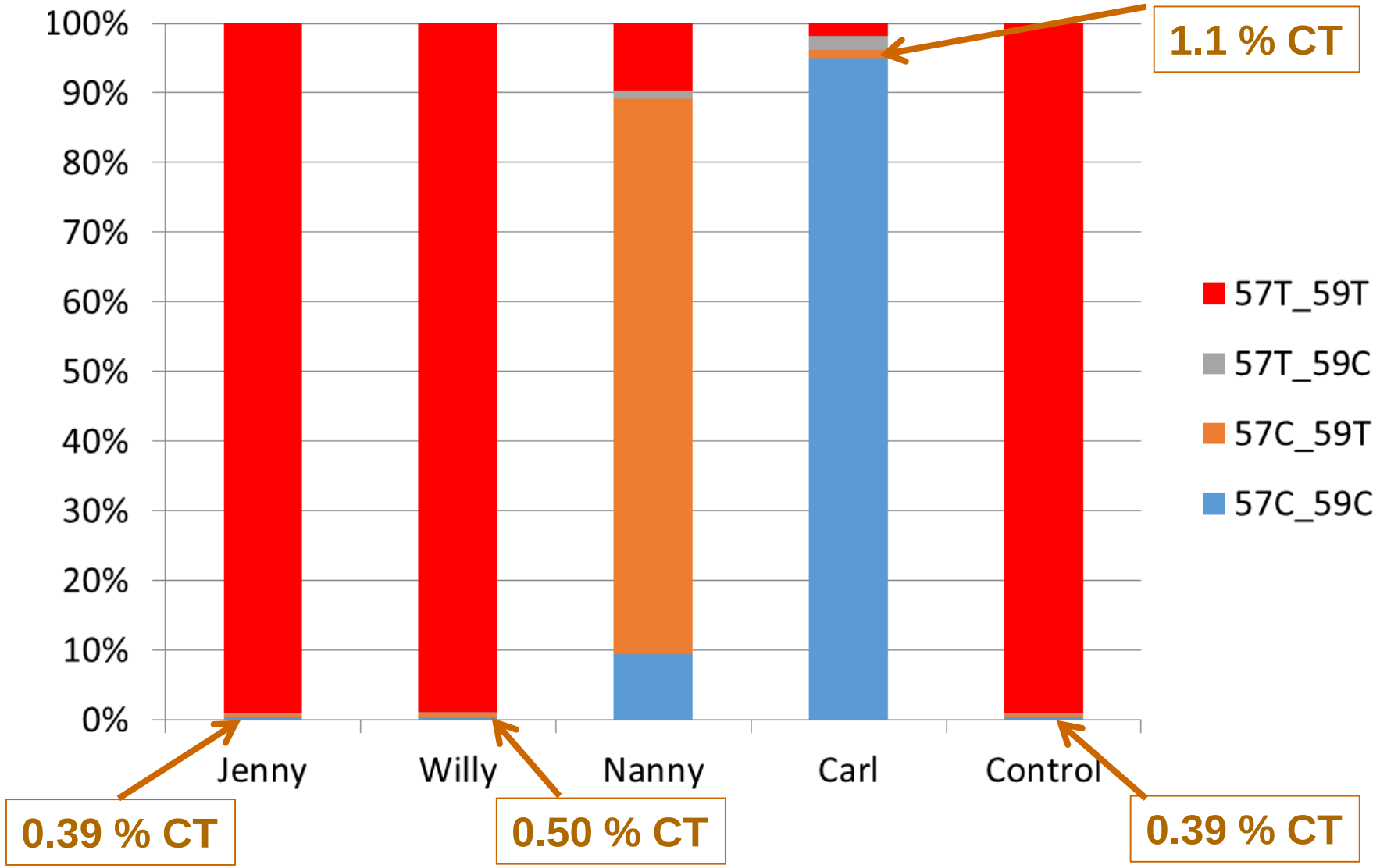
Immigration case - haploptype analysis



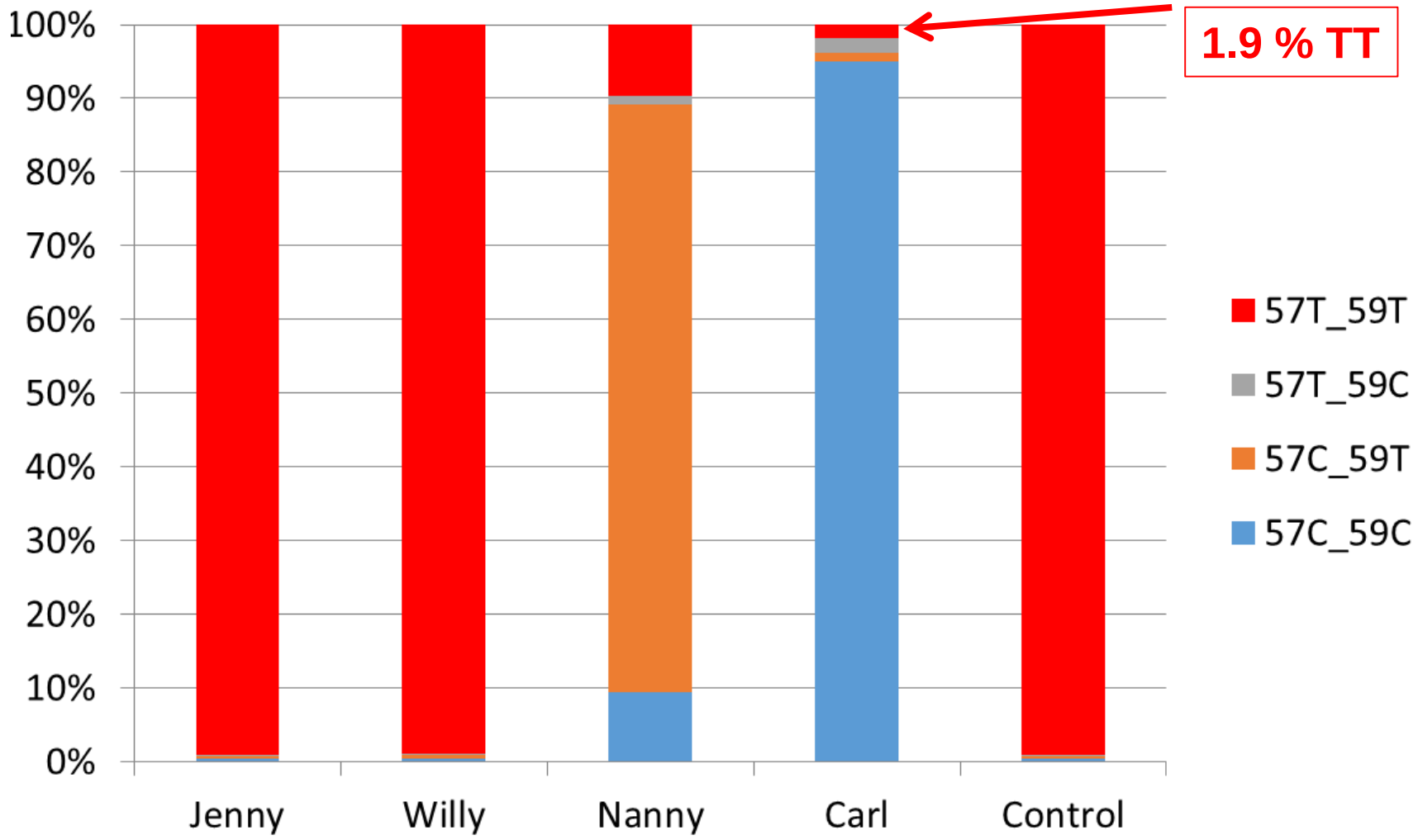
Immigration case - haploptype analysis



Immigration case - haploptype analysis



Immigration case - haploptype analysis



Summary Heteroplasmy Detection

MPS analysis allows detection of heteroplasmy invisible with Sanger even at 10% level

Heteroplasmy detection at the 1% level is possible with MPS (high coverages)

Better understanding of mutational processes and intermediate heteroplasmic states

Need more work to understand instrumental and molecular background

Haplotype analysis possible when more than one difference present

Need more work to understand mutational process and haplotype distribution

Limitations: background signal that cannot be overcome by increasing coverages

Acknowledgements

EUROFORGEN Network of Excellence FP7-SEC-2011-285487

FWF Der Wissenschaftsfonds.

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

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"

EMPOP

Mayra Eduardoff
Catarina Xavier
Christina Strobl
Gabriela Huber

Martin Bodner
Cordula Berger
Harald Niederstätter

POPH QUALIFICAR É CRESCER. QREN QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL GOVERNO DA REPÚBLICA PORTUGUESA UNIÃO EUROPEIA Fundo Social Europeu

Collaborators
Arne Dür (Innsbruck)
Jodi Irwin, Rebecca Just (FBI)
Bruce Budowle, Jonathan King (UNTSC)
Lilian Andrea Casas (Colombia)
EMPOP collaborators

eurofins
Burkhard Rolf
Kerstin Koop

ThermoFisher SCIENTIFIC
Robert Lagacé
Sharon Wootton
Joseph Chang