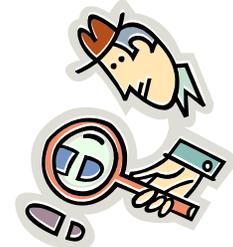
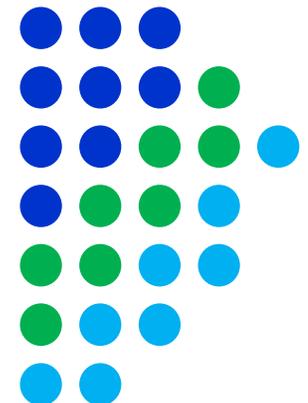




Mitchell M. Holland, Ph.D.
Professor, Biochem & MolBio
Former Director, Forensic Science Program
Eberly College of Science
Penn State University, University Park, PA



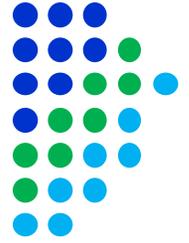
DCMPS of mtDNA Heteroplasmy: An Established Tool for Forensic Investigations



NJ State Police
Forensic Technology Center
8 Oct 2019



<http://forensics.psu.edu/research/dr.-mitchell-holland>

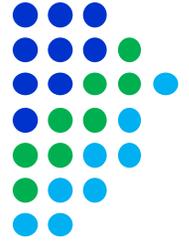


Next Generation Sequencing (NGS) of mtDNA

... a Massively Parallel Sequencing (MPS) Approach to
mtDNA Analysis

A recent search of the literature identified >29,000 published
articles on MPS applications

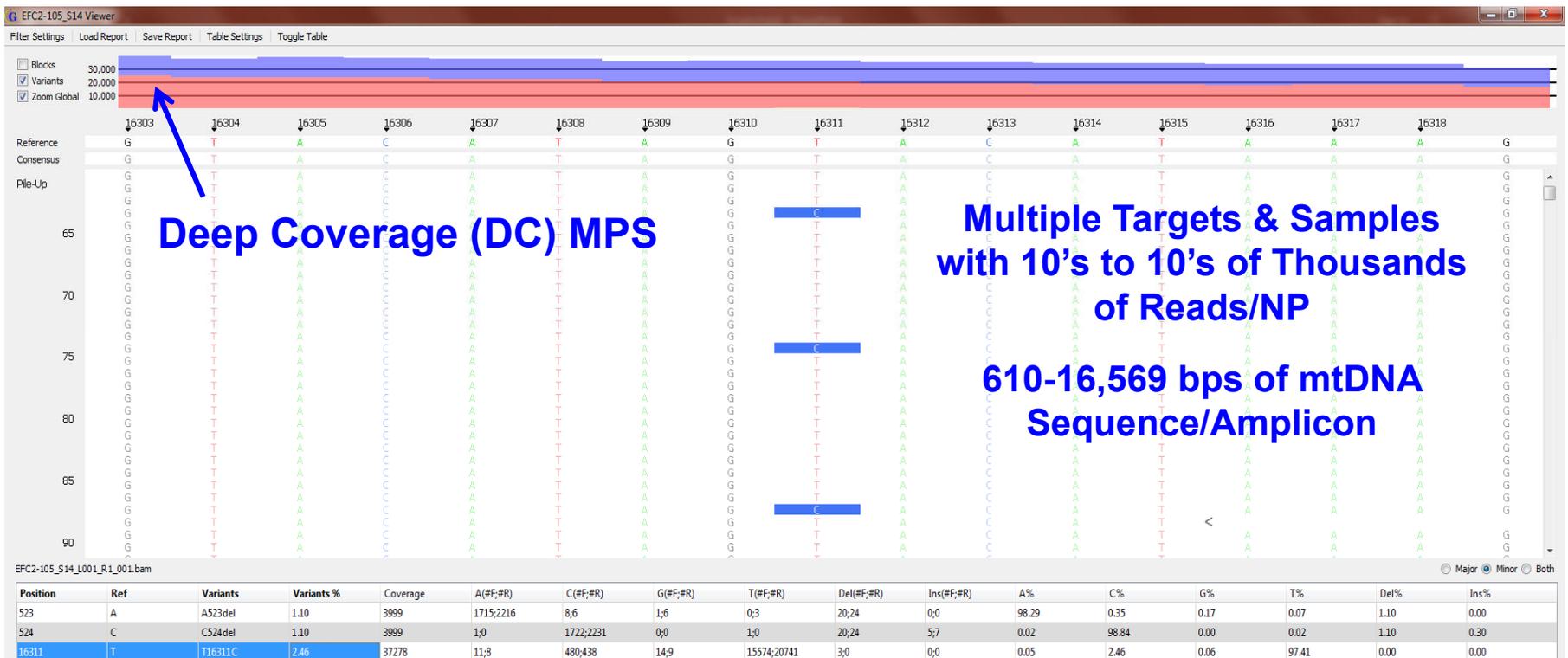
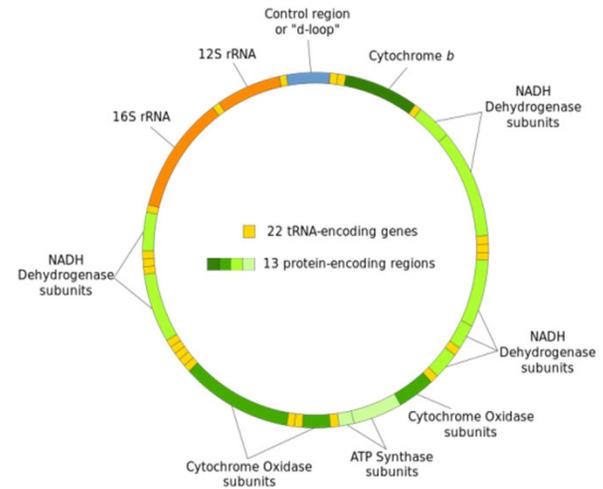
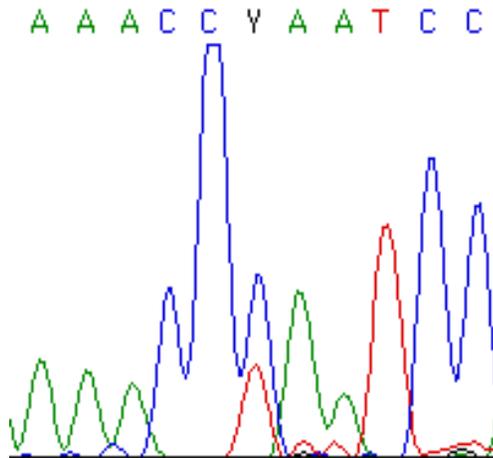
>750 publications on MPS mtDNA
>85 publications on forensic MPS mtDNA



September 23rd	September 24th	September 25th	September 26th	September 27th
8:30 am - 5:00 pm	<p data-bbox="661 755 1291 803">The Future is Now for MPS mtDNA Analysis</p> <div data-bbox="661 836 777 950"></div> <p data-bbox="798 868 1092 925">Michael Brandhagen <i>Co-Chair</i> Forensic DNA Scientist, FBI Laboratory</p> <div data-bbox="661 990 777 1104"></div> <p data-bbox="798 1023 1386 1096">Mitch Holland <i>Co-Chair</i> Associate Professor, Biochemistry and Molecular Biology, Eberly College of Science, Pennsylvania State University</p> <p data-bbox="661 1161 1428 1291"><i>The purpose of this workshop is to educate the community on the availability of complete systems for MPS analysis of mtDNA; including enrichment approaches, library preparation methods, instrument choices, and analysis software solutions.</i></p>			Location TBA

One Amplicon Target & Sample with One Read/NP

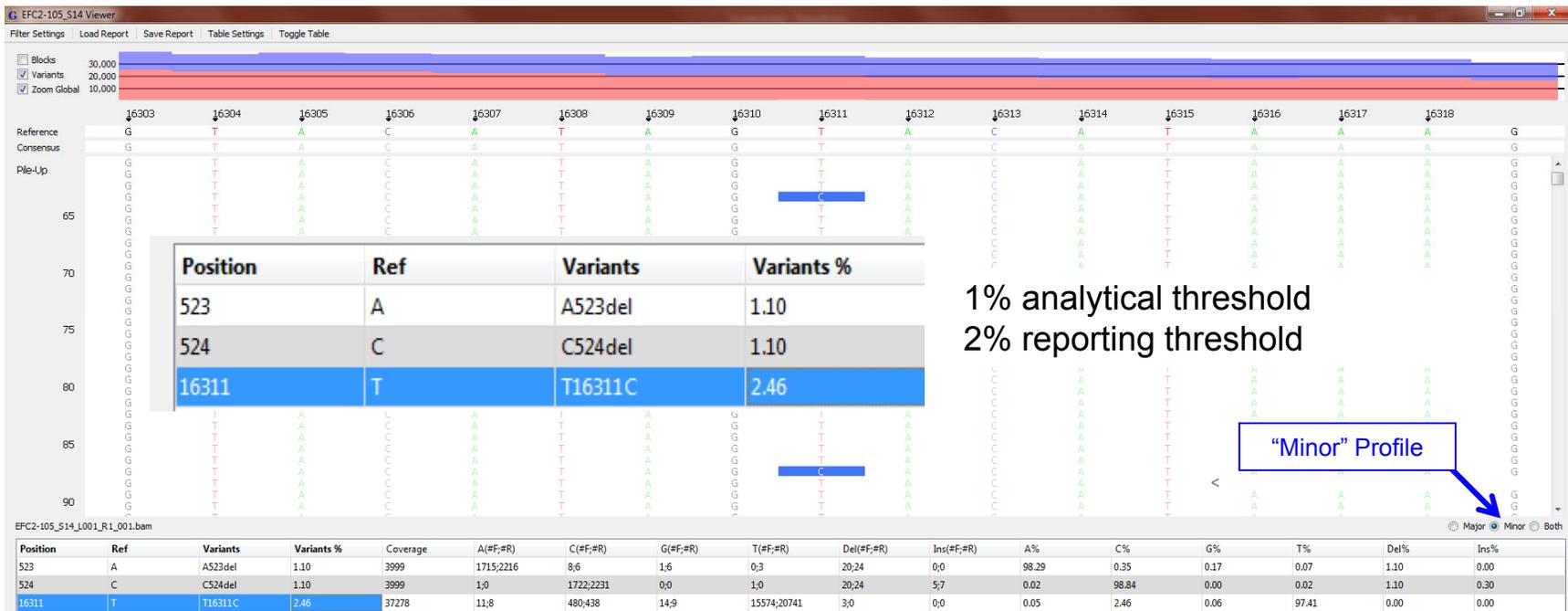
100-450 bps of mtDNA Sequence





MPS can resolve heteroplasmy

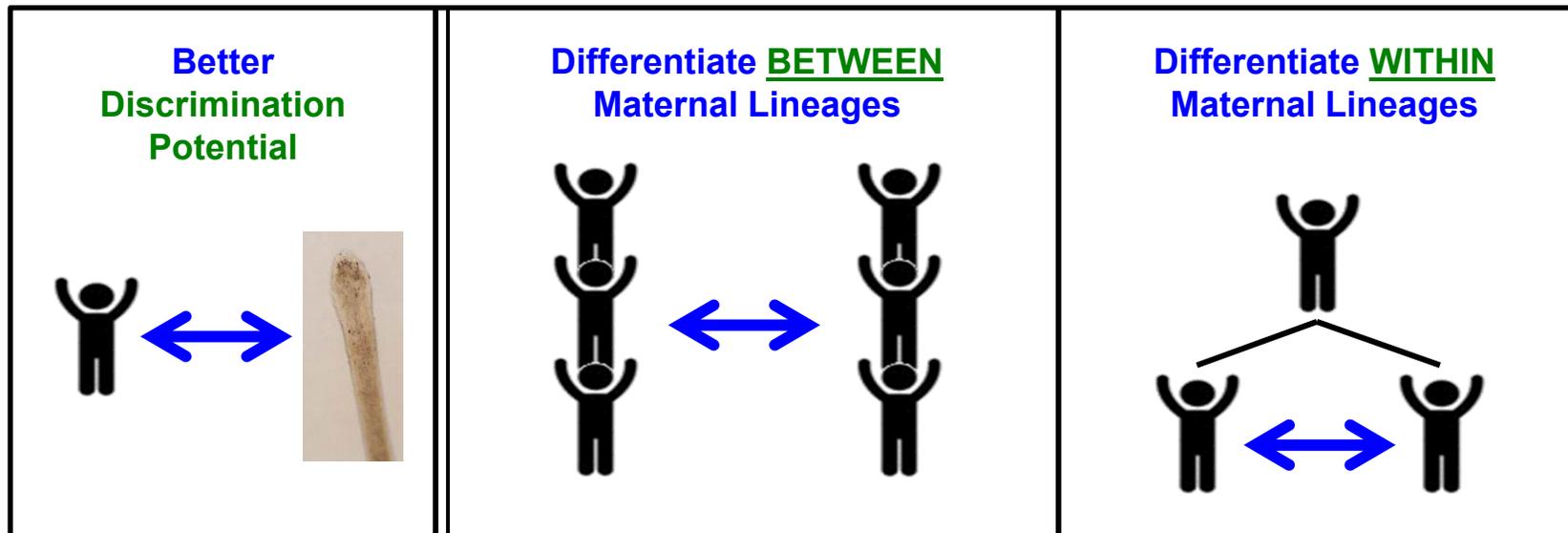
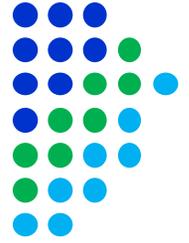
Threshold of 2%



3/30 C variants = 10% Minor Variant Frequency (MVF)



2000 total read minimum/np
40 read minimum for minor variants
2% reporting threshold
(1% analytical threshold)
Balance ratio filters



DCMPS of mtDNA heteroplasmy allows us to accomplish these goals



44% of Mother-Child Pairs were Differentiated

genes

Article

Deep-Coverage MPS Analysis of Heteroplasmic Variants within the mtGenome Allows for Frequent Differentiation of Maternal Relatives

Mitchell M. Holland ^{1*}, Kateryna D. Makova ² and Jennifer A. McElhoo ¹

¹ Department of Biochemistry & Molecular Biology, Forensic Science Program, Eberly College of Science, Pennsylvania State University, University Park, PA 16802, USA; jam760@psu.edu
² Department of Biology, Eberly College of Science, Pennsylvania State University, University Park, PA 16802, USA; kmakova@bx.psu.edu
* Correspondence: mmh20@psu.edu

Received: 1 January 2018; Accepted: 20 February 2018; Published: 26 February 2018



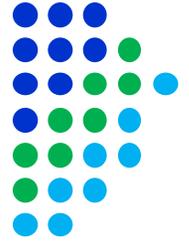
Table 2. Differentiating heteroplasmy for 17 of the 39 mother-child pairs; buccal cell (Bu) and blood (Bl) samples. Frequency of heteroplasmy at each np in percentage (%), with the minor variant annotated as the letter after the np; i.e., T2746C is heteroplasmy at np 2746, with 20.11% of the C variant. Coverage and read distribution (forward reads, #For, compared to reverse reads, #Rev) are provided for each np of differentiating heteroplasmy, along with the gene annotation and whether sites in protein coding genes result in a synonymous change (Y) or not (N). The gene annotations include: CR = control region, 12S & 16S = 12S & 16S rRNAs, ATP6 = ATP synthase 6, ND = NADH dehydrogenase, tRNA^{thr} = tRNA for threonine, and CO = cytochrome oxidase. Metadata for samples without the heteroplasmy are provided to illustrate that read percentages are clearly below reporting threshold and that coverage was adequate for this assessment.

Mother-Child Pair	Nucleotide Position	Sample Number	Major Allele	Coverage (#For:#Rev Reads)	Major Frequency (%)	Minor Allele	Coverage (#For:#Rev Reads)	Minor Frequency (%)	Gene Annotation	Synonymous (Y or N)
T2746C		Mother - Bu (693)	T	2920:6014	79.67	C	655:1600	20.11	16S	
		Child - Bu (677)	T	4838:14038	99.64	C	1:9	0.053		
		Mother - Bl (M207)	T	14187:14328	80.3	C	3440:3528	19.62		
		Child - Bl (M207-C)	T	24044:24176	99.88	C	6:12	0.037		
3	T9179C	Mother - Bu (1134)	T	3063:5076	85.02	C	538:892	14.93	ATP6	N (Val to Ala)
		Child - Bu (1099)	T	6651:8730	99.82	C	8:7	0.097		
		Mother - Bl (M502G)	T	16583:20269	87.14	C	2468:2934	12.77	ATP6	N (Val to Ala)
		Child - Bl (M501)	T	38769:44060	99.81	C	32:24	0.067		
4	G14040A	Mother - Bu (659)	G	5770:4227	92.01	A	474:381	7.86	ND5	Y (Gln)
		Child - Bu (722)	G	20789:16141	99.86	A	8:12	0.054		
		Mother - Bl (M242)	G	13200:12992	94.07	A	831:811	5.89	ND5	Y (Gln)
		Child - Bl (M242-C)	G	10355:10087	99.88	A	5:5	0.049		

... heteroplasmy must be observed in both tissues of one relative but not the other



Metadata: Differentiating



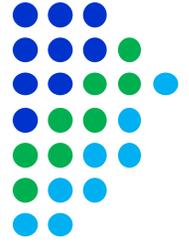
A total of **21 sites of heteroplasmy** were observed in the 17 mother-child pairs, with **no duplicates**, so some pairs had multiple sites of differentiating heteroplasmy

For two pairs, the **mother and child each** had a site of differentiating heteroplasmy

For a third pair, the **mother had one site and the child two sites** of differentiating heteroplasmy



Metadata: Differentiating



Of the 21 sites, 5 were observed in the control region, the most often targeted region of the mtgenome, and **16 sites were observed in the coding region**, further illustrating the value of sequencing the entire mtgenome

However, **5 of the 16 sites** in the coding region (~31%) resulted in **non-synonymous changes** to the protein sequence of the gene locus, so policy questions regarding potential connections to medical information will need to be answered

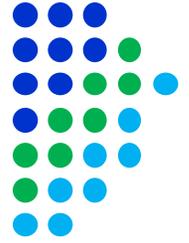


Table 1. **Shared heteroplasmy for 12 of the 39 mother-child pairs.** Frequency of heteroplasmy at each nucleotide position (np) in percentage (%), with the minor variant annotated as the letter after the np; i.e., A1656A (2.11%) is heteroplasmy at np 1656, with the first A as the reference sequence and the second A as the minor variant at 2.11% of the reads. Underlined letters following the np of heteroplasmy indicate a primary haplotype change between the mother and child. In two instances, heteroplasmy was not detected (ND) in a sample.

Mother-Child Pair	Sample Number	Shared Heteroplasmy
1	Mother-Bu (807)	A16183G (7.32%)
	Child-Bu (803)	A16183G (6.89%)
	Mother-BI (M490)	A16183G (2.81%)
	Child-BI (M490-C)	A16183G (2.46%)
2	Mother-Bu (618)	T16189C (7.74%)
	Child-Bu (606)	T16189C (11.07%)
	Mother-BI (M249)	T16189C (2.81%)
	Child-BI (M249-C)	T16189C (9.92%)
3	Mother-Bu (704)	T6152C (7.23%)
	Child-Bu (630)	T6152C (16.37%)
	Mother-BI (M234)	T6152C (5.04%)
	Child-BI (M234-C)	T6152C (16.48%)
4	Mother-Bu (762)	T10873C (2.53%)
	Child-Bu (702)	T10873C (6.66%)
	Mother-BI (M210)	ND
	Child-BI (M210-C)	T10873C (5.40%)

Overall, 22/39 pairs (~56%) had either shared (10), differentiating (5) or both (7) types of heteroplasmy

A total of **14 sites** of shared heteroplasmy (one family with three), with **no duplicates**, and only one site in common with the pairs exhibiting differentiating heteroplasmy (np 16093)

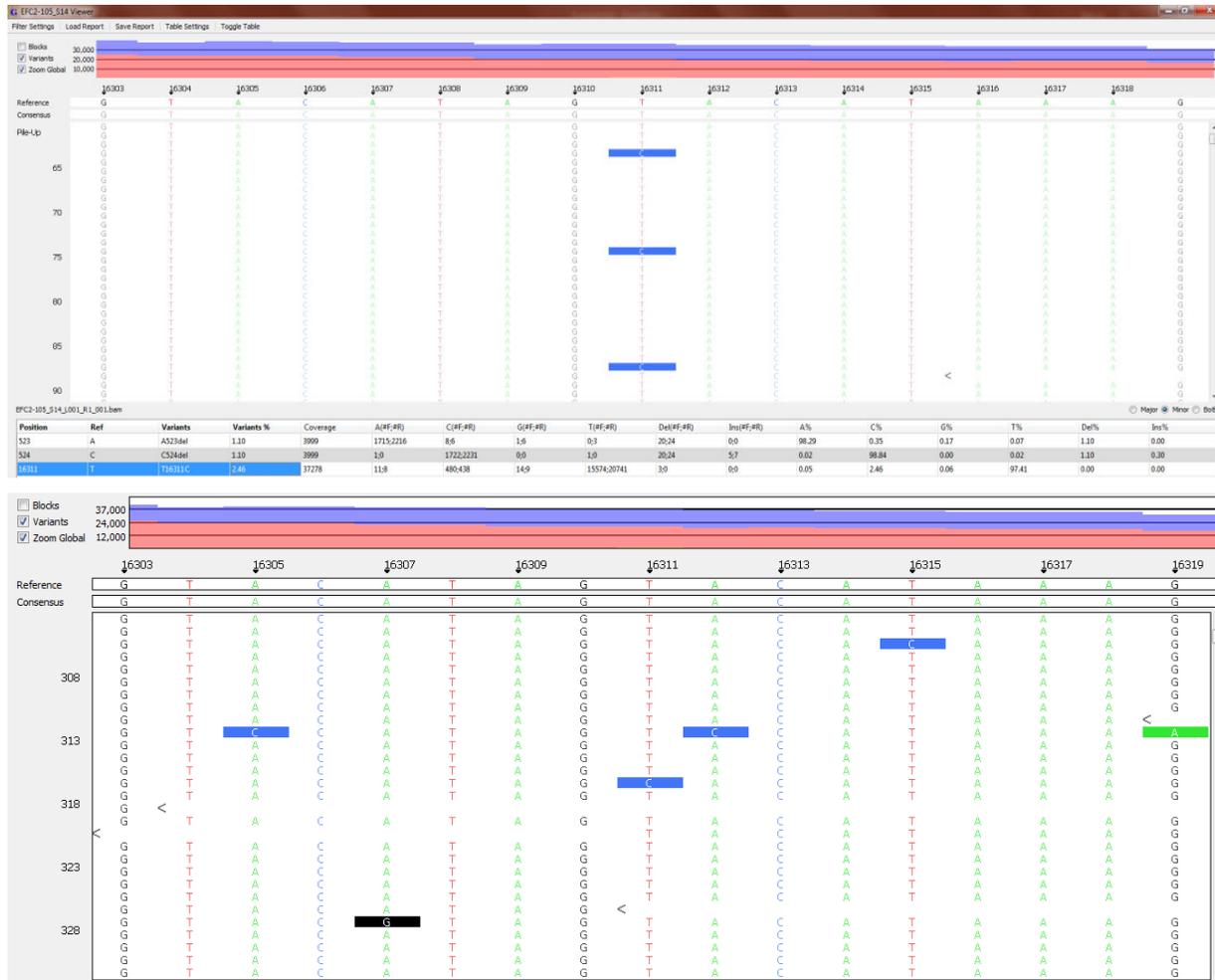


... are any of the shared or differentiating sites of heteroplasmy due to background noise/error in the MPS process, or to numts??



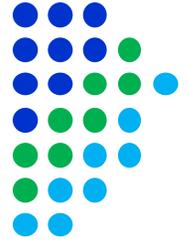


“Noise” in the System





“Noise” in the System



A.

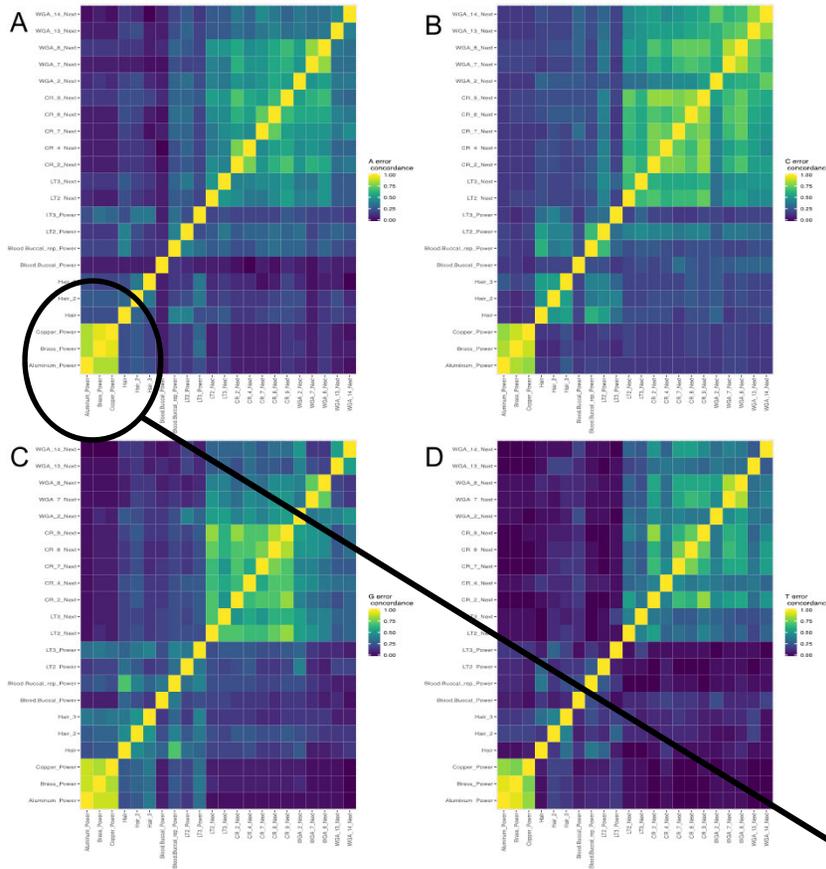
	Metals-Power	Hair-Power	Blood/Buccal-Power	LowTemp-Power	LowTemp-Next	Buccal-Next
Total error	0.485±0.0049	0.325±0.1005	0.231±0.0663	0.297±0.0694	0.205±0.0021	0.182±0.0645
A error	0.118±0.0009	0.100±0.0417	0.052±0.0230	0.048±0.0035	0.070±0.0476	0.037±0.0121
C error	0.121±0.0006	0.085±0.0280	0.080±0.0191	0.069±0.0189	0.079±0.0297	0.060±0.0205
G error	0.131±0.0033	0.048±0.0227	0.044±0.0097	0.059±0.0013	0.061±0.0023	0.047±0.0204
T error	0.115±0.0038	0.091±0.0261	0.053±0.0145	0.049±0.0110	0.065±0.0243	0.039±0.0139

B.

	Buccal-Next CR	Buccal-Next mtgenome
Total error	0.158±0.0720	0.166±0.0745
A error	0.032±0.0148	0.036±0.0164
C error	0.057±0.0273	0.063±0.0292
G error	0.036±0.0180	0.036±0.0171
T error	0.032±0.0126	0.030±0.0124

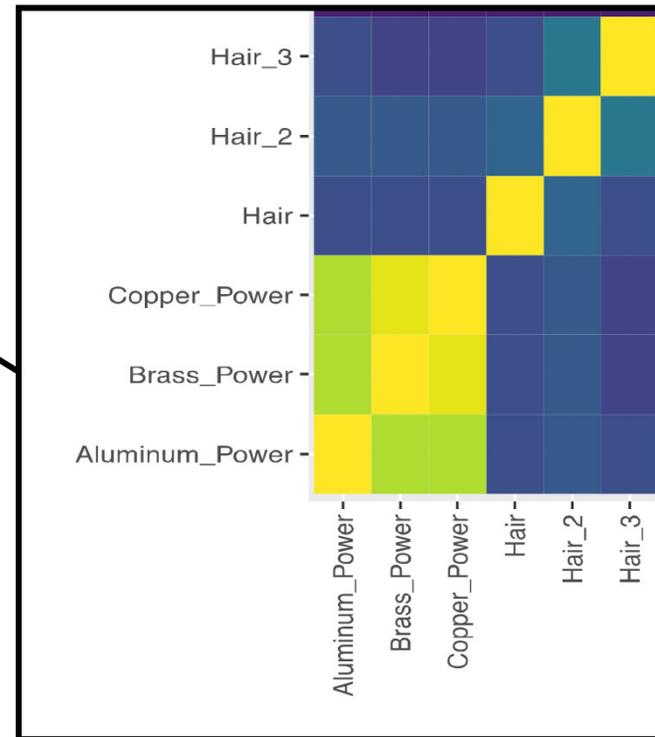
Background noise or error ranged from 0.030% to 0.131% across the four nucleotides.

Samples with increasing DNA damage saw increases in error.



Comparison of sites with greatest error between MiSeq runs (i.e., datasets)

Dark Blue = zero concordance
Yellow = 100% concordance



DNA recovered from the surface of ammunition components (Holland et al., FSIG 2019) gave the highest level of concordance



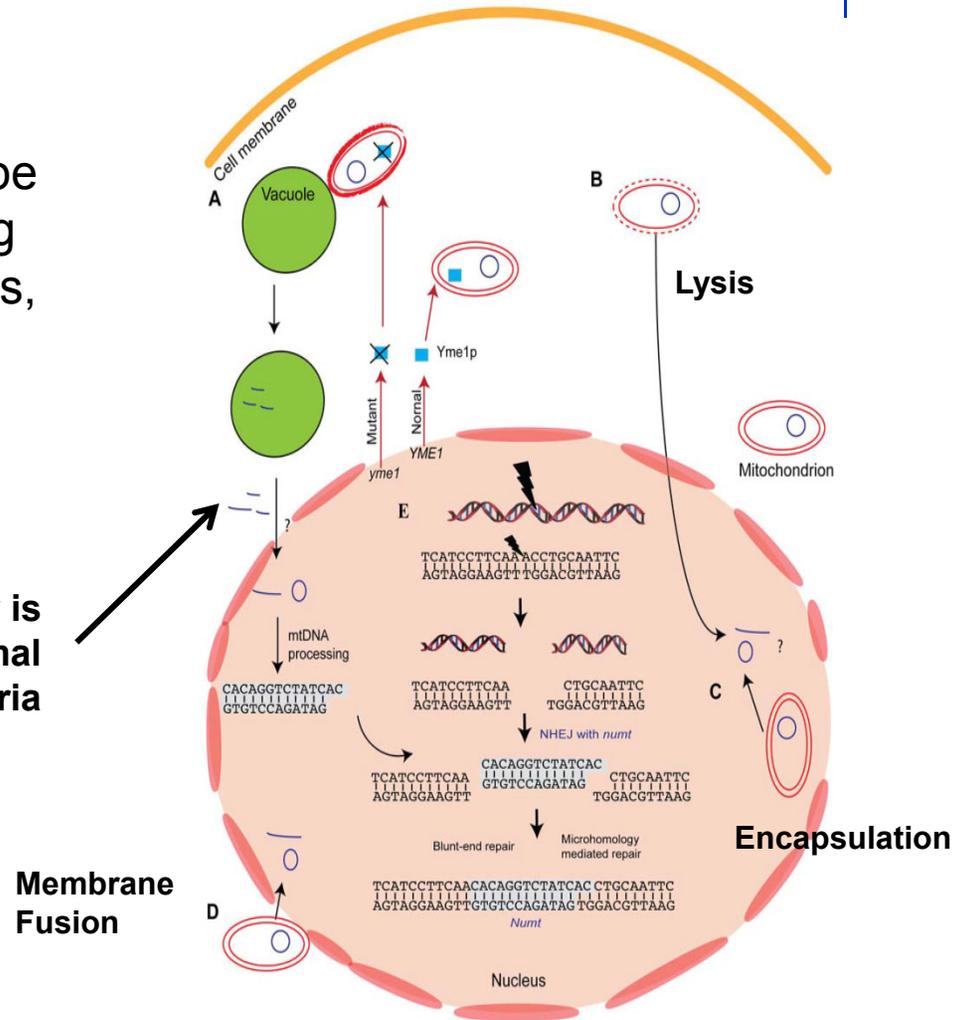
NUMTs

- = nuclear mitochondrial DNA segments *
- = nuclear mitochondrial sequences
- = nuclear mitochondrial segments
- = nuclear mitochondrial insertions



Estimate that ~750 NUMTs can be found in the nugenome, including some entirely intact mitogenomes, with ~4 unique NUMTs per individual

Most supported pathway is degradation of abnormal mitochondria



OPEN ACCESS Freely available online

PLoS GENETICS

Review

Molecular Poltergeists: Mitochondrial DNA Copies (*numts*) in Sequenced Nuclear Genomes

Einat Hazkani-Covo^{1†}, Raymond M. Zeller^{1,2}, William Martin³

1 National Evolutionary Synthesis Center, Durham, North Carolina, United States of America, 2 Mathematics Undergraduate Program, Duke University, Durham, North Carolina, United States of America, 3 Institut für Botanik III, Heinrich-Heine Universität Düsseldorf, Düsseldorf, Germany



Mitochondrion 46 (2019) 302–306

Contents lists available at ScienceDirect

Mitochondrion

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

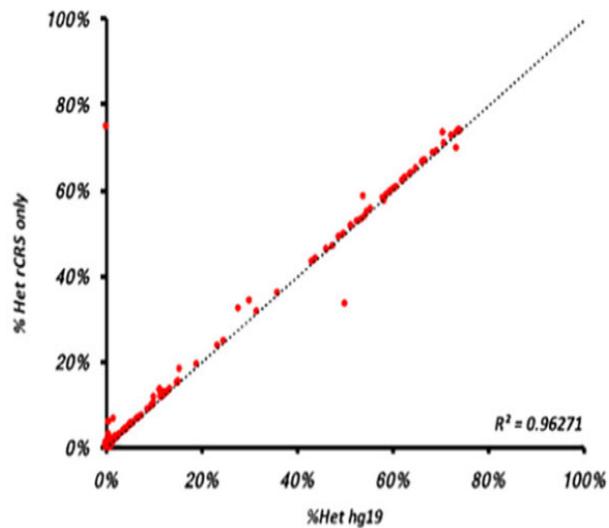
Short communication

Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution

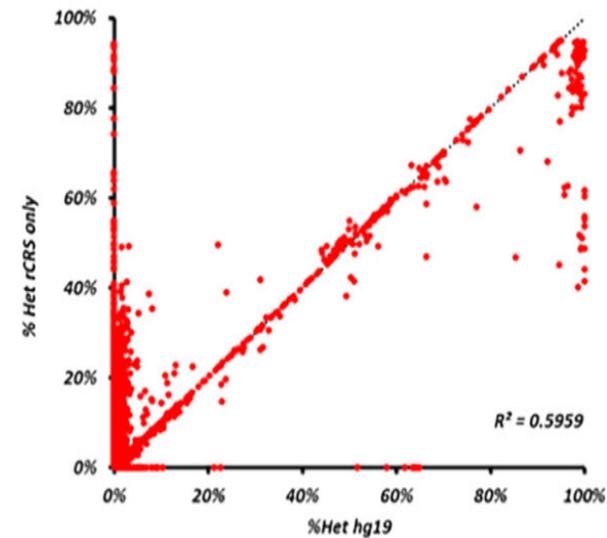
Mauro Santibanez-Koref^a, Helen Griffin^a, Douglass M. Turnbull^b, Patrick F. Chinnery^c, Mary Herbert^{b,d}, Gavin Hudson^{b,d,*}

^a Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK
^b The Wellcome Centre for Mitochondrial Research, Newcastle University, Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK
^c MRC Mitochondrial Biology Unit, Wellcome Trust/MRC Building, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0XY, UK
^d Newcastle Fertility Centre, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK

Heteroplasmy observed in whole mitogenome MPS data is dependent on enrichment method



2 amplicon approach



180 amplicon approach

As amplicon size decreases, the number of minor variants with discrepant frequencies can increase, with the majority of variant frequencies increasing when aligned to the rCRS only



Short communication

Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution



Mauro Santibanez-Koref^a, Helen Griffin^a, Douglass M. Turnbull^b, Patrick F. Chinnery^c, Mary Herbert^{b,d}, Gavin Hudson^{b,d,*}

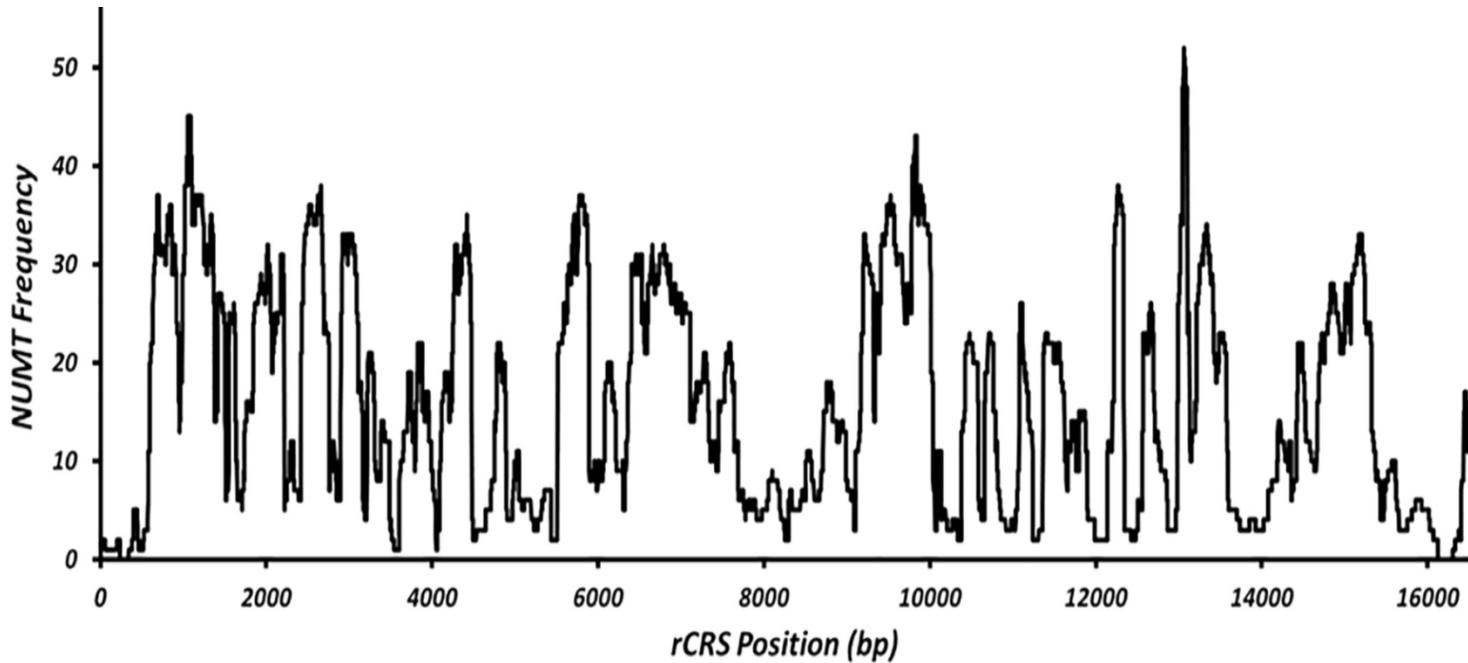
^a Institute of Genetic Medicine, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK

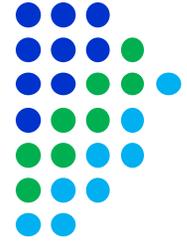
^b The Wellcome Centre for Mitochondrial Research, Newcastle University, Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK

^c MRC Mitochondrial Biology Unit, Wellcome Trust/MRC Building, Cambridge Biomedical Campus, Hills Road, Cambridge CB2 0XY, UK

^d Newcastle Fertility Centre, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK

NUMT frequency is relatively low in the control region compared to sites across the coding region





Rates of DNA Duplication and Mitochondrial DNA Insertion in the Human Genome

Douda Bensasson, Marcus W. Feldman, Dmitri A. Petrov

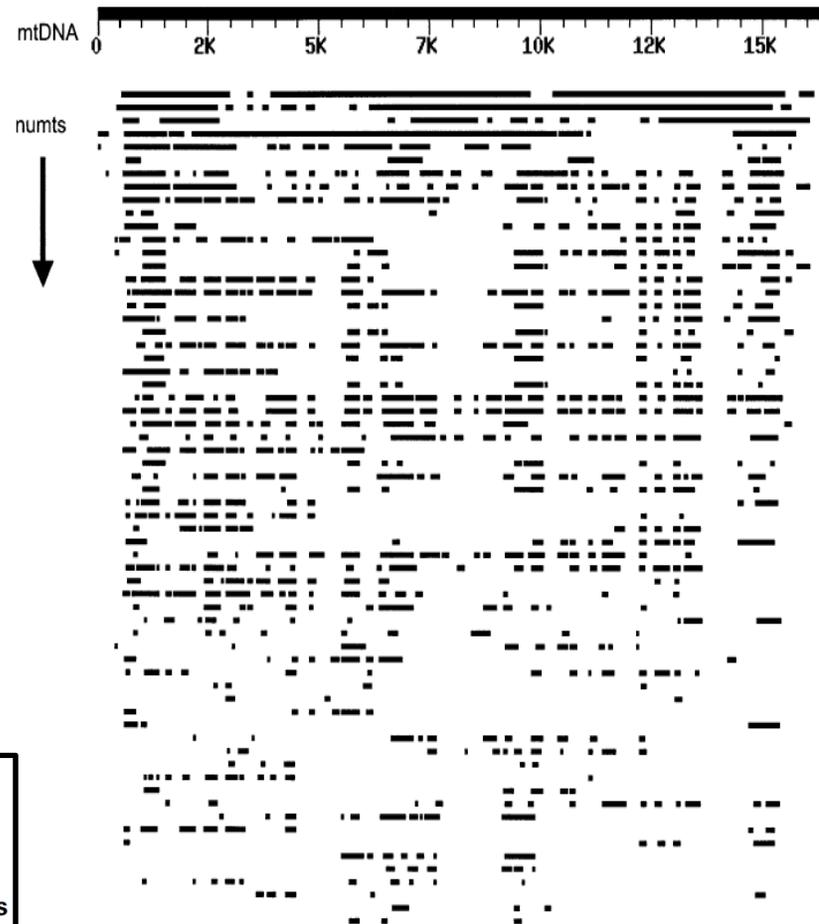
School of Biological Sciences, Stanford University, 371 Serra Mall, Stanford, CA 94305, USA

JOURNAL OF **MOLECULAR
EVOLUTION**

© Springer-Verlag New York Inc. 2003

NUMT copy number can be relatively high, but still far lower than actual mtDNA targets, so chances are low of detecting NUMT sequences

Bottom line: amplify larger targets and lower amounts of template DNA, and ultimately filter out phylogenetically inconsistent SNPs



Electrophoresis 2018, 39, 2785–2797

Joseph David Ring^{1,2}
Kimberly Sturk-Andreaggi^{1,2}
Michelle Alyse Peck^{1,2*}
Charla Marshall^{1,2}

Research Article

Bioinformatic removal of NUMT-associated variants in mitotiling next-generation sequencing data from whole blood samples

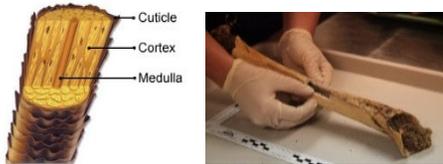
¹Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), DE, United States
²ARP Sciences, LLC, Rockville, MD, United States



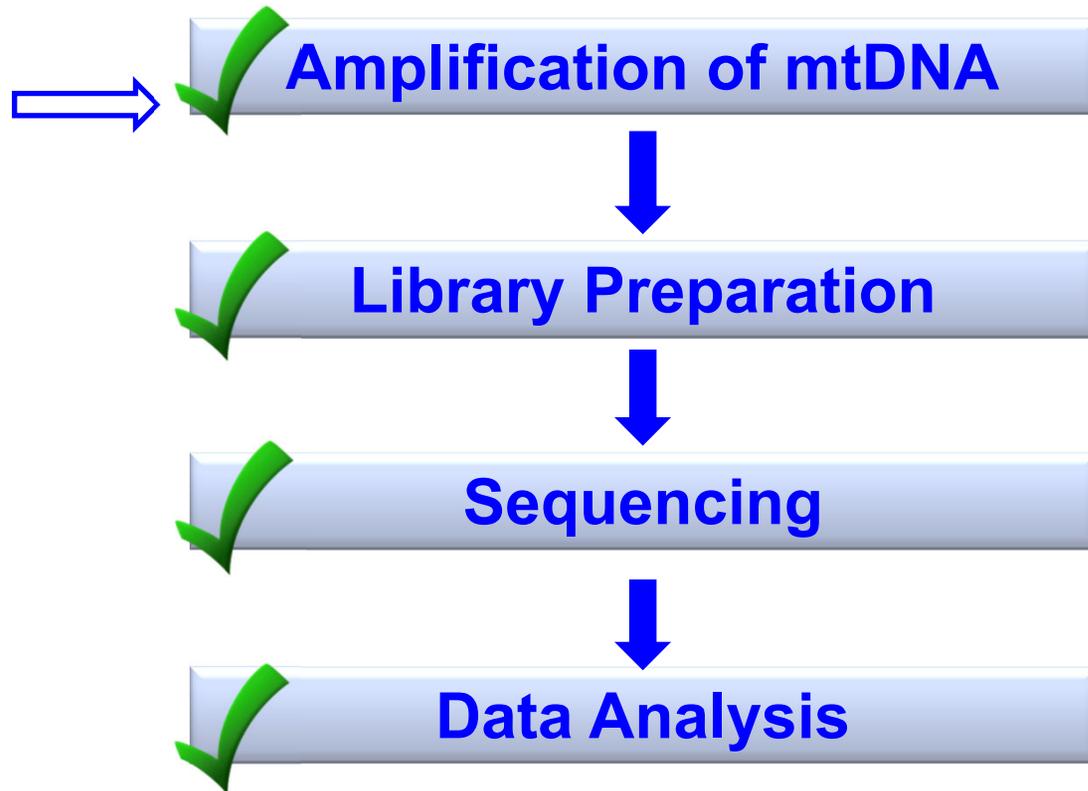
mtDNA Workflow

MPS

**DNA
Extraction**



DNA extraction is the same, but in many instances the amount of extract needed for amplification is reduced





Identification of Marija Krucifiksa Kozulić



Sidney Gaston-Sanchez, Dragan Primorac, Šimun Anđelinović, Željana Bašić, Ivana Kružić, Alan Bosnar, Anja Petaros, Miran Čoklo, Jennifer Daniels-Higginbotham, Charla Marshall

Archbishop Ivan Devčić, Roman Catholic Archdiocese of Rijeka and Sister Dobroslava Mlakić, Catholic Order of the Daughters of the Sacred Heart of Jesus, Rijeka

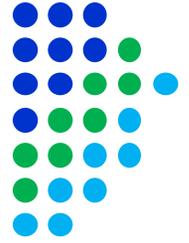
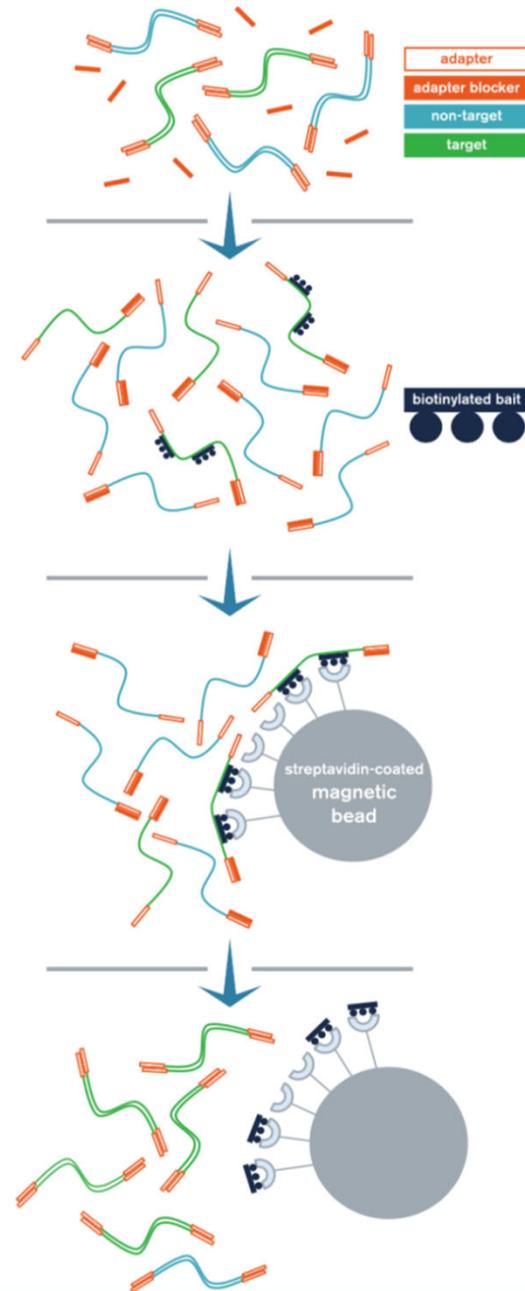
Sister **Marija Krucifiksa Kozulić** was a nun from Rijeka who dedicated her life to helping the poor and less fortunate. She is currently under consideration for beautification by the Vatican, which requires the identification of her remains.

Sister Kozulić died in 1922 and was buried in a tomb along with other nuns belonging to the Society of Sisters of the Sacred Heart of Jesus, including her biological sister, **Tereza Kozulić**.

Experts in varying disciplines such as forensic anthropology, pathology, and forensic molecular biology have contributed to the ongoing identification efforts.



Capture method using a custom designed bait cocktail developed by AFDIL for the entire mitogenome



Forensic Science International: Genetics 31 (2017) 198–206

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

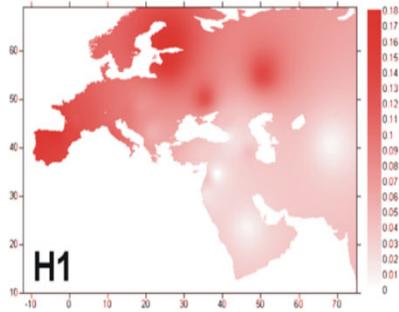
Research paper

Performance evaluation of a mitogenome capture and Illumina sequencing protocol using non-probative, case-type skeletal samples: Implications for the use of a positive control in a next-generation sequencing procedure

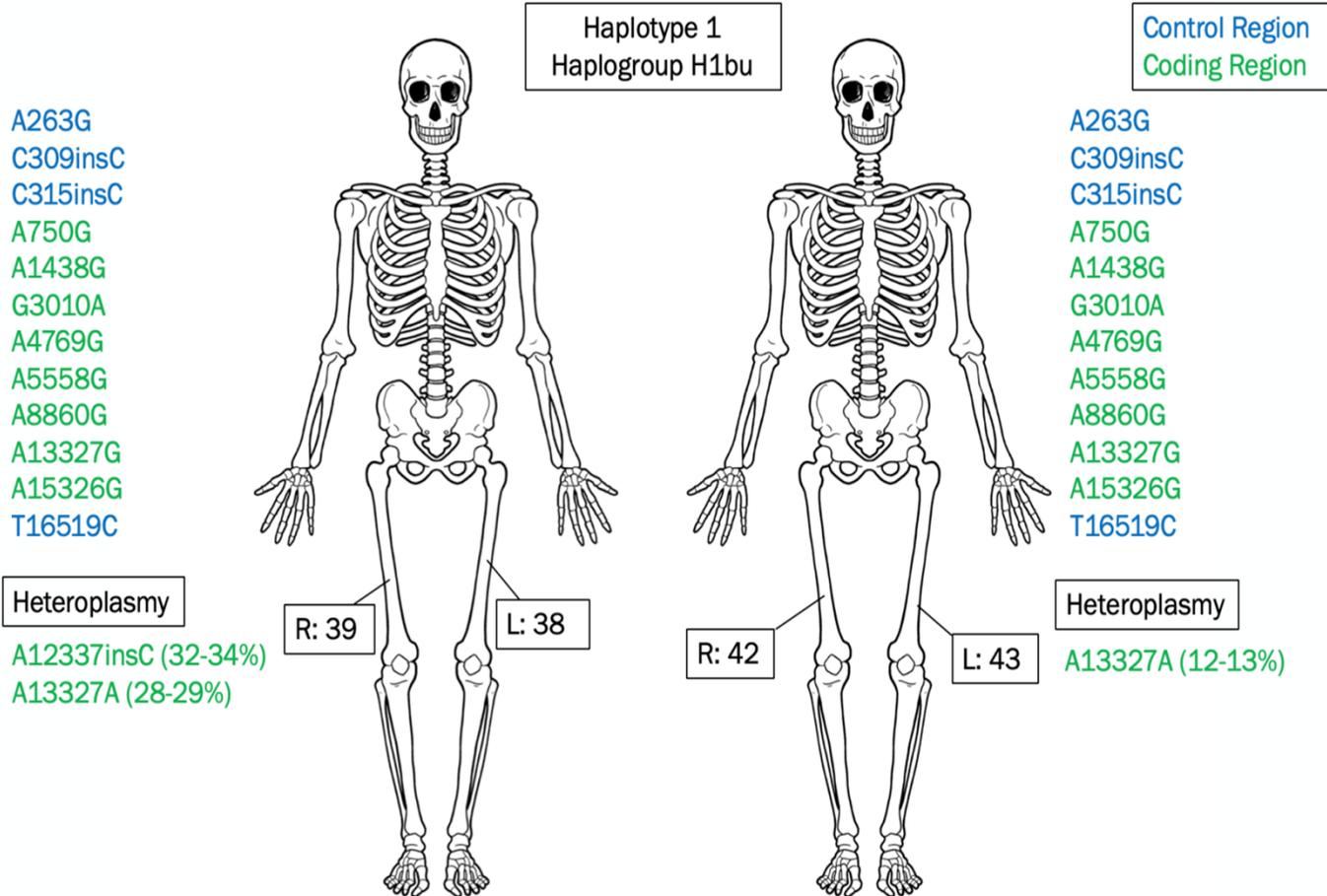
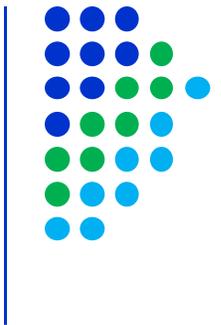
Charla Marshall^{a,b,*}, Kimberly Sturk-Andreaggi^{a,b}, Jennifer Daniels-Higginbotham^{a,b}, Robert Sean Oliver^{a,b}, Suzanne Barritt-Ross^{a,b}, Timothy P. McMahon^a

^a Annual Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), Department of Defense DNA Operations, 115 Purple Heart Dr., Dover AFB, DE 19902, United States

^b AFD Sciences, LLC, 9210 Corporate Blvd., Rockville, MD 20850, United States*

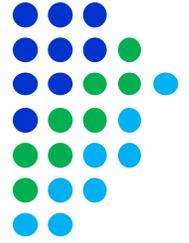


b subclade is common in Eastern Europe



13327 is in the ND5 gene, with the A>G SNP causing a non-syn change from T>A

No known disease state associated with the change



**Mitochondrial DNA
sequence heteroplasmy in
the Grand Duke of Russia
Georgij Romanov
establishes the authenticity
of the remains of Tsar
Nicholas II**

Pavel L. Ivanov¹, Mark J. Wadhams²,
Rhonda K. Roby², Mitchell M. Holland²,
Victor W. Weedn² & Thomas J. Parsons²
nature genetics volume 12 april 1996

Likelihood Ratio

$$LR = \frac{p(E1/R) \times p(E2/R)}{p(E1/R') \times p(E2/R')}$$



p(E1/R) = the probability of the evidence (haplotype match between Marija and Tereza Kozulić) given the hypothesis that the remains are those of Marija Kozulić

E2 = the probability of co-occurrence of heteroplasmy

R' = given the hypothesis that the remains are unrelated

Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA

Boris Rebolledo-Jaramillo^{a,1}, Marcia Shu-Wei Su^{b,1}, Nicholas Stoler^a, Jennifer A. McElhoe^c, Ben Dickins^d, Daniel Blankenberg^e, Thorfinn S. Korneliusen^f, Francesca Chiaromonte^g, Rasmus Nielsen^h, Mitchell M. Holland^f, Ian Paul^g, Anton Nekrutenko^{h,2}, and Kateryna D. Makova^{h,2}

Departments of ^aBiochemistry and Molecular Biology, ^bBiology, and ^cStatistics, ^fForensic Science Program, and ^dDepartment of Pediatrics, College of Medicine, Pennsylvania State University, University Park, PA 16801; ^eSchool of Science and Technology, Nottingham Trent University, Nottingham NG1 4BU, United Kingdom; and ^gDepartment of Integrative Biology, University of California, Berkeley, CA 94720

Edited by Michael Lynch, Indiana University, Bloomington, IN, and approved September 8, 2014 (received for review May 20, 2014)

Likelihood Ratio

$$LR = \frac{p(E1/R) \times p(E2/R)}{p(E1/R') \times p(E2/R')}$$



Calculations are dependent on:

Haplotype frequencies (search, 1366 mitogenome profiles)

Generational differences (known, 2)

Mutation rates (empirical, 1/640)

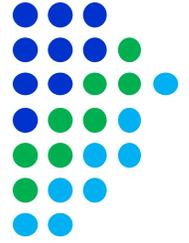
Heteroplasmy rates & sites (empirical, 5.6/10 & estimated 1/500)

Generations to fixation (estimated, 1/20)

**EMPOP has 1366 mitogenome sequences in the database:
Confidence Limit from Zero Proportion = 0.00219 (1 in 457)**



Likelihood Ratio

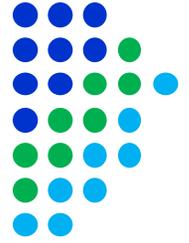


$$LR = \frac{p(E1/R) \times p(E2/R)}{p(E1/R') \times p(E2/R')}$$

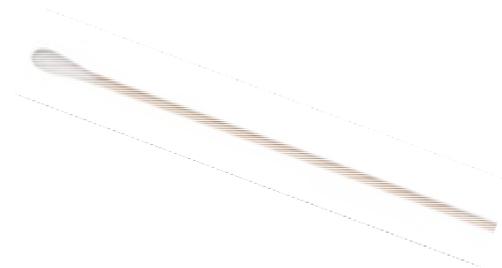
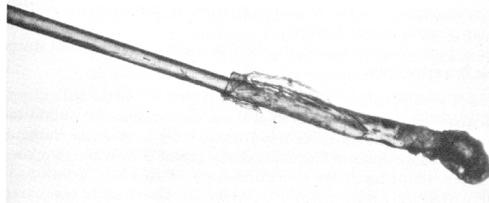


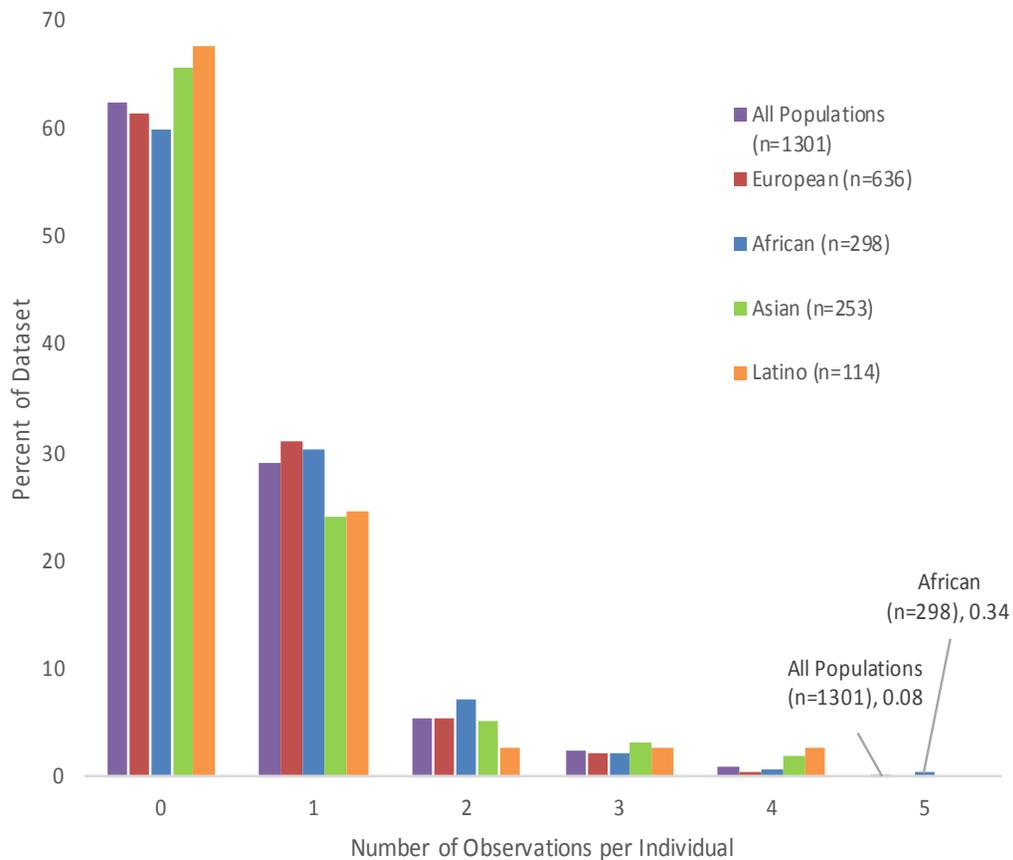
$$= (455)(807) = 367,185$$

Therefore, the DNA evidence is ~367,185 times more likely if one of the two set of remains are those of Marija Kozulić

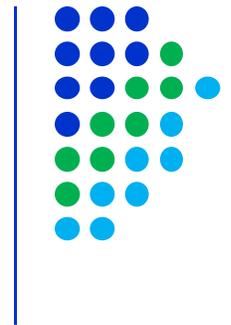
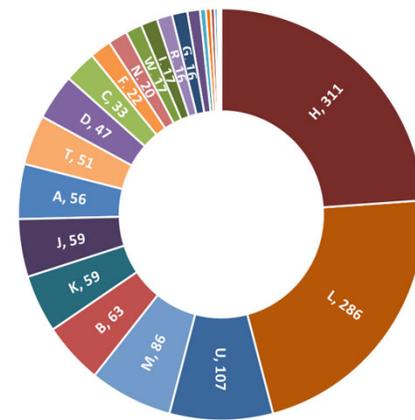


What if we just have an evidence
to reference match?



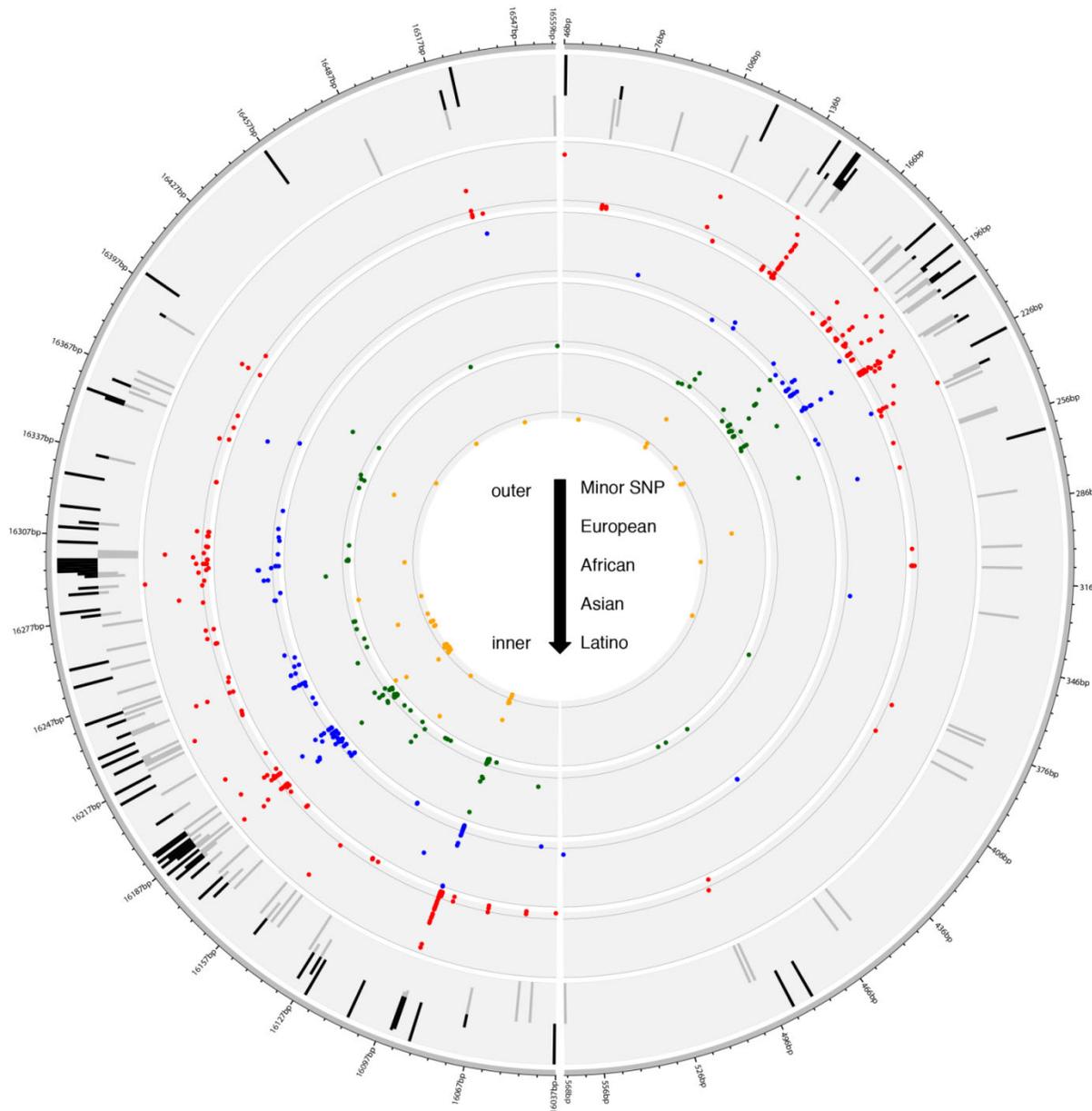
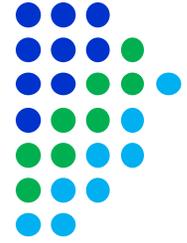


Rates of Heteroplasmy on a per Individual Basis



Overall, 37.7% of the **combined** population of 1301 individuals of **European, African, Asian, and Latino** decent exhibited a site or sites of heteroplasmy. The rate of heteroplasmy is highly similar across the four population groups, representing 22 major haplogroups, the majority of which fall within H and L.

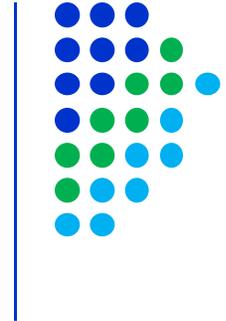
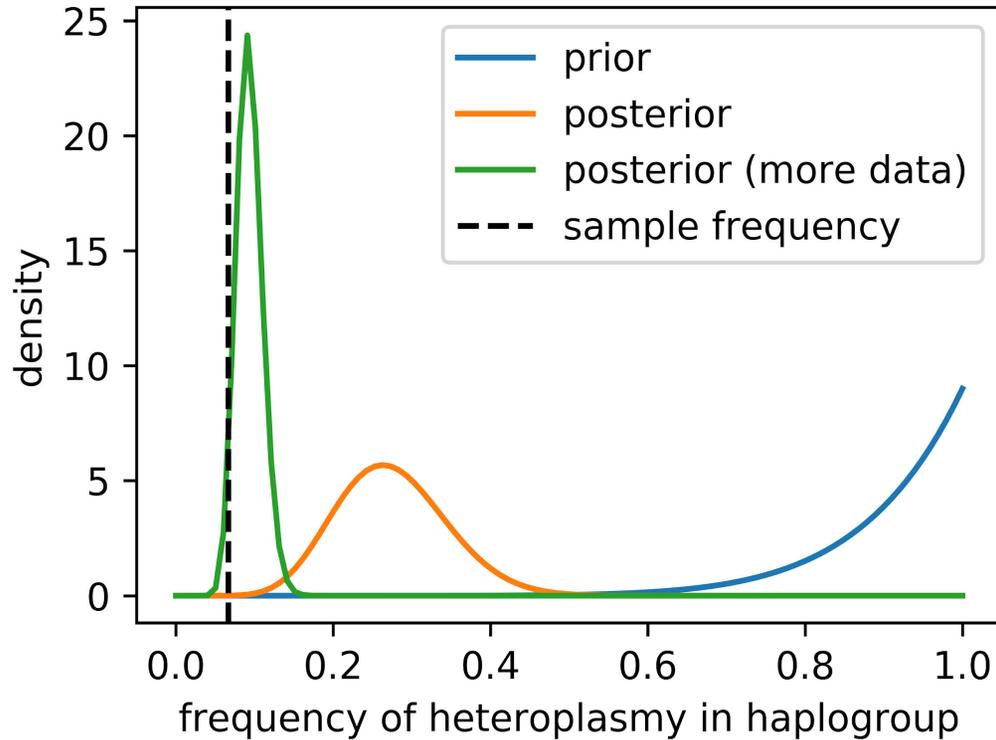
Therefore, heteroplasmy should be observed, on average, in more than one in three forensic cases.



Sites of heteroplasmy across the CR, with numbering according to the rCRS. Sites are identified for **European** (n=636), **African** (n=298), **Asian** (n=253), and **Latino** (n=114) population groups.

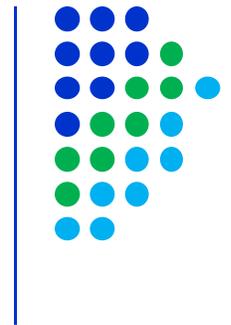
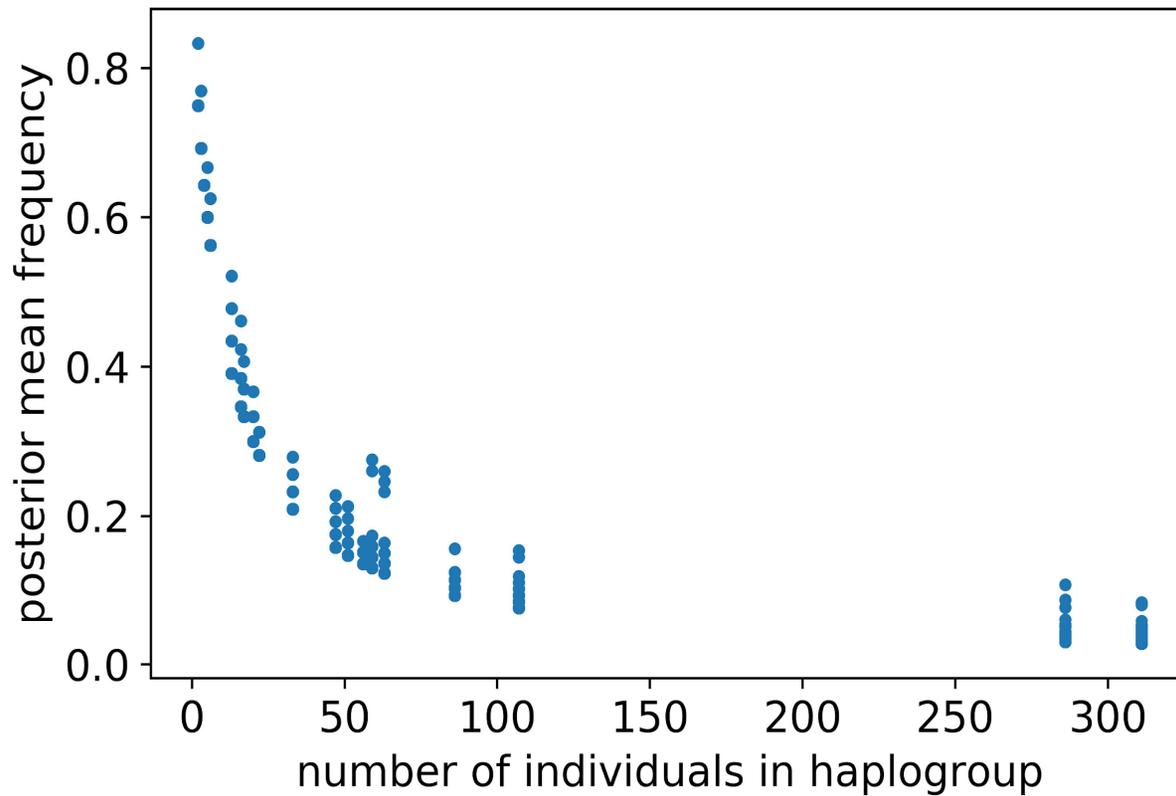
Inner doublet rings represent 2-10% heteroplasmy, with %'s above 10% moving outward for each group.

The outer ring reflects the direction of the minor variant, with **Gray** bars the SNP and **Black** bars the reference sequence. The length of the bar represents the proportion of the minor variant when both scenarios are observed.



Schematic of Bayesian inference of heteroplasmy frequencies in a haplogroup. The prior distribution on the frequency of heteroplasmy in a haplogroup (blue) is skewed toward higher, or more conservative frequencies. The orange distribution shows the posterior distribution after observing two heteroplasmy in 30 observations, and the green distribution shows the posterior distribution after observing 20 heteroplasmy in 300 observations. The sample frequency of both samples is 6.7% (dashed vertical line).

This model illustrates how the posterior distribution approaches the sample frequency as more data is collected.

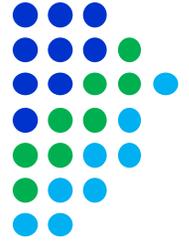


Relationship between posterior mean heteroplasmy frequency and major haplogroup sample size. Each point represents the estimate of heteroplasmy occurrence rate in at a particular site in a particular major haplogroup. Estimates are shown only for site-haplogroup combinations at which heteroplasmy was observed at least once.

Frequencies decrease as the number of individuals in a population group increases.



Bottom Line



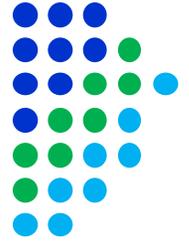
The majority of heteroplasmies are between 2-10%, strongly supporting MPS as the method of choice for mtDNA sequence analysis

Heteroplasmic sites are not evenly distributed, suggesting that selection plays a role in accumulation of heteroplasmies in the CR

The direction of heteroplasmy may be impacted by the location of the site, with the minor variant predominantly a SNP in void regions



Bottom Line



With the current dataset, co-occurrence of heteroplasmy will increase the LR by a factor of 10 or greater

With additional data, co-occurrence of heteroplasmy should increase the LR by a factor of 100-1000

Ultimately, the LR should be in the range of 100,000-1,000,000 for an mtDNA match when considering a haplotype match with the co-occurrence of heteroplasmy



**Other
Consultants &
Collaborators:
AFDIL, ATF,
Mitotyping, OH
BCI, UC
Berkeley**



**Walther Parson (Innsbruck) &
Ann Gross (MN BCA)
Consultants**

**Jen McElhoe,
Research Associate (NIJ)**

**Charity Holland,
Research Assistant (NIJ)**

Master's Students:

**Rachel Bonds (oxidative damage, ammunition, NIJ)
Sidney Gaston-Sanchez (deamination damage & Sister
Marija, NIJ)**

**Sam Jarvis (soil characterization)
Grace Marino (CT scanning of skeletal elements)
Therese Mandracchia (fired ammunition)
Lauren Canale (Hair mtGenome, NIJ)**

UG Students:

**Dana Dirnberger (PowerSeq WGM, mtqPCR)
Troy Adams (epigenetic age predictions)**

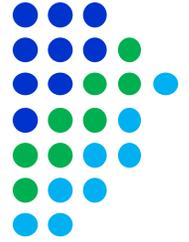


**Jamie Gallimore (drift in hair, NIJ)
Elena Zavala (bone extraction, ForenSeq)
Molly Rathbun (deamination damage, NIJ)
Erica Pack (GeneMarker® HTS, NIJ)
Emmy Demchak (hetero rates in pop groups, NIJ)**





Thanks!!



Verogen

Cydne Holt, Kathy Stephens, etc



SoftGenetics – GeneMarker® HTS

John Fasnacht, Teresa Snyder-Leiby, etc



Promega

Doug Storts, Spencer Hermanson, Jeff Shaw, Margaret Ewing



National Institute of Justice (NIJ 2014-DN-BX-K022)

National Institute of Justice (NIJ 2015-DN-BX-K025)

National Institute of Justice (NIJ 2016-DN-BX-0171)



Eberly College of Science, Department of Biochemistry & Molecular Biology, Forensic Science Program at Penn State



Daisy says hello!

mmh20@psu.edu

