Ancient DNA Analysis of St. Mary's City Lead Coffin Burials

October 23 2016 working paper

David Reich[1,2,3], Kristin Stewardson[1,2], Iosif Lazaridis[1], Swapan Mallick[1,2,3], Nadin Rohland[1] and Douglas Owsley[4]

[1] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA
[3] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4] Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA

To whom correspondence and request for data should be addressed: David Reich (reich@genetics.med.harvard.edu) and Douglas Owsley (owsleyd@si.edu)

Overview

The goal of this report is to document the screening for ancient DNA of bone samples from each of three Lead Coffin Burials from Historic St. Mary's City.

Methods

The screening of these samples was based on a modified version of protocols previously used in several published papers from the Reich laboratory at Harvard Medical School.[1-4] Table 1 shows the results. Briefly:

(1) A Dremel tool with a sterile disposable drill bit was used to take 43-82 mg of powder from each sample after cleaning the surface with a disposable sanding disk.

(2) DNA from the powder was extracted for each sample.[5]

(3) 30uL or 3uL of DNA extract was converted into barcoded non-UDG-treated Illumina sequencing libraries, a total of five libraries for the three samples.[6] Two of the libraries passed standard wet laboratory quality control (the others had evidence of inhibition).

(4) The two libraries passing quality control were hybridized to oligonucleotide probes targeting the mitochondrial genome.[1,7] The enriched libraries were sequenced on an Illumina NextSeq 500 instrument using 2×76 bp reads. Identifying barcodes and adapters were trimmed, read pairs with at least 15 base pairs of overlapping sequences were merged, and the merged sequences were mapped to the RSRS mitochondrial DNA reference genome,[8] using the Burrows Wheeler Aligner[9] (bwa) and the command samse (v0.6.1). Clusters of molecules duplicated during the laboratory process were identified by reads with matching start and end position and orientation, and were represented using the single read with highest quality. A mitochondrial consensus sequence was built and a haplogroup was called as previously described.[1]

(5) The two libraries passing quality control were hybridized to oligonucleotide probes targeting approximately 1.2 million single nucleotide polymorphisms (SNPs)[10] as previously described.[3] Sequenced was performed and data processed in the same way as for the mitochondrial analysis with the exception that we mapped to the whole genome, using the reference sequence *hg19*.

Table 1: Results on the three St. Mary's City individuals

Library ID	skeletal code	Wet lab quality control	Skeletal element	Further details	DNA Extract used	% damage in first base	mtDNA coverage	mtDNA haplogroup	mtDNA match rate to consensus	Sex	SNPs covered out of 1.2 million targeted
S2099.E1.L1	HSMC-99-100/HSMC#1	Pass	Proximal right Tibia	Adult male	30 ul	24%	234	T2b4	99.6%	M	326,801
S2098.E1.L1	HSMC-99-200/HSMC#2	Inhibited	Distal right Tibia	Adult female	30 ul	..	..	..	..	..	..
S2098.E1.L4	HSMC-99-200/HSMC#2	Inibited	Distal right Tibia	Adult female	3 ul	..	..	..	..	..	..
S2097.E1.L1	HSMC-99-300/HSMC#3	Inhibited	Right temporal	Infant	30 ul	..	..	..	..	..	..
S2097.E1.L4	HSMC-99-300/HSMC#3	Pass	Right temporal	Infant	30 ul	19%	419	V	99.1%	M	618,843

Note: Libraries S2098.E1.L2, S2098.E1.L3, S2097.E1.L2, and S2097.E1.L3 are not listed because of reagent failures.

Results

(1) Summary of screening for each of the three individuals

- A single library was made for HSMC-99-100/HSMC#1 (the adult male). This library passed quality control in the sense that at least 24% of cytosines in the last nucleotide of sequences were misread as thymines (authentic ancient DNA is expected to have at least 10%[11]), the match rate to the mitochondrial consensus sequence is 99.6% indicating minimal contamination, and the coverage on autosomal targets is 326,801 SNPs.

- Two libraries were successfully made for HSMC-99-200/HSMC#2 (the adult female), which were both inhibited or failed amplification. We stopped after two attempts.

- Two libraries were successfully made for HSMC-99-300/HSMC#3, the second of which gave results consistent with authentic ancient DNA. A fraction 19% of cytosines in the last nucleotide of sequences were misread as thymines, the match rate to the mitochondrial consensus sequence is 99.9% indicating minimal contamination, and the coverage on autosomal targets is 618,843 SNPs.

(2) The adult male and infant are genetically male and have minimal contamination
The observed ratio of X chromosome to Y chromosome SNP coverage for the two samples that gave working ancient DNA (the adult male and the infant) were both consistent with male sex. Because males have a single X chromosome it is possible to use heterozygosity on chromosome X to quantify contamination (using the ANGSD tool).[12] There was no evidence of contamination for either individual (both Z<2 standard errors from zero). Point estimates were 1.6% contamination for the adult male and 0.7% for the infant male.

(2) Mitochondrial DNA haplogroups are different for the two individuals
Mitochondrial DNA is inherited from mother to child. The mitochondrial DNA haplogroups inferred based on the consensus sequences are T2b4 for the adult male, and V for the infant. These are both consistent with Western European ancestry. The fact that the two haplogroups are different excludes the individuals from being full siblings.

(3) Y chromosome haplogroups are consistent with being the same for the two individuals

- The Y chromosome haplogroup call for the adult male is R1b1a2a1a (supported by the P311:18248698A>G and P310:18907236A>C mutations).

- The Y chromosome haplogroup call for the infant male is a more resolved version of that for the adult male, that is, R1b1a2a1a2 (supported by the P312:22157311C->A mutation).

These two individuals belong to a haplogroup that is common in people of western European descent and so the results are consistent with but do not prove a patrilineal relationship.

(4) Nuclear analysis implies a father-son pair

Principal Component Analysis using *smartpca* in EIGENSOFT[13] and comparison to individuals genotyped on the Affymetrix Human Origins array[1] indicate that both individuals are Western European in ancestry, clustering closely with present-day English individuals (Fig. 1). We used PLINK[14] to test the relatedness of the sample pair, and find that PI_HAT is 0.4274, consistent with being first-degree relatives.

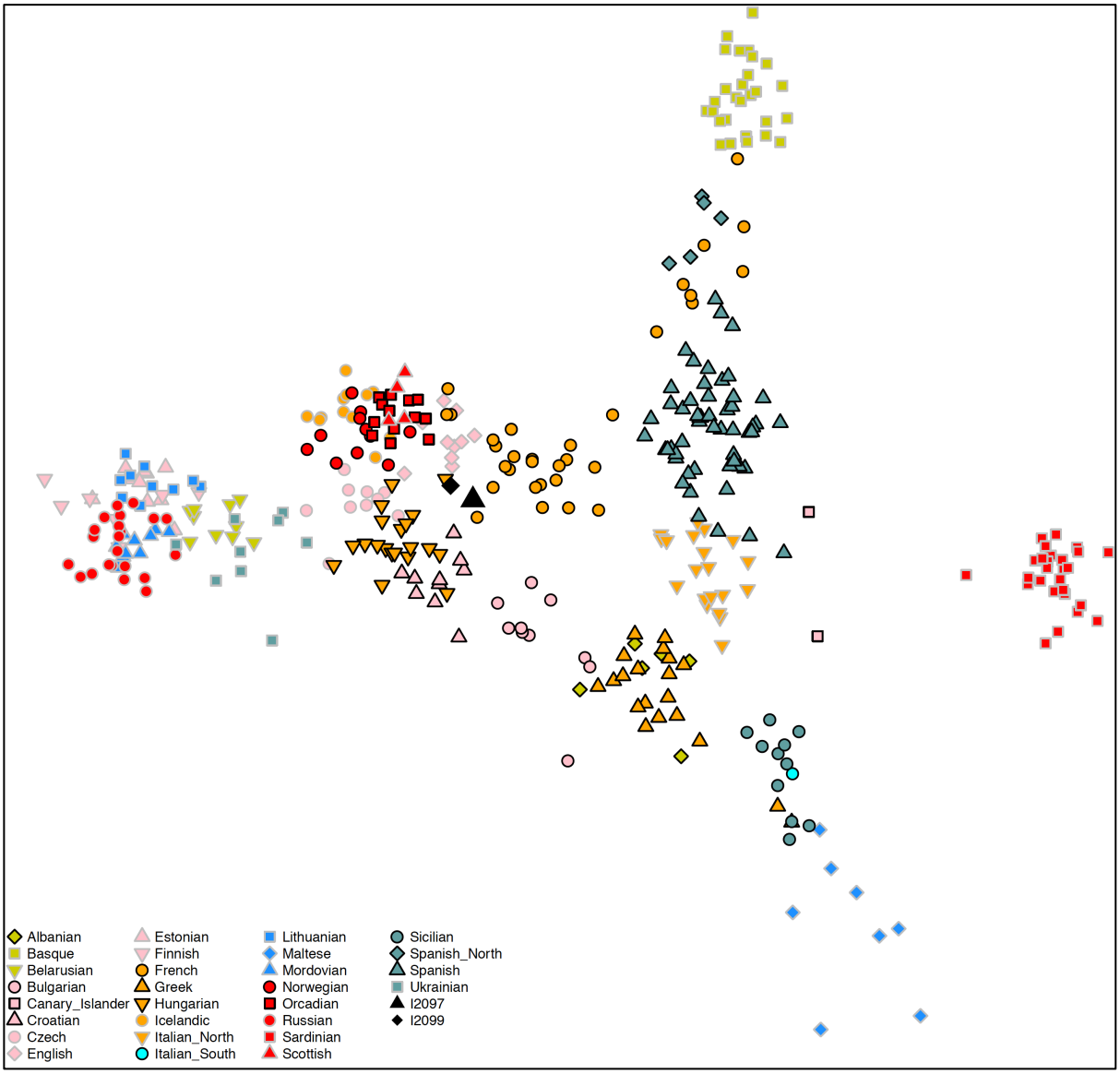

Figure 1: Principal components analysis. We projected the two ancient individuals onto the first two principal components of 392 present-day European individuals[2,15,16]. The two colonial individuals cluster closely to each other at the edge of the cluster of 10 present-day English individuals.

Conclusion

Taken together, these results suggest that HSMC-99-100/HSMC#1, the adult male identified as Philip Calvert, is a first degree relative of HSMC-99-300/HSMC#3, the unknown infant who is genetically identified as a male. The two individuals do not share the same mitochondrial DNA sequence which excludes them as full siblings. Thus, they are a father and son, consistent with their matching Y chromosome haplogroups.

References

1 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, doi:10.1038/nature14317 (2015).

2 Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature*, doi:10.1038/nature19310 (2016).

3 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, doi:10.1038/nature16152 (2015).

4 Skoglund, P. *et al.* Genomic insights into the peopling of the Southwest Pacific. *Nature*, doi:10.1038/nature19844 (2016).

5 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15758-15763, doi:10.1073/pnas.1314445110 (2013).

6 Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**, 20130624, doi:10.1098/rstb.2013.0624 (2015).

7 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

8 Behar, D. M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *American journal of human genetics* **90**, 675-684, doi:10.1016/j.ajhg.2012.03.002 (2012).

9 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

10 Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, doi:10.1038/nature14558 (2015).

11 Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PloS one* **7**, e34131, doi:10.1371/journal.pone.0034131 (2012).

12 Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics* **15**, 356, doi:10.1186/s12859-014-0356-4 (2014).

13 Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS genetics* **2**, e190, doi:10.1371/journal.pgen.0020190 (2006).

14 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).

15 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 http://www.nature.com/nature/journal/v513/n7518/abs/nature13673.html - supplementary-information (2014).

16 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).