Title

Neandertal signatures in modern human mitochondrial genome haplogroups?

Authors

Renata C. Ferreira[1⁑], Camila R. Rodrigues[2⁑], James R. Broach[3] and Marcelo R. S. Briones[4]*

Affiliations

[1]Department of Neurology and Neurosurgery, Federal University of São Paulo, São Paulo, SP, Brazil, CEP 04039032.

[2]Graduate Program in Microbiology and Immunology, Federal University of São Paulo, São Paulo, SP, Brazil, CEP 04039032.

[3]Institute for Personalized Medicine, Department of Biochemistry, Pennsylvania State University College of Medicine, 500 University Drive, 17033, Hershey PA, USA.

[4]Department of Health Informatics, Federal University of São Paulo, São Paulo, SP, Brazil, CEP 04039032.

⁑Both share first author status

*Corresponding author: Marcelo R. S. Briones

Escola Paulista de Medicina – UNIFESP, Edifício de Pesquisas II, L4E

Rua Pedro de Toledo, 669, CEP 04039-032, São Paulo, SP, Brazil. Tel: +55 11 5576-4537. E-mail: marcelo.briones@unifesp.br and briones.marcelo.rs@gmail.com

Short title: Neandertal signatures in human mtDNA

Abstract

Genetic contributions of Neandertals to the modern human genome have been evidenced by comparative analyses of present day human genomes and paleogenomes. The current data indicates that Neandertal introgression is higher in Asians and Europeans and lower in African lines of descent. Neandertal signatures in extant human genomes are attributed to intercrosses between Neandertals and ancient *Homo sapiens* lineages, or archaic Anatomically Modern Humans (AMH) that migrated from Africa into the Middle East and Europe in the last 50,000 years. It has been proposed however that there is no contribution of Neandertal mitochondrial DNA to contemporary human genomes. Here we show that the modern human mitochondrial genome might contain potential 66 Neandertal signatures, or Neandertal single nucleotide variants (N-SNVs) of which 7 are associated with traits such as cycling vomiting syndrome and Alzheimers' and Parkinsons' diseases and 2 N-SNVs associated with intelligence quotient. Principal component analysis and bootscan tests suggest rare recombination events. Also, contrary to what is observed in the nuclear genome, African mitochondrial haplogoups have more potential Neandertal signatures than Asian and European haplogroups. Based on our results we hypothesize that although most intercrosses occurred between Neandertal males and archaic AMH females, crosses between archaic AMH males and Neandertal females were extremely rare and with rare recombination events thus leaving few marks (66 out of 16,569bp) in present day mitochondrial genomes of human populations.

Introduction

Comparative analyses of present day human genomes and paleogenomes led to the proposal that Neandertals contributed to the modern makeup of the human genome (Green et al., 2010; Ko, 2016; Prüfer et al., 2014; Simonti et al., 2016). The current interpretation of genomic signatures is that Neandertal contributions are different in European, East Asian and African lines of descent, with a higher frequency of Neandertal segments in Asians and Europeans and lower frequencies in Africans (Green et al., 2010). Intercrosses between Neandertals and ancient *Homo sapiens* lineages, or archaic Anatomically Modern Humans (AMH) who migrated from Africa into the Middle East and Europe in the last 50,000 years might explain the presence of Neandertal signatures in extant human genomes (Prüfer et al., 2014; Simonti et al., 2016). The spatio-temporal overlap of Neandertals and AMH is estimated to be approximately 22,000 years since the first AMH arrived in Europe around 50,000 years ago and the last Neandertal remains (in Spain) date back to 28,000 years (Gibbons, 2010; Higham et al., 2014).

Although there is evidence for Neandertal contributions to the modern nuclear genome, it has been proposed that there is no Neandertal contribution to contemporary human mitochondrial genomes (mitogenomes) (Serre et al., 2004). Because of mitogenome matrilineal inheritance this implies that the intercrosses occurred exclusively between Neandertal males and AMH females or that crosses between AMH males and Neandertal females were extremely rare. Another possibility is that crosses between AMH males and Neandertal females produced such unfavorable trait combinations, due to mitonuclear incompatibility (Hill and Johnson, 2013), that none of their descendants left marks in present day human populations.

To address this problem we searched for mitochondrial single nucleotide variants (SNVs), or signatures, in contemporary human mitogenomes that were also present in Neandertals but not in archaic AMH mitogenomes. For this we selected archaic AMH samples with approximately the same age as Neandertal mitogenomes, thus representing, in theory, an archaic AMH ensemble that could have overlapped with Neandertals in Europe, Middle East and Central Asia. This also might alleviate the effect of homoplasies because Neanderthal and archaic AMH sequences have approximately the same age.

Results

Distribution of Neandertal SNVs

An alignment of 102 sequences was produced which contained the mitogenomes of 41 archaic AMH (Table 1), 9 *Homo sapiens neanderthalensis* (Table 1) and 52 contemporary *Homo sapiens sapiens* with representatives of the major worldwide mitochondrial haplogoups (table S1). This alignment contained the revised Cambrige Reference Sequence (rCRS) used as numbering reference for all polymorphisms identified (Andrews et al., 1999). We found 66 positions in which contemporary *Homo sapiens sapiens* were identical with at least one *Homo sapiens neanderthalensis* sequence position (Fig 1, data files S1 and S2). In 13 positions just a subset of contemporary *Homo sapiens sapiens* were identical with at least one *Homo sapiens neanderthalensis* sequence position and at least one sequence of archaic AMH. In 175 positions *Homo sapiens neanderthalensis* differed from contemporary *Homo sapiens sapiens* and archaic AMH. In 11 positions the archaic AMH differed from other sequences and in 653 positions the contemporary *Homo sapiens sapiens* presented expected variants among haplogroups that were not relevant for this analysis. The total variants are detailed in data S1.

To depict the distribution of Neandertal SNVs, or N-SNVs, in different human mitochondrial haplogroups, we constructed a heat map of N-SNVs (Fig. 2). Most N-SNVs are concentrated in the D-loop, followed by 12SrDNA and 16SrDNA. Among tRNAs N-SNVs were found only in isoleucine, asparagine and cysteine tRNAs. In *COX2* only one N-SNV was found in haplogroup R0a.

Of the 66 N-SNVs identified, 20 are common to modern African and Eurasian haplogroups, 25 are exclusive to African haplogroups and 21 are exclusive to Eurasian haplogroups. In Fig. 3 the distribution of N-SNVs is depicted in the human mitogenome map. N-SNVs exclusive to Eurasian haplogroups are represented by red bars and N-SNVs found in African haplogroups are represented by green bars.

The distribution of N-SNVs in modern haplogroups and archaic AMH can be summarized as six introgression routes with predicted consequences (Fig. 4). In hypothesis (1) SNVs conserved between Neandertals, archaic AMH and present day humans should be observed in all sequences; in (2) if the introgression crosses occurred between Neandertals and archaic AMH prior to divergence that separated archaic AMH from lineages of African and Eurasian haplogroups it is expected that N-SNVs would be present in all sequences, as observed for 13 N-SNVs. In (3) the crosses would have occurred only with archaic AMH who did not contribute to present day haplogroups while hypotheses (4), (5) and (6) would represent crosses between Neandertals and lineages who contributed to present day

haplogroups. The 66 N-SNVs are consistent with these hypotheses, which exclude their presence in archaic AMH and thus are likely signals of crosses between these two human subspecies. More importantly no SNVs are identical between Neandertals and archaic AMH at the exclusion of present day mitogenomes of either African and Eurasian haplogroups which suggests that the presence of N-SNVs must be a signal of horizontal transfer or a very high number of reverse/convergent substitutions.

Disease associated N-SNVs

Among the 66 N-SNVs, 7 are associated with diseases as depicted in Table 2. Of note 4 of these disease associated N-SNVs were observed in African haplogroups (L0, L1, L2, L3, L4, L5 and L6) and 3 were observed exclusively in Eurasian haplogroups. The most common diseases associated with N-SNVs are neurological disorders and tumors. One N-SNVs, in position 15,043 and associated with depression, was also found in one archaic AMH. Although not considered a *bona fide* N-SNV based on our exclusion criteria, it is relevant because chronic depression has been associated with Neandertal introgression (Racimo et al., 2017). Among protein coding genes an important N-SNV was found in *ND2* (position 5,460) causing an amino acid change from alanine to threonine that has been associated with Alzheimers' and Parkinsons' diseases due to its high prevalence in the brains of Alzheimers' and Parkinsons' patients (Lin et al., 1992; Petruzzella et al., 1992; Schnopp et al., 1996). This amino acid substitution changes a nonpolar amino acid to a polar amino acid, which promotes destabilizing effects in the encoded NADH dehydrogenase. Nonpolar to polar amino acid changes are associated with amyloid diseases (Alvares et al., 2013; Podoly et al., 2010; Volkenstein, 1965). Other disease associated N- SNVs are located in the D-loop, 16S rRNA and tRNA-Cys (Table 2).

Haplogroups of paleogenomes

Mitogenomes of Neandertals and archaic AMH were classified in haplogroups according to sequence similarity with extant human mitogenomes using Haplogrep 2 (Weissensteiner et al., 2016) (Table 1). 83% of archaic AMH mitogenomes belong to haplogroup U (45% haplogroup U5 and 16% to haplogroup U2) which is consistent as U being the oldest European haplogroup. Among the 9 Neandertal mitogenomes, the 7 more recent genomes can be classified as haplogroup H1 (European) while the two oldest can be classified as haplogroup L (African) (Table 2). N-SNVs at positions 16,278 and 16,298 are associated with intelligence quotient (Thomas et al., 1998). N-SNV 16,278 is found in African haplogroups (L0, L1, L2, L5 and L6) and two Eurasian haplogroups (X3, U2c and P2) and in all Neandertal sequences while N-SNV 16.298 is found only in Eurasian-Native American haplogroups (V1, V2, M8, C1, C4, C7 and Z1) and only in the Altai Neandertal.

PCAs of Neandertal and Human mitogenomes

Principal component analysis (PCA) of the whole mitochondrial genome shows four clusters: (1) the modern haplogroups including the ancient *H. sapiens* (Table 2), (2) the L haplogroup cluster, (3) the Neandertal Altai-Mezmaskaya L-like cluster (Table 2) and (4) the Neandertal H-like group (Fig. 5). However the PCA of the segment corresponding to the ribosomal RNA gene proximal half produces a pattern that approximates the L haplogroup cluster to the Neandertal Altai-Mezmaskaya L-like cluster suggesting the introgression point. The PCA of ribosomal RNA gene distal half suggests an opposite pattern with L cluster closer to Neandertal although not as close as shown in Fig. 5B.

Bootscan analysis of mitogenomes

We tested potential recombination in our dataset with bootscan. We used a different set of parental sequences depending on the query mitogenome (Fig. 6). The bootscan analysis indicates that there are potential recombination points. Upon deeper analysis we observed that bootscan considers the Neandertal specific signatures, such as in L haplogroups, as recombination points. Although the bootscan putative recombination segments are above the bootstrap threshold we do not consider this as definitive evidence of recombination since the segments between the Neandertal signatures are almost identical. Bootscan analysis did exclude Human-Neandertal recombintation in rCRS sequence (fig. S1). Sensivity of bootscan to substitution models and alignment methods was assessed by comparing the same set query-parentals with different parameters (fig. S2), revealing minor profile alterations. The alignment parameters are not so critical in this case because the sequences are extremely conserved (918 polymorphic positions in 16,565bp). Although indels are present in the alignments, 99% are located near the H promoter in the D-loop region. These are automatically excluded in phylogeny inference algorithms and therefore have no weight in bootscan results. The "positional homology" is therefore solid, particularly in coding domains and regions without repeats in non-coding domains. The Neandertal signatures are in unambiguously aligned segments.

Discussion

It can be argued that the Neandertal signatures are in fact character states conserved since the last common ancestor of Neandertals and present day *Homo sapiens* (e.g. *Homo erectus*but this would not be consistent with the absence of these signatures in ancient *H. sapiens* mitochondrial genomes (Fig. 1). Alternatively, the Neandertal signatures here described could be a consequence of random events. Because the random chance of the same nucleotide in a given position is 0.25 and there are 918 polymorphic positions in which 66 are Neandertal specific, it would roughly fit the 0.08 chance expectation. However, the chance that the 66 positions are simultaneously identical by chance alone would be roughly 0.25^{66} which is far less than observed here. Also some authors, based on populational data, propose that a single observed change in comparative mitogenomics roughly correspond to 10,000 years of divergence (Sykes, 2001). The human mitogenome clock, as estimated from ancient DNA, ranges from 2.14×10^{-8} to 2.74×10^{-8} substitutions per site per year (Fu et al., 2013, 2014; Posth et al., 2016; Rieux et al., 2014) which gives approximately 4.14 substitutions in the mitogenome every 10,000 years. These estimates suggest that for the 66 SNVs observed here to be products of homoplasies (reverse, parallel or convergent substitutions) the elapsed time would have had to be significantly higher that what is estimated for *Homo sapiens sapiens* versus *Homo sapiens neanderthalensis* divergence.

A back to Africa hypothesis has been proposed in which humans from Eurasia returned to Africa and impacted a wide range of sub-Saharan populations (Henn et al., 2012). Our data suggest that Neandertal signatures might be present in all major African haplogroups which supports the “Back to Africa” contribution to the modern mitochondrial African pool.

Our observations suggest that crosses between AMH males and Neandertal females left significantly less descendants than the reverse crosses (Neandertal males and AMH females), which seems to be the dominant pattern. Although it has been generally accepted that recombination does not occur in the human mitochondrial genome, evidence of mitochondrial

recombination has been reported (Kraytsberg et al., 2004; Schwartz and Vissing, 2002). A scenario with complete absence of recombination presents a problem to explain how the human mitochondrial genome would escape the Miller ratchet and therefore avoiding its predicted "genetic meltdown" (Felsenstein, 1974). It has been shown that even minimal recombination is sufficient to allow the escape from the Miller ratchet (Dokianakis and Ladoukakis, 2014) and this could be the case of the human mitochondrial genome.

Our data is compatible with a scenario in which the AMH-Neandertal crosses occur in Europeans, East Asians and African lines of descent. However, in the African haplogroups the crosses between AMH males and Neandertal females would have a higher frequency than in European lines of descent, where the reverse crosses would be predominant. Based on the comparison of Neandertal signatures in nuclear and mitochondrial genome haplogroups we hypothesize that the African lines of descent would have a higher female Neandertal contribution whereas European lines of descent would have higher male Neandertal contribution. The fact that AMH and Neandertals crossed and produced fertile descendants is evidence that they belong to the same species (Ko, 2016) and thus indicate that *Homo sapiens* emerged independently in Africa, Europe and Asia (Horai et al., 1995). The intercrosses of these three *Homo sapiens* subgroups, and other even deeper ancestors such as Denisovans, in its different proportions and specific signatures, would have produced the extant human genomes.

The analyses here presented suggest that Neandertal genomic signatures might have been a product of rare mtDNA recombination events. Although there is evidence supporting mtDNA recombination its weight in phylogenies remain controversial. Some authors contend that due to its high mutation rate reverse compensatory mutations can be confounded with recombination in mtDNA. Our data is consistent with a scenario in which mtDNA recombination occurred. These recombination events were probably extremely rare which might explain the observed small number of potential Neandertal signatures in modern human mitochondrial genomes.

Materials and Methods

Experimental Design

Comparative analysis of the mitochondrial DNA from present day humans (52 sequences), ancient *Homo sapiens*, or archaic AMH (41 sequences) and Neandertals (9 sequences) were aligned to identify single nucleotide variants. The 52 sequences of present day human mtDNA, representing all major mitochondrial haplogroups (table S1), were selected from the PhyloTREE database (Oven and Kayser, 2009) and downloaded from GenBank. Paleogenomes mtDNA sequences were downloaded from GenBank for 42 archaic AMH (*H. sapiens*) and 9 Neandertal mtDNA (Table 1). Neandertal signatures, or N-SNVs, were determined as positions that were identical between present day sequences and Neandertal sequences at the exclusion of archaic AMH.

Sequence assembly and alignments

The Ust-Ishim sequence was assembled using reads downloaded from Study PRJEB6622 at the European Nucleotide Archive (EMBL-EBI) and assembled using the CLC Genomics Workbench 7 program (https://www.qiagenbioinformatics.com). To maintain the reference

numbering, sequences were aligned to the revised Cambridge Reference Sequence (rRCS; GenBank accession number NC012920) (Andrews et al., 1999), totalizing 103 sequences using the map to reference option implemented in Geneious 10 program (Kearse et al., 2012). Variants were called using Geneious 10 program. A total of 918 polymorphic positions were found. Neandertals, ancient and modern humans were screened for disease associations at MitoMap (http://www.mitomap.org/MITOMAP) (Lott et al., 2013).

Phylogenetic inference

Position specific similarities between modern haplogroups and Neandertals were depicted by cladograms for each of the single 66 variant positions present only in Neandertals and modern humans and excluding archaic AMH, were generated using parsimony heuristic search implemented in PAUP v4.1a152 with the default parameters (Swofford, 2001). Proximity of mitochondrial haplogoups in Ancient *H. sapiens* and Neandertals were inferred using Haplogrep 2.1.0 (Weissensteiner et al., 2016). All 66 cladograms, corresponding to each N-SNV, are available upon request.

Recombination analysis

Potential recombination between Neandertals and ancient *H. sapiens* sequences was inferred by a phylogenetic based method implemented by manual bootscan in the Recombination Detection Program (RDP) v.4.87. Parameters for bootscan analysis were: window size = 200; step size = 20; bootstrap replicates = 1,000; cutoff percentage = 70; use neighbor joining trees; calculate binomial p-value; model option = Kimura 1980 (Kimura, 1980). For each analysis, a single alignment was created which included the modern haplogroup, all 9 Neandertal and all 6 Ancient *H. sapiens* sequences. When rCRS was used as query, two sets of possible parental sequences were selected: either Neandertals Mezmaiskaya and Altai and ancient *H. sapiens* Fumane and Ust Ishim or only Neandertals Feldhofer1, Mezmaiskaya and Vindija 33.16. For haplogroups L0d1a and L3d3b possible parental sequences were Neandertals Feldhofer1, Mezmaiskaya and Vindija 33.16 and Ancient *H. sapiens* Kostenki 14, Fumane, Doni Vestonice 14 and Tianyuan. For haplogroups M29a and R0a possible parental sequences were Neandertals Mezmaiskaya and Altai and Ancient *H. sapiens* Kostenki 14 and Doni Vestonice 14. For haplogroup N1b1a3 possible parental sequences were Neandertals Feldhofer1 and Vindija 33.16 and Ancient *H. sapiens* Kostenki 14 and Doni Vestonice 14.

Statistical analysis

For variants (SNVs) calling three different datasets were used: (1) the whole mitogenome from the 102 sequences alignment; (2) the 128 to 315bp fragment and (3) the 6,950 to 7,660 fragment of the same alignment. All fasta alignments were processed using the MSA2VCF software to generate the VCF files (Lindenbaum, 2015). The options used on msa2vcf were: --haploid --output. To convert the VCF files to Plink format we used the vcftools package (Danecek et al., 2011). Whole mitogenome alignment with 103 sequences had 785 SNPs (positions containing gaps in at least one sequence were excluded from the analysis). Both 128 – 315 and 6,950-7,660 fragments had 24 SNPs.

Principal component analysis was performed using the PLINK software v1.90b4 (Purcell et al., 2007). PCA figure plotting was made using Genesis PCA and admixture plot viewer (http://www.bioinf.wits.ac.za/software/genesis/). The first two principal components were chosen for the Neandertal - *H. sapiens* comparison.

References

Alvares, R. D. A., Tulumello, D. V., Macdonald, P. M., Deber, C. M., and Prosser, R. S. (2013). Effects of a polar amino acid substitution on helix formation and aggregate size along the detergent-induced peptide folding pathway. *Biochim. Biophys. Acta BBA - Biomembr.* 1828, 373–381. doi:10.1016/j.bbamem.2012.09.024.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23, 147. doi:10.1038/13779.

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., et al. (2009). Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science* 325, 318–321. doi:10.1126/science.1174462.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330.

Dokianakis, E., and Ladoukakis, E. D. (2014). Different degree of paternal mtDNA leakage between male and female progeny in interspecific Drosophila crosses. *Ecol. Evol.* 4, 2633–2641. doi:10.1002/ece3.1069.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* 78, 737–756.

Formicola, V., Pontrandolfi, A., and Svoboda, J. (2001). The Upper Paleolithic triple burial of Dolní Věstonice: Pathology and funerary behavior. *Am. J. Phys. Anthropol.* 115, 372–379. doi:10.1002/ajpa.1093.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514, 445–449. doi:10.1038/nature13810.

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., et al. (2013). A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Curr. Biol.* 23, 553–559. doi:10.1016/j.cub.2013.02.044.

Gansauge, M.-T., and Meyer, M. (2014). Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.*, gr.174201.114. doi:10.1101/gr.174201.114.

Gibbons, A. (2010). Close Encounters of the Prehistoric Kind. *Science* 328, 680–684. doi:10.1126/science.328.5979.680.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A Draft Sequence of the Neandertal Genome. *Science* 328, 710–722. doi:10.1126/science.1188021.

Green, R. E., Malaspinas, A.-S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., et al. (2008). A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. *Cell* 134, 416–426. doi:10.1016/j.cell.2008.06.021.

Henn, B. M., Botigué, L. R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J. K., et al. (2012). Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLOS Genet.* 8, e1002397. doi:10.1371/journal.pgen.1002397.

Hervella, M., Svensson, E. M., Alberdi, A., Günther, T., Izagirre, N., Munters, A. R., et al. (2016). The mitogenome of a 35,000-year-old *Homo sapiens* from Europe supports a Palaeolithic back-migration to Africa. *Sci. Rep.* 6, 25501. doi:10.1038/srep25501.

Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., et al. (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* 512, 306–309. doi:10.1038/nature13621.

Hill, G. E., and Johnson, J. D. (2013). The mitonuclear compatibility hypothesis of sexual selection. *Proc R Soc B* 280, 20131314. doi:10.1098/rspb.2013.1314.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K., and Takahata, N. (1995). Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. U. S. A.* 92, 532–536.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi:10.1093/bioinformatics/bts199.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120.

Ko, K. H. (2016). Hominin interbreeding and the evolution of human variation. *J. Biol. Res.-Thessalon.* 23, 17. doi:10.1186/s40709-016-0054-7.

Kraytsberg, Y., Schwartz, M., Brown, T. A., Ebralidse, K., Kunz, W. S., Clayton, D. A., et al. (2004). Recombination of Human Mitochondrial DNA. *Science* 304, 981–981. doi:10.1126/science.1096342.

Lin, F.-H., Lin, R., Wisniewski, H. M., Hwang, Y.-W., Grundke-Iqbal, I., Healy-Louie, G., et al. (1992). Detection of point mutations in codon 331 of mitochondrial NADH dehydrogenase subunit 2 in alzheimer's brains. *Biochem. Biophys. Res.*

Commun. 182, 238–246. doi:10.1016/S0006-291X(05)80136-6.

Lindenbaum, P. (2015). JVarkit: java-based utilities for Bioinformatics. doi:10.6084/m9.figshare.1425030.v1.

Lott, M. T., Leipzig, J. N., Derbeneva, O., Xie, H. M., Chalkia, D., Sarmady, M., et al. (2013). mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr. Protoc. Bioinforma.* 44, 1.23.1-1.23.26. doi:10.1002/0471250953.bi0123s44.

Oven, M. van, and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30, E386–E394. doi:10.1002/humu.20921.

Petruzzella, V., Chen, X., and Schon, E. A. (1992). Is a point mutation in the mitochondrial ND2 gene associated with alzheimer's disease? *Biochem. Biophys. Res. Commun.* 186, 491–497. doi:10.1016/S0006-291X(05)80834-4.

Podoly, E., Hanin, G., and Soreq, H. (2010). Alanine-to-threonine substitutions and amyloid diseases: Butyrylcholinesterase as a case study. *Chem. Biol. Interact.* 187, 64–71. doi:10.1016/j.cbi.2010.01.003.

Posth, C., Renaud, G., Mittnik, A., Drucker, D. G., Rougier, H., Cupillard, C., et al. (2016). Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* 26, 827–833. doi:10.1016/j.cub.2016.01.037.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43–49. doi:10.1038/nature12886.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795.

Racimo, F., Marnetto, D., and Huerta-Sánchez, E. (2017). Signatures of Archaic Adaptive Introgression in Present-Day Human Populations. *Mol. Biol. Evol.* 34, 296–317. doi:10.1093/molbev/msw216.

Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L. A., Warmuth, V., et al. (2014). Improved Calibration of the Human Mitochondrial Clock Using Ancient Genomes. *Mol. Biol. Evol.* 31, 2780–2792. doi:10.1093/molbev/msu222.

Schnopp, N. M., Kösel, S., Egensperger, R., and Graeber, M. B. (1996). Regional heterogeneity of mtDNA heteroplasmy in parkinsonian brain. *Clin. Neuropathol.* 15, 348–352.

Schwartz, M., and Vissing, J. (2002). Paternal Inheritance of Mitochondrial DNA. *N.*

Engl. J. Med. 347, 576–580. doi:10.1056/NEJMoa020350.

Serre, D., Langaney, A., Chech, M., Teschler-Nicola, M., Paunovic, M., Mennecier, P., et al. (2004). No Evidence of Neandertal mtDNA Contribution to Early Modern Humans. *PLOS Biol.* 2, e57. doi:10.1371/journal.pbio.0020057.

Simonti, C. N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D. S., Chisholm, R. L., et al. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. *Science* 351, 737–741. doi:10.1126/science.aad2149.

Swofford, D. L. (2001). *PAUP*: Phylogenetic Analysis Using Parsimony (and other methods) 4.0.b5*.

Sykes, B. (2001). *The Seven Daughters of Eve | W. W. Norton & Company, New York, NY.* Available at: http://books.wwnorton.com/books/detail.aspx?ID=4893 [Accessed May 17, 2018].

Thomas, M. G., Miller, K. W. P., and Mascie-Taylor, C. G. N. (1998). Mitochondrial DNA and IQ in Europe. *Intelligence* 26, 167–173. doi:10.1016/S0160-2896(99)80060-9.

Volkenstein, M. V. (1965). Coding of Polar and Non-polar Amino-acids. *Nature* 207, 294–295. doi:10.1038/207294a0.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., et al. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44, W58–W63. doi:10.1093/nar/gkw233.

Acknowledgments

The authors thank Prof. Dajiang Liu (Dept. Public Health Sciences, Penn State College of Medicine) and Prof. Rongling Wu (Center for Statistical Genetics, Penn State University) for critical reading of the manuscript. This work was supported by grants to M.R.S.B. from FAPESP, Brazil (2013/07838-0 and 2014/25602-6) and CNPq, Brazil (303905/2013- 1). R.C.F received a CNPq postdoctoral fellowship (206445/2014-8) and C.R.R. a CAPES (Brazil) MSc fellowship. **Author contributions:** R.C.F. and C.R.R. planned and performed analyses. M.R.S.B. performed preliminary analysis and wrote the manuscript. R.C.F., C.R.R., J.R.B and M.R.S.B. discussed data and analyses and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data and files of analyses here presented are available upon request to the corresponding author.

Table 1. Archaic AMH and Neandertal mitogenomes used in this study. Haplogroup inference quality as calculated by Haplogrep 2 (Weissensteiner et al., 2016).**

Archaic AMH	Haplogroup	Quality** (%)	Age (years)	Location	Reference	GenBank/ENA*
Berry Au Bac 1	U5b1a	98.78	7,160-7,319	France	(Posth et al., 2016)	KU534977
Bockstein	U5b1d1	98.72	8,016-8,329	Germany	(Posth et al., 2016)	KU534973
Cuiry Les Chaudardes 1	U5b1b	97.01	8,050-8,360	France	(Posth et al., 2016)	KU534975
Ofnet	U5b1d1	98.72	8,159-8,424	Germany	(Posth et al., 2016)	KU534974
Felsdach	U5a2c	97.43	8,380-8,980	Germany	(Posth et al., 2016)	KU534954
Hohlenstein Stadel	U5b2c1	94.77	8,446-8,809	Germany	(Posth et al., 2016)	KU534979
Falkenstein	U5b2a	92.19	8,993-9,409	Germany	(Posth et al., 2016)	KU534980
Mareuil Les Meaux 1	U5a2+16362	100	9,080-9,500	France	(Posth et al., 2016)	KU534959
Les Closeaux 3	U5a2	98.27	9,580-10,230	France	(Posth et al., 2016)	KU534958
Ranchot 88	U5b1	95.95	9,933-10,235	France	(Posth et al., 2016)	KU534978
Iboussieres 31-2	U5b1+16189	99.01	10,140	France	(Posth et al., 2016)	KU534976
Iboussieres 25-1	U5b2a	93.92	10,140	France	(Posth et al., 2016)	KU534981
Iboussieres 39	U5b2b	92.40	11,600-12,040	France	(Posth et al., 2016)	KU534972
Hohle Fels 10	U8a	96.92	12,700	Germany	(Posth et al., 2016)	KU534961
Rochedane	U5b2b	97.01	12,830-13,090	France	(Posth et al., 2016)	KU534971
Burkhardtshohle	U8a	96.92	14,150-15,080	Germany	(Posth et al., 2016)	KU534960
Hohle Fels 79	U8a	96.92	14,270-15,070	Germany	(Posth et al., 2016)	KU534962
Brillenhohle	U8a	97.66	14,400-15,120	Germany	(Posth et al., 2016)	KU534947
Oberkassel 998	U5b1+	99.45	14,000	Germany	(Hervella et al., 2016)	KC521457
Goyet Q-2	U8a	96.92	14,780-15,230	Belgium	(Posth et al., 2016)	KU534963
Rigney 1	U2'3'4'7'8'9	87.68	15,240-15,690	France	(Posth et al., 2016)	KU534957
Hohle Fels 49	U8a	96.92	15,568-16,250	Germany	(Posth et al., 2016)	KU534964
Paglicci 71	U5b2b	97.97	18,197-18,973	Italy	(Posth et al., 2016)	KU534950
Dolni Vestonice 43	U5	97.44	25,000	Czech Rep.	(Posth et al., 2016)	KU534970
Goyet Q56-16	U2	91.40	26,040-26,600	Belgium	(Posth et al., 2016)	KU534965
Goyet 2878-21	U5	92.05	26,269-27,055	Belgium	(Posth et al., 2016)	KU534955
Goyet Q376-19	U2	89.43	27,310-27,720	Belgium	(Posth et al., 2016)	KU534967
Goyet Q55-2	U2	85.86	27,310-27,730	Belgium	(Posth et al., 2016)	KU534948
La Rochette	M	92.38	27,400-27,784	France	(Posth et al., 2016)	KU534951
Goyet Q53-1	U2	93.23	27,720-28,230	Belgium	(Posth et al., 2016)	KU534966
Paglicci 108	U2'3'4'7'8'9	92.40	27,831-28,961	Italy	(Posth et al., 2016)	KU534968
Paglicci 133	U8c	99.41	28,000-29,000	Italy	(Posth et al., 2016)	KU534956
Dolni Vestonice 16	U5	97.94	29,386-30,567	Czech Rep.	(Posth et al., 2016)	KU534949
Doni Vestonice 14	U5	100	31,000	Czech Rep.	(Formicola et al., 2001)	KC521458
Cioclovina 1	U	96.72	32,519-33,905	Romania	(Posth et al., 2016)	KU534969
Goyet Q376-3	M	88.67	33,140-33940	Belgium	(Posth et al., 2016)	KU534953
Goyet Q116-1	M	93.64	34,430-35,160	Belgium	(Posth et al., 2016)	KU534952
Kostenki 14	U2	92.69	36-39,000	Russia	(Hervella et al., 2016)	FN600416*
Fumane 2	R	94.43	39-41,000	Italy	(Hervella et al., 2016)	KP718913
Tianyuan	B4’5	91.17	40,000	China	(Fu et al., 2013)	KC417443
Ust-Ishim	R	87.89	45,000	West Siberia	(Hervella et al., 2016)	PRJEB6622*
Neandertals						
Vindija 33.16	H1e	52.74	38,000	Croatia	(Green et al., 2008)	AM948965
Vindija 33.17	H1e	52.75	Not dated	Croatia	(Gansauge and Meyer, 2014)	KJ533544
Vindija 33.19	H1e	52.74	Not dated	Croatia	(Gansauge and Meyer, 2014)	KJ533545
Vindija 33.25	H1as	52.84	Not dated	Croatia	(Briggs et al., 2009)	FM865410
El Sidron 1253	H1e	52.81	39,000	Spain	(Briggs et al., 2009; Hervella et al., 2016)	FM865409
Feldhofer 1	H1as	52.84	40,000	Germany	(Briggs et al., 2009)	FM865407
Feldhofer 2	H1e	52.81	40,000	Germany	(Briggs et al., 2009)	FM865408
Altai Neandertal	L1’2’3’4’5’6	55.96	50,000	Siberia	(Prüfer et al., 2014)	KC879692
Mezmaskaya 1	L1’2’3’4’5’6	52.12	65,000	Russia	(Briggs et al., 2009)	FM865411

Table 2. Disease associated Neandertal SNVs. Associations as compiled and summarized in MITOMAP (Lott et al., 2013).

Position	**195**	**3,010**	**5,460**	**5,821**	**16,093**	**16,183**	**16,519**
Region / gene	D-loop	16S rRNA	*ND2*	tRNA-Cys	D-loop	D-loop	D-loop
DNA change	T>C	G>A	G>A	G>A	T>C	A>C	T>C
Type	Transition	Transition	Transition	Transition	Transition	Transversion	Transition
Codon position	-	-	1	-	-	-	-
Codon effect	-	-	Non Syn.	-	-	-	-
Codon change	-	-	GCC>ACC	-	-	-	-
Protein change	-	-	Ala>Thr	-	-	-	-
Disease associations	Bipolar disorder / melanoma patients	Cyclic Vomiting Syndrome with Migraine	Alzheimer`s disease / Parkinson`s disease	Deafness helper mutation	Cyclic Vomiting Syndrome	Melanoma patients	Cyclic Vomiting Syndrome with Migraine /metastasis / glioblastoma, gastric, lung, ovarian, prostate tumors
Modern haplogroups	L2a1f, L4a1, L5a1a, L6a, W1, K, J2a2a, M2b	H1a1, J1c, D4	L0a1b1, L4a1, W1, Q1	C7	L4a1, R0a, K1, H3, T1, A, C1a	X1a, X3, U1a1d, B2, M2b, M29a, O	L1c3a, L2a1f, L6a, H15, H1a1, H3, I1, K, K1, K2a2a, R0a1a3, R1a, U2c, V1a, W1, X1a, X3, B2, C4, C7, E1, F1a1, G1a1, M20, M29a, M2b, M3b, N1b1a3, N2a, O, P2
Neandertal	Altai	Altai El Sidron 1253 Feldhofer 1 Feldhofer 2 Mezmaiskaya1 Vindija 33.16 Vindija 33.17 Vindija 33.19 Vindija 33.25	Altai El Sidron 1253 Feldhofer 1 Feldhofer 2 Mezmaiskaya1 Vindija 33.16 Vindija 33.17 Vindija 33.19 Vindija 33.25	Altai El Sidron 1253 Feldhofer 1 Feldhofer 2 Mezmaiskaya1 Vindija 33.16 Vindija 33.17 Vindija 33.19 Vindija 33.25	Feldhofer1 Vin. 33.25	El Sidron 1253 Feldhofer 1 Feldhofer 2 Mezmaiskaya1 Vindija 33.16 Vindija 33.17 Vindija 33.19 Vindija 33.25	Altai El Sidron 1253 Feldhofer 1 Feldhofer 2 Mezmaiskaya1 Vindija 33.16 Vindija 33.17 Vindija 33.19 Vindija 33.25

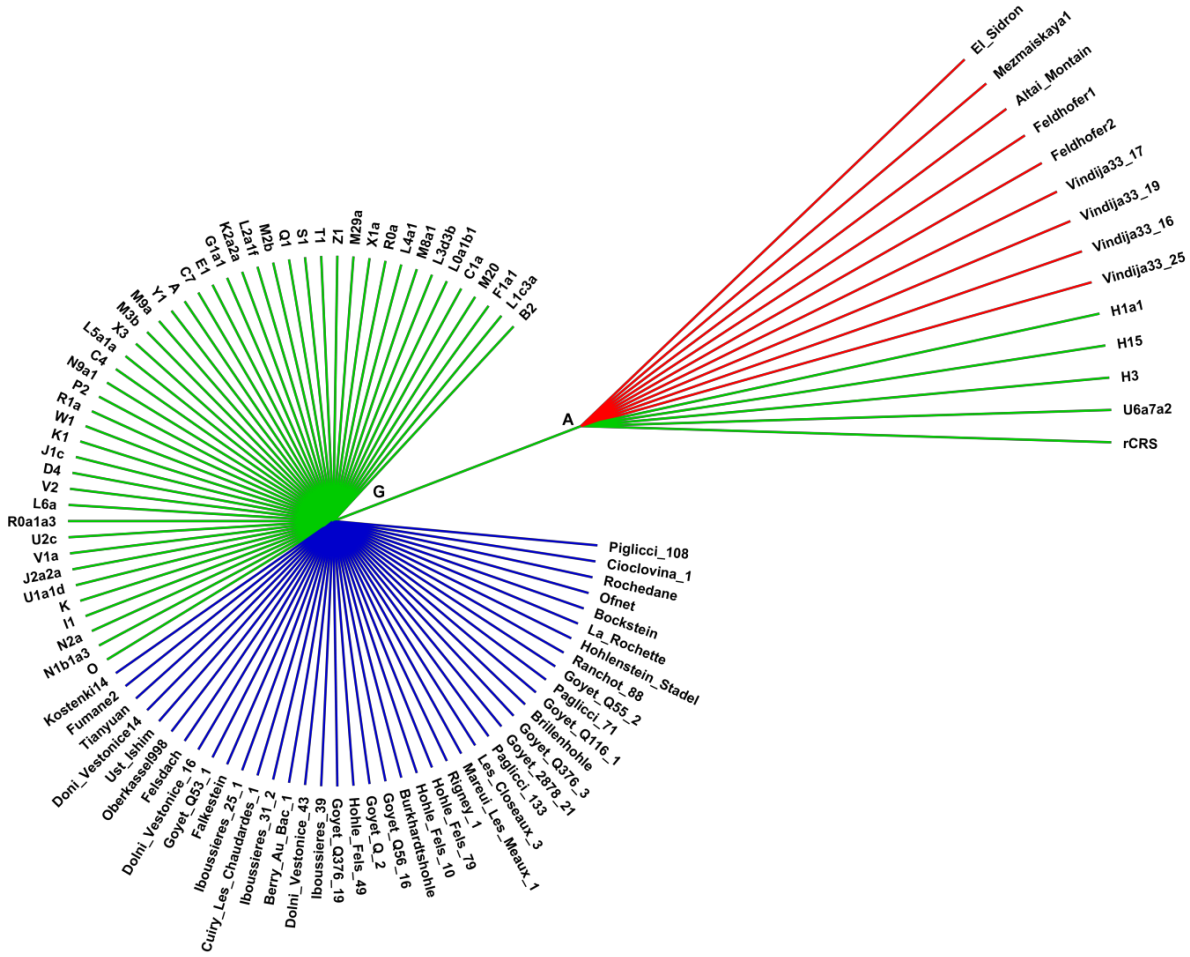

Fig. 1. Cladogram of mitochondrial genome position 2,706. This N-SNV is a G-A transition in the 16S rRNA gene. This genetic signature (Neandertal signature 2706G) is present in all Neandertal sequences (Red branches) and in European haplogroups U (U5a7a2) and H (H1a1, H3, H15), including the Revised Cambridge Reference Sequence (rCRS, haplogroup H2a2a1). The Neandertal signature 2706G is absent in all of the archaic AMH (Anatomically Modern Humans from Europe in temporal overlap with Neandertals - Blue branches) and other modern mitochondrial haplogroups. Position numbering corresponds to rCRS positions. This cladogram is representative of the 66 cladograms generated for each N-SNV giving similar topologies and which are available upon request.

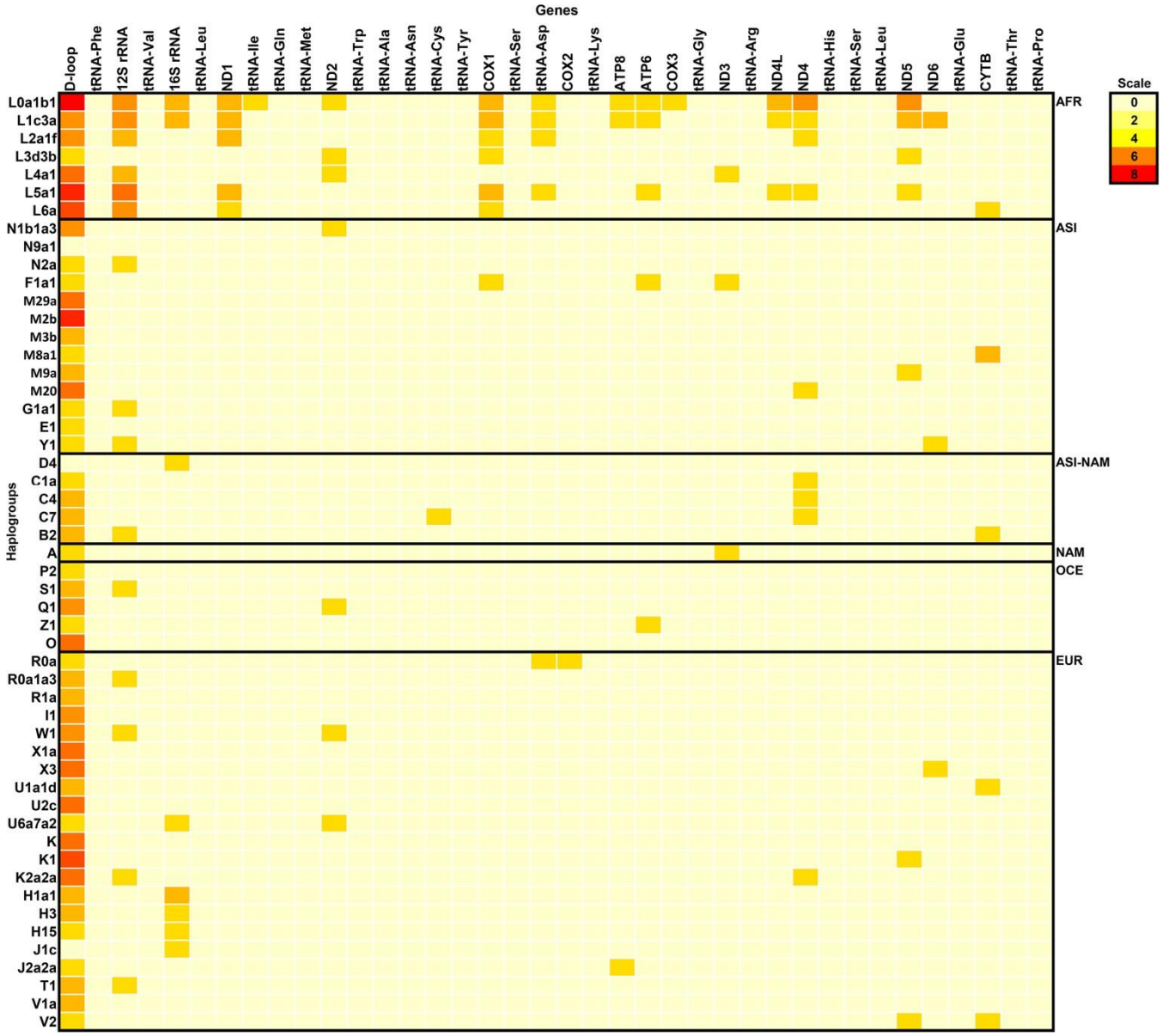

Fig. 2. Heatmap of Neandertal signatures along the mitochondrial genome in different haplogroups. The color scale indicates the number of Neandertal signatures (N-SNVs) present in modern human mitochondrial haplogoups. These signatures are absent in Ancient *H. sapiens* whose time range overlapped with Neandertals in Europe (from approximately 50,000 to 28,000 years ago, Supplementary Table S2). AFR=African, ASI=Asian, MDE=Middle East, NAM=Native American, OCE=Oceania and EUR=European.

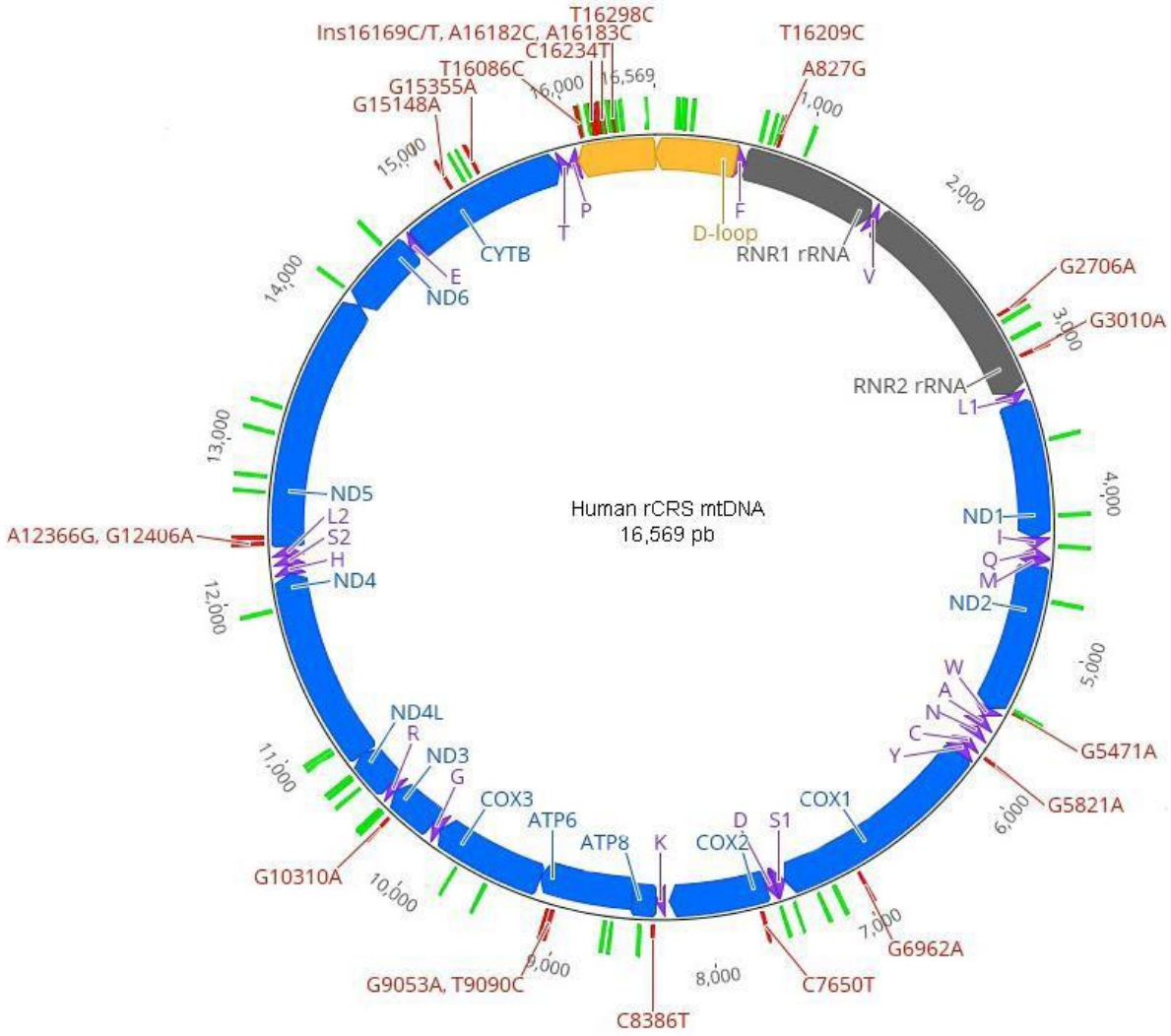

Fig. 3. Position of Neandertal SNVs (N-SNVs) in the mitochondrial genome. Blue bars indicate protein coding genes, gray bars indicate ribosomal RNA genes, purple arrows indicate tRNAs, yellow bars indicate the D-loop, green bars indicate Neandertal SNVs present in African haplogoups and red bars indicate N-SNVs exclusive of Eurasian haplogroups.

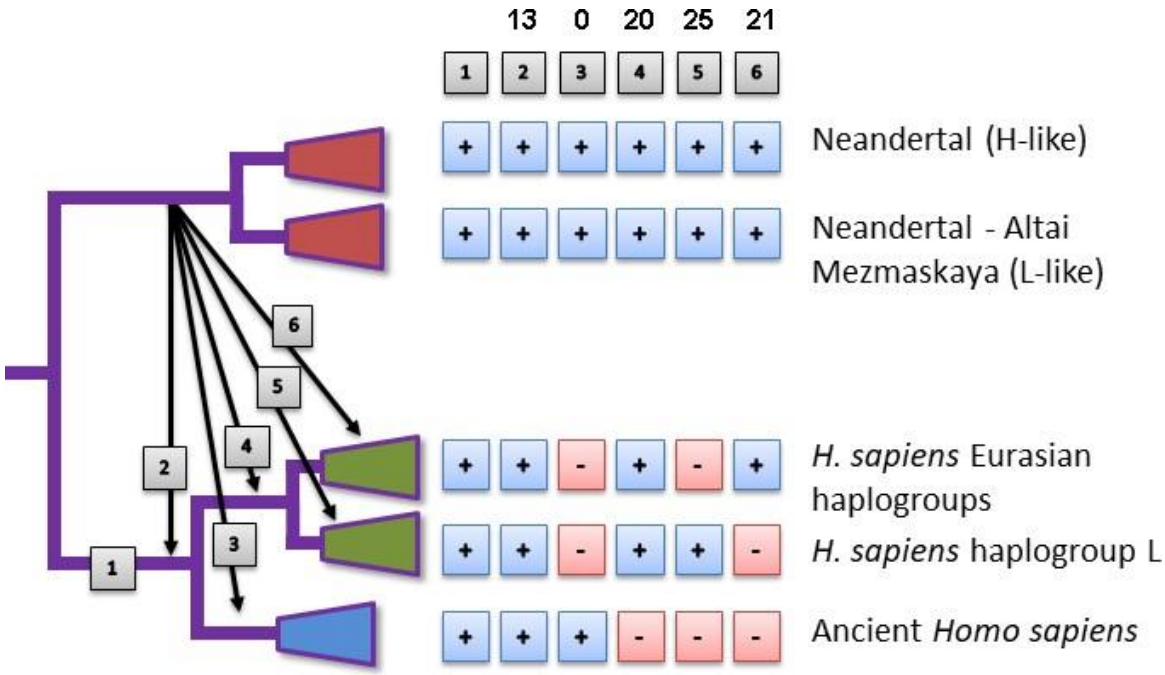

Fig. 4. Possible Neandertal introgression routes and mitogenome SNVs distribuitons. The diagram depicts six possible introgression routes, or intercross events, between Neandertals, archaic AMH and lineages leading to present day Homo sapiens haplogroups. Hypotheses 1 and 2 would produce plesiomorphic SNVs while hypothesis 3 would exclude present day haplogroups. Hyptheses 4, 5 and 6 would produce patterns in present day haplogroups consistent with more recent crosses between Neandertals and ancestors of present day haplogroups.

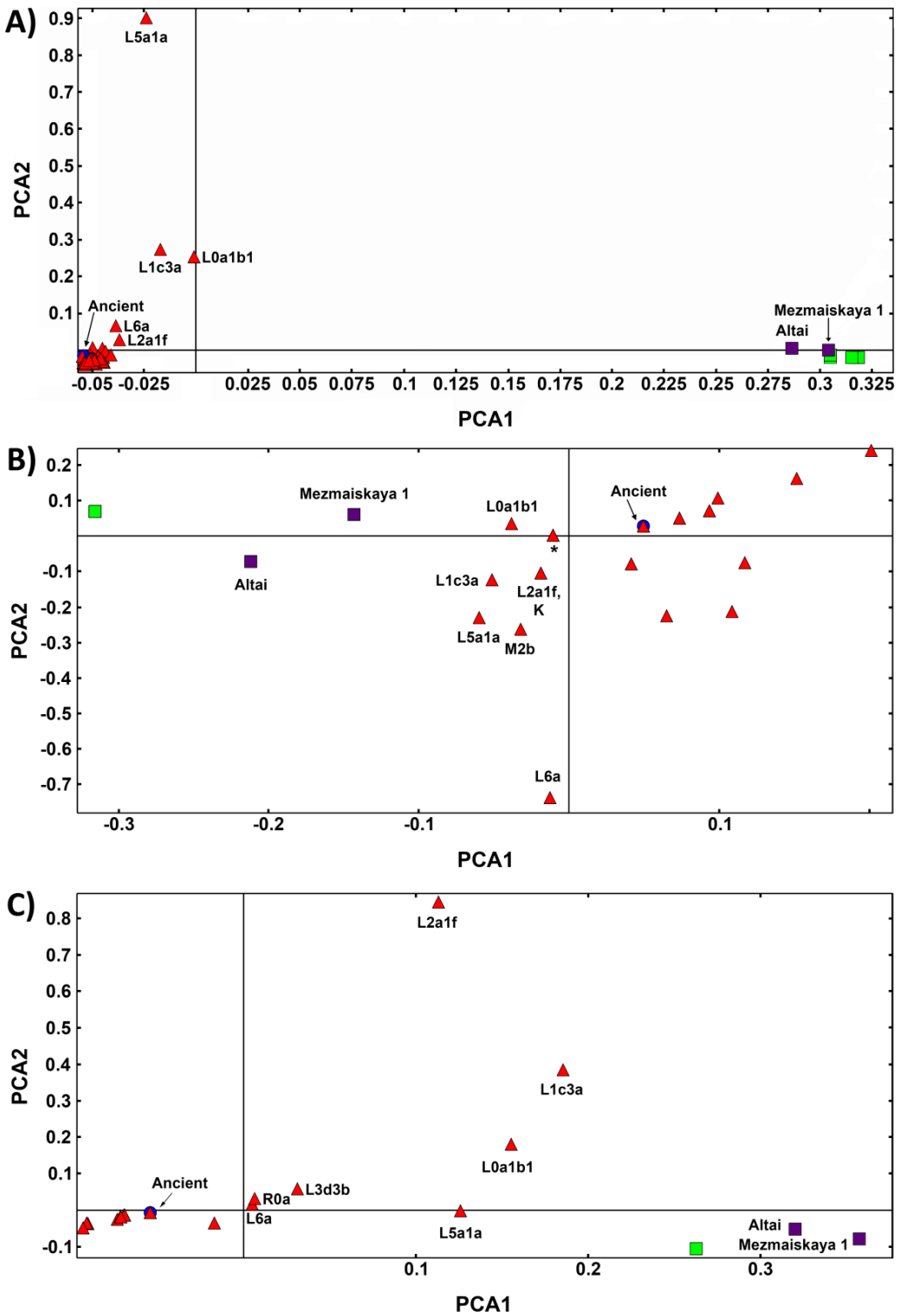

Fig. 5. Principal Component Analysis (PCA) of Human and Neandertal mitochondrial genomes. In (A) PCA results for all 16,569 positions of 103 mitogenomes being 9 Neandertals (green squares for H-like sequences and purple squares for L-like sequences), 42 archaic AMH (blue circles) and 53 modern *H. sapiens* haplogroups (red triangles). X-Axis denotes the value for PC1, while y-Axis denotes values for PC2. Each dot in the figure represents one or more individuals. In (B) PCA results for the segment comprising positions 128 to 315 extracted from the 103 mitogenomes alignment and in (C) PCA results for the segment comprising positions 6,950 to 7,660 extracted from the 103 mitogenomes alignment.

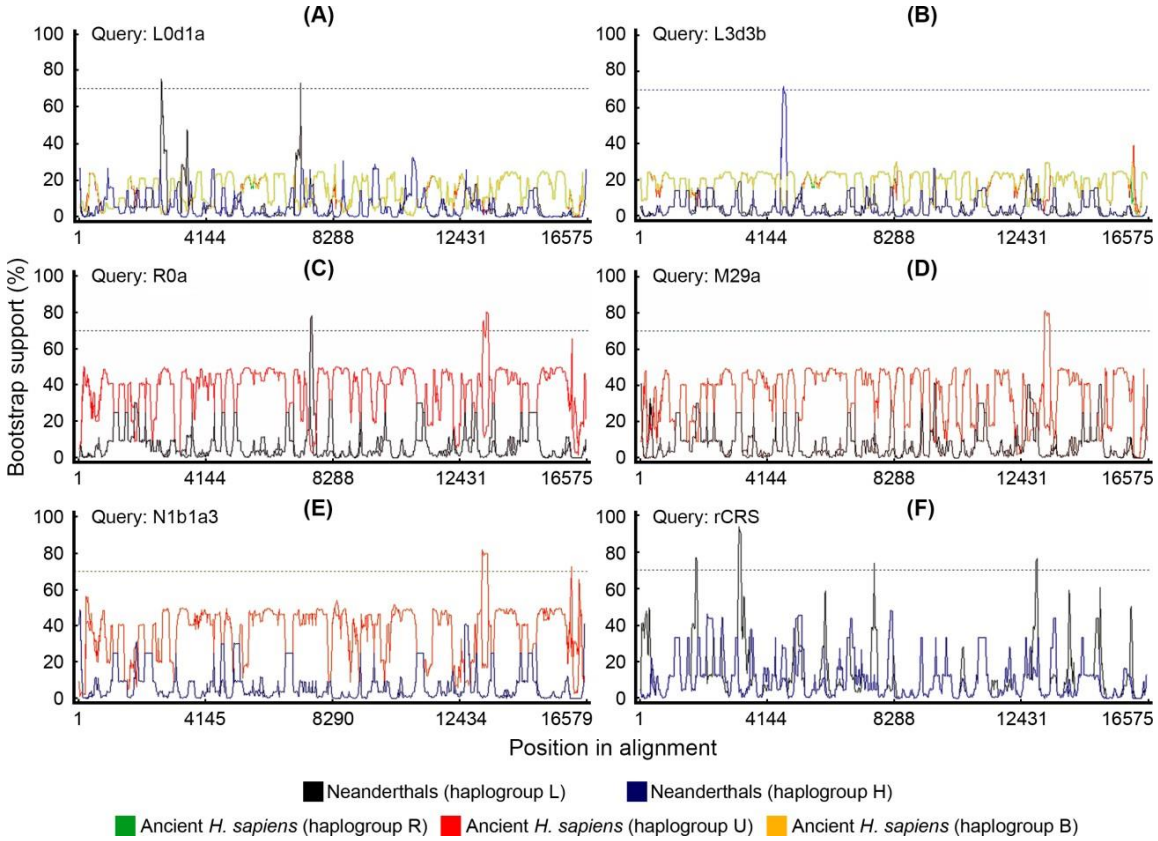

Fig. 6. Bootscan recombination analysis. Possible recombination points in were detected using different modern haplogroups as queries and Ancient *H. sapiens* or Neandertals as putative parental sequences.