Supporting Information

Skoglund et al. 10.1073/pnas.1318934111

SI Materials and Methods

Postmortem Degradation Score Distributions in Published Datasets. We obtained sequence data from three previous studies (1–3). We remapped the data which originated from HiFi polymerase amplification by Skoglund et al. (3) using Burrows-Wheeler Aligner (BWA) 0.5.9 (4) with the seed region disabled (parameters: −l 16500, −n 0.01, −o 2) and formed consensus sequences of reads with identical outer mapping coordinates (5). We used the calmd program in the samtools (6) suite to recompute the MD field (containing alignment information, such as mismatches) for all datasets. To obtain postmortem degradation score (PMDS) distributions, we excluded positions with base quality < 20 and clipped alignments and alignments with gaps (these filters were used regardless of whether PMDS thresholds were applied). For Neandertal data mapped with ANFO (1), we required a mapping quality of 90, whereas we required a mapping quality of 30 for all other datasets.

We mapped two previously obtained contaminated Neandertal sets (7) to the rCRS using BWA with the same parameters as above. Because the data were obtained using a 454 technology that did not produce base quality scores, we assumed a base quality score of 20 at all positions. To estimate contamination, we used 7 fixed transversions as well as 70 fixed transitions between Neandertal and modern human mitochondria [excluding positions where Neandertals had Thymine (T) and modern humans had Cytosine (C) and where Neandertals had Adenine (A) and modern humans had Guanine (G)]. This procedure was to increase the number of informative positions for these low-coverage datasets; for higher-coverage datasets, we restricted the analysis to only seven transversions (see below).

In Silico Contamination Experiments. In the mtDNA contamination experiment, we did not exclude any reads because of mapping quality but required a base quality of at least 20 for PMDS computation and a depth at each site of at least 20 sequence reads. We estimated modern contamination in the Vindija 33.16 mtDNA by identifying seven transversion polymorphisms that are fixed between 6 complete Neandertal mtDNAs (8, 9) and the mtDNAs of 311 modern humans (10, 11).

For the autosomal analysis, we pooled sequence data from three Neolithic hunter-gatherers, randomly choosing a single sequence read at positions where more than one aligned sequence was present, and we successively added more contaminating sequences from a present day French individual (20%, 50%, and 90%). We extracted SNPs using sequences with mapping quality of at least 30 and bases with base quality of at least 30 both with and without filtering for PMDS ≥ 3 (positions with base quality < 30 were excluded from PMDS computation). The number of SNPs ranged from 46,492 (no contamination or PMDS threshold) to 464,920 (90% contamination and no PMDS threshold). We used all SNPs to perform principal component analysis using EIGENSOFT 4.0 (12) with a reference dataset of 504 individuals from Europe and the Levant (13–15), which was haploidized (a single allele was randomly chosen from each diploid individual) like in ref. 16 before analysis. The French individual who was the source of the sequence data used for artificial contamination was excluded from the principal component analysis. One SNP from each pair with r^2 value ≥ 0.2 was excluded from the principal component analysis. We performed Procrustes transformation to the principal component 1–principal component 2 configuration obtained using only the reference dataset like in ref. 3.

Capture and Sequencing of Okladnikov 2 mtDNA. DNA was extracted from about 200 mg bone from a humerus shaft found in Okladnikov Cave in the Altai Mountains in the 1980s as described (17), with one modification: the DNA elution from the silica particles was done using TE (Tris-EDTA) buffer with 0.05% Tween 20. Siliconized tubes were used for long-term storage of the DNA extract. For the sampling of the bone, dentist drill bits were used together with an NSK E-Max Micromotor System. Before mtDNA enrichment, 25 μL DNA extract were turned into a sequencing library using a modified 454 library preparation protocol as described (9, 18), with the exception that Illumina p5 and p7 adapters were used as described elsewhere (19). The p7 adapter was modified to avoid contamination from other libraries carrying a unique 7-bp barcode.

The primer extension capture (PEC) mtDNA enrichment protocol was performed as described previously (9) using the exact same four 144-plex and 143-plex mixes used to retrieve complete Neandertal, Denisovan, and Pleistocene modern human mtDNAs previously (7, 9, 20). The PEC primers are 5′-biotinylated and consist of a universal 12-base 5′ spacer sequence (CAAGGACATCCG) followed by a specific primer sequence. To design the specific primer sequences, Primer3 (21) was used to find all possible primer sites on the light strand of the Vindija 33.16 Neandertal mtDNA sequence (8).

After the final spin column purification into 50 μL TE buffer, the PEC products were sequenced on the Illumina GAII platform. For this purpose, a PCR primer pair was constructed that is complementary to the p5 and p7 adapters on the 3′ ends. The Illumina/Solexa library was then amplified in a 100-μL reaction containing 50 μL Phusion High-Fidelity Master Mix, 500 nM each Solexa primer, and 10 μL PEC product template. Annealing temperature was 60 °C, and a total of 10 cycles of PCR was performed. The amplified products were spin column-purified and quantified on an Agilent 2100 Bioanalyzer DNA 1000 Chip. The PEC product was diluted and sequenced according to Illumina GAII protocols on a paired end run with a total of 76 cycles.

The sequencing run was analyzed starting from raw images using the Illumina Genome Analyzer pipeline 1.3.2. The first five sequencing cycles were used for cluster identification. After standard base calling, reads of the PhiX 174 control lane were aligned to the corresponding reference sequence to obtain a training dataset for the base caller Ibis (22). Raw sequences obtained from Ibis for the two paired end reads of each sequencing cluster were merged (including adapter removal), requiring at least 11-nt overlap between the two reads. For bases in the overlap, quality scores were summed up. In cases where different bases were called, the base with the higher-quality score was chosen.

SI Results

Comparison of PMDS Analysis with Simple Filtering Based on Mismatches. We compared the fraction of retained authentic sequences as a function of their enrichment relative to contamination for the PMDS approach with a simpler previously suggested approach, where a sequence is retained if it displays a C→T mismatch within, for example, 1–15 bp of the 5′ terminus of the sequence (3). We found that, compared with the basic method, the PMDS approach offers better tradeoff between the enrichment and the fraction-retained endogenous sequences as well as being able to enrich for authentic DNA far beyond what is possible with the simple method of observing a mismatch (Fig. S3). Whereas the basic method is able to achieve its maximum 20-fold enrichment by restricting to sequences with a mismatch in the first position (assuming contamination similar to the 100-y-old Australian sample and 40,000-y-old Neandertal endogenous sequences), the PMDS method is able to achieve, for instance, a 5,000-fold enrichment while retaining ∼10% of the authentic sequence reads in the same data (Fig. S3).

Model for Single-Stranded Libraries. In single-stranded library protocols, the only empirically observed mismatches caused by C deamination events are C → T mismatches at both ends of sequences (23). However, these mismatches could be affected by single-stranded overhangs in the original template molecule from either end. Unlike double-stranded library amplifications, it is not possible to differentiate between deamination caused by 5′ and 3′ single-stranded overhangs in data from single-stranded library preparation. We, thus, consider two possible processes giving rise to postmortem nucleotide polymorphisms D_z and D_y, where z and y denote distance to the 5′ and 3′ termini, respectively. Here, the probability of match is then the sum of the following exclusive events: no damage (from either end), no error, and no polymorphism; a misincorporation has been reverted because of sequence error (two terms: one for D_z and one for D_y); and finally, a polymorphism is reverted because of error conditional on no misincorporation:

$$\begin{aligned}P(Match|z,y) = {} & (1-\pi)\times(1-\varepsilon)\times(1-D_z)\times\left(\mathbf{1-D_y}\right)+(1-\pi)\\ & \times\varepsilon\times D_z\times\left(\mathbf{1-D_y}\right)+(1-\pi)\times\varepsilon\times D_y\times\left(\mathbf{1-D_z}\right)\\ & +\pi\times\varepsilon\times(1-D_z)\times\left(\mathbf{1-D_y}\right).\end{aligned} \quad [S1]$$

Adjusting Base Qualities for PMD. Selecting for nucleotide misincorporations amplifies a problem that already plagues ancient genomics: genotyping short reads with nucleotide misincorporations. PMD-aware consensus sequence calling has been developed for mtDNA sequencing studies (9) but has not been widely adopted for genome-wide sequence data (24), and no software for large-scale genomic data is currently available. One possibility is to adjust the base qualities for the probability of PMD (25), which allows their incorporation into all downstream analyses.

We incorporated the PMD model used here into genotyping by probabilistic adjustment of the base quality scores (25) (*SI Materials and Methods*). Given the parameter D_z of the rate of nucleotide misincorporations, we can adjust the base quality scores according to the position-specific probability of PMD (25). Specifically, C/G reference sites, where a T/A is observed in the sequence read, could be erroneously inferred as a true polymorphism because of either a sequencing error or a nucleotide misincorporation. In the same notation as shown in *Materials and Methods*, an adjusted phred-scaled base quality score Q_{adjusted} is then (25)

$$Q_{\text{adjusted}} = -10\times\log_{10}\left(1-(1{-}D_z)\times\left(1-10^{-Q/10}\right)\right). \quad [S2]$$

We tested the ability of this approach to alleviate problems with nucleotide misincorporations using the Vindija 33.16 mitochondrial data (1). First, the sequence data were remapped to the revised Cambridge human reference sequence (rCRS) using BWA with the same parameters as above, because ANFO alignments contained a portion of soft clipped regions that would be ignored during base quality adjustment. Second, we called genotypes for the entire mitochondrion using samtools (v0.1.16) pileup (6) in haploid mode, requiring a phred-scaled consensus quality of at least 30. Third, we counted the fraction of genotypes where there was C/T or G/A ambiguity, stratifying by different thresholds on the PMDS statistic, and found that this approach substantially increases the number of sites that can be reliably called (Fig. S9).

Analysis of the Okladnikov 2 mtDNA Assembly. Using positions that separate the Okladnikov 2 majority allele from 311 modern human mtDNAs (*Material and Methods*), we found that, of 1,318 informative fragments, 149 fragments were putative contaminants (10.2%; 95% confidence interval = 8.7–11.7%), whereas for 3,908 sequences that had a positive PMDS, only 17 of 1,301 informative fragments were putative contaminants (1.3%; 95% confidence interval = 0.7–1.9%).

Average consensus support in the final assembly was 97.0%, with 92% of all positions achieving support of 90% or higher. We observed only one difference between the 15-fold coverage consensus inferred using only sequences with PMDS ≥ 0 and the 25-fold unfiltered data: a T residue at the position corresponding to rCRS position 1,845 in the PMDS ≥ 0 assembly, where the unfiltered (contaminated) assembly displayed a C residue. We aligned the consensus sequences of Okladnikov 2 to complete mitochondrial sequences from 53 modern humans (10), 6 Neandertals (8, 9), 2 Denisovans (20, 26), 1 gorilla, 2 Bonobos, and 3 chimpanzees using MUSCLE (27) with default parameters. In this alignment, we found that all other 67 mitochondria also displayed the C residue, indicating that this position could be either private to Okladnikov 2 or caused by C deamination. However, when increasing the PMDS threshold to three, this position was not reliably called. Hence, we took the conservative approach of excluding this position from the final consensus. We also observed one position that differed between the Okladnikov 2 sequence obtained in this study and the sequence previously obtained by targeting the mtDNA control region detailed in the work by Krause et al. (28). At this position (corresponding to 16,148 in rCRS), the consensus called was a C, whereas the allele determined previously was a T.

To investigate the effect on the consensus calls of the 10% contamination rate in the unfiltered data, we repeated the procedure with the full contaminated data and investigated at which positions we had either gained or lost a reliable call by restricting to sequences with PMDS ≥ 0. There were 88 of 16,566 positions that were not reliably called for the PMDS ≥ 0 data (0.53%) compared with 24 of 16,563 positions in the unfiltered data (0.14%). In total, seven called sites were gained by restricting to PMDS ≥ 0, all of which were polymorphic or fixed differences in a set of mtDNA sequences from 6 Neandertal and 53 modern humans (Table S1). In contrast, among the calls that were lost by PMDS ≥ 0, only 4 of 68 sites (5.9%) were polymorphic or fixed differences. These observations are consistent with the prediction that the removal of contamination by the PMDS approach enables base calling at evolutionary informative sites that are otherwise ambiguous because of both contaminant and endogenous alleles being present. In contrast, the reduction from 25- to 15-fold coverage in the final PMDS ≥ 0 assembly resulted in random loss of sites without bias for those sites that are polymorphic or fixed differences in the set of 60 hominin mtDNAs.

To investigate the effect of reference sequence choice for assembly, we also aligned Okladnikov 2 sequences to the rCRS and observed two differences. The first difference is a CA dinucleotide repeat starting at position 514 in rCRS that displays from three (in Gorilla and the six previously analyzed Neandertals) to seven repeats (in an Aboriginal Australian; AF346965) in our alignment of 67 hominin mitochondria. Here, the Vindija 33.16 assembly had three repeats, consistent with the state in the six previously analyzed Neandertals (but not Denisovans, who have four repeats), whereas the rCRS assembly had five repeats, a state found in several modern human mtDNAs. The second difference is at position 16,263 in rCRS: the Vindija 33.16-based assembly is undetermined at two positions, where the rCRS-based assembly displays CT. After these two position, the six Neandertals previously analyzed also display an A insertion not seen in other sequences. The positions are further complicated by the PCR-amplified region discussed in the work by Krause et al. (28), which shows that the sequence CTA differs from the rCRS-based assembly. We, therefore, chose the conservative approach of masking these two positions in the final assembly based on Vi33.16 and excluding the possible A insertion.

Because the relationship between Okladnikov 2, Mezmaiskaya 1, and the five Western Neandertal mtDNAs differed from previous studies, we replicated the phylogenetic analysis using a 304-bp alignment sequenced in several additional Neandertals (28–36) and added three modern human sequences: Denisova 1, chimpanzee, and Bonobo sequences. In this analysis of 304 bp (Fig. S6), the deeper divergences between Neandertal mtDNAs remain unresolved, similar to the result found in ref. 28. However, the Okladnikov 2 sequence obtained here is nearly identical with the fragment sequenced by a previous study (28) for 304 bp, showing that the uncalled positions in the new sequence are not responsible for the well-supported topology of the Mezmaiskaya 1 being basal to the Okladnikov 2 sequence and the other complete Neandertal mtDNA genomes.

Validation of the Neandertal mtDNA Topology. To confirm that the topology of Neandertal mtDNA genomes, where Mezmaiskaya 1 is basal to Western Neandertals and Okladnikov 2, was not affected by the PMDS filtering of Okladnikov 2, we obtained shotgun sequences from Mezmaiskaya 1, Vindija 33.16, Vindija 33.25, and Vindija 33.36 from the study by Green et al. (1). We realigned mtDNA sequences from these four individuals using BWA with identical parameters as for Okladnikov 2 and downsampled these four datasets and the Okladnikov 2 data to exactly 2,500 randomly chosen sequences. We then excluded sequences with a PMDS below zero and called consensus sequences using the *mpileup* and *vcf2fq* tools in the samtools suite with default parameters. We reconstructed an mtDNA sequence phylogeny exactly as in the main analysis and found that the topology of Neandertal mtDNA sequences was exactly the same as in the main analysis, with the Mezmaiskaya 1 sequence basal to the Okladnikov 2 sequence and the Vindija Neandertals (Fig. S7).

1. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328(5979): 710–722.
2. Rasmussen M, et al. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334(6052):94–98.
3. Skoglund P, et al. (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336(6080):466–469.
4. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
5. Kircher M (2012) *Analysis of High-Throughput Ancient DNA Sequencing Data. Ancient DNA* (Springer, Berlin), pp 197–228.
6. Li H, et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
7. Krause J, et al. (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* 20(3):231–236.
8. Green RE, et al. (2008) A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134(3):416–426.
9. Briggs AW, et al. (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325(5938):318–321.
10. Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708–713.
11. Green RE, et al. (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* 444(7117):330–336.
12. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190.
13. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100–1104.
14. Surakka I, et al. (2010) Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging. *Genome Res* 20(10):1344–1351.
15. Altshuler DM, et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311):52–58.
16. Skoglund P, Jakobsson M (2011) Archaic human ancestry in East Asia. *Proc Natl Acad Sci USA* 108(45):18301–18306.
17. Rohland N, Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* 42(3):343–352.
18. Margulies M, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380.
19. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):t5448.
20. Krause J, et al. (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464(7290):894–897.
21. Rozen S, Skaletsky H (1999) *Primer3 on the WWW for General Users and for Biologist Programmers. Bioinformatics Methods and Protocols* (Springer, Berlin), pp 365–386.
22. Kircher M, Stenzel U, Kelso J (2009) Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol* 10(8):R83.
23. Meyer M, et al. (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222–226.
24. Rasmussen M, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757–762.
25. Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L (2013) mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.
26. Reich D, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327):1053–1060.
27. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
28. Krause J, et al. (2007) Neanderthals in central Asia and Siberia. *Nature* 449(7164): 902–904.
29. Krings M, et al. (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90(1):19–30.
30. Krings M, et al. (2000) A view of Neandertal genetic diversity. *Nat Genet* 26(2):144–146.
31. Ovchinnikov IV, et al. (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404(6777):490–493.
32. Lalueza-Fox C, et al. (2006) Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals. *Curr Biol* 16(16):R629–R630.
33. Lalueza-Fox C, et al. (2005) Neandertal evolutionary genetics: Mitochondrial DNA data from the iberian peninsula. *Mol Biol Evol* 22(4):1077–1081.
34. Orlando L, et al. (2006) Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr Biol* 16(11):R400–R402.
35. Caramelli D, et al. (2006) A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Curr Biol* 16(16):R630–R632.
36. Dalén L, et al. (2012) Partial genetic turnover in neandertals: Continuity in the East and population replacement in the West. *Mol Biol Evol* 29(8):1893–1897.

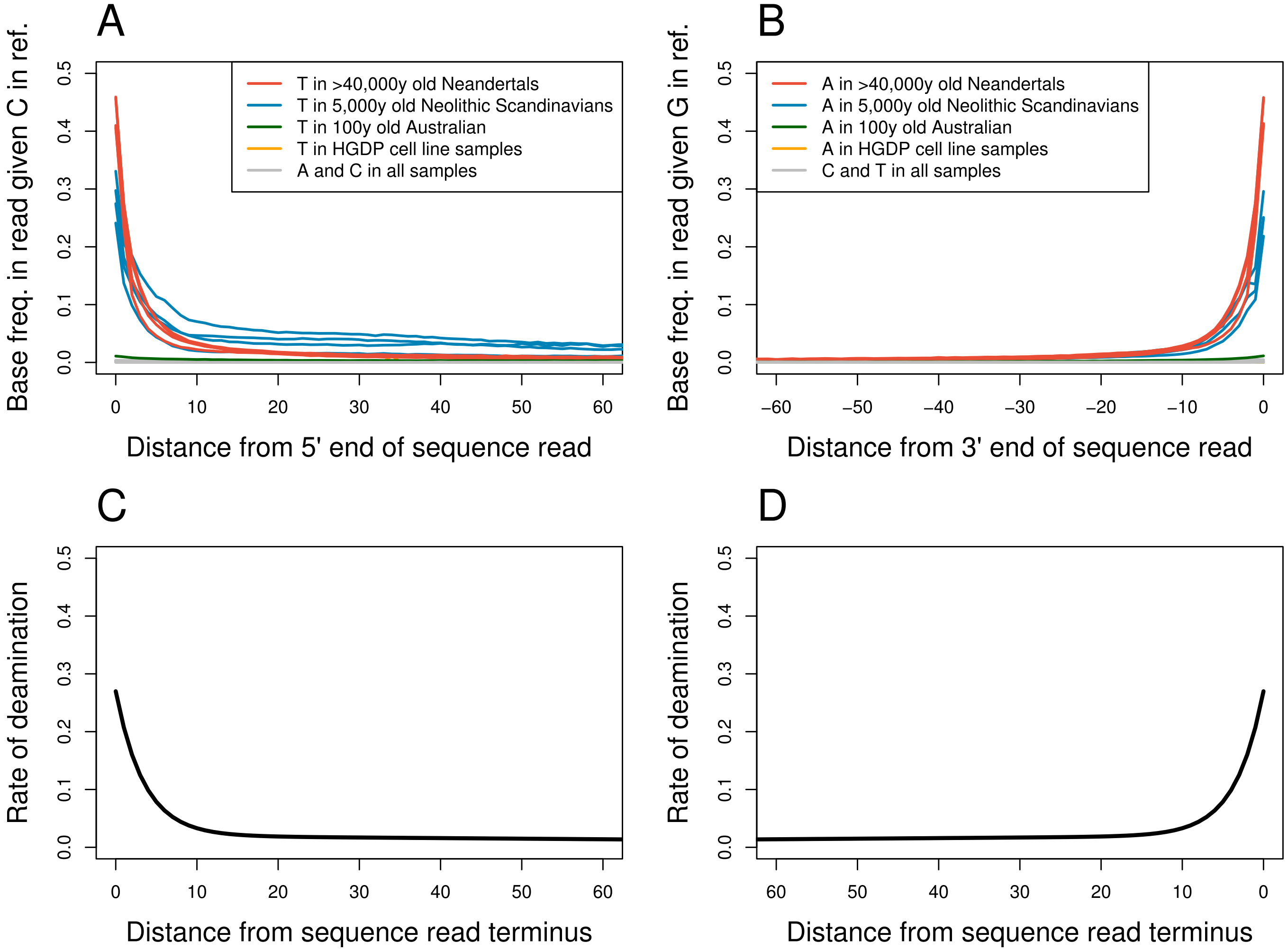

Fig. S1. Nucleotide misincorporation patterns arising from C deamination in ancient DNA sequences. (*A* and *B*) Empirical patterns in datasets analyzed in this study. (*C* and *D*) Rate of deamination assumed in the PMDS model in this study.

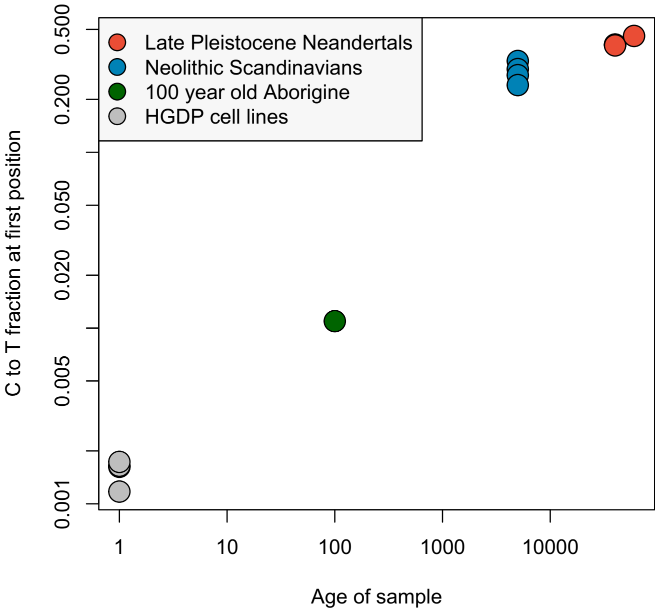

Fig. S2. Relationship between age and observed deamination fraction (C to T mismatches) in the 5′ end of sequence data analyzed in this study. Blue, Neolithic Scandinavians; gray, HGDP cell lines from present day individuals; green, 100-y-old Australian Aborigine hair sample; red, Pleistocene Neandertals. Note the log scale for both the *x* and *y* axes.

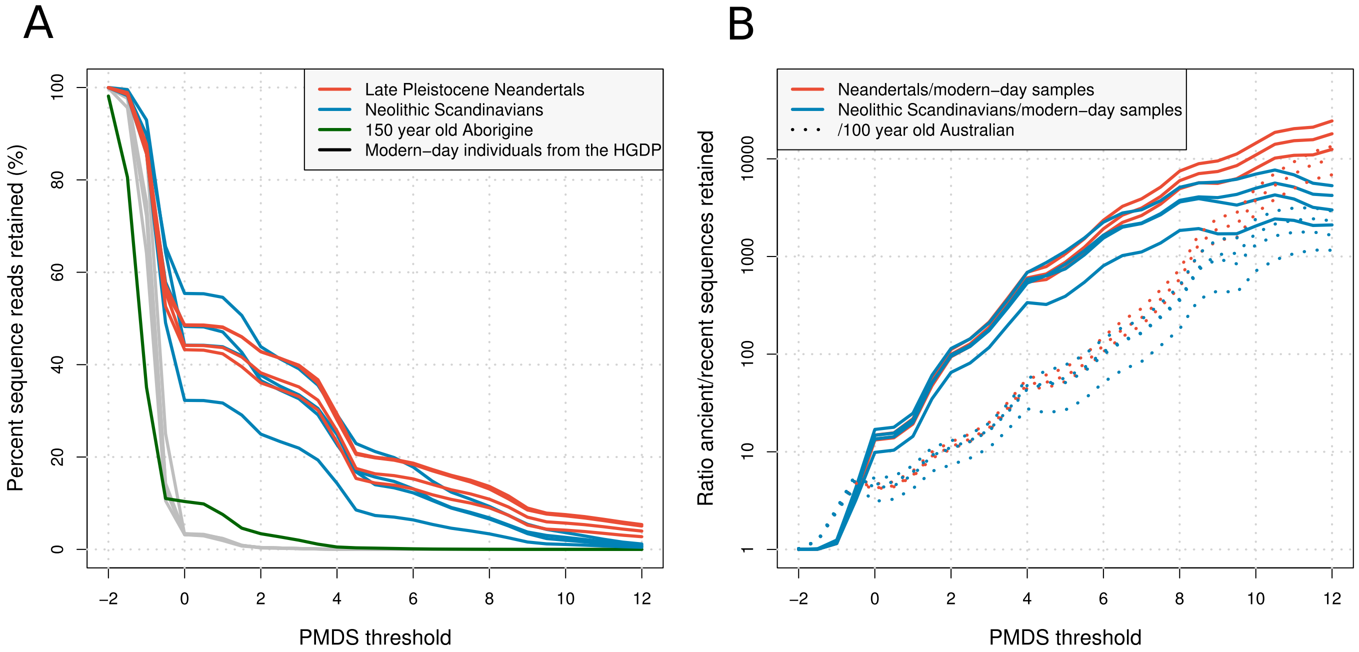

Fig. S3. (*A*) The amount of retained data as a function of different thresholds on PMDS. (*B*) The enrichment of ancient over more recent sequence reads as a function of chosen PMDS threshold. In our example data, a PMDS threshold of ∼2 would achieve a 100-fold enrichment of Neandertal and Neolithic Scandinavian sequences over present day (contaminating) sequences, and a PMDS threshold of ∼6 would achieve the same 100-fold enrichment when contaminating the Neandertal or Neolithic Scandinavian data with similar sequences from the 100-y-old remains of the Australian individual.

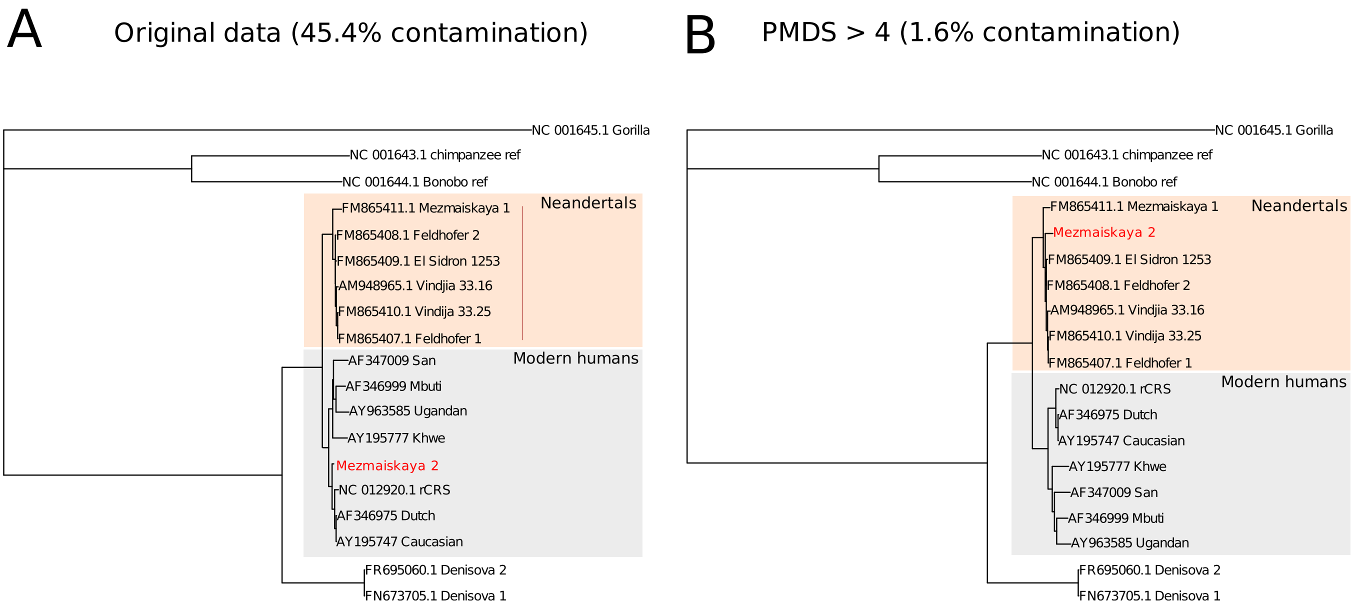

Fig. S4. Inferred gene genealogy of contaminated Mezmaiskaya 2 mtDNA sequence data (*A*) before and (*B*) after PMDS filtering. A consensus mtDNA sequence for Mezmaiskaya 2 was called using the haploid model of samtools pileup either using all sequences or restricting to sequences with a PMDS ≥ 4. The consensus sequences were subsequently aligned to mtDNA genome sequences from seven modern humans, six Neandertals, two Denisovans, one Bonobo, one chimpanzee, and one gorilla. A neighbor-joining tree was reconstructed for both datasets.

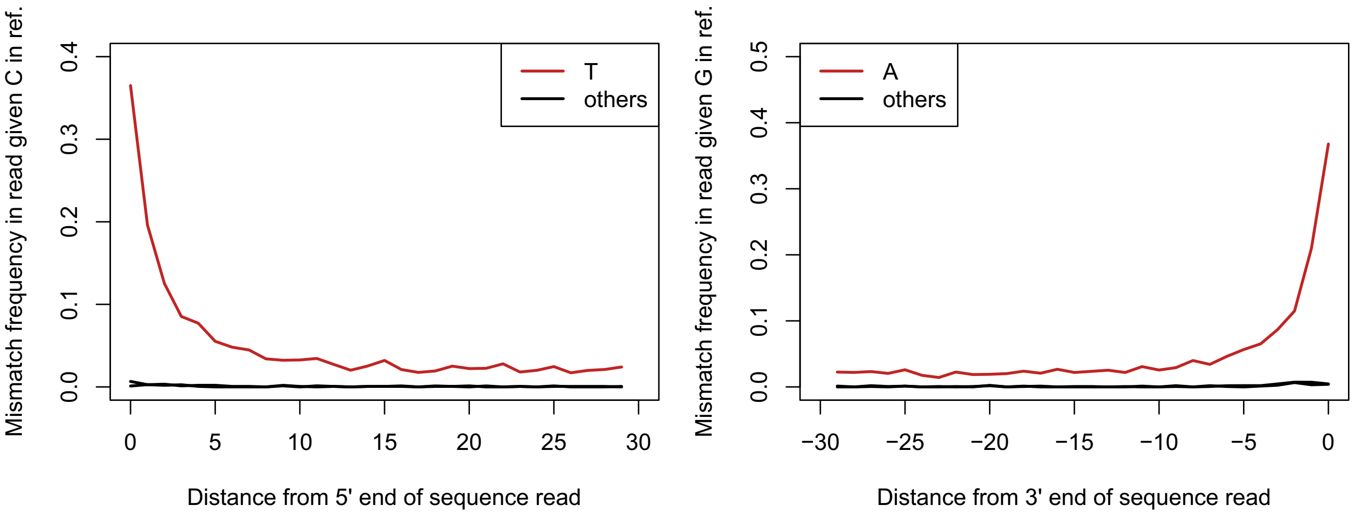

Fig. S5. Nucleotide misincorporation patterns in Okladnikov 2 mtDNA sequences. We show the fraction of given nucleotides in the Okladnikov 2 sequence data at positions where the rCRS reference sequence had (*Left*) a C residue or (*Right*) an A residue.

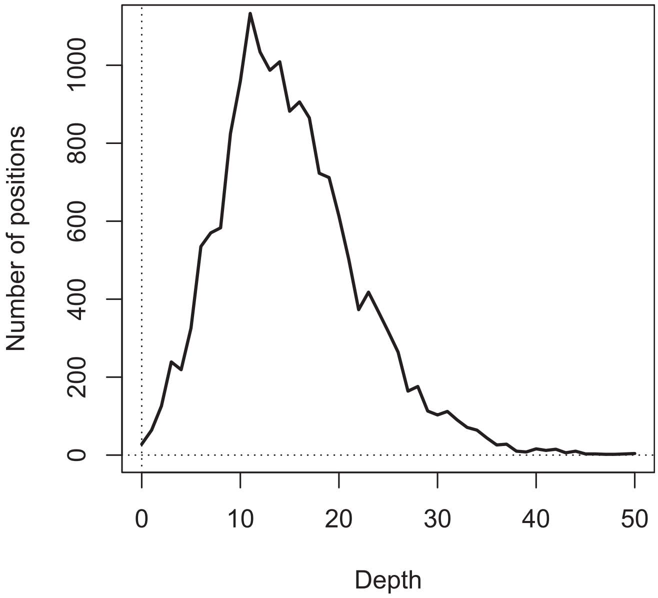

Fig. S6. Sequencing depth distribution for the Okladnikov 2 mitochondrial genome. Only sequences with PMDS ≥ 0 were used.

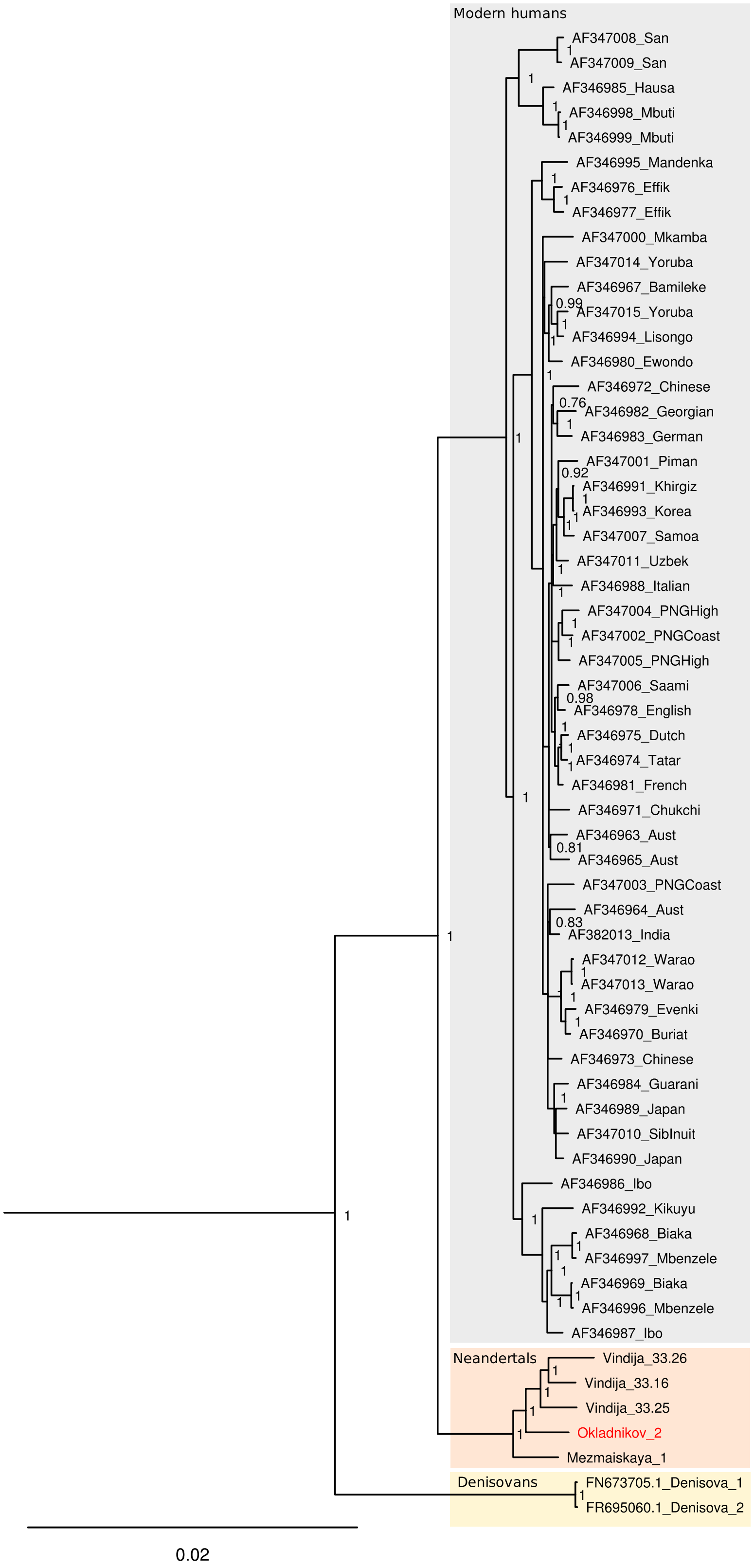

Fig. S7. The position of Okladnikov 2 in the Neandertal mtDNA tree topology is not biased by the PMDS filtering. Mitochondrial sequences from Okladnikov 2, Mezmaiskaya 1, Vindija 33.16, Vindija 33.25, and Vindija 33.26 were down-sampled to 2,500 randomly sampled sequences, of which sequences with negative PMDSs were excluded. A Bayesian tree was reconstructed, and the Neandertal mtDNA sequence tree topology was identical to the one in the main analysis, where only Okladnikov 2 was subjected to PMDS filtering.

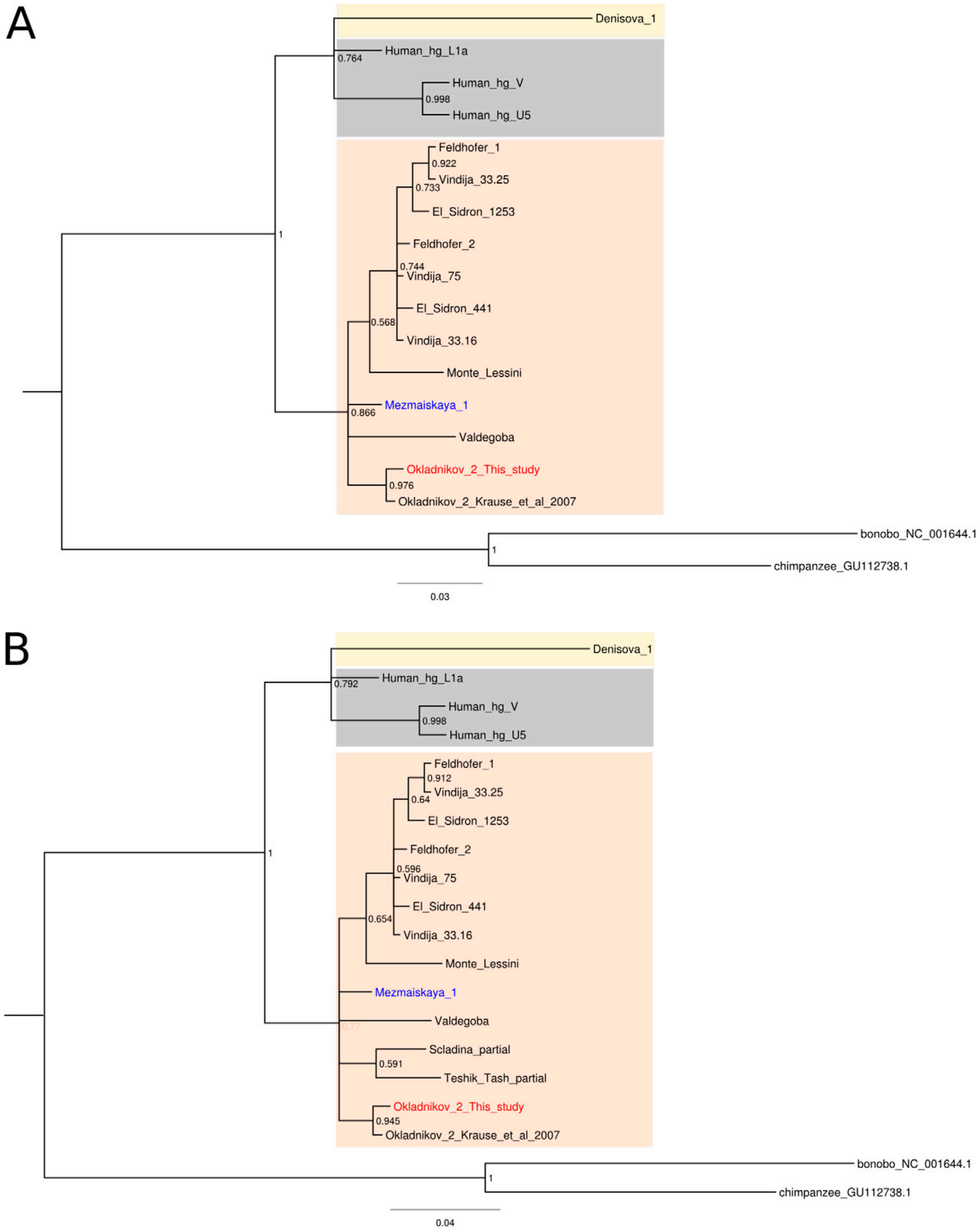

Fig. S8. Phylogenetic analysis of the hypervariable region. The Okladnikov 2 sequence obtained here falls unambiguously with the previously sequenced short fragment. However, the relationship with Mezmaiskaya 1 is not resolved based on this restricted region. *A* is restricted to sequences covering 304 bp, whereas *B* also includes the partial Scladina and Teshik Tash sequences (for which the flanking missing sequence was marked as missing data).

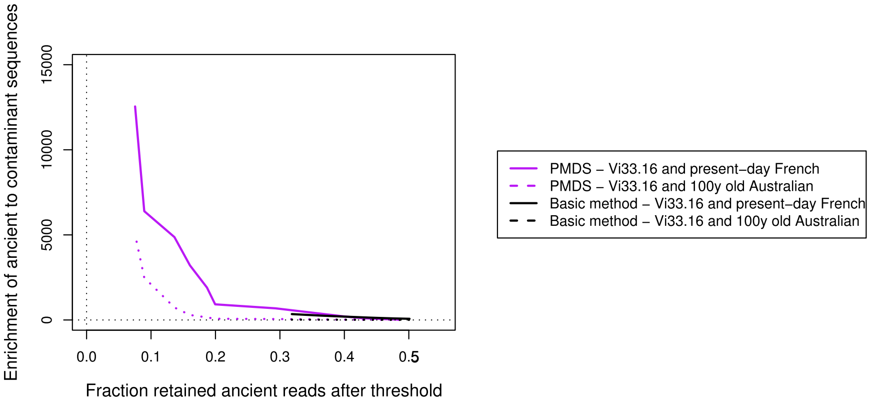

Fig. S9. Comparison of PMDS filtering with a basic method of extracting reads with a C→T/G→A mismatch in the terminus. The PMDS results are based on PMDS thresholds of 0 (least enrichment) to 10 (greater enrichment). The basic method results are based on extracting reads with a C→T mismatch at most *n* bp from the 5′ terminus, where *n* ranged from 15 (less enrichment) to 1 (greatest enrichment). Higher values of *n* result in less-efficient enrichment. For both methods, we excluded bases with quality < 20.

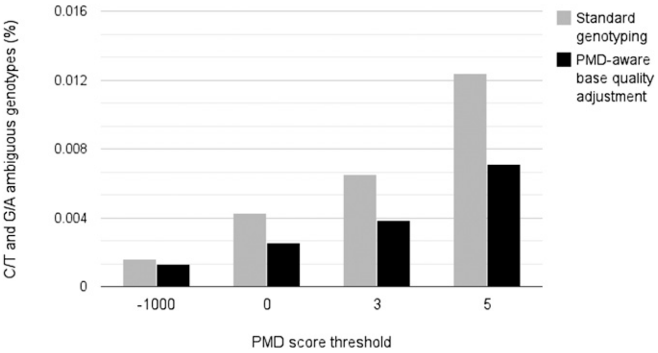

Fig. S10. Genotyping ancient DNA using base qualities adjusted for PMD. Fraction C/T and G/A genotypes obtained using standard samtools pileup haploid genotyping (gray) and base qualities adjusted using a model of nucleotide misincorporation (black) are displayed.

Table S1. Characterization of base calls gained and lost in the PMDS ≥ 0 assembly of Okladnikov 2

State in 6 Neandertal and 53 modern human mtDNA sequences	Calls gained in the PMDS ≥ 0 assembly	Calls lost in the PMDS ≥ 0 assembly
Not polymorphic	0	64
Fixed differences	2	2
Polymorphic in Neandertals only	1	0
Polymorphic in modern humans only	3	1
Shared polymorphisms	1	1
Total polymorphisms	7/7 (100%)	4/68 (5.9%)

We present the number of positions for each category that is polymorphic in a set of 6 previously published Neandertal mtDNAs and 53 modern human mtDNAs.