Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/7/e1700186/DC1

Supplementary Materials for

A fourth Denisovan individual

Viviane Slon, Bence Viola, Gabriel Renaud, Marie-Theres Gansauge, Stefano Benazzi, Susanna Sawyer, Jean-Jacques Hublin, Michael V. Shunkov, Anatoly P. Derevianko, Janet Kelso, Kay Prüfer, Matthias Meyer, Svante Pääbo

Published 7 July 2017, *Sci. Adv.* **3**, e1700186 (2017)
DOI: 10.1126/sciadv.1700186

This PDF file includes:

- table S4. Percentage and number of sequences matching the derived state of mitochondrial genomes from three hominin groups.
- table S5. Number of base differences between the *Denisova 2* mtDNA and other mtDNAs.
- table S6. Estimates of DNA divergence between low-coverage archaic genomes and the high-coverage genomes of a Denisovan, a Neandertal, and 12 present-day humans.
- table S7. Estimates of DNA divergence between subsampled genomes and the high-coverage genomes of a Denisovan, a Neandertal, and 12 present-day humans.
- table S8. Sharing of derived alleles between *Denisova 2* and present-day humans.
- table S9. Denisova-specific nonsynonymous coding changes corroborated by sequences from *Denisova 2*, *4*, or *8*.
- References (*70–82*)

section S1. The *Denisova 2* specimen

Morphology

The crown of this left lower second deciduous molar is almost completely worn away. A thin rim between 2.7 mm (on the buccal and lingual sides) and 1.6 mm (mesially) in height is all that remains of the enamel. The only feature of crown morphology visible is a small remnant of the buccal groove. The root is mostly resorbed, its length (from the enamel-dentine junction [EDJ]) is between 1.3 mm (distally) and 3.9 mm (mesiobuccally). The pulp cavity is exposed. Surface preservation is in general good, but a 4.2x3 mm piece of enamel is broken away on the buccal surface just distal of the buccal groove, and a smaller piece from the mesiolingual corner.

Description

Few morphological features are visible due to the extreme wear of the tooth. It is clear that the tooth was large, even though only the cervical diameters could be taken.

The wear surface is separated by a ridge at about a third of the mesiodistal length into two angled areas, a smaller and flatter mesial one, and a larger and concave distal one. The mesial area is sloping upwards from its mesiobuccal corner distolingually. The surface is relatively flat and polished, with small striae visible at 20x magnification. The striae are mainly running mesiolingually to distobuccally. The ridge separating the two parts of the occlusal surface starts near the buccal groove, and becomes more pronounced in the center of the tooth. It is placed a bit more mesially there, and then runs slightly distally towards the lingual margin. The distal part of the surface is sloping down its mesiobuccal corner, its deepest part is just lingually of the distal interproximal facet. The lingual face of the crown is somewhat raised. The surface is rough and strongly pitted.

An approximately 5x2.3 mm interproximal facet (IPF) is visible in the buccal two-thirds of the distal aspect of the tooth. The IPF extends from the EDJ to the occlusal surface, and thus is placed far down. Lingually, a slight enamel extension towards the cleft between the mesial and distal root is visible.

The pulp cavity shows five diverticles entering the crown, three buccally and two lingually. The deepest is the mesiolingual diverticle, followed by the larger mesiobuccal one. Two mesial root stumps are present, the mesiobuccal one is about 3.9 mm from the EDJ, the mesiolingual one about 3.4 mm.

Comparative morphology

Due to the extreme wear of this tooth, its interpretation is not easy and there is even disagreement about its position. Turner (*29*) determined it as a lower right dm_1, while Shpakova (*30, 70*) and the present study identify it as a dm_2. The determination as dm_2 is supported by the large size, and the absence of a *tuberculum molare* (or molar tubercle of Zuckerkandl) including the strong apical deviation of the cervical line in its area.

Lower deciduous second molars of Neandertals are characterized among else by their relatively large size, their elliptical occlusal outline, and the presence of a midtrigonid crest (*71, 72*). Sadly, none of these traits can be studied on this specimen, except for a very tentative comparison of size. The mesiodistal and buccolingual diameters that can be measured on the crown are minimum estimates, as deciduous teeth bulge sharply from the cervix outwards. We compared cervical diameters (*73*) of *Denisova 2* to a sample of recent and Upper Palaeolithic (UP) modern humans and Neandertals (table S1, fig. S3). *Denisova 2* is much larger than the recent comparative sample, and somewhat larger than the UP sample, but falls into the range of variation of Neandertal dm_2s.

Age determination

The extensive resorption of the roots of *Denisova 2* indicates that it was exfoliated naturally. Root resorption is a much less reliable indicator of age than crown formation, but it does give a rough estimate of age. Moorrees and colleagues (*74*) published data on root resorption of deciduous teeth in a contemporary sample from Ohio; in their sample exfoliation of the dm_2 occurred at 11.1±1.1 years in females and 11.6±1.2 years in males. Very little data is available about geographic and temporal variation in deciduous tooth root resorption, but Nanda (*75*) showed that root resorption in Indian children is significantly slower than in the Moorrees et al. (*74*) sample. In Indian children, dm_2 roots were still not completely resorbed (average score 6.6 of 8, where 0 is complete root and 8 is exfoliation, no standard deviations given) at the age of 12, when the study ended. Ubelaker's (*76*) charts show exfoliation between 10 and 11 years (±30 months), thus supporting a rough age estimate of 10-12 years.

section S2. Authenticating the DNA sequences

Nucleotide substitutions typical of ancient DNA

Cytosine (C) bases at the single-stranded ends of ancient DNA fragments tend to undergo deamination. The resulting uracil residues are read by DNA polymerases as thymines (T), yielding C to T substitutions compared to a reference genome (*36*). As this damage accumulates over time, ancient DNA fragments aligned to a reference sequence tend to carry high frequencies of apparent C to T substitutions at their 5'- and 3'- ends. In contrast, modern DNA shows much lower frequencies of terminal C to T substitutions (*68*), allowing us to use these frequencies to determine whether ancient molecules are present in a library (*55*). In the DNA library we prepared following the standard protocol (*32*), 9.4% and 11.1% of fragments starting or ending at positions where the reference base is a C carried a T, respectively (table S3, fig. S5). These elevated frequencies indicate that ancient DNA molecules are present in the library.

A recently-developed library preparation protocol ('U selection') takes advantage of the presence of uracil bases in ancient DNA by physically isolating fragments carrying such bases, in order to enrich for authentic ancient DNA molecules (*34*). Here we applied a modified version of this protocol ('mini-U-selection'), targeting uracil bases at the 3' ends of fragments (see Methods). As above, the elevated frequencies of C to T substitutions to the reference genome (7.3% and 53.8% at the 5' and 3' ends of fragments, respectively) are indicative of the presence of ancient DNA molecules in this library (table S3, fig. S5).

As expected, the frequency of C to T substitutions at the 3' end of fragments in the library enriched for terminal uracil residues is higher than in the library prepared with the standard protocol. Given that the 'mini-U-selection' protocol specifically enriches for ancient molecules, one would expect a higher frequency of C to T changes to occur on the 5' ends of fragments as well, however these frequencies are similar in the two libraries. One possible explanation for this discrepancy is an inefficiency in the selection for uracil-containing fragments, resulting in some DNA fragments with terminal uracil bases to remain immobilized on the beads. To test this, the 'beads' fraction was sequenced on 10% of a MiSeq (Illumina) lane. Of the fragments starting or ending at positions where the reference base is a C, 2.44% (236/9,864) and 0.9% (79/8,779) carried Ts at their 5'- and 3'- ends, respectively, indicating that some putatively ancient molecules were not incorporated into the final library. A partial inefficiency in the selection process would also result in the presence of DNA fragments that do not contain a terminal uracil in the final library, accounting for a 3'-end C to T substitution frequency of less than 100%. Other

factors, such as the introduction of different amounts of contaminating human DNA during the laboratory procedures, or inhibition of enzymes during the library preparation steps, could also contribute to differences in the frequency of terminal C to T substitutions in the different libraries prepared from the same DNA extract.

Estimating contamination by present-day human DNA

Three approaches were taken to estimate the amount of contamination by present-day modern human DNA in the libraries prepared from the *Denisova 2* specimen. First, we identified 172 positions at which the reconstructed mitochondrial genome of *Denisova 2* differs from all mtDNA sequences in a dataset of 311 present-day humans (*57*). We then realigned the *Denisova 2* sequences to the revised Cambridge human mitochondrial reference sequence (rCRS), to avoid omitting modern human mtDNA fragments which are too divergent from the Denisovan mitochondrial reference genome to map to it (*17*). The proportion of fragments which overlap these positions and carry the base typical of present-day humans constitutes the contamination estimate, with 95% exact binomial confidence intervals. Where one of the possible bases at informative positions was a C or a G, alignments on the forward or reverse strands, respectively, were ignored. Out of 7,751 DNA fragments from all libraries overlapping these positions, 4,433 carried the base typical of present-day humans, resulting in a contamination estimate of 57.2% (95% CI: 56.1-58.3%). The contamination estimate for the library prepared with enrichment for uracil-containing DNA fragments was 38.6% (CI: 36.7-40.6%) while for the standard library it was 65.5% (CI: 64.2-66.8%). After filtering for DNA fragments carrying C to T substitutions, 52 out of 1,552, *i.e.,* 3.4% (CI: 2.5-4.4%), carried the present-day human base in the combined data; 2.2% (CI: 1.4-3.3%) and 6.0% (CI: 4.0-8.6%) in the uracil-selected and standard libraries, respectively.

Second, in order to assess contamination by present-day human DNA in our nuclear data, we isolated fragments likely to be of ancient origin based on the presence of C to T substitution to the reference genome on their 5'- or 3'-ends, and computed the frequencies of C to T substitutions on their opposite end (*39*). These 'conditional' substitution frequencies serve as a proxy for the frequency expected in endogenous ancient fragments. Thus, if the frequencies of C to T substitutions in all DNA fragments in a library are lower than the 'conditional' frequencies, this reflects contamination with present-day human DNA. The contamination estimate was calculated by subtracting from 1 the ratio of the average 'conditional' substitution frequency to the average terminal C to T frequency observed in all sequences (*12*). The terminal C to T substitution frequencies in the combined dataset increased from an average of

20.4% in all fragments, to an average of 50.3% after filtering for C to T substitutions at the opposite end of the molecules, resulting in a rough contamination estimate of ~60%. Consistent with the mtDNA contamination estimates, the estimate of contamination by present-day human DNA was lower in the library enriched for fragments carrying a 3'-uracil (~50%) than in the library prepared without this enrichment strategy (~75%) (table S3).

Third, we determined the sex of the *Denisova 2* specimen by counting the number of fragments assigned to each autosome and the X chromosome, and dividing these by the number of bases on the chromosomes to which we map ancient fragments. When only sequences presenting a C to T substitution to the reference genome on one of their ends were used, the ratio between the number of fragments per base mapping to the X chromosome and the number of fragments per base mapping to the autosomes is 1.06, indicating that the *Denisova 2* specimen comes from a female (fig. S7). Among the DNA fragments not carrying any apparent terminal C to T substitution, this ratio is 0.90, consistent with the presence of human DNA contamination. The sex determination allows us to estimate male contamination in the libraries by dividing the number of fragments mapping to the Y chromosome by the number of such fragments expected had the individual been a male. The latter was estimated by multiplying the number of aligned fragments in the whole genome by the number of 'mappable' positions in the Y chromosome, and then dividing that product by the size of the 'mappable' genome (*7*). Using this approach, the present-day male DNA contamination is estimated to 48.0% (95% CI: 47.3-48.7) in the combined data, and to 47.8% and 48.0% in the library enriched and not enriched for terminal uracils, respectively. When restricting the analysis to sequences carrying a terminal C to T substitution to the reference genome, we observed 71 sequences that map to the Y chromosome while 1,474 sequences would be expected if the specimen originated from a male, yielding an estimate of 4.8% for putative male DNA contamination.

section S3. The mtDNA sequence of *Denisova 2*

Preliminary phylogenetic inference based on mtDNA 'diagnostic' positions

To test whether the *Denisova 2* mtDNA sequences more closely resemble the mitochondrial genome of a modern human, a Neandertal or a Denisovan, reads originating from capture of mtDNA fragments (library A4928) were aligned to the revised Cambridge human mitochondrial reference sequence (rCRS) and compared to three sets of 'diagnostic' positions, at which the mitochondrial genomes of these groups differ from one another. The first set, comprised of modern human-specific variants, includes 14 positions at which the mitochondrial genomes of 311 present-day humans (*57*) differ from the mitochondrial genomes of three Denisovans (*2, 16, 17*), ten Neandertals (*7, 34, 57-59*), one Middle Pleistocene hominin (*39*) and one chimpanzee (*60*). The second set is constituted of 38 positions at which the three Denisovan mtDNA sequences differ from all others; and the third includes 19 positions at which the Neandertal mitochondrial genomes differ from all other mtDNA sequences.

The state of sequences overlapping these diagnostic sites were used to attribute the mitochondrial sequences from *Denisova 2* to one of these three hominin mtDNA types (table S4). 82.4% of sequences overlapping diagnostic positions from the first set carried the modern human-specific variant. By comparison, 16.9% of sequences matched the Denisova-specific variants and no sequences carried the Neandertal-specific variants. As this elevated proportion of matching the modern-human state could be compatible with both the presence of ancient modern human mtDNA and the presence of contamination by present-day human DNA in the library, we restricted subsequent analyses to sequences carrying a terminal C to T substitution compared to the reference genome, as such damaged fragments are more likely to be of ancient origin (*39*). Out of 24 sequences overlapping diagnostic positions from the first set, only one (4.2%) matched the modern human state, suggesting that the modern human-like mtDNA sequences in the library are not of ancient origin. In contrast, 93.8% of such sequences overlapping the second set of diagnostic positions matched the Denisovan-specific variant. We conclude from this preliminary analysis that the ancient mtDNA fragments in the library originated from a mitochondrial genome of the Denisovan type.

Reconstructing the mitochondrial genome sequence of Denisova 2

21,537 mtDNA fragments, which carry an apparent C to T substitution to the *Denisova 3* mtDNA genome on the first three or last three terminal positions (for sequences generated by the standard single-stranded library preparation protocol) or at their first two or last two positions (when generated while

enriching for 3' uracil residues), were yielded by shotgun sequencing and by mitochondrial capture. After converting Ts at these terminal positions to Ns if the reference *Denisova 3* mtDNA genome carries a Cs, we used these fragments to reconstruct the mitochondrial genome sequence of *Denisova 2* with an average coverage of 51-fold (fig. S6). 27 positions were not called, 13 of which were covered by two or fewer sequences. Fewer than two-third of sequences overlapping the remaining uncalled positions carried an identical base. The 13 uncalled bases at or near the start and end of the reference genome (positions 1-8, 16,566-16,570) were filled-in manually, as these are conserved in all mitochondrial genomes in our comparative dataset. This dataset is comprised of the mtDNAs of 311 present-day humans (*57*), ten Neandertals (*7, 34, 57-59*), three Denisovans (*2, 16, 17*), a Middle Pleistocene hominin from Spain (*39*), one chimpanzee (*Pan troglodytes*, NC_001643), one gorilla (*Gorilla gorilla*, NC_001645) and one orangutan (*Pongo pygmaeus*, NC_001646) (*60*). 14 positions (positions 15, 18, 195, 197, 307, 572, 4910, 5383, 5896, 9529, 14009, 16088, 16113 and 16189) remain uncalled in the final mtDNA sequence of *Denisova 2*.

The same approach was used to reconstruct an mtDNA consensus genome using the sequences in our dataset that do not carry terminal C to T substitutions typical of ancient molecules, *i.e.,* which are thought to originate from contamination (see Supplementary Materials section B). 1,853 positions in the mitochondrial genome were overlapped by sequences carrying different bases, suggesting the presence of more than one source of contaminating mtDNA. The consensus mtDNA falls within modern human variation and corresponds to haplogroup H, as determined using Haplogrep2 (*77, 78*). Haplogroup H constitutes the most common mitochondrial DNA haplogroup in Europe, which also occurs frequently in western, central and southern Asia (*79*).

section S4. Nuclear DNA sequences from *Denisova 2*

Detecting ancient gene flow

The 'D' statistic, which tests for significant differences in the sharing of derived alleles between individuals, can be used to retrace past admixture events (*1, 80*). To test whether the *Denisova 2* individual shares more derived alleles with some present-day individuals than with others (B panel from (*7*)), we computed 'D' statistics using mistar (https://github.com/grenaud/mistartools). For the high-coverage genomes, the homozygous genotype was selected at homozygous sites and a random allele was selected at heterozygous ones; while for *Denisova 2*, a random allele was selected at all sites. CpG sites were excluded.

To determine whether the 47Mb of sequencing data we recovered from *Denisova 2* would be sufficient to retrace ancient admixture events using this statistic, we randomly subsampled the high-coverage *Denisova 3* genome (*3*) to 47Mb using SAMtools (*52*). We then computed the statistic D(H1,H2,subsampled *Denisova 3*,chimpanzee) using present-day humans from a variety of populations (*7*) as individuals H1 and H2. For each pair of H1 and H2, we compared our result to the one obtained using all the data yielded from *Denisova 3*, as reported by (*3*). All non-significant Z-scores ($|Z| < 2$) obtained using the full dataset remained so using the subset of sequences. However, most significant Z-scores from the high-coverage genome became non-significant when using the smaller dataset (fig. S9). This suggests that our power to carry out this test utilizing sequences from *Denisova 2* is limited.

We computed the statistic D(H1,H2,*Denisova 2*,chimpanzee) to test whether the *Denisova 2* individual shares more derived alleles with some present-day populations than with others (table S8), while noting that our results may partially be driven by residual present-day contamination from individuals outside sub-Saharan Africa (see Supplementary Materials section B). All statistics of the form D(Non-African,African,*Denisova 2*,chimpanzee) point to a higher affinity of *Denisova 2* sequences to non-African individuals (table S8). Although such a pattern is expected from a Denisovan individual (*2, 3*), we were in most cases unable to detect it using a subset of sequences from the high-coverage *Denisova 3* genome (fig. S9), in which contamination has been estimated to be less than 0.5% (*3*). Likewise, we found a higher percentage of shared derived allele between *Denisova 2* and a Dinka individual, compared to any other African individual (table S8). To investigate this, we performed the 28 pairwise comparisons of the form D(African,Dinka,non-African,chimpanzee) using positions covered by the *Denisova 2* sequences. All but two showed a significantly higher sharing of derived alleles between the non-African individual and the Dinka individual, pointing to a possible effect of non-African present-day contamination on our results.

To investigate whether the signal of super-archaic admixture detected in the high-coverage *Denisova 3* genome by comparison with the high-coverage Neandertal genome (*7*) would also be detectable in the *Denisova 2* sequences, we again utilized a subset of the *Denisova 3* sequences spanning 47Mb. We compared the number of sequences carrying an ancestral allele at positions where present-day humans are fixed or nearly-fixed for a derived allele (*81*), between the subsampled *Denisova 3* dataset and the high-quality genotype calls (using a random allele at heterozygous sites) of that same individual at the same positions (*3*). Allele frequencies in present-day humans were determined using Phase 1 of the 1000 Genomes Project, retaining only sites passing the 'strict' masking scheme (*81*). The significant difference at sites where present-day humans are fixed for a derived allele (fig. S10B) suggests that sequencing errors affecting single reads artificially increase the signal of super-archaic admixture, hindering us from drawing conclusions from this test when carried out using the *Denisova 2* sequences (fig. S10A).

Denisova-specific genetic changes

Using homozygous derived positions in the high-coverage genome of the *Denisova 3* individual, at which present-day humans carry a fixed ancestral base, a catalog of 692,818 single nucleotide changes that may be specific to the Denisovan lineage has previously been assembled (*3*). We investigated the state of sequences from *Denisova 2*, as well as of sequences from *Denisova 4* and *Denisova 8* produced elsewhere (*17*), overlapping this catalog to assess which positions are likely to be Denisovan-specific and which are polymorphic among Denisovans. A total of 10,712 sequences from *Denisova 2*, 126 from *Denisova 4* and 3,134 from *Denisova 8* overlapped positions in the catalog. Similarly to the results of the lineage attribution analysis (Fig. 3A, fig. S8), 53.8%, 76.2% and 66.6% of sequences carry the same base as *Denisova 3*, respectively. Many of the positions in the catalog are thus not fixed derived in Denisovans, which is not unexpected as it was assembled using the genome of a single Denisovan individual (*3*). Out of the positions that are corroborated as being fixed derived in Denisovans by the sequences of *Denisova 2*, *Denisova 4* or *Denisova 8*, 27 are non-synonymous coding changes. We retrieved information on genes encoded in such positions from the UCSC Genome Browser (*82*) and found that these affect a variety of genes, including some involved in signaling, spermatogenesis and immune response (table S9). In 19 other positions where the *Denisova 3* genome carries a homozygous derived allele that causes a non-synonymous changes, at least one of the low-coverage Denisovans carries the human-like (fixed ancestral) base.

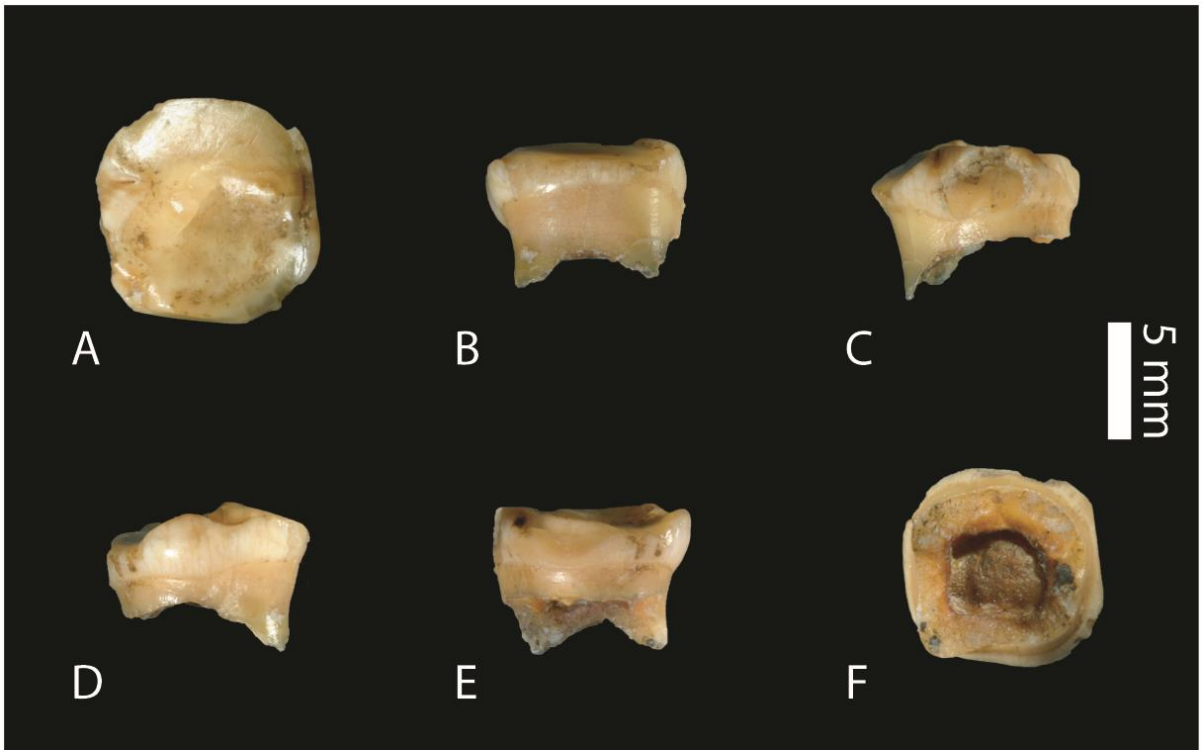

fig. S1. The *Denisova 2* specimen. Photographs of the *Denisova 2* deciduous lower second molar in (**A**) occlusal, (**B**) mesial, (**C**) buccal, (**D**) lingual, (**E**) distal, and (**F**) apical views.

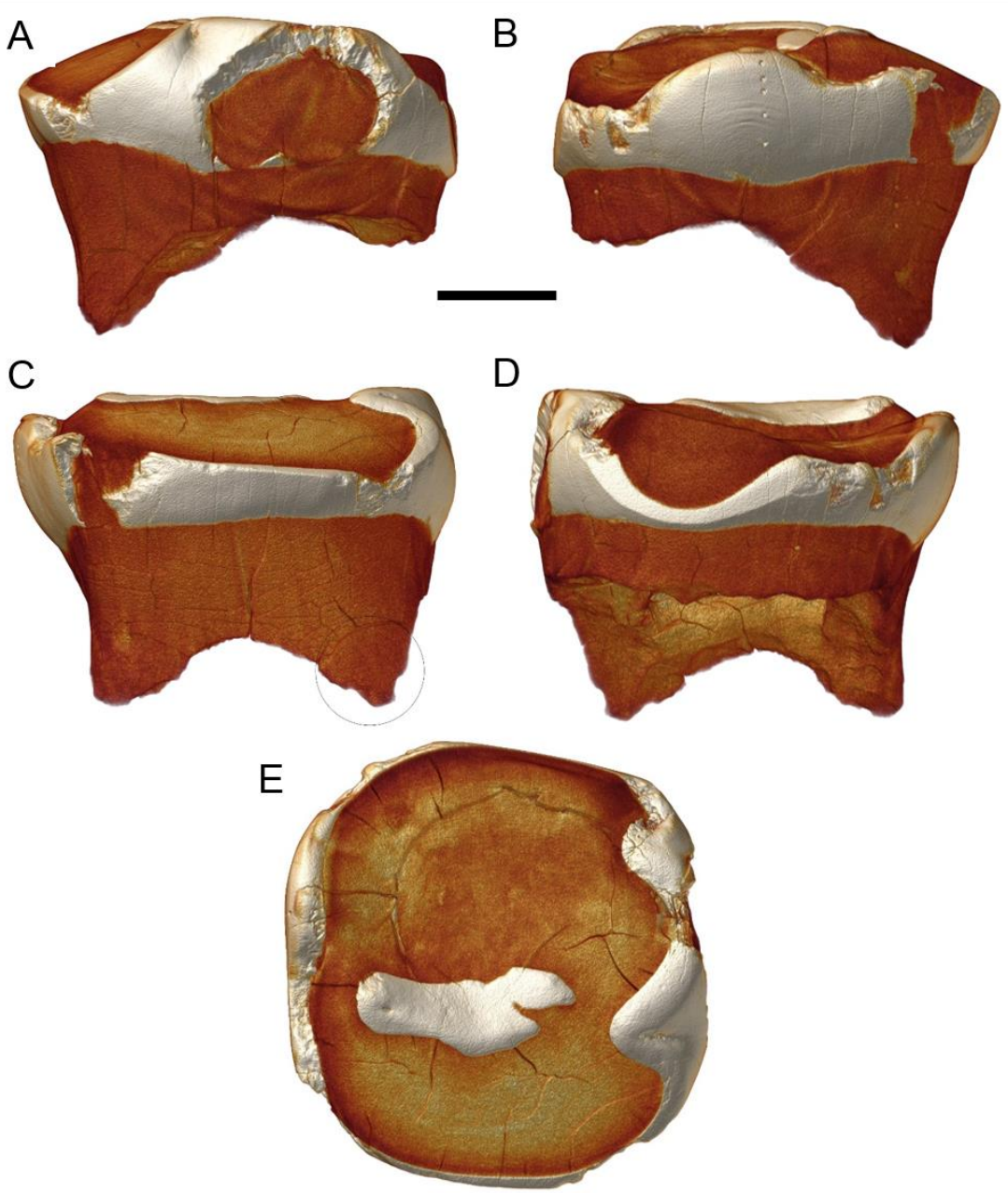

fig. S2. The *Denisova 2* specimen. µCT reconstruction of the *Denisova 2* deciduous lower second molar in (**A**) buccal, (**B**) lingual, (**C**) mesial, (**D**) distal, and (**E**) occlusal views. The scale for all panels is 2.5mm. The area sampled for ancient DNA analyses is marked by a gray circle in panel C.

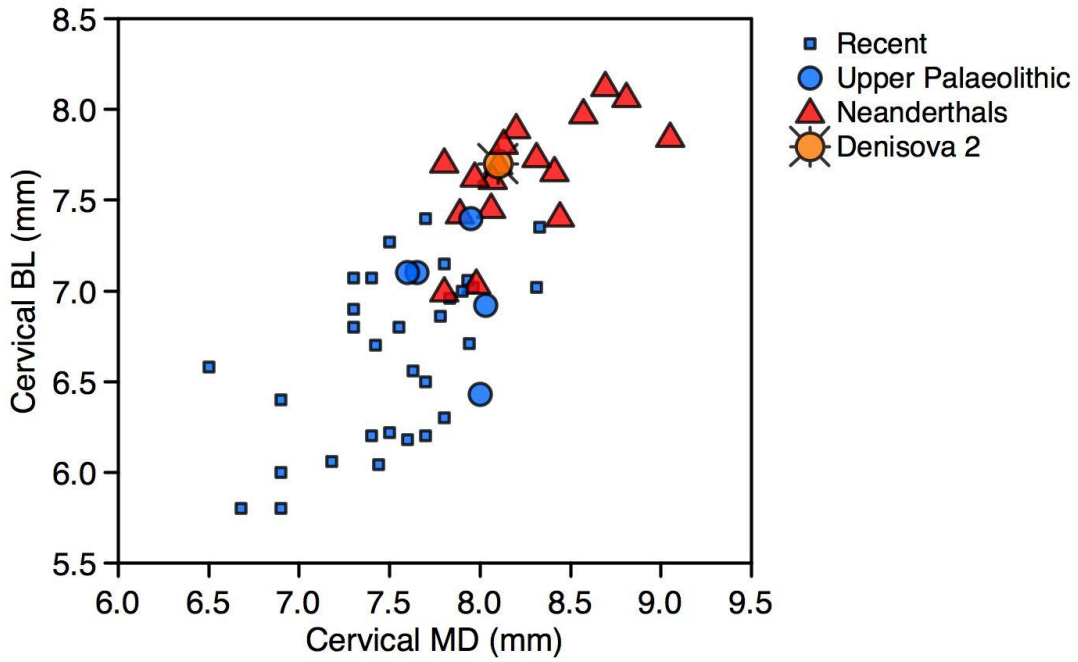

fig. S3. Biplot of cervical buccolingual (BL) and mesiodistal (MD) diameters.

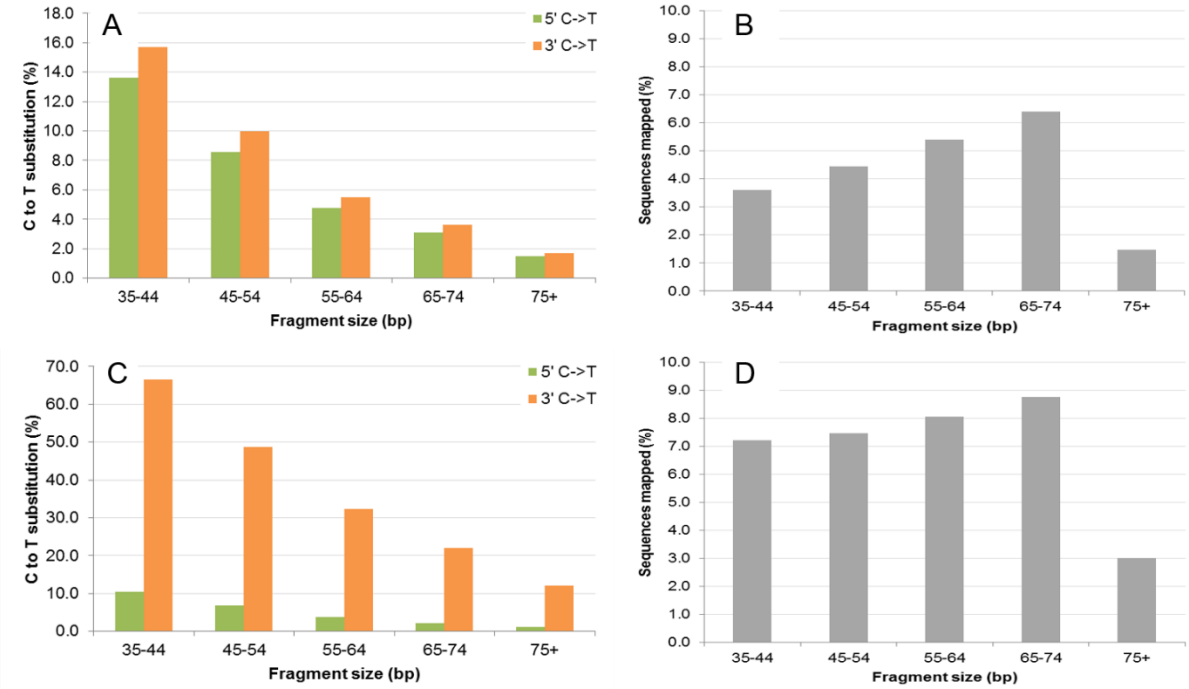

fig. S4. Characteristic of DNA fragments from the two libraries prepared from the *Denisova 2* specimen, per size bin. The frequency of cytosine (C) to thymine (T) substitutions and the percentage of sequences mapped to the human reference genome (out of sequences of the relevant size bin) for library A4891 (panels **A** and **B**) and for library A4934 (panels **C** and **D**) are shown. Note that the Y-axis has a different scale in panels A and C. In library A4934 (panel C), the frequency of C to T substitutions is markedly higher on the 3'- than on the 5'-end of sequences, as DNA fragments carrying a uracil at their 3'-ends were enriched for during library preparation. For both libraries, the frequency of nucleotide substitutions typical of ancient DNA decreases as the length of the DNA fragment increases, and the percentage of usable sequences is reduced in sequences longer than 74bp. Subsequent analyses were thus restricted to sequences shorter than 75 bp.

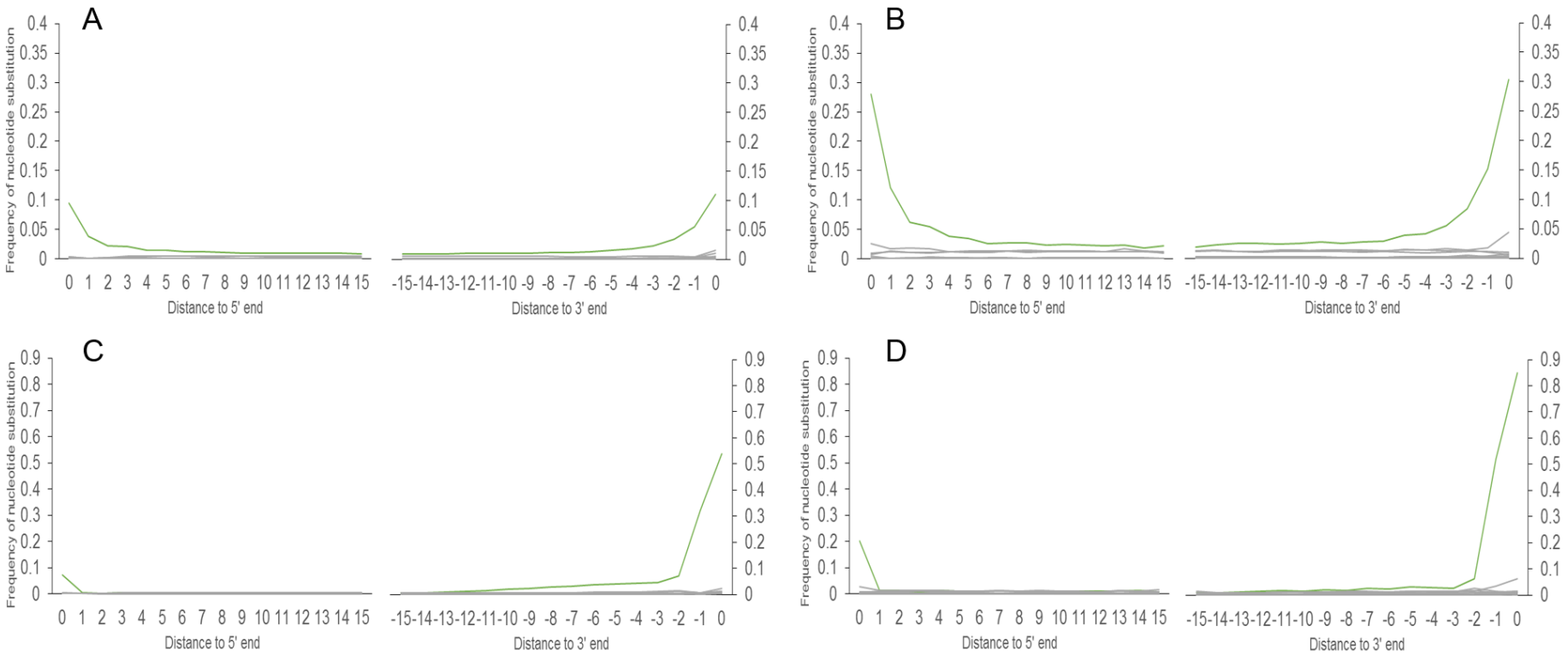

fig. S5. Frequency of nucleotide substitutions in the *Denisova 2* sequences. Cytosine to thymine substitutions are marked in green, all other types of substitutions in gray. (**A**) Nuclear DNA sequences from library A4891; (**B**) mtDNA sequences originating from shotgun sequencing (library A4891) and from capture for mitochondrial sequences (library A4928); (**C**) Nuclear DNA sequences from library A4934; (**D**) mtDNA sequences from library A4934. In library A4934 (panels C-D), the frequency of cytosine to thymine substitutions is markedly higher on the 3'- than on the 5'-end of sequences, as DNA fragments carrying a uracil at their 3'-ends were enriched for during library preparation. Note that the Y-axis scale differs between panels A-B and panels C-D.

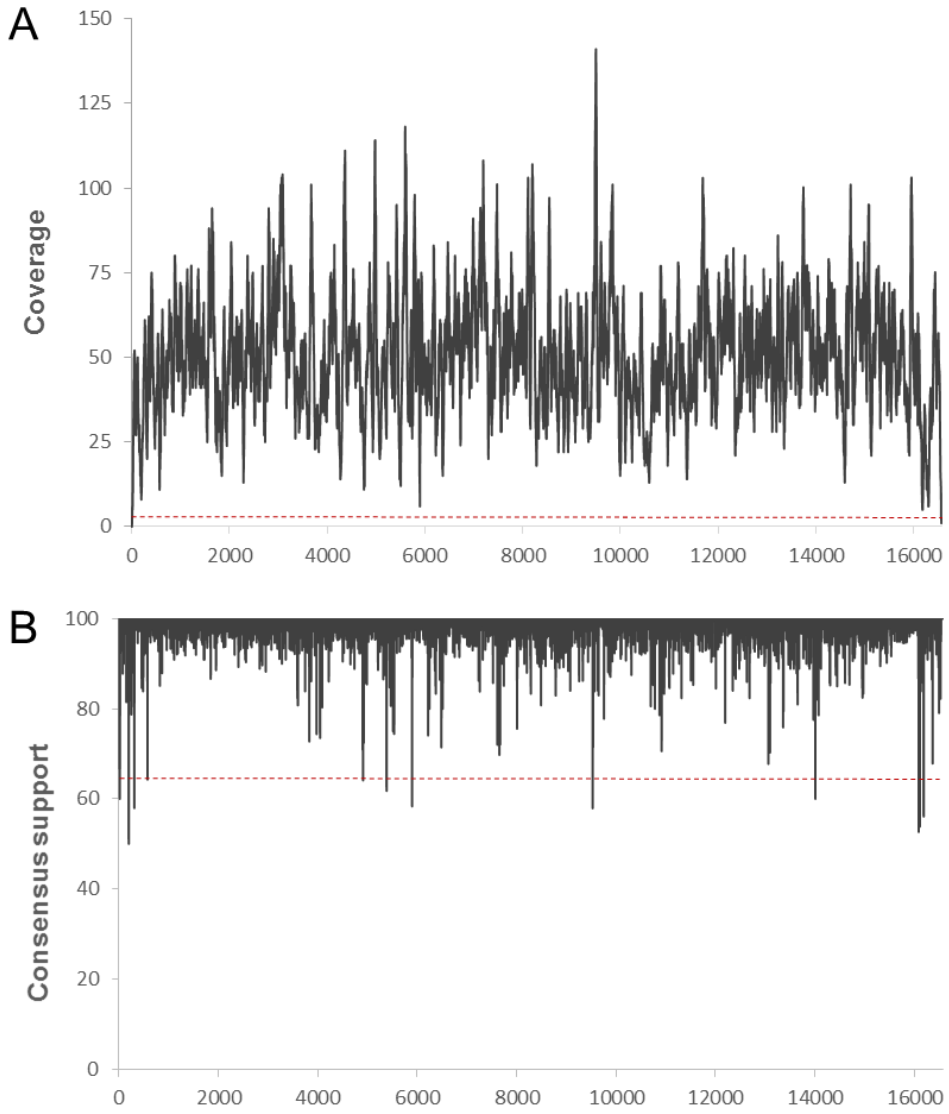

fig. S6. Reconstructing the *Denisova 2* mitochondrial genome. Sequences mapped to the *Denisova 3* mitochondrial genome and carrying C to T substitutions to it were utilized to determine the mtDNA sequence of *Denisova 2*. The number of sequences overlapping each position in the mitochondrial genome (coverage) (**A**) and the percentage of sequences carrying an identical base (consensus support) (**B**) are shown. The dashed red lines indicate the minimum coverage and minimum support required to call a consensus base, respectively.

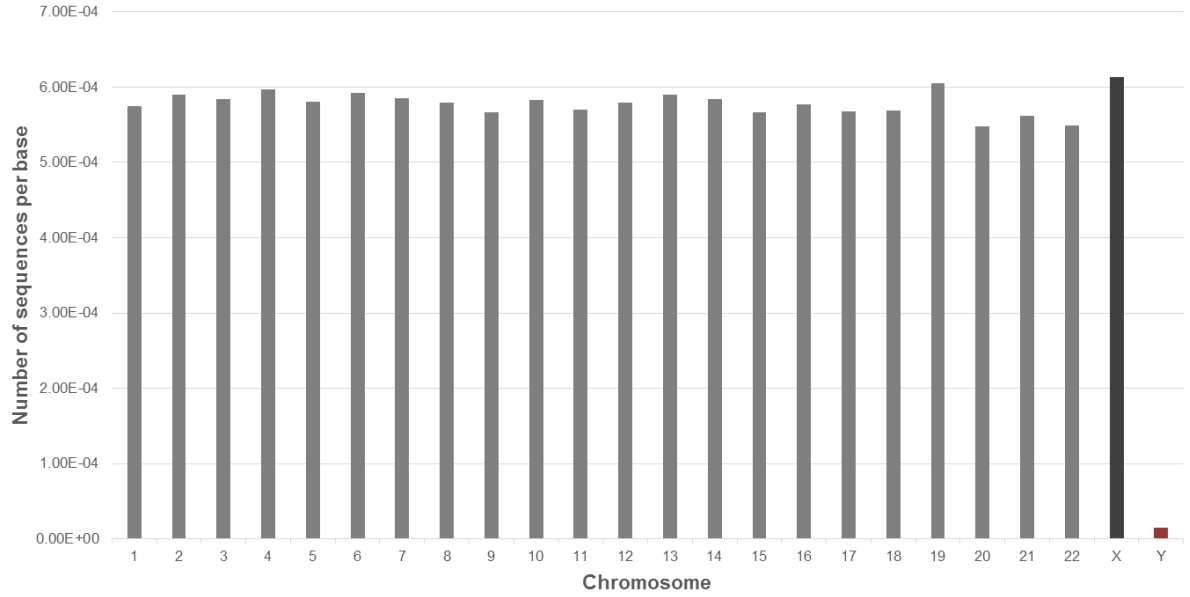

fig. S7. Sex determination of *Denisova 2*. The number of mapped sequences was divided by the number of positions within the alignability track to determine the number of sequences per base, for each chromosome. Only sequences presenting a C to T substitution relative to the reference genome at one of their ends were used. The ratio between the number of sequences per base mapping to the X chromosome (dark gray) and the number of sequences per base mapping to the autosomes (light gray) is 1.06, indicating a female. The low number of sequences mapping to the Y chromosome (red) is evidence for residual male contamination in the sample.

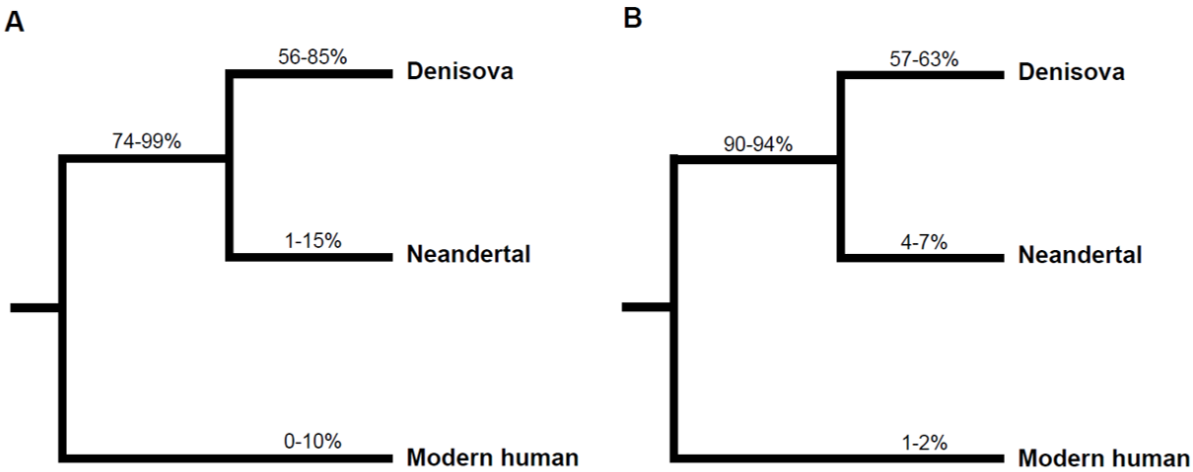

fig. S8. Lineage attribution of two other low-coverage Denisovan individuals. 95% binomial confidence intervals of the proportion of sequences sharing the derived state with the Denisovan, Neandertal or modern human branch for (**A**) *Denisova 4* and (**B**) *Denisova 8* are shown. Comparative sequence data was taken from (*3*) for the high-coverage Denisovan, (*7*) for the Neandertal and present-day human, and from (*17*) for *Denisova 4* and *8*.

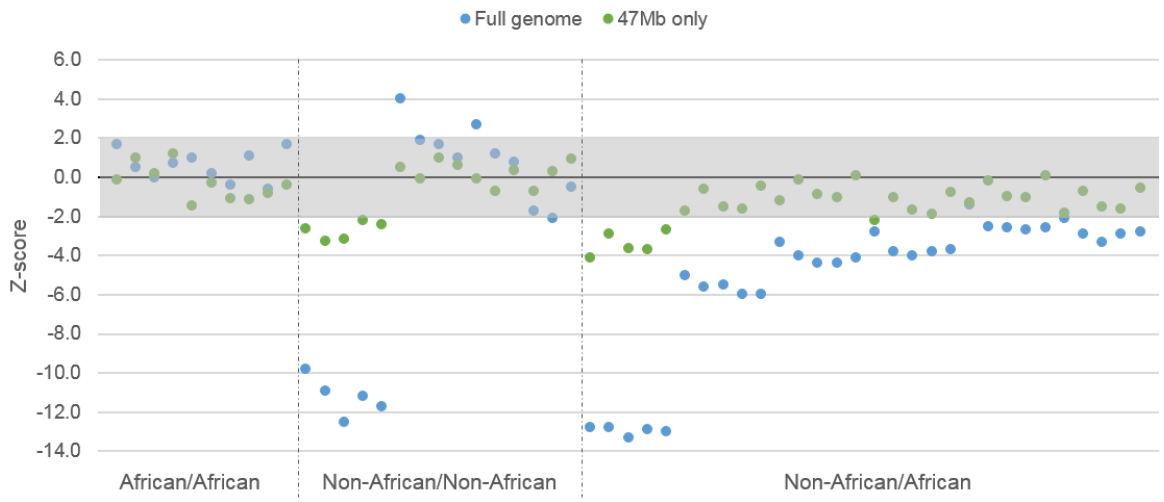

fig. S9. Testing our power to detect the sharing of derived alleles between Denisovans and present-day humans. Z-scores for the statistic D(H1,H2,*Denisova 3*,chimpanzee) are shown using all sequences from the high-coverage *Denisova 3* genome (blue) and using a random subset of sequences covering 47Mb (green). Z-scores for the full *Denisova 3* dataset were taken from table S24 in (*3*). Each score was multiplied by (-1), as the numerator for the statistic D was computed as (BABA-ABBA) there, but as (ABBA-BABA) in the current study. Non-significant Z-scores (|Z|<2) are marked in the shaded gray area.

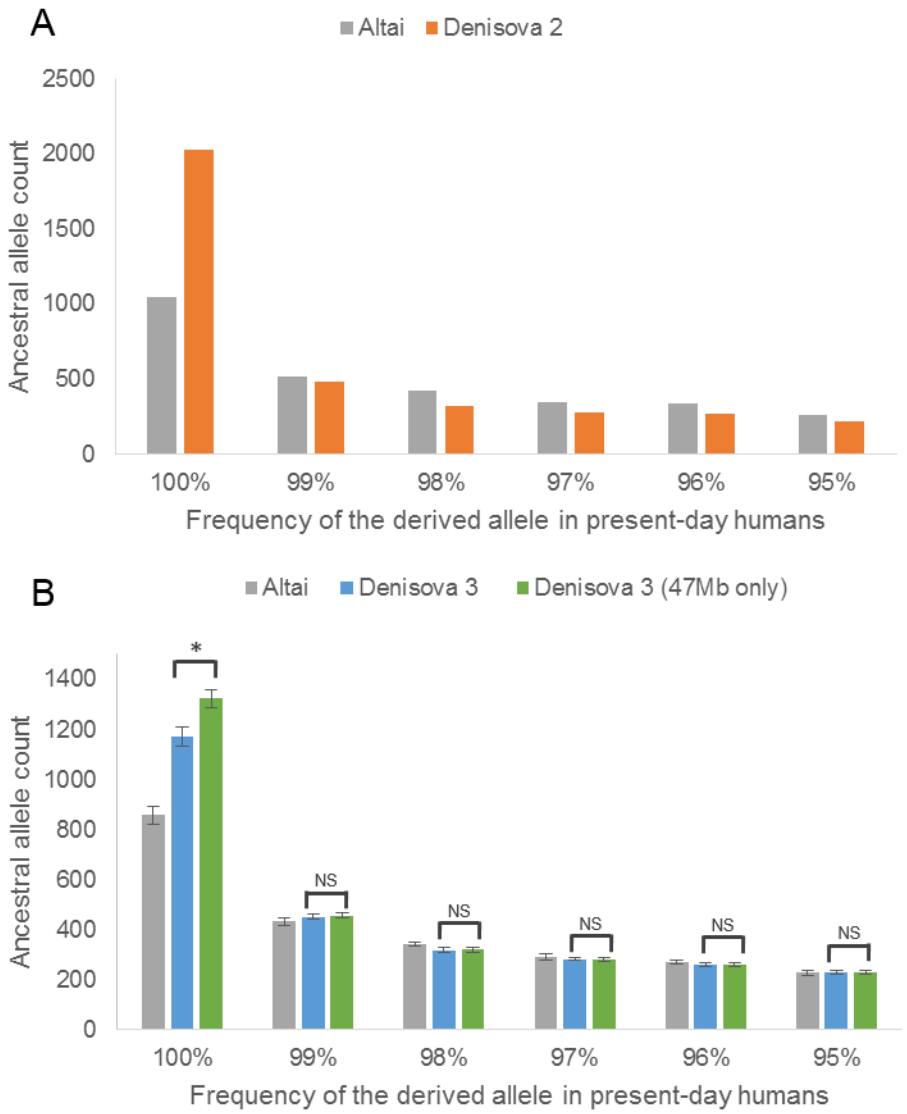

fig. S10. Ancestral allele counts in Denisovan and Neandertal sequences at sites that are fixed or nearly fixed for a derived allele in present-day humans. The frequency of a derived allele in present-day humans was taken from the 1000 Genomes panel (*81*). (**A**) There is an excess of ancestral alleles in *Denisova 2* (orange) compared to the high-coverage Altai Neandertal (*7*) (gray) at sites which are fixed derived in modern humans. (**B**) This signal is significantly increased when using only a subset of reads from the high-coverage *Denisova 3* genome (green), compared to using the high-quality genotypes of the same individual (*3*) (blue) at the same positions. The mean values (±standard deviation) from five subsampling schemes are represented. Differences were tested using the Mann-Whitney U test with a significance level of 0.05 (* - significant; NS - non-significant).

table S1. Metric comparisons of cervical and maximum mesiodistal and buccolingual diameters of *Denisova 2*, Neandertals, and recent and Upper Paleolithic modern humans.

	md	bl	cerv md	cerv bl
Recent	9.96±0.29 (N=34)	8.44±0.38 (N=35)	7.51±0.44 (N=31)	6.68±0.47 (N=31)
Denisova 2	10.3*	9.3*	8.1	7.7
Neandertals	10.49±0.78 (N=21)	9.56±0.57 (N=21)	8.25±0.35 (N=18)	7.63±0.33 (N=18)
Upper Palaeolithic	10.28±0.69 (N=27)	9.01±0.54 (N=27)	7.92±0.26 (N=6)	6.99±0.36 (N=5)

md – mesiodistal; bl –buccolingual; cerv md – cervical mesiodistal; cerv bl – cervical buccolingual.
* - indicates that the number constitutes a minimum measurement.

table S2. Characteristics of the DNA libraries prepared from *Denisova 2*.

Library ID	Library preparation method	Number of molecules in the library	Indexed library ID	Captured library ID	Aligned to the human reference genome					Aligned to the *Denisova 3* mtDNA genome		
					Number of raw sequences	Number of mapped sequences	Number of unique mapped sequences	Number of sequences with terminal C to T substitution	Number of unique sequences with terminal C to T substitution	Number of raw sequences	Number of unique mapped sequences	Number of unique sequences with terminal C to T substitution
A4881	ssDNA	7.27E+08	A4891	-	700,706,507	12,054,275	9,924,793	403,829	333,718	5,999,250	31,695	6,397
				A4928	-	-	-	-	-	139,652	39,551	5,878
A4934	Mini-U-selection	1.87E+08	A4944	-	144,912,352	4,441,871	1,288,012	668,621	194,397	3,482,260	4,136	2,436
			A4945	-	162,480,537	4,835,989	1,329,954	733,162	201,078	3,902,662	4,131	2,424
			A4946	-	136,172,058	4,134,104	1,284,584	621,888	194,020	2,885,235	4,081	2,483
			A4947	-	159,963,730	4,956,500	1,020,832	752,983	154,871	3,869,897	3,194	1,919

ssDNA – single-stranded DNA; C – cytosine; T – thymine; U – uracil; mtDNA – mitochondrial DNA.

The number of molecules was estimated using quantitative PCR for library A4881 and by digital droplet PCR for A4934. Library A4934 was indexed and amplified in four separate reactions. The number of sequences was computed after combining reads from all sequencing runs. For nuclear data, unique sequences (*i.e.*, following the removal of PCR duplicates) were tallied only if they were between 35 and 75 bases long, and mapped to the human reference genome (within the alignability track utilized) with a mapping quality higher than 30. Sequences were considered to contain a terminal C to T substitution if such a change occurred at the 5'- or 3'-end of a DNA fragment. For mitochondrial DNA, unique sequences mapping to the *Denisova 3* mtDNA were counted if they were between 30 and 75 bases long. Sequences were considered to contain a terminal C to T substitution if such a change occurred in the first or last three positions of reads originating from the single-stranded DNA library preparation (A4891 and A4928); and in the first or last two positions for reads from the libraries enriched for uracil-containing DNA fragments (A4944-A4947).

table S3. Frequencies of terminal C-to-T substitutions to the human reference genome. The percentage of sequences carrying a terminal C to T substitution is shown before and after filtering for the presence of a terminal C to T substitution on the other end of the fragment ('conditional' substitution), for each type of library and for the combined dataset. The contamination estimate is based on the ratio between the average C to T substitution frequency before and after the conditional filtering.

	All sequences		Conditional substitution		
Library preparation method	C to T on 5' end [%]	C to T on 3' end [%]	C to T on 5' end [%]	C to T on 3' end [%]	Contamination estimate
ssDNA	9.4	11.1	37.1	43.2	74.5%
Mini-U-selection	7.3	53.8	26.3	93.4	49.0%
Combined dataset	8.8	32.0	28.4	72.2	59.4%

ssDNA – single-stranded DNA; C – cytosine; T – thymine; U - uracil.

table S4. Percentage and number of sequences matching the derived state of mitochondrial genomes from three hominin groups. Sequences from *Denisova 2*, which overlap three sets of phylogenetically diagnostic positions, were utilized. Results are shown before and after filtering for sequences with a terminal C to T substitution.

		All mitochondrial DNA sequences		Sequences with a terminal C to T substitution	
Test group	Number of group-specific positions	Sequences matching test group [%] *(observations)*	Sequences matching alternative state [%] *(observations)*	Sequences matching test group [%] *(observations)*	Sequences matching alternative state [%] *(observations)*
Modern human	14	82.4 *(450/546)*	17.6 *(96/546)*	4.2 *(1/24)*	95.8 *(23/24)*
Denisova	38	16.9 *(242/1,436)*	83.1 *(1,194/1,436)*	93.8 *(30/32)*	6.3 *(2/32)*
Neandertal	19	0.0 *(0/836)*	100.0 *(836/836)*	0.0 *(0/48)*	100.0 *(48/48)*

C – cytosine; T – thymine

table S5. Number of base differences between the *Denisova 2* mtDNA and other mtDNAs.
Comparative sequence data was taken from [a] (*16*); [b] (*2*); [c] (*17*); [d] (*7*); [e] (*58*); [f] (*59*); [g] (*57*); [h] (*34*); [i] (*54*); [j] (*56*); [k] (*55*); [l] (*4*); [m] (*6*); [n] (*39*); [o] (*60*).

Comparative data	Accession code	Number of base differences	Comparative data	Accession code	Number of base differences
Denisovans			**Present-day modern humans**		
Denisova 3 (Siberia) [a]	NC013993	72	Chinese [i]	AF346972	353
Denisova 4 (Siberia) [b]	FR695060	70	French [i]	AF346981	347
Denisova 8 (Siberia) [c]	KT780370	29	Papuan [i]	AF347004	354
Neandertals			San [i]	AF347008	345
Altai (Siberia) [d]	KC879692	335	Yoruba [i]	AF347014	349
El Sidron 1253 (Spain) [e]	FM865409	343	**Ancient modern humans**		
Feldhofer 1 (Germany) [e]	FM865407	345	Dolni Vestonice 13 (Czech Republic) [j]	KC521459	345
Feldhofer 2 (Germany) [e]	FM865408	337	Dolni Vestonice 14 (Czech Republic) [j]	KC521458	345
Mezmaiskaya 1 (Caucasus) [e]	FM865411	339	Kostenki 14 (Russia) [k]	FM600416	347
Okladnikov 2 (Siberia) [f]	KF982693	330	Tianyuan (China) [l]	KC417443	345
Vindija 33.16 (Croatia) [g]	AM948965	345	Ust'-Ishim (Siberia) [m]	PRJEB6622	342
Vindija 33.17 (Croatia) [h]	KJ533544	344	**Middle Pleistocene hominins**		
Vindija 33.19 (Croatia) [h]	KJ533545	345	Sima de los Huesos (Spain) [n]	NC023100	185
Vindija 33.25 (Croatia) [e]	FM865410	345	**Chimpanzee** [o]	NC001643	1,451

table S6. Estimates of DNA divergence between low-coverage archaic genomes and the high-coverage genomes of a Denisovan, a Neandertal, and 12 present-day humans. The number of bases in the human genome covered by the low-coverage archaic genomes following our filtering scheme is noted in brackets. For each pair of genomes, the percentage of substitutions inferred to have occurred on the branch from the human-chimpanzee ancestral sequences to the high-coverage genome after the split from the low-coverage genome is shown. 95% confidence intervals shown in parentheses were calculated by jackknifing over 5Mb windows.

Low-coverage archaic genome [data available]	High-coverage genome													
	Denisovan	Neandertal	Yoruba	San	Dinka	Mbuti	Mandenka	French	Sardinian	Han	Dai	Karitiana	Australian	Papuan
Denisova 2 [47Mb]	5.9 *(5.6-6.2)*	9.4 *(9.0-9.9)*	11.4 *(10.9-11.8)*	11.6 *(11.1-12.1)*	11.5 *(11.0-12.0)*	11.4 *(10.9-11.9)*	11.6 *(11.1-12.1)*	11.0 *(10.5-11.5)*	11.2 *(10.7-11.7)*	11.2 *(10.7-11.6)*	11.2 *(10.7-11.7)*	11.0 *(10.5-11.4)*	10.9 *(10.5-11.4)*	11.1 *(10.6-11.5)*
Denisova 4 [0.5Mb]	4.3 *(2.3-6.7)*	9.1 *(6.3-12.2)*	12.4 *(9.2-15.7)*	11.8 *(8.5-15.2)*	11.1 *(7.9-14.5)*	11.2 *(8.2-14.5)*	13.5 *(10.1-17.0)*	11.9 *(8.8-15.2)*	11.5 *(8.5-14.7)*	11.7 *(8.6-14.9)*	10.1 *(7.0-13.5)*	11.9 *(8.8-15.2)*	9.9 *(6.9-12.9)*	11.0 *(7.9-14.2)*
Denisova 8 [13Mb]	4.3 *(4.0-4.7)*	8.6 *(8.0-9.1)*	11.2 *(10.5-11.9)*	11.7 *(11.0-12.5)*	11.5 *(10.8-12.3)*	11.7 *(11.0-12.4)*	11.5 *(10.8-12.2)*	11.4 *(10.7-12.1)*	11.1 *(10.4-11.8)*	11.4 *(10.7-12.0)*	11.1 *(10.4-11.8)*	11.6 *(10.9-12.3)*	11.3 *(10.7-12.1)*	11.4 *(10.7-12.1)*
Feldhofer 1 [0.1Mb]	13.8 *(7.5-20.8)*	5.7 *(1.7-10.8)*	16.6 *(10.3-23.3)*	13.7 *(8.0-19.8)*	13.8 *(8.2-20.0)*	13.0 *(7.0-19.2)*	14.5 *(8.3-20.9)*	9.1 *(4.1-14.8)*	10.7 *(5.6-16.1)*	9.1 *(4.2-14.2)*	13.0 *(7.2-19.4)*	13.8 *(7.3-21.0)*	15.1 *(8.6-22.0)*	11.3 *(5.7-17.2)*
Mezmaiskaya 1 [7Mb]	10.1 *(9.3-10.8)*	2.9 *(2.6-3.1)*	12.3 *(11.4-13.2)*	12.1 *(11.2-13.0)*	12.2 *(11.3-13.1)*	12.3 *(11.5-13.2)*	12.3 *(11.5-13.2)*	11.9 *(11.1-12.8)*	12.0 *(11.2-12.8)*	12.0 *(11.1-12.8)*	12.1 *(11.3-13.0)*	12.0 *(11.1-12.9)*	12.0 *(11.2-12.9)*	11.9 *(11.0-12.7)*
El Sidron 1253 [0.1Mb]	3.6 *(1.2-6.7)*	3.5 *(1.1-6.5)*	9.5 *(5.4-14.0)*	15.4 *(10.1-20.9)*	12.1 *(7.2-17.4)*	13.4 *(8.6-18.5)*	16.5 *(11.0-22.3)*	8.8 *(4.9-13.3)*	9.5 *(5.0-14.3)*	16.8 *(11.4-22.5)*	12.2 *(7.6-17.0)*	11.6 *(7.1-16.3)*	11.6 *(7.1-16.5)*	12.2 *(7.5-17.2)*
Vindija 33.16 [73Mb]	10.3 *(9.8-10.8)*	3.9 *(3.7-4.1)*	12.3 *(11.7-12.9)*	12.4 *(11.8-12.9)*	12.0 *(11.4-12.5)*	12.3 *(11.7-12.8)*	12.2 *(11.6-12.7)*	11.6 *(11.1-12.1)*	11.7 *(11.2-12.3)*	11.7 *(11.2-12.3)*	11.7 *(11.2-12.3)*	11.6 *(11.1-12.2)*	11.7 *(11.1-12.2)*	11.6 *(11.1-12.1)*
Vindija 33.25 [59Mb]	10.3 *(9.8-10.8)*	3.7 *(3.5-3.9)*	12.1 *(11.6-12.6)*	12.1 *(11.6-12.6)*	12.1 *(11.6-12.7)*	12.1 *(11.5-12.6)*	12.1 *(11.6-12.6)*	11.7 *(11.2-12.3)*	11.6 *(11.1-12.1)*	11.7 *(11.2-12.2)*	11.7 *(11.2-12.3)*	11.4 *(10.9-12.0)*	11.5 *(11.0-12.0)*	11.5 *(11.0-12.0)*
Vindija 33.26 [61Mb]	10.5 *(10.1-11.0)*	4.0 *(3.7-4.2)*	12.2 *(11.7-12.7)*	12.3 *(11.8-12.8)*	12.1 *(11.6-12.6)*	12.2 *(11.7-12.7)*	12.4 *(11.9-12.9)*	11.8 *(11.3-12.3)*	11.8 *(11.3-12.3)*	11.8 *(11.3-12.3)*	11.8 *(11.3-12.3)*	11.9 *(11.4-12.4)*	11.7 *(11.2-12.2)*	11.8 *(11.3-12.3)*

Mb - megabases

table S7. Estimates of DNA divergence between subsampled genomes and the high-coverage genomes of a Denisovan, a Neandertal, and 12 present-day humans. For each pair of genomes, the percentage of substitutions inferred to have occurred on the branch from the human-chimpanzee ancestral sequences to the high-coverage genome after the split from the subsampled genome is shown. 95% confidence intervals were calculated by jackknifing over 5Mb windows. The percentages were computed using only positions covered by at least one of the low-coverage archaic genomes from table S6.

Subsampled genome	High-coverage genome													
	Denisovan	Neandertal	Yoruba	San	Dinka	Mbuti	Mandenka	French	Sardinian	Han	Dai	Karitiana	Australian	Papuan
Denisovan	-	7.9 *(7.5-8.2)*	11.2 *(10.8-11.7)*	11.3 *(10.8-11.7)*	11.2 *(10.7-11.6)*	11.2 *(10.8-11.7)*	11.3 *(10.8-11.7)*	10.9 *(10.5-11.4)*	11.0 *(10.5-11.4)*	11.0 *(10.5-11.4)*	11.0 *(10.5-11.4)*	10.9 *(10.5-11.4)*	10.7 *(10.3-11.2)*	10.8 *(10.4-11.2)*
Neandertal	8.8 *(8.5-9.2)*	-	10.9 *(10.4-11.3)*	10.9 *(10.5-11.3)*	10.8 *(10.3-11.2)*	10.9 *(10.4-11.3)*	10.9 *(10.5-11.3)*	10.4 *(10.0-10.9)*	10.5 *(10.0-10.9)*	10.5 *(10.1-10.9)*	10.5 *(10.1-10.9)*	10.4 *(10.0-10.8)*	10.4 *(10.0-10.8)*	10.4 *(10.0-10.8)*
Yoruba	10.9 *(10.5-11.3)*	9.6 *(9.2-10.0)*	-	8.8 *(8.4-9.2)*	7.8 *(7.5-8.1)*	8.4 *(8.1-8.8)*	7.8 *(7.5-8.2)*	7.7 *(7.4-8.1)*	7.8 *(7.4-8.1)*	7.9 *(7.5-8.2)*	7.8 *(7.5-8.2)*	7.8 *(7.5-8.1)*	7.9 *(7.5-8.2)*	7.9 *(7.6-8.2)*
San	10.9 *(10.5-11.3)*	9.6 *(9.2-10.0)*	8.8 *(8.4-9.1)*	-	8.7 *(8.3-9.0)*	8.7 *(8.4-9.1)*	8.8 *(8.5-9.2)*	8.5 *(8.1-8.9)*	8.5 *(8.1-8.8)*	8.6 *(8.2-9.0)*	8.6 *(8.2-8.9)*	8.6 *(8.2-8.9)*	8.6 *(8.3-9.0)*	8.6 *(8.3-9.0)*
Dinka	10.9 *(10.5-11.3)*	9.5 *(9.2-9.9)*	7.9 *(7.6-8.2)*	8.8 *(8.4-9.1)*	-	8.3 *(8.0-8.7)*	7.9 *(7.6-8.2)*	7.5 *(7.2-7.9)*	7.5 *(7.2-7.9)*	7.6 *(7.3-8.0)*	7.6 *(7.3-7.9)*	7.6 *(7.3-7.9)*	7.6 *(7.3-8.0)*	7.7 *(7.3-8.0)*
Mbuti	10.9 *(10.5-11.3)*	9.6 *(9.2-10.0)*	8.4 *(8.1-8.8)*	8.8 *(8.4-9.1)*	8.2 *(7.9-8.6)*	-	8.4 *(8.1-8.8)*	8.3 *(7.9-8.6)*	8.2 *(7.9-8.6)*	8.3 *(8.0-8.7)*	8.3 *(7.9-8.6)*	8.3 *(7.9-8.6)*	8.3 *(8.0-8.7)*	8.3 *(8.0-8.7)*
Mandenka	10.9 *(10.4-11.3)*	9.6 *(9.2-9.9)*	7.8 *(7.5-8.1)*	8.8 *(8.4-9.2)*	7.8 *(7.5-8.1)*	8.4 *(8.1-8.8)*	-	7.6 *(7.3-7.9)*	7.7 *(7.4-8.0)*	7.8 *(7.5-8.2)*	7.8 *(7.5-8.1)*	7.7 *(7.4-8.1)*	7.8 *(7.5-8.1)*	7.8 *(7.5-8.2)*
French	10.8 *(10.4-11.3)*	9.4 *(9.0-9.8)*	8.0 *(7.7-8.3)*	8.8 *(8.4-9.2)*	7.7 *(7.4-8.1)*	8.5 *(8.2-8.9)*	7.9 *(7.6-8.3)*	-	5.8 *(5.6-6.1)*	6.3 *(6.1-6.6)*	6.3 *(6.0-6.6)*	6.2 *(5.9-6.5)*	6.5 *(6.2-6.8)*	6.5 *(6.2-6.7)*
Sardinian	10.8 *(10.4-11.2)*	9.3 *(9.0-9.7)*	7.9 *(7.6-8.3)*	8.7 *(8.3-9.1)*	7.7 *(7.3-8.0)*	*8.4* *(8.0-8.8)*	7.9 *(7.6-8.3)*	5.8 *(5.5-6.0)*	-	*6.4* *(6.2-6.7)*	6.3 *(6.1-6.6)*	*6.3* *(6.0-6.5)*	6.5 *(6.2-6.8)*	6.5 *(6.2-6.8)*
Han	10.7 *(10.3-11.2)*	9.3 *(8.9-9.7)*	8.0 *(7.6-8.3)*	8.7 *(8.4-9.1)*	7.7 *(7.4-8.0)*	8.4 *(8.1-8.8)*	8.0 *(7.6-8.3)*	6.2 *(5.9-6.5)*	6.4 *(6.1-6.6)*	-	5.4 *(5.2-5.7)*	5.7 *(5.5-6.0)*	6.1 *(5.9-6.4)*	6.1 *(5.9-6.4)*
Dai	10.8 *(10.4-11.2)*	9.4 *(9.0-9.7)*	8.0 *(7.6-8.3)*	8.8 *(8.4-9.2)*	*7.7* *(7.4-8.0)*	8.4 *(8.1-8.8)*	8.0 *(7.6-8.3)*	6.2 *(5.9-6.5)*	6.3 *(6.0-6.6)*	5.5 *(5.2-5.7)*	-	5.7 *(5.5-6.0)*	6.1 *(5.8-6.4)*	6.1 *(5.9-6.4)*
Karitiana	10.8 *(10.3-11.2)*	9.3 *(8.9-9.7)*	8.0 *(7.6-8.3)*	8.8 *(8.4-9.1)*	7.7 *(7.4-8.0)*	8.4 *(8.1-8.8)*	7.9 *(7.6-8.3)*	6.1 *(5.8-6.4)*	6.2 *(6.0-6.5)*	5.8 *(5.6-6.1)*	5.8 *(5.5-6.0)*	-	6.2 *(5.9-6.4)*	6.1 *(5.9-6.4)*
Australian	10.6 *(10.2-11.0)*	9.3 *(8.9-9.7)*	8.1 *(7.7-8.4)*	8.9 *(8.5-9.2)*	7.8 *(7.5-8.1)*	8.5 *(8.2-8.9)*	8.0 *(7.7-8.4)*	6.4 *(6.2-6.7)*	6.5 *(6.2-6.8)*	6.2 *(5.9-6.5)*	6.1 *(5.9-6.4)*	6.2 *(5.9-6.5)*	-	5.4 *(5.1-5.6)*
Papuan	10.7 *(10.2-11.1)*	9.3 *(8.9-9.7)*	8.1 *(7.8-8.5)*	8.9 *(8.5-9.2)*	7.8 *(7.5-8.1)*	8.5 *(8.2-8.9)*	8.1 *(7.7-8.4)*	6.4 *(6.1-6.7)*	6.5 *(6.2-6.8)*	6.2 *(6.0-6.5)*	6.2 *(5.9-6.5)*	6.2 *(5.9-6.5)*	5.4 *(5.1-5.6)*	-

table S8. Sharing of derived alleles between *Denisova 2* and present-day humans. The statistic D(H1,H2,*Denisova 2*,chimpanzee) computes the number of sites at which a derived allele from *Denisova 2* matches a present-day individual H1 (ABBA sites) or H2 (BABA sites) (*1, 80*).

H1	H2	ABBA sites	BABA sites	D	Z-score
Non-African/African					
Australian	Dinka	1214	1477	-0.098	**-5.1**
Australian	Mandenka	1151	1535	-0.143	**-7.5**
Australian	Mbuti	1204	1662	-0.160	**-8.6**
Australian	San	1272	1769	-0.163	**-9.0**
Australian	Yoruba	1151	1544	-0.146	**-7.6**
Dai	Dinka	1190	1339	-0.059	-3.0
Dai	Mandenka	1174	1442	-0.102	**-5.4**
Dai	Mbuti	1201	1542	-0.124	**-6.3**
Dai	San	1263	1648	-0.132	**-7.1**
Dai	Yoruba	1159	1436	-0.107	**-5.4**
French	Dinka	1228	1383	-0.059	-3.1
French	Mandenka	1170	1444	-0.105	**-5.7**
French	Mbuti	1203	1553	-0.127	**-6.6**
French	San	1257	1644	-0.133	**-7.5**
French	Yoruba	1205	1486	-0.104	**-5.6**
Han	Dinka	1180	1351	-0.068	-3.4
Han	Mandenka	1145	1432	-0.111	**-5.8**
Han	Mbuti	1163	1526	-0.135	**-6.9**
Han	San	1234	1638	-0.141	**-7.4**
Han	Yoruba	1150	1448	-0.115	**-5.7**
Karitiana	Dinka	1214	1416	-0.077	-3.4
Karitiana	Mandenka	1155	1472	-0.121	**-6.4**
Karitiana	Mbuti	1179	1570	-0.142	**-7.3**
Karitiana	San	1224	1658	-0.151	**-8.2**
Karitiana	Yoruba	1165	1494	-0.124	**-6.4**
Papuan	Dinka	1238	1470	-0.086	**-4.4**
Papuan	Mandenka	1165	1510	-0.129	**-7.0**
Papuan	Mbuti	1216	1640	-0.148	**-8.0**
Papuan	San	1259	1726	-0.156	**-8.6**
Papuan	Yoruba	1133	1493	-0.137	**-7.2**
Sardinian	Dinka	1150	1356	-0.082	**-4.2**
Sardinian	Mandenka	1125	1444	-0.124	**-6.5**
Sardinian	Mbuti	1198	1595	-0.142	**-7.4**
Sardinian	San	1208	1648	-0.154	**-8.4**
Sardinian	Yoruba	1141	1470	-0.126	**-6.6**
African/African					
Dinka	Mandenka	1230	1345	-0.045	-2.3
Dinka	Mbuti	1268	1462	-0.071	-3.6
Dinka	San	1312	1547	-0.082	**-4.5**
Dinka	Yoruba	1231	1358	-0.049	-2.6
Mandenka	Mbuti	1278	1351	-0.028	-1.4
Mandenka	San	1381	1498	-0.041	-2.2
Mandenka	Yoruba	1267	1277	-0.004	-0.2
Mbuti	San	1384	1426	-0.015	-0.8
Mbuti	Yoruba	1418	1354	0.023	1.2
San	Yoruba	1494	1388	0.037	1.9
Non-African/Non-African					
Australian	Dai	1015	1131	-0.054	-2.5
Australian	French	1091	1200	-0.048	-2.2
Australian	Han	1024	1119	-0.044	-2.0
Australian	Karitiana	1054	1121	-0.031	-1.4
Australian	Papuan	943	978	-0.018	-0.8
Australian	Sardinian	1058	1121	-0.029	-1.3
Dai	French	1043	1037	0.003	0.1
Dai	Han	904	885	0.011	0.4
Dai	Karitiana	974	921	0.028	1.2
Dai	Papuan	1091	1011	0.038	1.8
Dai	Sardinian	1053	999	0.026	1.2
French	Han	1077	1062	0.007	0.3
French	Karitiana	1069	1027	0.020	0.9
French	Papuan	1159	1083	0.034	1.6
French	Sardinian	993	944	0.025	1.1
Han	Karitiana	950	919	0.017	0.7
Han	Papuan	1091	1029	0.029	1.4
Han	Sardinian	1069	1037	0.015	0.7
Karitiana	Papuan	1070	1041	0.014	0.7
Karitiana	Sardinian	1056	1053	0.001	0.1
Papuan	Sardinian	1097	1124	-0.012	-0.6

The 'D' statistic is calculated as D=(ABBA-BABA)/ (ABBA+BABA); and the Z-score)is computed for $D \neq 0$. Highly significant scores ($|Z| > 4$) are highlighted in bold.

table S9. Denisova-specific nonsynonymous coding changes corroborated by sequences from *Denisova 2*, *4*, or *8*. Sequences were compared to the catalog of Denisova-specific single nucleotide changes (*3*). Information on putative function of the changes was taken from the UCSC Genome Browser (*82*).

Chromosome	Position	Affected gene	Putative function
2	18766036	NT5C1B	Catalyzes production of adenosine to regulate diverse physiologic processes
2	187372160	ZC3HF	Protects DRG1 from proteolytic degradation
2	206057945	PAR3L	Putative adapter protein involved in asymmetrical cell division and cell polarization processes
3	47467636	SCAP	Chaperone mediating the transport of sterol regulatory element binding proteins from the ER to the Golgi
4	124235172	SPATA5	Involved in morphological and functional mitochondrial transformations during spermatogenesis
5	55155404	IL31RA	Interleukin receptor
5	176798979	RGS14	Regulator of G-protein signaling
6	154763322	IPCEF1	Involved in regulation of signal transduction
6	167709632	UNC93A	N/A
7	158910097	VIPR2	Involved in smooth muscle relaxation, exocrine and endocrine secretion, and water and ion flux in lung and intestinal epithelia
8	10387071	PRSS55	Serine protease, primarily expressed in the Leydig and Sertoli cells of the testis and may be involved in male fertility
8	95850820	INTS8	Involved in the cleavage of small nuclear RNAs within the nucleus
9	135277474	TTF1	Transcription termination factor involved in ribosomal gene transcription
9	140139289	FAM166A	N/A
10	116048702	VWA2	Structural component in basement membranes or anchoring structures on scaffolds of collagen VII or fibrillin
11	45947993	GYLTL1B	Involved in metabolic process
11	123989362	VWA5A	May play a role as a tumor suppressor
13	25352441	RNF17	Based on similarity to a mouse gene, encodes a testis-specific protein
15	74426399	ISLR2	Involved in multicellular organismal development and regulation of axon extension
16	20809966	ERI2	Involved in metabolic process
16	66969396	CES2	May participate in fatty acyl and cholesterol ester metabolism, and putatively involved in the blood-brain barrier system
16	75298371	BCAR1	Involved in cellular migration, survival, transformation, and invasion
16	81198329	PKD1L2	May function as a component of cation channel pores
17	33448725	FNDC8	N/A
17	48596440	MYCBPAP	May play a role in spermatogenesis
19	44099400	IRGQ	Immunity-related GTPase
20	39802851	PLCG1	Involved in the intracellular transduction of tyrosine kinase activators