LETTER

doi:10.1038/nature17405

Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins

Matthias Meyer[1], Juan-Luis Arsuaga[2,3], Cesare de Filippo[1], Sarah Nagel[1], Ayinuer Aximu-Petri[1], Birgit Nickel[1], Ignacio Martínez[2,4], Ana Gracia[2,4], José María Bermúdez de Castro[5,6], Eudald Carbonell[7,8], Bence Viola[9], Janet Kelso[1], Kay Prüfer[1] & Svante Pääbo[1]

A unique assemblage of 28 hominin individuals, found in Sima de los Huesos in the Sierra de Atapuerca in Spain, has recently been dated to approximately 430,000 years ago[1]. An interesting question is how these Middle Pleistocene hominins were related to those who lived in the Late Pleistocene epoch, in particular to Neanderthals in western Eurasia and to Denisovans, a sister group of Neanderthals so far known only from southern Siberia. While the Sima de los Huesos hominins share some derived morphological features with Neanderthals, the mitochondrial genome retrieved from one individual from Sima de los Huesos is more closely related to the mitochondrial DNA of Denisovans than to that of Neanderthals[2]. However, since the mitochondrial DNA does not reveal the full picture of relationships among populations, we have investigated DNA preservation in several individuals found at Sima de los Huesos. Here we recover nuclear DNA sequences from two specimens, which show that the Sima de los Huesos hominins were related to Neanderthals rather than to Denisovans, indicating that the population divergence between Neanderthals and Denisovans predates 430,000 years ago. A mitochondrial DNA recovered from one of the specimens shares the previously described relationship to Denisovan mitochondrial DNAs, suggesting, among other possibilities, that the mitochondrial DNA gene pool of Neanderthals turned over later in their history.

When modern humans spread out of Africa and the Near East some 75,000–50,000 years ago, at least two archaic hominin groups, Neanderthals and Denisovans, inhabited Eurasia. While Neanderthals are known from an abundant fossil record in Europe and western and central Asia, Denisovan remains are currently only known from the Altai Mountains in southern Siberia[3,4]. However, Denisovan ancestry is detected in present-day human populations from Oceania, mainland Asia and in Native Americans[5], suggesting that they were once more widespread. High-quality genome sequences recovered from one Neanderthal and one Denisovan show that they were more closely related to each other than to modern humans[6,7] and that they diverged from a common ancestral population between 381,000 and 473,000 years ago[7] if a mutation rate of 0.5×10^{-9} per site per year is used.

The Middle Pleistocene fossils from Sima de los Huesos (SH) are relevant for the question of when and where the ancestral populations of Neanderthals and Denisovans lived, but their relationship to these later archaic groups is unclear. They share some derived dental and cranial features with Late Pleistocene Neanderthals, for example, a midfacial prognathism and some aspects of the supraorbital torus, the occipital bone and the glenoid cavity[1,8]. In apparent contrast to this, the mitochondrial (mt)DNA determined from one SH individual is more similar to an mtDNA ancestral to Denisovan than to Neanderthal mtDNAs[2]. However, the mtDNA is inherited as a single unit from mothers to offspring and does not necessarily reflect the overall relationship of individuals and populations. To clarify the relationships of the SH hominins to Neanderthals and Denisovans, we therefore set out to retrieve nuclear DNA from SH hominins. However, DNA preservation in these fossils is poor owing to their great age. Femur XIII, from which the SH mtDNA genome was sequenced, contains only small amounts of highly degraded endogenous DNA (30–45 base pairs (bp)) in a large excess of microbial DNA. To reconstruct its mtDNA genome, almost 2 g of bone had to be used to produce DNA libraries from which mtDNA fragments were isolated by hybridization capture. Furthermore, because of the presence of modern human DNA contamination, putatively endogenous sequences had to be identified on the basis of the presence of C to T substitutions that accumulate at the ends of DNA fragments over time owing to cytosine deamination[9], which are largely absent in recent human DNA that contaminates fossils[10,11].

To retrieve nuclear DNA sequences from femur XIII, we generated approximately 2.6 billion sequence reads from the library with the highest frequency of terminal C to T substitutions (library A2021 (ref. 2)). In addition, between 600 million and 900 million reads were collected from each of four new specimens that were recovered from the site for molecular analyses (Extended Data Table 1). These were an incisor (AT-5482), a femur fragment (AT-5431), a molar (AT-5444) and a scapula (AT-6672).

In addition to sequencing random fragments from these specimens, we also isolated mtDNA fragments from the four new specimens by hybridization capture. Between 1,419 and 3,742 unique mtDNA fragments of 30 bp or longer were retrieved (Extended Data Table 2). To investigate whether they represented endogenous DNA or present-day human contamination, we determined the frequency of C to T substitutions relative to the human mitochondrial genome at each position in the fragments. The fragments from femur AT-5431 carry 44% C to T substitutions at the 5′ ends and 41% at the 3′ ends, compatible with the presence of endogenous ancient mtDNA. The other three specimens do not show discernible evidence of deamination-induced substitutions. Because there were too few DNA fragments to reconstruct the complete mtDNA genome of femur AT-5431, we restricted further analyses to 'diagnostic' positions in the mtDNA genome where each lineage in the mtDNA tree differs from the other hominin lineages and from the chimpanzee. At positions where modern humans differ from Neanderthals, Denisovans, SH femur XIII and the chimpanzee, 41% (17 out of 41) of the mtDNA fragments share the modern human state, indicating that they are derived from present-day human contamination

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. [2]Centro de Investigación Sobre la Evolución y Comportamiento Humanos, Universidad Complutense de Madrid–Instituto de Salud Carlos III, 28029 Madrid, Spain. [3]Departamento de Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain. [4]Área de Paleontología, Departamento de Geografía y Geología, Universidad de Alcalá, Alcalá de Henares, 28871 Madrid, Spain. [5]Centro Nacional de Investigación sobre la Evolución Humana, Paseo Sierra de Atapuerca, 09002 Burgos, Spain. [6]Department of Anthropology, University College London, 14 Taviton Street, London WC1H 0BW, UK. [7]Institut Català de Paleoecologia Humana i Evolució Social, C/Marcel·lí Domingo s/n (Edifici W3), Campus Sescelades, 43007 Tarragona, Spain. [8]Àrea de Prehistòria, Departament d'Història i Història de l'Art, Universitat Rovira i Virgili, Facultat de Lletres, Avinguda de Catalunya, 35, 43002 Tarragona, Spain. [9]Department of Anthropology, University of Toronto, 19 Russell Street, Toronto, Ontario M5S 2S2, Canada.

Table 1 | Characteristics of the nuclear sequence alignments obtained from the five SH specimens

Specimen	Sequences ≥35 bp mapped (%)	C to T substitution frequency (%)				Aligned sequence in putatively deaminated fragments (bp)	Present-day human contamination (%)*	
				'Conditional'				
		5′ end	3′ end	5′ end	3′ end		All fragments	Putatively deaminated fragments
Femur XIII (AT2944)	0.02	5.3	8.2	32.6	41.9	225,329	88	63
Incisor (AT-5482)	0.25	12.9	15.9	60.4	58.5	2,015,167	86	21
Femur frag. (AT-5431)	0.09	21.5	22.3	63.6	55.8	1,186,442	76	18
Molar (AT-5444)	0.02	10.6	11.8	63.1	53.4	188,974	88	0
Scapula (AT-6672)	0.03	0.6	1.0	4.4	10.0	37,524	98	100

*Present-day human contamination is estimated here as the percentage of SH sequences that share alleles derived in 90% or more of present-day humans and ancestral in the two archaic high-coverage genomes and other primates.

(Extended Data Fig. 1). In contrast, among the five fragments that show evidence for deamination, none shares the human-derived state. Among the eight putatively deaminated fragments that overlap positions with variants specific to Denisovans and to femur XIII, all eight carry the derived variants present in both lineages. In addition, three out of five fragments carry variants specific to femur XIII only. In contrast, of nine mtDNA fragments overlapping positions diagnostic for Neanderthals, none carry the Neanderthal variants. We thus conclude that the mtDNA of femur AT-5431 is most closely related to the mtDNA of femur XIII.

The vast majority of endogenous DNA in the SH fossils is degraded to a size below 45 bp. To maximize the yield of DNA fragments we have therefore used fragments as short as 30 bp when reconstructing mtDNAs from these specimens[2,12], whereas DNA analyses from other archaic hominins have been restricted to fragments of 35 bp or longer[6,7]. We explored whether it might be possible to use DNA fragments as short as 30 bp to study also the nuclear genome in SH specimens; however, this resulted in 9–67% of the aligned DNA fragments appearing to be of microbial rather than of hominin origin. In contrast, no spurious alignments were detected at a length cut-off of ≥35 bp (Supplementary Information, section 1). Using the latter cut-off, between 0.02% and 0.25% of the DNA sequences determined from the fossils align to the human reference genome (Table 1). With the exception of the scapula, all specimens show C to T substitution frequencies between 5% and 22% at the terminal alignment positions. When conditioned on C to T substitutions at the other ends of fragments, they increase to between 53% and 64% for the incisor, the femur fragment and the molar, to 33% and 42% for femur XIII, and to 4% and 10% for the scapula (Table 1 and Extended Data Fig. 2), indicating that all five specimens carry mixtures of highly deaminated endogenous nuclear DNA and less deaminated human contamination (Supplementary Information, section 2).

Owing to the extremely small amounts of data available, assessment of nuclear DNA contamination cannot be achieved using existing approaches that require multi-fold coverage in at least parts of the nuclear genome[13,14]. We therefore used two alternative approaches to obtain estimates of contamination for the nuclear sequences (Supplementary Information, sections 2 and 3). The first approach, which compares deamination signals in all sequences to those carrying a C to T substitution at the opposing end, estimates human contamination to be >63% in all five specimens. The second approach estimates contamination as the percentage of sequences sharing the modern human state at sites where 90% or more of present-day humans differ from the chimpanzee and the two high-coverage archaic genomes. The contamination estimates from this approach are similarly high (Table 1), but decrease to 21% or less in three of the specimens (femur AT-5431, the incisor and the molar) when filtering for sequences showing terminal C to T substitutions indicative of deamination. Disregarding fragments without evidence for deamination, the amount of nuclear DNA sequence retrieved varies between 189 kb and 2.0 Mb for these three specimens (Table 1), all of which are male as inferred from the sequence coverage of chromosome X and the autosomes (Extended Data Fig. 3). The sex of the scapula and femur XIII cannot be confidently determined as a result of the high levels of present-day human contamination and the limited amount of data available.

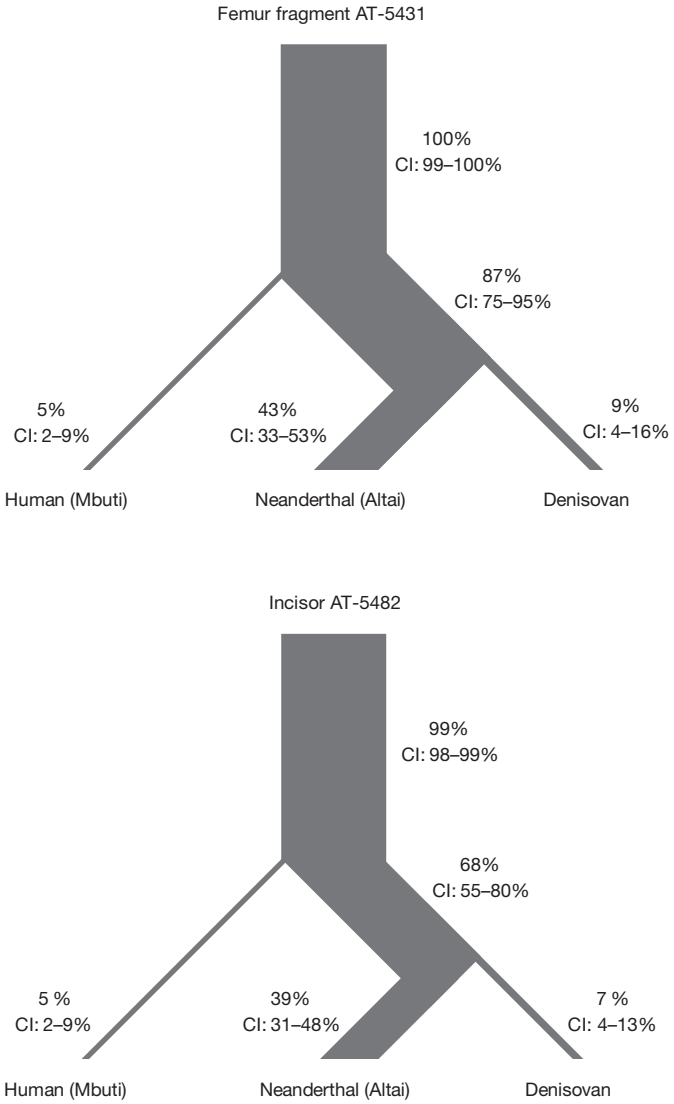


Figure 1 | Percentage of derived alleles shared between the SH specimen and the human, Neanderthal and Denisovan genomes. Ninety-five per cent binomial confidence intervals (CI) are indicated. The thickness of the branches is scaled by the extent of derived allele sharing. See Extended Data Fig. 4 for the total number of informative positions identified in the nuclear genome and Extended Data Table 3 for the number of sequences overlapping these positions.

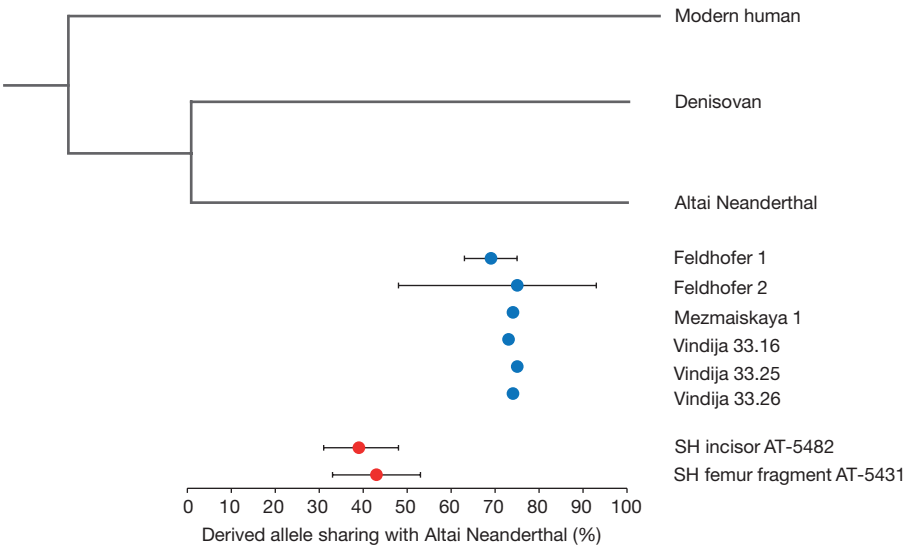


Figure 2 | Sharing of derived alleles with the Altai Neanderthal. Error bars, 95% confidence intervals.

To investigate how the SH hominins are related to modern humans, Neanderthals and Denisovans, we used the high-quality genome sequences of the Altai Neanderthal, the Denisovan finger bone and a present-day human individual from Africa (Mbuti, HGDP00982) to identify positions where one or more of these three genomes differ from those of the chimpanzee and other primates (bonobo, gorilla, orangutan, rhesus macaque) (Extended Data Fig. 4). We then estimated the percentages of all informative positions covered in each specimen that share the derived state for each branch in the tree relating the three genome sequences. For the femur fragment AT-5431 and the incisor, we find that 87% and 68%, respectively, of the positions on the common Neanderthal and Denisovan branch carry derived alleles; that 43% and 39%, respectively, of positions on the Neanderthal branch carry derived alleles; while 9% and 7%, respectively, on the Denisovan branch do so (Fig. 1). This indicates that the SH hominins are related to the ancestors of Neanderthals rather than Denisovans. The fraction of derived alleles shared with the Neanderthal genome is between two- and three-fold smaller when sequences without terminal C to T substitutions are also included (Extended Data Table 3), confirming that the signal linking the SH hominins to Neanderthals is derived from endogenous DNA fragments. These results are stable when present-day human individuals other than the Mbuti are used in the analysis (Extended Data Table 4). Similar to femur AT-5431 and the incisor, the molar also shows a greater sharing of derived alleles with the Neanderthal than the Denisovan genome, although not statistically significantly so, probably because of the small amount of data available (Extended Data Fig. 5). By comparison, the fraction of derived alleles shared with the Neanderthal genome for several Late Pleistocene Neanderthals sequenced to low-coverage[7,13] is between 69% and 75% (Fig. 2). Thus, DNA sequences of the SH hominins diverged more than twice as far back along the lineage from the Altai Neanderthal genome to its ancestor shared with the Denisovan genome than DNA sequences of the Late Pleistocene Neanderthals from Europe and the Caucasus.

Because nearly 30 hominin skeletons have been found in SH, it is likely that the specimens analysed here belong to different individuals. The nuclear DNA sequences of femur AT-5431 and the incisor show that they belonged to the Neanderthal evolutionary lineage, and the limited data available for the molar suggest that the same is true for this specimen. Thus, the results show that the SH hominins were early Neanderthals or closely related to the ancestors of Neanderthals after the divergence from a common ancestor shared with Denisovans. Although it is difficult to determine the age of Middle Pleistocene sites with certainty, geological dating methods[1], as well as the length of the branches in trees relating the mtDNAs from femur XIII and an SH cave bear to other mtDNAs[2,12], suggest an age of around 400,000 years for the SH fossils. This age is compatible with the population split time of 381,000–473,000 years ago estimated for Neanderthals and Denisovans on the basis of their nuclear genome sequences and using the human mutation rate of 0.5×10^{-9} per base pair per year[7]. This mutation rate also suggests that the population split between archaic and modern humans occurred between 550,000 and 765,000 years ago. Such an ancient separation of archaic and modern humans is difficult to reconcile with the suggestion that younger specimens often classified as *Homo heidelbergensis*, for example Arago or Petralona, belong to a population ancestral both to modern humans and to Neanderthals[15].

We further note that the SH hominins carry mtDNAs more closely related to those of Denisovans in Asia than Neanderthals, even though their nuclear genomes show that they are more closely related to Neanderthals. We have previously speculated that this discrepancy may be because the SH hominins carried two very divergent mtDNA lineages or that another hominin group contributed mtDNA both to the SH hominins and to Denisovans[2]. However, given that the SH hominins are early Neanderthals (or closely related to these), and assuming that the mtDNA they carried was typical of early Neanderthals, an additional possibility that appears reasonable is that the mtDNAs seen in Late Pleistocene Neanderthals were acquired by them later, presumably because of gene flow from Africa. It is possible that contacts between Africa and western Eurasia occurred in the Middle Pleistocene as indicated, for example, by the appearance of the Acheulean hand axe technology in Eurasia by 500,000 years ago[16] and by the spread of the so-called 'Mode 3' technology around 250,000 years ago[17]. Gene flow from Africa may perhaps also explain the absence of Neanderthal-derived morphological traits in some Middle Pleistocene specimens in Europe such as Ceprano and Mala Balanica[18,19]. Retrieval of further mtDNAs and, if possible, nuclear DNA from Middle Pleistocene fossils will be necessary to comprehensively address how Middle and Late Pleistocene hominins in Eurasia were related to each other.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 September 2015; accepted 2 February 2016.
Published online 14 March 2016.

1. Arsuaga, J. L. *et al.* Neandertal roots: cranial and chronological evidence from Sima de los Huesos. *Science* **344,** 1358–1363 (2014).
2. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505,** 403–406 (2014).

3. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053–1060 (2010).
4. Sawyer, S. *et al.* Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl Acad. Sci. USA* **112,** 15696–15700 (2015).
5. Qin, P. & Stoneking, M. Denisovan ancestry in East Eurasian and Native American populations. *Mol. Biol. Evol.* **32,** 2665–2674 (2015).
6. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
7. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
8. Arsuaga, J. L. *et al.* Postcranial morphology of the middle Pleistocene humans from Sima de los Huesos, Spain. *Proc. Natl Acad. Sci. USA* **112,** 11524–11529 (2015).
9. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104,** 14616–14621 (2007).
10. Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464,** 894–897 (2010).
11. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7,** e34131 (2012).
12. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110,** 15758–15763 (2013).
13. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
14. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334,** 94–98 (2011).
15. Stringer, C. The status of *Homo heidelbergensis* (Schoetensack 1908). *Evol. Anthropol.* **21,** 101–107 (2012).
16. Lycett, S. J. Understanding ancient hominin dispersals using artefactual data: a phylogeographic analysis of Acheulean handaxes. *PLoS ONE* **4,** e7404 (2009).
17. Lahr, M. M. & Foley, R. A. Towards a theory of modern human origins: Geography, demography, and diversity in recent human evolution. *Yb. Phys. Anthropol.* **41,** 137–176 (1998).
18. Dennell, R. W., Martinon-Torres, Bermúdez de Castro, J. M. Hominin variability, climatic instability and population demography in Middle Pleistocene Europe. *Quat. Sci. Rev.* **30,** 1511–1524 (2011).
19. Rink, W. J. *et al.* New radiometric ages for the BH-1 hominin from Balanica (Serbia): implications for understanding the role of the Balkans in Middle Pleistocene human evolution. *PLoS ONE* **8,** e54608 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Höber and A. Weihmann for help with sequencing the libraries, G. Renaud for processing the raw sequence data, S. Castellano and U. Stenzel for discussions and comments on the manuscript. Genetics work was funded by the Max Planck Society and its Presidential Innovation Fund. Field work at the Sierra de Atapuerca sites was funded by the Junta de Castilla y Leon, the Fundacion Atapuerca, the Spanish Ministerio de Ciencia e Innovacion (project CGL2009-12703-C03) and the Spanish Ministerio de Economia y Competitividad (project CGL2012-38434-C03).

Author Contributions M.M., J.-L.A. and S.P. directed the experimental work and wrote the manuscript. M.M. designed the laboratory experiments, which S.N., A.A. and B.N. performed. M.M., C.d.F., B.V., J.K. and K.P. analysed the data. J.-L.A., I.M., A.G., J.M.B. and E.C. excavated the fossil and provided archaeological expertise.

Author Information Sequences generated in this study have been deposited in the European Nucleotide Archive under study accession number PRJEB10597. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (mmeyer@eva.mpg.de).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Sampling, DNA extraction and library preparation. Over the past decade, several specimens from the SH excavation were exempted from the routine cleaning procedures for molecular analysis. Samples were removed from four of these specimens in a laboratory dedicated to ancient DNA work: (1) molar AT-5444, excavated in 2006 in Square T-16, Level LU-6, was sampled by drilling into one of the roots with a dentistry drill; (2) incisor AT-5482, removed from Square U-17, Level LU-6 in 2010, from which the tip of the root was removed with a scalpel, and powder was removed by drilling into the dentine; (3) scapula AT-6672, excavated in 2012 in Square U-16, Level LU-6, was sampled by drilling powder from the specimen; (4) femur fragment AT-5431, excavated in 2014 from Square T-12, Level LU-6, was sampled by scratching off thin slices of bone from the wall of the marrow cavity with a scalpel. Between 9 and 94 mg of material sample was used for DNA extraction by a silica-based method optimized for the recovery of short DNA fragments[12]. In an attempt to remove parts of the microbial DNA contamination, the sample from femur AT-5431 was treated with 0.5 M sodium phosphate buffer (pH 7.0) before DNA extraction[20]. Between 30% and 60% of the extracts were converted to DNA libraries using single-stranded library preparation[21]; in some cases an optimized version of the original method[20] was used (see Extended Data Table 1). The number of molecules in each library was determined by digital droplet PCR[22] and libraries were barcoded through amplification with pairs of indexed primers[23] using AccuPrime Pfx DNA polymerase (Life Technologies)[24]. Extraction and library negative controls were carried through DNA extraction, library preparation, mtDNA enrichment and sequenced alongside the SH samples to monitor laboratory contamination.

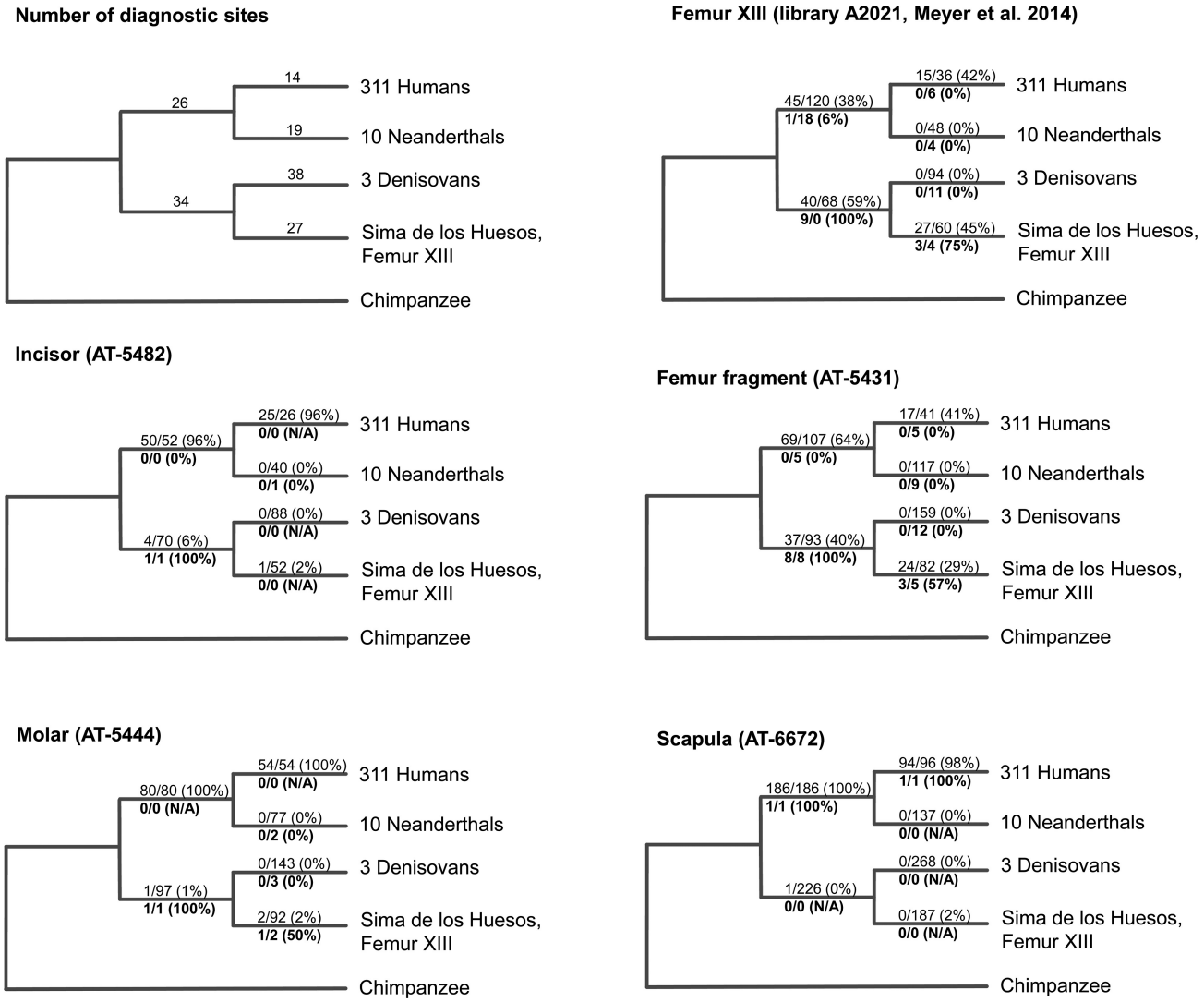
Mitochondrial DNA enrichment and analysis. Enrichment of mtDNA was performed in single or two successive rounds of hybridization capture following ref. 25 but using human mtDNA baits recovered from a microarray[26] and using reduced temperatures in the hybridization and wash steps as described elsewhere[2]. Enriched libraries were pooled with other libraries and sequenced on Illumina MiSeq and HiSeq 2500 platforms using 76 bp paired-end sequencing recipes for double-indexed sequencing[23]. Base calling was performed using Illumina Bustard software (MiSeq) or FreeIbis[27] (HiSeq). Sequences that did not perfectly match one of the expected index combinations were discarded. Overlapping paired-end reads were merged into single sequences to reconstruct full-length molecules[28]. Merged sequences $\geq$30 bp were aligned to the revised Cambridge reference human mtDNA sequence (rCRS; accession number NC_012920) using BWA[29] with 'ancient' parameters[6] requiring a mapping quality score $\geq$30. Duplicate sequences were removed by calling a consensus from sequences with identical alignment start and end coordinates (bam-rmdup; https://github.com/udo-stenzel/biohazard). Putatively deaminated fragments were identified on the basis of the presence of a C to T substitution to the reference genome in the first or last position.

To identify positions in the mtDNA genome that were diagnostic for each branch of the hominin tree, we aligned the mtDNA sequences from 311 worldwide present-day humans[30], 10 Neanderthals[7,31–33], 3 Denisovans[3,4], SH femur XIII[2] and the chimpanzee[34] to the rCRS using MAFFT[35]. To exclude errors due to cytosine deamination when counting the number of sequences supporting the respective branch, we used the fact that single-stranded library preparation is strand-specific, and we disregarded all sequences aligned to the reference genome in the orientation as sequenced if one of the two states of a diagnostic site was C, and sequences aligned in the reverse complement direction if one of the two states was G.

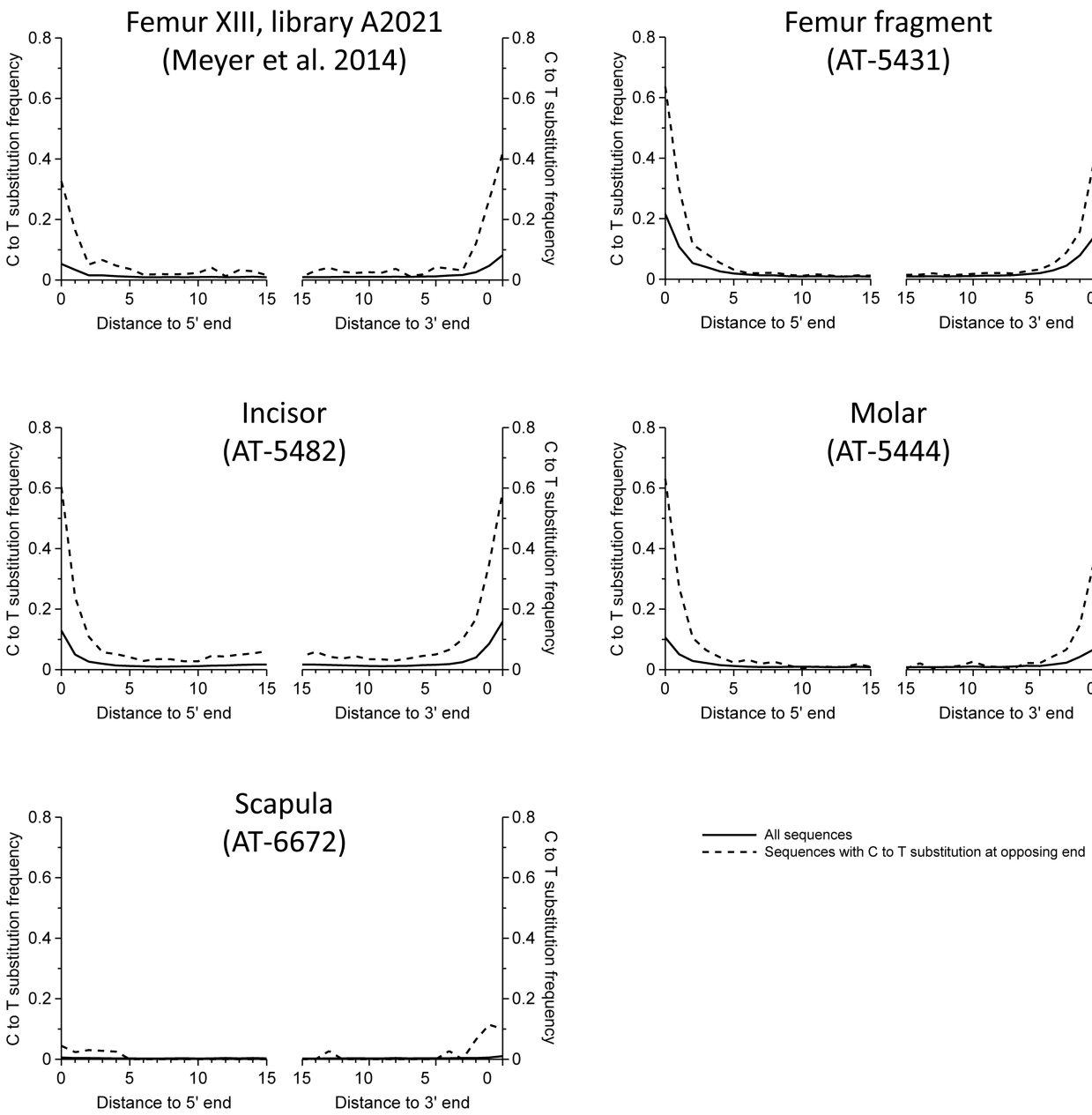
Nuclear DNA sequencing and raw data processing. In an initial attempt to obtain nuclear DNA sequences from an SH hominin, we sequenced eight lanes of a previously prepared library (library A2021) from femur XIII[2] on an Illumina HiSeq 2500 instrument using the same recipe as for the libraries enriched for mtDNA, but with shorter read length (2×50 bp). A 0.5% spike-in of a double-indexed φX174 control library (P7 index CGATTCG, P5 index CGATTCG) was added to the library to allow base calling with FreeIbis. The absence of sequence complexity in the index reads forced us to drop index filtering for this run. We then generated data from the other four specimens using between four and seven HiSeq lanes each (2×50 or 2×76 bp). We spiked-in between 1.5 and 20% of a pool of four φX174 libraries carrying AAAAAAA, CCCCCCC, GGGGGGG and TTTTTTT as P7 and P5 indices. This recovered the quality of the index reads and allowed us to filter for a perfect match to at least one of the two indices without losing large amounts of data (Extended Data Table 5). Further processing and mapping were performed as described for the capture-enriched libraries above, using only sequences of at least 35 bp for mapping to the human genome (hg19/GRCh37) and requiring a map quality score of 30 or greater.

Phylogenetic analysis. To place the SH specimens in the hominin phylogenetic tree, we identified phylogenetically informative positions where a randomly drawn allele from the Altai Neanderthal, Denisovan and/or Mbuti VCF files differed from all great apes (the chimpanzee (pantro2), the bonobo (panpan1), the gorilla (gorGor3), the orangutan (ponAbe2)) and the rhesus macaque (rheMac2). Orthologous ape and monkey outgroup sequences were extracted from pairwise lastz alignments to the human reference genome (hg19/GRCh37) provided by the University of California, Santa Cruz (UCSC) genome browser and in-house, and were required to show the identical base at informative positions. All sites were required to pass the map35_100% criteria described in supplementary section 5b of ref. 7. A file of all phylogenetically informative positions is available for download at http://bioinf.eva.mpg.de/sima/. To reduce the impact of cytosine deamination, all thymines within the first and last three positions of each sequence were masked out. This filter is less strict than that based on alignment orientation described above and thus retains more data. We also included sequence data of an early modern human[36], several Neanderthals[7,13] and two Denisovans[4] for comparison (Extended Data Fig. 4 and Extended Data Table 3).

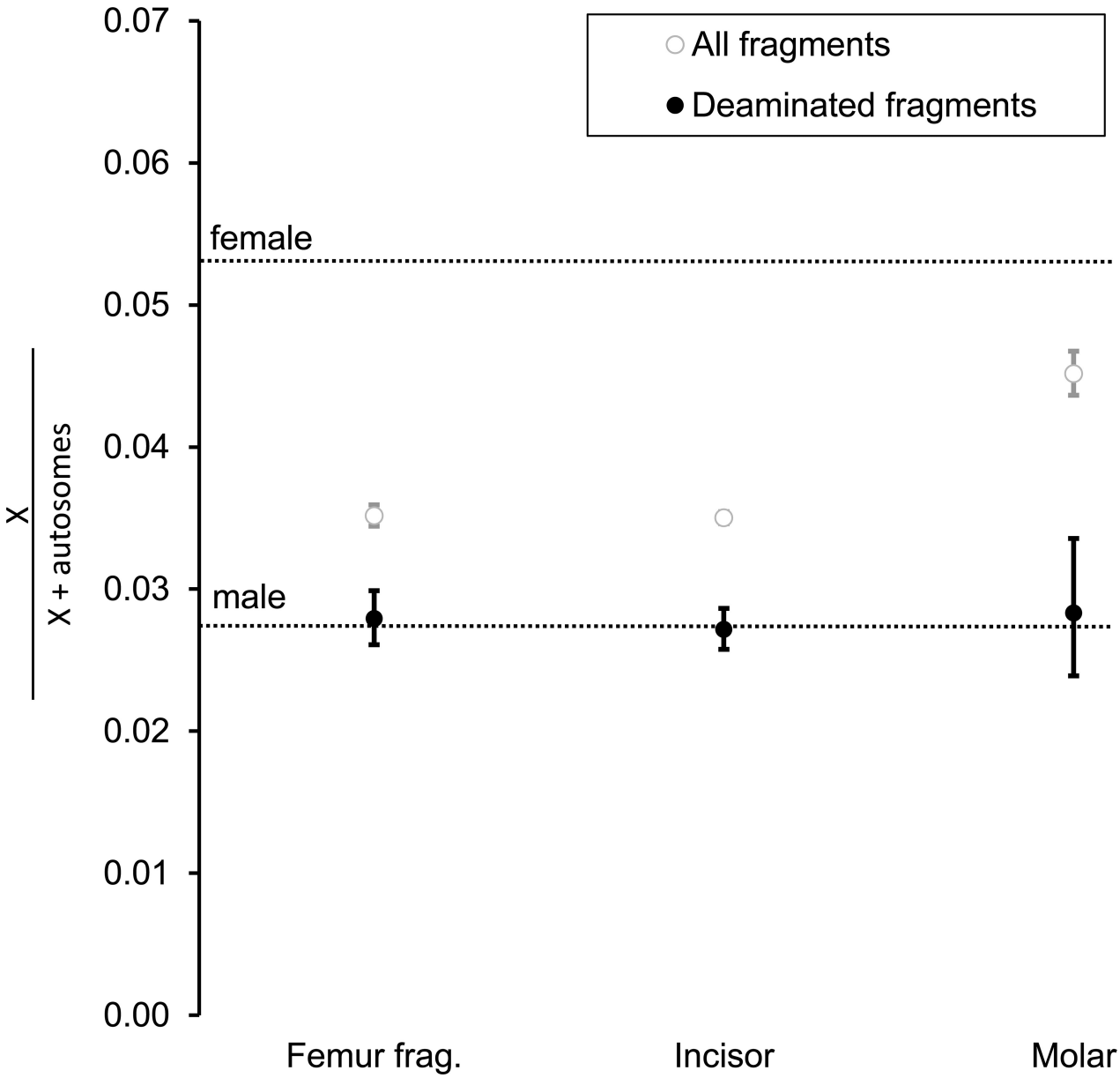
20. Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59,** 87–93 (2015).
21. Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8,** 737–748 (2013).
22. Slon, V., Glocke, I., Barkai, R., Gopher, A., Hershkovitz, I. & Meyer, M. Mammalian mitochondrial capture, a tool for rapid screening of DNA preservation in faunal and undiagnostic remains, and its application to Middle Pleistocene specimens from Qesem Cave (Israel). *Quat. Int.* http://dx.doi.org/10.1016/j.quaint.2015.03.039 (2015).
23. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40,** e3 (2012).
24. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52,** 87–94 (2012).
25. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5,** e14004 (2010).
26. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110,** 2223–2227 (2013).
27. Renaud, G., Kircher, M., Stenzel, U. & Kelso, J. freelbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29,** 1208–1209 (2013).
28. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42,** e141 (2014).
29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
30. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134,** 416–426 (2008).
31. Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325,** 318–321 (2009).
32. Gansauge, M. T. & Meyer, M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* **24,** 1543–1549 (2014).
33. Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111,** 2229–2234 (2014).
34. Horai, S. *et al.* Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* **35,** 32–43 (1992).
35. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
36. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).



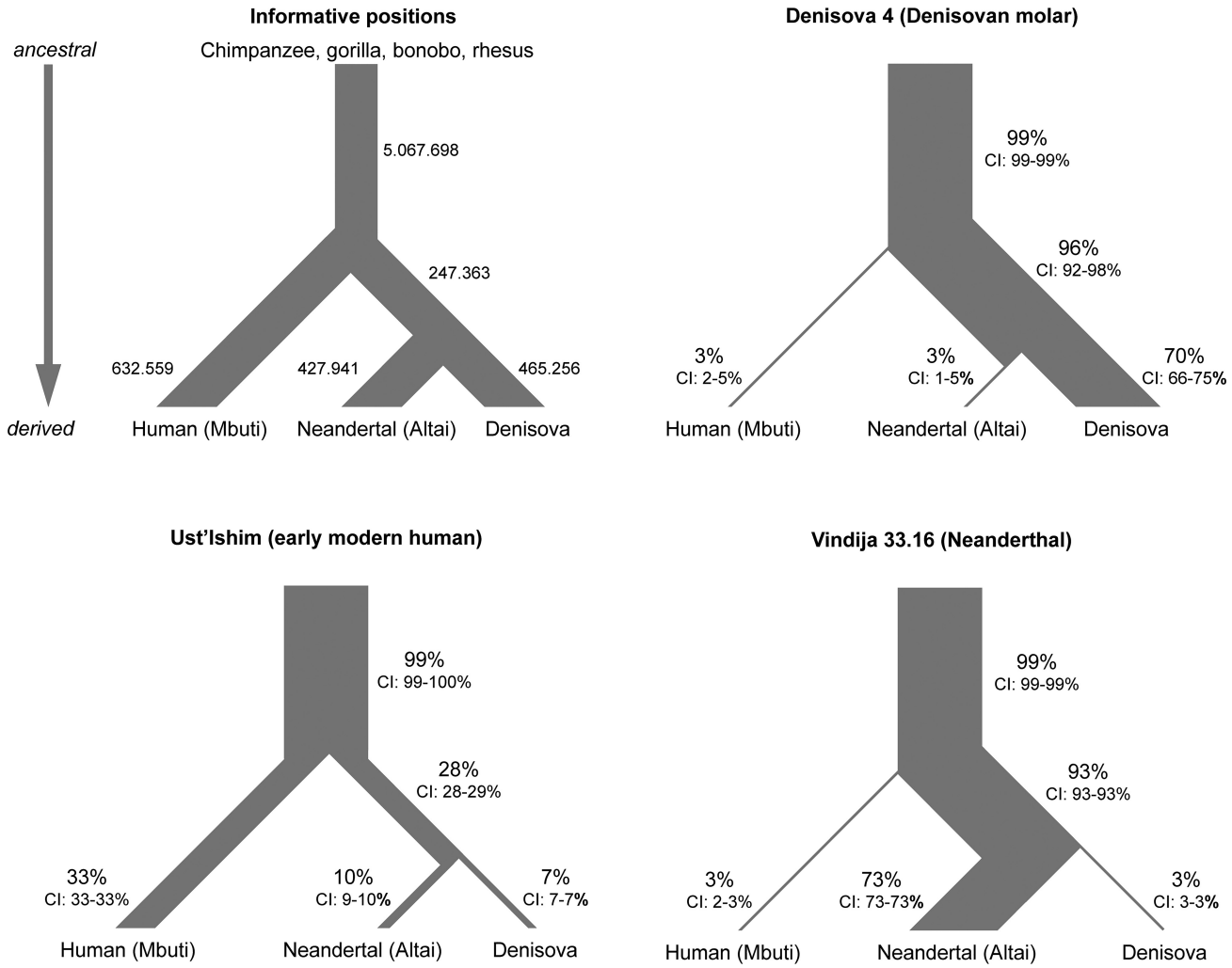
Extended Data Figure 1 | Sharing of derived alleles at diagnostic positions separating the hominin groups in the mitochondrial tree. The chimpanzee was used as outgroup to determine the ancestral state, which is shared with all individuals in the tree except those belonging to the labelled branch. Provided are the number of diagnostic sites available for this analysis (top left panel) as well as the number of sequences supporting the derived state, their percentage (in brackets) and the total number of observations. Numbers above the branch include all sequences whereas bold numbers below the branch are limited to sequences showing evidence of cytosine deamination. Published data from library A2021 of femur XIII were included in this analysis for comparison.



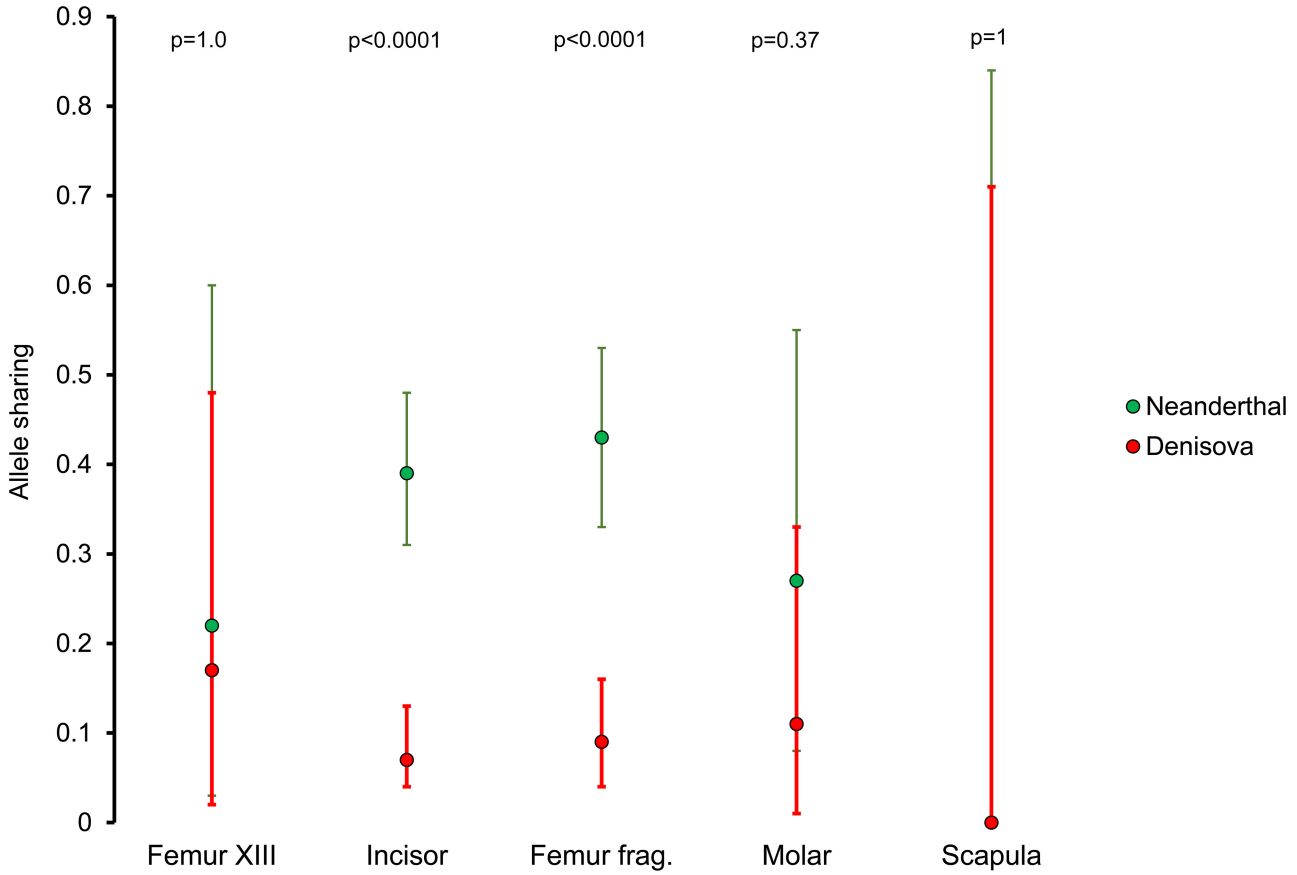
Extended Data Figure 2 | Frequency of C to T substitutions at the beginning and end of nuclear sequence alignments. Solid lines denote all sequences and dashed lines only those sequences carrying a C to T substitution at the opposing end.



Extended Data Figure 3 | Sex determination based on the number of sequences aligning to chromosome X and the autosomes. Ninety-five per cent binomial confidence intervals are provided as well as the expected ratios of X to (X + autosomal) sequences for male and female samples. The analysis was performed with and without enrichment of endogenous DNA by filtering for the presence of C to T substitutions at terminal alignment positions. Present-day human contamination in the unfiltered sequences appears to have been introduced at least partly by female individuals.



Extended Data Figure 4 | Number of informative positions identified for each branch of the tree. For comparison with the SH sequence data, we show the sharing of derived alleles at these positions using published sequence data from a Neanderthal (Vindija 33.16), a Denisovan individual (Denisova 4) and an early modern human (Ust'Ishim).



Extended Data Figure 5 | Derived allele sharing with the Neanderthal- and Denisovan-specific branches in deaminated DNA fragments from all five specimens from SH. Only sequences with a terminal C to T substitution were used in this analysis. Error bars, 95% confidence intervals. Significance was tested using Fisher's exact test (two-tailed).

Extended Data Table 1 | Overview of DNA extracts, libraries and shotgun sequences generated in three experiments

Exp.	Specimen	Material used [mg]	Fraction of extract used in library prep.	Library prep. protocol	Library ID	P7 / P5 index seq.	#Molecules in library acc. to dPCR/ qPCR(*)	#seq. generated	#seq. ≥35bp	#mapped seq.	#unique seq.	Deaminated sequences (unique)	
												#	Av. size
(Meyer et al. 2014)	Femur XIII (AT2944)	63	0.6	Ref. 21	A2021	ATGAGCA / CGACGGT	5.73E+09 (*)	2.55E+09	7.27E+08	1.52E+05	1.36E+05	2,772	41.7
1	Incisor (AT-5482)	9	0.3	Ref. 21	B2523	CGTAATC / AATAGTA	9.67E+08 (*)	8.73E+08	4.42E+08	1.10E+06	7.27E+05	30,950	40.5
1	Extraction blank		0.3	Ref. 21	B2524	CTATACG / CTGCGAC	3.78E+08 (*)						
1	Extraction blank		0.3	Ref. 21	B2529	TGGCGCT / TGGCAAT	4.23E+08 (*)						
1	Library blank			Ref. 21	B2531	ATCGTTC / CGAGATC	2.53E+08 (*)						
2	Femur frag. (AT-5431)	94	0.3	Ref. 20	R1848	AGACTCC / CCTAGGT	1.60E+09	7.98E+08	3.53E+08	3.08E+05	2.35E+05	17,265	40.5
2	Extraction blank		0.3	Ref. 20	R1858	CGCTATT / ACTATCA	4.44E+07						
2	Library blank			Ref. 20	R1861	GACCGAT / TAATGCG	2.13E+07						
3	Molar (AT-5444)	28	0.3	Ref. 20	R1753	TACTCGC / CTCGATG	2.34E+09	8.93E+08	4.89E+08	8.05E+04	6.95E+04	2,635	40.5
3	Scapula (AT-6672)	20	0.3	Ref. 20	R1754	AGCGCCA / GCTCGAA	1.14E+10	5.76E+08	3.49E+08	9.99E+04	9.75E+04	274	60.6
3	Extraction blank		0.3	Ref. 20	R1757	GTTGCAT / AACTCCG	2.45E+07						
3	Library blank			Ref. 20	R1758	ATCCTCT / TTGAAGT	9.15E+06						

Reported are the amount of sample used for DNA extraction, the fraction of DNA extract converted into DNA library, the number of molecules in each library as determined by digital PCR (dPCR) or quantitative PCR (qPCR), the protocol used for library preparation, index sequences of each library, the number of sequences generated from each library and the number of sequences retained after length filtering, mapping, duplicate removal and identification of putatively deaminated fragments.

Extended Data Table 2 | Characteristics of sequences obtained after mtDNA enrichment

Exp.	Specimen	Library ID	#sequences	#overlap-merged sequences ≥30bp	#mapped sequences (MQ ≥30)	#unique seq.	Dup. rate	#Seq. with C to T substitution in terminal positions	C to T substitution frequency [%] (#observations)		Human contamination [%] (95% C.I.)	
									5' end	3' end	All fragments	Deaminated fragments
(Meyer et al. 2014)	Femur XIII (AT2944)	A2021			257,277	2,592	97.3	687	46.4 (250/539)	47.8 (249/521)	41.7 (25.5-59.2)	0.0 (0.0-45.9)
1	Incisor (AT-5482)	B2523	1,172,042	773,601	62,659	1,419	44.0	47	4.7 (14/298)	6.7 (14/208)	96.2 (80.4-99.9)	N/A
1	Extraction blank	B2524	74,304	44,231	9,792	151	64.3	3	0.0 (0/30)	0.0 (0/29)		
1	Extraction blank	B2529	61,439	42,600	10,009	189	52.7	0	0.0 (0/47)	0.0 (0/39)		
1	Library blank	B2531	56,078	31,179	8,827	71	124.0	0	0.0 (0/15)	0.0 (0/16)		
2	Femur frag. (AT-5431)	R1848	38,797,693	25,567,161	678,518	3,188	207.7	731	44.2 (277/627)	41.1 (233/567)	41.5 (26.3-57.9)	0.0 (0.0-52.2)
2	Extraction blank	R1858	1,096,093	562,893	201,219	1,361	147.9	16	1.5 (4/261)	1.3 (3/229)		
2	Library blank	R1861	125,743	43,344	6,377	101	63.1	2	6.3 (1/16)	0 (0/15)		
3	Molar (AT-5444)	R1753	43,834,971	34,002,764	105,628	3,041	34.5	52	3.0 (15/498)	3.0 (20/666)	100.0 (93.4-100.0)	N/A
3	Scapula (AT-6672)	R1754	34,645,495	27,006,518	49,676	3,742	13.2	31	0.5 (4/784)	1.0 (7/721)	97.9 (92.7-99.7)	100.0 (2.5-100.0)
3	Extraction blank	R1757	296,666	103,470	7,608	358	21.1	3	1.5 (1/67)	0 (0/76)		
3	Library blank	R1758	565,241	156,407	10,261	58	166.7	3	0 (0/5)	0 (0/6)		

Extracts and sample libraries and their related controls were generated in three independent experiments as indicated. Provided are the total number of sequences generated from each library, the number of sequences ≥30 bp, the number of sequences ≥30 bp that could be mapped to the revised Cambridge reference sequence, the number of sequences remaining after duplicate removal and the average number of sequences collapsed into unique sequences during duplicate removal. Putatively deaminated fragments were identified by C to T substitutions at the terminal alignment positions, the frequency of which is denoted. Estimates of present-day human contamination and the respective 95% binomial confidence intervals were determined on the basis of the percentage of sharing of human-derived alleles (compare Extended Data Fig. 1). Previously published data from femur XIII (library A2021) are included for comparison.

Extended Data Table 3 | Fraction of derived alleles shared with the human, Neanderthal and Denisovan genomes and combinations thereof

a

	Included sequences	Human	Neanderthal	Denisova	Neanderthal-Denisova	Human-Neanderthal-Denisova	Human-Denisova	Neanderthal-Human
Femur XIII (AT-2944)	deaminated	0.23 (0.08-0.45)	0.22 (0.03-0.60)	0.17 (0.02-0.48)	0.67 (0.22-0.96)	0.98 (0.94-0.99)	0.20 (0.01-0.72)	1.00 (0.40-1.00)
	all	0.31 (0.29-0.33)	0.13 (0.11-0.15)	0.09 (0.07-0.10)	0.30 (0.27-0.34)	0.99 (0.98-0.99)	0.64 (0.58-0.70)	0.70 (0.65-0.76)
Incisor (AT-5482)	deaminated	0.05 (0.02-0.09)	0.39 (0.31-0.48)	0.07 (0.04-0.13)	0.68 (0.55-0.80)	0.99 (0.98-0.99)	0.29 (0.13-0.51)	0.82 (0.62-0.94)
	all	0.31 (0.30-0.32)	0.13 (0.12-0.14)	0.07 (0.06-0.07)	0.34 (0.33-0.36)	0.98 (0.98-0.99)	0.64 (0.61-0.67)	0.69 (0.67-0.72)
Femur frag. (AT-5431)	deaminated	0.05 (0.02-0.09)	0.43 (0.33-0.53)	0.09 (0.04-0.16)	0.87 (0.75-0.95)	1.00 (0.99-1.00)	0.39 (0.17-0.64)	0.67 (0.47-0.83)
	all	0.25 (0.23-0.27)	0.17 (0.15-0.19)	0.07 (0.06-0.09)	0.44 (0.41-0.47)	0.99 (0.99-0.99)	0.55 (0.50-0.60)	0.70 (0.66-0.74)
Molar (AT-5444)	deaminated	0.00 (0.00-0.23)	0.27 (0.08-0.55)	0.11 (0.01-0.33)	0.63 (0.25-0.92)	0.99 (0.97-1.00)	0.33 (0.01-0.91)	1.00 (0.03-1.00)
	all	0.29 (0.25-0.32)	0.12 (0.10-0.16)	0.08 (0.06-0.11)	0.33 (0.28-0.38)	0.99 (0.98-0.99)	0.61 (0.52-0.70)	0.67 (0.58-0.76)
Scapula (AT-6672)	deaminated	1.00 (0.03-1.00)	0.00 (0.00-0.84)	0.00 (0.00-0.71)	0.67 (0.09-0.99)	1.00 (0.83-1.00)	0.00 (0.00-0.98)	N/A
	all	0.32 (0.30-0.35)	0.11 (0.09-0.13)	0.06 (0.05-0.07)	0.29 (0.26-0.33)	0.99 (0.98-0.99)	0.62 (0.56-0.68)	0.71 (0.66-0.76)
Denisova 4	deaminated	0.03 (0.02-0.05)	0.03 (0.01-0.05)	0.70 (0.66-0.75)	0.96 (0.92-0.98)	0.99 (0.99-0.99)	0.86 (0.76-0.92)	0.12 (0.06-0.20)
Denisova 8	deaminated	0.02 (0.01-0.02)	0.06 (0.06-0.06)	0.59 (0.58-0.60)	0.92 (0.91-0.93)	1.00 (0.99-1.00)	0.88 (0.87-0.90)	0.14 (0.13-0.16)
Feldhofer 1	deaminated	0.05 (0.03-0.08)	0.69 (0.63-0.75)	0.03 (0.01-0.06)	0.93 (0.87-0.96)	0.99 (0.99-1.00)	0.09 (0.02-0.24)	0.86 (0.75-0.93)
Feldhofer 2	deaminated	0.00 (0.00-0.15)	0.75 (0.48-0.93)	0.06 (0.00-0.30)	0.86 (0.42-1.00)	0.99 (0.97-1.00)	0.33 (0.01-0.91)	1.00 (0.63-1.00)
Mezmaiskaya 1	deaminated	0.03 (0.03-0.03)	0.74 (0.73-0.74)	0.03 (0.03-0.03)	0.92 (0.92-0.92)	1.00 (1.00-1.00)	0.11 (0.11-0.12)	0.91 (0.91-0.91)
Vindija 33.16	deaminated	0.03 (0.02-0.03)	0.73 (0.73-0.73)	0.03 (0.03-0.03)	0.93 (0.93-0.93)	0.99 (0.99-0.99)	0.11 (0.11-0.11)	0.91 (0.91-0.92)
Vindija 33.25	deaminated	0.02 (0.02-0.02)	0.75 (0.74-0.75)	0.03 (0.03-0.03)	0.94 (0.93-0.94)	0.99 (0.99-0.99)	0.10 (0.10-0.10)	0.91 (0.91-0.92)
Vindija 33.26	deaminated	0.02 (0.02-0.02)	0.74 (0.74-0.74)	0.03 (0.03-0.03)	0.93 (0.93-0.94)	0.99 (0.99-0.99)	0.10 (0.10-0.10)	0.91 (0.91-0.92)
Ust'Ishim	deaminated	0.33 (0.33-0.33)	0.10 (0.09-0.10)	0.07 (0.07-0.07)	0.28 (0.28-0.29)	0.99 (0.99-0.99)	0.66 (0.66-0.67)	0.71 (0.71-0.71)

b

	Included sequences	Human	Neanderthal	Denisova	Neanderthal-Denisova	Human-Neanderthal-Denisova	Human-Denisova	Neanderthal-Human
Femur XIII (AT-2944)	deaminated	5/22	2/9	2/12	4/6	166/170	1/5	4/4
	all	529/1698	157/1220	111/1278	202/666	13680/13878	161/250	212/301
Incisor (AT-5482)	deaminated	10/199	51/131	10/137	41/60	1592/1617	7/24	22/27
	all	2725/8657	742/5763	412/6108	1134/3313	66296/67364	868/1357	1041/1498
Femur frag. (AT-5431)	deaminated	7/150	42/98	10/110	47/54	1278/1284	7/18	20/30
	all	651/2592	287/1666	143/1941	448/1012	20271/20522	211/386	325/464
Molar (AT-5444)	deaminated	0/14	4/15	2/19	5/8	165/166	1/3	1/1
	all	213/744	65/523	42/532	110/333	6081/6165	77/126	83/123
Scapula (AT-6672)	deaminated	1/1	0/2	0/3	2/3	20/20	0/1	0/0
	all	552/1699	127/1182	70/1198	198/682	13075/13256	166/268	214/301
Denisova 4	deaminated	20/580	12/460	320/454	230/240	4958/5000	72/84	12/98
Denisova 8	deaminated	262/15564	646/10780	6646/11274	5270/5720	127696/128310	2168/2456	380/2670
Feldhofer 1	deaminated	17/357	149/215	6/233	141/152	2799/2820	3/34	55/64
Feldhofer 2	deaminated	0/22	12/16	1/16	6/7	194/195	1/3	8/8
Mezmaiskaya 1	deaminated	9183/308630	154251/209649	7039/229075	111722/121511	2489810/2499964	5382/47214	50559/55532
Vindija 33.16	deaminated	5195/204696	97859/133612	4833/154912	72346/77903	1703087/1716488	3345/30304	33475/36604
Vindija 33.25	deaminated	4633/207608	101934/136687	4842/155443	74188/79239	1693259/1705720	3156/31131	33449/36574
Vindija 33.26	deaminated	4312/186298	91088/123221	4348/140193	66384/71122	1536122/1547680	2805/28107	30311/33196
Ust'Ishim	deaminated	84958/256372	16567/173398	12857/188709	28536/100640	2030409/2059155	26116/39296	31792/44800

a, Point estimates as well as 95% binomial confidence intervals. For the SH hominins, numbers are provided for all fragments ≥35 bp and the subset of fragments showing evidence of deamination.
b, Number of DNA fragments matching the derived state as well the total number of informative fragments.

Extended Data Table 4 | Derived allele sharing between putatively deaminated DNA fragments of the five SH specimens and all branches of the hominin evolutionary tree

		Present-day human used in analysis					
Branch	**Specimen**	**Mbuti**	**Yoruba**	**San**	**French**	**Han**	**Papuan**
Neandertal	FemurXIII	22.2% (2/9)	53.8% (7/13)	46.2% (6/13)	36.4% (4/11)	41.7% (5/12)	41.7% (5/12)
	Femur AT-5431	42.9% (42/98)	39.8% (39/98)	41.7% (40/96)	41.0% (41/100)	40.2% (39/97)	44.8% (43/96)
	Incisor	38.9% (51/131)	38.9% (49/126)	39.2% (51/130)	39.8% (51/128)	38.5% (47/122)	37.2% (45/121)
	Molar	26.7% (4/15)	28.6% (4/14)	28.6% (4/14)	26.7% (4/15)	30.8% (4/13)	21.4% (3/14)
	Scapula	0.0% (0/2)	0.0% (0/3)	0.0% (0/3)	0.0% (0/3)	0.0% (0/3)	0.0% (0/3)
Denisova	FemurXIII	16.7% (2/12)	14.3% (2/14)	12.5% (2/16)	14.3% (2/14)	6.7% (1/15)	12.5% (2/16)
	Femur AT-5431	9.1% (10/110)	5.9% (6/102)	8.1% (9/111)	6.4% (7/109)	6.5% (7/108)	7.2% (8/111)
	Incisor	7.3% (10/137)	9.1% (12/132)	8.5% (12/141)	8.1% (11/136)	7.4% (10/136)	8.1% (11/135)
	Molar	10.5% (2/19)	5.6% (1/18)	5.3% (1/19)	5.3% (1/19)	5.6% (1/18)	5.0% (1/20)
	Scapula	0.0% (0/3)	0.0% (0/3)	0.0% (0/4)	0.0% (0/4)	0.0% (0/4)	0.0% (0/3)
Human	FemurXIII	22.7% (5/22)	18.8% (3/16)	15.4% (4/26)	13.3% (2/15)	25.0% (4/16)	12.5% (2/16)
	Femur AT-5431	4.7% (7/150)	5.5% (7/127)	7.1% (11/154)	5.4% (8/147)	8.3% (12/144)	5.7% (8/140)
	Incisor	5.0% (10/199)	6.1% (12/197)	3.0% (6/198)	5.3% (10/188)	6.3% (12/190)	8.0% (14/176)
	Molar	0.0% (0/14)	0.0% (0/20)	0.0% (0/22)	0.0% (0/20)	0.0% (0/20)	0.0% (0/17)
	Scapula	100.0% (1/1)	33.3% (1/3)	100.0% (1/1)	20.0% (1/5)	33.3% (1/3)	50.0% (1/2)
Neandertal-Denisova	FemurXIII	66.7% (4/6)	75.0% (3/4)	71.4% (5/7)	81.8% (9/11)	72.7% (8/11)	71.4% (5/7)
	Femur AT-5431	87.0% (47/54)	81.8% (36/44)	83.3% (40/48)	85.7% (42/49)	82.2% (37/45)	80.0% (40/50)
	Incisor	68.3% (41/60)	73.5% (50/68)	74.6% (53/71)	71.9% (46/64)	76.2% (48/63)	71.4% (45/63)
	Molar	62.5% (5/8)	75.0% (6/8)	77.8% (7/9)	50.0% (4/8)	55.6% (5/9)	66.7% (4/6)
	Scapula	66.7% (2/3)	66.7% (2/3)	0.0% (0/1)	0.0% (0/1)	50.0% (1/2)	50.0% (1/2)
Human-Denisova	FemurXIII	20.0% (1/5)	50.0% (1/2)	33.3% (1/3)	33.3% (1/3)	50.0% (1/2)	50.0% (1/2)
	Femur AT-5431	38.9% (7/18)	34.5% (10/29)	31.8% (7/22)	40.0% (10/25)	40.0% (10/25)	39.1% (9/23)
	Incisor	29.2% (7/24)	19.2% (5/26)	21.7% (5/23)	26.9% (7/26)	23.1% (6/26)	20.0% (6/30)
	Molar	33.3% (1/3)	50.0% (2/4)	50.0% (2/4)	66.7% (2/3)	50.0% (2/4)	100.0% (2/2)
	Scapula	0.0% (0/1)	NA% (0/0)	NA% (0/0)	NA% (0/0)	NA% (0/0)	0.0% (0/1)
Neandertal-Human	FemurXIII	100.0% (4/4)	NA% (0/0)	100.0% (1/1)	100.0% (2/2)	100.0% (2/2)	100.0% (2/2)
	Femur AT-5431	66.7% (20/30)	68.8% (22/32)	67.7% (21/31)	77.8% (21/27)	77.4% (24/31)	63.3% (19/30)
	Incisor	81.5% (22/27)	77.4% (24/31)	70.0% (21/30)	80.0% (24/30)	75.8% (25/33)	67.6% (25/37)
	Molar	100.0% (1/1)	50.0% (1/2)	50.0% (1/2)	100.0% (1/1)	50.0% (1/2)	100.0% (2/2)
	Scapula	NA% (0/0)	NA% (0/0)	NA% (0/0)	NA% (0/0)	NA% (0/0)	NA% (0/0)
All	FemurXIII	97.6% (166/170)	96.0% (170/177)	97.1% (169/174)	96.3% (157/163)	97.0% (164/169)	97.1% (169/174)
	Femur AT-5431	99.5% (1278/1284)	99.5% (1299/1306)	99.4% (1298/1306)	99.5% (1303/1310)	99.5% (1297/1303)	99.5% (1299/1305)
	Incisor	98.5% (1592/1617)	98.3% (1603/1630)	98.3% (1605/1632)	98.4% (1621/1647)	98.2% (1591/1620)	98.4% (1620/1647)
	Molar	99.4% (165/166)	99.4% (168/169)	98.8% (169/171)	100.0% (173/173)	100.0% (169/169)	99.4% (172/173)
	Scapula	100.0% (20/20)	100.0% (21/21)	100.0% (22/22)	100.0% (23/23)	100.0% (22/22)	100.0% (21/21)

Phylogenetically informative sites were identified using six present-day human genomes. Provided are the fraction of shared derived alleles and the counts in brackets.

Extended Data Table 5 | Overview of the sequencing runs performed

Specimen	PhiX spike-in [%]	FlowCellID	#Sequencing cycles	Lane	One perfect index match to sample library [%]	Two perfect index matches to sample library [%]
Femur XIII (AT-2944), library A2021	0.5	130827_SN7001204_0226	2x 50	1	0.7	0.0
				2	15.8	0.0
				3	24.4	0.0
				4	37.0	2.0
				5	40.2	6.7
				6	38.9	2.7
				7	24.9	0.0
				8	10.5	0.0
Incisor (AT-5482)	1.5	140718_SN7001204_0282	2x 76	1	68.2	8.3
	4	150126_SN7001204_0332	2x 50	1	45.6	5.6
				2	79.7	29.1
	15	150127_SN7001204_0333	2x 76	2	85.1	69.2
	20	150212_SN7001204_0335	2x 76	2	85.1	69.2
	20	150212_SN7001204_0336	2x 50	1	60.9	23.5
				2	75.9	35.5
Femur fragment (AT-5431)	20	150312_SN7001204_0345	2x 50	1	82.2	61.4
				2	83.1	64.2
				3	83.1	67.5
				4	82.0	64.6
Molar (AT-5444)	20	150302_SN7001204_0343	2x 50	1	54.2	0.1
				2	64.7	2.1
		150423_SN7001204_0374	2x 50	1	83.7	66.0
				2	85.5	72.7
		150423_SN7001204_0375	2x 50	1	82.3	31.8
				2	82.1	28.4
Scapula (AT-6672)	20	150415_SN7001204_0369	2x 76	1	81.9	22.8
				2	82.5	49.1
		150420_SN7001204_0373	2x 76	1	79.6	33.7
				2	80.0	41.6

Provided are the percentage of PhiX spiked into each lane, the number of sequencing cycles performed and the success of index recognition.