LETTER

doi:10.1038/nature12788

A mitochondrial genome sequence of a hominin from Sima de los Huesos

Matthias Meyer[1], Qiaomei Fu[1,2], Ayinuer Aximu-Petri[1], Isabelle Glocke[1], Birgit Nickel[1], Juan-Luis Arsuaga[3,4], Ignacio Martínez[3,5], Ana Gracia[3,5], José María Bermúdez de Castro[6], Eudald Carbonell[7,8] & Svante Pääbo[1]

Excavations of a complex of caves in the Sierra de Atapuerca in northern Spain have unearthed hominin fossils that range in age from the early Pleistocene to the Holocene[1]. One of these sites, the 'Sima de los Huesos' ('pit of bones'), has yielded the world's largest assemblage of Middle Pleistocene hominin fossils[2,3], consisting of at least 28 individuals[4] dated to over 300,000 years ago[5]. The skeletal remains share a number of morphological features with fossils classified as *Homo heidelbergensis* and also display distinct Neanderthal-derived traits[6–8]. Here we determine an almost complete mitochondrial genome sequence of a hominin from Sima de los Huesos and show that it is closely related to the lineage leading to mitochondrial genomes of Denisovans[9,10], an eastern Eurasian sister group to Neanderthals. Our results pave the way for DNA research on hominins from the Middle Pleistocene.

The Sima de los Huesos site (see Fig. 1 for a map) is located at the foot of a 13 m vertical shaft, about 30 m below the surface and 500 m from the closest current entrance to the karst system[11]. Humidity at the site is close to saturation, temperature in the cave is constant around 10.6 °C and the fossils have been protected from major disturbances since deposition[12]. The Sima de los Huesos is also noteworthy because it has provided unique evidence of long-term DNA survival. DNA preservation in the site was first proposed based on enzymatic amplification of a few short mitochondrial DNA (mtDNA) fragments from Middle Pleistocene cave bear remains[13]. Recently, improvements in DNA extraction[14] and library preparation[10] techniques for highly degraded ancient DNA have enabled the retrieval of a complete mitochondrial genome of a cave bear (*Ursus deningeri*) found with the hominin remains in the cave[14]. DNA preservation for hundreds of thousands of years has otherwise been documented only under permafrost conditions[15,16].

Figure 1 | Location of the Middle Pleistocene site of Sima de los Huesos (yellow) as well as Late Pleistocene sites that have yielded Neanderthal DNA (red) and Denisovan DNA (blue).

To investigate whether DNA may also be preserved in the hominin remains, we obtained several samples of bone, totalling 1.95 g, by drilling holes into the breaks of a femur (Femur XIII, ref. 17) excavated in three parts, one in 1994 and the other two in 1999 (Fig. 2). DNA was isolated using a recently published silica-based method[14] and converted into 77 libraries for sequencing[10,18] (Extended Data Table 1). Following library amplification, we first characterized a subset of the libraries by shallow shotgun sequencing on Illumina's MiSeq platform (Extended Data Fig. 1). Overlapping paired-end reads were merged to reconstruct full-length molecule sequences and mapped against the human genome using Burrows–Wheeler alignment (BWA)[19]. For most libraries, fewer than 0.1% of the sequences could be confidently aligned to the human genome (Extended Data Table 2), but 21 libraries yielded proportions of aligned sequences that were high enough (between 0.1% and 8.4%) to investigate the frequencies of C to T substitutions at sequence ends, which are increased in authentic ancient DNA due to accelerated cytosine deamination in single-stranded overhangs[20–22]. However, in no case did C to T substitution frequencies exceed 3% at 5′ ends and 6% at 3′ ends (Extended Data Table 2), indicating that those libraries that are rich in human DNA are dominated by present-day human contamination.

We next enriched all libraries for mtDNA, using a probe set based on a present-day human sequence. An initial inspection of the isolated sequences revealed the closest similarities to the mtDNA of a Denisovan, an extinct archaic group related to Neanderthals[9]. Therefore, the libraries were additionally enriched with probes based on the Denisovan mtDNA[23]. Sequencing was performed on Illumina's HiSeq 2500 platform from both ends, and overlapping reads were merged and aligned to the human reference mtDNA. Sequences with identical start and end coordinates, which often represent amplification products of the same starting molecules, were fused to create consensus sequences, and sequences shorter than 30 base pairs (bp) were discarded. The enriched libraries yielded a sufficient number of mitochondrial sequences to estimate the frequencies of C to T substitutions. These varied widely among the libraries, ranging from 1% to 45% at 5′ ends, and from 2% to 47% at 3′ ends (Extended Data Table 1). In agreement with the shotgun

Figure 2 | Femur XIII reassembled from three parts after sampling. The natural fractures are visible in the proximal third of the femur.

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. [2]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. [3]Centro de Investigación Sobre la Evolución y Comportamiento Humanos, Universidad Complutense de Madrid–Instituto de Salud Carlos III, 28029 Madrid, Spain. [4]Departamento de Paleontología, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain. [5]Área de Paleontología, Depto. de Geografía y Geología, Universidad de Alcalá, Alcalá de Henares, 28871 Madrid, Spain. [6]Centro Nacional de Investigación sobre la Evolución Humana, Paseo Sierra de Atapuerca, 09002 Burgos, Spain. [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, Dept. d'Història i Història de l'Art, Univ. Rovira i Virgili, Fac. de Lletres, Av. Catalunya, 35, 43002 Tarragona, Spain.

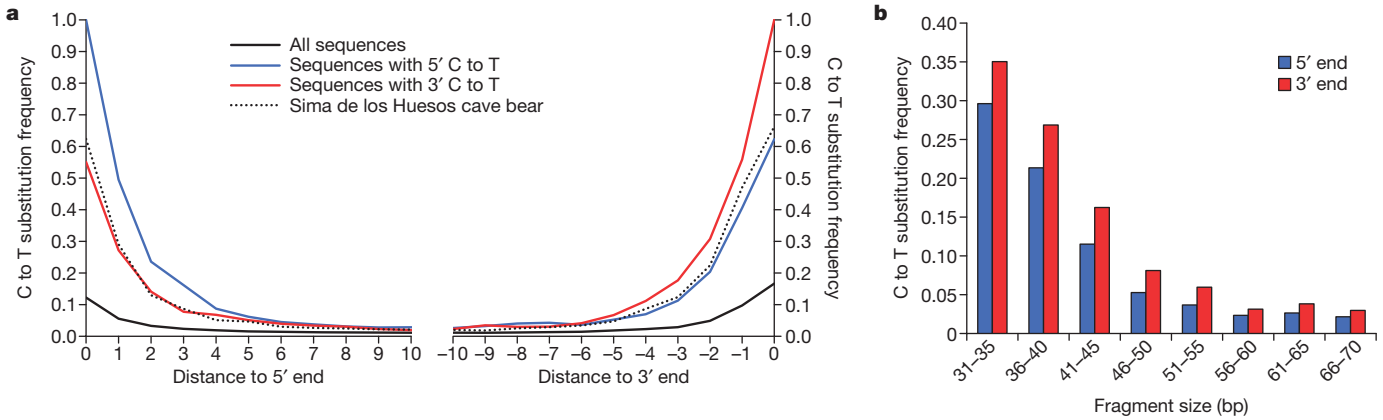


Figure 3 | Patterns of cytosine deamination in the libraries constructed from the Sima de los Huesos hominin femur. a, C to T substitution frequencies are shown for the terminal positions of the aligned sequences for all sequences (black), those sequences carrying a C to T substitutions at their 5′ ends (blue), at their 3′ ends (red), and for all Sima de los Huesos cave bear sequences from the *U. deningeri* sample[9] (dotted line). **b**, C to T substitution frequencies at the first and last base of sequences in different fragment length bins.

sequencing results, the libraries yielding the largest number of mitochondrial sequences exhibited very low terminal C to T substitution frequencies (≤3% and ≤6% at 5′ and 3′ ends, respectively; Extended Data Fig. 2) indicating that they are dominated by present-day human contamination. Libraries showing C to T substitution frequencies of less than 5% at either end were considered to be too contaminated and therefore disregarded in subsequent analyses.

Variation in C to T substitution frequencies among libraries suggest that two populations of sequences are present in the data, an endogenous population strongly affected by cytosine deamination and a contaminating population showing much less deamination. To test if this is the case, we determined the 5′ C to T substitution frequencies for sequences showing a 3′ C to T difference to the reference and vice versa, thereby enriching for putatively endogenous DNA. C to T substitution frequencies indeed increased to 55% at 5′ ends and 62% at 3′ ends, numbers that are close to those determined for the *U. deningeri* sample from Sima de los Huesos[14] (Fig. 3a). Furthermore, stratification of the deamination signal by fragment length shows that the endogenous DNA is primarily present among sequences that are shorter than 45 bp, again in agreement with the situation in the *U. deningeri* sample (Fig. 3b and Extended Data Fig. 3). Based on these results, we removed sequences longer than 45 bp and those that do not carry a terminal C to T substitution on either the 5′ or 3′ end (Extended Data Table 3). In addition, we applied a mapping quality filter to ensure unique placement of the sequences within the mtDNA genome and readjusted the alignment parameters to tolerate up to five C to T differences but no more than two other differences to the reference mtDNA sequence to discriminate against spurious alignments. Finally, T bases at the first and last three positions of each sequence were masked to reduce the impact of deamination-induced substitutions during consensus calling.

We first called consensus bases for 15,181 positions of the mitochondrial genome that were covered by 5 or more sequences of which at least 80% agreed. Average coverage across these positions was 21.8. However, such strict filtering increases the risk of ascertainment bias because residual modern human contamination as well as capture and mapping biases may lead to the exclusion of positions where the Sima de los Huesos specimen differs from the probes or the reference sequence. We therefore built a second more inclusive consensus by considering the three terminal positions while selecting sequences with C to T substitution and lowering the requirements for coverage and consensus agreement to 3 and >67%, respectively. This consensus encompasses 16,302 positions or ~98% of the human mitochondrial reference genome, with an average coverage of 31.6 (Extended Data Fig. 4). Third, to evaluate whether the use of Denisovan capture probes influence the results, we built a consensus using the strictest filtering criteria described above, but including only sequences isolated with present-day human mtDNA probes (Extended Data Fig. 5).

We reconstructed phylogenetic trees in a Bayesian statistical framework[24] using the three Sima de los Huesos consensus mtDNA sequences as well as the mtDNAs of present-day and ancient humans, Neanderthals, Denisovans, chimpanzees and bonobos. All three trees support a topology in which the Sima de los Huesos mtDNA shares a common ancestor with Denisovan mtDNAs to the exclusion of the other mtDNAs analysed with maximum posterior probability (Fig. 4 and Extended Data Fig. 6). As expected owing to its age, the branch leading to the Sima de los Huesos mtDNA is shorter than those leading to any of the other archaic or present-day humans. Using 13 directly dated ancient mtDNA sequences for calibration[25] and the three consensus sequences, we estimated the age of the Sima de los Huesos specimen based on the length of its mtDNA branch (Table 1). These dates vary between 0.15 to 0.64 million years with point estimates close to 400,000 years. This is in striking agreement with the point estimate of 409,000 years for the *U. deningeri* mtDNA[14]. We similarly estimated the divergence times of the major mitochondrial lineages (Table 1 and Extended Data Table 4) and find that the estimates for the divergence of the mtDNAs of the

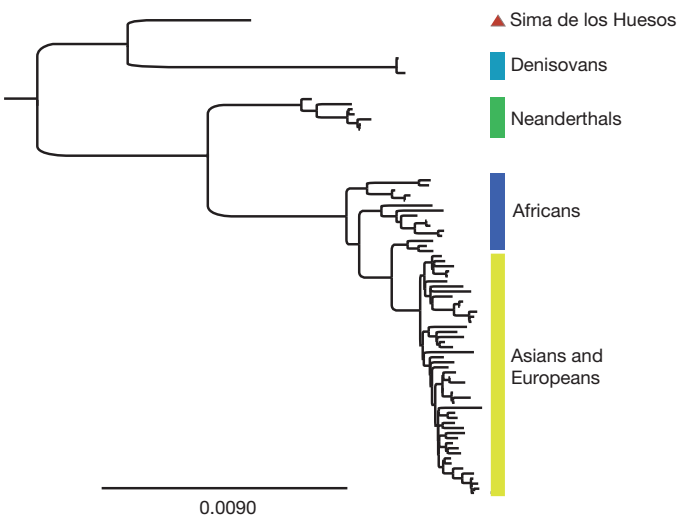


Figure 4 | Bayesian phylogenetic tree of hominin mitochondrial relationships based on the Sima de los Huesos mtDNA sequence determined using the inclusive filtering criteria. All nodes connecting the denoted hominin groups are supported with posterior probability of 1. The tree was rooted using chimpanzee and bonobo mtDNA genomes. The scale bar denotes substitutions per site.

Table 1 | Divergence times of the major hominin mtDNA lineages and the age of the Sima de los Huesos specimen as estimated by using three different filtering strategies for consensus calling

Mitochondrial lineages	Divergence dates and molecular age estimates in Myr BP (95% highest posterior density (HPD) intervals in brackets)		
	Strict filters	Inclusive filters	Strict filters enriched with human probes only
Sima de los Huesos age estimate	0.38 (0.18–0.60)	0.37 (0.16–0.59)	0.40 (0.15–0.64)
Human – Neanderthal	0.51 (0.35–0.69)	0.53 (0.36–0.72)	0.44 (0.29–0.61)
Sima de los Huesos/Denisova – Human/Neanderthal	0.99 (0.70–1.33)	1.04 (0.72–1.41)	0.83 (0.54–1.11)
Sima de los Huesos – Denisova	0.73 (0.50–1.01)	0.76 (0.51–1.06)	0.65 (0.40–0.92)

Sima de los Huesos hominin and Densiovans vary between 0.40 and 1.06 million years with point estimates around 700,000 years ago.

The fact that the Sima de los Huesos mtDNA shares a common ancestor with Denisovan rather than Neanderthal mtDNAs is unexpected in light of the fact that the Sima de los Huesos fossils carry Neanderthal-derived features (for example, in their dental, mandibular, midfacial, supraorbital and occipital morphology[2,6,7,26]). Denisovans were identified in 2010 based on DNA sequences retrieved from a manual phalanx and a molar found in southern Siberia[9,23]. Based on analyses of their nuclear genome[9,10] they are a sister group of Neanderthals, although the mtDNAs of Neanderthals and present-day humans share an mtDNA ancestor more recently with each other than with Denisovans[23]. This may be owing to either incomplete lineage sorting in the common ancestral populations of these groups or to gene flow into Denisovans from another archaic group[9].

Several evolutionary scenarios are compatible with the presence of a mtDNA sequence that falls on the Denisovan mtDNA lineage in a ~400,000-year-old hominin in western Europe. First, the Sima de los Huesos hominins may be closely related to the ancestors of Denisovans. However, this seems unlikely, because the presence of Denisovans in western Europe would indicate an extensive spatial overlap with Neanderthal ancestors, raising the question how the two groups could genetically diverge while overlapping in range. Furthermore, although almost no morphological information is available for Denisovans, a molar that carries Denisovan DNA is of exceptionally large size[9] and does not exhibit the cusp reduction seen in the Sima de los Huesos hominins[7]. Most importantly, the Sima de los Huesos specimen is so old that it probably predates the population split time between Denisovans and Neanderthals, which is estimated to one-half to two-thirds of the time to the split between Neanderthals and modern humans, which is estimated to be 170,000 to 700,000 years ago[9]. Second, it is possible that the Sima de los Huesos hominins represent a group distinct from both Neanderthals and Denisovans that later perhaps contributed the mtDNA to Denisovans. However, this scenario would imply the independent emergence of several Neanderthal-like morphological features in a group unrelated to Neanderthals. Third, the Sima de los Huesos hominins may be related to the population ancestral to both Neanderthals and Denisovans. Considering the age of the Sima de los Huesos remains and their incipient Neanderthal-like morphology, this scenario seems plausible to us, but it requires an explanation for the presence of two deeply divergent mtDNA lineages in the same archaic group, one that later recurred in Denisovans and one that became fixed in Neanderthals, respectively. A forth possible scenario is that gene flow from another hominin population brought the Denisova-like mtDNA into the Sima de los Huesos population or its ancestors. Such a hominin group might have also contributed mtDNA to the Densiovans in Asia[9,10]. Based on the fossil record, more than one evolutionary lineage may have existed in Europe during the Middle Pleistocene[27]. Several fossils have been found in Europe as well as in Africa and Asia that are close in time to Sima de los Huesos but do not exhibit clear Neanderthal traits. These fossils are often grouped into *H. heidelbergensis,* a taxon that is difficult to define[8,28,29], particularly with regard to whether the Sima de los Huesos hominins should be included[8]. Furthermore, there may have been relict populations of still earlier hominins, notably those classified as *Homo antecessor,* which share some morphological traits with Asian *Homo erectus*[30] and have been found just a few hundred metres away from Sima de los Huesos in Gran Dolina.

Although nuclear sequence data are needed to clarify the genetic relationship of the Sima de los Huesos hominins to Neanderthals and Densiovans, the mtDNA sequence establishes an unexpected link between Denisovans and the western European Middle Pleistocene fossil record. Future efforts will now focus on describing the mtDNA variation of the Sima de los Huesos hominins and retrieving nuclear DNA sequences from them. The latter will be a huge challenge given that almost two grams of bone were required to generate the mtDNA sequence even though several hundred copies of mtDNA exist per cell. Although preservation of DNA for such long periods of time may be favoured by unique preservation conditions in the Sima de los Huesos, the present results show that ancient DNA sequencing techniques have become sensitive enough to warrant further investigation of DNA survival at sites where Middle Pleistocene hominins are found.

METHODS SUMMARY

A detailed description of the methods used for data generation and analysis is provided in the Methods section.

Online Content Any additional Methods, 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 10 September; accepted 17 October 2013.
Published online 4 December 2013.

1. Carbonell, E. *et al.* The first hominin of Europe. *Nature* **452,** 465–469 (2008).
2. Arsuaga, J. L., Martinez, I., Gracia, A. & Lorenzo, C. The Sima de los Huesos crania (Sierra de Atapuerca, Spain). A comparative study. *J. Hum. Evol.* **33,** 219–281 (1997).
3. Arsuaga, J. L. *et al.* Size variation in Middle Pleistocene humans. *Science* **277,** 1086–1088 (1997).
4. Bermúdez de Castro, J. M. & Nicolas, M. E. Palaeodemography of the Atapuerca-SH Middle Pleistocene hominid sample. *J. Hum. Evol.* **33,** 333–355 (1997).
5. Bischoff, J. L. *et al.* Geology and preliminary dating of the hominid-bearing sedimentary fill of the Sima de los Huesos Chamber, Cueva Mayor of the Sierra de Atapuerca, Burgos, Spain. *J. Hum. Evol.* **33,** 129–154 (1997).
6. Martínez, I. & Arsuaga, J. L. The temporal bones from Sima de los Huesos Middle Pleistocene site (Sierra de Atapuerca, Spain). A phylogenetic approach. *J. Hum. Evol.* **33,** 283–318 (1997).
7. Martinón-Torres, M., Bermudez de Castro, J. M., Gomez-Robles, A., Prado-Simon, L. & Arsuaga, J. L. Morphological description and comparison of the dental remains from Atapuerca-Sima de los Huesos site (Spain). *J. Hum. Evol.* **62,** 7–58 (2012).
8. Stringer, C. The status of *Homo heidelbergensis* (Schoetensack 1908). *Evol. Anthropol.* **21,** 101–107 (2012).
9. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053–1060 (2010).
10. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
11. Ortega, A. I. *et al.* Evolution of multilevel caves in the Sierra de Atapuerca (Burgos, Spain) and its relation to human occupation. *Geomorphology* **196,** 122–137 (2013).
12. Arsuaga, J. L. *et al.* Sima de los Huesos (Sierra de Atapuerca, Spain). The site. *J. Hum. Evol.* **33,** 109–127 (1997).
13. Valdiosera, C. *et al.* Typing single polymorphic nucleotides in mitochondrial DNA as a way to access Middle Pleistocene DNA. *Biol. Lett.* **2,** 601–603 (2006).
14. 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).
15. Willerslev, E. *et al.* Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317,** 111–114 (2007).
16. Orlando, L. *et al.* Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499,** 74–78 (2013).

17. Carretero, J. M. *et al.* Stature estimation from complete long bones in the Middle Pleistocene humans from the Sima de los Huesos, Sierra de Atapuerca (Spain). *J. Hum. Evol.* **62,** 242–255 (2012).
18. Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8,** 737–748 (2013).
19. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
20. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104,** 14616–14621 (2007).
21. Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20,** 231–236 (2010).
22. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7,** e34131 (2012).
23. Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464,** 894–897 (2010).
24. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574 (2003).
25. Shapiro, B. *et al.* A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol. Biol. Evol.* **28,** 879–887 (2011).
26. Arsuaga, J. L., Martinez, I., Gracia, A., Carretero, J. M. & Carbonell, E. Three new human skulls from the Sima de los Huesos Middle Pleistocene site in Sierra de Atapuerca, Spain. *Nature* **362,** 534–537 (1993).
27. Arsuaga, J. L. Colloquium paper: terrestrial apes and phylogenetic trees. *Proc. Natl Acad. Sci. USA* **107** (Suppl. 2), 8910–8917 (2010).
28. Hublin, J. J. Out of Africa: Modern human origins special feature: The origin of Neandertals. *Proc. Natl Acad. Sci. USA* **106,** 16022–16027 (2009).
29. Mounier, A., Marchal, F. & Condemi, S. Is *Homo heidelbergensis* a distinct species? New insight on the Mauer mandible. *J. Hum. Evol.* **56,** 219–246 (2009).
30. Carbonell, E. *et al.* An Early Pleistocene hominin mandible from Atapuerca-TD6, Spain. *Proc. Natl Acad. Sci. USA* **102,** 5674–5678 (2005).

Acknowledgements We thank J. Dabney, M. Dannemann, C. de Filippo, S. Lippold, K. Prüfer, M. Slatkin, M. Stiller, C. Valdiosera and B. Viola for discussions and comments on the manuscript; G. Renaud and U. Stenzel for help with sequence data processing; B. Höber and A. Weihmann for performing the sequencing runs; M. Gansauge, P. Korlević, R. Rodríguez and I. Ureña for help in the laboratory; M. Schreiber for help with graphics; J. Trueba for providing the fossil image; M. Cruz Ortega for restoration of the fossil and the rest of the members of the Sima de los Huesos excavation team ?for decades of continuous efforts. Genetics work was funded by the Max Planck ?Society and its Presidential Innovation Fund. Field work at the Sierra de Atapuerca sites is funded by the Junta de Castilla y León and the Fundación Atapuerca. ?Research was supported by Spanish Ministerio de Ciencia e Innovación (project CGL2009-12703-C03) and Spanish Ministerio de Economía y Competitividad (project CGL2012-38434-C03).

Author Contributions M.M. designed the experiments and analysed the data; Q.F. performed phylogenetic analyses; A.A., I.G. and B.N. performed the experiments; J.-L.A., I.M., A.G., J.M.B. and E.C. excavated the fossil and provided expert archaeological and anthropological information; J.-L.A. and S.P. were involved in study design; and M.M., J.-L.A. and S.P. wrote the manuscript.

Author Information The Sima de los Huesos mtDNA consensus sequence (based on the inclusive filtering criteria) is deposited in GenBank under accession number KF683087. 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

Description of the femur, archaeological context and sampling. The largest femoral fragment (AT-999) was discovered in 1994 and represents the distal (lower) two thirds of the bone. The proximal (upper) third of the femur (AT-2943) was recovered in 1999. The third femoral fragment (AT-2944), also found in 1999, is a much smaller shaft fragment, which partially connects the two larger femoral fragments (see Fig. 2). All three fragments were found close to each other in square U-15 (the Sima de los Huesos excavation grid has squares of 0.5 m in length). U-15 excavation square is in the central area of the site and is particularly rich in human fossils, including complete skulls.

Sampling was performed by drilling small holes into the cortical tissue of all three bone fragments starting from pre-existing fractures. To reduce the impact of modern human contamination, approximately 1 mm of surface material was removed before drilling holes in the bone at low speed (1,000 r.p.m.) using a sterile dentistry drill. No damage was done to the outer surface of the femur.

DNA extraction, library preparation and shotgun sequencing. Using 1.95 g of bone material in total, 39 DNA extracts were made from between 25 and 75 mg of bone powder each using a recently published silica-based DNA extraction protocol optimized for the recovery of very short molecules from ancient biological material[14]. Two of these extracts were made from surface material, whereas all other extracts were generated from bone powder sampled from inside the bone. Extraction blank controls were carried alongside with the samples in each set of DNA extractions. As substantial pellets of undigested bone powder remained after extraction, we also generated re-extracts from 6 bone pellets to investigate whether additional DNA can be released by repeating the extraction. Libraries were prepared in sets of 16 using between 20 and 30 μl of sample or blank DNA extract (out of 50 μl total volume) and following a single-stranded library preparation protocol specifically developed for highly degraded ancient DNA[18]. One positive control and one blank control were included in each set of library preparations. No uracil-DNA glycosylase treatment was performed to preserve the C to T substitution patterns that are typical for sequences from ancient DNA[20].

The number of unique molecules in each library was estimated by quantitative PCR (qPCR), using primers hybridizing to the adaptor sequences[18]. All sample libraries ($n = 77$) yielded qPCR molecule counts between 1.9×10^9 and 2.4×10^{10}, with exception of the libraries prepared from re-extracted bone pellets, which returned numbers in the range of 1.3×10^8 to 1.1×10^9 (Extended Data Table 1). Molecule counts of the extraction and library preparation blanks ($n = 13$), which represent artefacts and library molecules derived from contamination with exogenous DNA, were consistently lower (in the range of 4.9×10^6 to 4.8×10^7). qPCR molecule counts thus indicate that substantial amounts of DNA reside in the bone and that most of this DNA is successfully released in a single round of DNA extraction.

Each library was divided into four aliquots of 12.5 μl, which were then amplified into PCR plateau in 100 μl reactions with AccuPrime Pfx DNA polymerase (Life Technologies) as described elsewhere[31]. During amplification, two unique index sequences were introduced into the adaptors of each library, following a double-indexing scheme described elsewhere[32]. Amplification products from the same library were pooled and purified using the MinElute PCR purification kit (Qiagen).

Aliquots from a subset of libraries were pooled and subjected to shallow shotgun sequencing using Illumina's MiSeq platform in double-index configuration ($2 \times 76 + 2 \times 7$ cycles)[32]. Raw sequence data were processed as described below. The fragment size distribution of the sequences before mapping shows a mode around 30 bp (Extended Data Fig. 1), indicating that small DNA fragments were efficiently extracted and converted into library molecules. Sequences were aligned against the human genome (GRCh37/1000 Genomes release) using BWA[19] with relaxed alignment parameters[10]. In most libraries, less than 0.1% of the sequences ≥ 35 bp mapped against the human genome with a mapping quality ≥ 30 (see Extended Data Table 2). Owing to the small numbers of aligned sequences, C to T substitution frequencies cannot be determined with confidence in most cases. A small number of libraries, most notably the ones generated from surface material, stand out with high fractions of aligned sequences (up to 8.4%). However, the C to T substitution frequencies in these libraries are extremely low, indicating that the vast majority of sequences are derived from modern human contamination. Based on these analyses, none of the libraries prepared from the femur is a suitable candidate for deeper shotgun sequencing.

Enrichment of mitochondrial DNA and sequencing. All sample and blank libraries were first enriched for mitochondrial DNA using present-day human mitochondrial probes synthesized on an oligonucleotide array (in 3-bp tiling density, using human mtDNA sequence NC_001807), following the method described in ref. 33, except that the hybridization and wash temperatures were lowered to 60 °C and 55 °C to facilitate enrichment of short library molecules[14]. After phylogenetic analyses showed that the mitochondrial genome of the Sima de los Huesos hominin is closer to Denisovans than modern humans, the libraries were also enriched using Denisovan mitochondrial probes. To construct these probes, 19 overlapping DNA fragments of approximately 1 kb were designed (GeneArt Fragments, Life Technologies) using the mitochondrial genome of the Denisovan manual phalanx[23] as reference. The fragments encompassed the following sequence coordinates: 319–1289, 1223–2191, 2101–3088, 3018–3950, 3897–4889, 4806–5763, 5688–6663, 6612–7601, 7529–8486, 8428–9418, 9371–10300, 10203–11156, 11085–12017, 11966–12931, 12881–13813, 13762–14706, 14641–15600, 15551–16503 and 16460–381. One of the fragments (8428–9418) failed several synthesis attempts and could not be included in the probe pool. The 18 successfully synthesized fragments were amplified with Q5 Hot Start High-Fidelity DNA Polymerase (NEB) according to the supplier's instructions using a 5′ biotinylated forward primer and an unmodified reverse primer. Amplified fragments were purified using solid-phase reversible immobilization (SPRI) beads as described elsewhere[34] and pooled in equimolar ratios. Bead capture was performed as described in ref. 35, but with lowered hybridization and wash temperatures as detailed above. Two successive rounds of hybridization enrichment were carried out with both probe sets.

The enriched libraries were combined into three pools and sequenced on Illumina's HiSeq 2500 platform in rapid mode, using recipes for paired-end sequencing with two index reads ($96 + 7 + 96 + 7$ or $76 + 7 + 76 + 7$ cycles)[32]. The first pool (including libraries B2949–B2994 enriched with present-day human probes) was sequenced together with libraries from another experiment (mitochondrial captures of ancient human samples), occupying 75% of two lanes of a flow cell. The second and third pools (including libraries A1543–A2045 enriched with present-day human probes and all libraries enriched with Denisovan probes, respectively) were sequenced on one lane of a flow cell each.

Raw data processing and mapping. Base calling was performed using Bustard (Illumina) or freeIbis[36]. Sequences that did not perfectly match one of the expected index combinations were discarded and full-length molecule sequences were reconstructed by overlap-merging of paired-end reads[37]. Merged sequences ≥ 30bp were aligned against the revised Cambridge reference sequence (NC_012920) using BWA[19] with the parameters '−n 5', which allows up to five mismatches, and '−l 16500', which turns off seeding. Sequences with identical alignment start and end coordinates were collapsed into single sequences by consensus calling[23].

Enrichment success and recovery of mitochondrial sequences. The efficiency of enrichment varied between the human and Denisovan probe sets, with 6.8% and 27.6% of the sequences ≥ 30 bp aligning to the human mitochondrial reference genome before duplicate removal (compare Extended Data Table 3). Each unique sequence is represented by 21 duplicates on average (but note that this value is deflated by few libraries yielding very large numbers of sequences; see Extended Data Table 1), indicating that the libraries were sequenced to exhaustion. There are remarkable differences in the number of unique sequences obtained from each library (Extended Data Table 1 and Extended Data Fig. 2), ranging from 122 to 719 for the blank controls, from 448 to 9,757 for the libraries prepared from re-extracted bone pellets, and from 1,529 to 773,319 for the regular sample libraries. Extended Data Table 3 summarizes the number of sequences retained in the sample libraries after each processing step.

In previous work on the cave bear sample from Sima de los Huesos, almost all sequenced fragments (94%) were ≤ 50 bp in length[14]. The fragment size distribution inferred from the hominin sequences exhibits a larger proportion of longer molecules (Extended Data Fig. 3), possibly reflecting contaminant sequences. When stratifying terminal C to T substitution frequencies by sequence length, we find a strong decline of the deamination signal with length (Fig. 3b), indicating that the pronounced tail of long sequences is due to modern human contamination.

Basic filters applied to the sequences before consensus calling. We used the following set of filters to decrease the load of modern human contamination and to eliminate spurious alignments before consensus calling. The number of sequences retained after each step of filtering is provided in Extended Data Table 3. First, we excluded libraries without substantial signals of cytosine deamination (<5.0% terminal C to T substitution frequencies). Second, for mapping the sequences with BWA we allowed up to 5 mismatches and one insertion or deletion to prevent the loss of sequences with several damage-derived C to T substitutions, but these parameters are extremely permissive and do not sufficiently discriminative against spurious alignments. We therefore removed sequences showing more than two differences to the human mitochondrial reference genome that cannot be explained by cytosine deamination (that is, sequences with more than two non-C-to-T substitutions in the orientation as sequenced). In addition, we limited the maximum number of acceptable differences to the reference to five, counting also insertions or deletions. Third, sequences with a mapping quality of less than 30 were removed to ensure secure placement within the mitochondrial genome. Fourth, sequences longer than 45 bp were removed, because they are particularly rich in contamination (see Fig. 3b).

Overview of the consensus calling procedure. As the patterns of cytosine deamination showed that all libraries are substantially contaminated with modern human DNA, we used terminal C to T substitutions to enrich for endogenous sequences before consensus calling. The simplest approach for identifying sequences with deamination-induced C to T substitutions is by comparison to the human mitochondrial reference genome. However, human contaminants will occasionally show true C to T differences to the reference due to sequence divergence. To reduce carryover of such sequences, we developed another approach where we isolated sequences with a terminal T (or a T in the first three or last three positions; see below) if 80% or more of all sequences covering the respective position in the mitochondrial genome show a C. This procedure accounts for the fact that a C to T change is not indicative of cytosine deamination if it is shared by many other sequences, irrespective of the state of the reference. Information about the state of all sequences at each position of the mitochondrial genome was obtained using the 'mpileup' command implemented in SAMtools[38].

To reduce the effect of damage-induced C to T substitutions during consensus calling, we next converted Ts to Ns in the first and last three positions of each sequence. In these positions, cytosines are converted to uracils with frequencies $\gg$ 10% (see Fig. 3a). We again took the state of all other sequences into account, only converting T to N if at least one other sequence showed a C at the respective position.

After masking terminal C to T substitutions, the 'mpileup' command was used again to convert the BAM alignment file into a position-based tabular format. This table was used to determine (1) the coverage at each position, (2) the consensus base (based on a simple majority vote), and (3) the percentage of sequences supporting the majority base ('consensus support'). Ns were disregarded in all three measures. As BWA does not account for the circularity of the mitochondrial genome, mapping, filtering and consensus calling were repeated using a modified reference sequence where 1 kb of sequence was moved from start to end. This way we obtained the same measures also for the first and last bases of the mitochondrial genome.

Constructing consensus sequences under different filtering regimes. Accurate reconstruction of the mitochondrial genome sequence of the hominin femur sample is complicated by the high background of modern human contamination and the short size of endogenous DNA fragments. Short fragments are less efficiently enriched in hybridization capture[14], even more so if their sequences show differences to the capture probe. In addition, mapping bias may reduce the probability of identifying endogenous sequences if they differ from the reference sequence. Both mapping bias and contamination, if not effectively removed, would make the hominin consensus sequence more similar to modern human mitochondrial DNA. Capture bias goes in the same direction for the sequences enriched with present-day human probes, but is expected to increase similarity with Denisovans when using Denisovan probes. As the effects of modern human contamination, capture bias and mapping bias are expected to increase with the stringency of filtering (that is, with more stringent cutoffs for coverage and consensus support), we reconstructed the consensus sequence using three different filtering strategies to test whether filtering influences the results of the phylogenetic analyses.

The first consensus sequence is based on a very strict filtering regime and includes the positions that can be determined with highest confidence. Sequences were filtered for C to T substitutions based on the first or last base of each sequence only, thus using the positions providing most power to discriminate between contaminant and endogenous sequences. We then required a minimum coverage of 5 and a consensus support $>=$ 80% in order to call a consensus base. After visual inspection of the sequence alignments we removed three stretches of C-rich homopolymer sequence from the consensus (positions 286–315, 956–965 and 16180–16193 according to the revised Cambridge reference sequence (rCRS) coordinate system), because they are difficult to resolve with sequences enriched for C to T substitutions. With this procedure, 15,181 bp of the mitochondrial genome (~92% of the reference sequence) could be determined. Each determined position is covered 21.8 times on average. Coverage distribution along the mitochondrial genome is provided in Extended Data Fig. 4.

The second consensus sequence is more inclusive and recovers a larger fraction of the mitochondrial genome. We filtered the sequences for C to T substitutions using the first and last three bases of each sequence, thereby increasing the number of sequences available for consensus calling from 10,160 to 15,528 (Extended Data Table 3). In addition, we lowered the threshold for consensus calling to a minimum coverage of 3 and the required consensus support to be >67%. Three C-rich homopolymer stretches were removed as described above. Using this less stringent approach, the number of determined bases increases to 16,302 (~98% of the mitochondrial genome) and average coverage is 31.6 for the positions that were determined (compare Extended Data Fig. 4).

The third consensus sequence was generated to test whether phylogenetic analyses are influenced by the use of Denisovan capture probes. For this purpose we reprocessed the sequence data from the start, using only sequences generated in capture experiments with present-day human probes. A consensus was then called using the high-confidence criteria described above, yielding base calls for 13,157 positions of the mitochondrial genome determined with 16.3-fold average coverage. The sequence coverage of the mitochondrial genome obtained from the enrichments with present-day human probes (and for comparison with Denisovan probes) is shown in Extended Data Fig. 5.

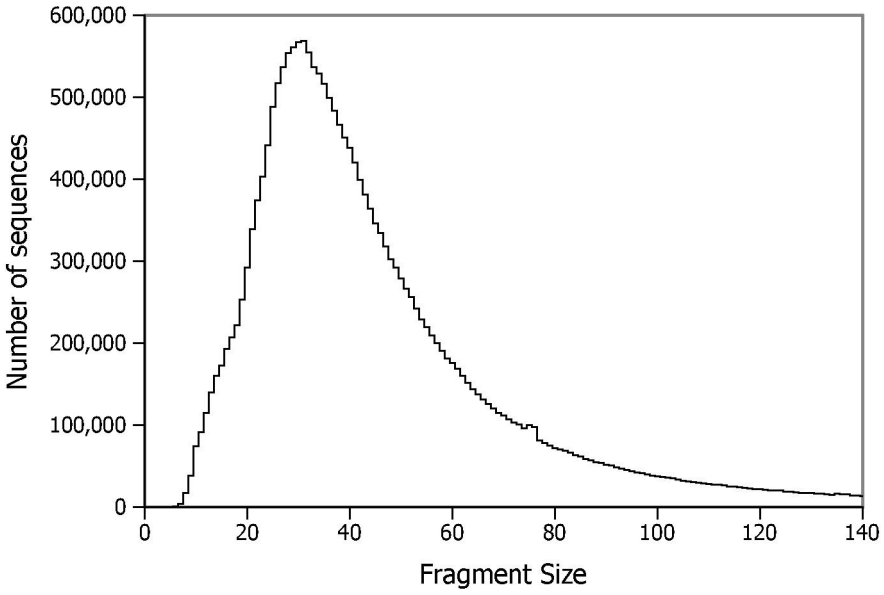
Phylogenetic reconstructions. Multiple sequence alignments were generated separately for each of the Sima de los Huesos consensus sequences using complete mitochondrial genome sequences of a worldwide panel of 54 present-day humans[39], 9 ancient humans[40], 7 Neanderthals (6 described in literature[41] and one deposited in GenBank with accession number KC879692), 2 Denosivans[9,23], 22 bonobos[42] and 24 chimpanzees[43] using MAFFT[44]. After removing the D-loop (rCRS positions 16023–577), we selected the general time reversible substitution model with invariant sites and a gamma distributed correction for rate heterogeneity (GTR + I + Γ) as suggested by MODELTEST[45]. Phylogenetic trees were reconstructed in a Bayesian statistical framework using MrBayes[24]. We performed four independent runs of Markov Chain Monte Carlo (MCMC) sampling with 30,000,000 generations, respectively. In each run, the first 3,000,000 generations were discarded as burn-in. All four consensus trees show the same topology of the major mitochondrial lineages (Fig. 4 and Extended Data Fig. 6) and group the Sima de los Huesos sequence with Denisovans with a posterior probability of 1.

We further estimated the divergence times among major mitochondrial lineages as well as the age of the Sima de los Huesos mitochondrial sequence based on its branch length in a Bayesian statistical framework[25] as implemented in BEAST[46]. For this analysis we used the same data set as described above in the MrBayes section. To inform the molecular clock rate estimate we used nine ancient modern human and four Neanderthal complete mitochondrial genome sequences from radiocarbon dated specimens[40]. For the ancient individuals of unknown age we used uniform priors ranging from 0 to 1,000,000 years BP. Two different models of rate variation among branches were tested: a strict clock and an uncorrelated lognormal-distributed relaxed clock, both under a constant size and a Bayesian skyline coalescent tree prior. For each of these four analyses, two Markov chain Monte Carlo (MCMC) runs of 30,000,000 generations with samples taken every 1,000 generations were performed, respectively. The first 6,000,000 iterations were discarded as burn-in and the remaining were combined using LogCombiner, resulting in a total of 48,000,000 generations per analysis to ensure sufficient sampling of parameters with effective sample sizes (ESS) of >200. When comparing the strict versus the relaxed clock model using Bayes Factors test[47], we found strong support in favour of the relaxed clock model ($\log_{10}$ BF $\geq$ 1.13 for all three consensus sequences). The constant size coalescent could not be rejected over the Bayesian skyline coalescent ($\log_{10}$ BF $\leq$ 0.39 for all three consensus sequences). We therefore used the analysis based on the relaxed clock model and the constant size coalescent before proceed estimating the divergence times among various clades as reported in Table 1 and Extended Data Table 4.

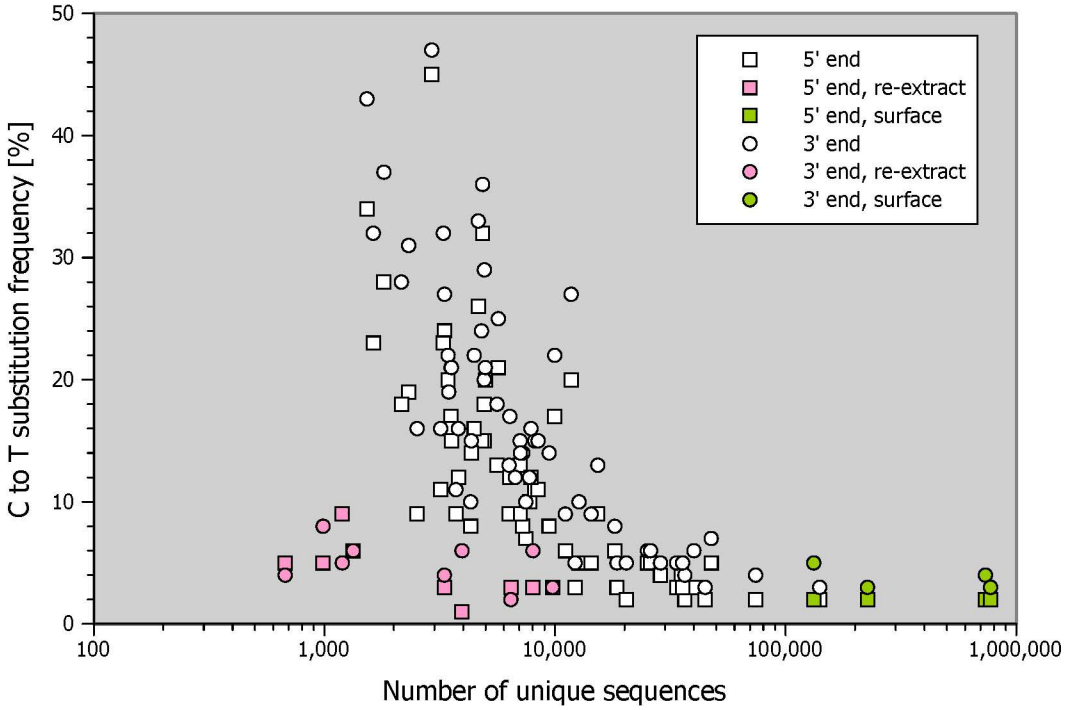
The divergence time estimates are stable and independent of which Sima de los Huesos consensus sequence was used (see the consensus calling section above). The 95% highest posterior density (HPD) intervals of all estimates are in agreement and include previous estimates based on mtDNA sequences, for example, for the divergence between humans and chimpanzees 4.2–5.2 Myr ago[48], between Denisovans and modern humans 0.6–1.3 Myr ago[23,40], between Neanderthal and modern humans 0.3–0.6 Myr ago[41,49], and between chimpanzees and bonobos 1.5–2.1 Myr ago[50] as well as the time to the most recent common ancestor of all humans around 120,000–236,000 years ago[40,51].

31. 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).
32. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40,** e3 (2012).
33. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110,** 2223–2227 (2013).
34. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22,** 939–946 (2012).
35. Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5,** e14004 (2010).
36. Renaud, G., Kircher, M., Stenzel, U. & Kelso, J. freelbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29,** 1208–1209 (2013).
37. Kircher, M. Analysis of high-throughput ancient DNA sequencing data. *Methods Mol. Biol.* **840,** 197–228 (2012).
38. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
39. Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708–713 (2000).
40. Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23,** 553–559 (2013).

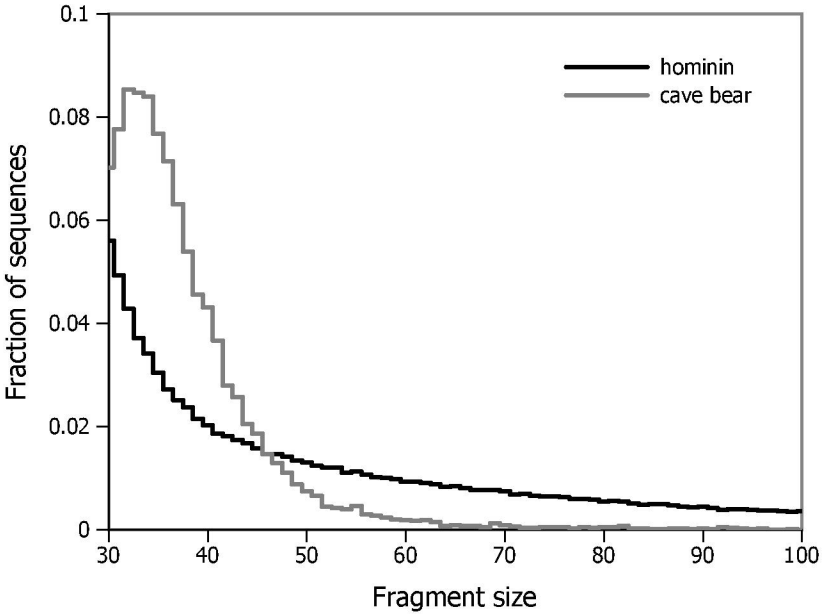
41. Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325,** 318–321 (2009).
42. Zsurka, G. *et al.* Distinct patterns of mitochondrial genome diversity in bonobos (*Pan paniscus*) and humans. *BMC Evol. Biol.* **10,** 270 (2010).
43. Bjork, A., Liu, W., Wertheim, J. O., Hahn, B. H. & Worobey, M. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol. Biol. Evol.* **28,** 615–623 (2011).
44. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 (2013).
45. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14,** 817–818 (1998).
46. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7,** 214 (2007).
47. Kass, R. E. & Raftery, A. E. Bayes Factors. *J. Am. Stat. Assoc.* **90,** 773–795 (1995).
48. Horai, S. *et al.* Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* **35,** 32–43 (1992).
49. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134,** 416–426 (2008).
50. Stone, A. C. *et al.* More reliable estimates of divergence times in *Pan* using complete mtDNA sequences and accounting for population structure. *Phil. Trans. R. Soc. Lond. B* **365,** 3277–3288 (2010).
51. Soares, P. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84,** 740–759 (2009).



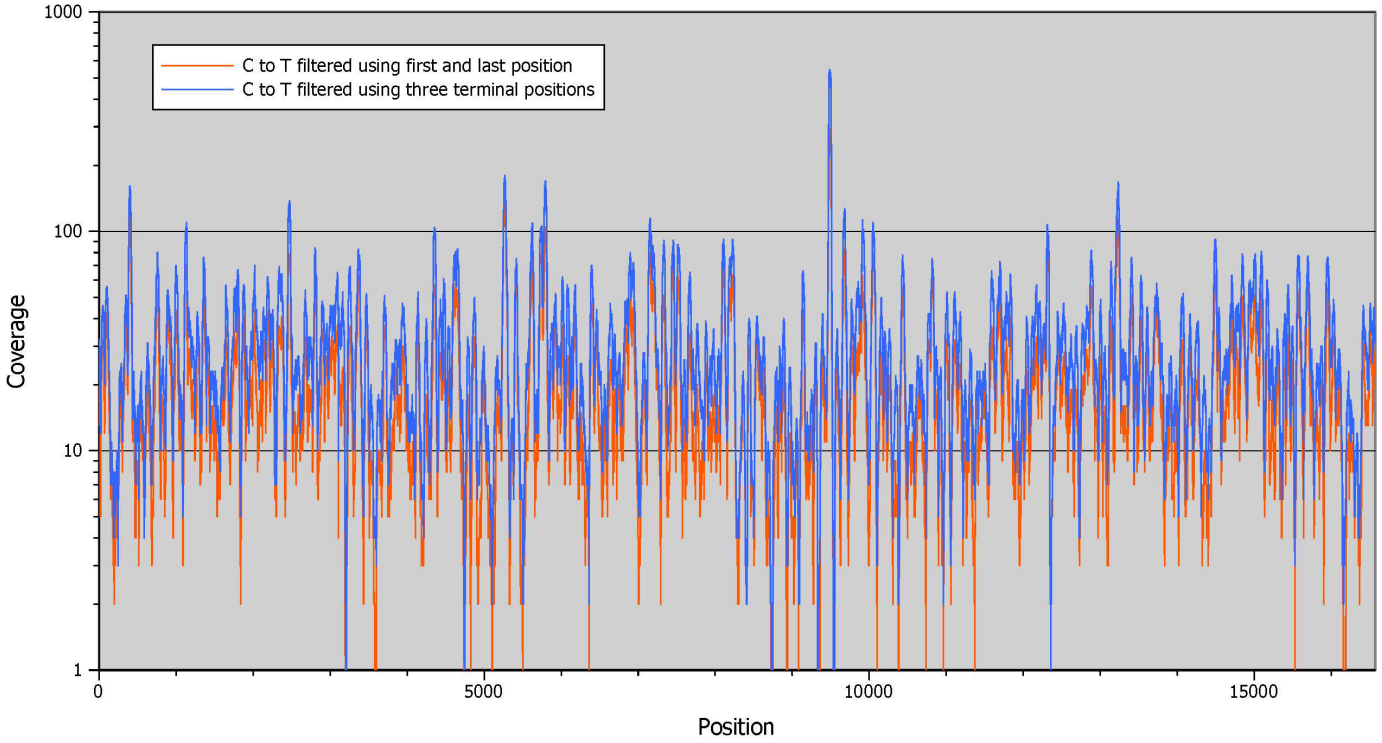
Extended Data Figure 1 | Size distribution of all overlap-merged sequences generated by shotgun sequencing (before mapping).



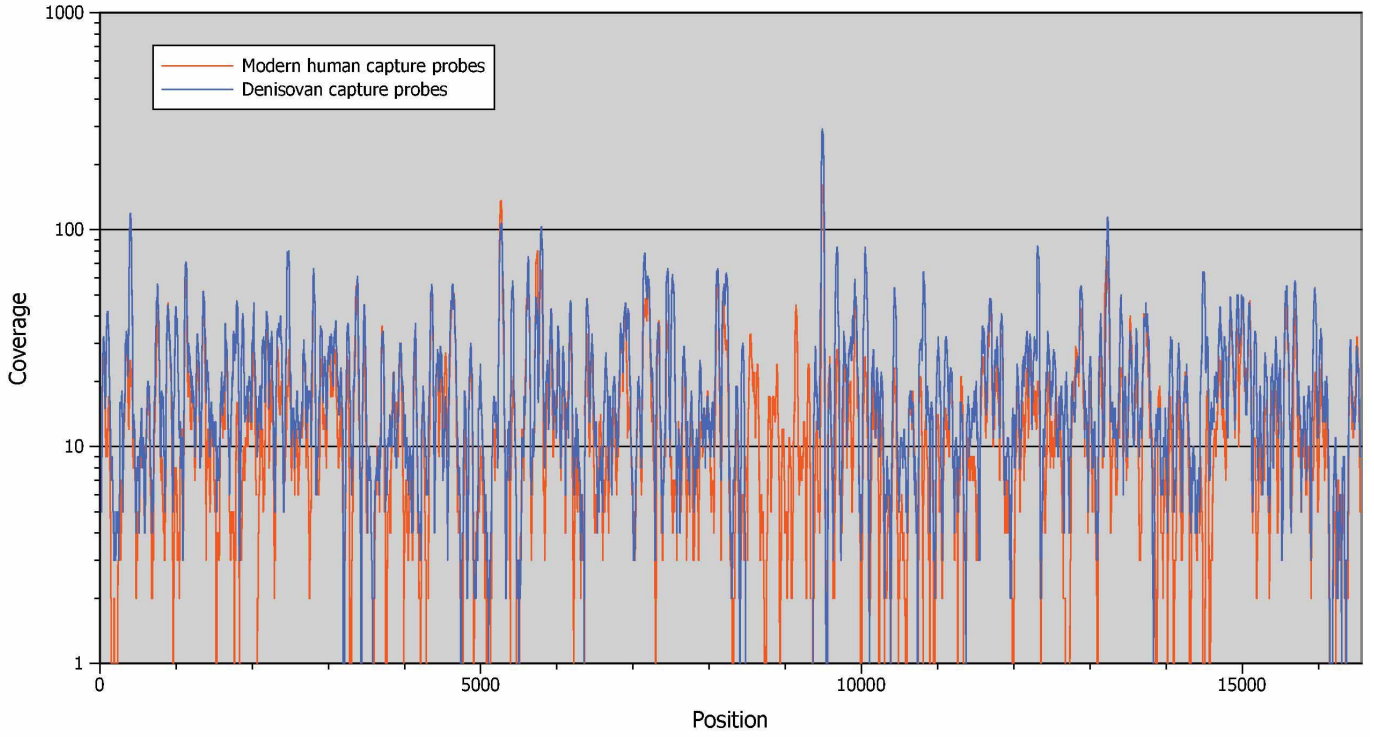
Extended Data Figure 2 | 5′ and 3′ C to T substitution frequencies plotted against the number of unique mitochondrial sequences retrieved from each sample library. Libraries prepared from re-extracted pellets or surface material are highlighted in colour.



Extended Data Figure 3 | Sequence length distribution of unique sequences. The distribution obtained from the Sima de los Huesos cave bear is shown for comparison.



Extended Data Figure 4 | Sequence coverage of the mitochondrial genome obtained from sequences with terminal C to T substitutions.



Extended Data Figure 5 | Sequence coverage of the mitochondrial genome plotted separately for both capture probe sets used (based on sequences with a C to T substitution at the first or last alignment position).

Extended Data Figure 6 | Complete view of the mid-point rooted phylogenetic tree constructed with a Bayesian approach under a GTR + I + Γ model of sequence evolution using the Sima de los Huesos consensus sequence generated with inclusive filters as well as 54 present-day humans, 9 ancient humans, 7 Neanderthals, 2 Denosivans, 22 bonobos and 24 chimpanzees. The posterior probabilities are provided for the major nodes.

Extended Data Table 1 | Characteristics of all libraries prepared for this study

Library set	Extracted material	DNA extract ID	Bone material used for DNA extr. [mg]	Library ID	Fraction of extract used for library prep.	qPCR molecule count	Unique mtDNA sequences (>= 30bp)	Average number of sequence replicates	C to T substitution frequencies [%] 5' / 3'	Sequences retained after filtering for terminal C to T subst.
1	sample	E1385	25	B2949	0.4	1.40E+10	3,717	72	9 / 11	80
1	sample	E1386	33	B2950	0.4	1.36E+10	4,307	69	8 / 10	78
1	sample	E1387	39	B2951	0.4	1.33E+10	7,754	75	10 / 12	175
1	sample	E1388	47	B2952	0.5	1.60E+10	3,317	70	24 / 27	217
1	sample	E1389	48	B2953	0.5	2.15E+10	4,936	71	15 / 20	212
1	sample	E1390	65	B2954	0.4	1.91E+10	3,280	70	23 / 32	247
1	blank	E1391		B2955	0.4	3.58E+07	433	895	N.R. / N.R.	
1	sample	E1392	42	B2956	0.5	7.81E+09	12,712	73	5 / 5	
1	sample	E1393	47	B2957	0.5	1.43E+10	8,168	87	11 / 15	232
1	sample	E1394	75	B2958	0.45	1.46E+10	7,059	84	13 / 15	188
1	sample	E1394	75	B2959	0.45	1.29E+10	8,447	83	11 / 15	214
1	blank	E1352		B2960	0.5	2.84E+07	342	876	N.R. / N.R.	
1	sample	E1348	26	B2961	0.5	5.41E+09	36,547	64	2 / 4	
1	sample	E1349	25	B2962	0.5	9.98E+09	7,466	78	7 / 10	133
1		blank		B2963		1.59E+07	156	715	N.R. / N.R.	
2	sample	E1467	44	B2980	0.5	1.03E+10	35,264	49	4 / 5	
2	sample	E1468	58	B2981	0.5	2.02E+10	40,006	42	3 / 6	
2	sample	E1469	45	B2982	0.5	1.53E+10	25,167	52	5 / 6	
2	blank	E1470		B2983	0.5	3.34E+07	362	571	N.R. / N.R.	
2	sample	E1467	44	B2984	0.5	1.15E+10	33,655	41	3 / 5	
2	sample	E1468	58	B2985	0.5	1.43E+10	35,718	48	3 / 5	
2	sample	E1469	45	B2986	0.5	1.74E+10	25,957	51	5 / 6	335
2	sample	E1344	31	B2987	0.6	2.02E+10	4,984	40	20 / 21	272
2	sample	E1345	30	B2988	0.5	1.23E+10	6,725	56	9 / 12	159
2	sample	E1348	26	B2989	0.5	9.85E+09	44,760	46	2 / 3	
2	sample	E1349	25	B2990	0.5	9.10E+09	6,323	65	9 / 13	168
2	sample	E1350	34	B2991	0.5	8.42E+09	4,799	90	15 / 24	189
2	sample	E1345	30	B2992	0.5	1.29E+10	7,251	63	8 / 14	186
2	sample	E1350	34	B2993	0.5	1.11E+10	5,606	93	13 / 18	223
2		blank		B2994		9.82E+06	168	2031	N.R. / N.R.	
3	pellet E1388	E1511		A1543	0.5	1.68E+08	676	925	5 / 4	
3	pellet E1390	E1512		A1544	0.5	1.09E+09	1,333	268	6 / 6	11
3	pellet E1350	E1513		A1545	0.5	2.98E+08	987	850	5 / 8	8
3	pellet E1468	E1514		A1546	0.5	9.06E+08	9,757	90	3 / 3	
3	pellet E1348	E1515		A1547	0.5	1.49E+08	3,313	414	3 / 4	
3	sample	E1516	50	A1548	0.5	1.36E+10	28,670	26	4 / 5	
3	sample	E1517	46	A1549	0.5	2.08E+10	7,867	39	12 / 16	255
3	sample	E1518	50	A1550	0.5	2.27E+10	7,076	31	9 / 14	170
3	sample	E1519	48	A1551	0.5	1.54E+10	4,852	48	32 / 36	455
3	sample	E1520	50	A1552	0.5	2.42E+10	11,749	28	20 / 27	726
3	sample	E1521	50	A1553	0.5	1.67E+10	140,535	10	2 / 3	
3	sample	E1522	62	A1554	0.5	1.19E+10	47,557	22	5 / 7	
3	sample	E1523	46	A1555	0.5	1.56E+10	9,972	35	17 / 22	487
3	blank	E1524		A1556	0.5	3.87E+07	719	264	1 / N.R.	
3		blank		A1557		1.85E+07	*not sequenced*			
4		blank		A2001		3.32E+07	153	1340	N.R. / N.R.	
4	pellet E1388	E1511		A2002	0.5	1.32E+08	487	1100	7 / N.R.	7
4	pellet E1390	E1512		A2003	0.5	5.35E+08	697	502	4 / N.R.	
4	pellet E1350	E1513		A2004	0.5	1.48E+08	448	1143	4 / N.R.	
4	pellet E1468	E1514		A2005	0.5	3.79E+08	3,951	182	1 / 6	
4	pellet E1348	E1515		A2006	0.5	7.61E+07	1,196	905	9 / 5	
4	sample	E1516	50	A2007	0.5	3.23E+09	12,219	65	3 / 5	
4	sample	E1517	46	A2008	0.5	3.81E+09	3,193	130	11 / 16	79
4	sample	E1518	50	A2009	0.5	5.75E+09	2,524	73	9 / 16	59
4	sample	E1519	48	A2010	0.5	4.30E+09	1,812	81	28 / 37	164
4	sample	E1520	50	A2011	0.5	7.29E+09	4,940	54	18 / 29	244
4	sample	E1521	50	A2012	0.5	8.42E+09	74,091	20	2 / 4	
4	sample	E1522	62	A2013	0.5	3.72E+09	14,340	55	5 / 9	139
4	sample	E1523	46	A2014	0.5	4.69E+09	3,542	102	17 / 21	152
4	blank	E1524		A2015	0.5	4.79E+07	372	460	N.R. / N.R.	
5		blank		A2016		9.30E+06	203	1019	N.R. / N.R.	
5	sample	E1534	47	A2017	0.6	1.26E+10	4,335	64	14 / 15	135
5	sample	E1535	47	A2018	0.6	6.81E+09	3,437	76	20 / 22	185
5	sample	E1536	50	A2019	0.6	4.09E+09	4,464	80	16 / 22	190
5	sample	E1537	53	A2020	0.6	1.18E+10	6,371	42	12 / 17	241
5	sample	E1538	63	A2021	0.6	5.73E+09	2,921	88	45 / 47	395
5	sample	E1539	50	A2022	0.6	7.89E+09	4,648	65	26 / 33	414
5	sample	E1540	43	A2023	0.6	9.56E+09	5,679	46	21 / 25	350
5	sample	E1541	49	A2024	0.6	4.14E+09	18,565	51	3 / 5	
5	sample	E1542	44	A2025	0.6	4.81E+09	18,168	48	6 / 8	286
5	sample	E1543	43	A2026	0.6	1.00E+10	15,328	34	9 / 13	403
5	surface	E1544	56	A2027	0.6	4.37E+09	132,351	14	2 / 5	
5	surface	E1545	65	A2028	0.6	1.09E+10	734,511	3	2 / 4	
5	pellet E1522	E1546		A2029	0.6	6.70E+08	8,024	98	3 / 6	
5	blank	E1547		A2030	0.6	2.39E+07	309	489	N.R. / N.R.	
6		blank		A2031		4.92E+06	122	1581	N.R. / N.R.	
6	sample	E1534	47	A2032	0.4	6.06E+09	3,810	81	12 / 16	112
6	sample	E1535	47	A2033	0.4	3.08E+09	2,320	124	19 / 31	137
6	sample	E1536	50	A2034	0.4	3.08E+09	3,461	103	16 / 19	137
6	sample	E1537	53	A2035	0.4	5.04E+09	3,553	86	15 / 21	149
6	sample	E1538	63	A2036	0.4	3.20E+09	1,529	118	34 / 43	139
6	sample	E1539	50	A2037	0.4	1.92E+09	1,629	165	23 / 32	132
6	sample	E1540	43	A2038	0.4	2.88E+09	2,157	115	18 / 28	119
6	sample	E1541	49	A2039	0.4	7.59E+09	20,365	40	2 / 5	
6	sample	E1542	44	A2040	0.4	3.25E+09	11,083	70	6 / 9	163
6	sample	E1543	43	A2041	0.4	6.07E+09	9,430	59	8 / 14	209
6	surface	E1544	56	A2042	0.4	8.83E+09	226,317	8	2 / 3	
6	surface	E1545	65	A2043	0.4	1.71E+10	773,319	3	2 / 3	
6	pellet E1522	E1546		A2044	0.4	5.01E+08	6,435	155	3 / 2	
6	blank	E1547		A2045	0.4	1.02E+07	262	750	N.R. / N.R.	

Substitution frequencies are not reported (N.R.) if based on fewer than 100 observations.

Extended Data Table 2 | Results from shallow shotgun sequencing of a subset of libraries

Extracted material	library ID	total number of sequences (overlap-merged, >= 35bp)	mapped sequences (map quality >= 30)	fraction mapped [%]	5' C to T substitution frequency [%] (observations)	3' C to T substitution frequency [%] (observations)
pellet E1388	A1543	330,819	535	0.162	1 (1/110)	4 (3/70)
pellet E1390	A1544	609,472	200	0.033	0 (0/45)	0 (0/27)
pellet E1350	A1545	382,645	307	0.080	0 (0/63)	6 (3/51)
pellet E1468	A1546	562,458	1,375	0.244	3 (7/248)	2 (4/180)
pellet E1348	A1547	382,039	794	0.208	1 (2/141)	1 (1/126)
sample	A1548	569,592	535	0.094	1 (1/104)	2 (1/54)
sample	A1549	383,570	82	0.021	5 (1/19)	10 (1/10)
sample	A1550	556,828	45	0.008	0 (0/6)	0 (0/2)
sample	A1551	521,648	60	0.012	0 (0/10)	0 (0/6)
sample	A1552	492,761	107	0.022	10 (2/20)	0 (0/9)
sample	A1553	391,037	2,575	0.659	1 (7/477)	3 (8/301)
sample	A1554	522,604	934	0.179	3 (5/185)	6 (6/102)
sample	A1555	439,135	73	0.017	0 (0/14)	29 (2/7)
pellet E1388	A2002	80,590	111	0.138	0 (0/22)	0 (0/9)
pellet E1390	A2003	176,798	47	0.027	0 (0/6)	0 (0/5)
pellet E1350	A2004	80,519	86	0.107	0 (0/12)	8 (1/12)
pellet E1468	A2005	153,555	441	0.287	0 (0/78)	0 (0/52)
pellet E1348	A2006	21,876	41	0.187	0 (0/6)	0 (0/3)
sample	A2007	198,123	169	0.085	0 (0/38)	0 (0/26)
sample	A2008	182,797	44	0.024	0 (0/8)	0 (0/5)
sample	A2009	271,679	19	0.007	0 (0/3)	0 (0/3)
sample	A2010	222,503	23	0.010	0 (0/6)	0 (0/2)
sample	A2011	193,697	50	0.026	0 (0/8)	0 (0/5)
sample	A2012	239,484	1,559	0.651	1 (4/291)	3 (5/187)
sample	A2013	171,305	308	0.180	2 (1/51)	0 (0/26)
sample	A2014	242,819	43	0.018	11 (1/9)	0 (0/6)
sample	A2017	172,196	21	0.012	0 (0/5)	0 (0/4)
sample	A2018	182,640	33	0.018	0 (0/6)	22 (2/9)
sample	A2019	241,673	50	0.021	0 (0/7)	17 (1/6)
sample	A2020	188,983	51	0.027	10 (1/10)	0 (0/6)
sample	A2021	148,940	21	0.014	50 (1/2)	50 (1/2)
sample	A2022	179,754	52	0.029	0 (0/13)	0 (0/4)
sample	A2023	169,802	64	0.038	0 (0/8)	0 (0/9)
sample	A2024	171,019	240	0.140	0 (0/51)	0 (0/25)
sample	A2025	200,408	462	0.231	1 (1/75)	2 (1/61)
sample	A2026	158,056	130	0.082	0 (0/25)	0 (0/16)
surface	A2027	204,421	3,977	1.945	2 (14/814)	3 (12/353)
surface	A2028	210,913	17,647	8.367	2 (55/3379)	3 (49/1790)
pellet E1522	A2029	178,297	654	0.367	0 (0/129)	0 (0/74)
sample	A2032	172,846	24	0.014	0 (0/2)	0 (0/2)
sample	A2033	159,571	11	0.007	0 (0/1)	0 (0/0)
sample	A2034	194,795	36	0.018	25 (2/8)	0 (0/4)
sample	A2035	196,685	62	0.032	8 (1/12)	0 (0/10)
sample	A2036	121,160	23	0.019	0 (0/7)	20 (1/5)
sample	A2037	158,477	45	0.028	0 (0/8)	0 (0/2)
sample	A2038	212,279	75	0.035	0 (0/20)	14 (1/7)
sample	A2039	185,602	247	0.133	0 (0/47)	3 (1/32)
sample	A2040	147,584	365	0.247	2 (1/61)	0 (0/47)
sample	A2041	174,738	168	0.096	3 (1/30)	15 (3/20)
surface	A2042	194,501	3,522	1.811	1 (6/666)	2 (9/422)
surface	A2043	189,332	14,305	7.556	1 (34/2770)	2 (44/1903)
pellet E1522	A2044	167,925	534	0.318	2 (2/101)	0 (0/71)

Extended Data Table 3 | Number of sequences retained in the sample libraries after each step of processing and filtering

Filter	Enrichment with human capture probes	Enrichment with Denisovan capture probes	Combined
All sequences	402,938,161	154,005,123	556,943,284
Overlap-merged	367,593,202	146,852,057	514,445,259
Length >= 30bp	320,441,187	124,694,112	445,135,299
Mapped	21,634,872	34,388,775	56,023,647
Unique	2,171,801		2,805,919
In libraries with terminal C to T substitution frequencies >= 5.0%	219,766		299,093
<= 2 mismatches other than C to T	197,019		264,319
Mapping quality >=30	196,536		263,302
Length <= 45	82,239		105,070
Terminal C to T substitution / C to T substitution in first or last three alignment positions	6,706		10,160 / 15,528

Sequences obtained from the blank controls are not included. All filters are additive in the order indicated.

Extended Data Table 4 | Inferred time to the most recent common ancestor (TMRCA) of the modern human, Neanderthal, chimpanzee and bonobo mtDNAs, as well as divergence estimates for human/chimpanzee and bonobo/chimpanzee mtDNA (continuation of Table 1)

Mitochondrial lineages	Divergence dates and molecular age estimates in million years before present [95% HPD interval in brackets]		
	Strict filters	Inclusive filters	Strict filters, enriched with human probes only
TMRCA modern humans	0.21 [0.15-0.29]	0.21 [0.15-0.30]	0.18 [0.12-0.25]
TMRCA Neanderthals	0.15 [0.11-0.20]	0.15 [0.11-0.19]	0.13 [0.10-0.17]
TMRCA chimpanzees	0.86 [0.59-1.16]	0.89 [0.61-1.22]	0.73 [0.48-1.04]
TMRCA bonobos	0.37 [0.24-0.50]	0.36 [0.24-0.50]	0.30 [0.19-0.43]
Human - Chimpanzee	6.18 [4.33-8.45]	6.44 [4.40-8.82]	5.43 [3.54-7.64]
Chimpanzee - Bonobo	1.83 [1.27-2.48]	1.91 [1.30-2.61]	1.61 [1.06-2.27]