

University of Wollongong

Research Online

---

Faculty of Science, Medicine and Health -  
Papers: part A

Faculty of Science, Medicine and Health

---

1-1-2017

## Neandertal and Denisovan DNA from Pleistocene sediments

Viviane Slon

*Max Planck Institute for Evolutionary Anthropology*

Charlotte Hopfe

*Max Planck Institute for Evolutionary Anthropology*

Clemens Weiß

*Max Planck Institute for Developmental Biology*

Fabrizio Mafessoni

*Max Planck Institute for Evolutionary Anthropology*

Marco de la Rasilla

*University of Oviedo*

*See next page for additional authors*

Follow this and additional works at: <https://ro.uow.edu.au/smhpapers>



Part of the [Medicine and Health Sciences Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

### Recommended Citation

Slon, Viviane; Hopfe, Charlotte; Weiß, Clemens; Mafessoni, Fabrizio; de la Rasilla, Marco; Lalueza-Fox, Carles; Rosas, Antonio; Soressi, Marie; Knul, Monika; Miller, Rebecca; Jacobs, Zenobia; Li, Bo; and Roberts, Richard G., "Neandertal and Denisovan DNA from Pleistocene sediments" (2017). *Faculty of Science, Medicine and Health - Papers: part A*. 4653.  
<https://ro.uow.edu.au/smhpapers/4653>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

## Neandertal and Denisovan DNA from Pleistocene sediments

### Abstract

Although a rich record of Pleistocene human-associated archaeological assemblages exists, the scarcity of hominin fossils often impedes the understanding of which hominins occupied a site. Using targeted enrichment of mitochondrial DNA, we show that cave sediments represent a rich source of ancient mammalian DNA that often includes traces of hominin DNA, even at sites and in layers where no hominin remains have been discovered. By automation-assisted screening of numerous sediment samples, we detected Neandertal DNA in eight archaeological layers from four caves in Eurasia. In Denisova Cave, we retrieved Denisovan DNA in a Middle Pleistocene layer near the bottom of the stratigraphy. Our work opens the possibility of detecting the presence of hominin groups at sites and in areas where no skeletal remains are found.

### Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

### Publication Details

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Jacobs, Z., Li, B., Roberts, R. G. et al (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, 356 (6338), 605-608.

### Authors

Viviane Slon, Charlotte Hopfe, Clemens Weiß, Fabrizio Mafessoni, Marco de la Rasilla, Carles Lalueza-Fox, Antonio Rosas, Marie Soressi, Monika Knul, Rebecca Miller, Zenobia Jacobs, Bo Li, and Richard G. Roberts

## **Title: Neandertal and Denisovan DNA from Pleistocene sediments**

**Authors:** Viviane Slon<sup>1,\*</sup>, Charlotte Hopfe<sup>1</sup>, Clemens L. Weiß<sup>2</sup>, Fabrizio Mafessoni<sup>1</sup>, Marco de la Rasilla<sup>3</sup>, Carles Lalueza-Fox<sup>4</sup>, Antonio Rosas<sup>5</sup>, Marie Soressi<sup>6,7</sup>, Monika V. Knul<sup>8</sup>, Rebecca Miller<sup>9</sup>, John R. Stewart<sup>8</sup>, Anatoly P. Derevianko<sup>10,11</sup>, Zenobia Jacobs<sup>12,13</sup>, Bo Li<sup>12</sup>, Richard G. Roberts<sup>12,13</sup>, Michael V. Shunkov<sup>10,14</sup>, Henry de Lumley<sup>15,16</sup>, Christian Perrenoud<sup>15</sup>, Ivan Gušić<sup>17</sup>, Željko Kućan<sup>17</sup>, Pavao Rudan<sup>17</sup>, Ayinuer Aximu-Petri<sup>1</sup>, Elena Essel<sup>1</sup>, Sarah Nagel<sup>1</sup>, Birgit Nickel<sup>1</sup>, Anna Schmidt<sup>1</sup>, Kay Prüfer<sup>1</sup>, Janet Kelso<sup>1</sup>, Hernán A. Burbano<sup>2</sup>, Svante Pääbo<sup>1</sup>,  
Matthias Meyer<sup>1,\*</sup>

### **Affiliations:**

<sup>1</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

<sup>2</sup>Research Group for Ancient Genomics and Evolution, Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany

<sup>3</sup>Área de Prehistoria, Department of History, Universidad de Oviedo, Calle Teniente Alfonso Martínez s/n, 33011 Oviedo, Spain

<sup>4</sup>Institute of Evolutionary Biology (UPF-CSIC), 08003 Barcelona, Spain

<sup>5</sup>Departamento de Paleobiología, Museo Nacional de Ciencias Naturales, CSIC, 28006 Madrid, Spain

<sup>6</sup>Faculty of Archaeology, Leiden University, 2333CC Leiden, The Netherlands

<sup>7</sup>Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology,  
Deutscher Platz 6, 04103 Leipzig, Germany

<sup>8</sup>School of Applied Sciences, Bournemouth University, Poole, Dorset, UK

<sup>9</sup>Service de Préhistoire, Université de Liège, 4000 Liège, Belgium

<sup>10</sup>Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch,  
Novosibirsk, RU-630090, Russia

<sup>11</sup>Altai State University, Barnaul, RU-656049, Russia

<sup>12</sup>Centre for Archaeological Science, School of Earth and Environmental Sciences, University of  
Wollongong, Wollongong, New South Wales 2522, Australia

<sup>13</sup>ARC Centre of Excellence for Australian Biodiversity and Heritage, University of  
Wollongong, Wollongong, New South Wales 2522, Australia

<sup>14</sup>Novosibirsk National Research State University, Novosibirsk, RU-630090, Russia

<sup>15</sup>Centre Européen de Recherches Préhistoriques de Tautavel, 66720 Tautavel, France

<sup>16</sup>Institut de Paléontologie Humaine, 75013 Paris, France

<sup>17</sup>Anthropology Center of the Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia

\*Correspondence to: [viviane\\_slon@eva.mpg.de](mailto:viviane_slon@eva.mpg.de) or [mmeyer@eva.mpg.de](mailto:mmeyer@eva.mpg.de).

**Abstract:** Although a rich record of Pleistocene human-associated archaeological assemblages exists, the scarcity of hominin fossils often impedes the understanding of which hominins occupied a site. Using targeted enrichment of mitochondrial DNA we show that cave sediments represent a rich source of ancient mammalian DNA that often includes traces of hominin DNA, even at sites and in layers where no hominin remains have been discovered. By automation-assisted screening of numerous sediment samples we detect Neandertal DNA in eight archaeological layers from four caves in Eurasia. In Denisova Cave we retrieved Denisovan DNA in a Middle Pleistocene layer near the bottom of the stratigraphy. Our work opens the possibility to detect the presence of hominin groups at sites and in areas where no skeletal remains are found.

**One Sentence Summary:** DNA from archaic humans can be retrieved from Late and Middle Pleistocene sediments, even in the absence of their skeletal remains.

**Main Text:**

DNA recovered from ancient hominin remains enriches our understanding of human evolution and dispersal (*e.g.* (1) and references therein), and has, for example, resulted in the discovery of the Denisovans, a previously unknown group of archaic hominins in Asia who were distantly related to Neandertals (2-4). However, hominin fossils are rare. We therefore decided to investigate whether hominin DNA may survive in sediments at archaeological sites in the absence of macroscopically visible skeletal remains.

Mineral and organic components in sediments can bind DNA (*e.g.* (5-8)) (Figs. S1-S3) and the amplification of short stretches of mitochondrial (mt) or chloroplast DNA from sediments by the polymerase chain reaction (PCR) has been used to demonstrate the past presence of animals and plants at several sites (*e.g.* (9-14)). More recently, DNA extracted from sediments has been converted to DNA libraries, from which DNA fragments were sequenced directly (“shotgun” sequencing) (15, 16). This approach is preferable to PCR as it allows the entire sequence of DNA fragments to be determined. This is important as it makes it possible to detect cytosine (C) to thymine (T) substitutions near the ends of DNA fragments, which are caused by the deamination of cytosine bases (17) and indicate that the DNA is of ancient origin (18-20). However, the abundance of bacterial DNA in sediments and the difficulty in assigning short nuclear DNA sequences to mammalian taxa limit the utility of shotgun sequencing for analyzing DNA from sediments.

**Isolating DNA from Pleistocene cave sediments**

To investigate whether ancient mammalian DNA, especially of archaic humans, may be preserved in Pleistocene cave sediments, we collected 85 samples from seven archaeological

sites with known hominin occupation, varying in age between ~14 thousand years ago (kya) and  $\geq 550$  kya (Data file S1) (8). Some samples were collected specifically for the purpose of this study: 4 from Les Cottés (France), 5 from Trou Al'Wesse (Belgium), 1 from El Sidrón (Spain), 1 from Vindija Cave (Croatia), 3 from Denisova Cave (Russia) and 13 from Caune de l'Arago (France). The other samples, 49 from Denisova Cave and 9 from Chagyrskaya Cave (Russia), had been collected previously for luminescence dating. The latter two sites are located in the Altai Mountains, where remains of both Neandertals and Denisovans have been uncovered (3, 21). We extracted DNA from between 38 and 160 milligrams of each sample and converted aliquots of the DNA to single-stranded DNA libraries (8, 22, 23). All libraries were shotgun sequenced and analyzed using a taxonomic binning approach (8). Whereas most of the DNA sequences (79.1%-96.1%) remained unidentified, the majority of those that could be identified were assigned to microorganisms and between 0.05% and 10% to mammals (Figs. S7-S15).

#### Enrichment of mammalian mtDNA

To determine the taxonomic composition of the mammalian DNA in the sediments, we isolated DNA fragments bearing similarities to mammalian mtDNAs by hybridization capture using probes for 242 mitochondrial genomes, including human mtDNA (8, 24). MtDNA is useful for this purpose because it is present in higher copy numbers than nuclear DNA in most eukaryotic cells and is phylogenetically informative in spite of its small size due to its fast rate of evolution in mammals. Between 3,535 and 3.2 million DNA fragments were sequenced per library (Data file S2), of which between 14 and 50,114 could be assigned to mammalian families with a strategy for taxonomic identification of short and damaged DNA fragments (8) (Fig. S18). To assess whether the sequences were of ancient origin, we evaluated them for the presence of C to

T substitutions at their 5'- and 3'-ends (17, 18) (see Fig. S19 for an example). Additionally, we computed the variance of coverage across the mitochondrial genome for each taxon to test whether sequences mapped randomly across the reference genome (Fig. S20), as would be expected for sequences that are genuinely derived from the taxon they are assigned to. With the exception of 46 sequences from a single sample from Les Cottés, which were originally attributed to procaviids but that mapped only to one restricted region of the genome (Fig. S21), this analysis lent support to the correct taxonomic classification of the sequences we obtained.

Of the 52 sediment samples from the Late Pleistocene, 47 contained mtDNA fragments from at least one family showing evidence of ancient DNA-like damage, while 14 out of 33 Middle Pleistocene samples did so (Figs. 1, S22). Overall, we detected ancient mtDNA fragments from 12 mammalian families, of which the most common were hyaenids, bovids, equids, cervids and canids (Data file S3, Figs. S23-S32). These taxa are all present in the zooarchaeological records of the sites as reconstructed from faunal remains (Fig. S33).

We exploited the known genetic variation within these families to determine the affinity of the sequences we obtained to specific species (8) (Data file S3). In all libraries containing elephantid DNA, the majority (71-100%) of sequences matched variants found in the mtDNAs of woolly mammoths, a species that became extinct in Eurasia during the Holocene (25), but not in other elephantids. Likewise, sequences attributed to rhinocerotids most often carried variants specific to the woolly rhinoceros branch (54-100% support), thought to have become extinct at the end of the Late Pleistocene (25), and show little support (0-6%) for other rhinoceros lineages. In ~70% of libraries containing hyaenid mtDNA, the sequences matched variants of the extinct cave hyena and/or the spotted hyena which exists today only in Africa (26). Lastly, 90% of ursid mtDNA sequences retrieved from Vindija Cave carried variants matching *Ursus ingressus*, an



Eastern European cave bear lineage which became extinct approximately 25,000 years ago (27, 28).

Extraction and DNA library preparation negative controls contained between 32 and 359 mammalian mtDNA sequences. These sequences do not exhibit damage patterns typical of ancient DNA and they originate from common contaminants (24, 29-31), predominantly human DNA, as well as DNA of bovids, canids and suids (Fig. S34).

### Targeting hominin DNA

Among the samples analyzed, the only site that yielded sequences from putatively deaminated DNA fragments that could be assigned to hominids (or hominins assuming that no other great apes were present at the sites analyzed here) was El Sidrón. This site differs from the others in that no ancient faunal DNA was identified there (Fig. 1), consistent with the almost complete absence of animal remains at the site (32). To test whether animal mtDNA was too abundant at other sites to detect small traces of hominin mtDNA, we repeated the hybridization capture for all DNA libraries using probes targeting exclusively human mtDNA (8). Between 4,915 and 2.8 million DNA fragments were sequenced per library, out of which between 0 and 8,822 were unique hominin sequences passing our filtering scheme (8). Between 10 and 165 hominin mtDNA sequences showing substitutions typical of ancient DNA were obtained from 15 sediment samples from four sites (Data file S4). To generate sufficient data for phylogenetic analyses, we prepared DNA extracts from additional subsamples of 10 of these samples and used automated liquid handling to generate 102 DNA libraries from these as well as the original extracts (Data file S1, Fig. S22). After enriching for human mtDNA and merging all sequences

from a given sediment sample, 9 samples yielded a sufficient number of deaminated hominin mtDNA fragments (between 168 and 13,207) for further analyses (Data file S4).

### Identifying Neandertal and Denisovan mtDNA

We identified “diagnostic” positions in the mtDNA genome that are inferred to have changed on each branch of a phylogenetic tree relating modern humans, Neandertals, Denisovans and a ~430,000-year-old hominin from Sima de los Huesos (8, 33). For eight sediment samples from El Sidrón, Trou Al’Wesse, Chagyrskaya Cave and Denisova Cave, the Neandertal state is shared by 87-98% of sequences overlapping positions diagnostic for Neandertal mtDNA, whereas the modern human, Denisovan and Sima de los Huesos branches are supported by 4-11%, 0-2% and 0-2% of sequences, respectively. In the ninth sample, collected in layer 15 of the East Gallery in Denisova Cave, 84% (16/19) of sequences carry Denisovan-specific variants, compared to 0% (0/10), 5% (1/19) and 0% (0/23) for the modern human, Neandertal and Sima de los Huesos variants, respectively, pointing to a Denisovan origin for these mtDNA fragments (Data file S4, Fig. S40). We note that none of the hominin sequences present in the extraction or library preparation negative controls carry variants specific to the Neandertal, Denisovan or Sima de los Huesos branches (Data file S4).

The average sequence coverage of the mitochondrial genome varied between 0.4- and 44-fold among the nine samples. To be able to reconstruct phylogenetic trees using these sequences, we called a consensus base at positions covered by at least two deaminated fragments and required more than two-thirds of fragments to carry an identical base (34). These relatively permissive parameters were chosen to avoid discarding samples that produced very small numbers of hominin sequences and allowed us to reconstruct between 8% and 99% of the

mtDNA genome (Table S3). Phylogenetic trees relating each of the reconstructed mtDNA genomes to those of modern and ancient individuals (8) (Table S5) show that they all fall within the genetic variation or close to known mtDNA genomes of Neandertals or Denisovans (Figs. 2, S41-S49).

#### Single vs. multiple sources of hominin mtDNA

We next aimed to assess whether mtDNA fragments from more than one individual are present in a given sediment sample. For this purpose, we identified positions in the mitochondrial genome that are covered by at least ten sequences exhibiting evidence of deamination. Three samples have sufficient data for this analysis (Fig. S50). At each of these positions, nearly all sequences from a sample collected in the Main Gallery of Denisova Cave carry the same base, suggesting that the DNA may derive from a single individual. In contrast, sequences from the El Sidrón sample support two different bases at a single position, as is the case for a second sample from Denisova Cave. Thus, at least two mtDNA genomes seem to be present in both these samples (Fig. S51). The fact that the variable position in the latter sample is a known variant among Neandertal mtDNAs supports the conclusion that more than one Neandertal contributed DNA to it (Table S7).

We then developed a maximum-likelihood approach to infer the number of mtDNA components also in low-coverage data (8) (Fig. S52), allowing us to investigate this issue in four additional samples. We detect only one ancient mtDNA type in the sample from Chagyrskaya Cave and in two other samples from Denisova Cave, while a fifth sample from that site contains mtDNA from at least two ancient individuals (Table S9).

### DNA yields from sediments

To assess how much DNA can be recovered from sediment compared to skeletal elements, we counted the number of mtDNA fragments retrieved per milligram of bone (2, 21, 35-38) or sediment originating from the same layers at three archaeological sites. The number of hominin mtDNA fragments retrieved from bone ranges from 28 to 9,142 per milligram, compared to between 34 and 4,490 mammalian mtDNA fragments in sediment (Table S10). Thus, surprisingly large quantities of DNA can survive in cave sediments. We note that most of the ancient taxa we identified are middle- to large-sized (Fig. 1), consistent with larger animals leaving more of their DNA in sediments.

The hominin DNA is present in similar concentrations among subsamples of sediment removed from larger samples (Fig. S53). This suggests that in most cases, the DNA is not concentrated in larger spots but spread relatively evenly within the sediment, which is compatible with it originating from excreta or the decay of soft tissue (9, 39, 40). One exception is a sample from the Main Gallery of Denisova Cave, from which one subsample contains over 500 times more hominin mtDNA fragments than others. As the mtDNA retrieved from it may originate from a single Neandertal (Tables S7, S9), we hypothesize that this is due to an unrecognized small bone or tooth fragment in the subsample. Despite its high content of hominin DNA, the library remains dominated by DNA from other mammals, as only ~7.5% of sequences were attributed to hominins following its enrichment with the mammalian mtDNAs probes. Nonetheless, if such microscopic fragments can be identified and isolated, they may represent a source of hominin DNA sufficiently devoid of other mammalian DNA to allow for analyses of the nuclear genome.

### DNA movement across layers

Post-depositional mixing of particles or a saturation of the sediments by large amounts of DNA can potentially lead to movements of DNA between layers in a stratigraphy (40-42). At the sites investigated here, the overall consistency between the taxa identified from DNA and the archaeological records (Fig. S33) suggests the integrity of the spatial distribution of DNA. In Chagyrskaya Cave for example, we recovered abundant mammalian mtDNA fragments showing degradation patterns typical of ancient DNA in layers rich in osseous and lithic assemblages, while no ancient mammalian DNA was identified in an archaeologically sterile layer underneath (43). Additionally, mtDNA sequences attributed to the woolly mammoth and woolly rhinoceros were identified in Late Pleistocene layers, yet they are absent from the layer which postdates the presumed time of extinction of these taxa (25) (Data file S3, Fig. S24). This implies that little or no movement of mtDNA fragments occurred downwards or upwards in Chagyrskaya Cave. However, as local conditions may affect the extent to which DNA can move in a stratigraphy, these need to be assessed at each archaeological site before the DNA recovered can be linked to a specific layer. This may be best achieved by dense sampling in and around layers of interest.

### Conclusions

We show that mtDNA can be efficiently retrieved from many Late and some Middle Pleistocene cave sediments using hybridization capture (Fig. 1). Encouragingly, this is possible also for samples that were stored at room temperature for several years (8). Sediment samples collected for dating, site formation analyses or the reconstruction of ancient environments at sites where excavations are now completed can thus be used for genetic studies.

The mtDNA genomes reconstructed from sediments of four archaeological sites recapitulate a large part of the mitochondrial diversity of Pleistocene hominins hitherto reconstructed from skeletal remains (Fig. 2). The recovery of Neandertal mtDNA from El Sidrón, Chagyrskaya Cave and Layer 11.4 of the East Gallery of Denisova Cave is in agreement with previous findings of Neandertal remains at those sites and in those layers (21, 32, 44). At Trou Al'Wesse, where we find Neandertal mtDNA, no hominin remains have been found in the Pleistocene layers. However, Late Mousterian artefacts and animal bones with cut-marks support the use of the site by Neandertals (45). In Denisova Cave, we detect Neandertal mtDNA in layers with Middle Paleolithic stone tools in the Main Gallery (46), in which no Neandertal remains have been found. In the East Gallery, we identify Denisovan as well as Neandertal mtDNA lower in the stratigraphy than where skeletal remains of archaic humans have been discovered (Fig. 3), indicating the repeated presence of both groups in the region.

The absence of identifiable ancient DNA in Middle Pleistocene layers in Caune de l'Arago and Chagyrskaya Cave is not surprising given their age (>300 kya). Although compared to other animals, hominins constitute a rare taxon at most sites, we were able to detect Neandertal DNA in the sediments of four of the six sites containing Late Pleistocene layers. For the remaining two sites, Vindija Cave and Les Cottés, only one and four samples, respectively, were available for this study, suggesting that extensive sampling is necessary at each site to ensure that hominin DNA is detected if present. Fortunately, the automation of laboratory procedures to generate DNA libraries and isolate DNA by hybridization capture (8) now makes it possible to undertake large-scale studies of DNA in sediments. This is likely to shed light on the genetic affiliations of the occupants of large numbers of archaeological sites where no human remains are found.

## References and Notes:

1. M. Slatkin, F. Racimo, Ancient DNA and human history. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6380-6387 (2016).
2. J. Krause, Q. Fu, J. M. Good, B. Viola, M. V. Shunkov, A. P. Derevianko, S. Pääbo, The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894-897 (2010).
3. D. Reich, R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. Johnson, T. Maricic, J. M. Good, T. Marques-Bonet, C. Alkan, Q. Fu, S. Mallick, H. Li, M. Meyer, E. E. Eichler, M. Stoneking, M. Richards, S. Talamo, M. V. Shunkov, A. P. Derevianko, J. J. Hublin, J. Kelso, M. Slatkin, S. Pääbo, Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).
4. M. Meyer, M. Kircher, M. T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andres, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226 (2012).
5. M. P. Greaves, M. J. Wilson, The adsorption of nucleic acids by montmorillonite. *Soil Biol. Biochem.* **1**, 317-323 (1969).
6. M. G. Lorenz, W. Wackernagel, Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. *Appl. Environ. Microbiol.* **53**, 2948-2952 (1987).

7. C. Crecchio, G. Stotzky, Binding of DNA on humic acids: Effect on transformation of *Bacillus subtilis* and resistance to DNase. *Soil. Biol. Biochem.* **30**, 1061-1067 (1998).
8. Materials and methods are available as supplementary materials at the Science website.
9. E. Willerslev, A. J. Hansen, J. Binladen, T. B. Brand, M. T. Gilbert, B. Shapiro, M. Bunce, C. Wiuf, D. A. Gilichinsky, A. Cooper, Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**, 791-795 (2003).
10. M. Hofreiter, J. I. Mead, P. Martin, H. N. Poinar, Molecular cloning. *Curr Biol* **13**, R693-695 (2003).
11. E. Willerslev, E. Cappellini, W. Boomsma, R. Nielsen, M. B. Hebsgaard, T. B. Brand, M. Hofreiter, M. Bunce, H. N. Poinar, D. Dahl-Jensen, S. Johnsen, J. P. Steffensen, O. Bennike, J. L. Schwenninger, R. Nathan, S. Armitage, C. J. de Hoog, V. Alfimov, M. Christl, J. Beer, R. Muscheler, J. Barker, M. Sharp, K. E. Penkman, J. Haile, P. Taberlet, M. T. Gilbert, A. Casoli, E. Campani, M. J. Collins, Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science* **317**, 111-114 (2007).
12. J. Haile, D. G. Froese, R. D. Macphee, R. G. Roberts, L. J. Arnold, A. V. Reyes, M. Rasmussen, R. Nielsen, B. W. Brook, S. Robinson, M. Demuro, M. T. Gilbert, K. Munch, J. J. Austin, A. Cooper, I. Barnes, P. Moller, E. Willerslev, Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22352-22357 (2009).
13. D. C. Murray, S. G. Pearson, R. Fullagar, B. M. Chase, J. Houston, J. Atchison, N. E. White, M. I. Bellgard, E. Clarke, M. Macphail, M. T. P. Gilbert, J. Haile, M. Bunce, High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Sci. Rev.* **58**, 135-145 (2012).



14. M. C. Lydolph, J. Jacobsen, P. Arctander, M. T. Gilbert, D. A. Gilichinsky, A. J. Hansen, E. Willerslev, L. Lange, Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Appl. Environ. Microbiol.* **71**, 1012-1017 (2005).
15. R. W. Graham, S. Belmecheri, K. Choy, B. J. Culleton, L. J. Davies, D. Froese, P. D. Heintzman, C. Hritz, J. D. Kapp, L. A. Newsom, R. Rawcliffe, E. Saulnier-Talbot, B. Shapiro, Y. Wang, J. W. Williams, M. J. Wooller, Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9310-9314 (2016).
16. M. W. Pedersen, A. Ruter, C. Schweger, H. Friebe, R. A. Staff, K. K. Kjeldsen, M. L. Mendoza, A. B. Beaudoin, C. Zutter, N. K. Larsen, B. A. Potter, R. Nielsen, R. A. Rainville, L. Orlando, D. J. Meltzer, K. H. Kjaer, E. Willerslev, Postglacial viability and colonization in North America's ice-free corridor. *Nature* **537**, 45-49 (2016).
17. A. W. Briggs, U. Stenzel, P. L. Johnson, R. E. Green, J. Kelso, K. Prüfer, M. Meyer, J. Krause, M. T. Ronan, M. Lachmann, S. Pääbo, Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616-14621 (2007).
18. J. Krause, A. W. Briggs, M. Kircher, T. Maricic, N. Zwyns, A. Derevianko, S. Pääbo, A Complete mtDNA Genome of an Early Modern Human from Kostenki, Russia. *Curr. Biol.* **20**, 231-236 (2010).
19. K. Prüfer, M. Meyer, Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans". *Science* **347**, 835 (2015).
20. C. L. Weiss, M. Dannemann, K. Prufer, H. A. Burbano, Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data. *Elife* **4**, e10005 (2015).

21. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49 (2014).
22. M. T. Gansauge, M. Meyer, Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737-748 (2013).
23. M. T. Gansauge, T. Gerber, I. Glocke, P. Korlevic, L. Lippik, S. Nagel, L. M. Riehl, A. Schmidt, M. Meyer, Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.*, gkx033 (2017).
24. V. Slon, I. Glocke, R. Barkai, A. Gopher, I. Hershkovitz, M. Meyer, 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). *Quatern. Int.* **398**, 210-218 (2016).
25. Y. V. Kuzmin, Extinction of the woolly mammoth (*Mammuthus primigenius*) and woolly rhinoceros (*Coelodonta antiquitatis*) in Eurasia: Review of chronological and environmental issues. *Boreas* **39**, 247-261 (2010).
26. N. Rohland, J. L. Pollack, D. Nagel, C. Beauval, J. Airvaux, S. Paabo, M. Hofreiter, The population history of extant and extinct hyenas. *Mol. Biol. Evol.* **22**, 2435-2443 (2005).

27. S. C. Münzel, M. Stiller, M. Hofreiter, A. Mittnik, N. J. Conard, H. Bocherens, Pleistocene bears in the Swabian Jura (Germany): Genetic replacement, ecological displacement, extinctions and survival. *Quatern. Int.* **245**, 225-237 (2011).
28. P. Wojtal, J. Wilczynski, A. Nadachowski, S. C. Münzel, Gravettian hunting and exploitation of bears in Central Europe. *Quatern. Int.* **359**, 58-71 (2015).
29. S. Pääbo, H. Poinar, D. Serre, V. Jaenicke-Despres, J. Hebler, N. Rohland, M. Kuch, J. Krause, L. Vigilant, M. Hofreiter, Genetic analyses from ancient DNA. *Annu. Rev. Genet.* **38**, 645-679 (2004).
30. J. A. Leonard, O. Shanks, M. Hofreiter, E. Kreuz, L. Hodges, W. Ream, R. K. Wayne, R. C. Fleischer, Animal DNA in PCR reagents plagues ancient DNA research. *J. Archaeol. Sci.* **34**, 1361-1366 (2007).
31. M. T. P. Gilbert, H. J. Bandelt, M. Hofreiter, I. Barnes, Assessing ancient DNA studies. *Trends Ecol. Evol.* **20**, 541-544 (2005).
32. A. Rosas, C. Martinez-Maza, M. Bastir, A. Garcia-Tabernerero, C. Lalueza-Fox, R. Huguet, J. E. Ortiz, R. Julia, V. Soler, T. de Torres, E. Martinez, J. C. Canaveras, S. Sanchez-Moral, S. Cuezva, J. Lario, D. Santamaria, M. de la Rasilla, J. Fortea, Paleobiology and comparative morphology of a late Neandertal sample from El Sidrón, Asturias, Spain. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19266-19271 (2006).
33. M. Meyer, J. L. Arsuaga, S. Nagel, A. Aximu-Petri, I. Martinez, A. Gracia, J. M. Bermúdez de Castro, E. Carbonell, J. Kelso, K. Prüfer, S. Pääbo, Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016).

34. M. Meyer, Q. Fu, A. Aximu-Petri, I. Glocke, B. Nickel, J. L. Arsuaga, I. Martinez, A. Gracia, J. M. Bermúdez de Castro, E. Carbonell, S. Pääbo, A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406 (2014).
35. S. Brown, T. F. Higham, V. Slon, S. Pääbo, M. Meyer, K. Douka, F. Brock, D. Comeskey, N. Procopio, M. V. Shunkov, A. Derevianko, M. Buckley, Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **6**, 23559 (2016).
36. R. E. Green, A. S. Malaspinas, J. Krause, A. W. Briggs, P. L. Johnson, C. Uhler, M. Meyer, J. M. Good, T. Maricic, U. Stenzel, K. Prufer, M. Siebauer, H. A. Burbano, M. Ronan, J. M. Rothberg, M. Egholm, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, M. Wikstrom, L. Laakkonen, J. Kelso, M. Slatkin, S. Pääbo, A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426 (2008).
37. M. T. Gansauge, M. Meyer, Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res.* **24**, 1543-1549 (2014).
38. A. W. Briggs, J. M. Good, R. E. Green, J. Krause, T. Maricic, U. Stenzel, C. Lalueza-Fox, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, R. Schmitz, V. B. Doronichev, L. V. Golovanova, M. de la Rasilla, J. Fortea, A. Rosas, S. Pääbo, Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321 (2009).
39. H. N. Poinar, M. Hofreiter, W. G. Spaulding, P. S. Martin, B. A. Stankiewicz, H. Bland, R. P. Evershed, G. Possnert, S. Paabo, Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**, 402-406 (1998).

40. J. Haile, R. Holdaway, K. Oliver, M. Bunce, M. T. Gilbert, R. Nielsen, K. Munch, S. Y. Ho, B. Shapiro, E. Willerslev, Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Mol. Biol. Evol.* **24**, 982-989 (2007).
41. L. J. Arnold, R. G. Roberts, R. D. E. Macphee, J. S. Haile, F. Brock, P. Moller, D. G. Froese, A. N. Tikhonov, A. R. Chivas, M. T. P. Gilbert, E. Willerslev, Paper II - Dirt, dates and DNA: OSL and radiocarbon chronologies of perennially frozen sediments in Siberia, and their implications for sedimentary ancient DNA studies. *Boreas* **40**, 417-445 (2011).
42. K. Andersen, K. L. Bird, M. Rasmussen, J. Haile, H. Breuning-Madsen, K. H. Kjaer, L. Orlando, M. T. Gilbert, E. Willerslev, Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* **21**, 1966-1979 (2012).
43. N. Rudaya, S. Vasiliev, B. Viola, S. Talamo, S. Markin, Palaeoenvironments during the period of the Neanderthals settlement in Chagyrskaya cave (Altai Mountains, Russia). *Palaeogeography, Palaeoclimatology, Palaeoecology* **467**, 265-276 (2017).
44. B. T. Viola, S. V. Markin, A. P. Buzhilova, M. B. Mednikova, M. V. Dobrovolskaya, A. Le Cabec, M. V. Shunkov, A. P. Dereviank, J. J. Hublin, New Neanderthal remains from Chagyrskaya Cave (Altai Mountains, Russian Federation). *Am. J. Phys. Anthropol.* **147**, 293-294 (2012).
45. R. Miller, J. S. Stewart, M. Otte, Résultats préliminaires de l'étude de la séquence paléolithique au Trou Al'Wesse (comm. de Modave). *Notae Praehistoricae* **27**, 41-49 (2007).

46. A. P. Derevianko, *Recent Discoveries in the Altai: Issues on the Evolution of Homo Sapiens* (Institute of Archaeology and Ethnography SB RAS Press, Novosibirsk, 2012).
47. A. M. Moigne, M. R. Palombo, V. Belda, D. Heriech-Briki, S. Kacimi, F. Lacomat, M. A. d. Lumley, J. Moutoussamy, F. Rivals, J. Quilès, A. Testu, Les faunes de grands mammifères de la Caune de l'Arago (Tautavel) dans le cadre biochronologique des faunes du Pléistocène moyen italien. *L'anthropologie* **110**, 788-831 (2006).
48. L. Byrne, Lithic tools from Arago cave, Tautavel (Pyrenees-Orientales, France): behavioural continuity or raw material determinism during the Middle Pleistocene? *J. Archaeol. Sci.* **31**, 351-364 (2004).
49. H. d. Lumley, *Caune de l'Arago, Tautavel-en-Roussillon Pyrénées-Orientales, France, Tome I* (CNRS éditions, Paris, 2014).
50. M. A. d. Lumley, L'homme de Tautavel. Un Homo erectus européen évolué. Homo erectus tautavelensis *L'Anthropologie* **119**, 303-348 (2015).
51. D. Barsky, Les industries lithiques de la Caune de l'Arago dans leur contexte stratigraphique. *Comptes Rendus Palevol.* **12**, 305-325 (2013).
52. A. M. Moigne, D. R. Barsky, "Large mammal assemblages from Lower Palaeolithic sites in France: La Caune de l'Arago, Terra-Amata, Orgnac 3 and Cagny L'Épinette" in *The role of early humans in the accumulation of European Lower and Middle Paleolithic bone assemblages* (Monographien des Römisch-Germanischen Zentralmuseums, Mainz, 1999).
53. L. Lebreton, E. Desclaux, C. Hanquet, A. M. Moigne, C. Perrenoud, Environmental context of the Caune de l'Arago Acheulean occupations (Tautavel, France), new insights from microvertebrates in Q–R levels. *Quatern. Int.* **411**, 182–192 (2016).

54. C. Falgueres, Q. Shao, F. Han, J. J. Bahain, M. Richard, C. Perrenoud, A. M. Moigne, D. Lumley, New ESR and U-series dating at Caune de l'Arago, France: A key-site for European Middle Pleistocene. *Quat. Geochronol.* **30**, 547-553 (2015).
55. A. P. Derevianko, M. V. Shunkov, S. V. Markin, *The Dynamics of the Paleolithic Industries in Africa and Eurasia in the Late Pleistocene and the Issue of the Homo sapiens Origin* (Institute of Archaeology and Ethnography SB RAS Press, Novosibirsk, 2014).
56. A. P. Derevianko, S. V. Markin, V. S. Zykin, V. S. Zykina, V. S. Zazhigin, A. O. Sizikova, E. P. Solotchina, L. G. Smolyaninova, A. S. Antipov, Chagyrskaya Cave: A Middle Paleolithic Site In The Altai. *Archaeology, Ethnology and Anthropology of Eurasia* **41**, 2-27 (2013).
57. S. K. Vasiliev, Large Mammal Fauna from the Pleistocene Deposits of Chagyrskaya Cave Northwestern Altai (based on 2007–2011 Excavations). *Archaeology, Ethnology and Anthropology of Eurasia* **41**, 28-44 (2013).
58. M. B. Mednikova, An archaic human ulna from Chagyrskaya Cave, Altai: morphology and taxonomy. *Archaeology Ethnology and Anthropology of Eurasia* **41**, 66-77 (2013).
59. B. Viola, S. V. Markin, A. Zenin, M. V. Shunkov, A. P. Derevianko, "Late Pleistocene hominins from the Altai Mountains, Russia" in *Characteristic features of the Middle to Upper Paleolithic Transition in Eurasia*, A. P. Derevianko, M. V. Shunkov, Eds. (Institute of Archaeology and Ethnography SB RAS, Novosibirsk, 2011), pp. 207-213.
60. A. Derevianko, S. Markin, S. Gladyshev, K. Kolobova, Excavations at the Chagyrskaya Cave, Russia: a Neanderthal Middle Palaeolithic industry in Northern Asia. *Antiquity* **345**, (2015).

61. A. Derevianko, *The Paleolithic of Siberia: New Discoveries and Interpretations* (University of Illinois Press, Illinois, 1998).
62. A. P. Derevianko, M. V. Shunkov, P. V. Volkov, A paleolithic bracelet from Denisova Cave. *Archaeology, Ethnology and Anthropology of Eurasia* **34**, 13-25 (2008).
63. M. B. Mednikova, A Proximal Pedal Phalanx of a Paleolithic Hominin from Denisova Cave, Altai. *Archaeology, Ethnology and Anthropology of Eurasia* **39**, 129-138 (2011).
64. S. Sawyer, G. Renaud, B. Viola, J. J. Hublin, M. T. Gansauge, M. V. Shunkov, A. Derevianko, K. Prüfer, J. Kelso, S. Pääbo, Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15696–15700 (2015).
65. C. G. Turner, “Paleolithic teeth of the Central Siberian Altai Mountains” in *Chronostratigraphy of the Paleolithic in North Central, East Asia and America*, A. P. Derevianko, Ed. (Institute of History, Philology and Philosophy, Siberian Branch of the USSR Academy of Sciences, Novosibirsk, 1990), pp. 239-243.
66. E. G. Shpakova, A. P. Derevianko, The interpretation of odontological features of Pleistocene human remains from the Altai. *Archaeology, Ethnology and Anthropology of Eurasia* **1**, 125-138 (2000).
67. S. K. Vasiliev, M. V. Shunkov, M. B. Kozlikin, Preliminary Results for the Balance of Megafauna from Pleistocene Layers of the East Gallery, Denisova Cave. *Problemy archeologii, etnografii, antropologii Sibiri i sopredelnykh territorij* **19**, 32-38 (2013).
68. G. Baryshnikov, Large mammals and Neanderthal paleoecology in the Altai mountains (Central Asia, Russia). *Préhistoire Européenne* **14**, 49-66 (1999).
69. A. K. Agadjanian, The dynamics of bioresources and activity of the paleolithic man, using the example of northwestern Altai Mountains. *Paleontol. J.* **40**, S482-S493 (2006).



70. N. V. Vorobieva, D. Y. Sherbakov, A. S. Druzhkova, R. Stanyon, A. A. Tsybankov, S. K. Vasil'ev, M. V. Shunkov, V. A. Trifonov, A. S. Graphodatsky, Genotyping of *Capreolus pygargus* Fossil DNA from Denisova Cave Reveals Phylogenetic Relationships between Ancient and Modern Populations. *Plos One* **6**, e24045 (2011).
71. A. A. Lissovsky, N. V. Serdyuk, Identification of Late Pleistocene pikas (*Ochotona*, Lagomorpha, Mammalia) of the alpina-hyperborea group from Denisova Cave (Altai) on the basis of the anterior lower Premolar(P-3). *Paleontol. J.* **38**, 680-686 (2004).
72. A. P. Derevianko, S. A. Laukhin, O. A. Kulikov, Z. N. Gnibidenko, M. V. Shunkov, First Middle Pleistocene age determinations of the Paleolithic in the Altai Mountains. *Dokl. Akad. Nauk.* **326**, 497–501 (1992).
73. V. K. Vlasov, O. A. Kulikov, Radiothermoluminescence dating and applications to Pleistocene sediments. *Phys. Chem. Miner.* **16**, 551–558 (1989).
74. M. Roussel, M. Sorresi, “Une nouvelle séquence du Paléolithique supérieur ancien aux marges sud-ouest du Bassin parisien : les Cottés dans la Vienne” in *Le Paléolithique supérieur ancien de l'Europe du Nord-ouest*, P. Bodu, L. Chehmana, L. Klaric, L. Mevel, S. Soriano, N. Teyssandier, Eds. (Mémoire de la Société préhistorique française, Paris, 2013), pp. 283-298.
75. M. Sorresi, M. Roussel, W. Rendu, J. Primault, S. Rigaud, J. P. Texier, D. Richter, S. Talamo, F. Pioquin, B. Larmignat, C. Tavormina, J. J. Hublin, “Les Cottés : nouveaux travaux sur l'un des gisements de référence pour la transition Paléolithique moyen-supérieur” in *Préhistoire entre Vienne et Charente : hommes et sociétés du Paléolithique*, J. Buisson-Catil, J. Primault, Eds. (Association des publications chauvinoises, Chauvigny, 2010), pp. 221-234.

76. M. Frouin, F. Ploquin, M. Soressi, W. Rendu, R. Macchiarelli, A. El Albani, A. Meunier, Clay minerals of late Pleistocene sites (Jonzac and Les Cottés, SW France): Applications of X-ray diffraction analyses to local paleoclimatic and paleoenvironmental reconstructions. *Quatern. Int.* **302**, 184-198 (2013).
77. E. Patte, Le crâne Aurignacien des Cottés. *L'anthropologie* **58**, 450-471 (1954).
78. F. Welker, M. Soressi, W. Rendu, J. J. Hublin, M. Collins, Using ZooMS to identify fragmentary bone from the Late Middle/Early Upper Palaeolithic sequence of Les Cottés, France. *J. Archaeol. Sci.* **54**, 279-286 (2015).
79. S. Talamo, M. Soressi, M. Roussel, M. Richards, J. J. Hublin, A radiocarbon chronology for the complete Middle to Upper Palaeolithic transitional sequence of Les Cottés (France). *J. Archaeol. Sci.* **39**, 175-183 (2012).
80. Z. Jacobs, B. Li, N. Jankowski, M. Soressi, Testing of a single grain OSL chronology across the Middle to Upper Palaeolithic transition at Les Cottés (France). *J. Archaeol. Sci.* **54**, 110-122 (2015).
81. R. Miller, N. Zwyns, M. Otte, Le site du Trou Al'Wesse (comm. de Modave): Campagne de Fouille 2004. *Notae Praehistoricae* **24**, 109-116 (2004).
82. R. Miller, J. Stewart, M. Knul, Y. Waersegers, P. Noiret, K. Wilkinson, The Middle to Upper Paleolithic transition at Trou Al'Wesse: A preliminary overview of stratigraphic units 17 to 15. *Notae Praehistoricae* **35**, 25-34 (2015).
83. R. Miller, M. Otte, J. Stewart, Nouvelles découvertes de la séquence holocène du Trou Al'Wesse: fouilles 2010. *Notae Praehistoricae* **30**, 35-42 (2010).
84. R. Miller, J. Stewart, N. Zwyns, M. Otte, "The stratified Early to Late Mesolithic sequence at Trou Al'Wesse (Modave, Belgium)" in *Chronology and Evolution within the*

- Mesolithic of North-West Europe: Proceedings of an International Meeting, Brussels, May 30th-June 1st, 2007*, P. Crombé, M. Van Strydonck, J. Sergant, M. Boudin, M. Bats, Eds. (Cambridge Scholars Publishing, England, 2009).
85. F. Collin, P. Masy, M. Tinant, La Grotte du Trou Al'Wesse (Province de Liège): Fouilles et découvertes de 1993. *Notae Praehistoricae* **13**, 21-25 (1993).
  86. P. Masy, La sépulture collective néolithique du Trou Al'Wesse à Modave (province de Liège). *Bulletin des Chercheurs de la Wallonie* **33**, 81-99 (1993).
  87. R. Miller, N. Zwyns, J. Stewart, M. Toussaint, M. Otte, Trou Al'Wesse : Campagne de fouilles 2006. *Notae Praehistoricae* **26**, 103-108 (2006).
  88. V. K. Lagerholm, E. Sandoval-Castellanos, A. Vaniscotte, O. Potapova, T. Tomek, Z. M. Bocheński, P. J. Shepherd, N. Barton, M.-C. Van Dyck, R. Miller, J. Höglund, N. G. Yoccoz, L. Dalén, J. R. Stewart, Range shifts or extinction? Ancient DNA and distribution modelling reveal past and future responses to climate warming in cold-adapted birds. *Global Change Biol.* **23**, 1425-1435 (2016).
  89. S. Brace, M. Ruddy, R. Miller, D. C. Schreve, J. R. Stewart, I. Barnes, The colonization history of British water vole (*Arvicola amphibius* (Linnaeus, 1758)): origins and development of the Celtic fringe. *Proc. R. Soc. B.* **283**, 20160130 (2016).
  90. S. Brace, E. Palkopoulou, L. Dalen, A. M. Lister, R. Miller, M. Otte, M. Germonpre, S. P. E. Blockley, J. R. Stewart, I. Barnes, Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20532-20536 (2012).
  91. M. Meiri, A. M. Lister, T. F. G. Higham, J. R. Stewart, L. G. Straus, H. Obermaier, M. R. G. Morales, A. B. Marin-Arroyo, I. Barnes, Late-glacial recolonization and

- phylogeography of European red deer (*Cervus elaphus* L.). *Mol. Ecol.* **22**, 4711-4722 (2013).
92. M. Otte, F. Collin, R. Miller, K. Engesser, Nouvelles datations du Trou Al'Wesse dans son contexte régional. *Notae Praehistoricae* **18**, 45-50 (1998).
93. J. C. M. Ahern, I. Karavanic, M. Paunovic, I. Jankovic, F. H. Smith, New discoveries and interpretations of hominid fossils and artifacts from Vindija Cave, Croatia. *J. Hum. Evol.* **46**, 27-67 (2004).
94. I. Karavanic, F. H. Smith, The middle/upper Paleolithic interface and the relationship of Neanderthals and early modern humans in the Hrvatsko Zagorje, Croatia. *J. Hum. Evol.* **34**, 223-248 (1998).
95. I. Jankovic, I. Karavanic, J. C. M. Ahern, D. Brajkovic, J. M. Lenardic, F. H. Smith, Vindija cave and the modern human peopling of Europe. *Collegium Antropol.* **30**, 457-466 (2006).
96. M. Malez, F. H. Smith, J. Radovic, D. Rukavina, Upper Pleistocene Hominids from Vindija, Croatia, Yugoslavia. *Curr. Anthropol.* **21**, 365-367 (1980).
97. F. H. Smith, D. C. Boyd, M. Malez, Additional Upper Pleistocene Hominid Remains from Vindija Cave. *Am. J. Phys. Anthropol.* **66**, 230-230 (1985).
98. M. H. Wolpoff, F. H. Smith, M. Malez, J. Radovic, D. Rukavina, Upper Pleistocene Human Remains from Vindija Cave, Croatia, Yugoslavia. *Am. J. Phys. Anthropol.* **54**, 499-545 (1981).
99. F. H. Smith, J. C. Ahern, Brief Communication - Additional Cranial Remains from Vindija-Cave, Croatia. *Am. J. Phys. Anthropol.* **93**, 275-280 (1994).

100. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).
101. E. M. Wild, M. Paunovic, G. Rabeder, I. Steffan, P. Steier, Age determination of fossil bones from the Vindija Neanderthal site in Croatia. *Radiocarbon* **43**, 1021-1028 (2001).
102. P. T. Miracle, J. M. Lenardic, D. Brajkovic, Last glacial climates, "Refugia", and faunal change in Southeastern Europe: Mammalian assemblages from Veternica, Velika pecina, and Vindija caves (Croatia). *Quatern. Int.* **212**, 137-148 (2010).
103. M. Hofreiter, C. Capelli, M. Krings, L. Waits, N. Conard, S. Munzel, G. Rabeder, D. Nagel, M. Paunovic, G. Jambresic, S. Meyer, G. Weiss, S. Paabo, Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears. *Mol. Biol. Evol.* **19**, 1244-1250 (2002).
104. M. Stiller, M. Molak, S. Prost, G. Rabeder, G. Baryshnikov, W. Rosendahl, S. Munzel, H. Bocherens, A. Grandal-d'Anglade, B. Hilpert, M. Germonpre, O. Stasyk, R. Pinhasi, A. Tintori, N. Rohland, E. Mohandesan, S. Y. W. Ho, M. Hofreiter, M. Knapp,

- Mitochondrial DNA diversity and evolution of the Pleistocene cave bear complex. *Quatern. Int.* **339**, 224-231 (2014).
105. D. Brajković, P. T. Miracle, “Middle Palaeolithic and Early Upper Palaeolithic subsistence practices at Vindija Cave, Croatia” in *The Palaeolithic of the Balkans*, A. Darlas, D. Mihailović, Eds. (Archaeopress, Oxford, 2008), pp. 107-116.
  106. K. Zaremba-Niedzwiedzka, S. G. E. Andersson, No Ancient DNA Damage in Actinobacteria from the Neanderthal Bone. *Plos One* **8**, e62799 (2013).
  107. M. Krings, C. Capelli, F. Tschentscher, H. Geisert, S. Meyer, A. von Haeseler, K. Grossschmidt, G. Possnert, M. Paunovic, S. Paabo, A view of Neandertal genetic diversity. *Nature Genet.* **26**, 144-146 (2000).
  108. J. Fortea, M. de la Rasilla, E. Martínez, S. Sánchez-Moral, J. C. Cañaveras, S. Cuezva, A. Rosas, V. Soler, R. Julia, T. de Torres, J. E. Ortiz, J. Castro, E. Badal, J. Altuna, J. Alonso, La cueva de El Sidrón (Borines, Piloña, Asturias): primeros resultados. *Estudios Geológicos* **59**, 159-179 (2003).
  109. D. Santamaría, J. Fortea, M. de La Rasilla, L. Martínez, E. Martínez, J. C. Cañaveras, S. Sanchez-Moral, A. Rosas, A. Estalrich, A. Garcia-Taberner, C. Lalueza-Fox, The technological and typological behaviour of a Neanderthal group from El Sidrón cave (Asturias, Spain). *Oxford Journal of Archaeology* **29**, 119-148 (2010).
  110. C. Lalueza-Fox, A. Rosas, A. Estalrich, E. Gigli, P. F. Campos, A. Garcia-Taberner, S. Garcia-Vargas, F. Sanchez-Quinto, O. Ramirez, S. Civit, M. Bastir, R. Huguet, D. Santamaria, M. T. P. Gilbert, E. Willerslev, M. de la Rasilla, Genetic evidence for patrilocal mating behavior among Neandertal groups. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 250-253 (2011).

111. A. Rosas, A. Estalrich, S. Garcia-Vargas, A. Garcia-Tabernerero, R. Huguet, C. Lalueza-Fox, M. de la Rasilla, Identification of Neandertal individuals in fragmentary fossil assemblages by means of tooth associations: The case of El Sidrón (Asturias, Spain). *Comptes Rendus Palevol.* **12**, 279-291 (2013).
112. S. Castellano, G. Parra, F. A. Sanchez-Quinto, F. Racimo, M. Kuhlwilm, M. Kircher, S. Sawyer, Q. Fu, A. Heinze, B. Nickel, J. Dabney, M. Siebauer, L. White, H. A. Burbano, G. Renaud, U. Stenzel, C. Lalueza-Fox, M. de la Rasilla, A. Rosas, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Meyer, J. Kelso, A. M. Andres, S. Pääbo, Patterns of coding variation in the complete exomes of three Neandertals. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6666-6671 (2014).
113. M. Kuhlwilm, I. Gronau, M. J. Hubisz, C. de Filippo, J. Prado-Martinez, M. Kircher, Q. Fu, H. A. Burbano, C. Lalueza-Fox, M. de la Rasilla, A. Rosas, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, T. Marques-Bonet, A. M. Andres, B. Viola, S. Pääbo, M. Meyer, A. Siepel, S. Castellano, Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530**, 429-433 (2016).
114. F. L. Mendez, G. D. Poznik, S. Castellano, C. D. Bustamante, The Divergence of Neandertal and Modern Human Y Chromosomes. *Am. J. Hum. Genet.* **98**, 728-734 (2016).
115. C. Lalueza-Fox, M. L. Sampietro, D. Caramelli, Y. Puder, M. Lari, F. Calafell, C. Martinez-Maza, M. Bastir, J. Fortea, M. de la Rasilla, J. Bertranpetit, A. Rosas, Neandertal evolutionary genetics: Mitochondrial DNA data from the Iberian Peninsula. *Mol. Biol. Evol.* **22**, 1077-1081 (2005).

116. T. de Torres, J. E. Ortiz, R. Grun, S. Eggins, H. Valladas, N. Mercier, N. Tisnerat-Laborde, R. Julia, V. Soler, E. Martinez, S. Sanchez-Moral, J. C. Canaveras, J. Lario, E. Badal, C. Lalueza-Fox, A. Rosas, D. Santamaria, M. de la Rasilla, J. Fortea, Dating of the hominid (*Homo neanderthalensis*) remains accumulation from El Sidrón Cave (Piloña, Asturias, North Spain): an example of a multi-methodological approach to the dating of Upper Pleistocene sites. *Archaeometry* **52**, 680-705 (2010).
117. R. E. Wood, T. F. G. Higham, T. de Torres, N. Tisnerat-Laborde, H. Valladas, J. E. Ortiz, C. Lalueza-Fox, S. Sanchez-Moral, J. C. Canaveras, A. Rosas, D. Santamaria, M. de la Rasilla, A new date for the Neanderthals from El Sidrón (Asturias, Northern Spain). *Archaeometry* **55**, 148-158 (2013).
118. P. Cai, Q. Huang, X. Zhang, H. Chen, Adsorption of DNA on clay minerals and various colloidal particles from an Alfisol. *Soil Biol. Biochem.* **38**, 471-476 (2006).
119. A. Ogram, G. S. Sayler, D. Gustin, R. J. Lewis, DNA Adsorption to Soils and Sediments. *Environ. Sci. Technol.* **22**, 982-984 (1988).
120. A. V. Ogram, M. L. Mathot, J. B. Harsh, J. Boyle, C. A. Pettigrew, Effects of DNA Polymer Length on Its Adsorption to Soils. *Appl. Environ. Microb.* **60**, 393-396 (1994).
121. R. E. Green, A. W. Briggs, J. Krause, K. Prufer, H. A. Burbano, M. Siebauer, M. Lachmann, S. Paabo, The Neandertal genome and ancient DNA authenticity. *EMBO J.* **28**, 2494-2502 (2009).
122. M. E. Allentoft, M. Collins, D. Harker, J. Haile, C. L. Oskam, M. L. Hale, P. F. Campos, J. A. Samaniego, M. T. Gilbert, E. Willerslev, G. Zhang, R. P. Scofield, R. N. Holdaway, M. Bunce, The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Biol. Sci.* **279**, 4724-4733 (2012).



123. J. Dabney, M. Knapp, I. Glocke, M. T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, S. Pääbo, J. L. Arsuaga, M. Meyer, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15758-15763 (2013).
124. C. Giguët-Covex, J. Pansu, F. Arnaud, P. J. Rey, C. Griggo, L. Gielly, I. Domaizon, E. Coissac, F. David, P. Choler, J. Poulénard, P. Taberlet, Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* **5**, 3211 (2014).
125. S. A. Bulat, M. Lubeck, I. A. Alekhina, D. F. Jensen, I. M. Knudsen, P. S. Lubeck, Identification of a universally primed-PCR-derived sequence-characterized amplified region marker for an antagonistic strain of *Clonostachys rosea* and development of a strain-specific PCR detection assay. *Appl. Environ. Microbiol.* **66**, 4758-4763 (2000).
126. A. C. Boere, W. I. Rijpstra, G. J. De Lange, J. S. Sinninghe Damste, M. J. Coolen, Preservation potential of ancient plankton DNA in Pleistocene marine sediments. *Geobiology* **9**, 377-393 (2011).
127. P. Korlevic, T. Gerber, M. T. Gansauge, M. Hajdinjak, S. Nagel, A. Aximu-Petri, M. Meyer, Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87-93 (2015).
128. J. Dabney, M. Meyer, 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).
129. M. Kircher, S. Sawyer, M. Meyer, Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).

130. M. M. Deangelis, D. G. Wang, T. L. Hawkins, Solid-Phase Reversible Immobilization for the Isolation of Pcr Products. *Nucleic Acids Res.* **23**, 4742-4743 (1995).
131. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb prot5448 (2010).
132. G. Renaud, M. Kircher, U. Stenzel, J. Kelso, freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208-1209 (2013).
133. G. Renaud, U. Stenzel, J. Kelso, leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
134. A. Herbig, F. Maixner, K. I. Bos, A. Zink, J. Krause, D. H. Huson, MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. *bioRxiv* (2016).
135. D. H. Huson, A. F. Auch, J. Qi, S. C. Schuster, MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377-386 (2007).
136. Q. Fu, M. Meyer, X. Gao, U. Stenzel, H. A. Burbano, J. Kelso, S. Pääbo, DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2223-2227 (2013).
137. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939-946 (2012).
138. T. Maricic, M. Whitten, S. Pääbo, Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *Plos One* **5**, e14004 (2010).
139. T. Lindahl, Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715 (1993).

140. M. Hofreiter, V. Jaenicke, D. Serre, A. von Haeseler, S. Paabo, DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793-4799 (2001).
141. S. Sawyer, J. Krause, K. Guschanski, V. Savolainen, S. Pääbo, Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131 (2012).
142. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. T. Gilbert, Y. Wang, M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D. Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K. Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Gronnow, M. Meldgaard, C. Andreasen, S. A. Fedorova, L. P. Osipova, T. F. Higham, C. B. Ramsey, T. V. Hansen, F. C. Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Ponten, R. Villems, R. Nielsen, A. Krogh, J. Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762 (2010).
143. J. P. Noonan, M. Hofreiter, D. Smith, J. R. Priest, N. Rohland, G. Rabeder, J. Krause, J. C. Detter, S. Paabo, E. M. Rubin, Genomic sequencing of Pleistocene cave bears. *Science* **309**, 597-600 (2005).
144. M. L. Carpenter, J. D. Buenrostro, C. Valdiosera, H. Schroeder, M. E. Allentoft, M. Sikora, M. Rasmussen, S. Gravel, S. Guillen, G. Nekhrizov, K. Leshtakov, D. Dimitrova, N. Theodossiev, D. Pettener, D. Luiselli, K. Sandoval, A. Moreno-Estrada, Y. R. Li, J. Wang, M. T. P. Gilbert, E. Willerslev, W. J. Greenleaf, C. D. Bustamante, Pulling out the

- 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *Am. J. Hum. Genet.* **93**, 852-864 (2013).
145. C. Gamba, E. R. Jones, M. D. Teasdale, R. L. McLaughlin, G. Gonzalez-Fortes, V. Mattiangeli, L. Domboroczki, I. Kovari, I. Pap, A. Anders, A. Whittle, J. Dani, P. Raczky, T. F. Higham, M. Hofreiter, D. G. Bradley, R. Pinhasi, Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
146. K. D. Pruitt, G. R. Brown, S. M. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, C. M. Farrell, J. Hart, M. J. Landrum, K. M. McGarvey, M. R. Murphy, N. A. O'Leary, S. Pujar, B. Rajput, S. H. Rangwala, L. D. Riddick, A. Shkeda, H. Sun, P. Tamez, R. E. Tully, C. Wallin, D. Webb, J. Weber, W. Wu, M. DiCuccio, P. Kitts, D. R. Maglott, T. D. Murphy, J. M. Ostell, RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* **42**, D756-763 (2014).
147. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
148. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
149. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, S. Genome Project Data Processing, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
150. R. C. Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2016).
151. D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, E. W. Sayers, GenBank. *Nucleic Acids Res.* **41**, D36-42 (2013).

152. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).
153. E. Willerslev, M. T. Gilbert, J. Binladen, S. Y. Ho, P. F. Campos, A. Ratan, L. P. Tomsho, R. R. da Fonseca, A. Sher, T. V. Kuznetsova, M. Nowak-Kemp, T. L. Roth, W. Miller, S. C. Schuster, Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evol. Biol.* **9**, 95 (2009).
154. C. J. Edwards, D. A. Magee, S. D. Park, P. A. McGettigan, A. J. Lohan, A. Murphy, E. K. Finlay, B. Shapiro, A. T. Chamberlain, M. B. Richards, D. G. Bradley, B. J. Loftus, D. E. MacHugh, A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*). *PLoS One* **5**, e9255 (2010).
155. A. K. Agadjanian, "Problems of reconstruction of paleoenvironment and conditions of the habitability of the Ancient Man by the example of Northwestern Altai" in *Biosphere Origin and Evolution*, N. Dobretsov, N. Kolchanov, A. Rozanov, G. Zavarzin, Eds. (Springer, 2008).
156. S. G. Turner, N. D. Ovodov, O. V. Pavlova, *Animal Teeth and Human Tools: A Taphonomic Odyssey in Ice Age Siberia* (Cambridge University Press, Cambridge, 2013).
157. W. Rendu, S. Renou, in *Les Cottés (Vienne). Rapport de fouille programmée*, M. Soressi, Ed. (Report deposited at the Service régional de l'archéologie de Poitou-Charentes, 2015), pp. 101-131.
158. K. Tamura, G. Stecher, D. Peterson, A. Filipowski, S. Kumar, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729 (2013).

159. A. Immel, D. G. Drucker, M. Bonazzi, T. K. Jahnke, S. C. Munzel, V. J. Schuenemann, A. Herbig, C. J. Kind, J. Krause, Mitochondrial Genomes of Giant Deers Suggest their Late Survival in Central Europe. *Sci. Rep.* **5**, 10853 (2015).

### **Acknowledgments:**

This study has received funding from the Max Planck Society; the Max-Planck-Förderstiftung (grant P.S.EVANLOMP to S.P.); the European Research Council (ERC) (grant AMD-694707-3 to S.P.); the Australian Research Council (fellowships FT150100138 to Z.J., FT140100384 to B.L. and FL130100116 to R.G.R.); and the Russian Science Foundation (project No. 14-50-00036 to A.P.D. and M.V.S.). The authors declare no conflicts of interest. We wish to thank M.-T. Gansauge, B. Höber and A. Weihmann for laboratory work; A. Hübner, M. Petr, U. Stenzel and B. Vernot for their input on data analysis; V. Aldeias and S. Gur-Arieh for advice on the preliminary study; P. Korlević and C. Piot for graphics; L. Jauregui for comments on the text; V. Wiebe for translations; and T. Devière, E. Flannery, T. Higham, J. Krause, T. Marčić, M. Morley, C. Stringer and B. Viola for their roles in sample acquisition. V.S., C.H., A.A.-P., E.E., S.N., B.N. and A.S. performed the laboratory work; V.S., C.L.W., F.M., K.P., J.K., H.A.B., S.P. and M.M. analyzed the data; M.d.I.R., C.L.-F., A.R., M.S., M.V.K., R.M., J.R.S., A.P.D, Z.J., B.L., R.G.R., M.V.S., H.d.L., C.P., I.G., Ž.K and P.R. provided samples and site-specific expertise; V.S., S.P. and M.M. designed the study and wrote the manuscript with input from all authors. Sequencing data generated for this study are available in the European Nucleotide Archive (ENA) under accession number PRJEB18629. The multiple sequence alignment files used for phylogenetic reconstructions were deposited in the Dryad Digital Repository (<http://dx.doi.org/10.5061/dryad.m2dk7>).

**Fig. 1. Ancient taxa detected in Late Pleistocene (LP) and Middle Pleistocene (MP) sediment samples from seven sites.** For each time period, the fraction of samples containing DNA fragments which could be assigned to a mammalian family and authenticated to be of ancient origin is indicated. The shaded symbols representing each family are not to scale.

**Fig. 2. Cladogram relating mtDNA genomes reconstructed from sediment samples to those of modern and ancient individuals.** The branches leading to mtDNA genomes reconstructed from sediments (dashed lines) were superimposed on a neighbor-joining tree relating the previously determined mtDNA genomes of ancient and present-day humans (purple), Neandertals (orange), Denisovans (green) and the Sima de los Huesos hominin (blue) (Table S5). Discrete phylogenetic trees relating each of the mtDNAs reconstructed here and the comparative data are shown in Figs. S41-S49.

**Fig. 3. Hominin mtDNAs along the stratigraphy of the East Gallery in Denisova Cave.** Layer numbers are noted in gray. The layers of origin for sediment samples and skeletal remains yielding Neandertal (orange) and Denisovan (green) mtDNA genomes are denoted. For details on these and other hominin skeletal remains from other parts of the cave see (8).

**Supplementary Materials:**

Materials and Methods

Figures S1-S53

Tables S1-S10

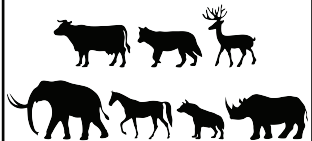
References (47-159)

Data files S1-S4



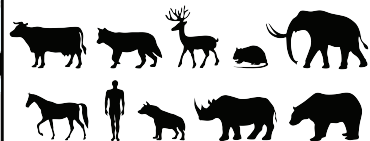
### Les Cottés

LP:4/4



### Trou Al'Wesse

LP:5/5



### Chagyrskaya Cave

LP:6/7 + MP:0/2



### El Sidrón

LP:1/1



### Vindija Cave

LP:1/1



### Caune de l'Arago

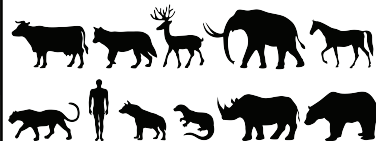
MP:0/13



Bovidae Canidae Cervidae Cricetidae Elephantidae Equidae Felidae Hominidae Hyaeinidae Mustelidae Rhinocerotidae Ursidae


### Denisova Cave

LP:30/34 + MP:14/18



Ancient and present-day modern humans

Trou Al'Wesse Stratum 17b 


Denisova 5 ("Altai") 


Denisova (East) Layer 14 


Denisova (Main) Layer 17 

Denisova (Main) Layer 19.1 

Denisova (East) Layer 11.4 

Denisova (Main) Layer 14.3 


Mezmaiskaya 1 


Okladnikov 2 

Chagyrskaya Unit 6C1 

El Sidron Stratum III 

Feldhofer 2 


El Sidron 1253 

Feldhofer 1 

Vindija 33.25 


Vindija 33.17 

Vindija 33.16 


Vindija 33.19 

Sima de los Huesos 

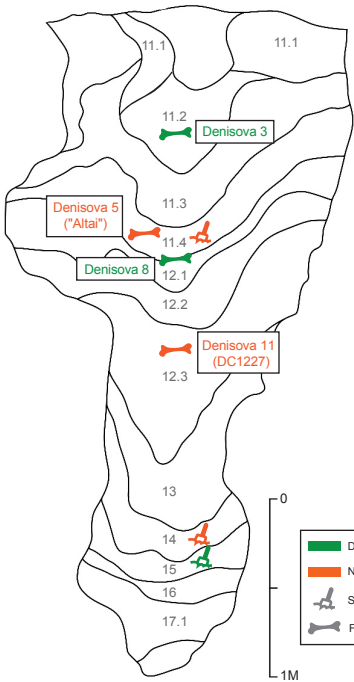
Denisova (East) Layer 15 

Denisova 8 

Denisova 3 

Denisova 4 







## Supplementary Materials for

### Neandertal and Denisovan DNA from Pleistocene sediments

Viviane Slon, Charlotte Hopfe, Clemens L. Weiß, Fabrizio Mafessoni, Marco de la Rasilla, Carles Lalueza-Fox, Antonio Rosas, Marie Soressi, Monika V. Knul, Rebecca Miller, John R. Stewart, Anatoly P. Derevianko, Zenobia Jacobs, Bo Li, Richard G. Roberts, Michael V. Shunkov, Henry de Lumley, Christian Perrenoud, Ivan Gušić, Željko Kućan, Pavao Rudan, Ayinuer Aximu-Petri, Elena Essel, Sarah Nagel, Birgit Nickel, Anna Schmidt, Kay Prüfer, Janet Kelso, Hernán A. Burbano, Svante Pääbo, Matthias Meyer

correspondence to: [viviane\\_slon@eva.mpg.de](mailto:viviane_slon@eva.mpg.de) or [mmeyer@eva.mpg.de](mailto:mmeyer@eva.mpg.de)

#### **This PDF file includes:**

Materials and Methods  
Figs. S1 to S53  
Tables S1 to S10  
Captions for Data files S1 to S4

#### **Other Supplementary Materials for this manuscript includes the following:**

Data files S1 to S4

# Table of Contents

## **Materials**

Sample collection .....	4
Caune de l' Arago (France) .....	4
Chagyrskaya Cave (Russia) .....	6
Denisova Cave (Russia) .....	7
Les Cottés (France) .....	9
Trou Al'Wesse (Belgium) .....	10
Vindija Cave (Croatia) .....	11
El Sidrón (Spain) .....	13

## **Methods**

### **DNA extraction**

Evaluating whether DNA binds to different sediment components .....	14
Determining the DNA binding capacity of clay and lime .....	15
Testing DNA extraction methods .....	17
Extracting DNA from archaeological sediments .....	19

### **DNA library preparation .....**

20

### **Shotgun sequencing and data analysis .....**

22

### **Enrichment for mtDNA fragments by hybridization capture .....**

24

### **Analysis of mammalian mtDNA capture data**

Taxonomic classification of mammalian mtDNA sequences in simulated datasets .....	27
Sequencing and raw data processing .....	30
Taxonomic classification of mammalian mtDNA sequences from archaeological samples ..	31
Identification of ancient sequences based on damage-derived base substitutions .....	32
Uniformity of sequence coverage along the mtDNA genome .....	33
Phylogenetic inferences at the species level .....	34
Comparisons with the zooarchaeological record .....	35
Testing how much of the variation in a layer is represented by a single sediment sample ...	35

Testing the homogeneity of taxa identification in repeated sampling .....	36
Testing the effect of sequencing depth .....	37
Testing for biases introduced during hybridization capture .....	37
Comparing the yields of DNA between sediment and skeletal remains .....	38
<b>Analysis of human mtDNA capture data</b>	
Sequencing and raw data processing .....	40
Identification and authentication of hominin sequences .....	40
Phylogenetic inferences using ‘diagnostic’ positions .....	41
Reconstructing mtDNA consensus sequences .....	42
Reconstructing phylogenetic trees .....	42
Estimating contamination by present-day human DNA .....	43
Determining whether sequences originated from more than one mtDNA genome .....	44
<b><u>Supplementary Figures S1-S53</u></b> .....	52
<b><u>Supplementary Tables S1-S10</u></b> .....	106
<b><u>Captions for Data files S1-S4</u></b> .....	117

## Materials

### Sample collection

The sediment samples from Caune de l'Arago (samples 1-13), Les Cottés (samples 75-78) and Trou Al'Wesse (samples 79-83) were collected during the 2015 excavation season and stored at -20°C. The sample from El Sidrón (sample 85) was collected in 2007 and stored at -20°C since. At these sites, sediments were sampled while wearing gloves, using sterile instruments and after the removal of surface material. The samples from Chagyrskaya Cave (samples 14-22) and 49 samples from Denisova Cave (samples 23-27, 29-33, 35, 37-74) were collected as part of another project during the 2014 season using a hand-held auger and stored at room temperature until shipped to the laboratory in Leipzig where they were stored at 4°C. The remaining samples from Denisova Cave (samples 28, 34 and 36) and the sample from Vindija Cave (sample 84) were obtained in 2011 and 2007, respectively, and stored at room temperature. The sediments were collected in containers of up to 50ml in volume, and no particular measures were taken to prevent the mixing of the material collected in each container. Details on the sampling locations are provided in Data file S1.

### Caune de l'Arago (France)

The site of Caune de l'Arago is located in the Pyrénées-Orientales department in southern France. The ~15m thick sequence of the site has been divided into four main stratigraphic complexes (lower, middle, upper and terminal), the middle one being the richest in archaeological material. The stratigraphy stretches between layers dated to 690 thousand years ago (kya) at the bottom of the lower complex and 70kya at the upper complex,

alternating periods of dry and cold climate with humid temperate ones. Evidence for hominin presence at the site have been identified in 15 layers, ranging between 550kya and 400kya (47-50). The lithic industry (Mode 2) appears to be generally constant, with no obvious changes in raw material choices between the Lower and Middle Paleolithic sequences, although differences in the types of tools made have been hypothesized to correspond with the use of the site as either a short- or long-term camp (48, 51).

To date, 149 hominin remains have been discovered at the site, representing at least 30 individuals. The distinct archaic morphological features of these remains account for an attribution to a newly-defined group of hominins, *Homo erectus tautavelensis* (50).

Studies of faunal remains at the site have revealed the presence of a variety of large mammals, including various ungulates (*e.g.* argali mouflon, wild horse, red deer, thar, reindeer, fallow deer, bison, rhinoceros and primitive muskox) and carnivores (*e.g.* Deningeri bear, wolf, lynx, fox, dhole and wild cat). Some of the earliest known remains of brown bear and steppe bison in western Europe have been uncovered at the site (47, 52). Birds and small vertebrate species have been identified as well (53).

Thirteen sediment samples from the middle stratigraphic complex were collected from the site (Data file S1). The majority of samples originate from layer G, a layer with a mean age of  $438\pm 31$  kya, which yielded the holotype of *Homo erectus tautavelensis*, the Arago XXI skull (50, 54). Other samples originate from layer J, where herbivore teeth have been dated by ESR/U-series to between  $342\pm 30$  and  $410\pm 42$  kya; and from layer Q which is correlated to the beginning of Marine Isotopic Stage 14, around 550kya (54).



### Chagyrskaya Cave (Russia)

Chagyrskaya Cave is located near the Charysh River within the Altai Mountains (Siberia, Russia). Seven stratigraphical units have been defined at the site, from Middle and Late Pleistocene layers (Units 5-7) to Holocene ones (Units 1-4). The Middle Paleolithic lithic industry is composed of a Mousteroid variant dubbed “Sibiryachikha facies”, to date identified only in Chagyrskaya Cave and the nearby Okladnikov Cave (55, 56). Based on a combined study of pollen records and large mammal assemblages, it has been suggested that the site was used by Neandertals as a long-term hunting camp during periods of dry climate, and later abandoned when the environment became warm and humid (43, 57).

Over 50 hominin remains, some of which could be attributed to Neandertals based on their morphology, have so far been found in Units 5 and 6 (43). These include an ulna fragment in Unit 6A; a deciduous canine and an atlas fragment in Unit 6B; and a mandible fragment and two isolated teeth from Unit 6C (44, 58, 59).

Although less than 5% of the osseous assemblage in Units 5-7 can be recognized to the species levels, 35 mammalian species have been identified. Skeletal remains are abundant in Units 5-6, and have been attributed among others to carnivores (*e.g.* dog, fox, cave hyena, weasel and cave lion), ungulates (such as reindeer, gazelle, goat and sheep), megafauna (woolly mammoth, woolly rhinoceros, horse and bison), small mammals (*e.g.* rodents, squirrel and pika), fish and birds. Unit 7 is markedly poor in remains, containing an order of magnitude fewer bone fragments than Units 5-6 (43, 56, 57).

We analyze 9 sediment samples from Chagyrskaya Cave, spanning Units 3, 5, 6 and 7 (Data file S1). While Unit 3 may be as young as the Holocene, dating of Units 5-6 by optically- or infrared-stimulated luminescence point to an age range of between  $47\pm 3$

and  $59\pm 3$ ky, in concordance with the dating of bison bones from these layers by radiocarbon, which yielded mostly ages beyond the limit of the method (*i.e.*  $>49$ ky). Unit 7 appears to be far older than the overlying deposits and has been dated to  $304\pm 22$ ky (43, 56, 60). The sediment sample from Unit 6C2 was collected from near the cave mouth, whereas all of the other samples were collected from the sediment profile exposed 5m inside the cave.

### Denisova Cave (Russia)

Denisova Cave is situated in the Anui River basin, in the Altai mountain range of Siberia (Russia). Ongoing excavations in the three galleries of the cave have so far revealed several distinct cultural layers associated with the Middle and Upper Paleolithic. Changes in the material culture can be observed throughout the stratigraphy, from an Early Middle Paleolithic lithic industry in the lower layers (Layers 21 and 22 in the Main Gallery and Layer 17 in the East Gallery) to an assemblage including bone and stone tools and ornaments in the uppermost strata associated with the Upper Paleolithic (Layers 9 and 11 in the Main and East Galleries) (55, 61, 62).

Genetic, archaeological and morphological studies indicate that modern humans, Neandertals and Denisovans lived in or in the vicinity of the cave. While no modern human skeletal remains have been identified to date, a proximal toe phalanx (63) (“Denisova 5” or “Altai Neanderthal”) from layer 11.4 of the East Gallery yielded the first high-coverage genome of a Neandertal (21); and an undiagnostic bone fragment (“Denisova 11” or “DC1227”) from layer 12 (sub-layer unknown) of the East Gallery was identified as hominin by collagen peptide mass fingerprinting and later found to have

a Neandertal-like mitochondrial (mt) genome (35). Additionally, three skeletal remains have been attributed to the Denisovans: a proximal hand phalanx (“Denisova 3”) from layer 11.2 in the East Gallery, whose genetic makeup defined the Denisovans as an archaic hominin group closely related to the Neandertals (2-4); an upper molar (“Denisova 4”) from layer 11.1 of the South Gallery (3, 64); and an upper molar (“Denisova 8”) from the transition between layers 11.4 and 12 in the East Gallery (64). A hominin deciduous lower molar (“Denisova 2”) was discovered in layer 22.1 in the Main Gallery (65, 66); however, given the absence of comparative morphological data for deciduous Denisovan teeth and without genetic data, its attribution to a hominin group is undetermined.

The vast majority of skeletal remains recovered at the site are too fragmentary to allow their identification by morphology (35, 67, 68). Nonetheless, analyses of the morphologically informative faunal remains from the Pleistocene layers have led to the identification of approximately 50 distinct taxa. The large mammalian assemblage includes carnivores (*e.g.* fox, bear, cave hyena and cave lion), ungulates and proboscideans (such as red deer, roe deer, horse, wild sheep, woolly rhinoceros and woolly mammoth). Smaller mammals include bats, voles, pikas, shrews, marmots and martens. Birds, reptiles, amphibians and fish have also been identified at the site (67-71).

We here analyze 28 sediment samples from the East Gallery and 24 from the Main Gallery (Data file S1). Sediment samples from the East Gallery encompass Layers 8 to 17.2, and preliminary optical ages on individual sedimentary quartz and potassium-rich feldspar grains from the same sediment samples suggest an age range of ~14 kya (Layer 8) to ~240 kya (Layer 17). Sediment samples from the Main Gallery were

obtained from Layers 9 to 22.3 and preliminary optical ages suggest an age range of ~25 kya (Layer 9) to >200 kya (Layer 22). These ages are supported by mostly infinite radiocarbon ages for bones with and without cut-marks from Layers 11 and 12 (3, 35) and radiation-induced thermoluminescence ages on quartz grains from Layers 14, 21 and 22 in the Main Gallery (72, 73).

### Les Cottés (France)

The site of Les Cottés is located in France, in the southwest part of the Parisian Basin. To date, eight archaeological strata have been defined. These are constituted of cultural layers bearing artifacts attributed to the Mousterian, Châtelperronian, Proto-Aurignacian and Early Aurignacian industries, interspaced with sterile layers. Thus, the deposits of the site span the disappearance of Neandertals and the appearance of modern humans in the area (74, 75). Analysis of the clay minerals in the sediment of the site suggests a cold environment with intermittent periods of temperate climate (76).

Skeletal remains of a single adult modern human individual were discovered at the site in the late 19<sup>th</sup> century and are thought to originate from the Early Aurignacian period (77). Renewed analyses of this partial skeleton are ongoing (75).

Faunal remains at the site have recently been investigated based on their morphology and by proteomic fingerprinting using mass-spectrometry. The most commonly identified taxa are reindeer, horse and bison; although remains of hyenas, rhinoceros, mammoths and foxes are also present (75, 76, 78). Modifications on these skeletal remains can be attributed to both carnivore and anthropogenic activities (75, 78).

Four samples from Les Cottés were analyzed: two from a layer characterized by Châtelperronian artifacts (US06) and two from the layer containing Mousterian ones (US08) (Data file S1). The Châtelperronian industry was radiocarbon dated to between 41.0 and 43.4 cal kyr BP (95.4% confidence interval), and the Mousterian industry to between 41.6 and 46.2 cal kyr BP (95.4% confidence interval) (79, 80). Optically stimulated luminescence dating of individual grains of quartz from the same layers provided a weighted mean age of  $43.1 \pm 2.2$  kyr ( $1\sigma$ ) for the Châtelperronian industry (consistent with the calibrated radiocarbon ages) and a slightly older weighted mean age of  $51.3 \pm 3.0$  kyr ( $1\sigma$ ) for the Mousterian industry; the latter is supported by an optical age of  $49.2 \pm 3.6$  kyr ( $1\sigma$ ) obtained using a multiple-aliquot MET-pIRIR procedure for potassium-rich feldspar grains (80).

#### Trou Al'Wesse (Belgium)

The site of Trou Al'Wesse is located near the village of Modave in the province of Liège, Belgium. Ongoing excavations have allowed the determination of 19 strata, yielding evidence for repeated occupations of the site by humans from the Late Pleistocene to the Middle Ages. Notably, the site points to a turnover from an earlier Neandertal occupation to a later one by modern humans, as indicated by a replacement of the Mousterian industry found in Stratum 17 at the base of the sequence by an Aurignacian material culture in Stratum 15 and inside the cave (45, 81, 82).

Modern human remains have been found in the upper archaeological strata dated to the Holocene. These include right mandibular fragments most probably originating from Stratum 7a and redeposited in Stratum AC (83), a cranial fragment from Stratum 4

(84) and isolated teeth from Strata 4b-delta and 4b-LaH dated to the Mesolithic period; a Neolithic burial yielding cranial fragments of at least nine individuals found in association with ceramic artifacts (85, 86); and a human tooth from the Bronze Age or Middle Ages from Stratum 2 (87). To date, no human remains have been found in the Pleistocene layers.

Faunal remains are abundant in the Pleistocene layers, including those that are associated with Mousterian tools (Stratum 17). Mammalian faunal remains from Stratum 17 have been attributed to equids, cervids, ursids, bovids, mustelids, insectivores, voles and lemmings, rhinoceroses and hyenas (45, 82). Previous genetic studies conducted on skeletal remains from the site focused on ptarmigans (88), water voles (89), collared lemmings (90) and red deer specimens (91) collected from layers younger than the ones sampled for the present study; and a previous attempt at amplifying DNA from the sediment by PCR was unsuccessful (12).

Five samples were collected for DNA analyses from the layers containing artifacts pertaining to the Mousterian lithic industry (Strata 17a and 17b) (Data file S1). The only bone dated by radiocarbon from this layer yielded a calibrated date of between 42.3 and 49.6kya (uncalibrated: OxA-7497, 41,100±2,300) (82, 92).

### Vindija Cave (Croatia)

The site of Vindija Cave is situated in the Hrvatsko Zagorje area in Croatia. Fourteen major stratigraphical units have been defined and include layers from the Holocene (Layers A-C) and the Pleistocene (Layers D-N), the latter spanning the interface between

the Middle and Upper Paleolithic material cultures and the replacement of Neandertals by modern humans in the region (93-95).

To date, over 100 hominin skeletal remains have been uncovered at the site. Studies of their morphology indicate that they pertain to both Neandertals (in Layers D-I) and modern humans (Layer D) (93-99). Additionally, genetic data have been obtained from several of the Neandertal individuals. Nuclear DNA retrieved from Vi33.16 (Layer G3), Vi33.25 (Layer I) and Vi33.26 (Layer G) were used to reconstruct the first draft of the Neandertal genome sequence (100). Additionally, the full mitochondrial genomes of the above-mentioned individuals (36, 38, 100), as well as of Vi33.17 (Layer I) and Vi33.19 (Layer G3), have been reconstructed (37).

Among the faunal assemblages of the Pleistocene layers, bear remains are the most abundant (~80%) and are ubiquitous throughout the stratigraphy (95, 101, 102). Two of these have been identified based on their mitochondrial DNA as cave bears typical of Eastern Europe (*Ursus ingressus*) (103, 104). Other taxa detected in the site include rodents, lagomorphs, carnivores (*e.g.* cave hyenas, mustelids and canids) and ungulates (*e.g.* reindeer, bison and Merck's rhinoceros) (102, 105). Previous analyses of a sediment sample from the site focused solely on its bacterial component (106).

We included in our experiments a single sediment sample collected in Layer G3 (Data file S1), which is characterized by a majority of Mousterian artifacts with some Upper Paleolithic elements (95). Radiocarbon dating of a Neandertal bone from this layer yielded an age greater than 42,000 years (107), while a cave bear bone was dated using U/Th to 41,000 +1000/-900 years BP (101).

### El Sidrón (Spain)

El Sidrón is a cave site situated in the Asturias region in Northern Spain. The site is atypical in the Pleistocene archaeological record, as the skeletal assemblage is vastly dominated by archaic human remains, probably accumulated in the cave in a single event (32, 108). Lithic artifacts found at the site originate from a classic Mousterian industry (109).

Over 2,500 Neandertal skeletal elements have been uncovered at the site, all from a single stratigraphical unit (Stratum III), and attributed to a minimum of thirteen individuals of both sexes and different ages (32, 108, 110, 111). Genetic analyses of some of the Neandertal remains from the site have been conducted using mtDNA, exome, chromosome 21 and Y-chromosome data (38, 110, 112-115).

The very few non-human remains from the site originate from a red deer, a large bovid and few small mammals and gastropods, and there is little evidence for carnivore or rodent activity on the skeletal remains (32).

We analyzed one sediment sample, collected in the Neandertal-bearing unit (Data file S1). The first radiocarbon dating of skeletal remains from the site yielded uncalibrated ages ranging between  $37,300 \pm 830$  and  $40,840 \pm 1,200$  years, resulting in an average calibrated age of  $43,129 \pm 129$  years (115). Subsequent dating of other artifacts ranged between  $\sim 35$  and  $49$  kya (116), while one bone (OxA-21 776) was dated to  $48,400 \pm 3,200$  years BP (uncalibrated) following an ultrafiltration pre-treatment protocol (117). Other methods (U/TH, OSL, ESR and AAR) resulted in an age estimate of more than 38,000 years (116).



## Methods

### DNA extraction

#### Evaluating whether DNA binds to different sediment components

Sediments are a conglomerate of different inorganic and organic components as well as air, water and living organisms (6). Previous experiments have shown that DNA can bind rapidly (< 2 hours) to both the mineral and the organic components of sediments, albeit the amount of DNA adsorbed depends on a range of factors, such as the concentration of salts and the pH (5-7, 118, 119). It has been shown that shorter DNA fragments bind more efficiently than longer ones to different types of sediments, however these shorter fragments were ~2.7k base pairs (bp) in size (120), roughly 50 times longer than ancient DNA fragments recovered from ancient bones and teeth (18, 29, 34, 121, 122). We thus aimed to determine to which extent very short DNA fragments in conditions mimicking body fluids can undergo stable interactions with some of the most abundant mineral inorganic components of sediment: clay (montmorillonite), lime, sand (quartz) and feldspar (oligoclase).

For this purpose we diluted 2µg (4µl) of a pool of DNA fragments (50bp ladder, Thermo Fisher Scientific) in 996µl Ringer solution (Serumwerk) supplemented with calcium (130mM Na<sup>+</sup>, 5.4mM K<sup>+</sup>, 1.0mM Mg<sup>2+</sup>, 2.4mM Ca<sup>2+</sup>, 115mM Cl<sup>-</sup>, 27mM acetate), thereby matching the ion concentration in human blood, and added 150mg of each material. After rotating overnight at room temperature, the samples were spun down using a table top centrifuge at 16,363 x g for 1:30 minutes and the supernatant was taken off without disturbing the pellet. Then, three 5-minutes washes with 1ml EBT buffer

(10mM Tris-HCl pH 8.0, 0.05% Tween-20) were performed to simulate water movement that may release DNA bound to sediment. The supernatant of each wash step was retained. To release any DNA that may have bound, an elution step with 0.5M sodium phosphate at pH 7.0 (Alfa Aesar) was carried out by rotating the samples for 15 minutes at room temperature. The supernatant of the binding step, washes I and III as well as the eluate were purified from undesired salts using the MinElute PCR purification kit (Qiagen) with plugged-in extension reservoirs to allow for large volumes (123). The purified DNA was separated by agarose gel electrophoresis at 120V for 1 hour using a 2% LE agarose gel (Biozym) containing SYBR Safe dye (Thermo Fisher Scientific). To determine whether the Ringer solution, EBT or sodium phosphate buffer interfere with the purification, three controls were generated by adding the same quantity of DNA to each buffer and purifying these as well.

As can be seen in Fig. S1, no ladder bands can be detected in the supernatant lanes of any of the tested sediment components, showing that DNA bound to each of the materials. In the EBT washes, some but not all DNA was released from sand, clay and oligoclase. In the extraction phase with sodium phosphate, DNA was released from all materials. In conclusion, all the components bound DNA under the test conditions, suggesting that sediment can undergo stable interactions with DNA fragments of varying lengths, enabling the preservation of DNA over time.

#### Determining the DNA binding capacity of clay and lime

To determine how much DNA can bind to different types of sediment under physiological conditions, we first generated large quantities of a 100bp PCR product

using a pUC19 plasmid as template for amplification with a primer pair of the sequences AATAGCACCGCCTACATACCTC (forward) and AAATCGTCTTGAGTCCAACCCG (reverse primer). After purification, 2 $\mu$ g of the product were added to 150mg, 110mg, 70mg or 30mg of clay and lime, respectively. As clay has previously been shown to bind high amounts of DNA (100 $\mu$ g DNA per milligram clay) at pH 5 (5), its binding capacity was additionally tested on 15mg, 5mg and 1mg of material. As above, DNA binding was carried out in Ringer solution supplemented with calcium ions during overnight rotation. The minerals were then washed three times with EBT. DNA was extracted by incubation in 1ml of 0.5M EDTA, pH 8.0 (AppliChem) overnight at room temperature. Following the purification of 100 $\mu$ l of each eluate using the MinElute PCR purification kit (Qiagen), the DNA was separated and visualized on an agarose gel as described above. A 100bp band was expected to be visible only at DNA/mineral ratios exceeding the binding capacity of the sediment component.

All of the DNA added bound to between 5mg and 150mg of clay, but not to 1mg of clay, as indicated by a 100bp band in the supernatant lane at this ratio (Fig. S2). This suggests that the binding capacity of clay for DNA under blood ionic conditions lies below 2  $\mu$ g DNA per 5mg of clay, or 400ng DNA per milligram. However, in line with the observations made in the experiment above, some of the bound DNA was released in the wash steps, indicating that DNA binding to clay is weak in solutions of low ionic strength. For lime, no DNA could be detected in the supernatant when added to between 70mg and 150mg of material, indicating that all the DNA bound to it at these concentrations. At a ratio of 2 $\mu$ g DNA to 30mg lime, approximately half the DNA failed to bind, as can be inferred by the roughly equally strong DNA bands recovered in the

supernatant and extraction lanes (Fig. S3). This suggests that the binding capacity limit of DNA to lime under physiological conditions is around 2 $\mu$ g DNA per 70mg of lime or 28.5ng of DNA per milligram. Very little DNA is released from lime in the wash steps, indicating that DNA forms stable interactions with lime also in solutions of low ionic strength.

#### Testing DNA extraction methods

To identify an extraction protocol that would be most suitable to recover ancient DNA from sediment, four methods were compared in respect to their ability to recover short DNA fragments from clay and lime. To do so, 2 $\mu$ g (4 $\mu$ l) of an Ultra Low Range (ULR) DNA ladder (Thermo Fisher Scientific) was bound to 150mg of each material in 996 $\mu$ l Ringer solution with added calcium. This ladder was chosen to mimic ancient DNA molecules that are typically of very short fragment size (*18, 29, 34, 121, 122*). Following the binding reaction by overnight rotation, the supernatant was taken off the pellet, which was then washed three times with EBT. Subsequent steps varied between the extraction protocols, as following:

(i) Phosphate extracts were produced by adding 1ml of 0.5M NaHPO<sub>4</sub> (Alfa Aesar), resuspending by vortexing and rotating at room temperature for 15 minutes. This is similar to what was described in (*124*) for the release of DNA from lake sediments, except that we used 0.5M sodium phosphate at pH 7.0 instead of 0.12M at pH  $\approx$ 8. The samples were then spun down and the extract taken off. This step was repeated twice to produce extracts I and II.

(ii) The DNA was extracted by adding 1ml of 0.5M EDTA at pH 8.0 (AppliChem) to the materials. After vortexing, suspensions were rotated overnight and subsequently spun down to recover the final extract. We note that elsewhere (6), DNA was released from sand using 0.4M EDTA in combination with different detergents.

(iii) To test a method originally designed for DNA extraction from modern sediment (125) and previously used in ancient DNA studies (9), 994 $\mu$ l of the “Bulat” lysis buffer (50mM EDTA (pH 8.0), 50mM Tris-HCl (pH 8.0), 150mM NaCl, 2.5% N-Lauroylsarcosine and 500mM mercaptoethanol) and 6 $\mu$ l proteinase K (100 $\mu$ g/ $\mu$ l, Sigma) were added to the material and rotated overnight at 37°C. The samples were then spun down and the extract taken off.

(iv) The PowerMax Soil DNA Isolation Kit (MoBio), which was used for DNA isolation from marine sediments in a previous study (126), was used following the manufacturer’s protocol except that reaction volumes were scaled down by a factor of 16.2. We used only 150mg of input material (instead of ~ 2.5g), 926 $\mu$ l of PowerBead solution and approximately 50 disruption grains. This mixture was vortexed for 1 minute, then 74 $\mu$ l of C1 solution was added and we vortexed again for 30 seconds. Subsequently the samples were mixed for 20 minutes at 1,500 rpm on an MKR 13 Thermo Shaker (HLC Ditabis) and then spun down to recover the extract.

For all tested methods, the extracted DNA was purified using the MinElute PCR purification kit (Qiagen) with an extension reservoir (123) and half the product was visualized on a 1% LE / 3% Sieve agarose gel (120V for 1:10 hours). A favorable method would be identified by stronger bands in the extraction lane, as well as by the recovery of shorter DNA molecules. To show that DNA binding had not occurred on the tube itself, a

negative control was subjected to the same binding and washing conditions and was extracted using the sodium phosphate buffer.

Of the four extraction methods tested, sodium phosphate, EDTA and the PowerMax Soil DNA Isolation Kit all successfully retrieved DNA from clay, as can be inferred from the appearance of bands in their respective extraction lanes. Based on the intensity of the bands, the EDTA method and the two phosphate extractions released roughly an equal amount of DNA, and performed better than the soil extraction kit (Fig. S4). All four methods released DNA bound to lime, with the faintest bands obtained using the commercial extraction kit. The shortest DNA fragments were recovered using the “Bulat” buffer, down to 25bp. While the DNA bands extracted with EDTA or sodium phosphate look similar in intensity, they only recover DNA as short as 35bp (Fig. S5). In the no substrate controls, DNA was recovered in the supernatant or wash I (Figs. S4, S5), indicating that no binding of DNA to tube walls had occurred to a detectable extent.

#### Extracting DNA from archaeological sediments

Only the EDTA-based and sodium phosphate-based methods successfully and efficiently retrieved DNA from both clay and lime. On this basis, we decided to apply the EDTA-based protocol routinely used in our laboratory to recover DNA fragments from skeletal remains (*I23, I27*) to extract DNA from the archaeological sediment samples. DNA was extracted from between 38 and 160mg of sediment.

Details on the input amounts used for DNA extraction and the negative controls included are provided in Data file S1. An overview of the laboratory procedures applied is presented in Fig. S6.

## DNA library preparation

Aliquots of 5 $\mu$ l or 10 $\mu$ l (10 or 20% of the extract) were converted into single-stranded DNA libraries, which were prepared either manually as described (22, 127), or using a modified version of the method (23) implemented on an automated liquid handling platform (Bravo NGS workstation B, Agilent Technologies). For use with the liquid handling system, reaction steps involving magnetic beads were performed by pipetting bead suspensions up and down every 5 minutes instead of keeping beads suspended by interval mixing on a thermoshaker. Library preparation was performed for 48 samples in parallel using 96-well microplates. Small quantities of a control oligonucleotide were spiked into the sample DNA to monitor potential inefficiencies in library preparation due to the presence of inhibitory substances. The total number of molecules in each library, as well as the number of spiked-in internal control fragments, were assessed by qPCR (22) or digital-droplet PCR (24). Each library was amplified to plateau (128) and tagged with two indexes (127, 129). Amplification products were purified with the MinElute PCR purification kit (Qiagen) and eluted in 30 $\mu$ l TE buffer (1M Tris-HCl, pH 8.0, 0.5M EDTA pH 8.0) if the libraries had been prepared manually. Libraries prepared in microplates were purified on the liquid handling system using SPRI technology (130) as follows: PCR products were mixed thoroughly with an equal volume of a DNA binding bead suspension containing 1 mg/ml Sera-Mag carboxylate-modified magnetic particles (GE Healthcare), 38% PEG-8000 (w/v), 1M NaCl, 0.05% Tween-20, 1mM EDTA and 10mM Tris-HCl, pH 8.0. Beads were pelleted on a magnetic rack and the supernatant discarded. Beads were then washed twice with 80% ethanol without breaking the pellet,

dried for 10 min at 37°C and resuspended in 20µl TE buffer. Beads were pelleted on a magnetic rack and the eluate transferred to a new microplate.

Details on the input amounts used for DNA library preparation, the type of library preparation performed and the negative controls included are provided in Data file S1.



## Shotgun sequencing and data analysis

To obtain a general overview of the taxa comprised in the DNA libraries prior to the performance of any target enrichment, all libraries were shotgun sequenced and analyzed using a taxonomic binning approach. For this, libraries were pooled and heteroduplexes were removed by a one-cycle PCR reaction using Herculase II Fusion DNA polymerase (Agilent Technologies) (128) with primers IS5 and IS6 (131). After determining the concentration of DNA in the pools using a DNA-1000 chip (Agilent Technologies), 76-cycle, paired-end sequencing was carried out with parameters suited to double-indexed libraries (129) on MiSeq or HiSeq 2500 v3 platforms (Illumina). Base calling was carried out using Bustard (Illumina) or freeIbis (132). Adapters were clipped and overlapping forward and reverse reads were merged using leeHom (133). Only sequences carrying the exact expected indexing combinations were retained for downstream processing. For further analysis we established a minimal length cutoff of 30 base pairs to decrease the number of spurious alignments.

Sequences were aligned to the full non-redundant nucleotide (nt) collection of the National Center for Biotechnology Information (NCBI) (downloaded January 2015) using the MEGAN alignment tool (MALT, version 0.0.12) (134) in BlastN mode. MALT uses a BlastN-like algorithm to generate local alignments of all obtained sequences to matching DNA sequences in a given database. After generating alignments, a lowest common ancestor (LCA) algorithm (135) is used to bin sequences along the NCBI taxonomy. The LCA assigns a sequence to the node in the taxonomy that is the lowest common ancestor of all nodes, which have alignments with bit-scores within 10% of the

best alignment. This means that sequences with unspecific alignments over multiple taxa end up with assignments at higher taxonomic levels. MALT saves the alignments as well as the binning information in an RMA file, which can then be processed and inspected using the Metagenome Analyzer MEGAN (version 5.11.3) (135). MEGAN allows visual inspection of how sequences are binned along the taxonomy, and provides a command line interface (CLI) for basic operations involving RMA files. Here, it was used to extract the number of sequences that are assigned to or below taxa from 7 different high taxonomic levels from the three domains of cellular life. We also kept track and reported the number of sequences which had no alignments to any comparative sequence in the database. Assignments to the domain Bacteria were divided into the three most abundant phyla (Actinobacteria, Firmicutes and Proteobacteria); assignments to the domain Eukarya were divided into fungi (Fungi), mammals (Mammalia) and plants (Viridiplantae); assignments to the domain Archaea were not subdivided into more specific taxonomic levels. Other taxa were represented by few sequences (less than 5% in all non-blank samples). Sequence counts were further analyzed using the statistical computing environment R. Results for samples are shown in Figs. S7-S15, and for extraction and library preparation negative controls in Fig. S16 and Fig. S17, respectively.

## Enrichment for mtDNA fragments by hybridization capture

Aliquots of each library were enriched for mtDNA fragments using two sets of synthetic 52-mer probes. The first encompasses 242 mammalian mtDNA genomes (24) and the second spans the full mitochondrial genome of a present-day human. The latter probe set was designed in 1 bp tiling using the Revised Cambridge Reference Sequence with the 'N' removed at position 3107, synthesized on an array, and converted into a probe library as described elsewhere (136). Probe libraries were amplified using Herculase II Fusion DNA polymerase (Agilent Technologies) with oligonucleotides APL2 and APL6 (136) as PCR primers. APL2 carries a 5' biotin to allow immobilization of the amplified probes on beads. Amplified probes libraries were purified using SPRI technology (130) as described in (137), except that the concentration of PEG-8000 in the MagNA suspension was adjusted to 38% (w/v) and a 1:1 ratio of MagNA suspension and sample was used.

Hybridization capture was performed in two successive rounds using an on-bead hybridization protocol (138) with modifications as follows: On a magnetic rack, 20 $\mu$ l of Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher Scientific), multiplied by the number of samples, were washed twice in 1000 $\mu$ l BWT buffer (1M NaCl, 10mM Tris, 1mM EDTA, pH 8.0, 0.05% Tween-20). The bead pellet was resuspended in BWT buffer, and 5 $\mu$ l of a 100ng/ $\mu$ l solution containing the biotinylated probes (500ng bait per reaction) were added. Reactions were mixed properly by vortexing and rotated for 15 minutes at room temperature. Beads were then pelleted using a magnetic rack and the supernatant was discarded. Beads were washed once with 1ml TT buffer (1mM Tris-HCl pH 8.0, 0.01% Tween-20) and twice with 1ml melt solution (125mM NaOH, 0.05%

Tween-20), with a 5-minutes incubation at room temperature in each melt step to remove the non-biotinylated strand. After an additional wash with 1ml TT buffer, beads were resuspended in 20µl BWT buffer, multiplied by the number of samples, and dispensed into a 96-well semi-skirted PCR plate.

In a separate 96-well reaction plate, the input DNA (1µg in the first round and 0.5µg in the second round of capture) was added to 0.5µL of each of four 500µM blocking oligos (BO4: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Pho; BO6: CAAGCAGAAGACGGCATAACGAGAT-Pho; BO8: GTGTAGATCTCGGTGGTCGCCGTATCATT-Pho; BO11: GGAAGAGCGTCGTGTAGGGAAAGAGTGT-Pho), 25µl of 2x HI-RPM hybridization buffer (Agilent Technologies) and 5µl of blocking agent (Agilent Technologies). Water was added for a total reaction volume of 50µl. After vortexing, the reactions were incubated at 95°C for 3 minutes and then at 37°C for at least 5 minutes. The reactions were added to the 96-well PCR plate containing the probes and incubated with gentle rotation at 65°C for 24-72 hours.

The following steps were performed either manually or using the liquid handling system. Post-hybridization washes were performed as described (136). Beads were then resuspended in 22µl melt solution and incubated at room temperature for 5 minutes. The supernatant was taken off the beads and added to a mixture of 5µl Sera-Mag carboxylate-modified magnetic particles (GE Healthcare), which had been washed twice in 500µl TE buffer and resuspended at a concentration of 50mg/µl, and 3µl 3M sodium acetate (pH 5.2). 90µl of 100% ethanol were added to each reaction and the plate was rotated for 15 minutes at room temperature. After discarding the supernatant by pelleting the beads on a

magnetic rack, the beads were washed twice in 150µl of either 70% ethanol in manual preparation or 80% ethanol in the automated procedure without breaking the pellet, dried for 4 minutes at 37°C, and resuspended in 30µl TT buffer to yield the capture eluate.

As described elsewhere (136), 1µl of the capture eluate was used as template for a qPCR assay to evaluate the recovery of molecules. The remaining 29µl were amplified using Herculase II Fusion DNA polymerase (Agilent Technologies) with primers IS5 and IS6 (131). Purification of the amplification products was carried out using SPRI technology as described above, either manually or using the liquid handling system.

## Analysis of mammalian mtDNA capture data

### Taxonomic classification of mammalian mtDNA sequences in simulated datasets

Ancient DNA datasets typically exhibit three characteristics that differentiate them from data generated from present-day organisms. First, DNA molecules become fragmented over time, resulting in the retrieval of relatively short molecules from ancient samples (18, 29, 121, 122). Second, chemical damage to the DNA fragments leads to nucleotide substitutions. Most commonly, cytosine (C) bases undergo deamination, resulting in uracil residues which are then read by DNA polymerases as thymine (T) bases (17, 139, 140). In the absence of an enzymatic treatment to remove uracil bases, C to T substitutions occur in high frequencies at the ends of sequences and in lower frequencies throughout them (17, 22, 37, 141). Third, some exceptions notwithstanding (e.g. (3, 21, 142)), usually only small amounts of DNA endogenous to the source material can be retrieved, whereas the vast majority of DNA fragments sequenced originates from environmental contaminants (e.g. (127, 143-145)).

To test whether existing software for taxa identification would be suitable to analyze ancient DNA, we generated three simulated datasets. Five mammalian mitochondrial genomes (of a Neandertal, a woolly mammoth, a hyena, a cow and a pig) and 114 bacterial genomes of various genera (Table S1) were randomly fragmented into sequences between 35 and 100 bp long. Each simulated dataset contained a total of 100,000 sequences composed of a mixture of mammalian mtDNA and/or bacterial DNA sequences - either 100% mammalian, 99% bacterial and 1% mammalian, or 100%

bacterial DNA (Fig. S18A). We then introduced to each of the three datasets varying levels of nucleotide substitutions, as follows:

(A) No C to T changes;

(B) 50% of Cs at the first and last positions of sequences from all genomes were transformed to Ts;

(C) 50% of Cs at the first and last positions of sequences from the Neandertal, the woolly mammoth and the hyena mtDNAs were altered to Ts;

(D) 50% of Cs at the first and last positions of sequences from the Neandertal mtDNA, the woolly mammoth mtDNA, the hyena mtDNA and the bacterial genomes were changed to Ts;

(E) 50% of Cs at the first and last positions and 10% of Cs at remaining positions of sequences from all genomes were transformed to Ts.

The simulated sequences were aligned to a non-redundant database of 796 placental mammalian mitochondrial genomes from the Reference Sequence (RefSeq) collection of NCBI (downloaded January 2016) (*146*) using nucleotide BLAST (BlastN) (*147*) with default parameters. The output was parsed using the lowest common ancestor (LCA) algorithm implemented in MEGAN (*135*) version 5.10.7. This algorithm assigns sequences matching only one specific taxon to that group, while those matching several genomes are assigned to the most recent common ancestor of these taxa. For a match to be retained, we required the alignment score to be at least 35 and within the top 10% of the highest alignment score for that sequence. For a taxon to be deemed present in the dataset, we required it to be represented by more than two sequences and by at least 1% of the total number of identifiable hits.

Using these parameters on the first simulated dataset containing only mammalian mtDNA sequences, between 71,208 and 85,549 sequences passed the filters and were utilized for taxonomic identification at the family level. Even when nucleotide changes were introduced (simulations B, C and E), no sequences were erroneously assigned to an unexpected family and the percentage of sequences assigned to each taxon closely matched the input data. For the second dataset, in which 1% of the sequences originated from mammalian mtDNA and 99% from bacterial genomes, between 714 and 852 sequences were attributed to a mammalian family. Four out of five families were correctly identified, however the small number of Neandertal mtDNA sequences used as input (10 sequences) was insufficient to identify the presence of Hominidae in the simulated dataset. None of the bacterial DNA sequences were misidentified as mammalian DNA, as corroborated by the analysis of the third dataset comprised only of bacterial DNA, where none of the sequences were assigned to a mammalian taxon (Fig. S18B).

Two issues hindered us from using MEGAN to assign sequences at a lower taxonomic level than biological families. First, when attempting to assign DNA fragments at the genus level, fewer sequences passed the filtering scheme and were identified to be of mammalian origin (57,405-68,730 sequences from the first dataset and 577-682 from the second one), markedly reducing the amount of data available for further analyses. Of those, 1.03% and 1.21%, respectively, were incorrectly assigned when C to T substitutions were introduced throughout the sequences (simulation E), making the taxa identification of ancient DNA-like fragments less reliable. Second, the lowest common ancestor algorithm implemented in MEGAN utilizes the NCBI taxonomy



to assign sequences to a given taxon, while discarding the assignment to an ancestor taxon if both matches are significant (135). This may introduce errors in taxa identification when the NCBI classification is debatable. For example, Neandertals are classified by NCBI as a subspecies of modern humans. Thus, a sequence matching equally well the Neandertal and modern human reference mitochondrial genomes will be assigned specifically to Neandertals, instead of being placed higher up in the taxonomy (*i.e.* *Homo* or Hominidae). Using the MEGAN attribution of sequences at a higher taxonomic level circumvents such unresolved classifications at the species or genus levels.

In conclusion, we find that the parameters utilized here are sufficiently stringent to make false assignments of ancient mammalian mtDNA sequences at the family level unlikely, but DNA from taxa present in very small proportions may remain undiscovered.

#### Sequencing and raw data processing

To generate sequencing data from the archaeological sediment samples, DNA libraries enriched for mammalian mtDNA fragments were pooled and heteroduplices were removed by a one-cycle PCR reaction using Herculase II Fusion DNA polymerase (Agilent Technologies) (128) with primers IS5 and IS6 (131). The concentration of DNA in the pools was assessed using a DNA-1000 chip (Agilent Technologies). Sequencing was carried out by 76-cycle paired-end runs with parameters suited to double-indexed libraries (129) on a MiSeq platform (Illumina). Base calling was performed using Bustard (Illumina). Adapters were clipped and overlapping forward and reverse reads were

merged using leeHom (133). Only sequences carrying the exact expected index combinations were retained for downstream processing.

In order to retain only DNA fragments bearing a resemblance to mammalian mtDNA, sequences originating from mammalian mitochondrial capture were aligned to the 242 mitochondrial genomes represented in the capture probes (24) using BWA (148) with permissive parameters (seeding turned off and allowing for more mismatches and indels than by default) (4). Unmapped sequences and those shorter than 35bp were discarded using SAMtools (149). PCR duplicates were removed by collapsing perfect sequence replicates into one sequence. To mitigate the influence of sporadic sequencing errors, sequences were retained only if they were seen at least twice.

#### Taxonomic classification of mammalian mtDNA sequences from archaeological samples

The sequences were compared using nucleotide BLAST (BlastN) (147) to a non-redundant database of 796 full mitochondrial genomes of placental mammals, downloaded from the RefSeq database of NCBI (146) in January 2016. We used the lowest common ancestor algorithm implemented in MEGAN (version 5.10.7) (135) to assign sequences to taxa at the family level. As above, a match was retained only if its alignment score was at least 35 and within the top 10% of the highest score achieved for that sequence. Taxa were concluded to be present in a library only when more than two sequences and at least 1% of the total number of identifiable sequences were assigned to them.

For each detected taxon, sequences attributed to it were realigned to a reference mitochondrial genome from that family using BWA as above (4, 148). A list of the

reference genomes used is provided in Table S2. PCR duplicates were identified based on identical start and end positions and collapsed into single sequences using bam-rmdup (<https://bitbucket.org/ustenzel/biohazard>). SAMtools (149) was used to filter for sequences longer than 35bp and with mapping quality of at least 25. The average length of retained sequences was computed using a custom Perl script.

#### Identification of ancient sequences based on damage-derived base substitutions

Cytosine (C) bases at the single-stranded ends of DNA fragments have a tendency to undergo deamination to uracil bases, which are recognized by DNA polymerases as thymines (T) (17, 139, 140). When sequences are aligned to a reference genome, this results in substitutions from Cs to Ts near the start and end of sequence alignments. As this damage accumulates over time, ancient DNA sequences are likely to carry elevated frequencies of terminal C to T changes (more than 10% of terminal Cs are read as Ts in samples older than 500 years), while DNA younger than 100 years displays little or no evidence of deamination (18, 19, 141).

To quantify the level of deamination-induced base damage in the mammalian mtDNA sequences we obtained from sediments, we determined the frequencies of C to T substitutions at the 5' and 3' ends of alignments using an in-house Perl script. We performed this analysis independently for each taxon in each library (see Fig. S19 for an example). Given that the sediment samples we analyzed were all dated to at least ~14kya, we required the frequency of C to T substitutions at the 5'- and 3'-ends of sequences to be significantly higher than 10% for a taxon to be identified as ancient (19), as determined using an exact binomial test implemented in R (version 3.3.1) (150).

### Uniformity of sequence coverage along the mtDNA genome

Since the preparation of DNA libraries retains a random subset of molecules present in an extract, the sequences generated are expected to map randomly across a relevant reference genome. In contrast, the use of an incorrect reference genome is more likely to result in the mapping of sequences to specific parts of it, such as regions that are conserved among mammalian mtDNAs. To verify whether the sequences we assigned to a given taxon mapped randomly, we computed the variance of coverage across positions in the reference mitochondrial genome in libraries where a taxon presented degradation patterns typical of ancient DNA, and excluded outliers with high variance (higher than one standard deviation from the mean), *i.e.* taxa where sequences mapped only to a restricted area in the genome.

Among all taxa in all libraries, only sequences attributed to Procaviidae, which were detected in a single library from Les Cottés (library R4070), showed an unusually high variance of coverage (Fig. S20). All of the 46 sequences mapped to a single 60-bases long region in the reference genome. In comparison, the 45 sequences attributed to Elephantidae from the same library covered 1,644 bases spread across the reference genome, with not more than 3 sequences covering any given position (Fig. S21). Comparison of the 46 putatively Procaviidae sequences to the full NCBI database (October 2016) using BLAST (147) revealed that only 6 match the cape rock hyrax genome best (*i.e.* with the highest score). In contrast, 14 matched best a spotted hyena mtDNA different from the one represented in our non-redundant reference database, 1

matched both and 25 matched these and other genomes equally well. We thus conclude that these sequences were misassigned to the Procaviidae family.

### Phylogenetic inferences at the species level

To identify ancient taxa in our datasets to a lower taxonomic level, we used publicly available full mtDNA genomes from extant and extinct species pertaining to each family to define branch-specific sequence differences, *i.e.* bases inferred to have changed in only one branch of a phylogenetic tree relating several taxa from that family. For this, full mitochondrial genomes were downloaded from GenBank (151) and aligned using MAFFT (152) (see Table S2 for the list of genomes and their accession codes). Branch-specific sequence differences were determined by identifying positions where one group of mtDNAs carried one base, while all others carried another (33). The state of sequences overlapping these sets of “diagnostic” positions was assessed using a custom-made Perl script. To diminish the influence of damage-derived substitutions (17), the first three and last three positions of sequences were ignored if they carried a T (aligned in the orientation as sequenced) or an A (aligned in reverse complement orientation). We then computed the percentage of sequences matching the state specific for each taxon.

Such an approach is useful to attribute sequences to a taxon when relatively little data has been generated (33), however the level of resolution in this analysis depends on the availability of comparative data. For example, full mitochondrial genomes are available for five extant and one extinct species of Rhinocerotidae, allowing us to identify between 257 and 362 positions in their mtDNAs where one species carries one base, while all others share another (Data file S3). These sets of positions are defined based on

only one mitochondrial genome sequence per species; if more genomes become available, the sets are likely to become smaller and to better reflect the fixed differences between these lineages. Moreover, while we are able to compare sequences to the mtDNA genome of the woolly rhinoceros (153), we are unable to identify other extinct species (*e.g.* the giant rhinoceros or Merck's rhinoceros) for which no sequencing data have been made available to date. Lastly, the identification of phylogenetically informative positions is dependent on the presence of differences between groups of interest. We are unable, for example, to distinguish between aurochs and modern cattle as their mtDNA sequences do not produce a reciprocally monophyletic genealogy (154).

#### Comparisons with the zooarchaeological record

For each site for which ancient DNA fragments were retrieved from the sediments, we compared the taxa identified based on their genetic material to the faunal composition determined from the zooarchaeological record, as described in (43, 57, 60) for Chagyrskaya Cave; in (67-69, 155, 156) for Denisova Cave; in (76, 78, 157) for Les Cottés; in (45, 82) for Trou Al'Wesse; in (102, 105) for Vindija Cave; and in (32) for the site of El Sidrón.

#### Testing how much of the variation in a layer is represented by a single sediment sample

To evaluate whether multiple samples are needed to represent the taxonomic composition in an archaeological layer, we compared the ancient taxa detected in four layers from which 3 or 4 different samples were collected (Data file S1). Given that the libraries prepared from these samples were exhaustively sequenced, as indicated by an average

duplication rate of at least 4 (Data file S2), sequencing depth is not expected to affect this comparison. Following the processing of the first sample from each layer, we counted the number of ancient taxa that were newly identified after analyzing the remaining samples. The majority of taxa were already detected in the first sample, suggesting that a single sample can be sufficient to represent the overall faunal composition identifiable by DNA (Fig. S35).

#### Testing the homogeneity of taxa identification in repeated sampling

To test whether the taxa identified are consistent among different subsamples taken from a larger sediment sample, we extracted DNA from two additional subsamples of four samples (samples 16, 75, 79 and 84) originating from four different sites. The amount of sediment per subsample was similar to the amount taken for the first screen ( $\pm 2$ mg) (Data file S1). We then processed these additional subsamples as was done for the first ones. One library (L5671) that failed during the hybridization capture procedure was disregarded from further analyses.

There is a good concordance in the taxa identified among different subsamples of the same sediment sample. The taxa identified in all the subsamples of a given larger sample were represented by between 88% and 100% of sequences. However, taxa identified based on relatively few sequences in a certain subsample may not be identified in others (e.g. Ursidae in sample 79 and Suidae in sample 84) (Fig. S36). This is probably due to the filtering scheme we implemented to avoid false assignments, by which a taxon was deemed present only if at least 1% of sequences in the dataset were attributed to it. As the percentage of sequences assigned to the different taxa varied across subsamples

(Fig. S36), we opted to express our results in terms of detection of taxa in a given layer, rather than the determination of their relative abundances.

#### Testing the effect of sequencing depth

To test whether sequencing depth affects the determination of taxonomic compositions, we used SAMtools (*149*) to sample randomly ~500,000 sequences from six libraries, from which 1.8 to 3.2 million DNA fragments were originally sequenced (Data file S2). The downsampled libraries were processed as described above, and the output was compared to the results obtained when using all sequences. The number of sequences identified as mammalian mtDNA in the downsampled libraries constituted between 13% and 65% of such sequences in the full datasets. For four of the libraries (R3231, R4069, R4100 and R4490), the taxa determined are identical between the two sets. In the remaining two libraries, the taxon represented by the lowest number of sequences in the full dataset is missing from the downsampled one (Canidae in R4073 and Felidae in R3852) (Fig. S37). Our inability to detect the least abundant taxon in a downsampled dataset most probably results from the filtering scheme we implemented to minimize false assignments, requiring at least 1% of sequences to be attributed to a given family to infer its presence.

#### Testing for biases introduced during hybridization capture

To test whether the hybridization capture (*24*) influences the taxonomic composition detected, we processed data obtained by shotgun sequencing of 13 DNA libraries (extraction set 4, Data file S1) as was done following capture. The results were compared



to the taxa composition determined following the enrichment of the same libraries for mammalian mtDNA fragments.

For six samples (sample numbers 20-22, 75, 77, 78) and the library preparation negative control (sample 97), no mammalian mtDNA sequences were identified in the shotgun data. In the other cases, few mammalian mtDNA sequences were detected, amounting to between 0.08% and 0.13% of sequences identified in the capture data for the samples (sample numbers 16-19, 76) and 0.86% of sequences for the extraction negative control (sample 89) (Fig. S38). The small amount of mtDNA sequences produced by shotgun sequencing points to the impracticality of using our analytical approach without prior enrichment for mammalian mtDNA fragments.

In all cases, the taxa determined from the shotgun data constituted a subset of the taxa identified following capture. The taxon represented by the highest number of sequences in the capture data was always identified also in the shotgun dataset (Hyaenidae for the samples and Hominidae for the negative control). However, taxa to which relatively few sequences were attributed in the capture data were not detected in the shotgun data (*e.g.* Elephantidae in sample 17 or all taxa except Hyaenidae in sample 76). No taxa were identified solely in the shotgun data (Fig. S38).

#### Comparing the yields of DNA between sediment and skeletal remains

We aimed to assess how the yield of mtDNA sequences recovered from sediment compares to the yield from skeletal elements. To do so, we compared the number of unique mtDNA sequences retrieved from hominin bones in previous studies (2, 21, 35-38) to the number of mammalian mtDNA sequences (all taxa combined) recovered from

sediment samples originating from the same layers at three archaeological sites. These numbers were then normalized by the amount of material used in DNA extraction (Data files S1, S2).

## Analysis of human mtDNA capture data

### Sequencing and raw data processing

Since hominin sequences cannot be detected in mammalian mtDNA capture data if they contribute to less than 1% of the mammalian sequences (see above for details on the data processing), we enriched the DNA libraries again using probes spanning the full mitochondrial genome of a present-day human (136, 138) (Fig. S6). The sequencing of libraries and raw data processing were carried out as described above for the libraries enriched for mammalian mtDNA fragments.

### Identification and authentication of hominin sequences

Sequences originating from enrichment for human mtDNA fragments were mapped to the revised Cambridge reference human mitochondrial genome sequence (NC\_012920) using BWA (148) with permissive parameters (4). Unmapped sequences and sequences shorter than 35bp or with a mapping quality lower than 25 were removed by using SAMtools (149), and PCR duplicates were removed using bam-rmdup (<https://bitbucket.org/ustenzel/biohazard>). After comparing the sequences to the non-redundant mammalian mtDNA database and parsing the output with MEGAN (135, 147) as described above, only sequences attributed to Hominidae were retained. We note that since no great apes other than those on the *Homo* lineage are expected to be present during the Middle and Late Pleistocene at the sites we investigated, we assume that any hominid DNA fragment we recover originated from hominins. The average length of sequences which passed our filtering scheme was evaluated using an in-house Perl script.

When more than one library was prepared from a sample (Data file S1), sequencing data were merged using SAMtools (149). Sequences were evaluated for the presence of terminal C to T substitutions (17) to the reference genome as described above. We required that the sequences present terminal C to T substitutions to the reference genome at their 5'- and 3'-ends at a frequency significantly higher than 10% for ancient hominin DNA to be deemed present (19), as tested using an exact binomial test in R (version 3.3.1) (150). Libraries from which sequences mapped only to a restricted area in the reference genome, as indicated by an unusually high variance of coverage (higher than one standard deviation from the mean), were excluded (Fig. S39).

#### Phylogenetic inferences using “diagnostic” positions

To define sets of positions differentiating between the mtDNAs of modern humans, Neandertals, Denisovans and the Sima de los Huesos (SH) hominin, we required that either 99% (for modern humans) or all (for Neandertals and Denisovans) of the mtDNAs pertaining to that group carry an identical base. Branch-specific variants were identified as those where the mtDNAs of one or more of these groups differed from a chimpanzee mtDNA and the mtDNAs of all other groups (33). To minimize the effect of deamination on this analysis, Ts were ignored if they were within the first three and last three alignment positions (in the orientation as sequenced). The number of DNA sequences matching the variant specific to each branch was computed using an in-house Perl script. A list of the genomes used to define the diagnostic positions is presented in Table S2.

### Reconstructing mtDNA consensus genomes

We reconstructed mtDNA genomes from sediment samples using only sequences presenting a C to T difference to the reference genome at their 5'- or 3'-ends, and by calling bases by a majority vote (34). To prevent damage-derived substitutions from affecting this procedure, Ts on forward strands and As on reverse ones in the three first and three last positions of a sequence were converted to Ns. To maximize the usability of the hominin DNA fragments, we used relatively relaxed parameters for reconstructing the mtDNA genomes. A base was called if it was covered by at least two sequences and if more than two-thirds of sequences carried an identical base. As such an approach may be prone to errors being introduced to the consensus genome (due to sequencing error, residual contamination and/or the presence of multiple ancient mtDNA genomes in the sample), we also attempted to use a stricter scheme for consensus calling, where we required a minimum coverage of three sequences for a base to be called. However, as this approach left very few base calls for two of the samples (samples 44 and 82; Table S3), we used the permissive approach for tree reconstruction. The number of pairwise base differences among the mtDNA consensus genomes reconstructed from sediments was computed in MEGA 6.0 (158), based on a multiple sequence alignment file created using MAFFT v7.271 (152) (Table S4).

### Reconstructing phylogenetic trees

Phylogenetic trees were reconstructed using the neighbor-joining method (158) implemented in MEGA 6.0 (158), by comparing each of the reconstructed mtDNA genomes to ancient and present-day hominin mtDNAs (Table S5). Multiple sequence

alignment files were created using MAFFT v7.271 (152). Evolutionary distances were computed based on the number of base differences between genomes, after excluding alignment positions with missing data in any of them. Support for each node was assessed by generating 500 bootstrap replicate trees.

We note that the fact that only partial mtDNA genomes were reconstructed from some of the sediment samples (Table S3) resulted in only 178 overlapping positions between them, hindering us from combining all nine mtDNAs in one phylogenetic analysis. To circumvent this issue, we reconstructed a phylogenetic tree relating only the mtDNA genomes in our comparative dataset (Table S5), using the software and parameters described above. We then used the branches leading to each of the mtDNA genomes reconstructed from sediment samples in the above-mentioned trees (Figs. S41-S49) to infer their position in a phylogenetic tree relating all of the mtDNAs in our dataset (Fig. 2).

#### Estimating contamination by present-day human DNA

To evaluate the extent of contamination by present-day human mtDNA in the datasets from which we reconstructed mtDNA genomes, we compared each of these mtDNA genomes to a panel of 311 present-day human mtDNAs and identified positions where the former differs from all of the latter (36). Between 3 and 70 such positions were determined (Table S6). For each sample, the percentage of sequences overlapping these positions that match the present-day human state constitutes the estimate of contamination by human mtDNA. Contamination estimates among sequences showing evidence of deamination are smaller than 10% for all samples.

## Determining whether sequences originated from more than one mtDNA genome

### *(i) Observing variable positions within each library*

It is possible that DNA from more than one archaic individual is present in any given sediment sample from which we reconstructed an mtDNA genome. To determine whether more than one mtDNA type can be distinguished in each dataset, we investigated positions in the mitochondrial genome covered by at least 10 sequences from putatively deaminated fragments, and computed the percentage of sequences carrying an identical base at these positions (Fig. S50). When two different bases were observed at a given position, to increase our confidence that the variability at those positions is not due to sequencing errors, deamination or contamination, we required each base to be seen at least twice in both sequencing orientations. We do note that at ~10-fold coverage, this approach would remove variants that deviate strongly from a 50:50 representation. Additionally, we tested whether the positions at which we observed more than one base are known variants among Neandertal or Denisovan mtDNAs in our comparative dataset (Table S5).

### *ii) Maximum likelihood approach*

To determine whether one or more mtDNA types are present in the datasets containing relatively few sequences, we developed a maximum likelihood approach to estimate the proportion of each mitochondrial genome contributing to a sample, as well as the divergence between the different components when multiple mtDNAs are distinguishable. The underlying assumption of the model is that the proportions of DNA

fragments contributed to the sample by different mtDNA genomes will be reflected in the proportions of sequences carrying different bases at variable positions within that dataset. Hence, our method is based on the intuitive idea that the frequency spectrum of the variable positions within each sample would reflect the proportions of the different mtDNAs in the sample.

For a specific combination of  $n$  individuals with different mtDNA genomes, the number of sequences supporting each base follows a multinomial distribution  $M(\pi_c)$ , with the probabilities of the multinomial  $\pi_c = \{\pi_{c1}, \pi_{c2}, \pi_{c3}, \pi_{c4}\}$  depending on the fraction of sequences contributed by each individual (called “mtDNA components” henceforth)  $P = \{p1, \dots, pn\}$ . At any given position, not all mtDNA components need to differ in sequence, and some will carry identical bases. At any given position, we therefore need to consider each possible partitioning  $c$  of the components. For example, for two components  $s1$  and  $s2$ , there are two possible partitions ( $s1=s2$  and  $s1 \neq s2$ ), depending on whether the two mtDNA sequences show identical or two different bases at the position under consideration. For three components, five possible partitions exist ( $s1=s2=s3$ ,  $s1=s2 \neq s3$ ,  $s1 \neq s2 = s3$ ,  $s1 \neq s2 \neq s3$ ,  $s1=s3 \neq s2$ ). In general, for a number  $n$  of components there are  $B_n$  possible partitions of the components, where  $B$  indicates the Bell number for  $n$  elements. The probabilities  $\pi_c$  of the multinomial distribution corresponding to a given partition  $c$  are computed as the sum of the proportions of identical sequences. Thus, in the case of two components, the probabilities associated with the two multinomials corresponding to partitions  $s1=s2$  and  $s1 \neq s2$  are  $\pi_1 = \{p1+p2, 0, 0, 0\}$  and  $\pi_2 = \{p1, p2, 0, 0\}$  respectively; while for three components, the multinomials associated with the partitions listed above are  $\pi_1 = \{p1+p2+p3, 0, 0, 0\}$ ,  $\pi_2 = \{p1+p2, p3, 0, 0\}$ ,  $\pi_3 = \{p1, p2+p3, 0, 0\}$ ,



$\pi_4=\{p1,p2,p3,p4\}$  and  $\pi_5=\{p1+p3,p2,0,0\}$ , respectively. Note that we do not consider the identity of each sequence, *i.e.* its state relative to the reference. Hence, the likelihood of each partition  $i$  has to be calculated and summed up over all permutations of  $\pi_i$ .

In addition to the frequencies of the different components, we include as free parameters estimated from the model the probabilities  $Q=\{q1..q_c..q_{Bn}\}$  of each possible partitioning  $c$  (associated to a distribution  $\pi_c$ ). If only two components are present in the dataset,  $Q=q_c$  corresponds to the divergence between the two mitochondrial genomes (*i.e.* the proportion of positions at which they differ). Since errors and mis-mapped sequences introduce variation, they could appear as additional components. We therefore included error in the model by estimating a parameter  $p_{err}$ , representing the probability that a sequence at a given position supports a false state. Hence, at each position we consider all possible configurations of observed sequences arising when a number  $n_{err}$  of sequences supports a false state. We assume a uniform distribution of errors across positions, hence for a given position, the probability of all configurations with  $n_{err}$  are weighted by the probability of having  $n_{err}$  in  $m_j$  sequences overlapping position  $j$ . Specifically, this distribution follows a binomial distribution with probability  $p_{err}$  and size parameter  $m_i$ .

We note that assuming a uniform distribution of errors across positions potentially underestimates the amount of errors, due to the susceptibility of certain bases or positions to specific types of damage (*e.g.* deamination (17)) or sequencing errors. In order to minimize the impact of deamination on our results, Ts on forward strands and As on reverse ones were ignored. Summarizing, for a model with  $k$  components, the free parameters are the probability of error  $p_{err}$ , the frequency of the different components  $P$  and the probabilities of each partition  $c$  ( $Q$ ).

Hence, the log likelihood can be written as:

$$L_k = \sum_{i \in \text{position}} \log \sum_{c \in C_k} \sum_{r \in R(C_k)} \sum_{n_{err}=0}^{nseq_i} \sum_{e \in E} q_c M(\pi_{c,r} | \bar{x}_i - \overline{xerr}_e) \binom{nseq_i}{nseq_i - n_{err}} P_{err}^{n_{err}} (1 - P_{err})^{nseq_i - n_{err}}$$

where  $nseq_i$  indicates the total count of sequences overlapping position  $i$  and  $n_{err}$  the number of errors. These quantities are equal to the sum over the counts of sequences  $\bar{x}_i$  and of the errors  $\overline{xerr}_e$  for each of the four different bases, *i.e.*  $nseq = \sum \bar{x}_i$  and  $n_{err} = \sum \overline{xerr}_e$ , respectively. The subscript  $e$  indicates one of the possible configurations of errors  $E(n_{err}, i)$  for a position  $i$  with counts  $\bar{x}_i$  and a number of errors  $n_{err}$ . Finally, we remind that we indicate with  $M$  a multinomial with probabilities  $\pi_{c,r}$ , where  $r$  indicates a specific permutation of the probabilities  $\pi_c$ .

In order to estimate the number of components  $k$  present in a sample, we calculated the likelihood that either one, two or three components contributed to the observed sequences. We then compared these models using a likelihood ratio test and by calculating the relative likelihood (relL) of the best model compared to the second best one. This maximum-likelihood approach was implemented in R (150).

To evaluate the efficiency of this method, we first simulated datasets of 15,000 independent positions with varying average coverage and with different proportions of two components, with a per position pairwise difference of 1% and an error probability of 1%, as shown in Fig. S52. We quantified the proportion of simulations (out of 100 iterations for each tested parameter) in which the number of components  $k$  was estimated correctly in the presence of two components (*i.e.* a model with  $k=2$  had lower AIC value than a model with  $k=1$  and  $k=3$ ) at an average coverage of 3-, 5- and 10-fold (Fig. S52A); as well as the proportion of correct inferences for a fixed proportion of the second

component (10%) at increasing divergence between the two components (Fig. S52B). For both manipulations, the proportion of correct inferences is above 95% under a wide range of parameters. For example, even when a simulated second component is introduced at small proportions (5%) or its divergence from the first component is low (1%), a 5-fold coverage is sufficient to correctly infer the presence of the second component, also in the presence of errors (1%). At a lower average coverage of 3-fold, the second mtDNA component is only correctly inferred if it is present at a higher proportion (10%) and/or if its divergence from the major component is higher (1.5%) (Fig. S52A-B).

In order to test the sensitivity to errors, we calculated the proportion of simulations in which a null model including only errors is rejected, when the dataset has been generated by a single mitochondrial genome. In this case, we observe that even at 3-fold coverage we overestimate the number of components in less than 2% of the simulations when errors are in the order of 1% (Fig. S52C). We do note that our simulations were generated assuming a uniform error distribution, therefore this does not exclude that strongly correlated errors or mis-mapped sequences might lead to an overestimation of the number of mitochondrial components.

Hence, in order to be conservative against overestimating the number of components due to correlated errors, we set an additional cutoff of 1% on the frequency of the estimated components, *i.e.* a component is not considered as a genuine additional mtDNA if its frequency is not at least 1%. Since we did not test for more than three components and we expect that the model does not have sufficient power to distinguish between exactly three and more than three different mtDNAs, we interpret an estimated  $k=3$  as the presence of multiple (three or more) mtDNAs in the sample. Lastly, we used

only datasets where the average mtDNA genome coverage was 2.5-fold or more, as the power of the model sharply decreases at lower coverages and as samples with lower coverage tested in exploratory analyses never supported the presence of more than one mtDNA.

We tested the method on a subset of eighteen samples from which animal mtDNA sequences attributed to three taxa were obtained, encompassing a large range of average mtDNA genome coverages (Table S8). For all tested samples, models with multiple components had higher support than a model with a single component, in which only the probability of error is estimated. For five out of eighteen samples (Elephantidae sequences in sample 61, Hyaenidae sequences in sample 83, and Ursidae sequences in samples 27, 32 and 39), the model with  $k=3$  indicating three or more mitochondrial components is supported over a model with only two mtDNAs. In the former two cases, however, the estimated frequencies of both minor components are low (below 1%), suggesting that the number of mtDNA genomes present in these samples are overestimated, possibly due to correlated errors. Hence the Elephantidae and Hyaenidae sequences in samples 61 and 83, respectively, originated for their vast majority and possibly in their entirety from a single mtDNA source of the relevant taxon.

We then applied this approach to samples from which hominin mtDNAs were reconstructed and where the average coverage was  $\sim 2.5$ -fold or more (Table S3). Only putatively deaminated sequences were considered, as described above for the consensus calling of the mtDNA genomes. Additionally, Ts on forward strands and As on reverse strands at any position in the sequence were ignored where the reference genome is a C or a G, respectively, as were bases with a quality lower than 30 and indels. Since

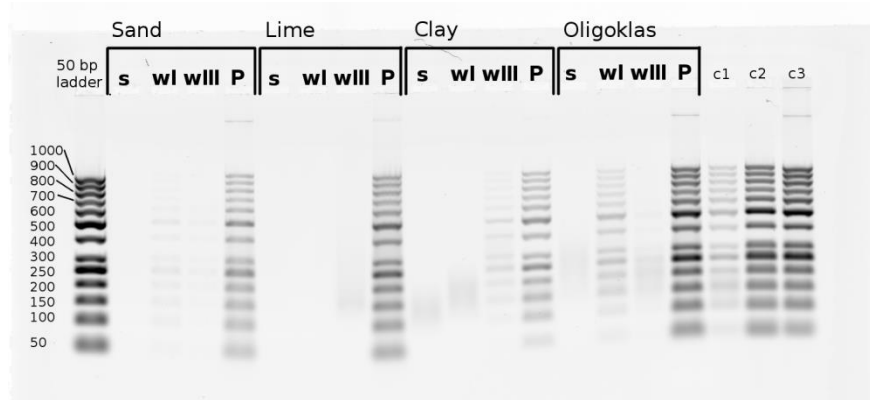
contamination might contribute present-day human DNA fragments to samples containing ancient hominin ones and appear as an additional mtDNA type, we expect to observe at least two mtDNA components in all cases where human contamination is detectable. We thus repeated the analysis after excluding positions where an observed variant is more likely to be from a contaminant, *i.e.* excluding the 63 positions in the mitochondrial genome where all of the 311 present-day humans in our comparative dataset carry a different base than all of the ten Neandertals (Table S5) (36). If the second component identified by the maximum likelihood model originates from contamination, one would expect a weaker support for this second component when excluding such positions than when considering all positions in the mtDNA genome.

The best supported model in all cases suggests the presence of multiple mtDNA components when using all available positions (Table S9). In two cases (sample 35 collected in Layer 11.4 of the East Gallery in Denisova Cave and sample 85 from El Sidrón), the exclusion of positions differing between modern human and Neandertal mtDNAs results in the removal of one of the mtDNA components estimated by the model. For the former, this reduces the number of estimated components to one, indicating the contribution of only one ancient mtDNA genome to the sequences obtained from the sample, in addition to the presence of contaminating DNA. In contrast, for the sample from El Sidrón, two mtDNA types remain distinguishable even after removing the putative contaminating component, suggesting that more than one ancient individual contributed to this sample.

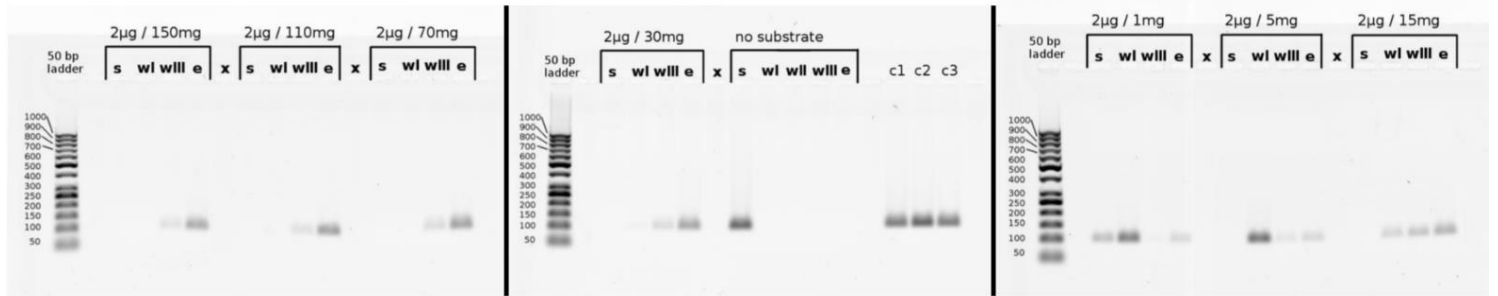
In the remaining cases (samples 19, 42, 60, 61 and 63), even when positions indicative of contamination are excluded, the number of mtDNA components in the best

supported model remains constant and suggests the presence of two mtDNAs (including for sample 60 where the fraction of the third component is below 1%). However, in the samples from Chagyrskaya Cave (sample 19) and from Layers 14.3 and 17 of the Main Gallery in Denisova Cave (samples 60 and 61), the exclusion of these positions results in a significant decrease in the inferred proportion of the second component, as is the case for the sample from El Sidrón (sample 85) (Table S9). The reduction in the inferred proportion of the second mtDNA type following the exclusion of positions likely to vary due to the presence of present-day human DNA suggests that the detected second component originated from contamination. We thus conclude that the sequences in samples 19, 60 and 61 originate from a mixture of a single ancient mtDNA type and contamination by present-day human mtDNA fragments. For sample 60, which yielded the highest number of deaminated mtDNA fragments (Data file S4, Fig. S53, Table S3), this is consistent with the observation that at any position in the mtDNA genome, nearly all sequences carry an identical base (Fig. S50D).

In contrast, the inferred proportion of the second mtDNA component detected in the samples from Layer 14 of the East Gallery and from Layer 19.1 of the Main Gallery of Denisova Cave (samples 42 and 63, respectively) is unaffected by the type of positions used in the analysis, coinciding with the low levels of contamination estimated for these samples (Table S6). In these two cases, we thus find it likely that both mtDNA components are derived from ancient individuals (Table S9).

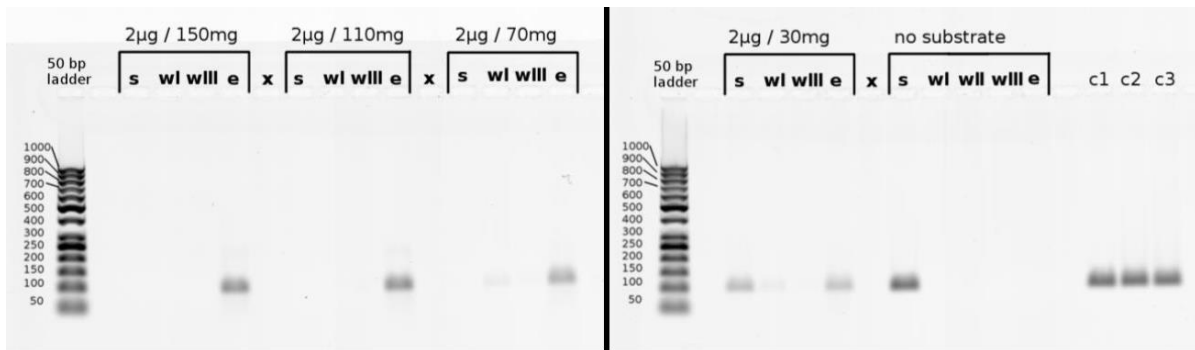


**Fig. S1. Binding of 2 $\mu$ g DNA to sand, lime, clay and oligoclase.** s - supernatant; w I & III – EBT washes I and III; P – extraction with 0.5M sodium phosphate; c1, c2, c3 - purification controls with Ringer solution and calcium, EBT and sodium phosphate buffers, respectively

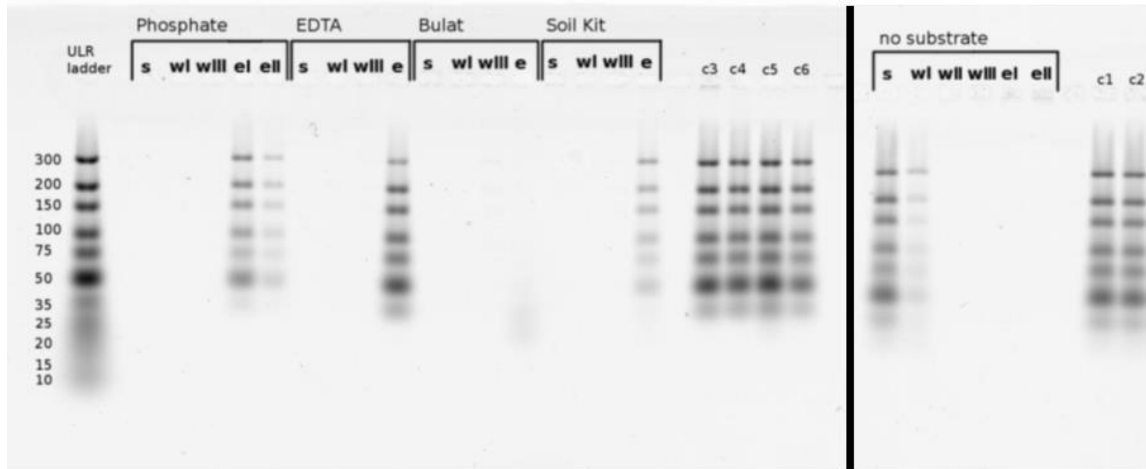


**Fig. S2. DNA binding capacity of clay.** s – supernatant ; w I, II, III – EBT washes I, II , III; e – extraction with 0.5M EDTA; c1, c2, c3 - purification controls with Ringer solution with calcium, EBT and EDTA, respectively.

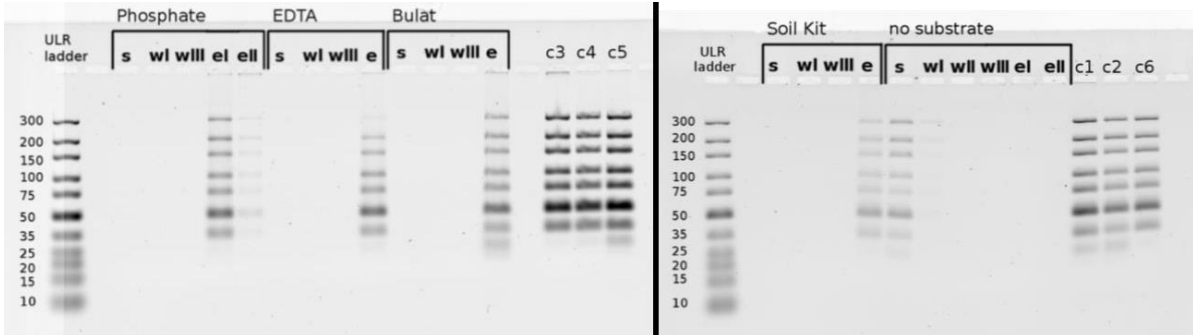




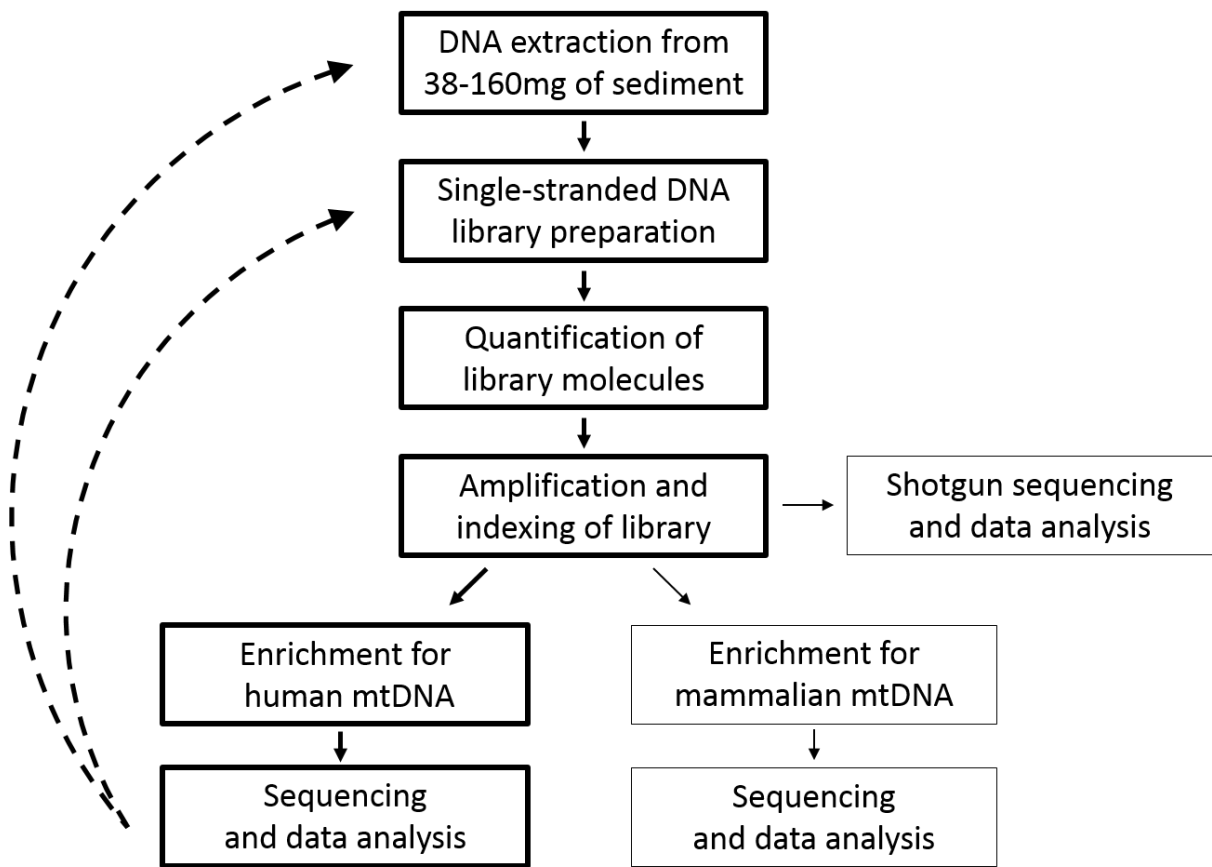
**Fig. S3. DNA binding capacity of lime.** s – supernatant ; w I, II, III – EBT washes I, II , III; e – extraction with 0.5M EDTA; c1, c2, c3 - purification controls with Ringer solution with calcium, EBT and EDTA, respectively.



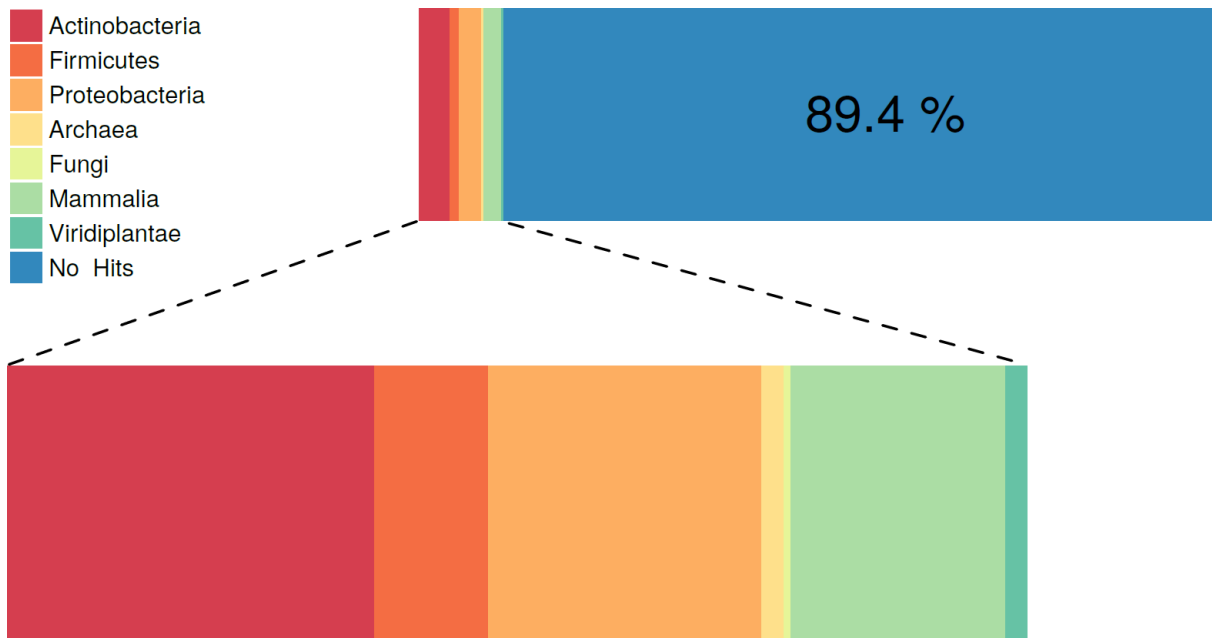
**Fig. S4. Comparison of different extraction methods on clay.** s - supernatant ; w I, II, III – EBT washes I, II, III; e – extract ; purification controls: c1 - Ringer solution with calcium; c2 – EBT buffer; c3 – sodium phosphate buffer; c4 – EDTA; c5 – “Bulat” buffer; c6 - buffer of the PowerMax Soil DNA Isolation Kit.



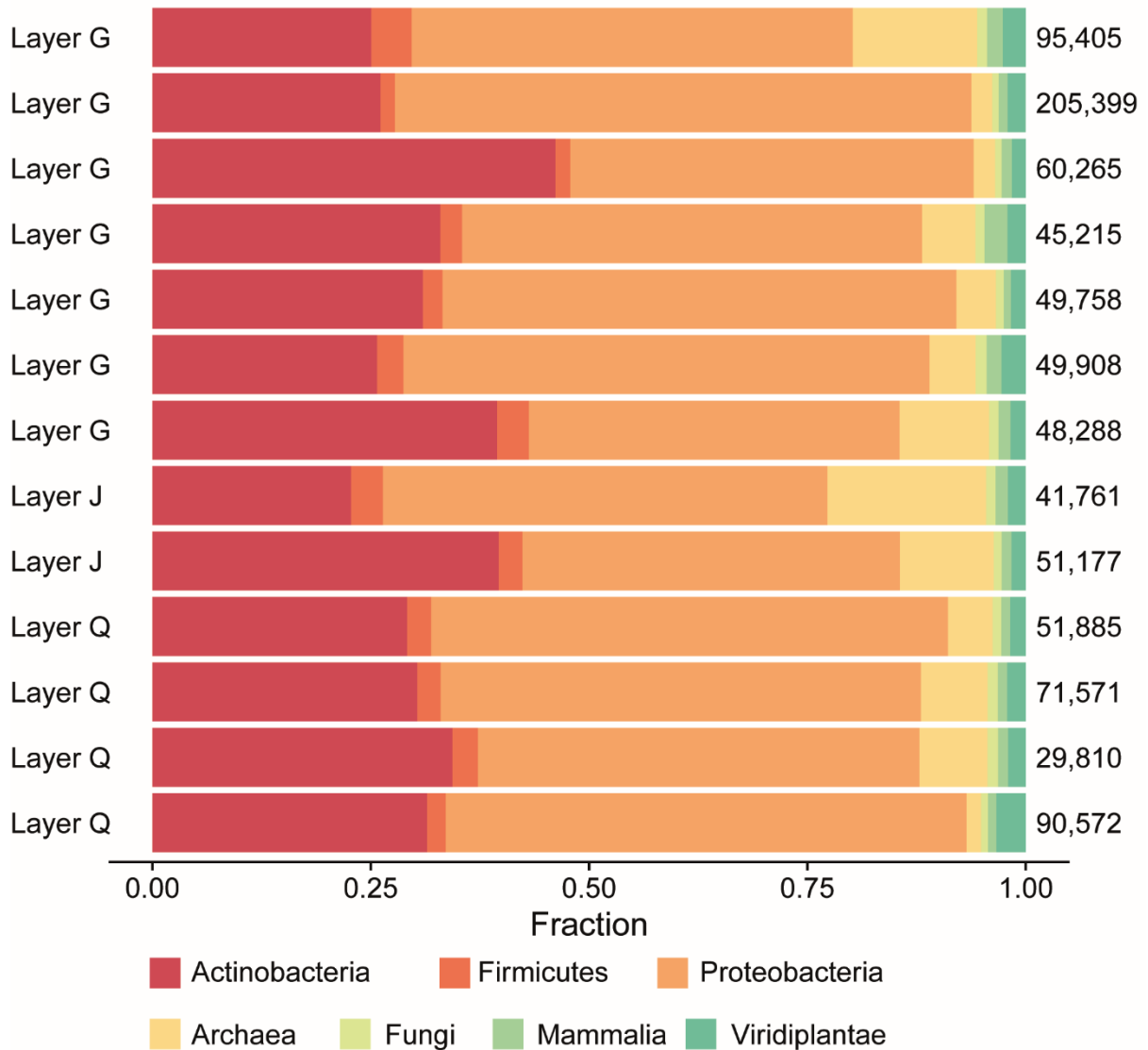
**Fig. S5. Comparison of different extraction methods on lime.** s - supernatant ; w I, II, III – EBT washes I, II, III; e – extract ; purification controls: c1 - Ringer solution with calcium; c2 – EBT buffer; c3 – sodium phosphate buffer; c4 – EDTA; c5 – “Bulat” buffer; c6 - buffer of the PowerMax Soil DNA Isolation Kit.



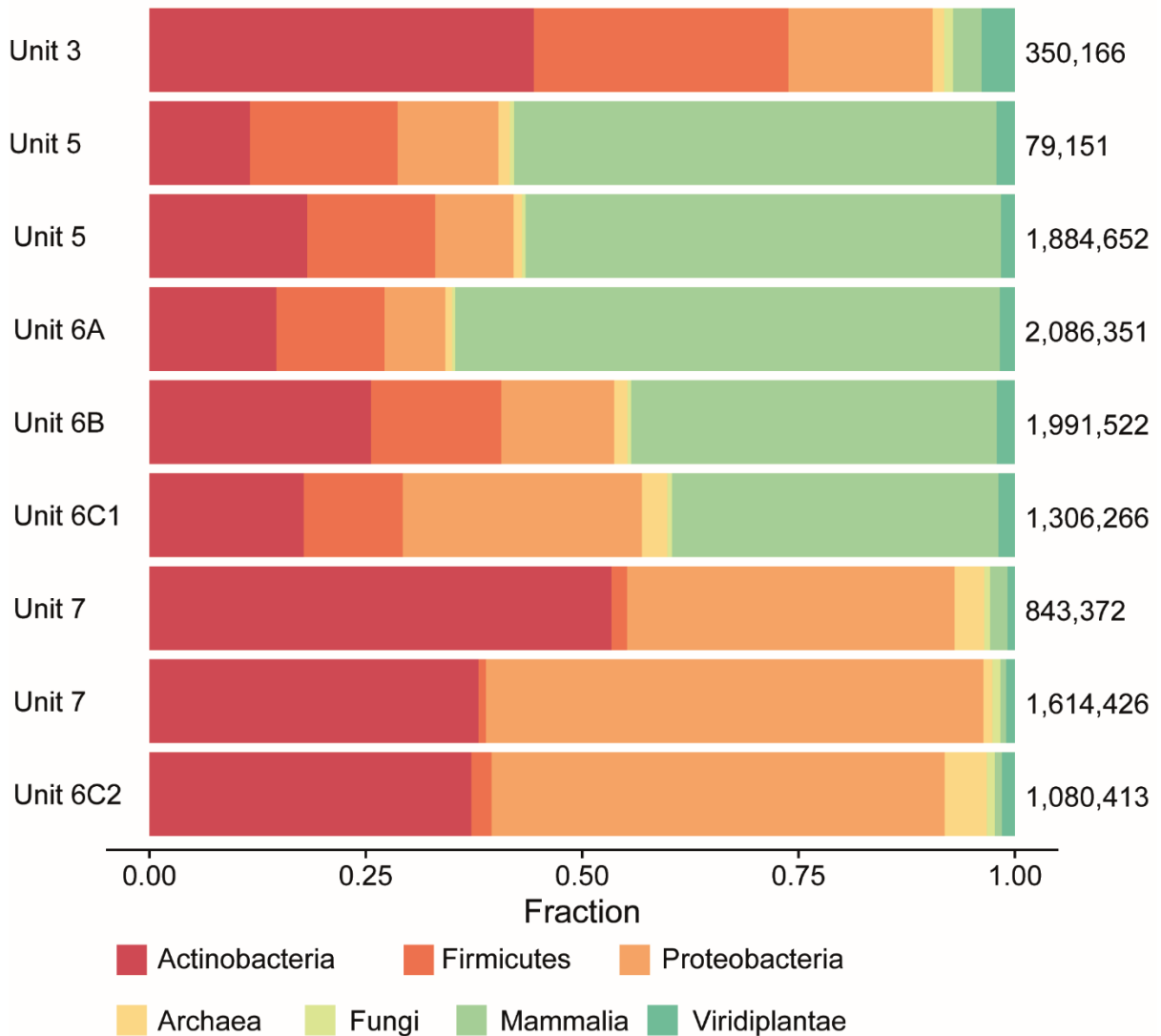
**Fig. S6. Overview of the methods applied to archaeological sediment samples.** Thin lines indicate that the procedure was carried out only once during the first screening of samples. Dashed arrows indicate the repetition of steps carried out on selected samples showing positive results (see Fig. S22).



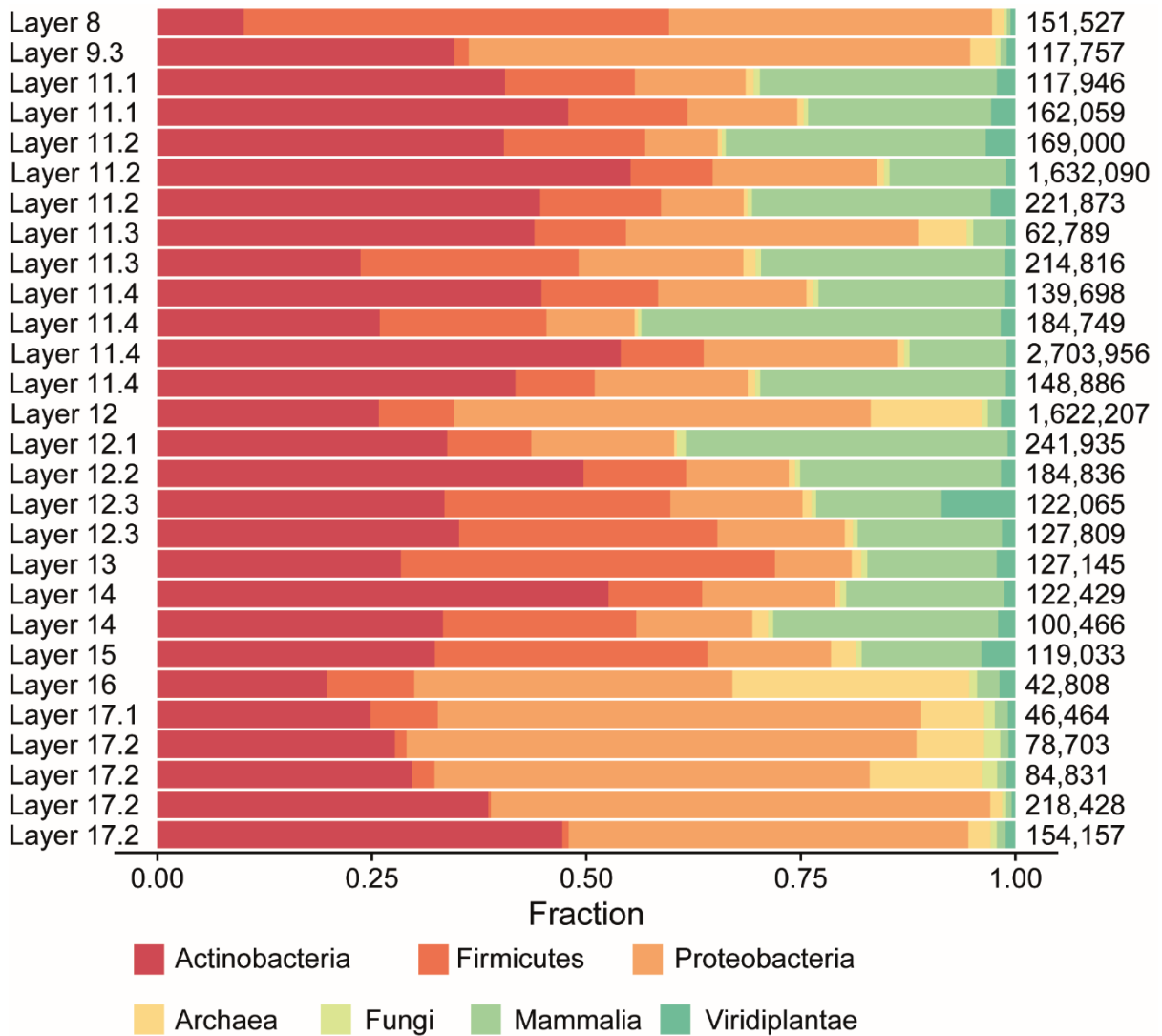
**Fig. S7. Taxonomical assignments of shotgun sequences from 7 archaeological sites (85 libraries) aligned to the full non-redundant NCBI nucleotide collection.** The upper bar shows the proportion of sequence counts assigned to 7 different taxonomic levels (using the lowest common ancestor approach), as well as sequences with no hits in the database. The zoom-in shows the distribution of counts only for sequences that aligned to any of the 7 different taxonomic levels.



**Fig. S8. Taxonomical assignments of shotgun sequences from Caune de l'Arago (samples 1-13).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library (sorted from top to bottom by archaeological layer). The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in this archaeological site varies from 3.9% to 10.2%.

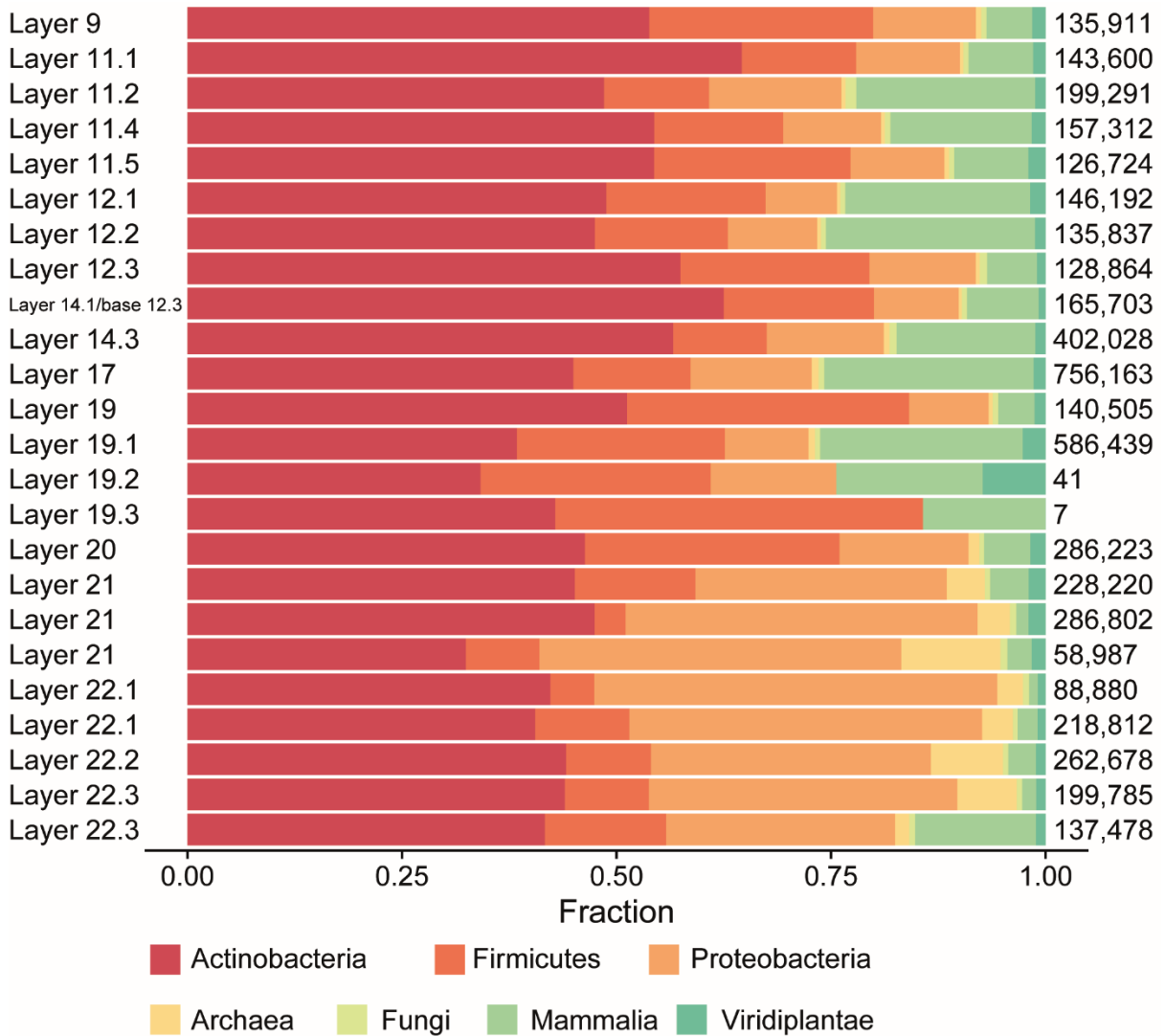


**Fig. S9. Taxonomical assignments of shotgun sequences from Chagyrskaya Cave (samples 14-22).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library (sorted from top to bottom by archaeological layer). The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in this archaeological site varies from 7.4% to 16%.

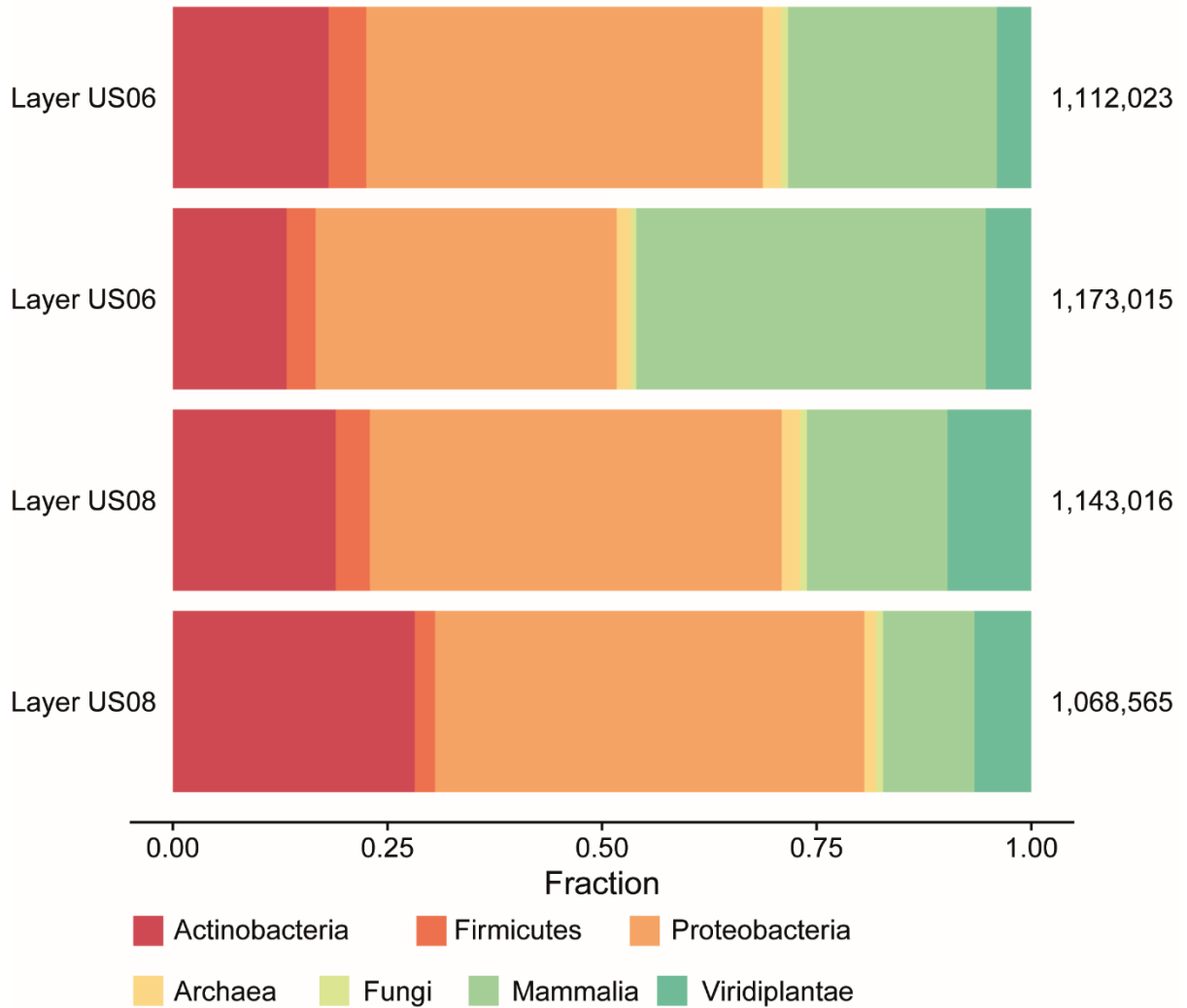


**Fig. S10. Taxonomical assignments of shotgun sequences from the East Gallery of Denisova Cave (samples 23-50).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library (sorted from top to bottom by archaeological layer). The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in this archaeological site varies from 4.9% to 21%.

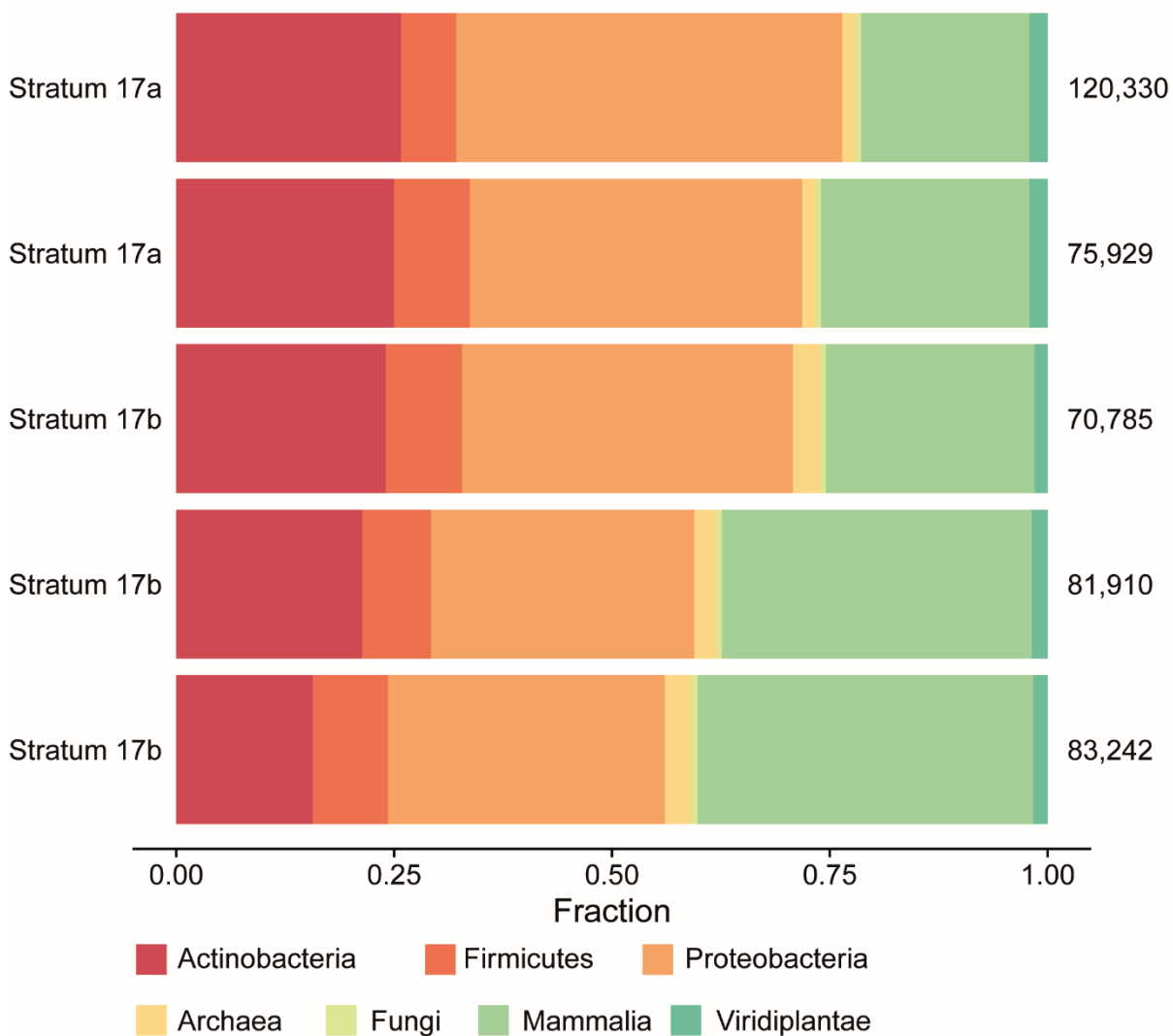




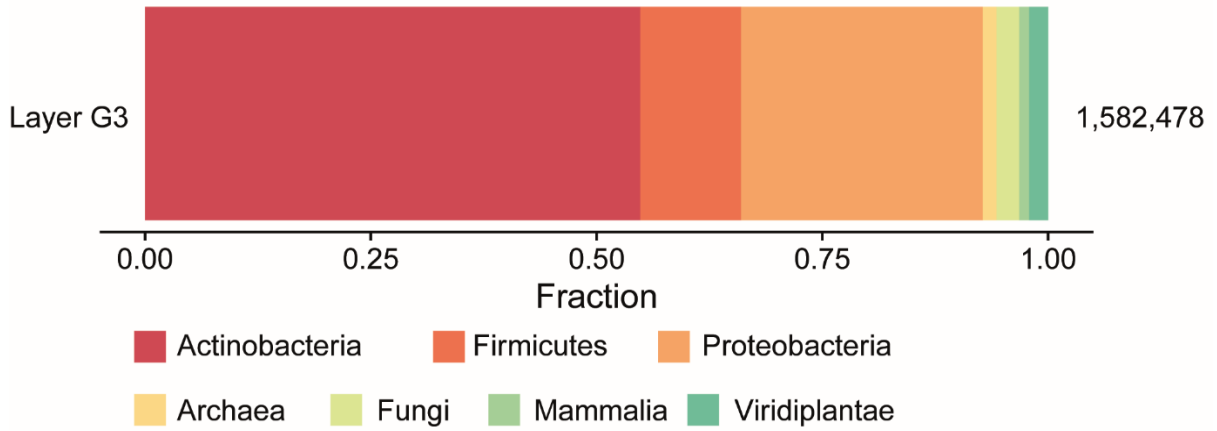
**Fig. S11. Taxonomical assignments of shotgun sequences from the Main Gallery of Denisova Cave (samples 51-74).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library (sorted from top to bottom by archaeological layer). The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in this archaeological site varies from 6.1% to 16.5%.



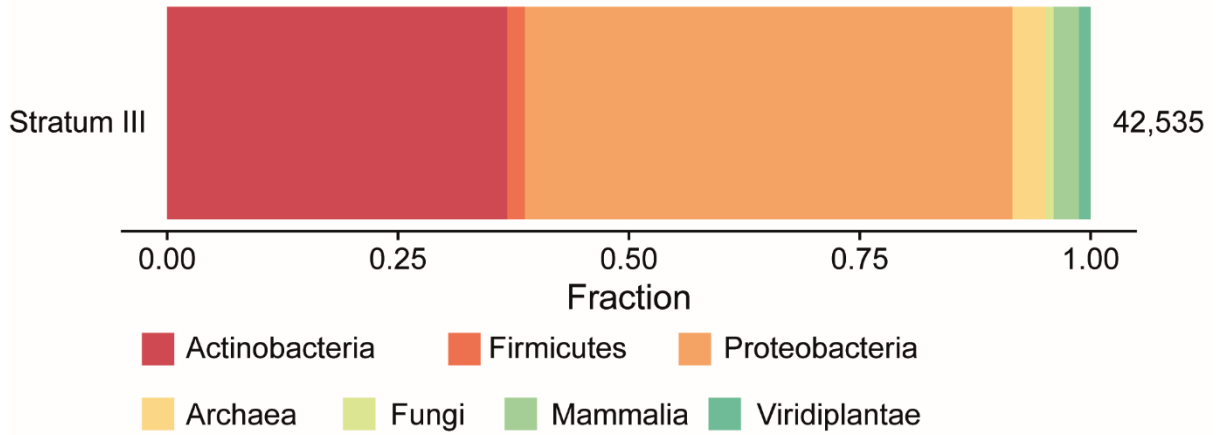
**Fig. S12. Taxonomical assignments of shotgun sequences from Les Cottés (samples 75-78).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library (sorted from top to bottom by archaeological layer). The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in this archaeological site varies from 7.9% to 10.2%.



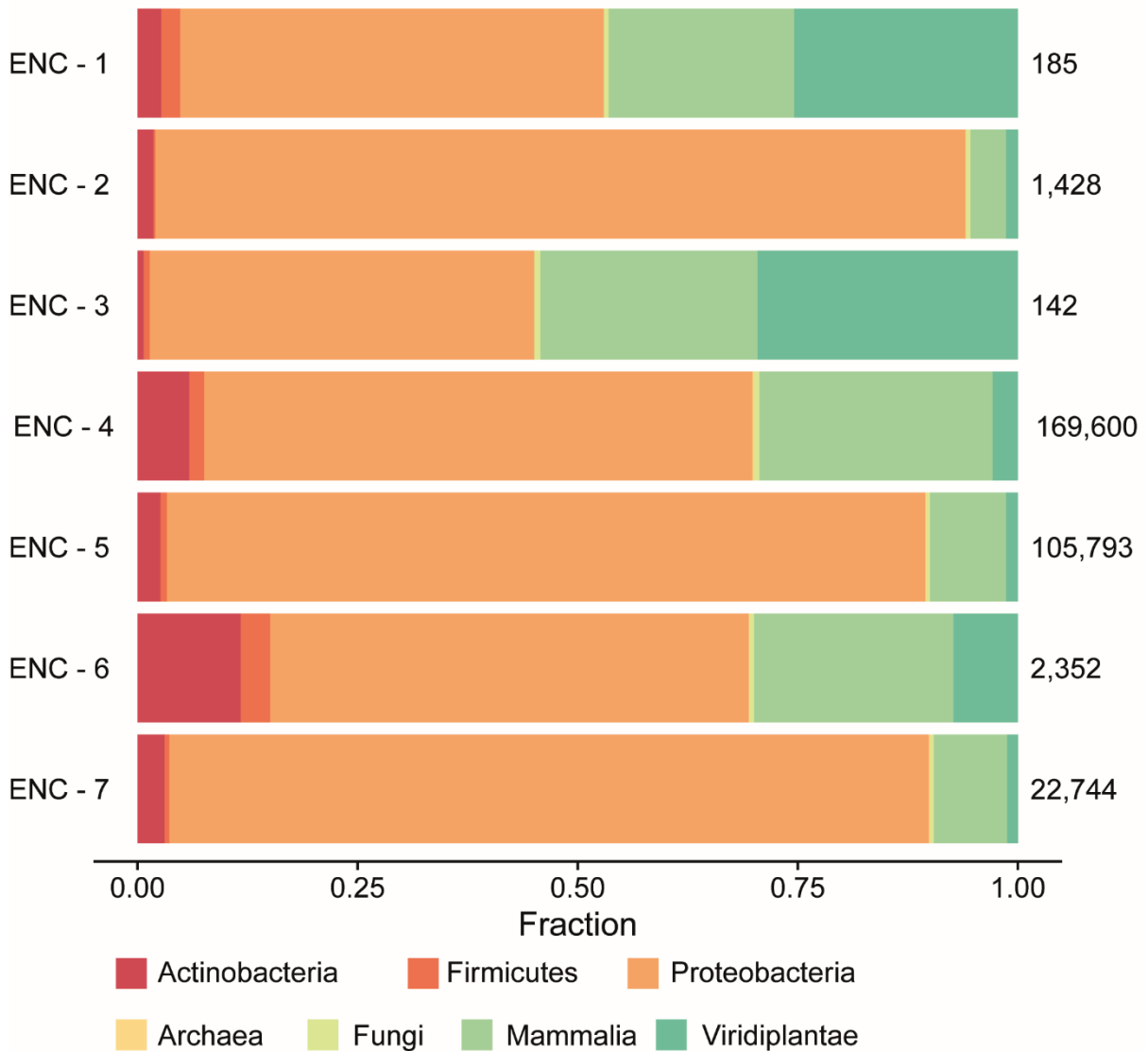
**Fig. S13. Taxonomical assignments of shotgun sequences from Trou Al'Wesse (samples 79-83).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library (sorted from top to bottom by archaeological layer). The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in this archaeological site varies from 6.9% to 9.2%.



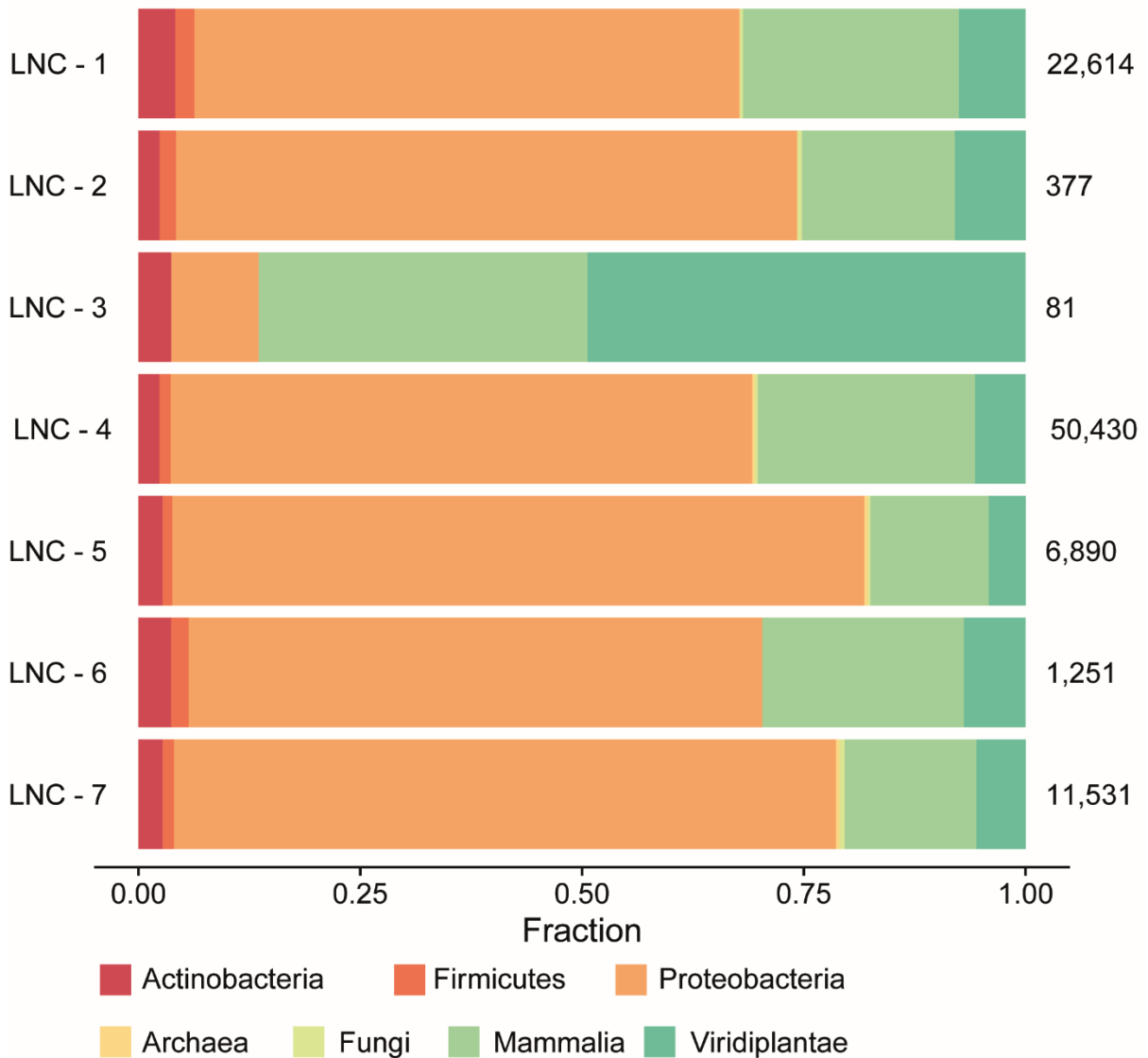
**Fig. S14. Taxonomical assignments of shotgun sequences from Vindija Cave (sample 84).** The bar shows the relative fraction of sequences with alignments to 7 different taxonomic levels. The number on the right corresponds to the number of sequences with hits to the database. The percentage of sequences aligned to the database in the library from this archaeological site was 12.5%.



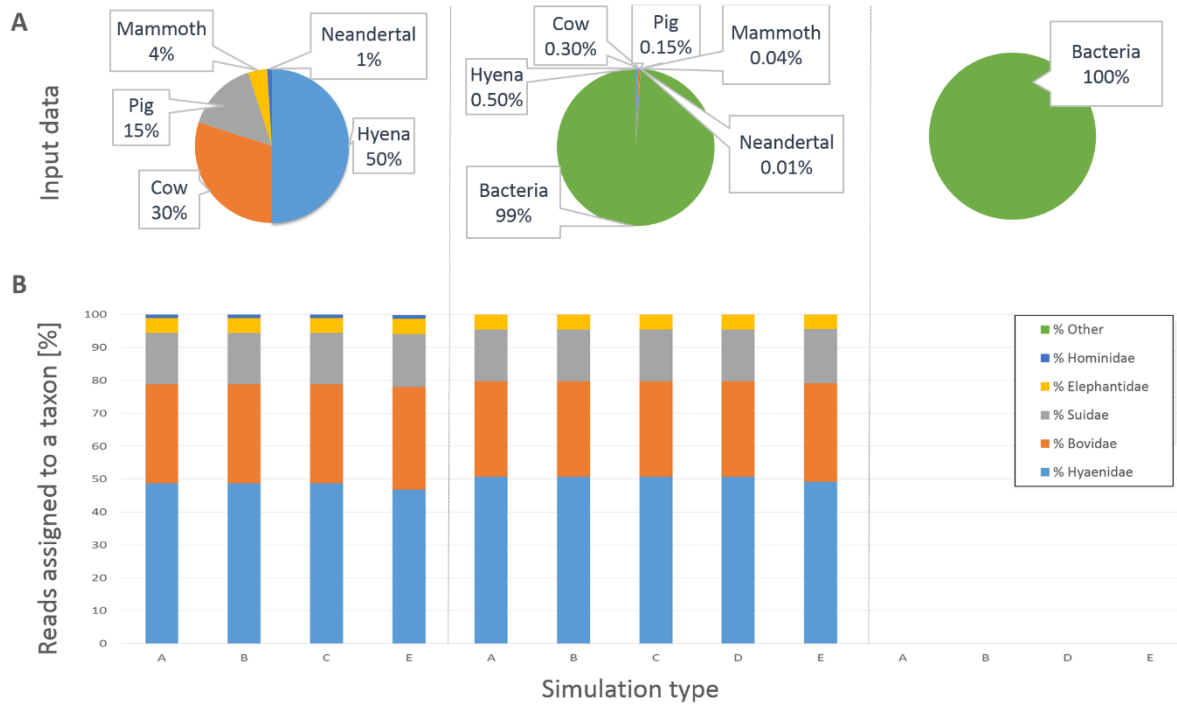
**Fig. S15. Taxonomical assignments of shotgun sequences from El Sidrón (sample 85).** The bar shows the relative fraction of sequences with alignments to 7 different taxonomic levels. The number on the right corresponds to the number of sequences with hits to the database. The percentage of sequences aligned to the database in the library from this archaeological site was 5.7%.



**Fig. S16. Taxonomical assignments of shotgun sequences from extraction negative controls (ENC; samples 86-92).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library prepared from an extraction blank. The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in these control libraries varies from 13% to 32%. The library for ENC-8 (sample 93) yielded no sequences.

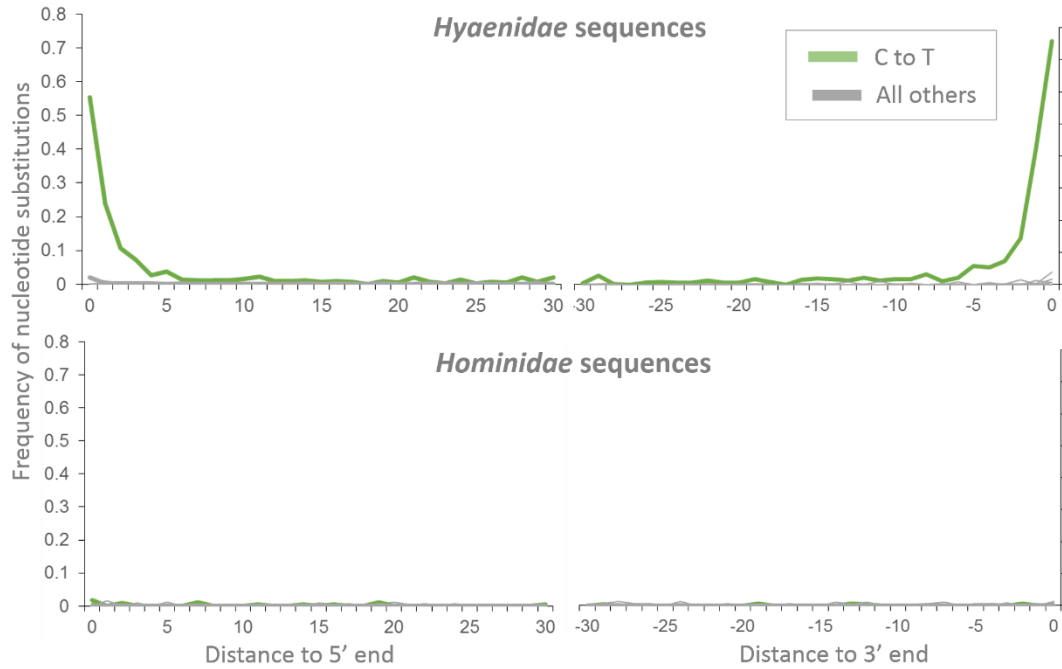


**Fig. S17. Taxonomical assignments of shotgun sequences from library preparation negative controls (LNC; samples 94-100).** The bars show the relative fraction of sequences with alignments to 7 different taxonomic levels. Each bar corresponds to a library preparation negative control. The numbers on the right correspond to the number of sequences with hits to the database for each library. The percentage of sequences aligned to the database in these control libraries varies from 10.8% to 25.5%. The library for LNC-8 (sample 101) yielded no sequences.

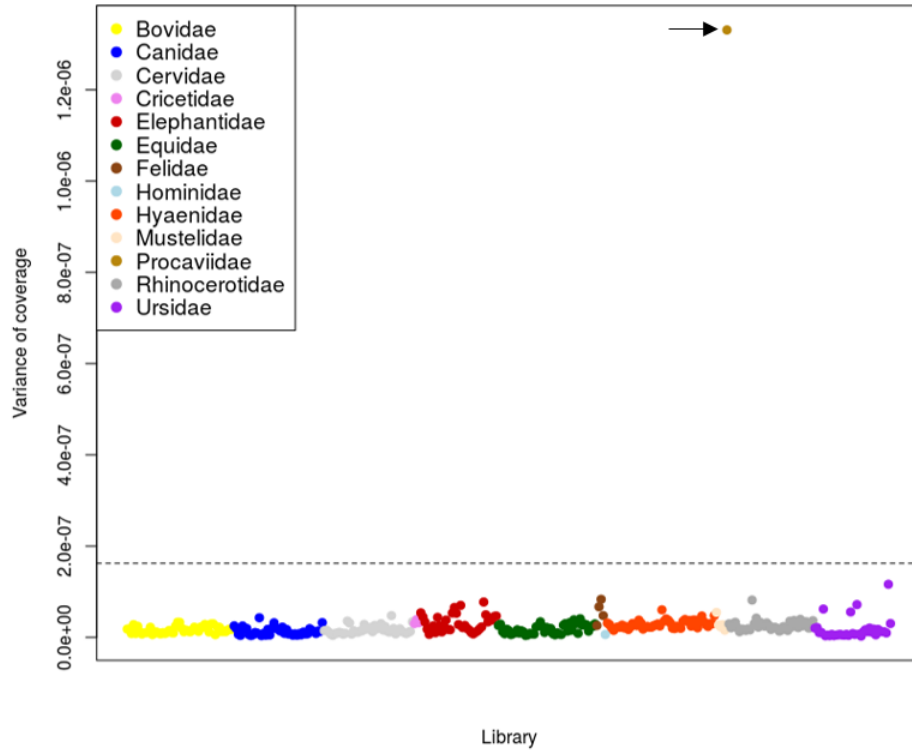


**Fig. S18. Taxa identification in simulated datasets.** (A) Three input datasets were utilized to test the accuracy of taxa identification. Each set was comprised of 100,000 sequences between 35 and 100 base pairs in length, and composed of a mixture of mammalian mtDNA and/or bacterial DNA. (B) The percentage of sequences assigned to each family is shown, in the absence of cytosine (C) to thymine (T) substitutions (simulation A), with 50% of terminal Cs converted to Ts in some or all taxa (simulations B-D), and with additionally 10% of Cs at any position converted to Ts (simulation E).

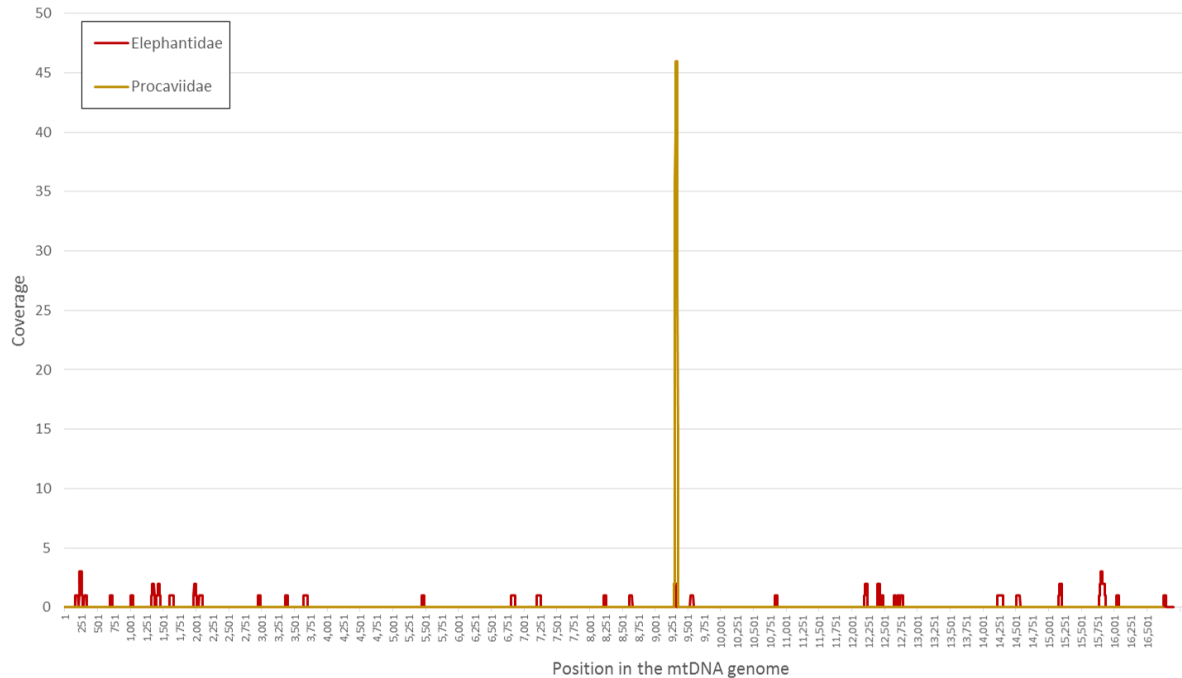




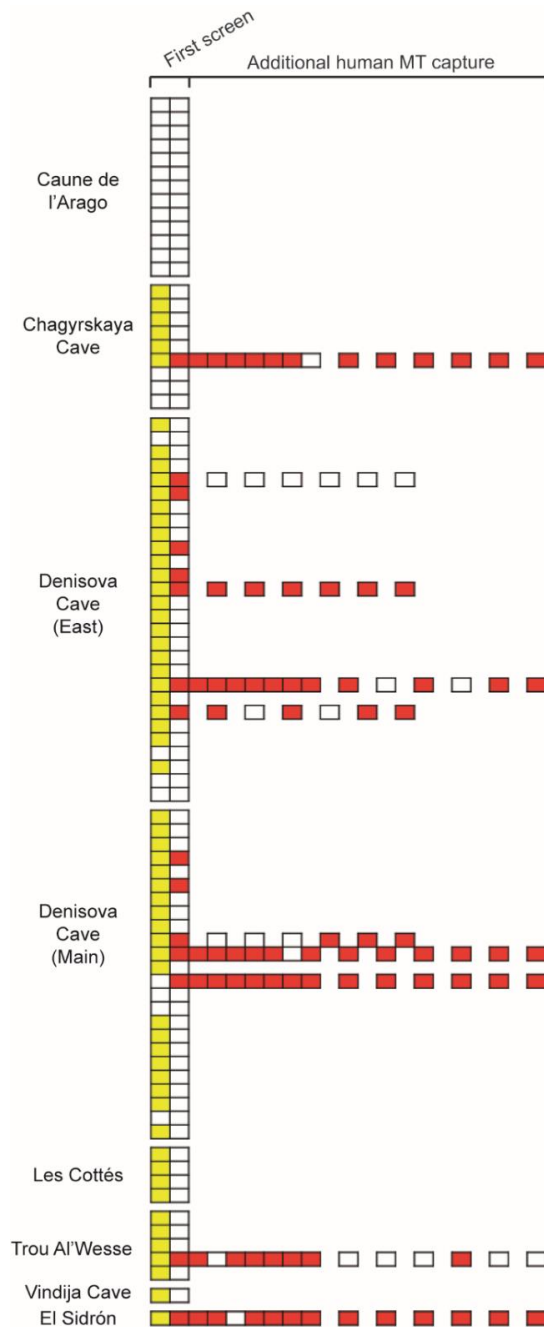
**Fig. S19. Nucleotide substitutions in Hyaenidae (top) and Hominidae sequences (bottom) obtained from the same library (library R4071).** While the hyena sequences carry terminal cytosine (C) to thymine (T) substitutions (green) typical of ancient DNA, the hominin ones likely originate from contamination.



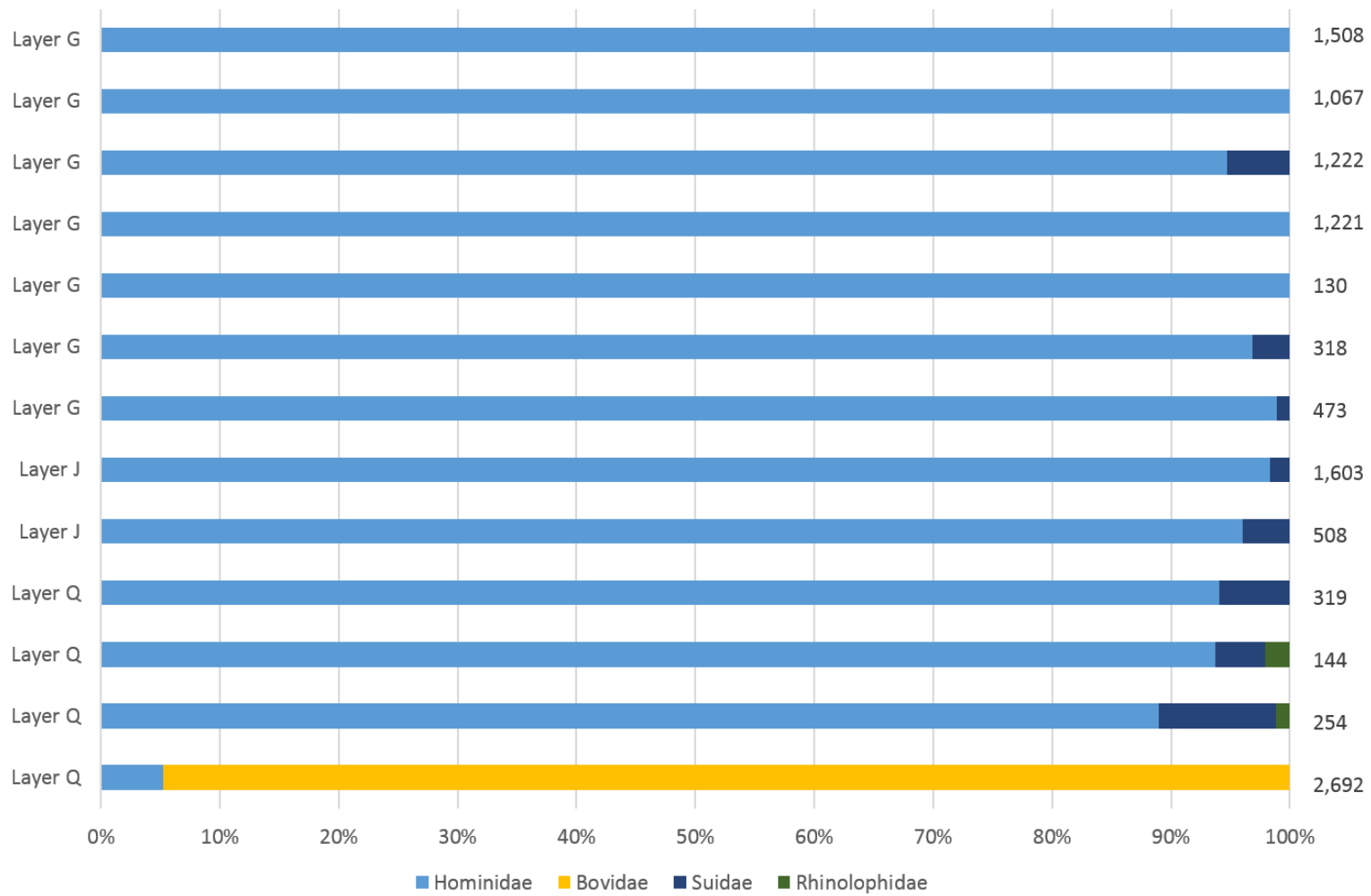
**Fig. S20. Variance of the mtDNA genome coverage in libraries containing sequences with elevated terminal damage-derived C to T substitutions, per taxa.** The dashed line represents one standard deviation from the mean. Sequences attributed to Procaviidae (arrow) have an unusually high variance, driven by the mapping of all sequences to a restricted region of the reference genome (see Fig. S21).



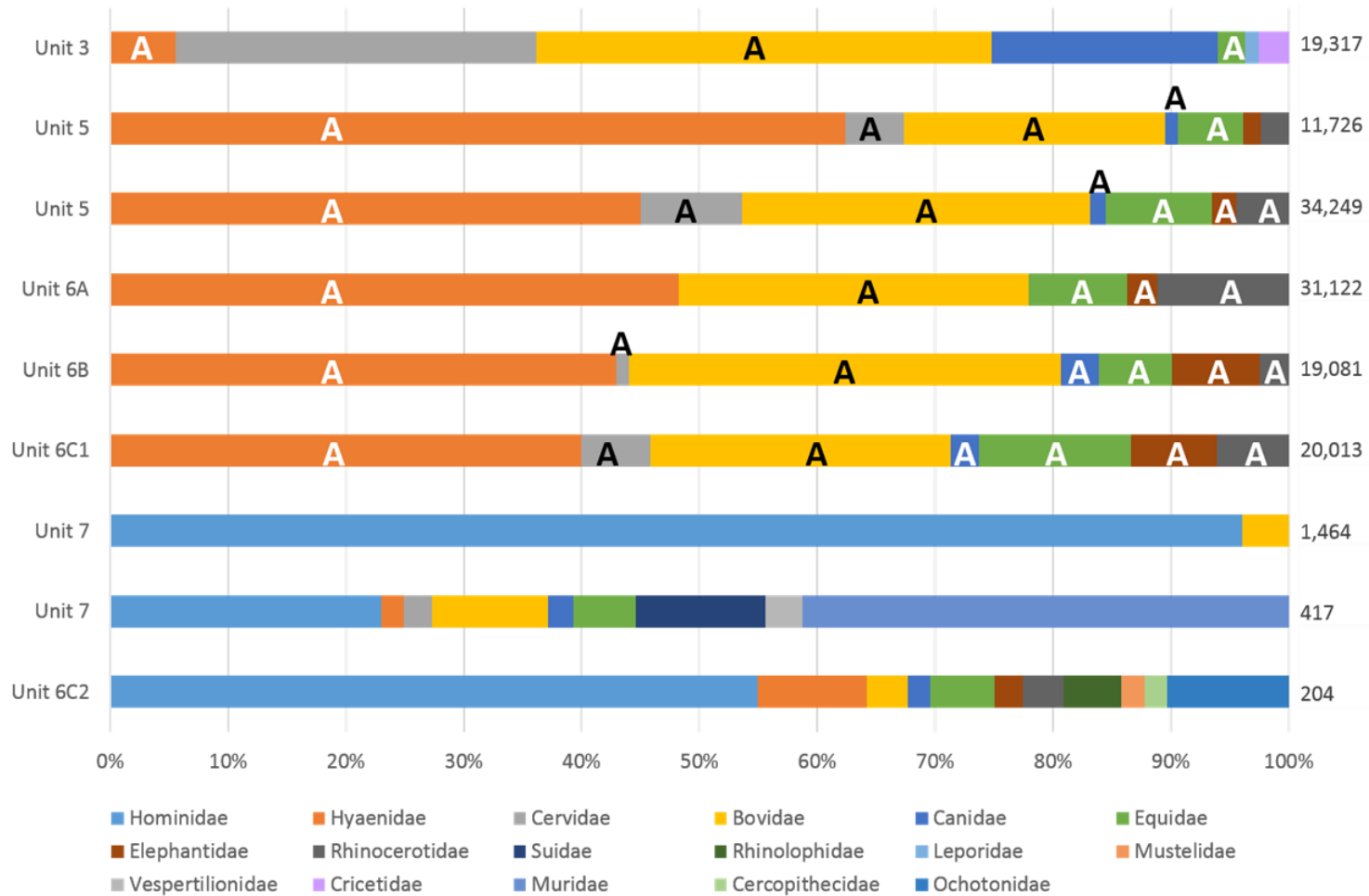
**Fig. S21. Number of sequences from library R4070 attributed to Elephantidae (red) and Procaviidae (gold) mapping to each position in their respective reference genome. 45 and 46 sequences were identified as originating from these families, yielding an average coverage of 0.12- and 0.11-fold, respectively. Note that while all Procaviidae sequences map to one restricted 60-bases long region in the genome, the Elephantidae ones are distributed across the reference genome.**



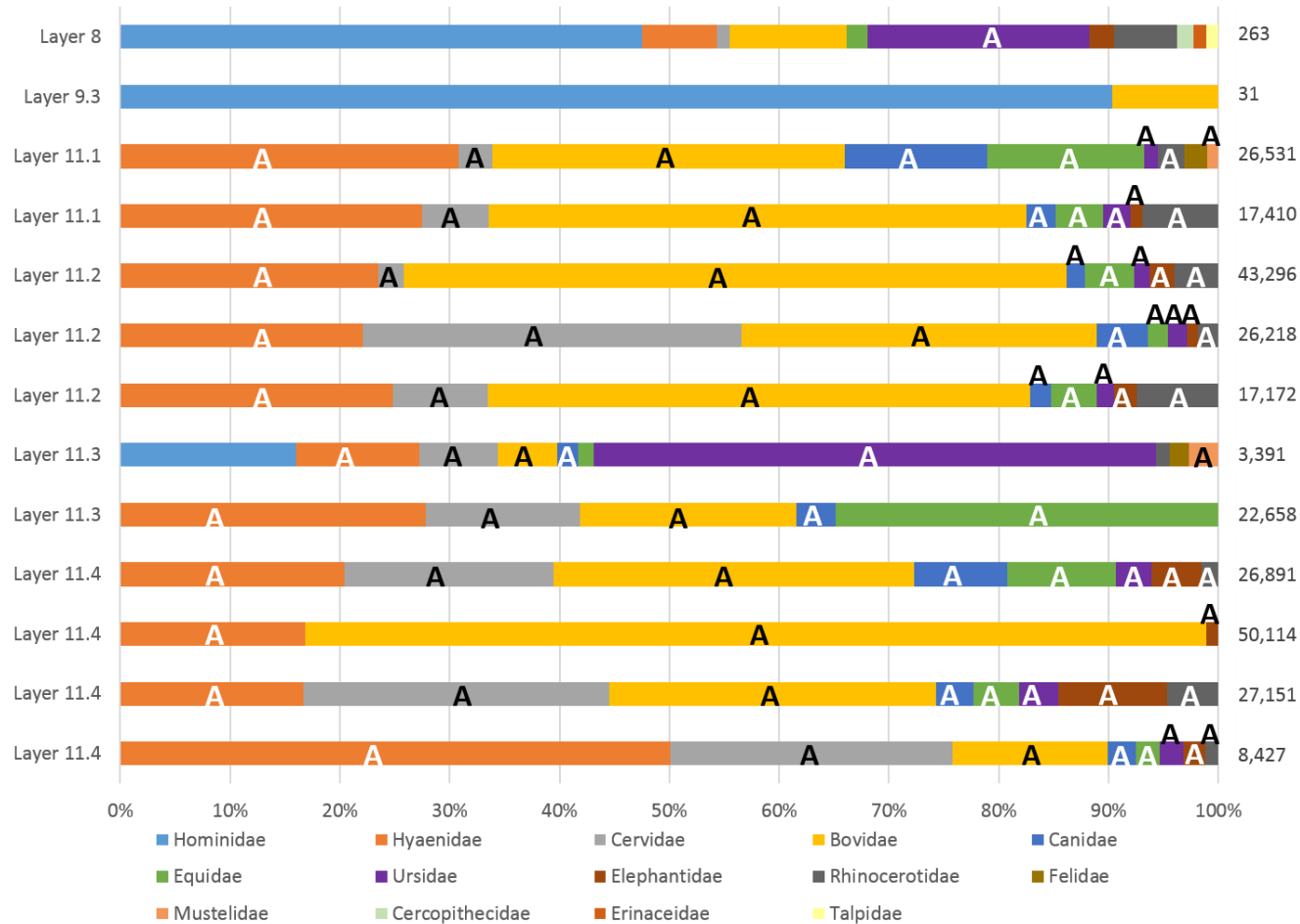
**Fig. S22. Preservation of ancient DNA in sediment samples. Each row represents one sample.** The 85 samples included in the study are organized from top to bottom along the stratigraphy of each site (see Data file S1). Libraries enriched for mammalian mtDNA are represented by the first column, where a yellow filling indicates that the library contains mammalian mtDNA sequences bearing substitutions typical of ancient DNA (all taxa combined, including hominins). All other columns represent libraries enriched for human mtDNA, where a red filling symbolizes the presence of ancient hominin mtDNA. Libraries prepared from the same extract are represented as one continuous block along the horizontal axis.



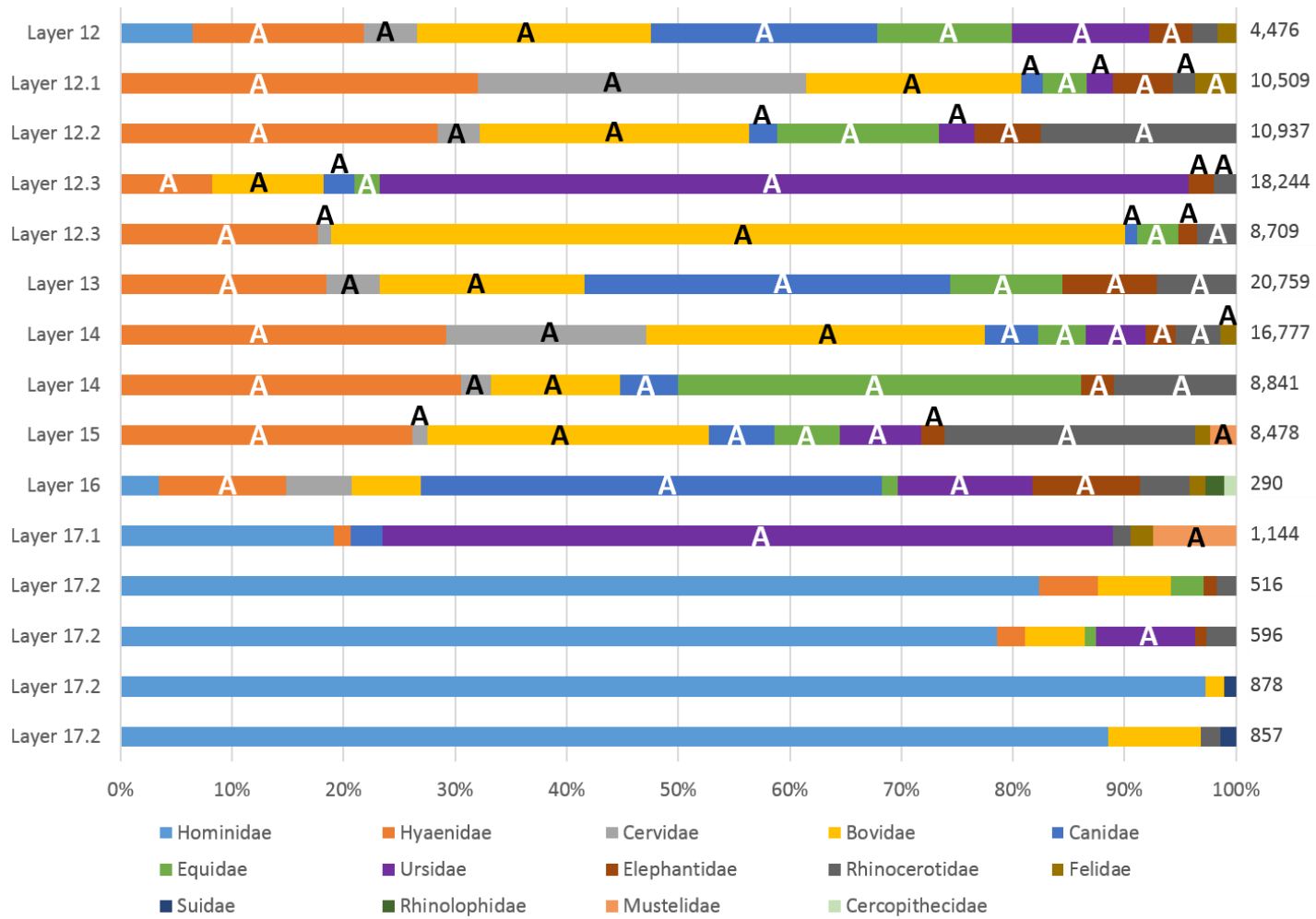
**Fig. S23. Taxa identification in sediment from Caune de l'Arago (samples 1-13).** The layers from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right.



**Fig. S24. Taxa identification in sediment from Chagyrskaya Cave (samples 14-22).** The stratigraphic units from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right. An 'A' indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences.

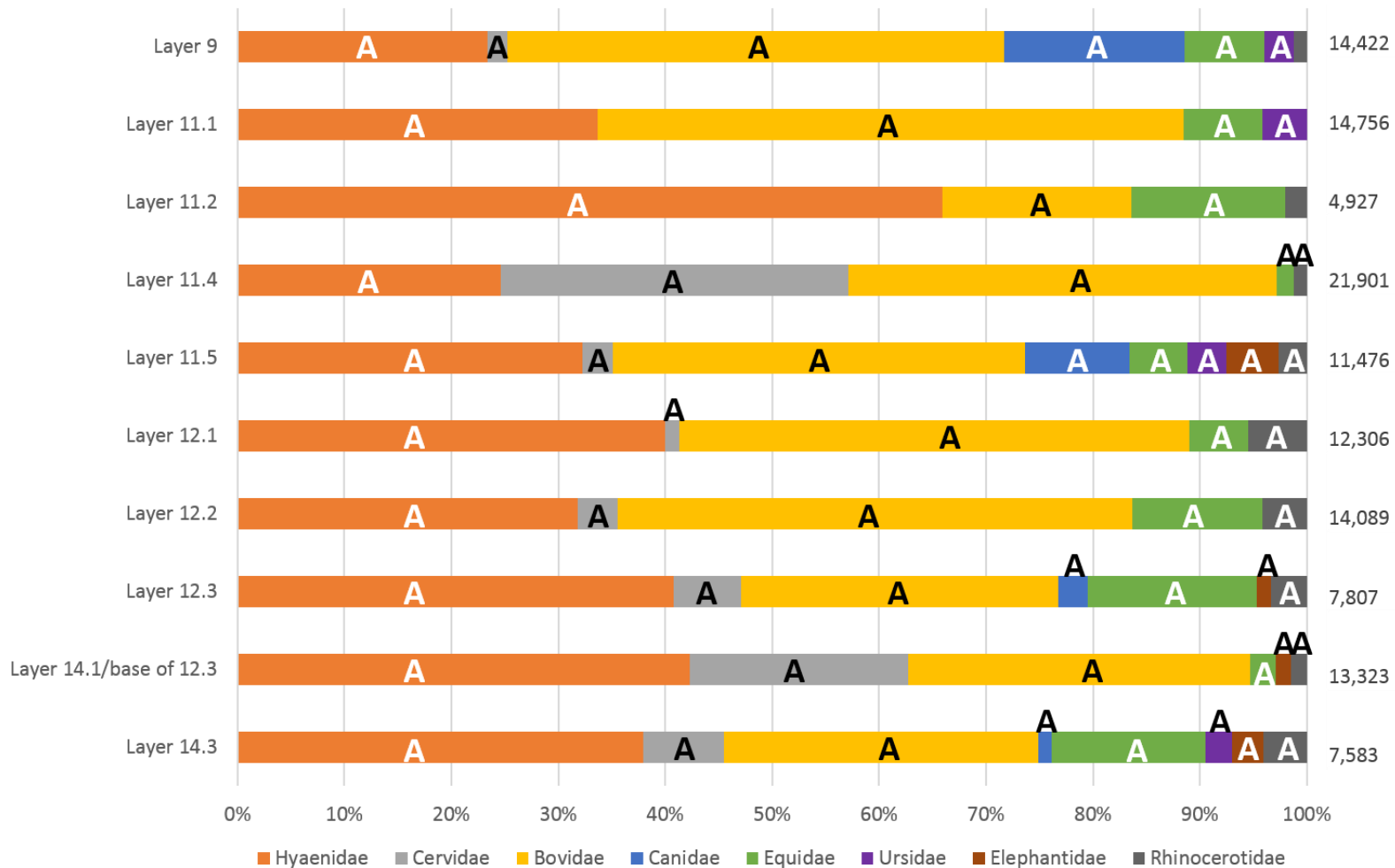


**Fig. S25. Taxa identification in sediment from the upper layers of the East Gallery of Denisova Cave (samples 23-35).** The layers from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right. An 'A' indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences.

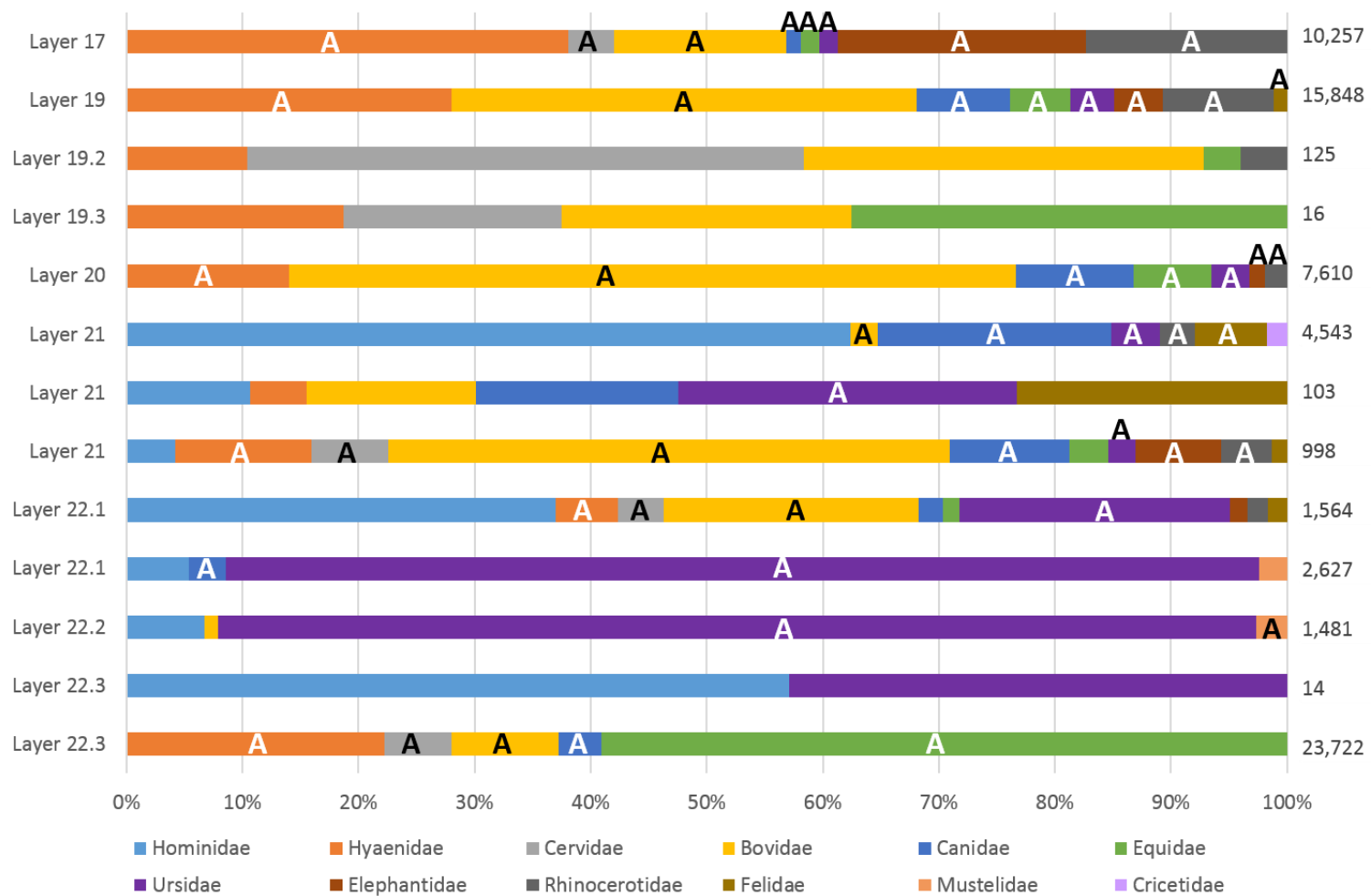


**Fig. S26. Taxa identification in sediment from the lower layers of the East Gallery of Denisova Cave (samples 36-50).** The layers from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right. An 'A' indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences.

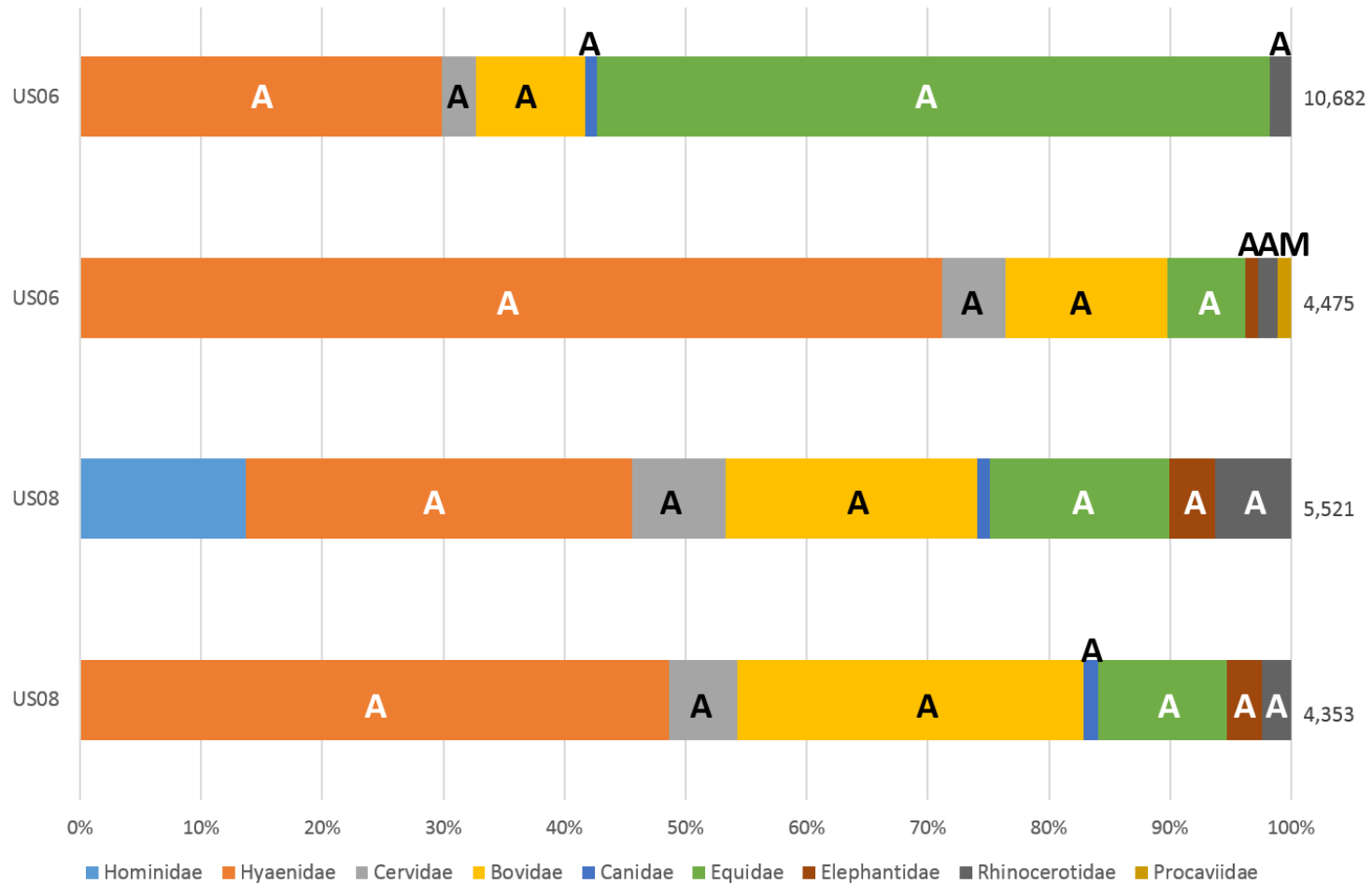




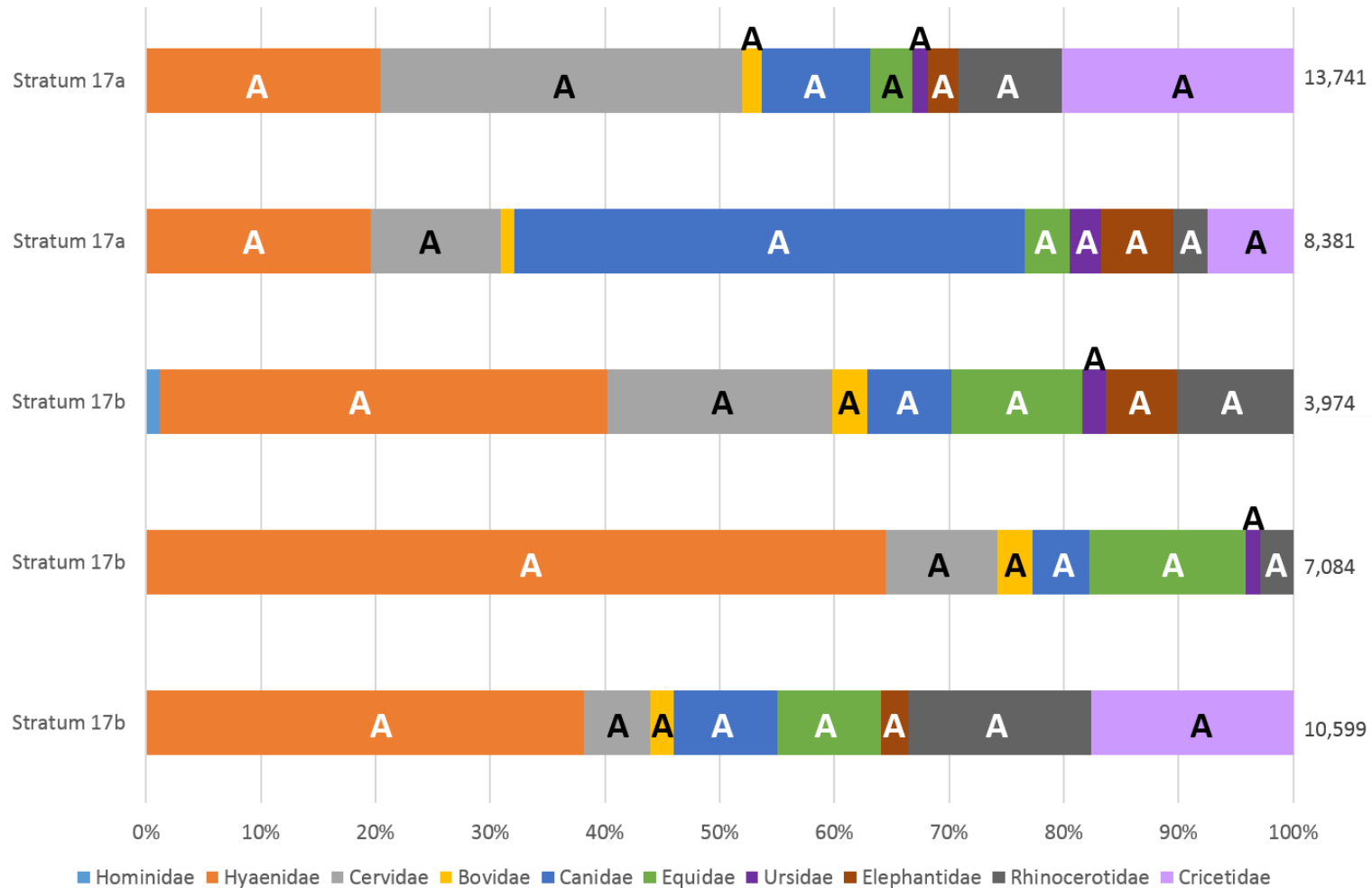
**Fig. S27. Taxa identification in sediment from the upper layers of the Main Gallery of Denisova Cave (samples 51-60).** The layers from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right. An 'A' indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences.



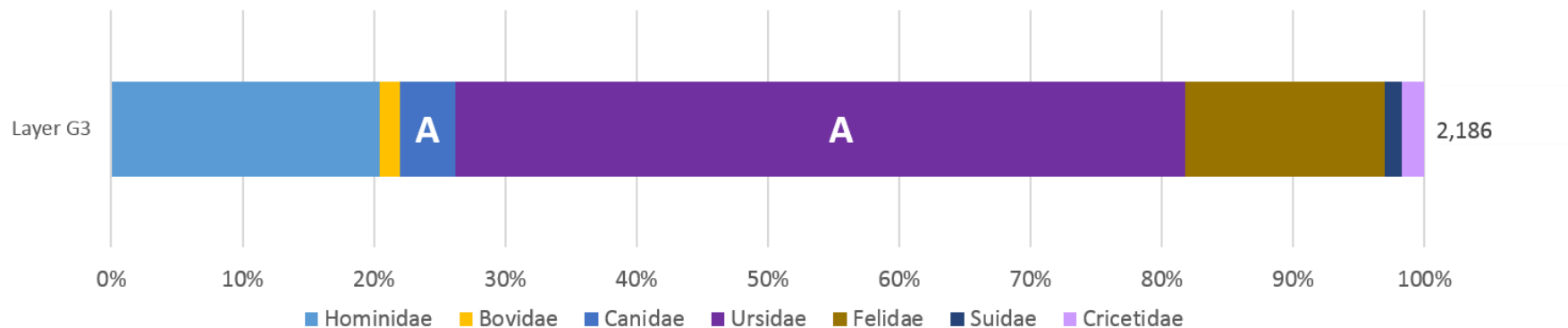
**Fig. S28. Taxa identification in sediment from the lower layers of the Main Gallery of Denisova Cave (samples 61-62, 64-74).** The layers from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right. An 'A' indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences. The library for sample 63 contained no sequences longer than 35bp.



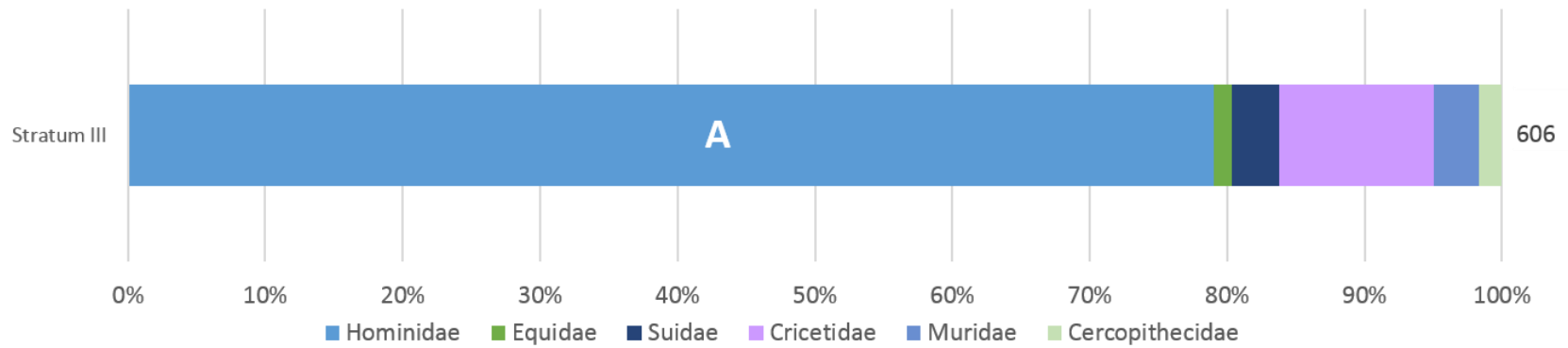
**Fig. S29. Taxa identification in sediment from Les Cottés (samples 75-78).** The layers from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right. An ‘A’ indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences. The ‘M’ indicates that sequences only mapped to a single 60-bases long region in the reference genome, suggesting that they originate from another organism and were misassigned to Procaviidae (see Figs. S20, S21).



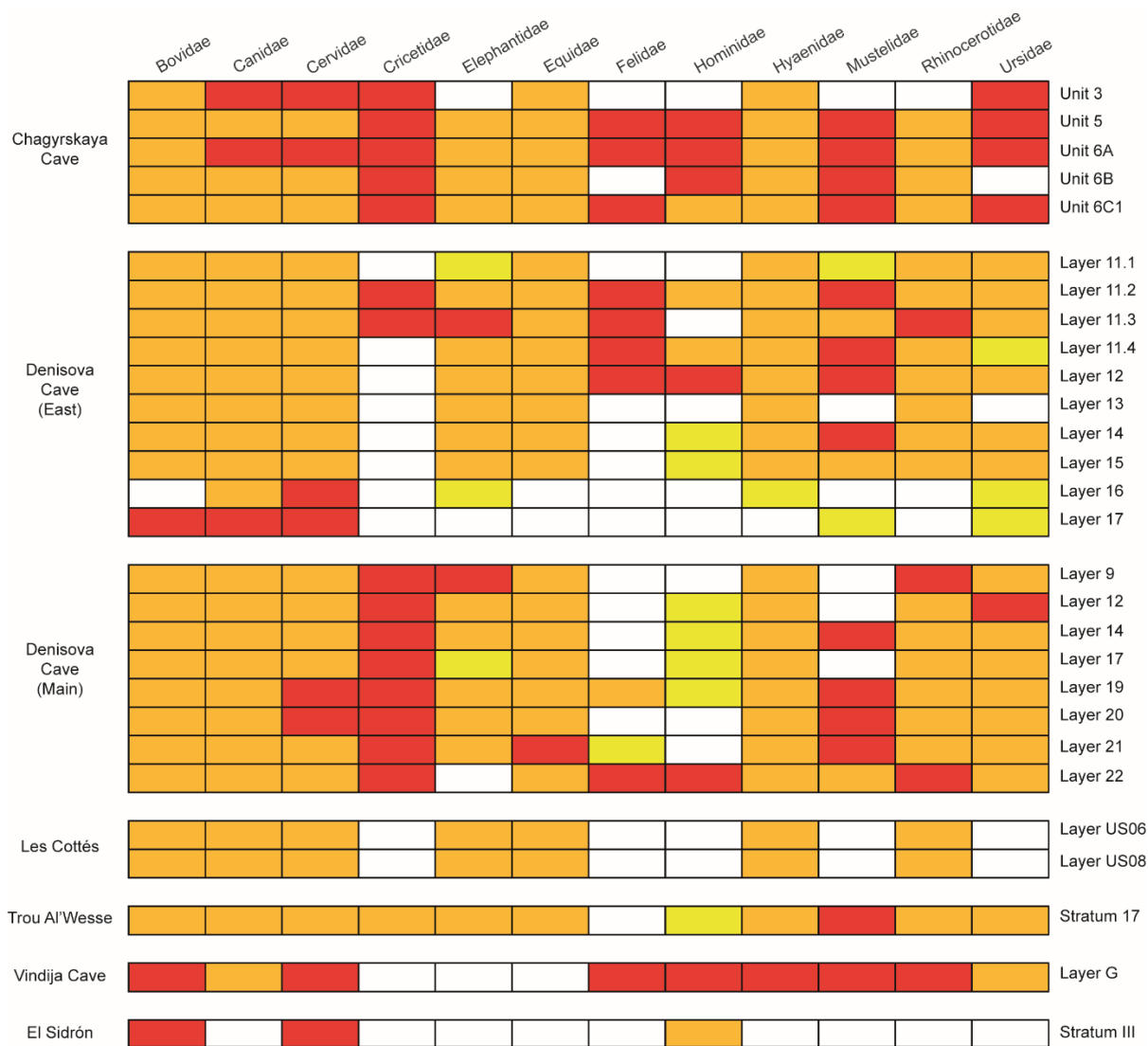
**Fig. S30. Taxa identification in sediment from Trou Al'Wesse (samples 79-83).** The stratigraphic units from which the samples are taken are noted on the left, the number of sequences identified as mammalian mtDNA in the libraries are marked on the right. An 'A' indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences.



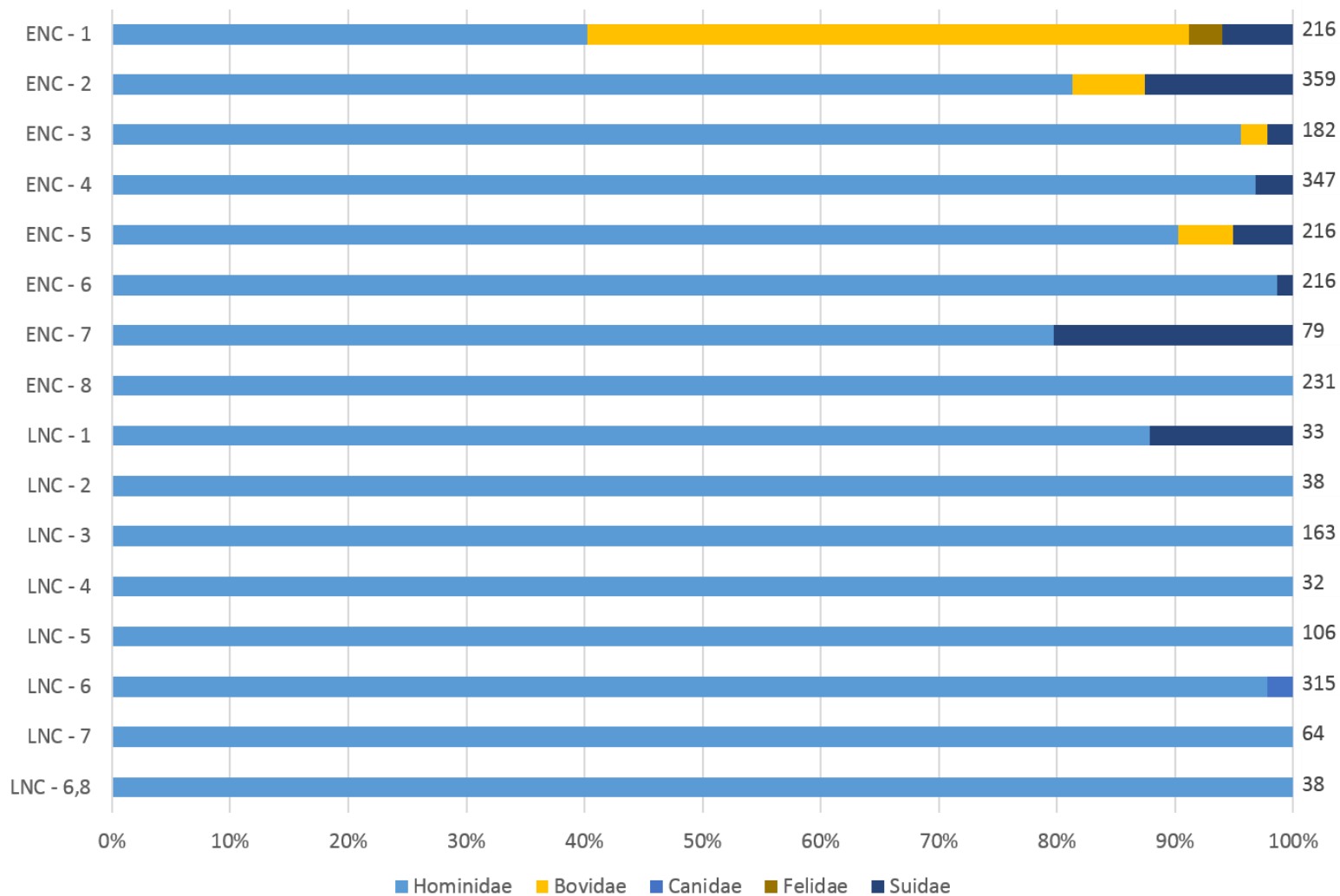
**Fig. S31. Taxa identification in sediment from Vindija Cave (sample 84).** The layer from which the sample was taken is noted on the left, the number of sequences identified as mammalian mtDNA in the library is marked on the right. An ‘A’ indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences.



**Fig. S32. Taxa identification in sediment from El Sidrón (sample 85).** The stratigraphic unit from which the sample was taken is noted on the left, the number of sequences identified as mammalian mtDNA in the library is marked on the right. An ‘A’ indicates that the taxon was identified as ancient in origin based on the presence of damage-derived substitutions at the ends of sequences.

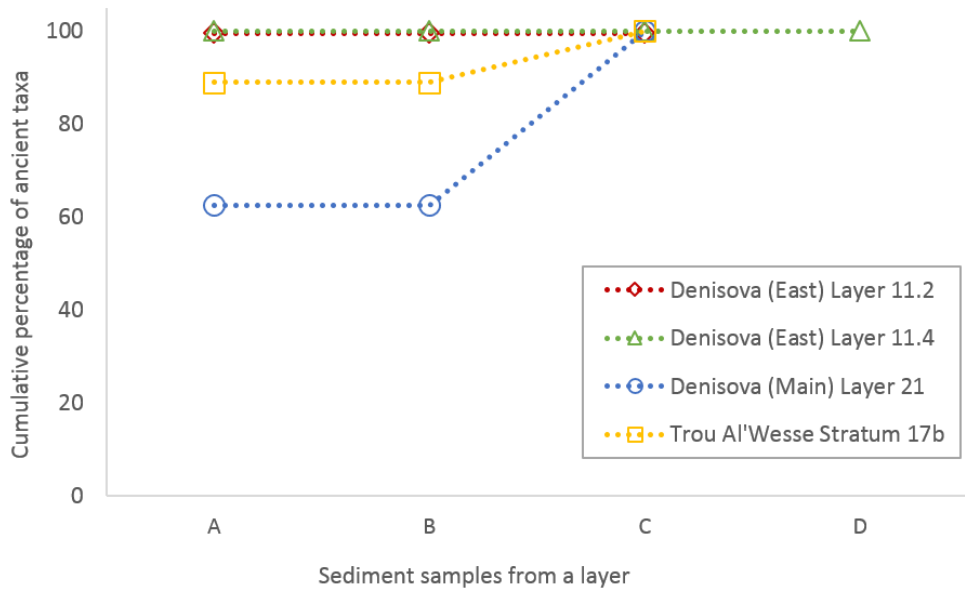


**Fig. S33. Comparing the taxonomic composition as determined by DNA retrieved from sediments to the zooarchaeological record.** Each row represents an archaeological layer, each column one of the 12 mammalian families from which ancient DNA sequences were detected at any of the sites (Fig. 2). A yellow filling indicates that ancient mtDNA fragments attributed to that family were recovered from the sediment sample(s); a red filling symbolizes the presence of relevant taxa in the fossil record; and an orange filling shows that the family was identified by both approaches. Other taxa identified based on their skeletal remains and which were not represented by DNA fragments in any of our samples are not shown. Archaeological layers from which no ancient DNA fragments were retrieved or for which no zooarchaeological analysis was available were disregarded, and archaeological sub-layers were combined where necessary to allow for comparisons with the literature. Comparative data was taken from (32, 43, 45, 57, 60, 67-69, 76, 78, 82, 102, 105, 155-157).

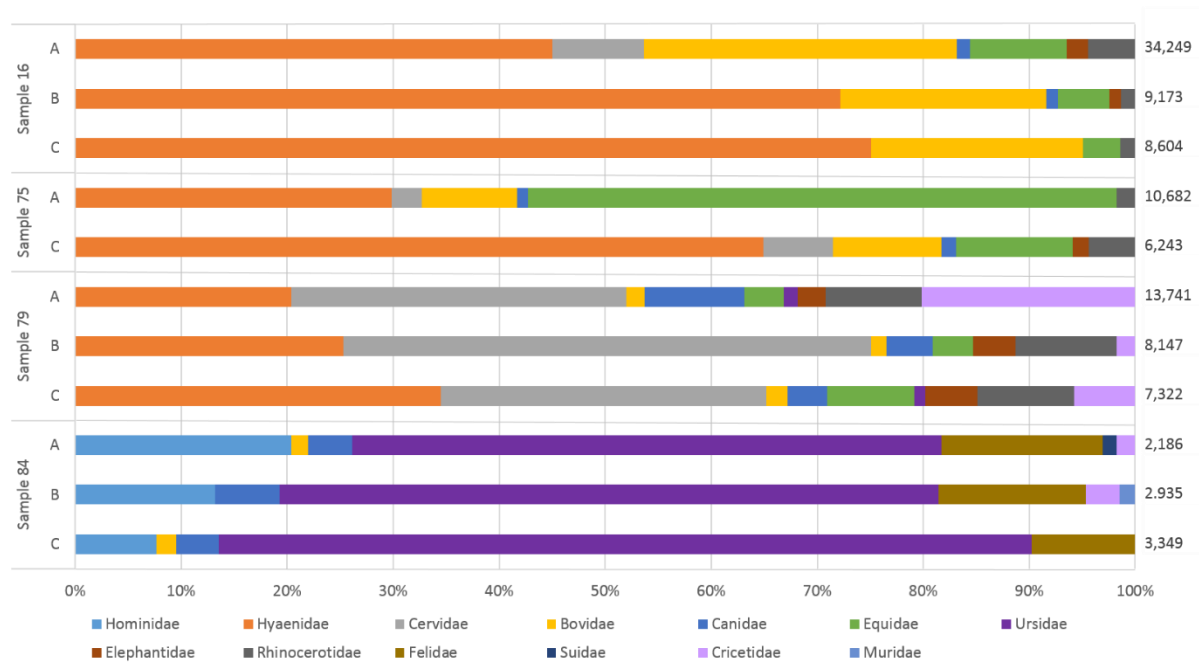


**Fig. S34. Taxa identification in extraction negative controls (ENC; samples 86-93) and library preparation negative controls (LNC; samples 94-101).** The number of sequences identified as mammalian mtDNA in the libraries are marked on the right.

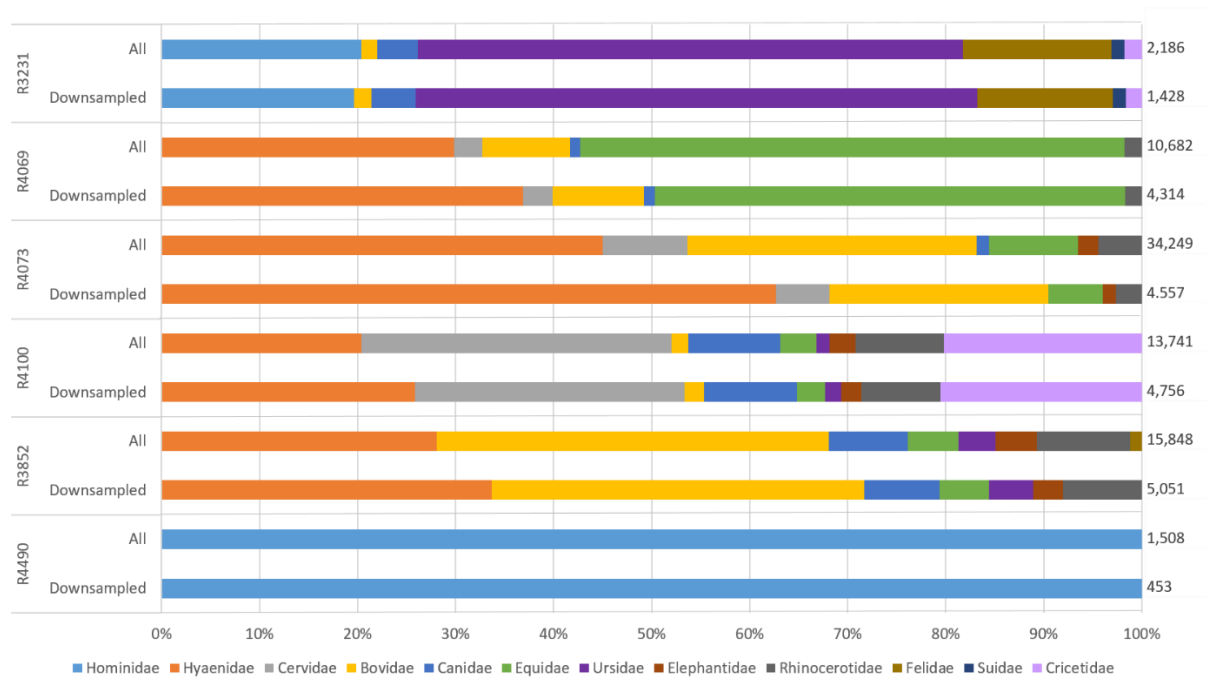




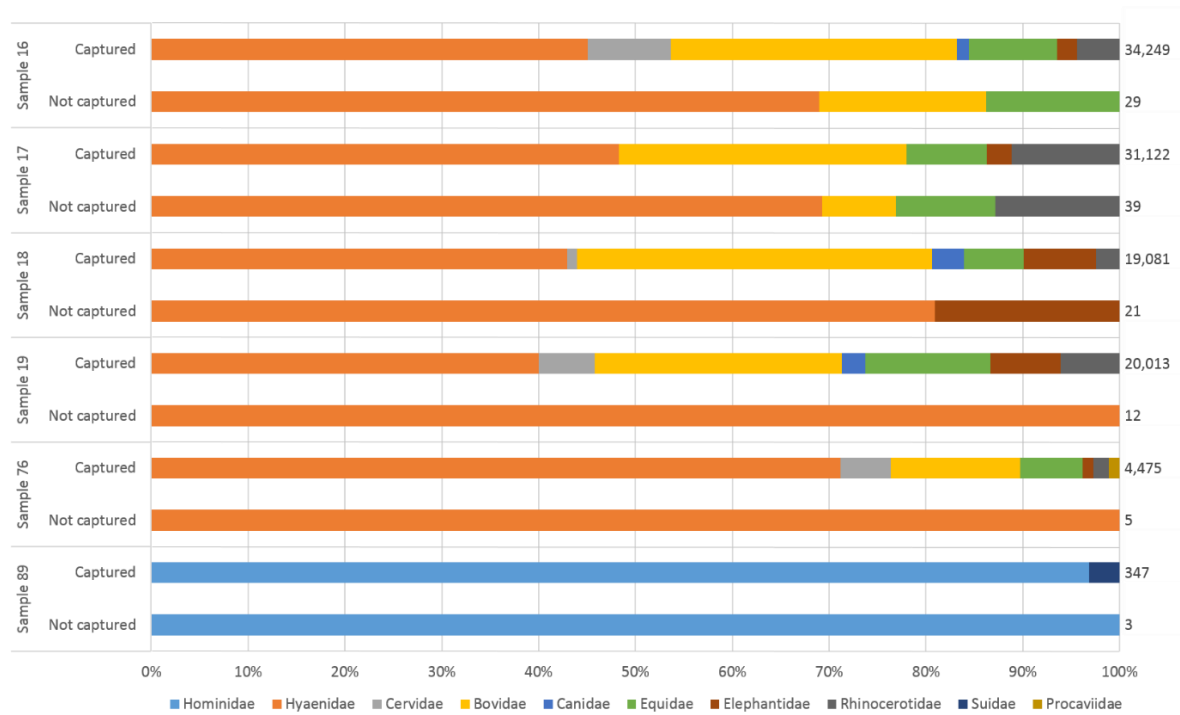
**Fig. S35. Cumulative percentage of ancient taxa identified in multiple samples from a given layer.**



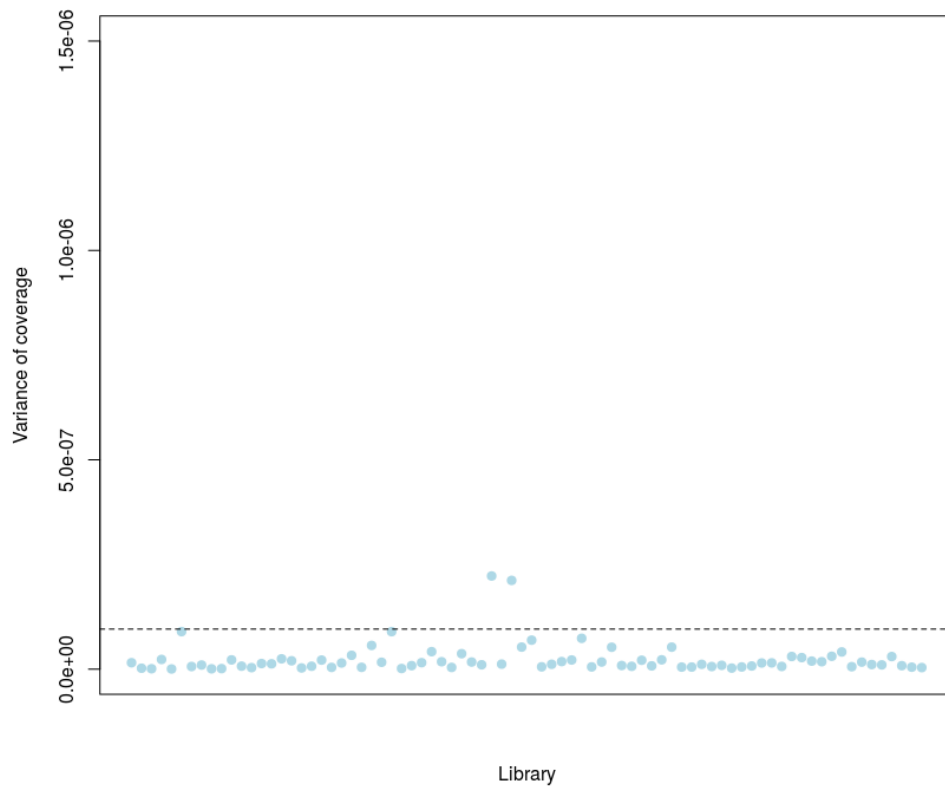
**Fig. S36. Taxa identification in three subsamples (A-C) of four sediment samples.** The number of sequences identified as mammalian mtDNA in the libraries are marked on the right.



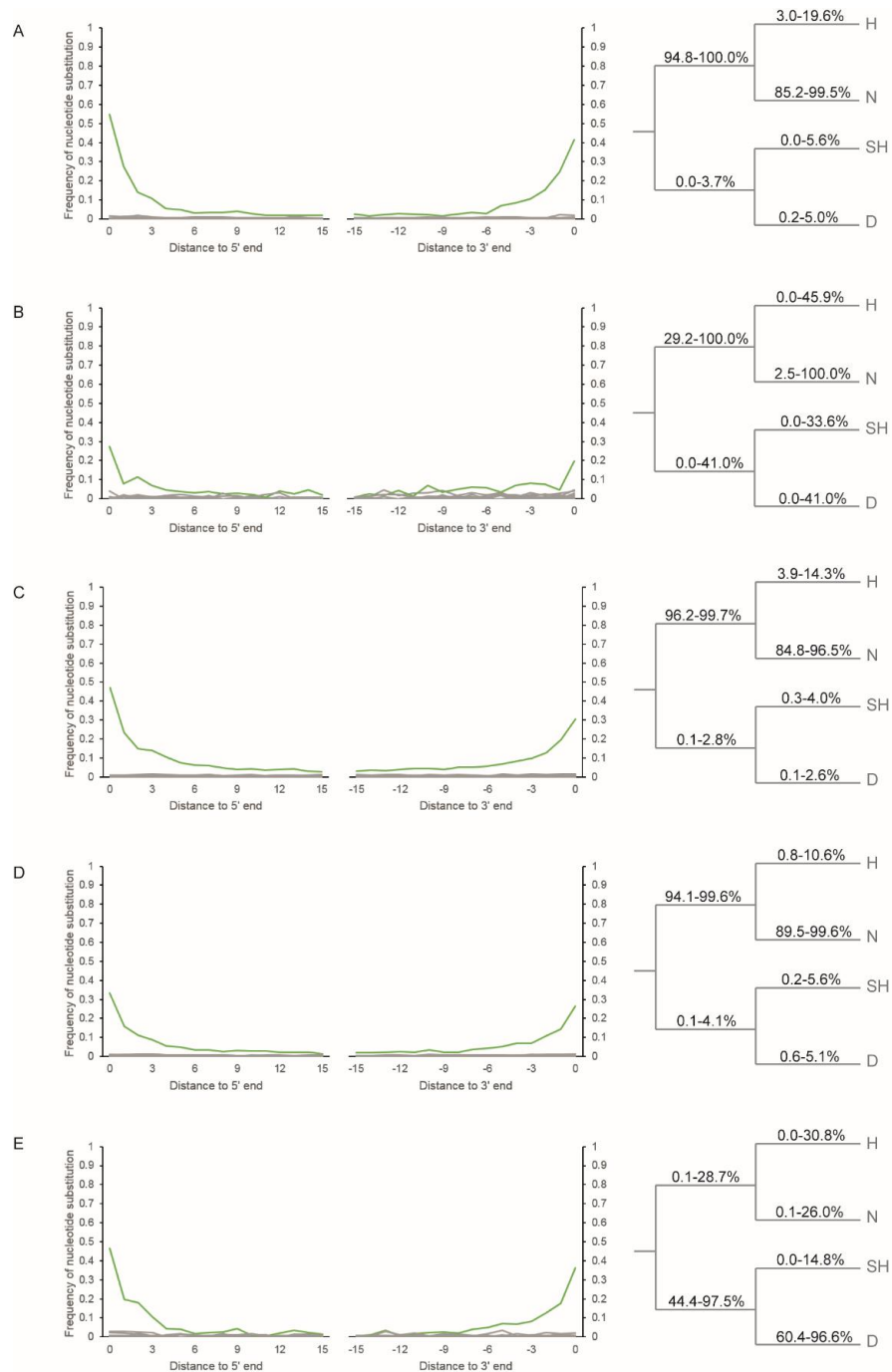
**Fig. S37. Taxa identification in six libraries, using all sequences generated and following their downsampling to ~500,000 sequences.** The number of sequences identified as mammalian mtDNA in the libraries are marked on the right.



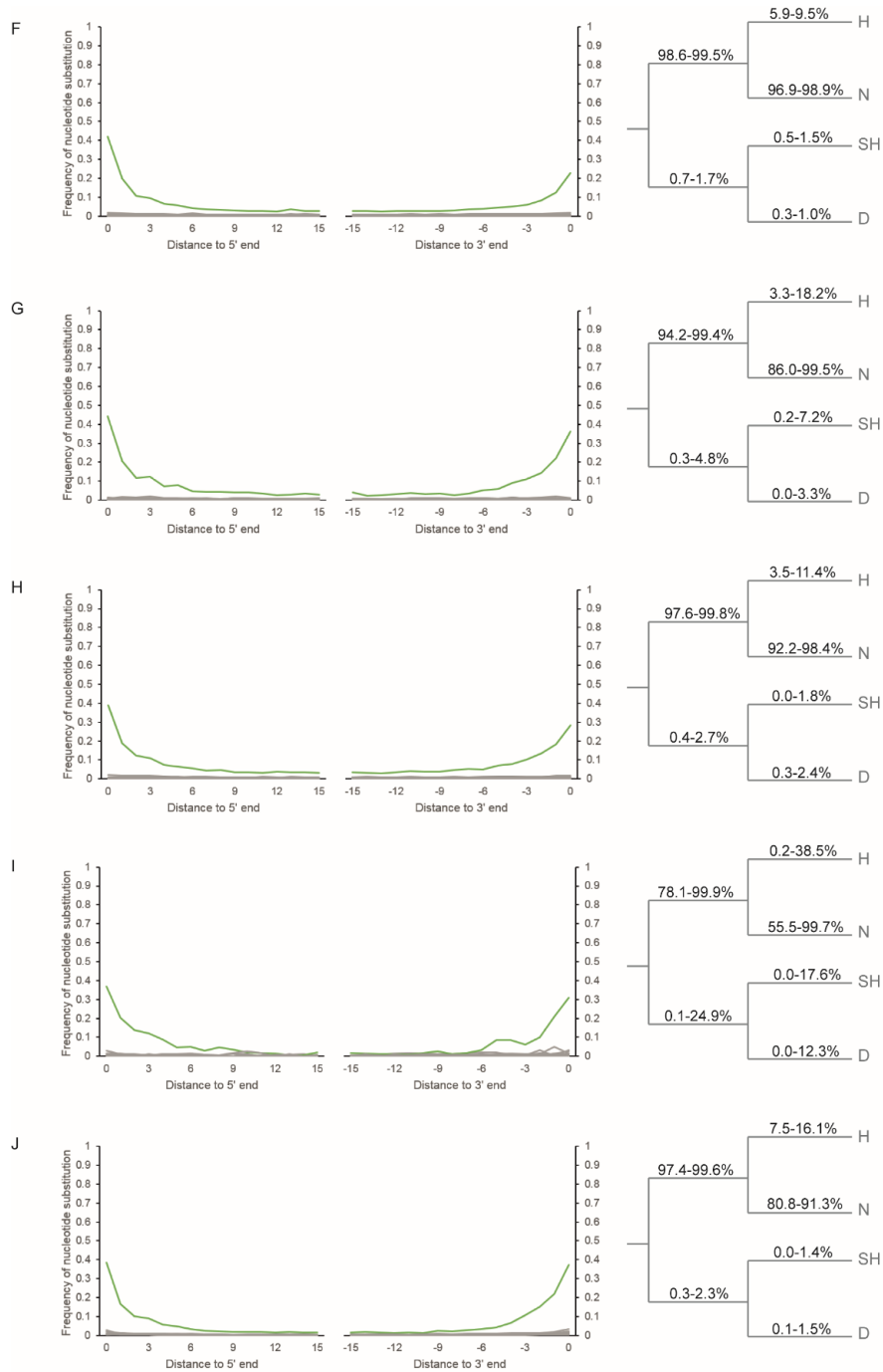
**Fig. S38. Taxa identification in six libraries, using data generated with and without hybridization capture for mammalian mtDNA fragments.** The number of sequences identified as mammalian mtDNA in the libraries are marked on the right.



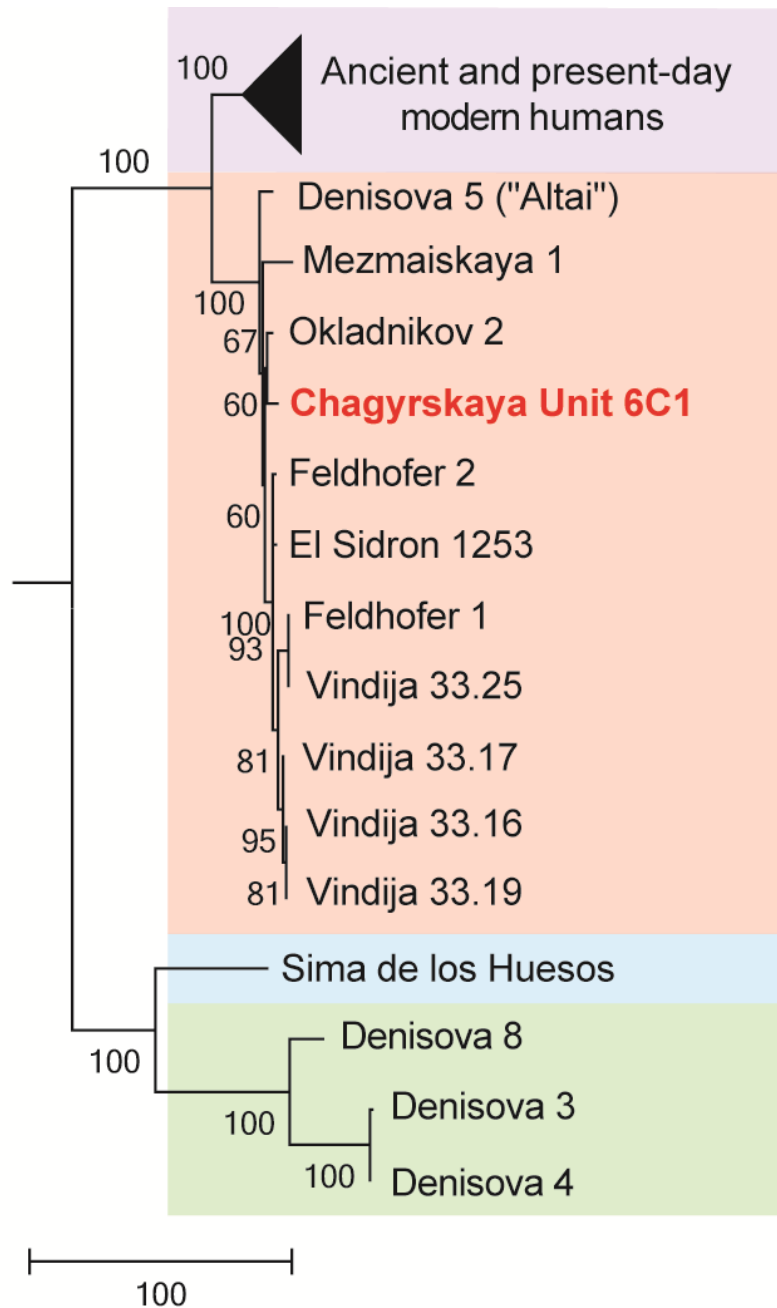
**Fig. S39. Variance of the mtDNA genome coverage in libraries containing hominin sequences.** The dashed line represents one standard deviation from the mean. Sequences from two libraries (R4050 and R4052) showed a higher variance than the average and were disregarded in further analyses.



**Fig. S40. Authentication of ancient hominin sequences.** In the left panel, frequencies of nucleotide substitutions at and near the terminal alignment positions are plotted (cytosine to thymine in green, all other substitutions in gray). In the right panel, 95% binomial confidence intervals of the percentage of putatively deaminated sequences matching variants specific to each branch in the hominin mtDNA tree are noted (H – human; N – Neandertal; SH – Sima de los Huesos; D - Denisovan). (A) Sample 19, Unit 6C1 of Chagyrskaya Cave; (B) Sample 27, Layer 11.2 of Denisova Cave (East Gallery); (C) Sample 35, Layer 11.4 of Denisova Cave (East Gallery); (D) Sample 42, Layer 14 of Denisova Cave (East Gallery); (E) Sample 44, Layer 15 of Denisova Cave (East Gallery).

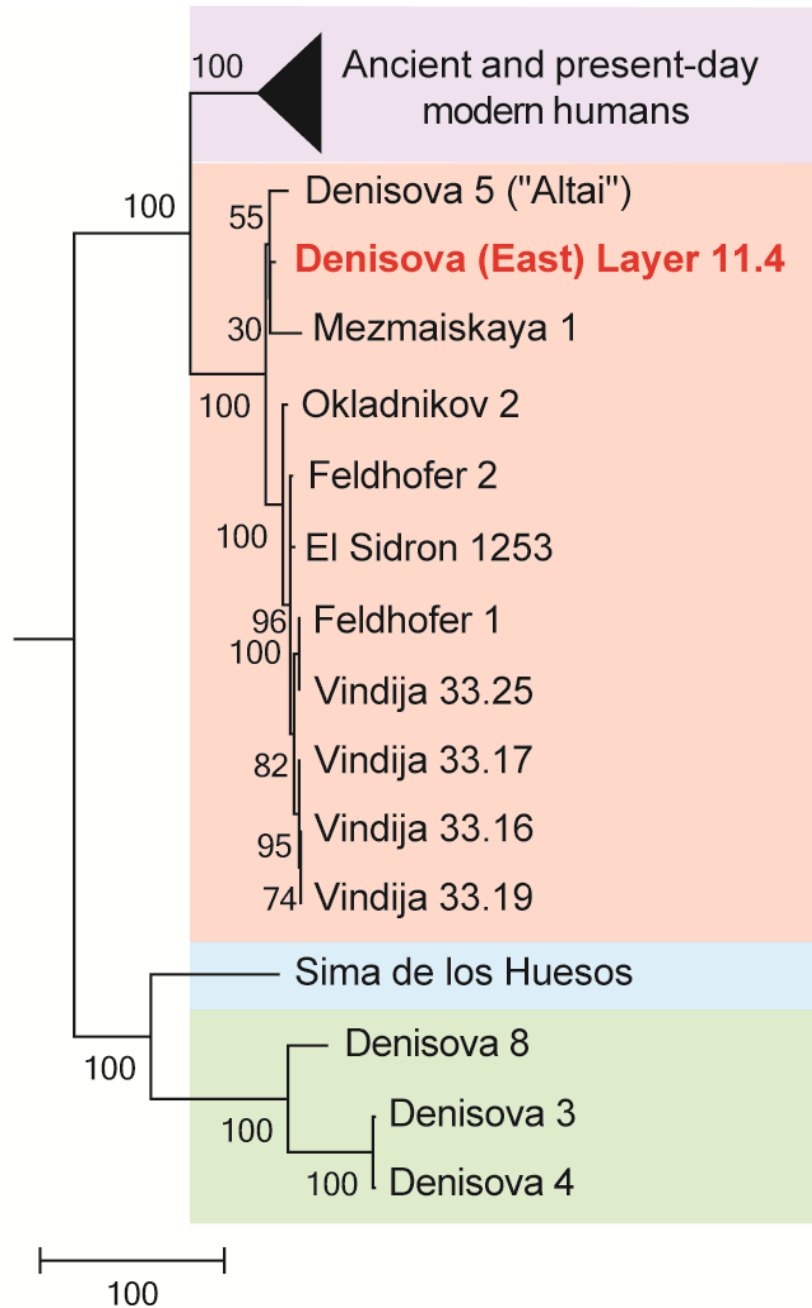


**Fig. S40 (continued).** (F) Sample 60, Layer 14.3 of Denisova Cave (Main Gallery); (G) Sample 61, Layer 17 of Denisova Cave (Main Gallery); (H) Sample 63, Layer 19.1 of Denisova Cave (Main Gallery); (I) Sample 82, Stratum 17b of Trou Al'Wesse; (J) Sample 85, Stratum III of El Sidrón.

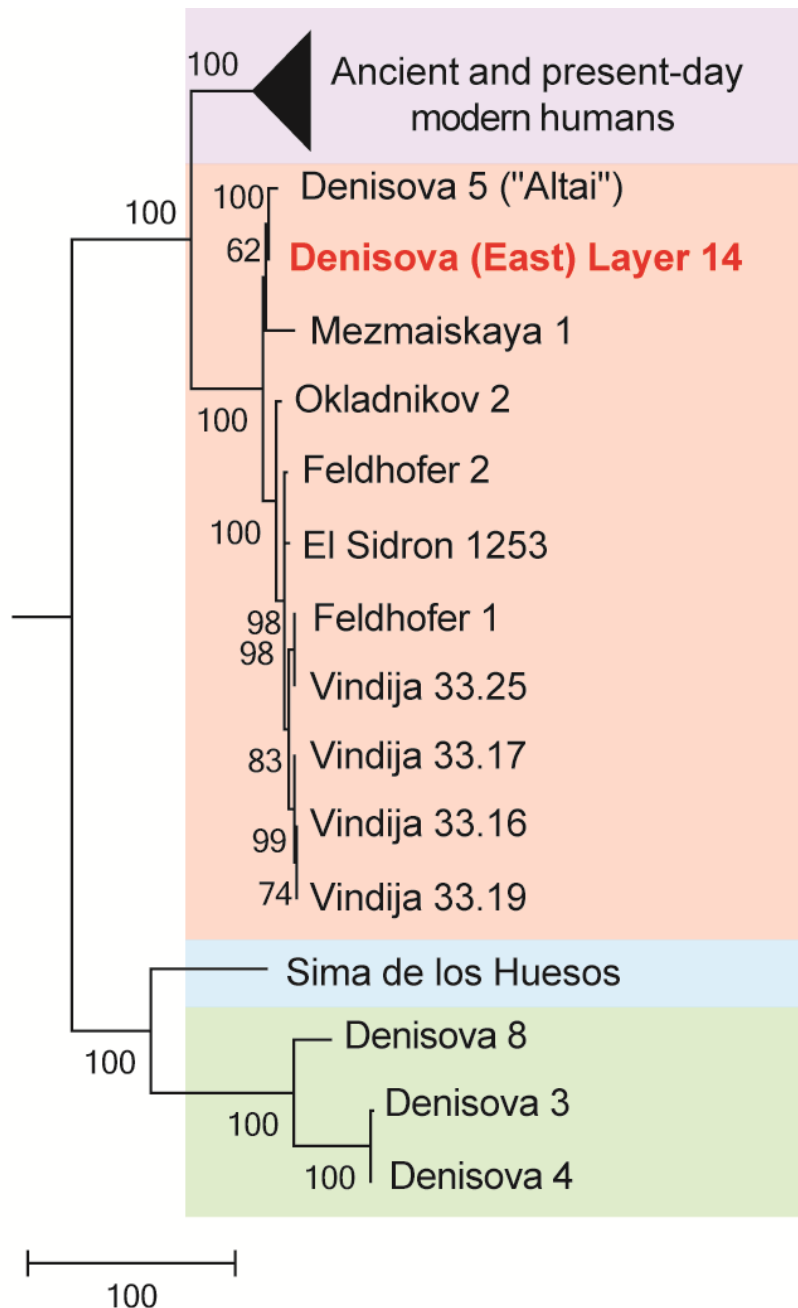


**Fig. S41. Neighbor-joining tree relating the mtDNA reconstructed from sample 19, collected in Unit 6C1 of Chagyrskaya Cave, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Neandertal mtDNAs (orange), and outside the variation of ancient and present-day modern human (purple), Denisovan (green) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 10,075 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.

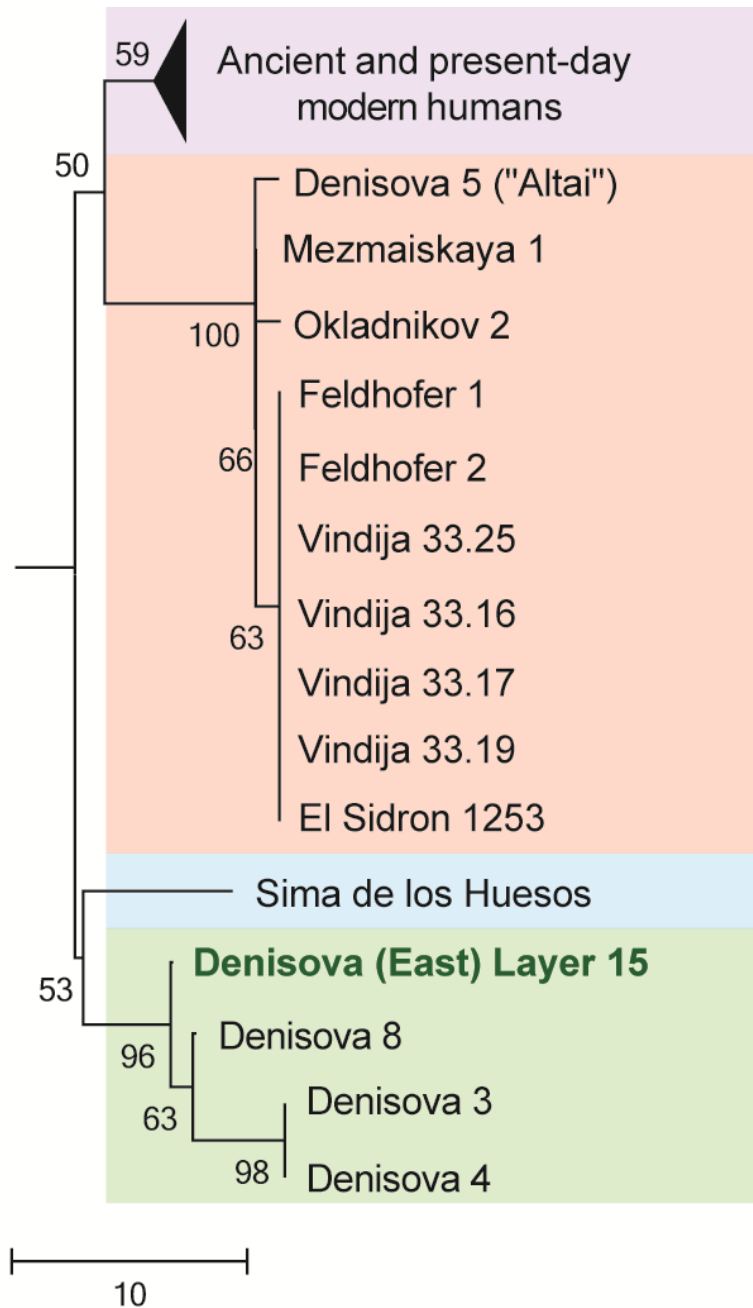




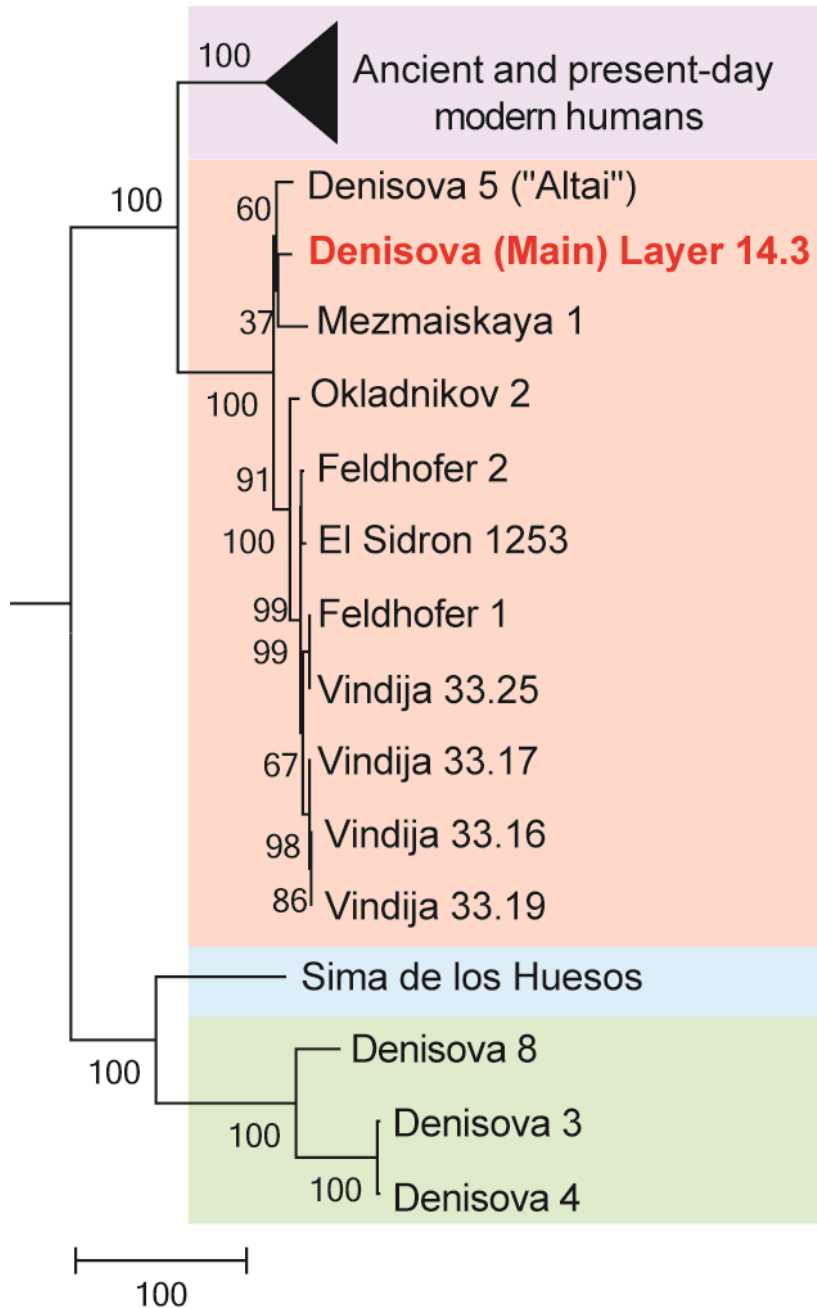
**Fig. S42. Neighbor-joining tree relating the mtDNA reconstructed from sample 35, collected in Layer 11.4 in the East Gallery of Denisova Cave, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Neandertal mtDNAs (orange), and outside the variation of ancient and present-day modern human (purple), Denisovan (green) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 14,616 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.



**Fig. S43. Neighbor-joining tree relating the mtDNA reconstructed from sample 42, collected in Layer 14 in the East Gallery of Denisova Cave, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Neandertal mtDNAs (orange), and outside the variation of ancient and present-day modern human (purple), Denisovan (green) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 12,744 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.

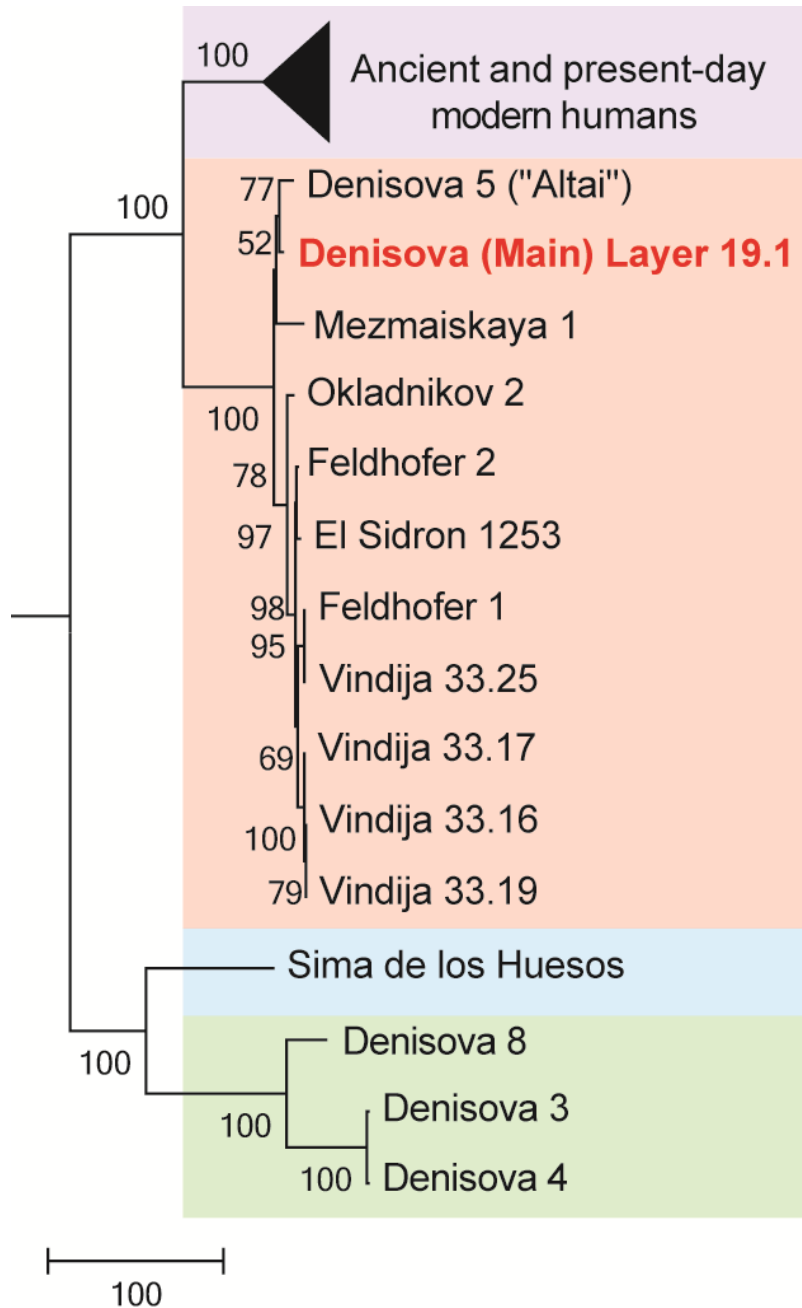


**Fig. S44. Neighbor-joining tree relating the mtDNA reconstructed from sample 44, collected in Layer 15 in the East Gallery of Denisova Cave, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Denisovan mtDNAs (green), and outside the variation of ancient and present-day modern human (purple), Neandertal (orange) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 1,283 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.

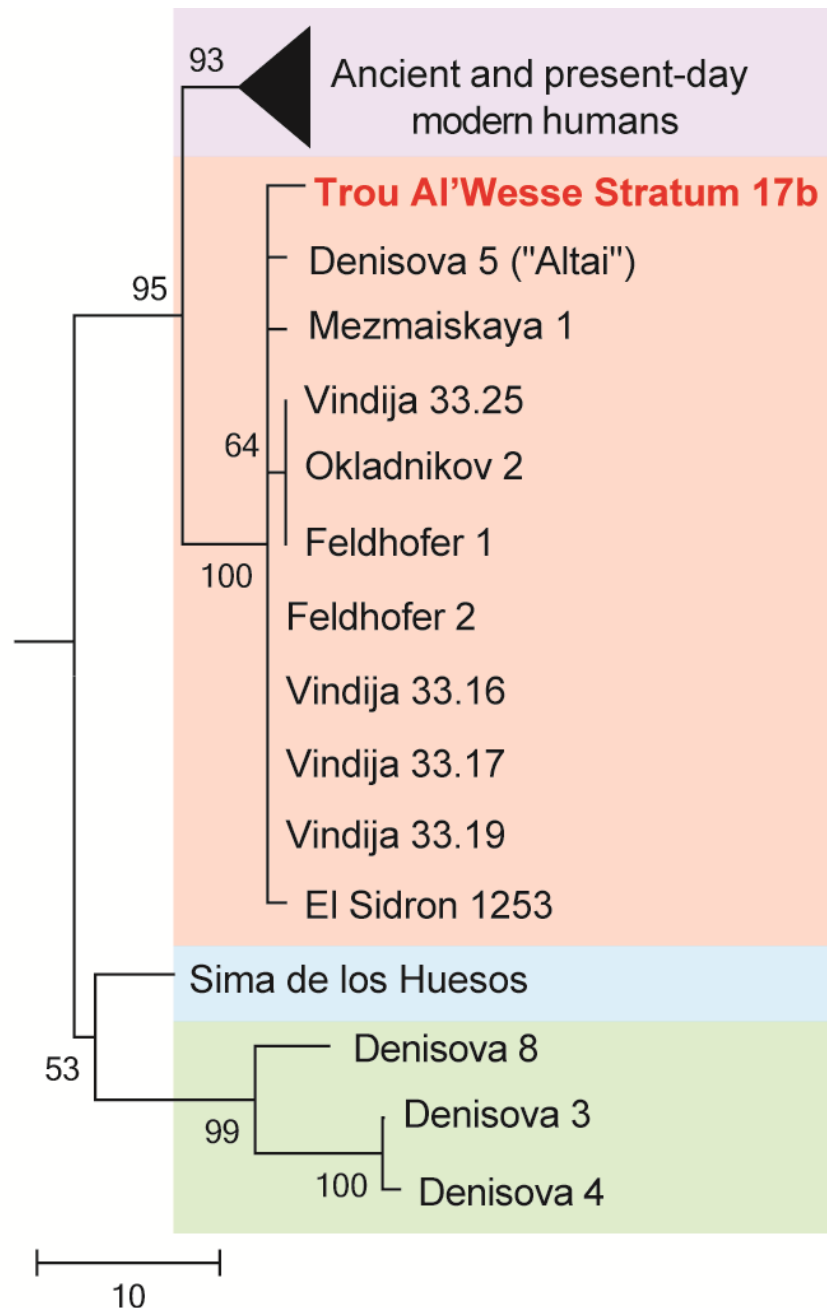


**Fig. S45. Neighbor-joining tree relating the mtDNA reconstructed from sample 60, collected in Layer 14.3 in the Main Gallery of Denisova Cave, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Neandertal mtDNAs (orange), and outside the variation of ancient and present-day modern human (purple), Denisovan (green) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 16,007 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.

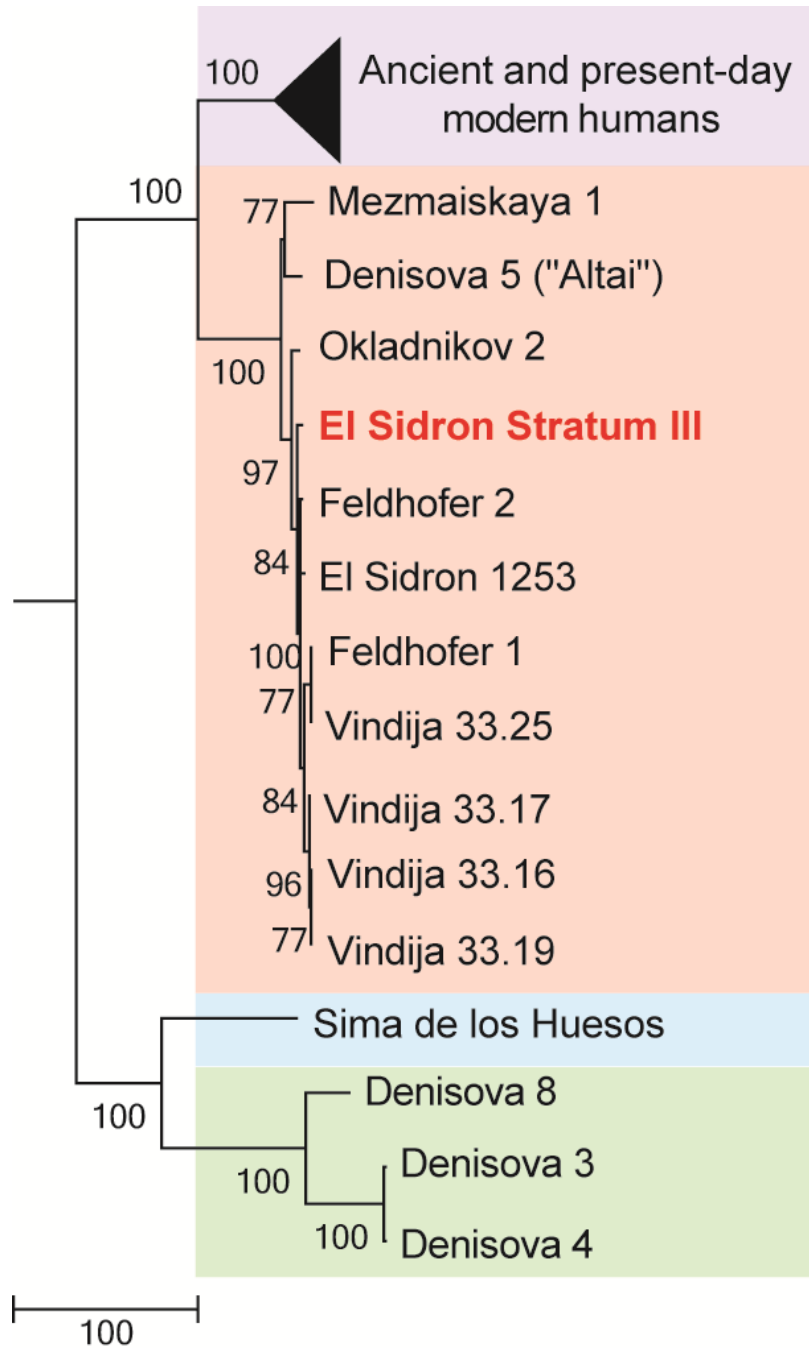




**Fig. S47. Neighbor-joining tree relating the mtDNA reconstructed from sample 63, collected in Layer 19.1 in the Main Gallery of Denisova Cave, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Neandertal mtDNAs (orange), and outside the variation of ancient and present-day modern human (purple), Denisovan (green) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 15,377 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.

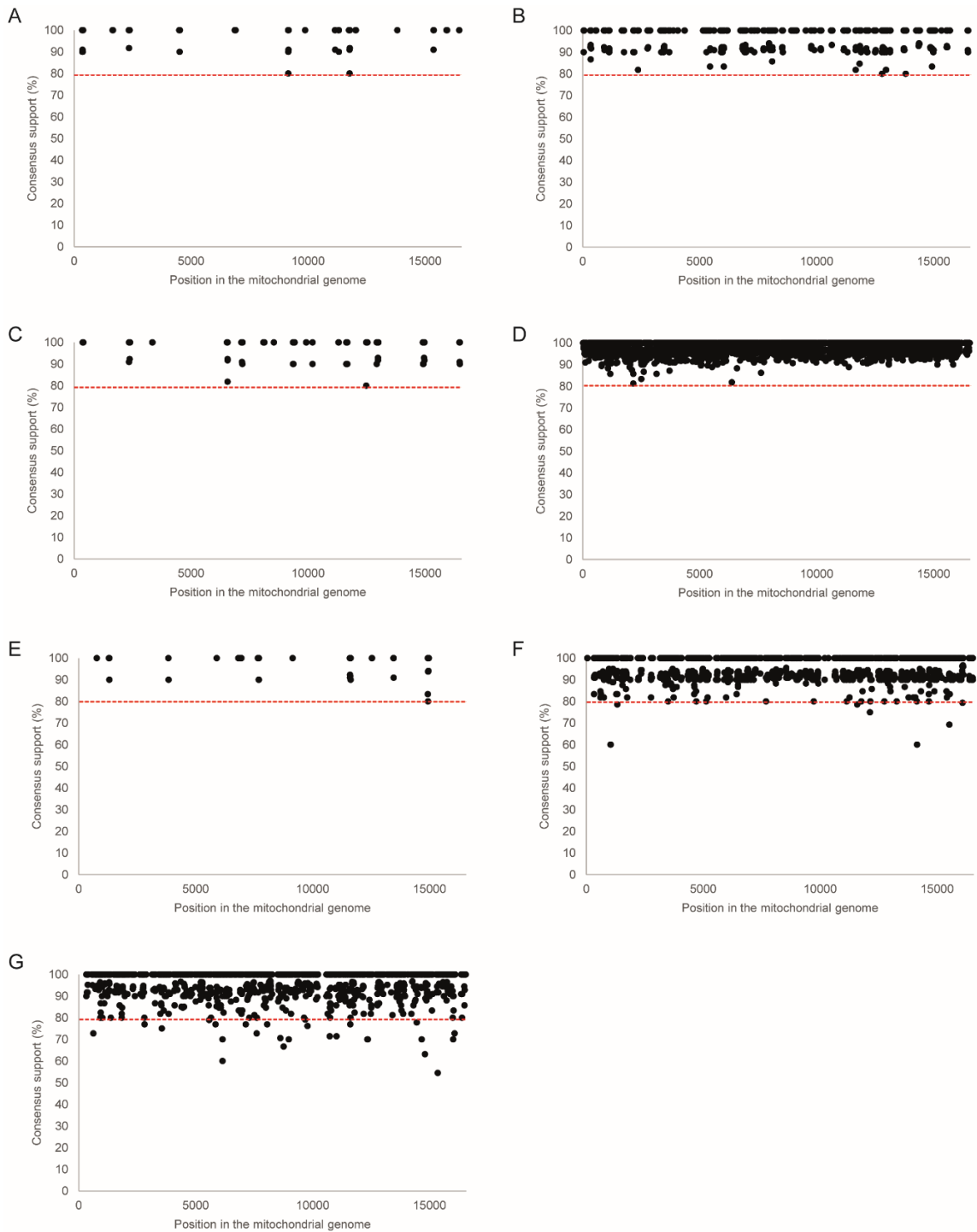


**Fig. S48. Neighbor-joining tree relating the mtDNA reconstructed from sample 82, collected in Stratum 17b of Trou Al'Wesse, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Neandertal mtDNAs (orange), and outside the variation of ancient and present-day modern human (purple), Denisovan (green) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 1,525 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.



**Fig. S49. Neighbor-joining tree relating the mtDNA reconstructed from sample 85, collected in Stratum III of El Sidrón, to other ancient and present-day mtDNAs.** The mtDNA clusters with previously determined Neandertal mtDNAs (orange), and outside the variation of ancient and present-day modern human (purple), Denisovan (green) and the SH hominin (blue) mtDNAs. A chimpanzee mtDNA (not shown) was used to root the tree. Missing positions and gaps in any of the mtDNAs were ignored in all others, leaving 15,462 positions in the analysis. The percentage of replicate trees (out of 500 bootstrap replications) in which taxa clustered together are noted adjacent to each branch. The tree is drawn to scale, with branch lengths measured in the number of base differences per genome. Details on the comparative data are presented in Table S5.

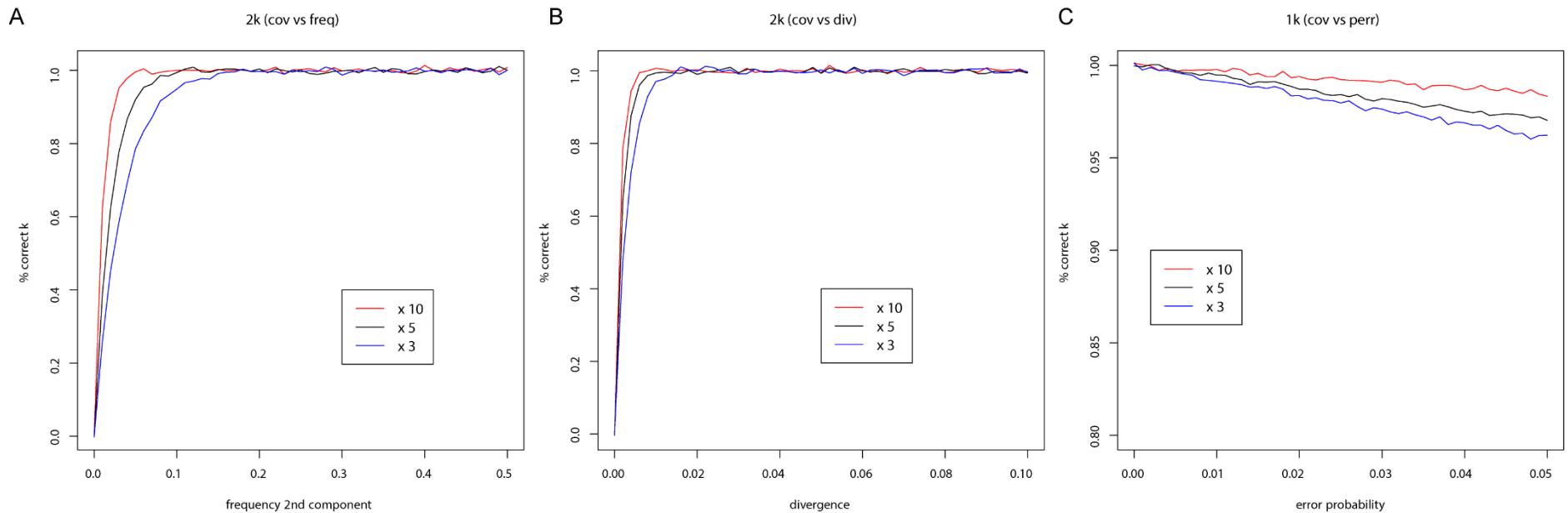




**Fig. S50. Percentage of sequences carrying an identical base at positions in the mitochondrial genome overlapped by at least 10 putatively deaminated sequences.** (A) Sample 19, Unit 6C1 of Chagyrskaya Cave; (B) Sample 35, Layer 11.4 of Denisova Cave (East Gallery); (C) Sample 42, Layer 14 of Denisova Cave (East Gallery); (D) Sample 60, Layer 14.3 of Denisova Cave (Main Gallery); (E) Sample 61, Layer 17 of Denisova Cave (Main Gallery); (F) Sample 63, Layer 19.1 of Denisova Cave (Main Gallery); (G) Sample 85, Stratum III of El Sidrón.

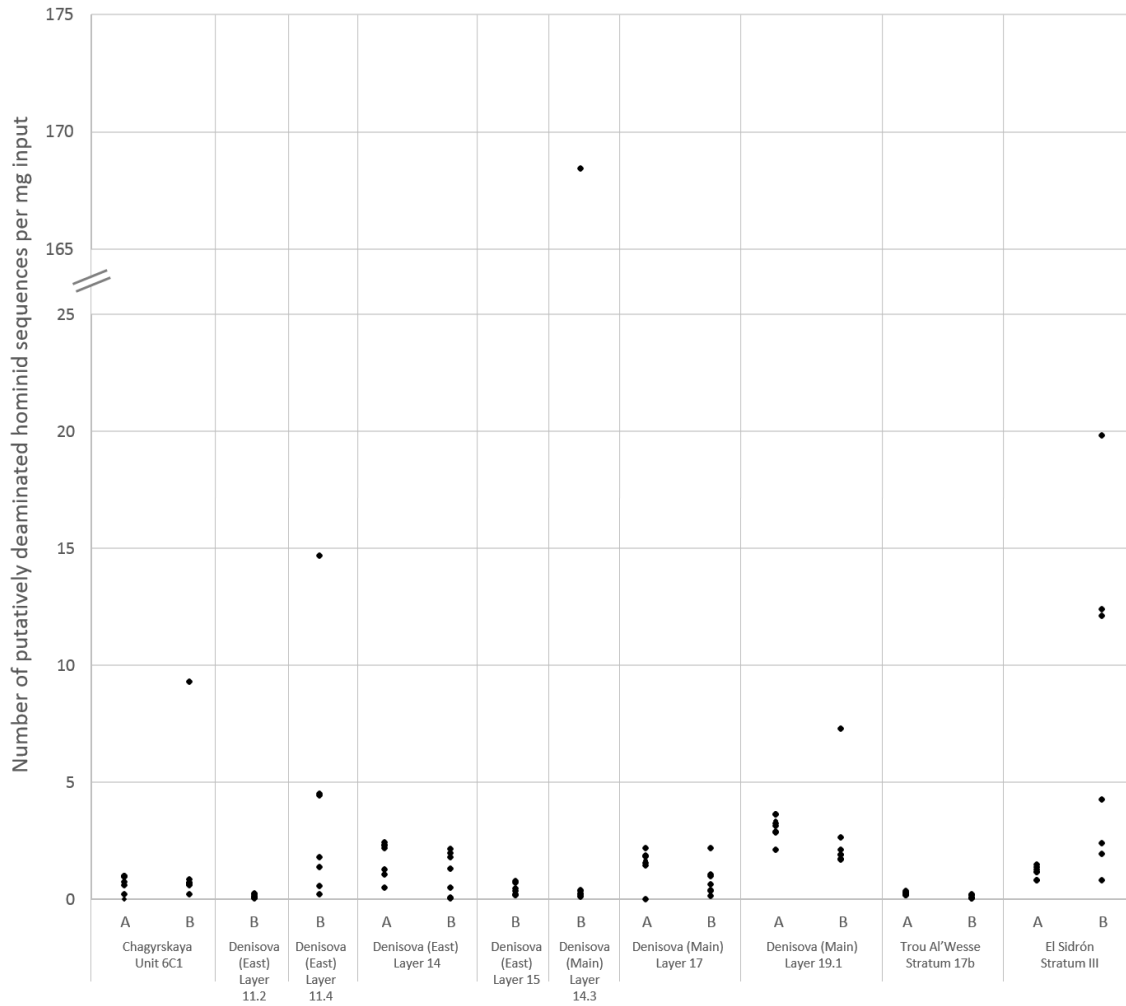


**Fig. S51. Sequences overlapping variable positions (yellow) in the mitochondrial genome.** Sequences are aligned to the revised Cambridge reference sequence mtDNA genome (noted at the top of each panel). On forward strands, bases are marked in uppercase letters if they differ from the reference genome and as dots if they are identical to it. On reverse strands, bases are marked as lowercase letters or commas, respectively. Visualization of the sequence alignments was carried out using SAMtools (149). Top: sample 63, Layer 19.1 of Denisova Cave (Main Gallery); Bottom: sample 85, Stratum III of El Sidrón.



**Fig. S52. Proportion of correct estimates of the number of components contributing to a simulated genome of length 15,000bp.**

To test the robustness of the method, (A) the frequency of the minor component, (B) its divergence from the other simulated component or (C) the error probability of a given sequence per position were varied. For all simulations, the number of sequences is assumed to follow a discretized normal distribution with average coverage as indicated in the caption (10-fold in red, 5-fold in blue, 3-fold in black) and standard deviation equal to  $0.2x$  average coverage. The error probability and the pairwise differences between two components are assumed to be 1% in all plots in which these parameters are not explicitly varied. In (B) the frequency of the second component is 10%.



**Fig. S53. Number of unique putatively deaminated hominin sequences yielded per milligram of sediment used for DNA extraction.** For each sample, this number is reported for all libraries prepared from a single DNA extract (A) when more than one such library was produced, and/or for libraries prepared from different extracts, i.e. from different subsamples of the same sediment sample (B).

**Table S1. List of 119 genomes used to generate the simulated datasets. Mammalian mitochondrial genomes are marked in bold.**

Bacterial genomes and mammalian mitochondrial genomes		
<i>Abiotrophia defectiva</i> ATCC 49176	<i>Dolosigranulum pigrum</i> ATCC 51524	<i>Neisseria bacilliformis</i> ATCC BAA-1200
<i>Achromobacter xylosoxidans</i> A8	<b>Domestic pig</b> (KC469586.1)	<i>Ochrobactrum anthropi</i> ATCC 49188
<i>Acinetobacter baumannii</i> AB0057	<i>Eggerthella lenta</i> DSM 2243	<i>Olsenella profusa</i> F0195
<i>Actinobaculum PI sp. oral taxon 183</i> F0552	<i>Eikenella corrodens</i> ATCC 23834	<i>Oribacterium sinus</i> F0268
<i>Actinomyces cardiffensis</i> F0333	<i>Enterobacter cancerogenus</i> ATCC 35316	<i>Paenibacillus sp. oral taxon 786</i> D14
<i>Afipia broomeae</i> ATCC 49717	<i>Enterococcus casseliflavus</i> EC30	<i>Parascardovia denticolens</i> DSM 10105
<i>Aggregatibacter actinomycetemcomitans</i> D17P-2	<i>Erysipelothrix tonsillarum</i> DSM 14972	<i>Parvimonas micra</i> ATCC 33270
<i>Agrobacterium tumefaciens str. C58</i>	<i>Escherichia coli</i> BW2952	<i>Peptoniphilus indolicus</i> ATCC 29427
<i>Alloiococcus otitis</i> ATCC 51267	<i>Eubacterium infirmum</i> F0142	<i>Peptostreptococcacea sp. oral taxon 113</i> W5053
<i>Alloscardovia omnicoles</i> DSM 21503	<i>Filifactor alocis</i> ATCC 35896	<i>Peptostreptococcus anaerobius</i> 653-L
<i>Anaerococcus lactolyticus</i> ATCC 51172	<i>Finegoldia magna</i> BVS033A4	<i>Porphyromonas asaccharolytica</i> DSM 20707
<i>Anaeroglobus geminatus</i> F0357	<i>Fusobacterium gonidiaformans</i> 3-1-5R	<i>Prevotella baroniae</i> DSM 16972
<i>Arcanobacterium haemolyticum</i> DSM 20595	<i>Gardnerella vaginalis</i> ATCC 14019	<i>Propionibacterium acidifaciens</i> DSM 21887
<i>Atopobium minutum</i> 10063974	<i>Gemella bergeriae</i> ATCC 700627	<i>Proteus mirabilis</i> HI4320_4363
<i>Bacillus anthracis</i> A0248	<i>Granulicatella adiacens</i> ATCC 49175	<i>Pseudomonas aeruginosa</i> LESB58
<i>Bacteroides pyogenes</i> DSM 20611	<i>Haemophilus aegyptius</i> ATCC 11116	<i>Pseudoramibacter alactolyticus</i> ATCC 23263
<i>Bacteroidetes bacterium oral taxon taxon 274</i> F0058	<i>Helicobacter pylori</i> India7	<i>Pyramidobacter piscolens</i> W5455, DSM 21147
<i>Bartonella schoenbuchensis</i> m07a	<i>Johnsonella ignava</i> ATCC 51276	<i>Ralstonia pickettii</i> 12D
<i>Bifidobacterium animalis subsp. Lactis</i> CNCM I-2494	<i>Jonquetella anthropi</i> E3_33 E1	<i>Rhodobacter capsulatus</i> SB 1003
<i>Bordetella pertussis</i> Tohama I	<i>Kingella denitrificans</i> ATCC 33394	<i>Rothia aerea</i> F0474
<i>Bradyrhizobium elkanii</i> USDA 76	<i>Klebsiella pneumoniae</i> Kp342	<i>Sanguibacter keddieii</i> DSM 10542
<i>Brevundimonas diminuta</i> ATCC 11568	<i>Kytococcus sedentarius</i> DSM 20547	<i>Scardovia inopinata</i> F0304
<i>Bulleidia extracta</i> W1219	<i>Lachnospiraceae bacterium</i> ACC2	<i>Selenomonas artemidis</i> DSM 19719
<i>Burkholderia cepacia</i> GG4	<i>Lactobacillus acidophilus</i> NCFM	<i>Shuttleworthia satelles</i> DSM 14600
<i>Candidate division SR1 bacterium</i> MGEHA	<i>Lactococcus lactis subsp. lactis</i> II1403	<i>Simonsiella muelleri</i> ATCC 29453
<i>Campylobacter concisus strain</i> 13826	<i>Lautropia mirabilis</i> ATCC 51599	<i>Slackia exigua</i> ATCC 700122
<i>Capnocytophaga gingivalis</i> ATCC 33624	<i>Leptotrichia buccalis</i> ATCC 14201	<i>Solobacterium moorei</i> W5408
<i>Cardiobacterium hominis</i> ATCC 15826	<i>Listeria monocytogenes</i> 08-5578	<b>Spotted hyena</b> (JF894377.1)
<i>Catonella morbi</i> ATCC 51271	<i>Lysinibacillus fusiformis</i> ZC1	<i>Staphylococcus aureus subsp. aureus</i> JH1
<i>Centipeda periodontii</i> DSM 2778	<i>Megasphaera micronuciformis</i> F0359	<i>Stenotrophomonas maltophilia</i> K279a
<i>Chloroflexi bacterium oral taxon 439 isolate</i> Chl2	<i>Mesorhizobium loti</i> MAFF303099	<i>Streptococcus agalactiae</i> NEM316
<i>Clostridiales [F-1][G-1] sp. oral taxon 876</i> F0540	<i>Methanobrevibacter oralis</i> JMR01	<i>Synergistetes sp. SGP1</i>
<i>Comamonas testosteroni</i> KF-1	<i>Microbacterium sp. oral taxon 186</i> F0373	<i>Tannerella forsythia</i> ATCC 43037
<i>Corynebacterium diphtheriae</i> NCTC 13129	<i>Mitsuokella multacida</i> DSM 20544	<i>Treponema denticola</i> ATCC 35405
<b>Cow</b> (DQ124371.1)	<i>Mobiluncus mulieris</i> ATCC 35243	<i>Turicella otitidis</i> ATCC 51513
<i>Cronobacter sakazakii</i> ATCC BAA-894	<i>Mogibacterium timidum</i> ATCC 33093	<i>Variovorax paradoxus</i> S110
<i>Cryptobacterium curtum</i> DSM 15641	<i>Moraxella catarrhalis</i> RH4	<i>Veillonella atypica</i> ACS-134-V-Col7a
<i>Delftia acidovorans</i> SPH-1	<i>Mycobacterium leprae</i> Br4923	<b>Woolly mammoth</b> (EU153448.1)
<i>Desulfobulbus sp. oral taxon 041</i> Dsb2	<i>Mycoplasma fermentans</i> M64	<i>Yersinia pestis Antiqua</i> JGI
<i>Dialister invisus</i> DSM 15470	<b>Neandertal</b> (KC879692.1)	

**Table S2. mtDNA genomes used to determine “diagnostic” positions differentiating between groups within taxa. The mtDNAs used as a reference genome for the realignment of sequences assigned to each taxa are marked in bold.**

Taxon	Accession Code	Taxon	Accession code	Taxon	Accession code	Taxon	Accession code	
<b>Bovidae</b>		<b>Canidae</b>		<b>Cercopithecidae</b>		<b>Elephantidae</b>		
<i>Bison bison</i> (American bison)	EU177871.1	<i>Canis latrans</i> (coyote)	DQ480509.1	<i>Macaca mulatta</i> (Rhesus macaque)	<b>NC_005943.1</b>	<i>Elephas maximus</i> (Asian elephant)	<b>NC_005129.2</b>	
	GU946978.1		DQ480511.1	<b>Cervidae</b>	EF588275.2			
	GU946985.1		KF661096.1	<i>Alces alces</i> (moose)	JN632595.1	AJ428946.1		
	GU946991.1		KT448275.1	<i>Capreolus capreolus</i> (European roe deer)	KJ681480.1	<i>Loxodonta africana</i> (African savannah elephant)	DQ316069.1	
	GU947002.1		KT448276.1		KJ681481.1		NC_000934.1	
	GU947003.1		KT448277.1		KJ681482.1		AB443879.1	
	GU947004.1		KF661080.1		KJ681483.1		JN673263.1	
	KF661088.1	KJ681484.1	JN673264.1					
<i>Bos primigenius</i> (auroch)	GU985279.1	KF661081.1	KJ681485.1	<i>Loxodonta cyclotis</i> (African forest elephant)	KJ557423.1			
	KF525852.1	KF661095.1	KJ681486.1	KJ557424.1				
<i>Bos taurus</i> (cow)	<b>NC_006853.1</b>	<i>Canis lupus</i> (grey wolf)	KF661042.1	<i>Capreolus pygargus</i> (Siberian roe deer)	KJ681487.1	<i>Mammuthus primigenius</i> (woolly mammoth)	DQ316067.1	
	EU177819.1		KF661048.1		JN632610.1		JF912200.1	
	EU177847.1		KF661052.1		KJ681492.1		NC_007596.2	
	EU177833.1		KF661066		KJ681493.1		EU153446.1	
	EU177868.1		<b>NC_002008.4</b>		KJ681494.1		EU153448.1	
	EU177849.1		KF661082.1		KJ681495.1		EU153451.1	
	EU177870.1		KF661083.1		<i>Cervus elaphus</i> (red deer)		<b>NC_007704.2</b>	EU153452.1
<i>Capra hircus</i> (goat)	KJ192209.1	<i>Canis lupus familiaris</i> (dog)	DQ480491.1	<i>Dama dama</i> (fallow deer)	JN632629.1	<i>Mammuth americanum</i> (mastodon)	NC_009574.1	
	KM093871.1		DQ480499.1	(*)	<i>Equidae</i>			
	KM360063.1		DQ480492.1	(*)		<i>Equus asinus</i> (donkey)	X97337.1	
	KR059152.1		DQ480489.1	<b>Cricetidae</b>	<i>Equus caballus</i> (horse)	KT182635.1		
	KR059172.1		EU408268.1	<i>Cricetulus griseus</i> (Chinese hamster)	NC_007936.1	<i>Equus hemionus onager</i> (wild ass)	KT757760.1	
	KR059213.1		<i>Lycaon pictus</i> (African wild dog)		KT448283.1		NC_027418.1	KT596764.1
	KR059211.1		<i>Speothos venaticus</i> (Bush dog)		KT598692.1		NC_013276.1	KT757740.1
	KR059225.1		<i>Vulpes lagopus</i> (arctic fox)	KT448286.1	<b>NC_015241.1</b>	KT757761.1		
	<i>Ovis aries</i> (sheep)	HM236174.1	<i>Vulpes vulpes</i> (red fox)	AM181037.1	<i>Microtus fortis fortis</i> (Reed vole)	NC_029477.1	KT368725.1	
		HM236176.1		GQ374180.1	<i>Myodes rufocanus</i> (Gray red-backed vole)		KT368725.1	
HM236179.1		JN711443.1			Early Middle Pleistocene horse	KT757763.1		
HM236180.1		KP342452.1			<i>Equus hemionus onager</i> (wild ass)	HM118851.1		
HM236182.1		KT448287.1			<i>Equus ovodovi</i> (Sussemionus)	JX312730.1		
KF302456.1						JX312734.1		
KF938321.1								
<i>Procapra gutturosa</i> (gazelle)	JN632689.1							
<i>Saiga tatarica</i> (Saiga antelope)	JN632700.1							

(\*) Giant elk mtDNA genomes are from (159).

**Table S2 (continued).**

Taxon	Accession Code	Taxon	Accession code	Taxon	Accession code	Taxon	Accession code
<b>Erinaceidae</b>		<b>Hyaenidae</b>		<b>Ursidae</b>			
<i>Erinaceus europaeus</i> (hedgehog)	NC_002080.2	<i>Crocuta crocuta</i> (spotted hyena)	JF894377.1		NC_003427.1		FN390866.1
<b>Felidae</b>		<i>Crocuta crocuta (spelaea)</i> (cave hyena)	NC_020670.1		EU497665.1		FN390867.1
<i>Acinonyx jubatus</i> (cheetah)	AF344830.1		JF894379.1		HQ685901.1		FN390855.1
	AY463959.1	<i>Hyaena hyaena</i> (striped hyena)	JF894376.1		HQ685902.1		EU327344.1
	KP202271.1	<b>Leporidae</b>			HQ685903.1		FN390865.1
	KR132579.1	<i>Oryctolagus cuniculus</i> (rabbit)	NC_001913.1		HQ685909.1		FN390868.1
<i>Felis catus</i> (domestic cat)	NC_001700.1	<b>Muridae</b>		<i>Ursus arctos</i> (brown bear)			FN390871.1
<i>Felis chaus</i> (jungle cat)	KP202274.1	<i>Mus musculus</i> (house mouse)	NC_005089.1		HQ685916.1		FN390847.1
<i>Felis sylvestris</i> (wild cat)	KP202278.1	<b>Mustelidae</b>			HQ685927.1		FN390852.1
<i>Lynx lynx</i> (lynx)	KM982549.1		AM711901.1		HQ685929.1		FN390852.1
	KP202283.1	<i>Gulo gulo</i> (wolverine)	KF415127.1		HQ685942.1		FN390872.1
	KR132581.1		KR611313.1		HQ685951.1		FN390849.1
	KR919624.1	<i>Martes martes</i> (pine marten)	NC_021749.1		HQ685957.1		FN390850.1
<i>Panthera leo</i> (lion)	KP202262.1	<i>Meles meles</i> (European badger)	AM711900.1	<i>Ursus deningeri</i> (Deninger's bear)	KF437625.2		FN390866.1
	KR132589.1	<i>Mustela erminae</i> (Stoat)	KM091450.1	<i>Ursus deningeri kudarensis</i> (Kudaro cave bear)	FN390863.1		FN390867.1
<i>Panthera pardus</i> (leopard)	EF551002.1	<i>Mustela eversmannii</i> (Steppe polecat)	KT224449.1		FN390864.1		FN390855.1
	KP202265.1	<i>Mustela nivalis</i> (least weasel)	HM106319.1		FN390860.1	<i>Ursus spelaeus</i> (cave bear)	EU327344.1
<b>Hominidae</b>		<i>Mustela putorius</i> (European polecat)	HM106318.1		FN390856.1		FN390865.1
Denisovans	NC_01399331	<b>Ochotonidae</b>			FN390857.1		FN390868.1
	FR695060	<i>Ochotona princeps</i> (American pika)	NC_005358.1		FN390858.1		FN390871.1
	KT780370.1	<b>Procaviidae</b>			NC_011112.1		FN390847.1
Middle Pleistocene hominin	NC_023100.1	<i>Procavia capensis</i> (Cape rock hyrax)	NC_004919.1		FN390845.1		FN390852.1
Neandertals	FM865411.1	<b>Rhinocerotidae</b>			FN390846.1		FN390872.1
	FM865407.1	<i>Ceratotherium simum</i> (white rhinoceros)	Y07726.1		FN390859.1		FN390849.1
	FM865408.1	<i>Coelodonta antiquitatis</i> (woolly rhinoceros)	FJ905813.1	<i>Ursus ingressus</i> (cave bear)	FN390869.1		FN390850.1
	FM865410.1	<i>Dicerorhinus sumatrensis</i> (Sumatran)	NC_012684.1		FN390870.1		FN390851.1
	AM948965.1	<i>Diceros bicornis</i> (black rhinoceros)	FJ905814.1		FN390862.1		FN390866.1
	KJ533544.1	<i>Rhinoceros sondaicus</i> (Javan rhino)	FJ905815.1		FN390842.1		FN390867.1
	KJ533545.1	<i>Rhinoceros unicornis</i> (Indian rhino)	X97336.1		FN390853.1	FN390855.1	
	FM865409.1	<b>Rhinolophidae</b>			FN390854.1	EU327344.1	
	KC879692.1	<i>Rhinolophus macrotis</i> (horseshoe bat)	NC_026460.1		FN390843.1	FN390865.1	
	KF982693.1	<b>Suidae</b>			FN390844.1	FN390868.1	
	<i>Pan troglodytes</i> (chimpanzee)	NC_001643.1	<i>Sus scrofa domesticus</i> (domestic pig)	NC_012095.1		FN390861.1	FN390871.1
	Present-day modern humans	NC_012920.1 (**)	<b>Talpidae</b>			FN390848.1	FN390847.1
		<i>Talpa europaea</i> (European mole)	NC_002391.1				
		<b>Vespertilionidae</b>					
		<i>Nyctalus noctula</i> (common noctule)	NC_027237.1				

(\*\*) The mtDNA genomes of 311 present-day modern humans from a variety of geographic origins were used (36).

**Table S3. Reconstructing mtDNA genomes using hominin sequences bearing damage-derived substitutions typical of ancient DNA.** A base was called if it was covered by at least two sequences in the permissive approach, or by at least three sequences in the stricter one. In both schemes, a base was called only if more than two-thirds of sequences overlapping the position carried an identical base. The consensus mtDNA genomes called using the permissive strategy were used in downstream analyses.

Sample	Origin	Number of sequences with terminal C to T	Average coverage (-fold)	Number of resolved bases	
				Permissive approach	Strict approach
19	Chagyrskaya Unit 6C1	1,009	2.7	10,209	7,363
35	Denisova (East) Layer 11.4	1,778	5.5	14,843	13,446
42	Denisova (East) Layer 14	1,199	3.6	12,946	10,430
44	Denisova (East) Layer 15	168	0.4	1,290	284
60	Denisova (Main) Layer 14.3	13,207	43.6	16,362	16,325
61	Denisova (Main) Layer 17	1,109	3.3	12,273	9,593
63	Denisova (Main) Layer 19.1	2,957	8.9	15,673	15,048
82	Trou Al'Wesse Stratum 17b	185	0.5	1,539	340
85	El Sidrón Stratum III	3,730	11.6	15,762	15,219

C- cytosine; T- thymine



**Table S4. Number of base differences among the mtDNA genomes reconstructed from sediment samples.** For each comparison, missing positions in either of the two genomes were disregarded.

	Chagyrskaya Unit 6C1	Denisova (East) Layer 11.4	Denisova (East) Layer 14	Denisova (East) Layer 15	Denisova (Main) Layer 14.3	Denisova (Main) Layer 17	Denisova (Main) Layer 19.1	Trou Al'Wesse Stratum 17b
Denisova (East) Layer 11.4	8	-						
Denisova (East) Layer 14	4	3	-					
Denisova (East) Layer 15	8	11	12	-				
Denisova (Main) Layer 14.3	10	10	7	14	-			
Denisova (Main) Layer 17	4	6	3	11	6	-		
Denisova (Main) Layer 19.1	7	6	2	13	5	5	-	
Trou Al'Wesse Stratum 17b	1	4	3	3	4	2	5	-
El Sidrón Stratum III	8	19	12	14	24	15	20	2

**Table S5. Comparative data used for reconstructing phylogenetic trees.**

	Accession code		Accession Code
<b>Ancient modern humans</b>		<b>Present-day modern humans</b>	
Boshan 11 (China)	KC521454	Australian	AY289066
Dolni Vestonice 13 (Czech Republic)	KC521459	Chinese	AF346973
Dolni Vestonice 14 (Czech Republic)	KC521458	German	AF346983
Iceman (Austro-Italian border)	EU810403	Indian	AF346966
Kostenki 14 (Russia)	FM600416	Italian	AY882393
Loschbour (Luxembourg)	KC521455	Native American	AY195759
Oberkassel 998 (Germany)	KC521457	Papua New Guinean (Highland)	AY289085
Saqqaq (Greenland)	EU725621	San	AF347008
Tianyuan (China)	KC417443	Taiwanese Indian	AY289098
Ust'-Ishim (Siberia)	PRJEB6622	Yoruba	AF347015
<b>Neandertals</b>		<b>Denisovans</b>	
Denisova 5 ("Altai") (Siberia)	KC879692	Denisova 3 (Siberia)	NC013993
El Sidrón 1253 (Spain)	FM865409	Denisova 4 (Siberia)	FR695060
Feldhofer 1 (Germany)	FM865407	Denisova 8 (Siberia)	KT780370
Feldhofer 2 (Germany)	FM865408	<b>Middle Pleistocene hominin</b>	
Mezmaiskaya 1 (Caucasus)	FM865411	Sima de los Huesos (SH) (Spain)	NC023100
Okladnikov 2 (Siberia)	KF982693	<b>Chimpanzee</b>	
Vindija 33.16 (Croatia)	AM948965		
Vindija 33.17 (Croatia)	KJ533544		
Vindija 33.19 (Croatia)	KJ533545		
Vindija 33.25 (Croatia)	FM865410		

**Table S6. Estimates of present-day human contamination in the libraries from which mtDNA genomes were reconstructed.** The contamination is estimated as the percentage of sequences matching the present-day human state at positions where the reconstructed mtDNA differs from a panel of 311 human mtDNAs. 95% exact binomial confidence intervals are shown.

	Number of positions unique to the test genome	All sequences			Sequences with terminal C to T		
		Number of sequences matching test mtDNA	Number of sequences matching human mtDNA	Contamination estimate (95% CI)	Number of sequences matching test mtDNA	Number of sequences matching human mtDNA	Contamination estimate (95% CI)
Chagyrskaya Unit 6C1	22	101	47	31.8 (24.4-39.9)	30	1	3.2 (0.1-16.7)
Denisova (East) Layer 11.4	49	584	38	6.1 (4.4-8.3)	109	1	0.9 (0.0-5.0)
Denisova (East) Layer 14	40	363	185	33.8 (29.8-37.9)	65	0	0.0 (0.0-5.5)
Denisova (East) Layer 15	3	7	0	0.0 (0.0-41.0)	6	0	0.0 (0.0-46.0)
Denisova (Main) Layer 14.3	70	6,268	53	0.8 (0.6-1.1)	873	8	0.9 (0.4-1.8)
Denisova (Main) Layer 17	35	228	62	21.4 (16.8-26.6)	47	3	6.0 (1.3-16.5)
Denisova (Main) Layer 19.1	58	1,110	75	6.3 (5.0-7.9)	194	0	0.0 (0.0-1.9)
Trou Al'Wesse Stratum 17b	6	8	3	27.3 (6.0-61.0)	4	0	0.0 (0.0-60.2)
El Sidrón Stratum III	59	892	969	52.1 (49.8-54.4)	188	16	7.8 (4.5-12.4)

C- cytosine; T- thymine

**Table S7. Variable positions in datasets from which mtDNA genomes were reconstructed.** To define variable positions within each sample, the following additive filters were applied: positions were required to be covered by 10 putatively deaminated sequences or more; fewer than 80% of sequences overlapping the position carried an identical base; both variants are present on at least two sequences on both forward and reverse strands. Additionally, we show whether the positions retained are known variants within Neandertal or Denisovan mtDNAs in our comparative datasets (see Table S5).

Sample	Origin	Number of positions with coverage $\geq 10$	Number of positions with support $< 80\%$	Number of positions with variants represented in both orientations	Known variant among Neandertal /Denisovan mtDNAs
19	Chagyrskaya Unit 6C1	287	0	NA	NA
35	Denisova (East) Layer 11.4	2,100	0	NA	NA
42	Denisova (East) Layer 14	395	0	NA	NA
44	Denisova (East) Layer 15	0	NA	NA	NA
60	Denisova (Main) Layer 14.3	15,669	0	NA	NA
61	Denisova (Main) Layer 17	172	0	NA	NA
63	Denisova (Main) Layer 19.1	7,331	9	1	yes
82	Trou Al'Wesse Stratum 17b	0	NA	NA	NA
85	El Sidrón Stratum III	9,354	27	1	no

NA – not applicable

**Table S8. Maximum likelihood estimates of the number of mtDNA components  $k$  contributing to datasets comprised of Elephantidae, Hyaenidae or Ursidae sequences.** The first columns contain observed statistics for each sample (average mtDNA coverage; the number of variable positions and their proportion among all positions with at least one overlapping sequence; and the frequency of the bases alternative to the most common one averaged across all variable positions). ‘logL1k’, ‘logL2k’ and ‘logL3k’ are the log likelihoods of the maximum likelihood estimate for models including one, two or three mtDNA components. The fraction of the rarest component is shown for the model with  $k=2$  and for  $k=3$  when this model is supported. ‘divfreq2k’ indicates the estimated divergence between the two components in a model with  $k=2$ . The total number of components inferred to be present in a dataset is given by the model with the lowest AIC value, as indicated in the column ‘modelAIC’; and ‘reL’ indicates the p-value for a likelihood ratio test assessing whether the best model has a stronger support than the second best one. The last column shows the interpretation for the number of mtDNA components in a sample by comparing the likelihoods of the different models following the removal of components for which the estimated frequency is below 1%.

Taxon	Sample	Average coverage	Variable positions (number)	Variable positions (proportion)	Observed frequency minor state	Fraction rarest component							Interpretation of the mtDNA components	
						logL1k	logL2k	logL3k	2k	3k	divfreq2k	modelAIC		reL
Elephantidae	19	9.2	40	0.001	0.160	-607.43	-269.01	-267.303	0.013	-	0.070	2k	0.000000	2 mtDNAs
	41	9.2	55	0.001	0.174	-698.8	-340.63	-342.485	0.015	-	0.049	2k	0.000000	2 mtDNAs
	61	16.0	57	0.001	0.101	-826.62	-362.81	-357.084	0.007	0.010	0.113	3k	0.065680	1 mtDNA
	80	3.5	12	0.001	0.241	-178.8	-86.52	-85.0958	0.079	-	0.014	2k	0.000000	2 mtDNAs
Hyaenidae	27	75.0	101	0.000	0.013	-971.69	-604.09	-618.586	0.005	-	0.043	2k	0.000000	1 mtDNA
	83	27	102	0.001	0.072	-1406.48	-665.89	-616.582	0.004	0.002	0.240	3k	0.000000	1 mtDNA
Ursidae	26	3.3	23	0.001	0.258	-332.09	-156.33	-156.442	0.025	-	0.052	2k	0.000000	2 mtDNAs
	27	4.0	35	0.001	0.233	-596.14	-247.54	-240.43	0.08	0.204	0.009	3k	0.016382	3 or more mtDNAs
	29	2.5	16	0.001	0.304	-232.69	-108.6	-107.665	0.053	-	0.033	2k	0.000000	2 mtDNAs
	32	4.0	127	0.003	0.287	-1626.43	-769.99	-738.825	0.125	0.177	0.041	3k	0.000000	3 or more mtDNAs
	34	4.2	176	0.005	0.287	-2325.18	-985.44	-987.512	0.202	-	0.031	2k	0.000000	2 mtDNAs
	37	2.5	14	0.001	0.295	-215.38	-95.14	-94.2013	0.063	-	0.020	2k	0.000000	2 mtDNAs
	39	49.9	198	0.000	0.045	-3808.58	-1436.52	-1270.32	0.017	0.033	0.053	3k	0.000000	3 or more mtDNAs
	42	4.8	121	0.003	0.257	-1596.21	-696.96	-701.167	0.107	-	0.034	2k	0.000000	2 mtDNAs
	51	3.5	16	0.001	0.280	-249.77	-115.03	-116.441	0.078	-	0.011	2k	0.000000	2 mtDNAs
	52	4.8	29	0.001	0.158	-440.41	-191	-190.641	0.049	-	0.021	2k	0.000000	2 mtDNAs
	55	2.8	26	0.001	0.333	-423.91	-181.32	-183.063	0.299	-	0.008	2k	0.000000	2 mtDNAs
62	2.9	39	0.001	0.319	-580.26	-259.6	-264.822	0.116	-	0.014	2k	0.000000	2 mtDNAs	

**Table S9. Maximum likelihood estimates of the number of mtDNA components k contributing to datasets comprised of putatively deaminated hominin sequences.** All fields in the table are analogous to Table S8. In order to account for contamination by present-day human DNA, estimates were performed using all positions in the mitochondrial genome or when removing positions at which present-day human mtDNAs differ from Neandertal ones (indicated as ‘-fixed diff’). The interpretation for the number of mtDNA components in a sample was carried out by comparing the likelihoods of the different models following the removal of components for which the estimated frequency is below 1%. A component was attributed to a human contaminant when the exclusion of positions differing between modern human and Neandertal mtDNAs resulted in the removal of one of the mtDNA components estimated by the model or in a significantly lower inferred frequency of the minor component. The latter significance is estimated by considering the range of frequencies for which a likelihood ratio test would be significant ( $\log L(\text{minfreq2k-fixed diff}, \theta_{\text{all}}) - \log L(\text{minfreq2kall}, \theta_{\text{all}}) > -1.92$ ), where  $\theta_i$  indicates all parameters estimated on dataset  $i$ , and  $\text{minfreq2k}_i$  the fraction of the rarest component for the model with  $k=2$ .

Sample	Origin	Type	Average coverage	Variable positions (number)	Variable positions (proportion)	Observed frequency minor state	logL1k	logL2k	logL3k	Fraction rarest component			Interpretation of the mtDNA components	
										2k	3k	modelAIC		reL
19	Chagyrskaya Unit 6C1	all	3.2	22	0.0010	0.316	-209.21	-188.7	-189.93	0.227	-	2k	0.000000	1 ancient mtDNA type + contamination
		-fixed diff	3.2	19	0.0008	0.296	-172.47	-162.02	-162.38	0.087	-	2k	0.000213	
35	Denisova (East) Layer 11.4	all	5.6	58	0.0005	0.213	-406.18	-402	-404.6	0.025	-	2k	0.112648	1 ancient mtDNA type + contamination
		-fixed diff	5.6	53	0.0005	0.207	-377.68	-377.7	-384.98	-	-	1k	0.131882	
42	Denisova (East) Layer 14	all	3.9	48	0.0010	0.290	-382.47	-352.79	-361.49	0.068	-	2k	0.000000	2 ancient mtDNA types
		-fixed diff	3.9	41	0.0009	0.274	-343.81	-309.57	-316.93	0.067	-	2k	0.000000	
60	Denisova (Main) Layer 14.3	all	42.7	308	0.0009	0.034	-3460.00	-1860.00	-1710.00	0.013	0.000	3k	0.000000	1 ancient mtDNA type + contamination
		-fixed diff	42.8	290	0.0009	0.031	-3158.30	-1679.32	-1596.41	0.009	0.001	3k	0.000000	
61	Denisova (Main) Layer 17	all	3.6	48	0.0006	0.323	-358.3	-334.28	-340.55	0.493	-	2k	0.000000	1 ancient mtDNA type + contamination
		-fixed diff	3.7	40	0.0005	0.317	-307.85	-291.9	-295.49	0.105	-	2k	0.000001	
63	Denisova (Main) Layer 19.1	all	8.9	106	0.0007	0.150	-784.95	-707.72	-707.02	0.059	-	2k	0.000000	2 ancient mtDNA types
		-fixed diff	8.9	100	0.0006	0.147	-754.15	-676.46	-683.15	0.050	-	2k	0.000000	
85	El Sidrón Stratum III	all	11.6	133	0.0010	0.190	-1130.62	-962.84	-925.95	0.053	0.072	3k	0.000000	2 ancient mtDNA types + contamination
		-fixed diff	11.6	107	0.0009	0.173	-909.98	-770.46	-769.5	0.035	-	2k	0.000000	

**Table S10. Number of mtDNA fragments retrieved from bones or from sediments, per milligram of material used for DNA extraction.** The numbers of unique sequences reported represent hominin mtDNA in bone samples and mtDNA of any mammalian taxa in sediment samples. References for the comparative data are noted in parentheses

Origin	Bone				Sediment			
	Sample	Number of unique mtDNA sequences	Input (mg)	Number of unique mtDNA sequences per mg	Sample	Number of unique mtDNA sequences	Input (mg)	Number of unique mtDNA sequences per mg
Denisova (East) Layer 11.2	Denisova 3 (2)	30,443	30	1,015	27	140,187	42.3	3,314
					28	89,522	151.9	589
					29	63,235	56.2	1,125
Denisova (East) Layer 11.4	Denisova 5 (“Altai”) (21)	268,551	38	7,067	32	93,201	52.2	1,785
					33	171,053	38.1	4,490
					34	74,108	152.8	485
					35	49,274	59.1	834
Denisova (East) Layer 12	Denisova 11 (35)	282,502	30.9	9,142	36	13,013	159.8	81
					37	77,496	67.3	1,152
					38	68,274	55.1	1,238
					39	42,341	55.6	762
					40	52,085	50.7	1,028
Vindija Cave Layer G3	Vi33.16 (36)	8,341	300	28	84	5,204	153.5	34
	Vi33.19 (37)	16,799	58	290				
El Sidrón Stratum III	Sid 1253 (38)	18,111	223	81	85	3,012	77.9	39

## Captions for Data files S1-S4

**Data file S1. Sediment samples analysed in this study.** Details on the sampling locations, amounts of material used for DNA extraction, the types of DNA libraries prepared and the number of molecules in each of them are presented.

**Data file S2. Characteristics of the DNA libraries enriched for mammalian mtDNA fragments.**

**Data file S3. Evaluating the authenticity of mtDNA fragments from 22 mammalian families.** For each family, the number of unique sequences attributed to it from each sample, their average length and the frequencies of nucleotide substitutions typical of ancient DNA in these sequences are reported. We tested whether the latter are significantly higher than 10% using an exact binomial test. For taxa determined to be ancient, the percentages of sequences matching different groups or species at phylogenetically informative positions are computed. L- length; MQ – mapping quality; bp – base pairs; C – cytosine; T – thymine; NA – not applicable.

**Data file S4. Characteristics of the DNA libraries enriched for human mtDNA fragments and evaluation of the authenticity of the obtained sequences.** For each library and for merged datasets pertaining to specific samples, the number of unique hominin sequences, their average length and the frequencies of nucleotide substitutions typical of ancient DNA are shown. We tested whether the latter are significantly higher than 10% using an exact binomial test. The number and percentage of sequences matching variants specific to each branch of a phylogenetic tree relating four hominin groups are reported, using all sequences in a samples and after retaining only those exhibiting terminal C to T substitutions. L- length; MQ – mapping quality; bp – base pairs; C – cytosine; T – thymine; NA – not applicable.