doi:10.1038/nature26151

Contents:

SI1: Information about the archaeological sites and specimens....2

SI2: Sampling, library preparation and initial contamination estimates....9

SI3: Sequencing, data processing and quality....23

SI4: Contamination estimates....32

SI5: Uniparental markers – mitochondrial DNA genomes and Y chromosomes of Neandertals....38

SI6: Basic set of filters for nuclear analyses....50

SI7: Principal Component Analysis, lineage attribution and divergence estimates....59

SI8: Split times of Neandertals and the neighbour-joining tree of nuclear genomes....69

SI9: Neandertal Population Relationships....75

SI10: Proximity to the introgressing Neandertals in present-day and ancient modern humans....86

SI11: No evidence of a recent early modern human admixture into late Neandertals....101

SI12: Assessing the catalogue of human-specific fixed derived changes using the late Neandertals....106

Supplementary Information 1

Information about the archaeological sites and specimens

Les Cottés (France)

Les Cottés is a small limestone cave located in the Southwestern part of the Parisian Basin in France. The site was discovered in the late nineteenth century, and skeletal remains of an anatomically modern human were found in an Upper Palaeolithic context at the entrance of the cave[1]. A diamicton of mostly centimetre-sized limestone clasts and a sandy clay matrix preserve Late Middle and Early Upper Palaeolithic occupations in the entrance of the cave[1]. The sequence contains five major stratigraphical units: one Mousterian, one Châtelperronian, one Proto-Aurignacian and two Early/Middle Aurignacian[2]. The Mousterian and the transitional Châtelperronian industries have been attributed to Neandertals, whereas Proto-Aurignacian and Early Aurignacian have been associated with the first early modern humans in Europe. The Mousterian and Châtelperronian in Les Cottés were separated from the Proto-Aurignacian layers with a sterile deposit that spans approximately 1,000 years. Radiometric dates on bones and on sediment indicate that the site was occupied from at least 45,000 years ago up until around 35,000 years ago[3,4]. Therefore, Les Cottés is one of the few sites with a complete sequence covering the Middle to Early Upper Palaeolithic period in Europe characterized by the disappearance of Neandertals and the arrival of early modern humans.

The tooth studied here was excavated on August 19th 2008 by M. Soressi and her team. As with any other faunal remain bigger than 2.5 cm, the exact location and original context of the tooth was recorded using a hand-held computer connected to a total station and a unique identification number (*Z4-1514*) was attributed to it. The tooth was washed with water, dried and stored in a zip-locked plastic bag in a non-air conditioned storage room. It was recognized as a likely human tooth by W. Rendu and S. Renou (PACEA, Talence, France) in 2012 and has been - from then on - stored in a fridge.

The radiocarbon dating on the root was prepared at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, using the pre-treatment method established in Talamo and Richards[5]. The tooth resulted in a calendar age ranging from 43,410 to 42,920 cal BP with one standard deviation error and between 43,740 to 42,720 cal BP with two standard deviation errors (MAMS-26196).

Troisième caverne of Goyet (Belgium)

The Troisième caverne of Goyet is a part of a large cave system in the Mosan Basin in Belgium, and was excavated in the second half of the nineteenth and the beginning of the twentieth century, as well as recently at the end of the twentieth century. The Mosan Basin is located in Southern Belgium and has yielded numerous Neandertal remains[6]. The main excavation at the Troisième caverne of Goyet in 1868 revealed Palaeolithic industries[7]. The following studies found rich archaeological evidence of human occupations from the Mousterian, the transitional Lincombian-Ranisian-Jerzmanowician (LRJ) industry, as well as the Upper Palaeolithic, represented by the Aurignacian, the Gravettian and the Magdalenian industries[8]. As the early excavations of the site represented a mix of material from different periods, a reassessment of the faunal and human material from the site started in 2008. Recent radiocarbon dating results, in combination with isotopic analyses[9] and obtained mitochondrial DNA sequences[10,11], assigned specimens either to late Neandertals or to modern humans.

A minimum number of four different adolescent or adult Neandertal individuals, one child and one neonate were identified in the Goyet collection[11,12]. The fragment of the right femur *Goyet Q56-1* analysed in this study was previously characterized as having a Neandertal mitochondrial genome and was directly radiocarbon dated to 43,000 – 42,080 cal BP (two standard deviation errors, GrA-46170)[11]. Given the lack of field data from the Troisième caverne of Goyet and the regional chronocultural context, it is impossible to securely assign Goyet Q56-1 to either the Mousterian or LRJ[11].

Spy (Belgium)

The cave of Spy is one of the richest prehistoric sites in Belgium and located only 20 kilometres from the Troisième caverne of Goyet in the Mosan Basin. The first human remains were discovered in stratigraphic context and in association with lithic material in 1886[13]. Since then, numerous excavations have been carried out at the site[14]. Mousterian, LRJ and Aurignacian industries have been identified at Spy[8]. The two incomplete adult Neandertal skeletons, *Spy I* and *Spy II*, were found in the deepest level of the terrace of the Spy cave[13].

Spy 94a was identified during the reassessment of the Spy collections in 2009[15]. It is an upper right molar (M3) with an associated alveolar bone that was directly radiocarbon dated to 39,150-37,880 cal BP (one standard deviation error, GrA-32623) and attributed to the *Spy I* Neandertal[15]. Recent re-dating of both Spy Neandertals makes them the youngest Neandertals identified in Northwestern Europe and contemporaneous with the transitional LRJ industry.

Vindija cave (Croatia)

The Vindija cave is located in the Northwestern part of Croatia. From the beginning of the excavations in the 1970s, Vindija cave has yielded numerous hominin remains from the Middle and Upper Palaeolithic[16-19]. There are 14 stratigraphical layers at the site, labelled A through N, with the upper layers corresponding to the Holocene time period. Layers D-N correspond to the Pleistocene and are important for understanding the disappearance of Neandertals and the arrival of early modern humans in this region[20-22]. More than 100 hominin remains were found in the Pleistocene layers of Vindija cave and attributed to both Neandertals and modern humans based on their morphology[16-19]. Low-coverage nuclear genomes were retrieved from *Vindija 33.16* (directly radiocarbon dated to 44,821-40,780 cal BP, one standard deviation error[23]), *Vindija 33.25* (originates from the layer I of the Vindija cave and did not have enough collagen preserved to be directly dated) and *Vindija 33.26* (directly radiocarbon dated to 48,426-47,013 cal BP, one standard deviation error, OxA-V-2291-18)[24]. Recently, a high coverage genome was obtained from *Vindija 33.19*, directly radiocarbon dated to 45,300 ± 2,300 BP before calibration (OxA-32278)[25] and with a date close to the limit of radiocarbon dating after calibration with one standard deviation error.

The morphologically undiagnostic bone fragment *Vindija 87* (Vi87) was found in layer G1 of the Vindija cave. The G1 layer of Vindija cave was considered to be one of the last Neandertal occupations in Europe based on the direct radiocarbon dating of two Neandertal specimens recovered from this layer[26]. Therefore, we removed 574 mg of bone powder from *Vindija 87* using a sterile dentistry drill and sent it to the Oxford Radiocarbon Accelerator Unit for direct radiocarbon dating. Only 4.2 mg of collagen was extracted, indicating a low preservation. However, the material was sufficient to obtain a minimum age for *Vindija 87,* of at least ~44000 ^{14}C years/~47,000 cal BP (OxA-X-2634-52).

Mezmaiskaya cave (Russia)

Mezmaiskaya cave, located 1310 m above sea level in the Azish-Tau Ridge (Lago-Naki highland, Krasnodar Kray, Russia), records the longest and continuous Late Middle Palaeolithic (LMP) and Upper Palaeolithic (UP) sedimentary sequence in the Northwestern Caucasus, reaching a maximum depth of 5 m and yielded tens of thousands of lithic and organic artefacts, and a very rich faunal assemblage. Since 1987, L. Golovanova excavated more than 80 m^2 in the site[27-29]. Currently, based on section Z11F11 (the deepest section inside the cave) from the 2016 excavation, the stratigraphic sequence of Mezmaiskaya cave consists of 6 Holocene and 20 Pleistocene strata, including 8 stratified UP layers dating from

c. 12 to 39 ka cal BP (from top to bottom): 1-3, 1-4, 1A1, 1A1/1A2, 1A2, 1B1, 1B2, and 1C, beneath which is found a sterile stratum 1D, lying at the LMP to UP transition and containing volcanic ash. The LMP sequence consists of 7 layers dating from c. 40,000 to 70,000 BP (from top to bottom): 2, 2A, 2B1, 2B2, 2B3, 2B4, and 3. The lowest Pleistocene layers (4–7) contain no archaeological material.

The Neandertal fossils discovered at Mezmaiskaya cave are: an almost-complete skeleton of a neonate (*Mezmaiskaya 1;* about 2 weeks after birth), which was recovered in 1993 in anatomical position in square M-26 in the lowermost 3–5 cm of Layer 3, the oldest LMP layer; an isolated permanent tooth (*Mezmaiskaya 3*), which was found later and is also from Layer 3; and 24 cranial fragments of an infant (*Mezmaiskaya 2*, 1–2 years of age), which were found in 1994 in square N-19 in a pit that originates in Layer 2, the youngest LMP level, and penetrates into the lower LMP layers 2A, 2B1, and 2B2.

The pit, in which skull fragments of *Mezmaiskaya 2* were recovered, was 80 cm wide and 40 cm deep, had well defined outlines, and was filled with sediments of Layer 2, and partly with sediments of layers 2A and 2B1, 2B2 that intruded into the pit probably as a result of erosion. The pit was adjacent to and partly intruded by a huge limestone block that resulted from a collapse of the cave ceiling. The *Mezmaiskaya 2* cranial fragments have been found in different depths within the pit. The uppermost fragments were lying on the sloping surface of the pit wall near the upper edge of the pit (depth of 93-96 cm below the datum; fragments 1-7), while the lowermost fragments were recovered from the bottom of the pit (depth 116 cm below the datum; fragments 8 and 9). Other fragments were found at intermediate depths between the upper and lower groups. The cranial fragments belong to the frontal, as well as left and right parietal bones, which are connected to each other but have some post-mortem deformation. The presence of the fronto-parietal suture suggests the age of the child within 1-2 years after birth. Only a natural limestone fragment and one artefact (a small flint shatter) were found on a sloping surface of the pit together with the cranial remains. The fragmentary character of the *Mezmaiskaya 2* cranial remains and their dispersed stratigraphic position within the pit, in which they were found, allow us to suggest that the pit is not an intentional burial.

Since *Mezmaiskaya 2* is stratigraphically overlying *Mezmaiskaya 1*, *Mezmaiskaya 2* should be younger than *Mezmaiskaya 1* (~70–60 ka $BP_{ESR/LU}$ suggested for Layer 3 on the basis of ESR dating[30]). This is in line with the age ~42-37 ka $BP_{ESR/LU}$ for Layer 2. The direct ultrafiltration radiocarbon dating of the *Mezmaiskaya 2* specimen produced a result of 39,700 ± 1,100 ^{14}C BP (OxA-21839;[31]: Table 1 and SOM Table S3 in Pinhasi *et al.*[31]) calibrated to

44,600-42,960 cal BP (one standard deviation error) and 45,600-42,300 cal BP (two standard deviation errors). The *Mezmaiskaya 2* specimen was very well preserved, and yielded 14.6% collagen by weight (in modern unadulterated bone, 20% by weight is collagen) and the C:N atomic ratio was 3.2 (in modern bone, this is 3.2). There is therefore no reason to doubt the accuracy of this result given the preservation state of the specimen, and the direct date of *Mezmaiskaya 2* indicates that the younger radiocarbon date for the *Mezmaiskaya 1* specimen obtained previously[32] is incorrect. This new date is further supported by the branch shortening estimate from the *Mezmaiskaya 1* mitochondrial genome which gives an age range of between 64,756 and 139,751 years ago (Table S5.2, Supplementary Information S5). The ultrafiltration radiocarbon dating results demonstrate that the *Mezmaiskaya 2* infant from Mezmaiskaya cave represents the youngest known Neandertal fossil from the Caucasus known to date, indicating that it is likely that Neandertals did not survive at Mezmaiskaya after 36.8ka cal BP (two standard deviation errors) or 39.39ka cal BP (one standard deviation error) based on the models of all of the available AMS and conventional radiocarbon ages of Mezmaiskaya cave[31].

References for SI1:

1 Soressi, M. *et al.* (Association des Publications Chauvinoises, 2010).

2 Roussel, M. & Soressi, M. in *Le Paléolithique supérieur ancien de l'Europe du Nord-Ouest. Réflexions et synthèses à partir d'un projet collectif de recherche sur le centre et le sud du Bassin parisien; actes du colloque de Sens (15-18 Avril 2009).* 283-298 (Société Préhistorique Française).

3 Talamo, S., Soressi, M., Roussel, M., Richards, M. & Hublin, J.-J. A radiocarbon chronology for the complete Middle to Upper Palaeolithic transitional sequence of Les Cottés (France). *Journal of Archaeological Science* **39**, 175-183, doi:10.1016/j.jas.2011.09.019 (2012).

4 Jacobs, Z., Li, B., Jankowski, N. & Soressi, M. Testing of a single grain OSL chronology across the Middle to Upper Palaeolithic transition at Les Cottés (France). *Journal of Archaeological Science* **54**, 110-122, doi:10.1016/j.jas.2014.11.020 (2015).

5 Talamo, S. & Richards, M. A comparison of bone pretreatment methods for AMS dating of samples> 30,000 BP. *Radiocarbon* **53**, 443-449 (2011).

6 Toussaint, M. & Pirson, S. Neandertal studies in Belgium: 2000–2005. *Periodicum biologorum* **108**, 373-387 (2006).

7 Dupont, E. *L'Homme pendant les âges de la pierre dans les environs de Dinant-sur-Meuse*. Vol. 2ème édition (C. Muquardt Ed., 1872).

8 Flas, D. *La transition du Paléolithique moyen au supérieur dans la plaine septentrionale de l'Europe. Les problématiques du Lincombien-Ranisien-Jerzmanowicien*, Université de Liège, Belgium, (2006).

9 Wißing, C. *et al.* Isotopic evidence for dietary ecology of Late Neandertals in North-Western Europe. *Quat. Int* **10** (2015).

10 Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr Biol* **26**, 827-833, doi:10.1016/j.cub.2016.01.037 (2016).

11 Rougier, H. *et al.* Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci Rep* **6**, 29005, doi:10.1038/srep29005 (2016).

12 Rougier, H. *et al.* The Troisième caverne of Goyet (Belgium): An exceptional site with both Neandertal and Upper Paleolithic human remains. *PESHE* **5**, 211 (2016).

13 Fraipont, J. & Lohest, M. *La race humaine de Néanderthal ou de Canstadt en Belgique: recherches ethnographiques sur des ossements humains découverts dans les dépôts quaternaires d'une grotte à Spy et détermination de leur âge géologique*. Vol. 7 (Vanderpoorten, 1887).

14 Rougier, H. *et al.* Collections de la Grotte de Spy:(re) découvertes et inventaire anthropologique. *Notae praehistoricae* **24**, 181-190 (2004).

15 Semal, P. *et al.* New data on the late Neandertals: direct dating of the Belgian Spy fossils. *Am J Phys Anthropol* **138**, 421-428, doi:10.1002/ajpa.20954 (2009).

16 Malez, M., Smith, F. H., Radovcic, J. & Rukavina, D. Upper Pleistocene hominids from Vindija, Croatia, Yugoslavia. *Current Anthropology* **21**, 365-367 (1980).

17 Smith, F. H., Boyd, D. C. & Malez, M. Additional upper pleistocene human remains from Vindija cave, Croatia, Yugoslavia. *American Journal of Physical Anthropology* **68**, 375-383 (1985).

18 Smith, F. H. & Ahern, J. C. Additional cranial remains from Vindija cave, Croatia. *American Journal of Physical Anthropology* **93**, 275-280 (1994).

19 Wolpoff, M. H., Smith, F. H., Malez, M., Radovčić, J. & Rukavina, D. Upper pleistocene human remains from Vindija cave, Croatia, Yugoslavia. *American Journal of Physical Anthropology* **54**, 499-545 (1981).

20 Ahern, J. C. M., Karavanić, I., Paunović, M., Janković, I. & Smith, F. H. New discoveries and interpretations of hominid fossils and artifacts from Vindija Cave,

Croatia. *Journal of Human Evolution* **46**, 27-67, doi:10.1016/j.jhevol.2003.09.010 (2004).

21 Karavanić, I. & Smith, F. H. The Middle/Upper Paleolithic interface and the relationship of Neanderthals and early modern humans in the Hrvatsko Zagorje, Croatia. *Journal of Human Evolution* **34**, 223-248 (1998).

22 Janković, I. *et al.* Vindija Cave and the modern human peopling of Europe. *Collegium antropologicum* **30**, 457-466 (2006).

23 Serre, D. *et al.* No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol* **2**, e57 (2004).

24 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

25 Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, doi:10.1126/science.aao1887 (2017).

26 Higham, T., Ramsey, C. B., Karavanić, I., Smith, F. H. & Trinkaus, E. Revised direct radiocarbon dating of the Vindija G1 Upper Paleolithic Neandertals. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 553-557 (2006).

27 Golovanova, L. V. & Doronichev, V. B. The Early Upper Paleolithic of the Caucasus in the West Eurasian Context. *L'Aurignacien de la Grotte Yaftehet son Contexte (Fouilles 2005-2008)* eds Otte M, Shidrang S, Flas D (Editions ERAUL 132, Univ de Liège, Liège, Belgium), 137–160 (2012).

28 Golovanova, L. *et al.* The epipaleolithic of the Caucasus after the last glacial maximum. *Quaternary International* **337**, 189-224 (2014).

29 Golovanova, L. V. Les hommes de Néandertal du Caucase du Nord: Entre l'ouest et l'est. *L'Anthropologie* **119**, 254-301 (2015).

30 Skinner, A. R. *et al.* ESR dating at Mezmaiskaya Cave, Russia. *Appl Radiat Isot* **62**, 219-224, doi:10.1016/j.apradiso.2004.08.008 (2005).

31 Pinhasi, R., Higham, T. F., Golovanova, L. V. & Doronichev, V. B. Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *Proc Natl Acad Sci U S A* **108**, 8611-8616, doi:10.1073/pnas.1018938108 (2011).

32 Ovchinnikov, I. V. *et al.* Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**, 490-493 (2000).

Supplementary Information 2

Sampling, library preparation and initial contamination estimates

Pre-treatment of the bone/tooth powder and DNA extraction

All specimens were sampled in dedicated clean room facilities, either at the Royal Belgium Institute for Natural Sciences in Brussels (*Spy 94a*) or at the Max Planck Institute for Evolutionary Anthropology in Leipzig (all other specimens). A thin layer of surface was removed from the chosen sampling area of each specimen using a sterile dentistry drill. Between 28 mg and 104 mg of tooth or bone powder was obtained by drilling and split evenly into two *aliquots* (Table S2.1). One powder *aliquot* was extracted using a silica-based method[1] in the implementation of Korlević *et al.*[2] with no prior pre-treatment of the powder ("untreated" extraction). The second powder aliquot was treated with 0.5% sodium hypochlorite solution (Roth)[2] prior to the DNA extraction in an attempt to remove some of the microbial and present-day human DNA contamination that can be frequently found in ancient specimens[3-5]. *Les Cottés Z4-1514* was sampled at two different spots on three separate occasions. Two powder *aliquots* were obtained by drilling into the same spot at the cementoenamel junction of the tooth (and later converted to extracts E1254 and E2889, Table S2.1), while the third powder aliquot was taken from the dentin at the apex of the root directly below the cementum (extract E3142). The second and the third powder *aliquots* of *Les Cottés Z4-1514* were treated with 0.5M sodium phosphate buffer prior to 0.5% sodium hypochlorite treatment[2].

Generation of a first set of DNA libraries, mitochondrial DNA (mtDNA) enrichment and sequencing

Five or ten µL (10 or 20% of the total volume) of each extract were converted into DNA libraries using a single-stranded method[6] with modifications described by Korlević *et al.*[2]. The first libraries prepared from *Vindija 87, Goyet Q56-1* and *Les Cottés Z4-1514* were treated with *E. coli* uracil-DNA-glycosylase (UDG) and *E. coli* endonuclease VIII (Endo VIII) in order to remove uracils from the interior of ancient DNA molecules[7,8]. All other libraries were prepared without this treatment in order to maximize the recovery of endogenous DNA fragments[9]. Except the first extract of *Les Cottés Z4-1514,* a control oligonucleotide was spiked into each aliquot of DNA extract used for library preparation, in order to monitor the efficiency of library preparation[10]. The total number of molecules in each

library was determined by digital droplet PCR (ddPCR) (QX200 system, Bio-Rad) by using 1 µL of a 5,000-fold library dilution in EB buffer (10 mM Tris-HCl pH 8.0, 0.05% Tween 20) as template in an Eva Green emulsion PCR assay (QX200 ddPCR EvaGreen Supermix, Bio-Rad) with primers IS7 and IS8[11]. The number of library molecules derived from the spiked-in oligonucleotide molecules was determined by digital droplet PCR by using 1 µL of an undiluted library as a template in a Droplet PCR Supermix assay (Bio-Rad) with primers IS7[11] and CL107[6] and the probe CL118[10]. The libraries were amplified into plateau with AccuPrime Pfx DNA polymerase (Life Technologies)[12] and labelled with two unique indexes[2,13]. Fifty µL of each amplified library were purified using the MinElute PCR purification kit (Qiagen) and eluted in 30 µL TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). DNA concentrations of the amplified libraries were determined using a NanoDrop 1000 Spectrophotometer.

The amplified libraries were pooled with libraries from other experiments and heteroduplices were removed in a one-cycle PCR reaction using Herculase II Fusion DNA polymerase (Agilent Technologies)[12] with primers IS5 and IS6[10,14]. An aliquot of each amplified library was additionally enriched for the hominin mitochondrial DNA (mtDNA) using a bead-based hybridization method[15-17] and modern human mtDNA as a bait. The pools of libraries were sequenced on Illumina MiSeq and HiSeq 2500 platforms in a double index configuration (2x76 cycles)[13]. Base calling was done using *Bustard* (Illumina) for the MiSeq runs and *FreeIbis*[18] for the HiSeq runs.

Sequence data processing

Adapters were trimmed and overlapping paired-end reads were merged into single sequences using *leeHom*[19]. The Burrows-Wheeler Aligner (BWA, version: 0.5.10-evan.9-1-g44db244; https://github.com/mpieva/network-aware-bwa)[20] was used to align the shotgun data to the modified human reference GRCh37 from the 1000 Genomes project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/) that additionally includes the revised Cambridge Reference Sequence (NC_01290), the 1000 Genomes Phase 2 decoy sequences, the ΦX174 genome (NC_001422) and the human herpesvirus (NC_007605). Sequences obtained after mtDNA capture were aligned to the revised Cambridge Reference Sequence (NC_01290) using BWA. BWA parameters were adjusted for ancient DNA ("-n 0.01 –o 2 –l 16500"), to allow for more mismatches and indels and to turn off the seeding[8]. Only sequence reads showing perfect matches to one of the expected index combinations were retained for subsequent analyses. PCR duplicates were

removed using *bam-rmdup* (version: 0.6.3; https://bitbucket.org/ustenzel/biohazard) by calling a consensus from sequences with identical alignment start and end coordinates. SAMtools (version: 1.3.1)[21] was used to filter for fragments that were longer than 35 base pairs and that had a mapping quality of at least 25.

Assessing the impact of hypochlorite treatment of bone/tooth powder

As determined by digital droplet PCR, hypochlorite treatment considerably reduced the number of DNA fragments recovered in the libraries (Table S2.1). This was mainly driven by a loss of non-endogenous DNA, as indicated by the substantial increase in the proportion of DNA sequences aligning to the human reference genome after hypochlorite treatment (Fig. 1B, Table S2.1). The increase in the percentage of the aligned sequences longer than 35 base pairs ranged from 5.6-fold for *Goyet Q56-1* (from 4.22% to 23.98%), 6.1-fold for *Vindija 87* (from 0.76% to 4.62%), 6.9-fold for *Mezmaiskaya 2* (from 1.45% to 9.99%), 53.5-fold for *Spy 94a* (from 0.7% to 37.43%) to 161-fold for *Les Cottés Z4-1514* (from 0.15% to 24.21%). A chi-square (χ^2) test was used to determine significant differences (denoted with ** in Fig.1B for $\alpha << 0.01$) to the untreated powder.

Using the number of unique library molecules (determined by ddPCR) and the fraction of library molecules that yield "informative" sequences (sequences of at least 35bp that could be mapped to the reference genome) it is possible to predict the genomic coverage that could be obtained by sequencing a library to exhaustion[10] (Table S2.1). Predicted genomic coverage is 1.3 to 12.6 times lower in the libraries prepared from hypochlorite-treated sample powder. However, the stark increase in the proportion of informative sequences after hypochlorite treatment substantially reduces the costs of genome sequencing due to the depletion of microbial DNA, making the generation of the low coverage genomes economically feasible.

We also observed that omitting the UDG/endonuclease VIII treatment for *Vindija 87* extract led to an additional increase in the percentage of sequences aligning to the human reference genome (from 4.6 to 27.2%), indicating that the endogenous DNA in this specimen is more prone to uracil excision and strand cleavage than the microbial DNA, presumably because the latter is of more recent origin[9]. In contrast, omitting uracil removal for the *Goyet Q56-1* extract did not substantially change the percentage of mapped sequences, indicating that most microbial contamination in this bone is as damaged as its endogenous DNA.

Authentication of ancient DNA in the initial libraries based on the nucleotide substitutions

Deamination of cytosine (C) to uracil (U) residues, which occurs primarily at single-stranded overhangs in ancient DNA molecules, leads to C-to-T substitution in ancient DNA sequence alignments, which are particularly frequent close to the alignment ends[7]. Elevated C-to-T substitutions can therefore provide evidence for the presence of authentic ancient DNA in specimens[5,23]. This also applies to DNA treated with UDG and endonuclease VIII, as UDG does not efficiently excise uracils from molecule ends[8]. In libraries prepared from untreated sample powder, C-to-T substitution frequencies range from 4.0% to 44.6% at the 5'-ends of molecules and between 12.8% and 41.8% of 3'-ends (Table S2.2). For *Les Cottés Z4-1514* and *Spy 94a* C-to-T substitution frequencies increase substantially if sequences carry a C-to-T substitution at the opposing end ("conditional" substitution frequencies; see Table S2.2), indicating that both endogenous ancient DNA as well present-day human DNA contamination[24] are present. After hypochlorite treatment of the bone/tooth powder, C-to-T substitutions ranged from 8.0% to 57.5% on the 5'-ends and from 28.0% to 64.3% on the 3'-ends of fragments and remain relatively stable after filtering for the presence of C-to-T substitutions at the opposing ends. This analysis, though limited in resolution, thus provides no evidence for the presence of human DNA contamination after hypochlorite treatment.

Phylogenetic inferences based on "diagnostic" positions and mtDNA contamination estimates

To estimate the proportion of present-day human DNA contamination in the libraries, we studied the state of fragments enriched for the hominin mtDNA that overlapped positions that are "diagnostic" for each branch in the mtDNA tree relating present-day humans, Neandertals, Denisovans and the hominin from Sima de los Huesos[25]. To diminish the influence of deamination-derived substitutions, all forward strands were ignored (in the orientation as sequenced) if the informative state was a C and all reverse strands were ignored if the informative state was a G. We determined the percentage of derived variants supporting the state diagnostic for each branch, using all unique mtDNA fragments, as well as only those fragments that had a terminal C-to-T difference to the reference genome for the UDG/endonuclease VIII treated libraries or the fragments with a C-to-T difference within the first three and/or the last three alignment positions for the libraries that were not treated with UDG/endonuclease VIII.

When using all fragments, between 50% and 99.33% fragments supported the Neandertal branch and between 0.48% and 58.93% fragments supported the modern human branch in the libraries prepared from the extracts with no prior pre-treatment of the bone/tooth powder (Table S2.3 and Table S2.4). When restricting the analysis to putatively deaminated DNA fragments, the support of the Neandertal branch increased over 69% for *Spy 94a* and to over 98% for all other individuals, compatible with the human DNA in the libraries originating from the recent contamination that is less deaminated than the endogenous Neandertal DNA. However, we observed a substantial decrease in the proportion of present-day human DNA contamination in all fragments after 0.5% hypochlorite treatment of the same bone/tooth powder (Tables S2.3, S2.4 and S2.5). The decrease was from 74.85% to 6.64% for *Spy 94a*, from 58.93% to 1.17% for *Les Cottés Z4-1514*, from 9.87% to 5.25% for *Goyet Q56-1* and from 1.47% to 0.26% for *Mezmaiskaya 2*, respectively. A chi-square (χ^2) test was used to determine significant differences (denoted with * in Fig.1C) compared to the untreated powder. The only specimen where we observed a slight increase in the point estimate of the proportion of present day human DNA contamination after 0.5% hypochlorite treatment was *Vindija 87* (from 0.52% to 0.85%). However, this increase was statistically not significant (χ^2 test equals to 0.858 with one degree of freedom, p-value = 0.3542).

Production of additional libraries from selected extracts and shotgun sequencing

Based on the assessment of the nuclear DNA content of the initial libraries, as well as the proportion of present-day human DNA contamination, we selected the extracts with the high percentage of endogenous DNA and low levels of present-day human DNA contamination to produce additional single-stranded DNA libraries (Table S2.6)[2,6]. All of the additional libraries have been produced without the UDG and endonuclease VIII treatment and by using ten µL of extract per library as input in library preparation and as described above. A total of 23 libraries from five late Neandertals were generated and sequenced on 50 lanes of the Illumina HiSeq 2500 platform in rapid mode, using double indexing configuration (2x76 bp)[13]. Libraries were either sequenced individually, *i.e.* one library per HiSeq lane, or pooled together for the sequencing (Table S2.6). In both cases heteroduplices were removed prior to the sequencing by one-cycle PCR with Herculase II Fusion DNA polymerase (Agilent)[12] and primers IS5 and IS6[11,14].

For the libraries that were sequenced individually, between 10% and 30% of the pool of four ΦX174 libraries with AAAAAAA, TTTTTTT, GGGGGGG and CCCCCCC as P5 and P7 indices was spiked into the sequencing library (Table S2.6) in order to retain the

complexity of index reads for base calling with *FreeIbis*[187,25]. For all sequencing runs that contained pools of four or more libraries, 0.5% of a regular double-indexed ΦX174 control library was spiked-in prior to sequencing (Table S2.6). Base calling for all sequencing runs was done using *FreeIbis*[18]. Adapters were clipped and overlapping forward and reverse reads were merged into single sequences using *leeHom*[19].

References for SI2:

1 Dabney, J. *et al.* 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, doi:10.1073/pnas.1314445110 (2013).

2 Korlević, P. *et al.* Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87-93, doi:10.2144/000114320 (2015).

3 Pääbo, S. *et al.* Genetic analyses from ancient DNA. *Annu Rev Genet* **38**, 645-679, doi:10.1146/annurev.genet.37.110801.143214 (2004).

4 Gilbert, M. T., Bandelt, H. J., Hofreiter, M. & Barnes, I. Assessing ancient DNA studies. *Trends Ecol Evol* **20**, 541-544, doi:10.1016/j.tree.2005.07.005 (2005).

5 Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* **20**, 231-236, doi:10.1016/j.cub.2009.11.068 (2010).

6 Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* **8**, 737-748, doi:10.1038/nprot.2013.038 (2013).

7 Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* **104**, 14616-14621, doi:10.1073/pnas.0704665104 (2007).

8 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

9 Gansauge, M. T. & Meyer, M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res* **24**, 1543-1549, doi:10.1101/gr.174201.114 (2014).

10 Glocke, I. & Meyer, M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. *Genome Research* (2017).

11 Meyer, M. & Kircher, M. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* **2010**, pdb.prot5448, doi:10.1101/pdb.prot5448 (2010).

12 Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87-94, doi:10.2144/000113809 (2012).

13 Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**, e3, doi:10.1093/nar/gkr771 (2012).

14 Gansauge, M. T. *et al.* Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res*, doi:10.1093/nar/gkx033 (2017).

15 Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004, doi:10.1371/journal.pone.0014004 (2010).

16 Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

17 Welker, F. *et al.* Palaeoproteomic evidence identifies archaic hominins associated with the Chatelperronian at the Grotte du Renne. *Proc Natl Acad Sci U S A* **113**, 11162-11167, doi:10.1073/pnas.1605834113 (2016).

18 Renaud, G., Kircher, M., Stenzel, U. & Kelso, J. freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208-1209, doi:10.1093/bioinformatics/btt117 (2013).

19 Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res* **42**, e141, doi:10.1093/nar/gku699 (2014).

20 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

21 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

22 Gansauge, M. T. & Meyer, M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res* **24**, 1543-1549, doi:10.1101/gr.174201.114 (2014).

23 Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131, doi:10.1371/journal.pone.0034131 (2012).

24 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

25 Meyer, M. *et al.* Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504-507, doi:10.1038/nature17405 (2016).

Table S2.1 Characteristics of the first set of DNA libraries prepared from the five specimens. Summary statistics of the shotgun data obtained from untreated bone/tooth powder and after 0.5% hypochlorite treatment (highlighted in grey).

Specimen	Amount of powder used for DNA extraction (mg)	Pre-treatment of the powder	Extract ID	Library ID	Volume of extract used for library preparation (µL)	UDG/Endo VIII treatment	Total number of DNA molecules in the library (ddPCR)	Number of spike-in molecules in the library (ddPCR)	Number of sequences generated	Number of mapped sequences (≥35bp, MQ≥25)	% Sequences (≥35bp) mapped	Estimated genomic coverage in the library	Estimated genomic coverage in the extract
Vindija 87	**57.5**	no	E2549	A9098	**10**	yes	1.95E+10	9.00E+05	1.04E+06	4.95E+03	**0.76**	**1.39**	**6.93**
	46.2	0.5% hypochlorite	E2550	A9099	**10**	yes	1.13E+09	6.35E+05	1.43E+06	3.26E+04	**4.62**	**0.38**	**1.92**
				R5005	**20**	no	4.70E+08	9.15E+05	5.86E+05	6.53E+04	**27.18**	**1.08**	**2.71**
Goyet Q56-1	**57.1**	no	E2555	A9104	**10**	yes	1.11E+10	7.45E+05	1.17E+06	3.60E+04	**4.22**	**6.73**	**33.64**
	41.8	0.5% sodium hypochlorite	E2556	A9105	**10**	yes	5.15E+08	8.05E+05	1.52E+06	2.27E+05	**23.98**	**1.51**	**7.57**
				R5006	**20**	no	2.59E+08	5.45E+05	4.74E+05	7.31E+04	**29.31**	**1.07**	**2.68**
Mezmaiskaya 2	**28.0**	no	E2829	R1916	**10**	no	1.31E+10	1.12E+06	1.20E+06	8.61E+03	**1.45**	**1.58**	**7.91**
	29.0	0.5% sodium hypochlorite	E2830	R1917	**10**	no	1.09E+09	9.70E+05	1.31E+06	6.49E+04	**9.99**	**0.91**	**4.56**
Les Cottés Z4-1514	**19.0**	no	E1254	L9451	**5**	yes	2.44E+10	-	2.00E+06	1.02E+03	**0.14**	**0.35**	**3.51**
	16.0	sodium phosphate, 0.5% sodium hypochlorite	E2889	R5022	**10**	no	3.72E+08	1.08E+06	4.70E+05	6.64E+04	**24.21**	**0.83**	**4.13**
	9.4	sodium phosphate, 0.5% sodium hypochlorite	E3142	A9309	**10**	no	9.30E+07	9.00E+05	9.72E+06	2.16E+06	**63.98**	**0.48**	**2.39**
Spy 94a	**13.8**	no	E3290	A9336	**10**	no	1.36E+09	7.75E+05	2.86E+06	8.95E+03	**0.70**	**0.08**	**0.40**
	14.5	0.5% sodium hypochlorite	E3342	R5556	**10**	no	1.41E+08	1.00E+06	1.33E+07	1.74E+06	**37.43**	**0.39**	**1.94**

Table S2.2 Frequencies of C-to-T substitutions at terminal positions of the sequence alignments for the first set of libraries generated by shotgun sequencing. The C-to-T substitution frequencies are determined on the mapped fragments longer than 35 base pairs with mapping quality of at least 25 (MQ≥25) reported in the Table S2.1 and calculated 95% binomial confidence intervals are provided in brackets. Libraries prepared from extracts after 0.5% hypochlorite treatment of bone/tooth powder are highlighted in grey.

Specimen	Pre-treatment of the powder	Library ID	All fragments		Fragments with C-to-T substitutions at the opposing end	
			5' C→T (%) [95% CI]	3' C → T (%) [95% CI]	5' C→T (%) [95% CI]	3' C → T (%) [95% CI]
Vindija 87	no	A9098	27.2 *[24.3 - 30.1]*	41.8 *[38.7 - 44.9]*	25.6 *[17.7 - 35.4]*	37.7 *[26.6 - 50.3]*
	0.5% hypochlorite	A9099	28.1 *[26.9 - 29.2]*	45.1 *[43.9 - 46.3]*	26.7 *[23.2 - 30.2]*	47.2 *[41.7 - 52.2]*
		R5005	57.4 *[56.6 - 58.2]*	64.3 *[63.4 - 65.1]*	55.7 *[53.4 - 58.0]*	61.8 *[59.5 - 64.2]*
Goyet Q56-1	no	A9104	7.1 *[6.5 - 7.7]*	29.9 *[28.6 - 31.3]*	7.7 *[5.1 - 11.4]*	33.3 *[21.6 - 44.0]*
	0.5% sodium hypochlorite	A9105	8.0 *[7.8 - 8.3]*	28.0 *[27.5 - 28.5]*	8.5 *[7.3 - 9.9]*	31.0 *[26.8 - 35.0]*
		R5006	24.0 *[23.4 - 24.7]*	48.9 *[47.7 - 50.2]*	25.0 *[22.0 - 28.3]*	50.6 *[45.4 - 55.7]*
Mezmaiskaya 2	no	R1916	44.6 *[42.4 - 46.7]*	40.8 *[38.4 - 43.3]*	37.0 *[28.9 - 44.4]*	31.4 *[24.9 - 38.7]*
	0.5% sodium hypochlorite	R1917	43.2 *[42.4 - 44.0]*	41.4 *[40.5 - 42.2]*	40.5 *[37.7 - 43.4]*	40.5 *[37.7 - 43.3]*
Les Cottés Z4-1514	no	L9451	4.0 *[1.9 - 8.0]*	12.8 *[8.3 - 19.1]*	50 *[15 -85]*	100 *[64.6 -100]*
	sodium phosphate, 0.5% sodium hypochlorite	R5022	51.9 *[51.1 - 52.7]*	55.9 *[56.0 - 56.8]*	49.1 *[46.7 - 51.4]*	54.4 *[51.8 - 56.8]*
	sodium phosphate, 0.5% sodium hypochlorite	A9309	57.5 *[57.3 - 57.6]*	54.6 *[54.5 - 54.8]*	54.6 *[52.1 - 53.4]*	52.0 *[51.6 - 52.4]*
Spy 94a	no	A9336	18.0 *[16.2 - 19.7]*	21.1 *[19.0 - 23.2]*	31.9 *[22.1 - 43.6]*	34.4 *[23.9 - 46.6]*
	0.5% sodium hypochlorite	R5556	29.7 *[29.6 - 29.9]*	41.1 *[40.9 - 41.3]*	28.1 *[27.5 - 28.7]*	39.6 *[38.9 - 40.3]*

doi:10.1038/nature26151

Table S2.3 Proportion and number of mitochondrial DNA fragments in the first set of libraries matching the derived state at positions diagnostic for the human branch in the hominin mitochondrial tree. Fragments obtained by human mtDNA captures of the first set of libraries from five specimens, which overlap 14 positions where all 311 present-day humans differ from 18 Neandertals, 3 Denisovans, one hominin from Sima de los Huesos and a chimpanzee, were utilized. The results are shown for all fragments and only those fragments with terminal C-to-T substitutions. The number of fragments supporting the derived variant and the total number of observations are provided in brackets. Libraries prepared from extracts after 0.5% hypochlorite treatment of bone/tooth powder are highlighted in grey.

Specimen	Pre-treatment of the powder	Library ID	All fragments		Fragments with terminal C-to-T substitutions	
			%human [observations]	%all others [observations]	%human [observations]	%all others [observations]
Vindija 87	no	A9098	0.48 *[6/1,247]*	99.52 *[1,241/1,247]*	0 *[0/128]*	100 *[128/128]*
	0.5% hypochlorite	A9099	0.79 *[9/1,140]*	99.21 *[5,076/1,140]*	0 *[0/112]*	100 *[112/112]*
		R5005	0.85 *[13/1,527]*	99.15 *[1,514/1,527]*	0 *[0/422]*	100 *[422/422]*
Goyet Q56-1	no	A9104	9.87 *[54/578]*	90.14 *[521/578]*	0 *[0/29]*	100 *[29/29]*
	0.5% sodium hypochlorite	A9105	5.65 *[34/602]*	94.35 *[568/602]*	0 *[0/34]*	100 *[34/34]*
		R5006	5.24 *[20/382]*	94.76 *[362/382]*	0 *[0/67]*	100 *[67/67]*
Mezmaiskaya 2	no	R1916	1.47 *[47/3,201]*	98.45 *[3,154/3,201]*	0.31 *[2/187]*	99.46 *[3,154/3,201]*
	0.5% sodium hypochlorite	R1917	0.26 *[9/3,451]*	99.71 *[3,442/3,451]*	0.28 *[9/3,451]*	99.69 *[3,442/3,451]*
Les Cottés Z4-1514	no	L9451	58.93 *[33/56]*	41.07 *[23/56]*	0 *[0/7]*	100 *[7/7]*
	sodium phosphate, 0.5% sodium hypochlorite	R5022	1.17 *[3/256]*	98.83 *[253/256]*	0 *[0/63]*	100 *[63/63]*
	sodium phosphate, 0.5% sodium hypochlorite	A9309	4.12 *[8/194]*	95.88 *[186/194]*	0 *[0/52]*	100 *[52/52]*
Spy 94a	no	A9336	74.85 *[128/171]*	25.15 *[43/171]*	11.11 *[2/18]*	88.89 *[16/18]*
	0.5% sodium hypochlorite	R5556	6.64 *[377/5,678]*	93.36 *[5,298/5,678]*	0.79 *[6/761]*	99.21 *[755/761]*

doi:10.1038/nature26151

Table S2.4 Proportion and number of mitochondrial DNA fragments in the first set of libraries matching the derived state at positions diagnostic for the Neandertal branch in the hominin mitochondrial tree. Fragments obtained by human mtDNA captures of the first set of libraries from the five specimens, which overlap 19 positions where all 18 Neandertals differ from 311 present-day humans, 3 Denisovans, one hominin from Sima de los Huesos and a chimpanzee, were utilized. The results are shown for all fragments and only those fragments with terminal C-to-T substitutions. The number of fragments supporting the derived variant and the total number of observations are provided in brackets. Libraries prepared from extracts after 0.5% hypochlorite treatment of bone/tooth powder are highlighted in grey.

Specimen	Pre-treatment of the powder	Library ID	All fragments		Fragments with terminal C-to-T substitutions	
			%Neandertal [observations]	%all others [observations]	%Neandertal [observations]	%all others [observations]
Vindija 87	no	A9098	99.33 *[1,780/1,792]*	0.67 *[12/1,792]*	100 *[250/250]*	0 *[0/250]*
	0.5% hypochlorite	A9099	99.03 *[1,630/1,646]*	0.97 *[16/1,646]*	99.17 *[240/242]*	0.83 *[2/242]*
		R5005	99.0 *[2,271/2,294]*	1.0 *[23/2,294]*	99.53 *[845/849]*	0.47 *[4/849]*
Goyet Q56-1	no	A9104	86.04 *[635/738]*	13.96 *[103738]*	95.83 *[46/48]*	4.6 *[2/48]*
	0.5% sodium hypochlorite	A9105	93.17 *[778/835]*	6.83 *[58/835]*	98.73 *[78/79]*	1.27 *[1/79]*
		R5006	91.88 *[600/653]*	8.12 *[53/653]*	98.53 *[134/136]*	1.47 *[2/136]*
Mezmaiskaya 2	no	R1916	98.05 *[5,034/5,134]*	1.95 *[100/5,134]*	98.80 *[1,321/1,337]*	1.20 *[16/1,337]*
	0.5% sodium hypochlorite	R1917	99.09 *[4,892/4,937]*	0.91 *[45/4,937]*	99.60 *[1,256/1,261]*	0.40 *[5/1,261]*
Les Cottés Z4-1514	no	L9451	50.0 *[32/64]*	50.0 *[32/64]*	100 *[10/10]*	0 *[0/10]*
	sodium phosphate, 0.5% sodium hypochlorite	R5022	98.57 *[414/420]*	1.43 *[6/420]*	100 *[148/148]*	0 *[0/148]*
	sodium phosphate, 0.5%sodium hypochlorite	A9309	94.44 *[272/288]*	5.56 *[16/288*	98.06 *[101/103]*	1.94 *[2/103]*
Spy 94a	no	A9336	13.94 *[2/129]*	86.06 *[127/129]*	69.23 *[9/13]*	30.77 *[4/13]*
	0.5% sodium hypochlorite	R5556	92.60 *[7,249/7828]*	7.40 *[579/7,828]*	99.17 *[1,551/1,564]*	0.83 *[13/1,564]*

doi:10.1038/nature26151

Table S2.5 Proportion and number of mitochondrial DNA fragments in the first set of libraries matching the modern human or the Neandertal state. Fragments obtained by human mtDNA captures of the first set of libraries from the five specimens, which overlap 63 positions where all 18 Neandertals differ from 311 present-day humans, were utilized. The results are shown for all fragments and only those fragments with terminal C-to-T substitutions. The number of fragments supporting the derived variant and the total number of observations are provided in brackets. Libraries prepared from extracts after 0.5% hypochlorite treatment of bone/tooth powder are highlighted in grey.

Specimen	Pre-treatment of the powder	Library ID	All fragments		Fragments with terminal C-to-T substitutions	
			%Neandertal [observations]	%Human [observations]	%Neandertal [observations]	%Human [observations]
Vindija 87	no	A9098	99.48 *[5,570/5,599]*	0.52 *[29/5,599]*	99.85 *[656/657]*	0.15 *[1/657]*
	0.5% hypochlorite	A9099	99.06 *[5,076/5,119]*	0.84 *[43/5,119]*	99.67 *[612/614]*	0.33 *[2/614]*
		R5005	99.06 *[6,850/6,909]*	0.85 *[59/6,909]*	99.52 *[2,095/2,105]*	0.48 *[10/2,105]*
Goyet Q56-1	no	A9104	88.23 *[2,173/2,463]*	11.77 *[290/2,463]*	97.84 *[136/139]*	2.16 *[3/139]*
	0.5% sodium hypochlorite	A9105	93.99 *[2,476/2,633]*	5.96 *[157/2,633]*	99.48 *[191/192]*	0.52 *[1/192]*
		R5006	93.99 *[1,766/1,878]*	5.96 *[112/1,878]*	98.77 *[322/326]*	1.23 *[4/326]*
Mezmaiskaya 2	no	R1916	98.45 *[14,607/14,837]*	1.55 *[230/14,837]*	99.33 *[3,275/3,296]*	0.64 *[21/3,296]*
	0.5% sodium hypochlorite	R1917	99.41 *[15,056/15,142]*	0.57 *[86/15,142]*	99.53 *[3,407/3,412]*	0.44 *[15/3,412]*
Les Cottés Z4-1514	no	L9451	59.03 *[134/227]*	40.97 *[93/227]*	100 *[17/17]*	0 *[0/17]*
	sodium phosphate, 0.5% sodium hypochlorite	R5022	97.74 *[1,211/1,239]*	2.26 *[28/1,239]*	98.64 *[364/369]*	1.36 *[5/369]*
	sodium phosphate, 0.5%sodium hypochlorite	A9309	95.64 *[834/836]*	4.36 *[38/872]*	97.46 *[269/276]*	2.54 *[7/276]*
Spy 94a	no	A9336	16.18 *[140/865]*	83.82 *[725/865]*	69.44 *[25/36]*	30.56 *[11/36]*
	0.5% sodium hypochlorite	R5556	93.72 *[642/685]*	6.28 *[43/685]*	96.79 *[151/156]*	3.21 *[5/156]*

Table S2.6 Sequencing runs with the final set of 23 late Neandertal libraries. The amounts of the spiked-in *ΦX174* per lane ranged from 10% to 30% for the libraries that were sequenced individually (*i.e.* one library per HiSeq lane) and 0.5% for the library pools.

Sequencing Run ID	**Lane(s)**	**ΦX174 [%]**	**Indexed library ID**
SN7001204_0381_AH2Y73BCXX_R_PEdi_A9230	1, 2	20	A9230
SN7001204_0483_BHG73FBCXX_R_PEdi_A9230	1, 2	20	A9230
SN7001204_0421_AH5FMGBCXX_R_PEdi_A9290_A9291	1	20	A9290
	2	20	A9291
SN7001204_0446_BH5HFCBCXX_R_PEdi_A9350	1, 2	20	A9350
SN7001204_0484_AHJJY7BCXX_R_PEdi_A9415_1	1, 2	0.5	R5029, R5046, A9252, A9253
SN7001204_0485_BHG5WTBCXX_R_PEdi_A9415_2	1, 2	0.5	R5029, R5046, A9252, A9253
SN7001204_0486_AHJJYTBCXX_R_PEdi_A9415_3	1, 2	0.5	R5029, R5046, A9252, A9253
SN7001204_0487_BHJKJLBCXX_R_PEdi_A9415_4	1, 2	0.5	R5029, R5046, A9252, A9253
SN7001204_0491_AHJJYLBCXX_R_PEdi_A9426	1, 2	0.5	A9309, A9393, A9394, A9395, A9420
SN7001204_0307_AHA44VADXX_R_PEdi_A9121_A9122	1	20	A9121
	2	20	A9122
SN7001204_0308_BHA450ADXX_R_PEdi_A9122	1	10	A9122
	2	30	A9122
SN7001204_0380_BHKYTKADXX_R_PEdi_A9229	1, 2	20	A9229
SN7001204_0445_AH5HCTBCXX_R_PEdi_A9349_2	1, 2	20	A9349
SN7001204_0339_BHBEUWADXX_R_PEdi_A9180_1	1, 2	20	A9180
SN7001204_0340_AHBEVDADXX_R_PEdi_A9180_2	1, 2	20	A9180
SN7001204_0341_BHBE1WADXX_R_PEdi_A9180_3	1, 2	20	A9180
SN7001204_0480_AHG7F3BCXX_R_PEdi_A9180_1	1, 2	20	A9180
SN7001204_0481_BHG72HBCXX_R_PEdi_A9180_2	1, 2	20	A9180
SN7001204_0418_AH5HHGBCXX_R_PEdi_A9288	1, 2	20	A9288
SN7001204_0422_BH5HGMBCXX_R_PEdi_A9289	1, 2	20	A9289
SN7001204_0379_AHKYGMADXX_R_PEdi_A9228	1, 2	20	A9228
SN7001204_0482_AHG7J5BCXX_R_PEdi_A9228	1, 2	20	A9228
SN7001204_0441_AH5HGGBCXX_R_PEdi_A9348	1, 2	20	A9348
SN7001204_0489_AHJJL3BCXX_R_PEdi_A9427_1	1, 2	0.5	R5556, A9416, A9417, A9418, A9419
SN7001204_0490_BHJJWLBCXX_R_PEdi_A9427_2	1, 2	0.5	R5556, A9416, A9417, A9418, A9419

Supplementary Information 3

Sequencing, data processing and quality

Demultiplexing sequencing runs with the *jivebunny* algorithm

Sequencing libraries are barcoded with a combination of two unique, seven base-pair long index sequences[1], which are selected from a pool of several hundreds of indices. The large number of combinations and the pairwise nucleotide differences ensure that two different libraries rarely end up with the identical index combination. Identifying unrelated index combinations in the sequencing pool would enable us to remove these as contamination from the sequencing data of other indexing libraries. Existing software for demultiplexing allows a fixed number of mismatches per index read and discards all fragments that do not match a known index pair. Such a method would remove contamination from a different indexing library efficiently when applying stringent filtering criteria, *e.g.* a very low number of mismatches, but would also discard valuable, non-contaminant fragments that have a higher number of mismatches due to an increase in the sequencing error rate. While other, more robust methods for assigning the index sequence to a known index exist, they do not have the ability to remove the contamination from other indexing libraries[2]. Therefore, we developed a two-step demultiplexing algorithm called *jivebunny* (https://bioinf.eva.mpg.de/jivebunny).

In the initial step, a subset of reads is analysed to compute the most likely composition of index combinations in the sequencing pool. The estimated relative abundance of index combinations can directly serve as a quality control. This allows to assess the evenness of library pooling before sequencing and to detect contaminant sequencing libraries. In the final step, the estimated relative abundances serve as prior for the maximum posterior probability classification of each read into read groups. We then computed a single quality score from the posterior probabilities and retained only the fragments that were assigned to the correct library based on their index sequences for all of the downstream analyses.

Alignment to the human reference genome and duplicate removal

We used the Burrows-Wheeler Aligner (BWA, version: 0.5.10-evan.9-1-g44db244; https://github.com/mpieva/network-aware-bwa) to align the fragments from all sequencing runs to the modified human reference genome (GRCh37/1000 Genomes release; ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/) using the same parameters as described in Supplementary Information 2. PCR

duplicates were removed by calling a consensus from the fragments with identical alignment start and end coordinates using *bam-rmdup* (version: 0.6.3; https://bitbucket.org/ustenzel/biohazard).

We merged the data from all libraries for each individual after duplicate removal and filtered for fragments that were longer than 35 base pairs using SAMtools (version: 1.3.1)[3]. All of the downstream analyses were further restricted to the fragments that had a mapping quality of at least 25 (MQ ≥ 25) and overlapped highly mappable regions of the human genome (Map35_100% of Prüfer *et al.*[4]). Tables S3.1, S3.2 and Extended Data Table 1 summarize the number of filter-passed fragments mapped to the human genome per individual for each library and for the final merged dataset.

Ancient DNA substitution patterns in the final dataset and fragment size distribution

For each individual we analysed substitution patterns along the fragments for all libraries separately and for the final dataset by counting the number of substitutions relative to the human reference genome (Tables S3.3 and S3.4). As the majority of the libraries were produced without the UDG and endonuclease VIII treatment, the C-to-T substitution frequencies ranged between 17.4% and 56.5% on the 5'-ends, and between 36.6% and 60.8% on the 3'-ends of the alignment (Tables S3.3 and S3.4, Extended Data Fig. 1). The frequencies of C-to-T substitutions remained stable after filtering for sequences with a C-to-T at the opposing end ('conditional' substitutions) (Tables S3.3 and S3.4, Extended Data Fig. 1), indicating that the majority of the data comes from one population of sequences with consistent substitution patterns[5]. In addition to substitutions at the ends, elevated frequencies of C-to-T substitutions of between 1.6% (*Goyet Q56-1*) and 5.9% (*Mezmaiskaya 2*) were also found within fragments (Extended Data Fig. 1). In order to enrich for endogenous ancient DNA for downstream analyses, we used elevated C-to-T substitutions relative to the reference genome at the first and/or the last two positions for the UDG/endonuclease VIII treated libraries and on the first three and/or the last three positions of the alignment for the libraries without UDG/endonuclease VIII treatment (the number of putatively deaminated fragments per library is reported in Table S3.2 and in the complete dataset in the Extended Data Table 1).

The average fragment length was comparable between all fragments and deaminated fragments and ranged from 45 base pairs to 65 base pairs depending on the specimen (Tables S3.1 and S3.2). The fragment size distributions for all mapped fragments and for putatively deaminated fragments are shown in Extended Data Fig. 2. Base compositions for mapped

fragments as a function of fragment length for all five individuals showed that the GC-content was below genome average for all fragment sizes. Additionally, Ts were more common than As in fragments of all sizes.

Observed differences between mapping quality 25 (MQ ≥ 25) and 37 (MQ ≥ 37)

In BWA the mapping quality is a discrete value for representing the quality of the alignment of the fragment to the reference genome[6]. When aligning Neandertal fragments to the modern human reference, a higher number of mismatches than in alignments of modern human DNA fragments is expected due to the divergence between the two hominin groups. An additional increase in the number of mismatches is caused by damage-derived substitutions characteristic for ancient DNA[7]. As the majority of our sequencing data originates from the libraries without UDG and endonuclease VIII treatment, C-to-T substitutions were at higher frequency not only at the alignment start and end, but throughout fragments as well. Requiring a mapping quality of 37 restricts the number of allowed mismatches to the reference genome and could therefore preferentially exclude fragments that show a high amount of deamination-derived substitutions, *i.e.* true endogenous fragments. However, when requiring mapping quality of 25, we increased the number of allowed mismatches to the reference genome and therefore included more fragments with deamination patterns (Table S3.65). We observed that by requiring mapping quality of 25 we retained between 0.46% and 3.26% more of putatively deaminated fragments when compared to the mapping quality of 37 (Table S3.5). Therefore, all of the subsequent analyses were restricted to the fragments with mapping quality of at least 25.

Obtained coverage of the nuclear genomes and sex determination

In order to determine the obtained coverage of the nuclear genomes we counted the number of bases with a base quality of at least 30 (BQ ≥ 30) in the fragments that overlapped highly mappable regions of the autosomes of the human genome (Map35_100% of Prüfer *et al.*[4]) and divided it by the total length of those regions. For *Les Cottés Z4-1514* we obtained 2.7-fold genome-wide coverage, 2.2-fold for *Goyet Q56-1*, 1.7-fold for *Mezmaiskaya 2*, 1.3-fold for *Vindija 87* and 1-fold for *Spy 94a*. After restricting the analyses to putatively deaminated fragments, the nuclear coverage was 0.2-fold for *Spy 94a*, 0.3-fold for *Goyet Q56-1*, 0.5-fold for *Vindija 87*, 0.6-fold for *Mezmaiskaya 2* and 1-fold for *Les Cottés Z4-1514*.

We determined the sex of the five Neandertal individuals by counting the number of fragments that aligned to the X chromosome and the autosomes. Based on the expected ratios

of X to (X + autosomal) fragments for male and female individuals we concluded that *Les Cottés Z4-1514*, *Goyet Q56-1* and *Vindija 87* were females, whereas *Mezmaiskaya 2* and *Spy 94a* were males (Extended Data Fig. 3). These results were concordant after restricting the analyses to deaminated fragments (Extended Data Fig. 3).

References for SI3:

1 Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**, e3, doi:10.1093/nar/gkr771 (2012).

2 Renaud, G., Stenzel, U., Maricic, T., Wiebe, V. & Kelso, J. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**, 770-772, doi:10.1093/bioinformatics/btu719 (2015).

3 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

4 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

5 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

6 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

7 Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* **104**, 14616-14621, doi:10.1073/pnas.0704665104 (2007).

Table S3.1 Final set of libraries for each of the five late Neandertal specimens and the amount of data generated. The majority of the libraries were not treated with UDG and endonuclease VIII in order to enrich the libraries for fragments with apparent C-to-T substitutions at their ends, *i.e.* endogenous fragments. Two UDG/endonuclease VIII treated libraries are denoted with * next to the library ID.

Specimen	Library ID	Number of sequenced fragments	Number of fragments ≥ 35 bp	Number of mapped fragments ≥ 35bp, MQ≥25	Number of mapped fragments ≥ 35bp, MQ≥25, Map35_100%	Number of unique fragments ≥ 35bp, MQ≥25, Map35_100%	% mapped fragments ≥ 35bp, MQ≥25	Average duplication rate	Average fragment length ≥ 35bp, MQ≥25	Obtained nuclear coverage
Les Cottés Z4-1514	A9230	711,761,175	413,434,047	99,375,141	90,463,605	35,424,399	**24.04**	2.55	49.5	0.76
	A9290	105,154,977	60,774,029	14,175,490	12,894,194	10,002,308	**23.33**	1.29	50.6	0.22
	A9291	103,634,816	46,884,650	10,323,454	9,370,775	5,439,170	**22.02**	1.72	52.4	0.12
	A9350	510,282,968	279,319,029	62,030,860	55,912,045	24,500,132	**22.21**	2.28	49.2	0.53
	A9309	62,105,233	27,695,289	13,556,977	12,260,934	7,999,080	**48.95**	1.53	53.9	0.19
	A9393	68,225,206	35,672,322	18,725,744	16,941,074	9,560,711	**52.49**	1.77	54.0	0.22
	A9394	64,252,134	33,551,703	17,787,655	16,092,420	9,051,852	**53.02**	1.78	54.2	0.21
	A9395	72,369,539	39,881,973	20,935,433	18,926,260	9,619,169	**52.49**	1.97	54.1	0.23
	A9420	78,322,081	41,768,370	22,397,498	20,273,165	9,739,677	**53.62**	2.08	54.0	0.23
Goyet Q56-1	A9122*	442,851,453	264,640,703	62,929,107	56,155,799	36,959,387	**23.78**	1.52	59.3	0.95
	A9229	260,745,804	166,984,330	40,330,160	35,917,589	23,170,496	**24.15**	1.55	66.3	0.66
	A9349	214,201,738	143,972,193	37,478,583	32,509,057	20,385,136	**26.03**	1.59	65.6	0.57
Mezmaiskaya 2	A9180	1,457,126,645	711,656,639	70,998,067	64,216,210	39,979,655	**9.98**	1.61	50.7	0.92
	A9288	559,876,880	328,784,774	34,986,505	31,621,445	17,817,710	**10.64**	1.77	52.5	0.42
	A9289	554,124,486	315,860,721	33,173,312	29,978,177	16,609,709	**10.50**	1.80	52.7	0.40
Vindija 87	A9121*	145,582,747	68,547,308	3,115,759	2,831,116	2,686,238	**4.55**	1.05	44.6	0.05
	A9228	596,242,509	317,977,873	65,742,483	59,650,713	36,813,537	**20.68**	1.62	47.3	0.78
	A9348	611,721,252	291,480,375	53,641,469	47,903,681	20,535,201	**18.40**	2.33	47.2	0.44
Spy 94a	R5556	192,624,453	85,384,785	14,511,814	21,995,017	12,195,918	**28.41**	1.80	48.0	0.26
	A9416	141,451,601	50,192,163	12,928,722	13,190,957	9,019,085	**28.91**	1.46	47.8	0.19
	A9417	125,028,470	44,955,189	14,258,930	11,738,269	8,531,565	**28.76**	1.38	47.9	0.18
	A9418	134,250,911	48,516,613	14,392,814	12,962,870	9,072,138	**29.39**	1.43	47.8	0.19
	A9419	139,856,628	49,546,872	24,259,504	13,075,750	8,017,847	**29.05**	1.63	47.9	0.17

Table S3.2 Number of fragments with C-to-T substitutions relative to the human reference genome per library for each of the five Neandertal specimens. We used elevated C-to-T substitutions relative to the reference genome at the first and/or the last two positions for the UDG/endonuclease VIII treated libraries (denoted with * next to the library ID) and on the first three and/or the last three positions of the alignment for the libraries without this treatment to select for putatively deaminated fragments.

Specimen	Library ID	Number of mapped deaminated fragments ≥ 35bp, MQ≥25	Number of mapped deaminated fragments ≥ 35bp, MQ≥25, Map35_100%	Average length of deaminated fragments	Obtained nuclear coverage ≥ 35bp, MQ≥25, Map35_100% with deaminated fragments only
Les Cottés Z4-1514	A9230	13,029,067	11,884,134	49.7	0.260
	A9290	3,658,005	3,336,083	50.8	0.075
	A9291	1,951,457	1,773,783	52.6	0.041
	A9350	9,009,417	8,224,061	49.4	0.179
	A9309	3,203,596	2,902,525	54.1	0.069
	A9393	3,806,905	3,448,782	54.2	0.082
	A9394	3,606,714	3,266,905	54.4	0.078
	A9395	3,839,802	3,476,083	54.3	0.083
	A9420	3,862,001	3,497,832	54.2	0.083
Goyet Q56-1	A9122*	2,646,598	2,363,948	58.2	0.060
	A9229	4,441,349	3,958,310	65.2	0.110
	A9349	3,976,994	3,537,449	64.7	0.099
Mezmaiskaya 2	A9180	13,869,276	12,581,544	50.9	0.290
	A9288	6,278,972	5,681,222	52.8	0.140
	A9289	8,816,559	5,260,895	53.1	0.130
Vindija 87	A9121*	481,258	442,293	44.5	0.009
	A9228	15,867,248	14,453,198	47.3	0.310
	A9348	8,720,373	7,935,151	47.2	0.170
Spy 94a	R5556	2,926,437	2,665,006	48.3	0.042
	A9416	2,095,735	1,913,112	48.0	0.039
	A9417	1,978,423	1,804,604	48.2	0.042
	A9418	2,106,835	1,923,209	48.0	0.037
	A9419	1,862,139	1,698,455	48.1	0.058

Table S3.3 Frequencies of C-to-T substitutions at terminal positions of the sequence alignments per library for each of the five late Neandertals. Percentages of fragments carrying terminal C-to-T substitutions relative to the human reference genome are shown for all fragments and for fragments that have a C-to-T substitution at the opposing end of a fragment ('conditional' substitutions). Utilized are mapped fragments longer than 35 base pairs with mapping quality of at least 25, reported in the Tables S3.1 and S3.2, and calculated 95% binomial confidence intervals are provided in brackets.

Specimen	Library ID	All fragments		Fragments with a C-to-T at the opposing end	
		5'-end C→T substitutions (%) [95% CI]	3'-end C→T substitutions (%) [95% CI]	5'-end C→T substitutions (%) [95% CI]	3'-end C→T substitutions (%) [95% CI]
Les Cottés Z4-1514	A9230	53.1 [53.0-53.1]	56 [55.9-56.0]	50.5 [50.4-50.6]	53.6 [53.4-53.7]
	A9290	51.5 [51.4-51.5]	61 [61.0-61.1]	48.9 [48.7-49.1]	58.5 [58.3-58.7]
	A9291	50.8 [50.7-50.9]	62.2 [62.1-62.4]	48.4 [48.1-48.7]	59.8 [59.5-60.2]
	A9350	51 [51.0-51.1]	54 [53.9-54.0]	48.3 [48.2-48.4]	51.3 [51.2-51.5]
	A9309	57.5 [57.4-57.5]	54.8 [54.8-54.9]	54.5 [54.3-54.7]	52.1 [51.9-52.3]
	A9393	57.3 [57.2-57.4]	54.4 [54.4-54.5]	54.5 [54.5-54.7]	52.1 [51.9-52.3]
	A9394	57.6 [57.5-57.6]	55.2 [55.1-55.3]	54.7 [54.5-54.9]	52.6 [52.4-52.9]
	A9395	57.6 [57.5-57.7]	55 [55.0-55.1]	54.8 [54.6-55.0]	52.5 [52.3-52.6]
	A9420	58.1 [58.0-58.1]	53.8 [53.8-53.9]	55.1 [54.9-55.3]	51.3 [51.1-51.5]
Goyet Q56-1	A9122*	8 [8.0-8.0]	28.5 [28.4-28.5]	8.6 [8.5-8.7]	30.6 [30.3-30.9]
	A9229	24.6 [24.6-24.6]	48.8 [48.7-48.9]	24.8 [24.7-25.0]	50.4 [50.1-50.7]
	A9349	24.6 [24.6-24.6]	44.3 [44.2-44.4]	25 [24.8-25.2]	45.9 [45.6-46.2]
Mezmaiskaya 2	A9180	43.7 [43.7-43.8]	41.5 [41.5-41.5]	40.5 [40.3-40.6]	38.4 [38.2-38.5]
	A9288	45.1 [45.0-45.1]	51.7 [51.6-51.7]	41.9 [41.7-42.0]	48.7 [48.5-48.9]
	A9289	44.6 [44.5-44.6]	52.7 [52.7-52.8]	41.4 [41.2-41.6]	49.7 [49.5-49.9]
Vindija 87	A9121*	26.6 [26.4-26.7]	45.8 [45.7-45.9]	26.3 [25.9-26.7]	45.4 [44.8-46.0]
	A9228	57.7 [57.7-57.7]	64.6 [64.5-64.6]	54.4 [54.3-54.5]	61.7 [61.6-61.8]
	A9248	57.3 [57.3-57.4]	56.8 [56.7-56.8]	54.1 [54.0-54.2]	53.9 [53.7-54.0]
Spy 94a	R5556	30.5 [30.4-30.6]	41.6 [41.5-41.7]	28.6 [28.4-28.8]	39.8 [39.5-40.1]
	A9416	30 [30.0-30.1]	39.6 [39.5-39.6]	28.2 [28.0-28.5]	37.8 [37.5-38.1]
	A9417	29.7 [29.7-29.8]	38.8 [38.7-38.9]	27.9 [27.6-28.1]	36.9 [36.6-37.2]
	A9418	29.8 [29.7-29.9]	39.8 [39.8-39.9]	27.9 [27.7-28.2]	38.2 [37.9-38.5]
	A9419	30.1 [30.0-30.2]	39.5 [39.4-39.6]	28.2 [27.9-28.4]	37.6 [37.3-37.9]

doi:10.1038/nature26151

Table S3.4 Frequencies of C-to-T substitutions at terminal positions of the sequence alignments in the final dataset for each of the five late Neandertals. Percentages of fragments carrying terminal C-to-T substitutions relative to the human reference genome are shown for all fragments and for fragments that have a C-to-T substitution at the opposing end of a fragment ('conditional' substitutions). Utilized are mapped fragments longer than 35 base pairs with mapping quality of at least 25, reported in the Tables S3.1 and S3.2, and calculated 95% binomial confidence intervals are provided in brackets.

Specimen	All fragments		Fragments with a C-to-T at the opposing end	
	5'-end C→T substitutions (%) [95% CI]	3'-end C→T substitutions (%) [95% CI]	5'-end C→T substitutions (%) [95% CI]	3'-end C→T substitutions (%) [95% CI]
Les Cottés Z4-1514	54.1 [54.1-54.2]	55.6 [55.6-55.6]	51.5 [51.4-51.6]	53.0 [53.0-53.1]
Goyet Q56-1	17.4 [17.4-17.4]	36.6 [36.6-36.7]	18.5 [18.4-18.6]	43.5 [43.3-43.6]
Mezmaiskaya 2	44.2 [44.2-44.3]	45.8 [45.8-45.9]	41.0 [40.9-41.1]	42.7 [42.6-42.8]
Vindija 87	56.5 [56.4-56.5]	60.8 [60.7-60.8]	53.4 [53.4-53.5]	58.4 [58.4-58.4]
Spy 94a	30.1 [30.0-30.1]	40.0 [39.9-40.0]	28.2 [28.1-28.3]	38.2 [38.1-38.3]

Table S3.5 Observed differences in the number of retained fragments between mapping quality of 25 (MQ ≥ 25) and mapping quality of 37 (MQ ≥ 37). Of the fragments that were filtered out by requiring mapping quality of 37, but retained with the mapping quality of 25, between 0.46% and 3.26% of fragments that are retained have C-to-T substitutions relative to the reference genome, *i.e.* are more likely to originate from true endogenous DNA.

Specimen	All fragments			Fragments with terminal C-to-T substitutions			
	Number of unique fragments ≥ 35bp, MQ≥25, Map35_100%	**Number of unique fragments ≥ 35bp, MQ≥37, Map35_100%**	**Difference all fragments**	**Number of unique fragments ≥ 35bp, MQ≥25, Map35_100%**	**Number of unique fragments ≥ 35bp, MQ≥37, Map35_100%**	**Difference deaminated fragments**	**[%] difference deaminated fragments/ difference all fragments** (MQ25 deam/MQ25 all) – (MQ37 deam/MQ37 all)
Les Cottés Z4-1514	121,325,944	114,178,142	7,147,802	41,804,332	36,668,878	5,135,454	**2.34**
Goyet Q56-1	80,513,596	79,480,443	1,033,153	9,859,280	9,364,696	494,584	**0.46**
Mezmaiskaya 2	74,403,383	68,211,593	6,191,790	23,521,855	19,681,787	3,840,068	**2.76**
Vindija 87	60,030,516	55,074,214	4,956,302	22,828,056	19,148,039	3,680,017	**3.26**
Spy 94a	46,834,696	45,177,626	1,657,070	10,003,560	9,004,149	999,411	**1.43**

Supplementary Information 4

Contamination estimates

We used four complementary methods to estimate the proportion of present-day human DNA contamination in the shotgun data of five late Neandertals. The proportion of present-day human DNA contamination in each dataset, with 95% binomial confidence intervals, are summarized in Table S4.1.

Mitochondrial contamination estimates

We estimated the proportion of mitochondrial DNA contamination by present-day human DNA in our datasets by using two different sets of positions. We first re-aligned the shotgun data of all HiSeq runs to the revised Cambridge Reference Sequence (rCRS, NC_012920) using BWA[1] and as described in Supplementary Information 2. After removing PCR duplicates with *bam-rmdup* (version: 0.6.3; https://bitbucket.org/ustenzel/biohazard), only mapped fragments with a length greater than 35 base pairs and a mapping quality of at least 25 ($L \geq 35$bp, $MQ \geq 25$) were retained for the analyses. We then counted the number of fragments that overlapped 63 positions where 18 published Neandertal mitochondrial genomes[2-8] differ from those of all 311 present-day humans sampled from world-wide populations[2].

In the second approach, we determined the positions in the reconstructed mtDNA genomes of each of the five Neandertals (see Supplementary Information 5 on details of the mitochondrial sequence reconstruction) that are specific ("diagnostic") for that Neandertal when compared to 311 present-day human mtDNA genomes. For *Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87* and *Spy 94a* we determined 80, 83, 79, 81 and 83 such positions, respectively. We then counted how many of the mtDNA fragments overlapping these positions support the modern human state and how many support the Neandertal state.

Again, to mitigate the effect of deamination (as described in Supplementary Information 2), for both approaches we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G[9]. Based on contamination by human mitochondrial fragments we estimated a contamination rate of between 0.50% (*Mezmaiskaya 2*) and 5.06% (*Goyet Q56-1*) using all fragments longer than 35 base pairs with a mapping quality of at least 25, and between 0.39% (*Mezmaiskaya 2*) and 1.3% (*Les Cottés Z4-1514*) among putatively deaminated fragments (Table S4.1).

Autosomal contamination estimates

We estimated the extent of present-day human DNA contamination on the autosomes in the five late Neandertals using the maximum likelihood based approach described in Green *et al.*[4]. This method co-estimates contamination, sequencing error and two population parameters. It is based on the idea that the contaminant contributes derived alleles as found in present-day humans to the dataset at positions where the ancient individual carries ancestral alleles. We estimated the proportion of autosomal contamination by using the fragments equal or longer than 35 base pairs with mapping quality of at least 25 that covered informative positions in the nuclear genome where humans carry a fixed derived variant when compared to the great apes. Only highly mappable regions of the genome (Map35_100% of Prüfer *et al.*[5]) were considered and bases with a quality of at least 30. Autosomal contamination estimates among all fragments were 0.18% (95% CI: 0-0.41%) for *Les Cottés Z4-1514*, 0.89% (95% CI: 0.69-1.11%) for *Goyet Q56-1*, 0.83% (95% CI: 0.1-52%) for *Mezmaiskaya 2*, 1.15% (95% CI: 0-2.37%) for *Vindija 87* and 1.75% (95% CI: 0.58-2.84%) for *Spy 92a* (Table S4.1).

Estimating autosomal contamination using an ancestry model

Assuming that any autosomal contamination is from non-African individuals, we can estimate the contamination proportion in different Neandertals using an ancestry model where they each trace a portion of their genome either from a high-coverage uncontaminated Neandertal, or from a population related to present-day non-Africans. We used *qpAdm*[10] that is a generalization of f_4-ratio estimates, to build a model that leveraged the observation that Eastern African Dinka individuals are more closely related to the source population of non-Africans than most Western African populations and central African rainforest hunter-gatherers are. We built a two-source *qpAdm* model where one part of the ancestry was modelled as being most closely related to the high-coverage *Vindija 33.19* Neandertal genome (which has negligible contamination[11]), and the other source of ancestry was modelled as being most closely related to the Dinka population. To estimate proportions of the genome contributed from each of these two sources we also used a set of outgroups modelled as being less closely related to the genome of each target low-coverage Neandertal populations (than either the *Vindija 33.19* high-coverage genome[11] or the putative non-African contamination): the *Denisovan* high-coverage genome[12], the *Altai* Neandertal high-coverage genome[5], Western African Mende and Yoruba, and central African Biaka and Mbuti. Neither of these African populations show any evidence of substantial recent gene flow from Western Eurasia. The modern human genomes were taken from the Simons Genome Diversity Project[13], and

we restricted to bi-allelic transversion SNPs between the 300 individuals in the SGDP and the *Altai* Neandertal and *Denisovan* genome for all estimates.

We found that the proposed model provided an adequate fit for all tested low-coverage Neandertal genomes (P>0.08) except for *Vindija 87* (P=0.0003). One possibility is that this is due to the *Vindija 87* sample originating from the same individual as the high-coverage *Vindija 33.19* genome used as a source in the analysis. We estimate detectable contamination both in *Spy 92a* (4.1%; 95% CI: 3.9-4.2%) and *Mezmaiskaya 1* (3.1%; 95% CI: 2.9-3.2%). We estimate 1.8% contamination in *Vindija 87* (we obtained ~0% jackknife standard error for this test, due to the source genome and the test sample being from the same individual) but and upper 95% CI limit of maximum 1.1% contamination in the remaining *Les Cottés Z4-1514*, *Mezmaiskaya 2*, and *Goyet Q56-1* genomes. When using the damage-restricted data, we find that the evidence of substantial contamination in two individuals is reduced to the same putatively negligible levels as the other individuals, with *Spy 92a* now being estimated as having 1.0% contamination (95% CI: 0.8-1.2%), and *Mezmaiskaya 1* as having 1.2% contamination (95% CI: 1.0-1.4%).

Y chromosome contamination estimates for female individuals

Based on the number of fragments aligning to the X chromosome and the autosomes, we concluded that *Les Cottés Z4-1514, Goyet Q56-1* and *Vindija 87* were females (see Supplementary Information 3 for more information). This allows us to estimate the proportion of male contamination for these three individuals. We counted the number of fragments aligning to the unique regions of the Y chromosome and divided it by the number of fragments that would be expected if the individual was a male. The latter is calculated based on the number of fragments aligned to the whole genome multiplied by the number of the highly mappable positions on the Y chromosome, and then divided by the total length of the genome in highly mappable regions (Map35_100% of Prüfer *et al.*[5]). We expected 183,046 Y fragments for *Les Cottés Z4-1514* if the individual was a male, 123,599 Y fragments for *Goyet Q56-1* and 90,536 Y fragments for *Vindija 87* among all fragments longer than 35 base pairs with a mapping quality of at least 25. Among fragments with C-to-T substitutions at the first three and/or last three positions we expected 57,167 Y fragments for *Les Cottés Z4-1514*, 13,483 Y fragments for *Goyet Q56-1* and 31,217 Y fragments for *Vindija 87*.

The observed number of Y chromosomal fragments among all fragments was 2,881 for *Les Cottés Z4-1514*, 1,593 for *Goyet Q56-1* and 1,561 for *Vindija 87*, giving the estimates of male contamination of 1.57% (95% CI: 1.52-1.63%), 1.29% (95% CI: 1.23-1.35%) and

2.60% (95% CI: 2.51-2.70%), respectively. The observed number of Y chromosomal fragments among putatively deaminated fragments was 746 for for *Les Cottés Z4-1514*, 167 for *Goyet Q56-1* and 382 for *Vindija 87*, resulting in the estimates of male contamination of 1.30% (95% CI: 1.22-1.40%), 1.24% (95% CI: 1.07-1.44%) and 1.22% (95% CI: 1.1-1.35%), respectively (Table S4.1).

References for SI4:

1 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

2 Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426, doi:10.1016/j.cell.2008.06.021 (2008).

3 Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321, doi:10.1126/science.1174462 (2009).

4 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

5 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

6 Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A* **111**, 2229-2234, doi:10.1073/pnas.1318934111 (2014).

7 Brown, S. *et al.* Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci Rep* **6**, 23559, doi:10.1038/srep23559 (2016).

8 Rougier, H. *et al.* Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci Rep* **6**, 29005, doi:10.1038/srep29005 (2016).

9 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

10 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207-211, doi:10.1038/nature14317 (2015).

11 Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, doi:10.1126/science.aao1887 (2017).

12 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

13 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206, doi:10.1038/nature18964 (2016).

Table S4.1 Proportion of present-day human DNA contamination among all and putatively deaminated fragments of the final datasets of *Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87* and *Spy 94a.* Utilized are mapped fragments longer than 35 base pairs with mapping quality of at least 25, reported in the Tables S3.1, S3.2 and S5.1. 95% binomial confidence intervals (CI) are reported in brackets for each of the estimates.

	Nuclear DNA contamination estimates				Mitochondrial DNA contamination estimates			
	Autosomal contamination estimates		Sex-based contamination estimates for female individuals		311 present-day humans ≠ 18 Neandertals		311 present-day humans ≠ diagnostic positions of the reconstructed mitochondrial genome	
Individual	% of all fragments matching derived states in present-day humans [95% CI]	% of deaminated fragments matching derived states in present-day humans [95% CI]	% Y chromosome contamination in all fragments [95% CI]	% Y chromosome contamination in deaminated fragments [95% CI]	% of all fragments matching present-day human state [95% CI]	% of deaminated fragments matching present-day human state [95% CI]	% of all fragments matching present-day human state [95% CI]	% of deaminated fragments matching present-day human state [95% CI]
Les Cottés Z4-1514	0.18 *[0.00-0.42]*	0.00 *[0.00-4.86]*	1.57 *[1.52-1.63]*	1.30 *[1.22-1.40]*	2.07 *[1.76-2.44]*	1.30 *[0.89-1.9]*	1.36 *[1.14-1.63]*	0.98 *[0.66-1.44]*
Goyet Q56-1	0.89 *[0.69-1.11]*	0.00 *[0.00-12.21]*	1.29 *[1.23-1.35]*	1.24 *[1.07-1.44]*	5.06 *[4.45-5.76]*	1.05 *[0.48-2.28]*	4.97 *[4.41-5.55]*	1.08 *[0.55-2.12]*
Mezmaiskaya 2	0.83 *[0.00-1.52]*	0.81 *[0.00-9.21]*	na	na	0.52 *[0.44-0.62]*	0.42 *[0.28-0.63]*	0.50 *[0.42-0.59]*	0.39 *[0.26-0.57]*
Vindija 87	1.15 *[0.00-2.37]*	0.41 *[0.00-9.80]*	2.60 *[2.51-2.70]*	1.22 *[1.11-1.35]*	0.53 *[0.43-0.64]*	0.44 *[0.30-0.65]*	0.46 *[0.39-0.56]*	0.39 *[0.27-0.57]*
Spy 94a	1.75 *[0.58-2.84]*	0.51 *[0.00-17.51]*	na	na	4.55 *[3.70-5.60]*	2.35 *[1.24-4.40]*	4.41 *[3.66-5.30]*	1.61 *[0.82-3.14]*

Supplementary Information 5

Uniparental markers – mitochondrial DNA genomes and Y chromosomes of Neandertals

Reconstruction of the mitochrondrial DNA (mtDNA) genomes of late Neandertals

In order to reconstruct full mitochondrial genomes of five late Neandertals, we re-aligned the shotgun data of all HiSeq runs to the *Vindija 33.16* mitochondrial genome (AM948965)[1] using BWA[2]. Because BWA does not successfully align fragments at the beginning and the end of a circular genome, we added 250 base pairs from the beginning of the *Vindija 33.16* mitochondrial genome to its end in order to get equal coverage of the fragments across the mtDNA[3]. PCR duplicates were removed using *bam-rmdup* (version: 0.6.3; https://bitbucket.org/ustenzel/biohazard) and only mapped fragments with a length greater than 35 base pairs and a mapping quality of at least 25 were retained for the analyses (L ≥ 35bp, MQ ≥ 25).

Between 29,146 and 304,914 unique mitochondrial fragments were obtained for the five late Neandertals, resulting in per individual average coverage of the mtDNA ranging from 82-fold to 988-fold (Table S3.1). We reconstructed full mitochondrial genomes of these five late Neandertals by calling a consensus base at each position along the genome that was covered by at least three fragments and where at least two-thirds of the fragments had an identical base[4]. We converted Ts on the forward strands and As on the reverse strands in the first three and the last three positions of a fragment into Ns to prevent deamination-derived substitutions influencing calling of a consensus base. For each individual, the mtDNA genome reconstructed from all fragments was identical to the one that was reconstructed from deaminated fragments only. We were unable to resolve the position 310 in the mtDNA genome of *Mezmaiskaya 2* despite its high coverage (318-fold). This position is in the C-homopolymer stretch of the mtDNA which is known to be problematic for the alignment. The mitochondrial genome of *Goyet Q56-1* was reconstructed as a part of a previous study[5] and was identical to the mtDNA that we reconstructed here from the shotgun data of this individual.

Pairwise differences between Neandertal mtDNAs and Maximum Parsimony tree

We aligned the reconstructed mitochondrial genomes of the five late Neandertals to the mtDNA genomes of 18 Neandertals[1,3,5-10], 311 present-day humans[11], 10 ancient modern

humans[12-17], three Denisovans[18-20], a hominin from Sima de los Huesos[4] and a chimpanzee[21] using MAFFT[22]. The number of pairwise differences among mitochondrial genomes was calculated using MEGA6[23] and we performed maximum parsimony analysis using the *Parsimony ratchet* as implemented in the R package *phangorn*[24] in order to visualize the relationship of the Neandertal mtDNA sequences as a phylogenetic tree (Fig. S5.1). The reconstructed mitochondrial genome of *Vindija 87* was identical to the mitochondrial genomes of *Vindija 33.16*[1] and *Vindija 33.19*[7] (Table S5.2). Interestingly, the mitochondrial genome of *Spy 94a* differed by only one substitution from the mtDNA genomes of *Goyet Q56-1, Goyet Q305-7* and *Goyet Q374a-1*[5]. The mitochondrial DNA genome of the *Mezmaiskaya 2* individual fell within the variation of western Neandertals with the least number of pairwise differences to *Feldhofer 2*, whereas the mitochondrial genome of *Les Cottés Z4-1514* had the least number of differences to the *DC1227* Neandertal from the Denisova cave[9] and the Neandertal from the Okladnikov cave[8]. Thus, the availability of these new mitochondrial genomes challenges the previously proposed division between the eastern and western Neandertal mtDNA genomes in the late surviving Neandertals[25].

The most recent common ancestor of Neandertal mtDNAs

For determining the time to the most recent common ancestor (TMRCA) of all Neandertal mitochondrial genomes, we used a Bayesian phylogenetic analysis that takes advantage of the known radiocarbon dates of a portion of the Neandertal individuals and ancient modern humans as calibration points for the molecular clock (Table S5.3). First, we subset the alignment that we used to calculate pairwise differences to 54 present-day humans in order to reduce the computational load. The poly-C stretches at positions 303-315 and 16,182-16,193 were excluded from the analysis following Duggan *et al.*[26] and the analysis was subset to the non-D-loop region (positions 577-16,023; following Fu *et al.*[15]) in order to reduce the bias of misalignments. We determined the best-fitting substitution model using *jModelTest2*[27] to be TrN[28] with invariable sites (+I) and rate variation among sites (+G). A Bayesian phylogenetic analysis was conducted using BEAST v2.4.5[29]: we set the substitution model to the one determined by *jModelTest2* and used a marginal likelihood estimation (MLE) analysis[30,31] in order to choose the best fitting clock model (strict clock or uncorrelated relaxed lognormal clock) and tree model (constant population size or Bayesian skyline). For all mitochondrial genomes with known radiocarbon dates, we set their tip date to the radiocarbon date point estimate and used their 95% confidence interval as the boundaries of a uniform prior (Table S5.3). We ran each model combination for 30 million steps and additional 100 x 300,000

steps for path sampling during MLE analysis. Following the scale of Kass and Raftery[32], the combination of a strict clock model and Bayesian skyline as a tree model was supported decisively by the sequencing data over the other model combinations ($\log_{10}$ BF > 4.5). For this model, we ran four independent analyses for each 75 million steps and subsequently combined them for the final maximum clade credibility annotation of the tree.

When following the analysis of Posth *et al.*[10], which used the coding region of mitochondrial genomes, and set the fixed mutation rate to $1.56*10^{-8}$ substitutions/bp/year as it was determined using ten ancient modern humans[15], we obtained a slightly older age for the TMRCA of all Neandertal mitochondrial sequences of 273,452 years ago (95% HPD: 229,373-317,694 years ago). However, when following the approach of Fu *et al.*[15], and used both the ten ancient modern humans and the eleven Neandertals with known radiocarbon dates for estimating the mutation rate, we estimated a faster mutation rate of $2.62*10^{-8}$ substitutions/bp/year (95% HPD: $2.36*10^{-8}$ – $2.89*10^{-8}$ substitutions/bp/year) for the complete molecule, and $2.10*10^{-8}$ (95% HPD: $1.86*10^{-8}$ – $2.35*10^{-8}$ substitutions/bp/year) for the coding region, and $1.49*10^{-7}$ substitutions/bp/year (95% HPD: $1.16*10^{-7}$ – $1.83*10^{-7}$ substitutions/bp/year) for the control region. Using the faster mutation rate (estimated in the Table S5.4), we subsequently determined a younger date for the TMRCA of all Neandertal mtDNA coding region sequences of 215,093 years ago (95% HPD: 176,240 – 254,290 years ago), as well as then younger molecular ages for the non-radiocarbon dated Neandertals (Table S5.4). As we were not able to confidentially estimate the Neandertal mutation rate just using Neandertal mitochondrial sequences with known radiocarbon dates (and as previously reported in Posth *et al.*[10]), this estimated mutation rate of both lineages has to be considered with caution as it reflects an average of the mutation rates on the modern human and Neandertal mtDNA lineages. The dates obtained using the mutation rate estimated from human mitochondrial genomes[15] should therefore be considered as an upper boundary for Neandertals.

Y chromosomes of Neandertals

For reconstructing the phylogeny of the two male late Neandertal individuals, *Mezmaiskaya 2* and *Spy 94a,* we applied the same filtering steps for the sequencing data as described in Supplementary Information 3 and Supplementary Information 6, and only used fragments that showed patterns of deamination at the alignment ends (putatively deaminated fragments). For the comparison with modern humans, we added the sequencing data of 175 male individuals of the Simons Genome Diversity Panel (SGDP[33]) and the sequencing data of two individuals

with haplogroup A00[34], to-date the most divergent human Y chromosome lineage[35]. We merged the closely related A00 individuals to gain coverage comparable to the SGDP sequencing data (19 – 61-fold average coverage per sample) and processed all modern samples following Barbieri *et al.*[36]: after duplicate removal, indels were re-aligned and base quality values were re-calibrated using GATK[37]. GATK's *UnifiedGenotyper* was used to call genotypes assuming haploidy. For the detection of variants in the Neandertal individuals, we closely followed the processing of the first published Neandertal Y chromosome analysis[35] and used the described parameters to call bases and inferred genotypes using samtools v1.4 and bcftools v1.4 in the regions of the Y-chromosome that were defined as callable by Poznik *et al.*[38]. We masked all genotypes which had an unusually high coverage or a high number of observed bases that were inconsistent with the consensus allele as missing. The alleles of the chimp Y chromosome that aligned to the human Y chromosome were added together with the only published Y chromosome sequences of a Neandertal from *El Sidrón*[35]. From this merged sequencing data set, we only retained variations that were transversions in order to avoid a bias due to the higher number of C-to-T substitutions throughout the non-USER treated sequencing libraries of *Mezmaiskaya 2* and *Spy 94a*. We further restricted these 13,832 SNPs to sites for which at least one of the three Neandertal individuals was genotyped or more than 50% of the present-day human individuals were typed while the QUAL value of the SNP was >= 30. This further filtering step left 13,771 SNPs for analysis.

Despite these stringent filtering the Neandertal samples only covered a fraction of the 13,771 SNPs as a result of their low coverage and the restriction to deaminated fragments. For example, while *Mezmaiskaya 2* had sequencing information for 2,251 SNPs and *Spy 94a* for 1,230 SNPs, the exome capture sequencing data of *El Sidrón* only overlapped 40 SNPs. Due to this low number of informative alleles a meaningful comparison of *El Sidrón* to the other two Neandertals or the present-day human individuals was not possible. We therefore excluded *El Sidrón* from the subsequent analysis.

On the remaining data set, we constructed a neighbour-joining tree based on pairwise differences and allowing for pairwise deletion using the R package *ape*[39] in order to visualize the Y chromosome phylogeny using the chimp sequence as an outgroup (Fig. 2B). Both *Mezmaiskaya 2* and *Spy 94a* formed a monophyletic clade and fell outside the variation of Y chromosomes of present-day humans, including A00.

References for SI5:

1 Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426, doi:10.1016/j.cell.2008.06.021 (2008).

2 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).

3 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

4 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

5 Rougier, H. *et al.* Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci Rep* **6**, 29005, doi:10.1038/srep29005 (2016).

6 Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321, doi:10.1126/science.1174462 (2009).

7 Gansauge, M. T. & Meyer, M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res* **24**, 1543-1549, doi:10.1101/gr.174201.114 (2014).

8 Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A* **111**, 2229-2234, doi:10.1073/pnas.1318934111 (2014).

9 Brown, S. *et al.* Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci Rep* **6**, 23559, doi:10.1038/srep23559 (2016).

10 Posth, C. *et al.* Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat Commun* **8**, 16046, doi:10.1038/ncomms16046 (2017).

11 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

12 Ermini, L. *et al.* Complete Mitochondrial Genome Sequence of the Tyrolean Iceman. *Current Biology* **18**, 1687-1693, doi:https://doi.org/10.1016/j.cub.2008.09.028 (2008).

13 Gilbert, M. T. P. *et al.* Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. *Science* **320**, 1787 (2008).

14 Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* **20**, 231-236, doi:10.1016/j.cub.2009.11.068 (2010).

15 Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* **23**, 553-559, doi:10.1016/j.cub.2013.02.044 (2013).

16 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, doi:10.1038/nature13810 (2014).

17 Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

18 Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894-897, doi:10.1038/nature08976 (2010).

19 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

20 Sawyer, S. *et al.* Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc Natl Acad Sci U S A* **112**, 15696-15700, doi:10.1073/pnas.1519905112 (2015).

21 Horai, S. *et al.* Man's place in hominoidea revealed by mitochondrial DNA genealogy. *Journal of Molecular Evolution* **35**, 32-43, doi:10.1007/BF00160258 (1992).

22 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

23 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729, doi:10.1093/molbev/mst197 (2013).

24 Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593, doi:10.1093/bioinformatics/btq706 (2011).

25 Dalen, L. *et al.* Partial genetic turnover in neandertals: continuity in the East and population replacement in the West. *Mol Biol Evol* **29**, 1893-1897, doi:10.1093/molbev/mss074 (2012).

26 Duggan, A. T. *et al.* Maternal history of Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the Austronesian expansion. *The American Journal of Human Genetics* **94**, 721-733 (2014).

27 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and high-performance computing. *Nature methods* **9**, 772-772, doi:10.1038/nmeth.2109 (2012).

28 Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526, doi:10.1093/oxfordjournals.molbev.a040023 (1993).

29 Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **10**, e1003537, doi:10.1371/journal.pcbi.1003537 (2014).

30 Baele, G. *et al.* Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. *Molecular Biology and Evolution* **29**, 2157-2167, doi:10.1093/molbev/mss084 (2012).

31 Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A. & Lemey, P. Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. *Molecular Biology and Evolution* **30**, 239-243, doi:10.1093/molbev/mss243 (2013).

32 Kass, R. E. & Raftery, A. E. Bayes factors. *Journal of the american statistical association* **90**, 773-795 (1995).

33 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206, doi:10.1038/nature18964 (2016).

34 Karmin, M. *et al.* A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome research* **25**, 459-466 (2015).

35 Mendez, F. L., Poznik, G. D., Castellano, S. & Bustamante, C. D. The Divergence of Neandertal and Modern Human Y Chromosomes. *Am J Hum Genet* **98**, 728-734, doi:10.1016/j.ajhg.2016.02.023 (2016).

36 Barbieri, C. *et al.* Refining the Y chromosome phylogeny with southern African sequences. *Human genetics* **135**, 541 (2016).

37 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303 (2010).

38 Poznik, G. D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562-565 (2013).

39 Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290 (2004).

40 Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720-730 (2013).

doi:10.1038/nature26151

Table S5.1 Number of unique mitochondrial fragments longer than 35 base pairs with mapping quality of at least 25 that were recovered from the shotgun data of five late Neandertals after re-alignment to the *Vindija 33.16* mitochondrial genome, average coverage and average fragment length. The numbers for all fragments and for fragments with terminal C-to-T substitutions are reported.

Specimen	*All fragments*				*Fragments with terminal C-to-T substitutions*		
	Number of mapped fragments ≥ 35bp, MQ≥25	Number of unique fragments ≥ 35bp, MQ≥25	Average coverage	Average fragment length	Number of fragments ≥ 35bp, MQ≥25	Average coverage	Average fragment length
Les Cottés Z4-1514	105,878	105,874	309.58	48.6	40,696	120.87	49.2
Goyet Q56-1	48,848	48,843	165.69	56.4	7,853	26.33	55.5
Mezmaiskaya 2	304,964	304,914	988.26	53.9	90,396	291.36	53.4
Vindija 87	288,751	288,743	839.21	48.3	111,599	323.19	48.0
Spy 94a	29,146	29,146	81.55	46.5	7,945	22.57	47.1

Table S5.2 Number of pairwise nucleotide differences between the mtDNA genomes of 17 previously published Neandertals (in brown), five late Neandertals (in red), three Denisovans (in green) and the Chimpanzee (in black). All positions containing gaps and missing data were removed from all individuals, leaving in total 16120 positions in the final dataset that was used for this comparison.

Individual1 /Individual2	rCRS	Mezmaiskaya 1	Feldhofer 1	Feldhofer 2	Vindija 33.25	Vindija 33.16	Vindija 33.17	Vindija 33.19	El Sidrón 1253	Altai Neandertal	Okladnikov 2	DC1227	Goyet Q57-2	Goyet Q57-3	Goyet Q305-4	Goyet Q305-7	Goyet Q374a-1	Goyet Q56-1	Les Cottés Z4-1514	Mezmaiskaya 2	Spy 94a	Vindija 87	Hohlenstein Stadel	Denisova 3	Denisova 2	Denisova 8	Sima de los Huesos
Mezmaiskaya 1	164																										
Feldhofer 1	167	37																									
Feldhofer 2	163	33	8																								
Vindija 33.25	167	37	0	8																							
Vindija 33.16	168	38	9	7	9																						
Vindija 33.17	167	37	8	6	8	1																					
Vindija 33.19	168	38	9	7	9	0	1																				
El Sidrón 1253	164	34	9	3	9	8	7	8																			
Altai Neandertal	156	28	35	31	35	36	35	36	32																		
Okladnikov 2	159	31	14	12	14	17	16	17	13	29																	
DC1227	158	26	11	7	11	12	11	12	8	24	5																
Goyet Q57-2	168	38	1	9	1	10	9	10	10	36	15	12															
Goyet Q57-3	168	38	1	9	1	10	9	10	10	36	15	12	0														
Goyet Q305-4	164	34	9	1	9	8	7	8	4	32	13	8	10	10													
Goyet Q305-7	171	41	12	10	12	5	4	5	11	39	20	15	13	13	11												
Goyet Q374a-1	171	41	12	10	12	5	4	5	11	39	20	15	13	13	11	0											
Goyet Q56-1	171	41	12	10	12	5	4	5	11	39	20	15	13	13	11	0	0										
Les Cottés Z4-1514	174	46	31	27	31	32	31	32	28	44	25	20	32	32	28	35	35	35									
Mezmaiskaya 2	164	34	9	3	9	8	7	8	4	32	13	8	10	10	4	11	11	11	28								
Spy 94a	170	40	11	9	11	4	3	4	10	38	19	14	12	12	10	1	1	1	34	10							
Vindija 87	168	38	9	7	9	0	1	0	8	36	17	12	10	10	8	5	5	5	32	8	4						
Hohlenstein-Stadel	149	77	81	79	81	82	81	82	80	72	74	73	81	81	80	83	83	83	92	80	82	82					
Denisova 3	331	320	327	321	327	328	327	328	324	314	319	316	328	328	320	331	331	331	326	324	330	328	310				
Denisova 2	331	320	327	321	327	328	327	328	324	314	319	316	328	328	320	331	331	331	326	324	330	328	310	2			
Denisova 8	304	298	300	294	300	301	300	301	297	290	292	289	301	301	293	304	304	304	299	297	303	301	295	77	75		
Sima de los Huesos	267	269	276	270	276	277	276	277	273	263	268	265	277	277	271	278	278	278	281	273	277	277	270	205	207	183	
Chimpanzee	1388	1368	1369	1367	1369	1370	1371	1370	1370	1369	1365	1364	1368	1368	1368	1373	1373	1373	1365	1368	1372	1370	1337	1395	1395	1393	1379

Table S5.3 Mitochondrial genomes of ancient modern humans and Neandertals with radiocarbon dates that were used in BEAST analyses. All radiocarbon dates were calibrated using OxCal 4.2[40].

Hominin group	Individual	mtDNA accession no.	Radiocarbon date: Point estimate	Radiocarbon date: 95% confidence interval	Publication of the date
ancient modern humans	Ust'Ishim		45020	42560 - 47480	Fu et al, 2014
	Tianyuan	KC417443	39008	37761 - 40254	Fu et al, 2013
	Kostenki 14	FN600416	37470	36260 - 38680	Krause et al, 2010a
	Dolní Věstonice 13	KC521459	30870	30670 - 31070	Fu et al, 2013
	Dolní Věstonice 14	KC521458	30870	30670 - 31070	Fu et al, 2013
	Oberkassel 998	KC521457	14130	13758 - 14501	Fu et al, 2013
	BS11	KC521454	8050	7940 - 8160	Fu et al, 2013
	Loschbour	KC521455	8050	7940 - 8160	Fu et al, 2013
	Iceman	EU810403	5191	5067 - 5315	Ermini et al, 2008
	Saqqaq Eskimo	EU725621	4504	4423 - 4585	Gilbert et al, 2008
Neandertals	Goyet Q305-4	KX198087	44290	43430 - 45150	Rougier et al, 2016
	Mezmaiskaya 2	MG025537	43780	42960 - 44600	Pinhasi et al, 2011
	Feldhofer 1	FM865407	43710	42670 - 44750	Schmitz et al, 2002
	Vindija 33.16	AM948965	43710	39240 - 48180	Serre et al, 2004
	Feldhofer 2	FM865408	43265	42190 - 43350	Schmitz et al, 2002
	Les Cottés Z4-1514	MG025536	43230	42720 - 43740	this study
	Goyet Q56-1	KX198082	42540	42080 - 43000	Rougier et al, 2016
	Goyet Q57-3	KX198083	42430	41960 - 42900	Rougier et al, 2016
	Goyet Q57-2	KX198088	41210	40620 - 41800	Rougier et al, 2016
	Spy 94a	MG025538	38515	37880 - 39150	Semal et al, 2009

doi:10.1038/nature26151

Table S5.4 Mean ages and 95% HPD intervals of the major splits in the hominin mtDNA phylogeny as reported by Posth *et al.*[10] or estimated by BEAST using different substitution rates. HST – Hohlenstein-Stadel.

Analysis	mtDNA lineage	Mean age	lower 95% HPD	upper 95% HPD
Posth *et al.*[10]	Modern humans-Neandertals	412,930	360,230	467,720
Fixed substitution rate	Modern humans-Neandertals	428,096	379,630	479,052
Estimated substitution rate	Modern humans-Neandertals	329,780	279,826	382,774
Posth *et al.*[10]	HST-Altai branch Neandertals	267,770	218,980	306,080
Fixed substitution rate	HST-Altai branch Neandertals	273,452	229,373	317,695
Estimated substitution rate	HST-Altai branch Neandertals	215,093	176,240	254,290
Posth *et al.*[10]	Altai-rest of the Neandertals	160,480	125,410	198,800
Fixed substitution rate	Altai-rest of the Neandertals	172,811	140,779	208,015
Estimated substitution rate	Altai-rest of the Neandertals	139,379	112,367	168,080
Posth *et al.*[10]	San-modern humans	146,730	123,650	169,520
Fixed substitution rate	San-modern humans	150,120	124,825	177,067
Estimated substitution rate	San-modern humans	114,834	92,641	138,604

Table S5.5 Mean age estimates and 95% HPD intervals for non-radiocarbon dated Neandertals when using either a fixed substitution rate or a substitution rate estimated from the ancient modern humans and Neandertals with known radiocarbon dates.

Individual (GenBank accession number)	Fixed mutation rate		Estimated mutation rate	
	Mean age	95% HPD interval	Mean age	95% HPD interval
Altai Neandertal (KC879692)	142,020	102,741-181,694	116,264	85,726-147,577
DC1227 (KU131206)	101,303	78,806-125,015	86,126	69,417-104,499
El Sidron 1253 (FM865409)	56,670	40,460-76,371	53,279	39,866-66,731
Goyet Q305-7 (KX198086)	40,307	32,211-46,522	40,917	35,084-45,651
Goyet Q374a-1 (KX198085)	40,312	32,422-46,685	40,924	35,092-45,690
Goyet Q57-1 (KX198082)	44,288	34,133-55,551	44,010	36,731-52,493
Hohlenstein-Stadel (KY751400)	128,512	70,800-187,099	106,658	61,415-148,939
Mezmaiskaya 1 (FM865411)	101,304	64,756-139,751	85,930	57,244-114,232
Okladnikov 2 (KF982693)	94,141	68,805-118,995	80,806	62,764-100,502
Vindija 33.17 (KJ533544)	52,293	44,606-60,124	50,559	44,253-56,721
Vindija 33.25 (FM865410)	44,332	34,458-55,538	44,042	36,797-52,698

Supplementary Information 6

Basic set of filters for nuclear analyses

Deamination of cytosine (C) to uracil (U) residues occurs primarily at single-stranded DNA overhangs and leaves characteristic C-to-T substitutions in ancient DNA sequence alignments, which are particularly frequent close to alignment ends[1]. These elevated C-to-T substitution frequencies are used to provide evidence for the presence of authentic ancient DNA in specimens[2,3]. The molecules of the majority of ancient genomes published so far have been treated with a combination of *E. coli* uracil-DNA-glycosylase (UDG) and *E.coli* endonuclease VIII, which removes uracils from the interior parts of DNA molecules but leaves a proportion of uracils at the ends of the molecules unaffected. We omitted this step for the majority of the libraries from the five Neandertal specimens presented in this study in order to maximize the amounts of endogenous DNA fragments. Therefore, elevated C-to-T substitutions were found not only at the terminal positions of DNA fragments, but also throughout the length of the fragments at frequencies of between 2% and 5% (see Supplementary Information 3, Extended Data Fig. 1). Due to this, previous approaches of diminishing the effect of cytosine deamination by removing the bases from the ends of the fragments with elevated C-to-T substitutions[4] or reducing base quality scores at the terminal positions of fragments[5,6] prior to random read sampling or genotype calling could not be applied here. Thus, we investigated different approaches of mitigating the effects of cytosine deamination on the downstream genetic inferences.

Testing random sampling of bases and *D-statistics*

As the data presented in this study were of low coverage and contained elevated C-to-T substitutions, standard tools for genotype calling could not be utilized. Instead, we sampled a random fragment at each position in the genome. However, the correlations between individuals due to DNA damage could affect downstream analyses. In order to investigate spurious correlations stemming from the properties of the data, as well as the effects of different filtering schemes, we used *D-statistics*[7-9] that provides a robust measure of derived allele sharing between populations. If *W, X, Y* and *Z* are four populations or four different individuals for which we sampled an allele from a random fragment at each position in the genome, then we can count the number of times we encounter the site patterns ABBA and BABA. The count n_{BABA} is the number of alleles agreeing in populations W and Y, as well as

in X and Z (but different from W and Y), and n_{ABBA} is the number of alleles agreeing in populations W and Z, as well as in Y and X (but different from W and Z). *D-statistics* is then calculated as:

$$D(W,X;Y,Z) = \frac{nBABA - nABBA}{nBABA + nABBA}$$

A positive *D-value* is an indication of an elevated allele sharing between populations W and Y, or X and Z. However, even small biases in the data can cause highly significant, but spurious results if the tested genomes show little variation in the number of BABA and ABBA counts.

To investigate different sampling schemes and their effect on the downstream analyses, we used genomes of present-day humans from the B-panel of Prüfer *et al.*[5]. To mimic ancient DNA miscoding lesions, we artificially deaminated DNA fragments *in silico* by turning a fraction of Cs into Ts, as well as Gs into As. We simulated 2% deamination throughout fragments. All the analyses were performed by using *heffalump* (https://bitbucket.org/ustenzel/heffalump; see the section below on the *heffalump* file format and tools). All genomes were imported directly from BAM files by randomly picking one read at each site and *D-values* were calculated directly by the *heffalump* software. Standard errors were computed using a Weighted Block Jackknife[9,10] with equally sized blocks of 5 million base pairs (5 Mb) across all autosomes.

If we simulate deamination in two individuals and place them on opposite sides of the *D-statistics*, we expect a significant *D-value*. If we use a common pattern of "CCCC" and the effective size of mappable regions of the human genome is 2 Gb, there are around 400 million such sites. Given the assumed rate of deamination, this pattern turns into "CTCC" or "CCTC" once in fifty sites each, or into "CTTC" once in 2,500 sites. Together with the similarly behaving "GGGG" pattern, deamination is expected to result in 320,000 additional ABBA counts, which outnumber real ABBA and BABA counts between pairs of individuals and skew the real signal.

Therefore, to mitigate the effects of cytosine deamination, we considered different sampling schemes:

- *Individual-1* and *Individual-2* are the same individual for which random read sampling was performed twice independently.
- *Individual-d* had simulated deamination. This was achieved by random read sampling as above and then turning each C into a T and each G into an A with a 2% chance.

- *Individual-i* was sampled as above, but Ts in the sequencing reads were ignored and a different read covering the position was picked instead.
- *Individual-j* was a combination of *Individual-d* and *Individual-i* approaches where we first simulated deamination and then randomly picked a read while ignoring Ts and sampled a different read instead.
- *Individual-k* was sampled as above, but Ts in the sequencing reads were converted into Ns.
- *Individual-s* ignored all Cs and Ts in the forward reads, as well as Gs and As in the reverse reads, then sampled randomly from the remaining reads at each site and was therefore strand-sensitive.
- *Individual-u* crudely simulated the treatment with UDG by ignoring 2% of Cs in the sequencing reads. This simulated the deamination of Cs into Us, which were subsequently ignored as they would be incised from ancient DNA molecules treated with UDG.

We performed each analysis by using both transition and transversion polymorphisms and by using transversion polymorphisms only, as they are not affected by ancient DNA damage (Table S6.1). We expect that *D (Han-1, Han-2; Mandenka-1; Mandenka-2)* is indistinguishable from 0, as Han-1 and Han-2, as well as Mandenka-1 and Mandenka-2 are the same individual for which random read sampling at each position in the genome was done independently twice. When any of the sampling schemes listed above were applied only to a single individual in the *D-statistics*, there were no significant correlations between individuals and the *D-value* was still indistinguishable from 0. However, if two deaminated individuals were placed on the opposite sides of the *D-statistics*, as in the *D (Han-1, Han-d; Mandenka-d, Mandenka-2)*, the underlying tree was violated and deamination caused individuals to appear significantly related/admixed when they were not (Table S6.1). Furthermore, all of the sampling schemes, except random read sampling and then restricting to transversions, as well as simulation of UDG treatment, resulted in significant spurious correlations between samples that have been treated the same way (Table S6.1).

Basic set of filters for datasets used for nuclear analyses

As most of the data of late Neandertals presented in this study were not treated with UDG and endonuclease VIII, we performed random read sampling by picking a base at each position in the genome that was covered by at least one high quality fragment longer than 35 base pairs with mapping quality of at least 25 and that was within the highly mappable regions of the

genome (L≥35 bp, MQ≥25, BQ≥30 and Map35_100% of Prüfer *et al.*[5]). To diminish the impact of present-day human DNA contamination and enrich for the endogenous fragments[11], we further selected the fragments that showed C-to-T substitutions relative to the human reference genome at the first three and/or the last three positions, *i.e.* putatively deaminated fragments. Only for the libraries A9121 and A9122 of *Vindija 87* and *Goyet Q56-1* (Table S3.1) that were treated with UDG and endonuclease VIII we selected putatively deaminated fragments by requiring C-to-T substitutions at the first position and/or the last two positions of the sequence alignment.

We included newly generated single stranded DNA sequencing data of *Mezmaiskaya 1*[12], a ~60,000-70,000-year-old Neandertal from Russia, in our analyses for comparative purposes. The data of *Mezmaiskaya 1* were processed in exactly the same way as the data of *Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87* and *Spy 94a* (see Supplementary Information 3 for the processing details). As we determined that the proportion of present-day human DNA contamination in the newly generated *Mezmaiskaya 1* data was around 2% (Table S6.2), we selected the fragments that showed C-to-T substitutions relative to the human reference genome at the first three and/or the last three positions prior to the random read sampling for all of the downstream analyses.

We used the new *snpAD* genotype calls (http://cdna.eva.mpg.de/neandertal/Vindija/VCF/) for the high coverage genomes of *Altai*[5] and *Vindija 33.19*[12] Neandertals, as well as the *Denisovan*[4] individual. Furthermore, for the *D-statistics* in the form of *D(Neandertal*$_1$*, Neandertal*$_2$*, Neandertal*$_3$*, outgroup)* (see Supplementary Information 9), we applied random read sampling from the BAM files of the high coverage *Altai* and *Vindija 33.19* Neandertals in the same way as we did for the late Neandertals to avoid correlations between samples due to the differences in the sequence quality. For *Vindija 33.19* we used only the sequencing data stemming from the libraries that were not UDG and endonuclease VIII treated in order to better match the sequencing data of the late Neandertals and *Mezmaiskaya 1*[12]. The analyses were further restricted to the fragments that showed C-to-T substitutions with respect to the human reference genome in the first and/or the last three positions. For the *Altai* Neandertal, whose data come from libraries treated with UDG and endonuclease VIII[5], we retained fragments that showed C-to-T substitutions at the first and/or the last two positions.

In order to infer the relationship between the introgressing Neandertals in present-day humans and the Neandertals whose genomes have been sequenced (see Supplementary Information 10), we used *hetfa* files of 263 present-day human genomes of the Simons

Genome Diversity Panel (SGDP)[13]. In order to investigate whether there is a difference among Neandertals in their proximity to the introgressed Neandertal detected in ancient modern humans, we have used the new *snpAD* genotype calls (http://cdna.eva.mpg.de/neandertal/Vindija/VCF/) of the high-quality genomes of *Ust'-Ishim*, a ~45,000-year-old modern human from Siberia[14]; *Loschbour*, a ~8,000-year-old hunter-gatherer from Luxembourg[15]; and *LBK*, a ~7,000-year-old farmer from Stuttgart[15].

We used *heffalumps* (https://bitbucket.org/ustenzel/heffalump) to extract variable positions across all genomes into an input format for *AdmixTools* (version 4.1)[9] and *TreeMix*[16], or to export them into combined VCF files. We further restricted all of the analyses to bi-allelic sites in the genome covered by at least one low coverage Neandertal and transversion polymorphisms unless otherwise indicated.

Heffalump **file format and tools**

Most analyses of genetic variation need the genotype information of individuals at variant sites as input. Storing this information across the entire human genome as VCF files easily requires about 70 GB per sample when the VCF file is compressed or about 500 GB when uncompressed. Therefore, set operations like the intersection or the union of VCF files with a large number of samples require too much time and computational power when being performed repeatedly. In order to make these processes faster and easier, we implemented a new file format called *heffalump* (https://bitbucket.org/ustenzel/heffalump) that stores genotype calls compactly, and additionally provides a set of tools to operate on these files. A *heffalump* file stores variant calls, while compressing runs of invariant calls or missing data into a single code alongside the length of the run. Using this approach, we obtain a similar compactness to a file that only stores variant sites but without losing any information. Each individual is stored separately and merging them generates the list of variants on the fly. This file format design enables us to merge variant information of many individuals faster compared to using VCF files.

The *heffalump* tool supports the import of the data from commonly used file formats, *e. g.* BCF/VCF (used for most genotype calls), *hetfa* (used in the SGDP dataset)[13], MAF[17] (used for whole-genome alignments with Great Apes, which we use as outgroups), or BAM. When importing BAM files, *heffalump* randomly samples one base as representative for each site instead of genotype calling. Output formats are *Eigenstrat* (used by the *Eigensoft* tool suite), *TreeMix* and standard VCF format.

For our analyses, we used *heffalump* for randomly sampling alleles from BAM files for our low coverage individuals and imported the genotype information of the high coverage hominins from VCF files. Present-day human genotypes from the SGDP data set were directly imported from *hetfa* files, while alignments of the Great Ape reference genomes against the human reference genome were imported from MAF files. Data sets were merged using *heffalump* and written directly in the output format required for the dedicated analysis.

References for SI6:

1 Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* **104**, 14616-14621, doi:10.1073/pnas.0704665104 (2007).

2 Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr Biol* **20**, 231-236, doi:10.1016/j.cub.2009.11.068 (2010).

3 Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131, doi:10.1371/journal.pone.0034131 (2012).

4 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

5 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

6 Sawyer, S. *et al.* Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc Natl Acad Sci U S A* **112**, 15696-15700, doi:10.1073/pnas.1519905112 (2015).

7 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

8 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

9 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, doi:10.1534/genetics.112.145037 (2012).

10 Busing, F. M. T. A., Meijer, E. & Van Der Leeden, R. Delete-m jackknife for unequal m. *Statistics and Computing* **9**, 3-8, doi:Doi 10.1023/A:1008800423698 (1999).

11 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

12 Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, doi:10.1126/science.aao1887 (2017).

13 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206, doi:10.1038/nature18964 (2016).

14 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, doi:10.1038/nature13810 (2014).

15 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).

16 Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967, doi:10.1371/journal.pgen.1002967 (2012).

17 Rosenbloom, K. R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res* **43**, D670-681, doi:10.1093/nar/gku1177 (2015).

doi:10.1038/nature26151

Table S6.1 Testing of different sampling schemes and their effects on *D-statistics* inferences. In order to mitigate the effects of cytosine deamination in ancient DNA specimens we investigated different sampling schemes and their effects on *D-statistics* by simulating deamination on genomes of present-day individuals from the B-team of Prüfer *et al.*[5]. Standard errors are from a weighted block jackknife.

Individuals				Transitions and transversions			Transversions only		
W	X	Y	Z	Counts (nBABA-nABBA) / (nBABA+nABBA)	D ± SE (%)	P-value	Counts (nBABA-nABBA) / (nBABA+nABBA)	D ± SE (%)	P-value
Han-1	Han-2	Mandenka-2	Mandenka-1	181825/362882	0.21 ± 0.16	0.097	107904/215258	0.26 ± 0.21	0.117
Han-1	Han-2	Mandenka-d	Mandenka-1	189617/378985	0.07 ± 0.17	0.347	107389/214591	0.09 ± 0.23	0.351
Han-1	Han-2	Mandenka-i	Mandenka-1	207039/413157	0.22 ± 0.16	0.079	125369/250418	0.13 ± 0.21	0.269
Han-1	Han-2	Mandenka-j	Mandenka-1	206742/412961	0.13 ± 0.16	0.214	125549/250440	0.26 ± 0.20	0.097
Han-1	Han-2	Mandenka-k	Mandenka-1	135423/271039	-0.07 ± 0.20	0.358	80228/160397	0.04 ± 0.25	0.441
Han-1	Han-2	Mandenka-s	Mandenka-1	200040/399248	0.21 ± 0.16	0.094	122185/243844	0.22 ± 0.21	0.151
Han-1	Han-2	Mandenka-u	Mandenka-1	182049/363629	0.13 ± 0.16	0.216	108047/215648	0.21 ± 0.22	0.169
Han-1	Han-d	Mandenka-2	Mandenka-1	188395/377495	0.19 ± 0.16	0.120	107035/214270	-0.09 ± 0.22	0.335
Han-1	Han-i	Mandenka-2	Mandenka-1	204568/409555	-0.10 ± 0.16	0.264	122470/245075	-0.06 ± 0.21	0.395
Han-1	Han-j	Mandenka-2	Mandenka-1	203956/408076	-0.04 ± 0.16	0.399	121866/243806	-0.03 ± 0.21	0.442
Han-1	Han-k	Mandenka-2	Mandenka-1	136067/271955	0.07 ± 0.18	0.360	81169/161904	0.27 ± 0.24	0.134
Han-1	Han-s	Mandenka-2	Mandenka-1	198493/396600	0.10 ± 0.16	0.272	119828/239049	0.25 ± 0.21	0.114
Han-1	Han-u	Mandenka-2	Mandenka-1	181175/362513	-0.04 ± 0.17	0.397	107595/215096	0.04 ± 0.22	0.421
Han-1	Han-d	Mandenka-d	Mandenka-1	188395/377495	-22.15 ± 0.24	0.00E+00	106549/213557	-0.21 ± 0.23	0.172
Han-1	Han-i	Mandenka-i	Mandenka-1	208028/502309	-17.17 ± 0.16	0.00E+00	128701/307438	-16.28 ± 0.20	0.00E+00
Han-1	Han-j	Mandenka-j	Mandenka-1	207129/497974	-16.81 ± 0.17	0.00E+00	128047/304928	-16.01 ± 0.20	0.00E+00
Han-1	Han-k	Mandenka-k	Mandenka-1	94694/210275	-9.93 ± 0.22	0.00E+00	57950/126804	-8.60 ± 0.28	1.20E-203
Han-1	Han-s	Mandenka-s	Mandenka-1	202650/473916	-14.48 ± 0.20	0.00E+00	123307/298661	-17.43 ± 0.25	0.00E+00
Han-1	Han-u	Mandenka-u	Mandenka-1	181858/363084	-0.17 ± 0.18	0.167	108032/215388	0.31 ± 0.22	0.081

Individual-1 and *Individual-2* - the same individual for which random read sampling was performed twice; *Individual-d* – simulated deamination; *Individual-i* – Ts in the sequencing reads are ignored; *Individual-j* – a combination of *Individual-d* and *Individual-i*; *Individual-k* – Ts are converted into Ns; *Individual-s*- strand specific; *Individual-u* – UDG treatment.

doi:10.1038/nature26151

Table S6.2 Proportion of present-day human DNA contamination among all and putatively deaminated fragments of the comparative dataset of *Mezmaiskaya 1* Neandertal[12]. 95% binomial confidence intervals (CI) are reported in brackets for each of the estimates.

	Nuclear DNA contamination estimates			Mitochondrial DNA contamination estimates			
	Autosomal contamination estimates	Sex-based contamination estimates for female individuals		311 present-day humans ≠ 18 Neandertals		311 present-day humans ≠ diagnostic positions of the reconstructed mitochondrial genome	
Individual	**% of all fragments matching derived states in present-day humans [95% CI]**	**% Y chromosome contamination in all framents [95% CI]**	**% Y chromosome contamination in deaminated framents [95% CI]**	**% of all fragments matching present-day human state [95% CI]**	**% of deaminated fragments matching present-day human state [95% CI]**	**% of all fragments matching present-day human state [95% CI]**	**% of deaminated fragments matching present-day human state [95% CI]**
Mezmaiskaya 1	1.76 *[0.38-2.54]*	2.23 *[1.98-2.46]*	1.24 *[1.11-1.35]*	2.07 *[1.76-2.44]*	1.05 *[0.58-1.66]*	2.05 *[1.71-2.53]*	0.87 *[0.32-1.35]*

doi:10.1038/nature26151

Supplementary Information 7

Principal Component Analysis, lineage attribution and divergence estimates

Principal Component Analysis (PCA)

We carried out a Principal Component Analysis (PCA)[1,2] using the genomes of *Vindija 33.19*[3], the *Altai* Neandertal[4] and *Denisova*[5] to estimate the eigenvectors of the genetic variation and then projected the five late Neandertals and *Mezmaiskaya 1*[3] onto the plane that was defined by these eigenvectors, which allowed us to explore the relationship of low coverage Neandertals relative to the high coverage archaics. Only transversion polymorphisms and bi-allelic sites were considered for the analysis. Extended Data Fig. 4 shows that the first principal component separated *Altai* and *Vindija 33.19* Neandertals from the *Denisovan* individual, whereas the second principal component separated the *Altai* Neandertal from *Vindija 33.19*. All of the late Neandertals, as well as the older *Mezmaiskaya 1* individual, fell closer to *Vindija 33.19* than to the *Altai* Neandertal.

Lineage attribution and sharing of the derived alleles with *Altai* and *Vindija 33.19* Neandertals

In order to determine more precisely to which hominin group the nuclear genomes of *Les Cottés Z4-1514, Goyet Q56-1, Spy 94a, Mezmaiskaya 2* and *Vindija 87* were most closely related to on average, we first followed an approach introduced in Meyer *et al.*[6] based on the sharing of derived alleles with different hominin groups. We investigated the state of DNA fragments overlapping the positions at which the high coverage genomes of the *Altai* Neandertal[4], the *Vindija 33.19* Neandertal[3], the *Denisovan* individual[5] and a present-day African (Mbuti, HGDP00982)[4] differ from those of the great apes (chimpanzee, bonobo, gorilla and orangutan). We then calculated the proportion of fragments for each of the low coverage Neandertals that supported the derived state of each of the branches in the phylogenetic tree relating the four high coverage hominin genomes[6]. We included previously published low coverage nuclear data of *El Sidrón 1253, Feldhofer 1, Vindija 33.16, Vindija 33.25, Vindija 33.26*[7], *Denisova 4* and *Denisova 8*[8], as well as the newly generated data of *Mezmaiskaya 1*[3], in our analyses for comparative purposes. We applied the same filtering scheme to these data as for the five late Neandertals.

The analyses were carried out using all fragments and putatively deaminated fragments. For the data generated using double stranded DNA library preparation (*El Sidron*

1253, Feldhofer 1, Vindija 33.16, Vindija 33.25 and *Vindija 33.26*), deaminated fragments were identified by a C-to-T substitution to the reference genome on the first three positions on the 5'-end or a G-to-A substitution at the last three positions on the 3'-end. For the late Neandertal data generated using single stranded DNA library preparation, and the newly generated data of *Mezmaiskaya 1*, a terminal C-to-T substitution was required on the first three and/or the last three positions on both ends of the alignment (see Supplementary Information 6). For the UDG/endonuclease VIII treated data of *Denisova 4* and *Denisova 8*, we required a terminal C-to-T substitutions on the first and/or the last two positions of the alignment ends.

For the five late Neandertals we found that 93.50% (95% CI: 91.17-96.26%) of the positions on the Neandertal branch carried derived alleles, while 0.98% (95% CI: 0.73-1.76%) of the positions on the human branch carried derived alleles, and 0.84% (95% CI: 0.45-1.08%) of the position on the Denisovan branch carried derived alleles (Table S7.1 and Extended Data Table 3). These results were concordant when using all fragments and when using putatively deaminated fragments, and similar to those obtained for *Vindija 33.16, Vindija 33.25* and *Vindija 33.26* (Table S7.1 and Extended Data Table 3). Thus, we conclude that all five specimens originated from Neandertal individuals. Moreover, all of the late Neandertals shared more derived alleles with the *Vindija 33.19* branch (between 46.64% and 76.29%) than with the *Altai* Neandertal (between 7.38% and 19.01%), suggesting that all late Neandertals analysed here, as well as an older *Mezmaiskaya 1*, were genetically closer to *Vindija 33.19* than to the *Altai* Neandertal.

Estimating the average sequence divergence of the low coverage Neandertals using the triangulation method

We estimated the divergence of the five late Neandertal genomes along the lineage from the ancestor shared with the chimpanzee and the high coverage genomes of *Altai* and *Vindija 33.19* Neandertals, the *Denisovan* individual, or a present-day human from the B-panel of Prüfer *et al.*[4], using the triangulation method that was previously applied to a number of archaic genomes[4,5,7-9].

The Enredo-Pecan-Ortheus (EPO) 6-way primate genome alignments from Ensembl version 69[10,11] were used to infer the human-chimpanzee common ancestor sequence[4,9]. To exclude the ambiguously aligned regions, only the alignments where a single human and a single chimpanzee base were present in the EPO alignment were kept for the analysis (https://github.com/grenaud/epoParser). We selected a random allele from the genotype calls

of the high coverage individuals and picked a base at random from the high-quality fragments for the late Neandertals (see Supplementary Information 6). We then counted how many of the changes in the high coverage genomes likely occurred before or after the split from the low coverage genome. Standard errors were computed by a Weighted Block Jackknife[12] with a block size of 5 million base pairs (5 Mb) across all autosomes. Again, the analyses were carried out using all fragments and putatively deaminated fragments, restricting all analyses to transversion polymorphisms.

The results of the pairwise divergence for all fragments are shown in Table S7.2, and for deaminated fragments in Table S7.3, respectively. The fraction of derived alleles not shared with the *Altai* Neandertal genome was on average 2.28% (95% CI: 2.14-2.38%) among late Neandertals, compared to 1.8% (95% CI: 1.6-1.97%) not shared with *Vindija 33.19* (tables S7.3 and S7.4). Again, this would suggest that *Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87* and *Spy 94a* were genetically closer to *Vindija 33.19* than to the *Altai* Neandertal. This was also the case for the low-coverage data of *Vindija 33.16, Vindija 33.25* and *Vindija 33.26*. Furthermore, the estimated divergence among these nearly contemporaneous late Neandertals to the *Vindija 33.19* or the *Altai* Neandertal was approximately one-third of that estimated for the comparative data of present-day humans worldwide and approximately half what is observed among non-African individuals (Table S7.4).

As it was previously shown[3], *Vindija 33.19* and *Altai* Neandertal do not differ much in their estimated diversities, with *Altai* Neandertal having 1.58 heterozygous sites per 10,000 bases and *Vindija 33.19* having 1.62 heterozygous sites per 10,000 bases. This suggests that the effective population sizes for the two groups of Neandertals to which these individuals belonged were rather small and similar in size (around 3,000 individuals) even though they lived at least 60,000 years apart. As the late Neandertals presented in this study show similar patterns in the population genetic analyses as the roughly contemporaneous *Vindija 33.19* (see Supplementary Information 9 and Supplementary Information 10), we speculate that that their population sizes, as well as the number of deleterious derived variants would be in the same order of magnitude. This is supported by the analysis of the divergence of the late Neandertals, as well as *Mezmaiskaya 1*, and the three low-coverage Neandertals from Vindija to the high coverage genomes of the *Altai* Neandertal and *Vindija 33.19* along the lineage shared with the chimpanzee.

Vindija 87 and _Vindija 33.19_ likely belonged to the same individual

As the nuclear genomes of specimens *Vindija 87* and *Vindija 33.19* were highly similar and as they carried identical mtDNA genome sequences (Table S5.2, Supplementary Information 5), we investigated whether they belonged to the same individual. In order to exclude the possibility that even a parent-offspring relationship could explain the high similarity of the two samples, we computed the fraction of fragments for which *Vindija 87* presented the same state as *Vindija 33.19* or the *Altai* Neandertal, at positions where the two high coverage genomes were homozygous for different alleles.

Vindija 87 had the same state as *Vindija 33.19* in 97.7% of these sites. The proportion of sites that matched the *Altai* Neandertal rather than *Vindija 33.19* was only twice as high as the fraction of putative errors, *i.e.* fragments matching a third state not observed in any low coverage Neandertal. The percentage of sites that matched *Vindija 33.19* was much higher than what is expected by simulations and what was observed in other individuals from Vindija cave (*e.g.* 82.5% for *Vindija 33.16*, 81.4% for *Vindija 33.25* and 82.2% for *Vindija 33.26*), suggesting that *Vindija 33.19* and *Vindija 87* bones belonged to the same individual. Therefore, as the genome of *Vindija 33.19* is of a higher quality and higher coverage, we continued using it for all of the downstream analyses instead of the low coverage data of *Vindija 87*. Moreover, the specimen *Vindija 33.15*[13] shares 99.7% of the sites with *Vindija 33.19*, suggesting that it also comes from the same individual as *Vindija 33.19* and *Vindija 87*.

References for SI7:

1 Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190, doi:10.1371/journal.pgen.0020190 (2006).

2 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909, doi:http://www.nature.com/ng/journal/v38/n8/suppinfo/ng1847_S1.html (2006).

3 Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science, doi:10.1126/science.aao1887 (2017).

4 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

5 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

6 Meyer, M. *et al.* Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504-507, doi:10.1038/nature17405 (2016).

7 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

8 Sawyer, S. *et al.* Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc Natl Acad Sci U S A* **112**, 15696-15700, doi:10.1073/pnas.1519905112 (2015).

9 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

10 Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* **18**, 1814-1828, doi:10.1101/gr.076554.108 (2008).

11 Paten, B. *et al.* Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res* **18**, 1829-1843, doi:10.1101/gr.076521.108 (2008).

12 Busing, F. M. T. A., Meijer, E. & Van Der Leeden, R. Delete-m jackknife for unequal m. *Statistics and Computing* **9**, 3-8, doi:Doi 10.1023/A:1008800423698 (1999).

doi:10.1038/nature26151

Table S7.1 The fraction of derived alleles among all fragments that each of the low coverage individuals shares with the *Altai* Neandertal, *Vindija 33.19, Denisovan*, and a present day human genome. Fragments longer than 35 base pairs with mapping quality of at least 25 and within the highly mappable regions of the genome reported in the Table S3.1were utilized for this analysis. 95% binomial confidence intervals are provided in brackets.

	Human (%)	Neandertal (%)	Altai Neandertal (%)	Vindija 33.19 (%)	Denisova (%)	Neandertal-Denisova (%)	Human-Neandertal (%)	Human-Denisova (%)
Les Cottés Z4-1514	0.78 [0.77-0.80]	93.09 [93.03-93.16]	18.66 [18.48-18.84]	47.68 [47.47-47.88]	0.95 [0.92-0.97]	97.72 [97.67-97.76]	97.22 [97.14-97.29]	3.25 [3.16-3.34]
Goyet Q56-1	0.83 [0.81-0.85]	93.90 [93.83-93.97]	16.85 [16.66-17.04]	53.67 [53.45-53.89]	0.83 [0.81-0.85]	97.76 [97.71-97.81]	97.45 [97.37-97.53]	3.0 [2.91-3.10]
Spy 94a	1.71 [1.68-1.76]	91.30 [91.17-91.42]	16.67 [16.39-16.97]	51.84 [51.50-52.18]	1.04 [1.00-1.08]	95.82 [95.72-95.92]	96.52 [96.38-96.66]	4.99 [4.81-5.18]
Vindija 87	0.76 [0.73-0.78]	96.18 [96.11-96.26]	7.38 [7.20-7.56]	76.29 [76.29-76.80]	0.47 [0.45-0.49]	98.25 [98.19-98.30]	98.61 [98.53-98.68]	2.18 [2.08-2.29]
Mezmaiskaya 2	0.82 [0.80-0.84]	93.05 [92.97-93.14]	19.10 [18.87-19.33]	46.64 [46.39-46.90]	0.91 [0.88-0.93]	97.67 [97.61-97.73]	97.29 [97.19-97.38]	3.36 [3.25-3.48]
Mezmaiskaya 1	1.41 [1.38-1.44]	90.53 [90.42-90.63]	21.66 [21.40-21.92]	39.05 [38.78-39.32]	1.16 [1.13-1.19]	96.20 [96.12-96.28]	96.24 [96.12-96.36]	4.74 [4.60-4.88]
Vindija 33.16	2.48 [2.42-2.53]	93.09 [92.96-93.22]	15.58 [15.26-15.90]	58.87 [58.48-59.25]	2.22 [2.16-2.28]	96.18 [96.07-96.29]	96.38 [96.22-96.54]	4.39 [4.20-4.59]
Vindija 33.25	1.86 [1.81-1.91]	94.10 [93.98-94.22]	15.67 [15.35-16.00]	59.61 [59.23-60.0]	1.90 [1.84-1.96]	97.06 [96.96-97.16]	96.89 [96.73-97.04]	3.38 [3.21-3.56]
Vindija 33.26	2.14 [2.09-2.20]	93.60 [93.47-93.73]	15.73 [15.40-16.07]	58.27 [58.86-59.67]	1.98 [1.92-2.04]	96.71 [96.60-96.82]	96.66 [96.49-96.82]	3.83 [3.65-4.03]
Feldhofer 1	5.41 [3.74-7.76]	90.87 [86.57-93.89]	14.29 [7.95-24.34]	50.00 [40.65-59.35]	1.78 [0.82-3.83]	92.57 [88.11-95.45]	93.51 [85-68-97.19]	5.45 [1.87-14.85]
El Sidron 1253	1.92 [1.01-3.61]	94.80 [91.31-96.94]	16.67 [10.01-26.46]	36.89 [28.20-46.53]	4.32 [2.64-7.01]	96.27 [92.11-98.28]	95.83 [88.45-98.57]	2.00 [0.35-10.50]
Denisova 4	24.33 [23.57-25.11]	8.79 [8.10-9.53]	6.16 [5.18-7.31]	3.07 [2.45-3.83]	31.81 [30.84-32.80]	47.42 [45.94-48.90]	52.52 [50.27-54.75]	72.08 [69.88-74.19]
Denisova 8	6.80 [6.57-7.05]	7.25 [6.91-7.61]	2.73 [2.36-3.15]	1.85 [1.59-2.16]	51.69 [51.11-52.26]	82.81 [81.81-83.02]	22.46 [21.47-23.48]	85.20 [84.26-86.09]

Table S7.2 Estimates of pairwise DNA divergence between low-coverage Neandertals and the high coverage genomes of *Vindija 33.19*, the *Altai* Neandertal, *Denisova* and 12 present-day humans using all fragments. Reported is the proportion of substitutions inferred to have occurred from the human chimpanzee ancestor to the high coverage genome after the split from the low coverage genome. This proportion was calculated using all fragments longer than 35 base pairs with the mapping quality of at least 25 reported in the Table S3.1. Standard errors were calculated by Weighted Block Jackknife with a block size of 5 million base pairs (5 Mb) across all autosomes.

		Low coverage individual								
		Les Cottés Z4-1514	**Goyet Q56-1**	**Spy 94a**	**Vindija 87**	**Vindija 33.16**	**Vindija 33.25**	**Vindija 33.26**	**Mezmaiskaya 2**	**Mezmaiskaya 1**
High coverage individual	**Vindija 33.19 (%)**	2.03 ± 0.03	1.71 ± 0.02	2.16 ± 0.03	0.93 ± 0.02	1.48 ± 0.03	1.47 ± 0.03	1.45 ± 0.03	1.92 ± 0.03	2.45 ± 0.03
	Altai (%)	2.44 ± 0.03	2.28 ± 0.03	2.69 ± 0.03	2.35 ± 0.03	2.32 ± 0.04	2.27 ± 0.04	2.19 ± 0.04	2.32 ± 0.03	2.55 ± 0.03
	Denisova (%)	9.25 ± 0.06	9.12 ± 0.06	9.26 ± 0.07	9.10 ± 0.06	9.11 ± 0.08	9.16 ± 0.08	9.11 ± 0.08	9.11 ± 0.06	9.26 ± 0.06
	French (%)	11.45 ± 0.06	11.42 ± 0.06	11.12 ± 0.06	11.02 ± 0.06	11.20 ± 0.08	11.40 ± 0.08	11.13 ± 0.08	11.19 ± 0.06	11.17 ± 0.06
	Sardinian (%)	11.36 ± 0.06	11.36 ± 0.06	11.02 ± 0.06	10.96 ± 0.06	11.15 ± 0.08	11.27 ± 0.08	11.10 ± 0.08	11.12 ± 0.06	11.11 ± 0.06
	Han (%)	11.74 ± 0.06	11.70 ± 0.06	11.44 ± 0.07	11.36 ± 0.06	11.53 ± 0.08	11.64 ± 0.08	11.45 ± 0.08	11.51 ± 0.06	11.42 ± 0.06
	Dai (%)	11.64 ± 0.06	11.55 ± 0.06	11.32 ± 0.06	11.26 ± 0.06	11.43 ± 0.08	11.54 ± 0.08	11.26 ± 0.08	11.34 ± 0.06	11.29 ± 0.06
	Mixe (%)	11.49 ± 0.06	11.49 ± 0.06	11.25 ± 0.06	11.11 ± 0.06	11.29 ± 0.08	11.35 ± 0.08	11.19 ± 0.08	11.29 ± 0.06	11.21 ± 0.06
	Karitiana (%)	11.41 ± 0.06	11.39± 0.06	11.14 ± 0.06	10.96 ± 0.06	11.23 ± 0.08	11.31 ± 0.08	11.08 ± 0.09	11.18 ± 0.06	11.09 ± 0.06
	Papuan (%)	11.42 ± 0.06	11.42 ± 0.06	11.14 ± 0.06	11.04 ± 0.06	11.14 ± 0.08	11.31 ± 0.08	11.17 ± 0.08	11.21 ± 0.06	11.16 ± 0.06
	Australian (%)	11.36 ± 0.06	11.34 ± 0.06	11.06 ± 0.07	10.97 ± 0.06	11.12 ± 0.08	11.26 ± 0.08	11.09 ± 0.08	11.15 ± 0.06	11.09 ± 0.06
	Mandenka (%)	11.57 ± 0.06	11.53 ± 0.06	11.32 ± 0.07	11.23 ± 0.06	11.35 ± 0.08	11.52 ± 0.08	11.39 ± 0.09	11.36 ± 0.06	11.32 ± 0.06
	Dinka (%)	11.65 ± 0.06	11.61 ± 0.06	11.47 ± 0.06	11.32 ± 0.06	11.46 ± 0.08	11.55 ± 0.08	11.38 ± 0.08	11.44 ± 0.06	11.42 ± 0.06
	Yoruba (%)	11.58 ± 0.06	11.53 ± 0.05	11.35 ± 0.06	11.24 ± 0.06	11.41 ± 0.08	11.45 ± 0.08	11.43 ± 0.08	11.36 ± 0.06	11.38 ± 0.06
	Mbuti (%)	11.56 ± 0.06	11.50 ± 0.06	11.32 ± 0.06	11.23 ± 0.06	11.34 ± 0.08	11.44 ± 0.08	11.39 ± 0.09	11.34 ± 0.06	11.37 ± 0.06
	San (%)	11.59 ± 0.06	11.53 ± 0.06	11.35 ± 0.07	11.29 ± 0.06	11.50 ± 0.08	11.51 ± 0.08	11.36 ± 0.08	11.38 ± 0.06	11.39 ± 0.06

doi:10.1038/nature26151

Table S7.3 Estimates of pairwise DNA divergence between low-coverage Neandertals and the high coverage genomes of *Vindija 33.19*, the *Altai* Neandertal, *Denisova* and 12 present-day humans using putatively deaminated fragments. Reported is the proportion of substitutions inferred to have occurred from the human chimpanzee ancestor to the high coverage genome after the split from the low coverage genome. This proportion was calculated using fragments with terminal C-to-T substitutions relative to the reference genome longer than 35 base pairs with the mapping quality of at least 25 reported in the Table S3.2. Standard errors were calculated by Weighted Block Jackknife with a block size of 5 million base pairs (5 Mb) across all autosomes.

		Low coverage individual								
		Les Cottés Z4-1514	**Goyet Q56-1**	**Spy 94a**	**Vindija 87**	**Vindija 33.16**	**Vindija 33.25**	**Vindija 33.26**	**Mezmaiskaya 2**	**Mezmaiskaya 1**
High coverage individual	**Vindija 33.19 (%)**	1.94 ± 0.03	1.64 ± 0.04	1.78 ± 0.04	0.79 ± 0.02	1.38 ± 0.05	1.47 ± 0.05	1.41 ± 0.05	1.84 ± 0.03	2.17 ± 0.04
	Altai (%)	2.33 ± 0.04	2.25 ± 0.05	2.33 ± 0.05	2.18 ± 0.04	2.26 ± 0.06	2.32 ± 0.07	2.09 ± 0.07	2.20 ± 0.03	2.33 ± 0.04
	Denisova (%)	9.11 ± 0.07	9.03 ± 0.08	8.91 ± 0.10	8.90 ± 0.07	9.09 ± 0.12	9.26 ± 0.13	9.25 ± 0.14	8.97 ± 0.07	9.04 ± 0.08
	French (%)	11.01 ± 0.06	11.26 ± 0.08	10.81 ± 0.10	10.49 ± 0.07	10.87 ± 0.13	11.12 ± 0.14	10.69 ± 0.14	10.70 ± 0.07	10.78 ± 0.08
	Sardinian (%)	10.91 ± 0.06	11.20 ± 0.09	10.69 ± 0.10	10.43 ± 0.07	10.84 ± 0.13	11.22 ± 0.13	10.67 ± 0.15	10.60 ± 0.07	10.70 ± 0.08
	Han (%)	11.33 ± 0.06	11.53 ± 0.09	11.19 ± 0.10	10.87 ± 0.07	11.21 ± 0.14	11.39 ± 0.14	11.17 ± 0.15	11.03 ± 0.07	11.04 ± 0.08
	Dai (%)	11.19 ± 0.07	11.42 ± 0.11	10.60 ± 0.12	10.61 ± 0.09	11.13 ± 0.13	11.02 ± 0.14	10.79 ± 0.14	10.87 ± 0.08	10.90 ± 0.07
	Mixe (%)	11.03 ± 0.07	11.33 ± 0.09	10.88 ± 0.10	10.65 ± 0.07	11.06 ± 0.13	11.04 ± 0.13	10.88 ± 0.14	10.80 ± 0.07	10.74 ± 0.08
	Karitiana (%)	10.99 ± 0.07	11.25 ± 0.09	10.73 ± 0.10	10.46 ± 0.07	10.96 ± 0.13	11.09 ± 0.14	10.86 ± 0.15	10.69 ± 0.07	10.66 ± 0.08
	Papuan (%)	10.97 ± 0.06	11.24 ± 0.09	10.88 ± 0.10	10.52 ± 0.07	10.94 ± 0.13	11.02 ± 0.14	10.82 ± 0.14	10.75 ± 0.07	10.78 ± 0.08
	Australian (%)	10.94 ± 0.07	11.18 ± 0.09	10.77 ± 0.10	10.49 ± 0.07	10.85 ± 0.13	11.37 ± 0.13	10.75 ± 0.14	10.62 ± 0.07	10.72 ± 0.08
	Mandenka (%)	11.15 ± 0.07	11.36 ± 0.09	10.99 ± 0.10	10.79 ± 0.07	11.09 ± 0.13	11.49 ± 0.14	11.37 ± 0.15	10.87 ± 0.07	10.87 ± 0.08
	Dinka (%)	11.18 ± 0.06	11.46 ± 0.09	11.06 ± 0.10	10.85 ± 0.07	11.38 ± 0.13	11.42 ± 0.14	11.22 ± 0.14	11.02 ± 0.07	11.09 ± 0.08
	Yoruba (%)	11.20 ± 0.06	11.30 ± 0.08	11.10 ± 0.09	10.81 ± 0.07	11.25 ± 0.13	11.42 ± 0.14	11.27 ± 0.14	10.94 ± 0.07	11.01 ± 0.08
	Mbuti (%)	11.21 ± 0.07	11.26 ± 0.08	10.99 ± 0.09	10.81 ± 0.07	11.13 ± 0.13	11.53 ± 0.14	11.14 ± 0.14	11.00 ± 0.07	11.02 ± 0.08
	San (%)	11.25 ± 0.07	11.31 ± 0.08	10.95 ± 0.09	10.86 ± 0.07	11.36 ± 0.14	11.53 ± 0.14	11.15 ± 0.14	11.10 ± 0.07	11.05 ± 0.08

doi:10.1038/nature26151

Table S7.4 Estimates of pairwise DNA divergence for the comparative data of 12 present-day humans (B panel from Prüfer *et al.*[4], restricted to the positions in the genome where at least one low-coverage Neandertal had coverage). Standard errors were calculated by Weighted Block Jackknife with a block size of 5 million base pairs (5 Mb) across all autosomes.

Individual1/ Individual2	French	Sardinian	Han	Dai	Karitiana	Mixe	Australian	Papuan	Mandenka	Dinka	Mbuti	Yoruba	San
French (%)	Na	5.85 ± 0.06	6.63 ± 0.06	6.48 ± 0.06	6.24 ± 0.06	6.30 ± 0.06	6.54 ± 0.06	6.53 ± 0.06	7.81 ± 0.06	7.73 ± 0.06	8.42 ± 0.06	7.94 ± 0.06	8.69 ± 0.06
Sardinian (%)	5.85 ± 0.06	Na	6.76 ± 0.06	6.60 ± 0.06	6.35 ± 0.06	6.37 ± 0.06	6.60 ± 0.06	6.62 ± 0.06	7.82 ± 0.06	7.76 ± 0.06	8.39 ± 0.06	7.92 ± 0.06	8.69 ± 0.06
Han (%)	6.63 ± 0.06	6.76 ± 0.06	Na	5.71 ± 0.06	5.85 ± 0.06	5.83 ± 0.06	6.20 ± 0.06	6.25 ± 0.06	7.84 ± 0.06	7.70 ± 0.06	8.39 ± 0.06	7.95 ± 0.06	8.66 ± 0.06
Dai (%)	6.48 ± 0.06	6.60 ± 0.06	5.71 ± 0.06	Na	5.86 ± 0.06	5.84 ± 0.06	6.20 ± 0.06	6.20 ± 0.06	7.82 ± 0.06	7.64 ± 0.06	8.35 ± 0.06	7.91 ± 0.06	8.65 ± 0.06
Karitiana (%)	6.24 ± 0.06	6.35 ± 0.06	5.85 ± 0.06	5.86 ± 0.06	Na	4.96 ± 0.06	6.28 ± 0.06	6.30 ± 0.06	7.83 ± 0.06	7.74 ± 0.06	8.43 ± 0.06	7.92 ± 0.06	8.68 ± 0.06
Mixe (%)	6.30 ± 0.06	6.37 ± 0.06	5.83 ± 0.06	5.84 ± 0.06	4.96 ± 0.06	Na	6.31	6.32 ± 0.06	7.80 ± 0.06	7.68 ± 0.06	8.40 ± 0.06	7.91 ± 0.06	8.70 ± 0.06
Australian (%)	6.54 ± 0.06	6.60 ± 0.06	6.20 ± 0.06	6.20 ± 0.06	6.28 ± 0.06	6.31 ± 0.06	Na	5.49 ± 0.06	7.93 ± 0.06	7.82 ± 0.06	8.47 ± 0.06	8.05 ± 0.06	8.81 ± 0.06
Papuan (%)	6.53 ± 0.06	6.62 ± 0.06	6.25 ± 0.06	6.20 ± 0.06	6.30 ± 0.06	6.32 ± 0.06	5.49 ± 0.06	Na	7.93 ± 0.06	7.77 ± 0.06	8.47 ± 0.06	8.06 ± 0.06	8.79 ± 0.06
Mandenka (%)	7.81 ± 0.06	7.82 ± 0.06	7.84 ± 0.06	7.82 ± 0.06	7.83 ± 0.06	7.80 ± 0.06	7.93 ± 0.06	7.93 ± 0.06	Na	7.88 ± 0.06	8.39 ± 0.06	7.82 ± 0.06	8.78 ± 0.06
Dinka (%)	7.73 ± 0.06	7.76 ± 0.06	7.70 ± 0.06	7.64 ± 0.06	7.74 ± 0.06	7.68 ± 0.06	7.82 ± 0.06	7.77 ± 0.06	7.88 ± 0.06	Na	8.30 ± 0.06	7.94 ± 0.06	8.73 ± 0.06
Mbuti (%)	8.42 ± 0.06	8.39 ± 0.06	8.39 ± 0.06	8.35 ± 0.06	8.43 ± 0.06	8.40 ± 0.06	8.47 ± 0.06	8.47 ± 0.06	8.39 ± 0.06	8.30 ± 0.06	Na	8.49 ± 0.06	8.79 ± 0.06
Yoruba (%)	7.94 ± 0.06	7.92 ± 0.06	7.95 ± 0.06	7.91 ± 0.06	7.92 ± 0.06	7.91 ± 0.06	8.05 ± 0.06	8.06 ± 0.06	7.82 ± 0.06	7.94 ± 0.06	8.49 ± 0.06	Na	8.81 ± 0.06
San (%)	8.69 ± 0.06	8.69 ± 0.06	8.66 ± 0.06	8.65 ± 0.06	8.68 ± 0.06	8.70 ± 0.06	8.81 ± 0.06	8.79 ± 0.06	8.78 ± 0.06	8.73 ± 0.06	8.79 ± 0.06	8.81 ± 0.06	Na

doi:10.1038/nature26151

Table S7.5 Percentage of sites that match the *Vindija 33.19* or the *Altai* Neandertal state for the late Neandertal genomes and four samples from the Vindija cave that were previously reported[7,13]. Only sites at which *Vindija 33.19* and the *Altai* Neandertal are both homozygous and have a different state were considered in this analysis.

Specimen	Vindija 33.19 (%)	Altai Neandertal (%)	Other (%)
Vindija 33.15	99.7	0.3	0.0
Vindija 33.16	82.5	16.8	0.7
Vindija 33.25	81.4	18.1	0.5
Vindija 33.26	82.2	17.2	0.7
Vindija 87	97.1	2.2	0.7
Vindija 87 deaminated	97.7	1.6	0.7
Les Cottés Z4-1514	69.8	29.4	0.7
Les Cottés Z4-1514 deaminated	70.7	28.6	0.6
Goyet Q56-1	75.6	24.1	0.2
Goyet Q56-1 deaminated	75.3	24.3	0.4
Mezmaiskaya 2	69.3	29.7	1.0
Mezmaiskaya 2 deaminated	68.2	30.9	0.9
Spy 94a	76.0	23.4	0.7
Spy 94a deaminated	75.4	24.0	0.6

Supplementary Information 8

Split times of Neandertals and the neighbour-joining tree of nuclear genomes

Split times of late Neandertals and the high coverage genomes

In order to estimate the split times between the low coverage samples and the *Altai*[1] and *Vindija 33.19*[2] Neandertals, we conditioned on the heterozygous sites observed in one of the two high coverage individuals in turn, and then computed the fraction of sites that show the same derived allele in randomly sampled fragments of each of the low coverage individuals. The advantage of this statistic, called F(A|B)[1,3,4], is that its expected value only depends on the demography of the B-lineage on which the heterozygous sites are determined and the demography of the A-lineage does not affect the result. Hence, we can rely on the demographies estimated for the high-coverage genomes of the *Altai*[1] and *Vindija 33.19*[2] Neandertals. The calibration of the split times are sensitive to the mutation rate and generation times used, with the substantial source of uncertainty being the human mutation rate. In the form of an assumed average sequence divergence of humans and chimpanzees of 13 million years, we have used the mutation rate of 0.5×10^{-9} base pairs per year estimated from the comparison of parent-offspring trios[5,6]. We therefore caution that while the split times presented are consistent relative to one another, the absolute estimates are approximate because of the uncertainty in the models of population history used. Nevertheless, all of the split times that we obtained are consistent with the split time estimates using both *Altai* Neandertal and *Vindija 33.19* genomes.

We performed these analyses twice, once using all randomly sampled fragments, and once randomly sampling from deaminated fragments only (Extended Data Table 4), restricting the analyses to transversions in both cases to avoid recurrent mutations and mitigate the effect of deamination. The results were similar bewteen both sets of fragments for all individuals. However, for each individual, except *Goyet Q56-1*, we observe a small but consistent reduction in the F(A|B) values, resulting in a slightly older split times when only deaminated fragments were considered (Extended Data Table 4). This small reduction in F(A|B) for deaminated fragments cannot be explained by a difference in contamination between the two sets of fragments, since the split times from an African individual showed the same reduction. In contrast *Goyet Q56-1* showed a small increase in F(A|B) when considering deaminated fragments in comparison to the *Altai* and *Vindija 33.19* Neandertals, resulting in more recent split times. A reduction in contamination in the deaminated set of

fragments is a plausible explanation for this observation, since the African individual shows an opposite effect.

The F(A|B) values show, that all late Neandertals individuals are closer to *Vindija 33.19* than to *Altai* Neandertal. In particular, they all showed a split time to the *Altai* Neandertal that was about ~150,000 years ago, while only less than 70,000 years ago from *Vindija 33.19* (Extended Data Table 4). Remarkably, all samples except *Goyet Q56-1*, separated from the ancestor with the *Altai* Neandertal approximatively ~35,000 years before the birth of the *Altai* Neandertal, despite the inaccuracy of the F(A|B) for very short split times. This observation is consistent with the *Altai* Neandertal falling basal to all late Neandertals, and with an independent Western clade which separated early from the *Altai* Neandertal population ~150,000 years ago. The slightly different behavior of *Goyet Q56-1*, in conjuction with the observation that deaminated fragments provided a substantially different pattern, might therefore indicate that the signal provided by this individual was confounded by potential technical factors. A possible explanation for this difference could be the composition of the *Goyet Q56-1* data, which comes from libraries that were partially UDG/endonuclease VIII treated and partially not treated. This composition matches the high-coverage samples better than other samples because they were also generated to different extents from the libraries that were treated with UDG/endonuclease VIII, while the other late Neandertal libraries were not treated. In order to test this hypothesis we calculated the fraction of pairwise differences between the late Neandertal samples and either UDG or non-UDG treated randomly sampled fragments used for the high coverage *Vindija 33.19* genome (Table S8.1). *Goyet Q56-1* showed the smallest fraction of differences when compared to Vindija's UDG-treated fragments, and the highest ratio of pairwise differences when comparing non-UDG treated and UDG-treated fragments (p-value=0.0014). Hence, the small departure of *Goyet Q56-1* from the F(A|B) values of the other late Neandertals was likely an effect of the mixed composition of its DNA fragments.

When investigating the split times to the high coverage *Vindija 33.19* Neandertal, all samples showed very high F(A|B) values, indicating extremely recent split times. In particular, our estimates indicate that most samples separated from this lineage about ~15,000 years before the life of the *Vindija* individual. Due to the uncertainty in demography estimates and due to quality differences between samples, the F(A|B) statistic is not expected to give precise estimates for split times much more recent than 100,000 years. To accommodate some uncertainty in the estimates, we considered split times as identical when they were less than 5,000 years apart.

We also noticed that F(A|B) values support the hypothesis that *Vindija 33.19* and *Vindija 87* specimens belonged to the same individual. In fact, coalescent simulations for the calibration of the F(A|B) statistic indicated that two individuals sampled from the same population would show an F(A|B) value of 41%. Hence, a value of 47%, as observed when considering the randomly sampled fragments from the *Vindija 87* bone, would be slightly too high even for two siblings or a parent-offspring relationship, that would give an F(A|B) of 45.5%.

We also investigated the split times from the *Denisovan* individual and present-day humans (Extended Data Table 4). In both cases the estimates were consistent with previous analyses using high coverage ancient individuals, with dates for the two splits that are about ~400,000 and ~530,000 years ago, respectively.

A neighbour-joining tree of the archaic and present-day humans

We used the low coverage genomes of *Les Cottés Z4-1514, Goyet Q56-1, Spy 94a, Mezmaiskaya 2* and *Mezmaiskaya 1*, as well as the high coverage genomes of *Vindija 33.19,* the *Altai* Neandertal*, Denisova* and 12 present-day humans from Prüfer *et al.*[1] for constructing a neighbour-joining tree. We counted the total number of transversions between all pairs of individuals and the common ancestor of humans and chimpanzees. We observed that the low coverage Neandertals had a higher proportion of transversions than present-day humans and the high coverage archaics (Table S8.2), presumably due to more sequencing errors in low coverage genomes. We used this excess of transversions in each of the individuals to estimate individual error rates[1,7]. In order to determine what the error rate would be without any sequencing errors we used the modern human with the lowest error rate as a baseline and treated the excess of transversions in the low coverage Neandertals as errors. We then subtracted these errors from all the comparisons between individuals, *i.e.* corrected the individuals for errors.

We constructed a neighbour joining tree[8] based on the corrected pairwise number of transversions in windows of 5 Mb across all autosomes between all individuals and repeated the procedure for 1000 bootstrap replicates over the 5Mb windows to estimate the support for each branch (Fig. 3C). The tree was constructed as implemented in the R-package *phangorn*[9] and visualized with *FigTree* (version: v1.4.2) (http://tree.bio.ed.ac.uk/software/figtree/).

References for SI8:

1 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

2 Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, doi:10.1126/science.aao1887 (2017).

3 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

4 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

5 Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American Journal of Physical Anthropology* **128**, 415-423, doi:10.1002/ajpa.20188 (2005).

6 Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471-475, doi:10.1038/nature11396 (2012).

7 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

8 Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425, doi:10.1093/oxfordjournals.molbev.a040454 (1987).

9 Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593, doi:10.1093/bioinformatics/btq706 (2011).

doi:10.1038/nature26151

Table 8.1 Pairwise differences of the late Neandertal samples from *Vindija 33.19*, either restricting *Vindija 33.19 fragments* to UDG-treated or non-UDG treated libraries. PD_{noUDG}/ PD_{UDG} indicates the ratio between the two.

Individual	Pairwise-Differences UDG	PD_{noUDG}/ PD_{UDG}
Goyet Q56-1	0.009	2.12
Spy 94a	0.015	1.62
Vindija 87	0.015	1.47
Les Cottés Z4-1514	0.17	1.56
Mezmaiskaya 2	0.24	1.38

Table S8.2 Percentage of transversions between the human-chimpanzee ancestor and archaic and present-day human individuals.

Individual	% of transversions
Altai Neandertal	0.3903
Vindija 33.19	0.3911
Denisovan	0.3924
Les Cottés Z4-1514	0.523
Goyet Q56-1	0.4314
Mezmaiskaya 2	0.4539
Vindija 87	0.4291
Spy 94a	0.5363
Mezmaiskaya 1	0.5155
French	0.3922
Sardinian	0.3902
Han	0.3931
Dai	0.3927
Karitiana	0.3924
Mixe	0.3926
Australian	0.3923
Papuan	0.3923
Dinka	0.3927
Mbuti	0.3924
Yoruba	0.3924
Mandenka	0.3922
San	0.3926

Supplementary Information 9

Neandertal Population Relationships

Overview

The availability of the five late Neandertal genomes and of the previously published Neandertal genomes of the *Altai* Neandertal[1], *Vindija 33.19*[2], and *Mezmaiskaya 1*[2] enabled us to investigate the population relationship among these individuals. Applying *D-statistics*[3-5] on randomly sampled fragments, we were able to show that all of the Neandertals were significantly closer to *Vindija 33.19* than to the *Altai* Neandertal. This shows that the Altai Neandertal is not only likely oldest Neandertal individual among those compared here, but it also falls basal to other Neandertal individuals.

Having multiple Neandertals sampled from geographically close locations and from the same time period allowed us to investigate whether geography and time are predictors of the genetic relationship between Neandertals. The three Neandertals from Western Europe (*Goyet Q56-1*, *Spy 94a*, and *Les Cottés Z4-1514*) were genetically closer to the other Neandertals from the same geographical region and time, and showed most distant relationship to the Mezmaiskaya Neandertals. While the two Neandertals from Mezmaiskaya cave were geographically close, their inferred dates are at least ~25,000 years apart. *D-statistics* analysis indicated that the *Mezmaiskaya 2* Neandertal was genetically closer to all other late Neandertals than to the older *Mezmaiskaya 1* Neandertal, suggesting a population turnover in Mezmaiskaya cave.

Dataset for inferring the relationships between Neandertals

In order to infer the relationships between Neandertals we used *D-statistics*[3-5], which tests for significant differences in the sharing of derived alleles between individuals, and can be used to detect past admixture events. We analysed the low coverage Neandertal genomes of *Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2* and *Spy 94a* with the the low coverage genome of the *Mezmaiskaya 1* Neandertal[2] and the high coverage genomes of the *Altai* [1] and *Vindija 33.19*[2] Neandertals. We excluded *Vindija 87* from the analyses, as we concluded that it belonged to the same individual as *Vindija 33.19* (see Supplementary Information 7).

For the high coverage genomes of the *Altai* and *Vindija 33.19* Neandertals we used the *snpAD* genotype calls that passed the GC-corrected coverage filter and did not fall in the

tandem repeat regions[2]. For the late Neandertals and *Mezmaiskaya 1*, we randomly sampled an allele from fragments longer than 35 base pairs that had a mapping quality of at least 25 and that were within the highly mappable regions of the genome (MQ≥25, BQ≥30, Map35_100% of Prüfer *et al.*[1]; see Supplementary Information 6). To diminish the impact of present-day human DNA contamination and enrich for the endogenous fragments[6], we further selected the fragments that showed C-to-T substitutions relative to the human reference genome at the first three and/or last three positions, *i.e.* putatively deaminated fragments. We used the whole genome alignments of the chimpanzee, orangutan and rhesus macaque to the human reference genome[7-9] to infer the ancestral states for the analyses of *D(Altai Neandertal, Vindija 33.19; Neandertal X, outgroup)*. Furthermore, we used the genomes of the Dinka and Mbuti individuals from the Simons Genome Diversity Panel (SGDP)[10] as outgroups for the statistics of the form *D*($Neandertal_1$, $Neandertal_2$; $Neandertal_3$, *outgroup)*.

The standard errors were computed using a Weighted Block Jackknife[5,11] with equally sized blocks of 5 million base pairs (5 Mb) over all autosomes. We further restricted the analyses to bi-allelic sites in the genome covered by at least one low coverage Neandertal and transversion polymorphisms.

All late Neandertals and *Mezmaiskaya 1* were genetically closer to the *Vindija 33.19* than to the *Altai* Neandertal

By comparing the high coverage genomes of the *Altai* and *Vindija 33.19* Neandertals to the late Neandertals and *Mezmaiskaya 1*, we find that all late Neandertals and *Mezmaiskaya 1* shared significantly more derived alleles with *Vindija 33.19* than with the *Altai* Neandertal (Extended Data Table 2; *D-values* between 36% and 56%, with Z-scores between -36.63 and -68.97). The signal was consistent independent of whether Dinka or Mbuti were used as outgroups.

We investigated further, if the observation holds true when replacing *Vindija 33.19* with any other late Neandertal genome in the statistics. This comparison allows us to test, whether all late Neandertals were reciprocally genetically closer to each other than to the *Altai* Neandertal. However, these statistics involve genomes of different sequence qualities on opposite sides of the statistics that can cause spurious correlations in the *D-statistics*. To minimize the effect of this issue, we applied the same processing as for the late Neandertals to the high coverage *Altai* and *Vindija 33.19* Neandertals and sampled fragments at random from BAM files, effectively lowering the qualities of the high coverage genomes to similar levels as those of other individuals (Supplementary Information 3 and Supplementary Information

6). For *Vindija 33.19* we used sequencing data from non-UDG treated libraries in order to more closely resemble the sequencing data of the late Neandertals and *Mezmaiskaya 1*[2], which were mostly untreated. The analyses were further restricted to the fragments that showed C-to-T substitutions with respect to the human reference in the first and/or last three positions. Only UDG-treated libraries were sequenced for the *Altai* Neandertal and we retained fragments that showed C-to-T substitutions at the first and/or last two positions.

All of the pairwise combinations of the late Neandertals, *Vindija 33.19* and *Mezmaiskaya 1* showed the same *D-statistics* pattern in their relationship to the *Altai* Neandertal with the reprocessed high-coverage genomes. Interestingly, the results did not change when genotypes were used for the high coverage archaics (Table S9.1). All of the late Neandertals, including *Mezmaiskaya 1*, shared significantly more derived alleles with each other than with the *Altai* Neandertal (Z-score between -15.63 and -47.46). This observation suggests that the population of the *Altai* Neandertal separated first from populations of all late Neandertals and *Mezmaiskaya 1*. This results is in concordance with both the Neandertal divergence estimates in Supplementary Information 7, and the split time estimates in Supplementary Information 8.

Relationship of the late Neandertals to the high coverage *Vindija 33.19*

We further investigated, whether there was a difference in the relationship of the late Neandertals to the *Vindija 33.19* by computing the statistics in the form of $D(Neandertal_1, Neandertal_2, Vindija\ 33.19, Mbuti/Dinka)$. We find that *Goyet Q56-1* and *Spy 94a* shared more derived alleles with *Vindija 33.19* than any other Neandertal, whereas there was no significant difference between them, followed by the *Les Cottés Z4-1514* and *Mezmaiskaya 2*. However, all of the late Neandertals shared more derived alleles with *Vindija 33.19* than *Mezmaiskaya 1*, suggesting that all late Neandertals, including *Vindija 33.19*, form a clade compared to the *Mezmaiskaya 1* individual. (Table S9.2).

Relationships between *Goyet Q56-1, Spy 94a* and *Les Cottés Z4-1514*

We had two geographical clusters in our data set: Neandertals from Belgium and France (*Goyet Q56-1, Spy 94a* and *Les Cottés Z4-1514*) and Neandertals from the Mezmaiskaya cave in Russia. *Goyet Q56-1* and *Spy 94a* originate from two neighbouring archaeological sites that are part of a large cave system located in the Mosan Basin in Belgium[12,13] and radiocarbon dating placed them close in time (*Goyet Q56-1* was directly dated to ~43,000-42,080 cal BP[13], whereas *Spy 94a* was indirectly dated to 39,150-37,880 cal BP[12]). This proximity in space and

time is also reflected by their genetic relationship. When testing the *D-statistics* of *D(Neandertal*$_1$*, Neandertal*$_2$*, Goyet Q56-1, Mbuti)* and *D(Neandertal*$_1$*, Neandertal*$_2$*, Spy 94a, Mbuti), Goyet Q56-1* and *Spy 94a* shared significantly more derived alleles with each other than with any other Neandertal (Tables S9.3 and S9.4). Furthermore, they were equally close to *Les Cottés Z4-1514* and *Vindija 33.19*, but shared significantly less derived alleles with the Mezmaiskaya Neandertals, which is in concordance with the geographical distance between the sampling sites. The same was the case for *Les Cottés Z4-1514*, which is the most western specimen and which was directly radiocarbon dated to ~43,740-42,720 cal BP (see Supplementary Information 1 on the radiocarbon dating of *Les Cottés Z4-1514*). *Les Cottés Z4-1514* shared significantly more derived alleles with its geographically closest neighbours, *Goyet Q56-1* and *Spy 94a*, than with any other Neandertal, without being significantly closer to either of them (Table S9.5). Again, *Les Cottés Z4-1514* shared significantly less derived alleles with the Mezmaiskaya Neandertals than with *Vindija 33.19*.

Relationship of the late Neandertals to *Mezmaiskaya 1*

A second geographical cluster is the Mezmaiskaya cave in the Northern Caucasus, which yielded two distinct Neandertal specimens: *Mezmaiskaya 1*, a nearly complete Neandertal neonate indirectly dated to >60,000-70,000 BP, and *Mezmaiskaya 2*, a second Neandertal infant represented by skull fragments that were directly radiocarbon dated to ~44,430-42,640 cal BP[14,15]. Despite the same geographical location, the inferred dates of the two Neandertal specimens are at least 25,000 years apart.

Consistent with the nuclear and mitochondrial relationships (Supplementary Information 5), the late Neandertal *Mezmaiskaya 2* shared significantly more derived alleles with all other late Neandertals, including *Vindija 33.19*, than with the older *Mezmaiskaya 1* Neandertal (Table S9.6). Interestingly, the late Neandertal that was genetically most distant to *Mezmaiskaya 2* was *Les Cottés Z4-1514*, which is also separated by the largest geographic distance. There was no significant difference between all other late Neandertals in their genetic proximity to *Mezmaiskaya 2* (Table S9.6).

When comparing all these late Neandertals to the older *Mezmaiskaya 1* individual, we observed that they all shared significantly more derived alleles with each other than with *Mezmaiskaya 1* Neandertal (Table S9.7). This result indicates at least a partial population turnover in Mezmaiskaya cave.

Relationships between *Vindija* Neandertals

The third geographical cluster is Vindija cave in Croatia. We compared previously obtained three low coverage Neandertal genomes of *Vindija 33.16*, *Vindija 33.25* and *Vindija 33.26* to the high coverage archaics, late Neandertals and *Mezmaiskaya 1* using the same framework. All of the Vindija Neandertals for which nuclear genome sequences have been generated come from a relatively narrow time range. As for the late Neandertals and *Mezmaiskaya 1*, the low coverage genomes of *Vindija 33.16*, *Vindija 33.25* and *Vindija 33.26* shared significantly more derived alleles with the high coverage *Vindija 33.19* than with the *Altai* Neandertal. Furthermore, all of the Neandertals from Vindija cave shared significantly more derived alleles with each other than with any other Neandertal sequenced to date ($-2.2 \leq Z \leq -14.5$; Tables S9.8-S9.10), concordant with the high relatedness within the same geographical area and same time period, unlike the Neandertal individuals from Mezmaiskaya cave that come from the same location, but were separated by at least 25,000 years.

References for SI9:

1. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).
2. Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science, doi:10.1126/science.aao1887 (2017).
3. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).
4. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).
5. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, doi:10.1534/genetics.112.145037 (2012).
6. Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).
7. Chimpanzee, S. & Analysis, C. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69-87, doi:10.1038/nature04072 (2005).
8. Locke, D. P. *et al.* Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529-533, doi:10.1038/nature09687 (2011).
9. Gibbs, R. A. *et al.* Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. *Science* **316**, 222-234 (2007).

10 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206, doi:10.1038/nature18964 (2016).

11 Busing, F. M. T. A., Meijer, E. & Van Der Leeden, R. Delete-m jackknife for unequal m. *Statistics and Computing* **9**, 3-8, doi:Doi 10.1023/A:1008800423698 (1999).

12 Semal, P. *et al.* New data on the late Neandertals: direct dating of the Belgian Spy fossils. *Am J Phys Anthropol* **138**, 421-428, doi:10.1002/ajpa.20954 (2009).

13 Rougier, H. *et al.* Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci Rep* **6**, 29005, doi:10.1038/srep29005 (2016).

14 Pinhasi, R., Higham, T. F., Golovanova, L. V. & Doronichev, V. B. Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. *Proc Natl Acad Sci U S A* **108**, 8611-8616, doi:10.1073/pnas.1018938108 (2011).

15 Skinner, A. R. *et al.* ESR dating at Mezmaiskaya Cave, Russia. *Appl Radiat Isot* **62**, 219-224, doi:10.1016/j.apradiso.2004.08.008 (2005).

Table S9.1 Z-scores of the *D(Altai, Neandertal$_1$, Neandertal$_2$, Mbuti)* using genotypes of the *Altai* and *Vindija 33.19* Neandertals in the top pane, and random reads in the bottom pane. Only putatively deaminated fragments reported in the Table S3.2 were used for random read sampling of Neandertal individuals, and the analyses were restricted to transversion polymorphisms. Blue: Z-score < -2; Yellow: Z-score > 2

D (Altai genotypes, X, Y, Mbuti)						
X/Y	Goyet Q56-1	Les Cottés Z4-1514	Mezmaiskaya 1	Mezmaiskaya 2	Spy 94a	Vindija33.19 genotypes
Goyet Q56-1	NA	-38.45	-20.48	-28.86	-35.46	-41.31
Les Cottés Z4-1514	-36.76	NA	-25.55	-29.76	-37.00	-38.65
Mezmaiskaya1	-19.68	-26.24	NA	-19.02	-20.58	-25.59
Mezmaiskaya2	-24.94	-29.08	-20.14	NA	-24.82	-35.76
Spy 94a	-32.32	-39.82	-21.13	-24.95	NA	-39.04
Vindija33.19 genotypes	-57.22	-54.94	-33.59	-44.54	-52.79	NA
D (Altai random reads, X, Y, Mbuti)						
X/Y	Goyet Q56-1	Les Cottés Z4-1514	Mezmaiskaya 1	Mezmaiskaya 2	Spy 94a	Vindija33.19 random reads
Goyet Q56-1	NA	-32.43	-17.54	-25.14	-29.88	-37.90
Les Cottés Z4-1514	-32.79	NA	-21.68	-27.57	-32.39	-37.51
Mezmaiskaya1	-17.68	-21.80	NA	-16.98	-16.30	-23.38
Mezmaiskaya2	-21.92	-25.40	-15.63	NA	-20.95	-33.22
Spy 94a	-28.00	-35.04	-16.84	-22.32	NA	-35.18
Vindija33.19 random reads	-47.46	-47.43	-27.65	-39.79	-43.21	NA

doi:10.1038/nature26151

Table S9.2 Z-scores of *D(Neandertal$_1$, Neandertal$_2$, Vindija 33.19, Mbuti)* using the genotypes of the *Altai* and *Vindija 33.19* Neandertals in the top pane, and random reads of the *Altai* and *Vindija 33.19* Neandertals in the bottom pane. Only putatively deaminated fragments reported in the Table S3.2 were used for random read sampling of Neandertal individuals, and the analyses were restricted to transversion polymorphisms. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Vindija33.19, Mbuti)						
X/Y	Altai	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a
Altai	NA	-41.31	-38.65	-25.59	-35.76	-39.04
Goyet Q56-1	41.31	NA	9.30	15.25	9.62	2.10
Les Cottés	38.65	-9.30	NA	11.82	0.44	-7.19
Mezmaiskaya1	25.59	-15.25	-11.82	NA	-10.04	-13.05
Mezmaiskaya2	35.76	-9.62	-0.44	10.04	NA	-6.42
Spy 94a	39.04	-2.10	7.19	13.05	6.42	NA
D(X, Y, Vindija33.19, Mbuti)						
X/Y	Altai	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a
Altai	NA	-35.49	-36.78	-22.85	-33.12	-34.80
Goyet Q56-1	35.49	NA	8.24	13.13	8.89	1.40
Les Cottés	36.78	-8.24	NA	11.05	0.84	-6.92
Mezmaiskaya1	22.85	-13.13	-11.05	NA	-8.84	-11.77
Mezmaiskaya2	33.12	-8.89	-0.84	8.84	NA	-6.62
Spy 94a	34.80	-1.40	6.92	11.77	6.62	NA

Table S9.3 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Goyet Q56-1, Mbuti*) using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Goyet Q56-1, Mbuti)						
X/Y	Altai	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	-31.15	-16.63	-20.94	-27.65	-46.33
Les Cottés	31.15	NA	9.40	5.52	-4.23	-2.59
Mezmaiskaya1	16.63	-9.40	NA	-5.60	-7.85	-13.18
Mezmaiskaya2	20.94	-5.52	5.60	NA	-7.50	-9.89
Spy 94a	27.65	4.23	7.85	7.50	NA	3.65
Vindija33.19	46.33	2.59	13.18	9.89	-3.65	NA

doi:10.1038/nature26151

Table S9.4 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Spy 94a, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Spy 94a, Mbuti)						
X/Y	Altai	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Vindija33.19
Altai	NA	-29.24	-31.74	-15.92	-19.71	-42.72
Goyet Q56-1	29.24	NA	4.35	8.47	7.54	4.97
Les Cottés	31.74	-4.35	NA	8.42	5.26	-1.74
Mezmaiskaya1	15.92	-8.47	-8.42	NA	-2.00	-9.89
Mezmaiskaya2	19.71	-7.54	-5.26	2.00	NA	-7.72
Vindija33.19	42.72	-4.97	1.74	9.89	7.72	NA

Table S9.5 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Les Cottés Z4-1514, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, LesCottés, Mbuti)						
X/Y	Altai	Goyet Q56-1	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	-32.38	-21.80	-25.72	-33.83	-49.36
Goyet Q56-1	32.38	NA	9.15	8.83	0.35	5.90
Mezmaiskaya1	21.80	-9.15	NA	-3.97	-9.76	-11.42
Mezmaiskaya2	25.72	-8.83	3.97	NA	-6.92	-7.43
Spy 94a	33.83	-0.35	9.76	6.92	NA	5.48
Vindija33.19	49.36	-5.90	11.42	7.43	-5.48	NA

Table S9.6 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Mezmaiskaya 2, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Mezmaiskaya2, Mbuti)						
X/Y	Altai	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Spy 94a	Vindija33.19
Altai	NA	-25.14	-27.57	-16.98	-22.32	-39.79
Goyet Q56-1	25.14	NA	3.16	6.21	0.36	-1.06
Les Cottés	27.57	-3.16	NA	5.08	-2.14	-6.68
Mezmaiskaya1	16.98	-6.21	-5.08	NA	-4.79	-12.15
Spy 94a	22.32	-0.36	2.14	4.79	NA	-0.97
Vindija33.19	39.79	1.06	6.68	12.15	0.97	NA

Table S9.7 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Mezmaiskaya1, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Mezmaiskaya1, Mbuti)						
X/Y	Altai	Goyet Q56-1	Les Cottés	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	-16.50	-21.87	-16.11	-17.09	-27.78
Goyet Q56-1	16.50	NA	0.03	0.26	0.55	0.25
Les Cottés	21.87	-0.03	NA	0.69	-0.76	0.71
Mezmaiskaya2	16.11	-0.26	-0.69	NA	-1.99	-1.98
Spy 94a	17.09	-0.55	0.76	1.99	NA	1.64
Vindija33.19	27.78	-0.25	-0.71	1.98	-1.64	NA

Table S9.8 *D(Mezmaiskaya1, Mezmaiskaya2, Neandertal, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D(Mezmaiskaya1, Mezmaiskaya2, Neandertal, Mbuti)				Deaminated fragments		
W	X	Y	Z	D-value	Z-score	# of sites
Mezmaiskaya1	Mezmaiskaya2	Altai Neandertal	Mbuti	3.18	1.47	227,164
Mezmaiskaya1	Mezmaiskaya2	Vindija33.19	Mbuti	-16.91	-9.56	249,823
Mezmaiskaya1	Mezmaiskaya2	Les Cottés	Mbuti	-8.97	-4.22	162,386
Mezmaiskaya1	Mezmaiskaya2	Goyet Q56-1	Mbuti	-18.99	-5.98	64,665
Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Mbuti	-7.77	-2.13	63,169

Table S9.9 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Vindija 33.16, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Vindija 33.16, Mbuti)									
X/Y	Altai	Goyet Q56-1	Les_Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19	Vindija33.25	Vindija33.26
Altai	NA	-20.95	-18.28	-9.07	-18.06	-14.10	-45.16	-20.53	-19.31
Goyet Q56-1	20.95	NA	5.42	7.27	4.09	1.81	-2.73	-3.25	-2.47
Les_Cottés	18.28	-5.42	NA	2.49	-1.47	-2.22	-13.48	-5.73	-7.17
Mezmaiskaya1	9.07	-7.27	-2.49	NA	-3.73	-2.84	-14.55	-7.72	-6.41
Mezmaiskaya2	18.06	-4.09	1.47	3.73	NA	-2.74	-10.90	-4.95	-7.34
Spy 94a	14.10	-1.81	2.22	2.84	2.74	NA	-4.91	-2.16	-2.86
Vindija33.19	45.16	2.73	13.48	14.55	10.90	4.91	NA	-3.53	-3.07
Vindija33.25	20.53	3.25	5.73	7.72	4.95	2.16	3.53	NA	0.05
Vindija33.26	19.31	2.47	7.17	6.41	7.34	2.86	3.07	-0.05	NA

Table S9.10 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Vindija 33.26, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Vindija 33.26, Mbuti)									
X/Y	Altai	Goyet Q56-1	Les_Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.16	Vindija33.19	Vindija33.25
Altai	NA	-19.91	-19.61	-11.35	-17.51	-14.36	-19.62	-39.04	-16.49
Goyet Q56-1	19.91	NA	2.78	4.73	3.72	1.70	-2.38	-2.21	-2.91
Les_Cottés	19.61	-2.78	NA	2.35	-0.07	-1.77	-6.21	-10.43	-5.52
Mezmaiskaya1	11.35	-4.73	-2.35	NA	-2.64	-2.93	-6.06	-11.99	-7.14
Mezmaiskaya2	17.51	-3.72	0.07	2.64	NA	-1.56	-5.34	-7.64	-3.56
Spy 94a	14.36	-1.70	1.77	2.93	1.56	NA	-4.35	-4.13	-3.74
Vindija33.16	19.62	2.38	6.21	6.06	5.34	4.35	NA	3.86	0.07
Vindija33.19	39.04	2.21	10.43	11.99	7.64	4.13	-3.86	NA	-1.74
Vindija33.25	16.49	2.91	5.52	7.14	3.56	3.74	-0.07	1.74	NA

Table S9.11 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, Vindija 33.25, Mbuti)* using random reads for all individuals and restricting to transversions. Only putatively deaminated fragments reported in the Table S3.2 were used for Neandertal individuals. Blue: Z-score < -2; Yellow: Z-score > 2

D (X, Y, Vindija 33.25, Mbuti)									
X/Y	Altai	Goyet Q56-1	Les_Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.16	Vindija33.19	Vindija33.26
Altai	NA	-19.48	-19.06	-9.36	-15.85	-16.60	-18.51	-42.06	-15.34
Goyet Q56-1	19.48	NA	2.17	5.44	2.79	-1.23	-2.90	-3.88	-2.80
Les_Cottés	19.06	-2.17	NA	4.14	-0.44	-3.29	-4.76	-12.11	-4.72
Mezmaiskaya1	9.36	-5.44	-4.14	NA	-3.90	-3.57	-5.64	-14.30	-4.45
Mezmaiskaya2	15.85	-2.79	0.44	3.90	NA	-3.15	-5.32	-10.53	-3.05
Spy 94a	16.60	1.23	3.29	3.57	3.15	NA	-1.54	-2.69	-2.62
Vindija33.16	18.51	2.90	4.76	5.64	5.32	1.54	NA	3.70	0.11
Vindija33.19	42.06	3.88	12.11	14.30	10.53	2.69	-3.70	NA	-0.52
Vindija33.26	15.34	2.80	4.72	4.45	3.05	2.62	-0.11	0.52	NA

Supplementary Information 10

Proximity to the introgressing Neandertals in present-day and ancient modern humans

Overview

The availability of multiple Neandertal genomes allowed us to investigate whether one of these Neandertals is significantly closer to the Neandertal population that contributed one to three percent of the genomes of present-day humans living outside of Sub-Saharan Africa[1,2].

We find that the Neandertals who introgressed into the ancestors of present-day non-Africans are genetically closer to all late Neandertals, including *Vindija 33.19*[3], and the older *Mezmaiskaya 1*[3] Neandertal, than they are to the *Altai*[2] Neandertal. The genetic affinity between late Neandertals and present-day non-Africans can be explained by gene flow into the ancestors of all present-day non-Africans from a Neandertal population that was equidistant to all late Neandertals and already separated from the *Altai* Neandertal population. This observation is in agreement with the *Altai* Neandertal falling basal to all other Neandertals in our comparisons (Supplementary Information 7 – Supplementary Information 9).

Interestingly, *Ust'-Ishim*, the 45,000-year-old modern human from Siberia[4], which overlaps in time with late Neandertals, has the same relationship to late Neandertals as present-day non-Africans. This would suggest that even though *Ust'-Ishim* is not on the direct lineage to present-day human populations[4], he received gene flow from the same Neandertal population as present-day humans.

Dataset for inferring the relationship to the introgressed Neandertals in present-day and ancient modern humans

In order to infer the relationship between the introgressing Neandertals in present-day humans and the Neandertals whose genomes have been sequenced, we co-analysed the low coverage late Neandertal genomes presented in this study with the high coverage genomes of the *Altai*[2] and *Vindija 33.19* Neandertals[3], the low coverage genome of the *Mezmaiskaya 1* Neandertal[3], the high coverage genomes of the *Denisovan* individual[5] and 263 present-day humans of the Simons Genome Diversity Panel (SGDP)[6]. We used the new genotype calls (http://cdna.eva.mpg.de/neandertal/Vindija/VCF/) of the *Altai* Neandertal, *Denisova* and

Vindija 33.19 that passed the GC corrected coverage filter and did not fall in tandem repeat regions[3]. For the low coverage Neandertals we selected a base from a random fragment that had mapping quality of at least 25 and was within the highly alignable regions of the genome (Map35_100% of Prüfer *et al.*[2]). The analyses were performed separately for all fragments and putatively deaminated fragments. In order to investigate whether there was a difference among Neandertals in their proximity to the introgressed Neandertal detected in ancient modern humans, we used the high quality genotypes of *Ust'-Ishim*, a ~45,000-year-old modern human from Siberia[4]; *Loschbour*, a ~8,000-year-old hunter-gatherer from Luxembourg[7]; and *LBK*, a ~7,000-year-old farmer from Stuttgart[7].

As the majority of the low coverage Neandertal data comes from non-UDG treated libraries, elevated C-to-T substitution frequencies were observed throughout the read and not only at the alignment start and end positions (Supplementary Information 3). Therefore, we restricted all of our analyses to transversion polymorphisms[1]. We used *heffalumps* (https://bitbucket.org/ustenzel/heffalump) to extract variable positions across all genomes into an input format for *AdmixTools* (version 4.1)[8]. We further restricted the analyses to bi-allelic sites in the genome covered by at least one low coverage Neandertal and transversion polymorphisms.

We used *D-statistics* to infer the relationships between individuals[1,8,9] and computed standard errors using a Weighted Block Jackknife[8,10] over all autosomes with equally sized blocks of 5 million base pairs (5 Mb).

Comparison of present-day human populations to Neandertals

We first investigated the degree of Neandertal allele sharing among present-day human populations that we grouped according to their geographical origin[6], by calculating *D (Human*$_1$*, Human*$_2$*, Neandertal, Chimpanzee)*. We used the genome of the chimpanzee (*panTro2*) for inferring the ancestral state at any given site. All non-African populations shared significantly more derived alleles with Neandertals than African populations, as observed in previous studies[1,2], and irrespective of the Neandertal genome used in the analysis (Z-score > 11.08-20.53 for deaminated fragments in the late Neandertals, Tables S10.1-S10.3). We then used the statistic *D (Neandertal*$_1$*, Neandertal*$_2$*, African*$_1$*, African*$_2$*)* to test whether some African populations shared significantly more derived alleles with any of the Neandertals in our dataset. We find that the D is not significantly different from 0 for all Neandertal pairs when restricting the analyses to deaminated fragments (Table S10.4 and Table S10.5), indicating that present-day sub-Saharan Africans were equally close to all

Neandertal individuals analysed to date. This is consistent with the observation in Prüfer *et al.*[2] that sub-Saharan Africans form a clade compared to Neandertals.

When comparing pairs of present-day human populations, we identified a significantly higher proportion of Neandertal ancestry in East Asians than in Western Eurasians and South Asians for all analysed Neandertals. These results were consistent with findings of a higher Neandertal ancestry in East Asians[11-14]. The highest degree of allele sharing with Neandertals was observed in present-day humans from Oceania, likely due to the additional gene flow from Denisovans, a sister group of Neandertals, into these populations[5,9,15,16].

The introgressing Neandertals in present-day humans were genetically closer to all late Neandertals and *Mezmaiskaya 1* than to the *Altai* Neandertal

In order to investigate whether there is a difference among the Neandertals in our dataset in their proximity to the Neandertals who introgressed into modern human populations, we selected two non-African populations per their geographical origin and computed the *D-statistics* of the form *D (*$Neandertal_1$*,* $Neandertal_2$*, non-African, Mbuti)*. In all cases where both Neandertals in this test were either late Neandertals, *Vindija 33.19* or *Mezmaiskaya 1*, no significant differences were observed in the extent of derived alleles sharing with any present-day human population (Fig. 3B, and Tables S10.6-S10.8). In contrast, when one of the test Neandertals was the *Altai* Neandertal, the other Neandertal always shared more derived alleles with the present-day non-African human population, irrespective of their geographical origin (Fig. 3A and tables S10.6-S10.8).

Our findings suggest that the Neandertals that introgressed into the ancestors of present-day non-Africans were genetically closer to all late Neandertals than to the *Altai* Neandertal. A similar pattern emerges for the genomes of *Vindija 33.19* and *Mezmaiskaya 1* compared to the *Altai* Neandertal, indicating that the gene flow into the ancestors of present-day non-Africans occurred from a Neandertal population that was equally distant to all late Neandertals and *Mezmaiskaya 1*, and that formed after their split from the *Altai* Neandertal population around 150,000 years ago. This observation is consistent with our conclusion that all Neandertals analysed here were a clade with respect to the *Altai* Neandertal (Supplementary Information 8 and Supplementary Information 9).

Proximity to the introgressed Neandertals in ancient modern humans

We used the high coverage genomes of *Ust'-Ishim*, a 45,000-year-old modern human from Siberia[4]; *Loschbour*, a ~8,000-year-old hunter-gatherer from Luxembourg[7]; and *LBK*, a

~7,000-year-old farmer from Stuttgart[7] to investigate whether ancient modern humans differ in their relationship to Neandertals from present-day humans. This allows us to test whether there was a change in the Neandertal population which introgressed into modern humans over time. We note that obtaining genome-wide data of multiple Neandertals is necessary to study this question with sufficient resolution.

To test for significant differences among Neandertals in the sharing of derived alleles with ancient modern humans, we computed the *D-statistics* of the form *D (Neandertal*$_1$*, Neandertal*$_2$*, ancient modern human, Mbuti)*. Similarly to the patterns observed when testing present-day human populations, we found no significant differences in this test among late Neandertals, *Vindija 33.19* and *Mezmaiskaya 1*, while all of these shared more derived alleles with ancient modern humans than the *Altai* Neandertal (Table S10.9).

Therefore, even though the *Ust'-Ishim* individual overlapped in time with the late Neandertals presented in this study, it was not significantly closer to any of them, irrespective of their geographical origin. Furthermore, while *Ust'-Ishim* was not a direct ancestor of any present-day human population[4], his genome had a closer affinity to all late Neandertals compared to the *Altai* Neandertal, as was the case for present-day non-Africans. This suggests that the introgression event from Neandertals into modern humans was shared between the ancestors of *Ust'-Ishim* and those of present-day populations, consistent with the predicted age of the gene-flow of 50,000-60,000 years ago. Alternatively, these findings could indicate that the Neandertal population that introgressed into the population from which *Ust'-Ishim* is derived was equally distant to all late Neandertals as the Neandertal population that also introgressed into the ancestors of other ancient and present-day non-Africans. Overall, as the patterns of affinity to the late Neandertals were similar among ancient and present-day humans, we detect no changes in the Neandertal population that introgressed into modern humans over time within limits of our resolution.

References for SI10:

1 Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722, doi:10.1126/science.1188021 (2010).

2 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

3 Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, doi:10.1126/science.aao1887 (2017).

4 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, doi:10.1038/nature13810 (2014).

5 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

6 Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201-206, doi:10.1038/nature18964 (2016).

7 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).

8 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, doi:10.1534/genetics.112.145037 (2012).

9 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).

10 Busing, F. M. T. A., Meijer, E. & Van Der Leeden, R. Delete-m jackknife for unequal m. *Statistics and Computing* **9**, 3-8, doi:Doi 10.1023/A:1008800423698 (1999).

11 Kim, B. Y. & Lohmueller, K. E. Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. *Am J Hum Genet* **96**, 454-461, doi:10.1016/j.ajhg.2014.12.029 (2015).

12 Vernot, B. & Akey, J. M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017-1021, doi:10.1126/science.1245938 (2014).

13 Vernot, B. & Akey, J. M. Complex history of admixture between modern humans and Neandertals. *Am J Hum Genet* **96**, 448-453, doi:10.1016/j.ajhg.2015.01.006 (2015).

14 Wall, J. D. *et al.* Higher levels of neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**, 199-209, doi:10.1534/genetics.112.148213 (2013).

15 Vernot, B. *et al.* Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235-239, doi:10.1126/science.aad9416 (2016).

16 Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Curr Biol* **26**, 1241-1247, doi:10.1016/j.cub.2016.03.037 (2016).

doi:10.1038/nature26151

Table S10.1 Z-score of *D (Human$_1$, Human$_2$, Neandertal, Chimp)* for *Les Cottés Z4-1514, Goyet Q56-1* and *Spy 94a* using deaminated fragments and restricted to transversions. Blue indicates Z-score < -2, while yellow indicates Z-score > 2. CeAsSi – Central Asia and Siberia; SouAs – Southern Asia, WEurAs – Western Eurasia. Only putatively deaminated fragments reported in the Table S3.2 were utilized.

D (X, Y; Les Cottés Z4-1514, Chimp)							
X/Y	Africa	America	CeAsSi	EastAsia	Oceania	SouAs	WEurAs
Africa	NA	-14.54	-17.57	-16.50	-16.41	-18.77	-18.18
America	14.54	NA	-2.00	-1.46	-3.60	0.88	1.27
CeAsSi	17.57	2.00	NA	0.37	-2.86	3.27	3.08
EastAsia	16.50	1.46	-0.37	NA	-3.21	2.70	2.51
Oceania	16.41	3.60	2.86	3.21	NA	4.97	4.66
SouAs	18.77	-0.88	-3.27	-2.70	-4.97	NA	1.05
WEurAs	18.18	-1.27	-3.08	-2.51	-4.66	-1.05	NA
D (X, Y; Goyet_Q56-1, Chimp)							
X/Y	Africa	America	CeAsSi	EastAsia	Oceania	SouAs	WEurAs
Africa	NA	-11.44	-13.95	-12.89	-13.43	-14.75	-14.26
America	11.44	NA	-1.20	-0.91	-3.00	0.86	1.38
CeAsSi	13.95	1.20	NA	0.25	-2.77	2.59	2.82
EastAsia	12.89	0.91	-0.25	NA	-2.98	2.12	2.29
Oceania	13.43	3.00	2.77	2.98	NA	4.39	4.31
SouAs	14.75	-0.86	-2.59	-2.12	-4.39	NA	1.37
WEurAs	14.26	-1.38	-2.82	-2.29	-4.31	-1.37	NA
D (X, Y; Spy 94a, Chimp)							
X/Y	Africa	America	CeAsSi	EastAsia	Oceania	SouAs	WEurAs
Africa	NA	-11.08	-14.00	-13.09	-13.22	-14.26	-14.29
America	11.08	NA	-1.51	-1.35	-2.80	0.52	0.70
CeAsSi	14.00	1.51	NA	0.06	-2.45	2.57	2.31
EastAsia	13.09	1.35	-0.06	NA	-2.53	2.23	1.97
Oceania	13.22	2.80	2.45	2.53	NA	4.06	3.72
SouAs	14.26	-0.52	-2.57	-2.23	-4.06	NA	0.57
WEurAs	14.29	-0.70	-2.31	-1.97	-3.72	-0.57	NA

Table S10.2 Z-score of *D (Human$_1$, Human$_2$, Neandertal, Chimp)* for *Mezmaiskaya 2* and *Mezmaiskaya 1* using deaminated fragments and restricted to transversions. Blue indicates Z-score < -2, while yellow indicates Z-score > 2. CeAsSi – Central Asia and Siberia; SouAs – Southern Asia, WEurAs – Western Eurasia. Only putatively deaminated fragments reported in the Table S3.2 were utilized.

D (X, Y; Mezmaiskaya 2, Chimp)							
X/Y	Africa	America	CeAsSi	EastAsia	Oceania	SouAs	WEurAs
Africa	NA	-15.22	-17.87	-17.20	-16.87	-20.13	-20.53
America	15.22	NA	-1.60	-1.68	-3.74	0.46	0.79
CeAsSi	17.87	1.60	NA	-0.55	-3.34	2.27	2.16
EastAsia	17.20	1.68	0.55	NA	-3.22	2.39	2.16
Oceania	16.87	3.74	3.34	3.22	NA	4.74	4.35
SouAs	20.13	-0.46	-2.27	-2.39	-4.74	NA	0.75
WEurAs	20.53	-0.79	-2.16	-2.16	-4.35	-0.75	NA
D (X, Y; Mezmaiskaya 1, Chimp)							
X/Y	Africa	America	CeAsSi	EastAsia	Oceania	SouAs	WEurAs
Africa	NA	-13.91	-17.12	-16.19	-16.21	-18.58	-17.94
America	13.91	NA	-0.92	-1.00	-3.00	0.89	1.47
CeAsSi	17.12	0.92	NA	-0.32	-2.95	2.37	2.71
EastAsia	16.19	1.00	0.32	NA	-2.99	2.35	2.50
Oceania	16.21	3.00	2.95	2.99	NA	4.49	4.46
SouAs	18.58	-0.89	-2.37	-2.35	-4.49	NA	1.41
WEurAs	17.94	-1.47	-2.71	-2.50	-4.46	-1.41	NA

doi:10.1038/nature26151

Table S10.3 Z-score of *D (Human$_1$, Human$_2$, Neandertal, Chimp)* for the high coverage *Altai* and *Vindija 33.19* Neandertals, restricted to the sites in the genome that are covered by a least one of the late Neandertal low coverage genomes. CeAsSi – Central Asia and Siberia; SouAs – Southern Asia, WEurAs – Western Eurasia. Blue indicates Z-score < -2, while yellow indicates Z-score. Only putatively deaminated fragments reported in the Table S3.2 were utilized, amounting to in total 3,892,358 informative sites across genome.

D (X, Y; Altai, Chimp)							
X/Y	Africa	America	CeAsSi	EastAsia	Oceania	SouAs	WEurAs
Africa	NA	-12.69	-15.55	-14.66	-15.92	-17.54	-16.90
America	12.69	NA	-2.01	-1.81	-4.74	0.51	1.36
CeAsSi	15.55	2.01	NA	-0.15	-4.16	2.63	3.18
EastAsia	14.66	1.81	0.15	NA	-4.23	2.43	2.81
Oceania	15.92	4.74	4.16	4.23	NA	5.78	5.77
SouAs	17.54	-0.51	-2.63	-2.43	-5.78	NA	1.95
WEurAs	16.90	-1.36	-3.18	-2.81	-5.77	-1.95	NA
D (X, Y; Vindija 33.19, Chimp)							
X/Y	Africa	America	CeAsSi	EastAsia	Oceania	SouAs	WEurAs
Africa	NA	-14.14	-17.07	-16.04	-16.34	-18.64	-18.34
America	14.14	NA	-1.91	-1.71	-4.00	1.05	1.66
CeAsSi	17.07	1.91	NA	-0.15	-3.46	3.37	3.42
EastAsia	16.04	1.71	0.15	NA	-3.57	3.09	3.03
Oceania	16.34	4.00	3.46	3.57	NA	5.54	5.32
SouAs	18.64	-1.05	-3.37	-3.09	-5.54	NA	1.51
WEurAs	18.34	-1.66	-3.42	-3.03	-5.32	-1.51	NA

doi:10.1038/nature26151

Table S10.4 ***D (Neandertal₁, Neandertal₂, African₁, African₂)*** **for all fragments and deaminated fragments of late Neandertals and** ***Mezmaiskaya 1*****, restricted to transversions.** Only putatively deaminated fragments reported in the Table S3.2 were utilized, amounting to in total 3,892,358 informative sites across genomes.

D ($Neandertal_1$, $Neandertal_2$, $African_1$, $African_2$)				All fragments					Deaminated fragments				
W	X	Y	Z	D-value (%)	Z-score	BABA	ABBA	# of sites used	D-value (%)	Z-score	BABA	ABBA	# of sites used
Altai	Vindija33.19	Dinka	Yoruba	-0.27	-0.34	6,344	6,379	4,784,846	-0.27	-0.34	6,344	6,379	4,784,846
Altai	Vindija33.19	Dinka	Mbuti	-1.69	-1.30	6,676	6,897	4,786,590	-1.69	-1.30	6,676	6,897	4,786,590
Altai	Vindija33.19	Yoruba	Mbuti	-1.45	-1.00	6,754	6,943	4,786,378	-1.45	-1.00	6,754	6,943	4,786,378
Altai	Les Cottés	Dinka	Yoruba	-0.13	-0.16	6,037	6,053	4,230,006	-0.02	-0.03	3,656	3,658	2,878,157
Altai	Les Cottés	Dinka	Mbuti	-1.90	-1.47	6,293	6,531	4,231,248	-0.77	-0.67	3,830	3,887	2,878,885
Altai	Les Cottés	Yoruba	Mbuti	-1.76	-1.31	6,363	6,584	4,231,102	-0.74	-0.61	3,885	3,940	2,878,863
Altai	Goyet Q56-1	Dinka	Yoruba	0.04	0.06	6,095	6,090	4,170,659	0.90	0.83	1,508	1,480	1,150,925
Altai	Goyet Q56-1	Dinka	Mbuti	-1.68	-1.57	6,372	6,586	4,171,941	-1.42	-1.15	1,567	1,611	1,151,222
Altai	Goyet Q56-1	Yoruba	Mbuti	-1.73	-1.68	6,437	6,659	4,171,845	-2.26	-2.14	1,564	1,636	1,151,205
Altai	Mezmaiskaya1	Dinka	Yoruba	-0.42	-0.57	5,594	5,642	3,383,791	0.39	0.37	1,866	1,852	1,590,564
Altai	Mezmaiskaya1	Dinka	Mbuti	-4.33	-5.24	5,699	6,213	3,384,808	-1.91	-1.71	1,925	1,999	1,591,020
Altai	Mezmaiskaya1	Yoruba	Mbuti	-3.93	-4.12	5,773	6,243	3,384,701	-2.24	-1.97	1,942	2,030	1,590,942
Altai	Mezmaiskaya2	Dinka	Yoruba	1.51	2.05	5,283	5,126	3,807,901	1.30	1.38	2,452	2,389	2,004,604
Altai	Mezmaiskaya2	Dinka	Mbuti	0.61	0.74	5,563	5,495	3,809,156	1.05	1.05	2,595	2,541	2,005,237
Altai	Mezmaiskaya2	Yoruba	Mbuti	-0.82	-1.00	5,521	5,612	3,809,023	-0.19	-0.19	2,571	2,581	2,005,141
Altai	Spy 94a	Dinka	Yoruba	-1.33	-1.81	5,187	5,327	2,855,390	-1.11	-0.87	1,197	1,224	1,033,829
Altai	Spy 94a	Dinka	Mbuti	-5.52	-5.95	5,209	5,815	2,856,137	-2.08	-1.33	1,236	1,287	1,034,109
Altai	Spy 94a	Yoruba	Mbuti	-4.20	-3.84	5,306	5,768	2,856,081	-0.95	-0.53	1,271	1,293	1,034,105
Vindija33.19	Les Cottés	Dinka	Yoruba	0.03	0.03	4,100	4,098	4,228,074	0.18	0.16	2,416	2,407	2,877,506
Vindija33.19	Les Cottés	Dinka	Mbuti	-1.25	-1.16	4,274	4,381	4,229,306	0.05	0.05	2,541	2,538	2,878,223
Vindija33.19	Les Cottés	Yoruba	Mbuti	-1.21	-1.25	4,302	4,406	4,229,160	-0.09	-0.08	2,573	2,577	2,878,204
Vindija33.19	Goyet Q56-1	Dinka	Yoruba	0.60	0.66	3,785	3,739	4,169,229	0.97	0.68	881	863	1,150,711
Vindija33.19	Goyet Q56-1	Dinka	Mbuti	-1.01	-1.04	3,917	3,997	4,170,520	-0.76	-0.53	907	921	1,151,011
Vindija33.19	Goyet Q56-1	Yoruba	Mbuti	-1.56	-1.72	3,940	4,064	4,170,422	-1.66	-1.20	910	941	1,150,996
Vindija33.19	Mezmaiskaya1	Dinka	Yoruba	-0.44	-0.52	4,466	4,506	3,382,369	1.66	1.28	1,399	1,353	1,590,187
Vindija33.19	Mezmaiskaya1	Dinka	Mbuti	-4.06	-4.78	4,512	4,894	3,383,375	-0.21	-0.16	1,444	1,451	1,590,638
Vindija33.19	Mezmaiskaya1	Yoruba	Mbuti	-3.65	-4.69	4,576	4,923	3,383,255	-1.84	-1.53	1,428	1,481	1,590,557
Vindija33.19	Mezmaiskaya2	Dinka	Yoruba	2.21	2.37	3,693	3,533	3,806,419	2.16	1.72	1,679	1,608	2,004,143
Vindija33.19	Mezmaiskaya2	Dinka	Mbuti	3.31	1.91	3,977	3,730	3,807,626	3.35	1.82	1,811	1,697	2,004,749
Vindija33.19	Mezmaiskaya2	Yoruba	Mbuti	1.26	0.67	3,918	3,830	3,807,509	1.39	0.69	1,779	1,734	2,004,666
Vindija33.19	Spy 94a	Dinka	Yoruba	-1.51	-1.86	3,708	3,822	2,854,390	-1.24	-0.77	745	764	1,033,640
Vindija33.19	Spy 94a	Dinka	Mbuti	-6.32	-7.22	3,682	4,178	2,855,139	-1.24	-0.74	777	796	1,033,921
Vindija33.19	Spy 94a	Yoruba	Mbuti	-4.79	-6.16	3,742	4,118	2,855,094	-0.06	-0.04	789	789	1,033,918

doi:10.1038/nature26151

Table S10.5 Continuation of the Table S10.5 with the *D (Neandertal$_1$, Neandertal$_2$, African$_1$, African$_2$)* for all fragments and deaminated fragments of late Neandertals and *Mezmaiskaya 1*, restricted to transversions. Only putatively deaminated fragments reported in the Table S3.2 were utilized, amounting to in total 3,892,358 informative sites across genomes.

D (Neandertal$_1$, Neandertal$_2$, African$_1$, African$_2$)				All fragments					Deaminated fragments				
W	X	Y	Z	D-value (%)	Z-score	BABA	ABBA	# of sites used	D-value (%)	Z-score	BABA	ABBA	# of sites used
Les Cottés	Goyet Q56-1	Dinka	Yoruba	-0.38	-0.55	6,226	6,273	4,891,941	0.95	0.63	803	788	875,477
Les Cottés	Goyet Q56-1	Dinka	Mbuti	-0.56	-0.69	6,549	6,624	4,894,469	-0.21	-0.15	848	851	875,813
Les Cottés	Goyet Q56-1	Yoruba	Mbuti	-0.24	-0.33	6,638	6,671	4,894,240	-1.05	-0.70	845	863	875,801
Les Cottés	Mezmaiskaya1	Dinka	Yoruba	-0.58	-0.86	6,510	6,586	3,991,584	0.64	0.51	1,250	1,234	1,261,430
Les Cottés	Mezmaiskaya1	Dinka	Mbuti	-2.60	-3.66	6,740	7,102	3,993,649	-1.01	-0.81	1,286	1,312	1,261,972
Les Cottés	Mezmaiskaya1	Yoruba	Mbuti	-2.10	-3.26	6,796	7,090	3,993,415	-1.73	-1.34	1,289	1,335	1,261,906
Les Cottés	Mezmaiskaya2	Dinka	Yoruba	1.69	2.33	5,800	5,607	4,388,597	1.61	1.34	1,408	1,363	1,516,909
Les Cottés	Mezmaiskaya2	Dinka	Mbuti	2.90	2.13	6,253	5,906	4,390,787	2.16	1.53	1,516	1,453	1,517,486
Les Cottés	Mezmaiskaya2	Yoruba	Mbuti	1.36	0.98	6,184	6,025	4,390,516	0.74	0.52	1,500	1,479	1,517,427
Les Cottés	Spy 94a	Dinka	Yoruba	-1.16	-1.72	5,357	5,483	3,406,273	-3.58	-2.31	672	722	810,845
Les Cottés	Spy 94a	Dinka	Mbuti	-3.98	-6.05	5,465	5,919	3,408,034	-1.01	-0.72	720	734	811,192
Les Cottés	Spy 94a	Yoruba	Mbuti	-2.80	-4.45	5,556	5,877	3,407,805	2.53	1.80	737	701	811,171
Goyet Q56-1	Mezmaiskaya1	Dinka	Yoruba	0.02	0.03	6,496	6,494	3,885,011	2.42	1.29	501	478	483,403
Goyet Q56-1	Mezmaiskaya1	Dinka	Mbuti	-2.44	-3.97	6,648	6,981	3,886,986	1.51	0.81	519	503	483,614
Goyet Q56-1	Mezmaiskaya1	Yoruba	Mbuti	-2.46	-4.02	6,719	7,058	3,886,850	-0.87	-0.54	509	518	483,557
Goyet Q56-1	Mezmaiskaya2	Dinka	Yoruba	1.36	2.02	5,696	5,544	4,315,492	2.53	1.52	582	554	612,130
Goyet Q56-1	Mezmaiskaya2	Dinka	Mbuti	2.72	2.52	6,080	5,761	4,317,677	2.98	1.67	615	579	612,401
Goyet Q56-1	Mezmaiskaya2	Yoruba	Mbuti	1.42	1.39	6,103	5,937	4,317,446	0.51	0.33	601	595	612,390
Goyet Q56-1	Spy 94a	Dinka	Yoruba	-0.87	-1.38	5,047	5,136	3,292,764	2.94	1.15	252	238	312,081
Goyet Q56-1	Spy 94a	Dinka	Mbuti	-3.91	-6.02	5,106	5,521	3,294,424	1.88	0.83	262	252	312,193
Goyet Q56-1	Spy 94a	Yoruba	Mbuti	-2.97	-4.48	5,185	5,502	3,294,162	-1.05	-0.44	251	256	312,182
Mezmaiskaya2	Mezmaiskaya1	Dinka	Yoruba	1.50	2.27	5,627	5,461	3,513,553	-0.24	-0.17	774	777	845,479
Mezmaiskaya2	Mezmaiskaya1	Dinka	Mbuti	-4.47	-5.42	5,595	6,117	3,515,339	0.24	0.16	817	813	845,858
Mezmaiskaya2	Mezmaiskaya1	Yoruba	Mbuti	-3.04	-3.53	5,716	6,072	3,515,133	0.64	0.46	824	814	845,783
Mezmaiskaya2	Spy 94a	Dinka	Yoruba	-0.50	-0.86	5,292	5,346	2,753,086	-0.98	-0.54	413	422	451,028
Mezmaiskaya2	Spy 94a	Dinka	Mbuti	-0.72	-1.20	5,530	5,609	2,754,450	0.54	0.30	440	435	451,230
Mezmaiskaya2	Spy 94a	Yoruba	Mbuti	-0.16	-0.25	5,599	5,616	2,754,212	1.46	0.82	445	432	451,183
Mezmaiskaya1	Spy 94a	Dinka	Yoruba	-2.20	-3.34	4,697	4,909	2,962,805	0.05	0.03	455	455	539,115
Mezmaiskaya1	Spy 94a	Dinka	Mbuti	-6.11	-6.30	4,739	5,354	2,964,240	-1.14	-0.62	473	484	539,324
Mezmaiskaya1	Spy 94a	Yoruba	Mbuti	-3.91	-3.51	4,855	5,247	2,964,005	-1.14	-0.56	471	482	539,301

Table S10.6 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, non-African, Mbuti)* for deaminated fragments and restricted to transversions, for Western Eurasians and East Asians. Blue indicates Z-score < -2, yellow indicates Z-score > 2. Only putatively deaminated fragments reported in the Table S3.2 were utilized, amounting to in total 3,892,358 informative sites.

D (Nea1, Nea2, French, Mbuti) - Western Eurasia								
X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	6.44	-4.64	-5.53	-4.69	-4.79	-3.56	-5.43
Denisova	-6.44	NA	-7.20	-8.50	-10.20	-8.84	-6.10	-8.12
Goyet Q56-1	4.64	7.20	NA	1.68	0.36	1.04	0.56	1.93
Les Cottés	5.53	8.50	-1.68	NA	-0.57	0.42	-0.32	1.09
Mezmaiskaya1	4.69	10.20	-0.36	0.57	NA	-0.53	-1.11	0.88
Mezmaiskaya2	4.79	8.84	-1.04	-0.42	0.53	NA	0.37	0.46
Spy 94a	3.56	6.10	-0.56	0.32	1.11	-0.37	NA	0.85
Vindija33.19	5.43	8.12	-1.93	-1.09	-0.88	-0.46	-0.85	NA
D (Nea1, Nea2, Sardinian, Mbuti) - Western Eurasia								
X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	8.39	-3.37	-4.03	-3.26	-3.95	-2.46	-4.01
Denisova	-8.39	NA	-7.80	-9.54	-10.52	-10.04	-6.99	-9.38
Goyet Q56-1	3.37	7.80	NA	1.06	0.96	-0.40	0.60	0.53
Les Cottés	4.03	9.54	-1.06	NA	-0.68	-0.35	0.16	0.01
Mezmaiskaya1	3.26	10.52	-0.96	0.68	NA	-0.59	-1.27	-0.20
Mezmaiskaya2	3.95	10.04	0.40	0.35	0.59	NA	0.06	0.59
Spy 94a	2.46	6.99	-0.60	-0.16	1.27	-0.06	NA	0.36
Vindija33.19	4.01	9.38	-0.53	-0.01	0.20	-0.59	-0.36	NA
D (Nea1, Nea2, Han, Mbuti) - East Asia								
X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	5.67	-4.25	-3.84	-3.58	-3.07	-3.60	-4.39
Denisova	-5.67	NA	-7.03	-8.17	-8.63	-7.89	-6.69	-7.99
Goyet Q56-1	4.25	7.03	NA	0.56	-0.16	1.15	0.26	-0.17
Les Cottés	3.84	8.17	-0.56	NA	-1.41	0.36	-0.26	-1.27
Mezmaiskaya1	3.58	8.63	0.16	1.41	NA	-0.14	-1.24	-0.99
Mezmaiskaya2	3.07	7.89	-1.15	-0.36	0.14	NA	-0.87	-1.62
Spy 94a	3.60	6.69	-0.26	0.26	1.24	0.87	NA	0.77
Vindija33.19	4.39	7.99	0.17	1.27	0.99	1.62	-0.77	NA
D (Nea1, Nea2, Dai, Mbuti) - East Asia								
X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	7.52	-3.58	-3.59	-3.84	-4.10	-3.00	-4.87
Denisova	-7.52	NA	-7.03	-8.84	-10.13	-9.62	-6.71	-8.76
Goyet Q56-1	3.58	7.03	NA	1.35	0.10	0.79	1.03	1.17
Les Cottés	3.59	8.84	-1.35	NA	-1.31	-0.40	0.24	-0.35
Mezmaiskaya1	3.84	10.13	-0.10	1.31	NA	-1.20	-1.70	0.21
Mezmaiskaya2	4.10	9.62	-0.79	0.40	1.20	NA	-0.92	0.23
Spy 94a	3.00	6.71	-1.03	-0.24	1.70	0.92	NA	-0.14
Vindija33.19	4.87	8.76	-1.17	0.35	-0.21	-0.23	0.14	NA

Table S10.7 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, non-African, Mbuti)* for deaminated fragments and restricted to transversions, for Oceania and Americas. Blue indicates Z-score < -2, yellow indicates Z-score > 2. Only putatively deaminated fragments reported in the Table S3.2 were utilized, amounting to in total 3,892,358 informative sites.

D (Nea1, Nea2, Papuan, Mbuti) - Oceania

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	-0.08	-3.21	-3.47	-2.91	-4.01	-2.94	-4.08
Denisova	0.08	NA	-1.11	-1.23	-1.99	-2.12	-1.36	-1.01
Goyet Q56-1	3.21	1.11	NA	1.34	-0.59	-0.08	0.12	0.21
Les Cottés	3.47	1.23	-1.34	NA	-1.81	-0.86	-0.77	-0.24
Mezmaiskaya1	2.91	1.99	0.59	1.81	NA	-1.59	-1.60	-0.30
Mezmaiskaya2	4.01	2.12	0.08	0.86	1.59	NA	-0.73	0.11
Spy 94a	2.94	1.36	-0.12	0.77	1.60	0.73	NA	0.65
Vindija33.19	4.08	1.01	-0.21	0.24	0.30	-0.11	-0.65	NA

D (Nea1, Nea2, Bougainville, Mbuti) - Oceania

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	2.00	-4.41	-3.99	-3.70	-4.60	-2.97	-5.38
Denisova	-2.00	NA	-3.63	-2.84	-4.69	-4.15	-3.10	-3.22
Goyet Q56-1	4.41	3.63	NA	1.82	1.73	0.87	0.16	1.13
Les Cottés	3.99	2.84	-1.82	NA	-1.45	-0.50	-1.23	-0.69
Mezmaiskaya1	3.70	4.69	-1.73	1.45	NA	-1.50	-1.13	-1.11
Mezmaiskaya2	4.60	4.15	-0.87	0.50	1.50	NA	-0.51	0.66
Spy 94a	2.97	3.10	-0.16	1.23	1.13	0.51	NA	0.49
Vindija33.19	5.38	3.22	-1.13	0.69	1.11	-0.66	-0.49	NA

D (Nea1, Nea2, Karitiana, Mbuti) - America

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	6.12	-4.62	-3.19	-4.24	-3.97	-4.23	-5.63
Denisova	-6.12	NA	-6.06	-7.50	-8.38	-8.60	-6.73	-7.51
Goyet Q56-1	4.62	6.06	NA	0.73	-0.68	-0.40	-0.66	-0.05
Les Cottés	3.19	7.50	-0.73	NA	-1.88	-0.59	-0.45	-0.91
Mezmaiskaya1	4.24	8.38	0.68	1.88	NA	-0.53	-1.49	-0.03
Mezmaiskaya2	3.97	8.60	0.40	0.59	0.53	NA	-0.60	-0.20
Spy 94a	4.23	6.73	0.66	0.45	1.49	0.60	NA	0.60
Vindija33.19	5.63	7.51	0.05	0.91	0.03	0.20	-0.60	NA

D (Nea1, Nea2, Mixe, Mbuti) - America

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	7.41	-4.40	-4.07	-4.38	-4.89	-3.48	-5.39
Denisova	-7.41	NA	-7.05	-8.41	-9.53	-9.53	-6.79	-8.66
Goyet Q56-1	4.40	7.05	NA	0.13	0.83	-0.65	0.71	-0.07
Les Cottés	4.07	8.41	-0.13	NA	-1.47	-0.57	-0.10	-0.25
Mezmaiskaya1	4.38	9.53	-0.83	1.47	NA	-1.26	-1.37	-0.20
Mezmaiskaya2	4.89	9.53	0.65	0.57	1.26	NA	0.13	0.67
Spy 94a	3.48	6.79	-0.71	0.10	1.37	-0.13	NA	-0.36
Vindija33.19	5.39	8.66	0.07	0.25	0.20	-0.67	0.36	NA

doi:10.1038/nature26151

Table S10.7 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, non-African, Mbuti)* for deaminated fragments and restricted to transversions, for Central Asia and Siberia and South Asia. Blue indicates Z-score < -2, yellow indicates Z-score > 2. Only putatively deaminated fragments reported in the Table S3.2 were utilized, amounting to in total 3,892,358 informative sites across genomes.

D (Nea1, Nea2, Eskimo Naukan, Mbuti) - Central Asia and Siberia

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	7.43	-2.59	-4.00	-3.16	-2.96	-2.69	-4.18
Denisova	-7.43	NA	-6.64	-8.84	-9.68	-8.91	-6.66	-8.43
Goyet Q56-1	2.59	6.64	NA	0.02	0.36	-0.03	0.84	0.26
Les Cottés	4.00	8.84	-0.02	NA	-1.26	0.48	0.36	0.39
Mezmaiskaya1	3.16	9.68	-0.36	1.26	NA	-1.31	-1.94	-0.86
Mezmaiskaya2	2.96	8.91	0.03	-0.48	1.31	NA	-0.38	-0.50
Spy 94a	2.69	6.66	-0.84	-0.36	1.94	0.38	NA	0.39
Vindija33.19	4.18	8.43	-0.26	-0.39	0.86	0.50	-0.39	NA

D (Nea1, Nea2, Yakut, Mbuti) - Central Asia and Siberia

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	8.18	-3.41	-4.04	-3.53	-4.15	-3.19	-5.47
Denisova	-8.18	NA	-7.30	-9.12	-10.11	-9.65	-6.75	-9.32
Goyet Q56-1	3.41	7.30	NA	0.37	0.64	0.23	0.09	0.26
Les Cottés	4.04	9.12	-0.37	NA	-0.14	-0.76	0.15	-0.34
Mezmaiskaya1	3.53	10.11	-0.64	0.14	NA	-1.35	-1.49	-1.25
Mezmaiskaya2	4.15	9.65	-0.23	0.76	1.35	NA	1.03	0.74
Spy 94a	3.19	6.75	-0.09	-0.15	1.49	-1.03	NA	-0.05
Vindija33.19	5.47	9.32	-0.26	0.34	1.25	-0.74	0.05	NA

D (Nea1, Nea2, Kusunda, Mbuti) - South Asia

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	7.23	-3.24	-3.28	-2.90	-3.59	-2.14	-4.11
Denisova	-7.23	NA	-6.80	-8.43	-9.83	-8.93	-5.92	-8.33
Goyet Q56-1	3.24	6.80	NA	1.40	0.37	-0.18	0.23	0.43
Les Cottés	3.28	8.43	-1.40	NA	-0.52	0.57	-0.26	-0.35
Mezmaiskaya1	2.90	9.83	-0.37	0.52	NA	-0.74	-0.82	-1.09
Mezmaiskaya2	3.59	8.93	0.18	-0.57	0.74	NA	0.36	0.01
Spy 94a	2.14	5.92	-0.23	0.26	0.82	-0.36	NA	-0.22
Vindija33.19	4.11	8.33	-0.43	0.35	1.09	-0.01	0.22	NA

D (Nea1, Nea2, Yadava, Mbuti) - South Asia

X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	8.07	-2.91	-3.33	-3.46	-3.90	-2.33	-4.75
Denisova	-8.07	NA	-7.80	-9.10	-10.34	-10.37	-7.25	-9.12
Goyet Q56-1	2.91	7.80	NA	0.90	0.52	-0.34	-0.24	0.29
Les Cottés	3.33	9.10	-0.90	NA	-0.63	0.75	-1.10	-0.38
Mezmaiskaya1	3.46	10.34	-0.52	0.63	NA	-1.33	-1.70	-0.94
Mezmaiskaya2	3.90	10.37	0.34	-0.75	1.33	NA	0.20	0.04
Spy 94a	2.33	7.25	0.24	1.10	1.70	-0.20	NA	0.14
Vindija33.19	4.75	9.12	-0.29	0.38	0.94	-0.04	-0.14	NA

doi:10.1038/nature26151

Table S10.8 Z-scores of *D (Neandertal$_1$, Neandertal$_2$, ancient modern human, Mbuti)* for deaminated fragments and restricted to transversions. Blue - Z-score < -2,Yellow – Z-score > 2. Only putatively deaminated fragments reported in the Table S3.2 were utilized, amounting to in total 3,892,358 informative sites across genomes.

D (Nea1, Nea2, Ust'-Ishim, Mbuti)								
X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	7.30	-4.27	-3.64	-3.06	-3.88	-2.94	-4.75
Denisova	-7.30	NA	-7.09	-7.65	-9.04	-8.55	-6.72	-7.90
Goyet Q56-1	4.27	7.09	NA	1.56	1.30	1.68	-1.02	0.00
Les Cottés	3.64	7.65	-1.56	NA	-0.13	-0.51	-0.59	-1.37
Mezmaiskaya1	3.06	9.04	-1.30	0.13	NA	-1.01	-1.69	-1.98
Mezmaiskaya2	3.88	8.55	-1.68	0.51	1.01	NA	-0.31	-0.57
Spy 94a	2.94	6.72	1.02	0.59	1.69	0.31	NA	-0.23
Vindija33.19	4.75	7.90	0.00	1.37	1.98	0.57	0.23	NA
D (Nea1, Nea2, Loschbour, Mbuti)								
X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	7.66	-2.81	-2.96	-3.30	-3.72	-3.08	-3.67
Denisova	-7.66	NA	-7.70	-8.23	-8.67	-9.34	-6.12	-8.57
Goyet Q56-1	2.81	7.70	NA	1.81	0.26	0.01	-0.59	0.76
Les Cottés	2.96	8.23	-1.81	NA	-0.92	-0.93	-1.57	-0.72
Mezmaiskaya1	3.30	8.67	-0.26	0.92	NA	-0.22	-1.26	-0.32
Mezmaiskaya2	3.72	9.34	-0.01	0.93	0.22	NA	0.33	0.32
Spy 94a	3.08	6.12	0.59	1.57	1.26	-0.33	NA	0.88
Vindija33.19	3.67	8.57	-0.76	0.72	0.32	-0.32	-0.88	NA
D (Nea1, Nea2, LBK, Mbuti)								
X/Y	Altai	Denisova	Goyet Q56-1	Les Cottés	Mezmaiskaya1	Mezmaiskaya2	Spy 94a	Vindija33.19
Altai	NA	6.96	-3.07	-3.18	-3.96	-4.24	-2.89	-3.72
Denisova	-6.96	NA	-7.02	-8.31	-9.70	-8.29	-4.97	-7.92
Goyet Q56-1	3.07	7.02	NA	0.59	0.76	0.00	1.23	0.45
Les Cottés	3.18	8.31	-0.59	NA	-1.10	-0.02	0.53	0.74
Mezmaiskaya1	3.96	9.70	-0.76	1.10	NA	-0.61	-0.29	0.72
Mezmaiskaya2	4.24	8.29	0.00	0.02	0.61	NA	0.62	1.61
Spy 94a	2.89	4.97	-1.23	-0.53	0.29	-0.62	NA	0.42
Vindija33.19	3.72	7.92	-0.45	-0.74	-0.72	-1.61	-0.42	NA

Supplementary Information 11

No evidence of a recent early modern human admixture into late Neandertals

Late Neandertals presented in this study overlapped in time with the putative arrival of early modern humans in Eurasia. As it was recently shown that the admixture between Neandertals and modern humans was not limited to the ancestors of present-day humans and the Levant, but happened at later times and in Europe as well[1], we investigated if we can detect a recent gene flow in the other direction, *i.e.* from early modern humans into late Neandertals. By using simulation data (see the section on the simulation testing below) where we model admixture from modern humans into Neandertals, and with an ascertainment scheme in which both the *Denisovan* individual and the *Altai* Neandertal are fixed for the ancestral allele and at least half of the alleles in present-day African populations are derived, we apply the method as described in Moorjani *et al.*[2] and estimate a date of a recent early modern human (EMH) admixture into Neandertals (~10-100 generations ago). When we apply this method to the actual data sampled from Neandertals from this and previous studies, we find no evidence of recent early modern human admixture in late Neandertals (see the section below on Neandertal results).

Here, we describe our SNPs ascertainment scheme and whether this strategy works in simulations, and then we apply this scheme to the Neandertal data presented in this study. At each SNP in the genome, we consider genetic information from all Yoruba individuals from the 1000 Genomes Project[3] covered by at least three reads that pass a pre-defined set of filters. At these sites, we called majority alleles (drawing a random allele in the case when an equal number of reads supports both alleles). Furthermore, we restricted the analysis to sites in the genome where $\geq$ 24 Yoruba individuals as well as the *Altai* Neandertal and the *Denisovan* individual had allele calls (Map35_50% filter from Prüfer *et al.*[4]). We then selected sites where the *Denisovan* individual and the *Altai* Neandertal are fixed ancestral and more than half of the alleles are derived among the Yoruba individuals. The ancestral states were taken from the inferred ancestor of humans and chimpanzees (Ensembl Compara v64)[5,6]. In total, we got 648,209 SNPs, out of which 45,266 SNPs overlapped with the archaic admixture array of Fu *et al.*[1].

Simulation testing

To assess the ascertainment scheme of the EMH admixture, we performed coalescent simulations using ms[7] under various demographic models. We generated data for three populations that we chose to have demographic parameters roughly similar to what we expect for Neandertals, sub-Saharan Africans and non-Africans. We used an ascertainment scheme in which the *Denisovan* individual and the *Altai* Neandertal are both fixed for the ancestral allele and at least half of the alleles in the Africans are derived. For all SNPs that matched our ascertainment scheme, we performed an analysis using the statistic described previously[8] to estimate the date of EMH admixture.

We introduced a more complex demographic history that is loosely based on the model described in Gravel *et al.*[9]. We do not use the population sizes and the European-Asian population split time estimated by Gravel *et al.* (of 17,200–26,500 years before present), since these are incompatible with the recent observations that the Tianyuan individual from Asia, an ancient specimen dating to ~40,000 years before present, is already part of the lineage leading to present-day Asians[10]. We generated 100 African, 4 European, 4 Asian, 2 introgressing Neandertals, 2 *Altai*, 2 *Denisovan*, 2 ancient non-African, and 2 admixed Neandertal haploid sequences. We set the parameters as follows:

- Mutation rate = 1.5×10^{-8} per bp/generation
- Recombination rate = 2×10^{-8} per bp/generation
- The *Altai* Neandertal splits from the introgressing Neandertal 4,000 generations before present
- We sample the *Altai* Neandertal 2,400 generations before present
- Africans split from Neandertals 12,000 generations before present
- N_e of Neandertals is 2,500
- We introduce a population expansion in the common ancestor of present-day humans at 6,000 generations ago with N_e increasing from 7,000 to 14,000
- Non-Africans split from Africans at 3,000 generations before present and underwent a bottleneck until 2,200 generations before present, reducing N_e to 1,860
- Europeans and East Asians split 2,000 generations ago with a reduction in N_e of East Asians to 550, and N_e of Europeans to 1,032
- Both European and East Asian populations subsequently undergo expansions until the present ($t = 0$) where the present-day East Asians have an N_e of 45,300 and the present-day Europeans have an N_e of 33,800

- Gene flow from Neandertals into the ancestors of present-day non-Africans occurred at 2,200 generations before present
- The proportion of Neandertal ancestry in non-Africans is 0.03
- We sample EMH 1,500 generations before present
- The split time between Eurasian and EMH occurred 2,000 generations before present
- Gene flow from EMH into the ancestors of admixed Neandertals occurred at 1,700 generations before present or 1,610 generations before present
- The proportion of EMH ancestry in admixed Neandertals is 0.05
- We sample the admixed Neandertal 1,600 generations before present

Table S11.1 shows the estimated dates of EMH admixture based on the simulation data. For an expected number of 100 generations since the admixture, we obtain an estimate of 95 generations with a standard error of 16 generations. And for an expected number of 10 generations since the admixture, we obtain an estimate of 7 generations with a standard error of 4 generations. Therefore, our estimated results are in the range of the expected number of generations, indicating that our ascertainment strategy helps to discern recent admixture events.

Neandertal results indicate that there was no recent gene flow from early modern humans

With the 648,209 SNPs available after applying the ascertainment scheme described above, we calculated the average linkage disequilibrium (LD) over all pairs of SNPs within a 0.001 cM bin, fitting an exponential curve to the decay of LD based on the method described in Moorjani *et al.*[11]. We applied the method to all fragments longer than 35 bp with mapping quality of at least 25, that overlapped highly mappable regions of the genome (Map35_100%), and to putatively deaminated fragments (see Supplementary Information 6 for basic set of filters applied to all data). For comparative purposes, we also randomly sampled fragments for the two high coverage Neandertals in the same way as for the late Neandertals (described in the Supplementary Information 6). Using this approach, we obtained an estimate of the time of the Neandertal admixture event that ranges between 1,952 and 2,448 generations before the death of Neandertal individuals (Table S11.2), which massively exceeds 100 generations. Therefore, there is no evidence showing recent admixture from early modern human into these late Neandertals.

References for SI11:

1 Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216-219, doi:10.1038/nature14558 (2015).

2 Moorjani, P. *et al.* A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proceedings of the National Academy of Sciences* **113**, 5652-5657, doi:10.1073/pnas.1514696113 (2016).

3 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).

4 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

5 Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* **18**, 1814-1828, doi:10.1101/gr.076554.108 (2008).

6 Paten, B. *et al.* Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res* **18**, 1829-1843, doi:10.1101/gr.076521.108 (2008).

7 Hudson, R. R. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-338, doi:10.1093/bioinformatics/18.2.337 (2002).

8 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449, doi:10.1038/nature13810 (2014).

9 Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 11983-11988, doi:DOI 10.1073/pnas.1019276108 (2011).

10 Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci U S A* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

11 Moorjani, P. *et al.* A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 5652-5657, doi:10.1073/pnas.1514696113 (2016).

doi:10.1038/nature26151

Table S11.1 Number of generations prior to an early modern human admixture into late Neandertals used for simulations and obtained results.

Individual	Expected	Estimate obtained from the method	Std Error
Admixed Neandertal	100	95	16
Admixed Neandertal	10	7	4

Table S11.2 Estimates of a number of generations since an early modern admixture into an ancestor of a given Neandertal. Calculations are based on the random sampling of all fragments and putatively deaminated fragments only.

Individual	Estimate of generations since admixture with EMH	Std Error
Altai Neandertal deaminated	2,435	221
Altai Neandertal all	2,448	167
Vindija 33.19 deaminated	2,258	197
Vindija 33.19 all	2,230	277
Goyet Q56-1 deaminated	2,181	547
Goyet Q56-1 all	2,090	278
Les Cottés Z4-1514 deaminated	1,952	297
Les Cottés Z4-1514 all	1,952	289
Mezmaiskaya 1 deaminated	2,215	645
Mezmaiskaya 1 all	2,025	353
Mezmaiskaya 2 deaminated	2,356	368
Mezmaiskaya 2 all	2,098	314
Spy 94a deaminated	2,552	658
Spy 94a all	2,126	441

Supplementary Information 12

Assessing the catalogue of human-specific fixed derived changes using the late Neandertals

A catalog of 31,389 modern-human-specific fixed differences was previously identified using the genomes of present-day humans and the high quality *Altai* and *Denisova* genomes[1]. At these positions the genomes of 1,094 present-day humans from the 1000 Genomes Project[2] are homozygous for the derived allele, while the high coverage genomes of the *Altai* Neandertal[1] and the *Denisovan* individual[3] are homozygous for the ancestral allele. The availability of additional Neandertal genomes now makes it possible to refine this catalog by identifying sites where Neandertals may have been variable.

We first investigated how many of these changes are shared with the new high coverage genome of the *Vindija 33.19*[4] Neandertal. The *Vindija 33.19* genome is homozygous for the archaic state (*i.e.* is identical to the *Altai* Neandertal and the *Denisovan* individual) at 29,976 sites (Supplementary Data File 1). At the remaining 1,413 sites the *Vindija 33.19* genome is either heterozygous and one of the alleles matches the present-day human state (n = 624), or is homozygous and both alleles match the present-day human state (n = 787). There are only 2 sites in the genome of *Vindija 33.19* where it has a third state, not present in the other two archaics or present-day humans.

We then identified DNA fragments from the late Neandertals and *Mezmaiskaya 1*[1,4] that overlap these 29,976 positions in order to determine which positions are likely to be fixed, and which are polymorphic, in Neandertals. Accurate inference of the derived state in the late Neandertals is complicated by the low coverage available for each of the individuals. The average coverage of the late Neandertals and *Mezmaiskaya 1* ranges from 1 to 2.7-fold and even low levels of present-day human contamination and sequencing error could inflate the number of estimated non-archaic states. To address this we therefore considered only fragments that were deaminated, and likely of archaic origin, and we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G[5] (as described in Supplementary Information 2). We also counted only alleles matching either the ancestral or the derived states to reduce the impact of sequencing errors.

Between 5,707 and 24,566 deaminated DNA fragments from the late Neandertals and *Mezmaiskaya 1* overlapped 26,206 out of 29,976 informative positions in the catalogue, and between 97.12% and 99.01% have the ancestral allele (Table S12.1). When restricting the

analyses to putatively deaminated fragments longer than 35 base pairs (bp) with the mapping quality of at least 25, there were no sites at which all of the late Neandertals and *Mezmaiskaya 1* would be identical for a derived variant (Table S12.2). However, there were 6 sites where at least four Neandertals individuals agreed and had a derived variant as present-day humans (Tables S12.2 and S12.4).

Of the 96 fixed non-synonymous coding changes in 87 proteins that were identifed previously[1] we detect derived alleles for eight sites in at least one of the late Neandertals (Table S12.5). For two of these sites (1:101196790 (*VCAM1*) and 2:241463466 (*ANKMY1*)) we exclusively observe fragments carrying the derived state in three late Neandertals, suggsting that these sites were likely variable in Neandertals. Using the *Vindija 87* specimen, which derives from the same individual as the high coverage genome of *Vindija 33.19,* we estimate that 0.9% of the derived sites identified using only deaminated fragments are incorrectly assigned.

In conclusion, for the majority of the sites, and particularly for the non-synonymous changes that are observed in all high coverage archaics and fixed among the present-day humans, the late Neandertals and *Mezmaiskaya I* also carry the archaic state.

References for SI12:

1 Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).

2 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

3 Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).

4 Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, doi:10.1126/science.aao1887 (2017).

5 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

doi:10.1038/nature26151

Table 12.1 Proportion and number of nuclear DNA fragments in the late Neandertals and *Mezmaiskaya 1* matching the ancestral or the derived state at 29,976 positions where present-day humans from 1000 Genomes Project are fixed derived and the three high coverage archaics are ancestral. The results are shown for all fragments and only those fragments with terminal C-to-T substitutions (putatively deaminated fragments). In order to mitigate the effect of deamination (as described in Supplementary Information 2), we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G. The number of fragments supporting the archaic variant and the human variant, as well as the total number of observations are provided in brackets.

	All fragments			*Deaminated fragments*		
Individual	**# of fragments**	**% archaic [#observations]**	**% human [#observations]**	**# of fragments**	**% archaic [#observations]**	**% human [#observations]**
Goyet Q56-1	62,531	97.61 [30,231/30,972]	2.39 [741/30,972]	8,126	98.01 [3,500/3,571]	1.99 [71/3,571]
Les Cottés Z4-1514	74,122	97.54 [35,722/36,623]	2.46 [901/36,623]	24,566	97.69 [11,146/11,409]	2.31 [263/11,409]
Mezmaiskaya 1	38,799	96.09 [19,226/20,009]	3.91 [783/20,009]	10,044	97.50 [4,644/4,763]	2.50 [119/4,763]
Mezmaiskaya 2	42,729	97.07 [20,460/21,078]	2.93 [618/21,078]	12,612	97.12 [5,6895,858]	2.88 [169/5,858]
Spy 94a	26,264	95.45 [12,517/13,113]	4.55 [596/13,113]	5,707	97.96 [2,447/2,498]	2.04 [51/2,495]
Vindija 87	32,371	97.89 [15,937/16,281]	2.46 [344/16,281]	11,527	99.10 [5,388/5,437]	0.90 [49/5,437]

Table S12.2 Number of sites in the genome where late Neandertals and *Mezmaiskaya 1* carry a human derived allele. The positions in the genome are identified by using all fragments longer than 35 bp with mapping quality of at least 25 (L ≥ 35 bp, MQ ≥ 25) and deaminated fragments only. In order to mitigate the effect of deamination, we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G. Reported are the sites where only the human derived allele was observed in the fragments covering the site, *i.e.* no fragments carrying an archaic state were identified.

Number of individuals carrying a human derived allele	Number of sites all fragments	Number of sites deaminated fragments
2	287	196
3	151	42
4	91	9
5	29	0
6	10	0

Table S12.3 8 positions in the genome where all late Neandertals and *Mezmaiskaya 1* carry human derived alleles, identified by using all fragments longer than 35 bp with mapping quality of at least 25 (L ≥ 35 bp, MQ ≥ 25). In order to mitigate the effect of deamination, we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G. Reported are the sites where only the human derived allele was observed in the fragments covering the site, *i.e.* no fragments carrying an archaic state were identified.

Chromosome	Position
4	145,750,068
4	145,750,073
4	145,750,867
4	145,751,398
4	145,966,125
4	145,981,923
9	9,657,257
9	9,657,884

Table S12.4 6 positions in the genome where at least four late Neandertals and/or *Mezmaiskaya 1* carry human derived alleles identified by using deaminated fragments longer than 35 bp with mapping quality of at least 25 (L ≥ 35 bp, MQ ≥ 25). In order to mitigate the effect of deamination, we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G. Reported are the sites where only the human derived allele was observed in the fragments covering the site, *i.e.* no fragments carrying an archaic state were identified.

Chromosome	Position
4	145,866,762
4	145,966,125
6	29,483,833
11	29,863,432
11	112,953,400
13	52,191,663

doi:10.1038/nature26151

Table S12.5 Proportion and number of nuclear DNA fragments in the late Neandertals and *Mezmaiskaya 1* matching the ancestral or the derived state at 96 non-synonymous changes where present-day humans from 1000 Genomes Project are fixed derived and the three high coverage archaics are ancestral. The results are shown for all fragments and only those fragments with terminal C-to-T substitutions (putatively deaminated fragments). The number of fragments overlapping the informative sites are indicated. In order to mitigate the effect of deamination, we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G (strandedness filtering). The sites where all observed fragments are matching the archaic states are color-coded in red, the sites where at least one human derived allele is observed are indicated in yellow, and the sites where only human derived alleles are observed are indicated in green. The sites with no informative overlapping fragments that met the requirement of the strandedness filtering are denoted as 0/0.

			All fragments						Deaminated fragments					
chromosome: position	archaic allele	human allele	Goyet Q56-1	Les Cottés Z4-1514	Mezmaiskaya 1	Mezmaiskaya 2	Spy 94a	Vindija 87	Goyet Q56-1	Les Cottés Z4-1514	Mezmaiskaya 1	Mezmaiskaya 2	Spy 94a	Vindija 87
1:79106805	A	G	1/1	3/3	1/1	1/1	1/1	2/2	0/0	3/3	1/1	0/0	0/0	1/1
1:101196790	A	G	0/3	0/3	0/0	0/1	0/1	3/3	0/0	0/1	0/0	0/1	0/1	0/0
1:118558632	C	T	1/1	1/1	3/3	2/2	1/1	2/2	0/0	1/1	0/0	0/0	0/0	0/0
1:118634297	C	A	1/1	3/3	0/0	1/1	2/2	0/0	0/0	2/2	0/0	0/0	1/1	0/0
1:153751869	T	G	1/1	1/1	0/0	1/1	0/0	2/2	0/0	0/0	0/0	0/0	0/0	0/0
1:158648210	C	T	1/1	0/0	2/2	2/2	2/2	0/1	1/1	0/0	2/2	0/0	1/1	0/0
1:204966474	A	G	1/1	0/0	1/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
1:245582905	G	A	2/2	1/1	0/0	1/1	0/0	4/4	0/0	0/0	0/0	0/0	0/0	1/1
2:40657356	A	G	1/1	1/1	1/1	0/0	1/1	2/2	0/0	1/1	1/1	0/0	0/0	1/1
2:73438011	A	T	3/3	2/2	2/2	0/0	1/1	0/0	0/0	1/1	1/1	0/0	0/0	0/0
2:241463466	G	T	0/2	2/3	0/3	0/7	0/1	0/1	0/0	0/1	0/0	0/2	0/0	0/1
3:47469149	G	A	1/1	2/2	0/0	2/2	0/1	4/4	0/0	1/1	0/0	1/1	0/0	2/2
3:98073220	G	C	2/2	0/0	1/1	0/0	1/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0
3:98073475	C	G	1/1	4/4	2/2	0/0	1/1	1/1	0/0	1/1	0/0	0/0	0/0	1/1
4:2949274	C	G	1/1	1/1	0/0	2/2	1/1	1/1	0/0	1/1	0/0	0/0	0/0	0/0
4:5642249	T	C	0/0	0/0	3/3	3/3	1/2	0/0	0/0	0/0	1/1	0/0	0/0	0/0
4:89408223	A	G	1/1	0/0	0/0	4/4	1/1	0/0	0/0	0/0	0/0	1/1	1/1	0/0
4:89410317	G	A	4/4	0/0	3/3	2/2	0/0	1/1	0/0	0/0	0/0	1/1	0/0	0/0
5:54585213	T	C	1/1	2/2	0/0	0/0	1/1	0/0	1/1	2/2	0/0	0/0	0/0	0/0
5:71638807	A	G	0/0	0/0	0/0	2/2	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0
5:75591644	A	C	0/0	6/6	0/0	4/4	1/1	0/0	0/0	1/1	0/0	2/2	0/0	0/0
5:82837946	A	G	1/1	1/1	1/1	2/2	0/0	0/0	0/0	1/1	1/1	1/1	0/0	0/0
5:86564477	G	A	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
6:79577384	G	A	2/2	0/0	0/0	2/2	0/0	0/1	0/0	0/0	0/0	2/2	0/0	0/0

doi:10.1038/nature26151

6:100368868	A	G	0/0	2/2	0/1	2/2	0/0	1/1	0/0	1/1	0/1	0/0	0/0	1/1
6:109802500	G	C	0/0	2/2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
6:149918766	C	T	0/0	2/2	0/0	2/2	1/1	2/2	0/0	1/1	0/0	1/1	0/0	0/0
7:91793211	T	A	3/3	1/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
7:130418720	C	A	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
7:134642991	G	A	3/3	0/0	2/2	1/1	0/0	1/1	1/1	0/0	0/0	0/0	0/0	1/1
8:8869129	G	T	2/2	5/5	0/0	1/1	3/3	1/1	0/0	2/2	0/0	1/1	0/0	1/1
8:19316070	C	T	3/3	3/3	0/0	2/2	2/2	1/1	0/0	2/2	0/0	0/0	1/1	1/1
8:30557599	T	C	4/4	3/3	1/1	0/0	0/0	0/0	0/0	2/2	0/0	0/0	0/0	0/0
8:39537618	T	C	0/0	3/3	0/0	0/0	1/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0
8:39564352	A	G	0/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
8:53568742	T	C	1/1	2/2	1/1	2/2	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0
8:54975904	T	C	0/0	0/0	0/0	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0
8:145730809	T	G	1/1	0/0	0/0	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0
9:6606647	A	G	1/1	3/3	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
9:111929421	C	G	0/0	2/2	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
9:127113155	G	C	3/3	0/0	0/0	3/3	0/0	2/2	1/1	0/0	0/0	2/2	0/0	2/2
9:135275592	G	A	2/2	6/6	0/0	4/4	1/1	2/2	0/0	3/3	0/0	1/1	0/0	2/2
9:139836554	C	T	2/2	1/1	0/0	1/1	0/0	0/0	1/1	0/0	0/0	1/1	0/0	0/0
9:140139881	G	T	1/1	1/1	0/0	2/2	0/0	0/0	0/0	1/1	0/0	1/1	0/0	0/0
9:140507366	C	T	4/4	0/0	1/1	1/1	0/0	0/0	1/1	0/0	0/0	1/1	0/0	0/0
10:37508641	G	A	0/0	2/2	0/2	2/2	0/1	1/1	0/0	1/1	0/0	1/1	0/1	0/0
10:75000739	A	G	2/2	0/0	3/3	1/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0
10:102676434	A	G	0/0	3/3	3/3	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0
10:112660279	A	G	0/3	2/2	2/2	2/2	0/0	0/0	0/0	1/1	0/0	2/2	0/0	0/0
10:118321055	A	T	1/1	6/6	0/1	2/2	3/3	2/2	0/0	2/2	0/0	0/0	0/0	1/1
11:5530026	G	A	5/5	1/1	3/3	3/3	0/0	0/0	1/1	0/0	1/1	0/0	0/0	0/0
11:6654769	T	C	1/1	1/1	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
11:28119295	C	T	6/6	9/9	2/2	1/1	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0
11:59812212	A	C	1/1	0/0	0/1	2/2	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0
11:64884957	G	A	1/1	2/2	2/2	5/5	0/0	2/2	0/0	1/1	1/1	0/0	0/0	2/2
11:129772293	G	T	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	1/1
12:1937340	A	G	1/1	0/0	0/0	4/4	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0
12:6883790	A	G	0/0	1/1	1/1	1/1	0/0	1/1	0/0	0/0	0/0	1/1	0/0	1/1
12:46321732	C	T	2/2	3/3	3/3	2/2	1/1	3/3	0/0	2/2	0/0	1/1	0/0	1/1
13:84454655	C	A	3/3	0/0	1/1	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0
14:26918100	T	C	1/1	0/0	1/1	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0
14:76249759	G	A	2/2	1/1	3/3	1/1	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0
14:105517492	G	C	3/3	2/2	2/2	5/5	0/0	1/1	2/2	1/1	1/1	2/2	0/0	0/0
15:40912860	A	G	0/1	2/2	0/0	0/1	0/1	2/2	0/0	2/2	0/0	0/1	0/0	0/0
15:40915640	G	A	0/1	5/5	2/2	0/1	0/1	0/0	0/0	1/1	1/1	0/1	0/0	0/0
15:42985549	G	A	2/2	1/1	0/0	3/3	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0
15:48499987	A	G	1/1	0/0	1/1	2/2	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
15:81173308	A	G	1/1	0/0	1/1	1/1	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0
16:66947064	C	T	2/2	0/0	2/2	1/1	0/0	1/1	1/1	0/0	1/1	0/0	0/0	0/0
16:88804443	T	C	2/2	0/0	1/1	2/2	0/0	1/1	0/0	0/0	1/1	2/2	0/0	0/0

17:26919034	C	G	1/1	0/0	3/5	1/1	1/1	2/2	0/0	0/0	0/0	0/0	1/1	1/1
17:26919777	T	C	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
17:26925578	G	A	6/6	1/1	0/0	2/2	1/1	1/1	3/3	1/1	0/0	1/1	0/0	0/0
17:27959258	T	C	0/0	0/0	0/0	3/3	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0
17:35913918	G	A	2/2	2/3	3/3	1/2	0/0	1/1	1/1	0/0	1/1	1/2	1/1	0/0
17:41931199	T	C	1/1	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
17:62290457	T	C	2/2	1/1	1/1	0/0	1/1	0/0	0/0	1/1	0/0	0/0	0/0	0/0
17:73753035	A	G	1/1	1/1	0/0	1/1	0/0	0/0	0/0	1/1	0/0	1/1	0/0	0/0
17:73753305	A	G	1/1	1/1	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	1/1
17:80006980	A	G	0/0	0/0	0/0	1/1	0/1	0/0	0/0	0/0	0/0	0/0	0/0	0/0
18:19085379	G	A	3/3	3/3	0/0	0/0	0/0	2/2	0/0	1/1	0/0	0/0	0/0	0/0
19:2434033	T	A	3/3	1/1	1/1	1/1	0/0	1/1	1/1	1/1	1/1	1/1	0/0	1/1
19:3547315	G	C	2/2	0/0	2/2	4/4	2/2	0/0	0/0	0/0	2/2	3/3	0/0	0/0
20:33337529	C	A	1/1	0/0	0/0	5/5	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
20:44002597	T	G	3/3	1/1	2/2	1/1	0/0	0/0	2/2	1/1	1/1	0/0	0/0	0/0
20:44003932	G	A	1/1	2/2	2/2	0/0	0/0	1/1	0/0	0/0	1/1	0/0	0/0	0/0
21:34166190	T	A	5/5	1/1	1/1	1/2	1/1	1/1	2/2	1/1	1/1	1/2	1/1	0/0
21:43897491	T	G	0/0	3/3	2/2	0/0	1/1	0/0	0/0	1/1	1/1	0/0	1/1	0/0
22:40161572	G	C	4/4	4/4	2/2	1/1	2/2	0/0	1/1	2/2	0/0	0/0	1/1	0/0
22:40760978	C	T	2/2	1/1	0/0	0/0	1/1	0/0	0/0	1/1	0/0	0/0	0/0	0/0
X:23018785	T	C	0/0	1/1	1/1	1/1	1/2	1/1	0/0	0/0	0/0	0/0	0/0	1/1
X:36156570	A	G	1/1	4/5	0/0	2/2	2/2	2/2	0/0	3/3	0/0	0/0	0/0	1/1
X:76939325	C	G	1/1	2/2	2/2	1/1	1/1	1/1	1/1	2/2	2/2	0/0	1/1	0/0
X:131212487	C	G	1/1	0/0	2/2	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0	0/0
X:152128250	G	A	0/0	1/1	1/1	1/1	1/1	1/1	0/0	0/0	1/1	0/0	1/1	0/0
X:153543608	A	G	1/1	2/2	1/1	0/0	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0