Published in final edited form as:
Nature. 2018 March 29; 555(7698): 652–656. doi:10.1038/nature26151.

Reconstructing the Genetic History of Late Neandertals

Mateja Hajdinjak[1,*], **Qiaomei Fu**[2], **Alexander Hübner**[1], **Martin Petr**[1], **Fabrizio Mafessoni**[1], **Steffi Grote**[1], **Pontus Skoglund**[3], **Vagheesh Narasimham**[3], **Hélène Rougier**[4], **Isabelle Crevecoeur**[5], **Patrick Semal**[6], **Marie Soressi**[7,8], **Sahra Talamo**[8], **Jean-Jacques Hublin**[8], **Ivan Gušić**[9], **Željko Kućan**[9], **Pavao Rudan**[9], **Liubov V. Golovanova**[10], **Vladimir B. Doronichev**[10], **Cosimo Posth**[11,12], **Johannes Krause**[11,12], **Petra Korlević**[1], **Sarah Nagel**[1], **Birgit Nickel**[1], **Montgomery Slatkin**[13], **Nick Patterson**[3,14], **David Reich**[3,14,15], **Kay Prüfer**[1], **Matthias Meyer**[1], **Svante Pääbo**[1,*], and **Janet Kelso**[1,*]

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany [2]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China [3]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA [4]Department of Anthropology, California State University Northridge, Northridge, California 91330-8244, USA [5]Université de Bordeaux, CNRS, UMR 5199-PACEA, 33615 Pessac Cedex, France [6]Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium [7]Faculty of Archaeology, Leiden University, 2300 RA Leiden, The Netherlands [8]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany [9]Croatian Academy of Sciences and Arts, Zrinski trg 11, HR-10000 Zagreb, Croatia [10]ANO Laboratory of Prehistory 14 Linia 3-11, St. Petersburg 1990 34, Russia [11]Max Planck Institute for the Science of Human History, 07745 Jena, Germany [12]Institute for Archaeological Sciences, University of Tübingen, Rümelin Strasse 23, 72070 Tübingen, Germany [13]Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA [14]Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA [15]Howard Hughes Medical Institute, Seattle, Washington 98195, USA

Abstract

Although it is known that Neandertals contributed DNA to modern humans1,2, not much is known about the genetic diversity of Neandertals or the relationship between late Neandertal populations at the time when their last interactions with early modern humans occurred and before they eventually disappeared. Our ability to retrieve DNA from a larger number of Neandertal

*Correspondence and requests for materials should be addressed to M.H. (mateja_hajdinjak@eva.mpg.de), S.P. (paabo@eva.mpg.de) or J.Ke. (kelso@eva.mpg.de).

Author contribution: M.H., M.M. and S.P. conceived the study. M.Sl., N.P., D.R., K.P., M.M., S.P. and J.Ke. supervised the study. M.H., P.K., S.N. and B.N. performed ancient DNA lab work. H.R., I.C., P.Se., M.So., S.T., J.J.H., I.G., Ž.K., P.R., L.V.G., V.B.D., C.P., J.Kr. provided and analysed archaeological material. M.H., Q.F., A.H., M.P., F.M., S.G., P.Sk. and V.N. analysed ancient DNA data. M.H., M.M., S.P. and J.Ke. wrote the manuscript with the input of all co-authors.

Author information: Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper.

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

individuals has been limited by poor preservation of endogenous DNA3 and large amounts of microbial and present-day human DNA that contaminate Neandertal skeletal remains3–5. Here we use hypochlorite treatment6 of as little as 9 mg of bone or tooth powder to generate between 1- and 2.7-fold genomic coverage of five 39,000- to 47,000-year-old Neandertals (*i.e.* late Neandertals), thereby doubling the number of Neandertals for which genome sequences are available. Genetic similarity among late Neandertals is well predicted by their geographical location, and comparison to the genome of an older Neandertal from the Caucasus2,7 indicates that a population turnover is likely to have occurred, either in the Caucasus or throughout Europe, towards the end of Neandertal history. We find that the bulk of Neandertal gene flow into early modern humans originated from one or more source populations that diverged from the Neandertals studied here at least 70,000 years ago, but after they split from a previously sequenced Neandertal from Siberia2 ~150,000 years ago. Although four of these Neandertals post-date the putative arrival of early modern humans into Europe, we do not detect any recent gene flow from early modern humans in their ancestry.

The Middle to Upper Palaeolithic transition in Europe was characterized by major cultural and biological changes that coincided with the arrival of early modern humans and the disappearance of Neandertals8,9. Analysis of the first Neandertal genomes provided evidence for gene flow from Neandertals into modern humans between 50,000 and 60,000 years ago, resulting in around 2% of Neandertal DNA in the genomes of non-Africans today1,2,10. Additionally, genetic analyses of a ~39,000-42,000-year-old modern human from Romania, *Oase 1*, showed that interbreeding between Neandertals and modern humans also happened in Europe at a later point in time11. However, little is known about the diversity of late Neandertal populations across Europe and West Asia shortly before their disappearance, or about their relationship to the population that admixed with early modern humans. To date, only a handful of Neandertal remains have been identified with a sufficiently high content of endogenous DNA and low enough levels of microbial and human DNA contamination3 to allow analysis of larger parts of their genomes1,2,7, limiting our ability to study their genetic history.

In an attempt to make more Neandertal genomes available for population analyses, we identified five specimens with sufficient preservation of endogenous DNA to explore the possibility of nuclear genome sequencing (Fig. 1A and Supplementary Information 1): a fragment of a right femur (*Goyet Q56-1*) dated to 43,000-42,080 cal BP from the Troisième caverne of Goyet in Belgium12; an upper right molar (*Spy 94a*) that is associated with a maxillary fragment dated to 39,150-37,880 cal BP from the neighbouring Spy cave13; a tooth (*Les Cottés Z4-1514*) dated to 43,740-42,720 cal BP from Les Cottés cave14 in France; an undiagnostic bone fragment found in Vindija cave in Croatia (*Vindija 87*) dated to >44,000 uncalibrated BP; and a skull fragment of an infant from Mezmaiskaya cave (*Mezmaiskaya 2*) in the Russian Caucasus dated to 44,600-42,960 cal BP15.

We extracted DNA from between 9 and 58 mg of bone or tooth powder. Approximately half of the powder from each specimen was treated with 0.5% hypochlorite solution prior to DNA extraction in order to remove present-day human and microbial DNA contamination6. Hypochlorite treatment increased the proportion of DNA fragments mapping to the human

reference genome between 5.6- and 161-fold (Fig. 1B; Supplementary Information 2), and reduced present-day human contamination in four of the specimens between 2- and 18-fold (Fig. 1C; Supplementary Information 2). This substantial increase in the proportion of informative fragments made whole genome sequencing of these previously inaccessible specimens feasible.

We generated additional single-stranded DNA libraries from selected extracts and sequenced them to an average genomic coverage of 2.7-fold for *Les Cottés Z4-1514*, 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* (Extended Data Table 1; Supplementary Information 3). Estimates of mitochondrial and nuclear DNA contamination ranged from 0.52% to 5.06% and from 0.18% to 1.75%, respectively (Supplementary Information 4). Therefore, we restricted analyses to fragments carrying cytosine (C) to thymine (T) substitutions at their ends as these derive from deamination of cytosine to uracil and indicate that DNA molecules are of ancient origin16,17 (Extended Data Fig. 1 and 2). This reduced the mtDNA contamination to 0.39-1.61% and the autosomal contamination to 0-0.81% (Supplementary Information 4). To mitigate the influence of deamination on genetic inferences, we further restricted the analyses to only transversion polymorphisms (see Supplementary Information 6).

A phylogenetic tree (Fig. 2A) of the reconstructed complete mitochondrial genomes places these five specimens within the Neandertal mtDNA variation. The mtDNA of *Les Cottés Z4-1514* is more closely related to the mtDNAs of Neandertals from Okladnikov cave and Denisova cave located at the Eastern edge of the Neandertal range (Supplementary Information 5), whereas the mtDNA of *Mezmaiskaya 2* is most closely related to *Feldhofer 2* from Germany, challenging the previously proposed division between Eastern and Western mtDNAs in late-surviving Neandertals18. As inferred from the sequence coverage of the X chromosome and the autosomes, *Goyet Q56-1*, *Les Cottés Z4-1514* and *Vindija 87* were females, whereas *Mezmaiskaya 2* and *Spy 94a* were males (Extended Data Fig. 3). The Y chromosome sequences of both male individuals fall outside the known variation of present-day human Y chromosomes (Fig. 2B; Supplementary Information 5), as is the case for the Y chromosome of a Neandertal from El Sidrón, Spain19.

We analysed the genomes of the five late Neandertals with the previously published high-quality genomes of a ~120,000-year-old Neandertal from Siberia (*Altai* Neandertal)2 and a >45,000-year-old Neandertal from Croatia (*Vindija 33.19*)7, the low-coverage genome of a ~60,000-70,000-year-old Neandertal from Mezmaiskaya cave (*Mezmaiskaya 1*)2,7, the composite low-coverage genome of three Neandertals from Croatia1, the high-coverage genome of a *Denisovan* individual20, and a world-wide panel of present-day humans2,21. Based on sharing of derived alleles with the *Vindija 33.19* genome7, we find that the *Vindija 87* specimen originates from the same individual as *Vindija 33.19*. We therefore excluded *Vindija 87* from subsequent analyses (Supplementary Information 7).

A neighbour-joining tree based on the number of pairwise transversions between individuals shows that all Neandertals form a monophyletic clade relative to the *Denisovan* individual (Fig. 2C; Supplementary Information 7 and 8). The tree reflects an apparent age-related division among the Neandertals with the oldest specimen, the *Altai* Neandertal branching off

first, followed by *Mezmaiskaya 1*, while the late Neandertals form a clade. To assess the relationship of the late Neandertals to the high coverage genomes of the *Altai* and *Vindija 33.19* Neandertals, we used two related statistics that measure the fraction of derived allele-sharing between the genomes. We find that all late Neandertals and *Mezmaiskaya 1* share significantly more derived alleles with *Vindija 33.19* than with the *Altai* Neandertal ($-32.1 \leq Z \leq -58.1$; Extended Data Table 2; Supplementary Information 9) and that late Neandertals share on average 49.0% (95% CI: 44.2-54.2%) of the derived alleles separating *Vindija 33.19* from other high-coverage genomes in the analysis, whereas they share 17.2% (95% CI: 15.9-18.3%) of the alleles derived in the *Altai* Neandertal (Extended Data Table 3; Supplementary Information 7).

Obtaining genome-wide data of multiple late Neandertals from a broad geographical range allowed us to determine for the first time whether relatedness among Neandertals is correlated to their geographical proximity, as is the case for present-day humans22. In support of this, we find that the two Neandertals from Belgium share more derived alleles with each other than with any other Neandertal ($-3.65 \leq Z \leq -8.47$; Supplementary Information 9), and in turn more derived alleles with Neandertals from France and Croatia than with the late Neandertal from the Caucasus. Similarly, the four Neandertals from Vindija cave that come from a relatively narrow time range share more derived alleles with each other than with other Neandertals ($-2.2 \leq Z \leq -14.5$; Supplementary Information 9). Furthermore, specimens of similar age with the largest geographic distance between them (*Les Cottés Z4-1514* and *Mezmaiskaya 2*) shared the fewest derived alleles (Supplementary Information 9). In contrast, *Mezmaiskaya 2* shared more derived alleles with the other late Neandertals than with *Mezmaiskaya 1* ($-2.13 \leq Z \leq -9.56$; Supplementary Information 9), suggesting that there was a population turnover towards the end of Neandertal history. This turnover may have been the result of a population related to Western Neandertals replacing earlier Neandertals in the Caucasus, or the replacement of Neandertals in the West by a population related to *Mezmaiskaya 2*. The timing of this turnover coincides with dramatic climatic fluctuations during Marine Isotope Stage 3 between 60,000 and 24,000 years ago23, when extreme cold periods in Northern Europe may have triggered local extinction of Neandertal populations and subsequent re-colonization from refugia in South Europe or Western Asia24,25.

We estimated the population split times between each of the low-coverage Neandertals and the two high-coverage Neandertals by determining the fraction of sites at which each of the low coverage Neandertals shares a derived allele that occurs in the heterozygote state in one of the high-quality genomes (F(A|B) statistics1,20). This fraction was then used to estimate the population split times for each pair of Neandertals using previous inferences of how the Neandertal population sizes changed over time2,7. Due to the uncertainties in the mutation rate and generation times, we caution that while the times presented are likely to accurately reflect the relative ages of the population split times, the absolute estimates in years are approximate. We estimate that all late Neandertals separated from a common ancestor with the *Altai* Neandertal ~150,000 years ago (kya) (95% CI: 142-186 kya), and from a common ancestor with *Vindija 33.19* ~70 kya (95% CI: 58-72 kya; Extended Data Table 4, Supplementary Information 8). The estimates of the population split times from the common ancestors shared with the *Denisovan* and with modern humans are ~400 kya (95% CI:

367-484 kya) and ~530 kya (95% CI: 503-565 kya; Extended Data Table 4, Supplementary Information 8), respectively, consistent with previous estimates using the *Altai* and *Vindija 33.19* Neandertal genomes1,2,20.

To investigate whether any of the Neandertals sequenced to date is more closely related to the Neandertal population that contributed genetic material to modern humans, we compared the Neandertal genomes from this and previous studies to the genomes of 263 present-day humans21 as well as a number of early modern humans10,11,26,27. We find that all late Neandertals and the older *Mezmaiskaya 1* Neandertal share significantly more derived alleles with the introgressing Neandertals than the *Altai* Neandertal does ($-2.4 \leq Z \leq -5.6$; Fig. 3A and Supplementary Information 10), with no significant differences among them ($-0.1 \leq Z \leq 1.8$; Fig. 3B; Supplementary Information 10). Interestingly, this is also true for a ~45,000-year-old modern human from Siberia (*Ust'-Ishim*) (Fig. 3; Supplementary Information 10), who was contemporaneous with late Neandertals, but is not a direct ancestor of present-day humans10. Thus, the majority of gene flow into early modern humans appears to have originated from one or more Neandertal populations that diverged from other late Neandertals after their split from the *Altai* Neandertal about 150 kya, but before the split from *Mezmaiskaya 1* at least 90 kya (Extended Data Table 4). Due to the paucity of overlapping genetic data from *Oase 1*, whose genome revealed an unusually high percentage of Neandertal ancestry11, we were unable to resolve whether one of these late Neandertals was significantly closer than others to the introgressing Neandertal in *Oase 1*.

Interbreeding between Neandertals and early modern humans is likely to have occurred intermittently, presumably resulting in gene flow in both directions28. However, when we applied an approach that utilizes the extended length of haplotypes expected from recent introgression to the late Neandertals analysed, we find no indications of recent gene flow from early modern humans (Supplementary Information 11). We caution that given the small number of Neandertals analysed we cannot exclude that such gene flow occurred. However, it is striking that *Oase 1*, one of the two early modern humans that overlapped in time with late Neandertals, showed evidence for recent additional Neandertal introgression10,11 whereas none of the late Neandertals analysed here do. This may indicate that gene flow affected the ancestry of modern human populations more than it did Neandertals29. Further work is necessary to determine if this was the case. Our work demonstrates that the generation of genome sequences from a large number of archaic human individuals is now technically feasible, and opens the possibility to study Neandertal populations across their temporal and geographical range.

Online Methods

DNA extraction and library preparation

All specimens were sampled in clean room facilities dedicated to the analysis of ancient DNA. Between 28 mg and 104 mg of tooth or bone powder was obtained by drilling once into the physically cleaned part of the specimen and split evenly (Supplementary Information 2; Table S2.1). Approximately half of the powder was directly subjected to DNA extraction using a silica-based method31 as implemented by Korlević *et al.*6 ("untreated" sample), whereas the second half was treated with 0.5% sodium hypochlorite

solution6 prior to the DNA extraction in an attempt to remove some of the microbial and present-day human DNA contamination3–5. Five or ten μL of each extract were converted into single-stranded DNA libraries32 with the modifications as in Korlevic *et al.*6. The initial libraries of *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)15,20 to excise uracils, while all other libraries were prepared without this enzymatic treatment (Supplementary Information 2). The libraries were amplified into plateau33 and tagged with two sample-specific indices6,34. An aliquot of each amplified library was additionally enriched for the hominin mitochondrial DNA using a bead-based hybridization method and modern human mtDNA as a bait35–37. After analysing the first set of libraries, we selected the extracts with the highest proportion of endogenous DNA and the lowest levels of present-day human DNA contamination to produce final set of single-stranded DNA libraries (Supplementary Information 2; Table S2.6)6,32.

Genome sequencing and data processing

All libraries were initially sequenced together on Illumina's MiSeq and HiSeq 2500 platforms to determine their suitability for whole genome sequencing. Subsequently, 23 libraries from five Neandertal specimens were selected and sequenced on 50 lanes of the Illumina HiSeq 2500 platform in rapid mode, using double index configuration (2x76 bp)34 (Supplementary Information 2). Base calling was done using *Bustard* (Illumina) for the MiSeq runs and *FreeIbis*38 for the HiSeq runs. Adapters were trimmed and overlapping paired-end reads were merged into single sequences using *leeHom*39. The Burrows-Wheeler Aligner (BWA, version: 0.5.10-evan.9-1-g44db244; https://github.com/mpieva/network-aware-bwa)40 was used to align the shotgun data to the modified human reference *GRCh37* (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/) and to align the human mitochondrial DNA capture data to the revised Cambridge Reference Sequence (NC_01290) with parameters adjusted for ancient DNA ("-n 0.01 –o 2 –l 16500")20. We developed a two-step algorithm called *jivebunny* (https://bioinf.eva.mpg.de/jivebunny) for de-multiplexing of the sequencing runs (Supplementary Information 3), retaining only fragments that were assigned to the correct library based on their index sequences for all of the downstream analyses. PCR duplicates were removed using *bam-rmdup* (version: 0.6.3; https://bitbucket.org/ustenzel/biohazard) and fragments were filtered for read length (≥ 35 bp) and mapping quality (MQ≥25) using SAMtools (version: 1.3.1)41.

The deamination of cytosine (C) to uracil (U) residues leaves characteristic C-to-T substitution patterns in ancient DNA molecules, which are particularly close to the alignment ends15 and thus provide evidence for the presence of authentic ancient DNA in specimens5,42 (Supplementary Information 2 and 3; Extended Data Fig. 1). We evaluated the frequency of these nucleotide substitution patterns characteristic of ancient DNA using an in-house Perl script. For determining the 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.*2) and divided it by the total length of those regions. We determined the sex of

the Neandertal individuals by counting the number of fragments that aligned to the X chromosome and the autosomes (Extended Data Fig. 3).

Contamination estimates

We used four methods to estimate the proportion of present-day human DNA contamination in the final dataset (Supplementary Information 4). We estimated the proportion of mtDNA contamination by present-day human DNA 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 BWA40. We then counted the number of fragments overlapping 63 positions where 18 published Neandertal mitochondrial genomes1,2,12,43–46 differ from those of 311 present-day humans43. In the second approach, we determined positions in the reconstructed mtDNA genomes of each of the five Neandertals that are specific for that Neandertal when compared to 311 present-day humans. To mitigate the effect of deamination for both approaches we ignored the alignments on the forward or reverse strands at positions where the informative base was a C or a G17.

We estimated the extent of present-day human DNA contamination on the autosomes in the five late Neandertals using a maximum likelihood approach described in Green *et al.*1, based on all sites covering informative positions in the nuclear genome where humans carry a fixed derived variant when compared to the great apes. Additionally, we estimated levels of present-day human DNA contamination using an ancestry model where each low-coverage Neandertal traces a portion of its genome either from a high-coverage uncontaminated Neandertal, or from a modern human population related to present-day non-Africans. We built a two-source *qpAdm*47 model where one part of the ancestry was modelled as being closely related to the high-coverage *Vindija 33.19* Neandertal, and the other source of ancestry was modelled as being most closely related to the Dinka population of the Simons Genome Diversity Project (SGDP)21. We estimate the proportion of male contamination for the three female individuals, *Les Cottés Z4-1514, Goyet Q56-1* and *Vindija 87* by counting the number of fragments aligning to the unique regions of the Y chromosome and divided it with the number of fragments that would be expected if the individual was a male (Supplementary Information 4; Table S4.1).

mtDNA and Y chromosomes of Neandertals

We re-aligned the shotgun data of all HiSeq runs to the *Vindija 33.16* mitochondrial genome (AM948965)43 using BWA40 to reconstruct complete mitochondrial genomes of five Neandertals. A consensus base was called at each position covered by at least three fragments and where at least two-thirds of fragments had an identical base17 (Supplementary Information 5). We aligned the reconstructed mitochondrial genomes of the five late Neandertals to the mtDNA genomes of 18 Neandertals2,12,43–46,48,49, 311 present-day humans43, 10 ancient modern humans5,10,36,50–52, three Denisovans53–55, a hominin from Sima de los Huesos17 and a chimpanzee56 using MAFFT57 and calculated the number of pairwise differences among mitochondrial genomes using MEGA658. Bayesian phylogenetic analysis was performed using BEAST v2.4.559 to infer the time to the most recent common ancestor (TMRCA) of all Neandertal mitochondrial genomes (Supplementary Information 5). Radiocarbon dates of Neandertals and ancient modern

humans were used as calibration points for the molecular clock and we used *jModelTest2*60 and marginal likelihood estimation (MLE) analysis61,62 in order to choose the best fitting substitution, clock, and tree model (Supplementary Information 5).

For reconstructing the phylogeny of the Neandertal Y chromosomes, we compared the two male Neandertal individuals, *Mezmaiskaya 2* and *Spy 94a,* to 175 present-day human males from the SGDP21 and two present-day humans with haplogroup A0030. For processing the Y chromosomal data of Neandertal individuals presented here, we followed the processing of the previously published Neandertal Y chromosome analysis19 and used the described parameters to call bases and infer genotypes (Supplementary Information 5).

Set of filters for nuclear data analyses

As the data presented in this study were of low-coverage and contained elevated C-to-T substitutions, we investigated spurious correlations stemming from the properties of the data using the *D-statistics*1,54,63 (Supplementary Information 6). All of the sampling schemes, except random read sampling and then restricting to transversion polymorphisms, as well as simulation of UDG treatment, resulted in significant spurious correlations between individuals (Supplementary Information 6; Table S6.1). Therefore, we applied random read sampling at each position in the genome that was covered for the low-coverage individuals. To diminish the impact of present-day human DNA contamination and enrich for the endogenous fragments17, 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. The newly generated sequencing data of *Mezmaiskaya 17,* a ~60,000-70,000-year-old Neandertal from Russia, were processed in exactly the same way as the data of the other low-coverage individuals. For the high-coverage genomes of *Altai*2 and *Vindija 33.19*7 Neandertals, as well as the *Denisovan*20 individual, *snpAD* genotype calls (http://cdna.eva.mpg.de/neandertal/Vindija/VCF/) were used. For comparison to modern humans, we used *hetfa* files of 263 present-day human genomes of the SGDP21 and *snpAD* genotype calls of three high quality ancient modern humans, *Ust'-Ishim*, a ~45,000-year-old modern human from Siberia10; *Loschbour*, a ~8,000-year-old hunter-gatherer from Luxembourg26; and *LBK*, a ~7,000-year-old farmer from Stuttgart26. Variant sites across all genomes were extracted (https://bitbucket.org/ustenzel/heffalump) and converted into an input format for *AdmixTools* (version 4.1)63 or exported into combined VCF files. We further restricted all analyses to bi-allelic sites in the genome covered by at least one low-coverage Neandertal and to transversion polymorphisms.

Principal Component Analysis

We carried out a Principal Component Analysis (PCA)64,65 using genomes of *Vindija 33.19*7, *Altai* Neandertal2 and *Denisovan* individual20 to estimate the eigenvectors of the genetic variation and then projected the low-coverage Neandertals onto the defined plane. This allowed us to explore the relationship of low-coverage Neandertals relative to the high-coverage Neandertals (Extended Data Fig. 4).

Europe PMC Funders Author Manuscripts

Lineage attribution and average sequence divergence of late Neandertals

We followed an approach introduced in Meyer *et al.*66 based on the sharing of derived alleles with a certain hominin group 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* are most closely related to. We investigated the state of DNA fragments overlapping the positions at which the high-coverage genomes of the *Altai* Neandertal2, the *Vindija 33.19* Neandertal7, the *Denisovan* individual20 and a present-day African (Mbuti, HGDP00982)2 differ from those of the great apes (chimpanzee, bonobo, gorilla and orangutan). We then determined 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 genomes66 (Extended Data Table 3; Supplementary Information 7; Table S7.1).

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 the *Altai* and *Vindija 33.19* Neandertals, the *Denisovan* individual, or a present-day human from the B-panel of Prüfer *et al.*2, using the triangulation method previously applied to a number of ancient genomes1,2,20,54,55. We calculated how many of the substitutions inferred to have occurred from the human-chimpanzee ancestor to the high-coverage genomes occurred after the split from the low-coverage genome (Supplementary Information 7; Tables S7.2 and S7.3). Standard errors were computed by a Weighted Block Jackknife67 with a block size of 5 million base pairs (5 Mb) across all autosomes.

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

We conditioned on the heterozygous sites in the two high-coverage Neandertals, and then computed the fraction of sites that show the same derived allele in randomly sampled fragments of the low-coverage individuals in order to estimate the split times between the low-coverage Neandertals and the *Altai*2 and *Vindija 33.19*7 Neandertals (Extended Data Table 4; Supplementary Information 8). In this F(A|B) statistic1,2,20, expected value depends only on the demography of the high-coverage individuals from which heterozygous sites are determined. We performed these analyses on all randomly sampled fragments, and on deaminated fragments only (Extended Data Table 4; Supplementary Information 8).

We used the low-coverage nuclear genomes of *Les Cottés Z4-1514, Goyet Q56-1, Spy 94a, Mezmaiskaya 2* and *Mezmaiskaya 1*, the high-coverage genomes of *Vindija 33.19, Altai Neandertal, Denisovan* individual and 12 present-day humans from Prüfer *et al.*2 for constructing a neighbour-joining tree. We counted the total number of transversions between all pairs of individuals and the human-chimpanzee common ancestor2,54. We constructed a neighbour-joining tree68 based on the pairwise number of transversions in windows of 5 Mb across all autosomes between all individuals with 1000 bootstrap replications (Supplementary Information 8). The tree was constructed as implemented in the R-package *phangorn*69 and visualized with *FigTree* (version: v1.4.2) (http://tree.bio.ed.ac.uk/software/figtree/).

Europe PMC Funders Author Manuscripts

Inferring the relationships between Neandertals

We used *D-statistics*1,54,63 to investigate the population relationships among Neandertal individuals. We co-analysed all low-coverage Neandertal genomes with the high-coverage genomes of the *Altai2* and *Vindija 33.19*7 Neandertals. We used the whole genome alignments of the chimpanzee, orangutan and rhesus macaque to the human reference genome70–72 to infer the ancestral states for the analyses of *D(Altai, Vindija 33.19; Neandertal, outgroup).* Furthermore, we used the genomes of the Dinka and Mbuti individuals from the SGDP21 as outgroups for the statistics of $D(Neandertal_1, Neandertal_2, Neandertal_3, outgroup)$. The standard errors were computed using a Weighted Block Jackknife63,67 with equally sized blocks of 5 Mb over all autosomes. We further restricted these analyses to bi-allelic sites in the genome covered by at least one low-coverage Neandertal and transversion polymorphisms.

Inferring the relationship to the introgressed Neandertals in present-day and ancient modern humans

We analysed the low-coverage late Neandertal genomes together with the high-coverage genomes of the *Altai*2 and *Vindija 33.19* Neandertals7, the high-coverage genomes of the *Denisovan* individual20 and 263 present-day humans of the SGDP21 (Supplementary Information 10). We included *Ust'-Ishim*, a ~45,000-year-old modern human from Siberia10; *Loschbour*, a ~8,000-year-old hunter-gatherer from Luxembourg26; and *LBK*, a ~7,000-year-old farmer from Stuttgart26 to study the differences among Neandertals in their proximity to the introgressed Neandertal DNA detected in ancient modern humans. Analyses were restricted to transversion polymorphisms and to bi-allelic sites in the genome covered by at least one low-coverage Neandertal. As above, the *D-statistics* was used to infer the relationships between individuals1,54,63 and standard errors were computed using a Weighted Block Jackknife63,67 over all autosomes (block size: 5 Mb).

Early modern human gene flow into late Neandertals

We modelled admixture from modern humans into Neandertals with an ascertainment scheme in which both the *Denisova* and *Altai* Neanderal were fixed for the ancestral allele and at least half of the alleles in present-day African populations are derived. We applied the method as described previously73 and estimated a date of a recent early modern human (EMH) admixture into Neandertals to be ~10-100 generations ago. At each SNP in the genome, we considered data from all Yoruba individuals from the 1000 Genomes Project74 covered by at least three fragments that passed a pre-defined set of filters. Furthermore, we restricted the analysis to sites in the genome where ≥24 Yoruba individuals as well as the *Altai* Neandertal and *Denisovan* individual had allele calls (Map35_50% filter from Prüfer *et al.*2). The ancestral states were taken from the inferred ancestor of humans and chimpanzees (Ensembl Compara v64)75,76. We introduced a more complex demographic history that is loosely based on the model described in Gravel *et al.*77 (details in the Supplementary Information 11).

Europe PMC Funders Author Manuscripts

Data availability

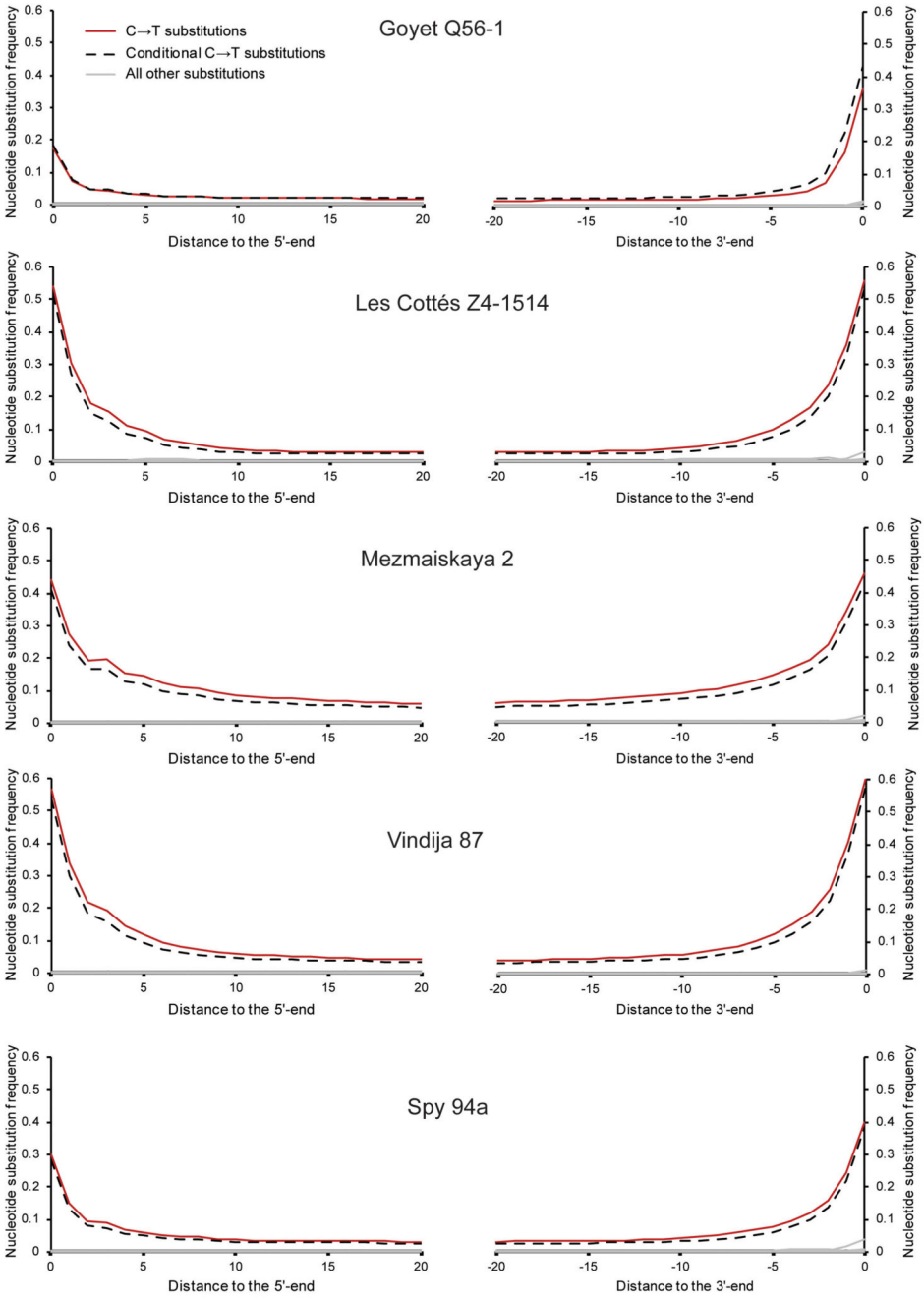
The aligned sequences have been deposited in the European Nucleotide Archive under accession numbers PRJEB21870 (*Goyet Q56-1*), PRJEB21875 (*Les Cottés Z4-1514*), PRJEB21881 (*Mezmaiskaya 2*), PRJEB21882 (*Vindija 87*) and PRJEB21883 (*Spy 94a*). The mitochondrial consensus sequences reported in this paper are available in GenBank with the accession codes MG025536-MG025540.

Code availability

All software packages used for analysis are cited and all packages are publicly available. Algorithms and packages are freely available at https://bioinf.eva.mpg.de/.

Extended Data

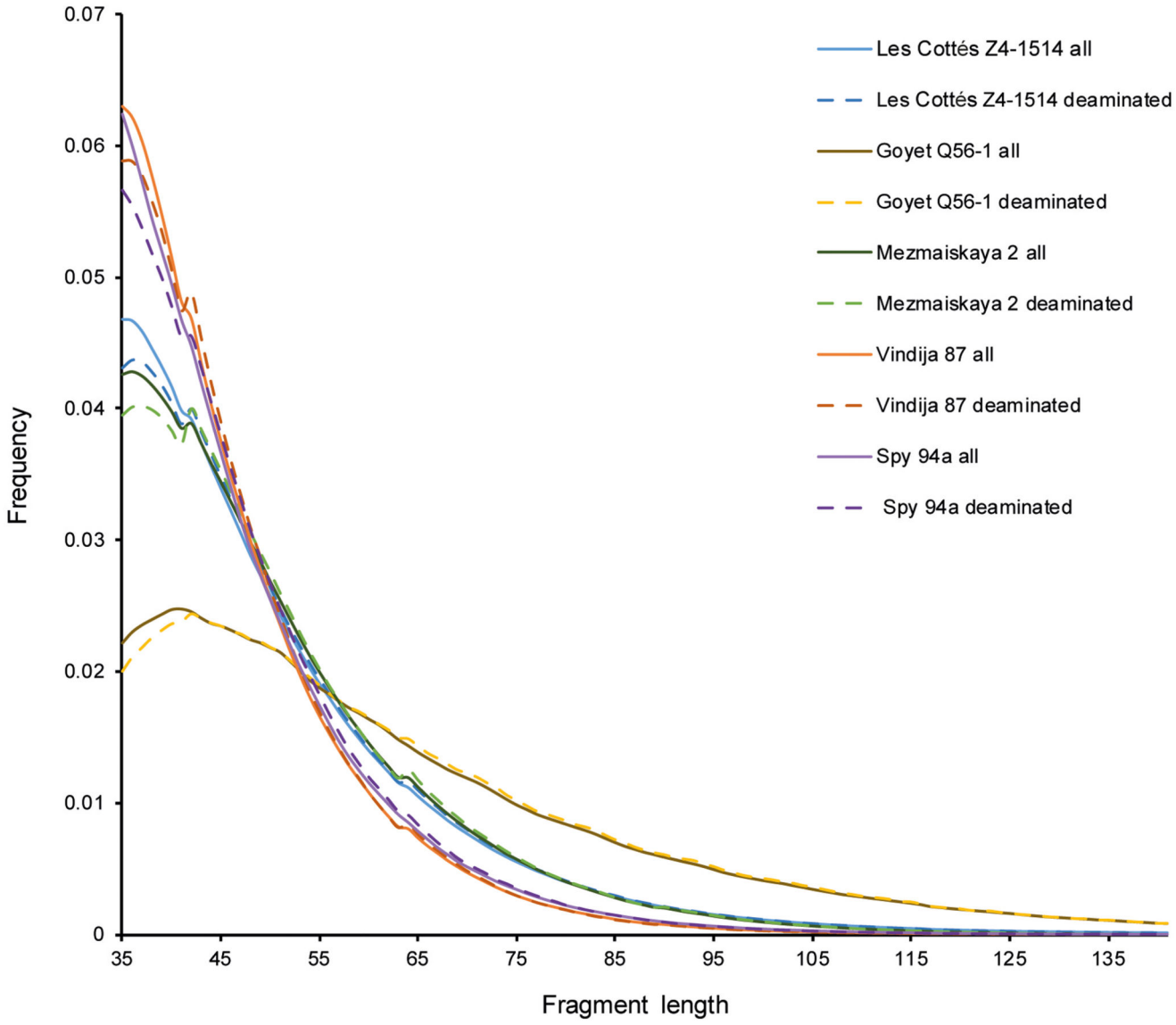
Europe PMC Funders Author Manuscripts



Extended Data Figure 1. Frequency of nucleotide substitutions at the beginning and the end of nuclear alignments for the final dataset of *Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87* and *Spy 94a*.

Only fragments of at least 35 base pairs *(bp)* that mapped to the human reference genome with a mapping quality of at least 25 (MQ ≥ 25) were used for this analysis. Solid lines depict all fragments and dashed lines the fragments that have a C-to-T substitution at the opposing end ('conditional' C-to-T substitutions). All other types of substitutions are marked in grey.

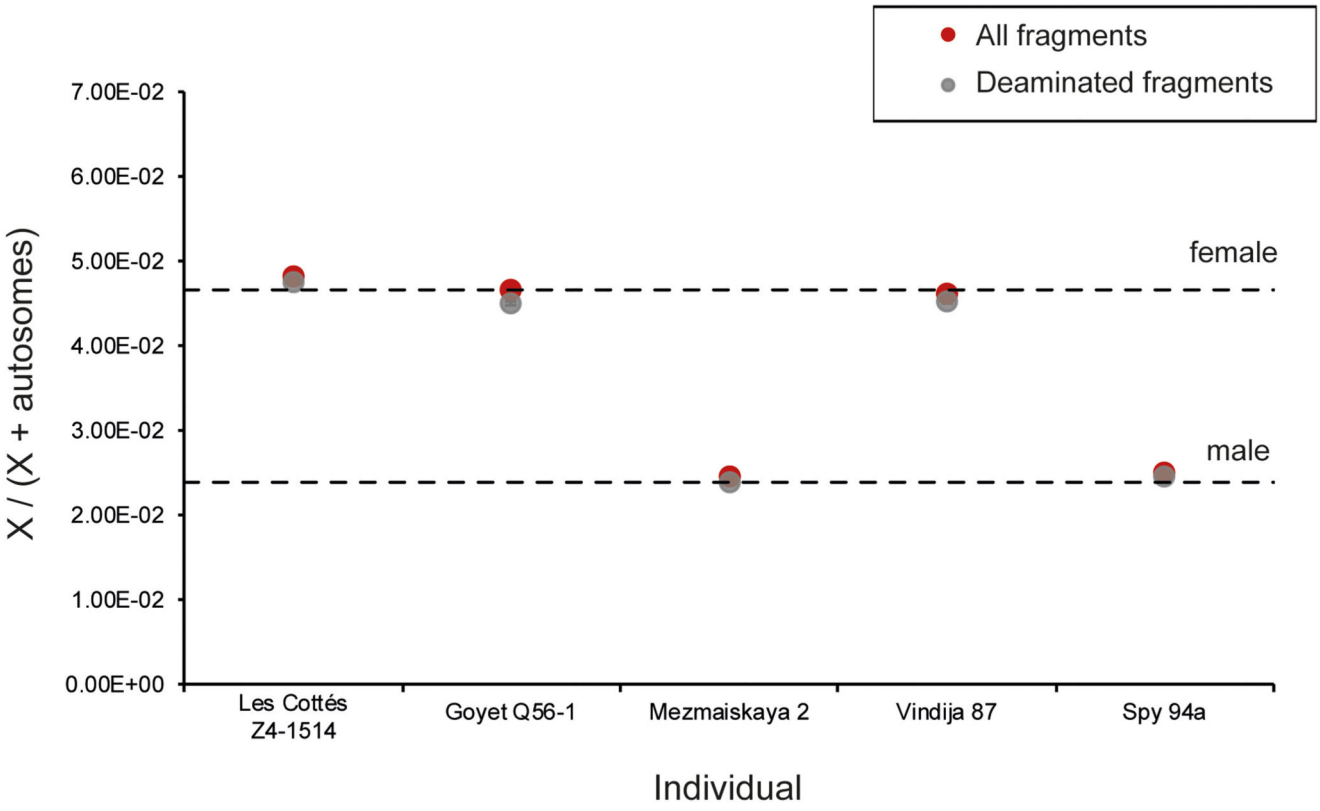
Europe PMC Funders Author Manuscripts



Extended Data Figure 2. Fragment size distribution of fragments longer than 35 base pairs (bp) mapped to the human reference genome with the mapping quality of at least 25 (MQ≥25) for each of the five late Neandertals.

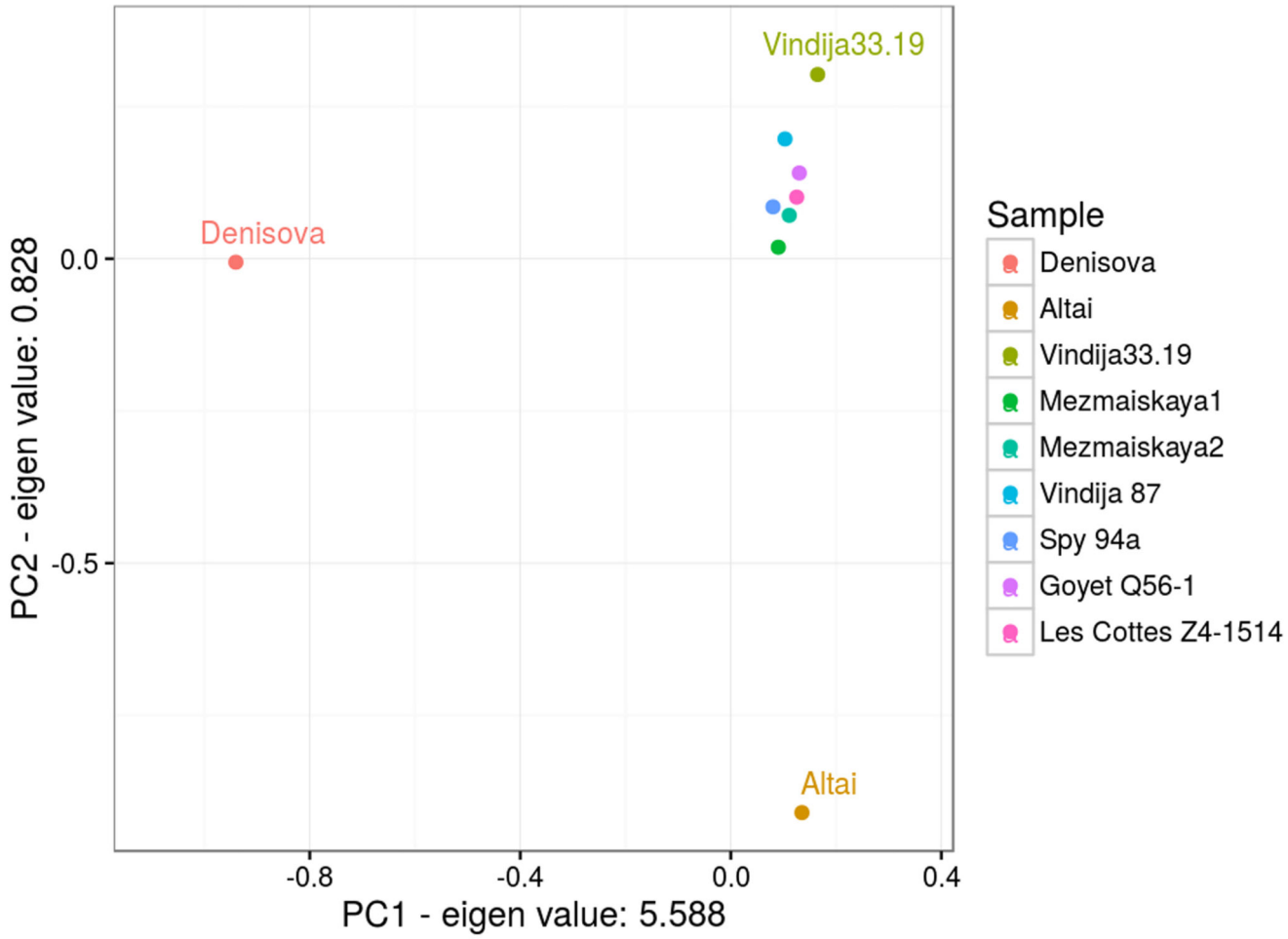
All fragments are depicted in solid lines and fragments with C-to-T substitutions to the reference genome (putatively deaminated fragments) are depicted with dashed lines.

Europe PMC Funders Author Manuscripts



Extended Data Figure 3. Sex determination based on the number of fragments aligning to the X chromosome and the autosomes.

The expected ratios of X to (X + autosomal) fragments for a female and a male individual are depicted as dashed lines. The results were concordant for all fragments (in red) and for deaminated fragments only (in grey).



Extended Data Figure 4. A Principal Component Analysis (PCA) of the genomes of *Vindija 33.19, Altai, Denisova*, five late Neandertals and *Mezmaiskaya 1*.

Genomes of the high coverage archaics were used to estimate the eigenvectors of the genetic variation and low coverage Neandertals were projected onto the plane. Only transversion polymorphisms and bi-allelic sites were considered for the analysis, amounting to a total of 1,010,417 sites as defined by the high coverage genomes.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Extended Data Table 1

Amount of data generated for *Les Cottés Z4-1514, Goyet Q56-1, Mezmaiskaya 2, Vindija 87* and *Spy 94a*.

Reported is the number of fragments after merging all of the sequencing libraries together. The obtained coverage of the nuclear genomes is determined by counting the number of bases with the base quality of at least 30 in fragments longer than 35 bp with the mapping quality of at least 25 that overlapped highly mappable regions of autosomes of the human genome and dividing that number by the total length of these regions.

Specimen	All fragments					Fragments with terminal C-to-T substitutions to the reference genome		
	Number of sequenced fragments	Number of fragments ≥ 35 bp	Number of mapped fragments ≥ 35 bp, MQ ≥ 25, Map35_100%	Number of unique fragments ≥ 35 bp, MQ ≥ 25, Map35_100%	Obtained nuclear coverage	Number of mapped fragments ≥ 35 bp, MQ ≥ 25, Map35_100%	Number of unique fragments ≥ 35 bp, MQ ≥ 25, Map35_100%	Obtained nuclear coverage
Les Cottés Z4-1514	1,776,108,129	978,981,412	253,134,472	121,336,498	2.71	59,777,476	41,804,332	1.00
Goyet Q56-1	917,798,995	575,597,226	124,582,445	80,515,019	2.18	13,585,646	9,859,280	0.27
Mezmaikaya 2	2,571,128,011	1,356,302,134	125,815,832	74,407,074	1.74	32,838,281	23,521,855	0.56
Vindija 87	1,353,546,508	678,005,556	110,385,510	60,034,976	1.27	32,597,516	22,828,056	0.49
Spy 94a	733,212,063	278,595,622	72,962,863	46,836,553	1.00	14,301,192	10,003,560	0.22

Europe PMC Funders Author Manuscripts

Extended Data Table 2

Relationship of the late Neandertals and *Mezmaiskaya 1* to the *Altai* and *Vindija 33.19* calculated as *D (Altai, Vindija 33.19, Neandertal, outgroup)* for all fragments and deaminated fragments, restricted to transversions.

All late Neandertals and *Mezmaiskaya 1* are significantly closer to *Vindija 33.19* than to the *Altai* Neandertal, irrespective of the used outgroup. Blue denotes Z-score << -2. The total number of informative sites that are transversions among Neandertals and where at least one late Neandertal has coverage is 1,567,449.

D (Altai, Vindija 33.19, Neandertal, outgroup)				All fragments				Deaminated fragments			
W	X	Y	Z	% D	Z-score	BABA	ABBA	% D	Z-score	BABA	ABBA
Altai	Vindija33.19	Les Cottés Z4-1514	Dinka	48.87	-55.31	30,276	10,399	49.08	-52.83	19,048	6,506
Altai	Vindija33.19	Les Cottés Z4-1514	Mbuti	49.08	-57.83	30,340	10,362	49.34	-55.25	19,077	6,471
Altai	Vindija33.19	Les Cottés Z4-1514	Chimp	49.16	-59.83	29,427	10,028	49.38	-56.69	18,597	6,302
Altai	Vindija33.19	Les Cottés Z4-1514	Orang	48.88	-60.25	28,149	9,664	49.15	-57.21	17,775	6,060
Altai	Vindija33.19	Les Cottés Z4-1514	Rhesus	48.35	-58.96	25,856	9,003	48.38	-55.77	16,281	5,664
Altai	Vindija33.19	Goyet Q56-1	Dinka	55.87	-66.92	33,208	9,403	56.32	-56.88	8,673	2,423
Altai	Vindija33.19	Goyet Q56-1	Mbuti	56.00	-68.27	33,269	9,383	56.32	-58.37	8,689	2,428
Altai	Vindija33.19	Goyet Q56-1	Chimp	55.96	-68.97	32,188	9,090	56.76	-59.25	8,485	2,340
Altai	Vindija33.19	Goyet Q56-1	Orang	55.68	-68.11	30,735	8,749	56.79	-59.39	8,125	2,239
Altai	Vindija33.19	Goyet Q56-1	Rhesus	55.40	-67.04	28,312	8,125	56.31	-58.09	7,483	2,092
Altai	Vindija33.19	Mezmaiskaya1	Dinka	36.58	-36.63	20,392	9,468	37.09	-32.69	8,379	3,845
Altai	Vindija33.19	Mezmaiskaya1	Mbuti	36.80	-37.33	20,516	9,479	37.43	-33.41	8,438	3,842
Altai	Vindija33.19	Mezmaiskaya1	Chimp	36.70	-38.66	19,860	9,197	37.00	-33.45	8,126	3,736
Altai	Vindija33.19	Mezmaiskaya1	Orang	36.18	-37.38	18,904	8,860	36.77	-32.27	7,758	3,586
Altai	Vindija33.19	Mezmaiskaya1	Rhesus	36.00	-37.77	17,433	8,204	36.36	-32.05	7,129	3,327
Altai	Vindija33.19	Mezmaiskaya2	Dinka	47.44	-42.97	26,802	9,553	46.84	-40.59	12,477	4,516
Altai	Vindija33.19	Mezmaiskaya2	Mbuti	47.83	-47.39	26,781	9,451	47.13	-44.43	12,471	4,481
Altai	Vindija33.19	Mezmaiskaya2	Chimp	47.92	-51.11	26,019	9,160	47.37	-47.06	12,233	4,369
Altai	Vindija33.19	Mezmaiskaya2	Orang	47.86	-53.48	24,921	8,788	47.07	-47.53	11,700	4,210
Altai	Vindija33.19	Mezmaiskaya2	Rhesus	47.33	-49.32	22,908	8,189	46.65	-44.74	10,799	3,929
Altai	Vindija33.19	Spy 94a	Dinka	54.86	-56.62	21,416	6,242	55.17	-50.39	6,558	1,895
Altai	Vindija33.19	Spy 94a	Mbuti	55.18	-59.76	21,453	6,196	55.53	-52.89	6,567	1,878
Altai	Vindija33.19	Spy 94a	Chimp	55.32	-60.58	20,814	5,988	55.39	-50.67	6,354	1,824
Altai	Vindija33.19	Spy 94a	Orang	55.01	-60.26	19,793	5,746	55.33	-50.10	6,052	1,740
Altai	Vindija33.19	Spy 94a	Rhesus	54.46	-58.94	18,179	5,360	54.93	-48.52	5,584	1,624

Europe PMC Funders Author Manuscripts

Extended Data Table 3

The fraction of derived alleles among putatively deaminated fragments that each of the low coverage individuals shares with the *Altai* Neandertal, *Vindija 33.19*, *Denisovan*, and a present day human genome.

The state of DNA fragments overlapping the positions at which the high coverage genomes of the *Altai* Neandertal, the *Vindija 33.19* Neandertal, the *Denisovan* individual and a present-day African (Mbuti, HGDP00982) differ from those of the great apes were investigated. Fragments longer than 35bp with mapping quality of at least 25 and within the highly mappable regions of the genome that had terminal C-to-T substitutions reported in the Table S3.2 were utilized. 95% binomial confidence intervals are provided in brackets.

	Human (%)	Neandertal (%)	Altai Neandertal (%)	Vindija 33.19 (%)	Denisovan (%)	Neandertal-Denisova (%)	Human-Neandertal (%)	Human-Denisova (%)
Les Cottés Z4-1514	0.67 [0.65-0.70]	93.01 [92.89-93.13]	18.00 [17.69-18.32]	46.41 [46.06-46.77]	0.86 [0.82-0.89]	97.90 [97.83-97.98]	97.30 [97.17-97.43]	3.22 [3.07-3.38]
Goyet Q56-1	0.60 [0.56-0.65]	94.38 [94.19-94.57]	16.47 [15.93-17.02]	53.55 [52.91-54.18]	0.80 [0.74-0.86]	98.17 [98.04-98.29]	97.72 [97.49-97.92]	2.52 [2.29-2.78]
Spy 94a	0.71 [0.66-0.77]	93.89 [93.65-94.12]	16.52 [15.90-17.17]	51.45 [50.70-52.21]	0.80 [0.73-0.87]	97.83 [97.66-97.99]	97.62 [97.35-97.87]	3.06 [2.75-3.39]
Vindija 87	0.36 [0.33-0.38]	97.08 [96.97-97.19]	6.81 [6.53-7.09]	75.93 [75.49-76.36]	0.37 [0.34-0.40]	99.05 [98.98-99.12]	98.91 [98.79-99.02]	1.38 [1.24-1.53]
Mezmaiskaya 2	0.65 [0.62-0.69]	92.78 [92.62-92.93]	17.91 [17.51-18.31]	44.65 [44.19-45.12]	0.83 [0.79-0.88]	97.74 [97.63-97.84]	97.53 [97.36-97.68]	3.38 [3.18-3.59]
Mezmaiskaya 1	0.76 [0.72-0.80]	91.87 [91.67-92.06]	20.30 [19.81-20.80]	38.38 [37.85-38.91]	0.97 [0.91-1.03]	97.61 [97.48-97.73]	96.80 [96.58-97.00]	3.73 [3.49-4.00]
Vindija 33.16	3.60 [3.50-3.71]	93.39 [93.18-93.59]	17.09 [16.55-17.65]	58.99 [58.36-59.63]	3.94 [3.81-4.08]	96.14 [95.95-96.32]	95.46 [95.15-95.74]	4.89 [4.56-5.24]
Vindija 33.25	3.04 [2.94-3.15]	94.17 [93.95-94.37]	16.69 [16.10-17.30]	60.81 [60.11-61.49]	3.41 [3.28-3.55]	96.53 [96.33-96.72]	95.79 [95.46-96.09]	3.77 [3.46-4.12]
Vindija 33.26	3.49 [3.38-3.62]	93.54 [93.31-93.77]	17.04 [16.43-17.67]	59.65 [58.94-60.36]	3.70 [3.55-3.85]	96.25 [96.04-96.45]	95.40 [95.04-95.72]	4.56 [4.21-4.95]
Feldhofer 1	4.71 [2.40-9.01]	94.38 [87.51-97.58]	20 [8.86-39.13]	50.00 [34.07-65.93]	1.80 [0.50-6.33]	90.63 [81.02-95.63]	96.00 [80.46-99.29]	0 [0-12.87]
El Sidron 1253	1.29 [0.44-3.72]	90.99 [84.21-95.03]	17.50 [8.75-31.95]	42.86 [30.02-56.73]	4.35 [2.22-8.34]	95.24 [88.39-98.13]	97.44 [86.82-99.55]	2.33 [0.41-12.06]
Denisova 4	2.63 [1.51-4.54]	2.31 [1.06-4.94]	2.13 [0.59-7.43]	2.04 [0.56-7.14]	71.43 [65.87-76.40]	96.15 [91.86-98.23]	10.00 [5.15-18.51]	90.32 [80.45-95.49]
Denisova 8	1.60 [1.39-1.85]	6.75 [6.13-7.43]	1.75 [1.23-2.48]	1.83 [1.37-2.45]	60.02 [58.92-61.12]	92.53 [91.66-93.32]	15.32 [13.72-17.06]	87.72 [86.01-89.25]

Europe PMC Funders Author Manuscripts

Extended Data Table 4

Time of separation of late Neandertals and *Mezmaiskaya 1* (A) from the ancestor with the high-coverage genomes of *Altai* and *Vindija 33.19* Neandertals, *Denisovan* individual and a present-day human (B), when measured in terms of time of split from the B individual (split A-B), or time from present (split-time + branch shortening (bs)).

Results reported for all fragments and deaminated fragments only (Tables S3.1 and S3.2). 95% confidence intervals for F(A|B) values are estimated via Weighted Block Jackknife, using blocks of 5 Mb. In order to obtain estimates of split times from present, we correct "split A-B" by adding the age of the high coverage B-individual estimated from branch-shortening (column split time + bs (kya)).

		All fragments				Deaminated fragments			
A	**B**	**% F(A|B)**	**Split A-B (ky)**	**Split time + bs (kya)**	**95% CI (kya)**	**% F(A|B)**	**Split A-B (ky)**	**Split time + bs (kya)**	**95% CI (kya)**
Les Cottés Z4-1514	Altai	35.2	35.0	157.5	144.6-160.7	34.7	37.8	160.2	156.4-163.5
Goyet Q56-1	Altai	35.8	22.3	144.7	142.8-157.6	35.9	22.2	144.6	141.5-159.3
Mezmaiskaya 2	Altai	35.4	33.8	156.2	143.9-159.7	34.2	40.2	162.6	158.7-166.5
Spy 94a	Altai	34.4	39.2	161.7	158.2-165.2	33.6	43.1	165.5	160.4-186.2
Vindija 87	Altai	35.2	35.4	157.8	144.7-161.1	34.6	38.3	160.8	156.2-164.8
Mezmaiskaya 1	Altai	34.8	37.4	159.9	156.1-163.0	34.4	42.9	165.3	161.1-178.6
Les Cottés Z4-1514	Vindija 33.19	35.3	15.1	66.9	66.1-68.7	34.3	18.3	70.0	67.9-72.4
Goyet Q56-1	Vindija 33.19	37.8	8.3	60.1	58.8-61.7	38.0	8.1	59.9	57.8-61.9
Mezmaiskaya 2	Vindija 33.19	35.6	14.3	66.1	64.3-67.7	34.3	18.3	70.0	67.4-74.5
Spy 94a	Vindija 33.19	36.9	10.4	62.2	60.6-64.3	37.2	9.9	62.7	59.1-64.8
Vindija 87	Vindija 33.19	46.8	0.0	51.8	51.8-51.8	45.0	0.0	61.7	51.8-51.8
Mezmaiskaya 1	Vindija 33.19	31.6	42.0	93.8	81.5-96.9	30.6	46.9	98.7	94.5-102.9
Les Cottés Z4-1514	Denisovan	12.6	333.1	405.1	381.9-448.2	12.2	346.7	418.7	390.4-470.8
Goyet Q56-1	Denisovan	12.8	328.6	400.6	377.3-438.0	13.1	322.4	394.3	367.1-434.0
Mezmaiskaya 2	Denisovan	12.4	338.2	410.2	387.0-458.4	11.9	362.0	434.0	398.9-483.8
Spy 94a	Denisovan	12.5	336.0	407.9	384.7-451.6	12.4	339.4	411.3	382.4-470.8
Vindija 87	Denisovan	12.5	337.1	409.1	385.8-455.0	12.2	346.7	418.7	390.9-470.2

A	B	All fragments				Deaminated fragments			
		% F(A|B)	Split A-B (ky)	Split time + bs (kya)	95% CI (kya)	% F(A|B)	Split A-B (ky)	Split time + bs (kya)	95% CI (kya)
Mezmaiskaya 1	Denisovan	12.3	341.6	413.6	389.2-463.5	12.3	345.0	417.0	388.7-471.4
Les Cottés Z4-1514	Mbuti	17.5	525.9	525.9	511.3-539.8	17.2	549.3	549.3	533.2-565.4
Goyet Q56-1	Mbuti	17.7	515.0	515.0	501.8-528.9	17.6	520.8	520.8	502.6-540.6
Mezmaiskaya 2	Mbuti	17.6	521.6	521.6	506.9-536.2	17.4	535.4	535.4	517.9-553.0
Spy 94a	Mbuti	18.0	502.6	502.6	488.7-517.2	17.3	542.7	542.7	518.6-566.1
Vindija 87	Mbuti	17.6	523.0	523.0	509.1-537.6	17.3	542.0	542.0	524.5-558.8
Mezmaiskaya 1	Mbuti	17.8	512.8	512.8	498.9-528.1	17.1	552.2	552.2	534.7-569.8

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank Antje Weihmann and Barbara Höber for their help with DNA sequencing, Udo Stenzel for computational support and advice for data analysis, Ronny Barr for the help with the graphics, Viviane Slon for helpful discussions and comments on the manuscript. Q.F. is funded in part by National Key R&D Program of China (2016YFE0203700), CAS (XDB13000000, QYZDB-SS W-DQC003, XDPB05), NSFC (91731303, 41672021, 41630102) and the Howard Hughes Medical Institute (grant no. 55008731). D.R. is supported by the U.S. National Science Foundation (grant BCS-1032255) and is an investigator of the Howard Hughes Medical Institute. This study was funded by the Max Planck Society and the European Research Council (grant agreement no. 694707 to S.P.). M.So. thanks Jacques B. owner of Les Cottés and the French Ministry of Culture for financial support and excavation permits.

References

1. Green RE, et al. A draft sequence of the Neandertal genome. Science. 2010; 328:710–722. DOI: 10.1126/science.1188021 [PubMed: 20448178]
2. Prüfer K, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014; 505:43–49. DOI: 10.1038/nature12886 [PubMed: 24352235]
3. Pääbo S, et al. Genetic analyses from ancient DNA. Annu Rev Genet. 2004; 38:645–679. DOI: 10.1146/annurev.genet.37.110801.143214 [PubMed: 15568989]
4. Gilbert MT, Bandelt HJ, Hofreiter M, Barnes I. Assessing ancient DNA studies. Trends Ecol Evol. 2005; 20:541–544. DOI: 10.1016/j.tree.2005.07.005 [PubMed: 16701432]
5. Krause J, et al. A complete mtDNA genome of an early modern human from Kostenki, Russia. Curr Biol. 2010; 20:231–236. DOI: 10.1016/j.cub.2009.11.068 [PubMed: 20045327]
6. Korlević P, et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. Biotechniques. 2015; 59:87–93. DOI: 10.2144/000114320 [PubMed: 26260087]
7. Prüfer K, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science. 2017; doi: 10.1126/science.aao1887
8. Hublin J-J. The modern human colonization of western Eurasia: when and where? Quaternary Science Reviews. 2015; 118:194–210. DOI: 10.1016/j.quascirev.2014.08.011

9. Higham T, et al. The timing and spatiotemporal patterning of Neanderthal disappearance. Nature. 2014; 512:306–309. DOI: 10.1038/nature13621 [PubMed: 25143113]
10. Fu Q, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature. 2014; 514:445–449. DOI: 10.1038/nature13810 [PubMed: 25341783]
11. Fu Q, et al. An early modern human from Romania with a recent Neanderthal ancestor. Nature. 2015; 524:216–219. DOI: 10.1038/nature14558 [PubMed: 26098372]
12. Rougier H, et al. Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. Sci Rep. 2016; 6doi: 10.1038/srep29005
13. Semal P, et al. New data on the late Neandertals: direct dating of the Belgian Spy fossils. Am J Phys Anthropol. 2009; 138:421–428. DOI: 10.1002/ajpa.20954 [PubMed: 19003923]
14. Soressi M, et al. Association des Publications Chauvinoises. 2010
15. Pinhasi R, Higham TF, Golovanova LV, Doronichev VB. Revised age of late Neanderthal occupation and the end of the Middle Paleolithic in the northern Caucasus. Proc Natl Acad Sci U S A. 2011; 108:8611–8616. DOI: 10.1073/pnas.1018938108 [PubMed: 21555570]
16. Briggs AW, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A. 2007; 104:14616–14621. DOI: 10.1073/pnas.0704665104 [PubMed: 17715061]
17. Meyer M, et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature. 2014; 505:403–406. DOI: 10.1038/nature12788 [PubMed: 24305051]
18. Dalen L, et al. Partial genetic turnover in neandertals: continuity in the East and population replacement in the West. Mol Biol Evol. 2012; 29:1893–1897. DOI: 10.1093/molbev/mss074 [PubMed: 22362080]
19. Mendez FL, Poznik GD, Castellano S, Bustamante CD. The Divergence of Neandertal and Modern Human Y Chromosomes. Am J Hum Genet. 2016; 98:728–734. DOI: 10.1016/j.ajhg.2016.02.023 [PubMed: 27058445]
20. Meyer M, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012; 338:222–226. DOI: 10.1126/science.1224344 [PubMed: 22936568]
21. Mallick S, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016; 538:201–206. DOI: 10.1038/nature18964 [PubMed: 27654912]
22. Cavalli-Sforza, LL, Menozzi, P, Piazza, A. The history and geography of human genes. Princeton university press; 1994.
23. Svensson A, et al. A 60 000 year Greenland stratigraphic ice core chronology. Climate of the Past. 2008; 4:47–57.
24. Hublin J-J, Roebroeks W. Ebb and flow or regional extinctions? On the character of Neandertal occupation of northern environments. Comptes Rendus Palevol. 2009; 8:503–509.
25. Müller UC, et al. The role of climate in the spread of modern humans into Europe. Quaternary Science Reviews. 2011; 30:273–279.
26. Lazaridis I, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014; 513:409–413. [PubMed: 25230663]
27. Fu Q, et al. The genetic history of Ice Age Europe. Nature. 2016; 534:200–205. DOI: 10.1038/nature17993 [PubMed: 27135931]
28. Kuhlwilm M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. Nature. 2016; 530:429–433. DOI: 10.1038/nature16544 [PubMed: 26886800]
29. Overmann KA, Coolidge FL. Human species and mating systems: Neandertal–Homo sapiens reproductive isolation and the archaeological and fossil records. Journal of Anthropological Sciences. 2013; 91:91–110. [PubMed: 24344097]
30. Karmin M, et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome research. 2015; 25:459–466. [PubMed: 25770088]
31. 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. 2013; 110:15758–15763. DOI: 10.1073/pnas.1314445110 [PubMed: 24019490]
32. Gansauge MT, Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat Protoc. 2013; 8:737–748. DOI: 10.1038/nprot.2013.038 [PubMed: 23493070]

33. 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. 2012; 52:87–94. DOI: 10.2144/000113809 [PubMed: 22313406]
34. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 2012; 40:e3.doi: 10.1093/nar/gkr771 [PubMed: 22021376]
35. Maricic T, Whitten M, Paabo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One. 2010; 5:e14004.doi: 10.1371/journal.pone.0014004 [PubMed: 21103372]
36. Fu Q, et al. DNA analysis of an early modern human from Tianyuan Cave, China. Proc Natl Acad Sci U S A. 2013; 110:2223–2227. DOI: 10.1073/pnas.1221359110 [PubMed: 23341637]
37. 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. 2016; 113:11162–11167. DOI: 10.1073/pnas.1605834113 [PubMed: 27638212]
38. Renaud G, Kircher M, Stenzel U, Kelso J. freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. Bioinformatics. 2013; 29:1208–1209. DOI: 10.1093/bioinformatics/btt117 [PubMed: 23471300]
39. Renaud G, Stenzel U, Kelso J. leeHom: adaptor trimming and merging for Illumina sequencing reads. Nucleic Acids Res. 2014; 42:e141.doi: 10.1093/nar/gku699 [PubMed: 25100869]
40. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. DOI: 10.1093/bioinformatics/btp698 [PubMed: 20080505]
41. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. DOI: 10.1093/bioinformatics/btp352 [PubMed: 19505943]
42. Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One. 2012; 7:e34131.doi: 10.1371/journal.pone.0034131 [PubMed: 22479540]
43. Green RE, et al. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell. 2008; 134:416–426. DOI: 10.1016/j.cell.2008.06.021 [PubMed: 18692465]
44. Briggs AW, et al. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science. 2009; 325:318–321. DOI: 10.1126/science.1174462 [PubMed: 19608918]
45. Skoglund P, et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc Natl Acad Sci U S A. 2014; 111:2229–2234. DOI: 10.1073/pnas.1318934111 [PubMed: 24469802]
46. Brown S, et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. Sci Rep. 2016; 6doi: 10.1038/srep23559
47. Haak W, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015; 522:207–211. DOI: 10.1038/nature14317 [PubMed: 25731166]
48. Gansauge MT, Meyer M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Res. 2014; 24:1543–1549. DOI: 10.1101/gr.174201.114 [PubMed: 25081630]
49. Posth C, et al. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. Nat Commun. 2017; 8doi: 10.1038/ncomms16046
50. Ermini L, et al. Complete Mitochondrial Genome Sequence of the Tyrolean Iceman. Current Biology. 2008; 18:1687–1693. DOI: 10.1016/j.cub.2008.09.028 [PubMed: 18976917]
51. Gilbert MTP, et al. Paleo-Eskimo mtDNA Genome Reveals Matrilineal Discontinuity in Greenland. Science. 2008; 320:1787. [PubMed: 18511654]
52. Fu Q, et al. A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol. 2013; 23:553–559. DOI: 10.1016/j.cub.2013.02.044 [PubMed: 23523248]
53. Krause J, et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature. 2010; 464:894–897. DOI: 10.1038/nature08976 [PubMed: 20336068]
54. Reich D, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 2010; 468:1053–1060. DOI: 10.1038/nature09710 [PubMed: 21179161]

55. Sawyer S, et al. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. Proc Natl Acad Sci U S A. 2015; 112:15696–15700. DOI: 10.1073/pnas.1519905112 [PubMed: 26630009]
56. Horai S, et al. Man's place in hominoidea revealed by mitochondrial DNA genealogy. Journal of Molecular Evolution. 1992; 35:32–43. DOI: 10.1007/BF00160258 [PubMed: 1518083]
57. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–780. DOI: 10.1093/molbev/mst010 [PubMed: 23329690]
58. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30:2725–2729. DOI: 10.1093/molbev/mst197 [PubMed: 24132122]
59. Bouckaert R, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014; 10:e1003537.doi: 10.1371/journal.pcbi.1003537 [PubMed: 24722319]
60. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and high-performance computing. Nature methods. 2012; 9:772–772. DOI: 10.1038/nmeth.2109
61. Baele G, et al. Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty. Molecular Biology and Evolution. 2012; 29:2157–2167. DOI: 10.1093/molbev/mss084 [PubMed: 22403239]
62. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. Molecular Biology and Evolution. 2013; 30:239–243. DOI: 10.1093/molbev/mss243 [PubMed: 23090976]
63. Patterson N, et al. Ancient admixture in human history. Genetics. 2012; 192:1065–1093. DOI: 10.1534/genetics.112.145037 [PubMed: 22960212]
64. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190.doi: 10.1371/journal.pgen.0020190 [PubMed: 17194218]
65. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]
66. Meyer M, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. Nature. 2016; 531:504–507. DOI: 10.1038/nature17405 [PubMed: 26976447]
67. Busing FMTA, Meijer E, Van Der Leeden R. Delete-m jackknife for unequal m. Statistics and Computing. 1999; 9:3–8. DOI: 10.1023/A:1008800423698
68. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987; 4:406–425. DOI: 10.1093/oxfordjournals.molbev.a040454 [PubMed: 3447015]
69. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011; 27:592–593. DOI: 10.1093/bioinformatics/btq706 [PubMed: 21169378]
70. Chimpanzee S, Analysis C. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005; 437:69–87. DOI: 10.1038/nature04072 [PubMed: 16136131]
71. Locke DP, et al. Comparative and demographic analysis of orang-utan genomes. Nature. 2011; 469:529–533. DOI: 10.1038/nature09687 [PubMed: 21270892]
72. Gibbs RA, et al. Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. Science. 2007; 316:222–234. [PubMed: 17431167]
73. 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. 2016; 113:5652–5657. DOI: 10.1073/pnas.1514696113
74. Genomes Project, C. et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. DOI: 10.1038/nature11632 [PubMed: 23128226]
75. Paten B, Herrero J, Beal K, Fitzgerald S, Birney E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. Genome Res. 2008; 18:1814–1828. DOI: 10.1101/gr.076554.108 [PubMed: 18849524]
76. Paten B, et al. Genome-wide nucleotide-level mammalian ancestor reconstruction. Genome Res. 2008; 18:1829–1843. DOI: 10.1101/gr.076521.108 [PubMed: 18849525]

77. Gravel S, et al. Demographic history and rare allele sharing among human populations. Proceedings of the National Academy of Sciences. 2011; 108:11983–11988. DOI: 10.1073/pnas.1019276108

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

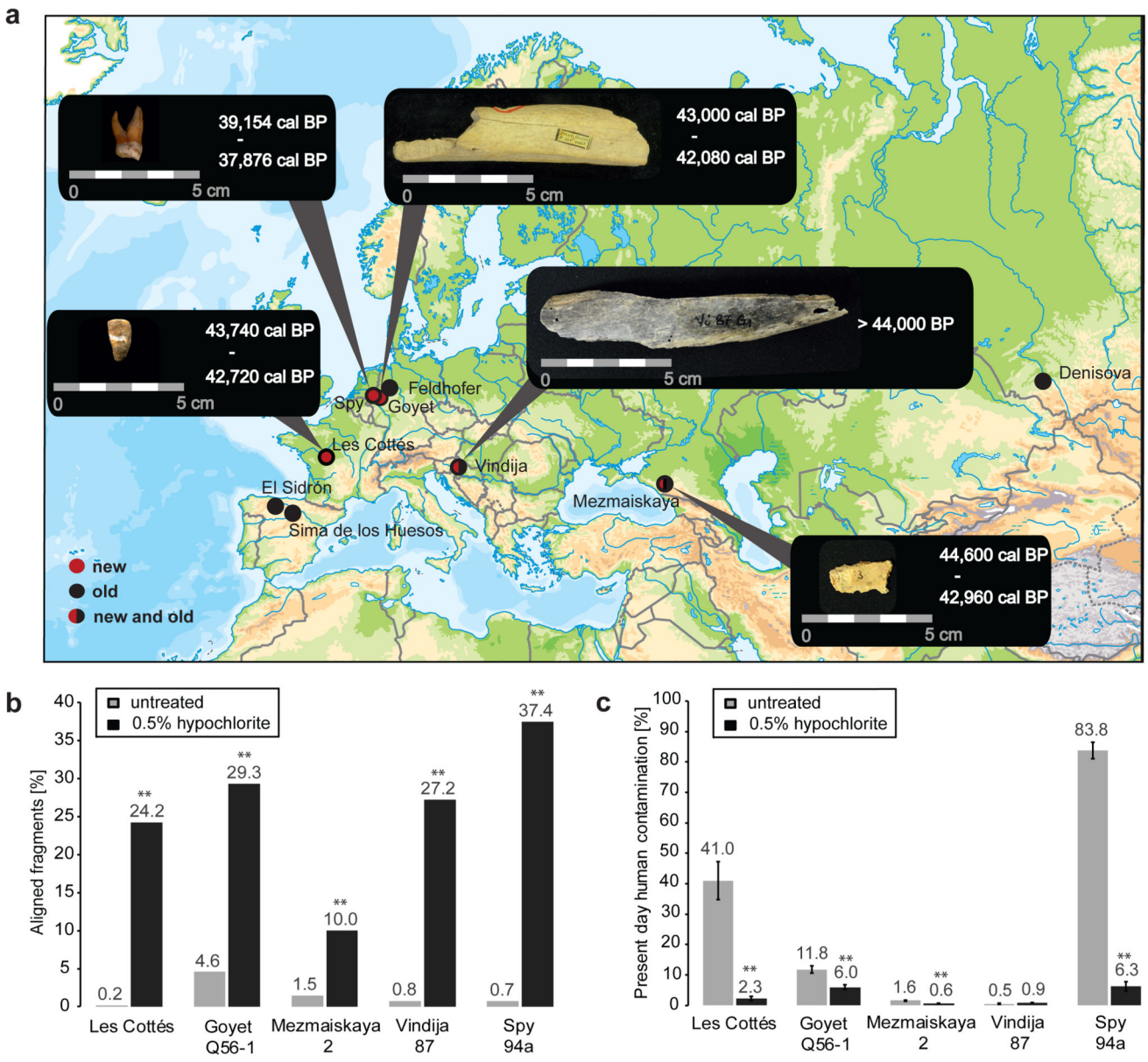


Figure 1. Specimen information and the effects of 0.5% hypochlorite treatment.

a, Location and age of the five late Neandertal specimens analysed in this study (new), and other Neandertal sites from which genome-wide data was obtained previously (old; map source Vectormapcollection). **b,** Proportion of DNA fragments aligned to the human reference genome in untreated bone/tooth powder and in powder treated with 0.5% sodium hypochlorite (Supplementary Table S2.1). **c,** Proportion of present-day human contamination (with 95% binomial confidence intervals) inferred from mitochondrial DNA fragments in treated and untreated sample material. Pearson's chi-squared test (two-sided) used for calculating significant differences (denoted with **, $\alpha << 0.001$) (Supplementary Table S2.5).

Europe PMC Funders Author Manuscripts

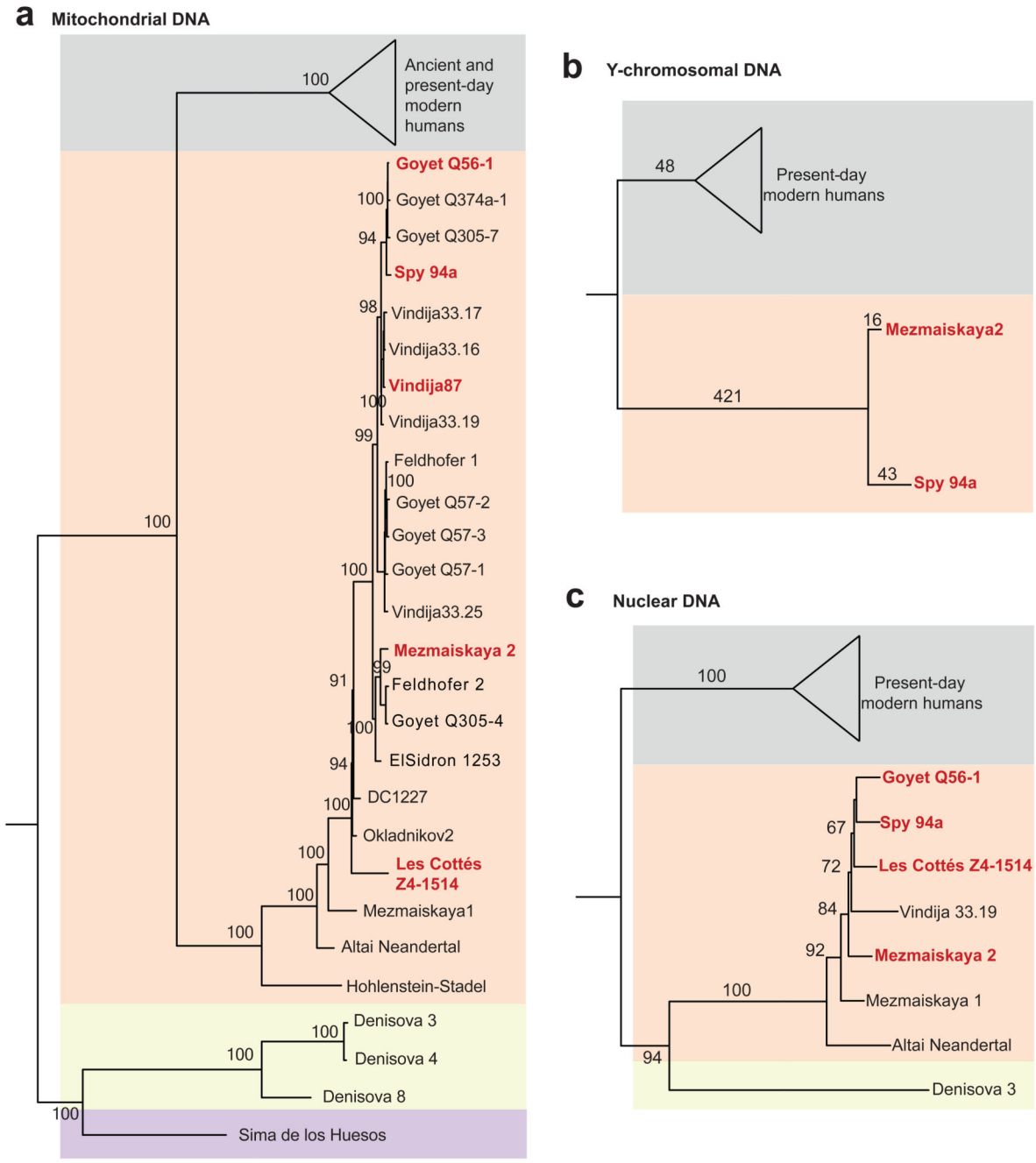


Figure 2. Phylogenetic relationships of late Neandertals.

a, Bayesian phylogenetic tree of mitochondrial genomes of 23 Neandertals, 3 Denisovans, 64 modern humans and a hominin from Sima de los Huesos. Reported are posterior probabilities for the branches. **b**, Neighbour-joining tree of Y chromosome sequences of *Mezmaiskaya 2*, *Spy 94a*, 175 present-day humans21 and two present-day humans carrying the A00 haplogroup30. Numbers of substitutions are denoted above the branches. **c**, Neighbour-joining tree of nuclear genomes based on autosomal transversions among late

Neandertals, *Vindija 33.19, Mezmaiskaya 1*, *Altai* Neandertal, *Denisovan* and 12 present-day humans. Reported are bootstrap support values after 1000 replications.

Europe PMC Funders Author Manuscripts

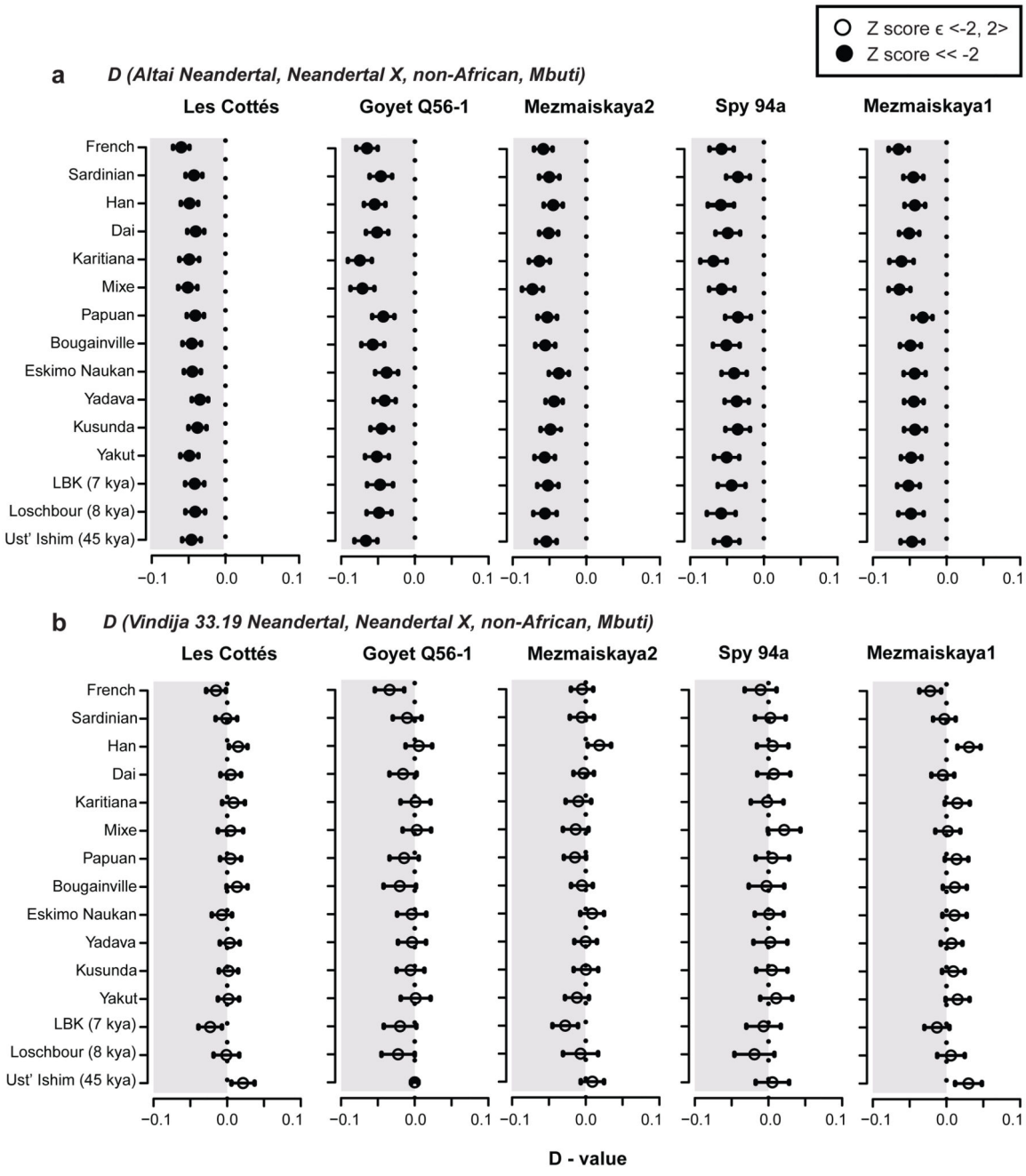


Figure 3. Proximity to the introgressing Neandertal populations in present-day and ancient humans calculated from the *D (Neandertal1, Neandertal2; non-African, African)*.

Three Mbuti individuals from SGDP were used as an outgroup and standard errors were calculated using a weighted block jackknife (Supplementary Information 10). Shaded grey region corresponds to the D-values < 0. **a,** All late Neandertals and the older *Mezmaiskaya 1* are significantly closer to the introgressing Neandertal population(s) than the *Altai Neandertal* **b**, There is no significant difference between late Neandertals, *Mezmaiskaya 1*

Europe PMC Funders Author Manuscripts

and *Vindija 33.19* in their proximity to the introgressing Neandertal population(s) in present-day and ancient humans.

Europe PMC Funders Author Manuscripts