An early modern human from Romania with a recent Neanderthal ancestor

Qiaomei Fu1,2,*, Mateja Hajdinjak3*, Oana Teodora Moldovan4, Silviu Constantin5, Swapan Mallick2,6,7, Pontus Skoglund2, Nick Patterson6, Nadin Rohland2, Iosif Lazaridis4, Birgit Nickel1, Bence Viola1,2,7,8, Kay Prüfer1, Matthias Meyer5, Janet Kelso3, David Reich2,6,9 & Svante Pääbo1

Neanderthals are thought to have disappeared in Europe approximately 39,000–41,000 years ago but they have contributed 1–3% of the DNA of present-day people in Eurasia1. Here we analyse DNA of a 37,000–42,000-year-old modern human from Peștera cu Oase, Romania. Although the specimen contains small amounts of human DNA, we use an enrichment strategy to isolate sites that are informative about its relationship to Neanderthals and present-day humans. We find that on the order of 6–9% of the genome of the Oase individual is derived from Neanderthals, more than any other modern human sequenced to date. Three chromosomal segments of Neanderthal ancestry are over 50 centimorgans in size, indicating that this individual had a Neanderthal ancestor as recently as four to six generations back. However, the Oase individual does not share more alleles with later Europeans than with East Asians, suggesting that the Oase population did not contribute substantially to later humans in Europe.

Between 45,000 and 35,000 years ago, anatomically modern humans spread across Europe, while the Neanderthals, present since before 300,000 years ago, disappeared. How this process occurred has long been debated2–5. Comparisons between the Neanderthal genome and the genomes of present-day humans have shown that Neanderthals contributed approximately 1–3% of the genomes of all people living today outside sub-Saharan Africa6 suggesting that human populations ancestral to all non-Africans mixed with Neanderthals. The size of segments of Neanderthal ancestry in present-day humans suggests that this occurred between 37,000 and 86,000 years ago6. However, where and how often this occurred is not understood. For example, Neanderthals share more alleles with East Asians and Native Americans than with Europeans, which may reflect additional interbreeding in the ancestors of eastern non-Africans6–8. Surprisingly, analyses of present-day genomes have not yielded any evidence that Neanderthals mixed with modern humans in Europe, despite the fact that Neanderthals were numerous there and cultural interactions between the two groups have been proposed9,10.

More direct insight into the interactions between modern and archaic humans can be obtained by studying genomes from modern humans who lived at a time when they could have met Neanderthals. Recent analyses of genomes from a 43,000–47,000-year-old modern human from western Siberia15 and a 36,000–39,000-year-old modern human from eastern Europe16 showed that Neanderthal gene flow into modern humans occurred before these individuals lived. The Siberian individual’s genome contained some segments of Neanderthal ancestry as large as 6 million base pairs (bp), suggesting that some Neanderthal gene flow could have occurred a few thousand years before his death15.

We report genome-wide data from a modern human mandible, Oase 1, found in 2002 in the Peștera cu Oase, Romania. The age of this specimen has been estimated to be ~37,000–42,000 years by direct radiocarbon dating17,18. Oase 1 is therefore one of the earliest modern humans in Europe. Its morphology is generally modern but some aspects are consistent with Neanderthal ancestry19–21. Subsequent excavations uncovered a cranium from another, probably contemporaneous individual, Oase 2, which also carries morphological traits that could reflect admixture with Neanderthals17,19.

We prepared two DNA extracts from 25 mg and 10 mg of bone powder removed from the inferior right ramus of Oase 1. We treated an aliquot of each of these extracts with Escherichia coli uracil-DNA glycosylase (UDG), an enzyme that removes uracils from the interior parts of DNA molecules, but leaves a proportion of uracils at the ends of the molecules unaffected. Uracil residues occur in DNA molecules as a result of deamination of cytosine residues, and are particularly prevalent at the ends of ancient DNA molecules22,23. Among the DNA fragments sequenced from these two extracts, 0.18% and 0.06%, respectively, could be mapped to the human reference genome. We prepared three additional DNA libraries from the extract containing 0.18% human-like molecules, but omitted the UDG treatment to increase the number of molecules in which terminal C-to-T substitutions could be seen and used to identify putatively ancient fragments. Because the fraction of endogenous DNA is so small, we used hybridization to DNA probes to isolate human DNA fragments from the libraries. Applying this strategy to the mitochondrial genome allowed the mitochondrial (mt)DNA from the five libraries to be sequenced to an average coverage of 803-fold (Supplementary Note 1).

At the ends of the DNA fragments, cytosine residues appeared as thymine residues relative to the human mtDNA reference in 21% of fragments, reflecting appreciable levels of cytosine deamination. This suggests that at least some of the human mtDNA is of ancient origin. We determined mtDNA consensus sequences in two ways: using all mtDNA fragments, and using only deaminated fragments that carry C-to-T substitutions at either end relative to the consensus mtDNA sequence based on these fragments, an approach known to enrich for endogenous DNA24–26. The mtDNA sequence based on all fragments clusters with present-day Europeans (Extended Data Fig. 1) (Supplementary Note 1). In contrast, the mtDNA sequence based on deaminated fragments is related to a large group of present-day Eurasian mtDNAs (haplogroup N) but diverges from these before they diverged from each other. This Oase 1 mtDNA carries a few private mutations on the basis of which its age can be estimated to be 36,330 years before present (14,520–56,450; 95% confidence interval). Using six positions at which the mtDNA sequence differs from at least 99% of 311 present-day humans, we estimate the contamination

1Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. 2Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. 3Department of Evolutionary Genomics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. 4Emil Racoviţă Institute of Speleology, Cluj Branch, 400006 Cluj, Romania. 5Emil Racoviţă Institute of Speleology, Department of Geospeleology and Paleontology, 010986 Bucharest 12, Romania. 6Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. 7Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. 8Department of Anthropology, University of Toronto, Toronto, Ontario, MSS 252, Canada. 9Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, 02115, USA.

*These authors contributed equally to this work.
among all mtDNA fragments to be 67% (95% confidence interval 65–69%). When we restrict to mtDNA fragments that carry terminal C-to-T substitutions, the contamination estimate is 4% (95% confidence interval of 2–9%) (Supplementary Note 1).

To isolate nuclear DNA from Oase 1, we used three sets of oligonucleotide probes that cover about two million sites that are single nucleotide polymorphisms (SNPs) in present-day humans and captured DNA molecules from the five libraries. Of the SNPs targeted, 51% ($n = 1,038,619$) were covered by at least one DNA fragment, and 13% ($n = 271,326$) were covered by at least one fragment with a terminal C-to-T substitution. To estimate nuclear DNA contamination, we tested whether Oase 1 DNA fragments with or without evidence of deamination share more alleles with present-day Europeans or with East Asians. We found that Europeans share significantly fewer alleles with the Oase 1 fragments that are deaminated than with Oase 1 fragments that are not, consistent with European contamination of the Oase 1 genome using three different statistics 7,29 (Supplementary Note 4). We then estimated the proportion of Neanderthal DNA in the Oase 1 individual when we compare him to four early modern humans: an 8,000-year-old individual from Luxembourg, and three individuals from Russia who vary in age between 24,000 and 45,000 years (3.6 ≤ $|Z| ≤ 6.8$; Extended Data Table 3). Thus, the Oase 1 individual appears to have carried more Neanderthal-like DNA than any other modern human analysed to date. This observation cannot be explained by residual present-day human contamination among the DNA fragments that carry terminal C-to-T substitutions, because all modern humans studied to date carry less Neanderthal ancestry than the Oase 1 genome, and thus contamination would lower, rather than increase, the apparent Neanderthal ancestry.

We estimated the proportion of Neanderthal DNA in the Oase 1 genome using three different statistics 7,29 (Supplementary Note 4). Although the results differ, they all yield point estimates between 6.0% and 9.4% (Table 1). For one of the statistics, none of the 90% confidence intervals for Neanderthal ancestry in the other modern present-day individuals from different populations using $D$-statistics, which provides a robust estimate of admixture almost regardless of how SNPs for analysis are chosen 27. We find that Oase 1 shared more alleles with present-day East Asians and Native Americans than with present-day Europeans, counter to what might naively be expected for an ancient individual from Europe (Fig. 1) ($5.2 ≤ |Z| ≤ 6.4$; Extended Data Table 1). However, it has been suggested that Europeans after the introduction of agriculture derive a part of their ancestry from a ‘basal Eurasian’ population that separated from the initial settlers of Europe and Asia before they split from each other 28. Therefore, we replaced present-day Europeans with Paleolithic and Mesolithic European individuals in these analyses. We then find that the Oase 1 individual shares equally many alleles with these early Europeans as with present-day East Asians and Native Americans (Fig. 1) ($|Z| ≤ 1.5$ in Extended Data Table 1). Restricting this analysis to transversion polymorphisms, which are not susceptible to errors induced by cytosine deamination, does not influence this result (Extended Data Table 2 and Supplementary Note 3). This suggests that the Oase 1 individual belonged to a population that did not contribute much, or not at all, to later Europeans. This contrasts, for example, with the ~36,000–39,000-year-old Kostenki 14 individual from western Russia, who was more closely related to later Europeans than to East Asians (1.9 ≤ $|Z| ≤ 13.7$; Extended Data Table 1) 18.

To assess whether the ancestors of the Oase 1 individual mixed with Neanderthals, we tested whether the Altai Neanderthal genome shares more alleles with the Oase 1 genome than with sub-Saharan Africans. We find this to be the case ($|Z| = 7.7$; Supplementary Note 4). We then asked whether the amount of Neanderthal ancestry in the Oase 1 genome is similar to that in present-day non-Africans. Surprisingly, the Neanderthal genome shares more alleles with the Oase 1 individual than it does with any present-day people in Eurasia that we tested, indicating that he carries more Neanderthal-like DNA than present-day people (5.0 ≤ $|Z| ≤ 8.2$; Extended Data Table 3). We also observe more Neanderthal-like alleles in the Oase 1 individual when we compare him to four early modern humans: an 8,000-year-old individual from Luxembourg, and three individuals from Russia who vary in age between 24,000 and 45,000 years (3.6 ≤ $|Z| ≤ 6.8$; Extended Data Table 3). Thus, the Oase 1 individual appears to have carried more Neanderthal-like DNA than any other modern human analysed to date. This observation cannot be explained by residual present-day human contamination among the DNA fragments that carry terminal C-to-T substitutions, because all modern humans studied to date carry less Neanderthal ancestry than the Oase 1 genome, and thus contamination would lower, rather than increase, the apparent Neanderthal ancestry.

We estimated the proportion of Neanderthal DNA in the Oase 1 genome using three different statistics 7,29 (Supplementary Note 4). Although the results differ, they all yield point estimates between 6.0% and 9.4% (Table 1). For one of the statistics, none of the 90% confidence intervals for Neanderthal ancestry in the other modern

---

**Table 1 | Estimated fraction of the Oase 1 genome that derives from Neanderthals**

| Sample          | Statistic 1 | Statistic 2 | Statistic 3 |
|-----------------|-------------|-------------|-------------|
|                 | $f_X(Denisova, Altai; Mbuti, X)$ | $f_X(Mbuti, Chimp; X, Denisova)$ | $f_X(X, Mbuti; Denisova, Chimp)$ |
| Oase 1          | 0.0%        | 0.0%        | 0.0%        |
| Ljubljanica     | 0.0%        | 0.0%        | 0.0%        |
| Kostenki 14     | 0.0%        | 0.0%        | 0.0%        |
| MA1             | 0.0%        | 0.0%        | 0.0%        |
| Loschbour       | 0.0%        | 0.0%        | 0.0%        |
| La Brana        | 0.0%        | 0.0%        | 0.0%        |
| Stuttgart       | 0.0%        | 0.0%        | 0.0%        |
| Han             | 0.0%        | 0.0%        | 0.0%        |
| Dai             | 0.0%        | 0.0%        | 0.0%        |
| French          | 0.0%        | 0.0%        | 0.0%        |

CI, confidence interval; s.e.m., standard error of the mean; negative values are truncated to 0%.

---

©2015 Macmillan Publishers Limited. All rights reserved
human samples overlap with the confidence interval in Oase 1. When we restrict analysis to transversion SNPs, the point estimates of Neanderthal ancestry are even higher (range of 8.4% to 11.3%) (Extended Data Table 4).

To study the spatial distribution of Neanderthal DNA across the Oase 1 genome, we designed capture probes for around 1.7 million nucleotide positions at which nearly all individuals in a sub-Saharan African population carry one allele whereas Neanderthal genomes carry a different allele. We used these probes to isolate DNA fragments from the Oase 1 individual. A total of 78,055 sites were covered by deaminated DNA fragments from the Oase 1 individual and were also covered by DNA fragments sequenced from the ~36,000–39,000-year-old Kostenki 14 individual from western Russia, the ~43,000–47,000-year-old individual from Ust’-Ishim in Siberia, and three present-day human genomes from China, France and Sudan (Supplementary Note 5). Because the Dinka from Sudan are thought to have little or no Neanderthal ancestry, we subtracted the number of alleles that match the Neanderthals in the Dinka individual (485) from the number in the other genomes to estimate the number of alleles attributable to Neanderthal ancestry. The resulting numbers of putative Neanderthal alleles are 3,746 in the Oase 1 individual, 1,586 and 1,121 in the Ust’-Ishim and Kostenki 14 individuals, respectively, and 1,322 and 1,033 in the Chinese and the European individuals (Extended Data Table 5). Thus, the Neanderthal contribution to the Oase 1 genome appears to be between 2.3- and 3.6-fold larger than to the other genomes analysed. Assuming that the Neanderthal contribution to the European individual is 2% (ref. 7), this suggests that 7.3% of the Oase 1 genome is of Neanderthal origin. When the numbers of alleles matching the Neanderthal genome are compared per chromosome (Extended Data Table 5), the highest numbers are always observed for the Oase 1 genome, except in the case of chromosome 21, in which the Ust’-Ishim individual carries a large segment of likely Neanderthal ancestry.

We plotted the positions of Neanderthal-like alleles across the Oase 1 genome (Fig. 2). We detect three segments that are over 50 centimorgans (cM) in size, suggesting that the Neanderthal contribution to the Oase 1 individual occurred so recently in his family tree that chromosomal segments of Neanderthal origin had little time to break up due to recombination. To estimate the date of the most recent Neanderthal contribution to the Oase 1 genome, we studied the size spans of seven segments of the genome that appeared to be recently derived from Neanderthals. Their genetic lengths suggest that the Oase 1 individual had a Neanderthal ancestor as a fourth-, fifth- or sixth-degree relative (Supplementary Note 5). This would predict that an average of 1.6% to 6.3% of the Oase 1 genome derived from a recent Neanderthal ancestor. Visual inspection of the Oase 1 genome suggests that in addition to these seven segments, other smaller segments also carry Neanderthal-like alleles (Fig. 2). When we remove the seven longest segments, the estimate of Neanderthal ancestry in Oase 1 drops from 7.3% to 4.8%, which is still around twice the 2.0–2.9% estimated for the French, Han, Kostenki and Ust’-Ishim individuals in this remaining part of the genome. This additional Neanderthal ancestry...
could reflect an older Neanderthal admixture into the ancestors of Oase 1, or that we failed to find all segments of recent Neanderthal ancestry. The Oase 1 genome shows that mixture between modern humans and Neanderthals was not limited to the first ancestors of present-day people to leave Africa, or to people in the Near East; it occurred later as well and probably in Europe. The fact that the Oase 1 individual had a Neanderthal ancestor removed by only four to six generations allows this Neanderthal admixture to be dated to less than 200 years before the time he lived. However, the absence of a clear relationship of the Oase 1 individual to later modern humans in Europe suggests that he may have been a member of an initial early modern human population that interbred with Neanderthals but did not contribute much to later European populations. To better understand the interactions between early modern and Neanderthal populations, it will be important to study other specimens that, like Oase 1, have been suggested to carry morphological traits suggestive of admixture with Neanderthals.

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

**Received 24 February 2015; accepted 5 May 2015.**

**Published online 22 June 2015.**

1. Higham, T. et al. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* 512, 306–309 (2014).
2. Trinkaus, E. et al. An early modern human from the Pestera cu Oase, Romania. *Proc. Natl Acad. Sci. USA* 100, 11231–11236 (2003).
3. Smith, F. H. et al. The assimilation model: modern human origins in Europe, and the extinction of Neandertals. *Quat. Int.* 137, 7–19 (2005).
4. Zilhão, J. Neandertals and moderns mixed, and it matters. *Evol. Anthropol.* 15, 183–195 (2006).
5. Hublin, J.-J. The modern human colonization of western Eurasia: when and where? *Quat. Sci. Rev.* 118, 194–210 (2014).
6. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* 328, 710–722 (2010).
7. Prüfer, K. et al. The complete genome sequence of a Neandertal from the Altai Mountains. *Nature* 505, 43–49 (2014).
8. Sankararaman, S. et al. The date of interbreeding between Neandertals and modern humans. *PloS Genet.* 8, e1002947 (2012).
9. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222–226 (2012).
10. Wall, J. D. et al. Higher levels of Neandertal ancestry in East Asians than in Europeans. *Genetics* 194, 199–209 (2013).
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 (2015).
12. Vernot, B. & Akey, J. M. Complex history of admixture between modern humans and Neandertals. *Am. J. Hum. Genet.* 96, 448–453 (2015).
13. d’Errico, F. et al. Neandertal acculturation in Western Europe? A critical review of the evidence and its interpretation. *Curr. Anthropol.* 39 (suppl.), S1–S44 (1998).
14. Mellars, P. The impossible coincidence: a single-species model for the origins of modern human behavior in Europe, *Evol. Anthropol.* 14, 12–27 (2005).
15. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514, 445–449 (2014).
16. Seguin-Orlando, A. et al. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346, 1113–1118 (2014).
17. Rougier, H. et al. Peštera cu Oase 2 and the cranial morphology of early modern Europeans. *Proc. Natl Acad. Sci. USA* 104, 1165–1170 (2007).
18. Constantin, S. et al. Reconstructing the evolution of cave systems as a key to understanding the taphonomy of fossil accumulations: the case of Ursililor Cave (Western Carpathians, Romania). *Quat. Int.* 339–340, 25–40 (2014).
19. Ackermann, R. R. Phenotypic traits of primate hybrids: Recognizing admixture in the fossil record. *Evolutionary Anthropology: Issues, News, and Reviews* 19 258–270 (2010).
20. Frayer, D. in *Continuity or Replacement? Controversies in Homo sapiens Evolution* (eds Brauer, G. & Smith, F. H.) (A.A. Balkema Publishers, 1992).
21. Trinkaus, E. et al. Life and Death in the Neandertal Oase. *A Setting for Modern Human Emergence in Europe* (Oxford Univ. Press, 2013).
22. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38, e87 (2010).
23. Fu, Q. et al. DNA analysis of an early modern human from Tiyanuan Cave, China. *Proc. Natl Acad. Sci. USA* 110, 2223–2227 (2013).
24. Sawyer, S. et al. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PloS ONE* 7, e34131 (2012).
25. Meyer, M. et al. A mitochondrial genome sequence of a hominin from Sima de los Huesos, *Nature* 505, 403–406 (2014).
26. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* 111, 2229–2234 (2014).
27. Patterson, N. et al. Ancient admixture in human history. *Genetics* 192, 1065–1093 (2012).
28. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413 (2014).
29. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060 (2010).
30. Trinkaus, E. European early modern humans and the fate of the Neandertals. *Proc. Natl Acad. Sci. USA* 104, 7367–7372 (2007).

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

**Acknowledgements** We thank E. Trinkaus for facilitating the sampling of Oase 1. J. Krause for help with sampling and initial DNA analyses, A. Aximu for help with DNA methylation in ancient DNA. Nucleic Acids Res.

**Author Contributions** N.P., K.P., M.M., J.K., D.R. and S.P. supervised the study. S.C. and D.R. is supported by US National Science Foundation HOMINID grant BCS-1032255, the Howard Hughes Medical Institute. The laboratory work was funded by the Presidential Innovation Fund of the Max Planck Society.

**Author Information** The aligned sequences have been deposited in the European Nucleotide Archive under accession number PRJEB8987. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R. (reich@genetics.med.harvard.edu) or S.P. (paabo@eva.mpg.de).
METHODS

DNA extraction and library preparation. We used a dentistry drill to remove two samples of bone powder from an area where a larger sample had previously been removed for carbon dating. We prepared two extracts (E1406, E1843) from 25 mg and 10 mg of bone powder, respectively, as described. We produced five libraries from the two extracts using a single-stranded library protocol and primer design. We amplified all libraries by PCR for 35 cycles using AccuPrime Pfx DNA polymerase (Life Technologies) and primers carrying library-specific indexes. We determined library concentrations using a NanoDrop 2000 spectrophotometer.

Sequencing and DNA capture. We shotgun sequenced the UDG-treated libraries A5252 and A5227 and found that they contained 0.06% and 0.18% human DNA, respectively. We used hybridization to oligonucleotide probes to enrich the libraries for subsets of the nuclear genome containing panels of known SNPs as described, except that each SNP was targeted by five 52-nucleotide probes: two immediately flanking the SNP on both sides, and two centred on the SNP containing one or the other alternate allele, respectively. We used four panels of probes.

Panel 1 “390k”: 394,577 SNPs, about 90% of which are on the Affymetrix Human Origins array. See ref. 36 for SNPs and probes.

Panel 2 “840k”: 842,630 SNPs constituting the rest of the SNPs on the Human Origins array, all SNPs on the Illumina 610-Quad array, all SNPs on the Affymetrix 50k array, and smaller numbers of SNPs chosen for other purposes. See Supplementary Data 1.

Panel 3 “100k”: 997,780 SNPs comprising all transversion polymorphisms seen in two Yoruba males from Nigeria sequenced to high coverage and transversion polymorphisms seen in the Altai Neanderthal genome. The design was restricted to SNPs that passed strict quality filters in the Neanderthal genome (Map35_99%) and had chimpanzee alleles available. Probes were designed from chimpanzee flanking sequences. See Supplementary Data 2.

Panel 4 “Archaic”: This panel contains SNPs where the West-African Yoruba population carry a high frequency of one allele while at least one archaic individual carries an alternative allele. To determine Yoruba allele frequencies, we examined data from all Yoruba individuals from the 1000 Genomes Project covered by at least three sequences passing filters. At these sites we called majority alleles (drawing a random allele in the case of equal numbers of reads supporting both alleles). We furthermore restricted the analysis to sites at which ≥24 Yoruba individuals as well as the Altai Neanderthal and Denísovan genomes had allele calls (Map35_50% filter). We then selected sites at which at most one alternative allele is seen among the Yoruba while at least one of four archaic genomes (Denísovan, Altai, Vindija and Mezmaiskaya Neanderthals) carry the alternative allele. Ancestral states were taken from the inferred ancestor of humans and chimpanzees (Ensembl Compara v.64). We used the following classes of sites. Class 1: 297,894 SNPs where Yoruba is derived and at least one ancestral allele is seen in the Altai, Vindija, Mezmaiskaya or Denísovan genomes. Class 2: sites where Yoruba alleles are all or nearly all ancestral and derived alleles are seen in archaic genomes. Since such derived alleles often arise due to errors in an archaic genome, we restricted this class to the following three cases: (1) 1,321,774 SNPs where the high-coverage Altai Neanderthal and/or Denísovan genomes are homozygous derived; (2) 523,041 SNPs where the Altai and/or Denísovan genomes are heterozygous but are not C-to-T or G-to-A substitutions relative to the ancestral allele; and (3) 30,735 SNPs that are homozygous ancestral in Altai and/or Denísovan and at least one copy of the derived allele is observed in the Mezmaiskaya or Vindija Neanderthal genomes, and the derived allele represents a transversion that is also seen in the Simons Genome Diversity Panel (https://www.simonsfoundation.org/life-sciences/simons-genome-diversity-project/). After eliminating SNPs where capture probes covered ambiguous bases in the human (hg19) and chimpanzee (pan tro2) genomes or overlapped for less than 35 nucleotides with mapable regions (Map35_50%), this left us with a set of 1,749,385 SNPs (see Supplementary Data 3).

Sequencing of capture products and data processing. We sequenced capture products using 2 × 75 bp reads on an Illumina HiSeq2500 or an Illumina NextSeq500. We de-multiplexed the reads allowing one mismatch in each of the two indices (Extended Data Table 6), and merged paired reads into sequenced fragments requiring an overlap of at least 15 bp (allowing one mismatch) using a modified form of SeqPrep (https://github.com/jstjohn/SeqPrep). We used the bases with the higher quality (and score) to represent the overlap region. After removing adapters, we mapped merged fragments to hg19 using BWA (v.0.6.1) using the ‘samse’ command. We identified duplicated fragments on the basis of sharing the same orientation and end positions, in which case we kept the fragment with the highest quality (Extended Data Table 7).

To focus on putatively deaminated fragments we used fragments with C-to-T substitutions relative to the hg19 human genome reference sequence in the first 5′ or last two 3′ bases for the UDG-treated libraries, and to fragments with C-to-T substitutions relative to hg19 in the terminal three bases at either end of fragments from non-UDG-treated libraries (Supplementary Note 1 and Extended Data Table 8).

Merging the Oase 1 data with genome sequences. At each SNP covered at least once in Oase 1, we selected the majority allele (in case of a tie, we picked a random allele). We then merged the Oase 1 data with 25 genomes of present-day humans sequenced to 24–42× coverage: the Altai Neanderthal, the Siberian Denisovan, a ~45,000-year-old modern human from Ust’ Ishim in Siberia, an ~8,000-year-old Melosilicid individual from Loschbour Cave, Luxembourg, and ~7,000-year-old early farmer from Stuttgart, Germany (Extended Data Table 9). All the genotype calls for the five deeply sequenced ancient genomes were performed in the same way. We restricted analyses to sites with a minimum root-mean-square mapping quality (MAPQ) of 30 in the 30 genomes. We added lower coverage shotgun data from the ~36,000-year-old Kostenki 14 from Russia, the ~24,000-year-old Mal’ta Siberian individual from Russia, an 8,000-year-old Melosilicid individual from La Brâna Cave, Spain, a Neanderthal from Mezmaiskaya in Russia, and a pool of three Neanderthals from Vindija Cave in Croatia. For these samples, we restricted to fragments with a map quality of MAPQ ≥ 37 to match the filter for the low-coverage Oase 1 data (Extended Data Table 9).

Population genetic analyses. To determine the relationship of Oase 1 to other modern humans, we used D-statistics to evaluate whether sets of four tested samples are consistent with being related to one another according to an unrooted tree (Supplementary Note 3). We used D-statistics and FST-statistic ratios to test both whether there is excess archaic ancestry in Oase 1 compared with other modern humans, and to estimate proportions of Neanderthal ancestry (Supplementary Note 4). We studied the genomic distribution of alleles that are likely to derive from Neanderthals in the sense of being shared with Neanderthal but either absent or at very low frequency in West Africans. We used the spatial distribution of these sites to identify stretches of likely Neanderthal ancestry in several individuals including Oase 1. We also used these data to estimate the number of generations since the most recent Neanderthal ancestor of Oase 1 (Supplementary Note 5).

31. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl Acad. Sci. USA 110, 15758–15763 (2013).
32. Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols 8, 737–748 (2013).
33. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl Acad. Sci. USA 104, 14616–14621 (2007).
34. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. Biotechniques 52, 87–94 (2012).
35. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. 40, e3 (2012).
36. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211 (2015).
37. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012).
38. 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 (2008).
39. Paten, B. et al. Genome-wide nucleotide-level mammalian ancestor reconstruction. Genome Res. 18, 1829–1843 (2008).
40. Raghavan, M. et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505, 87–91 (2014).
41. Olaide, I. et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature 507, 225–228 (2014).
Extended Data Figure 1 | Mitochondrial DNA tree for Oase 1 and other modern humans. The consensus sequences for all Oase 1 fragments and for deaminated fragments are shown. The tree is rooted with a Neanderthal mtDNA (Vindija33.25).
### Extended Data Table 1 | Allele sharing between early modern humans and other humans

| Non-African<sub>1</sub>       | Non-African<sub>2</sub>       | Oase 1 D | Oase 1 Z | Ust'-Ishim D | Ust'-Ishim Z | Kostenki 14 D | Kostenki 14 Z |
|-------------------------------|-------------------------------|---------|---------|-------------|-------------|--------------|--------------|
| Oase 1                        | Ust'-Ishim                    |         |         | -0.0037     | -4.1        | -0.0033      | -3.8         |
| Oase 1                        | Kostenki 14                   |         |         | -0.0031     | -5.1        | -0.0092      | -9.8         |
| Oase 1                        | MA1                           |         |         | -0.0032     | -3.5        | -0.0101      | -12.2        |
| Oase 1                        | Loschbour                     |         |         | -0.0032     | -3.9        | -0.0011      | -1.6         |
| Oase 1                        | East Asian                    |         |         | -0.0027     | -3.8        | -0.0009      | -1.4         |
| Oase 1                        | Native American               |         |         | -0.0030     | -4.1        | -0.0039      | -5.5         |
| Ust'-Ishim                    | Kostenki 14                   | -0.0005 | -0.6    |             |             | -0.0059      | -6.4         |
| Ust'-Ishim                    | MA1                           | -0.0007 | -0.8    |             |             | -0.0068      | -8.5         |
| Ust'-Ishim                    | Loschbour                     | 0.0002  | 0.3     |             |             | 0.0022       | 3.3          |
| Ust'-Ishim                    | East Asian                    | 0.0000  | -0.1    |             |             | 0.0006       | 0.8          |
| Ust'-Ishim                    | Native American               | -0.0007 | -1.0    |             |             | -0.0006      | -0.8         |
| Kostenki 14                   | MA1                           | -0.0004 | -0.6    | 0.0003      | 0.4         |              |              |
| Kostenki 14                   | Loschbour                     | 0.0007  | 1.0     | 0.0006      | 0.8         |              |              |
| Kostenki 14                   | East Asian                    | 0.0004  | 0.6     | 0.0011      | 1.6         |              |              |
| Kostenki 14                   | Native American               | -0.0002 | -0.3    | 0.0008      | 1.1         |              |              |
| MA1                           | Loschbour                     | 0.0012  | 1.7     | 0.0005      | 0.7         | -0.0012      | -1.5         |
| MA1                           | East Asian                    | 0.0008  | 1.2     | 0.0007      | 1.1         | 0.0079       | 10.6         |
| MA1                           | Native American               | 0.0001  | 0.1     | 0.0004      | 0.6         | 0.0051       | 7.0          |
| Loschbour                     | East Asian                    | -0.0002 | -0.4    | 0.0005      | 0.9         | 0.0090       | 13.7         |
| Loschbour                     | Native American               | -0.0009 | -1.5    | 0.0002      | 0.3         | 0.0062       | 9.0          |
| East Asian                    | Native American               | -0.0006 | -1.6    | -0.0003     | -0.8        | -0.0028      | -6.6         |

We compute $D(\text{Non-African}_1, \text{Non-African}_2; \text{Early Modern Human}, \text{African})$ to test whether an early modern human (Oase 1, Ust'-Ishim, or Kostenki 14) shares more alleles with Non-African<sub>1</sub> (in which case the statistic is positive) or Non-African<sub>2</sub> (negative). We use a pool of six sub-Saharan African genomes (2 Mbuti, 2 Yoruba, 2 Dinka) as an outgroup; a pool of four genomes (2 French, 2 Sardinians) to represent Europeans; a pool of four genomes (2 Han, 2 Dai) to represent East Asians; and a pool of three genomes (2 Karitiana, 1 Mixe) to represent Native Americans. Results are based on 242,122 transition and transversion SNPs covered by at least one deaminated fragment in Oase 1, and covered in all other samples, although not necessarily MA1. For analyses involving MA1, a subset of 176,569 transversion SNPs was analysed.
Extended Data Table 2 | Allele sharing between early modern humans and other humans (transversions only)

| Non-African₁ | Non-African₂ | Oase 1 D | Z   | Ust’-Ishim D | Z   | Kostenki 14 D | Z   |
|-------------|-------------|---------|-----|--------------|-----|---------------|-----|
| Oase 1      | Ust’-Ishim  |        |     | -0.0031      | -3.3|                |     |
| Oase 1      | Kostenki 14 |        |     | -0.0026      | -2.9| -0.0071       | -6.5|
| Oase 1      | MA1         |        |     | -0.0023      | -2.6| -0.0081       | -8.8|
| Oase 1      | Loschbour   |        |     | -0.0013      | -1.9| 0.0007        | 1.0 |
| Oase 1      | East Asian  |        |     | -0.0019      | -2.7| -0.0018       | -2.3|
| Oase 1      | Native American |    |     |               |     |               |     |
| Ust’-Ishim  | Kostenki 14 |        | -0.0012 | -1.4 |        |     |
| Ust’-Ishim  | MA1         |        | -0.0006 | -0.7 |        |     |
| Ust’-Ishim  | Loschbour   |        | 0.0003  | 0.4  |        |     |
| Ust’-Ishim  | East Asian  |        | 0.0005  | 0.7  |        |     |
| Ust’-Ishim  | Native American |    | -0.0003 | -0.4 |        |     |
| Kostenki 14 | MA1         |        | 0.0001  | 0.1  | 0.0002  | 0.3 |
| Kostenki 14 | Loschbour   |        | 0.0015  | 2.0  | 0.0008  | 1.1 |
| Kostenki 14 | East Asian  |        | 0.0017  | 2.3  | 0.0017  | 2.5 |
| Kostenki 14 | Native American |    | 0.0009  | 1.2  | 0.0012  | 1.6 |
| MA1         | Loschbour   |        | 0.0019  | 2.2  | 0.0010  | 1.3 |
| MA1         | East Asian  |        | 0.0011  | 1.4  | 0.0013  | 1.9 |
| MA1         | Native American |    | 0.0006  | 0.7  | 0.0007  | 1.1 |
| Loschbour   | East Asian  |        | 0.0001  | 0.2  | 0.0009  | 1.5 |
| Loschbour   | Native American |    | -0.0006 | -0.9 | 0.0004  | 0.6 |
| East Asian  | Native American |    | -0.0008 | -1.7 | -0.0006 | -1.3 |
| European    | Oase 1      |        | -0.0023 | -3.3 |        |     |
| European    | Ust’-Ishim  |        | -0.0035 | -5.1 | -0.0035 | -5.2 |
| European    | Kostenki 14 |        | -0.0033 | -4.5 | -0.0033 | -5.2 |
| European    | MA1         |        | -0.0020 | -3.6 | -0.0027 | -5.1 |
| European    | Loschbour   |        | -0.0018 | -3.6 | -0.0018 | -4.0 |
| European    | East Asian  |        | -0.0026 | -4.8 | -0.0023 | -5.2 |
| European    | Native American |    | -0.0009 | -1.7 | -0.0010 | -2.2 |
| European    | Stuttgart   |        | 0.0005  | 0.7  | 0.0041  | 4.7 |
| Stuttgart   | Oase 1      |        |        |     | 0.0007  | 8.2 |
| Stuttgart   | Ust’-Ishim  |        | -0.0014 | -1.8 |        |     |
| Stuttgart   | Kostenki 14 |        | -0.0026 | -3.3 | -0.0025 | -3.5 |
| Stuttgart   | MA1         |        | -0.0026 | -3.1 | -0.0023 | -3.2 |
| Stuttgart   | Loschbour   |        | -0.0011 | -1.6 | -0.0017 | -2.8 |
| Stuttgart   | East Asian  |        | -0.0010 | -1.4 | -0.0008 | -1.3 |
| Stuttgart   | Native American |    | -0.0017 | -2.4 | -0.0013 | -2.2 |

We compute $D_{\text{Non-African₁, Non-African₂}}$ (Early Modern Human, African), to test whether an early modern human (Oase 1, Ust’-Ishim or Kostenki 14) shares more alleles with Non-African₁ (in which case the statistic is positive) or Non-African₂ (negative). We use a pool of six sub-Saharan African genomes (2 Mbuti, 2 Yoruba, 2 Dinka) as an outgroup; a pool of four genomes (2 French, 2 Sardinians) to represent Europeans; a pool of four genomes (2 Han, 2 Dai) to represent East Asians; and a pool of three genomes (2 Karitiana, 1 Mixe) to represent Native Americans. Statistics are as in Extended Data Table 1 but are based on 106,004 transversion SNPs covered by at least one deaminated fragment in Oase 1 and that also have coverage for all other samples, although not necessarily MA1. For analyses involving MA1, a subset of 76,715 transversion SNPs is analysed.
### Extended Data Table 3 | Testing whether archaic genomes share more alleles with Oase 1 than with other modern humans

| Test       | Sites  | Archaic = Altai |               | Archaic = Denisovan |               |
|------------|--------|-----------------|---------------|---------------------|---------------|
|            |        | Chimp           | Mbuti         | Chimp               | Mbuti         |
|            |        | D Z             | D Z           | D Z                 | D Z           |
| Han        | 115,300| -0.0036         | -5.1          | -0.0071              | -7.6          | -0.0014 | -2.2  | -0.0049 | -6.3  |
| Dai        | 115,300| -0.0035         | -5.0          | -0.0077              | -8.2          | -0.0013 | -2.1  | -0.0056 | -7.0  |
| Karitiana  | 115,300| -0.0032         | -4.3          | -0.0063              | -6.9          | -0.0008 | -1.3  | -0.0040 | -5.3  |
| French     | 115,300| -0.0049         | -6.9          | -0.0074              | -8.2          | -0.0021 | -3.4  | -0.0047 | -6.2  |
| Sardinian  | 115,300| -0.0038         | -5.1          | -0.0071              | -7.8          | -0.0016 | -2.5  | -0.0050 | -6.5  |
| Papuan     | 115,300| -0.0026         | -3.6          | -0.0051              | -5.4          | 0.0009  | 1.5   | -0.0016 | -2.1  |
| Ust’-Ishim | 115,100| -0.0026         | -3.6          | -0.0052              | -5.5          | -0.0009 | -1.5  | -0.0035 | -4.4  |
| Kostenki14 | 108,100| -0.0032         | -4.1          | -0.0059              | -6.0          | -0.0017 | -2.4  | -0.0044 | -5.3  |
| MA1        | 83,200 | -0.0031         | -3.6          | -0.0050              | -4.7          | -0.0007 | -0.9  | -0.0028 | -2.8  |
| Loschbour  | 114,300| -0.0043         | -5.7          | -0.0066              | -6.8          | -0.0019 | -2.9  | -0.0043 | -5.3  |
| LaBrana    | 111,000| -0.0033         | -4.2          | -0.0072              | -7.3          | -0.0008 | -1.2  | -0.0047 | -5.4  |
| Stuttgart  | 114,000| -0.0037         | -5.1          | -0.0066              | -7.1          | -0.0013 | -2.1  | -0.0042 | -5.6  |

The statistic $D_{(Test, Oase 1; Archaic, Outgroup)}$ is negative if the archaic genomes share more alleles with Oase 1 than with a test sample. The outgroups are either chimpanzee or a sub-Saharan African (Mbuti).
## Extended Data Table 4 | Estimated fraction of the Oase 1 genome that derives from Neanderthals

| Sample         | \( f_4(Den,to,Alt,Alt;Nbuti) \) | \( 1 - f_4(Nbuti,Chimp;X,Denis) \) | \( f_4(X,Nbuti;Denisova,Chimp) \) |
|----------------|---------------------------------|---------------------------------|---------------------------------|
|                | Prop.   | S.E.   | 90% CI | Prop.   | S.E.   | 90% CI | Prop.   | S.E.   | 90% CI |
| Oase 1         | 11.3%   | 2.8%   | 6.7%-16% | 10.9%   | 1.6%   | 8.3%-13.6% | 8.4%   | 2.7%   | 4.0%-12.9% |
| Ust'-Ishim     | 2.9%    | 1.2%   | 1.0%-4.9% | 6.0%    | 0.8%   | 4.7%-7.4%  | 4.2%   | 1.5%   | 1.8%-6.6%  |
| Kostenki 14    | 3.0%    | 1.4%   | 0.7%-5.3% | 3.0%    | 0.9%   | 1.6%-4.5%  | 6.2%   | 1.6%   | 3.6%-8.7%  |
| MA1            | 1.5%    | 1.5%   | 0.0%-4.0% | 3.6%    | 1.0%   | 1.9%-5.2%  | 5.5%   | 1.6%   | 2.8%-8.2%  |
| Loschbour      | 1.1%    | 1.2%   | 0.0%-3.1% | 4.8%    | 0.9%   | 3.3%-6.2%  | 3.6%   | 1.5%   | 1.2%-6.1%  |
| LaBrana        | 3.7%    | 1.3%   | 1.4%-5.9% | 2.4%    | 0.9%   | 0.9%-3.8%  | 4.8%   | 1.5%   | 2.4%-7.2%  |
| Stuttgart      | 2.8%    | 1.2%   | 0.8%-4.8% | 3.4%    | 0.9%   | 2.0%-4.9%  | 3.8%   | 1.5%   | 1.4%-6.2%  |
| Han            | 1.0%    | 1.3%   | 0.0%-3.1% | 2.8%    | 0.9%   | 1.3%-4.2%  | 3.6%   | 1.5%   | 1.2%-6.1%  |
| Dai            | 2.1%    | 1.2%   | 0.2%-4.0% | 1.3%    | 0.9%   | 0.0%-2.8%  | 3.8%   | 1.5%   | 1.4%-6.2%  |
| French         | 1.6%    | 1.2%   | 0.0%-3.5% | 3.3%    | 0.9%   | 1.9%-4.7%  | 2.7%   | 1.5%   | 0.3%-5.2%  |
| Sardinian      | 2.7%    | 1.2%   | 0.8%-4.7% | 2.3%    | 0.9%   | 0.8%-3.7%  | 3.7%   | 1.4%   | 1.3%-6.1%  |

Estimates are as in Table 1 but restrict to transversions. Present-day human genomes are from a data set reported previously. ©2015 Macmillan Publishers Limited. All rights reserved
Extended Data Table 5  |  Counts of putative Neanderthal alleles in six modern humans

| Chr | Sites | Neanderthal allele counts | Neanderthal ancestry |
|-----|-------|---------------------------|----------------------|
|     |       | Oase 1 | Ust'-Ishim | Kostenki 14 | Han | French | Dinka | Oase 1 | Ust'-Ishim | Kostenki 14 | Han | French |
| 1   | 6740  | 323    | 196       | 148        | 129 | 117    | 25    | 6.70%  | 3.84%      | 2.77%      | 2.34%  | 2.07%  |
| 2   | 7112  | 294    | 145       | 121        | 188 | 199    | 29    | 5.65%  | 2.47%      | 1.96%      | 3.39%  | 3.62%  |
| 3   | 5417  | 177    | 102       | 96         | 74  | 98     | 28    | 4.17%  | 2.07%      | 1.90%      | 1.29%  | 1.96%  |
| 4   | 4495  | 359    | 86        | 63         | 141 | 96     | 42    | 10.69% | 1.48%      | 0.71%      | 3.34%  | 1.82%  |
| 5   | 4330  | 446    | 108       | 66         | 103 | 95     | 23    | 14.80% | 2.97%      | 1.50%      | 2.80%  | 2.52%  |
| 6   | 4549  | 324    | 155       | 167        | 142 | 138    | 73    | 8.36%  | 2.73%      | 3.13%      | 2.30%  | 2.16%  |
| 7   | 4422  | 147    | 68        | 65         | 102 | 72     | 34    | 3.87%  | 1.16%      | 1.06%      | 2.33%  | 1.30%  |
| 8   | 4322  | 131    | 132       | 72         | 35  | 38     | 14    | 4.10%  | 4.14%      | 2.03%      | 0.74%  | 0.84%  |
| 9   | 3107  | 500    | 69        | 120        | 118 | 49     | 15    | 23.65% | 2.63%      | 5.12%      | 5.02%  | 1.66%  |
| 10  | 4009  | 147    | 139       | 67         | 131 | 86     | 22    | 4.72%  | 4.42%      | 1.70%      | 4.12%  | 2.42%  |
| 11  | 4193  | 153    | 93        | 88         | 81  | 73     | 26    | 4.59%  | 2.42%      | 2.24%      | 1.99%  | 1.70%  |
| 12  | 3456  | 456    | 160       | 54         | 125 | 93     | 10    | 19.55% | 6.58%      | 1.93%      | 5.04%  | 3.64%  |
| 13  | 2457  | 96     | 81        | 33         | 54  | 30     | 18    | 4.81%  | 3.89%      | 0.93%      | 2.22%  | 0.74%  |
| 14  | 2390  | 85     | 27        | 52         | 50  | 52     | 13    | 4.56%  | 0.89%      | 2.47%      | 2.35%  | 2.47%  |
| 15  | 2327  | 73     | 78        | 47         | 38  | 32     | 5     | 4.43%  | 4.75%      | 2.73%      | 2.15%  | 1.76%  |
| 16  | 3139  | 90     | 121       | 68         | 43  | 39     | 8     | 3.96%  | 5.45%      | 2.90%      | 1.69%  | 1.50%  |
| 17  | 2543  | 72     | 89        | 37         | 85  | 75     | 56    | 0.95%  | 1.97%      | -1.13%     | 1.73%  | 1.13%  |
| 18  | 2305  | 57     | 58        | 59         | 27  | 29     | 5     | 3.42%  | 3.48%      | 3.55%      | 1.45%  | 1.58%  |
| 19  | 1769  | 79     | 49        | 33         | 43  | 35     | 12    | 5.74%  | 3.17%      | 1.80%      | 2.66%  | 1.97%  |
| 20  | 2492  | 107    | 29        | 62         | 56  | 43     | 12    | 5.78%  | 1.03%      | 3.04%      | 2.68%  | 1.88%  |
| 21  | 1026  | 36     | 53        | 22         | 8   | 11     | 10    | 3.84%  | 6.35%      | 1.77%      | -0.30% | 0.15%  |
| 22  | 1455  | 79     | 33        | 66         | 34  | 18     | 5     | 7.71%  | 2.92%      | 6.35%      | 3.02%  | 1.35%  |
| All | 78055 | 4231   | 2071      | 1606       | 1807 | 1518   | 485   | 7.27%  | 3.08%      | 2.18%      | 2.57%  | “2%”   |

The analysis is based on 78,055 sites covered by at least one deaminated fragment in Oase 1. To convert the counts to estimates of ancestry, we subtract the Dinka count as an estimate of the false positive rate and divide by the number of sites covered (as indicated for the whole genome on the bottom). This gives the rate of alleles per screened site on this chromosome for this individual. We then multiply this quantity by 2%/1.32% to recalibrate the 1.32% seen genome-wide in the French to an assumed 2% genome-wide Neanderthal ancestry in the French.”
| Library | Extract | UDG treatment | Index 1 | Index 2 | Extract used (µl) | Sequencing results | All fragments | Deaminated fragments |
|---------|---------|---------------|---------|---------|------------------|-------------------|---------------|---------------------|
|         |         |               |         |         |                  | Sequences going into alignment | Sequences ≥35bp mapped | After dup. removal | Cov.-age  | % C→T  | % C→T  | % C→T  | Cov.-age  | % 5' end | % 3' end | % 5' end | % 3' end |
| A5227   | E1406   | Yes           | ACTGCG  | AACTCCG | 8                | 206,982           | 118,976       | 34,486   | 112       | 8       | 19       | 5       | 19       | 36       |
| A5252   | E1843   | Yes           | GTAGGC  | TGAAGT  | 40               | 74,384            | 46,394        | 31,368   | 114       | 7       | 25       | 5       | 18       | 55       |
| A9032   | E1406   | No            | ATAACG  | ACTATCA | 6                | 9,321,903        | 5,904,210    | 51,810   | 178       | 20      | 21       | 12      | 31       | 39       |
| A9033   | E1406   | No            | AATAGGA| ACCAACT | 6                | 7,932,271        | 4,816,314    | 55,878   | 193       | 21      | 20       | 13      | 36       | 38       |
| A9034   | E1406   | No            | ATCCAGA| AACTCCG | 6                | 10,422,467       | 6,861,634    | 59,883   | 207       | 20      | 20       | 14      | 35       | 38       |
|         |         |               |         |         |                  | 27,958,007       | 17,747,528   | 233,425  | 803       | 17      | 21       | 49      | 30       | 39       |
Extended Data Table 7 | Sequencing metrics on the five libraries for the four capture probe panels

| Library | Panel   | No. target SNPs | Fragments going into alignment | Fragments mapped to genome | Fragments on target after dup. removal and MAPQ37 filter | % SNPs hit at least once | Average coverage on SNPs |
|---------|---------|----------------|--------------------------------|---------------------------|--------------------------------------------------------|--------------------------|---------------------------|
| A9032   | 390k    | 393,577        | 10,849,144                     | 2,235,955                 | 133,564                                                | 26.5%                    | 0.34                      |
| A9033   | 390k    | 393,577        | 17,159,085                     | 2,808,704                 | 73,824                                                 | 15.9%                    | 0.19                      |
| A9034   | 390k    | 393,577        | 16,902,935                     | 3,256,438                 | 142,520                                                | 27.7%                    | 0.36                      |
| A5227   | 390k    | 393,577        | 63,441,719                     | 22,124,247                | 195,161                                                | 36.0%                    | 0.5                       |
| A5252   | 390k    | 393,577        | 60,181,844                     | 14,278,978                | 180,626                                                | 33.3%                    | 0.46                      |
| All 5   | 390k    | 393,577        | 168,534,727                    | 44,704,322                | 724,653                                                | 73.0%                    | 1.84                      |
| A9032   | 840k    | 842,630        | 25,105,625                     | 3,801,435                 | 178,015                                                | 17.6%                    | 0.21                      |
| A9033   | 840k    | 842,630        | 29,196,969                     | 4,655,434                 | 183,093                                                | 17.9%                    | 0.22                      |
| A9034   | 840k    | 842,630        | 35,780,652                     | 5,968,851                 | 200,767                                                | 19.3%                    | 0.24                      |
| A5227   | 840k    | 842,630        | 28,209,496                     | 4,276,439                 | 152,411                                                | 15.3%                    | 0.18                      |
| A5252   | 840k    | 842,630        | 20,286,540                     | 1,630,343                 | 106,943                                                | 11.2%                    | 0.13                      |
| All 5   | 840k    | 842,630        | 138,579,282                    | 20,332,502                | 818,648                                                | 51.7%                    | 0.97                      |
| A9032   | 1000k   | 997,780        | 26,088,835                     | 2,964,094                 | 159,162                                                | 13.5%                    | 0.16                      |
| A9033   | 1000k   | 997,780        | 26,641,358                     | 4,490,372                 | 158,614                                                | 13.3%                    | 0.16                      |
| A9034   | 1000k   | 997,780        | 28,795,043                     | 4,985,140                 | 154,177                                                | 13.0%                    | 0.15                      |
| A5227   | 1000k   | 997,780        | 25,848,311                     | 4,395,413                 | 71,537                                                 | 6.4%                     | 0.07                      |
| A5252   | 1000k   | 997,780        | 25,691,323                     | 2,254,636                 | 53,932                                                 | 5.0%                     | 0.05                      |
| All 5   | 1000k   | 997,780        | 133,064,870                    | 19,089,655                | 596,107                                                | 36.1%                    | 0.6                       |
| A9032   | Archaic | 1,749,385      | 19,329,832                     | 2,086,208                 | 205,095                                                | 10.0%                    | 0.12                      |
| A9033   | Archaic | 1,749,385      | 24,629,023                     | 2,768,355                 | 237,818                                                | 11.4%                    | 0.14                      |
| A9034   | Archaic | 1,749,385      | 31,200,466                     | 3,783,805                 | 257,351                                                | 12.2%                    | 0.15                      |
| A5227   | Archaic | 1,749,385      | 27,659,125                     | 3,606,375                 | 195,356                                                | 9.6%                     | 0.11                      |
| A5252   | Archaic | 1,749,385      | 31,472,143                     | 2,435,080                 | 136,637                                                | 6.8%                     | 0.08                      |
| All 5   | Archaic | 1,749,385      | 134,290,589                    | 14,679,823                | 1,022,046                                              | 34.6%                    | 0.58                      |
| A9032   | Combined | 3,801,245     | 81,373,436                    | 11,087,692                | 719,146                                                | 15.5%                    | 0.19                      |
| A9033   | Combined | 3,801,245     | 97,626,435                    | 14,722,865                | 698,890                                                | 15.1%                    | 0.18                      |
| A9034   | Combined | 3,801,245     | 112,679,096                   | 17,994,234                | 806,589                                                | 17.0%                    | 0.21                      |
| A5227   | Combined | 3,801,245     | 145,158,651                   | 34,402,474                | 666,195                                                | 14.2%                    | 0.18                      |
| A5252   | Combined | 3,801,245     | 137,631,850                   | 20,599,037                | 531,873                                                | 11.4%                    | 0.14                      |
| All 5   | Combined | 3,801,245     | 574,469,468                   | 98,806,302                | 3,406,685                                              | 45.3%                    | 0.90                      |
### Extended Data Table 8 | Effect of filtering on amount of nuclear data available

| Panel          | Target SNPs | All fragments | Deaminated fragments only |
|----------------|-------------|---------------|---------------------------|
|                |             | No. SNPs hit ≥1× | % SNPs hit ≥1× | Average coverage | No. SNPs hit ≥1× | % SNPs hit ≥1× | Average coverage |
| Panels 1-3     | 2,051,902   | 1,038,619      | 50.6%        | 1.03            | 271,326          | 13.2%        | 0.16             |
| Panel 4 subset* | 954,849     | 361,681        | 37.9%        | 0.69            | 87,803           | 9.2%         | 0.11             |
| Panels 1-4     | 3,801,245   | 1,685,891      | 44.4%        | 0.85            | 426,027          | 11.2%        | 0.13             |

Note that numbers differ from Extended Data Table 7 because only sites with base quality ≥20 were used.

* The Panel 4 subset excludes the sites where only the Denisovan genome differs from the African panel.
Extended Data Table 9 | Genomes merged with the Oase 1 data

| Sample ID       | Human     | Data type     | Mean | UDG-treated          |
|-----------------|-----------|---------------|------|----------------------|
| Oase1           | Modern    | Low coverage  | Capture | Mix of library types |
| Vindija         | Archaic   | Low coverage  | 1.3   | No                   |
| Mezmaiskaya     | Archaic   | Low coverage  | 0.5   | Yes                  |
| Altai           | Archaic   | High coverage | 52    | Yes                  |
| Denisova        | Archaic   | High coverage | 31    | Yes                  |
| Kostenki14      | Modern    | Low coverage  | 2.4   | Mix of library types |
| MA1             | Modern    | Low coverage  | 1     | No                   |
| LaBrana         | Modern    | Low coverage  | 3.4   | No                   |
| Loschbour       | Modern    | High coverage | 22    | Yes                  |
| Stuttgart       | Modern    | High coverage | 19    | Yes                  |
| Ust’-Ishim      | Modern    | High coverage | 42    | Yes                  |
| Dinka_A         | Modern    | High coverage | 28    | ..                   |
| French_A        | Modern    | High coverage | 27    | ..                   |
| Papuan_A        | Modern    | High coverage | 26    | ..                   |
| Sardinian_A     | Modern    | High coverage | 25    | ..                   |
| Han_A           | Modern    | High coverage | 28    | ..                   |
| Yoruba_A        | Modern    | High coverage | 32    | ..                   |
| Karitiana_A     | Modern    | High coverage | 26    | ..                   |
| San_A           | Modern    | High coverage | 33    | ..                   |
| Mandenka_A      | Modern    | High coverage | 25    | ..                   |
| Dair_A          | Modern    | High coverage | 28    | ..                   |
| Mbuti_A         | Modern    | High coverage | 24    | ..                   |
| Dair_B          | Modern    | High coverage | 37    | ..                   |
| French_B        | Modern    | High coverage | 42    | ..                   |
| Han_B           | Modern    | High coverage | 35    | ..                   |
| Mandenka_B      | Modern    | High coverage | 37    | ..                   |
| Mbuti_B         | Modern    | High coverage | 37    | ..                   |
| Papuan_B        | Modern    | High coverage | 42    | ..                   |
| San_B           | Modern    | High coverage | 38    | ..                   |
| Sardinian_B     | Modern    | High coverage | 38    | ..                   |
| Yoruba_B        | Modern    | High coverage | 39    | ..                   |
| Karitiana_B     | Modern    | High coverage | 35    | ..                   |
| Mixe_B          | Modern    | High coverage | 42    | ..                   |
| Australian_B1   | Modern    | High coverage | 42    | ..                   |
| Australian_B2   | Modern    | High coverage | 37    | ..                   |
| Dinka_B         | Modern    | High coverage | 35    | ..                   |

For the 25 present-day humans, individuals ending with a subscript 'A' are from ‘Panel A’ reported in ref. 9 and individuals with a subscript ‘B’ are from ‘Panel B’ reported in ref. 7. Unless otherwise specified, we used Panel B individuals.