Paleogenetic investigations of hominin diversity and dispersals in Eurasian prehistory

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Cosimo Posth

aus Prato, Italien

Tübingen

2016

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	10.02.2017
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Johannes Krause
2. Berichterstatter:	Prof. Nicholas J. Conard, PhD

Acknowledgements

I would like to express my gratitude to several people who, over the last four years, supported me throughout this doctorate marathon.
To Johannes Krause, always source of inspiration and reliable supervisor. I am truly glad to carry on working together.
To Nick Conard, for supporting me during a wonderful PhD experience in Germany.
To Wolfgang Haak, for the translation of the summary and many fruitful discussions and to Adam Powell, not only for the working nights. I am lucky to have encountered both of you before the end of my dissertation.
To Maria Spyrou, for the thorough proofreading, your warmth and our stimulating relationship. You have all my admiration.
To Michal and Immel, for all private and professional exchanges, Kirsti and Alexander for the helpful suggestions and Peltzer, Alissa, Verena, Ash, Kerttu for the fun atmosphere and activities in Tübingen.
To the team at the INA in Tübingen and at the MPI SHH in Jena, it was and it is an honor to be part of these groups.
To my boys: Mohammed, Ali, Alvise and Cris, each of you represented a reference point during those years.
To the members of various WGs where I lived and learned most of the German culture.
To my sister Betta who, with my nephews Sirio and Elmo, gives me reasons to go back home more often and makes me a proud uncle.
To my sister Vero, not only geographically close, who constantly introduces me to the latest art scene.
To aunt Grazia, for our lovely calls and her joyful character.
And to Carla and Enrico, for all you taught and are still teaching me.

It has been an intense and exciting run. Enjoy the read!

Table of Contents

Abbreviations

A	Adenine
ABC	Approximate Bayesian Computation
aDNA	ancient DNA
bp	base pairs
C	Cytosine
°C	degree Celsius
D-loop	mitochondrial displacement loop
DNA	deoxyribonucleic acid
G	Guanine
GUI	graphical user interface
hg	haplogroup
HPD	high probability density
HVR-I	mitochondrial hypervariable region 1
ka	thousand years ago
LGM	Last Glacial Maximum
M	million
MSMC	multiple sequentially Makovian coalescent
mtDNA	mitochondrial DNA
nDNA	nuclear DNA
NGS	next generation sequencing
PCA	principal component analysis
PCR	polymerase chain reaction
qPCR	quantitative PCR
SNP	single nucleotide polymorphism
T	Thymine
TMRCA	time to the most recent common ancestor
UDG	Uracil DNA glycosylase
UV	ultraviolet

Summary

Ancient DNA (aDNA) is able to provide genetic snapshots into the human past that can be linked together to study evolutionary processes and demographic patterns impossible to uncover with the study of modern-day DNA alone. In this thesis I make use of major methodological “game changers” in the field of aDNA in order to reconstruct complete mitochondrial DNA (mtDNA), as well as genome-wide nuclear data (nDNA) from ancient human specimens. The combination of next generation sequencing (NGS) technologies and target enrichment paired up with sampling of different anatomical elements, enabled me to expand the amount of analyzable hominin remains ranging from Pleistocene European hunter-gatherers to Holocene farmers in Remote Oceania. I first investigated the mtDNA of late Neandertals from Goyet cave in Belgium and of an archaic femur from Hohlenstein-Stadel in southwest Germany to explore the changes in genetic diversity of this extinct hominins through time and provide the temporal interval for a putative African gene flow event into Neandertal populations. In addition, I carried out two studies that explored demographic changes in European Upper Paleolithic and Mesolithic hunter-gatherers by means of both mtDNA and nDNA, and revealed population structure and unexpected genetic turnovers in Ice Age Europe. By expanding the temporal and geographical distribution of genomic data it was able to infer population movements in European prehistory and compare them to climatic and archaeological records over almost 40,000 years. While the formation of some genetic clusters tightly matches to the associated archaeological changes across Europe, other major genomic transformations seem to be more influenced by environmental fluctuations. In the last project, I contributed in producing aDNA of four individuals among the first settlers of Vanuatu and Tonga in the Southwest Pacific. Sampling the petrous portion of their temporal bones allowed me to retrieve genomic data from climatic conditions unfavorable for DNA preservation. Those genomes highlight the role of previously unknown dispersals in shaping the ancestry of present-day people in Remote Oceania. Here, I take a time trip to shed light into the genetic history of our ancestors and closest extinct relatives.

Zusammenfassung

Die Analyse alter DNA (aDNA) erlaubt Einblicke in die genetische Vergangenheit der Menschheit und hiermit die Möglichkeit Entwicklungsprozesse und demografische Muster zu untersuchen, welche aus Sicht der Genetik rezenter Populationen verborgen bleiben. In meiner Doktorarbeit wende ich neueste methodische Errungenschaften an, um ganze mitochondriale Genome sowie genom-weite Daten prähistorischer Individuen zu rekonstruieren. Die gezielte Anreicherung menschlicher DNA durch Hybridisierungmethoden gepaart mit *Next Generation* Sequenzierungs-technologie erlaubte es mir die Anzahl erfolgreich analysierter menschlicher Skelettreste um altsteinzeitliche Jaeger-Sammler sowie die der frühesten Siedler des südwestlichen Pazifiks zu erweitern. Zunächst gelang mir die erfolgreiche genetische Untersuchung von späten Neandertalern aus der Goyet Höhle in Belgien sowie eines archaischen Femur aus der Hohlenstein-Stadel Höhle, SW Deutschland. Diese beiden Studien ermöglichten einen diachronen Einblick in die genetische Diversität archaischer Homininen und lieferten zudem eine mögliche Zeitspanne für genetischen Einfluss aus Afrika in Neandertaler-Gruppen Europas. In zwei weiteren Arbeiten zu mitochondrialer sowie Kern-DNA konnte ich überraschenderweise mehrfach genetische und demografische Umwälzungen in Europas Altsteinzeit nachweisen. Eine sehr gute geographische sowie zeitliche Auflösung an Fundmaterial der frühen Besiedlungsgeschichte Europas erlaubte es mir zudem die genetischen Daten mit archäologischem Fundkontext sowie geologischen bzw. paläo-klimatischen Daten abzugleichen und hiermit die Siedlungsdynamik früher Europäer zu erfassen. Interessanterweise lassen sich genetische *cluster* mit archäologischen Sachkulturen bzw. deren Änderungen durch die Zeit verbinden, während andere Veränderungen eher in Einklang mit den dramatischen Klimafluktuationen der letzten 40,000 Jahre gebracht werden können. In meinem letzten Kapitel beschreibe ich meinen Beitrag zu einer Arbeit, welche die Erstbesiedlung der südpazifischen Inseln Vanuatu und Tonga untersucht hat. Uns gelang die Generierung von genomweiten Daten aus den Felsenbeinen von vier Individuen aus tropischen Regionen – ein Unterfangen,

das bis dahin unmöglich schien. Die genetischen Ergebnisse werfen neues Licht auf die Besiedlungsgeschichte und die genetische Komplexität in der Genese der Bewohner der südwestpazifischen Inseln in Ozeanien. Mit dieser Arbeit gelingt mir ein zeitlich wie geographisch weit gefächerter Einblick in die genetische Herkunft unserer Vorfahren sowie eng mit uns verwandter, aber heute ausgestorbener, menschlicher Entwicklungslinien.

List of publications

The following five publications are included and discussed in this dissertation:

1. **C. Posth**, G. Renaud, A. Mittnik, D. G. Drucker, H. Rougier, C. Cupillard, F. Valentin, C. Thevenet, A. Furtwangler, C. Wissing, M. Francken, M. Malina, M. Bolus, M. Lari, E. Gigli, G. Capecchi, I. Crevecoeur, C. Beauval, D. Flas, M. Germonpre, J. van der Plicht, R. Cottiaux, B. Gely, A. Ronchitelli, K. Wehrberger, D. Grigorescu, J. Svoboda, P. Semal, D. Caramelli, H. Bocherens, K. Harvati, N. J. Conard, W. Haak, A. Powell and J. Krause (2016). "Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe." *Current Biology* 26(6): 827-833.

2. Q. Fu, **C. Posth***, M. Hajdinjak*, M. Petr, S. Mallick, D. Fernandes, A. Furtwangler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gely, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Paabo and D. Reich (2016). "The genetic history of Ice Age Europe." *Nature* 534(7606): 200-205.
   *equal contributors

3. H. Rougier, I. Crevecoeur, C. Beauval, **C. Posth**, D. Flas, C. Wissing, A. Furtwangler, M. Germonpre, A. Gomez-Olivencia, P. Semal, J. van der Plicht, H. Bocherens and J. Krause (2016). "Neandertal cannibalism and Neandertal bones used as tools in Northern Europe." *Scientific Reports* **6**: 29005.

4. P. Skoglund, **C. Posth**, K. Sirak, M. Spriggs, F. Valentin, S. Bedford, G. R. Clark, C. Reepmeyer, F. Petchey, D. Fernandes, Q. Fu, E. Harney, M. Lipson, S. Mallick, M. Novak, N. Rohland, K. Stewardson, S. Abdullah, M. P. Cox, F. R. Friedlaender, J. S. Friedlaender, T. Kivisild, G. Koki, P. Kusuma, D. A. Merriwether, F.-X. Ricaut, J. T. S. Wee, N. Patterson, J. Krause, R. Pinhasi and D. Reich (2016). "Genomic insights into the peopling of the Southwest Pacific." *Nature* 538(7626): 510-513.

5. **C. Posth**, C. Wißing, K. Kitagawa, L. Pagani, L. van Holstein, K. Wehrberger, N. J. Conard, C. J. Kind, H. Bocherens, J. Krause (2016). "Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neandertals". Review in *Nature Communications*.

Reprints were made available with permission of the respective publishers.

Own contributions

1. (Posth et al., 2016a) I performed lab work for 31 of the 35 mtDNA sequences newly reported in this study, two by the colleague Alissa Mittnik and two by the student assistant Anja Furtwängler, the latter whom I supervised. I reconstructed mtDNAs together with Gabriel Renaud, now at the Center for Geogenetics in Copenhagen and Wolfgang Haak, team leader at the Max Plank Institute for the Science of Human History (MPI SHH) in Jena. I conducted data analyses, generated figures and tables and wrote the manuscript with Adam Powell, team leader at the (MPI SHH) in Jena, under the supervision of Johannes Krause.

2. (Fu et al., 2016) I performed lab work for 22 of the 38 individuals with nuclear DNA data newly reported in this study, partially supported by the student assistant Anja Furtwängler, edited the main text and co-wrote Section 1 and Section 2 in the Supplementary Information.

3. (Rougier et al., 2016) I performed lab work for the ten mtDNA sequences reported in this study with the support from the student assistant Anja Furtwängler. I conducted genetic data analyses, wrote related method section, edited the main text and generated Figure 2. I also created Table S7, Figures S4, S17, S18 and S19 and wrote Note S7 in the Supplementary Information, under the supervision of Johannes Krause.

4. (Skoglund et al., 2016) I performed lab work for one of the four individuals with nuclear DNA data newly reported in this study, edited the main text and Supplementary Information.

5. (Posth et al., 2016b) I performed lab work for the mtDNA sequence reported in this study, conducted data analyses, generated figures (except Supplementary Figure S1 and S2) and tables (except Supplementary Table S1 and S8) and wrote the manuscript, with the input of all authors and under the supervision of Johannes Krause.

1 Introduction

1.1 Game changers in ancient DNA research

In recent years, next generation sequencing technologies applied to ancient DNA (aDNA) dramatically expanded our understanding of archaic hominin evolution and modern human population dynamics through time. From the analyses of short mitochondrial DNA fragments (e.g. Krings et al., 1997) we rapidly moved to the study of entire ancient hominin genomes (e.g. Prufer et al., 2014) that allowed the characterization of the genetic makeup of modern humans and our closest extinct relatives across the Middle and Late Pleistocene (Meyer et al., 2016). Rapid technological developments within the field of aDNA exponentially enlarged the sample size of remains from which genome-wide data can be retrieved, making paleogenomic studies not limited only to the analyses of a few specimens with outstanding DNA preservation. Major methodological innovations involved the identification of anatomical elements with higher proportion of endogenous DNA content, the design of specific protocols to extract and enrich for specific genomic DNA portions and the construction of computational pipelines appositely designed for aDNA processing. The combination of those improved techniques expanded the research field, and allowed to analyze specimens from periods further back in time and from a wider geographical range, even where environmental conditions heavily affect aDNA preservation. In this section, I highlight the major methodological advances that allowed overcoming the main challenges in retrieving authentic genomic data from hominin specimens spanning from archaic times and cold climates to more recent periods and tropical environments.

1.1.1 High-throughput DNA sequencing

For the vast majority of the time since it was established, the aDNA field has relied on PCR (polymerase chain reaction) methods to target and amplify genetic regions of interest. The most commonly analyzed region in human DNA is the ~360bp HVR-I (hypervariable region 1) in the non-coding D-loop portion of the

mtDNA. The main reasons for this choice are that mtDNA is inherited from mother to offspring and it does not recombine and it is present in multiple copies within every mitochondrion and several mitochondria are present in each cell, providing a higher copy count per sample in comparison to nuclear DNA. Moreover, HVR-I is a highly polymorphic region and distinct substitutions allow the assignment of even a short sequence to distinct mtDNA haplogroups (hg). This method is still in use for assessing maternal relatedness for example within a cemetery (Alt et al., 2014) or to monitor changes in hg frequencies through time (Brandt et al., 2013). However, its resolution is limited by the few informative positions available in respect to the whole mtDNA with its total of ~16,500 bp or when compared to the around three billion nucleotides in the nuclear genome. Additionally, there are two major limitation factors associated to the use of PCR technology. First, a typical feature of aDNA is its high fragmentation to average fragment lengths often below 60bp (Sawyer et al., 2012), while regions targeted with PCR are usually longer than that, effectively favoring the retrieval of better preserved present-day human contaminant molecules (Krause et al., 2010b). Second, despite stringent authentication criteria were introduced in the early 2000s (Cooper and Poinar, 2000, Paabo et al., 2004), no quantitative methods are available to estimate levels of modern-day human contamination in isolated DNA fragments.

The first major transformation in the field happened around 10 years ago with the application of NGS technologies to aDNA. While early studies relied on the Roche/454 platform (Green et al., 2006), the leading sequencing platform currently is Illumina for which several aDNA custom protocols have been adapted (Meyer and Kircher, 2010, Gansauge and Meyer, 2013, Gansauge and Meyer, 2014, Rohland et al., 2015). Following those methodologies genetic libraries are built by ligating two known oligonucleotides at the end of each aDNA molecule allowing the retrieval of the entire genetic information contained in an aDNA extract, therefore not limited to single targeted regions. Moreover, each library gets an individual and unique dual or quadruple index/barcode combination that permits to distinguish one library from the other (Kircher et al., 2012, Rohland et

al., 2015) and avoid post-amplification cross-contaminations (Green et al., 2010). After sequencing several hundred million of sequences (also called reads) per run, authentication criteria other than sequence length profiles have been established (Green et al., 2009). Once mapped against a reference sequence, aDNA reads in contrast to modern contaminants were found to exhibit typical substitution patterns towards the molecule ends caused by post mortem damage (Briggs et al., 2007, Krause et al., 2010a), a feature that tend to accumulate over time (Sawyer et al., 2012). Moreover, methods that estimate the percentage of human contamination have been introduced mainly for mtDNA and X chromosome in males. Making use of their haploid state, DNA coming from different sources can be clearly detected (Fu et al., 2013a, Korneliussen et al., 2014, Renaud et al., 2015). Using NGS data not fulfilling the described guidelines resulted in publications in high impact journals to be questioned as the main conclusions might have been heavily affected by contamination (Prufer and Meyer, 2015, Weiss et al., 2015).

The application of NGS technology paired up with the aforementioned authentication criteria rapidly boosted the field, starting with the production of complete mtDNAs (Green et al., 2008, Krause et al., 2010a) to entire genomes of archaic (Green et al., 2010, Reich et al., 2010) and ancient modern humans (Rasmussen et al., 2010). Those later studies were produced through a “shotgun” sequencing approach where library molecules are randomly sequenced without the implementation of any targeting technique. Genetic libraries with exceptional preservation in terms of percentage of hominin endogenous DNA further allowed to produce full ancient genomes with high coverage (above 20-fold) in order to reliably perform diploid calls at a quality comparable to that of present-day human genomes (Meyer et al., 2012, Fu et al., 2014, Lazaridis et al., 2014, Prufer et al., 2014).

1.1.2 Fishing out

The vast majority of human remains exhibit minimal proportions of endogenous DNA after being buried in the ground for thousands of years. DNA extracted from

human teeth and bones is usually a pool of DNA from different organisms where human DNA is often represented by less than 1% of the total molecules. Therefore, in most cases a deep shotgun DNA sequencing approach is not economical and this results in excluding the vast majority of screened specimens from further analyses (Allentoft et al., 2015). Alternative approaches have therefore been designed to target specific regions in the human genome, to fish them out from genetic libraries and wash out the non-target molecules. Bead-capture techniques were designed to capture entire mtDNA sequences (Briggs et al., 2009, Maricic et al., 2010) as well as methods aiming to enrich wider portions of the nuclear genome through array hybridization (Burbano et al., 2010) or in-solution enrichment (Fu et al., 2013b). The latter technique was successfully adapted to capture up to ~390,000 informative single nucleotide polymorphisms (SNPs) scattered across the genome of 69 individuals who lived in Europe between 8 ka and 3 ka (Haak et al., 2015). In a follow-up up study that analyzed 230 ancient Western Eurasians for signatures of positive selection, it was shown that capture was around 50 times more efficient than the shotgun sequencing approach and therefore more cost efficient (Mathieson et al., 2015).

1.1.3 The sweet spot

Despite great advances in sequencing technologies and sophisticated laboratory procedures to extract and increase the proportion of human DNA, retrieving DNA from extremely old specimens, such as the Middle Pleistocene hominin from Sima de los Heusos (Meyer et al., 2014, Meyer et al., 2016) or from hot and humid environments, such a tropical climates (Schroeder et al., 2015) is still an extremely challenging task. DNA degrades over time and gets fragmented into short molecules. This process is enhanced by the presence of water that causes hydrolytic reactions, mainly as depurinations, forming abasic sites where the DNA sugar-phosphate backbone tends to break (Dabney et al., 2013b). Moreover, an inverse correlation between temperature and DNA half-life (i.e. the time in which DNA molecule of a certain length get fragmented) was confirmed studying the decay kinetics in mtDNA sequences of extinct New Zealand moa

(Allentoft et al., 2012). Those studies highlight the limitations in analyzing aDNA from humid and warm climates where water and heat increase the degradation rate and reduce the chances of aDNA retrieval. It is probably not surprising that the vast majority of genome-wide hominin data was obtained from archaeological sites located at high latitudes (Slatkin and Racimo, 2016) where colder climatic conditions favor aDNA preservation. However, going forward in time from the past, the fossil record tends to increase not only in the number of excavated individuals but also in their skeletal completeness. Comparative DNA analyses between different skeletal elements belonging to the same individual showed that a cranial element, the petrous portion of the temporal bone, consistently preserved the highest percentage of human DNA (Gamba et al., 2014). In particular within the petrous bone the densest part, the inner ear, was revealed to be the area with the highest DNA yields, potentially enabling aDNA retrieval even from tropical regions (Pinhasi et al., 2015). The implementation of such sampling procedures coupled with nDNA capture, provided recently the opportunity to obtain genome-wide data of 44 Near Easterners spanning from Natufian hunter-gatherers to Bronze Age farmers (Lazaridis et al., 2016).

The joint application of NGS technologies, capture techniques and, when available, sampling of the petrous bone drastically advanced aDNA research making paleogenomic investigations further back in time and beyond temperate frontiers possible.

1.2 Archaeogenetic perspectives on Paleolithic hominins

1.2.1 Eurasian archaic humans

The first aDNA study of archaic hominins was performed on the right humerus of the Neandertal-type specimen Feldhofer1 found in 1856 in western Germany (Krings et al., 1997). PCR and cloning techniques were implemented to sequence its mtDNA HVR-I resulting in a sequence that falls outside of modern human mtDNA variation. In the following years several studies reported additional Neandertal-like HVR-I confirming its distinctiveness from modern-day

as well as early modern humans thus excluding the possibility of a large proportion of Neandertal admixture into modern humans (Currat and Excoffier, 2004, Serre et al., 2004). Since the advent of NGS technologies (excluding for now the work in this thesis) eleven entire Neandertal mtDNAs have been reconstructed (Green et al., 2008, Briggs et al., 2009, Gansauge and Meyer, 2014, Prufer et al., 2014, Skoglund et al., 2014) depicting an overall lower mtDNA diversity than found in modern humans, even when accounting for possible biases in temporal sampling (Briggs et al., 2009). In the absence of a Neandertal-specific mtDNA mutation rate, ancient modern human and Neandertal mtDNAs from dated individuals were used as tip calibration points to assess a combined molecular clock resulting in an estimate of ~400 ka for the divergence time between Neandertal and modern human mtDNAs (Fu et al., 2013a, Rieux et al., 2014). In 2010, the first draft of the Neandertal genome was produced by combining DNA sequences from three individuals from Vindija cave in Croatia (Green et al., 2010). The main finding of this study was a greater genetic affinity of Neandertals with modern-day humans in Eurasia than with sub-Saharan Africans, suggesting Neandertal admixture into the ancestors of all non-Africans. In the same year, a distinct mtDNA sequence of an unknown hominin from Denisova cave in the Altai Mountains was discovered (Krause et al., 2010b). Its divergence time from the Neandertal-modern human branch was estimated to be around one million years ago, suggesting a distinct migration for the Denisova hominin ancestors out of Africa. Shotgun sequencing DNA extracted from the same finger bone from which the mtDNA sequence was retrieved, provided a low coverage genome of this individual (Reich et al., 2010). The amount of data available together with a similar mtDNA from a molar found at the same site was enough to designate those specimens as belonging to a new hominin group called “Denisovans”. Intriguingly, a substantial genetic contribution of this population was identified in present-day New Guineans, Australians and Mamanwa (Philippines) far away from the Siberian Denisova cave and implying a broad geographical extension of Denisovans (Reich et al., 2011). Moreover, from genome-wide data it became evident that Denisovans

were a sister group of Neandertals sharing a common ancestors more recently with Neandertals than with modern humans (Reich et al., 2010). In following studies high coverage genomes for both the Denisovan finger bone and a Neandertal individual from Denisova cave confirmed the previously described evolutionary history for the two archaic groups and defined more accurately the time to the most recent common ancestor (TMRCA) between archaic and modern humans at 765 - 550 ka (Meyer et al., 2012, Prufer et al., 2014). This was calculated using a human mutation rate of 0.5×10^{-9} per bp per year based on direct substitution measurement per generation (Kong et al., 2012), that is twice as slow as the one calculated with a human-chimpanzee split time set at 6.5 M years (Green et al., 2010) and which was later supported in a study that analyzed a 45,000-year-old genome from Siberia, using the age of the fossil and the missing substitutions from present-day human genomes (Fu et al., 2014). Furthermore, the recently measured divergence time between Neandertal and modern human Y chromosomes widely confirmed the population split time estimated from autosomal data (Mendez et al., 2016). On the other hand, as previously mentioned, the equivalent mtDNA coalescence time resulted in almost twofold younger divergence time and is thus incompatible with the nuclear estimates. Moreover, while the mtDNA of Middle Pleistocene individuals from Sima de los Huesos in Spain (~430 ka) fell on the Denisovan lineage (Meyer et al., 2014), their nDNA showed closer genetic affiliations to Neandertals possibly resembling early forms of this hominin group (Meyer et al., 2016). This led the authors to hypothesize that the “typical” Late Pleistocene Neandertal mtDNA lineage was acquired via a gene flow event from Africa (Meyer et al., 2016), summarized here in Figure 1. Indeed, assuming that the Denisovan-like mtDNA type was initially common among early Neandertals in Eurasia (e.g. Sima de los Huesos), it would have then been largely replaced by an introgressing mtDNA. The admixture event should be subsequent to both the TMRCA between Late Pleistocene Neandertal and modern human mtDNA (~400 ka) and the nDNA split of the Sima de los Huesos hominins within the Neandertal population (at least ~430 ka). An interesting observation came from the re-analyses of the high

coverage Neandertal genome from the Altai region that compared to chromosome 21 capture data of El Sidrón and Vindija Neandertals revealed the presence of a small proportion (0.1 - 2.1 %) of genetic admixture in Eastern Neandertals coming from an early divergent modern human African population (Kuhlwilm et al., 2016). The temporal range of the interbreeding event was estimated to 100 - 230 ka based on recombination rate and length of African haplotypes or after ~110 ka (167 – 68 ka) representing the separation time of Altai Neandertal from El Sidrón/Vindija branch. However, analyses were performed on individuals with quantitatively and qualitatively different genomic data that might have affected the dating calculations. The Max Planck Institute for Evolutionary Anthropology in Leipzig just announced the sequencing of a new high coverage Neandertal genome that might possibly help in assessing the distribution of African gene flow into Neandertal populations.

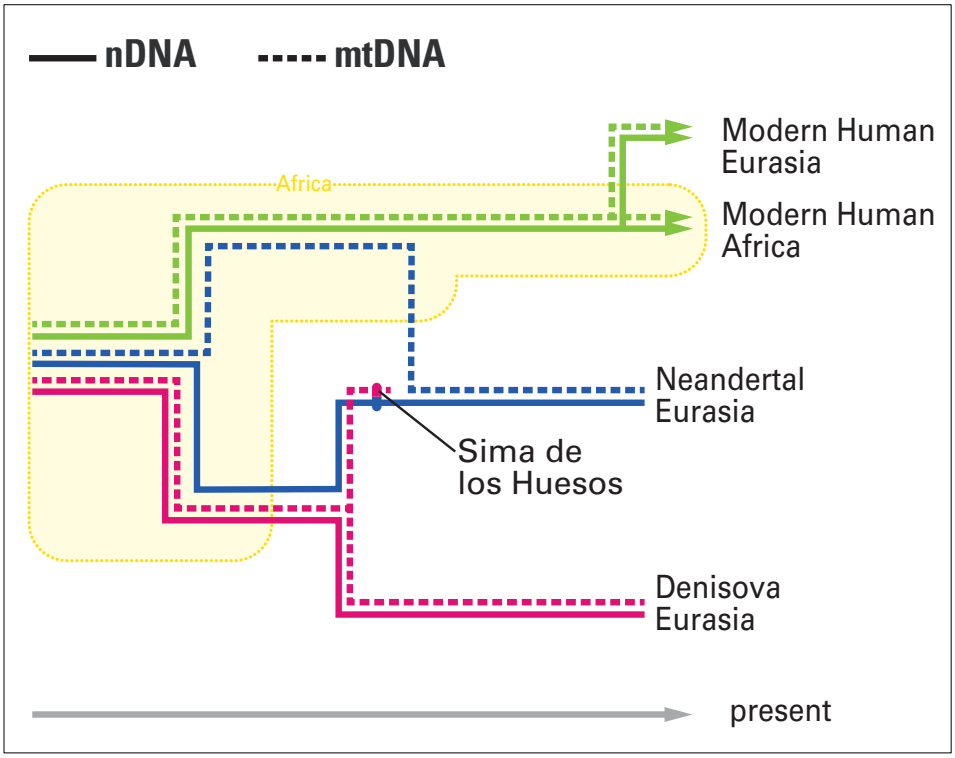


Figure 1: Schematic comparison of archaic and modern human mtDNA and nDNA phylogenies. The reconciling evolutionary scenario includes a mtDNA introgression from Africa into Neandertals. Graphical support by Annette Günzel.

1.2.2 Non-African modern humans

A highly debated topic among the fields of archaeology, paleoanthropology and genetics concerns the out of Africa and following dispersals of modern humans around the globe (Scally and Durbin, 2012). A pivotal early study identified Africa as the modern human homeland since it is where the deepest divergent mtDNA lineages are found today (Cann et al., 1987). Broadly, two sets of models that describe subsequent dispersal(s) of our species outside Africa have been proposed depending on how many expansions reaching Eurasia and Australasia are interpreted as successful i.e. contributed to at least part of the genetic make-up of modern-day people. The first category includes the Single Dispersal models according to which a single major expansion from Africa left descendants in the rest of the world (Oppenheimer, 2012, Mellars et al., 2013). Those phylogeographic studies addressed the different hypotheses by analyzing the distribution of present-day African and non-African mtDNA haplogroups linking them to their coalescence ages using a mtDNA mutation rate calibrated with the chimpanzee-human split time (Soares et al., 2009). An aDNA work used instead ancient radiocarbon dated mtDNA genomes as tip calibration points to obtain a ~1.6-fold higher substitution rate (Fu et al., 2013a). This dated the split of hg L3, the lineage giving rise to all non-African mtDNA hgs (M and N, which includes R hg), to 95 - 62 ka that was interpreted as a conservative upper bound for the out of Africa event. However, mtDNA represents a single genetic locus that follows a tree-like phylogeny without recombination while the demographic history of non-Africans might have included gene flow events not necessarily detectable through the study of mtDNA alone (Groucutt et al., 2015).

Those arguments are in accord with the second model category summarized in Multiple Dispersal models, which argue for an earlier out of Africa migration contributing to the genetic composition of at least some non-African populations (Lahr and Foley, 1998, Reyes-Centeno et al., 2014). Mayor support for this scenario comes from archaeological findings that attested the presence of modern humans in the Near East (Qafzeh and Skhul) and in the Arabian Peninsula (Jebel Faya) prior to the suggested major out of Africa event (Armitage

et al., 2011). In 2016 three publications were released in the same issue of *Nature* addressing those questions using a large series of worldwide present-day human genomes sequenced at high coverage. While two of the studies report evidence of a single out of Africa expansion (Malaspinas et al., 2016, Mallick et al., 2016) the third one identifies at least 2% of Austro-Melanesian ancestry tracing back to an earlier migration (Pagani et al., 2016), showing how the out of Africa dispersal debates are not yet concluded.

Within Upper Paleolithic Europe, the archaeological record is limited in human skeletal remains, only a few of which were genetically investigated when this thesis project started. The earliest possible evidence of modern humans in Europe was found in Grotta del Cavallo (southern Italy) associated to the Uluzzian "transitional" technocomplex and dated to ~45 ka (Benazzi et al., 2011). This old age confirmed a prolonged temporal and spatial overlapping between modern humans and Neandertals at least until ~40 ka, the suggested end of Mousterian and Châtelperronian cultures (Higham et al., 2014). In this roughly 5,000-year time frame (45 - 40 ka) archaeological remains assigned to the distinctive modern human Aurignacian culture are found throughout Europe such as in Geissenklösterle (Germany) (Higham et al., 2012), Willendorf II (Austria) (Nigst et al., 2014) and Kent's Cavern (UK), although contested (Higham et al., 2011, White and Pettitt, 2012). The oldest European directly dated skeletal remain attributed to *Homo sapiens* is a mandible discovered at Peștera cu Oase in Romania with an age of ~40 ka but lacks of archaeological assignment (Trinkaus et al., 2003). Interestingly, the specimen was initially described as harboring a mosaic of modern human and Neandertal derived morphological features and successive genomic analyses confirmed a higher proportion of Neandertal admixture in this individual who, however, did not significantly contributed to the genetic ancestry of present-day Europeans (Fu et al., 2015). The interpretation of Oase1 seen as an initial unsuccessful offshoot of modern humans into Europe is also supported by its mtDNA sequence, which branches off basal from the N lineage and before any other ancient or modern-day individual (Fu et al., 2015). A basal hg R falling within present-day mtDNA

from a tooth found in Grotta di Fumane (northern Italy) associated to the Protoaurignacian culture and stratigraphically dated to 41 - 38.5 ka (Benazzi et al., 2015), thus slightly younger than the early Aurignacian sites mentioned before. Kostenki14 (Russia) dated to around 38.7 - 36.2 ka was also firstly analyzed for the complete mtDNA (Krause et al., 2010a) and afterwards for the entire genome suggesting that population structure seen in Europe today dates back at least to the specimen's age (Seguin-Orlando et al., 2014). Additional NGS data were generated from European hunter-gatherers dated to before or after the Last Glacial Maximum (LGM) (Bollongino et al., 2013, Fu et al., 2013a, Lazaridis et al., 2014, Olalde et al., 2014, Jones et al., 2015) for a total of 20 complete mtDNAs and six shotgun genomes prior of publishing the works presented here. In conclusion, the sparse genetic data initially available for the Upper Paleolithic was insufficient to study population dynamics over ~40,000 years of European pre-history and the potential interplay of profound climatic and archaeological changes on the genetic landscape of early Europeans.

1.3 Peopling of Remote Oceania

A fascinating and poorly explored geographical region in aDNA research is the Southwest Pacific, the last unpopulated land to be reached ~3 ka by agriculturalists associated to the Lapita culture (Sheppard et al., 2015). Despite being a more recent time compared to the deep European pre-history, as mentioned before, tropical climate represents an adverse factor for the retrieval of aDNA. Therefore, paleogenetic studies in the region have been limited so far to the analyses of isolated mtDNA haplotypes, thus providing limited resolution to address question about human dispersals (Matisoo-Smith, 2015). Genome-wide research on modern-day people of Remote Oceania (i.e. southern and eastern of the Solomon Islands) revealed them as a genetic mixture of mainly an East Asian component and a minimum one forth Austro-Melanesian (Papuan) related ancestry (Kayser et al., 2008, Wollstein et al., 2010). Largely debated is the origin and timing of this admixture event. The first peopling of Remote Oceania has been described with two competing models: in the "Express Train" model

(Diamond, 1988, Gray and Jordan, 2000) people of ultimate East Asian origin reached Remote Oceania for the first time without a major exchange with Austro-Melanesian populations encountered *en route* to the southwest Pacific; instead, according to the "Slow Boat" or "Entangled Bank" model (Oppenheimer and Richards, 2001) those initial voyagers mixed extensively with Papuans on their way towards Remote Oceania. The admixture processes between the two genetic components possibly included sex-biased population mixtures as highlighted from the study of uniparental markers (mtDNA and Ychr) (Kayser, 2010). Genome-wide data from the first settlers is thus crucial to discriminate between the two conflicting models and to better comprehend the homogenization dynamics occurring during the peopling of the Southwest Pacific.

2 Aims

The main goal of this thesis is the application of techniques specifically optimized for obtaining ancient DNA from hominin fossils to be able to explore the genetic diversity existing in different times of the past and infer population movements in archaic and modern human prehistory.

Neandertals dominated Pleistocene Europe for several hundred thousand years before the first arrival of modern humans ~45,000 years ago. Two main questions are tackled though my work, first, which was the genetic diversity of late Neandertals when they encountered and admixed with modern humans? And second, were there any previous genetic interactions between different hominin groups? Seeking for answers to those issues I focused on reconstructing complete mitochondrial genomes from late as well as earlier Neandertals to propose an evolutionary scenario that may reconcile existing and contrasting phylogenies.

On the other hand, the genetic composition of early modern humans in Europe is understudied thus preventing a wide comprehension of the dispersal processes outside Africa. Indeed, most of ancient DNA studies focused on the genetic transformations from farming introduction onwards, while Europeans relied on foraging lifestyle for the vast majority of their existence. Which was the genetic make up of pre-Neolithic hunter-gatherers and how did it vary thought time? Here, I worked to enlarge the sample size both temporally and geographically of Upper Paleolithic and Mesolithic European mitochondrial and nuclear data to investigate population expansion and contractions possibly correlated to climatic and archaeological changes.

Finally, aDNA preservation is not only affected by time but also by climatic conditions that have prevented so far the retrieval of authentic genome-wide data from tropical environments such as islands of the Southwest Pacific. In this study, I attempted a combined application of petrous bone sampling, nuclear capture and NGS sequencing on remains from the first settlers of Remote Oceania and interrogated demographic changes and migration pulses in this region of the world.

3 Methods

I personally collected most of the analyzed human specimens presented in this thesis from curators at local institutions around Europe. Laboratory work was performed in the ancient DNA facilities of both the Institute for Archaeological Sciences at the University of Tübingen and the Max Planck Institute for the Science of Human History in Jena. In both institutions two separated areas were used for pre-amplification work and post-PCR processing called “clean room” and “modern lab”, respectively. In the clean room area strict procedures were followed to minimize potential further contamination of the human remains with modern human DNA. Laboratory investigators wore protective suites, two or three pairs of gloves, masks and goggles. UV light irradiation was applied in all laboratory areas, hoods and human remains previous sampling in an apposite sampling hood. All reaction mixes were prepared in a DNA-free hood where DNA samples were not introduced, whereas extraction, library preparation and indexing PCR steps were performed in a DNA hood. Furthermore, stricter clean room de-contamination measures were applied throughout my doctorate period. All tubes as well as all reagents not containing enzymes or detergents were UV treated for 30 minutes in a crosslinker and extraction funnels were bleached and rinsed with distilled water. Moreover Eppendorf Biopur tubes were used in the initial bone powder extraction and preparation of every reaction mix. After the introduction of the aforementioned procedures, I empirically observed a reduction of the total unique mtDNA sequences in library and extraction negative controls from over 200 down to 10 molecules in mtDNA capture experiments.

Human remains were photographed before and after sampling them with a dentist drill using diamond bits rotated at low speed to avoid heating that might harm DNA content. DNA extraction was performed applying a protocol appositely designed to retrieve short molecules typical of aDNA (Dabney et al., 2013a). After an enzymatic lysis and decalcification of the bone powder, DNA is bound to a large volume of binding buffer and spun through a column containing a silica membrane, which is then washed before the final DNA elution. An important limitation factor in this first chemical step is the amount of bone powder that can

be processed in each extraction. Despite the original protocol was designed to extract between 30 mg and 50 mg of bone powder I was able to successfully process up to 130 mg of material (see Posth *et al.* 2016b). However, higher amount of bone powder or samples with greater collagen content would clog the silica column preventing the possibility to generate a more concentrated and complex DNA extract.

I applied several library preparation techniques optimized for low concentration DNA extracts characteristic of ancient samples. For the vast majority of libraries a double stranded construction method was used (Meyer and Kircher, 2010) either with or without a preliminary pretreatment with UDG and endonuclease VIII in order to remove uracil residues (deaminated cytosines) accumulated along DNA molecules as post-mortem damage (Briggs et al., 2010). Furthermore, I supervised the student assistant Anja Furtwängler in establishing and applying the single stranded library protocol (Gansauge and Meyer, 2013) aimed to retrieve a higher DNA complexity in poorly preserved samples. Two versions of this method were implemented, one without correcting for post mortem damage and the other using USER enzyme (*E. coli* UDG) that effectively maintains damage only in the first two and last two sequence positions, a feature that can be used as aDNA authenticity criterion (Meyer et al., 2012).

Positive controls (cave bear) and negative controls (water) were carried along for both extraction and library preparation for a total of four controls each library batch and their DNA concentration was quantified with qPCR in order to evaluate experiment effectiveness. Library and extraction DNA blanks were furthermore taken along in subsequent laboratory steps and finally sequenced to obtain information on baseline contamination levels throughout the workflow.

The entire volume of DNA libraries was then indexed with an unique double barcode combination (Kircher et al., 2012) through a 10-cycle indexing PCR using the PfuTubo Cx Hotstart Polymerase that is able to read through uracil as thymine resulting in the typical aDNA damage patterns C to T at 5’ and G to A at 3’ ends for double stranded libraries (Briggs et al., 2007) and C to T at both molecule termini for single stranded DNA libraries (Meyer et al., 2012). I

furthermore modified the published indexing PCR protocol (Meyer and Kircher, 2010) adding two changes to the index primer design. First, I slightly changed the sequence of both the P7 primer by adding two nucleotides (GC) at the end of the oligonucleotide (AATGATACGGCGACCACCGAGATCTACAC-index-ACACTCTT TCCCTACACGACGC) and the P5 primer by removing four nucleotides (TCTT) at the end of the oligonucleotide (AATGATACGGCGACCACCGAGATCTACAC-index-ACACTCTTTCCCTACACGACGC) to favor their annealing to the complementary adaptor sequence with terminal GC bonds. In library qPCR assays, the corresponding modified version of the amplification primers (IS8 and IS7) resulted in narrower and more overlapping melting peaks for the library standards. Second, together with a former student from Tübingen, Florian Aldehoff, we created an index combination of 8bp instead of the previous 6-7bp to make use of the full length of the index sequence provided by Illumina platforms reducing the risk of cross talks between different indexes. Moreover, indexes were designed with an equal representation of A, C, G, T in each position of the index sequence in order to have a balanced nucleotide composition in both index reads for multiplexed sequencing runs.

The following step was an amplification of the indexed libraries to generate many copies of each molecule reaching a DNA concentration below the enzyme saturation limit to avoid the formation of heteroduplex products (Ruano and Kidd, 1992). AccuPrime Pfx DNA Polymerase was used in initial experiments because it was shown leading to higher percentage of endogenous DNA without introducing dramatic length and GC-content biases, as observed for other polymerases (Dabney and Meyer, 2012). However, using this enzyme we observed through Agilent 2100 Bioanalyzer quantification the formation of PCR artifacts consisting of DNA spikes at high concentration distanced ~35-40bp. This problem was especially remarkable for low concentrated libraries precluding the possibility of an accurate concentration measurement prior sequencing. The whole working group then switched to the use of Herculase II Fusion DNA Polymerase for indexed library amplification, which prevented artifact formation still providing comparable performances (Dabney and Meyer, 2012).

An aliquot of the amplified libraries was sequenced via shotgun sequencing for an average of ~1 - 5 M reads on an Illumina next generation sequencer to gain information on the percentage of human DNA within the library. Another aliquot was instead used for capture experiments aiming to target and enrich either the entire mitochondrial genome or hundred thousands of SNPs within the nuclear genome. For mtDNA capture experiments I followed a published protocol (Maricic et al., 2010) extracting DNA from my own saliva to produce two ~8,500bp long range PCR products encompassing the entire mtDNA and shearing the products into 150 - 500 bp fragments via ultrasonication. The latter were biotinylated, bound to magnetic beads and used as single stranded DNA baits to anneal to complementary ancient mtDNA fragments. After incubation, hybridized molecules were isolated with a magnet, eluted and re-amplified. For nuclear DNA enrichment instead, I established in the laboratories of Tübingen and Jena an in-solution capture protocol that was designed and initially described by the Max Planck Institute for Evolutionary Anthropology in Leipzig (Fu et al., 2013b). Up to ~1.24 M SNPs distributed across the human genome (Fu et al., 2015) were synthetized each over four probes on multiple Agilent arrays, cleaved off and converted into probe libraries. Those were amplified, transformed into biotinylated single stranded baits and used for hybridization capture (Fu et al., 2013b). The full list of targeted SNPs includes markers genotyped on thousands of modern-day individuals from diverse worldwide populations (e.g. Human Origins dataset) serving as comparative datasets for population genetic analyses (Patterson et al., 2012, Mathieson et al., 2015).

Captured libraries were sequenced on Illumina sequencing platforms and DNA reads were initially processed following previously described bioinformatics steps specifically designed for aDNA (Kircher, 2012), afterward replaced by the GUI user-friendly EAGER pipeline developed and improved by members of our group (Peltzer et al., 2016). Project specific genetic analyses are described individually in the supplementary information of each publication in Appendix.

4 Results

4.1 Mitogenome of an archaic femur from Hohlenstein-Stadel, Germany

C. Posth, C. Wißing, K. Kitagawa, L. Pagani, L. van Holstein, K. Wehrberger, N. J. Conard, C. J. Kind, H. Bocherens, J. Krause (2016). "Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neandertals". Review in *Nature Communications*.

Synopsis

This manuscript provides support to an evolutionary scenario that resolves the observed discrepancy between nDNA and mtDNA phylogenies of archaic and modern humans by reporting two key findings: the reconstruction of the deepest divergent Neandertal mtDNA lineage known to date and the analytical exploration of a putative Neandertal mtDNA replacement.

First, I report DNA of a hominin femur (HST) presenting archaic morphology from the Hohlenstein-Stadel cave in southwestern Germany. The preservation conditions of the specimen have not yet allowed for analysing the nDNA. However, it was feasible to reconstruct its entire mtDNA (mitogenome) by combining an iterative likelihood method coupled with a visual inspection of the aligned mtDNA sequences. The level of modern human mtDNA contamination was estimated with three different approaches that reported consistent values ~10%. In a phylogenetic tree including modern human, Neandertal and Denisovan mtDNAs, HST represents a novel lineage that diverges basal on the Neandertal branch and displays a considerable branch shortening. By implementing a Bayesian statistic framework, the split time of HST from all other Neandertal mtDNAs was dated to ~270 ka (95% high probability density [HPD] 316 - 219 ka), which is ~120 ka older than the most divergent modern human mtDNAs. Moreover, radiocarbon dating attempted directly on the femur resulted in inconsistencies suggesting that the hominin specimen might be beyond the detection limits of the ^{14}C method. I thus attempted a molecular date for the HST mtDNA by calculating its phylogenetic branch length to an age of 124 ka (95%

HPD 183 - 62 ka). Moreover, the analysis of the highly divergent archaic femur's mtDNA suggests a greater genetic diversity in Middle Pleistocene Neandertals, followed by a decline in effective population size of Neandertals until a possible demographic growth preceding their disappearance.

Second, I further explored the inconsistency in the nDNA and mtDNA phylogenies of archaic versus modern humans. While nDNA indicates Denisovans being a sister group of Neandertals after separating from the modern human lineage, mtDNA of archaic hominins suggests an alternative evolutionary scenario. The higher similarity and younger divergence time of Neandertal and modern human mtDNAs compared to Denisovans have been explained with a gene flow event from Africa into Neandertal populations. HST mtDNA allowed us to constrain the boundaries for the time of the putative African introgression into Neandertals between 410 ka and 270 ka (95% HPD 467 - 219 ka). Over such a time interval, an entire mtDNA replacement was shown feasible even if the introgressing lineage represented a few percent of the initial Neandertal gene pool.

4.2 Mitochondrial DNA diversity in late Neandertals from Goyet, Belgium

H. Rougier, I. Crevecoeur, C. Beauval, **C. Posth**, D. Flas, C. Wissing, A. Furtwangler, M. Germonpre, A. Gomez-Olivencia, P. Semal, J. van der Plicht, H. Bocherens and J. Krause (2016). "Neandertal cannibalism and Neandertal bones used as tools in Northern Europe." *Scientific Reports* **6**: 29005.

Synopsis

This publication provides evidence for Neandertal cannibalism and presents multiple Neandertal bones used as stone tools at the Troisième caverne of Goyet (Goyet cave) in Belgium. A multidisciplinary approach was implemented to re-analyze the archaeological collection excavated around 150 years ago that includes an almost continuous temporal transect with material attributed to Middle and Upper Paleolithic cultures (Mousterian, Lincombian-Ranisian-

Jerzmanowician, Aurignacian, Gravettian and Magdalenian). Taphonomic and morphological examination was performed to identify remains of possible Neandertal origin. Ten bone fragments were directly radiocarbon dated to 45.5 - 40.5 ka, thus overlapping in time with the presence of early modern humans in Europe. I conducted genetic analyses on ten specimens in order to confirm their taxonomic assignment with the implementation of a mtDNA capture technique. All remains provided Neandertal mtDNA sequences distinctive from modern humans and seven complete mtDNA genomes were successfully reconstructed. These almost doubled the number of whole Neandertal mtDNAs recovered to date and phylogenetic analyses confirmed the low genetic diversity of European late Neandertals. Despite their limited mtDNA diversity, a significant behavioral variability was observed among late Neandertal mortuary practices in Northern Europe. Several of the 99 identified Neandertal remains showed anthropogenic modifications, such as cutmarks and percussion marks providing evidence for butchery activities never found before in Northern European Neandertals.

4.3 Mitochondrial DNA composition of pre-Neolithic Europeans

C. Posth, G. Renaud, A. Mittnik, D. G. Drucker, H. Rougier, C. Cupillard, F. Valentin, C. Thevenet, A. Furtwangler, C. Wissing, M. Francken, M. Malina, M. Bolus, M. Lari, E. Gigli, G. Capecchi, I. Crevecoeur, C. Beauval, D. Flas, M. Germonpre, J. van der Plicht, R. Cottiaux, B. Gely, A. Ronchitelli, K. Wehrberger, D. Grigorescu, J. Svoboda, P. Semal, D. Caramelli, H. Bocherens, K. Harvati, N. J. Conard, W. Haak, A. Powell and J. Krause (2016). "Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe." *Current Biology* 26(6): 827-833.

Synopsis

In this study I generated 35 complete mitochondrial genomes from Upper Paleolithic and Mesolithic individuals spanning almost 30,000 years of European prehistory (from 35 ka to 7 ka). Using mtDNA capture techniques in combination

with high-throughput DNA sequencing I could thus almost triple the amount of ancient European hunter-gatherer sequences available to date. This larger paleogenetic investigation provided two majors unexpected findings about the genetic diversity of pre-Neolithic Europe.

First, three individuals dating from 35 ka to 27 ka years ago from Belgium and France were found to belong to mtDNA hg M that, together with N, represent the two almost unique mtDNA clades present outside Africa today. Haplogroup M is almost absent in contemporary Europeans, while distributed at high frequency in present-day Asian, Australian and Native American populations. Our finding provides new evidence on non-African dispersal processes. Making use of the enlarged ancient mtDNA dataset of 66 radiocarbon dated sequences we applied Bayesian phylogenetic methods to recalculate the mtDNA mutation rate (2.74 x 10^{-8} [95% HPD 2.44 - 3.01 x 10^{-8}] mutation/site/year) and more accurately re-date the diversification time of the two non-African clades, N and M, to ~50 ka (±5,000 years). On the other hand, the oldest undisputed evidence of early modern humans in Europe and Australia were dated to at least 45 ka. Therefore our estimated molecular date supports a major, single and thus rapid dispersal of all non-Africans ancestors after 55 ka carrying both M and N hgs not only across eastern Eurasia but also into western Eurasia.

Second, coalescent modeling and approximate Bayesian computation (ABC) on mtDNAs of 55 pre-Neolithic Europeans were combined to explicitly test the potential influence of dramatic Late Pleistocene environmental changes on the demographic history of European hunter-gatherers. From the six tested models the best supported by the data was the one including an LGM mtDNA bottleneck that possibly displaced hg M from Europe. Furthermore, the favored model depicted a previously unknown population turnover at 14.5 ka, the start of a period named Late Glacial and characterized by severe climatic oscillations towards the end of the Late Pleistocene. The observed genetic transformations through time suggest that climate might have influenced European hunter-gatherer demography.

4.4 Genomic structure of Upper Paleolithic and Mesolithic Europe

Q. Fu, **C. Posth***, M. Hajdinjak*, M. Petr, S. Mallick, D. Fernandes, A. Furtwangler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gely, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Paabo and D. Reich (2016). "The genetic history of Ice Age Europe." *Nature* 534(7606): 200-205.
*equal contributors

Synopsis

In this work nDNA of 38 Paleolithic and Mesolithic Europeans (mtDNAs of 24 of them were previously reported in Posth *et al.* 2016) was enriched for 390,000 - 3.7 M SNPs and combined with 13 previously published Eurasian shotgun genomes, making a total of 51 early modern humans analyzed at genome-wide level. It was first observed that from 45 ka to present-day the proportion of Neandertal admixture in Western Eurasian steadily decreased from 3-6% to around 2%. The depletion of Neandertal DNA was higher in proximity of functional regions in the genome, suggesting negative selection operating against the introgressed DNA.

Second, we investigated the population genomic history of Eurasian hunter-gatherers spanning around 40,000 years. To avoid relying on general archaeological grouping alone, the studied individuals were initially clustered on the base of their shared genetic drift after diverging from a common outgroup using f_3-statistics. Five main genetic clusters were identified, which included most

of the individuals except the ones admixed between clusters or who formed independent basal branches. The oldest group was the “Věstonice Cluster” represented by 14 individuals from 34 - 26 ka spanning from Czech Republic, Belgium, Austria and Italy all associated to the Gravettian culture. The “Mal’ta Cluster” was composed of three Russian individuals dated from 24 ka to 17 ka. The “El Mirón Cluster” included seven individuals from Spain, France, Belgium and Germany dated to 19 - 14 ka and associated to the Magdalenian culture. The “Villabruna Cluster” formed of 15 individuals from Spain, France, Luxemburg, Germany, Italy, Switzerland and Hungary spanning 14 - 7 ka was associated with Azilian, Epipaleolithic, Epigravettian and Mesolithic cultures. And the “Satsurblia Cluster” composed of two Georgian individuals from 13 – 10 ka.

Analyzing from the oldest individuals going forward in time, no substantial contribution to the modern-day Eurasian gene pool derived from either of two initial modern humans genetically analyzed, i.e. the ~45,000-year-old Ust’-Ishim individual from Siberia and the ~40,000-year-old Oase1 individual from Romania. On the contrary, all analyzed individuals in our series from ~37 ka (the age of Kostenki14) to ~7 ka (the youngest individual), shared part of their ancestry with modern-day people in Europe. In particular, from about 37 ka to 14 ka all specimens presented DNA that seems to originate from a unique founder population since its divergence from the lineage leading to the 24 ka Siberian Mal’ta1 individual. Within Europe, a ~35,000-year-old individual from Belgium (GoyetQ116-1), chronologically assigned to the Aurignacian culture, was found to belong to an early branch different from Kostenki14. The latter shared most of its ancestry with members of the Gravettian-related Věstonice Cluster, while the GoyetQ116-1 branch has no detectable contribution to that cluster, disappearing from our genomic record for almost 15,000 years. At the end of the LGM, this component however reappeared in the El Mirón Cluster in Spain, before spreading across central Europe possibly associated with the Magdalenian culture. After 14 ka, during the Late Glacial major warming period, all individuals belonging to the Villabruna Cluster showed a genetic affinity to present-day Near Eastern populations. This archaeological phase in Europe is not accompanied

with major cultural transitions. However, the Near Eastern-related genetic component might have appeared not necessarily only through long-distance migrations but also via population turnovers within Europe. Finally, some but not all hunter-gatherers from the Villabruna Cluster showed a closer relationship to modern-day East Asians, providing evidences of interactions between western and eastern Eurasian hunter-gatherer groups dating back to the Paleolithic time.

4.5 Genomic origins of the first people in the Southwest Pacific

P. Skoglund, **C. Posth**, K. Sirak, M. Spriggs, F. Valentin, S. Bedford, G. R. Clark, C. Reepmeyer, F. Petchey, D. Fernandes, Q. Fu, E. Harney, M. Lipson, S. Mallick, M. Novak, N. Rohland, K. Stewardson, S. Abdullah, M. P. Cox, F. R. Friedlaender, J. S. Friedlaender, T. Kivisild, G. Koki, P. Kusuma, D. A. Merriwether, F.-X. Ricaut, J. T. S. Wee, N. Patterson, J. Krause, R. Pinhasi and D. Reich (2016). "Genomic insights into the peopling of the Southwest Pacific." *Nature* 538(7626): 510-513.

Synopsis

This publication reports genome-wide data of individuals from Remote Oceania associated to the farming Lapita culture, spanning between ~3 ka to ~2.5 ka. Nuclear data was obtained from capturing for 1.24 M nDNA SNPs of the total DNA extracted from petrous bones of four individuals, three from Vanuatu and one from Tonga. This data was co-studied with 778 modern-day individuals from East Asia and Oceania genotyped for ~600,000 overlapping Human Origins SNPs. First, in the ancient samples we found no detectable evidence of Austro-Melanesian-related ancestry, supporting the “Express train” model of a fast settlement of Remote Oceania by an un-admixed East Asian-related population. This was also confirmed by the almost absent Denisovan contribution into Lapita people, contrary to Austro-Melanesian populations as well as present-day islanders from Remote Oceania.

Second, present-day populations from the Southwest Pacific are not simply

descending from people associated to the initial Lapita movement but subsequent Papuan-related admixtures shaped their genetic makeup. Making use of the known human genome recombination rate, we estimated the admixture date between Papuan and East Asian-related genetic components to 2 - 1.2 ka. Finally, a significant asymmetry in the proportion of the two genetic ancestries was identified between autosomal and sex chromosomes in modern-day people from Remote Oceania. The East Asian-related component is in excess on the X chromosome, suggesting that males mainly mediated the additional dispersal waves that carried Papuan ancestry into the Southwest Pacific.

5 Discussion

5.1 Ancient DNA retrieval

Preservation of aDNA is the fundamental prerequisite for any paleogenetic study. With the aid of advanced molecular biology techniques, in this thesis, I first explore the preservation of DNA in hominin remains spanning from the Middle Paleolithic temperate Europe to 2.5 ka tropical Southwest Pacific. Several challenges exist to retrieve authentic endogenous DNA across this wide temporal and environmental range. Moreover, most of the specimens analyzed here were excavated many decades ago and not following today's archaeological standards, therefore contamination with present-day human DNA represents a significant limitation.

The archaic human femur (HST) from Hohlenstein-Stadel (Germany) presented in Posth *et al.* 2016b, was excavated in 1937 and possibly extensive washing and handling resulted in a high proportion of modern human DNA contamination, which was estimated representing around 10% of the isolated mtDNA fragments. When reconstructing the HST mtDNA, I observed a substantial bias introduced by the reference sequence used as mapping scaffold. Mapping accuracy was limited by high damage pattern (50% at the molecule termini) and the notably short fragments retrieved (43bp on average). A lower rate of miss-mapping with contaminant reads was achieved aligning all sequenced reads against a reference that is phylogenetically closer to the investigated mtDNA, the Reconstructed Neandertal Reference Sequence (RNRS) (Behar et al., 2012). By making use of the mtDNA phylogenetic branch shortening the specimen was molecularly dated to 124 ka (95% HPD 183 - 62 ka), the second oldest mtDNA sequence among Neandertals after the Altai Neandertal, which I dated to 130 ka (95% HPD 172 - 88 ka), consistent with genome-wide calculations (Prufer et al., 2014).

In Rougier *et al.* 2016 a multidisciplinary approach was implemented to assess the taxonomic assignment of human remains found at Goyet cave (Belgium) mainly during the 1868 excavation, which lacks a detailed archaeological

documentation. Due to the fragmentary status of many Paleolithic hominin fossil collections, aDNA is nowadays an often used tool for the molecular identification of morphologically ambiguous Neandertal specimens, especially in cases of temporal and/or spatial overlapping with modern humans or Denisovans (Krause et al., 2007, Brown et al., 2016, Talamo et al., 2016). Hominin bones from Goyet cave were dated to 45.5 - 40.5 ka, thus temporally coexisting with the earliest modern humans in Europe. Phylogenetic analyses on mtDNA sequences of ten specimens confirmed their attribution to Neandertal, and their positioning in the mtDNA tree was maintained even when restricting the analyses to only damaged fragments, supposed to be uniquely endogenous to the individual (Skoglund et al., 2014). The mtDNA preservation of seven out of ten remains was remarkably good, also when compared to some younger modern human remains found in the same cave and genetically analyzed in Posth *et al.* 2016a. Signs of butchery activities were shown on several Neandertal bones and, therefore, defleshing perhaps favored specimen preservation (White, 1992).

In Posth *et al.* 2016a and Fu *et al.* 2016, a partially overlapping set of individuals were investigated using newly generated mtDNA as well as genomic data for a total of 49 European hunter-gatherers spanning between 35 ka and 7 ka. DNA preservation varied greatly and possibly linked to the distinct taphonomic histories of the specimens.

A common feature across the genetic collection was instead a significant level of human contamination as measured with both mtDNA (Fu et al., 2013a, Renaud et al., 2015) and nDNA in males (Korneliussen et al., 2014). Those values did not necessarily coincide, as expected from a different ratio of mitochondrial over nuclear DNA contents between endogenous and contaminant DNA. Complete mtDNA sequences were confidently reconstructed despite estimated contamination level rising in some cases up to 18%. Facilitated by the haploid state of the mtDNA in combination with high average coverage values (at least 10-fold), we were able to reliably call nucleotide positions even when clearly affected by modern-day human contamination. On the contrary, an X chromosome heterozygosity rate in males above 2.5% was considered

excessively high in order to perform reliable nDNA analyses. In the absence of standard methods to estimate nuclear contamination in females, all nDNA data from female individuals and males above the indicated threshold were restricted to damaged sequences. This procedure relies on retaining only fragments that exhibit the typical aDNA damage pattern at their ends in order to decrease contamination at the price of substantial data reduction (Meyer et al., 2014).

Finally, in Skoglund *et al.* 2016 three ~3000-year-old individuals from Vanuatu (Teouma site) and one ~2500-year-old individual from Tonga (Telasiu site) in Remote Oceania provided high quality genome-wide data. My previous attempts to obtain aDNA extracted from several teeth and cortical portions of long bones from the same archaeological site in Tonga repeatedly failed. On the contrary, sampling the inner ear of the petrous bone paired up with in-solution capture resulted in the retrieval of 140,000 – 230,000 usable genome-wide SNPs to genetically investigate the first peopling of Remote Oceania. The low ratio of sequences aligning to the Y chromosome compared to the ones mapping on both sex chromosomes (X + Y) (Skoglund et al., 2013) revealed that all four individuals were females. As mentioned before, in the absence of a reliable method to estimate nuclear contamination in females, aDNA authenticity was confirmed by comparing the individual positioning on a principal component analysis (PCA) before and after the restriction to only damaged fragments. Data from the same individuals clustered closely together and did not overlap with any modern-day population suggesting minimal levels of contamination in the analyzed DNA.

Therefore, petrous bone sampling in combination with NGS and capture techniques seem to represent the principal game changers that will allow the extension of paleogenetic investigations into adverse - for aDNA retrieval - tropical environments.

5.2 Pre-historic dispersals

Despite the increasing amount of autosomal DNA data retrieved from archaic humans, the analyses of non-recombining loci such as mtDNA can be useful in defining the temporal upper bound of population splits in the past (Veeramah and Hammer, 2014). In Posth *et al.* 2016b I made use of this property to infer the temporal interval for the putative African gene flow into Neandertal populations responsible for providing the typical Late Pleistocene Neandertal mtDNA. This lineage is closely related to modern human mtDNA while Denisovans form an outgroup clade with the Middle Pleistocene early Neandertals from Sima del los Huesos.

Instead, nuclear DNA undergoes recombination in every generation and the resulting incomplete lineage sorting produces conflicting genealogies between different genomic loci. However, by analyzing millions of nucleotide positions across the genome more precise phylogenies can be drawn than with a single locus such as mtDNA. For archaic hominins, phylogenetic inferences based on nDNA attested Denisovans as sister group of Neandertals after diverging from an ancestor shared with modern humans. Therefore, genome-wide phylogenetic relationships contrast the mtDNA tree topology. An evolutionary scenario that can reconcile the described inconsistency is the one reported here in Figure 1 where the mtDNA found in Late Pleistocene Neandertals replaced the pre-existing Denisovan-like lineage in Neandertal populations across Eurasia (Meyer et al., 2016). Such introgressing lineage most parsimoniously originated in Africa, as it shows a closest phylogenetic relationship to modern human mtDNAs. However, a genetic admixture between African and European hominins should have left traces not only in their mtDNA but also in their nuclear genome. Indeed, genetic traces of a modern human-like component (up to 2.1%) have been detected in the high coverage Altai Neandertal genome (Kuhlwilm et al., 2016). This genetic material was certainly acquired after the Denisovan/Neandertal nDNA divergence [473 - 381 ka (Prufer et al., 2014)] and the introgression event

was further constrained to 230 - 100 ka based on the coalescence age of African haplotypes.

Finding HST mtDNA in Europe, branching off basal on the Late Pleistocene Neandertal lineage, suggests that by the time of its divergence the mtDNA introgression event had already happened. We dated this spit time to ~270 ka (95% HDP 316 - 219 ka) therefore providing a lower time boundary for the African gene flow into Neandertals. The calculated date interval only marginally overlaps with the Kuhlwilm *et al.* 2016's estimate thus the genetic admixture that provided the Late Pleistocene Neandertal mtDNA might have been an additional, earlier episode. In fact the dispersal event of African hominins into Eurasia described in Posth *et al.* 2016b is estimated to occur during the Middle Pleistocene in a temporal interval between 410 ka and 270 ka (460 - 219 ka including upper and lower 95% HPD).

In Posth *et al.* 2016a, non-African dispersal processes are investigated through the analyses of ancient European and modern worldwide mtDNAs. Present-day non-Africans belong to two major mtDNA clades M and N, which TMRCAs were estimated to ~50 ka and ~60 ka, respectively (Behar et al., 2012). However, while hg N is present everywhere outside Africa, hg M is widespread in Asia, Australia and America but absent in Europe today (Kivisild, 2015). The uneven geographical distribution of the two clades led authors to suggest two non-African dispersal events, an early spreading of hg M throughout Asia via a southern route followed by a second diffusion of hg N via a northern route (Maca-Meyer et al., 2001). This model has been contested with the proposal of an alternative scenario in which modern humans undertook a single rapid dispersal into Eurasia first towards Asia and then back into Western Eurasia, after a lengthy pause during which hg M vanished (Macaulay et al., 2005).

Finding hg M in ancient Europeans overturns both dispersal models and instead suggests that succeeding demographic processes within Europe are responsible for the present-day lack of hg M. Moreover, making use of 66 ancient ^{14}C dated mtDNAs as calibration points on both hgs M and N lineages a new mtDNA mutation rate was calculated with narrower 95% HDP intervals than in previous

studies. As a result the TMRCAs of both hgs M and N tightly overlap in time spanning from 44 ka to 55 ka. These young dates, together with the presence of modern humans in the European and Australian archaeological records at least 45 ka (Bowler et al., 2003, Benazzi et al., 2011), supports the model of a single, rapid dispersal after 55 ka for all non-African populations containing hgs N and M, not only across Asia but also into Europe.

Despite mtDNA providing a reliable upper bound for the time of population split in the past, genetic data is unsuitable for assessing the geographic location where the dispersal began. The calculated TMRCAs are effectively dating the start of the mtDNA diversification of all present-day non-Africans. It is assumed that this process happened outside Africa because M and N mtDNA branches most likely originated there after the split from the East African L3 lineage. However, by using genetic data alone it is not possible to exclude a scenario where the diversification initiated within Africa in a substructured population. Only combining genetic with archaeological and paleoenvironmental data would allow to narrow down from where this major dispersal process commenced.

Another caveat of this result concerns archaeological remains of modern humans found in Asia and elsewhere before 55 ka (e.g. Liu et al., 2015). A renowned limitation of mtDNA is that, due to coalescence processes, many mtDNA lineages from the past will not leave descendants through time. Therefore, individuals found outside Africa who potentially originated from previous expansions did not contribute descendants to the present-day non-African mtDNA pool (Pagani et al., 2016). Interestingly, two recently published papers on modern-day worldwide genome analyses broadly confirmed our mtDNA finding of a single ancestry for all people outside Africa whose oldest genetic structure dates back to ~50 ka (Malaspinas et al., 2016, Mallick et al., 2016).

Within Europe itself, Posth *et al.* 2016a and Fu *et al.* 2016 found indications of population continuity in the Upper Paleolithic from at least ~37 ka to ~14 ka in both mtDNA and nDNA data, without evidence of substantial contribution from other ancestral sources. However, both studies described evidences of a previously unknown population turnover at the beginning of the Late Glacial ~14

ka. This period in Europe is characterized by drastic climatic upheavals starting with warming phases (Bølling-Allerød interstadial) followed by abrupt cooling (Younger Dryas stadial) (Heiri et al., 2014). During the early warm phases of the Late Glacial several Holarctic megafaunal species went extinct (Cooper et al., 2015) while a shift in European hunter-gatherer mtDNA hg frequency is observed in Posth *et al.* 2016a. The demographic model best supported by the data is consistent with a replacement of the existing Upper Paleolithic population by one from a different source. However, mtDNA only depicts matrilinear population dynamics and, therefore, cannot exclude some level of genetic continuity as well as provide information about the origins of the incoming population. In Fu *et al.* 2016 genome-wide data confirmed the mtDNA signal and further extended the argument. All individuals belonging to the Villabruna Cluster dated from 14 ka onwards shared genetic affinity with modern-day Near Easterners. Two alternative scenarios could explain the observed pattern. Either a long-distance migration from the Near East into Europe happening 6,000 years earlier than the Neolithic dispersal (Lazaridis et al., 2014), or a double population expansion from somewhere around the eastern Mediterranean region both towards Europe and the Near East drawing together their genetic ancestry.

The last dispersal process dealt with in this thesis and described in Skoglund *et al.* 2016 is the one that led into Remote Oceania, among the last major expansions of modern humans into an unpopulated geographical region. The first settlers are associated to the Lapita culture, which was a farming society equipped with out-rigger canoes able to cross long stretches of the Pacific Ocean at least by ~3 ka. Archaeological and linguistic evidence placed the ancestral homeland of this population in East Asia (Bellwood, 2005, Gray et al., 2009) before spreading along the north coast of Papua New Guinea until the Solomon Islands and finally reaching the islands of Remote Oceania. Genetically, all people living today in the region share a large proportion of their ancestry not only with East Asians but also with Papuans (Kayser et al., 2008, Wollstein et al., 2010). The widespread distribution of Papuan-related ancestry in Remote Oceania populations led the authors of Kayser *et al.* 2008 and Wollstein *et al.*

2010 to support the Slow Boat model according to which the two major genetic components mixed somewhere around the Island Southeast Asia region before dispersing in the Southwest Pacific (Oppenheimer and Richards, 2001). This model is in disagreement with our genomic data analyses of four individuals from Vanuatu and Tonga dated to 3 - 2.5 ka and associated to the Lapita culture. In those genomes, no Papuan ancestry contribution was detected, which instead supports the Fast Train model, according to which people first arrived in Remote Oceania without substantial admixture with Austro-Melanesian populations encountered along their journey. Those results indicate that at least another major and previously undetected dispersal was responsible for the spread of Papuan-related ancestry across the Southwest Pacific. Dating these additional population movements by genetic means relies on many assumptions (e.g. recombination rate and generation time) thus providing a large 95% confidence interval of ~800 years for the Austro-Melanesian-related admixture date (2 - 1.2 ka). Moreover distinct Southwest Pacific islands might have received the Papuan contribution at different times as described by the diachronic predominance of Melanesian-like craniometric features in the fossil record (Valentin et al., 2016). Finally, a lower proportion of Papuan affinity than the Lapita-like genetic component is identified on the X chromosome compared to the autosomal DNA of present-day Southwest Pacific islanders. Females carry two X chromosomes while males only one, thus females contribute two-thirds of the X chromosomes to a population but only half of the autosomes like males. Therefore, those results suggest that males may have mediated secondary migration streams distributing Papuan ancestry throughout Remote Oceania from western Melanesia.

5.3 Hominin genetic diversity

Genetic diversity represents the set of traits defining the genetic makeup of a specific group. This can vary over time and positively correlates with the effective population size of such a group. In population genetics, a method to calculate

diversity is through the Watterson's estimator (*θw*), which measures the number of polymorphic sites within a population accounting for the sample size of the analyzed individuals. *θw* was previously calculated for three Denisovan mtDNAs (Sawyer et al., 2015) despite belonging to at least two different temporal phases of occupation at Denisova cave. In Posth *et al.* 2016b the same analysis was repeated on Neandertals to evaluate genetic diversity among this group, including six of the seven complete mtDNAs newly released in Rougier *et al.* 2016 for a total of 17 sequences spanning from western Europe to the Altai Mountains. The *θw* was notably lower than in Denisovans, consistent with a low genetic diversity in late Neandertals (Briggs et al., 2009). However, when including the highly divergent HST mtDNA within the Neandertal group, *θw* almost doubles reaching a level comparable to Denisovans and present-day Europeans thus indicating that the Neandertal genetic diversity was higher than previously thought. Making use of the enlarged dataset of 18 complete mtDNAs, changes in Neandertal effective population size (*Ne*) through time were explored in a Skyline plot [implemented in BEAST (Drummond and Rambaut, 2007)]. This depicts a Middle Pleistocene *Ne* that progressively declined during the Late Pleistocene. Analyses performed on the high coverage Altai Neandertal genome with the multiple sequentially Makovian coalescent (MSMC) method reported similar *Ne* reduction patterns through time (Prufer et al., 2014). Interestingly, the Skyline reconstruction also highlights an increase in effective population size towards the end of the Neandertal temporal range, consistent with the results obtained from chromosome 21 of the ~44,000-year-old Vindija Neandertal (Kuhlwilm et al., 2016). Hence, during the period of coexistence between modern humans and Neandertals in Europe, the latter group might have actually experienced a population expansion just before disappearing from the archaeological record. Further studies focusing on the demographic changes and interactions between archaic and modern humans would be of pivotal importance to better comprehend the processes that led to the Neandertal extinction.

The genetic diversity of early modern humans from their arrival in Europe ~45 ka until the Neolithic transition 8 - 7 ka is described in Posth *et al*. 2016a and Fu *et*

al. 2016. In Figure 2 of this thesis, the mtDNA of 70 pre-Neolithic hunter-gatherers from Europe and Caucasus are aligned to reconstruct a maximum parsimony tree. As mentioned before, among all analyzed individuals spanning the time periods between ~40 ka and the LGM onset (~25 ka) both major non-African mtDNA clades, M and N, are identified in Europe. While individuals belonging to hg N are distributed across different sub-haplogroups such as basal R, basal U, U2'3'4'7'8'9, U5 and U6, all four individuals carrying the hg M fall on the same phylogenetic branch, which has, to date, not been described in modern-day individuals. The diversity of pre-LGM hg M in Europe was therefore lower than among hg N, a fact that possibly contributed to its disappearance after an LGM population bottleneck, as described in the best-supported demographic model in Posth *et al.* 2016a. During this cold and dry temporal interval (~25 – 19.5 ka) European hunter-gatherers likely underwent a range contraction into warmer climatic refugia (Gamble et al., 2004, Stewart and Stringer, 2012). The proposed areas of human retreat include mostly Mediterranean regions such as Franco-Cantabria, Southern Italy and the Balkans, as well as East European Plains (Soares et al., 2010). At the beginning of the post-LGM phase (19.5 - 14.5 ka) a genetic component, initially identified in a 35,000-year-old individual from Belgium (GoyetQ116-1) associated to the Aurignacian culture, reappeared in the El Mirón individual from Cantabria (Spain) after being largely displaced by the Gravettian-related Věstonice Cluster across Europe. The Iberian individual provides the name to the El Mirón Cluster, which is possibly linked to the spread of the Magdalenian culture into central Europe from southwestern European climatic refugia (Straus, 2013). Notably, Magdalenian individuals from the French and Swabian Jura as well as from Goyet cave, all dated to ~15 ka, contain a higher proportion of the GoyetQ116-1 component than El Mirón. Moreover, despite GoyetQ116-1 genomic signature resurfaced at the end of the LGM, its mtDNA sequence belonging to the M branch has not yet been found in post-LGM individuals. The latest appearance of mtDNA hg M in our genetic record is at ~27 ka in both La Rochette (southern France) and Ostuni1 (southern Italy). Taken together, the two previous points suggest that the Aurignacian genetic

component specifically related to GoyetQ116-1 might have survived in different climatic refugia other than Iberia.

Remarkable is the genetic/temporal transect that emerges from Goyet cave where the Aurignacian-related GoyetQ116-1 component is greatly substituted by the Gravettian-related Věstonice Cluster before its resurgence as part of the Magdalenian-related El Mirón Cluster. Taken together this highlights the importance of regional investigations through the Upper Paleolithic and Mesolithic to better comprehend processes occurring at a broader geographical scale.

Towards the end of the Late Pleistocene, a profound genetic transformation occurred across Europe, which was detected through a shift in the mtDNA composition and at the nDNA level through the formation of the Villabruna Cluster. All analyzed individuals after ~14 ka are in fact characterized by a vast representation of mtDNA hg U5 and by a genetic affinity to modern-day Near Easterners, which is absent in previous time periods. No clear discontinuities are observed in the material culture at this archeological horizon such as the transition from Magdalenian to Azilian cultures (Valentin, 2008). Instead, it correlates in time to the start of the Late Glacial identified by the Bølling-Allerød interstadial, the first abrupt warming phase after the LGM (Weaver et al., 2003, Heiri et al., 2014). During this period, a drastic transformation in floral composition possibly resulted in the diffusion of forests towards Northern Europe (Stewart and Lister, 2001) and the Eurasian extinction of megafauna such as the wholly rhinoceros and the cave lion (Stuart and Lister, 2012). However, while the described patterns are linked with environmental changes, the influence of modern human activities is still poorly understood. The genomic results presented here are consistent with a period of human mobility favored by the retreatment of the ice sheets and the opening of ecological corridors that might have enhanced contacts between previously separated groups. The appearance of the Near Eastern genetic component in central Europe might therefore represent the spread of people from southeastern European refugia but the lack of genome-wide data of Late Pleistocene hunter-gatherers from that region has

not yet allowed for the identification of the genetic identity of the putative source population.

Two hunter-gatherer genomes from Georgia (Satsurblia and Kotias) dated to 13 - 10 ka (Jones et al., 2015) and grouped in the Satsurblia cluster present a higher affinity to the Villabruna Cluster than with previous Europeans despite not being direct representatives of the incoming population. Their mtDNA as well as of two Mesolithic individuals from Theopetra in Greece (Hofmanova et al., 2016) belong to hgs H and K (Figure 2), found in central Europe so far only from the early Neolithic onwards (Haak et al., 2010). The presence of those markers in Balkan hunter-gatherers but not in central Europe before the Neolithic suggests the existence of further population structure, whereby Europe hosted groups with different genetic affinity to the Near East. After a long period of isolation, the Late Glacial population turnover discovered here was thus responsible for drawing together European and Near Eastern ancestries. Additional hunter-gatherer mitochondrial and whole genome data, particularly from eastern Mediterranean regions are of key importance to explore the complete genetic diversity that existed in Europe at the end of the Pleistocene.

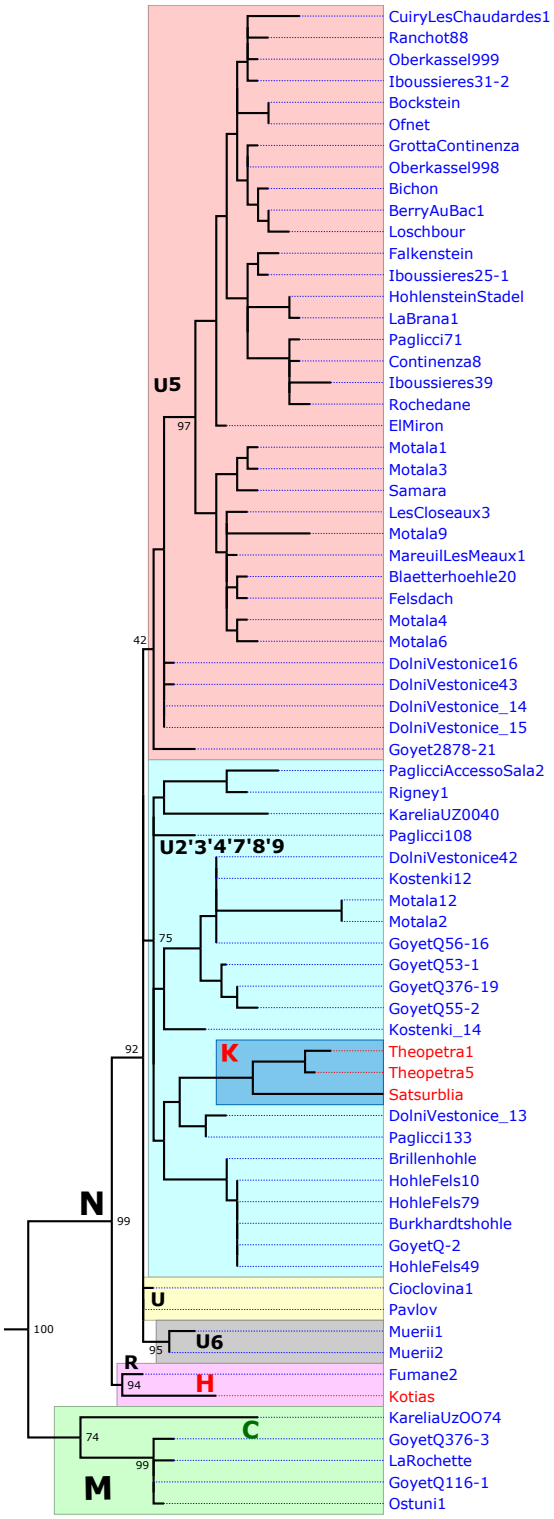


Figure 2: Maximum parsimony tree of 70 pre-Neolithic European and Caucasian mtDNAs. European hunter-gatherers (excluding Balkans) are shown in blue while Caucasian and Balkan hunter-gatherers in red print. The tree is built in MEGA6 (Tamura et al., 2013) with 95% partial deletion, 1000 iterations as bootstrap support and rooted with a basal African mtDNA (not shown).

6 Outlook

The field of aDNA applied to the study of human population history is exponentially growing both in terms of individuals analyzed, quality of data and geographical regions form where genomic information can be retrieved. This would have not been possible without a constant improvement of specific techniques from skeletal sampling to sequencing of entire ancient genomes. Computational pipelines designed for aDNA are also in fast development in order to account for characteristic DNA damage in ancient sequence data and to detect subtle genetic signals hidden in the human past. All five studies described in this thesis have been conducted through the application of NGS technology in combination with targeted enrichment of either the complete mtDNA or selected SNPs across the human genome. The latter, while representing a powerful tool to capture the often miniscule amount of genetic material still preserved in ancient remains, is limited to known genetic variants for which hybridization probes are designed. An alternative strategy would be to enrich for contiguous regions or entire chromosomes but capture biases and sequence costs are possibly prohibitive for justifying the effort. Ultimately, the only way to reconstruct the entire diversity present in an ancient individual is through sequencing its whole genome. Depending on the percentage of human DNA and its complexity the economic investment may vary but it surely provides a whole range of additional information. Particularly for archaic humans, where only one Neandertal and one Denisovan high coverage genome are currently available, enlarging the sample size would allow a deeper understanding of the intragroup population dynamics as well as the interactions with other hominins. Genomic studies on European hunter-gatherers would also benefit from a denser geographic and temporal sampling in order to identify reservoirs of genetic diversity during severe climatic fluctuations in association to cultural changes. Finally, the retrieval of additional aDNA from tropical environments such as the Southwest Pacific will help to retrace demographic events in that region and provide further insights into waves of population dispersal during human prehistory in Remote Oceania.

References

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Gilbert, M. T., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N. and Bunce, M. (2012) 'The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils', *Proc Biol Sci,* 279(1748), pp. 4724-33.

Allentoft, M. E., Sikora, M., Sjogren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlstrom, T., Vinner, L., Malaspinas, A. S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Della Casa, P., Dabrowski, P., Duffy, P. R., Ebel, A. V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolar, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R., Moiseyev, V., Paja, L., Palfi, G., Pokutta, D., Pospieszny, L., Price, T. D., Saag, L., Sablin, M., Shishlina, N., Smrcka, V., Soenov, V. I., Szeverenyi, V., Toth, G., Trifanova, S. V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sicheritz-Ponten, T., Brunak, S., Nielsen, R., Kristiansen, K. and Willerslev, E. (2015) 'Population genomics of Bronze Age Eurasia', *Nature,* 522(7555), pp. 167-72.

Alt, K. W., Knipper, C., Peters, D., Muller, W., Maurer, A. F., Kollig, I., Nicklisch, N., Muller, C., Karimnia, S., Brandt, G., Roth, C., Rosner, M., Mende, B., Schone, B. R., Vida, T. and von Freeden, U. (2014) 'Lombards on the move--an integrative study of the migration period cemetery at Szolad, Hungary', *PLoS One,* 9(11), pp. e110793.

Armitage, S. J., Jasim, S. A., Marks, A. E., Parker, A. G., Usik, V. I. and Uerpmann, H. P. (2011) 'The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia', *Science,* 331(6016), pp. 453-6.

Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogvali, E. L., Silva, N. M., Kivisild, T., Torroni, A. and Villems, R. (2012) 'A "Copernican" reassessment of the human mitochondrial DNA tree from its root', *Am J Hum Genet,* 90(4), pp. 675-84.

Bellwood, P. (2005) 'First farmers', *The Origins of Agricultural Societies*.

Benazzi, S., Douka, K., Fornai, C., Bauer, C. C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., Condemi, S., Ronchitelli, A., Harvati, K. and Weber, G. W. (2011) 'Early dispersal of modern humans in Europe and implications for Neanderthal behaviour', *Nature,* 479(7374), pp. 525-8.

Benazzi, S., Slon, V., Talamo, S., Negrino, F., Peresani, M., Bailey, S. E., Sawyer, S., Panetta, D., Vicino, G., Starnini, E., Mannino, M. A., Salvadori, P. A., Meyer, M., Paabo, S. and Hublin, J. J. (2015) 'Archaeology. The makers of the Protoaurignacian and implications for Neandertal extinction', *Science,* 348(6236), pp. 793-6.

Bollongino, R., Nehlich, O., Richards, M. P., Orschiedt, J., Thomas, M. G., Sell, C., Fajkosova, Z., Powell, A. and Burger, J. (2013) '2000 years of parallel societies in Stone Age Central Europe', *Science,* 342(6157), pp. 479-81.

Bowler, J. M., Johnston, H., Olley, J. M., Prescott, J. R., Roberts, R. G., Shawcross, W. and Spooner, N. A. (2003) 'New ages for human occupation and climatic change at Lake Mungo, Australia', *Nature,* 421(6925), pp. 837-40.

Brandt, G., Haak, W., Adler, C. J., Roth, C., Szecsenyi-Nagy, A., Karimnia, S., Moller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., Dresely, V., Nicklisch, N., Pickrell, J. K., Sirocko, F., Reich, D., Cooper, A., Alt, K. W. and Genographic, C. (2013) 'Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity', *Science,* 342(6155), pp. 257-61.

Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M. and Paabo, S. (2007) 'Patterns of damage in genomic DNA sequences from a Neandertal', *Proc Natl Acad Sci U S A,* 104(37), pp. 14616-21.

Briggs, A. W., Good, J. M., Green, R. E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V. B., Golovanova, L. V., de la Rasilla, M., Fortea, J., Rosas, A. and Paabo, S. (2009) 'Targeted retrieval and analysis of five Neandertal mtDNA genomes', *Science,* 325(5938), pp. 318-21.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M. and Paabo, S. (2010) 'Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA', *Nucleic Acids Res,* 38(6), pp. e87.

Brown, S., Higham, T., Slon, V., Paabo, S., Meyer, M., Douka, K., Brock, F., Comeskey, D., Procopio, N., Shunkov, M., Derevianko, A. and Buckley, M. (2016) 'Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis', *Sci Rep,* 6, pp. 23559.

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J. and Paabo, S. (2010) 'Targeted investigation of the Neandertal genome by array-based sequence capture', *Science,* 328(5979), pp. 723-5.

Cann, R. L., Stoneking, M. and Wilson, A. C. (1987) 'Mitochondrial DNA and human evolution', *Nature,* 325(6099), pp. 31-6.

Cooper, A. and Poinar, H. N. (2000) 'Ancient DNA: Do it right or not at ALL', *Science,* 289(5482), pp. 1139-1139.

Cooper, A., Turney, C., Hughen, K. A., Brook, B. W., McDonald, H. G. and Bradshaw, C. J. (2015) 'PALEOECOLOGY. Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover', *Science,* 349(6248), pp. 602-6.

Currat, M. and Excoffier, L. (2004) 'Modern humans did not admix with Neanderthals during their range expansion into Europe', *PLoS Biol,* 2(12), pp. e421.

Dabney, J. and Meyer, M. (2012) 'Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries', *Biotechniques,* 52(2), pp. 87-94.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M. T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J. L. and Meyer, M. (2013a) 'Complete

mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments', *Proc Natl Acad Sci U S A,* 110(39), pp. 15758-63.

Dabney, J., Meyer, M. and Paabo, S. (2013b) 'Ancient DNA damage', *Cold Spring Harb Perspect Biol,* 5(7), pp. 15758-63.

Diamond, J. M. (1988) 'Express train to Polynesia', *Nature,* 336, pp. 307-308.

Drummond, A. J. and Rambaut, A. (2007) 'BEAST: Bayesian evolutionary analysis by sampling trees', *BMC Evol Biol,* 7, pp. 214.

Fu, Q., Mittnik, A., Johnson, P. L., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Paabo, S. and Krause, J. (2013a) 'A revised timescale for human evolution based on ancient mitochondrial genomes', *Curr Biol,* 23(7), pp. 553-9.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., Johnson, P. L., Aximu-Petri, A., Prufer, K., de Filippo, C., Meyer, M., Zwyns, N., Salazar-Garcia, D. C., Kuzmin, Y. V., Keates, S. G., Kosintsev, P. A., Razhev, D. I., Richards, M. P., Peristov, N. V., Lachmann, M., Douka, K., Higham, T. F., Slatkin, M., Hublin, J. J., Reich, D., Kelso, J., Viola, T. B. and Paabo, S. (2014) 'Genome sequence of a 45,000-year-old modern human from western Siberia', *Nature,* 514(7523), pp. 445-9.

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., Viola, B., Prufer, K., Meyer, M., Kelso, J., Reich, D. and Paabo, S. (2015) 'An early modern human from Romania with a recent Neanderthal ancestor', *Nature,* 524(7564), pp. 216-9.

Fu, Q. M., Meyer, M., Gao, X., Stenzel, U., Burbano, H. A., Kelso, J. and Paabo, S. (2013b) 'DNA analysis of an early modern human from Tianyuan Cave, China', *Proceedings of the National Academy of Sciences of the United States of America,* 110(6), pp. 2223-2227.

Fu, Q. M., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwangler, A., Haak, W., Meyer, M., Mittnik, A., Nickel, B., Peltzer, A., Rohland, N., Slon, V., Talamo, S., Lazaridis, I., Lipson, M., Mathieson, I., Schiffels, S., Skoglund, P., Derevianko, A. P., Drozdov, N., Slavinsky, V., Tsybankov, A., Cremonesi, R. G., Mallegni, F., Gely, B., Vacca, E., Morales, M. R. G., Straus, L. G., Neugebauer-Maresch, C., Teschler-Nicola, M., Constantin, S., Moldovan, O. T., Benazzi, S., Peresani, M., Coppola, D., Lari, M., Ricci, S., Ronchitelli, A., Valentin, F., Thevenet, C., Wehrberger, K., Grigorescu, D., Rougier, H., Crevecoeur, I., Flas, D., Semal, P., Mannino, M. A., Cupillard, C., Bocherens, H., Conard, N. J., Harvati, K., Moiseyev, V., Drucker, D. G., Svoboda, J., Richards, M. P., Caramelli, D., Pinhasi, R., Kelso, J., Patterson, N., Krause, J., Paabo, S. and Reich, D. (2016) 'The genetic history of Ice Age Europe', *Nature,* 534(7606), pp. 200-+.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboroczki, L., Kovari, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F., Hofreiter, M., Bradley, D. G. and Pinhasi, R. (2014) 'Genome flux and stasis in a five millennium transect of European prehistory', *Nat Commun,* 5, pp. 5257.

Gamble, C., Davies, W., Pettitt, P. and Richards, M. (2004) 'Climate change and evolving human diversity in Europe during the last glacial', *Philos Trans R Soc Lond B Biol Sci,* 359(1442), pp. 243-53; discussion 253-4.

Gansauge, M. T. and Meyer, M. (2013) 'Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA', *Nat Protoc,* 8(4), pp. 737-48.

Gansauge, M. T. and Meyer, M. (2014) 'Selective enrichment of damaged DNA molecules for ancient genome sequencing', *Genome Res,* 24(9), pp. 1543-9.

Gray, R. D. and Jordan, F. M. (2000) 'Language trees support the express-train sequence of Austronesian expansion', *Nature,* 405(6790), pp. 1052-5.

Gray, R. D., Drummond, A. J. and Greenhill, S. J. (2009) 'Language phylogenies reveal expansion pulses and pauses in Pacific settlement', *Science,* 323(5913), pp. 479-83.

Green, R. E., Krause, J., Ptak, S. E., Briggs, A. W., Ronan, M. T., Simons, J. F., Du, L., Egholm, M., Rothberg, J. M., Paunovic, M. and Paabo, S. (2006) 'Analysis of one million base pairs of Neanderthal DNA', *Nature,* 444(7117), pp. 330-6.

Green, R. E., Malaspinas, A. S., Krause, J., Briggs, A. W., Johnson, P. L., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prufer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Wikstrom, M., Laakkonen, L., Kelso, J., Slatkin, M. and Paabo, S. (2008) 'A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing', *Cell,* 134(3), pp. 416-26.

Green, R. E., Briggs, A. W., Krause, J., Prufer, K., Burbano, H. A., Siebauer, M., Lachmann, M. and Paabo, S. (2009) 'The Neandertal genome and ancient DNA authenticity', *EMBO J,* 28(17), pp. 2494-502.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H., Hansen, N. F., Durand, E. Y., Malaspinas, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prufer, K., Meyer, M., Burbano, H. A., Good, J. M., Schultz, R., Aximu-Petri, A., Butthof, A., Hober, B., Hoffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E. S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V. B., Golovanova, L. V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R. W., Johnson, P. L., Eichler, E. E., Falush, D., Birney, E., Mullikin, J. C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D. and Paabo, S. (2010) 'A draft sequence of the Neandertal genome', *Science,* 328(5979), pp. 710-22.

Groucutt, H. S., Petraglia, M. D., Bailey, G., Scerri, E. M., Parton, A., Clark-Balzan, L., Jennings, R. P., Lewis, L., Blinkhorn, J., Drake, N. A., Breeze, P. S., Inglis, R. H., Deves, M. H., Meredith-Williams, M., Boivin, N., Thomas, M. G. and Scally, A. (2015) 'Rethinking the dispersal of Homo sapiens out of Africa', *Evol Anthropol,* 24(4), pp. 149-64.

Haak, W., Balanovsky, O., Sanchez, J. J., Koshel, S., Zaporozhchenko, V., Adler, C. J., Der Sarkissian, C. S., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., Villems, R., Meller, H., Alt, K. W., Cooper, A. and Members of the Genographic, C. (2010) 'Ancient DNA from European early neolithic farmers reveals their near eastern affinities', *PLoS Biol,* 8(11), pp. e1000536.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Banffy, E., Economou, C., Francken, M., Friederich, S., Pena, R. G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S. L., Risch, R., Rojo Guerra, M. A., Roth, C., Szecsenyi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K. W. and Reich, D. (2015) 'Massive migration from the steppe was a source for Indo-European languages in Europe', *Nature,* 522(7555), pp. 207-11.

Heiri, O., Brooks, S. J., Renssen, H., Bedford, A., Hazekamp, M., Ilyashuk, B., Jeffers, E. S., Lang, B., Kirilova, E., Kuiper, S., Millet, L., Samartin, S., Toth, M., Verbruggen, F., Watson, J. E., van Asch, N., Lammertsma, E., Amon, L., Birks, H. H., Birks, H. J., Mortensen, M. F., Hoek, W. Z., Magyari, E., Munoz Sobrino, C., Seppa, H., Tinner, W., Tonkov, S., Veski, S. and Lotter, A. F. (2014) 'Validation of climate model-inferred regional temperature change for late-glacial Europe', *Nat Commun,* 5, pp. 4914.

Higham, T., Compton, T., Stringer, C., Jacobi, R., Shapiro, B., Trinkaus, E., Chandler, B., Groning, F., Collins, C., Hillson, S., O'Higgins, P., FitzGerald, C. and Fagan, M. (2011) 'The earliest evidence for anatomically modern humans in northwestern Europe', *Nature,* 479(7374), pp. 521-4.

Higham, T., Basell, L., Jacobi, R., Wood, R., Ramsey, C. B. and Conard, N. J. (2012) 'Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: the radiocarbon chronology of Geissenklosterle', *J Hum Evol,* 62(6), pp. 664-76.

Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruiz, C., Bergman, C., Boitard, C., Boscato, P., Caparros, M., Conard, N. J., Draily, C., Froment, A., Galvan, B., Gambassini, P., Garcia-Moreno, A., Grimaldi, S., Haesaerts, P., Holt, B., Iriarte-Chiapusso, M. J., Jelinek, A., Jorda Pardo, J. F., Maillo-Fernandez, J. M., Marom, A., Maroto, J., Menendez, M., Metz, L., Morin, E., Moroni, A., Negrino, F., Panagopoulou, E., Peresani, M., Pirson, S., de la Rasilla, M., Riel-Salvatore, J., Ronchitelli, A., Santamaria, D., Semal, P., Slimak, L., Soler, J., Soler, N., Villaluenga, A., Pinhasi, R. and Jacobi, R. (2014) 'The timing and spatiotemporal patterning of Neanderthal disappearance', *Nature,* 512(7514), pp. 306-9.

Hofmanova, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Diez-Del-Molino, D., van Dorp, L., Lopez, S., Kousathanas, A., Link, V., Kirsanow, K., Cassidy, L. M., Martiniano, R., Strobel, M., Scheu, A., Kotsakis, K., Halstead, P., Triantaphyllou, S., Kyparissi-Apostolika, N., Urem-Kotsou, D., Ziota, C., Adaktylou, F., Gopalan, S., Bobo, D. M., Winkelbach, L., Blocher, J., Unterlander, M., Leuenberger, C., Cilingiroglu, C., Horejs, B., Gerritsen, F., Shennan, S. J., Bradley, D. G., Currat, M., Veeramah, K. R., Wegmann, D., Thomas, M. G., Papageorgopoulou, C. and Burger, J. (2016) 'Early farmers from across Europe directly descended from Neolithic Aegeans', *Proc Natl Acad Sci U S A*.

Jones, E. R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R. L., Gallego Llorente, M., Cassidy, L. M., Gamba, C., Meshveliani, T., Bar-Yosef, O., Muller, W., Belfer-Cohen, A., Matskevich, Z., Jakeli, N.,

Higham, T. F., Currat, M., Lordkipanidze, D., Hofreiter, M., Manica, A., Pinhasi, R. and Bradley, D. G. (2015) 'Upper Palaeolithic genomes reveal deep roots of modern Eurasians', *Nat Commun,* 6, pp. 8912.

Kayser, M., Lao, O., Saar, K., Brauer, S., Wang, X., Nurnberg, P., Trent, R. J. and Stoneking, M. (2008) 'Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians', *Am J Hum Genet,* 82(1), pp. 194-8.

Kayser, M. (2010) 'The human genetic history of Oceania: near and remote views of dispersal', *Curr Biol,* 20(4), pp. R194-201.

Kircher, M. (2012) 'Analysis of high-throughput ancient DNA sequencing data', *Methods Mol Biol,* 840, pp. 197-228.

Kircher, M., Sawyer, S. and Meyer, M. (2012) 'Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform', *Nucleic Acids Res,* 40(1), pp. e3.

Kivisild, T. (2015) 'Maternal ancestry and population history from whole mitochondrial genomes', *Investig Genet,* 6, pp. 3.

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., Thorsteinsdottir, U. and Stefansson, K. (2012) 'Rate of de novo mutations and the importance of father's age to disease risk', *Nature,* 488(7412), pp. 471-5.

Korneliussen, T. S., Albrechtsen, A. and Nielsen, R. (2014) 'ANGSD: Analysis of Next Generation Sequencing Data', *Bmc Bioinformatics,* 15.

Krause, J., Orlando, L., Serre, D., Viola, B., Prufer, K., Richards, M. P., Hublin, J. J., Hanni, C., Derevianko, A. P. and Paabo, S. (2007) 'Neanderthals in central Asia and Siberia', *Nature,* 449(7164), pp. 902-4.

Krause, J., Briggs, A. W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A. and Paabo, S. (2010a) 'A complete mtDNA genome of an early modern human from Kostenki, Russia', *Curr Biol,* 20(3), pp. 231-6.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P. and Paabo, S. (2010b) 'The complete mitochondrial DNA genome of an unknown hominin from southern Siberia', *Nature,* 464(7290), pp. 894-7.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M. and Paabo, S. (1997) 'Neandertal DNA sequences and the origin of modern humans', *Cell,* 90(1), pp. 19-30.

Kuhlwilm, M., Gronau, I., Hubisz, M. J., de Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H. A., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Marques-Bonet, T., Andres, A. M., Viola, B., Paabo, S., Meyer, M., Siepel, A. and Castellano, S. (2016) 'Ancient gene flow from early modern humans into Eastern Neanderthals', *Nature,* 530(7591), pp. 429-33.

Lahr, M. M. and Foley, R. A. (1998) 'Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution', *Am J Phys Anthropol,* Suppl 27, pp. 137-76.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J. G., Castellano, S., Lipson, M., Berger, B., Economou, C.,

Bollongino, R., Fu, Q., Bos, K. I., Nordenfelt, S., Li, H., de Filippo, C., Prufer, K., Sawyer, S., Posth, C., Haak, W., Hallgren, F., Fornander, E., Rohland, N., Delsate, D., Francken, M., Guinet, J. M., Wahl, J., Ayodo, G., Babiker, H. A., Bailliet, G., Balanovska, E., Balanovsky, O., Barrantes, R., Bedoya, G., Ben-Ami, H., Bene, J., Berrada, F., Bravi, C. M., Brisighelli, F., Busby, G. B., Cali, F., Churnosov, M., Cole, D. E., Corach, D., Damba, L., van Driem, G., Dryomov, S., Dugoujon, J. M., Fedorova, S. A., Gallego Romero, I., Gubina, M., Hammer, M., Henn, B. M., Hervig, T., Hodoglugil, U., Jha, A. R., Karachanak-Yankova, S., Khusainova, R., Khusnutdinova, E., Kittles, R., Kivisild, T., Klitz, W., Kucinskas, V., Kushniarevich, A., Laredj, L., Litvinov, S., Loukidis, T., Mahley, R. W., Melegh, B., Metspalu, E., Molina, J., Mountain, J., Nakkalajarvi, K., Nesheva, D., Nyambo, T., Osipova, L., Parik, J., Platonov, F., Posukh, O., Romano, V., Rothhammer, F., Rudan, I., Ruizbakiev, R., Sahakyan, H., Sajantila, A., Salas, A., Starikovskaya, E. B., Tarekegn, A., Toncheva, D., Turdikulova, S., Uktveryte, I., Utevska, O., Vasquez, R., Villena, M., Voevoda, M., Winkler, C. A., Yepiskoposyan, L., Zalloua, P., Zemunik, T., Cooper, A., Capelli, C., Thomas, M. G., Ruiz-Linares, A., Tishkoff, S. A., Singh, L., Thangaraj, K., Villems, R., Comas, D., Sukernik, R., Metspalu, M., Meyer, M., Eichler, E. E., Burger, J., Slatkin, M., Paabo, S., Kelso, J., Reich, D. and Krause, J. (2014) 'Ancient human genomes suggest three ancestral populations for present-day Europeans', *Nature,* 513(7518), pp. 409-13.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D. C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., Connell, S., Stewardson, K., Harney, E., Fu, Q. M., Gonzalez-Fortes, G., Jones, E. R., Roodenberg, S. A., Lengyel, G., Bocquentin, F., Gasparian, B., Monge, J. M., Gregg, M., Eshed, V., Mizrahi, A. S., Meiklejohn, C., Gerritsen, F., Bejenaru, L., Bluher, M., Campbell, A., Cavalleri, G., Comas, D., Froguel, P., Gilbert, E., Kerr, S. M., Kovacs, P., Krause, J., McGettigan, D., Merrigan, M., Merriwether, D. A., O'Reilly, S., Richards, M. B., Semino, O., Shamoon-Pour, M., Stefanescu, G., Stumvoll, M., Tonjes, A., Torroni, A., Wilson, J. F., Yengo, L., Hovhannisyan, N. A., Patterson, N., Pinhasi, R. and Reich, D. (2016) 'Genomic insights into the origin of farming in the ancient Near East', *Nature,* 536(7617), pp. 419-+.

Liu, W., Martinon-Torres, M., Cai, Y. J., Xing, S., Tong, H. W., Pei, S. W., Sier, M. J., Wu, X. H., Edwards, R. L., Cheng, H., Li, Y. Y., Yang, X. X., de Castro, J. M. and Wu, X. J. (2015) 'The earliest unequivocally modern humans in southern China', *Nature,* 526(7575), pp. 696-9.

Maca-Meyer, N., Gonzalez, A. M., Larruga, J. M., Flores, C. and Cabrera, V. M. (2001) 'Major genomic mitochondrial lineages delineate early human expansions', *BMC Genet,* 2, pp. 13.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N. K., Raja, J. M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H. J., Oppenheimer, S., Torroni, A. and Richards, M. (2005) 'Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes', *Science,* 308(5724), pp. 1034-6.

Malaspinas, A. S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., Bergstrom, A., Athanasiadis, G., Cheng, J. Y., Crawford, J. E., Heupink, T. H., Macholdt, E., Peischl, S., Rasmussen, S., Schiffels, S., Subramanian, S., Wright, J. L., Albrechtsen, A., Barbieri, C., Dupanloup, I., Eriksson, A., Margaryan, A., Moltke, I., Pugach, I., Korneliussen, T. S., Levkivskyi, I. P., Moreno-Mayar, J. V., Ni, S., Racimo, F., Sikora, M., Xue, Y., Aghakhanian, F. A., Brucato, N., Brunak, S., Campos, P. F., Clark, W., Ellingvag, S., Fourmile, G., Gerbault, P., Injie, D., Koki, G., Leavesley, M., Logan, B., Lynch, A., Matisoo-Smith, E. A., McAllister, P. J., Mentzer, A. J., Metspalu, M., Migliano, A. B., Murgha, L., Phipps, M. E., Pomat, W., Reynolds, D., Ricaut, F. X., Siba, P., Thomas, M. G., Wales, T., Wall, C. M., Oppenheimer, S. J., Tyler-Smith, C., Durbin, R., Dortch, J., Manica, A., Schierup, M. H., Foley, R. A., Lahr, M. M., Bowern, C., Wall, J. D., Mailund, T., Stoneking, M., Nielsen, R., Sandhu, M. S., Excoffier, L., Lambert, D. M. and Willerslev, E. (2016) 'A genomic history of Aboriginal Australia', *Nature,* 538(7624), pp. 207-214.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., Spence, J. P., Song, Y. S., Poletti, G., Balloux, F., van Driem, G., de Knijff, P., Romero, I. G., Jha, A. R., Behar, D. M., Bravi, C. M., Capelli, C., Hervig, T., Moreno-Estrada, A., Posukh, O. L., Balanovska, E., Balanovsky, O., Karachanak-Yankova, S., Sahakyan, H., Toncheva, D., Yepiskoposyan, L., Tyler-Smith, C., Xue, Y., Abdullah, M. S., Ruiz-Linares, A., Beall, C. M., Di Rienzo, A., Jeong, C., Starikovskaya, E. B., Metspalu, E., Parik, J., Villems, R., Henn, B. M., Hodoglugil, U., Mahley, R., Sajantila, A., Stamatoyannopoulos, G., Wee, J. T., Khusainova, R., Khusnutdinova, E., Litvinov, S., Ayodo, G., Comas, D., Hammer, M. F., Kivisild, T., Klitz, W., Winkler, C. A., Labuda, D., Bamshad, M., Jorde, L. B., Tishkoff, S. A., Watkins, W. S., Metspalu, M., Dryomov, S., Sukernik, R., Singh, L., Thangaraj, K., Paabo, S., Kelso, J., Patterson, N. and Reich, D. (2016) 'The Simons Genome Diversity Project: 300 genomes from 142 diverse populations', *Nature,* 538(7624), pp. 201-206.

Maricic, T., Whitten, M. and Paabo, S. (2010) 'Multiplexed DNA sequence capture of mitochondrial genomes using PCR products', *PLoS One,* 5(11), pp. e14004.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E. R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J. L., de Castro, J. M., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M. A., Roodenberg, J., Verges, J. M., Krause, J., Cooper, A., Alt, K. W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R. and Reich, D. (2015) 'Genome-wide patterns of selection in 230 ancient Eurasians', *Nature,* 528(7583), pp. 499-503.

Matisoo-Smith, E. (2015) 'Ancient DNA and the human settlement of the Pacific: a review', *J Hum Evol,* 79, pp. 93-104.

Mellars, P., Gori, K. C., Carr, M., Soares, P. A. and Richards, M. B. (2013) 'Genetic and archaeological perspectives on the initial modern human colonization of southern Asia', *Proc Natl Acad Sci U S A,* 110(26), pp. 10699-704.

Mendez, F. L., Poznik, G. D., Castellano, S. and Bustamante, C. D. (2016) 'The Divergence of Neandertal and Modern Human Y Chromosomes', *Am J Hum Genet,* 98(4), pp. 728-34.

Meyer, M. and Kircher, M. (2010) 'Illumina sequencing library preparation for highly multiplexed target capture and sequencing', *Cold Spring Harb Protoc,* 2010(6), pp. pdb prot5448.

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prufer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., Briggs, A. W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M. F., Shunkov, M. V., Derevianko, A. P., Patterson, N., Andres, A. M., Eichler, E. E., Slatkin, M., Reich, D., Kelso, J. and Paabo, S. (2012) 'A high-coverage genome sequence from an archaic Denisovan individual', *Science,* 338(6104), pp. 222-6.

Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J. L., Martinez, I., Gracia, A., de Castro, J. M., Carbonell, E. and Paabo, S. (2014) 'A mitochondrial genome sequence of a hominin from Sima de los Huesos', *Nature,* 505(7483), pp. 403-6.

Meyer, M., Arsuaga, J. L., de Filippo, C., Nagel, S., Aximu-Petri, A., Nickel, B., Martinez, I., Gracia, A., Bermudez de Castro, J. M., Carbonell, E., Viola, B., Kelso, J., Prufer, K. and Paabo, S. (2016) 'Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins', *Nature,* 531(7595), pp. 504-7.

Nigst, P. R., Haesaerts, P., Damblon, F., Frank-Fellner, C., Mallol, C., Viola, B., Gotzinger, M., Niven, L., Trnka, G. and Hublin, J. J. (2014) 'Early modern human settlement of Europe north of the Alps occurred 43,500 years ago in a cold steppe-type environment', *Proc Natl Acad Sci U S A,* 111(40), pp. 14394-9.

Olalde, I., Allentoft, M. E., Sanchez-Quinto, F., Santpere, G., Chiang, C. W., DeGiorgio, M., Prado-Martinez, J., Rodriguez, J. A., Rasmussen, S., Quilez, J., Ramirez, O., Marigorta, U. M., Fernandez-Callejo, M., Prada, M. E., Encinas, J. M., Nielsen, R., Netea, M. G., Novembre, J., Sturm, R. A., Sabeti, P., Marques-Bonet, T., Navarro, A., Willerslev, E. and Lalueza-Fox, C. (2014) 'Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European', *Nature,* 507(7491), pp. 225-8.

Oppenheimer, S. (2012) 'A single southern exit of modern humans from Africa: Before or after Toba?', *Quaternary International,* 258, pp. 88-99.

Oppenheimer, S. J. and Richards, M. (2001) 'Polynesian origins. Slow boat to Melanesia?', *Nature,* 410(6825), pp. 166-7.

Paabo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L. and Hofreiter, M. (2004) 'Genetic analyses from ancient DNA', *Annual Review of Genetics,* 38, pp. 645-679.

Pagani, L., Lawson, D. J., Jagoda, E., Morseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L., Wall, J. D., Cardona, A., Magi, R., Sayres, M. A., Kaewert, S., Inchley, C., Scheib, C. L., Jarve, M., Karmin, M., Jacobs, G. S., Antao, T., Iliescu, F. M., Kushniarevich, A., Ayub, Q., Tyler-Smith, C., Xue, Y., Yunusbayev, B., Tambets, K., Mallick, C. B., Saag, L., Pocheshkhova, E., Andriadze, G., Muller, C., Westaway, M. C., Lambert, D. M., Zoraqi, G.,

Turdikulova, S., Dalimova, D., Sabitov, Z., Sultana, G. N., Lachance, J., Tishkoff, S., Momynaliev, K., Isakova, J., Damba, L. D., Gubina, M., Nymadawa, P., Evseeva, I., Atramentova, L., Utevska, O., Ricaut, F. X., Brucato, N., Sudoyo, H., Letellier, T., Cox, M. P., Barashkov, N. A., Skaro, V., Mulahasanovic, L., Primorac, D., Sahakyan, H., Mormina, M., Eichstaedt, C. A., Lichman, D. V., Abdullah, S., Chaubey, G., Wee, J. T., Mihailov, E., Karunas, A., Litvinov, S., Khusainova, R., Ekomasova, N., Akhmetova, V., Khidiyatova, I., Marjanovic, D., Yepiskoposyan, L., Behar, D. M., Balanovska, E., Metspalu, A., Derenko, M., Malyarchuk, B., Voevoda, M., Fedorova, S. A., Osipova, L. P., Lahr, M. M., Gerbault, P., Leavesley, M., Migliano, A. B., Petraglia, M., Balanovsky, O., Khusnutdinova, E. K., Metspalu, E., Thomas, M. G., Manica, A., Nielsen, R., Villems, R., Willerslev, E., Kivisild, T. and Metspalu, M. (2016) 'Genomic analyses inform on migration events during the peopling of Eurasia', *Nature*.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D. (2012) 'Ancient admixture in human history', *Genetics,* 192(3), pp. 1065-93.

Peltzer, A., Jager, G., Herbig, A., Seitz, A., Kniep, C., Krause, J. and Nieselt, K. (2016) 'EAGER: efficient ancient genome reconstruction', *Genome Biol,* 17(1), pp. 60.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietrusewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H. and Hofreiter, M. (2015) 'Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone', *PLoS One,* 10(6), pp. e0129102.

Posth, C., Renaud, G., Mittnik, A., Drucker, D. G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwangler, A., Wissing, C., Francken, M., Malina, M., Bolus, M., Lari, M., Gigli, E., Capecchi, G., Crevecoeur, I., Beauval, C., Flas, D., Germonpre, M., van der Plicht, J., Cottiaux, R., Gely, B., Ronchitelli, A., Wehrberger, K., Grigorescu, D., Svoboda, J., Semal, P., Caramelli, D., Bocherens, H., Harvati, K., Conard, N. J., Haak, W., Powell, A. and Krause, J. (2016a) 'Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe', *Curr Biol,* 26(6), pp. 827-33.

Posth, C., Wißing, C., Kitagawa, K., Pagani, L., van Holstein, L., Wehrberger, K., Conard, N. J., Kind, C. J., Bocherens, H. and Krause, J. (2016b) 'Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neandertals', *Review in Nature Communications*.

Prufer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., Siebauer, M., Theunert, C., Tandon, A., Moorjani, P., Pickrell, J., Mullikin, J. C., Vohr, S. H., Green, R. E., Hellmann, I., Johnson, P. L., Blanche, H., Cann, H., Kitzman, J. O., Shendure, J., Eichler, E. E., Lein, E. S., Bakken, T. E., Golovanova, L. V., Doronichev, V. B., Shunkov, M. V., Derevianko, A. P., Viola, B., Slatkin, M., Reich, D., Kelso, J. and Paabo, S. (2014) 'The complete genome sequence of a Neanderthal from the Altai Mountains', *Nature,* 505(7481), pp. 43-9.

Prufer, K. and Meyer, M. (2015) 'Comment on "Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans"', *Science,* 347(6224).

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M. T., Wang, Y., Raghavan, M., Campos, P. F., Kamp, H. M., Wilson, A. S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E. D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T. L., Gronnow, B., Meldgaard, M., Andreasen, C., Fedorova, S. A., Osipova, L. P., Higham, T. F., Ramsey, C. B., Hansen, T. V., Nielsen, F. C., Crawford, M. H., Brunak, S., Sicheritz-Ponten, T., Villems, R., Nielsen, R., Krogh, A., Wang, J. and Willerslev, E. (2010) 'Ancient human genome sequence of an extinct Palaeo-Eskimo', *Nature,* 463(7282), pp. 757-62.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M. V., Derevianko, A. P., Hublin, J. J., Kelso, J., Slatkin, M. and Paabo, S. (2010) 'Genetic history of an archaic hominin group from Denisova Cave in Siberia', *Nature,* 468(7327), pp. 1053-60.

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., Ko, A. M., Ko, Y. C., Jinam, T. A., Phipps, M. E., Saitou, N., Wollstein, A., Kayser, M., Paabo, S. and Stoneking, M. (2011) 'Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania', *Am J Hum Genet,* 89(4), pp. 516-28.

Renaud, G., Slon, V., Duggan, A. T. and Kelso, J. (2015) 'Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA', *Genome Biol,* 16(1), pp. 224.

Reyes-Centeno, H., Ghirotto, S., Detroit, F., Grimaud-Herve, D., Barbujani, G. and Harvati, K. (2014) 'Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia', *Proc Natl Acad Sci U S A,* 111(20), pp. 7248-53.

Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L. A., Warmuth, V., Ruiz-Linares, A., Manica, A. and Balloux, F. (2014) 'Improved calibration of the human mitochondrial clock using ancient genomes', *Mol Biol Evol,* 31(10), pp. 2780-92.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. and Reich, D. (2015) 'Partial uracil-DNA-glycosylase treatment for screening of ancient DNA', *Philos Trans R Soc Lond B Biol Sci,* 370(1660), pp. 20130624.

Rougier, H., Crevecoeur, I., Beauval, C., Posth, C., Flas, D., Wissing, C., Furtwangler, A., Germonpre, M., Gomez-Olivencia, A., Semal, P., van der Plicht, J., Bocherens, H. and Krause, J. (2016) 'Neandertal cannibalism and Neandertal bones used as tools in Northern Europe', *Sci Rep,* 6, pp. 29005.

Ruano, G. and Kidd, K. K. (1992) 'Modeling of heteroduplex formation during PCR from mixtures of DNA templates', *PCR Methods Appl,* 2(2), pp. 112-6.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. and Paabo, S. (2012) 'Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA', *PLoS One,* 7(3), pp. e34131.

Sawyer, S., Renaud, G., Viola, B., Hublin, J. J., Gansauge, M. T., Shunkov, M. V., Derevianko, A. P., Prufer, K., Kelso, J. and Paabo, S. (2015) 'Nuclear and mitochondrial DNA sequences from two Denisovan individuals', *Proc Natl Acad Sci U S A,* 112(51), pp. 15696-700.

Scally, A. and Durbin, R. (2012) 'Revising the human mutation rate: implications for understanding human evolution', *Nat Rev Genet,* 13(10), pp. 745-53.

Schroeder, H., Avila-Arcos, M. C., Malaspinas, A. S., Poznik, G. D., Sandoval-Velasco, M., Carpenter, M. L., Moreno-Mayar, J. V., Sikora, M., Johnson, P. L., Allentoft, M. E., Samaniego, J. A., Haviser, J. B., Dee, M. W., Stafford, T. W., Jr., Salas, A., Orlando, L., Willerslev, E., Bustamante, C. D. and Gilbert, M. T. (2015) 'Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean', *Proc Natl Acad Sci U S A,* 112(12), pp. 3669-73.

Seguin-Orlando, A., Korneliussen, T. S., Sikora, M., Malaspinas, A. S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., Goebel, T., Westaway, M., Lambert, D., Khartanovich, V., Wall, J. D., Nigst, P. R., Foley, R. A., Lahr, M. M., Nielsen, R., Orlando, L. and Willerslev, E. (2014) 'Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years', *Science,* 346(6213), pp. 1113-8.

Serre, D., Langaney, A., Chech, M., Teschler-Nicola, M., Paunovic, M., Mennecier, P., Hofreiter, M., Possnert, G. and Paabo, S. (2004) 'No evidence of Neandertal mtDNA contribution to early modern humans', *PLoS Biol,* 2(3), pp. E57.

Sheppard, P. J., Chiu, S. and Walter, R. (2015) 'Re-dating Lapita movement into remote Oceania', *Journal of Pacific Archaeology–Vol,* 6(1).

Skoglund, P., Storå, J., Götherström, A. and Jakobsson, M. (2013) 'Accurate sex identification of ancient human remains using DNA shotgun sequencing', *Journal of Archaeological Science,* 40(12), pp. 4477-4482.

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Paabo, S., Krause, J. and Jakobsson, M. (2014) 'Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal', *Proc Natl Acad Sci U S A,* 111(6), pp. 2229-34.

Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G. R., Reepmeyer, C., Petchey, F., Fernandes, D., Fu, Q., Harney, E., Lipson, M., Mallick, S., Novak, M., Rohland, N., Stewardson, K., Abdullah, S., Cox, M. P., Friedlaender, F. R., Friedlaender, J. S., Kivisild, T., Koki, G., Kusuma, P., Merriwether, D. A., Ricaut, F.-X., Wee, J. T. S., Patterson, N., Krause, J., Pinhasi, R. and Reich, D. (2016) 'Genomic insights into the peopling of the Southwest Pacific', *Nature,* 538(7626), pp. 510-513.

Slatkin, M. and Racimo, F. (2016) 'Ancient DNA and human history', *Proc Natl Acad Sci U S A,* 113(23), pp. 6380-7.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V. and Richards, M. B. (2009) 'Correcting for purifying selection: an improved human mitochondrial molecular clock', *Am J Hum Genet,* 84(6), pp. 740-59.

Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H. J., Torroni, A. and Richards, M. B. (2010) 'The archaeogenetics of Europe', *Curr Biol,* 20(4), pp. R174-83.

Stewart, J. R. and Lister, A. M. (2001) 'Cryptic northern refugia and the origins of the modern biota', *Trends in Ecology & Evolution,* 16(11), pp. 608-613.

Stewart, J. R. and Stringer, C. B. (2012) 'Human evolution out of Africa: the role of refugia and climate change', *Science,* 335(6074), pp. 1317-21.

Straus, L. G. (2013) 'After the deep freeze: confronting "Magdalenian" realities in Cantabrian Spain and beyond', *Journal of Archaeological Method and Theory,* 20(2), pp. 236-255.

Stuart, A. J. and Lister, A. M. (2012) 'Extinction chronology of the woolly rhinoceros Coelodonta antiquitatis in the context of late Quaternary megafaunal extinctions in northern Eurasia', *Quaternary Science Reviews,* 51, pp. 1-17.

Talamo, S., Hajdinjak, M., Mannino, M. A., Fasani, L., Welker, F., Martini, F., Romagnoli, F., Zorzin, R., Meyer, M. and Hublin, J. J. (2016) 'Direct radiocarbon dating and genetic analyses on the purported Neanderthal mandible from the Monti Lessini (Italy)', *Sci Rep,* 6, pp. 29144.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis version 6.0', *Mol Biol Evol,* 30(12), pp. 2725-9.

Trinkaus, E., Moldovan, O., Milota, S., Bilgar, A., Sarcina, L., Athreya, S., Bailey, S. E., Rodrigo, R., Mircea, G., Higham, T., Ramsey, C. B. and van der Plicht, J. (2003) 'An early modern human from the Pestera cu Oase, Romania', *Proc Natl Acad Sci U S A,* 100(20), pp. 11231-6.

Valentin, B. (2008) *Jalons pour une paléohistoire des derniers chasseurs, XIVe-VIe millénaire avant J.-C.* Publications de la Sorbonne.

Valentin, F., Detroit, F., Spriggs, M. J. and Bedford, S. (2016) 'Early Lapita skeletons from Vanuatu show Polynesian craniofacial shape: Implications for Remote Oceanic settlement and Lapita origins', *Proc Natl Acad Sci U S A,* 113(2), pp. 292-7.

Veeramah, K. R. and Hammer, M. F. (2014) 'The impact of whole-genome sequencing on the reconstruction of human population history', *Nat Rev Genet,* 15(3), pp. 149-62.

Weaver, A. J., Saenko, O. A., Clark, P. U. and Mitrovica, J. X. (2003) 'Meltwater pulse 1A from Antarctica as a trigger of the bolling-allerod warm interval', *Science,* 299(5613), pp. 1709-1713.

Weiss, C. L., Dannemann, M., Prufer, K. and Burbano, H. A. (2015) 'Contesting the presence of wheat in the British Isles 8,000 years ago by assessing ancient DNA authenticity from low-coverage data', *Elife,* 4.

White, M. and Pettitt, P. (2012) 'Ancient digs and modern myths: The age and context of the Kent's Cavern 4 maxilla and the earliest Homo sapiens specimens in Europe', *European Journal of Archaeology,* 15(3), pp. 392-420.

White, T. D. (1992) *Prehistoric cannibalism at Mancos 5MTUMR-2346.* Princeton, N.J.: Princeton University Press.

Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R. J., Nurnberg, P., Stoneking, M. and Kayser, M. (2010) 'Demographic history of Oceania inferred from genome-wide data', *Curr Biol,* 20(22), pp. 1983-92.

Figures

Appendix

1. **C. Posth**, G. Renaud, A. Mittnik, D. G. Drucker, H. Rougier, C. Cupillard, F. Valentin, C. Thevenet, A. Furtwangler, C. Wissing, M. Francken, M. Malina, M. Bolus, M. Lari, E. Gigli, G. Capecchi, I. Crevecoeur, C. Beauval, D. Flas, M. Germonpre, J. van der Plicht, R. Cottiaux, B. Gely, A. Ronchitelli, K. Wehrberger, D. Grigorescu, J. Svoboda, P. Semal, D. Caramelli, H. Bocherens, K. Harvati, N. J. Conard, W. Haak, A. Powell and J. Krause (2016). "Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe." *Current Biology* 26(6): 827-833. Including Supplementary Information.

2. Q. Fu, **C. Posth***, M. Hajdinjak*, M. Petr, S. Mallick, D. Fernandes, A. Furtwangler, W. Haak, M. Meyer, A. Mittnik, B. Nickel, A. Peltzer, N. Rohland, V. Slon, S. Talamo, I. Lazaridis, M. Lipson, I. Mathieson, S. Schiffels, P. Skoglund, A. P. Derevianko, N. Drozdov, V. Slavinsky, A. Tsybankov, R. G. Cremonesi, F. Mallegni, B. Gely, E. Vacca, M. R. G. Morales, L. G. Straus, C. Neugebauer-Maresch, M. Teschler-Nicola, S. Constantin, O. T. Moldovan, S. Benazzi, M. Peresani, D. Coppola, M. Lari, S. Ricci, A. Ronchitelli, F. Valentin, C. Thevenet, K. Wehrberger, D. Grigorescu, H. Rougier, I. Crevecoeur, D. Flas, P. Semal, M. A. Mannino, C. Cupillard, H. Bocherens, N. J. Conard, K. Harvati, V. Moiseyev, D. G. Drucker, J. Svoboda, M. P. Richards, D. Caramelli, R. Pinhasi, J. Kelso, N. Patterson, J. Krause, S. Paabo and D. Reich (2016). "The genetic history of Ice Age Europe." *Nature* 534(7606): 200-205. Including Extended Data and Supplementary Information.
   *equal contributors

3. H. Rougier, I. Crevecoeur, C. Beauval, **C. Posth**, D. Flas, C. Wissing, A. Furtwangler, M. Germonpre, A. Gomez-Olivencia, P. Semal, J. van der Plicht, H. Bocherens and J. Krause (2016). "Neandertal cannibalism and Neandertal bones used as tools in Northern Europe." *Scientific Reports* **6**: 29005. Including Supplementary Information.

4. P. Skoglund, **C. Posth**, K. Sirak, M. Spriggs, F. Valentin, S. Bedford, G. R. Clark, C. Reepmeyer, F. Petchey, D. Fernandes, Q. Fu, E. Harney, M. Lipson, S. Mallick, M. Novak, N. Rohland, K. Stewardson, S. Abdullah, M. P. Cox, F. R. Friedlaender, J. S. Friedlaender, T. Kivisild, G. Koki, P. Kusuma, D. A. Merriwether, F.-X. Ricaut, J. T. S. Wee, N. Patterson, J. Krause, R. Pinhasi and D. Reich (2016). "Genomic insights into the peopling of the Southwest Pacific." *Nature* 538(7626): 510-513. Including Extended Data and Supplementary Information.

5. **C. Posth**, C. Wißing, K. Kitagawa, L. Pagani, L. van Holstein, K. Wehrberger, N. J. Conard, C. J. Kind, H. Bocherens, J. Krause (2016). "Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neandertals". Review in *Nature Communications*. Including Supplementary Information.

Report

Current Biology

Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe

Highlights

- Newly generated pre-Neolithic European mtDNA genomes triple the number available
- Clade M found for the first time in Europe, prior to the Last Glacial Maximum bottleneck
- Rapid single dispersal of all non-Africans less than 55,000 years ago
- Previously unknown major population shift in Europe at the end of the Pleistocene

Authors

Cosimo Posth, Gabriel Renaud, Alissa Mittnik, ..., Wolfgang Haak, Adam Powell, Johannes Krause

Correspondence

posth@shh.mpg.de (C.P.), powell@shh.mpg.de (A.P.), krause@shh.mpg.de (J.K.)

In Brief

Posth et al. recover 35 new mitochondrial genomes from Late Pleistocene and early Holocene European hunter-gatherers. Major human mtDNA haplogroup M, absent in contemporary Europeans, is discovered in several pre-Last Glacial Maximum individuals. Demographic modeling reveals a major population turnover during the Late Glacial ~14,500 years ago.

Accession Numbers

KU534947–KU534981

Posth et al., 2016, Current Biology *26*, 1–7
March 21, 2016
http://dx.doi.org/10.1016/j.cub.2016.01.037

CellPress

Please cite this article in press as: Posth et al., Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe, Current Biology (2016), http://dx.doi.org/10.1016/j.cub.2016.01.037

Current Biology
Report

Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe

Cosimo Posth,[1,28,*] Gabriel Renaud,[2] Alissa Mittnik,[1,3] Dorothée G. Drucker,[4] Hélène Rougier,[5] Christophe Cupillard,[6,7] Frédérique Valentin,[8] Corinne Thevenet,[9] Anja Furtwängler,[1] Christoph Wißing,[4] Michael Francken,[10] Maria Malina,[11] Michael Bolus,[11] Martina Lari,[12] Elena Gigli,[12] Giulia Capecchi,[13] Isabelle Crevecoeur,[14] Cédric Beauval,[15] Damien Flas,[16] Mietje Germonpré,[17] Johannes van der Plicht,[18,19] Richard Cottiaux,[9] Bernard Gély,[20] Annamaria Ronchitelli,[13] Kurt Wehrberger,[21] Dan Grigorescu,[22] Jiří Svoboda,[23,24] Patrick Semal,[17] David Caramelli,[12] Hervé Bocherens,[4,25] Katerina Harvati,[10,25] Nicholas J. Conard,[25,26] Wolfgang Haak,[3,27] Adam Powell,[3,*] and Johannes Krause[1,3,25,*]

[1]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Rümelinstraße 23, 72070 Tübingen, Germany
[2]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[3]Max Planck Institute for the Science of Human History, Kahlaische Straße 10, 07745 Jena, Germany
[4]Department of Geosciences, Biogeology, University of Tübingen, Hölderlinstraße 12, 72074 Tübingen, Germany
[5]Department of Anthropology, California State University Northridge, 18111 Nordhoff Street, Northridge, CA 91330-8244, USA
[6]Service Régional d'Archéologie de Franche-Comté, 7 Rue Charles Nodier, 25043 Besançon Cedex, France
[7]Laboratoire de Chrono-Environnement, UMR 6249 du CNRS, UFR des Sciences et Techniques, 16 Route de Gray, 25030 Besançon Cedex, France
[8]CNRS/UMR 7041 ArScAn MAE, 21 Allée de l'Université, 92023 Nanterre, France
[9]INRAP/UMR 8215 Trajectoires, 21 Allée de l'Université, 92023 Nanterre, France
[10]Institute for Archaeological Sciences, Paleoanthropology, University of Tübingen, Rümelinstraße 23, 72070 Tübingen, Germany
[11]Heidelberg Academy of Sciences and Humanities, Research Center "The Role of Culture in Early Expansions of Humans" at the University of Tübingen, Rümelinstraße 23, 72070 Tübingen, Germany
[12]Dipartimento di Biologia, Università di Firenze, Via del Proconsolo 12, 50122 Florence, Italy
[13]Dipartimento di Scienze Fisiche, della Terra e dell'Ambiente, U.R. Preistoria e Antropologia, Università degli Studi di Siena, Via Laterina 8, 53100 Siena, Italy
[14]CNRS, UMR 5199, PACEA, A3P, Université de Bordeaux, Allée Geoffroy Saint Hilaire, CS 50023, 33615 Pessac Cedex, France
[15]Archéosphère, 2 Rue des Noyers, 11500 Quirbajou, France
[16]TRACES, UMR 5608, Université Toulouse Jean Jaurès, Maison de la Recherche, 5 Allée Antonio Machado, 31058 Toulouse Cedex 9, France
[17]Royal Belgian Institute of Natural Sciences, 29 Vautier Street, 1000 Brussels, Belgium
[18]Centre for Isotope Research, Groningen University, Nijenborgh 4, 9747 AG Groningen, the Netherlands
[19]Faculty of Archaeology, Leiden University, PO Box 9514, 2300 RA Leiden, the Netherlands
[20]Direction Régionale des Affaires Culturelles Rhône-Alpes, Le Grenier d'Abondance 6, Quai Saint-Vincent, 69283 Lyon Cedex 01, France
[21]Ulmer Museum, Marktplatz 9, 89073 Ulm, Germany
[22]Department of Geology, Faculty of Geology and Geophysics, University of Bucharest, Bulevardul Nicolae Balcescu 1, 01041 Bucharest, Romania
[23]Department of Anthropology, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic
[24]Institute of Archaeology at Brno, Academy of Science of the Czech Republic, 69129 Dolní Věstonice, Czech Republic
[25]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72072 Tübingen, Germany
[26]Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Schloss Hohentübingen, 72070 Tübingen, Germany
[27]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia
[28]Present address: Max Planck Institute for the Science of Human History, Kahlaische Straße 10, 07745 Jena, Germany
*Correspondence: posth@shh.mpg.de (C.P.), powell@shh.mpg.de (A.P.), krause@shh.mpg.de (J.K.)
http://dx.doi.org/10.1016/j.cub.2016.01.037

SUMMARY

How modern humans dispersed into Eurasia and Australasia, including the number of separate expansions and their timings, is highly debated [1, 2]. Two categories of models are proposed for the dispersal of non-Africans: (1) single dispersal, i.e., a single major diffusion of modern humans across Eurasia and Australasia [3–5]; and (2) multiple dispersal, i.e., additional earlier population expansions that may have contributed to the genetic diversity of some present-day humans outside of Africa [6–9]. Many variants of these models focus largely on Asia and Australasia, neglecting human dispersal into Europe, thus explaining only a subset of the entire colonization process outside of Africa [3–5, 8, 9]. The genetic diversity of the first modern humans who spread into Europe during the Late Pleistocene and the impact of subsequent climatic events on their demography are largely unknown. Here we analyze 55 complete human mitochondrial genomes (mtDNAs) of hunter-gatherers spanning ~35,000 years of European prehistory. We

Please cite this article in press as: Posth et al., Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe, Current Biology (2016), http://dx.doi.org/10.1016/j.cub.2016.01.037

Figure 1. Late Pleistocene and Early Holocene Archeological Sites and Hunter-Gatherer mtDNA Haplogroups

(A) Pre-LGM dispersal of non-African populations, carrying both M and N lineages (hgs R, U, U5, and U2′3′4′7′8′9 belong to the N clade, distinct from the M clade).

(B) Post-LGM re-expansion in Europe while ice sheets retracted.

(C) Late Glacial shift in mtDNA hg frequency.

(D) Holocene hunter-gatherer mtDNA, mainly belonging to hg U5.

See also Table S1, Table S2, Table S4, and the Supplemental Experimental Procedures.

unexpectedly find mtDNA lineage M in individuals prior to the Last Glacial Maximum (LGM). This lineage is absent in contemporary Europeans, although it is found at high frequency in modern Asians, Australasians, and Native Americans. Dating the most recent common ancestor of each of the modern non-African mtDNA clades reveals their single, late, and rapid dispersal less than 55,000 years ago. Demographic modeling not only indicates an LGM genetic bottleneck, but also provides surprising evidence of a major population turnover in Europe around 14,500 years ago during the Late Glacial, a period of climatic instability at the end of the Pleistocene.

RESULTS AND DISCUSSION

Genetic studies of human mitochondrial DNA (mtDNA) show that all present-day non-Africans belong to two basal mtDNA haplogroups (hgs), M and N [10]. The time to the most recent common ancestor (TMRCA) of each of these two clades has been estimated independently at around 50,000 years ago (50 ka) (95% confidence interval [CI], 53–46 ka) and 59 ka (95% CI, 64–54 ka), respectively [11]. However, whereas present-day Asians, Australasians, and Native Americans carry both M and N mtDNA hgs, modern individuals with European ancestry lack almost completely lineages of the M clade [12]. The different spatial distributions and TMRCA estimates of these two ancestral clades have been interpreted as evidence of an early spread of modern humans carrying hg M into Asia, perhaps via a southern route, followed by a later non-African diffusion of the N clade, perhaps via a northern route [7]. However, an alternative model proposes a rapid and single dispersal across Eurasia, with Asia being reached first, whereas Western Eurasia would have been settled only after a hiatus, during which hg M was lost [4].

Little is known about the genetic makeup of the first European hunter-gatherers, who likely arrived ~45 ka [13], or about the subsequent population dynamics during the nearly 40,000 years spanning from the Late Pleistocene to the Neolithic transition [14]. Here, we reconstructed 35 complete or nearly complete mtDNAs (from 11× to 1,860× average coverage) of ancient

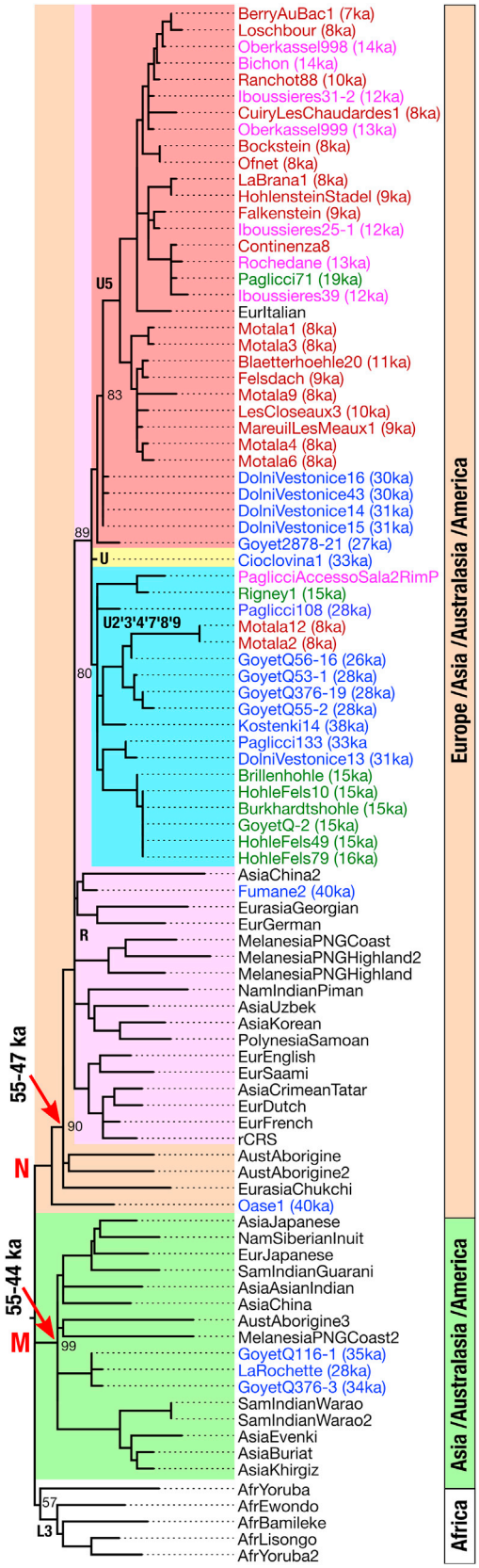


Figure 2. Maximum Parsimony Tree of Present-Day Human and 55 Pre-Neolithic mtDNA Genomes

Pre-LGM samples are shown in blue, Post-LGM in green, Late Glacial in magenta, Holocene hunter-gatherers in red, and present-day individuals in black print. Average values of ^{14}C dates are reported next to each specimen when available. Red arrows indicate divergence times of M and N clades. Hg M is almost absent in present-day individuals with European ancestry. Oase1 represents a pre-N lineage. The tree is rooted with one Neanderthal and 16 deeply divergent African mtDNAs (not shown). See also Figure S1 and the Supplemental Experimental Procedures.

modern human individuals from Italy, Germany, Belgium, France, Czech Republic, and Romania, spanning in age from 35 to 7 ka (Figure 1; Table S1). Hybridization capture of mtDNA in combination with high-throughput sequencing technologies [15] allowed us to evaluate typical DNA damage patterns and average fragment length [16] as criteria for authentication of ancient DNA (Supplemental Experimental Procedures). Both features were taken into account in an iterative probabilistic approach [17] that jointly estimates present-day human contamination and reconstructs mtDNA sequences (Table S2). Combining 311 modern and 66 ancient dated worldwide mtDNA genomes (both new and from the literature; Table S3), we used Bayesian phylogenetic methods [18] to estimate the mutation rate and hg coalescence times. Further, we combined our 35 new mtDNA genomes with 20 previously published ancient European mtDNAs for a total of 55 pre-Neolithic sequences (Table S4) and explicitly tested scenarios of the early population history of Europe using coalescent demographic modeling paired with approximate Bayesian computation (ABC) [19] (Supplemental Experimental Procedures).

Hg assignment of the authenticated mtDNAs confirmed that the vast majority of Late Pleistocene and Early Holocene individuals belonged to the U lineage, which is a subgroup of the N clade [20] (Figures 2 and S1). We also found a basal U lineage that had no derived position leading to known sub-hgs in a 33,000-year-old Romanian individual. Surprisingly, three hunter-gatherers from Belgium and France dating to between 35 and 28 ka carried mtDNA hg M, today found predominantly in Asia, Australasia, and the Americas, although it is almost absent in extant populations with European ancestry [12].

We used 66 ancient dated mtDNAs as tip calibration points in BEAST v1.8.1 [18] in combination with 311 modern worldwide mtDNA sequences to reduce the possible impact of sample biases (Table S3 and Supplemental Experimental Procedures) in estimating the mtDNA mutation rate and hg M and N divergence times. Strict and uncorrelated lognormal relaxed clocks were tested, under both a constant size and a Bayesian skyline tree prior. The four analyses returned mtDNA mutation rates (Table 1) consistent with previously published rates using similar methodology [21, 22]. The Bayesian skyline, in combination with strict rate variation among branches, performed best according to stepping-stone and path sampling methods [23] and highest effective sample size (ESS) values, giving a best estimate of the mutation rate of 2.74×10^{-8} (95% highest posterior density [HPD], $2.44–3.01 \times 10^{-8}$) mutation/site/year. This model allowed us to refine time estimates for the TMRCA of the basal non-African clades M and N of circa 49 ka (95% HPD, 54.8–43.6 ka) and 51 ka (95% HPD, 55.1–46.9 ka),

Table 1. Haplogroup Divergence Times and mtDNA Mutation Rate

			Divergence Time			Log Marginal Likelihood	
Tree Prior	Clock	Statistic	TMRCA hg M	TMRCA hg N	Clock Rate Whole mtDNA	Stepping-Stone Sampling	Path Sampling
Constant	strict	mean	58,869	57,482	2.62×10^{-8}	−48,759	−48,754
		median	58,578	57,181	2.62×10^{-8}		
		95% HPD	68,163–50,380	64,363–51,387	$2.30–2.93 \times 10^{-8}$		
		ESS	585	445	651		
Constant	relaxed	mean	58,961	58,531	2.67×10^{-8}	−48,755	−48,751
		median	58,507	58,207	2.67×10^{-8}		
		95% HPD	70,389–49,125	66,398–51,664	$2.30–3.04 \times 10^{-8}$		
		ESS	354	416	431		
Skyline	strict	mean	49,106	50,562	2.74×10^{-8}	−48,577	−48,571
		median	48,837	50,317	2.74×10^{-8}		
		95% HPD	54,780–43,598	55,138–46,892	$2.44–3.01 \times 10^{-8}$		
		ESS	741	799	863		
Skyline	relaxed	mean	48,005	50,179	2.77×10^{-8}	−48,550	−48,546
		median	47,695	50,021	2.77×10^{-8}		
		95% HPD	53,917–43,054	54,189–46,483	$2.47–3.07 \times 10^{-8}$		
		ESS	251	285	348		

The values reported are obtained in BEAST [18] using 377 worldwide mtDNAs, 66 of which come from ancient dated human remains. A Bayesian skyline tree prior in combination with strict rate variation between branches performed better than the other three tested models according to higher log marginal likelihood estimates (compared to the constant tree prior models) and effective sample size (ESS) values. HPD, highest posterior density. See also Table S3 and the Supplemental Experimental Procedures.

respectively (Table 1; Figure 2; Supplemental Experimental Procedures).

The observed mtDNA hg variation through time, including the apparent loss of hg M in Europe, suggests a genetic bottleneck that may have been influenced by climatic events (Figure 3). This period of European prehistory was accompanied by severe climatic fluctuations, such as the Last Glacial Maximum (LGM, 25 to 19.5 ka) and, at the end of the Pleistocene, the Bølling-Allerød interstadial followed by the stadial Younger Dryas—a period we refer to as the Late Glacial (14.5 to 11.5 ka) [24, 25]. These climatic changes have been proposed as a driver of the range contraction to refugia in many species [26], including modern humans, for whom there is absence of evidence of northwestern European occupation during the LGM [25, 27]. We used coalescent modeling paired with ABC [19] to test a range of explicit models of European hunter-gatherer demography (Figure S2; Table S6), using the complete set of 55 pre-Neolithic ancient mtDNA genomes (Table S4). The best-fitting model (Figure 3 and 2b in Figure S2) strongly supports maternal population continuity through the LGM, albeit as a single genetic bottleneck, before being replaced by a new incoming population at the onset of the Late Glacial 14.5 ka (model posterior probability, $P_{2b} = 0.807$). Based on the estimated parameter values of this model (Table S5), we infer that this surviving population diverged from the ancestral one around 29 ka (95% HPD, 36–25 ka), prior to the beginning of the LGM.

The new hunter-gatherer mtDNA genomes reported here approximately triple the available sequences from pre-Neolithic Europe. One novel finding, that three out of 18 European pre-LGM hunter-gatherers carry a previously undescribed basal mtDNA lineage M (Figure 1A), has important implications for the timing of the dispersal of modern humans into Eurasia. Excluding a ~40,000-year-old Romanian individual known not to have contributed notably to the modern European gene pool [28], our BEAST analyses give a TMRCA for clades M and N from 44 to 55 ka, respectively. Our estimated dates, together with the oldest accepted archeological evidence for the presence of early modern humans in Australia and Europe (both dated to at least 45 ka [13, 29]), are consistent with a model of a single, late, and therefore rapid dispersal of a source population containing both M and N hgs, which contributed all the mitochondrial diversity of present-day non-Africans (cf. [7]). Human individuals whose ancestries trace back to potential earlier expansion(s) outside Africa [30, 31] are thus unlikely to have left any surviving mtDNA descendants.

Phylogeographic inference based solely on mtDNA has limitations [2], but information from single loci can provide meaningful constraints on models of human prehistory. In particular, the fact that hg M has never previously been found in Europe is generally interpreted as an important limitation for the proposed scenarios of non-African population dispersals [4, 7]. According to the most popular model [4], an early expansion occurred before the M and N diversification with a subsequent loss of M in only the population ancestral to Europeans. Our evidence for the existence of hg M in Late Pleistocene Europe revises this scenario. It suggests that the loss of hg M may be due to population dynamics that occurred later within Europe itself. The expansion either occurred before the diversification of M and N, with subsequent migration bringing both lineages into Europe, or the dispersal was later, occurring after their TMRCAs. Contrary to recent findings [11], though similar to a previous study [32], our two TMRCAs are almost identically dated, suggesting a single major dispersal after 55 ka for all non-African populations, including Europe. The genetic evidence of pre-LGM hg M

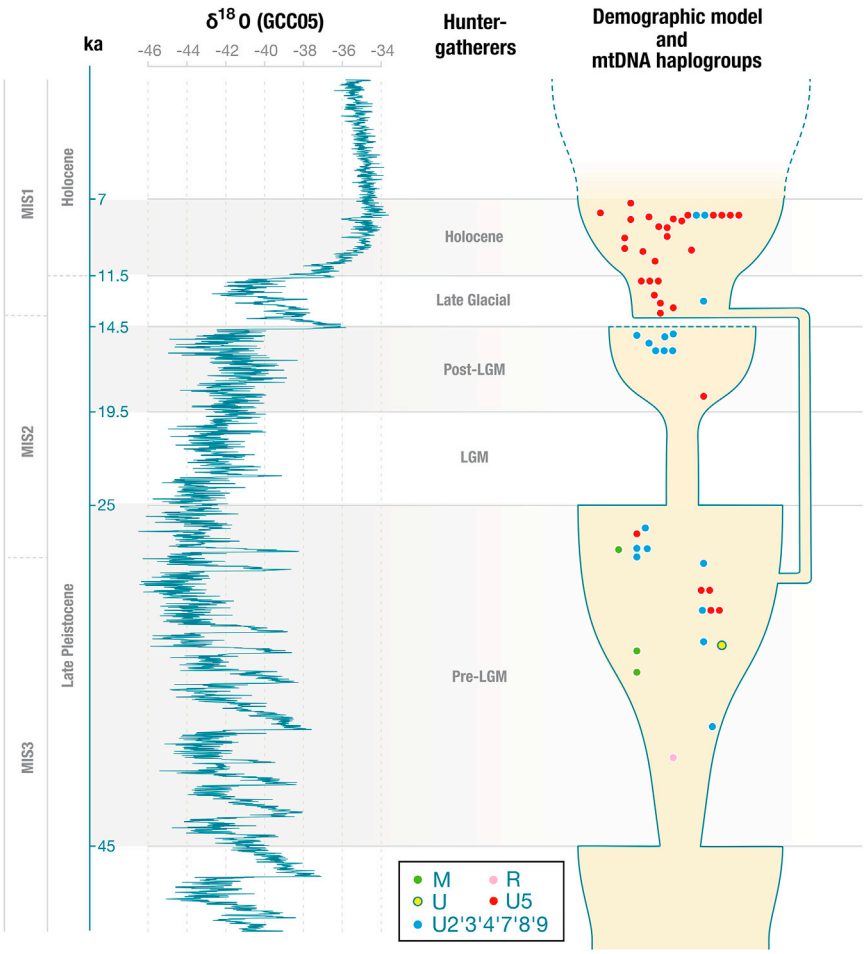


Figure 3. Late Pleistocene and Early Holocene Climatic Fluctuations and European Hunter-Gatherer Demography

On the left is the NGRIP $\delta^{18}O$ climate record, and on the right is an illustration of the best-supported demographic model (2b in Figure S2). Each colored point gives the mtDNA hg of the 55 dated pre-Neolithic individuals used in the coalescent modeling analysis. West-East site locations for each sample are approximated. See also Figure S2, Table S4, and the Supplemental Experimental Procedures.

indicates that this lineage reached Western Europe by at least 35 ka (GoyetQ116-1), either alongside the first arrival of N or later. The reconstructed phylogeny (Figure 2) with both basal N and M lineages in Late Pleistocene Europe possibly mirrors the inferred back migration into Africa, which has been suggested by the existence of hgs U6 and M1 in modern-day North Africans [33]. Therefore, the major modern human dispersal described here after 55 ka might have affected not only non-Africans, but also African populations to some extent.

The potential impact of climatic events on the demography, and thus the genetic diversity of early Europeans, has previously been difficult to quantify, but it likely had consequences for the relative components of ancient ancestry in modern-day populations [14]. Our demographic modeling reveals a dynamic history of hunter-gatherers, including a previously unknown major population shift during the Late Glacial interstadial (the Bølling-Allerød, ~14.5 ka). Under our best-fitting model (Figure 3 and 2b in Figure S2), the small initial founder population of Europe slowly grows up until ~25 ka and survives with smaller size in LGM climatic refugia (25–19.5 ka) [25] before re-expanding as the ice sheets retract (Figure 1B). Although this model supports population continuity from pre- to post-LGM, the genetic bottleneck is consistent with the apparent loss of hg M in the post-LGM. The subsequent Late Glacial period is characterized by drastic climatic fluctuations, beginning with an abrupt warming during the Bølling-Allerød interstadial and followed by a similarly drastic period of cooling during the Younger Dryas [24]. Globally, the early warming phases of the Late Glacial are strongly associated with substantial demographic changes, including extinctions of several megafaunal species [34] and the first expansion of modern humans into the Americas [35]. In European hunter-gatherers, our model best explains this period of upheaval as a replacement of the post-LGM maternal population by one from another source. Although the exact origin for this later population is unknown, the inferred demographic history (Figure 3 and 2b in Figure S2) suggests that it descended from another, separate LGM refugium. On the basis of mtDNA alone, we cannot rule out some degree of genomic continuity throughout the Late Pleistocene and early Holocene hunter-gatherer populations, and thus into present-day Europeans [14]. For this reason, we interpret our model as capturing the maternal signal of a major population shift, rather than a complete replacement. Ancient nuclear DNA data and additional geographically and temporally distributed specimens may provide a more comprehensive picture, possibly identifying the source and ancestry of these later incoming hunter-gatherers.

In conclusion, the large dataset presented here allowed us to provide a late upper bound on the major dispersal of all non-Africans and to uncover unexpected population dynamics of European hunter-gatherers. The Late Glacial event that we identify here is the oldest in an accumulating list of major European population turnovers revealed by ancient mtDNA [20].

ACCESSION NUMBERS

The accession numbers for the 35 mtDNA genome sequences reported in this paper are GenBank: KU534947–KU534981.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.01.037.

AUTHOR CONTRIBUTIONS

D.G.D., H.R., C.C., F.V., C.T., M.F., M.M., M.B., M.L., E.G., G.C. I.C., C.B., D.F., M.G., J.v.d.P., R.C., B.G., A.R., K.W., H.B., D.G., J.S., D.C., P.S., K.H., and N.J.C. provided archeological material and related information. C.P., A.M., A.F., and C.W. performed laboratory work. C.P., G.R., W.H., A.P., and J.K. analyzed genetic data. C.P., W.H., A.P., and J.K. wrote the manuscript with input from all co-authors.

ACKNOWLEDGMENTS

We are grateful for comments from Stephan Schiffels, Hugo Reyes-Centeno, David Reich, Maria Spyrou, Henrike Heyne, Heidi Colleran, and members of the Department of Archaeogenetics of the Max Planck Institute for the Science of Human History, as well as the three anonymous reviewers. We thank Pontus Skoglund, Qiaomei Fu, Viviane Slon, and Eppie Ruth-Jones for access to unpublished data, Martyna Molak, Alexander Peltzer, Marek Dynowski, and Judith Beier for technical support, and Annette Günzel for graphical support. We further thank the Soprintendenza Archeologia della Puglia, which authorized and supported the excavations at Grotta Paglicci, and Professor A. Palma di Cesnola for fieldwork and studies over the years. Part of this work was performed on the computational resource bwGRiD Cluster Tübingen funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC. J.K. and C.P. were supported by the Baden Württemberg Foundation, J.K. and A.M. by the DFG grant KR 4015/1-1, and K.H. by the European Research Council (ERC StG 283503). The Goyet project led by H.R. was funded by the Wenner-Gren Foundation (grant no. 7837), the College of Social and Behavioral Sciences of CSUN, and the RBINS.

Received: December 17, 2015
Revised: January 13, 2016
Accepted: January 14, 2016
Published: February 4, 2016; corrected online February 22, 2016

REFERENCES

1. Scally, A., and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. Nat. Rev. Genet. *13*, 745–753.
2. Groucutt, H.S., Petraglia, M.D., Bailey, G., Scerri, E.M., Parton, A., Clark-Balzan, L., Jennings, R.P., Lewis, L., Blinkhorn, J., Drake, N.A., et al. (2015). Rethinking the dispersal of Homo sapiens out of Africa. Evol. Anthropol. *24*, 149–164.
3. Mellars, P., Gori, K.C., Carr, M., Soares, P.A., and Richards, M.B. (2013). Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. Proc. Natl. Acad. Sci. USA *110*, 10699–10704.
4. Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science *308*, 1034–1036.
5. Oppenheimer, S. (2012). A single southern exit of modern humans from Africa: before or after Toba? Quat. Int. *258*, 88–99.
6. Lahr, M.M., and Foley, R.A. (1998). Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. Am. J. Phys. Anthropol. (*Suppl 27*), 137–176.
7. Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C., and Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. BMC Genet. *2*, 13.
8. Reyes-Centeno, H., Ghirotto, S., Détroit, F., Grimaud-Hervé, D., Barbujani, G., and Harvati, K. (2014). Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. Proc. Natl. Acad. Sci. USA *111*, 7248–7253.
9. Armitage, S.J., Jasim, S.A., Marks, A.E., Parker, A.G., Usik, V.I., and Uerpmann, H.P. (2011). The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia. Science *331*, 453–456.
10. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, E386–E394.
11. Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am. J. Hum. Genet. *90*, 675–684.
12. Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. Investig. Genet. *6*, 3.
13. Benazzi, S., Douka, K., Fornai, C., Bauer, C.C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., et al. (2011). Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. Nature *479*, 525–528.
14. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.
15. Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS ONE *5*, e14004.
16. Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., et al. (2009). Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science *325*, 318–321.
17. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.
18. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214.
19. Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025–2035.
20. Brandt, G., Szécsényi-Nagy, A., Roth, C., Alt, K.W., and Haak, W. (2015). Human paleogenetics of Europe–the known knowns and the known unknowns. J. Hum. Evol. *79*, 73–92.
21. Fu, Q., Mittnik, A., Johnson, P.L., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559.
22. Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L.A., Warmuth, V., Ruiz-Linares, A., Manica, A., and Balloux, F. (2014). Improved calibration of the human mitochondrial clock using ancient genomes. Mol. Biol. Evol. *31*, 2780–2792.
23. Baele, G., Lemey, P., and Vansteelandt, S. (2013). Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. BMC Bioinformatics *14*, 85.
24. Heiri, O., Brooks, S.J., Renssen, H., Bedford, A., Hazekamp, M., Ilyashuk, B., Jeffers, E.S., Lang, B., Kirilova, E., Kuiper, S., et al. (2014). Validation of climate model-inferred regional temperature change for late-glacial Europe. Nat. Commun. *5*, 4914.
25. Gamble, C., Davies, W., Pettitt, P., and Richards, M. (2004). Climate change and evolving human diversity in Europe during the last glacial. Philos. Trans. R. Soc. Lond. B Biol. Sci. *359*, 243–253, discussion 253–244.
26. Stewart, J.R., Lister, A.M., Barnes, I., and Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. Proc. Biol. Sci. *277*, 661–671.
27. Stewart, J.R., and Stringer, C.B. (2012). Human evolution out of Africa: the role of refugia and climate change. Science *335*, 1317–1321.
28. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216–219.

29. Bowler, J.M., Johnston, H., Olley, J.M., Prescott, J.R., Roberts, R.G., Shawcross, W., and Spooner, N.A. (2003). New ages for human occupation and climatic change at Lake Mungo, Australia. Nature *421*, 837–840.

30. Liu, W., Martinón-Torres, M., Cai, Y.J., Xing, S., Tong, H.W., Pei, S.W., Sier, M.J., Wu, X.H., Edwards, R.L., Cheng, H., et al. (2015). The earliest unequivocally modern humans in southern China. Nature *526*, 696–699.

31. Tassi, F., Ghirotto, S., Mezzavilla, M., Vilaça, S.T., De Santi, L., and Barbujani, G. (2015). Early modern human dispersal from Africa: genomic evidence for multiple waves of migration. Investig. Genet. *6*, 13.

32. Forster, P., Torroni, A., Renfrew, C., and Röhl, A. (2001). Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol. Biol. Evol. *18*, 1864–1881.

33. Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D.M., Dugoujon, J.M., et al. (2006). The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science *314*, 1767–1770.

34. Cooper, A., Turney, C., Hughen, K.A., Brook, B.W., McDonald, H.G., and Bradshaw, C.J. (2015). PALEOECOLOGY. Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. Science *349*, 602–606.

35. Meltzer, D.J. (2009). First Peoples in a New World: Colonizing Ice Age America (University of California Press).

Current Biology, Volume *26*

Supplemental Information

Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe

Cosimo Posth, Gabriel Renaud, Alissa Mittnik, Dorothée G. Drucker, Hélène Rougier, Christophe Cupillard, Frédérique Valentin, Corinne Thevenet, Anja Furtwängler, Christoph Wißing, Michael Francken, Maria Malina, Michael Bolus, Martina Lari, Elena Gigli, Giulia Capecchi, Isabelle Crevecoeur, Cédric Beauval, Damien Flas, Mietje Germonpré, Johannes van der Plicht, Richard Cottiaux, Bernard Gély, Annamaria Ronchitelli, Kurt Wehrberger, Dan Grigorescu, Jiří Svoboda, Patrick Semal, David Caramelli, Hervé Bocherens, Katerina Harvati, Nicholas J. Conard, Wolfgang Haak, Adam Powell, and Johannes Krause

Supplemental Data

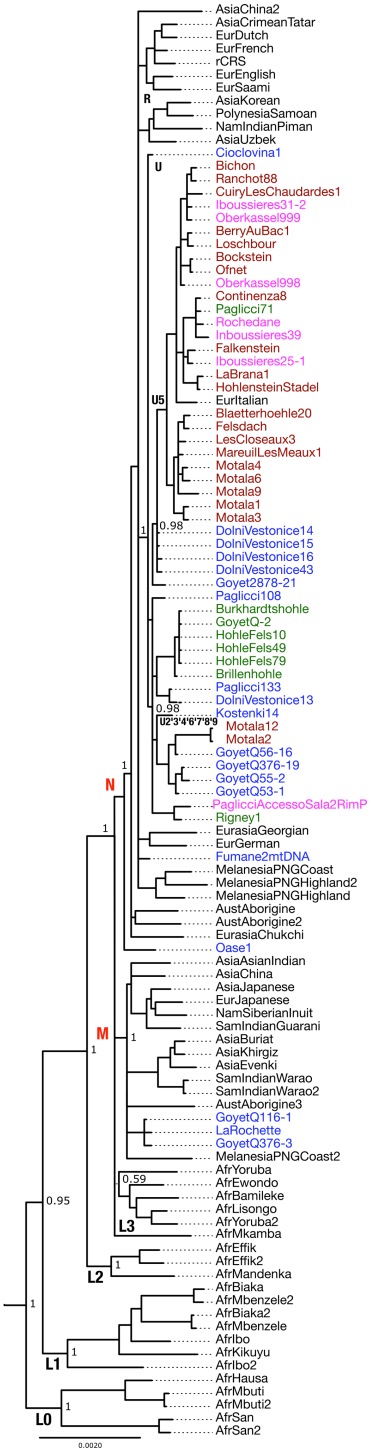


Figure S1.

Bayesian phylogenetic tree of 54 present-day human and 55 Pre-Neolithic mtDNA genomes (related to Figure 2). The tree shows the same general topology of the Maximum Parsimony tree (Figure 2) with posterior probability of 1 at both M and N nodes. Pre-LGM samples are shown in blue, Post-LGM in green, Late Glacial in magenta, Holocene hunter-gatherers in red. Oase1 represents a pre-N lineage.

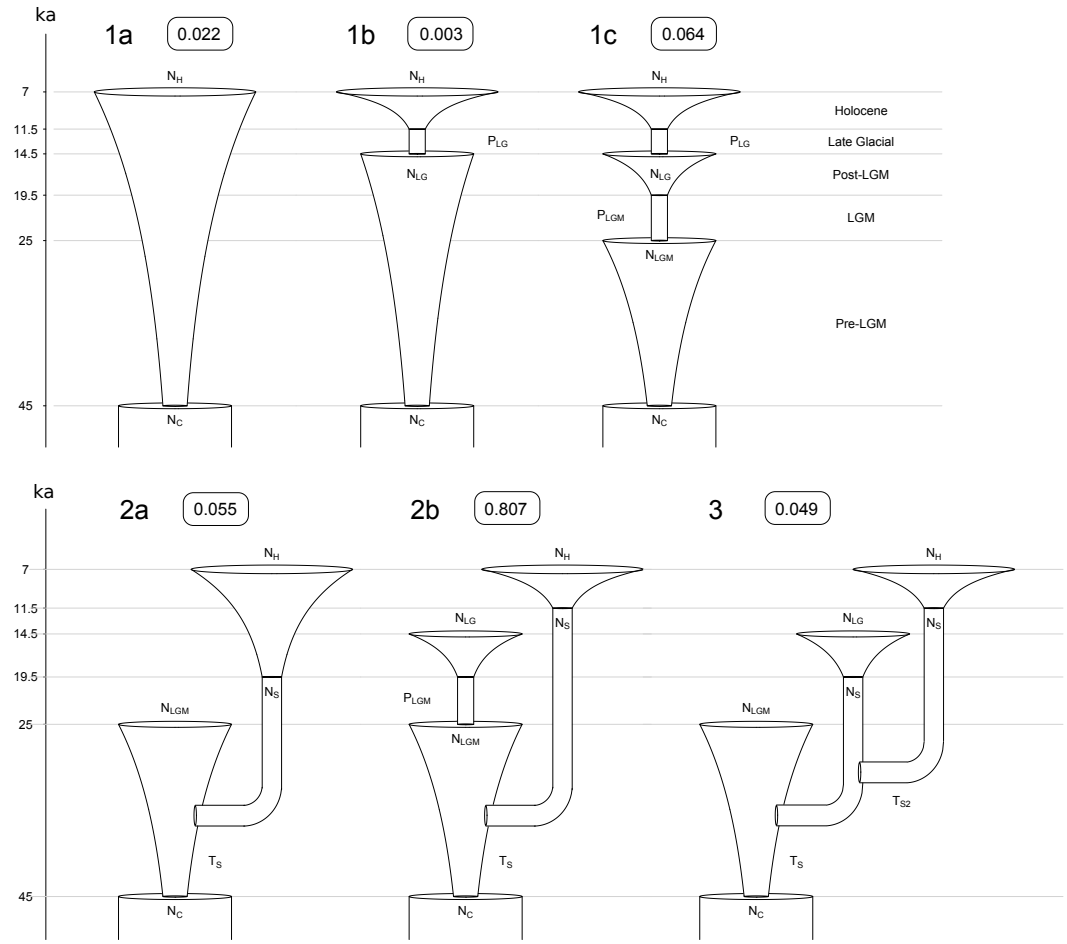


Figure S2.

Schematic of the demographic models used for coalescent simulations (related to Figure 3). The six named demographic models described in the Supplemental Experimental Procedures, with estimated model posterior probabilities given in the box in each subplot. Model 2b is identified as by far the best-fitting model, suggesting that the European hunter-gatherer population survived through an LGM bottleneck but was then largely replaced during the Late Glacial around 14.5 ka.

Table S1.

Archeological sample information (related to Figure 1). For each sample genetically analyzed, the origin, anatomical element, radiocarbon date and culture are reported (see "Archeological site information" in Supplemental Experimental Procedures). * Direct: radiocarbon date of the individual genetically analyzed; Layer: radiocarbon date of the associated stratigraphic layer.

Sample ID	Inventory name	Site (location)	Country	Geographical coordinates	Element	Date 14C (years BP) *	Source dating	Associated Culture
BerryAuBac1	BRB-1 (BVT 353)	Le Vieux Tordoir (Berry-au-Bac)	France	N 49° 24' E 03° 54'	Radius	6325±35(SacA-5455) direct	This study	Mesolithic
Bockstein	Varia I.1	Bockstein Höhle	Germany	N 48° 33' E 10° 09'	Tooth	7350±70 (UtC-7887) - 7460±60 (UtC-6796) layer	[S1]	Mesolithic
Brillenhohle	OSUT 5827	Brillenhöhle	Germany	N 48° 24' E 09° 46'	Cranium	12470±65 (OxA11054)- 12535±50 (OxA23414) direct	[S2]	Magdalenian
Burkhardtshohle	Tü 33/32 420	Burkhardtshöhle	Germany	N 48° 32' E 09° 35'	Cranium	12450±110 (ETH-7613) direct	[S3]	Magdalenian
Cioclovina1	Cioclovina1	Peştera Cioclovina Uscata˘	Romania	N 45° 35' E 23° 08'	Cranium	28510±170 (OxA-15527) direct	[S4]	Unassigned
CuiryLesChaudardes1	CRC-1	Les Fontinettes (Cuiry-lès-Chaudardes)	France	N 49° 23' E 03° 46'	Tibia	7400±60 (GrA-28268) direct	This study	Mesolithic
DolniVestonice16	DV 16	Dolní Věstonice	Czech Republic	N 48° 53' E 16° 39'	Femur	25740±210 (GrN15277) - 25570±280 (GrN15276) layer	[S5]	Gravettian
DolniVestonice43	DV 43	Dolní Věstonice	Czech Republic	N 48° 53' E 16° 39'	Femur	25740±210 (GrN15277) - 25570±280 (GrN15276) layer	[S5]	Gravettian
Falkenstein	OSUT 2306	Falkensteiner Höhle	Germany	N 48° 06' E 09° 04'	Fibula	8185±80 (ETH-7615) direct	[S6]	Mesolithic
Felsdach	F-I 28	Felsdach Inzigofen	Germany	N 48° 04' E 09° 10'	Tooth	7770 ± 120 (B-933) layer	[S7]	Mesolithic
Goyet2878-21	2878-21	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Clavicle	22360±110 (GrA-62455) direct	This study	Gravettian
GoyetQ-2	Q-2	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Humerus	12650±50 (GrA-46168) direct	This study	Magdalenian
GoyetQ116-1	Q116-1	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Humerus	30880 +170, -160 (GrA-46175) direct	This study	Aurignacian
GoyetQ376-19	Q376-19	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Humerus	23260 +110, -100 (GrA-54026) direct	This study	Gravettian
GoyetQ376-3	Q376-3	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Humerus	29370 +180, -170 (GrA-60034) direct	This study	Aurignacian
GoyetQ53-1	Q53-1	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Fibula	23920±100 (GrA-46169) direct	This study	Gravettian
GoyetQ55-2	Q55-2	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Fibula	23270 +120, -110 (GrA-54031) direct	This study	Gravettian
GoyetQ56-16	Q56-16	Troisième caverne (Goyet)	Belgium	N 50° 26' E 05° 00'	Fibula	22100±100 (GrA-59991) direct	This study	Gravettian
HohleFels10	HF 10 Ic 405	Hohle Fels	Germany	N 48° 22' E 09° 45'	Femur	12770±220 (H5312-4907) - 13085±95 (H5119-4601) layer	[S8]	Magdalenian
HohleFels49	HF 49 Ib1 66	Hohle Fels	Germany	N 48° 22' E 09° 45'	Femur	12770±220 (H5312-4907) - 13085±95 (H5119-4601) layer	[S8]	Magdalenian
HohleFels79	HF 79 IIb 876	Hohle Fels	Germany	N 48° 22' E 09° 45'	Cranium	12490±70 (MAMS-25564) direct	This study	Magdalenian
HohlensteinStadel	OSUT 5830 a-c	Hohlenstein-Stadel	Germany	N 48° 32' E 10° 10'	Vertebra	7835±80 (ETH-5732) direct	[S9]	Mesolithic
Iboussieres25-1	IBS-6 (25-1)	Aven des Iboussières (Malataverne)	France	N 44° 29' E 04° 45'	Coxal	10140±50 (GrA-43700) layer	This study	Epipaleolithic
Iboussieres31-2	IBS-8 (31-2)	Aven des Iboussières (Malataverne)	France	N 44° 29' E 04° 45'	Vertebra	10140±50 (GrA-43700) layer	This study	Epipaleolithic
Iboussieres39	IBS-9 (39)	Aven des Iboussières (Malataverne)	France	N 44° 29' E 04° 45'	Femur	10140±50 (GrA-43700) direct	This study	Epipaleolithic
LaRochette	OSUT 7074	La Rochette (saint Léon-sur-Vézère)	France	N 45° 00' E 01° 05'	Ulna	23630±130 (OxA11053)- 23400±110 (OxA23413) direct	[S2]	Gravettian
LesCloseaux3	LCX-13 (s3 b1)	Les Closeaux (Rueil-Malmaison)	France	N 48° 52' E 02° 11'	Femur	8870±130 (OxA-7109(Ly-612)) direct	[S10]	Mesolithic
MareuilLesMeaux1	MLM-1 (burial 11)	Les Vignolles (Mareuil-lès-Meaux)	France	N 48° 55' E 02° 51'	Femur	8320±90 (GrN-27225) direct	[S10]	Mesolithic
Ofnet	OSUT 4043	Große Ofnet Höhle	Germany	N 48° 49' E 10° 27'	Tooth	7480±80(OxA1574) layer	[S9]	Mesolithic
Paglicci108	PA108 (21B-1)	Grotta Paglicci	Italy	N 41° 40' E 15° 34'	Phalanx	23470±370 (F-52) layer	[S11]	Gravettian
Paglicci133	PA133 (23C-2)	Grotta Paglicci	Italy	N 41° 40' E 15° 34'	Tooth	28100±400 (UTC-1414) - 29300±600(UtC-1789) layer	[S12]	Gravettian
Paglicci71	PA71 (8A)	Grotta Paglicci	Italy	N 41° 40' E 15° 34'	Patella	15460±220 (F-66) layer	[S13]	Epigravettian
Ranchot88	RAN88D9AOC154	L'abri des "Cabônes"	France	N 47° 09' E 05° 43'	Cranium	8985±40 (GrA-38019) direct	This study	Mesolithic
Rigney1	Rigney 1	Rigney 1 (Rigney)	France	N 47° 23' E 06° 10'	Mandible	12930 ± 55 Ly-6515 (OxA) direct	[S14]	Magdalenian
Rochedane	RochedaneA' 69	Rochedane (Villars-sous-Dampjoux)	France	N 47° 20' E 06° 45'	Mandible	11120±50 (GrA-41739) direct	[S14]	Epipaleolithic

Table S2.

Summarized mtDNA capture results (related to Figure 1 and Figure 2). Schmutzi [S15] output comprising mtDNA average coverage, average length, deamination at molecule termini, contamination estimate. MtDNA haplogroup was assigned with Haplogrep. The last column provides the number of unassigned positions (Ns) over the complete mtDNA (16569bp).

Sample ID	Library ID	mtDNA (fold)	Insert size (bp)	5' deamination (%)	3' deamination (%)	contamination estimate (%), first iteration (low - high)	contamination estimate (%), final iteration (low - high)	mtDNA haplogroup	Ns (bp)
BerryAuBac1	GA261	106.8	55.4	59.5	57.6	5.5 (5 - 6)	1 (0 - 2)	U5b1a	1
Bockstein	GA89	267.0	50.4	32.4	32.5	6 (5.5 - 6.5)	1 (1 - 1)	U5b1d1	1
Brillenhohle	GA79	18.6	54.4	23.6	22.2	29.5 (27.5 - 31.5)	4 (3 - 5)	U8a	64
Burkhardtshohle	Burk	45.0	65.5	30.6	30.8	29.5 (28.5 - 30.5)	13 (12 - 14)	U8a	1
Cioclovina1	Cioclovina	19.2	64.6	26.3	23.5	28 (26 - 30)	3 (2 - 4)	U	2
CuiryLesChaudardes1	GA113	16.9	54.3	60.4	55.8	4.5 (3.5 - 5.5)	1 (0 - 2)	U5b1b	2
DolniVestonice16	DV16	14.5	80.9	27.3	29.0	26 (23.5 - 28.5)	1 (0 - 2)	U5	30
DolniVestonice43	DV43	46.3	57.0	46.0	44.4	12 (11 - 13)	1 (0 - 2)	U5	1
Falkenstein	FL	599.7	56.1	37.0	37.1	4 (3.5 - 4.5)	1 (1 - 1)	U5b2a	1
Felsdach	Fels	55.2	50.6	33.7	34.2	28 (27.5 - 28.5)	13 (12 - 14)	U5a2c	17
Goyet2878-21	GA232	21.1	54.4	31.8	32.1	21.5 (20 - 23)	3 (2 - 4)	U5	12
GoyetQ-2	GA231	406.0	49.8	38.0	36.7	10.5 (10 - 11)	1 (1 - 1)	U8a	0
GoyetQ116-1	GA63	56.0	54.3	29.2	28.9	14 (13 - 15)	2 (2 - 2)	M	3
GoyetQ376-19	Q376-19	42.8	52.2	37.3	35.1	15.5 (14.5 - 16.5)	4 (3 - 5)	U2	3
GoyetQ376-3	Q376-3	45.7	60.6	24.8	23.2	31 (30 - 32)	12 (11 - 13)	M	1
GoyetQ53-1	Q53-1	48.3	60.6	31.5	31.9	25.5 (24.5 - 26.5)	12 (11 - 13)	U2	2
GoyetQ55-2	Q55-2	17.0	76.8	22.2	23.6	42 (40 - 44)	10 (8 - 12)	U2	31
GoyetQ56-16	Q56-16	45.1	72.6	22.2	24.5	36.5 (35.5 - 37.5)	4 (4 - 4)	U2	1
HohleFels10	GA81	106.2	52.0	40.0	41.1	7.5 (7 - 8)	1 (0 - 2)	U8a	1
HohleFels49	GA82	364.2	64.9	35.9	36.1	11 (10.5 - 11.5)	1 (1 - 1)	U8a	0
HohleFels79	GA90	42.2	52.2	33.7	32.7	22 (21 - 23)	3 (3 - 3)	U8a	1
HohlensteinStadel	VE	33.5	59.4	25.8	20.9	26.5 (25.5 - 27.5)	16 (15 - 17)	U5b2c1	1
Iboussieres25-1	MA121	27.1	73.6	40.7	41.7	11 (9.5 - 12.5)	4 (3 - 5)	U5b2a	2
Iboussieres31-2	MA123	13.6	68.0	44.2	42.7	19 (17 - 21)	4 (3 - 5)	U5b1	4
Iboussieres39	GA77	30.9	49.7	61.1	59.7	9.5 (8.5 - 10.5)	3 (2 - 4)	U5b2b	1
LaRochette	LaRochette	40.1	55.7	31.9	31.6	24 (23 - 25)	14 (13 - 15)	M	2
LesCloseaux3	LCX	19.0	62.9	39.4	37.2	28 (26.5 - 29.5)	12 (10 - 14)	U5a2	3
MareuilLesMeaux1	MLM	18.9	63.8	44.7	40.0	31 (29.5 - 32.5)	16 (14 - 18)	U5a2	5
Ofnet	GA93	1859.3	53.4	31.1	30.4	9.5 (9 - 10)	1 (1 - 1)	U5b1d1	0
Paglicci108	B1	19.5	54.4	45.0	34.2	21.5 (20 - 23)	6 (5 - 7)	U2'3'4'7'8'9	4
Paglicci133	C2	27.7	51.4	54.4	53.3	17.5 (16.5 - 18.5)	7 (6 - 8)	U8c	6
Paglicci71	FA	11.6	52.8	60.0	60.6	10.5 (9 - 12)	5 (4 - 6)	U5b2b	23
Ranchot88	GA262	86.3	47.7	53.7	49.4	0 (0 - 0.5)	4 (3 - 5)	U5b1	1
Rigney1	Rigney1	41.4	62.4	32.8	30.6	26 (25 - 27)	10 (9 - 11)	U2'3'4'7'8'9	1
Rochedane	GA127	104.4	47.8	42.7	43.4	12 (11.5 - 12.5)	1 (0 - 2)	U5b2b	1

Table S3.

Dated ancient mtDNA genomes (related to Table 1). 66 ancient radiocarbon dated samples (except the archeologically dated Stuttgart sample) used as calibration points in BEAST [S16] to estimate the mtDNA molecular clock. *^{14}C dates newly reported in this study were calibrated using Reimer et al. [S17] calibration curve and reported in calibrated years before present (cal BP).

Sample ID	Country	Date interval (cal BP)*	Data source
Ajvide52	Sweden	4600-4900	[S18]
Ajvide58	Sweden	4600-4900	[S18]
Ajvide70	Sweden	4600-4900	[S18]
BerryAuBac1	France	7169-7319	This study
BLA10*	Germany	5355-5481	[S19]
BLA11*	Germany	5862-5982	[S19]
BLA13*	Germany	5411-5615	[S19]
BLA14*	Germany	5589-5617	[S19]
BLA20*	Germany	10594-10710	[S19]
BLA7*	Germany	5646-5686	[S19]
Bockstein	Germany	8016 - 8329	This study
Boshan11	China	8040-8320	[S20]
BRA1	Spain	7690-7940	[S21]
Brillenhohle	Germany	14440-15120	This study
Burkhardtshohle	Germany	14150 – 15080	This study
Cioclovina1	Romania	32519-33905	This study
CuiryLesChaudardes1	France	8050-8360	This study
DolniVestonice13	Czech Republic	31070-31240	[S20]
DolniVestonice14	Czech Republic	31070-31240	[S20]
DolniVestonice16	Czech Republic	29386 - 30567	This study
Falkenstein	Germany	8993 - 9409	This study
Felsdach	Germany	8380-8980	This study
Fumane2	Italy	38500-41110	[S22]
Gökhem2	Sweden	4750-5050	[S18]
Gökhem5	Sweden	4750-5050	[S18]
Gökhem7	Sweden	4750-5050	[S18]
Goyet2878-21	Belgium	26269 - 27055	This study
GoyetQ-2	Belgium	14780-15230	This study
GoyetQ116-1	Belgium	34430-35160	This study
GoyetQ376-19	Belgium	27310-27720	This study
GoyetQ376-3	Belgium	33140-33940	This study
GoyetQ53-1	Belgium	27720-28230	This study
GoyetQ55-2	Belgium	27310-27730	This study
GoyetQ56-16	Belgium	26040-26600	This study
HohleFels49	Germany	15568 - 16250	This study
HohleFels79	Germany	14270-15070	This study
HohlensteinStadel	Germany	8446 - 8809	This study
Iboussieres39	France	11600-12040	This study
Iceman	Italy	5100-5350	[S23]
Ire8	Sweden	4150-5100	[S18]
Kostenki14	Russia	37320-38650	[S24]
LaRochette	France	27400-27784	This study
LesCloseaux3	France	9580-10230	This study
Loschbour	Luxembourg	7927-8181	[S25]
MA-1	Russia	23891-24423	[S26]
MareuilLesMeaux1	France	9080-9500	This study
Motala1	Sweden	7530-8375	[S25]
Motala12	Sweden	7530-8375	[S25]
Motala2	Sweden	7530-8375	[S25]
Motala3	Sweden	7530-8375	[S25]
Motala4	Sweden	7530-8375	[S25]
Motala6	Sweden	7530-8375	[S25]
Motala9	Sweden	7530-8375	[S25]
Oase1	Romania	37000-42000	[S27]
Oberkassel998	Germany	13870-14170	[S20]
Oberkassel999	Germany	13290-13570	[S20]
Ofnet	Germany	8159 - 8424	This study
Paglicci108	Italy	27831 – 28961	This study
Paglicci71	Italy	18197-18973	This study
Ranchot88	France	9933-10235	This study
Rigney1	France	15240-15690	This study
Rochedane	France	12830-13090	This study
Saqqaq	Greenland	3600-4170	[S28]
Stuttgart	Germany	6800-7150	[S25]
Tianyuan1301	China	38830-40120	[S20]
Ust'Ishim	Russia	43210-46880	[S29]

Table S4.

European pre-Neolithic dated mtDNA genomes (related to Figure 1 and Figure 3). 55 complete or almost complete mtDNA sequences of Late Pleistocene and early Holocene hunter-gatherers (sorted chronologically) used for coalescent demographic modeling. The last column provides average values of ^{14}C-dates reported in calibrated years before present (cal BP). ^Radiocarbon date not available in the study.

Sample ID	Country	Map sites (Figure 1)	Time period	Haplogroup	Publication	Date (cal BP)
Fumane2	Italy	Fumane	preLGM	R	[S22]	39805
Kostenki14	Russia	Kostenki	preLGM	U2	[S24]	37985
GoyetQ116-1	Belgium	Goyet	preLGM	M	This study	34795
GoyetQ376-3	Belgium	Goyet	preLGM	M	This study	33540
Cioclovina1	Romania	Cioclovina	preLGM	U	This study	33212
Paglicci133	Italy	Paglicci	preLGM	U8c	This study	33000
DolniVestonice13	Czech Republic	Dolni Vestonice	preLGM	U8	[S20]	31155
DolniVestonice14	Czech Republic	Dolni Vestonice	preLGM	U5	[S20]	31155
DolniVestonice15	Czech Republic	Dolni Vestonice	preLGM	U5	[S20]	31155
DolniVestonice16	Czech Republic	Dolni Vestonice	preLGM	U5	This study	29977
DolniVestonice43	Czech Republic	Dolni Vestonice	preLGM	U5	This study	29977
Paglicci108	Italy	Paglicci	preLGM	U2'3'4'7'8'9	This study	28396
GoyetQ53-1	Belgium	Goyet	preLGM	U2	This study	27975
LaRochette	France	La Rochette	preLGM	M	This study	27592
GoyetQ55-2	Belgium	Goyet	preLGM	U2	This study	27520
GoyetQ376-19	Belgium	Goyet	preLGM	U2	This study	27515
Goyet2878-21	Belgium	Goyet	preLGM	U5	This study	26662
GoyetQ56-16	Belgium	Goyet	preLGM	U2	This study	26320
Paglicci71	Italy	Paglicci	postLGM	U5b2b	This study	18585
HohleFels79	Germany	Swabian Jura	postLGM	U8a	This study	15909
HohleFels10	Germany	Swabian Jura	postLGM	U8a	This study	15470
HohleFels49	Germany	Swabian Jura	postLGM	U8a	This study	15470
Rigney1	France	French Jura	postLGM	U2'3'4'7'8'9	This study	15465
GoyetQ-2	Belgium	Goyet	postLGM	U8a	This study	15005
Brillenhohle	Germany	Swabian Jura	postLGM	U8a	This study	14780
Burkhardtshohle	Germany	Swabian Jura	postLGM	U8a	This study	14615
Oberkassel998	Germany	Oberkassel	LateGlacial	U5b1	[S20]	14020
Bichon	Switzerland	French Jura	LateGlacial	U5b1h	[S30]	13700
Oberkassel999	Germany	Oberkassel	LateGlacial	U5b1	[S20]	13430
Paglicci Accesso sala 2 Rim P	Italy	Paglicci	LateGlacial	U2'3'4'7'8'9	[S20]	13000^
Rochedane	France	French Jura	LateGlacial	U5b2b	This study	12960
Iboussieres39	France	Aven des Iboussières	LateGlacial	U5b2b	This study	11820
Iboussieres25-1	France	Aven des Iboussières	LateGlacial	U5b2a	This study	11820
Iboussieres31-2	France	Aven des Iboussières	LateGlacial	U5b1	This study	11820
BLA20*	Germany	Blätterhöhle	Holocene	U5a2c3*	[S19]	10652
Ranchot88	France	French Jura	Holocene	U5b1	This study	10084
Continenza 8	Italy	Continenza	Holocene	U5b2b1	[S20]	10000^
LesCloseaux3	France	Paris Basin	Holocene	U5a2	This study	9905
MareuilLesMeaux1	France	Paris Basin	Holocene	U5a2	This study	9290
Falkenstein	Germany	Swabian Jura	Holocene	U5b2a	This study	9201
Felsdach	Germany	Swabian Jura	Holocene	U5a2c	This study	8680
HohlensteinStadel	Germany	Swabian Jura	Holocene	U5b2c1	This study	8628
Ofnet	Germany	Swabian Jura	Holocene	U5b1d1	This study	8292
CuiryLesChaudardes1	France	Paris Basin	Holocene	U5b1b	This study	8205
Bockstein	Germany	Swabian Jura	Holocene	U5b1d1	This study	8173
Loschbour	Luxembourg	Loschbour	Holocene	U5b1a	[S20]	8054
Motala1	Sweden	Motala	Holocene	U5a1	[S25]	7953
Motala12	Sweden	Motala	Holocene	U2e1	[S25]	7953
Motala2	Sweden	Motala	Holocene	U2e1	[S25]	7953
Motala3	Sweden	Motala	Holocene	U5a1	[S25]	7953
Motala4	Sweden	Motala	Holocene	U5a2d	[S25]	7953
Motala6	Sweden	Motala	Holocene	U5a2d	[S25]	7953
Motala9	Sweden	Motala	Holocene	U5a2	[S25]	7953
BRA1	Spain	La Braña	Holocene	U5b2c1	[S21]	7815
BerryAuBac1	France	Paris Basin	Holocene	U5b1a	This study	7244

Table S5.

ABC estimated parameter posterior distributions for demographic model 2b in Figure S2 (related to Figure 3). Prior distributions are given, along with modes, medians and 95% highest posterior density (HPD) intervals for each parameter. Tolerance proportion $F\delta = 0.25\%$, i.e. 12500 retained from 5 million simulations. Prior distributions for the two additional parameters used in other models are given in gray.

	Prior (uniform)				HPD	
Parameter	**Minimum**	**Maximum**	**Mode**	**Median**	**2.5%**	**97.5%**
$\log(N_C)$	1	3.699	1.963	1.867	1.057	3.113
T_S	45 ka	25 ka	29394	29610	36420	25000
$\log(N_S)$	2	4.699	3.099	3.148	2.762	3.767
$\log(N_{LGM})$	2	4.699	3.329	3.202	2.248	4.036
P_{LGM}	0	1	0.169	0.305	0.074	0.909
$\log(N_{LG})$	2.699	4.699	2.802	3.048	2.713	4.407
$\log(N_H)$	2.699	5	4.265	4.062	2.871	4.928
T_{S2}	T_S	19.5 ka	-	-	-	-
P_{LG}	0	1	-	-	-	-

Table S6.

Summary statistics (related to Figure 3). Observed statistic values, ranked by median Kruskal-Wallis p-value per block for ABC model choice. Sample groups: 0 = Holocene; 1 = Late Glacial; 2 = Post-LGM; and 3 = Pre-LGM. h_i = number of haplotypes; H_i = haplotype diversity; π_i = mean number of pairwise differences; s_i = number of polymorphic sites; D_i = Tajima's D. Within-group statistics for the Pre-LGM sample group (group 3), in gray are excluded from the model choice procedure, as described in the Supplemental Experimental Procedures.

Summary statistic	Observed value	Kruskal Wallis test p-value
h_0	19	
h_1	7	~0
h_2	4	
h_3	15	
H_0	0.943	
H_1	0.857	~0
H_2	0.563	
H_3	0.930	
π_{0vs1}	14.211	
π_{0vs2}	19.470	
π_{0vs3}	19.104	~0
π_{1vs2}	17.714	
π_{1vs3}	19.286	
π_{2vs3}	16.680	
$F_{ST\ 0vs1}$	0.069	
$F_{ST\ 0vs2}$	0.415	
$F_{ST\ 0vs3}$	0.258	~0
$F_{ST\ 1vs2}$	0.416	
$F_{ST\ 1vs3}$	0.319	
$F_{ST\ 2vs3}$	0.323	
s_{all}	145	1E-301
π_{all}	16.776	1E-301
π_0	14.262	
π_1	12.191	5.052E-276
π_2	8.500	
π_3	14.083	
s_0	77	
s_1	36	4.999E-264
s_2	29	
s_3	54	
D_0	-1.349	
D_1	-0.977	2.771E-130
D_2	-1.271	
D_3	-0.572	
H_{all}	0.971	3.49E-128
h_{all}	45	8.310E-113
D_{all}	-1.704	1.840E-81

Supplemental Experimental Procedures

Ancient DNA

Sampling
Human specimens under investigation for ancient DNA (aDNA) (Table S1) were brought to "clean room" facilities at the Institute for Archaeological Sciences in Tübingen, cataloged (Sample ID assignment), photographed (and in presence of diagnostic morphological features also surface scanned) and stored in sample plastic bags. If not already powdered, both bones and teeth were irradiated for at least 1 hour under UV lights before starting with sample processing. Teeth were sampled by cutting each tooth transversally with a coping saw under the crown/root border in order to not damage informative enamel characteristics. Tooth dentine from inside either dental crowns or roots was powdered using a dentist drill with diamond bits rotated at low speed (below 15rpm). For bone sampling, the dentist drill was used with the same settings, initially to remove a thin layer of surface, and then to drill inside the bone. For each specimen between 30 mg and 120 mg of bone/tooth powder were sampled and used in DNA extraction.

DNA extraction and library preparation
DNA was extracted from bone and tooth powder following a published method especially designed for the retrieval of short DNA fragments typical of aDNA [S31], except for the samples DolniVestonice16 and DolniVestonice43 that were extracted with a different silica-based protocol [S32]. DNA was eluted in 100µl TET (10mM Tris-HCl, 1mM EDTA pH 8.0, 0.1% Tween20) and between 5 µl and 60 µl of DNA extract was transformed into a genetic library following a double stranded DNA protocol without any UDG treatment in order to keep damage pattern along DNA fragments [S33]. The concentration of not indexed libraries measured with qPCR varied form ~10E8 to ~10E9 copies/ul whereas extraction and library negative controls show 4 to 5 orders of magnitude lower concentration (not shown). Moreover, extraction and library positive controls (cave bear specimens) resulted in expected concentrations confirming the extraction and library preparation success. A unique combination of two indexes (6-8 base pair (bp)) was incorporated in each library molecule as sample specific DNA barcodes [S34]. Library aliquots before and after 10 cycles of indexing PCR were quantified with qPCR [S33] and the reactions success were estimated as total molecules after indexing over total molecules before indexing. Indexed libraries of each sample were subsequently amplified for different PCR cycles with AccuPrime Pfx DNA polymerase as reported in Schuenemann et al. [S35] in order to stay below the reaction plateau phase and avoid heteroduplex formation [S36]. Extraction and library negative controls were treated accordingly. Amplified products were then purified using MinElute spin columns following the manufacturer's protocol and quantified on an Agilent 2100 Bioanalyzer DNA 1000 chip.

Mitochondrial DNA enrichment and sequencing
Human mitochondrial DNA (mtDNA) was enriched through a bait-capture method described in Maricic et al. [S37]. Up to five amplified indexed libraries from different samples were pooled equimolarly to reach a total of 2000ng DNA and captured together. Hybridized molecules were eluted from streptavidin beads with 125mM NaOH and purified on MinElute spin columns. After qPCR quantification, pooled captured libraries yielded a concentration between ~10E5 (blanks) and ~10E7 copies/ul. A positive control library with known mtDNA preservation was captured in parallel to estimate the efficiency of the enrichment process. An additional amplification with AccuPrime Pfx DNA polymerase of each pool was performed as mentioned before. Enriched library pools were then quantified using an Agilent 2100 Bioanalyzer DNA 1000 chip, pooled in equal concentration and sequenced in different percentage on several HiSeq2500 via 2x100+8+8 cycles and an Illumina MiSeq run for 2x150+8+8 cycles. After sequencing, libraries yielded between 1 and 10 million reads.

DNA sequence pre-processing and mapping
Base call files (.bcl) were calculated by the Instrument Control Software's RTA, which employs Bustard as base caller [S38]. Bcl were converted in raw sequences that were demultiplexed and sorted in sample specific folders making use of the individual double barcodes. First, adaptors and index sequences were clipped off. Paired-end reads were then merged into single sequences if they overlapped by at least 10bp and choosing the base with higher sequencing quality if forward and reverse read disagreed [S39]. Only merged reads above 30bp length were mapped. Alignment of the selected sequences to the human

mitochondrial genome reference (rCRS) was performed using BWA with "–n 5" parameter [S40] in combination with a custom developed tool that takes into account the circularity of mtDNA [S41] and between ~0.2 and 50% of reads were mapped. After mapping, duplicate sequences with the same start and end positions were removed using a tool developed in-house that considers coordinates of both read termini [S41]. Reads with mapping quality below 30 were discarded with the samtools software package [S42] in order to consider only sequences with a secure placement within the mtDNA in subsequent analyses. Finally, we obtained an average coverage of mtDNA ranging between 11 and 1860-fold with an average fragment length of 49 to 80bp and deamination rate at read termini from 20 to 61% (Table S2).

Joint consensus call and contamination estimates

A probabilistic iterative method, schmutzi [S15], was used (parameters: "--logindel 1 --uselength") to jointly infer the endogenous consensus and to estimate present-day human contamination levels. Average deamination rates at the first bases of the 5' and 3' ends were computed using internal programs that are part of the software package. The program mitigates the impact of deamination on consensus calling. To identify endogenous bases, the software uses deamination rates and sequence length distributions of the fragments. Present-day human contamination estimates were performed using a non-redundant database of human mitochondrial genomes distributed with the software package. An initial contamination estimate was computed using deamination patterns. Since these contaminating DNA fragments can be longer than endogenous aDNA ones, there can be a discrepancy between a contamination estimate based on the fraction of contaminating fragments versus one based on the fraction of contaminating bases. Since the initial contamination estimate based on deamination patterns produces an estimate more consistent with the former definition, this estimate is refined in later iterations of the program to facilitate endogenous consensus calling as it relies on the latter definition. These subsequent iterations use the endogenous consensus and the aforementioned database of putative mitochondrial contaminant genomes. This method iteratively co-estimates the endogenous mitochondrial genome and the present-day human contamination rate. The predicted endogenous bases and insertions/deletions are produced with an error probability on a PHRED scale. Due to the relatively low coverage for some samples, no quality threshold was applied on the final predictions thus effectively taking the bases with the highest posterior probability. The following poly-C regions and mutational hotspot were masked on the final endogenous consensus sequence: 303-315, 515-522, 16519 [S43]. The resulting sequences were uploaded to HaploGrep [S44] to assign haplogroups and call mtDNA SNPs against the Reconstructed Sapiens Reference Sequence (RSRS) [S45] (Table S2).

Visual inspection and haplotype calls

Given the significance of the samples but also critical levels of contamination detected in some of them using schmutzi, we also inspected each assembly visually and made phylogeny informed haplotype calls for both the assumed endogenous portion of reads (the majority of damaged reads) and contaminating lineages that were identified by characteristic substitution profiles. The standardized procedure is described as follows:

i. We imported the bam.files into Geneious 8.1.7 (http://www.geneious.com) [S46] and re-assembled the reads against the Reconstructed Sapiens Reference Sequence (RSRS) [S45] with 5 iterations, which improved the assembly in problematic C-stretch regions of the HVS I and II.

ii. We used the automated variant caller of Geneious with a minimum variant frequency of 0.7 (or 0.6 for samples with high contamination), and a minimum coverage of 2X and exported the resulting variant (SNP) tables into Excel.

iii. We compared the variants to the SNPs reported in the online mtDNA phylogeny (mtDNA tree Build 16, 19 Feb 2014, http://www.phylotree.org) to ensure all phylogenetically expected SNPs are indeed observed. As per above, we masked notoriously troublesome sites at 309.1C(C), 315.1C, AC indels at 515-522, 16182C, 16183C, 16193.1C(C) and 16519 [S43].

iv. The assemblies were subsequently inspected visually, taking note of additional SNPs that might be presented but had been called automatically as the variant frequency fell below the 0.7 (0.6) cut-off. At the same time we paid particular attention to sites that could identify potential contaminating lineages. For example, as both ancient and contaminating lineages are in all cases of Eurasian origin, SNPs of the basal 'African' stem of the tree are shared and therefore present in nearly 100% of all reads covering the respective sites. For most samples (non-M haplogroup), this was also the case for the characteristic SNPs leading into macro-hg N (G8701A, C9540T, G10398A, C10873T, A15301G!), and R (T12705C, T16223C). For all reads covering SNP sites characteristic for hg U (A11467G, A12308G, G12372A) and

the respective diagnostic U sub-hg SNPs we usually also observed a fraction of reads that did not show the derived status at these sites. If the ratio of 'contaminating' reads exceeded 5%, it proved relatively easy to track down the profile of the contaminating lineage, for example by double-checking for reads with derived calls in diagnostic SNP sites for hg H (G73A, A11719G, T14766C, G2706A, T7028C). An alternative crude but effective check was to look at the HVSI regions for SNPs that are characteristic for other haplogroups (which often proved to be a reliable indicator) and then checking the expected diagnostic positions from the last common branch (in most cases branch R) upwards.

v. For these diagnostic sites we simply counted the number of ancestral and derived bases relative to the depth of coverage at that site. By averaging across all diagnostic sites that lead from the common branch R to a derived sub-haplogroup (e.g. 12 sites for sub-hg U5a1 or 5 sites into basal hg H), we have sufficient data points to calculate a rough but direct estimate for the support (majority or minority) of the respective lineages present in reads from an ancient sample, which is also a direct measure of support of endogenous vs. contaminating reads. For example, sample Paglicci108 is very likely a U2'3'4'7'8'9 lineage but the derived diagnostic SNPs are only present in 67.4% of the reads on average. At the same time, we observe reads that are derived at six characteristic SNPs diagnostic for hg H1 with an average 33.9% of the reads. This almost sums up perfectly, while the discrepancy could be explained by DNA damage that could potentially confound the precision of the signal by in- or deflating the calls for SNP, which follow the direction (C>T, G>A) of characteristic ancient DNA damage. Moreover, the authenticity of the mtDNA hg U2'3'4'7'8'9 rather than H1 is further confirmed by the coalescence age of the latter hg being considerably younger (~10 ka) [S45] than the age of the specimen (~ 28ka).

The SNP lists generated by schmutzi and visual inspection were finally compared for each sample and the detected inconsistencies were manually checked in the read alignments using Geneious. In total only 34 positions were edited in the schmutzi consensus sequences over the complete dataset of 35 complete mtDNAs (see below). Only 5 positions were converted into a different nucleotide based on hg phylogeny or contamination that was further estimated looking at reads also overlapping neighboring mutations. Insertions were shifted in 6 samples because of misalignment with the reference genome. Finally all the others 22 positions were mainly transitions (19 out of 22) in the form C to T or G to A that were converted into Ns because after visual inspection we couldn't rule out post-mortem-damage origin especially at low coverage (1-3fold).

For each sample we report below the assigned haplogroup (Table S2) in bold, their derived and missing SNPs compared to the RSRS reference and the manually edited positions in brackets.

- BerryAuBac1, **U5b1a**: T16192C!, C16519T
- Bockstein, **U5b1d1**: C16176T, C16519T, (525insAC)
- Brillenhohle, **U8a**: G1422A, C16519T, C9365T missing, (C14519N, 10107N, 14493N)
- Burkhardtshohle, **U8a**: C150T, G1422A, C16519T, C9365T missing, (A7055G)
- Cioclovina1, **U**: T15889C, C16519T, (G12372A)
- CuiryLesChaudardes1, **U5b1b**: 6056T, 6057T, T16092C, C16327T, C16519T, (7362N)
- DolniVestonice16, **U5**: G1462A, C16519T, (4N, 6N)
- DolniVestonice43, **U5**: T16231C, C16519T
- Falkenstein, **U5b2a**: A4732G, G8572A, A16171G, C16519T, T16189C! missing
- Felsdach, **U5a2c**: T152C!,735.1G, 11728T, C16192T, C16519T
- Goyet2878-21, **U5**: T3202C, C3612T, C13272T, A13299G, T16192C!, C16519T
- GoyetQ-2, **U8a**: C150T, G1422A, C16519T, C9365T missing
- GoyetQ116-1, **M**: G207A, C1556T, C6045T, C8619a, A11084G, T16297C
- GoyetQ376-19, **U2**: C152T!!, T217C, T5426C, G12406A, A12579G, T16092C, G16129c, T16189C!
- GoyetQ376-3, **M**: G207A, C1556T, A6040G, C6041T, C6045T, C8619a, A11084G, T16297C
- GoyetQ53-1, **U2**: C152T!!, T217C, T5426C, G12406A, A12579G, 16129G, T16189C!, (1N)
- GoyetQ55-2, **U2**: C152T!!, T217C, C4011T, C4013T, T5426C, G12406A, A12579G, T16092C, G16129c, T16189C!, (T16092C)
- GoyetQ56-16, **U2**: C152T!!, T217C, T5426C, G16129c, T16189C!, T16362C
- HohleFels10, **U8a**: C150T, G1422A, C16519T, C9365T missing
- HohleFels49, **U8a**: C150T, G1422A, C16519T, C9365T missing
- HohleFels79, **U8a**: C150T, G1422A, C16519T, C9365T missing
- HohlensteinStadel, **U5b2c1**: C16519T, (960.XC, T14182C)

- Iboussieres25-1, **U5b2a**: A4732G, C16114a, C16519T
- Iboussieres31-2, **U5b1**: 12398T, G13708A, C16519T, (6N, 11N, 16192.1T)
- Iboussieres39, **U5b2b**: 6027A, 8161T, T8167C, T13356C, C16519T
- LaRochette, **M**: G207A, C1556T, C6045T, C6164T, C8619a, A11084G, C12816T, (525insAC, 16297N)
- LesCloseaux3, **U5a2**: G54A, G7805A, C16519T
- MareuilLesMeaux1, **U5a2**: T16362C, C16519T, (6008N)
- Ofnet, **U5b1d1**: C16176T, C16519T, (525insAC)
- Paglicci108, **U2'3'4'7'8'9**: G1709A, G6260A, T9716C, C16256T, C16519T (C14766T)
- Paglicci133, **U8c**: C16519T, (6N, 11N)
- Paglicci71, **U5b2b**: C6910T, C16519T, (11N, 4467N, 5213N, 5844N, 7396N, 15050N, 16313N)
- Ranchot88, **U5b1**: G3531A,8276.1C, T16189C!, T16192C!, C16519T, (8276.1C)
- Rigney1, **U2'3'4'7'8'9**: C150T, T152C!, T6152C, T5999C, T10020C, A12082G, A12530G, G15466A, T16297C
- Rochedane, **U5b2b**: A8449G, T16086C, C16519T, (525insAC)

Phylogenetic analyses

The software MUSCLE [S47] was used to align complete or almost complete sequences of 56 Late Pleistocene and Early Holocene European human mtDNAs (including Oase1 [S27]) to 54 modern worldwide human mtDNAs [S48]. A Neanderthal mtDNA sequence (Vindija 33.25 FM865410.1) was included as a phylogenetic out-group for a total of 111 mtDNA sequences aligned. A phylogenetic tree (Figure 2) was built with the Maximum Parsimony method (SPR algorithm) using MEGA6 [S49]. The tree was constructed with 99% partial deletion for a total of 16371 positions in the final dataset. Moreover the phylogeny was tested with the bootstrap method with 1000 replications. The same multiple genome alignment, including missing sites but not gaps, was tested in Modelgenerator [S50], and GTR with invariant sites and gamma distributed correction for rate heterogeneity was found to be the substitution model that best fits the data (AIC1). This was selected in MrBayes [S51], which was used to reconstruct a Bayesian phylogenetic tree (Figure S1). Markov chain Monte Carlo (MCMC) was performed with 50,000,000 iterations and 10,000 sampling interval. After removing the first 10% of the generated trees as burn-in, the summarized tree was built, showing high posterior support at the major mitochondrial branches. Both trees were visualized and edited with FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/) and finally refined in Inkscape (https://inkscape.org/en/).

Phylogenetic relationships between 56 (55 plus Oase1) complete or almost complete mtDNA from European pre-Neolithic individuals were visualized through Maximum Parsimony and Bayesian trees, which both provided the same general topology (Figure 2, Figure S1). All ancient samples share a MRCA within the modern day mtDNA diversity with the exception of Oase1 [S27], which branches off basal to all other present-day N lineages. The great majority of samples (52 out of 55) belong to hg N mostly falling within sub-hgs U8, U5 and U2'3'4'7'8'9. Only Fumane2 [S22] represents a basal R lineage without any derived position leading to known sub-hgs, such as the ~45,000 years old individual from Ust'Ishim in Siberia [S29]. Finally, three pre-LGM individuals (LaRochette, GoyetQ116-1 and GoyetQ363-3) carry mtDNA hg M defined by four positions from the L3 split (T489C, C10400T, T14783C, G15043A). Looking more closely at the intragroup M phylogeny, these three pre-LGM individuals are all placed on a branch not yet seen in modern-day individuals, either in Europe or elsewhere. In particular GoyetQ116-1 and GoyetQ363-3 from Belgium dated ~34-35 ka, show 6 (G207A, C1556T, C6045T, C8619a, A11084G, T16297) and 8 (G207A, C1556T, A6040G, C6041T, C6045T, C8619a, A11084G, T16297) derived mutations from the M root, respectively, whereas LaRochette from South France dated ~27.5 ka presents 7 (G207A, C1556T, C6045T, C6164T, C8619a, A11084G, C12816T) derived mutations from the M root. LaRochette carries a possible additional back mutation at position C16297T! (16Ts and 4Cs), compared to the two Belgian sequences both carrying the T16297C mutation. The four Cs at this position cannot be explained by damage and are likely the result of contamination or heteroplasmy, therefore an unassigned base (N) was placed at that position in the final consensus sequence (see above). Despite their geographic and temporal separation all three pre-LGM individuals share five mutations (G207A, C1556T, C6045T, C8619a, A11084) from the MRCA of the basal M lineage. This finding suggests that this M branch arrived in Europe before 35 ka and survived at least until 27.5 ka, before the LGM started (~25 ka).

BEAST analyses

The human mtDNA mutation rate can be estimated with different methods. All rely on temporal calibration points in the past such as the split time of certain populations, assuming a known phylogeny [S45]. Ancient DNA provides the opportunity to use endogenous mtDNA of ancient radiocarbon dated specimens to tip calibrate the human mtDNA molecular clock [S20, S52]. Ancient sequences should show less derived positions than more recent ones as a sign of their antiquity because they had "less time" to accumulate mutations. Therefore sequence branch shortening can be used not only to confirm the authenticity of the results but also to date divergence events. The software package BEAST v1.8.1 [S16] was used to calculate, in a Bayesian statistic framework, the human mtDNA mutation rate with ancient dated sequences as multiple calibration points in order to estimate haplogroup divergence ages. Initially, a multiple genome alignment was performed with MUSCLE by combining 311 worldwide modern human mtDNA and 66 ancient dated mtDNAs (Table S3) for a total 377 sequences. Ancient mtDNAs were selected according to the following characteristics: i) directly ^{14}C dated specimens or associated findings or archeologically dated e.g. the early Neolithic farmer from Stuttgart; ii) excluding identical mtDNA sequences with same date; iii) over 99% of the entire mtDNA with assigned nucleotides; iv) age between ~46,800 and 3,600 years BP. After removal from the 377 mtDNA alignment of all positions containing gaps and missing data, a total of 15963 positions were retained. Modelgenerator v.85 was run indicating Tamura-Nei 93 with a fixed fraction of invariable sites and gamma distributed rates as the best-supported model for our dataset. These substitution and site heterogeneity models with 6 gamma categories were selected in BEAST. Tip dates were indicated as zero for modern sequences whereas sampling with individual priors was set for ancient sequences. For each of the 66 sequences, lower and upper values of a uniform prior distribution were given as confidence interval limits of the date estimate (Table S3). Two different models of rate variation among tree branches were investigated: a strict clock and an uncorrelated lognormal-distributed relaxed clock. In both cases an approximate continuous time Markov chain rate reference prior [S53] was chosen. Moreover, two tree priors were tested: a population Constant size and Bayesian Skyline coalescent with 20 as group number and piecewise-linear as the Skyline model. For the four different model combinations generated this way, we performed three MCMC runs with 50,000,000 iterations each, sampling every 10,000. Tracer v1.6 [S16] was used to visually inspect ESS values and chain convergence, in order to identify the percentages of each run that was discarded as chain burn-in (Table S5). For each model the three independent runs were combined using LogCombiner v1.8.1 resulting in 60,000,000 to 135,000,000 iterations and summarized in a Maximum Clade Credibility (MCC) tree using TreeAnnotator v1.8.1 (both programs included in the BEAST package). Model comparison and best support assessment was performed through a marginal likelihood estimation (MLE) using path sampling (PS) / stepping-stone sampling (SS) [S54]. Overall the Skyline tree prior resulted in higher likelihoods for both strict and relaxed molecular clock than the Constant size tree prior (Table 1). The latter provided in general older hg divergence times and a higher tree root height as previously reported [S20]. Moreover the Skyline prior, that models changes in population size through time, resulted not only in younger splitting times but also narrower confidence intervals. Despite having comparable likelihood to the strict clock, this parameter in combination with a lognormal relaxed clock required more iterations for the run to converge. A burn-in of 60% was manually selected instead of the default 10% applied for the other three combinations of parameters. This caused a strong reduction in the overall ESS values although the mutation rate and coalescence times were almost identical to the strict clock. In conclusion, the Skyline tree prior in combination with a strict clock mutation rate among different tree branches was overall best supported by the data and therefore this was favored to estimate the human mtDNA mutation rate and to assess haplogroup divergence times (Table 1).

The newly estimated mutation rate value overlapped with published rates obtained with similar methods but with a narrower confidence interval [S20, S52], possibly resulting from the higher number of calibration points used in this study. This allowed us to determine hg coalescence times more precisely especially for lineages with know rate heterogeneity, such as in the M clade [S45]. In the latter study, however, the authors reported almost identical values for hg M TMRCA (~50 ka) but an almost 10 k older hg N TMRCA (~ 59 ka) than the one we obtained. A general observation is that by using a Skyline tree prior in BEAST we allowed for changes in effective population size through time, resulting in younger mtDNA tree splits than with a Constant size tree prior (Table 5) or phylogenetic methods [S45]. However, including three Late Pleistocene M sequences as deep time anchors, gave us the opportunity to date the divergence time of all M lineages in the same way as for the N lineage. This resulted in similar TMRCAs for both N and M clades at around ~50 ka. We acknowledge the possible influence of sampling biases in our dataset of 377 ancient and modern mtDNA sequences both in terms of sample size and geographical-

temporal distribution [S55]. In order to keep the potential impact at a minimum we included a vast majority of modern worldwide mtDNAs (273 out of 311) that belong to the M and N clades. Moreover, the 66 ancient sequences represented not only European individuals before 7 ka but also mtDNAs from different regions and more recent time periods (Table S3) to reduce the aforementioned sampling biases also for the ancient sequences.

Coalescent demographic modeling with approximate Bayesian computation (ABC)

The 55 samples were divided into 4 chronologically-based sample groups: 'Pre-LGM' (~45 – 25 ka, n = 18), 'Post-LGM' (~19.5 – 14.5 ka, n = 8), 'Late Glacial' (~14.5 – 11.5 ka, n = 8) and 'Holocene' (~11.5 – 7 ka, n = 21) (Table S4). Coalescent simulations were performed using Bayes Serial Simcoal [S56], and all models assumed an intergeneration time of 25 years, a mutation rate of 2.74×10^{-8} per site per year (see main text and Table 1), a transition bias of 0.956 and a continuous gamma distribution of mutation rates across sites with parameter 0.205 [S57, S58]. We used approximate Bayesian computation (ABC) to both perform model choice [S58, S59] and estimate the parameters of the best-fitting model [S60].

Following an initial test phase, we considered six distinct demographic models: 1a, 1b, 1c, 2a, 2b and 3, illustrated schematically in Figure S2. All models start with a non-European ancestral maternal population of effective size 5000, and at 45 ka, a population of size N_C expands into Europe. In model 1a, this European population simply grows exponentially, unimpeded, up to size N_H at 7 ka. In model 1b the population grows exponentially to size N_{LG} at 14.5 ka. At this point, a Late Glacial bottleneck begins, where the population is reduced to a constant proportion P_{LG} of its current size. Following the end of the Late Glacial at 11.5 ka, the population then again grows exponentially, up to effective size N_H at 7 ka. Model 1c is the same, except that there is also a bottleneck during the LGM. In this model the original European population first grows to N_{LGM} at 25 ka, then is reduced to constant proportion P_{LGM} throughout the LGM. After the end of the LGM at 19.5 ka, model 1c is identical to model 1b. In model 2a, a split occurs in the pre-LGM population at time T_S, with the new population having constant size N_S. The original population then goes extinct at the beginning of the LGM, and the new population re-expands following the LGM, up to N_H at 7 ka. Model 2b is similar, except that the pre-LGM population survives through an LGM bottleneck up until the beginning of the Late Glacial, but is then replaced by the new population, expanding at the end of the Late Glacial. Model 3 assumes that both the LGM and the Late Glacial caused population replacements, with each new population diverging from the previous one (at times T_S and T_{S2}) prior to the replacement event. All parameter prior distributions are reported in Table S5.

Model choice procedure

Previous authors have suggested that posterior probabilities of multiple competing models may be estimated in the ABC framework by calculating the proportion of accepted simulations from each model below a certain tolerance δ in a common ranked pool [S61]. However, it has been demonstrated that this approach is not theoretically justified [S62], as the loss of information in reducing the data to summary statistics means that this approximation is not guaranteed to converge on the true Bayes factors. We performed model choice using the multinomial logistic regression approach introduced by Beaumont [S58], which treats a model indicator as a categorical variable and returns an estimate of the posterior probability for each model. In order to evaluate this procedure we nested it in a simulation-based power-analysis similar to that of Veeramah *et al.* [S59], where datasets generated under each of the six models defined above are used as pseudo-observed data. Firstly, for each model we simulated 10^6 datasets and 50 independent datasets to act as pseudo-observed data. Then, given a set of summary statistics S and a tolerance level δ (discussed below), we iterated through each of the 300 pseudo-observed datasets, each time performing the rejection and multinomial logistic regression steps (using the VGAM R package by Yee TW, https://www.stat.auckland.ac.nz/~yee/VGAM), using the highest posterior probability criterion to select the best model. Power was then approximated by the proportion of times this procedure identified the correct model.

The choice of appropriate statistics S in ABC model comparison (and in general) is still a difficult and unresolved problem, so we followed the heuristic approach implemented by Veeramah *et al*. [S59], in which a set of candidate summary statistics are ranked according to their p-value from a Kruskal-Wallis test applied to a random sample of 10,000 simulations from each model. The idea is that a lower p-value suggests a greater difference in the medians of the statistic distributions across models, and hence provides greater power to distinguish between models. In order to minimize bias for or against any particular model,

we group statistics by family, e.g. all within-group mean pairwise differences (π_i) are grouped, and rank these blocks using their median p-value. Global statistics (e.g. global mean pairwise differences π_{all}) are added as blocks individually. We iteratively added new summary statistic blocks from the ranked list (Table S6) to our set *S* in our analysis. Within-group statistics for the 'Pre-LGM' sample group were excluded *a priori* given that they appear in the same position in each model and thus have no significant between-model discriminatory power. The choice of tolerance δ in the ABC rejection algorithm is also difficult to fix objectively so, again following Veeramah *et al.* [S59], we iteratively repeated the analysis accepting between 1000 and 10,000 simulations in steps of 1000. Iterating across the number of summary statistics and the number of retained simulations we found the combination (*S*,δ) with maximal power to discriminate between all competing models.

The combination (*S*,δ) that maximizes the power (0.650) for this model choice procedure is δ = 5000 and *S* comprises the first 10 statistics blocks given in Table S6. Using these values applied to our observed data we find that model 2b has by far the highest posterior probability (0.807). All 6 model posterior probabilities are given in Figure S2.

Parameter estimation

Following the identification of 2b as the best-fitting model, a further four million coalescent simulations were run as previously described, giving five million in total. We performed ABC with local-linear multivariate weighted regression adjustment [S60], assuming a tolerance $F\delta$ = 0.25% so that 12,500 simulations were retained. Other tolerance values of 0.1% and 0.5% gave almost identical results. We used 5 within-group statistics (number of haplotypes, number of segregating sites, mean pairwise difference π, haplotype diversity *H* and Tajima's *D*) for each of the four chronological sample groups described above (see Table S6), and two between-group statistics (mean pairwise difference π and F_{ST}) for each pair, giving a total of 32 summary statistics. Observed values of each summary statistic are given in Table S6, and marginal posterior densities for each parameter are given in Table S5.

Archeological site information

The 35 Upper Paleolithic and Mesolithic human specimens, from which we sequenced complete mtDNA, were obtained from 21 archeological sites across Europe. For each human remains, a portion or the entire specimen was collected from archeological collections in Belgium, Romania, Italy, Czech Republic, France and Germany and sampled in the dedicated ancient DNA laboratory facility of the University of Tübingen, in Germany. All samples were directly or stratigraphically radiocarbon dated. Newly reported ^{14}C dates were calibrated using Reimer et al.'s [S17] calibration curve (Table S1). In the main text and in the Supplemental Experimental Procedures, the date unit reported is thousands of calibrated years before present (ka) unless otherwise specified.

Peştera Cioclovina Uscată (South Transylvania, Romania)

The calvaria Cioclovina1 was discovered in Uscată located in South Transylvania. The specimen was recently radiocarbon dated to 28,510 ± 170 ^{14}C years before present [S4], placing this individual among the oldest human specimens found in Central and South-Eastern Europe. The complete mtDNA genome recovered from a cranial fragment shows a basal phylogenetic position compared to all modern day humans belonging to haplogroup U. This confirms its affinity to modern humans at the mtDNA level as previously reported from geometric morphometric analyses [S63].

La Rochette (Dordogne, France)

Otto Hauser first excavated the Middle and Upper Paleolithic site of La Rochette close to Saint-Léon-sur-Vézère in 1910. He identified human occupations of the site spanning from the Mousterian to the Magdalenian. Human remains were originally assigned to a general Aurignacian layer. However, a revision of the stratigraphy [S64] reconsidered its division into three distinct horizons, i.e. Gravettian, Aurignacian and Chatelperronian. A first ^{14}C AMS dating of a human right ulna was performed in 2002 by Orschiedt and was repeated recently making use of more advanced ultrafiltration techniques [S2]. Both dates consistently fall around 23,000 ^{14}C years before present, associating this human remain to the Gravettian material culture horizon. The dated specimen was also sampled for ancient DNA investigation.

Dolní Věstonice (South Moravia, Czech Republic)
Excavations in the northern slope of the Pavlovské Hills started in 1924 and since then three Gravettian human burial sites have been discovered, namely the Dolní Věstonice I, Dolní Věstonice II and Pavlov I. In the two former sites, human fossils are numbered from DV1 to DV64 and in Pavlov I from 1 to 33. Ancient DNA analysis was performed on individuals DV16 and DV43 both coming from the Dolní Věstonice II site. Radiocarbon dating was performed on earth and charcoal associated to DV16, assigning it to around 30 ka. Both human remains were identified in close vicinity, and were suggested to be of similar age [S5].

Grotta Paglicci (Apulia, Italy)
Grotta Paglicci (Rignano Garganico - Foggia) is located in Apulia, the southeastern-most Italian region. The site comprises the present-day cave and a rock shelter that, in the past, was part of the same cave system and testifies to a human occupation during the Lower-Middle Paleolithic (Acheulean, Early Mousterian). The 12m-thick sequence excavated inside the cave yielded remains extending from the Aurignacian to the Final Epigravettian [S12]. Paglicci is also important for the several human remains uncovered along the Upper Paleolithic sequence and for the most ancient evidence of producing flour [S65]. The investigations were conducted by the Natural History Museum of Verona (F. Zorzi) from 1961 to 1971 and carried out by the University of Siena (first by A. Palma di Cesnola and currently by A. Ronchitelli) in collaboration with the Soprintendenza Archeologia della Puglia.
Radiocarbon dating was performed on archeological layers associated with modern human occupation. The three samples analyzed here belong to three distinct stratigraphic units. Paglicci71 found in layer 8A is associated with the Evolved Epigravettian culture, Paglicci108 in layer 21B with the Evolved Gravettian and Paglicci133 found in layer 23C2, between the Early Gravettian layer 23A and the Aurignacian layer 24A1 [S11, S12, S13]. The volcanic tephra Codola dated elsewhere to around 33 ka [S66] was found in layer 23C2.

Troisième Caverne (Goyet, Belgium)
The Troisième Caverne of Goyet in Belgium was excavated in the latter half of the 19th and beginning of the 20th century, and again at the end of the 1990s. The main excavations were performed in 1868 by Edouard Dupont who identified Paleolithic human occupations [S67] that were later attributed to the Middle and Upper Paleolithic (including Mousterian, Lincombian-Ranisian-Jerzmanowician, Aurignacian, Gravettian and Magdalenian material) as well as to the Neolithic and historic period [S68]. Starting in 2008, the reassessment of both the human and faunal collections from the site yielded new human remains. Due to the lack of detailed documentation of the excavated material, their association to a specific occupation was impossible and a multidisciplinary study of the human remains and their context was undertaken. Morphometric and taphonomic features, completed by the direct radiocarbon dating of the remains, were used to assign them to different periods. In combination with isotopic and genetic analyses [S69, S70] the results allowed for the specimens to be assigned to either late Neanderthals or modern humans. Here we analyzed the mtDNA genomes of two specimens directly dated to the Aurignacian, five to the Gravettian and one to the Magdalenian (Table S1).

Swabian Jura sites (Baden-Württemberg, Germany)
We were able to collect and analyze a total of nine human remains from seven cave sites of the Swabian Jura (Hohle Fels, Brillenhöhle, Burkhardtshöhle, Bockstein-Höhle, Falkensteiner Höhle, Hohlenstein-Stadel and Felsdach Inzigkofen). Those are located in Southwest Germany mainly along the Ach and Lone valleys formed by the homonymous Danube's tributaries. The entire region is composed of Jurassic limestone where karst landscapes lead to cave formations.

- Hohle Fels (Ach valley) is the worldwide famous Swabian Jura site for the discovery of an ivory 'Venus' figurine and bird bone and ivory flutes. These are both dated to the early Aurignacian period and are among the first figurative art and musical instruments ever found in Europe [S71, S72]. In the cave, five human remains have been discovered. Three of them were genetically analyzed in the present study. A left and a right femur fragments were found in close proximity in the Magdalenian stratigraphic layer and may belong to the same individual [S7]. Both were previously genetically characterized as belonging to haplogroup U, based on a short mtDNA amplicon [S6]. In the present study, we reconstructed complete mtDNA genomes for both specimens and a well-supported phylogenetic placement was assigned. The two specimens were found to have an identical haplogroup U8a (with private mutations C150T, G1422A, C16519T) supporting that, if not the same individual, they have a close maternal relationship. A skull

fragment was also identified bearing the same haplotype. Although it was initially assigned to the Gravettian layer [S73], we report here a direct radiocarbon date of the specimen to around 15 ka placing it also within the Magdalenian archeological horizon.

- Brillenhöhle (Ach Valley) was discovered in 1956 and excavated by Gustav Riek. A minimum number of four individuals have been assigned to the Magdalenian techno-complex. The parietal bone fragment genetically analyzed in the present work was directly dated to around 15 ka [S2]. This date is within the typical temporal range of the Magdalenian in the Swabian Jura [S7].
- Burkhardtshöhle (Westerheim, Württemberg) was first excavated in 1933-34 by Gustav Riek, who resumed the excavation in 1953 after the end of World War II. A total of five cranial fragments probably belonging to a single individual were discovered and radiocarbon dating confirmed their assignment to their Magdalenian timeframe [S3].
- Bockstein-Höhle (Lone valley) was discovered in 1883 and first excavated by Bürger. A double burial was identified with an almost complete female skeleton and an infant skeleton at the right end of the female's feet. The infant has been considered as the child of the female individual. Two anatomical elements of the infant were previously radiocarbon dated providing a consistent burial age of around 8 ka and associating it to the Late Mesolithic time period [S1]. The upper right second incisor tooth from the female was sampled and genetically analyzed.
- Falkensteiner Höhle is located close to the village of Bad Urach, southwest from the aforementioned Ach and Lone valleys. Around 40 human skeletal elements were identified during the excavation in 1933 and the analysis of the retrieved material was carried out in 1964 [S7]. As a result, the layer that yielded the human remains with associated microlithic stone tools has been radiocarbon dated to the Early Mesolithic [S74]. The same anatomical element investigated here was genetically analyzed in Bramanti et al. [S6]. Besides confirming the assignment to mtDNA hg U5b2, we were able to reconstruct its complete mtDNA and further define it as haplogroup U5b2a with additional private mutations (A4732G, G8572A, A16171G, C16519T).
- Hohlenstein-Stadel (Lone valley) contains evidence of hominid occupations during the Middle Paleolithic, Upper Paleolithic, Mesolithic and Neolithic. A burial formed by the skulls and cervical vertebras of three individuals (one male, one female and one infant) was discovered and dated to the Late Mesolithic [S9]. Here, we sampled one cervical vertebra for ancient DNA analysis.

Felsdach Inzigkofen (Upper Danube valley, Germany)

The site was discovered in 1965 close to the village of Beuron in the Upper Danube valley and excavated by Wolfgang Taute. A single wisdom tooth analyzed here was discovered in a layer dated to the initial Late Mesolithic phase also known as the Early Atlantic [S7, S74].

Große Ofnet Höhle (Franconian Jura, Germany)

The molar analyzed by ancient DNA belongs to an individual that was found in the Große Ofnet cave close to Nördlingen. It may belong to a burial site formed by 34 skulls of both sexes and different developmental ages, found in two burial pits [S9]. Lithic artifacts and faunal remains (deer canines, snail shells) were found in association with the human remains. All crania were facing west and presented cut-marks [S75]. Several skulls display traumatic injuries, suggesting interpersonal violence. All radiocarbon-dated individuals provided a consistent age around 8 ka and were assigned to the Late Mesolithic [S9].

French Jura sites (Franche-Comté, France)

Three analyzed human specimens were identified in three different prehistoric sites of the French Jura, located along the border between Eastern France and Western Switzerland. This medium size mountainous region is composed of Mesozoic limestone defining a karst landscape [S76].

- Cabônes rockshelter (Ranchot, Jura department) was excavated between 1978 and 1990 and yielded at least five Mesolithic human individuals [S77]. The one genetically analyzed is defined by a reconstructed right parietal bone dated to around 10 ka and assigned to the Middle Mesolithic period.
- Rigney 1 cave (Doubs department) was excavated at the beginning of the 1950s and a human mandible fragment was discovered during a rescue excavation that occurred in 1986 and 1987 at this Magdalenian site [S76]. This bone has been directly radiocarbon dated to around 15.5 ka [S14] and it was sampled for paleogenetic investigation.

• Rochedane rockshelter (Villars-sous-Dampjoux, Doubs department) was discovered at the end of the 19th century and was excavated by A. Thévenin between 1968 and 1976. This site is one of the most important prehistoric Late Glacial sequences in Eastern France [S14]. Several human bones have been found in the Mesolithic and the Final Paleolithic levels. The one genetically analyzed is a mandible dated to around 13 ka and assigned to the Epipaleolithic period.

Paris Basin (France)

The Paris Basin is a geological region in Northeast France formed by sedimentary rocks. The area is characterized by plateau plains where several Mesolithic human burials have been identified [S78].

• Les Closeaux at Rueil-Malmaison is a Mesolithic site located along the river Seine. In sector 3, a circular burial with a single and almost complete human skeleton was found. The individual was buried in a sitting position and has been directly radiocarbon dated to 9.9 ka [S10].

• Les Vignolles at Mareuil-lès-Meaux is an archeological site on the river Marne. The oldest inhumation that was discovered at the site is a single burial directly dated to the Mesolithic time [S10]. However, the presence of additional Neolithic burials and a Bronze Age necropolis attested to occupations of the site also during later periods [S79].

• Les Fontinettes at Cuiry-lès-Chaudardes is a multi-period site with an important early Neolithic (LBK) settlement. The Mesolithic burial was found on the edge of the site. The individual, in a squatting position, had a necklace of fish vertebrae and the burial contained three flint bladeless [S80]. The skeleton has been dated to around 8 ka.

• Le Vieux Tordoir at Berry-au-Bac is also a multi-period site, including an early Neolithic (LBK) settlement. One Mesolithic burial was discovered on the site. The deceased was buried in a seated position with ochre and a bone tool [S81]. The skeleton has been dated to around 7 ka.

Aven des Iboussières à Malataverne (Rhône-Alpes, France)

The site is located in the Drôme department of South France. A minimum number of eight human individuals were found at the Aven des Iboussières and were interpreted as burials of Epipaleolithic hunter-gatherers [S82]. We obtained complete mtDNA genomes from three individuals: Iboussieres25-1, Iboussieres31-2 and Iboussieres39. Only the latter, a left femur fragment, was directly radiocarbon dated to 11.8 ka. Nonetheless, the date overlaps with that of a faunal remain from the site (10210 ± 80 uncal BP; OxA5682) [S83], suggesting that all burials belong to the same temporal unit.

Climatic considerations and time definitions

The oldest possible evidence of early modern humans in Europe, found in Southern Italy, dates back to ~45 ka [S2]. Europeans relied uniquely on a foraging lifestyle for almost 40,000 years, from the first arrival to the spread of farming (~8 ka). This time spans the end of the Late Pleistocene (130 – 11.5 ka) and the beginning of the Holocene epochs (11.5ka – today) and represents around three-quarters of the time during which modern humans occupied Europe. This period (45 – 8 ka) was affected by severe climatic fluctuations and repeated environmental changes [S84]. Such phenomena are captured by the extended INTIMATE oxygen-18 oscillation curve [S85] (Figure 3) recorded by the North Greenland Ice-Core Project (NGRIP) [S86]. The observed values along this timeline correlate indirectly with the temperatures of the Northern Hemisphere and separate time into Marine Isotope Stages (MIS). Those represent relatively consistent warm interglacial (odd numbers) or cold glacial (even numbers) intervals alternating through time and numbered going back in time from the present day interglacial MIS1.

The time interval investigated in this study (45 – 7 ka) span across three climatically distinct periods: MIS3, MIS2 and MIS1 (from the past to the present). During the general warm MIS3 (~57 – 29 ka), there were dramatic fluctuations of climatic conditions spanning ~1,000 years intervals [S87]. It has been suggested that the spread of modern humans from the Near East into Europe was favored by a climatic warming phase during the MIS3 [S88]. Figure 1A depicts one of these relatively warm phase around 39 – 36 ka with sea level about 60 meters below present level [S84]. The mitigated climatic conditions reduced the ice sheet diffusion with a retraction of the steppe-tundra landscape towards Northern Europe [S84]. However environmental conditions kept on changing for the next millennia during an attested human occupation of Europe (Pre-LGM in Figure 3).

With the start of MIS2 (~29 ka), a general trend of temperature reduction began until the onset of the Last Glacial Maximum (LGM) ~25ka when all ice sheets reached their maximal extension since the last glacial

period and the sea level was approximately 130m lower than today [S89, S90] During this period, that lasted until ~19.5 ka, ice caps covered most of Northern Europe, the Alps and the Pyrenees mountain chains [S91, S92]. These cold and dry conditions throughout Europe prevented human occupation of North-Western Europe and likely forced migrations to climatic and environmental refugia [S90]. Proposed areas for human refugia during the LGM are the Mediterranean regions of Franco-Cantabria and Southern Italy, the Balkans and the East European Plain [S93]. Those geographical areas were less impacted by harsh climatic conditions and would have allowed human survival [S94].

With the end of the LGM, humans spread again into central Europe during a period we designated as the post-LGM (Figure 1B, ~19.5 – 14.5 ka), following a general retraction of the Alpine and Pyrenean ice caps and a fragmentation of the North European ice sheet into the Scandinavian and the British ice sheets [S95]. Sea levels were around 100m lower than currently [S96], exposing a vast area of land, known as Doggerland, which is now covered by the Southern North Sea. The bridge of land that connected Great Britain to main Europe lasted until ~8.5ka, when it was covered by the rising sea levels [S97].

The glacial MIS2 phase terminates with the end of the Pleistocene. This period we refer to as the Late Glacial (Figure 1C) is characterized by abrupt climatic variations starting with the warm Bølling/Allerød interstadial (14.5 – 12.7 ka) and ending with the cold Younger Dryas stadial (12.7 – 11.5ka) [S98]. During the first phase, temperatures rose dramatically over the MIS2 average, causing a radical change in European flora that triggered the spread of natural forestation towards Northern Europe [S99]. Those warmer climatic stages are also likely connected with the large extinction of Pleistocene megafauna species and a general turnover in mammalian diversity [S100, S101]. In the following Younger Dryas, the final stage of the last glacial, European average temperature sharply decreased within a few years [S102]. The impact of climatic and environmental fluctuations on human populations at the end of the Pleistocene is largely unknown. However, climate might have played a role in the mtDNA composition shift observed in this study with the beginning of the Late Glacial.

The Holocene starts around 11.5 ka within the MIS1 phase that is characterized by a warm and moist climate, which continues until today. Temperatures rose simultaneously with the sea level and by 7 ka almost all coastlines were in a similar location as nowadays (Figure 1D) [S97]. Holocene hunter-gatherers persisted with a distinct genetic signature at least until ~5 ka in Northern Europe [S18]. The Neolithic transition reached central Europe ~8 ka with the arrival of agricultural communities that drastically changed the existing European genetic makeup [S25].

Supplemental references:

S1. Wehrberger, K. (2000). „Der Streit ward definitiv beendet...” Eine mesolithische Bestattung aus der Bocksteinhöhle im Lonetal, Alb-Donau-Kreis. Zur Erinnerung an Ludwig Bürger (1844-1898). Archäologisches Korrespondenzblatt *30*, 15-31.

S2. Benazzi, S., Douka, K., Fornai, C., Bauer, C.C., Kullmer, O., Svoboda, J., Pap, I., Mallegni, F., Bayle, P., Coquerelle, M., et al. (2011). Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. Nature *479*, 525-528.

S3. Simon, U. (1993). Die Burkhardtshöhle - eine Magdalénienstation am Nordrand der Schwäbischen Alb. (Eberhard-Karls-Universität Tübingen: Magisterarbeit).

S4. Soficaru, A., Petrea, C., Dobos, A., and Trinkaus, E. (2007). The human cranium from the Pestera Cioclovina Uscata, Romania - Context, age, taphonorny, morphology, and paleopathology. Curr Anthropol *48*, 611-619.

S5. Trinkaus, E., and Svoboda, J. (2006). Early Modern Human Evolution in Central Europe: The People of Dolní Věstonice and Pavlov, Volume 12, (Oxford University Press).

S6. Bramanti, B., Thomas, M.G., Haak, W., Unterlaender, M., Jores, P., Tambets, K., Antanaitis-Jacobs, I., Haidle, M.N., Jankauskas, R., Kind, C.J., et al. (2009). Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. Science *326*, 137-140.

S7. Haas-Campen, S. (1993). Die menschlichen Skelettreste des Spätpleistozäns und Frühholozäns in Baden-Württemberg. In Geowissenschaftliche Fakultät, Institut für Humangenetik und Anthropologie. (Eberhard-Karls-Universität Tübingen: Magisterarbeit).

S8. Housley, R.A., Gamble, C.S., Street, M., and Pettitt, P.B. (1997). Radiocarbon evidence for the Lateglacial human recolonisation of Norther Europe. Proc. Prehist. Soc. *63*, 25-54.

S9. Orschiedt, J. (1999). Manipulationen an menschlichen Skelettresten. Taphonomische Prozesse, Sekundärbestattungen oder Kannibalismus. Dissertation. Urgeschichtliche Materialhefte *13*.

S10. Valentin, F., COTTIAUX R., C., B.-M., CONFALONIERI J., DELATTRE V., LANG L., LE GOFF I., LAWRENCE-DUBOVAC P., and C., V. (2008). Découvertes récentes d‘inhumations et d’incinération datées du Mésolithique en Ile de France. Revue Archéologique d’Ile-de-France, 21-42.

S11. Azzi, C.M., Bigliocca , L., and Piovan, F. (1974). Florence Radiocarbon Dates II. Radiocarbon *16*, 10-14.

S12. Palma di Cesnola, A. (2004). Paglicci. L’Aurignaziano e il Gravettiano antico, (Foggia: Claudio Grenzi Ed.).

S13. Azzi, C.M., Bigliocca , L., and Piovan, F. (1977). Florence Radiocarbon Dates III. Radiocarbon *19*, 165-169.

S14. Cupillard, C., Magny, M., Bocherens, H., Bridault, A., Begeot, C., Bichet, V., Bossuet, G., Drucker, D.G., Gauthier, E., Jouannic, G., et al. (2015). Changes in ecosystems, climate and societies in the Jura Mountains between 40 and 8 ka cal BP. Quaternary International *378*, 40-72.

S15. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol *16*, 224.

S16. Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol *7*, 214.

S17. Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2013). Intcal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years Cal Bp. Radiocarbon *55*, 1869-1887.

S18. Skoglund, P., Malmstrom, H., Omrak, A., Raghavan, M., Valdiosera, C., Gunther, T., Hall, P., Tambets, K., Parik, J., Sjogren, K.G., et al. (2014). Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science *344*, 747-750.

S19. Bollongino, R., Nehlich, O., Richards, M.P., Orschiedt, J., Thomas, M.G., Sell, C., Fajkosova, Z., Powell, A., and Burger, J. (2013). 2000 years of parallel societies in Stone Age Central Europe. Science *342*, 479-481.

S20. Fu, Q., Mittnik, A., Johnson, P.L., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr Biol *23*, 553-559.

S21. Sanchez-Quinto, F., Schroeder, H., Ramirez, O., Avila-Arcos, M.C., Pybus, M., Olalde, I., Velazquez, A.M., Marcos, M.E., Encinas, J.M., Bertranpetit, J., et al. (2012). Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Curr Biol *22*, 1494-1499.
S22. Benazzi, S., Slon, V., Talamo, S., Negrino, F., Peresani, M., Bailey, S.E., Sawyer, S., Panetta, D., Vicino, G., Starnini, E., et al. (2015). Archaeology. The makers of the Protoaurignacian and implications for Neandertal extinction. Science *348*, 793-796.
S23. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., et al. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. Curr Biol *18*, 1687-1693.
S24. Krause, J., Briggs, A.W., Kircher, M., Maricic, T., Zwyns, N., Derevianko, A., and Paabo, S. (2010). A complete mtDNA genome of an early modern human from Kostenki, Russia. Curr Biol *20*, 231-236.
S25. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409-413.
S26. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87-91.
S27. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature *524*, 216-219.
S28. Gilbert, M.T., Kivisild, T., Gronnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Gotherstrom, A., Campos, P.F., et al. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science *320*, 1787-1789.
S29. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prufer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445-449.
S30. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun *6*, 8912.
S31. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.L., et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A *110*, 15758-15763.
S32. Rohland, N., and Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. Nat Protoc *2*, 1756-1762.
S33. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc *2010*, pdb prot5448.
S34. Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res *40*, e3.
S35. Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jager, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., et al. (2013). Genome-wide comparison of medieval and modern Mycobacterium leprae. Science *341*, 179-183.
S36. Ruano, G., and Kidd, K.K. (1992). Modeling of heteroduplex formation during PCR from mixtures of DNA templates. PCR Methods Appl *2*, 112-116.
S37. Maricic, T., Whitten, M., and Paabo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One *5*, e14004.
S38. Kao, W.C., Stevens, K., and Song, Y.S. (2009). BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing. Genome Res *19*, 1884-1895.
S39. Kircher, M. (2012). Analysis of high-throughput ancient DNA sequencing data. Methods Mol Biol *840*, 197-228.
S40. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.
S41. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (subm.). EAGER: Efficient Ancient Genome Reconstruction. Genome Res.

S42. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
S43. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat *30*, E386-394.
S44. Kloss-Brandstatter, A., Pacher, D., Schonherr, S., Weissensteiner, H., Binna, R., Specht, G., and Kronenberg, F. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat *32*, 25-32.
S45. Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogvali, E.L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet *90*, 675-684.
S46. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.
S47. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.
S48. Ingman, M., Kaessmann, H., Paabo, S., and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. Nature *408*, 708-713.
S49. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol *30*, 2725-2729.
S50. Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and McLnerney, J.O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol *6*, 29.
S51. Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754-755.
S52. Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L.A., Warmuth, V., Ruiz-Linares, A., Manica, A., and Balloux, F. (2014). Improved calibration of the human mitochondrial clock using ancient genomes. Mol Biol Evol *31*, 2780-2792.
S53. Ferreira, M.A.R., and Suchard, M.A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. Can J Stat *36*, 355-368.
S54. Baele, G., Lemey, P., and Vansteelandt, S. (2013). Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. BMC Bioinformatics *14*, 85.
S55. Mellars, P., Gori, K.C., Carr, M., Soares, P.A., and Richards, M.B. (2013). Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. Proc Natl Acad Sci U S A *110*, 10699-10704.
S56. Anderson, C.N., Ramakrishnan, U., Chan, Y.L., and Hadly, E.A. (2005). Serial SimCoal: a population genetics model for data from multiple populations and points in time. Bioinformatics *21*, 1733-1734.
S57. Bandelt, H.J., Salas, A., and Bravi, C.M. (2006). What is a 'novel' mtDNA mutation--and does 'novelty' really matter? J Hum Genet *51*, 1073-1082.
S58. Beaumont, M. (2008). Joint determination of topology, divergence time and immigration in population trees. In Simulations, genetics and human prehistory, S. Matsumura, P. Forster and C. Renfrew, eds. (Cambridge: McDonald Institute for Archaeological Research), pp. 135-154.
S59. Veeramah, K.R., Wegmann, D., Woerner, A., Mendez, F.L., Watkins, J.C., Destro-Bisol, G., Soodyall, H., Louie, L., and Hammer, M.F. (2012). An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol Biol Evol *29*, 617-630.
S60. Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025-2035.
S61. Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., and Feldman, M.W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol *16*, 1791-1798.
S62. Robert, C.P., Cornuet, J.M., Marin, J.M., and Pillai, N.S. (2011). Lack of confidence in approximate Bayesian computation model choice. Proc Natl Acad Sci U S A *108*, 15112-15117.

S63. Harvati, K., Gunz, P., and Grigorescu, D. (2007). Cioclovina (Romania): affinities of an early modern European. J Hum Evol *53*, 732-746.
S64. Delporte, H. (1962). Le gisement paléolithique de la Rochette, commune de Saint-Léon-sur-Vézère, Dordogne. Gallia Préhistoire *V*, 1-22.
S65. Lippi, M.M., Foggi, B., Aranguren, B., Ronchitelli, A., and Revedin, A. (2015). Multistep food plant processing at Grotta Paglicci (Southern Italy) around 32,600 cal BP. P Natl Acad Sci USA *112*, 12075-12080.
S66. Giaccio, B., Isaia, R., Fedele, F.G., Di Canzio, E., Hoffecker, J., Ronchitelli, A., Sinitsyn, A.A., Anikovich, M., Lisitsyn, S.N., and Popov, V.V. (2008). The Campanian Ignimbrite and Codola tephra layers: Two temporal/stratigraphic markers for the Early Upper Palaeolithic in southern Italy and eastern Europe. J Volcanol Geoth Res *177*, 208-226.
S67. Dupont, E. (1872). L'Homme pendant les âges de la pierre dans les environs de Dinant-sur-Meuse, Volume 2ème édition, (Bruxelles: C. Muquardt Ed.).
S68. Flas, D. (2006). La transition du Paléolithique moyen au supérieur dans la plaine septentrionale de l'Europe. Les problématiques du Lincombien-Ranisien-Jerzmanowicien. Volume 119. (Université de Liège: PhD dissertation), p. 254.
S69. Wißing, C., Rougier, H., Crevecoeur, I., Germonpré, M., Naito, Y.I., Semal, P., and Bocherens, H. (in press). Isotopic evidence for dietary ecology of Late Neandertals in North-Western Europe. Quat Int.
S70. Rougier, H., Crevecoeur, I., Beauval, C., Posth, C., Flas, D., Wißing, C., Furtwängler, A., Germonpré, M., Gómez-Olivencia, A., Semal, P., et al. (subm.). Neandertal cannibalism at the Troisième caverne of Goyet (Belgium).
S71. Conard, N.J., Malina, M., and Munzel, S.C. (2009). New flutes document the earliest musical tradition in southwestern Germany. Nature *460*, 737-740.
S72. Conard, N.J. (2009). A female figurine from the basal Aurignacian of Hohle Fels Cave in southwestern Germany. Nature *459*, 248-252.
S73. Conard, N.J., and Bolus, M. (2003). Radiocarbon dating the appearance of modern humans and timing of cultural innovations in Europe: new results and new challenges. J Hum Evol *44*, 331-371.
S74. Holdermann, C.-S., and Kind, C.-J. (2009). Zeitwechsel in Schichten. Bedeutende Fundstellen an der oberen Donau. Eiszeit. Kunst und Kultur. Begleitband zur Großen Landesausstellung Eiszeit - Kunst und Kultur im Kunstgebäude Stuttgart, 332–335.
S75. Mollison, T. (1936). Zeichen gewaltsamer Verletzungen an den Ofnet-Schädeln. Anthrop. An. *13*, 79-88.
S76. Cupillard, C., Malgarini, R., and Fornage-Bontemps, S. (2013). Le Paléolithique supérieur ancien dan le quart nord-est de la France : l'exemple de la Franche-Comté. Environnement, chronologie et faciès culturels. Mémoire LVI de la Société préhistorique française, 351 – 363.
S77. Valentin, F. (1998). Les restes humains de l'abri des Cabônes à Ranchot (Jura). In : Les derniers chasseurs-cueilleurs (13000-5500 av. JC) dans le massif du Jura et ses marges. Centre Jurassien du Patrimoine, 185-186.
S78. Bosset, G., and Valentin, F. (2013). Mesolithic burial practices in the northern half of France: Isolated burials and their spatial organisation. Paris: Société préhistorique française, 207-216.
S79. Cottiaux, R., Delattre, V., Lawrence-Dunovac, P., and Durand, S. (2001). Les occupations néolitiques et protohistoriques du site Mareuil-lès-Meaux "les Vignolles" (Seine-et-Marne), résultats préliminares. Achtes des Journées Archéologiques d' Île-de-France, 60-63.
S80. Ilett, M. (1998). Cuiry-lès-Chaudardes "les Fontinettes". In Rapport de fouille programmée. (Université de Paris 1-Panthéon Sorbonne: Ministère de la Culture).
S81. Auxiette, G., and Hachem, L. (1989). Berry-au-Bac, "Le Chemin de la Pêcherie ouest", "le Vieux-Tordoir", "la culée". (Fouille Protohistorique de la Vallée de l'Aisne: Ministère de la Culture).
S82. Gely, B., and Morand, P. (1998). Les sépultures épipaléolithiques de l'aven des Iboussières à Malataverne (Drôme, France): Premiers résultats. Ardèche Archéologie *15*, 13-18.
S83. Chaix, L. (2005). Hétéroclite et éclectique: la faune épipaléolithique de l'Aven des Iboussières (Drôme, France). Antropologia-Arkeologia *57*, 411-420.
S84. Van Andel, T.H., and Tzedakis, P.C. (1996). Palaeolothic landscapes of Europe and environs, 150,00-25,000 years ago: an overview. Quaternary Science Reviews *15*, 481-500.

S85. Blockley, S.P.E., Lane, C.S., Hardiman, M., Rasmussen, S.O., Seierstad, I.K., Steffensen, J.P., Svensson, A., Lotter, A.F., Turney, C.S.M., and Bronk Ramsey, C. (2012). Synchronisation of palaeoenvironmental records over the last 60,000 years, and an extended INTIMATE event stratigraphy to 48,000 b2k. Quaternary Science Reviews *36*, 2-10.

S86. Andersen, K.K., Azuma, N., Barnola, J.M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H.B., Dahl-Jensen, D., Fischer, H., et al. (2004). High-resolution record of Northern Hemisphere climate extending into the last interglacial period. Nature *431*, 147-151.

S87. Huijzer, A.S., and Isarin, R.F.B. (1997). The reconstruction of past climates using multi-proxy evidence: an example of the Weichselian Pleniglacial in northwest and central Europe. Quaternary Science Reviews *16*, 513-533.

S88. Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science *308*, 1034-1036.

S89. Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W., and McCabe, A.M. (2009). The Last Glacial Maximum. Science *325*, 710-714.

S90. Gamble, C., Davies, W., Pettitt, P., and Richards, M. (2004). Climate change and evolving human diversity in Europe during the last glacial. Philos Trans R Soc Lond B Biol Sci *359*, 243-253; discussion 253-244.

S91. Carr, S.J., Holmes, R., Van der Meer, J.J.M., and Rose, J. (2006). The Last Glacial Maximum in the North Sea Basin: micromorphological evidence of extensive glaciation. J Quaternary Sci *21*, 131-153.

S92. Ivy-Ochs, S., Kerschner, H., Kubik, P.W., and Schluchter, C. (2006). Glacier response in the European Alps to Heinrich Event 1 cooling: the Gschnitz stadial. J Quaternary Sci *21*, 115-130.

S93. Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.J., Torroni, A., and Richards, M.B. (2010). The archaeogenetics of Europe. Curr Biol *20*, R174-183.

S94. Gamble, C., Davies, W., Pettitt, P., Hazelwood, L., and Richards, M. (2005). The Archaeological and Genetic Foundations of the European Population during the Late Glacial: Implications for. Cambridge Archaeological Journal *15*, 193.

S95. Clark, C.D., Evans, D.J.A., Khatwa, A., Bradwell, T., Jordan, C.J., Marsh, S.H., Mitchell, W.A., and Bateman, M.D. (2004). Map and GIS database of glacial landforms and features related to the last British Ice Sheet. Boreas *33*, 359-375.

S96. Weaver, A.J., Saenko, O.A., Clark, P.U., and Mitrovica, J.X. (2003). Meltwater pulse 1A from Antarctica as a trigger of the bolling-allerod warm interval. Science *299*, 1709-1713.

S97. Gaffney, V.L., Thomson, K., and Fitch, S. (2007). Mapping Doggerland: the Mesolithic landscapes of the southern North Sea, (Oxford: Archaeopress).

S98. Heiri, O., Brooks, S.J., Renssen, H., Bedford, A., Hazekamp, M., Ilyashuk, B., Jeffers, E.S., Lang, B., Kirilova, E., Kuiper, S., et al. (2014). Validation of climate model-inferred regional temperature change for late-glacial Europe. Nat Commun *5*, 4914.

S99. Stewart, J.R., and Lister, A.M. (2001). Cryptic northern refugia and the origins of the modern biota. Trends Ecol Evol *16*, 608-613.

S100. Stuart, A.J., and Lister, A.M. (2012). Extinction chronology of the woolly rhinoceros Coelodonta antiquitatis in the context of late Quaternary megafaunal extinctions in northern Eurasia. Quaternary Science Reviews *51*, 1-17.

S101. Cooper, A., Turney, C., Hughen, K.A., Brook, B.W., McDonald, H.G., and Bradshaw, C.J. (2015). PALEOECOLOGY. Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. Science *349*, 602-606.

S102. Severinghaus, J.P., and Brook, E.J. (1999). Abrupt climate change at the end of the last glacial period inferred from trapped air in polar Ice. Science *286*, 930-934.

ARTICLE

doi:10.1038/nature17993

The genetic history of Ice Age Europe

Qiaomei Fu[1,2,3], Cosimo Posth[4,5]*, Mateja Hajdinjak[3]*, Martin Petr[3], Swapan Mallick[2,6,7], Daniel Fernandes[8,9], Anja Furtwängler[4], Wolfgang Haak[5,10], Matthias Meyer[3], Alissa Mittnik[4,5], Birgit Nickel[3], Alexander Peltzer[4], Nadin Rohland[2], Viviane Slon[3], Sahra Talamo[11], Iosif Lazaridis[2], Mark Lipson[2], Iain Mathieson[2], Stephan Schiffels[5], Pontus Skoglund[2], Anatoly P. Derevianko[12,13], Nikolai Drozdov[12], Vyacheslav Slavinsky[12], Alexander Tsybankov[12], Renata Grifoni Cremonesi[14], Francesco Mallegni[15], Bernard Gély[16], Eligio Vacca[17], Manuel R. González Morales[18], Lawrence G. Straus[18,19], Christine Neugebauer-Maresch[20], Maria Teschler-Nicola[21,22], Silviu Constantin[23], Oana Teodora Moldovan[24], Stefano Benazzi[11,25], Marco Peresani[26], Donato Coppola[27,28], Martina Lari[29], Stefano Ricci[30], Annamaria Ronchitelli[30], Frédérique Valentin[31], Corinne Thevenet[32], Kurt Wehrberger[33], Dan Grigorescu[34], Hélène Rougier[35], Isabelle Crevecoeur[36], Damien Flas[37], Patrick Semal[38], Marcello A. Mannino[11,39], Christophe Cupillard[40,41], Hervé Bocherens[42,43], Nicholas J. Conard[43,44], Katerina Harvati[43,45], Vyacheslav Moiseyev[46], Dorothée G. Drucker[42], Jiří Svoboda[47,48], Michael P. Richards[11,49], David Caramelli[29], Ron Pinhasi[8], Janet Kelso[3], Nick Patterson[6], Johannes Krause[4,5,43]§, Svante Pääbo[3]§ & David Reich[2,6,7]§

Modern humans arrived in Europe ~45,000 years ago, but little is known about their genetic composition before the start of farming ~8,500 years ago. Here we analyse genome-wide data from 51 Eurasians from ~45,000–7,000 years ago. Over this time, the proportion of Neanderthal DNA decreased from 3–6% to around 2%, consistent with natural selection against Neanderthal variants in modern humans. Whereas there is no evidence of the earliest modern humans in Europe contributing to the genetic composition of present-day Europeans, all individuals between ~37,000 and ~14,000 years ago descended from a single founder population which forms part of the ancestry of present-day Europeans. An ~35,000-year-old individual from northwest Europe represents an early branch of this founder population which was then displaced across a broad region, before reappearing in southwest Europe at the height of the last Ice Age ~19,000 years ago. During the major warming period after ~14,000 years ago, a genetic component related to present-day Near Easterners became widespread in Europe. These results document how population turnover and migration have been recurring themes of European prehistory.

Modern humans arrived in Europe around 45,000 years ago and have lived there ever since, even during the Last Glacial Maximum 25,000–19,000 years ago when large parts of Europe were covered in ice[1]. A major question is how climatic fluctuations influenced the population history of Europe and to what extent changes in material cultures documented by archaeology corresponded to movements of people. To date, it has been difficult to address this question because genome-wide ancient DNA has been retrieved from just four Upper Palaeolithic individuals from Europe[2–4]. Here we assemble and analyse genome-wide data from 51 modern humans dating from 45,000 to 7,000 years ago (Extended Data Table 1; Supplementary Information section 1).

Ancient DNA retrieval

We extracted DNA from human remains in dedicated clean rooms[5], and transformed the extracts into Illumina sequencing libraries[6–8]. A major challenge in ancient DNA research is that the vast majority

[1]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [4]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, 72070 Tübingen, Germany. [5]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [6]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [7]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [8]School of Archaeology and Earth Institute, University College Dublin, Belfield, Dublin 4, Ireland. [9]CIAS, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal. [10]Australian Centre for Ancient DNA, School of Biological Sciences, The University of Adelaide, SA-5005 Adelaide, Australia. [11]Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [12]Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, 17 Novosibirsk, RU-630090, Russia. [13]Altai State University, Barnaul, RU-656049, Russia. [14]Dipartimento di Civiltà e Forme del Sapere, Università di Pisa, 56126 Pisa, Italy. [15]Department of Biology, University of Pisa, 56126 Pisa, Italy. [16]Direction régionale des affaires culturelles Rhône-Alpes, 69283 Lyon, Cedex 01, France. [17]Dipartimento di Biologia, Università degli Studi di Bari 'Aldo Moro', 70125 Bari, Italy. [18]Instituto Internacional de Investigaciones Prehistóricas, Universidad de Cantabria, 39005 Santander, Spain. [19]Department of Anthropology, MSC01 1040, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA. [20]Quaternary Archaeology, Institute for Oriental and European Archaeology, Austrian Academy of Sciences, 1010 Vienna, Austria. [21]Department of Anthropology, Natural History Museum Vienna, 1010 Vienna, Austria. [22]Department of Anthropology, University of Vienna, 1090 Vienna, Austria. [23]"Emil Racoviță" Institute of Speleology, 010986 Bucharest 12, Romania. [24]"Emil Racoviță" Institute of Speleology, Cluj Branch, 400006 Cluj, Romania. [25]Department of Cultural Heritage, University of Bologna, 48121 Ravenna, Italy. [26]Sezione di Scienze Preistoriche e Antropologiche, Dipartimento di Studi Umanistici, Università di Ferrara, 44100 Ferrara, Italy. [27]Università degli Studi di Bari 'Aldo Moro', 70125 Bari, Italy. [28]Museo di "Civiltà preclassiche della Murgia meridionale", 72017 Ostuni, Italy. [29]Dipartimento di Biologia, Università di Firenze, 50122 Florence, Italy. [30]Dipartimento di Scienze Fisiche, della Terra e dell'Ambiente, U.R. Preistoria e Antropologia, Università degli Studi di Siena, 53100 Siena, Italy. [31]CNRS/UMR 7041 ArScAn MAE, 92023 Nanterre, France. [32]INRAP/UMR 8215 Trajectoires 21, 92023 Nanterre, France. [33]Ulmer Museum, 89073 Ulm, Germany. [34]University of Bucharest, Faculty of Geology and Geophysics, Department of Geology, 01041 Bucharest, Romania. [35]Department of Anthropology, California State University Northridge, Northridge, California 91330-8244, USA. [36]Université de Bordeaux, CNRS, UMR 5199-PACEA, 33615 Pessac Cedex, France. [37]TRACES – UMR 5608, Université Toulouse Jean Jaurès, Maison de la Recherche, 31058 Toulouse Cedex 9, France. [38]Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium. [39]Department of Archaeology, School of Culture and Society, Aarhus University, 8270 Højbjerg, Denmark. [40]Service Régional d'Archéologie de Franche-Comté, 25043 Besançon Cedex, France. [41]Laboratoire Chronoenvironnement, UMR 6249 du CNRS, UFR des Sciences et Techniques, 25030 Besançon Cedex, France. [42]Department of Geosciences, Biogeology, University of Tübingen, 72074 Tübingen, Germany. [43]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72072 Tübingen, Germany. [44]Department of Early Prehistory and Quaternary Ecology, University of Tübingen, 72070 Tübingen, Germany. [45]Institute for Archaeological Sciences, Paleoanthropology, University of Tübingen, 72070 Tübingen, Germany. [46]Museum of Anthropology and Ethnography, Saint Petersburg 34, Russia. [47]Department of Anthropology, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic. [48]Institute of Archaeology at Brno, Academy of Science of the Czech Republic, 69129 Dolní Věstonice, Czech Republic. [49]Department of Archaeology, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

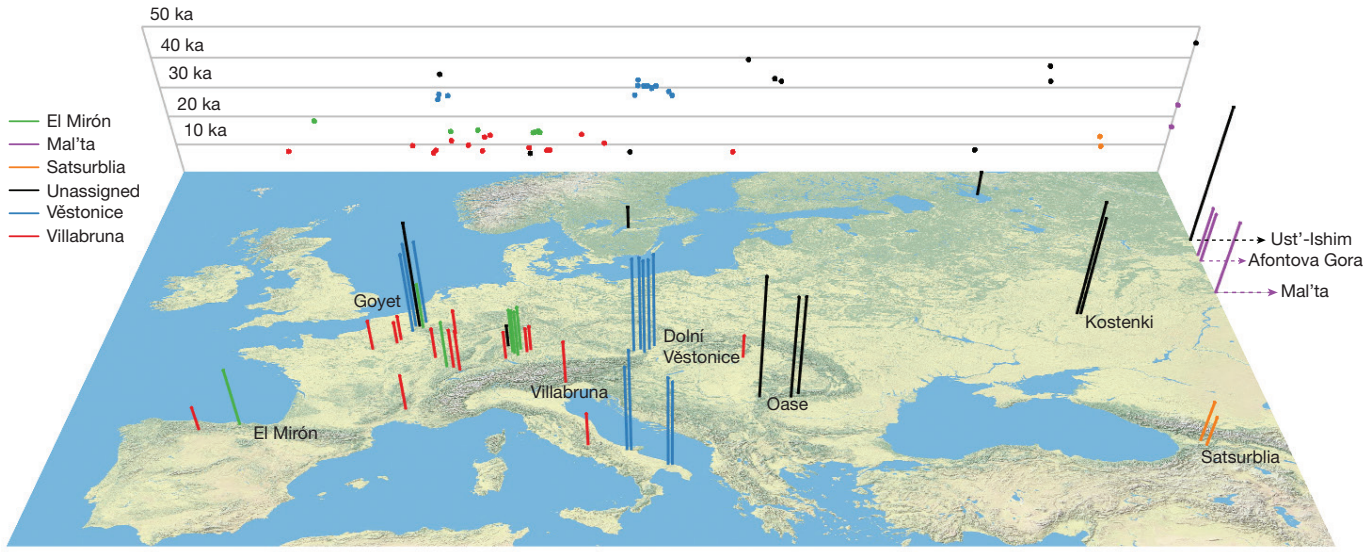


Figure 1 | Location and age of the 51 ancient modern humans. Each bar corresponds to an individual, the colour code designates the genetically defined cluster of individuals, and the height is proportional to age (the background grid shows a projection of longitude against age). To help in visualization, we add jitter for sites with multiple individuals from nearby locations. Four individuals from Siberia are plotted at the far eastern edge of the map. ka, thousand years ago.

of the DNA extracted from most specimens is of microbial origin, making random shotgun sequencing prohibitively expensive. We addressed this problem by enriching the libraries for between 390,000 and 3.7 million single nucleotide polymorphisms (SNPs) in the nuclear genome via hybridizing to pools of previously synthesized 52-base-pair oligonucleotide probes targeting these positions. This makes it possible to generate genome-wide data from samples with high percentages of microbial DNA that are not practical to study by shotgun sequencing[3,9]. We sequenced the isolated DNA fragments from both ends, and mapped the consensus sequences to the human genome (hg19), retaining fragments that overlapped the targeted SNPs. After removing fragments with identical start and end positions to eliminate duplicates produced during library amplification, we chose one fragment at random to represent each individual at each SNP.

Contamination from present-day human DNA is a danger in ancient DNA research. To address this, we took advantage of three characteristic features of ancient DNA (Supplementary Information section 2). First, for an uncontaminated specimen, we expect only a single mitochondrial DNA sequence to be present, allowing us to detect contamination as a mixture of mitochondrial sequences. Second, because males carry a single X chromosome, we can detect contamination in male specimens as polymorphisms on chromosome X[10]. Third, cytosines at the ends of genuine ancient DNA molecules are often deaminated, resulting in apparent cytosine to thymine substitutions[11], and thus we can filter out contaminating molecules by restricting analysis to those with evidence of such deamination[12]. For libraries from males with evidence of mitochondrial DNA contamination or X chromosomal contamination estimates >2.5%—as well as for all libraries from females—we restricted the analyses to sequences with evidence of cytosine deamination (Supplementary Information section 2). After merging libraries from the same individual and limiting to individuals with >4,000 targeted SNPs covered at least once, 38 individuals remained, which we merged with newly generated shotgun sequencing data from the *Karelia* individual[9] (2.0-fold coverage), and published data from ancient[2–4,7,13–19] and present-day humans[20]. The final data set includes 51 ancient modern humans, of which 16 had at least 790,000 SNPs covered (Fig. 1; Extended Data Table 1).

Natural selection reduced Neanderthal ancestry over time

We used two previously published statistics[3,7,21] to test if the proportion of Neanderthal ancestry in Eurasians changed over the last 45,000 years. Whereas on the order of 2% of present-day Eurasian DNA is of Neanderthal origin (Extended Data Table 2), the ancient modern human genomes carry significantly more Neanderthal DNA (Fig. 2) ($P \ll 10^{-12}$). Using one statistic, we estimate a decline from 4.3–5.7% from a time shortly after introgression to 1.1–2.2% in Eurasians today (Fig. 2). Using the other statistic, we estimate a decline from 3.2–4.2% to 1.8–2.3% (Extended Data Fig. 1 and Extended Data Table 3). Because all of the European individuals we analysed dating to between 37,000 and 14,000 years ago are consistent with descent from a single founding

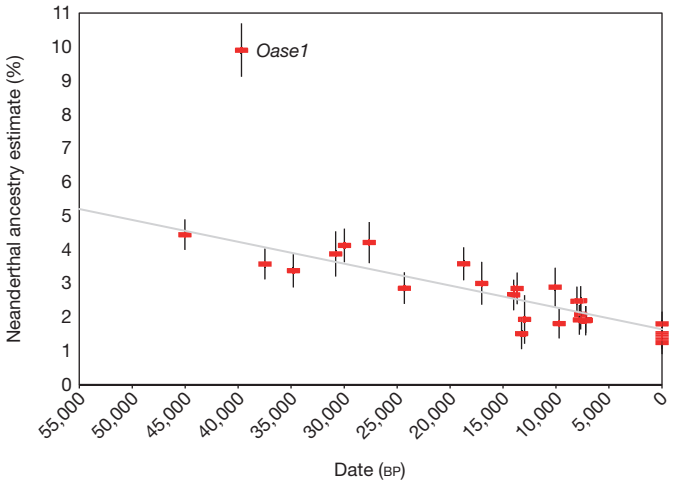


Figure 2 | Decrease of Neanderthal ancestry over time. Plot of radiocarbon date against Neanderthal ancestry for individuals with at least 200,000 SNPs covered, along with present-day Eurasians (standard errors are from a block jackknife). The least squares fit (grey) excludes the data from *Oase1* (an outlier with recent Neanderthal ancestry) and three present-day European populations (known to have less Neanderthal ancestry than east Asians). The slope is significantly negative for all eleven subsets of individuals we analysed ($10^{-29} < P < 10^{-11}$ based on a block jackknife) (Extended Data Table 3). BP, before present.

population, admixture with populations with lower Neanderthal ancestry cannot explain the steady decrease in Neanderthal-derived DNA that we detect during this period, showing that natural selection against Neanderthal DNA must have driven this phenomenon (Fig. 2). We also obtained an independent line of evidence for selection from our observation that the decrease in Neanderthal-derived alleles is more marked near genes than in less constrained regions of the genome ($P = 0.010$) (Extended Data Table 3; Supplementary Information section 3)[22–25].

Chromosome Y, mtDNA, and significant mutations

We used the proportion of sequences mapping to the Y chromosome to infer sex (Extended Data Table 4; Supplementary Information section 4), and determined Y chromosome haplogroups for the males. We were surprised to find haplogroup R1b in the ~14,000-year-old *Villabruna* individual from Italy. While the predominance of R1b in western Europe today owes its origin to Bronze Age migrations from the eastern European steppe[9], its presence in *Villabruna* and in a ~7,000-year-old farmer from Iberia[9] documents a deeper history of this haplotype in more western parts of Europe. Additional evidence of an early link between West and East comes from the *HERC2* locus, where a derived allele that is the primary driver of light eye colour in Europeans appears nearly simultaneously in specimens from Italy and the Caucasus ~14,000–13,000 years ago. Extended Data Table 5 presents results for additional alleles of biological importance. When analysing the mitochondrial genomes we noted the presence of haplogroup M in a ~27,000-year-old individual from southern Italy (*Ostuni1*) in agreement with the observation that this haplogroup, which today occurs in Asia and is absent in Europe, was present in pre-Last Glacial Maximum Europe and was subsequently lost[26]. We also find that the ~33,000-year-old *Muierii2* from Romania carries a basal version of haplogroup U6, in agreement with the hypothesis that the presence of derived versions of this haplogroup in North Africans today is due to back-migration from western Eurasia[27].

Genetic clustering of the ancient specimens

This data set provides an unprecedented opportunity to study the population history of Upper Palaeolithic Europe over more than 30,000 years. In order not to prejudice any association between genetic and archaeological groupings among the individuals studied, we first allowed the genetic data alone to drive the groupings of the specimens, and only afterward examined their associations with archaeological cultural complexes. We began by computing f_3-statistics[14] of the form $f_3(X, Y; Mbuti)$, which measure shared genetic drift between a pair of ancient individuals after divergence from an outgroup (here *Mbuti* from sub-Saharan Africa) (Fig. 3a and Extended Data Fig. 2). Through multi-dimensional scaling (MDS) analysis of this matrix (Fig. 3b), as well as through *D*-statistic analyses[28] (Supplementary Information section 5), we identify five clusters of individuals who share substantial amounts of genetic drift. We name these clusters after the oldest individual in each cluster with >1.0-fold coverage (Supplementary Information section 5; Extended Data Table 1). In contrast, we were not able to identify clear structure among the individuals studied based on model-based clustering[29,30], which may reflect the fact that many of the individuals are so ancient that present-day human variation is not very relevant to understanding their patterns of genetic differentiation[4,13]. The 'Věstonice Cluster' is composed of 14 pre-Last

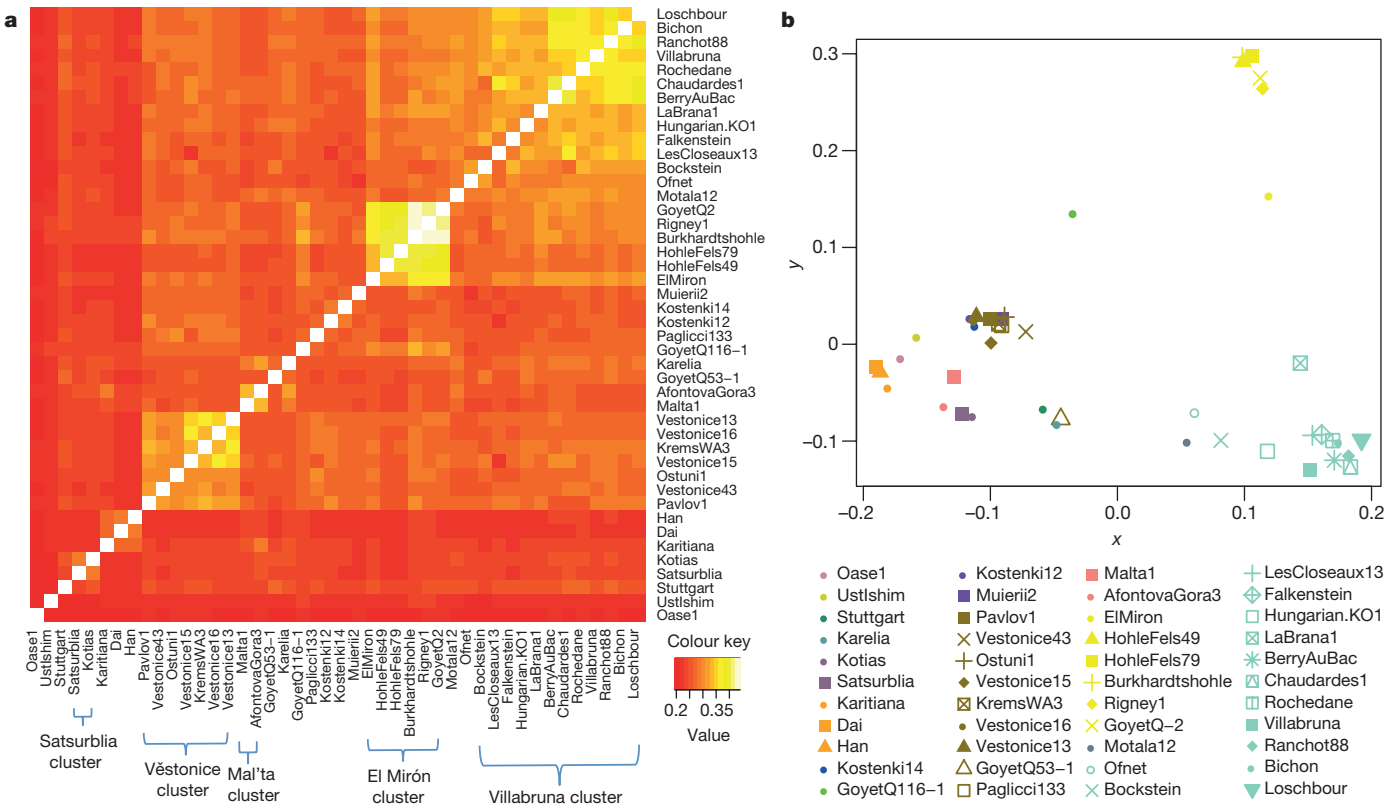


Figure 3 | Genetic clustering of the ancient modern humans. a, Shared genetic drift measured by $f_3(X,Y; Mbuti)$ among individuals with at least 30,000 SNPs covered (for *AfontovaGora3, ElMiron, Falkenstein, GoyetQ-2, GoyetQ53-1, HohleFels49, HohleFels79, LesCloseaux13, Ofnet, Ranchot88* and *Rigney1*, we use all sequences for higher resolution). Lighter colours indicate more shared drift. **b**, Multi-dimensional scaling (MDS) analysis, computed using the R software cmdscale package, highlights the main genetic groupings analysed in this study: Věstonice Cluster (brown), Mal'ta Cluster (pink), El Mirón Cluster (yellow), Villabruna Cluster (light green), and Satsurblia Cluster (dark purple). The affinity of *GoyetQ116-1* (dark green) to the El Mirón Cluster is evident in both views of the data.

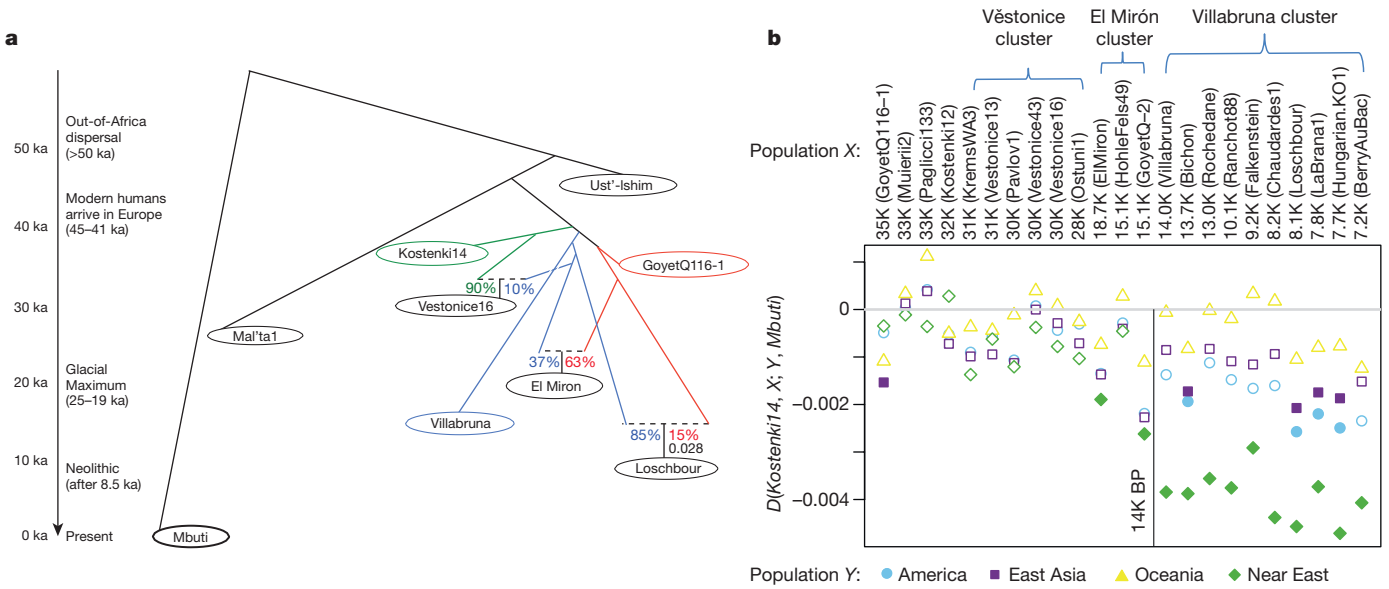


Figure 4 | Population history inferences. a, Admixture graph relating selected high coverage individuals. Dashed lines show inferred admixture events; the estimated mixture proportions fitted using the ADMIXTUREGRAPH software are labelled[28] (the estimated genetic drift on each branch is given in a version of this graph shown in Supplementary Information section 6). The individuals are positioned vertically based on their radiocarbon date, but we caution that the population split times are not accurately known. Colour is used to highlight important early branches of the European founder population: the *Kostenki14* lineage is modelled as the predominant contributor to the Věstonice Cluster (green); the *GoyetQ116-1* lineage as the predominant contributor to the El Mirón Cluster (red); and the *Villabruna* lineage as broadly represented across many clusters. **b**, Drawing together of European and Near Eastern populations ~14,000 years ago. Plot of affinity of each pre-Neolithic European population *X* to non-Africans outside Europe *Y* moving forward in time, comparing to *Kostenki14* as a baseline; values $Z < -3$ standard errors below zero are indicated with filled symbols (we restricted to individuals with >50,000 SNPs). We observe an affinity to Near Easterners beginning with the Villabruna Cluster, and another to east Asians that affects a subset of the Villabruna Cluster.

Glacial Maximum individuals from 34,000–26,000 years ago, who are all associated with the archaeologically defined Gravettian culture. The 'Mal'ta Cluster' is composed of three individuals who lived between 24,000–17,000 years ago from the Lake Baikal region of Siberia. The 'El Mirón Cluster' is composed of seven post-Last Glacial Maximum individuals from 19,000–14,000 years ago, who are all associated with the Magdalenian culture. The 'Villabruna Cluster' is composed of 15 post-Last Glacial Maximum individuals from 14,000–7,000 years ago, associated with the Azilian, Epipaleolithic and Mesolithic cultures. The 'Satsurblia Cluster' is composed of two individuals from 13,000–10,000 years ago from the southern Caucasus[2]. Ten individuals were not assigned to any cluster, either because they represented distinct early lineages (*Ust'-Ishim, Oase1, Kostenki14*, *GoyetQ116-1*, *Muierii2, Cioclovina1* and *Kostenki12*), because they were admixed between clusters (*Karelia* or *Motala12*), or because they were of very different ancestry (*Stuttgart*). To classify the ancestry of additional low coverage individuals, we built an admixture graph that fits the allele frequency correlation patterns among high-coverage individuals[28] (Fig. 4a; Supplementary Information section 6). We fit each low-coverage individual into the graph in turn, using all DNA fragments from these individuals, rather than just fragments with evidence of cytosine deamination, and account for contamination by modelling (Supplementary Information section 7).

A founding population for Europeans 37–14 ka

A previous genetic analysis of early modern humans in Europe using data from the ~37,000-year-old *Kostenki14* suggested that the population to which *Kostenki14* belonged harboured within it the three major lineages that exist in mixed form in Europe today[4,15]: (1) a lineage related to all later pre-Neolithic Europeans, (2) a 'Basal Eurasian' lineage that split from the ancestors of Europeans and east Asians before they separated from each other; and (3) a lineage related to the ~24,000-year-old *Mal'ta1* from Siberia. With our more extensive sampling of Ice Age Europe, we find no support for this. When we test whether the ~45,000-year-old *Ust'-Ishim*—an early Eurasian without any evidence of Basal Eurasian ancestry—shares more alleles with one test individual or another by computing statistics of the form $D(Test_1, Test_2; Ust'\text{-}Ishim, Mbuti)$, we find that the statistic is consistent with zero when the *Test* populations are any pre-Neolithic Europeans or present-day east Asians[3,13]. This would not be expected if some of the pre-Neolithic Europeans, including *Kostenki14,* had Basal Eurasian ancestry (Supplementary Information section 8). We also find no evidence for the suggestion that the *Mal'ta1* lineage contributed to Upper Palaeolithic Europeans[4], because when we compute the statistic $D(Test_1, Test_2; Mal'ta1, Mbuti)$, we find that the statistic is indistinguishable from zero when the *Test* populations are any pre-Neolithic Europeans beginning with *Kostenki14*, consistent with descent from a single founder population since separation from the lineage leading to *Mal'ta1* (Supplementary Information section 9). A corollary of this finding is that the widespread presence of *Mal'ta1*-related ancestry in present-day Europeans[15] is probably explained by migrations from the Eurasian steppe in the Neolithic and Bronze Age periods[9].

Resurfacing of a European lineage in the Glacial Maximum

Among the newly reported individuals, *GoyetQ116-1* from present-day Belgium is the oldest at ~35,000 years ago. This individual is similar to the ~37,000-year-old *Kostenki14* and all later individuals in that it shares more alleles with present-day Europeans (for example, *French*) than with east Asians (for example, *Han*). In contrast, *Ust'-Ishim* and *Oase1*, which predate *GoyetQ116-1* and *Kostenki14,* do not show any distinctive affinity to later Europeans (Extended Data Table 6). Thus, from about 37,000 years ago, populations in Europe shared at least some ancestry with present Europeans. However, *GoyetQ116-1* differs from *Kostenki14* and from all individuals of the succeeding Věstonice Cluster in that both f_3-statistics (Fig. 3; Extended Data Fig. 2) and *D*-statistics show that it shares more alleles with members of the El Mirón Cluster who lived 19,000–14,000 years ago than with other pre-Neolithic Europeans (Supplementary Information section 10). Thus, *GoyetQ116-1* has an affinity to individuals who lived more than 15,000 years later. While at least half of the ancestry of all

El Mirón Cluster individuals comes from the lineage represented by *GoyetQ116-1*, this proportion varies among individuals with the largest amount found outside Iberia ($Z = -4.8$) (Supplementary Information section 10).

Europe and the Near East drew together around 14 ka

Beginning around 14,000 years ago with the Villabruna Cluster, the strong affinity to *GoyetQ116-1* seen in El Mirón Cluster individuals who belong to the Late Glacial Magdalenian culture becomes greatly attenuated (Supplementary Information section 10). To test if this change might reflect gene flow from populations that did not descend from the >37,000-year-old European founder population, we computed statistics of the form *D(Early European, Later European; Y, Mbuti)* where *Y* are various present-day non-Africans. If no gene flow from exogenous populations occurred, this statistic is expected to be zero. Figure 4b shows that it is consistent with zero ($|Z| < 3$) for nearly all individuals dating to between about 37,000 and 14,000 years ago. However, beginning with the Villabruna Cluster, it becomes highly significantly negative in comparisons where the non-European population (*Y*) is Near Easterners (Fig. 4b; Extended Data Fig. 3; Supplementary Information section 11). This must reflect a contribution to the Villabruna Cluster from a lineage also found in present-day Near Easterners (Fig. 4b).

The Satsurblia Cluster individuals from the Caucasus dating to ~13,000–10,000 years ago[2] share more alleles with the Villabruna Cluster individuals than they do with earlier Europeans, indicating that they are related to the population that contributed new alleles to people in the Villabruna Cluster, although they cannot be the direct source of the gene flow. One reason for this is that the Satsurblia Cluster carries large amounts of Basal Eurasian ancestry while Villabruna Cluster individuals do not[2] (Supplementary Information section 12; Extended Data Fig. 4). One possible explanation for the sudden drawing together of the ancestry of Europe and the Near East at this time is long-distance migrations from the Near East into Europe. However, a plausible alternative is population structure, whereby Upper Palaeolithic Europe harboured multiple groups that differed in their relationship to the Near East, with the balance shifting among groups as a result of demographic changes after the Glacial Maximum.

The Villabruna Cluster is represented by the largest number of individuals in this study. This allows us to study heterogeneity within this cluster (Supplementary Information section 13). First, we detect differences in the degree of allele sharing with members of the El Mirón Cluster, as revealed by significant statistics of the form $D(Test_1, Test_2; El\ Mirón\ Cluster, Mbuti)$. Second, we detect an excess of allele sharing with east Asians in a subset of Villabruna Cluster individuals—beginning with an ~13,000-year-old individual from Switzerland—as revealed by significant statistics of the form $D(Test_1, Test_2; Han, Mbuti)$ (Fig. 4b and Extended Data Fig. 3). For example, *Han* Chinese share more alleles with two Villabruna Cluster individuals (*Loschbour* and *LaBrana1*) than they do with *Kostenki14*, as reflected in significantly negative statistics of the form *D(Kostenki14, Loschbour/LaBrana1; Han, Mbuti)*[4]. This statistic was originally interpreted as evidence of Basal Eurasian ancestry in *Kostenki14*. However, because this statistic is consistent with zero when *Han* is replaced with *Ust'-Ishim*, these findings cannot be driven by Basal Eurasian ancestry (as we discuss earlier), and must instead be driven by gene flow between populations related to east Asians and the ancestors of some Europeans (Supplementary Information section 8).

Conclusions

We show that the population history of pre-Neolithic Europe was complex in several respects. First, at least some of the initial modern humans to appear in Eurasia, exemplified by *Ust'-Ishim* and *Oase1*, failed to contribute appreciably to the current European gene pool[3,13]. Only from around 37,000 years ago do all the European individuals analysed share ancestry with present-day Europeans. Second, from the time of *Kostenki14* about 37,000 years ago until the time of the Villabruna Cluster about 14,000 years ago, all individuals seem to derive from a single ancestral population with no evidence of substantial genetic influx from elsewhere. It is interesting that during this time, the Mal'ta Cluster is not represented in any of the individuals we sampled from Europe. Thus, while individuals assigned to the Gravettian cultural complex in Europe are associated with the Věstonice Cluster, there is no genetic connection between them and the *Mal'ta1* individual in Siberia, despite the fact that Venus figurines are associated with both. This suggests that if this similarity is not a coincidence[31], it reflects diffusion of ideas rather than movements of people. Third, we find that *GoyetQ116-1* derives from a different deep branch of the European founder population than the Věstonice Cluster which became predominant in many places in Europe between 34,000 and 26,000 years ago including at Goyet. *GoyetQ116-1* is chronologically associated with the Aurignacian cultural complex. Thus, the subsequent spread of the Věstonice Cluster shows that the diffusion of the Gravettian cultural complex was mediated at least in part by population movements. Fourth, the population represented by *GoyetQ116-1* did not disappear, as its descendants became widespread again after ~19,000 years ago in the El Mirón Cluster when we detect them in Iberia. The El Mirón Cluster is associated with the Magdalenian culture and may represent a post-Glacial Maximum expansion from southwestern European refugia[32]. Fifth, beginning with the Villabruna Cluster at least ~14,000 years ago, all European individuals analysed show an affinity to the Near East. This correlates in time to the Bølling-Allerød interstadial, the first significant warming period after the Glacial Maximum[33]. Archaeologically, it correlates with cultural transitions within the Epigravettian in southern Europe[34] and the Magdalenian-to-Azilian transition in western Europe[35]. Thus, the appearance of the Villabruna Cluster may reflect migrations or population shifts within Europe at the end of the Ice Age, an observation that is also consistent with the evidence of mitochondrial DNA turnover[26,36]. One scenario that could explain these patterns is a population expansion from southeastern European or west Asian refugia after the Glacial Maximum, drawing together the genetic ancestry of Europe and the Near East. Sixth, within the Villabruna Cluster, some, but not all, individuals have an affinity to east Asians. An important direction for future work will be to generate similar ancient DNA data from southeastern Europe and the Near East to arrive at a more complete picture of the Upper Palaeolithic population history of western Eurasia.

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 18 December 2015; accepted 12 April 2016.
Published online 2 May 2016.

1. Gamble, C., Davies, W., Pettitt, P. & Richards, M. Climate change and evolving human diversity in Europe during the last glacial. *Philosoph. Trans. Royal Soc. B* **359,** 243–253 (2004).
2. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Commun.* **6,** 8912 (2015).
3. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
4. Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346,** 1113–1118 (2014).
5. 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).
6. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010,** pdb.prot5448 (2010).
7. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
8. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370,** 20130624 (2015).
9. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
10. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15,** 356 (2014).

11. Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20,** 231–236 (2010).
12. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336,** 466–469 (2012).
13. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514,** 445–449 (2014).
14. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
15. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
16. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507,** 225–228 (2014).
17. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature Commun.* **5,** 5257 (2014).
18. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
19. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
20. Mallick, S. *et al.* The landscape of human genome diversity. *Nature* (in the press).
21. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053–1060 (2010).
22. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507,** 354–357 (2014).
23. Vernot, B. & Akey, J. M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343,** 1017–1021 (2014).
24. Harris, K. & Nielsen, R. The genetic cost of Neanderthal introgression. *Genetics* http://dx.doi.org/10.1534/genetics.116.186890 (2016).
25. Juric, I., Aeschbacher, S. & Coop, G. The strength of selection against Neanderthal introgression. Preprint at http://dx.doi.org/10.1101/030148 (2015).
26. Posth, C. *et al.* Pleistocene mitochondrial genomes suggest a single major dispersal of Non-Africans and a Late Glacial population turnover in Europe. *Curr. Biol.* **26,** 827–833 (2016).
27. Olivieri, A. *et al.* The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* **314,** 1767–1770 (2006).
28. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
29. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).
30. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195,** 693–702 (2013).
31. Conard, N. J. A female figurine from the basal Aurignacian of Hohle Fels Cave in southwestern Germany. *Nature* **459,** 248–252 (2009).
32. Straus, L. G. After the deep freeze: confronting "Magdalenian" realities in Cantabrian Spain and beyond. *J. Archaeol. Method Theory* **20,** 236–255 (2013).
33. Weaver, A. J., Saenko, O. A., Clark, P. U. & Mitrovica, J. X. Meltwater pulse 1A from Antarctica as a trigger of the Bølling-Allerød warm interval. *Science* **299,** 1709–1713 (2003).
34. Montoya, C. & Peresani, M. Nouveaux éléments de diachronie dans l'Epigravettien récent des Préalpes de la Vénétie in *D'un monde à l'autre. Les systèmes lithiques pendant le Tardiglaciaire autour de la Méditerranée nord-occidentale* (eds Bracco, J.-P. & Montoya, C.) 123–138 (Mémoire Société Préhistorique Française, 2005).
35. Valentin, B. *Jalons pour une paléohistoire des derniers chasseurs (XIVe-VIe millénaires av. J.-C.)* (Cahiers archéologiques de Paris 1 – 1, Publications de la Sorbonne, 2008).
36. Pala, M. *et al.* Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am. J. Hum. Genet.* **90,** 915–924 (2012).

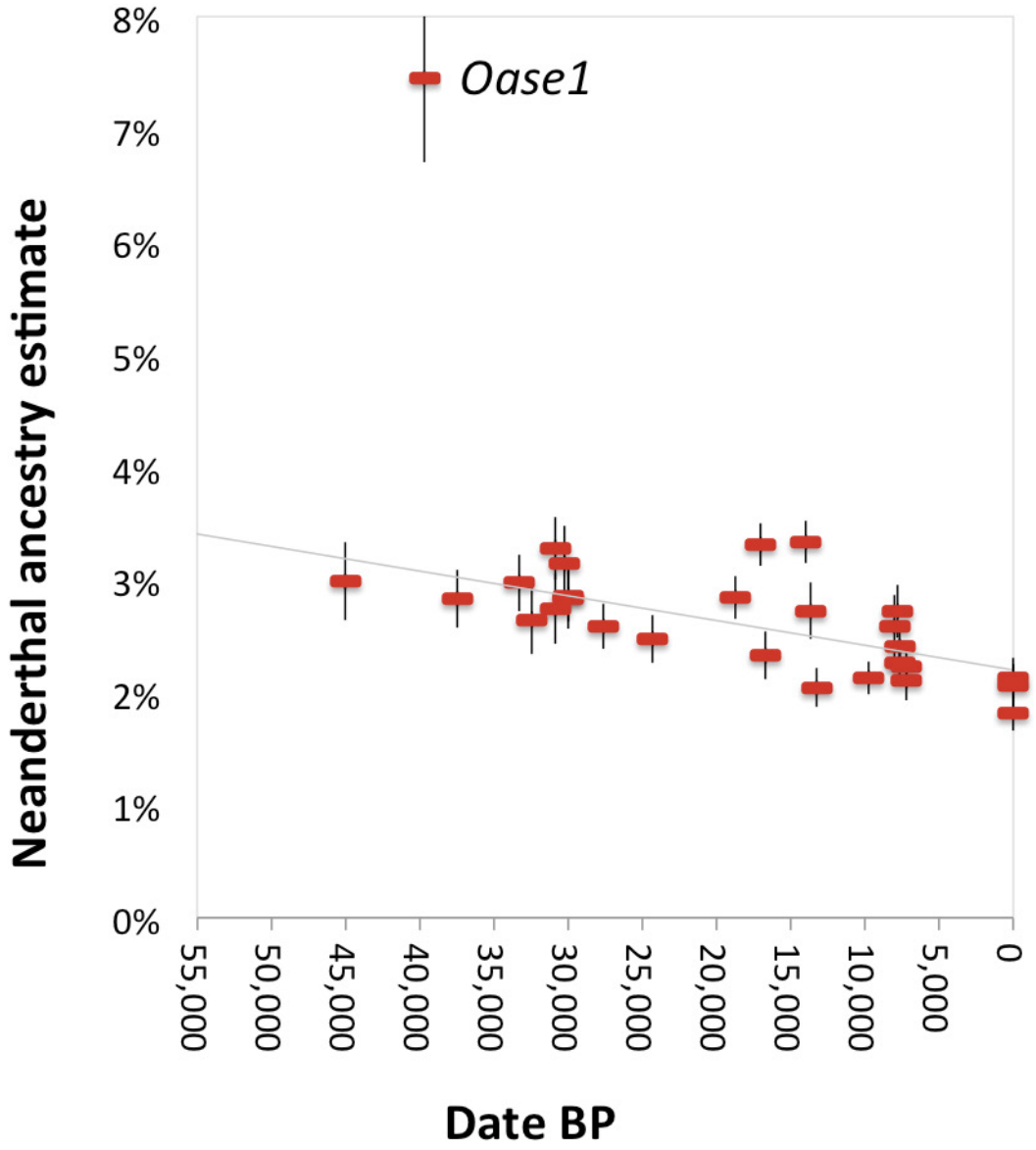
Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Alex, D. Meltzer, P. Moorjani, I. Olalde, S. Sankararaman and B. Viola for comments, K. Stewardson and E. Harney for sample screening, and F. Hallgren for sharing a radiocarbon date for *Motala12*. The Fig. 1 map is plotted using data available under the Open Database License © OpenStreetMap (http://www.openstreetmap.org/copyright). The Goyet project led by H.R. was funded by the Wenner-Gren Foundation (Gr. 7837), the College of Social and Behavioral Sciences of CSUN, the CSUN Competition for Research, Scholarship and Creative Activity Awards, and the RBINS. The excavation of the El Mirón Cave burial, led by L.G.S. and M.R.G.M., was supported by the Gobierno de Cantabria, the L.S.B. Leakey Foundation, the University of New Mexico, the Stone Age Research Fund (J. and R. Auel, principal donors), the town of Ramales de la Victoria and the Universidad de Cantabria. Excavations at Grotta Paglicci were performed by A. Palma di Cesnola in collaboration with the Soprintendenza Archeologia della Puglia (founded by MIUR and local Institutions). Research at Riparo Villabruna was supported by MIBACT and the Veneto Region. Q.F. was funded by the Special Foundation of the President of the Chinese Academy of Sciences (2015–2016), the Bureau of International Cooperation of the Chinese Academy of Sciences, the Chinese Academy of Sciences (XDA05130202), the National Natural Science Foundation of China (L1524016) and the Chinese Academy of Sciences Discipline Development Strategy Project (2015-DX-C-03). D.Fe was supported by an Irish Research Council grant (GOIPG/2013/36). I.M. was supported by a long-term fellowship from the Human Frontier Science Program LT001095/2014-L. P.Sk was supported by the Swedish Research Council (VR 2014-453). S.T., and M.P.R. were funded by the Max Planck Society. C.N.-M. was funded by FWF P-17258, P-19347, P-21660 and P-23612. S.C. and O.T.M. were funded by a 'Karsthives' Grant PCCE 31/2010 (CNCS-UEFISCDI, Romania). A.P.D., N.D., V.Sla and N.D. were funded by the Russian Science Foundation (project No.14-50-00036). M.A.M. was funded by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (grant number PIEF-GA-2008-219965). M.La and D.C. were funded by grants PRIN 2010-11 and 2010EL8TXP_003. C.C. and the research about the French Jura sites of Rochedane, Rigney and Ranchot was funded by the Collective Research Program (PCR) (2005-2008). K.H. was supported by the European Research Council (ERC StG 283503) and the Deutsche Forschungsgemeinschaft (DFG INST37/706-1FUGG, DFG FOR2237). D.G.D. was funded by the European Social Fund and Ministry of Science, Research and Arts of Baden-Württemberg. R.P. was funded by ERC starting grant ADNABIOARC (263441). J.Ke was funded by a grant from the Deutsche Forschungsgemeinschaft (SFB1052, project A02). J.Kr was funded by DFG grant KR 4015/1-1, the Baden Württemberg Foundation, and the Max Planck Society. S.P. were funded by the Max Planck Society and the Krekeler Foundation. D.R. was funded by NSF HOMINID grant BCS-1032255, NIH (NIGMS) grant GM100233, and the Howard Hughes Medical Institute.

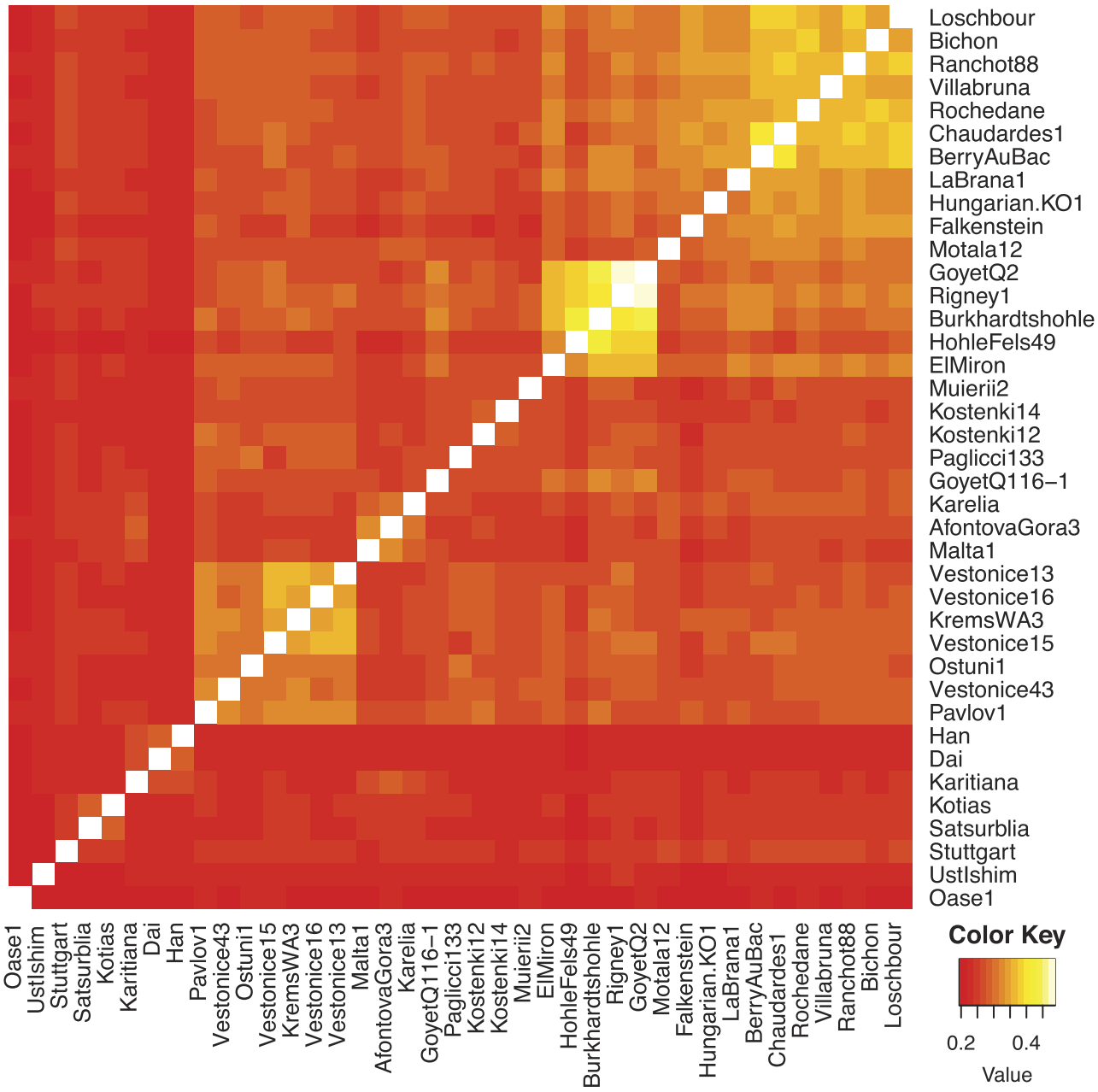
Author Contributions J.Kr, S.P. and D.R. conceived the idea for the study. Q.F., C.P., M.H., W.H., M.Me, V.Slo, R.G.C., A.P.D., N.D., V.Sla, A.T., F.M., B.G., E.V., M.R.G.M., L.G.S., C.N.-M., M.T.-N., S.C., O.T.M., S.B., M.Per, D.Co, M.La, S.R., A.R., F.V., C.T., K.W., D.G., H.R., I.C., D.Fl, P.Se, M.A.M., C.C., H.B., N.J.C., K.H., V.M., D.G.D., J.S., D.Ca, R.P., J.Kr, S.P. and D.R. assembled archaeological material. Q.F., C.P., M.H., D.Fe, A.F., W.H., M.Me, A.M., B.N., N.R., V.Slo, S.T., H.B., D.G.D., M.P.R., R.P., J.Kr, S.P. and D.R. performed or supervised wet laboratory work. Q.F., C.P., M.H., M.Pet, S.M., A.P., I.L., M.Li, I.M., S.S., P.Sk, J.Ke, N.P. and D.R. analysed data. Q.F., C.P., M.H., M.Pet, J.Ke, S.P. and D.R. wrote the manuscript and supplements.

Author Information The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB13123. 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).

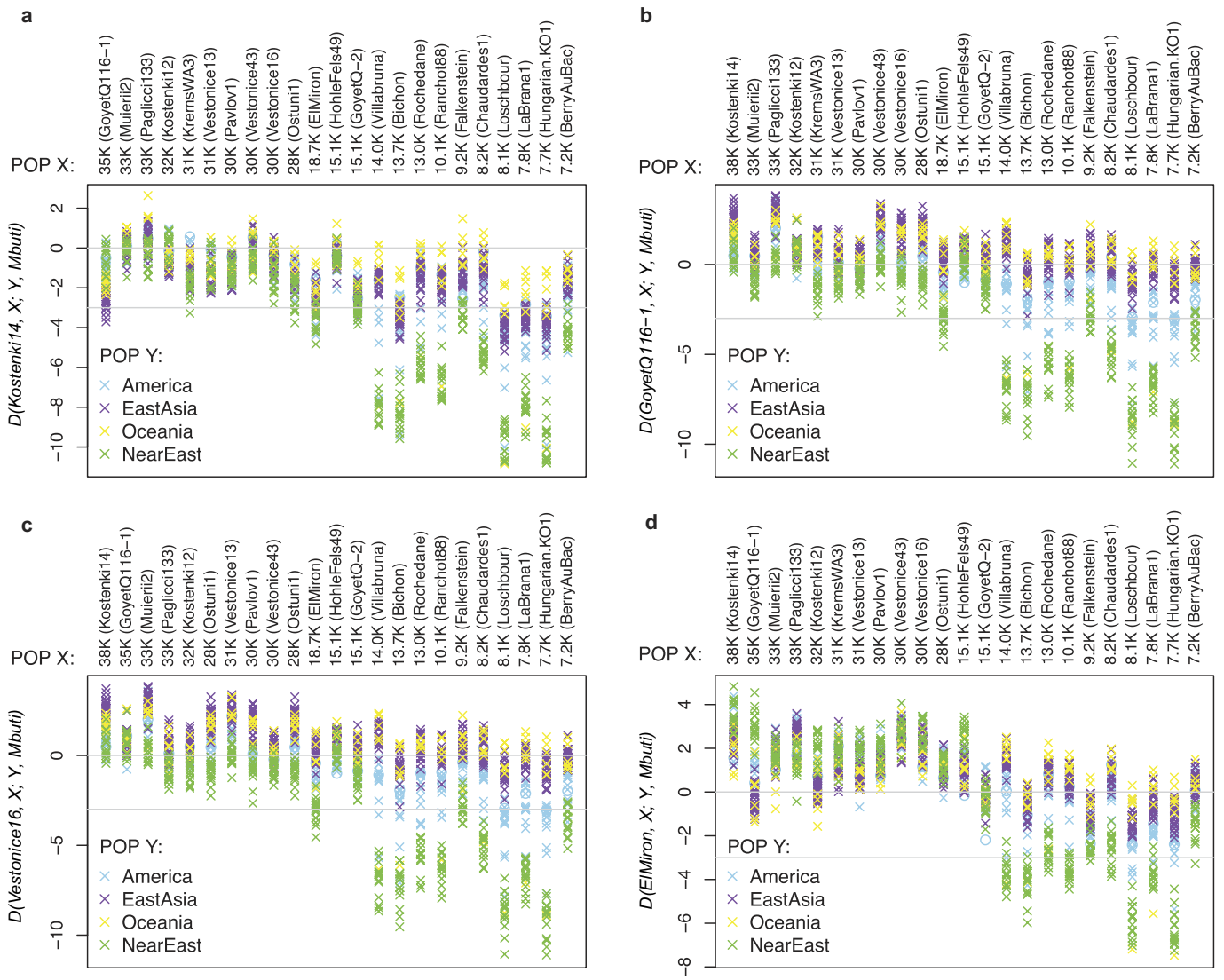
Reviewer Information *Nature* thanks C. Lalueza-Fox and the other anonymous reviewer(s) for their contribution to the peer review of this work.



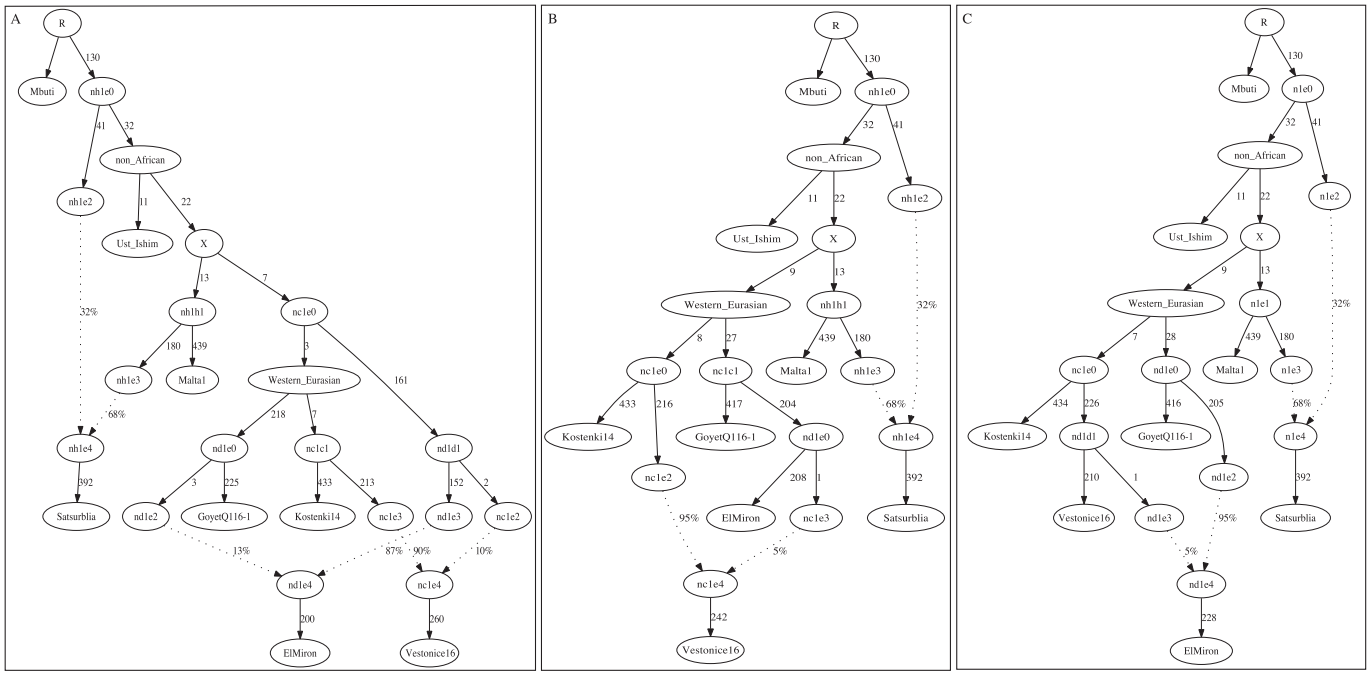
Extended Data Figure 1 | A decrease in Neanderthal ancestry in the last 45,000 years. This is similar to Fig. 2, except we use ancestry estimates from rates of alleles matching to Neanderthal rather than f_4-ratios, as described in Supplementary Information section 3. The least-squares fit excludes *Oase1* (as an outlier with recent Neanderthal ancestry) and Europeans (known to have reduced Neanderthal ancestry). The regression slope is significantly negative ($P = 0.00004$, Extended Data Table 3).



Extended Data Figure 2 | Heat matrix of pairwise $f_3(X, Y; Mbuti)$ for selected ancient individuals. Only individuals with at least 30,000 SNPs covered at least once are analysed.



Extended Data Figure 3 | Studying how the relatedness of non-European populations to pairs of European hunter-gatherers changes over time. Statistics were examined of the form *D(W, X; Y, Mbuti)*, with the *Z*-score given on the *y* axis, where *W* is an early European hunter-gatherer, *X* is another European hunter-gatherer (in chronological order on the *x* axis), and *Y* is a non-European population (see legend). **a**, *W* = *Kostenki14*. **b**, *W* = *GoyetQ116-1*. **c**, *W* = *Vestonice16*. **d**, *W* = *ElMiron*. |Z| > 3 scores are considered statistically significant (horizontal line). The similar Fig. 4b gives absolute *D*-statistic values rather than *Z*-scores (for W = *Kostenki14*) and uses pooled regions rather than individual populations *Y*.



Extended Data Figure 4 | Three admixture graph models that fit the data for *Satsurblia*, an Upper Palaeolithic individual from the Caucasus. These models use 127,057 SNPs covered in all populations. Estimated genetic drifts are given along the solid lines in units of f_2-distance (parts per thousand), and estimated mixture proportions are given along the dotted lines. All three models provide a fit to the allele frequency correlation data among *Mbuti*, *Ust'-Ishim*, *Kostenki14*, *Vestonice16*, *Malta1*, *ElMiron* and *Satsurblia* to within the limits of our resolution, in the sense that all empirical f_2-, f_3- and f_4-statistics relating the individuals are within three standard errors of the expectation of the model. Models in which *Satsurblia* is treated as unadmixed cannot be fit.

Extended Data Table 1 | The 51 ancient modern humans analysed in this study

Sample Code	Data source	Country	Lat.	Long.	Cal BP 95.4%	Date type (ref.)	Culture	Remain	SNP Panel	Sex	mtDNA haplogroup	Y chrom. haplogroup	Genetic Cluster	Damage restrict	Mean coverage+	SNPs covered
UstIshim	[13]	Russia	57.43	71.10	47,480-42,560	Direct-UF ([13])	Unassigned	Femur	Shotgun	M	R	K (xLT)	Unassigned	No	42	2,137,615
Oase1	[3]	Romania	45.12	21.90	41,640-37,580	Direct-UF ([37])	Unassigned	Mandible	Shotgun	M	N	F	Unassigned	Yes	0.156	285,076
Kostenki14*	New	Russia	51.23	39.30	38,680-36,260	Direct-UF ([38])	Unassigned	Tibia	3.7M	M	U2	C1b	Unassigned	No	16.1	1,774,156
GoyetQ116-1	New	Belgium	50.26	4.28	35,160-34,430	Direct-NotUF ([26])	Aurignacian	Humerus	1240k	M	M	C1a	Unassigned	No	1.046	846,983
Muierii2	New	Romania	45.11	23.46	33,760-32,840	Direct-UF ([39])	Unassigned	Temporal	3.7M	F	U6		Unassigned	Yes	0.049	98,618
Paglicci133	New	Italy	41.65	15.61	34,580-31,210	Layer ([40])	Gravettian	Tooth	1240k	M	U8c	I	Věstonice	No	0.041	82,330
Cioclovina1	New	Romania	45.35	23.84	33,090-31,780	Direct-UF ([41])	Unassigned	Cranium	1240k	M	U	CT	Unassigned	Yes	0.006	12,784
Kostenki12	New	Russia	51.23	39.30	32,990-31,840	Layer ([42])	Unassigned	Cranium	3.7M	M	U2	CT	Unassigned	No	0.03	61,228
KremsWA3	New	Austria	48.41	15.59	31,250-30,690	Layer ([43])	Gravettian	Cranium	1240K	M	U5		Věstonice	No	0.11	203,986
Vestonice13	New	Czech Republic	48.53	16.39	31,070-30,670	Layer ([44])	Gravettian	Femur	3.7M	M	U8c	CT(notIJK)	Věstonice	Yes	0.071	139,568
Vestonice15	New	Czech Republic	48.53	16.39	31,070-30,670	Layer ([44])	Gravettian	Femur	3.7M	M	U5	BT	Věstonice	Yes	0.015	30,900
Vestonice14	New	Czech Republic	48.53	16.39	31,070-30,670	Layer ([44])	Gravettian	Femur	390k	M	U		Věstonice	Yes	0.003	5,677
Pavlov1	New	Czech Republic	48.53	16.39	31,110-29,410	Layer ([44])	Gravettian	Femur	3.7M	M	U5	C1a2	Věstonice	Yes	0.028	57,005
Vestonice43	New	Czech Republic	48.53	16.39	30,710-29,310	Layer ([44])	Gravettian	Femur	3.7M	M	U	F	Věstonice	Yes	0.087	163,946
Vestonice16	New	Czech Republic	48.53	16.39	30,710-29,310	Layer ([44])	Gravettian	Femur	3.7M	M	U5	IJK	Věstonice	No	1.31	945,292
Ostuni2	New	Italy	40.73	17.57	29,310-28,640	Direct-UF (New)	Gravettian	Femur	3.7M	F	U2		Věstonice	Yes	0.008	17,017
GoyetQ53-1	New	Belgium	50.26	4.28	28,230-27,720	Direct-NotUF ([26])	Gravettian	Fibula	1240k	F	U2		Věstonice	Yes	0.006	12,567
Paglicci108	New	Italy	41.65	15.61	28,430-27,070	Layer ([45])	Gravettian	Phalanx	1240k	F	U2'3'4'7'8'9		Věstonice	Yes	0.002	4,330
Ostuni1	New	Italy	40.73	17.57	27,810-27,430	Direct-UF (New)	Gravettian	Tibia	3.7M	F	M		Věstonice	Yes	0.245	369,313
GoyetQ376-19	New	Belgium	50.26	4.28	27,720-27,310	Direct-NotUF ([26])	Gravettian	Humerus	1240k	F	U2		Věstonice	Yes	0.012	25,400
GoyetQ56-16	New	Belgium	50.26	4.28	26,600-26,040	Direct-NotUF ([26])	Gravettian	Fibula	1240k	F	U2		Věstonice	Yes	0.005	9,988
Malta1	[14]	Russia	52.9	103.5	24,520-24,090	Direct-UF ([14])	Unassigned	Humerus	Shotgun	M	U	R	Mal'ta	No	1.174	1439501
ElMiron	New	Spain	43.26	-3.45	18,830-18,610	Direct-UF ([26])	Magdalenian	Toe	3.7M	F	U5b		El Mirón	Yes	1.012	797,714
AfontovaGora3	New	Russia	56.05	92.87	16,930-16,490	Layer ([26])	Unassigned	Tooth	3.7M	F	R1b		Mal'ta	Yes	0.17	286,355
AfontovaGora2	[14]	Russia	56.05	92.87	16,930-16,490	Direct-UF ([14])	Unassigned	Humerus	Shotgun	M			Mal'ta	No	0.071	143,751
Rigney1	New	France	47.23	6.10	15,690-15,240	Direct-NotUF ([46])	Magdalenian	Mandible	1240k	F	U2'3'4'7'8'9		El Mirón	Yes	0.017	35,600
HohleFels49	New	Germany	48.22	9.45	16,000-14,260	Layer ([47])	Magdalenian	Femur	390k	M	U8a	I	El Mirón	Yes	0.033	63,151
GoyetQ-2	New	Belgium	50.26	4.28	15,230-14,780	Direct-NotUF ([26])	Magdalenian	Humerus	1240k	M	U8a	HIJK	El Mirón	Yes	0.035	72,263
Brillenhohle	New	Germany	48.24	9.46	15,120-14,440	Direct-UF ([48])	Magdalenian	Cranium	390k	M	U8a		El Mirón	Yes	0.006	13,459
HohleFels79	New	Germany	48.22	9.45	15,070-14,270	Direct-UF ([39])	Magdalenian	Cranium	390k	M	U8a		El Mirón	Yes	0.005	11,211
Burkhardtshohle	New	Germany	48.32	9.35	15,080-14,150	Direct-NotUF ([49])	Magdalenian	Cranium	1240k	M	U8a	I	El Mirón	Yes	0.018	38,376
Villabruna	New	Italy	46.15	12.21	14,180-13,780	Direct-UF ([50])	Epigravettian	Femur	3.7M	M	U5b2b	R1b1	Villabruna	No	3.137	1,215,433
Bichon	[2]	Switzerland	47.01	6.79	13,770-13,560	Direct-UF ([2])	Azilian	Petrous	Shotgun	M	U5b1h	I2	Villabruna	No	8.119	2,116,782
Satsurblia	[2]	Georgia	42.24	42.92	13,380-13,130	Direct-UF ([2])	Epigravettian	Petrous	Shotgun	M	K3	J2	Satsurblia	No	1.195	1,460,368
Rochedane	New	France	47.21	6.45	13,090-12,830	Direct-NotUF ([51])	Epipaleolithic	Mandible	1240k	M	U5b2b	I	Villabruna	No	0.131	237,390
Iboussieres39	New	France	44.29	4.46	12,040-11,410	Direct-NotUF ([26])	Epipaleolithic	Femur	390k	M	U5b2b		Villabruna	Yes	0.005	9,659
Continenza	New	Italy	41.96	13.54	11,200-10,510	Layer (New)	Mesolithic	Cranium	3.7M	F	U5b1		Villabruna	Yes	0.006	11,717
Ranchot88	New	France	47.91	5.43	10,240-9,930	Direct-NotUF ([26])	Mesolithic	Cranium	1240k	F	U5b1		Villabruna	Yes	0.322	414,863
LesCloseaux13	New	France	48.52	2.11	10,240-9,560	Direct-NotUF ([52])	Mesolithic	Femur	1240k	F	U5a2		Villabruna	Yes	0.004	8,635
Kotias	[2]	Georgia	42.13	43.12	9,890-9,550	Direct-UF ([2])	Mesolithic	Tooth	Shotgun	M	H13c	J	Satsurblia	No	12.157	2,133,968
Falkenstein	New	Germany	48.06	9.04	9,410-8,990	Direct-NotUF ([53])	Mesolithic	Fibula	390k	M	U5a2c	F	Villabruna	Yes	0.033	64,428
Karelia	[9]	Russia	61.65	35.65	8,800-7,950	Layer ([54])	Mesolithic	Tooth	Shotgun	M	C1g	R1a1	Unassigned	No	1.952	1,754,410
Bockstein	New	Germany	48.33	10.09	8,370-8,160	Layer ([55])	Mesolithic	Tooth	390k	F	U5b1d1		Villabruna	Yes	0.011	21,977
Ofnet	New	Germany	48.49	10.27	8,430-8,060	Layer ([56])	Mesolithic	Tooth	390k	F	U5b1d1		Villabruna	Yes	0.003	6,263
Chaudardes1	New	France	49.24	3.46	8,360-8,050	Direct-NotUF ([26])	Mesolithic	Tibia	1240k	M	U5b1b	I	Villabruna	Yes	0.046	92,657
Loschbour	[15]	Luxembourg	49.70	6.24	8,160-7,940	Direct-UF ([15])	Mesolithic	Tooth	Shotgun	M	U5b1a	I2a1b	Villabruna	No	20	2,091,584
LaBrana1	[16]	Spain	42.93	-5.35	7,940-7,690	Direct-UF ([57])	Mesolithic	Tooth	Shotgun	M	U5b2c1	C1a2	Villabruna	No	3.338	1,884,745
Hungarian.KO1	[17]	Hungarian	47.93	21.20	7,730-7,590	Direct-UF ([17])	Neolithic	Petrous	Shotgun	M	R3	I2a	Villabruna	No	1.1	1,410,303
Motala12	[15]	Sweden	58.54	15.05	7,670-7,580	Direct-UF (New)	Mesolithic	Tooth	Shotgun	M	U2e1	I2a1b*	Unassigned	No	2.185	1,874,519
BerryAuBac	New	France	49.24	3.54	7,320-7,170	Direct-NotUF ([26])	Mesolithic	Radius	1240k	M	U5b1a	I	Villabruna	No	0.027	54,690
Stuttgart	[15]	Germany	48.78	9.18	7,260-7,020	Direct-UF (New)	Early Neolithic	Tooth	Shotgun	F	T2c1d1		Unassigned	No	19	2,078,724

Refs 37–57 are cited in this table. All dates are obtained as described in Supplementary Information section 1. When an individual has a direct date from the same skeleton it is marked 'direct' followed by a hyphen to indicate whether the date is obtained by ultrafiltration ('UF') or without ('NotUF'). If the date is from the archaeological layer, the date type is marked as 'layer'. All the dates are calibrated using IntCal13 (ref. 58) and the OxCal4.2 program[59].

*Kostenki14 is represented in most analyses by our newly reported 16.1× capture data, but key analyses were repeated on the previously reported 2.8× shotgun data[4].

+Mean coverage is computed on the 3.7 million SNP targets.

37. Rougier, H. *et al.* Peştera cu Oase 2 and the cranial morphology of early modern Europeans. *Proc. Natl Acad. Sci. USA* **104,** 1165–1170 (2007).
38. Marom, A., McCullagh, J. S. O., Higham, T. F. G., Sinitsyn, A. A. & Hedges, R. E. M. Single amino acid radiocarbon dating of Upper Paleolithic modern humans. *Proc. Natl Acad. Sci. USA* **109,** 6878–6881 (2012).
39. Soficaru, A., Doboş, A. & Trinkaus, E. Early modern humans from the Peştera Muierii, Baia de Fier, Romania. *Proc. Natl Acad. Sci. USA* **103,** 17196–17201 (2006).
40. Palma di Cesnola, A. *Paglicci. L'Aurignaziano e il Gravettiano antico* (Claudio Grenzi, 2004).
41. Soficaru, A., Petrea, C., Doboş, A. & Trinkaus, E. The human cranium from the Peştera Cioclovina Uscată, Romania — context, age, taphonomy, morphology, and paleopathology. *Curr. Anthropol.* **48,** 611–619 (2007).
42. Sinitsyn, A. A. Les sépultures de Kostenki: chronologie, attribution culturelle, rite funéraire in *La spiritualité* (Otte M. (ed.)), Proceedings of UISPP conference, Liège, ERAUL **106,** 237–244 (2004).
43. Simon, U., Haendel, M., Einwoegerer, T. & Neugebauer-Maresch, C. The archaeological record of the Gravettian open air site Krems-Wachtberg. *Quat. Int.* **351,** 5–13 (2014).
44. Trinkaus, E. & Svoboda, J. *Early Modern Human Evolution in Central Europe: The People of Dolní Věstonice and Pavlov,* Vol. 12 (Oxford Univ. Press, 2006).
45. Azzi, C. M., Bigliocca, L. & Piovan, F. Florence radiocarbon dates II. *Radiocarbon* **16,** 10–14 (1974).
46. Cupillard, C. *et al.* Changes in ecosystems, climate and societies in the Jura Mountains between 40 and 8 ka cal BP. *Quat. Int.* **378,** 40–72 (2015).
47. Housley, R. A., Gamble, C. S., Street, M. & Pettitt, P. B. Radiocarbon evidence for the Lateglacial human recolonisation of Northern Europe. *Proc. Prehist. Soc.* **63,** 25–54 (1997).
48. Benazzi, S. *et al.* Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479,** 525–528 10.1038/nature10617 (2011).
49. Simon, U. Die Burkhardtshöhle — eine Magdalénienstation am Nordrand der Schwäbischen Alb, Magisterarbeit (1993).
50. Vercellotti, G., Alciati, G., Richards, M. P. & Formicola, V. The Late Upper Paleolithic skeleton Villabruna 1 (Italy): a source of data on biology and behavior of a 14.000 year-old hunter. *J. Anthropol. Sci.* **86,** 143–163 (2008).
51. Drucker, D. G., Bridault, A., Cupillard, C., Hujic, A. & Bocherens, H. Evolution of habitat and environment of red deer (*Cervus elaphus*) during the Late-glacial and early Holocene in eastern France (French Jura and the western Alps) using multi-isotope analysis ($\delta^{13}C$, $\delta^{15}N$, $\delta^{18}O$, $\delta^{34}S$) of archaeological remains. *Quat. Int.* **245,** 268–278 (2011).
52. Valentin, F. *et al.* Découvertes récentes d'inhumations et d'une incinération datées du Mésolithique en Île-de-France. *Revue Archéologique d'Île-de-France* **1,** 21–42 (2008).
53. Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326,** 137–140 (2009).
54. Price, T. D. & Jacobs, K. Olenii Ostrov — first radiocarbon dates from a major mesolithic cemetery Karelia, USSR. *Antiquity* **64,** 849–853 (1990).
55. Wehrberger, K. "Der Streit ward definitiv beendet..." Eine mesolithische Bestattung aus der Bocksteinhöhle im Lonetal, Alb-Donau-Kreis. Zur Erinnerung an Ludwig Bürger (1844–1898). *Archäologisches Korrespondenzblatt* **30,** 15–31 (2000).
56. Orschiedt, J. Manipulationen an menschlichen Skelettresten. Taphonomische Prozesse, Sekundärbestattungen oder Kannibalismus. Dissertation. Urgeschichtliche Materialhefte 13 (1999).
57. Sánchez-Quinto, F. *et al.* Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol.* **22,** 1494–1499 (2012).
58. Reimer, P. J. *et al.* IntCal13 and marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* **55,** 1869–1887 (2013).
59. Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55,** 720–730 (2013).

Extended Data Table 2 | Estimated proportion of Neanderthal ancestry

		f_4-ratios				Archaic Ancestry Informative SNPs					
Sample Code	**Age BP**	**SNPs**	**Est.**	**95% CI**		**SNPs**	**Est.**	**95% CI**		**Increase in Neanderthal ancestry with B**	**S.E.**
UstIshim	45,020	2,137,615	4.4%	3.6%	5.3%	778,774	3.0%	2.3%	3.7%	-0.9%	1.3%
Oase1	39,610	285,076	9.9%	8.4%	11.4%	59,854	7.5%	6.0%	8.9%	2.5%	1.8%
Kostenki14	37,470	1,774,156	3.6%	2.7%	4.4%	632,748	2.8%	2.3%	3.3%	-1.0%	1.0%
GoyetQ116-1	34,795	846,983	3.4%	2.4%	4.3%						
Muierii2	33,300	98,618	5.2%	3.0%	7.4%	22,189	3.0%	2.5%	3.5%	0.6%	1.1%
Paglicci133	32,895	82,330	4.1%	2.1%	6.0%						
Cioclovina1	32,435	12,784	4.1%	-1.1%	9.3%						
Kostenki12	32,415	61,228	1.9%	-0.7%	4.4%	13,385	2.6%	2.1%	3.2%	1.7%	1.5%
KremsWA3	30,970	203,986	3.9%	2.6%	5.2%						
Vestonice13	30,870	139,568	4.6%	2.6%	6.5%	35,983	3.3%	2.7%	3.8%	0.3%	1.3%
Vestonice15	30,870	30,900	4.3%	0.6%	7.9%	5,855	2.7%	2.1%	3.4%	-1.5%	1.3%
Vestonice14	30,870	5,677	2.6%	-5.9%	11.0%						
Pavlov1	30,260	57,005	4.4%	1.6%	7.1%	9,327	3.1%	2.5%	3.8%	0.7%	1.2%
Vestonice43	30,010	163,946	6.9%	5.2%	8.5%	38,749	2.9%	2.4%	3.3%	0.9%	0.9%
Vestonice16	30,010	945,292	4.1%	3.1%	5.1%	268,157	2.8%	2.3%	3.3%	-0.1%	1.0%
Ostuni2	28,975	17,017	1.6%	-3.2%	6.3%	2,746	2.3%	1.4%	3.1%	1.3%	1.6%
GoyetQ53-1	27,975	12,567	4.8%	-0.7%	10.3%						
Paglicci108	27,750	4,330	3.4%	-6.0%	12.7%						
Ostuni1	27,620	369,313	4.2%	3.0%	5.4%	88,449	2.6%	2.2%	3.0%	0.1%	0.9%
GoyetQ376-19	27,515	25,400	6.5%	2.7%	10.2%						
GoyetQ56-16	26,320	9,988	3.6%	-1.9%	9.1%						
Malta1	24,305	1,439,501	2.9%	1.9%	3.8%	437,187	2.5%	2.1%	2.9%	1.0%	0.8%
ElMiron	18,720	797,714	3.6%	2.6%	4.5%	250,071	2.8%	2.5%	3.2%	0.6%	0.9%
AfontovaGora3	16,710	286,355	3.0%	1.8%	4.2%	96,237	3.3%	2.9%	3.7%	-1.5%	1.0%
AfontovaGora2	16,710	143,751	2.2%	0.4%	4.0%	37,280	2.3%	1.9%	2.7%	-0.3%	0.9%
Rigney1	15,465	35,600	0.8%	-2.6%	4.2%						
HohleFels49	15,130	63,151	2.3%	-0.6%	5.2%						
GoyetQ-2	15,005	72,263	1.7%	-0.6%	4.0%						
Brillenhohle	14780	13,459	2.5%	-3.0%	8.1%						
HohleFels79	14,670	11,211	1.7%	-5.1%	8.5%						
Burkhardtshohle	14,615	38,376	1.7%	-1.6%	5.0%						
Villabruna	13,980	1,215,433	2.7%	1.8%	3.5%	425,148	3.3%	3.0%	3.7%	1.1%	0.9%
Bichon	13,665	2,116,782	2.9%	1.9%	3.8%	769,422	2.7%	2.2%	3.2%	0.7%	1.3%
Satsurblia	13,255	1,460,368	1.5%	0.6%	2.4%	542,561	2.0%	1.7%	2.4%	0.9%	0.6%
Rochedane	12,960	237,390	1.9%	0.5%	3.3%						
Iboussieres39	11,725	9,659	6.4%	-0.8%	13.7%						
Continenza	10,855	11,717	4.1%	-1.4%	9.6%	1,733	2.9%	1.8%	4.0%	-10.6%	4.4%
Ranchot88	10,085	414,863	2.9%	1.8%	4.0%						
LesCloseaux13	9,900	8,635	-	-9.7%	3.8%						
Kotias	9,720	2,133,968	1.8%	1.0%	2.7%	779,146	2.1%	1.8%	2.4%	0.7%	0.5%
Falkenstein	9,200	64,428	4.8%	1.7%	7.8%						
Karelia	8,375	1,754,410	1.9%	1.1%	2.7%	582,444	2.2%	1.9%	2.6%	-0.2%	0.7%
Bockstein	8,265	21,977	5.7%	1.0%	10.5%						
Ofnet	8,245	6,263	9.8%	1.4%	18.1%						
Chaudardes1	8,205	92,657	1.9%	-0.2%	3.9%						
Loschbour	8,050	2,091,584	2.5%	1.6%	3.3%	774,139	2.6%	2.0%	3.1%	2.7%	1.7%
LaBrana1	7,815	1,884,745	1.9%	1.1%	2.8%	642,231	2.7%	2.3%	3.2%	0.4%	0.8%
Hungarian.KO1	7,660	1,410,303	2.1%	1.2%	3.0%	439,408	2.4%	2.0%	2.8%	-0.1%	1.2%
Motala12	7,625	1,874,519	2.5%	1.6%	3.3%	655,685	2.3%	1.9%	2.7%	-0.1%	0.7%
BerryAuBac	7,245	54,690	2.5%	-0.2%	5.1%						
Stuttgart	7,140	2,078,724	1.9%	1.1%	2.7%	767,813	2.1%	1.8%	2.5%	0.0%	0.7%
Dai	0	2,144,502	1.4%	0.7%	2.1%	782,066	1.8%	1.5%	2.1%	1.4%	0.4%
Han	0	2,144,502	1.8%	1.1%	2.5%	782,164	2.1%	1.8%	2.5%	1.9%	0.7%
English	0	2,144,502	1.5%	0.8%	2.2%						
French	0	2,144,502	1.5%	0.9%	2.1%	782,386	1.7%	1.4%	1.9%	1.4%	0.6%
Sardinian	0	2,144,502	1.2%	0.6%	1.9%	782,351	1.7%	1.4%	2.0%	0.7%	0.5%
Karitiana	0					782,037	2.1%	1.7%	2.4%	1.5%	1.0%

Extended Data Table 3 | Significant correlation of Neanderthal ancestry estimate with specimen age

Subset of samples	N	P-value for date correlation	Decrease in ancestry per 10,000 years	Estimate of Neanderthal ancestry at different time points			
				0 years ago (present)	50,000 years ago	55,000 years ago	60,000 years ago
f_4-ratio estimates							
Core Set 1 (all ancient samples (except *Oase1*) + *Han* + *Dai*)	57	5×10^{-22}	0.48-0.73%	1.1-2.2%	4.0-5.4%	4.3-5.7%	4.5-6.0%
Subset of Core Set 1 (<32kya)	50	2×10^{-15}	0.59-0.98%	0.9-2.1%	4.5-6.4%	4.8-6.9%	5.1-7.4%
Subset of Core Set 1 (>32kya or <25kya)	44	4×10^{-18}	0.44-0.69%	1.0-2.2%	3.7-5.2%	4.0-5.5%	4.2-5.8%
Subset of Core Set 1 (>25kya or <14kya)	47	5×10^{-21}	0.48-0.73%	1.0-2.2%	3.9-5.3%	4.2-5.7%	4.5-6.0%
Subset of Core Set 1 (>14kya or present day)	37	2×10^{-18}	0.47-0.74%	1.1-2.4%	4.1-5.5%	4.3-5.8%	4.6-6.2%
Subset of Core Set 1 (only ancient samples)	50	4×10^{-15}	0.46-0.76%	1.0-2.3%	4.0-5.4%	4.3-5.8%	4.5-6.1%
Subset of Core Set 1 (individuals with >200,000 SNPs)	28	4×10^{-19}	0.46-0.71%	1.1-2.3%	3.9-5.3%	4.2-5.7%	4.4-6.0%
Modification of Core Set 1 (replace East Asians with Europeans)	58	2×10^{-23}	0.49-0.73%	1.1-2.3%	4.0-5.4%	4.3-5.8%	4.6-6.1%
All ancient samples including *Oase1* + *Han* + *Dai*	58	8×10^{-29}	0.57-0.81%	1.0-2.2%	4.3-5.7%	4.7-6.1%	5.0-6.5%
All ancient samples	51	1×10^{-20}	0.57-0.86%	0.9-2.2%	4.4-5.8%	4.7-6.2%	5.0-6.6%
All ancient samples except *Oase1* or *UstIshim*	49	8×10^{-12}	0.45-0.81%	1.0-2.3%	4.0-5.6%	4.2-6.0%	4.5-6.4%
Ancestry informative SNPs							
Core Set 2 (all ancient samples (except *Oase1*) + *Han* + *Dai* + *Karitiana*)	29	4×10^{-11}	0.21-0.39%	1.8-2.3%	3.1-4.0%	3.2-4.2%	3.3-4.3%
Subset of Core Set 2 (no *Han, Dai, Karitiana, Stuttgart*)	25	1×10^{-4}	0.11-0.36%	1.8-2.5%	2.9-3.8%	3.0-4.0%	3.0-4.1%
Subset of Core Set 2 (no *Han, Dai, Karitiana, Stuttgart, UstIshim*)	24	2×10^{-4}	0.11-0.37%	1.8-2.5%	2.9-3.8%	2.9-4.0%	3.0-4.2%

'Core set 1' used for the f_4-ratio analyses, refers to 50 ancient individuals (removing *Oase1* as an outlier) along with 7 east Asians (*Dai* and *Han*). 'Core set 2' used for the analyses of Neanderthal ancestry informative SNPs, refers to 26 ancient individuals (removing *Oase1*, *Han*, *Dai* and *Karitiana*).

Extended Data Table 4 | Sex determination for newly reported individuals

Sample	Target	Type	N_{auto}	N_X	N_Y	N_X/N_{auto}	N_Y/N_{auto}	X-rate	Y-rate	Sex
	1240k or 2.2M		1151240	49711	32681	0.0432	0.0284			
	390k		388745	1819	2242	0.0047	0.0058			
Kostenki14	2.2M	all	29633405	395534	262846	0.0133	0.0089	0.309	0.312	M
GoyetQ116-1	1240k	all	2122620	36391	22256	0.0171	0.0105	0.397	0.369	M
Cioclovina1	1240k	Damage	11521	184	125	0.0160	0.0108	0.370	0.382	M
Kostenki12	2.2M	Subset	63908	856	504	0.0134	0.0079	0.310	0.278	M
Muierii2	2.2M	Damage	81165	2177	8	0.0268	0.0001	0.621	0.003	F
Vestonice13	2.2M	Damage	119094	1578	1059	0.0133	0.0089	0.307	0.313	M
Vestonice15	2.2M	Damage	28762	338	227	0.0118	0.0079	0.272	0.278	M
Vestonice14	390k	Damage	4846	8	11	0.0017	0.0023	0.353	0.394	M
Vestonice43	2.2M	Damage	136933	1826	1204	0.0133	0.0088	0.309	0.310	M
Pavlov1	2.2M	Damage	54429	631	404	0.0116	0.0074	0.268	0.261	M
Vestonice16	2.2M	Subset	2433741	30463	20976	0.0125	0.0086	0.290	0.304	M
KremsWA3	1240k	all	235069	4119	2661	0.0175	0.0113	0.406	0.399	M
Ostuni2	2.2M	Damage	15749	138	1	0.0088	0.0001	0.203	0.002	F
Ostuni1	2.2M	Damage	427199	10868	47	0.0254	0.0001	0.589	0.004	F
Paglicci108	1240k	Damage	3883	124	2	0.0319	0.0005	0.740	0.018	F
GoyetQ53-1	1240k	Damage	10771	311	4	0.0289	0.0004	0.669	0.013	F
GoyetQ376-19	1240k	Damage	20052	680	10	0.0339	0.0005	0.785	0.018	F
GoyetQ56-16	1240k	Damage	8702	304	7	0.0349	0.0008	0.809	0.028	F
Paglicci133	1240k	Subset	81092	1641	983	0.0202	0.0121	0.469	0.427	M
ElMiron	2.2M	Damage	1765696	40647	196	0.0230	0.0001	0.533	0.004	F
HohleFels79	390k	Damage	10188	28	22	0.0027	0.0022	0.587	0.374	M
AfontovaGora3	2.2M	Damage	291798	8705	37	0.0298	0.0001	0.691	0.004	F
HohleFels49	390k	Damage	61051	113	111	0.0019	0.0018	0.396	0.315	M
Rigney1	1240k	Damage	32797	1131	9	0.0345	0.0003	0.799	0.010	F
GoyetQ-2	1240k	Damage	65563	1123	706	0.0171	0.0108	0.397	0.379	M
Brillenhohle	390k	Damage	12603	22	22	0.0017	0.0017	0.373	0.303	M
Burkhardtshohle	1240k	Damage	34207	563	407	0.0165	0.0119	0.381	0.419	M
Villabruna	2.2M	Subset	5505838	72055	52110	0.0131	0.0095	0.303	0.333	M
Rochedane	1240k	Subset	256325	4780	2830	0.0186	0.0110	0.432	0.389	M
Continenza	2.2M	Damage	10647	208	2	0.0195	0.0002	0.452	0.007	F
Iboussieres39	390k	Damage	8246	12	22	0.0015	0.0027	0.311	0.463	M
Ranchot88	1240k	Damage	594962	18520	119	0.0311	0.0002	0.721	0.007	F
LesCloseaux13	1240k	Damage	7326	275	2	0.0375	0.0003	0.869	0.010	F
Falkenstein	390k	Damage	58970	113	102	0.0019	0.0017	0.410	0.300	M
Bockstein	390k	Damage	20214	62	0	0.0031	0.0000	0.655	0.000	F
Ofnet	390k	Damage	5294	13	1	0.0025	0.0002	0.525	0.033	F
Chaudardes1	1240k	Damage	84052	1429	865	0.0170	0.0103	0.394	0.363	M
BerryAuBac	1240k	All	49670	902	554	0.0182	0.0112	0.421	0.393	M

*We restrict analysis to the 1240k target set for study of the 2.2M capture datasets.
Y-rate is the ratio of N_Y/N_{auto} divided by the same quantity for the genome-wide target set. Female sex (F) is inferred as Y-rate <0.05 and male sex (M) as Y-rate >0.

Extended Data Table 5 | Allele counts at SNPs affected by selection in individuals with >1-fold coverage

	SNP	*LCT* rs4988235	*SLC45A2* rs16891982	*SLC24A5* rs1426654	*EDAR* rs3827760	*HERC2* rs12913832
	Ancestral	G	C	G	A	A
	Derived	A	G	A	G	G
UstIshim	Coverage	31	46	52	42	50
	Derived allele frequency	0%	0%	2%	0%	0%
Kostenki14	Coverage	140	113	6	45	52
	Derived allele frequency	0%	2%	17%	0%	0%
GoyetQ116-1	Coverage	8	6	0	9	1
	Derived allele frequency	0%	0%	n/a	0%	0%
Vestonice16	Coverage	13	18	0	4	5
	Derived allele frequency	0%	6%		0%	0%
Malta1	Coverage	1	0	2	2	2
	Derived allele frequency	0%		0%	0%	0%
ElMiron	Coverage	2	10	0	7	5
	Derived allele frequency	0%	0%		0%	0%
Villabruna	Coverage	17	52	5	19	10
	Derived allele frequency	0%	0%	0%	0%	100%
Bichon	Coverage	11	4	25	16	9
	Derived allele frequency	0%	0%	0%	0%	33%
Satsurblia	Coverage	1	2	4	1	4
	Derived allele frequency	0%	0%	100%	0%	50%
Kotias	Coverage	16	22	13	20	15
	Derived allele frequency	0%	0%	100%	0%	0%
Loschbour	Coverage	19	18	20	17	21
	Derived allele frequency	0%	0%	0%	0%	100%
LaBrana1	Coverage	8	6	2	11	3
	Derived allele frequency	12%	0%	0%	0%	100%
Hungarian.KO1	Coverage	1	2	2	1	2
	Derived allele frequency	0%	0%	50%	0%	100%
Motala12	Coverage	2	0	3	3	1
	Derived allele frequency	0%		0%	33%	100%
Karelia	Coverage	1	9	4	0	1
	Derived allele frequency	0%	67%	0%		0%
Stuttgart	Coverage	25	21	15	29	21
	Derived allele frequency	0%	0%	100%	0%	0%

rs4988235 is responsible for lactase persistence in Europe[60,61]. The SNPs at *SLC24A5* and *SLC45A2* are responsible for light skin pigmentation. The SNP at *EDAR*[62,63] affects tooth morphology and hair thickness. The SNP at *HERC2* (refs 64, 65) is the primary determinant of light eye colour in present-day Europeans. We present the fraction of fragments overlapping each SNP that are derived; the observation of a low rate of derived alleles does not prove that the individual carried the allele, and instead may reflect sequencing error or ancient DNA damage. Sites highlighted in light grey were judged (based on the derived allele count) likely to be heterozygous for the derived allele, and dark grey sites are likely to be homozygous.

60. Enattah, N. S. *et al.* Identification of a variant associated with adult-type hypolactasia. *Nature Genet.* **30,** 233–237 (2002).
61. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74,** 1111–1120 (2004).
62. Fujimoto, A. *et al.* A scan for genetic determinants of human hair morphology: *EDAR* is associated with Asian hair thickness. *Hum. Mol. Genet.* **17,** 835–843 (2008).
63. Kimura, R. *et al.* A common variation in *EDAR* is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.* **85,** 528–535 (2009).
64. Sturm, R. A. *et al.* A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82,** 424–431 (2008).
65. Eiberg, H. *et al.* Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum. Genet.* **123,** 177–187 (2008).

Extended Data Table 6 | All European hunter-gatherers beginning with *Kostenki14* share genetic drift with present-day Europeans

Test	SNPs used	D-value	Z score
Ust'-Ishim	2,050,358	0.003	6.6
Oase1	278,785	0.005	10.6
Kostenki14	1,676,253	-0.002	-5.5
Muierii2	95,787	-0.004	-6.3
GoyetQ116-1	811,756	-0.004	-8.0
Kostenki12	59,850	-0.004	-5.1
Paglicci133	79,624	-0.004	-5.5
Vestonice13	136,598	-0.004	-7.1
Vestonice15	30,252	-0.006	-6.4
Vestonice16	914,141	-0.004	-9.1
Pavlov1	55,835	-0.005	-6.3
Vestonice43	160,463	-0.004	-6.9
KremsWA3	229,187	-0.005	-10.2
Ostuni1	360,347	-0.004	-8.6
Malta1	1,401,718	-0.005	-11.3
ElMiron	777,654	-0.007	-14.7
AfontovaGora2	141,073	-0.007	-13.6
AfontovaGora3	707,617	-0.006	-13.6
HohleFels49	62,816	-0.004	-5.2
Rigney1	34,445	-0.006	-6.1
GoyetQ-2	70,210	-0.006	-8.8
Burkhardtshohle	37,234	-0.006	-6.2
Villabruna	1,170,777	-0.010	-24.7
Bichon	2,034,069	-0.009	-23.6
Satsurblia	1,419,824	-0.005	-13.1
Rochedane	229,806	-0.011	-20.8
Ranchot88	402,274	-0.010	-21.8
Kotias	2,047,856	-0.006	-15.8
Falkenstein	64,043	-0.008	-11.6
Chaudardes1	90,047	-0.011	-16.0
Loschbour	2,037,082	-0.011	-25.4
LaBrana1	1,824,307	-0.009	-23.0
Motala12	1,816,201	-0.009	-23.8
Hungarian.KO1	1,372,801	-0.010	-26.5
Karelia	1,701,664	-0.009	-21.9
Stuttgart	2,023,939	-0.009	-23.9
BerryAuBac	53,028	-0.011	-14.0

The statistic *D(Han, Test; French, Mbuti)* was computed measuring whether present-day *French* share more alleles with *Han* or with a *Test* population (restricting to ancient individuals with at least 30,000 SNPs covered at least once). Present-day Europeans share significantly more genetic drift with European hunter-gatherers from *Kostenki14* onward than they do with *Han*. Thus, by the date of *Kostenki14*, there was already west Eurasian-specific genetic drift.

Table of Contents 1

Section 1
Archaeological information

Radiocarbon date calibrations used in this study
We recalibrated all radiocarbon dates using IntCal13[1] in OxCal4.2[2]. We give 95.4% confidence intervals for calibrated dates in years before present (where the present is defined as 1950 CE). We round calibrated dates to the nearest decade.

Previously published genome-wide ancient DNA data
We do not provide archaeological context summaries for genome-wide ancient DNA data sets that have previously been published. However, we list the dates here:

- *UstIshim* at 47,480-42,560 cal BP (OxA-25516: 41,400 ±1300 ^{14}C; OxA-30190: 41,400 ±1400 ^{14}C)[3]
  (direct date, using collagen ultrafiltration)

- *Oase1* at 41,640-37,580 cal BP (GrA-22810: 34,290, +970, -870 ^{14}C; OxA-11711: >35,200 ^{14}C; Combined ^{14}C Age: OxA-GrA: 34,950 +900 -890 ^{14}C)[4].
  (direct date, using collagen ultrafiltration)

- *Malta1* at 24,520-24,090 cal BP (UCIAMS-79666: 20,240 ± 60 ^{14}C)[5]
  (direct date, using collagen ultrafiltration)

- *AfontovaGora2* at 16,940-16,480 cal BP (UCIAMS-79661: 13,810 ± 35 ^{14}C)[5]
  (direct date, using collagen ultrafiltration)

- *Bichon* at 13,770-13,560 cal BP (OxA-27763: 11,855 ± 50 ^{14}C)[6]
  (direct date, using collagen ultrafiltration)

- *Satsurblia* at 13,380-13,130 cal BP (OxA-34632: 11,415 ± 50 ^{14}C)[6]
  (direct date, using collagen ultrafiltration)

- *Kotias* at 9,890-9,550 cal BP (RTT-5246: 8,665 ± 65 ^{14}C; OxA-28256: 8,745 ± 40 ^{14}C)[6]
  (direct date, using collagen ultrafiltration)

- *Loschbour* at 8,160-7,940 cal BP (OxA-7738: 7,205 ± 50 ^{14}C)[7]
  (direct date, using collagen without ultrafiltration)

- *LaBrana1* at 7,940-7,690 cal BP (Beta-226472: 6,980 ± 50 ^{14}C)[8]
  (direct date, using collagen ultrafiltration)

- *Hungarian.KO1* at 7,730-7,590 cal BP (OxA-23757: 6835 ± 34 ^{14}C)[9]
  (direct date, using collagen ultrafiltration)

- *Motala12* at 7,670-7,580 cal BP (Ua-51723-16.8: 6,773 ± 30 ^{14}C)[7]
  (direct date, using collagen ultrafiltration – the present study is the first report of this date, courtesy of Fredrik Hallgren)

- *Karelia* at 8,800-7,950 cal BP (OxA-1665, OxA-2266, OxA-1667, OxA1668, OxA-1669, OxA-2124, OxA-2125, OxA-1973 ^{14}C)[10]
  We genetically analysed individual UZOO-74/I0061 (MAE RAS collection number 5773–74, grave number 142), collected from the Mesolithic site Yuzhnyy Oleni' Ostrov (*Oleneostrovski' mogilnik* or *Deer Island cemetery*), Onega Lake, Karelia, Russian Federation (61°30'N 35°45'E). Table S1.1 gives the dates we used.
  (layer date, not from collagen ultrafiltration, based on other individuals in the graveyard)[11]

Table S1.1. Layer dates used for the *Karelia* sample

Lab number	burial no.	^{14}C years BP	Cal BP 95.4% probability
OxA-1665	57	7,280 ± 80	8,220-7,950
OxA-2266	57	7,350 ± 90	8,360-8,000
OxA-1667	80	7,330 ± 90	8,340-7,980
OxA-1668	80	7,560 ± 90	8,540-8,190
OxA-1669	80	7,560 ± 90	8,540-8,190
OxA-2124	89	7,280 ± 90	8,310-7,950
OxA-2125	85	7,510 ± 90	8,480-8,160
OxA-1973	108	7,750 ± 110	8,800-8,360

- *Stuttgart* at 7,260-7,020 cal BP (MAMS-24635: 6,246 ± 30 ^{14}C)[7]
  (direct date, using collagen ultrafiltration – the present study is the first report of this date)

Peştera Cioclovina Uscată (South Transylvania, Romania)
We analysed an occipital bone fragment from the cranium, which gave a date between the late phases of the Aurignacian and the beginning of the Gravettian. The remains were not associated with artifacts, so we consider its archaeological context to be "Unassigned":

- *Cioclovina1* at 33,090-31,780 cal BP (OxA-15527: 28,510 ± 170 ^{14}C)[12]
  (direct date, using collagen ultrafiltration)

La Rochette (Dordogne, France)
The site of La Rochette, in the proximity of Saint-Léon-sur-Vézère, was excavated beginning in 1910 by Otto Hauser. There is evidence of human occupation of the site spanning from the Middle Palaeolithic (Mousterian) until the Upper Palaeolithic (Magdalenian). A revision of the site stratigraphy divided the layer with human remains into Gravettian, Aurignacian and Châtelperronian periods[13]. The right ulna was redated in 2011, which confirmed a previous date from 2002 associating it with the Gravettian horizon[14]. The ancient DNA analysis was performed on an ulna, yielding data from fewer than 4,000 SNPs covered at least once after quality control. Thus, we did not use the data from this individual in our main analysis.

- *LaRochette* at 27,780-27,400 cal BP (OxA-23413: 23,400 ± 110 ^{14}C)[15]
  (direct date, using collagen ultrafiltration)

Dolní Věstonice and Pavlov (South Moravia, Czech Republic)
Three Gravettian burial sites were discovered beginning in 1924 on the northern slope of the Pavlovské Hills in South Moravia: Dolní Věstonice I, Dolní Věstonice II and Pavlov I. From Dolní Věstonice I we analysed the triple burial comprised of *Vestonice13* (a femur), *Vestonice14* (a femur) and *Vestonice15* (a femur). Associated charcoal gives a date of:

- *Vestonice13, Vestonice14, and Vestonice15* at 31,070-30,670 cal BP (GrN-14831: 26,640 ± 110 ^{14}C)[16]
  (layer date, based on associated charcoal)

The *Vestonice16* (a femur), *Vestonice42* (a femur) and *Vestonice43* (a femur) samples were retrieved close to Dolní Věstonice I, suggesting similar depositional age (*Vestonice42* provided fewer than 4,000 SNPs and we did not use it for whole genome analysis). We obtained two dates on charcoal associated with these specimens:

- *Vestonice16, Vestonice42,* and *Vestonice43* at 30,710-29,310 cal BP (GrN-15277: 25,740 ± 210 ^{14}C and GrN-15276: 25,570 ± 280 ^{14}C)[16]
  (layer dates, based on associated charcoal)

Charcoal from the Pavlov I site from which we analysed *Pavlov1* (a femur) gave a date of:

- *Pavlov1* at 31,110-29,410 cal BP (GrN-20391: 26,170 ± 450 ^{14}C)[16]
  (layer date, based on associated charcoal)

Grotta Paglicci (Rignano Garganico - Foggia, Italy)
Paglicci Cave is situated in the Apulia region in southeast Italy. Excavation began in 1961 led by F. Zorzi of the Natural History Museum of Verona, followed, in 1971, by A. Palma di Cesnola and then by A. Ronchitelli of the University of Siena in collaboration with the Soprintendenza Archeologia della Puglia. Human occupation at this site is attested during the Early Middle and throughout the entirety of the Upper Palaeolithic. Paglicci is also important for the presence of the only Palaeolithic wall paintings discovered so far in Italy and for the most ancient evidence of flour production[17]. The main stratigraphic sequence is 12 meters thick. The Upper Palaeolithic layers yielded several human remains spanning from the Early Gravettian to the Final Epigravettian[18]. The three human remains used in ancient DNA analyses came from three distinct stratigraphic units.

We performed genetic analysis on *Palgicci71*, a right patella discovered in layer 8 whose ^{14}C date confirms the archaeological attribution to the Evolved Epigravettian, although this sample was not used in our main analysis as it yielded fewer than 4,000 SNPs:

- *Paglicci71* at 19,250-18,210 cal BP (F-66: 15,460 ± 220 ^{14}C)[19]
  (layer date, based on associated charcoal)

We also performed genetic analysis on *Paglicci108*, a phalanx discovered in layer 21B. A date from charcoal is consistent with the Evolved Gravettian material found in this layer:

- *Paglicci108* at 28,430-27,070 cal BP (F-52: 23,470 ± 370 ^{14}C)[20]
  (layer date, based on associated charcoal)

We finally performed genetic analysis on *Paglicci133*, a tooth (M_3 dx) found in layer 23C2, which is not directly dated but which can be associated on the basis of its stratigraphic position and the associated artifacts to the Early Gravettian culture:

- *Paglicci133* was found in a layer (23C2) whose chronology can be derived from the occurrence of tephra Codola elsewhere dated to around 33,000 years ago[21]. Moreover this layer is located between layer 23A (Early Gravettian) with a date of 33,110-31,210 cal BP (UTC-1415: 28,100 ± 400 ^{14}C) and layer 24A1 (Aurignacian) with a date of 34,580-31,860 cal BP (UTC-1789: 29,300 ± 600 ^{14}C)[18].
  (both dates are based on charcoal in the layers above and below)

Troisième Caverne (Goyet, Belgium)

The Troisième Caverne of Goyet in Belgium was excavated between the second half of the 19th century and the beginning of the 20th century, and more recently in the last decade of the 20th century. Edouard Dupont led the main excavations in 1868 and revealed Palaeolithic industries[22]. Follow-up studies found evidence of human occupation represented by Mousterian material, Lincombian-Ranisian-Jerzmanowician material, and Upper Palaeolithic complexes, including Aurignacian, Gravettian and Magdalenian material[23]. In 2008, a study was initiated with the goal of revisiting the human and faunal collections, and resulted in the identification of numerous new human remains. However, due to the lack of detailed excavation data, it has been challenging to assign the human remains to specific stratigraphic horizons. Taphonomic and morphometric characteristics were therefore first evaluated to sort human remains in different groups. All the specimens we analysed were directly radiocarbon dated in order to confirm the temporal attribution[24]. Those results, in combination with isotopic and genetic analyses[25-27], allowed specimen assignment either to late Neanderthal or to modern human origin. For ancient DNA analysis we analysed two individuals dated to the Aurignacian period[24]: *GoyetQ116-1* (a humerus), and *GoyetQ376-3* (a humerus):

- *GoyetQ116-1* at 35,160-34,430 cal BP (GrA-46175: 30,880 + 170 -160 ^{14}C)[24] (direct date, without ultrafiltration)

- *GoyetQ376-3* at 33,940-33,140 cal BP (GrA-60034: 29,370 + 180 -170 ^{14}C)[24] (direct date, without ultrafiltration)

We also analysed five individuals dated to the Gravettian period[24]: *GoyetQ376-19* (a humerus), *GoyetQ53-1* (a fibula), *GoyetQ55-2* (a fibula), *GoyetQ56-16* (a fibula) and *Goyet2878-21* (a clavicle):

- *GoyetQ376-19* at 27,720-27,310 cal BP (GrA-54026: 23,260 + 110 -100 ^{14}C)[24] (direct date, without ultrafiltration)

- *GoyetQ53-1* at 28,230-27,720 cal BP (GrA-46169: 23,920 ± 100 ^{14}C)[24] (direct date, without ultrafiltration)

- *GoyetQ55-2* at 27,730-27,310 cal BP (GrA-54031: 23,270 + 120 -110 ^{14}C)[24] (direct date, without ultrafiltration)

- *GoyetQ56-16* at 26,600-26,040 cal BP (GrA-59991: 22,100 ± 100 ^{14}C)[24] (direct date, without ultrafiltration)

- *Goyet2878-21* at 27,060-26,270 cal BP (GrA-62455: 22,360 ± 110 ^{14}C)[24] (direct date, without ultrafiltration)

We finally analysed an individual dated to the Magdalenian period[24]: *GoyetQ-2* (a humerus).

- *GoyetQ-2* at 15,230-14,780 cal BP (GrA-46168: 12,650 ± 50 ^{14}C)[24] (direct date, without ultrafiltration)

Three samples (*GoyetQ376-3*, *GoyetQ55-2* and *Goyet2878-21*) were not used for genome-wide analyses because fewer than 4,000 SNPs were covered after quality control.

Swabian Jura sites (Baden-Württemberg, Germany)
We carried out ancient DNA analysis on seven specimens from six caves in the Swabian Jura region, Southwest Germany. The geology of this area was formed during the Jurassic period and is mainly comprised of limestone that favors the formation of caves. The Ach and Lone rivers form two valleys of the same name before they merge into the Danube.

(i) Hohle Fels (Ach valley): This site is famous for the discoveries of Upper Palaeolithic ivory figurines and flutes associated with the early Aurignacian[28,29]. We performed genetic analysis on two remains, *HohleFels49* (a femur) and *HohleFels79* (a cranial fragment) both dated to the Magdalenian period:

- *HohleFels49* at 16,000-14,260 cal BP (H5312-4907: 12,770 ± 220 ^{14}C and H5119-4601: 13,085 ± 95 ^{14}C)[30]
  (layers date, based on a direct dates on bones from collagen without ultrafiltration)

- *HohleFels79* at 15,070-14,270 cal BP (MAMS-25564: 12,490 ± 70 ^{14}C)[24]
  (direct date, using collagen ultrafiltration)

(ii) Brillenhöhle (Ach valley): An excavation led by Gustav Riek beginning in 1956 discovered at least four individuals, all assigned to the Magdalenian culture based on associated artifacts[31]. For ancient DNA analysis we analysed a parietal bone:

- *Brillenhohle* at 15,120-14,440 cal BP (OxA23414: 12,535 ± 50 ^{14}C)[15]
  (direct date, using collagen ultrafiltration)

(iii) Burkhardtshöhle (Westerheim, Württemberg): An excavation beginning in 1933-1934 and led by Gustav Riek found five cranial fragments, probably from the same individual. These specimens were associated with the Magdalenian culture and gave a date of:

- *Burkhardtshohle* at 15,080-14,150 cal BP (ETH-7613: 12,450 ± 110 ^{14}C)[32]
  (direct date, without collagen ultrafiltration)

(iv) Bockstein-Höhle (Lone valley): This cave yielded a double burial consisting of an infant and an adult. The adult was morphologically identified as female and the sex is confirmed here through genetic analysis of an incisor. Two skeletal elements of the infant were dated and linked the burial to the Late Mesolithic period:

- *Bockstein* at 8,370-8,160 cal BP (UtC-7887: 7,350 ± 70 ^{14}C and UtC-6796: 7,460 ± 60 ^{14}C)[33]
  (layer date, based on a direct date of a second human skeleton in the same layer, without collagen ultrafiltration)

(v) Falkensteiner Höhle: This cave is situated in a karstic area in the vicinity of Bad Urach. The excavations of the site started in 1933 and in later analyses, around 40 human remains were identified[31]. The same site yielded typical Mesolithic stone tools as well as the human fibula that we investigated genetically. The remains gave a date:

- *Falkenstein* at 9,410-8,990 cal BP (ETH-7615: 8,185 ± 80 ^{14}C)[34]
  (direct date, using collagen without ultrafiltration)

(vi) Hohlenstein-Stadel (Lone valley): Signs of human occupation at this site spanned from the Middle Palaeolithic to the Neolithic. A triple burial only represented by human skulls and cervical vertebrae was found in the cave and was dated to a late phase of the Mesolithic. We did not include this sample in our main analysis because fewer than 4,000 SNPs were covered at least once after quality control.

- *HohlensteinStadel* at 8,980-8,440 cal BP (ETH-5732: 7,835 ± 80 ^{14}C)[35]
  (direct date, using collagen without ultrafiltration)

Felsdach Inzigkofen (Upper Danube valley, Germany)
This site is located in the Upper Danube valley, in close proximity of Beuron village. Wolfgang Taute excavated the site where a 3rd molar was discovered in a layer associated to the early period of the Late Mesolithic, also called the Early Atlantic. We did not include this sample in our main analysis because it gave fewer than 4,000 SNPs after quality control.

- *Felsdach* at 8,980-8,380 cal BP (B-933: 7,770 ± 120 ^{14}C)[31,36]
  (layer date)

Große Ofnet Höhle (Franconian Jura, Germany)
Große Ofnet cave is found close to Nördlingen, in Bavaria. Within the cave, two sites yielded a total of 34 crania[35], all facing west and some presenting cut-marks or sign of injuries[37]. For ancient DNA analysis, we analysed a molar tooth that was excavated by Oscar Fraas in 1875-1876 and is part of the human osteological collection of the University of Tübingen (OSUT 4043). Here we quote a radiocarbon date that associates the remains to the Late Mesolithic:

- *Ofnet* at 8,430-8,060 cal BP (OxA1574: 7,480 ± 80 ^{14}C)[35]
  (direct date, using collagen without ultrafiltration)

French Jura sites (Franche-Comté, France)
This mountainous region is located between Eastern France and Western Switzerland. The entire area is composed of Mesozoic limestone, where a karstic landscape leads to cave formation[38]. We analysed samples from three sites.

(i) Cabônes rockshelter (Ranchot, Jura department): This site was discovered in the 1950s, and excavation occurred between 1978 and 1990. At least five individuals were assigned to the Mesolithic[39]. The specimen we analysed, *Ranchot88*, is a right parietal fragment:

- *Ranchot88* at 10,240-9,930 cal BP (GrA-38019: 8,985 ± 40 ^{14}C)[24]
  (direct date, using collagen without ultrafiltration)

(ii) Rigney 1 cave (Doubs department): We performed ancient DNA analysis on a mandible fragment recovered at Rigney 1 cave in 1986-87 during a rescue excavation. This site is linked to the Magdalenian material culture[40].

- *Rigney1* at 15,690-15,240 cal BP (Ly-6515(OxA): 12,930 ± 55 ^{14}C)[38]
  (direct date, using collagen without ultrafiltration)

(iii) Rochedane rockshelter (Villars-sous-Dampjoux, Doubs department): A. Thévenin excavated this site between 1968 and 1976 after its discovery at the end of the 19th century. The mandible we analysed was assigned to the EpiPalaeolithic culture.

- *Rochedane* at 13,090-12,830 cal BP (GrA-41739: 11,120 ± 50 ^{14}C)[38]
  (direct date, using collagen without ultrafiltration)

Paris Basin (France)
In this region northeast of Paris, several Mesolithic human remain were discovered[41]. The local geology is characterized by a plateau of sedimentary rocks. We analysed three sites.

(i) Les Closeaux at Rueil-Malmaison: A single human skeleton was buried in squatting position inside a circular pit along the river Seine.

- *LesCloseaux13* at 10,240-9,560 cal BP (OxA-7109 (Ly-612): 8,870±130 ^{14}C)[42]
  (direct date, using collagen without ultrafiltration)

(ii) Les Fontinettes at Cuiry-lès-Chaudardes: This Mesolithic burial, which we call *Chaudardes1*, was found at the side of a Neolithic settlement with a single individual in seated position accompanied by three flints and a necklace[43].

- *Chaudardes1* at 8,360-8,050 cal BP (GrA-28268: 7,400 ± 60 ^{14}C)[24]
  (direct date, using collagen without ultrafiltration)

(iii) Le Vieux Tordoir at Berry-au-Bac: This site revealed human occupation in several periods including a fortified Neolithic site. For ancient DNA analysis we studied a radius from a skeleton buried with a bone tool and covered with ochre[44].

- *BerryAuBac* at 7,320-7,170 cal BP (SacA-5455: 6,325 ± 35 ^{14}C)[24]
  (direct date, using collagen without ultrafiltration)

Aven des Iboussières à Malataverne (Rhône-Alpes, France)
In this site located in South France, at least eight human individuals were discovered and associated to the EpiPalaeolithic culture[45]. We studied the left femur of a juvenile.

- *Iboussieres39* at 12,040-11,410 cal BP (GrA-43700: 10,140 ± 50 ^{14}C)[24]
  (direct date, using collagen without ultrafiltration)

Peştera Muierilor (Romania)
Excavations at Muierilor Cave in the 1950s yielded human remains from up to three individuals. The *Muierii2* temporal bone that we analysed genetically gave a date of:

- *Muierii2* at 33,760-32,840 cal BP (OxA-16252: 29,110 ± 190 ^{14}C yr BP)[46]
  (direct date, using collagen ultrafiltration)

El Mirón Cave (Cantabria, Spain)
This large cave is located in the Cantabrian Cordillera, around 20 km from the present shore and equidistant between Bilbao and Santander. Since 1996, it has been excavated under the direction of L.G. Straus and M.R. González Morales, uncovering a nearly complete sequence of cultural occupations from the late Mousterian through the early Bronze Age, with 86 radiocarbon dates ranging from >46,000 years to 3,200 years uncal. BP[47]. Among the richest levels are those of the Cantabrian Lower Magdalenian with classic lithic and osseous artifacts

and engraved red deer scapulae, as well as rock art dated *terminus ante* and *post quem* to this period and intimately associated with the human burial—the first human interment of Magdalenian age to be discovered in the Iberian Peninsula. The "Red Lady of El Mirón" skeleton, which consists of about 100 individual elements (including the mandible, but missing the cranium and most of the large long bones), was recovered between 2010-2013[48]. The human bones, the layer in which they were buried, and the face of the large, engraved block against which the individual's back seems to have rested, were all stained or impregnated with specially prepared, non-local, red ochre (specular hematite). The genetic analysis was performed on a toe bone, and the radiocarbon date came from a fibula of the individual, who was a female, 35-40 years old at the time she died, relatively tall and robust, and in apparently in good health, with a mixed diet of terrestrial game, fish and plants.

- *ElMiron* at 18,830-18,610 cal BP (MAMS-14585: 15,460 ± 40 ^{14}C)[49]
  (direct date, using collagen ultrafiltration)

Villabruna (Sovramonte – Belluno, Italy)
The burial of Riparo Villabruna was discovered in 1988 by A. Broglio in the small rockshelter named Riparo Villabruna A in the Veneto Dolomites. It contains a partial skeleton with lower limbs severed at the distal femoral shafts associated with burial goods of the Epigravettian culture[50]. The date quoted here comes from the skull[51], whereas the genetic analysis is of a left femur. This individual bears the earlier known example of treatment of dental caries[52].

- *Villabruna* at 14,180-13,780 cal BP (KIA-27004: 12,140±70^{14}C)[51]
  (direct date, using collagen ultrafiltration)

Kostenki 14 – Markina Gora (Voronezh region, Russia)
The *Kostenki14* skeleton was found in 1954 in a crouched position inside an oval burial below the third cultural layer. The layer lacked diagnostic stone tools and thus its cultural assignment is unclear. The bones were partially covered with ochre and associated with a small number of flints as well as isolated faunal bones[53,54]. We performed ancient DNA analysis on a tibia from the skeleton.

- *Kostenki14* at 38,680-36,260 cal BP (OxA-X-2395-15: 33,250 ± 500 ^{14}C)[55]
  (HPLC-separated hydroxyproline fraction in collagen)

Kostenki 12 – Volkovskaya (Voronezh region, Russia)
The *Kostenki12* cranial bone belongs to a perinatal child, maybe deliberately buried, found in layer 1 of the Kostenki 12 – Volkovskaya site. This layer yielded an assemblage that has been attributed to the Gorodtsovian, an Early Upper Palaeolithic culture that is characteristic of the region, and that overlaps in time with the Early Gravettian (the Early Gravettian is documented at Kostenki 8 in layer 2). The association between the dated bone and the Gordotsovian material is not secure because of reworking of the deposit[56]. *Kostenki12* was not assigned to a genetic cluster:

- *Kostenki12* at 32,990-31,840 cal BP (GrA-5552: 28,500 ± 140 ^{14}C)[56]
  (layer date, based on associated charcoal)

Ostuni (Apulia, Italy)
The Grotta di Santa Maria di Agnano is a cave near the village of Ostuni, located at about 170 meters above sea level in a limestone formation (*Calcare di Altamura*) on the south-

eastern slopes of the Murge tablelands in central Apulia (southeast Italy). Excavations at this site led by one of the authors (D. Coppola) have demonstrated that the cave was occupied for tens of thousands of years from the Middle Palaeolithic until modern times, including Middle Palaeolithic (Mousterian) and Upper Palaeolithic (Aurignacian, Gravettian and Epigravettian) industries[57]. The excavation discovered two burials, which on stratigraphic and typological grounds are attributable to the Gravettian culture. The first of these burials (*Ostuni1*) is of a pregnant female of around 20 years of age, about to give birth[57]. The skeletons of both the mother and the fetus are remarkably well preserved and, to the best of our knowledge, constitute the oldest reported burial of a pregnant female in the world. The second burial (*Ostuni2*) contains the remains of an adult whose gender has not yet been established based on morphology (due to the partial excavation of this inhumation) but who genetically is female. Around their crania, both *Ostuni1* and *Ostuni2* had numerous ornaments made of perforated shells of the marine gastropod *Cyclope neritea*, similar to other Gravettian burials from Italy. We extracted collagen from the ribs of both individuals using the protocol of Talamo and Richards[58]. The radiocarbon dates for both samples are based on ultrafiltration, and give dates that confirm the assignment of both samples to the Gravettian:

- *Ostuni1* at 27,810-27,430 cal BP (MAMS-11449: 23,446 ± 107 ^{14}C)
  (direct date, using collagen ultrafiltration; the present study is the first report of this date)

- *Ostuni2* at 29,310-28,640 cal BP (MAMS-11450: 24,910 ± 125 ^{14}C)
  (direct date, using collagen ultrafiltration; the present study is the first report of this date)

Afontova Gora (Russia)
Afontova Gora 3 was found in 2014 during salvage excavations connected to bridge construction in the Yenisei River basin. Excavation area No. 24 cut into the third river terrasse and yielded a first cervical vertebra, a mandible, and five teeth from an *in situ* discovery in occupation layer 2. Preliminary analyses suggest that the mandible and five teeth may be associated with a 14-15 year old girl; the sex is confirmed in this study genetically. The mandible demonstrates a gracile structure: its longitudinal and transverse measurements fall into the category of "small" or "very small". The gracile mandible is quite different from the children of similar age recovered of Sungir-2 and Sungir-3, which might indicate different affinities. We performed genetic analysis on a left molar (M_2). We assume that the *AfontovaGora3* specimen is approximately contemporaneous with the *AfontovaGora2* sample found at the site in the 1920s, and so use that sample to provide a layer date:

- *AfontovaGora3* at 16,930-16,490 cal BP (UCIAMS-79661: 13,810 ± 35 ^{14}C)[5]
  (layer date, based on ultrafiltration of remains from the *AfontovaGora2* individual)

Grotta Continenza (Abruzzo, Italy)
Grotta Continenza is a cave at an elevation of 700 meters above sea level overlooking the Fucino plain in an area of Abruzzo, known as Marsica, in central Italy. It contains an outer and an inner chamber, occupied from the late Upper Palaeolithic to the Bronze Age, as well as during the Roman period[59,60]. The prehistoric deposits are stratified in a 9 meter sequence which from the most recent to the oldest is: Neolithic (spits 4 to 23), Castelnovian (spit 24), Sauveterrian (spits 25 to 29) and Late Epigravettian (spits 30 to 48) layers. Human remains are numerous throughout the Mesolithic deposits (spits 24 to 29). A cranial fragment recovered within the Mesolithic layers in stratigraphic spit 24 was selected for ancient DNA analysis. We obtained three radiocarbon dates on a bone from the top of the stratigraphic unit in which the cranial element analysed for DNA was recovered, as well as two bones from

immediately below. The method of bone collagen extraction that we adopted is that proposed by Talamo and Richards[58]. The dates suggest that the deposit from which the cranial fragment was recovered dates to 11,210-10,510 cal years BP. It can therefore be concluded that the specimen for which we obtained ancient DNA is Mesolithic. Despite being recovered in a stratigraphic spit associated with Castelnovian lithic industries, the radiocarbon dates suggest that a tentative attribution to the Sauveterrian is more feasible.

- *Continenza* at 11,200-10,510 calBP, based on a range of three dates from stratigraphic contexts immediately above and below:
  10,870-10,700 cal BP (MAMS-11444: 9,521 ± 31 ^{14}C)
  10,690-10,510 cal BP (MAMS-11445: 9,379 ± 30 ^{14}C)
  11,200-11,080 cal BP (MAMS-11448: 9,680 ± 32 ^{14}C)
  (layer date, based on collagen ultrafiltration of three bones – the present study is the first report of these dates)

Krems-Wachtberg (Lower Austria)

Between 2005-2015, the Austrian Academy of Sciences carried out an interdisciplinary research project at Wachtberg in Krems, Lower Austria. This site is found on the southern slope of a promontory between the Danube and Krems River. The profiles are up to 8 meters high and represent a significant record of the time between 40,000-20,000 years ago. At a depth of 5 meters, about 150 square meters of a well-preserved cultural layer with associated features including hearths and two burials were recovered. The living floor dates to 27.0 ^{14}C ka BP. Burial 1 is a unique double burial of newborns (individuals 1 and 2), while burial 2 (individual 3) is a single burial of an immature individual (its estimated age at death is around 3 months). All three individuals were buried in flexed position and embedded in red ochre. The sample for the present study was taken from the parietal bone of individual 3 (burial 2). The date is from charcoal on the living floor[61].

- *Krems-Wachtberg* at 31,250-30,690 cal BP (VERA-3941: 26,870 ± 220^{14}C)[61,62]
  (layer date, based on associated charcoal)

References

1 Reimer, P. J. *et al.* Intcal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years cal BP. *Radiocarbon* **55**, 1869-1887 (2013).
2 Ramsey, C. B. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720-730 (2013).
3 Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445-449 (2014).
4 Rougier, H. *et al.* Peştera cu Oase 2 and the cranial morphology of early modern Europeans. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1165-1170 (2007).
5 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91 (2014).
6 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Communications* **6**, 8912 (2015).
7 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413 (2014).
8 Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228 (2014).

9 Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications* **5**, 5257 (2014).
10 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522,** 207-211 (2015).
11 Price, T. D. & Jacobs, K. Olenii Ostrov - 1st radiocarbon dates from a major Mesolithic cemetery in Karelia, USSR. *Antiquity* **64**, 849-853 (1990).
12 Soficaru, A., Dobo**ş**, A. & Trinkaus, E. Early modern humans from the Pe**ş**tera Muierii, Baia de Fier, Romania. Proc. Natl Acad. Sci. USA **103,** 17196–17201 (2006).
13 Delporte, H. Le gisement paléolithique de la Rochette, commune de Saint-Léon-sur-Vézère, Dordogne. *Gallia Préhistoire* **V**, 1-22 (1962).
14 Orschiedt, J. Datation d'un vestige humain provenant de la Rochette (Saint Léon-sur-Vézère, Dordogne) par la méthode du carbone 14 en spectrométrie de masse. *Paléo* **14**, 239-240 (2002).
15 Benazzi, S. *et al.* Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature* **479**, 525-528 (2011).
16 Trinkaus, E. & Svoboda, J. *Early Modern Human Evolution in Central Europe: The People of Dolní Věstonice and Pavlov*. Vol. 12 (Oxford University Press, 2006).
17 Mariotti Lippi, M., Foggi, B., Aranguren, B., Ronchitelli, A. & Revedin, A. Multistep food plant processing at Grotta Paglicci (Southern Italy) around 32,600 cal B.P. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 12075-12080 (2015).
18 Palma di Cesnola, A. *Paglicci. L'Aurignaziano e il Gravettiano antico* (Claudio Grenzi Ed., 2004).
19 Azzi, C. M., Bigliocca, L. & Piovan, F. Florence Radiocarbon Dates III. *Radiocarbon* **19**, 165-169 (1977).
20 Azzi, C. M., Bigliocca, L. & Piovan, F. Florence Radiocarbon Dates II. *Radiocarbon* **16**, 10-14 (1974).
21 Giaccio, B. *et al.* The Campanian Ignimbrite and Codola tephra layers: Two temporal/stratigraphic markers for the Early Upper Palaeolithic in southern Italy and eastern Europe. *Journal of Volcanology and Geothermal Research* **177**, 208-226 (2008).
22 Dupont, E. *L'Homme pendant les âges de la pierre dans les environs de Dinant-sur-Meuse*, *2ème édition* (C. Muquardt Ed., 1872).
23 Flas, D. *La transition du Paléolithique moyen au supérieur dans la plaine septentrionale de l'Europe. Les problématiques du Lincombien-Ranisien-Jerzmanowicien*. PhD dissertation (Université de Liège, 2006).
24 Posth, C. *et al.* Pleistocene mitochondrial genomes suggest a single major dispersal of Non-Africans and a Late Glacial population turnover in Europe. *Current Biology* **26,** 827-833 (2016).
25 Rougier, H. *et al.* First evidence of Neandertal cannibalism in Northern Europe. (subm.).
26 Wißing, C. *et al.* Isotopic evidence for dietary ecology of Late Neandertals in North-Western Europe. *Quaternary International*, doi:10.1016/j.quaint.2015.09.091 (2015).
27 Rougier, H., *et al.* The first Upper Paleolithic human remains from Belgium: Aurignacian, Gravettian and Magdalenian fossils at the "Troisième caverne" of Goyet. *PaleoAnthropology* **A33** (2013).

28 Conard, N. J., Malina, M. & Munzel, S. C. New flutes document the earliest musical tradition in southwestern Germany. *Nature* **460**, 737-740 (2009).
29 Conard, N. J. A female figurine from the basal Aurignacian of Hohle Fels Cave in southwestern Germany. *Nature* **459**, 248-252 (2009).
30 Housley, R. A., Gamble, C. S., Street, M. & Pettitt, P. B. Radiocarbon evidence for the Lateglacial human recolonisation of Norther Europe. *Proc. Prehist. Soc.* **63**, 25-54 (1997).
31 Haas-Campen, S. *Die menschlichen Skelettreste des Spätpleistozäns und Frühholozäns in Baden-Württemberg*, Magisterarbeit (1993).
32 Simon, U. *Die Burkhardtshöhle - eine Magdalénienstation am Nordrand der Schwäbischen Alb*, Magisterarbeit (1993).
33 Wehrberger, K. "Der Streit ward definitiv beendet..." Eine mesolithische Bestattung aus der Bocksteinhöhle im Lonetal, Alb-Donau-Kreis. Zur Erinnerung an Ludwig Bürger (1844-1898). *Archäologisches Korrespondenzblatt* **30**, 15-31 (2000).
34 Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137-140 (2009).
35 Orschiedt, J. Manipulationen an menschlichen Skelettresten. Taphonomische Prozesse, Sekundärbestattungen oder Kannibalismus. Dissertation. *Urgeschichtliche Materialhefte* **13** (1999).
36 Holdermann, C.-S. & Kind, C.-J. Zeitwechsel in Schichten. Bedeutende Fundstellen an der oberen Donau. *Eiszeit. Kunst und Kultur. Begleitband zur Großen Landesausstellung Eiszeit - Kunst und Kultur im Kunstgebäude Stuttgart*, 332–335 (2009).
37 Mollison, T. Zeichen gewaltsamer Verletzungen an den Ofnet-Schädeln. *Anthrop. An.* **13**, 79-88 (1936).
38 Cupillard, C. *et al.* Changes in ecosystems, climate and societies in the Jura Mountains between 40 and 8 ka cal BP. *Quaternary International* **378**, 40-72 (2015).
39 Valentin, F. Les restes humains de l'abri des Cabônes à Ranchot (Jura). In : Les derniers chasseurs-cueilleurs (13000-5500 av. JC) dans le massif du Jura et ses marges. *Centre Jurassien du Patrimoine*, 185-186 (1998).
40 Cupillard, C., Malgarini, R. & Fornage-Bontemps, S. Le Paléolithique supérieur ancien dan le quart nord-est de la France : l'exemple de la Franche-Comté. Environnement, chronologie et faciès culturels. *Mémoire LVI de la Société préhistorique française*, 351-363 (2013).
41 Bosset, G. & Valentin, F. Mesolithic burial practices in the northern half of France: Isolated burials and their spatial organisation. *Paris: Société préhistorique française*, 207-216 (2013).
42 Valentin, F. et al. Découvertes récentes d'inhumations et d'une incinération datées du Mésolithique en Île-de-France. Revue Archéologique d'Île-de-France **1,** 21–42 (2008).
43 Ilett, M. Cuiry-lès-Chaudardes "les Fontinettes" (Ministère de la Culture, Université de Paris 1-Panthéon Sorbonne, 1998).
44 Auxiette, G. & Hachem, L. Berry-au-Bac, "Le Chemin de la Pêcherie ouest", "le Vieux-Tordoir", "la culée" (Ministère de la Culture, Fouille Protohistorique de la Vallée de l'Aisne, 1989).

45 Gely, B. & Morand, P. Les sépultures épipaléolithiques de l'aven des Iboussières à Malataverne (Drôme, France): Premiers résultats. *Ardèche Archéologie* **15**, 13-18 (1998).
46 Soficaru, A., Dobo**ş**, A. & Trinkaus, E. Early modern humans from the Peştera Muierii, Baia de Fier, Romania. Proc. Natl Acad. Sci. USA **103,** 17196–17201 (2006).
47 Straus, L. G., González Morales, M.R. (eds.) El Mirón Cave, Cantabrian Spain. *University of New Mexico Press* (2012).
48 Straus, L. G., Gonzalez Morales, M.R., Carretero, J.M. (eds.) The Red Lady of El Miron Cave': Lower Magdalenian Human Burial in Cantabrian Spain. *J. Archaeol. Sci.* **60**, 1-137 (2015).
49 Straus, L. G., Gonzalez Morales, M. R., Higham, T., Richards, M. & Talamo, S. Radiocarbon dating the late Upper Paleolithic of Cantabrian Spain: El Miron Cave Date List IV. *Radiocarbon* **57**, 183-188 (2015).
50 Aimar, A., et al. . Les Abris Villabruna dans la Vallèe du Cismon. *Preistoria Alpina* **28**, 227- 254 (1992).
51 Vercellotti, G., Alciati, G., Richards, M. P. & Formicola, V. The Late Upper Paleolithic skeleton Villabruna 1 (Italy): a source of data on biology and behavior of a 14.000 year-old hunter. *Journal of Anthropological Sciences* **86**, 143-163 (2008).
52 Oxilia, G. *et al.* Earliest evidence of dental caries manipulation in the Late Upper Palaeolithic. *Scientific Reports* **5**, 12150 (2015).
53 Sinitsyn, A. A. Kostenki 14 (markina gora): Data, problems, and perspectives. *Prehistoire Europeene*, 273-313 (1996).
54 Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current Biology* **20**, 231-236 (2010).
55 Marom, A., McCullagh, J. S., Higham, T. F., Sinitsyn, A. A. & Hedges, R. E. Single amino acid radiocarbon dating of Upper Paleolithic modern humans. *Proc Natl Acad Sci U S A* **109**, 6878-6881 (2012).
56 Sinitsyn, A. A. Les sépultures de Kostenki : chronologie, attribution culturelle, rite funéraire. *In OTTE M. (dir.), La spiritualité, Proceedings of UISPP conference, Liege, ERAUL* **106**, 237-244 (2004).
57 Coppola, D. Il riparo di Agnano nel paleolitico superiore. La sepoltura Ostuni ed i suoi simboli, vol.1. . *Università di Roma Tor Vergata, Roma.* (2012).
58 Talamo, S. & Richards, M. A comparison of bone pretreatment methods for AMS dating of samples >30,000 BP. *Radiocarbon* **53**, 443-449 (2011).
59 Serradimigni, M. Grotta Continenza (Trasacco, Prov. de l'Aquila). *Notiziario di Preistoria e Protostoria* **2014 -1.I, 12-14** (2014).
60 Grifoni Cremonesi, R., Serradimigni, M. & Usala, M. . in Il Fucino e le aree limitrofe nell'antichità. *Atti del III Convegno di Archeologia in ricordo di Walter Cianciusi, Avezzano* (2011).
61 Haendel, M., Einwoegerer, T., Simon, U. & Neugebauer-Maresch, C. Krems-Wachtberg excavations 2005-12: Main profiles, sampling, stratigraphy, and site formation. *Quaternary International* **351**, 38-49 (2014).
62 Einwogerer, T. *et al.* Upper Palaeolithic infant burials. *Nature* **444**, 285 (2006).

Section 2
Ancient DNA processing and quality control

DNA extraction and library preparation

All ancient DNA extracts were prepared using the method described in ref. [1], starting from between 28 and 350 mg of bone or tooth powder.

A total of 44 libraries generated at the Max Planck Institute for Evolutionary Anthropology (MPI-EVA) in Leipzig were prepared using a single-stranded protocol[2,3]. All but three were treated with uracil-DNA-glycosylase (UDG) from *Escherichia coli* and endonuclease (Endo VIII)[4]. These libraries retain characteristic ancient DNA damage at the terminal 5' nucleotide as well as the two terminal 3' nucleotides[2], and we denote them as "ss UDG". We denote the three MPI-EVA libraries prepared without UDG-treatment as "ss noUDG" (Table S2.1).

Table S2.1: Libraries prepared at MPI-EVA in Leipzig

Sample	Library	Ext-ract	% endo-genous	SNPs	mt cov	Match consens. Est.	Match consens. 95% CI	mt. Hg	Library type	Dam-age	Chr. X contam.	Library decision*	Ana-lyzed?
Vestonice13	A5280	E1500	1.94	2.2M	93	1	0.91-1	U8c	ss UDG	0.37	5.5%	Damage	Yes
Vestonice14	A5281	E1501	n/a	390k	114	1	0.9-1	H7	ss UDG	0.2	n/a	Damage	Yes
Vestonice14	A5282	E1798	0.37	390k	81	0.99	0.76-1	H7a1	ss UDG	0.34	n/a	Damage	Yes
Vestonice15	A5271	E1502	0.47	2.2M	88	1	0.91-1	U5	ss UDG	0.36	16.0%	Damage	Yes
Vestonice16	A5188	E1799	0.1	2.2M	97	1	0.91-1	U	ss UDG	0.22	27.4%	Fail	Yes
Vestonice16	A5272	E1503	1.57	2.2M	115	1	0.92-1	U	ss UDG	0.27	26.3%	Fail	Yes
Vestonice16	A5306	E1789	0.76	2.2M	108	1	0.92-1	U5	ss UDG	0.3	1.9%	Pass	Yes
Vestonice16	A5307	E1799	1.28	390k	121	0.99	0.92-1	U	ss UDG	0.22	23.6%	Fail	Yes
Vestonice42	A5284	E1504	0.28	390k	107	0.99	0.87-1	H	ss UDG	0.19	34.1%	Damage	No
Vestonice43	A5287	E1505	0.18	390k	85	0.99	0.84-1	U	ss UDG	0.3	33.3%	Fail	Yes
Vestonice43	A5308	E1800	0.89	2.2M	102	1	0.92-1	U5	ss UDG	0.33	9.0%	Damage	Yes
ElMiron	A5268	E1796	2.58	2.2M	104	1	0.92-1	U5b	ss UDG	0.33	n/a	Pass	Yes
ElMiron	A5279	E1796	2.73	2.2M	105	0.98	0.88-1	U5b	ss UDG	0.39	n/a	Pass	Yes
ElMiron	A5301	E1796	2.69	2.2M	133	0.99	0.9-1	U5b	ss UDG	0.34	n/a	Pass	Yes
Continenza	A5189	E1788	n/a	2.2M	33	0.99	0.66-1	U5b1	ss UDG	0.24	n/a	Damage	Yes
Continenza	A5206	E1788	0.03	2.2M	43	0.99	0.83-1	U5b1	ss UDG	0.58	n/a	Damage	Yes
Continenza	A5207	E1788	0.03	2.2M	43	0.99	0.8-1	U5b1	ss UDG	0.32	n/a	Damage	Yes
Continenza	A5213	E1788	0.03	390k	21	0.96	0.47-0.99	U5b1	ss UDG	0.26	n/a	Fail	Yes
Continenza	A5214	E1788	0.02	390k	26	0.84	0.23-0.95	U5b1	ss UDG	0.33	n/a	Fail	Yes
Kostenki12	A5212	E1794	1.81	390k	11	0.92	0.03-0.98	U	ss UDG	0.5	n/a	Fail	Yes
Kostenki12	A5273	E1785	0.07	2.2M	20	0.76	0.16-0.96	U2	ss UDG	0.32	11.8%	Fail	Yes
Kostenki12	A5275	E1785	0.14	390k	6	n/a	n/a	R	ss UDG	0.28	n/a	Fail	Yes
Kostenki12	A5276	E1648	0.66	2.2M	43	0.99	0.72-1	U2	ss UDG	0.32	n/a	Pass	Yes
Kostenki12	A5288	E1844	0.07	2.2M	40	0.98	0.6-1	U2	ss UDG	0.27	15.0%	Fail	Yes
Kostenki14	A5187	E1782	n/a	2.2M	89	1	0.88-1	U2	ss UDG	0.22	2.1%	Pass	Yes
Kostenki14	A5195	E1781	8.6	390k	120	0.97	0.86-1	U2	ss UDG	0.48	1.2%	Pass	Yes
Kostenki14	A5196	E1781	9.65	2.2M	125	1	0.93-1	U2	ss UDG	0.21	1.5%	Pass	Yes
Kostenki14	A5197	E1781	8.66	2.2M	112	0.99	0.88-1	U2	ss UDG	0.2	1.5%	Pass	Yes
Kostenki14	A5198	E1781	8.48	2.2M	122	1	0.89-1	U2	ss UDG	0.2	2.0%	Pass	Yes
Kostenki14	A5278	E1781	12.15	390k	110	0.97	0.84-1	U2	ss UDG	0.33	1.2%	Fail	Yes
Kostenki14	A5305	E1782	4.72	2.2M	113	1	0.91-1	U2	ss UDG	0.23	1.8%	Pass	Yes
Muierii2	A5270	E1407	0.67	390k	27	0.57	0.39-0.73	U	ss UDG	0.23	n/a	Fail	Yes
Muierii2	A5289	E1850	1.99	390k	25	0.66	0.44-0.79	U	ss UDG	0.23	n/a	Fail	Yes
Muierii2	A5299	E1851	1.09	2.2M	109	0.93	0.88-0.97	U6	ss UDG	0.23	n/a	Damage	Yes
Oberkassel	A5302	E1783	0.09	390k	62	0.99	0.75-1	U5b1	ss UDG	0.24	n/a	Damage	No
Oberkassel	A5303	E1783	0.08	390k	60	0.86	0.34-0.99	U5b1	ss UDG	0.25	n/a	Damage	No
Ostuni1	A5180	E1750	1.04	2.2M	32	0.69	0.52-0.84	M	ss UDG	0.41	n/a	Damage	Yes
Ostuni1	A5181	E1750	0.97	2.2M	44	0.88	0.74-0.94	M	ss UDG	0.41	n/a	Damage	Yes
Ostuni1	A5182	E1750	0.91	2.2M	61	0.87	0.77-0.93	M	ss UDG	0.4	n/a	Damage	Yes
Ostuni1	A5201	E1751	0.76	2.2M	93	0.89	0.8-0.93	M	ss UDG	0.33	n/a	Damage	Yes
Ostuni1	A5265	E1750	1.06	2.2M	48	0.77	0.65-0.86	M	ss UDG	0.4	n/a	Damage	Yes
Ostuni2	A5203	E1752	0.08	2.2M	67	0.83	0.56-0.94	U2	ss UDG	0.46	19.5%	Damage	Yes
Pavlov1	A5277	E1506	0.17	2.2M	83	0.99	0.89-1	U	ss UDG	0.34	25.3%	Damage	Yes
Pavlov1	A5304	E1801	0.09	2.2M	80	0.99	0.86-1	U5	ss UDG	0.35	11.8%	Damage	Yes
Villabruna	A5290	E1845	n/a	2.2M	92	1	0.91-1	U5b2b	ss noUDG	0.45	1.60%	Pass	Yes
Villabruna	A5294	E1849	n/a	2.2M	104	0.95	0.93-0.96	U5b2b	ss noUDG	0.37	5.90%	Fail	Yes
AfontovaGora3	L5121	E2670	1.36	2.2M	2504	0.99	0.97-0.99	R1b	ss noUDG	0.37	n/a	Pass	Yes

Note: Cells in red have mitochondrial contamination estimates that fail our criteria: <95% for the point estimate, or a ninety five percent confidence interval that includes values <85%. The final column indicates whether the sample was among the 51 used in the genome-wide analysis (some samples were not used because there were <4,000 SNPs after quality control).
* Some samples do not have an X chromosome contamination estimate because of likely female sex or because fewer than 200 SNPs were covered at least twice.
+ Libraries marked "Fail" are not used in analysis. For libraries marked "Damage," analyses are restricted to fragments showing deamination.

A total of 46 libraries generated at the University of Tübingen were prepared by both single-stranded and double-stranded protocols[5,6]. The single-stranded libraries were prepared using the same protocols as the libraries at MPI-EVA in Leipzig. The double-stranded libraries were prepared either without UDG-treatment ("ds noUDG"), or with a protocol that also removes deaminated cytosines at the final nucleotide ("ds UDG") (Table S2.2).

Table S2.2: Libraries prepared at the University of Tübingen

Sample	Library	Extract	Endogenous%	SNPs	mt cov	Match consens. Est.	Match consens. 95% CI	mt. Hg	Library type	Dam-age	*Chr. X contam.	Library decision+	Analyzed?
Iboussieres39	GA162	GX45	n/a	390k	31	0.92	0.85-0.96	U5b2b	ds UDG	0.04	39.7%	Fail	Yes
	GA77	GX45	0.5	390k					ds noUDG	0.43	n/a	Damage	Yes
Paglicci108	GA264	B1	4.23	1240k	20	0.93	0.86-0.98	U2'3'4'7'	ds noUDG	0.12	n/a	Damage	Yes
BerryAuBac	GA261	GX81	0.2	1240k	107	0.97	0.94-0.99	U5b1a	ds noUDG	0.46	-1.6%	Pass	Yes
Bockstein	GA165	GX37	n/a	390k	267	0.97	0.95-0.99	U5b1d1	ds UDG	0.03	n/a	Fail	Yes
	GA89	GX37	2.67	390k					ds noUDG	0.25	n/a	Pass	Yes
Brillenhohle	GA163	GX52	n/a	390k	19	0.9	0.82-0.95	U8a	ds UDG	0.01	26.2%	Fail	Yes
	GA79	GX52	1.08	390k					ds noUDG	0.16	14.3%	Damage	Yes
Burkhardtshohle	GA260	GX53	0.35	1240k	45	0.95	0.89-0.98	U8a	ds noUDG	0.26	15.7%	Damage	Yes
Paglicci133	GA252	C2	3.67	1240k	28	0.83	0.74-0.89	U8c	ds noUDG	0.41	-0.7%	Pass	Yes
	MA160	C2	n/a	1240k					ss noUDG	0.08	n/a	Fail	Yes
Cioclovina1	GA259	GX51	0.21	1240k	19	0.88	0.76-0.95	U	ds noUDG	0.20	15.9%	Damage	Yes
Chaudardes1	CRC	CRC	1.65	1240k	17	0.93	0.91-0.95	U5b1b	ds noUDG	0.35	18.3%	Damage	Yes
	MA169	CRC	n/a	1240k					ss UDG	0.28	n/a	Damage	Yes
Paglicci71	GA265	FA	0.27	1240k	12	0.83	0.68-0.92	U5b2b	ds noUDG	0.21	44.4%	Damage	No
Falkenstein	FLA	FL	5.45	390k	600	0.97	0.97-0.98	U5a2c	ds noUDG	0.23	7.7%	Damage	Yes
Falkenstein	GA53	FL	n/a	390k					ds UDG	0.02	9.6%	Fail	Yes
Falkenstein	GA54	FL	n/a	390k					ds UDG	0.02	6.3%	Fail	Yes
Felsdach	GA258	GX49	0.5	1240k	55	0.76	0.67-0.80	U5b2a	ds noUDG	0.04	47.2%	Damage	No
HohleFels49	GA164	GX55	n/a	390k	364	0.99	0.98-0.99	U8a	ds UDG	0.03	5.5%	Fail	Yes
	GA82	GX55	1.2	390k					ds noUDG	0.31	4.9%	Damage	Yes
HohleFels79	GA166	GX47	n/a	390k	42	0.98	0.95-1.00	U8a	ds UDG	0.02	4.2%	Fail	Yes
	GA90	GX47	1.52	390k					ds noUDG	0.19	n/a	Damage	Yes
HohlensteinStadel	MA162	VE	n/a	1240k	34	0.77	0.69-0.84	U5b2c1	ss noUDG	0.14	n/a	Damage	No
	MA171	VE	n/a	1240k					ss noUDG	0.15	n/a	Damage	No
LaRochette	GA253	LA	0.28	1240k	40	0.80	0.67-0.88	M	ds noUDG	0.09	7.0%	Damage	No
LesCloseaux13	GA256	GX43	0.17	1240k	19	0.98	0.91-1.00	U5a2	ds noUDG	0.27	n/a	Pass	Yes
Ofnet	GA167	GX50	n/a	390k	1859	1	1.00-1.00	U5b1d1	ds UDG	0.02	n/a	Fail	Yes
	GA93	GX50	1.25	390k					ds noUDG	0.21	n/a	Pass	Yes
GoyetQ116-1	GA63	GX58	4.88	1240k	56	0.99	0.97-1.00	M	ds noUDG	0.22	1.0%	Pass	Yes
	MA167	GX58	n/a	1240k					ss UDG	0.25	0.9%	Pass	Yes
GoyetQ-2	GA231	GX176	8.6	1240k	406	1	0.99-1.00	U8a	ds noUDG	0.30	5.5%	Damage	No
	MA166	GX176	n/a	1240k					ss UDG	0.28	3.9%	Damage	No
Goyet2878-21	GA248	GX177	1.76	1240k	21	0.99	0.95-1.00	U5	ds noUDG	0.16	n/a	Damage	No
	MA168	GX177	n/a	1240k					ss UDG	0.11	8.0%	Damage	No
GoyetQ376-19	GA250	GX60	0.9	1240k	43	0.92	0.86-0.95	U2	ds noUDG	0.13	n/a	Damage	Yes
	MA158	GX60	n/a	1240k					ss noUDG	0.26	27.6%	Damage	Yes
GoyetQ376-3	MA161	GX59	0.27	1240k	46	0.9	0.82-0.94	M	ss noUDG	0.17	n/a	Damage	No
	MA170	GX59	n/a	1240k					ss noUDG	0.24	-7.1%	Damage	No
GoyetQ53-1	GA251	GX64	0.62	1240k	48	0.82	0.73-0.88	U2	ds noUDG	0.13	32.4%	Damage	Yes
	MA159	GX64	n/a	1240k					ss noUDG	0.16	6.9%	Damage	Yes
GoyetQ55-2	GA254	GX62	0.1	1240k	17	0.87	0.78-0.92	U2	ds noUDG	0.16	n/a	Damage	No
GoyetQ56-16	GA255	GX63	0.19	1240k	45	0.84	0.77-0.89	U2	ds noUDG	0.18	24.6%	Damage	Yes
Ranchot88	GA262	GX83	20.65	1240k	86	0.99	0.96-1.00	U5b1	ds noUDG	0.39	n/a	Pass	Yes
Rigney1	GA263	GX89	0.53	1240k	41	0.9	0.86-0.93	U2'3'4'7'	ds noUDG	0.22	n/a	Damage	Yes
Rochedane	GA127	GX96	4.25	1240k	104	0.98	0.96-0.99	U5b2b	ds noUDG	0.39	5.6%	Damage	Yes
	MA165	GX96	n/a	1240k					ss UDG	0.38	1.6%	Pass	Yes

Note: Cells in red have mitochondrial contamination estimates that fail our criteria: <95% for the point estimate, or a ninety five percent confidence interval that includes values <85%.
The final column indicates whether the sample was among the 51 used in the genome-wide analysis (some samples were not used because there were <4,000 SNPs after quality control).
* Some samples do not have an X chromosome contamination estimate because of likely female sex or because fewer than 200 SNPs were covered at least twice.
+ Libraries marked "Fail" are not used in analysis. For libraries marked "Damage," analyses are restricted to fragments showing deamination.

Libraries generated at Harvard Medical School were prepared using double-stranded protocols and were UDG-treated in a way that retains some characteristic ancient DNA damage at the last nucleotide[7]. We denote these libraries as "ds partial UDG" in Table S2.3.

Table S2.3: Libraries prepared at Harvard Medical School

Sample	Library	Extract	Endogenous%	SNPs	mt cov	Match consens. Est.	Match consens. 95% CI	mt. Hg	Library type	Dam-age	Chr. X contam.	Library decision	Analyzed?
KremsWA3	S1577.E1.L2	S1577.E1	0.50	1240k	85	1.00	1.00-1.00	U5	ds partial UDG	0.084	Pass	n/a	Yes
	S1577.E1.L3	S1577.E1	0.39	1240k	43	1.00	1.00-1.00	U5	ds partial UDG	0.077	Pass	1.8%	Yes

In solution capture of mitochondrial DNA
We hybridized the libraries to oligonucleotide probes overlapping the mitochondrial DNA genome (mtDNA). We used the method of ref. [8] for the libraries from MPI-EVA and of ref. [9] for the libraries from Tübingen. We sequenced the enriched libraries on the Illumina MiSeq or HiSeq2500 platforms using a double index configuration (2×75bp or 2x100bp reads).

To analyse the mtDNA capture data for the MPI-EVA libraries, we demultiplexed the reads according to the expected index pairs, allowing one mismatch for each. We merged paired-end reads into a single fragment by requiring an overlap of at least 11 bp (with one mismatch allowed), using a modified version of SeqPrep[10] in which the base and quality score is determined by the read that has higher quality. After stripping adapters, we mapped merged fragments which we required to be at least 30bp in length to the mtDNA revised Cambridge Reference Sequence (rCRS) with BWA (v0.6.1) using the *samse* command. We identified duplicated fragments based on having the same orientation, start and end positions, and kept the highest quality fragment. We excluded fragments with mapping quality <30.

To analyse the mtDNA capture data for the University of Tübingen libraries, we clipped adapters and merged paired reads that overlapped by at least 10 bp using the program Clip&Merge[11]. We restricted to merged fragments, and filtered out ones less than 30 bp in length. We mapped fragments to the mtDNA revised Cambridge Reference Sequence (rCRS) and removed duplicates[11]. We excluded fragments with mapping quality below 30.

In solution capture of nuclear DNA for 38 samples
We hybridized libraries in solution to oligonucleotide probes synthesized by Agilent Technologies[8]. For 8 samples, we enriched for a targeted set of 394,577 SNPs (SNP Panel 1, "390k"), using the probe sequences specified in Supplementary Data 2 of ref. [12] (http://www.nature.com/nature/journal/v522/n7555/abs/nature14317.html#supplementary-information). For 16 samples, we enriched for ~1.24 million SNPs (SNP Panels 1 and 2, "1240k"), and for 14 samples we enriched for ~3.7 million SNPs (SNP Panels 1, 2, 3 and 4, "3.7M"), using the additional probe sequences specified in Supplementary Data 1, 2 and 3 of [13] (http://www.nature.com/nature/journal/v524/n7564/full/nature14558.html#supplementary-information). Extended Data Table 1, as well as Table S2.1, Table S2.2 and Table S2.3, specify which samples were enriched for which targeted set of SNPs.

Sequencing and alignment to the nuclear genome
We generated 2×75bp reads on Illumina HiSeq2500 or NextSeq500 instruments. We processed the fragments as for mtDNA, except we aligned to the human reference genome, *hg19,* and required an overlap of at least 15 bp. We mapped with the command bwa –n 0.01 and –l 16500.

Four tests for contamination

(1) We required all analysed libraries to have a damage profile consistent with ancient DNA
All single stranded libraries ("ss UDG" and "ss noUDG"), as well as all non-UDG-treated double stranded libraries ("ds noUDG") retain damage in the last nucleotide. We restricted analyses of such libraries to ones in which ≥10% of terminal nucleotides that are cytosines in the reference genome are read as thymines, as expected for authentic ancient DNA[14]. For the "ds partial UDG" protocol, we restricted analysis to libraries that had a rate of ≥3% cytosine-to-thymine substitutions in the terminal nucleotide[7]. To be conservative, for our whole

genome analyses, we did not use any data from double stranded UDG-treated libraries ("ds UDG"), as this protocol does not retain damage at the terminal nucleotide.

(2) We tested for contamination based on the match rate to the mtDNA consensus
We used mtDNA data to flag samples as possibly contaminated (in red in Table S2.1 and Table S2.2). We declared a library possibly contaminated if, after running the contamination estimator ContamMix[8], either of two criteria were met:
(i) The fraction of fragments matching the reconstructed consensus better than any of 311 worldwide mtDNA sequences used for comparison is <95%.
(ii) The ninety-five percent confident lower bound of the fraction of fragments matching the consensus better than any of 311 worldwide mtDNA sequences is <85%.

(3) We tested for contamination based on consistency of damaged and undamaged fragments
For each sample, we restricted to fragments with characteristic ancient DNA damage, and checked whether the population genetic affinities inferred from damaged fragments match those from the consensus of all fragments. In the case of mtDNA data, only *Vestonice14* showed a change in haplogroup comparing damaged fragments to all fragments, suggesting mtDNA contamination. Further evidence of contamination in *Vestonice14* comes from the fact that when we determine sex based on the proportion of Y chromosome fragments (Supplementary Information section 3), the sex for this individual switched from female when all fragments are analysed to male for damaged fragments only. The genetically determined sex of *Ostuni2, Paglicci108, GoyetQ53-1, GoyetQ376-19, GoyetQ56-16, GoyetQ55-2* and *Goyet2878-21* changes from male when all fragments are analysed to female for damaged fragments only.

(4) We tested for contamination based on the X chromosome polymorphism rate in males
Males have only one X chromosome and thus are not expected to be polymorphic in this part of their genome. This can be used to obtain a conservative estimate of contamination in males given sufficient X chromosome coverage[15,16]. We used the ANGSD software to run this test on all males where it gave good resolution (we only used X chromosome estimates for males for whom at least 200 SNPs covered at least twice). We considered libraries as effectively uncontaminated if their X chromosome contamination estimates were less than 2.5%.

Selecting libraries for analysis and restricting to damaged fragments
For nine male individuals, we identified libraries with no evidence of contamination by the four criteria above, and with high enough coverage to perform a chromosome X contamination estimate and to determine that the contamination was less than 2.5%. We used all fragments from these libraries. For the remaining individuals, we restricted to fragments carrying characteristic ancient DNA damage at their terminal ends, which is known to reduce contamination albeit at the cost of losing data[17,18] (Box S2.1). Specifically:

(i) We restricted to damaged fragments for all libraries with evidence of contamination according to criteria 1-4, and that we did not fail outright (Table S2.1 and Table S2.2). We failed libraries outright if (a) they had evidence of contamination and were from samples that had multiple libraries some of which had no evidence of contamination, or (b) if they were prepared with a double-stranded UDG-treated protocol ("ds UDG"), as damage is not retained for this class of libraries even in the last nucleotide.

(ii) We restricted to damaged fragments for libraries with evidence of contamination based on criteria 1-3, and for which we could not perform an X contamination estimate. While this is a severe step—for example, it means that we restricted to damaged fragments for all

female samples—we decided that it was the only safe thing to do in cases where we could not perform an X chromosome contamination assessment. In particular, we found that a substantial fraction of male samples with no evidence of contamination by criteria 1-3 gave evidence of substantial contamination by the X chromosome method.

After damage restriction, we merged fragments from all libraries from the same sample. At each SNP covered at least once, we used a randomly sampled fragment to determine an allele. Thus, we do not attempt to determine diploid genotypes, and none of the SNPs are assigned a heterozygous genotype. The final dataset is given in Extended Data Table 1, after restricting to samples with at least 4,000 SNPs hit at least once.

Box S2.1. Strategy used to retain damaged fragments for contaminated libraries

ss UDG-treated libraries: Restrict to fragments with C→T substitutions in the first position at the 5'-end and the last two positions at the 3'-end.

ss noUDG-treated libraries: Restrict to fragments with C→T substitutions in the first three positions at the 5'-end and the last three positions bases at the 3'-end.

ds UDG- partial treated libraries: Restrict to fragments with C→T substitutions in the first position at the 5'-end and G→A substitutions in the last position at the 3'-end.

ds noUDG-treated libraries: Restrict to fragments with C→T substitutions in the first three positions at the 5'-end, and G→A substitutions in the last three positions on the 3'-end.

ds UDG-treated libraries: Cannot restrict to damaged fragments so do not use.

References

1 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* **110**, 15758-15763, doi:10.1073/pnas.1314445110 (2013).

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

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

4 Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14616-14621, doi:10.1073/pnas.0704665104 (2007).

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

6 Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols* **2010**, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).

7 Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**, 20130624, doi:10.1098/rstb.2013.0624 (2015).

8 Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

9 Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PloS one* **5**, e14004, doi:10.1371/journal.pone.0014004 (2010).
10 John, J. S. *SeqPrep*, <https://github.com/jstjohn/SeqPrep> (2011).
11 Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome biology* **17**, 60, doi:10.1186/s13059-016-0918-z (2016).
12 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, doi:10.1038/nature14317 (2015).
13 Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, doi:10.1038/nature14558 (2015).
14 Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PloS one* **7**, e34131, doi:10.1371/journal.pone.0034131 (2012).
15 Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98, doi:10.1126/science.1211177 (2011).
16 Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics* **15**, 356, doi:10.1186/s12859-014-0356-4 (2014).
17 Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 2229-2234, doi:10.1073/pnas.1318934111 (2014).
18 Meyer, M. *et al.* A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**, 403-406, doi:10.1038/nature12788 (2014).

Section 3
Evidence for a decrease in Neanderthal ancestry over time

Change in Neanderthal ancestry over time as inferred from f_4-ratio statistics

We used f_4-ratio statistics[1] to estimate the proportion of Neanderthal ancestry in the 51 ancient Eurasian samples analysed in this study as well as in a handful of present-day Europeans and East Asians for comparison. We used the dataset of approximately 2.2 million SNPs (the SNPs captured in Panel 1, Panel 2 and Panel 3) (Supplementary Information section 2).

To estimate Neanderthal ancestry proportion in each individual, we used the previously reported f_4-ratio statistic shown in Equation S3.1[2-4]. Under the assumption that West and Central African populations are outgroups to East African *Dinka* and to the modern human ancestry in non-Africans today, this quantity provides an unbiased estimate of the proportion of modern human ancestry in a present-day non-African. We use 1 minus this quantity to estimate the Neanderthal ancestry. We compute a Block Jackknife on non-overlapping 5 centimorgan blocks to determine empirical standard errors[5]. There are a number of f_4-ratio statistics that have been proposed for estimating Neanderthal ancestry, and while each has different merits, some have large standard errors[4]. We use Equation S3.1 as we find empirically that it has relatively small standard errors.

$$Q(Test) = 1 - \frac{f_4(West\_and\_Central\_Africans, Chimp; Test, Archaic)}{f_4(West\_and\_Central\_Africans, Chimp; Dinka, Archaic)} \qquad \text{(Equation S3.1)}$$

We compute allele frequencies by pooling data from each of the following sets of samples:

West_and_Central_Africans: a pool of 9 samples from the Mbuti, Yoruba and Mende populations (S_Mbuti-1, S_Mbuti-2, S_Mbuti-3, B_Mbuti-4, S_Yoruba-1, S_Yoruba-2, S_Yoruba-3, S_Mende-1, S_Mende-2)

Dinka: a pool of 3 samples (S_Dinka-1, S_Dinka-2, B_Dinka-3)

Archaic: a pool of 2 samples (Altai Neanderthal and the Siberian Denisovan)

For *Test*, we analyse 51 ancient individuals along with 8 Europeans (S_French-1, S_French-2, B_French-3, S_English-1, S_English-2, S_Sardinian-1, S_Sardinian-2, B_Sardinian-3) and 7 East Asians (S_Dai-1, S_Dai-2, S_Dai-3, B_Dai-4, S_Han-1, S_Han-2, B_Han-3).

Extended Data Table 2 presents the results for all samples, while Figure 2 plots the results for the subset of individuals with at least 200,000 SNPs covered (these gave low enough standard errors to produce a visually clear plot). Figure 2 reveals a clear decrease in Neanderthal ancestry proportion over time. There is one outlier—*Oase1*—which has previously been shown to have a Neanderthal ancestor 4-6 generations back in its family tree[4], and which we estimate here had 9.9 ± 0.8% Neanderthal ancestry. Such an ancestor is recent enough that *Oase1* is not expected to be representative of the population in which he lived. Here we are interested in how Neanderthal ancestry proportion changed over time in populations that reached equilibrium; that is, in populations in which all individuals had approximately the same Neanderthal ancestry proportion. Thus, we remove *Oase1* for most of the analyses that

follow. This is conservative for testing whether there has been a decrease in the proportion of Neanderthal ancestry over time.

To test whether the slope of the fitted line is significantly negative, we needed to compute a standard error on the slope taking into account the fact that we had information from multiple individuals with different amounts of data and different extents of correlated history. To do this, we used a Weighted Block Jackknife. As this is a novel context in which to use a Weighted Block Jackknife for population history analysis, we describe this in more detail.

Let n_i be our estimate of the mean Neanderthal ancestry in sample i using the f_4-ratio. We estimate the covariance matrix $\boldsymbol{V}$ of the errors using a Weighted Block Jackknife[5,6] (with 5 centimorgan blocks and weights equal to the number of SNPs used). Set $\boldsymbol{Q} = \boldsymbol{V^1}$. If a is the ancestry proportion now (time=0) and s is the "slope" in units of ancestry per year, then we can model n_i as:

$$n_i = a + sd_i + e_i \qquad \text{(Equation S3.2)}$$

Here, d_i is the date in years (BP) of sample i and the error e_i has covariance $\boldsymbol{V}$. It is now natural to estimate a and s by minimizing:

$$L(a,s) = \sum_{i,j} Q_{ij}\,(n_i - a - sd_i)\big(n_j - a - sd_j\big) \qquad \text{(Equation S3.3)}$$

This is a generalized least squares problem. If we set $\boldsymbol{L_2}=max(\boldsymbol{L(a, s)})$ and $\boldsymbol{L_1}=max(\boldsymbol{L(a, 0)})$, then $(L_1\text{-}L_0)$ is approximately χ^2 with 1 degree of freedom under the null (s=0).

Extended Data Table 3 shows the results of the regression analysis for different subsets of samples. Our "Core Set 1" of individuals consists of 50 ancient modern humans (removing *Oase1* as an outlier) along with 7 East Asians that we use to represent present-day individuals. We use East Asians rather than Europeans to represent present-day individuals, since present-day East Asians are known to harbor more Neanderthal ancestry than Europeans[3,7,8]. Using East Asians to represent present-day people is conservative for a test that searches for evidence of a decrease in Neanderthal ancestry proportion over time.

For our Core Set 1 of samples, we observe a highly significant $P=5\times10^{-22}$) correlation of Neanderthal ancestry with sample date. We also observe highly statistical significant signals for a 11 alternate sample sets (always in the range ($10^{-29}<P<10^{-11}$):

- 5×10^{-22} for Core Set 1
- 2×10^{-15} remove the oldest samples (restrict to <32 kya)
- 4×10^{-18} remove most Věstonice Cluster samples (restrict to >32 kya or <25 kya)
- 5×10^{-21} remove the El Mirón and Mal'ta Cluster samples (restrict to >25 kya or <14 kya)
- 2×10^{-18} remove Villabruna Cluster, Neolithic samples (restrict to >14 kya or present-day)
- 4×10^{-15} remove present-day samples (restrict to ancient samples)
- 4×10^{-19} restrict to samples with >200,000 SNPs (same set of samples as in Figure 2)
- 2×10^{-23} replace East Asians with Europeans
- 8×10^{-29} add *Oase1*
- 1×10^{-20} restrict to ancient samples only (including *Oase1*)
- 8×10^{-12} restrict to ancient samples only (but excluding both *Oase1* and *UstIshim*)

Having established a significant decrease in Neanderthal over time, we used the parameters of the least squares fit to estimate the rate of decrease in Neanderthal ancestry. There is no theoretical reason to expect that the proportion of Neanderthal ancestry in modern humans decreased at a linear rate immediately after introgression. Indeed, reasonable models of selection against Neanderthal ancestry suggest that there might have been a much faster decrease in the proportion of Neanderthal ancestry under the pressure of natural selection for the first couple of dozens of generations after introgression[9]. However, the line is useful in providing a meaningful measurement of the rate of decrease in Neanderthal ancestry over most of the period since Neanderthals and modern humans interbred. By extrapolating to the time when introgression occurred, we can estimate what the Neanderthal ancestry proportion was some time after introgression, and can likely obtain a minimum estimate.

The results of the least squares fitting are summarizing in Extended Data Table 3. For the Core Set 1 of samples, we estimate a 0.48-0.73% decrease in Neanderthal ancestry per 10,000 years (95% confidence interval). We obtain similar intervals for other sample subsets.

To obtain an estimate of the proportion of Neanderthal DNA in the ancestral population some time after introgression occurred, we need to make an assumption about when Neanderthal introgression occurred. If we use a date of 55,000 years, the best estimate from the analyses of the *UstIshim*[10] and *Kostenki14*[11] genomes, we infer a 95% confidence interval of 4.3-5.7%. We also considered how uncertainty in the date of Neanderthal admixture—estimated to be 50,000-60,000 years ago in the analysis of the *UstIshim* genome—affects this estimate[10]. Assuming 50,000 years, we estimate 4.0-5.4%, and assuming 60,000 years, we estimate 4.5-6.0%). Thus, a conservative range is 4.0-6.0%.

The inferred 4.0-6.0% Neanderthal ancestry some time after introgression is substantially higher than the proportion of Neanderthal ancestry in Eurasian populations today, which in the same analyses we estimate to be 1.1-2.2%. This is higher than the values of ~3-3.5% considered by Juric et al.[12] and Harris and Nielsen[9]. An important direction for future work is to explore what distributions of selection coefficients and demographic scenarios might produce this inferred 2-4-fold magnitude of reduction of Neanderthal ancestry in present-day compared to ancient individuals.

Neanderthal ancestry estimates from ancestry informative SNPs

The f_4-ratio statistic analysis depends on assumptions about the deep historical relationships of the populations used in the statistic. To obtain an estimate of Neanderthal ancestry that is not strongly dependent on such assumptions, we analysed SNPs where Neanderthals carry an allele that differs from the great majority of present-day sub-Saharan Africans. Such alleles have a high chance of deriving from Neanderthal ancestors, and by counting these in each individual, it is possible to obtain a number that is linearly related to and highly sensitive to an individual's proportion of Neanderthal ancestry[3,4].

We used in-solution hybridization capture[13] to enrich the libraries from 15 individuals (the individuals listed in Table S2.1 in Supplementary Information section 2) for 1,749,385 SNPs where all Yoruba individuals in the 1000 Genomes Project are fixed or nearly fixed for one allele, and an archaic genome (a Denisovan or a Neanderthal) carries another allele[4]. For the analyses that follow, we focused on a subset of 783,747 specifically Neanderthal informative SNPs where a randomly selected allele from the high coverage Altai Neanderthal genome differs from the great majority of Yoruba in the 1000 Genomes project, and also from a randomly selected allele in the deeply sequenced Yoruba individual from "Panel B" of ref. [8].

We also further required that the analysed SNPs had an assignment of a B-statistic measuring the intensity of linked selection at the position[14].

We sequenced the enriched libraries on the Illumina HiSeq 2500 platform using a double index configuration (2 x 76bp), and spiked in an indexed ΦX174 control library in each sequencing run. We performed base calling using the machine-learning algorithm implemented in freeIBIS[15], merged overlapping pair-end reads[3], and mapped them to the human reference genome (*hg19*) using the Burrows-Wheeler Aligner (BWA)[16]. We adjusted BWA parameters to allow for more mismatches and indels and disabled seeding ("-n 0.01 –o 2 –l 16500"), as appropriate for analysis of error-prone ancient DNA sequences[3].

We restricted our analyses to merged fragments that had perfect matches to the expected index combinations. After removing fragments identified as being likely duplicates (having the same mapped orientation and start and stop positions) (https://github.com/udo-stenzel/biohazard), we retained for analysis fragments that were longer than 35bp and that had a mapping quality of at least 37. For some individuals, we further restricted to fragments with evidence of deamination (Extended Data Table 1). To do this for UDG-treated libraries, we retained fragments that showed C→T substitutions in the first base at the 5' and/or at the last two bases at the 3'-end relative to *hg19*. For the non-UDG treated libraries (from *AfontovaGora3*), we retained fragments with a C→T substitution at the first three and/or the last three bases. We merged data from all libraries from the same individual, and in Extended Data Table 2 report the number of fragments overlapping positions of interest and the number of SNPs covered at least once. We reprocessed the previously reported *Oase1*[2] data using the same workflow for consistency.

We co-analysed these data with shotgun sequence data from 12 ancient modern humans and a pool of three Vindija Neanderthals to represent the rate of matching to Altai for an individual with 100% Neanderthal ancestry (we obtained qualitatively similar results when replacing the Vindija Neanderthals with the Mezmaiskaya Neanderthal). We represented present-day humans using 7 individuals drawn from "Panel B" of ref.[8] (6 non-Africans and a Dinka East African to represent an individual with 0% Neanderthal ancestry).

To compute the percentage of alleles matching Neanderthal, we represented each individual by a single DNA fragment at each targeted SNP. For the high coverage individuals (*UstIshim* and the present-day humans), we randomly sampled one of the two alleles from the genotype file. For the low coverage individuals, we randomly sampled an allele using the same procedure described in Supplementary Information section 2.

We converted the fraction of alleles matching *Altai* for each *Test* individual into an estimate of Neanderthal ancestry by assuming that the proportion of Neanderthal ancestry can be estimated by how far a *Test* individual's rate of matching E_{Test} is along the path from an individual assumed to have 0% Neanderthal ancestry (E_{Dinka}, the value seen in a *Dinka* individual from East Africa), and 100% Neanderthal ancestry ($E_{Vindija}$, the value seen in then pool of three *Vindija* Neanderthals from Croatia[17]). Concretely, we used the equation:

$$R(Test) = \frac{E_{Test} - E_{Dinka}}{E_{Vindija} - E_{Dinka}} \qquad \text{(Equation S3.4)}$$

We computed a standard error on *R(Test)* using a Block Jackknife with 100 equal blocks[5].

Extended Data Table 2 and Extended Data Figure 1 reveal that *Oase1* has an elevated rate of alleles matching to Neanderthal. This replicates a previous finding[4] and is also qualitatively consistent with the findings from the f_4-ratio statistic analysis. We also replicate previous findings in estimating that *UstIshim* has an elevated rate of alleles matching Neanderthal compared to present-day humans[4].

We performed a least squares fit to the scatterplot of sample age versus estimated Neanderthal ancestry proportion. We repeated this procedure leaving out each of 100 equally sized contiguous chunks of SNPs in turn, and used jackknife statistics to compute a standard error on the two regression parameters (slope and intercept). We tested whether the slope is non-zero by evaluating the number of standard errors it is from zero. We then used this as a significance test for whether Neanderthal ancestry proportion has changed over time. We determined the intercept at 50,000, 55,000, and 60,000 years ago to obtain estimates of the proportion of Neanderthal ancestry shortly after introgression.

The results are shown in Extended Data Table 3. We observe a significant decrease in Neanderthal ancestry over time, whatever subset of samples we analyse, just as for the f_4 ratio analysis. For the "Core Set 2" of 29 samples (all the ancient samples for which we had data with the exception of *Oase1*, and adding in *Han*, *Dai*, and *Karitiana* to represent present-day Eurasians), we observe a significant ($P=4.0\times10^{-11}$) correlation of Neanderthal ancestry with sample date. We also observe significant signals for alternate sample sets:

- 4.0×10^{-11} for Core Set 2
- 1.1×10^{-4} remove *Han, Dai*, *Karitiana* and *Stuttgart*
- 1.6×10^{-4} remove *Han, Dai*, *Karitiana, Stuttgart, and UstIshim*

For the Core Set 2 group of samples, the least squares fit corresponds to an estimated 0.21-0.39% decrease in Neanderthal ancestry every 10,000 years (95% confidence interval based on a standard error from the Block Jackknife), and an extrapolated 3.2-4.2% Neanderthal ancestry 55,000 years ago. These results support the conclusion that the proportion of Neanderthal ancestry has decreased over time, consistent with the findings of the f_4-ratio analysis in the previous section, although the magnitude of the estimated reduction is smaller.

The reduction in Neanderthal ancestry has been faster closer to genes

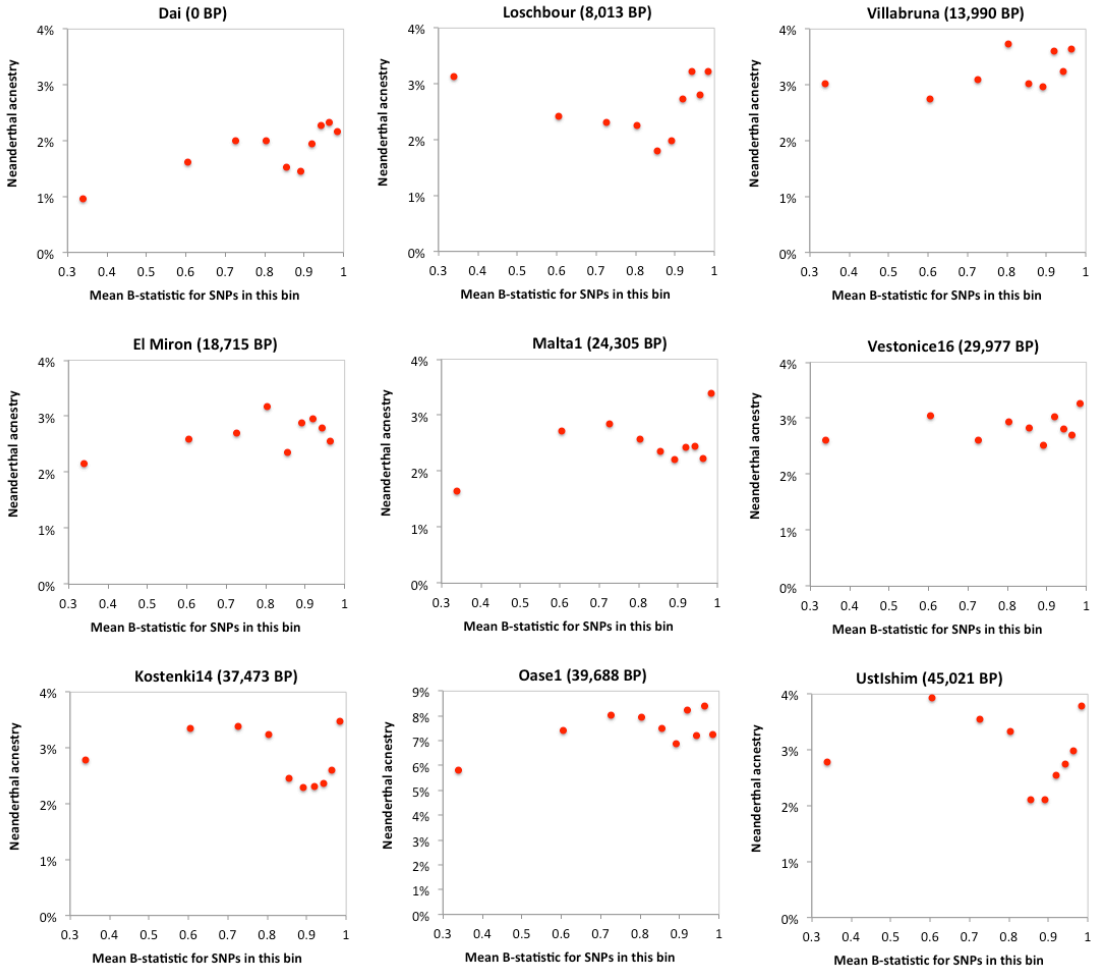
We took advantage of the power of ancestry informative SNPs to study whether there is evidence for a more rapid decrease in Neanderthal ancestry closer to genes. This pattern is predicted based on the reduced Neanderthal ancestry close to genes that is observed today[18,19].

To test whether the reduction in Neanderthal ancestry over time has been more rapid closer to genes, we stratified the genome based on a previously published *B* statistic, which provides a measure of the strength of loss of diversity in a region due to proximity to functionally important regions[14]. We divided the SNPs sites that we studied as being informative about ancestry into 10 bins of equal size based on B, ranging from most strongly constrained by selection (bin 1) to least constrained (bin 10). We calculated the proportion of Neanderthal ancestry in each bin, and converted it into an estimate of Neanderthal ancestry.

Figure S3.1 shows that if there is a trend in any sample toward an increase in Neanderthal ancestry with increasing B-statistic, it is very noisy. We fit a line to the individual-specific plots of Neanderthal ancestry against B-statistic and computed a standard error using Block Jackknife. The regression coefficients and standard errors are presented in Extended Data

Table 3. As expected from the noisy by-individual plots (Figure S3.1), the evidence for positive slopes is weak. The strongest evidence for a positive slope (as measured by a Z-score formed from the ratio of the estimate to the standard error), is Z=3.3 for *Dai*, followed by Z=2.7 for *Han*, Z=2.1 for *French*, and Z=1.6 for *Loschbour*. The strongest evidence for a negative slope is non-significant after multiple-hypothesis testing (Z=-2.4 for *Continenza*).

Figure S3.1. Empirical correlation of Neanderthal ancestry estimates to B-statistic bin. We show results for a selected set of relatively high coverage samples.



Despite the noisy individual estimates, we observed that the fitted slopes for the regressions against B-statistics tended to be positive (Extended Data Table 2), and that this was especially true for more recent samples. We were also motivated to study these patterns based on previous evidence for a significant depletion of Neanderthal ancestry in functional regions of present-day Europeans and East Asians based on Neanderthal introgression maps[18,19].

To test formally whether the dependence of Neanderthal ancestry on B-statistic is stronger for more recent periods, we carried out a least squares fit to the slopes of the "Main set" of samples in Extended Data Table 2, weighting the information from each sample by their empirical standard errors (which can also be read off of Extended Data Table 2). We used a Block Jackknife over the whole procedure to determine the standard error ("slope of slopes"). We observe a weakly significant reduction in the dependence on B-statistic with increasing sample date (Z=-2.32; P=0.010). These results are consistent with the theory that selection against Neanderthal alleles over time has been most intense closest to genes.

References

1 Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-494, doi:10.1038/nature08365 (2009).

2 Reich, D. *et al*. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:nature09710 [pii]

10.1038/nature09710 (2010).

3 Meyer, M. *et al*. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, doi:10.1126/science.1224344 (2012).

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

5 Kunsch, H. R. THE JACKKNIFE AND THE BOOTSTRAP FOR GENERAL STATIONARY OBSERVATIONS. *Annals of Statistics* **17**, 1217-1241, doi:10.1214/aos/1176347265 (1989).

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

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

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

9 Harris, K., Nielsen, R. The Genetic Cost of Neanderthal Introgression. *biorxiv.org* http://dx.doi.org/10.1101/030387 (2015).

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

11 Seguin-Orlando, A. *et al*. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, doi:10.1126/science.aaa0114 (2014).

12 Juric, I., Aeschbacher, S., Coop, C. The Strength of Selection Against Neanderthal Introgression. *biorxiv.org* http://dx.doi.org/10.1101/030148 (2015).

13 Fu, Q. *et al*. DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2223-2227, doi:10.1073/pnas.1221359110 (2013).

14 McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS genetics* **5**, e1000471, doi:10.1371/journal.pgen.1000471 (2009).

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

16 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

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

18 Sankararaman, S. *et al*. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-357, doi:10.1038/nature12961 (2014).

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

Section 4
Sex determination and Y chromosome analyses

Sex determination
For each individual, we extracted merged sequences of high mapping quality (MAPQ ≥ 37) using samtools 0.1.18. We counted the number of SNPs on the autosomal, chromosome X and chromosome Y targets covered by at least once sequence. We restricted this analysis to the smaller 1240k SNP target set for the samples captured for the 3.7M SNP target set (Extended Data Table 4).

It is tempting to interpret the ratios N_X/N_{auto} and N_Y/N_{auto} for each sample in Extended Data Table 4 as directly informative about sex. However, since the ratio of Autosome:X:Y targets is very different for the 1240k compared to the 390k SNP target set, we need to make an adjustment for the target set. We define "X-rate" as the $N_X/N_{autosomes}$ rate for a sample, divided by the expected value of this quantity based on the number of SNPs in the relevant target set (similarly for the "Y-rate"). For the 1240k capture, these normalizing quantities 0.0432 (X target set) and 0.0284 (Y target set). For the 390k capture set they are 0.0047 (X target set) and 0.0058 (Y target set).

We empirically observe two clusters in the X-rate to Y-rate scatterplot, corresponding to males and females (Figure S4.1). The X-rate histogram does not show a clear separation between males and females but the Y-rate histogram does. We ascribe the greater usefulness of the Y chromosome information to the fact that for true females, the expectation of the number of Y chromosome sequences is extremely low (because females do not have a Y chromosome), and has an extremely low standard deviation, and this makes its distribution simple to distinguish from the non-zero male expectation. In contrast, distinguishing between the expectations for the two sexes for the X chromosome sequences is more difficult because of the high empirical standard deviation for the male and female distributions, which means that the distributions are overlapping even though the means are different by a factor of two. We suspect that the reason why there are large empirical standard deviations when the true number of sequences in the library is not zero (Y chromosomes in females) is bias in the capture experiment in terms of which SNPs were effectively targeted.

Guided by the marginal histogram on the right of Figure S4.1, we determine genetic sex by the rule that Y-rate < 0.05 for a female and Y-rate > 0.2 for a male.

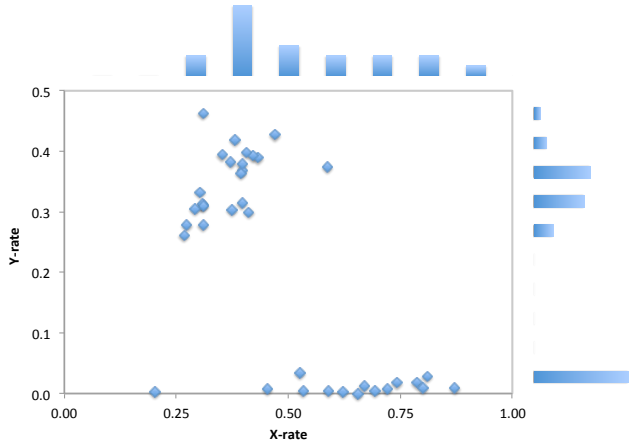


Figure S4.1 X-rate vs. Y-rate plot. There is better separation of the two sexes based on the Y-rate (see the marginal histogram).

Y chromosome haplogroup determination

For the male samples, we determined Y chromosome haplogroups by identifying the most derived allele upstream and the most ancestral allele downstream in the phylogenetic tree in the ISOGG database version 10.01 (http://www.isogg.org/tree). If the most derived Y chromosome SNP upstream was a C→T or G→A substitution (susceptible to ancient DNA damage), we required as least two derived SNPs to assign it to the haplogroup (otherwise, we assigned it to the upstream haplogroup). The results are shown in Table S4.1 and Table S4.2, and are shown in a summary column in Extended Data Table 1:

- We assign *Kostenki14* to haplogroup C1b, as previously described[1] (Table S4.1).
- We assign *GoyetQ116-1* to C1a.
- We assign *Vestonice16* to C1a2. Although our data suggests it carries the derived allele at an A>G SNP that is characteristic of C1b1a1, we find that it carries the ancestral allele at many SNPs that are characteristic of haplogroups upstream of C1b1a1 (i.e. C1b, C1b1, C1b1a) (Table S4.2). Thus, this site may be affected by a sequencing or database error and we ignore the information from it.
- We were surprised to assign *Villabruna* to R1b1 (Table S4.2). When we restrict to damaged sequences, we still assign it to R1b.

Table S4.1. Details of Y haplogroup SNPs for pre-Villabruna Cluster samples.
The most "upstream" ancestral allele observed for a sample in a group of related Y chromosomes is indicated in light red.

Haplogroup	SNP	Ancestral	Derived	Ypos37	Observed	Sequence Depth
Kostenki14	**C1b**					
C	M130	C	T	2734854	T	5
C	V199	C	A	2772928	A	2
C	V232	T	C	7629098	C	23
C	P255	G	A	8685038	A	10
C	V183	G	A	14263271	A	9
C	Page85	G	T	14924643	T	3
C	M216	C	T	15437564	T	3
C	P260	A	C	17286006	C	31
C1	F3393	C	A	23023974	A	2
C1b	F1370	G	C	8643365	C	12
GoyetQ116-1	**C1a**					
C	M130	C	T	2734854	T	1
C	V232	T	C	7629098	C	1
C1	F3393	C	A	23023974	A	2
C1a	CTS11043	G	T	22914979	T	3
Cioclovina1	**CT**					
CT	M5756	T	C	18948988	C	1
Kostenki12	**CT**					
CT	M5609	T	G	7738840	G	1
CT	M5611	C	T	7778691	T	1
CT	M5692	A	C	16325663	C	1
CT	M5712	A	C	17104433	C	1
CT	M5737	C	T	17897543	T	1
CT	CTS9948	G	C	19167672	C	1
Vestonice13	**CT**	**Not IJK**				
CT	CTS109	C	A	2733618	A	1
CT	CTS5318	G	T	16203547	T	1
CT	CTS6327	A	G	16822011	G	1
CT	CTS8243	C	T	17894575	T	1
CT	CTS9556	C	A	18961874	A	1
CT	Z17718	T	C	22263161	C	1
CT	Y1571	G	A	23234852	A	1
CT	M5831	A	C	28685341	C	1
IJ	P126	C	G	21225770	C	1
IJK	L16	G	A	7173143	G	1

Vestonice15	**BT**					
BT	PF1178	G	T	22460746	T	1
Vestonice43	**F**					
F	P145	G	A	8424089	A	1
F	P158	C	T	17493513	T	1
I	PF3641	T	C	7688470	T	1
I	FI4	G	T	8873160	G	1
I	CTS2193	G	T	14214481	G	1
I	CTS4848	C	T	15862842	C	1
I	CTS8963	C	T	18582617	C	1
I	CTS11540	C	T	23156725	C	1
Pavlov1	**I**					
IJK	L16	G	A	7173143	A	1
I	CTS4517	T	G	14986989	G	1
I	FGC2414	C	T	21155653	C	1
Vestonice16	**C1a2**					
C1a	CTS11043	G	T	22914979	T	2
C1a2	V20	G	A	6845955	A	3
C1a2	V86	G	A	6909957	A	1
C1b	F1370	G	C	8643365	G	2
C1b1	M356	C	G	2888203	C	1
C1b1a	K43	G	A	2889366	G	1
C1b1a	F930	C	A	7202706	C	3
C1b1a	K108	C	T	7281157	C	1
C1b1a	Z12437	G	A	7747597	G	4
C1b1a	Z12438	G	T	7821105	G	2
C1b1a	K129	G	A	8292050	G	2
C1b1a	Z12441	A	G	8373844	A	1
C1b1a	Z12442	G	A	8410393	G	1
C1b1a	K141	C	G	8583426	C	2
C1b1a	Z12443	G	A	8635324	G	4
C1b1a	Z12447	G	T	13592515	G	2
C1b1a	K187	G	T	14069571	G	1
C1b1a	Z12450	C	A	14653473	C	2
C1b1a	Z12459	T	C	16549378	T	1
C1b1a	K280	T	C	17237260	T	1
C1b1a	Z12463	C	T	17631240	C	2
C1b1a	Z12464	G	A	18381850	G	2
C1b1a	K319	T	C	18602855	T	2
C1b1a	K396	A	T	22475806	A	1
C1b1a	K414	G	T	23131625	G	3
C1b1a	K415	A	G	23150146	A	2
C1b1a	K417	A	C	23156792	A	1
C1b1a	K426	G	A	23273888	G	2
C1b1a	K435	G	A	23630857	G	3
C1b1a1	K231	A	G	15545270	G	1
Paglicci133	**I**					
I	CTS674	C	T	6943522	T	1
I	CTS9269	C	T	18789763	T	1
I	FGC2416	G	T	7642823	G	1
I	CTS8300	T	A	17924382	T	1
I	PF3815	G	T	21841289	G	2
I1	L80	A	G	14640715	A	1
I1	M253	C	T	15022707	C	1
I1	L81	A	C	22513726	A	1
I1	M307.2	G	A	22750951	G	1
I	FGC2416	G	T	7642823	G	1
I	CTS8300	T	A	17924382	T	1
I	PF3815	G	T	21841289	G	2
HohleFels49	**I**					
I	CTS674	C	T	6943522	T	1
I	CTS9269	C	T	18789763	T	1
GovetQ2	**HIJK**					
HIJK	F929	C	T	7202703	T	1
I	PF3837	G	A	22573702	A	1
Burkhardtshohle	**I**					
I	CTS5650	A	G	16415916	G	1

Table S4.2. Details of Y haplogroup SNPs in Villabruna Cluster samples.

Haplogroup	SNP	Ancestral	Derived	Ypos37	Observed	Read Depth
Villabruna	**R1b1**	**All fragments (subset)**				
R	P224	C	T	17285993	T	7
R	M734	C	T	18066156	T	5
R	P285	C	A	19267344	A	2
R	P227	G	C	21409706	C	7
R1	P294	G	C	7570822	C	2
R1	P242	G	A	7647357	A	2
R1	P238	G	A	7771131	A	6
R1	P245	T	C	8633545	C	2
R1	M173	A	C	15026424	C	1
R1	P286	C	T	17716251	T	6
R1	P236	C	G	17782178	G	10
R1	M306	C	A	22750583	A	5
R1b	M343	C	A	2887824	A	10
R1b1	M415	C	A	9170545	A	2
R1b1	L278	C	T	18914441	T	1
Villabruna	**R1b**	**Damaged fragments**				
R	P224	C	T	17285993	T	4
R	M734	C	T	18066156	T	3
R	P285	C	A	19267344	A	1
R	P280	C	G	21843090	G	1
R1	P294	G	C	7570822	C	2
R1	P238	G	A	7771131	A	2
R1	P286	C	T	17716251	T	1
R1	P236	C	G	17782178	G	1
R1b	M343	C	A	2887824	A	3
Rochedane	**I**					
IJK	L16	G	A	7173143	A	1
I	PF3641	T	C	7688470	C	1
I	CTS4088	T	C	15389836	C	1
I	CTS7593	G	A	17548890	A	1
I	CTS8420	C	A	18018313	A	1
I	CTS10058	A	G	19233673	G	1
I	PF3800	A	G	21402723	G	1
I	PF3803	A	G	21452125	G	1
I2	L68	C	T	18700150	T	1
Falkenstein	**F**					
CF	P143	G	A	14197867	A	1
F	P187	G	T	9108252	T	1
CuiryLesChaudardes1	**I**					
I	PF3640	T	A	7681156	A	1
I	CTS5764	A	G	16471254	G	1
I	Z16987	A	G	22243817	G	1
BerryAuBac	**I**					
HIJK	F929	C	T	7202703	T	1
I	CTS2514	T	C	14337364	C	1
I	PF3836	T	G	22525421	G	1

References

1 Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, doi:10.1126/science.aaa0114 (2014).

Section 5
Genetic clustering of ancient samples

Overview
We assembled ancient DNA data on a total of 51 ancient samples (Extended Data Table 1). To prepare these data, we combined previously published data from 14 samples (*Malta1*[1], *AfontovaGora2*[1], *UstIshim*[2], *Oase1*[3], *Loschbour*[4], *LaBrana1*[5], *Hungarian.KO1*[6], *Motala12*[4], *Karelia*[7], *Stuttgart*[4], *Satsurblia*[8], *Kotias*[8], *Bichon*[8], and *Kostenki14*[9]), with our new capture data for 38 samples. One sample overlapped between the two collections, *Kostenki14*. For most of our analyses, we use the 16.1-fold coverage data that we newly report in this study, rather than the 2.8-fold coverage published data[9] (we confirmed that the key scientific results were consistent between the two independently collected *Kostenki14* datasets).

Figure 1 shows the geographical distribution of the samples. Below we list the 51 samples by time period, with samples having more than >0.1x coverage highlighted in bold, and samples with fewer than 10,000 SNPs covered underlined.

Early Upper Palaeolithic (>33,000 BP): ***UstIshim***, ***Oase1***, ***Kostenki14***, ***GoyetQ116-1***, *Cioclovina1, Kostenki12*, *Muierii2*

Middle Upper Palaeolithic (33,000-24,000 BP): *Vestonice13, Vestonice15, Vestonice14, Vestonice43, Pavlov1,* ***Vestonice16****, Palgicci133,* ***KremsWA3****, Ostuni2,* ***Ostuni1****, Paglicci108, GoyetQ53-1, GoyetQ376-19, GoyetQ56-16, Malta1*

Late Upper Palaeolithic and Mesolithic (19,000-6,000 BP): ***ElMiron****, AfontovaGora2, AfontovaGora3, HohleFels79, HohleFels49, Rigney1, GoyetQ-2, Brillenhohle, Burkhardtshohle,* ***Villabruna****,* ***Bichon****,* ***Satsurblia****,* ***Rochedane****, Continenza, Iboussieres39,* ***Ranchot88****, LesCloseaux13,* ***Kotias****, Falkenstein, Bockstein, Ofnet, Chaudardes1,* ***Loschbour****,* ***LaBrana1****,* ***Hungarian.KO1****, BerryAuBac,* ***Motala12****,* ***Karelia***

Early Neolithic (7,000 BP): ***Stuttgart***

Outgroup f_3-statistics
We computed statistics of the form $f_3(X, Y; Mbuti)$, which measure the shared genetic drift between populations X and Y since their separation from an outgroup (*Mbuti*)[1]. For the version of this analysis shown in Figure 3A, we did not restrict to damaged sequences for *AfontovaGora3, ElMiron, Falkenstein, GoyetQ-2, GoyetQ53-1, HohleFels79, HohleFels49, Les Closeaux13, Ofnet, Ranchot88* and *Rigney1*. While this allows some present-day human contamination into the data, the increased size of the dataset produces a clearer view of the shared ancestry. A version of this analysis that restricts to damaged sequences only, and to samples with at least 30,000 SNPs, is shown in Extended Data Figure 2.

The shared genetic drift analysis based on these f_3-statistics reveals several apparent clusters (note that this is not the full list of samples we assign to these clusters in Extended Data Table 1, as in the analyses that follow we are also able to assign additional individuals):

- "Věstonice Cluster": *Vestonice13, Vestonice15, Vestonice14, Vestonice43, Vestonice16, KremsWA3, Ostuni1, Pavlov*
- "El Mirón Cluster": *Burkhardtshohle, ElMiron, GoyetQ-2, HohleFels79, HohleFels49* and *Rigney1*

- "Villabruna Cluster": *BerryAuBac, Bichon, Bockstein, Chaudardes1, Falkenstein, Hungarian.KO1 LaBrana1, LesCloseaux13, Loschbour, Ofnet, Ranchot88, Rochedane* and *Villabruna*.
- "Mal'ta Cluster": *AfontovaGora3* and *Malta1*.
- "Satsurblia Cluster": *Satsurblia, Kotias*

"Věstonice Cluster" (these individuals all lived 34,000-26,000 BP)

We used *D*-statistics to test formally if various pairs of pre-Neolithic samples are consistent with being clades with respect to the other samples and outgroups. We first identified a large group of samples that were consistent with being a clade with samples from the site of Dolní Věstonice. Specifically, *Vestonice13, Vestonice15, Vestonice43, KremsWA3, Ostuni1*, and *Pavlov1* share more alleles with *Vestonice16* (our highest coverage sample that we use to represent it for many analyses) than with other pre-Neolithic samples as revealed by statistics of the form *D(X, Y; Vestonice_Cluster, Mbuti)*. This is consistent with the patterns in Figure 3A and Extended Data Fig. 2 (Table S5.1). We also identified three subgroups within the "Věstonice Cluster":

(a) "Věstonice Central European Cluster" (Vestonice_CE_C): Vestonice, KremsWA3, Pavlov

Statistics of the form $D(Vestonice\_Cluster_1, Vestonice\_Cluster_2; Vestonice\_Cluster_3, Mbuti)$, where $Vestonice\_Cluster_3$ is *Vestonice13, Vestonice15, Vestonice16, Vestonice43, KremsWA3, Ostuni1* or *Pavlov1* in turn, suggest that (*Vestonice16, Vestonice13, Vestonice15, KremsWA3*) form a subgroup (Table S5.2). These statistics also suggest that *Vestonice43* and *Pavlov1* are more distantly related, although one statistic, *D(Vestonice16, KremsWA3; Vestonice43, Mbuti),* is significantly less than 0, providing some evidence against this simple model (Z=-3.6).

(b) "Věstonice Italian Cluster" (Vestonice_I_C) – Ostuni1, Ostuni2 and Palgicci133

The Italian high coverage sample *Ostuni1*, while in the "Věstonice Cluster", is a definite outgroup compared to all samples in the "Věstonice Central European Cluster" (Table S5.2). There is also *D*-statistic evidence that the three Italian samples sub-cluster (Table S5.3).

(c) "Vestonice Goyet Cluster"(Vestonice_Goyet_C) – Multiple samples from Goyet cave

Samples from Goyet cave in Belgium dating from 28,000-26,000 BP are part of the broader Věstonice Cluster*: GoyetQ376-19, Goyet53-1* and *Goyet56-16*. Notably, not all Goyet cave samples are from this cluster: the earlier ~35,000 BP *GoyetQ116-1*, and the later ~15,000 BP *GoyetQ-2*, have very different genetic affinities.

Table S5.1 Z-score of *D(X, Y; Věstonice Cluster, Mbuti)*

D(X, Y; Vestonice13, Mbuti) Vestonice13: 139,568 SNPs											
X/Y	Han	Ust	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	0.6	3.3	-12.9	-11.4	-25.5	-17.1	-12.4	-14.4	-13.1	-12.6
UstIshim	-0.6	NA	1.9	-11.5	-10.3	-23.9	-15.6	-11.4	-12.7	-11.4	-11.4
Oase1	-3.3	-1.9	NA	-8.9	-7.8	-18.8	-9.8	-8.7	-8.3	-8.5	-9.1
Kostenki14	12.9	11.5	8.9	NA	0.6	-13.7	-5.7	-0.1	-0.3	0.6	0.5
GoyetQ116-1	11.4	10.3	7.8	-0.6	NA	-13.5	-6.2	-0.8	-0.9	-0.4	-0.2
Vestonice16	25.5	23.9	18.8	13.7	13.5	NA	5.9	13.2	13.8	14.3	13.7
Ostuni1	17.1	15.6	9.8	5.7	6.2	-5.9	NA	4.6	5.5	6.7	6.2
ElMiron	12.4	11.4	8.7	0.1	0.8	-13.2	-4.6	NA	-0.3	0.8	0.8
Villabruna	14.4	12.7	8.3	0.3	0.9	-13.8	-5.5	0.3	NA	1	0.9
Loschbour	13.1	11.4	8.5	-0.6	0.4	-14.3	-6.7	-0.8	-1	NA	-0.2
LaBrana1	12.6	11.4	9.1	-0.5	0.2	-13.7	-6.2	-0.8	-0.9	0.2	NA
Malta1	7.9	7.2	4.9	-4	-3.4	-15.6	-7.9	-3.9	-4.4	-3.1	-3.6
D(X, Y; Vestonice15, Mbuti) Vestonice15: 30,900 SNPs											
X/Y	Han	Ust	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	0.5	0.8	-7.5	-7.2	-19.5	-10.5	-7.4	-10.1	-11.5	-7.2
UstIshim	-0.5	NA	0.4	-7.3	-7	-19	-10.2	-7.5	-9	-9.8	-6.6
Oase1	-0.8	-0.4	NA	-4.6	-3.9	-10.7	-6.3	-4.3	-5.9	-6.6	-4.2

Kostenki14	7.5	7.3	4.6	NA	0.1	-11.1	-3.2	0	-1.6	-1.8	0.7
GoyetQ116-1	7.2	7	3.9	-0.1	NA	-11.4	-5	-0.2	-1.9	-2.1	1
Vestonice16	19.5	19	10.7	11.1	11.4	NA	5.4	11	10.2	10.7	12.2
Ostuni1	10.5	10.2	6.3	3.2	5	-5.4	NA	3.9	3	2.6	4.6
ElMiron	7.4	7.5	4.3	0	0.2	-11	-3.9	NA	-1.4	-1.9	1.7
Villabruna	10.1	9	5.9	1.6	1.9	-10.2	-3	1.4	NA	0.2	2.9
Loschbour	11.5	9.8	6.6	1.8	2.1	-10.7	-2.6	1.9	-0.2	NA	3.2
LaBrana1	7.2	6.6	4.2	-0.7	-1	-12.2	-4.6	-1.7	-2.9	-3.2	NA
Malta1	5.9	5.1	3.1	-0.7	-0.7	-10.7	-3.5	-0.8	-2.8	-2.7	-0.2

D(X, Y; Vestonice43, Mbuti) Vestonice43: 163,946 SNPs

X/Y	Han	Ust	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	0.3	2.2	-14.2	-12.1	-19.6	-19.4	-15.6	-17.7	-18.3	-14.7
UstIshim	-0.3	NA	2.1	-13.3	-10.9	-17.4	-16.7	-14.1	-15.6	-16.1	-12.7
Oase1	-2.2	-2.1	NA	-9.8	-9.1	-13.3	-12.4	-10	-12.6	-12	-8.8
Kostenki14	14.2	13.3	9.8	NA	1.7	-5.1	-5.7	-2.5	-2.8	-2.8	0.7
GoyetQ116-1	12.1	10.9	9.1	-1.7	NA	-6.1	-7.4	-3	-3.5	-3.9	-0.6
Vestonice16	19.6	17.4	13.3	5.1	6.1	NA	-0.9	3.9	3.2	3.3	6.2
Ostuni1	19.4	16.7	12.4	5.7	7.4	0.9	NA	4.6	3.8	4.3	6.7
ElMiron	15.6	14.1	10	2.5	3	-3.9	-4.6	NA	-0.9	-0.3	2.8
Villabruna	17.7	15.6	12.6	2.8	3.5	-3.2	-3.8	0.9	NA	0	3.7
Loschbour	18.3	16.1	12	2.8	3.9	-3.3	-4.3	0.3	0	NA	4.2
LaBrana1	14.7	12.7	8.8	-0.7	0.6	-6.2	-6.7	-2.8	-3.7	-4.2	NA
Malta1	8	7.2	6.7	-5.2	-3.8	-9.9	-9.3	-6	-7	-7.7	-4.3

D(X, Y; Vestonice16, Mbuti) Vestonice16: 945,292 SNPs

X/Y	Han	Ust	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	-1.2	3.9	-16.3	-14.4	NA	-24.6	-18.6	-19	-18.5	-16
UstIshim	1.2	NA	4.4	-12.5	-11	NA	-20.2	-14.8	-14.7	-14.5	-12.1
Oase1	-3.9	-4.4	NA	-15	-13.3	NA	-18.5	-15.9	-14.7	-15.9	-13.5
Kostenki14	16.3	12.5	15	NA	0.9	NA	-8.7	-2	-1.5	-1.7	0.6
GoyetQ116-1	14.4	11	13.3	-0.9	NA	NA	-9.5	-2.8	-2.5	-3.2	-0.5
Ostuni1	24.6	20.2	18.5	8.7	9.5	NA	NA	7.4	8.1	7.8	9.7
ElMiron	18.6	14.8	15.9	2	2.8	NA	-7.4	NA	0.9	0.3	3.4
Villabruna	19	14.7	14.7	1.5	2.5	NA	-8.1	-0.9	NA	-0.6	2.9
Loschbour	18.5	14.5	15.9	1.7	3.2	NA	-7.8	-0.3	0.6	NA	3.2
LaBrana1	16	12.1	13.5	-0.6	0.5	NA	-9.7	-3.4	-2.9	-3.2	NA
Malta1	11.8	8.4	10.4	-4.6	-2.6	NA	-11.9	-6.3	-5.6	-6.1	-3.8

D(X, Y; Pavlov1, Mbuti) Pavlov1: 57,005 SNPs

X/Y	Han	Ust	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	0.3	0.5	-8.6	-9.5	-16.7	-13.2	-10.7	-12.2	-14	-11.3
UstIshim	-0.3	NA	0.2	-8.1	-9.2	-16.7	-12.6	-10.4	-11.6	-12.3	-10.4
Oase1	-0.5	-0.2	NA	-3.1	-3.4	-8.6	-7	-4.2	-6.1	-6.3	-5.6
Kostenki14	8.6	8.1	3.1	NA	-0.7	-7.6	-5.7	-1.9	-2.6	-2.6	-1.7
GoyetQ116-1	9.5	9.2	3.4	0.7	NA	-7.3	-4.7	-0.8	-2	-2.3	-0.9
Vestonice16	16.7	16.7	8.6	7.6	7.3	NA	0.7	5.2	4.9	5.7	6.2
Ostuni1	13.2	12.6	7	5.7	4.7	-0.7	NA	4.5	3.7	4.2	5
ElMiron	10.7	10.4	4.2	1.9	0.8	-5.2	-4.5	NA	-0.8	-0.8	0.3
Villabruna	12.2	11.6	6.1	2.6	2	-4.9	-3.7	0.8	NA	0.2	1.4
Loschbour	14	12.3	6.3	2.6	2.3	-5.7	-4.2	0.8	-0.2	NA	1.1
LaBrana1	11.3	10.4	5.6	1.7	0.9	-6.2	-5	-0.3	-1.4	-1.1	NA
Malta1	7.3	6.8	2.7	-0.3	-1.7	-8.1	-6.3	-2.7	-4.5	-4	-2.7

D(X, Y; KremsWA3, Mbuti) KremsWA3: 203,986 SNPs

X/Y	Han	Ust	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	-0.3	2.4	-14.3	-13.5	-26.5	-20.1	-16.2	-18.4	-19.6	-15.9
UstIshim	0.3	NA	3	-11.9	-11.1	-24.8	-16.5	-14.4	-16.3	-15.9	-13.8
Oase1	-2.4	-3	NA	-8.1	-9.6	-18	-10.4	-11.2	-10.7	-11.6	-9.9
Kostenki14	14.3	11.9	8.1	NA	-0.5	-13.9	-7	-2.7	-3.6	-3.2	-1.6
GoyetQ116-1	13.5	11.1	9.6	0.5	NA	-13.1	-6.7	-2.3	-2.8	-2.8	-1
Vestonice16	26.5	24.8	18	13.9	13.1	NA	6.3	11.6	12.1	12.7	13.3
Ostuni1	20.1	16.5	10.4	7	6.7	-6.3	NA	4.1	3.5	5.1	5.9
ElMiron	16.2	14.4	11.2	2.7	2.3	-11.6	-4.1	NA	-0.7	-0.2	1.2
Villabruna	18.4	16.3	10.7	3.6	2.8	-12.1	-3.5	0.7	NA	0.5	2.4
Loschbour	19.6	15.9	11.6	3.2	2.8	-12.7	-5.1	0.2	-0.5	NA	2.1
LaBrana1	15.9	13.8	9.9	1.6	1	-13.3	-5.9	-1.2	-2.4	-2.1	NA
Malta1	9.5	7.5	5.6	-4.5	-3.1	-17.9	-7.8	-5.4	-7.3	-7.4	-5.7

D(X, Y; Ostuni1, Mbuti) Ostuni1: 369,313 SNPs

X/Y	Han	Ust	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	0.9	3.4	-13.4	-13	-23.3	NA	-17.5	-18.5	-17.3	-16.1
UstIshim	-0.9	NA	1.8	-12.2	-12.3	-21.9	NA	-15.4	-16.4	-14.9	-14.2
Oase1	-3.4	-1.8	NA	-9.8	-10.5	-17.4	NA	-13	-13.3	-13	-12
Kostenki14	13.4	12.2	9.8	NA	-0.4	-10	NA	-3.5	-3.9	-2.4	-1.9
GoyetQ116-1	13	12.3	10.5	0.4	NA	-8.8	NA	-2.7	-3	-2.3	-1.2
Vestonice16	23.3	21.9	17.4	10	8.8	NA	NA	6.7	6.1	7.9	9
Ostuni1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ElMiron	17.5	15.4	13	3.5	2.7	-6.7	NA	NA	-0.5	0.7	1.7
Villabruna	18.5	16.4	13.3	3.9	3	-6.1	NA	0.5	NA	1.7	2.6
Loschbour	17.3	14.9	13	2.4	2.3	-7.9	NA	-0.7	-1.7	NA	1
LaBrana1	16.1	14.2	12	1.9	1.2	-9	NA	-1.7	-2.6	-1	NA
Malta1	9	8	6.4	-3.9	-3.3	-13.8	NA	-6.6	-7.3	-6.7	-6.3

Table S5.2 Z-score *D(Vestonice_Cluster$_1$, Vestonice_Cluster$_2$; Vestonice_Cluster$_3$, Mbuti)*

D(X, Y; Vestonice13, Mbuti) Vestonice13: 139,568 SNPs						
X/Y	Vestonice15	Vestonice43	Pavlov1	Vestonice16	Ostuni1	KremsWA3
Vestonice15	NA	1.8	-0.4	0.7	0.4	0.4
Vestonice43	-1.8	NA	-1.3	-6.8	0.3	-2.8
Pavlov1	0.4	1.3	NA	-2.5	0.8	-2.8
Vestonice16	-0.7	6.8	2.5	NA	5.9	-1.3
Ostuni1	-0.4	-0.3	-0.8	-5.9	NA	-5.3
KremsWA3	-0.4	2.8	2.8	1.3	5.3	NA
D(X, Y; Vestonice15, Mbuti) Vestonice15: 30,900 SNPs						
X/Y	Vestonice13	Vestonice43	Pavlov1	Vestonice16	Ostuni1	KremsWA3
Vestonice13	NA	2.1	1.4	-1	1.9	0.1
Vestonice43	-2.1	NA	-1	-5.4	-1.6	-2.5
Pavlov1	-1.4	1	NA	-2.7	-1.9	-0.1
Vestonice16	1	5.4	2.7	NA	5.4	1.3
Ostuni1	-1.9	1.6	1.9	-5.4	NA	-1
KremsWA3	-0.1	2.5	0.1	-1.3	1	NA
D(X, Y; Vestonice43, Mbuti) Vestonice43: 163,946 SNPs						
X/Y	Vestonice13	Vestonice15	Pavlov1	Vestonice16	Ostuni1	KremsWA3
Vestonice13	NA	0.3	-1	0.6	-0.3	0.6
Vestonice15	-0.3	NA	1	-0.4	-0.2	0.4
Pavlov1	1	-1	NA	3.2	1.6	-0.7
Vestonice16	-0.6	0.4	-3.2	NA	-0.9	-3.6
Ostuni1	0.3	0.2	-1.6	0.9	NA	-2.2
KremsWA3	-0.6	-0.4	0.7	3.6	2.2	NA
D(X, Y; Pavlov1, Mbuti) Pavlov1: 57,005 SNPs						
X/Y	Vestonice13	Vestonice15	Vestonice43	Vestonice16	Ostuni1	KremsWA3
Vestonice13	NA	1.8	0.2	1.2	2.2	-0.2
Vestonice15	-1.8	NA	2	-0.6	-0.7	-0.9
Vestonice43	-0.2	-2	NA	1.8	1.2	-0.7
Vestonice16	-1.2	0.6	-1.8	NA	0.7	-1.5
Ostuni1	-2.2	0.7	-1.2	-0.7	NA	-2
KremsWA3	0.2	0.9	0.7	1.5	2	NA
D(X, Y; Vestonice16, Mbuti) Vestonice16: 945,292 SNPs						
X/Y	Vestonice13	Vestonice15	Vestonice43	Pavlov1	Ostuni1	KremsWA3
Vestonice13	NA	-1.8	7.4	3.8	6.7	2.2
Vestonice15	1.8	NA	5.1	2	7.9	3.4
Vestonice43	-7.4	-5.1	NA	-1.4	-0.5	-6.9
Pavlov1	-3.8	-2	1.4	NA	1.7	-3.7
Ostuni1	-6.7	-7.9	0.5	-1.7	NA	-5.5
KremsWA3	-2.2	-3.4	6.9	3.7	5.5	NA
D(X, Y; Ostuni1, Mbuti) Ostuni1: 369,313 SNPs						
X/Y	Vestonice13	Vestonice15	Vestonice43	Pavlov1	Vestonice16	KremsWA3
Vestonice13	NA	1.7	-0.5	1.4	0.7	1.5
Vestonice15	-1.7	NA	1.4	1.2	2.6	1.3
Vestonice43	0.5	-1.4	NA	-0.4	0.4	-0.3
Pavlov1	-1.4	-1.2	0.4	NA	1	-0.7
Vestonice16	-0.7	-2.6	-0.4	-1	NA	0.8
KremsWA3	-1.5	-1.3	0.3	0.7	-0.8	NA
D(X, Y; KremsWA3, Mbuti) KremsWA3: 203,986 SNPs						
X/Y	Vestonice13	Vestonice15	Vestonice43	Pavlov1	Vestonice16	Ostuni1
Vestonice13	NA	-0.3	3.3	2.6	3.5	6.7
Vestonice15	0.3	NA	2.9	-0.7	2.1	2.3
Vestonice43	-3.3	-2.9	NA	0	-3	1.9
Pavlov1	-2.6	0.7	0	NA	-2	1.4
Vestonice16	-3.5	-2.1	3	2	NA	6.3
Ostuni1	-6.7	-2.3	-1.9	-1.4	-6.3	NA

"El Mirón Cluster" (ElMiron_C) (these individuals all lived 19,000-14,000 BP)

Brillenhohle, *Burkhardtshohle*, *GoyetQ-2*, *HohleFels79*, *HohleFels49* and *Rigney1* share more alleles with *ElMiron* than with other pre-Neolithic Europeans as revealed by statistics like *D(X, Y; El_Miron_Cluster, Mbuti)* (Table S5.4), consistent with outgroup f_3-statistics (Figure 3A and Extended Data Figure 2). The *D*-statistics also suggest that *ElMiron*—the earliest and most southern of the samples—is an outgroup to the later set of samples, as revealed by statistics like *D(El_Miron_Cluster$_1$, El_Miron_Cluster$_2$; El_Miron_Cluster$_3$, Mbuti)* (Table S5.5). We designate the entire group as the "El Mirón Cluster" and designate the subgroup excluding *ElMiron* as the "El Mirón Non-Iberian Cluster" (ElMiron_NI_C).

Table S5.3 Z-score of *D(X, Y; Ostuni2/Paglicci133, Mbuti)*

D(X, Y; Ostuni2, Mbuti) Ostuni2: 17,017 SNPs

X/Y	Han	Ust	Oase1	Kost14	Q116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	0.1	0.4	-6.2	-6.2	-8.5	-11.9	-7.4	-5.6	-6.5	-5.2
UstIshim	-0.1	NA	0.5	-6	-5.8	-8.3	-10.2	-6.5	-5.3	-5.7	-4.8
Oase1	-0.4	-0.5	NA	-3.6	-2.8	-4	-4.2	-3.9	-3.2	-3.4	-3.8
Kostenki14	6.2	6	3.6	NA	-0.1	-2.3	-5.4	-0.9	-0.2	1.1	1.4
GoyetQ116-1	6.2	5.8	2.8	0.1	NA	-2.3	-4.4	-0.4	0.8	1.4	1.4
Vestonice16	8.5	8.3	4	2.3	2.3	NA	-2.9	1.2	2.6	3.6	3.6
Ostuni1	11.9	10.2	4.2	5.4	4.4	2.9	NA	4.4	5.8	6.5	4.9
ElMiron	7.4	6.5	3.9	0.9	0.4	-1.2	-4.4	NA	1.8	2	2.4
Villabruna	5.6	5.3	3.2	0.2	-0.8	-2.6	-5.8	-1.8	NA	0.5	0.8
Loschbour	6.5	5.7	3.4	-1.1	-1.4	-3.6	-6.5	-2	-0.5	NA	0.3
LaBrana1	5.2	4.8	3.8	-1.4	-1.4	-3.6	-4.9	-2.4	-0.8	-0.3	NA
Malta1	3.6	3.3	2.2	-2.3	-2.1	-3.9	-4.9	-2.5	-1.4	-1.1	-1

D(X, Y; Paglicci133, Mbuti) Paglicci133: 82,330 SNPs

X/Y	Han	Ust	Oase1	Kost14	Q116-1	Vestonice16	Ostuni1	ElMiron	Villabruna	Loschbour	LaBrana1
Han	NA	1.5	1.8	-9.2	-9.5	-13.3	-12.7	-9.4	-11.4	-11.9	-10.5
UstIshim	-1.5	NA	1.1	-9.2	-9.5	-13.3	-11.7	-8.9	-11.5	-11.8	-9.9
Oase1	-1.8	-1.1	NA	-5.1	-6.5	-7	-7.2	-4.3	-7.5	-6.9	-5.2
Kostenki14	9.2	9.2	5.1	NA	-0.6	-4.6	-5.6	-0.7	-1.8	-1.4	-0.7
GoyetQ116-1	9.5	9.5	6.5	0.6	NA	-3.3	-4.9	0.9	-0.7	-0.5	0.7
Vestonice16	13.3	13.3	7	4.6	3.3	NA	-1.8	3.5	2.7	3	4.2
Ostuni1	12.7	11.7	7.2	5.6	4.9	1.8	NA	5	3.5	4.5	4.6
ElMiron	9.4	8.9	4.3	0.7	-0.9	-3.5	-5	NA	-1.7	-1.6	-0.4
Villabruna	11.4	11.5	7.5	1.8	0.7	-2.7	-3.5	1.7	NA	0.7	1
Loschbour	11.9	11.8	6.9	1.4	0.5	-3	-4.5	1.6	-0.7	NA	0.2
LaBrana1	10.5	9.9	5.2	0.7	-0.7	-4.2	-4.6	0.4	-1	-0.2	NA
Malta1	6	6.1	3.5	-2.9	-2.6	-6.4	-7.4	-3.4	-4.7	-5	-4.6

Table S5.4 Z-score of *D(X, Y; El_Miron_Cluster, Mbuti)*

D(X, Y; ElMiron, Mbuti) HohleFels79: 797,714 SNPs

X/Y	Han	Ust	Oase1	Kost14	GoyetQ116-	Vestonice1	Ostuni	ElMiro	Villabruna	Loschbour	LaBrana
Han	NA	0.8	3.8	-12.2	-22.5	-17	-17.4	-21.9	-25.9	-30.5	-33.5
UstIshim	-0.8	NA	3.2	-11.2	-19.7	-15.7	-15.3	-20.4	-23	-26.3	-27.8
Oase1	-3.8	-3.2	NA	-10.6	-17.9	-16.2	-13.8	-13.2	-20.5	-24.6	-23.6
Kostenki14	12.2	11.2	10.6	NA	-11	-5.2	-5	-13.4	-13.5	-16.4	-18.6
GoyetQ116-1	22.5	19.7	17.9	11	NA	6.1	4.5	-8.4	-1.8	-5.2	-5.3
Vestonice16	17	15.7	16.2	5.2	-6.1	NA	-0.4	-12.3	-8.3	-12.1	-12.5
Ostuni1	17.4	15.3	13.8	5	-4.5	0.4	NA	-9.3	-8.1	-11.3	-11.3
Villabruna	25.9	23	20.5	13.5	1.8	8.3	8.1	-7.1	NA	-4.1	-4.1
Loschbour	30.5	26.3	24.6	16.4	5.2	12.1	11.3	-5.5	4.1	NA	-0.3
LaBrana1	33.5	27.8	23.6	18.6	5.3	12.5	11.3	-6.8	4.1	0.3	NA
Malta1	9.1	7.8	8.9	-2.1	-11.9	-7.5	-7	-12.5	-15	-18.5	-19.9

D(X, Y; HohleFels79, Mbuti) HohleFels79: 11,211 SNPs

X/Y	Han	Ust	Oase1	Kost14	GoyetQ116-	Vestonice1	Ostuni	ElMiro	Villabruna	Loschbour	LaBrana
Han	NA	0.1	0.5	-3.3	-8.7	-3.7	-2.5	-12.2	-4.4	-8.7	-7.7
UstIshim	-0.1	NA	0.8	-3.4	-7.9	-3.7	-2.3	-11	-3.9	-7.9	-6.8
Oase1	-0.5	-0.8	NA	-2.5	-3.3	-2	-1.2	-5	-1.2	-3.9	-3.7
Kostenki14	3.3	3.4	2.5	NA	-3.6	-0.6	0	-6.9	-0.7	-3.9	-3
GoyetQ116-1	8.7	7.9	3.3	3.6	NA	4.4	2.5	-3.1	3.9	1.4	0.9
Vestonice16	3.7	3.7	2	0.6	-4.4	NA	0.7	-7	-0.3	-3.6	-2.9
Ostuni1	2.5	2.3	1.2	0	-2.5	-0.7	NA	-5.4	-1	-3.4	-2.5
ElMiron	12.2	11	5	6.9	3.1	7	5.4	NA	6.6	4.5	3.8
Villabruna	4.4	3.9	1.2	0.7	-3.9	0.3	1	-6.6	NA	-3.5	-2.8
Loschbour	8.7	7.9	3.9	3.9	-1.4	3.6	3.4	-4.5	3.5	NA	-0.1
LaBrana1	7.7	6.8	3.7	3	-0.9	2.9	2.5	-3.8	2.8	0.1	NA
Malta1	3.2	2.4	2.4	-0.1	-3.2	0.3	-0.6	-5.6	-0.2	-3	-2.5

D(X, Y; HohleFels49, Mbuti) HohleFels49: 63,151 SNPs

X/Y	Han	Ust	Oase1	Kost14	GoyetQ116-	Vestonice1	Ostuni	ElMiro	Villabruna	Loschbour	LaBrana
Han	NA	0.2	1.6	-6.2	-16	-8	-7.1	-23.7	-11.1	-17	-15.5
UstIshim	-0.2	NA	1.9	-5.7	-14.1	-7.4	-6.7	-21.7	-9.9	-14.6	-13.4
Oase1	-1.6	-1.9	NA	-4.2	-9.9	-5.7	-4	-12.3	-6.3	-9.2	-8.6
Kostenki14	6.2	5.7	4.2	NA	-8.7	-0.8	-2.8	-14.4	-4	-7.3	-7
GoyetQ116-1	16	14.1	9.9	8.7	NA	6.6	5.2	-6.9	4.8	1.8	2.1
Vestonice16	8	7.4	5.7	0.8	-6.6	NA	-0.2	-14.4	-2.6	-6	-5
Ostuni1	7.1	6.7	4	2.8	-5.2	0.2	NA	-10	-0.5	-3.9	-3.2
ElMiron	23.7	21.7	12.3	14.4	6.9	14.4	10	NA	10.9	9.1	8.9
Villabruna	11.1	9.9	6.3	4	-4.8	2.6	0.5	-10.9	NA	-3.8	-3
Loschbour	17	14.6	9.2	7.3	-1.8	6	3.9	-9.1	3.8	NA	-0.2
LaBrana1	15.5	13.4	8.6	7	-2.1	5	3.2	-8.9	3	0.2	NA
Malta1	4.4	4.1	2.7	-1.3	-8.1	-1.9	-2.4	-14	-3.6	-7.4	-6.9

D(X, Y; Rigney1, Mbuti) Rigney1: 35,600 SNPs

X/Y	Han	Ust	Oase1	Kost14	GoyetQ116-	Vestonice1	Ostuni	ElMiro	Villabruna	Loschbour	LaBrana
Han	NA	-0.5	2.2	-5.1	-11.1	-6.5	-6.2	-18.7	-10.9	-14.9	-14.7
UstIshim	0.5	NA	2.2	-4.3	-10.1	-5.9	-5.3	-17	-9.9	-12.8	-12.8
Oase1	-2.2	-2.2	NA	-3.6	-7.3	-4.1	-3.6	-9	-6.8	-7.6	-7.8
Kostenki14	5.1	4.3	3.6	NA	-5.2	-1.9	-1.9	-11.5	-4.8	-7.3	-7.7
GoyetQ116-1	11.1	10.1	7.3	5.2	NA	4.3	2.1	-6.3	0.5	-1.2	-1.5
Vestonice16	6.5	5.9	4.1	1.9	-4.3	NA	-1.6	-10.8	-3.3	-5.1	-5.3
Ostuni1	6.2	5.3	3.6	1.9	-2.1	1.6	NA	-7.7	-1.7	-2.9	-3.3

ElMiron	18.7	17	9	11.5	6.3	10.8	7.7	NA	7	6.4	5.5
Villabruna	10.9	9.9	6.8	4.8	-0.5	3.3	1.7	-7	NA	-2.5	-2.9
Loschbour	14.9	12.8	7.6	7.3	1.2	5.1	2.9	-6.4	2.5	NA	-1
LaBrana1	14.7	12.8	7.8	7.7	1.5	5.3	3.3	-5.5	2.9	1	NA
Malta1	4	3.2	2.6	-0.4	-6.7	-2.5	-2.2	-9.9	-5.4	-7.8	-7.4
D(X, Y; GoyetQ-2, Mbuti) GoyetQ-2: 72,263											
X/Y	Han	Ust	Oase1	Kost14	GoyetQ116-	Vestonice1	Ostuni	ElMiro	Villabruna	Loschbour	LaBrana
Han	NA	0.8	0.4	-6.3	-18.7	-9	-6.6	-25.4	-12.7	-19.2	-19.2
UstIshim	-0.8	NA	-0.2	-6.1	-16.9	-8.9	-6.8	-23.2	-12	-16.4	-16.9
Oase1	-0.4	0.2	NA	-3.5	-8.1	-5.3	-5.1	-12.9	-7.2	-9	-8.5
Kostenki14	6.3	6.1	3.5	NA	-10.9	-1.8	-2.5	-15.2	-4.9	-8.5	-10.2
GoyetQ116-1	18.7	16.9	8.1	10.9	NA	7.3	4.9	-6.8	5.2	3.8	1.5
Vestonice16	9	8.9	5.3	1.8	-7.3	NA	-0.8	-13.7	-3.1	-6.1	-6.9
Ostuni1	6.6	6.8	5.1	2.5	-4.9	0.8	NA	-9.9	-1.6	-3.3	-5.7
ElMiron	25.4	23.2	12.9	15.2	6.8	13.7	9.9	NA	11.7	10.8	7.6
Villabruna	12.7	12	7.2	4.9	-5.2	3.1	1.6	-11.7	NA	-3.1	-4.9
Loschbour	19.2	16.4	9	8.5	-3.8	6.1	3.3	-10.8	3.1	NA	-2.7
LaBrana1	19.2	16.9	8.5	10.2	-1.5	6.9	5.7	-7.6	4.9	2.7	NA
Malta1	7.3	6.6	3.7	0.9	-8	-1.1	-1.1	-12.8	-3.7	-6.6	-7.9
D(X, Y; Burkhardtshohle, Mbuti) Burkhardtshohle: 38,376 SNPs											
X/Y	Han	Ust	Oase1	Kost14	GoyetQ116-	Vestonice1	Ostuni	ElMiro	Villabruna	Loschbour	LaBrana
Han	NA	-0.2	2.1	-4.9	-15.2	-8	-5.9	-21.2	-9.4	-14.5	-14.3
UstIshim	0.2	NA	2.5	-3.9	-13.5	-7	-5.5	-18.9	-8.4	-12.2	-12.3
Oase1	-2.1	-2.5	NA	-3.8	-8.5	-4.6	-3.8	-10.4	-6	-8.1	-7.6
Kostenki14	4.9	3.9	3.8	NA	-8.5	-3	-1.2	-13.2	-3.3	-6.6	-7.4
GoyetQ116-1	15.2	13.5	8.5	8.5	NA	5.6	3.9	-5.6	5.2	2.7	1.3
Vestonice16	8	7	4.6	3	-5.6	NA	0.6	-11.4	-0.9	-4.3	-4.6
Ostuni1	5.9	5.5	3.8	1.2	-3.9	-0.6	NA	-8.8	-0.6	-2.5	-3.1
ElMiron	21.2	18.9	10.4	13.2	5.6	11.4	8.8	NA	11.1	9.6	7.4
Villabruna	9.4	8.4	6	3.3	-5.2	0.9	0.6	-11.1	NA	-3.5	-4.2
Loschbour	14.5	12.2	8.1	6.6	-2.7	4.3	2.5	-9.6	3.5	NA	-1.1
LaBrana1	14.3	12.3	7.6	7.4	-1.3	4.6	3.1	-7.4	4.2	1.1	NA
Malta1	4.5	3.8	4.3	0	-7.5	-2.5	-1.4	-11.3	-3.3	-6.9	-7.3
D(X, Y; Brillenhohle, Mbuti) Brillenhohle: 13,459 SNPs											
X/Y	Han	Ust	Oase1	Kost14	GoyetQ116-	Vestonice1	Ostuni	ElMiro	Villabruna	Loschbour	LaBrana
Han	NA	1	1.7	-2	-7.6	-3.5	-4.2	-11.2	-6.3	-9.7	-8.8
UstIshim	-1	NA	0.6	-3	-7.6	-3.8	-4.2	-11.4	-6.6	-8.9	-8.8
Oase1	-1.7	-0.6	NA	-2.2	-5.1	-1.8	-3.2	-6.7	-4	-6.1	-4.9
Kostenki14	2	3	2.2	NA	-4.4	-1.2	-2	-6.7	-3.7	-5.1	-5.9
GoyetQ116-1	7.6	7.6	5.1	4.4	NA	3.5	1.8	-2.8	1.9	0.1	-0.3
Vestonice16	3.5	3.8	1.8	1.2	-3.5	NA	-0.7	-6.7	-1.7	-4	-4.2
Ostuni1	4.2	4.2	3.2	2	-1.8	0.7	NA	-4.2	-0.6	-2.8	-2.4
ElMiron	11.2	11.4	6.7	6.7	2.8	6.7	4.2	NA	5.2	3.6	3.1
Villabruna	6.3	6.6	4	3.7	-1.9	1.7	0.6	-5.2	NA	-2.3	-2.1
Loschbour	9.7	8.9	6.1	5.1	-0.1	4	2.8	-3.6	2.3	NA	-0.5
LaBrana1	8.8	8.8	4.9	5.9	0.3	4.2	2.4	-3.1	2.1	0.5	NA
Malta1	1.7	1.9	1.8	-0.5	-3.9	-1.3	-1.5	-7	-3.1	-4.3	-4.4

Table S5.5 Z-score *D(El_Miron_Cluster$_1$, El_Miron_Cluster$_2$; El_Miron_Cluster$_3$, Mbuti)*

D(X, Y; ElMiron, Mbuti) HohleFels79: 797,714 SNPs						
X/Y	HohleFels79	HohleFels49	Rigney1	GoyetQ-2	Brillenhohle	Burkhardtshohle
HohleFels79	NA	0.5	0.8	1.6	0.7	1.2
HohleFels49	-0.5	NA	-1.2	-0.9	0.6	0.9
Rigney1	-0.8	1.2	NA	0.4	-0.9	0
GoyetQ-2	-1.6	0.9	-0.4	NA	0.3	-0.1
Brillenhohle	-0.7	-0.6	0.9	-0.3	NA	0.8
Burkhardtshohle	-1.2	-0.9	0	0.1	-0.8	NA
D(X, Y; HohleFels79, Mbuti) HohleFels79: 11,211 SNPs						
X/Y	ElMiron	HohleFels49	Rigney1	GoyetQ-2	Brillenhohle	Burkhardtshohle
ElMiron	NA	-2.9	-0.4	0.2	0.5	0.3
HohleFels49	2.9	NA	-0.4	0.8	1.5	0.2
Rigney1	0.4	0.4	NA	0	1.3	0
GoyetQ-2	-0.2	-0.8	0	NA	0.8	-1.1
Brillenhohle	-0.5	-1.5	-1.3	-0.8	NA	1.5
Burkhardtshohle	-0.3	-0.2	0	1.1	-1.5	NA
D(X, Y; HohleFels49, Mbuti) HohleFels49: 63,151 SNPs						
X/Y	ElMiron	HohleFels79	Rigney1	GoyetQ-2	Brillenhohle	Burkhardtshohle
ElMiron	NA	-3.3	-4.2	-3.4	-3	-2.5
HohleFels79	3.3	NA	-0.5	1.3	-0.4	0.4
Rigney1	4.2	0.5	NA	1.8	0.6	0.1
GoyetQ-2	3.4	-1.3	-1.8	NA	-0.1	0.4
Brillenhohle	3	0.4	-0.6	0.1	NA	0.7
Burkhardtshohle	2.5	-0.4	-0.1	-0.4	-0.7	NA
D(X, Y; Rigney1, Mbuti) Rigney1: 35,600 SNPs						
X/Y	ElMiron	HohleFels79	HohleFels49	GoyetQ-2	Brillenhohle	Burkhardtshohle

ElMiron	NA	-1.1	-3.1	-5.1	-1.9	-2.1
HohleFels79	1.1	NA	-0.1	-1	1.2	-0.5
HohleFels49	3.1	0.1	NA	0.2	-0.2	0.3
GoyetQ-2	5.1	1	-0.2	NA	0	-0.3
Brillenhohle	1.9	-1.2	0.2	0	NA	0
Burkhardtshohle	2.1	0.5	-0.3	0.3	0	NA
D(X, Y; GoyetQ-2, Mbuti) GoyetQ-2: 72,263 SNPs						
X/Y	ElMiron	HohleFels79	HohleFels49	Rigney1	Brillenhohle	Burkhardtshohle
ElMiron	NA	-1.5	-2.4	-5.3	-2.5	-2.9
HohleFels79	1.5	NA	0.5	-1.3	1.3	-0.6
HohleFels49	2.4	-0.5	NA	-1.7	0.2	1
Rigney1	5.3	1.3	1.7	NA	0	-0.4
Brillenhohle	2.5	-1.3	-0.2	0	NA	-0.6
Burkhardtshohle	2.9	0.6	-1	0.4	0.6	NA
D(X, Y; Brillenhohle, Mbuti) Brillenhohle: 13,459 SNPs						
X/Y	ElMiron	HohleFels79	HohleFels49	Rigney1	GoyetQ-2	Burkhardtshohle
ElMiron	NA	-0.2	-3.5	-1.1	-2.8	0.3
HohleFels79	0.2	NA	-1.8	0	0	2.1
HohleFels49	3.5	1.8	NA	-0.7	0.3	0
Rigney1	1.1	0	0.7	NA	0	0
GoyetQ-2	2.8	0	-0.3	0	NA	-1
Burkhardtshohle	-0.3	-2.1	0	0	1	NA
D(X, Y; Burkhardtshohle, Mbuti) Burkhardtshohle: 38,376 SNPs						
X/Y	ElMiron	HohleFels79	HohleFels49	Rigney1	GoyetQ-2	Brillenhohle
ElMiron	NA	-0.9	-3.5	-2	-2.6	-0.6
HohleFels79	0.9	NA	0.2	0	0.5	0.8
HohleFels49	3.5	-0.2	NA	0.1	0.6	-0.8
Rigney1	2	0	-0.1	NA	0	0
GoyetQ-2	2.6	-0.5	-0.6	0	NA	-0.5
Brillenhohle	0.6	-0.8	0.8	0	0.5	NA

"Villabruna Cluster" (Villabruna_C) (these individuals all lived 14,000-7,000 BP)

This genetic grouping includes *BerryAuBac*, *Bichon*, *Bockstein*, *Chaudardes1*, *Falkenstein*, *Hungarian.KO1*, *LaBrana1*, *LesCloseaux13*, *Loschbour*, *Ranchot88*, *Rochedane*, *Ofnet*, and *Villabruna*, all of whom lived after around 14,000 BP. This is the largest single grouping of samples in our dataset. Because of the fact that there are many samples, and that a number of them have low coverage, we cannot co-analyse them at the set of SNPs that are covered at least once in all samples, as this would leave us with too few SNPs. In Table S5.6, we therefore analyse each sample in turn along with a selected set of high coverage samples:

- *Kostenki14*
- *GoyetQ116-1*
- *Vestonice16* (as a high coverage representative of the Věstonice Cluster)
- *ElMiron* (as a high coverage representative of the El Mirón Cluster)
- *Malta1* (as a high coverage representative of the Mal'ta Cluster)
- *Loschbour*
- *LaBrana1*
- *Hungarian.KO1*
- *Motala12*

Based on analysis of statistic like *D(X, Y; Villabruna Cluster, Mbuti)*, we find that *BerryAuBac, Bichon, Bockstein, Chaudardes1, Falkenstein, Ranchot88, Rochedane,* and *Villabruna* all show a high degree of allele sharing with Mesolithic Western Europeans including *Loschbour* and *LaBrana1*, which are sometimes also called "Western Hunter Gatherers"[4] (Table S5.6). We view all these samples as closely related, along with *Hungarian.KO1* which clusters with them despite being from an Early Neolithic context[6].

When we include the lower coverage *Continenza* and *Bockstein* (each with 10,000-30,000 SNPs), they too show a genetic affinity to "Western Hunter Gatherers" (end of Table S5.6). We do not include *LesCloseaux13* and *Ofnet* in some analyses involving the Villabruna Cluster, since after restricting to damaged sequences, both have fewer than 10,000 SNPs.

We designate this combined set of samples as a broad "Villabruna Cluster" (Table S5.6). Within this cluster, *LaBrana1* and *Hungarian.KO1* are possible outgroups to the main cluster as revealed by outgroup-f_3 analysis (Figure S5.1) as well as statistics like *D(Villabruna_Cluster$_1$, Villabruna_Cluster$_2$; Villabruna_Cluster$_3$, Mbuti)* (Table S5.7). We caution, however, that the libraries from these two samples are both non-UDG-treated (the majority of samples analysed in this study are represented by UDG-treated libraries), and this could artifactually be contributing to an inference of them being outgroups. *Falkenstein* is difficult to place, as it is genetically closer to *Chaudardes1* than it is to *Villabruna* (Table S5.7). We do not attempt to more finely place *Continenza* and *Bockstein* because of their low coverage. There are complexities in the population structure within the Villabruna Cluster, which we explore in more detail in Supplementary Information section 13.

Figure S5.1 Heat matrix of *f_3(X, Y; Mbuti)* for selected Villabruna Cluster samples. We restrict the analysis to samples with at least 30,000 SNPs covered at least once.

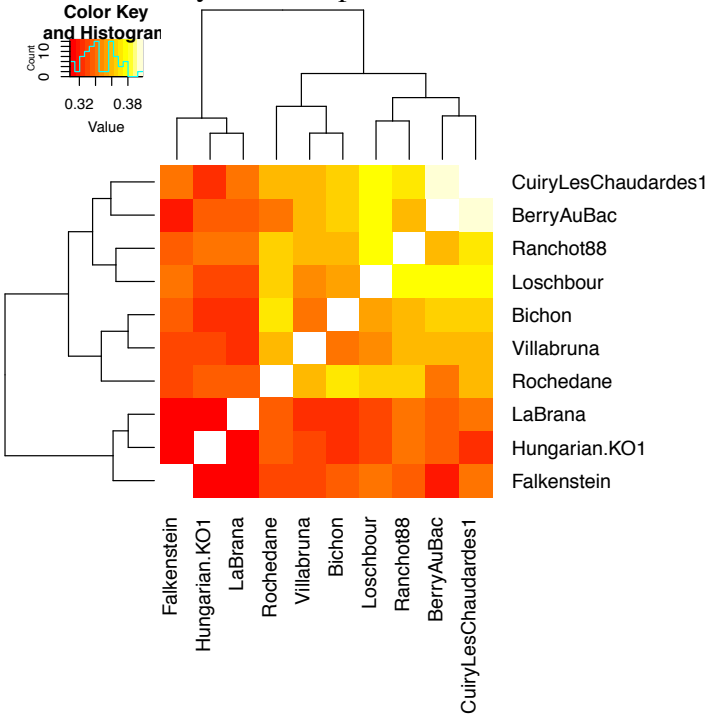


Table S5.6 Z-score of *D(X, Y; Villabruna_Cluster, Mbuti)*

D(X, Y; Villabruna, Mbuti) Villabruna: 1,215,433 SNPs								
X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-1.8	-5.3	-14.4	-29.5	-21.4	-24.2	-16.8
GoyetQ116-1	1.8	NA	-3.4	-12.4	-29.4	-19.4	-22.4	-15.2
Vestonice16	5.3	3.4	NA	-9.2	-25.1	-16.7	-18.5	-12.1
ElMiron	14.4	12.4	9.2	NA	-15.1	-6.6	-9.3	-1.8
Loschbour	29.5	29.4	25.1	15.1	NA	9.1	5.9	13.9
LaBrana1	21.4	19.4	16.7	6.6	-9.1	NA	-3	5.1
Hungarian.KO1	24.2	22.4	18.5	9.3	-5.9	3	NA	8
Motala12	16.8	15.2	12.1	1.8	-13.9	-5.1	-8	NA
Malta1	-1.8	-2.2	-6.6	-15	-31.1	-21.6	-23.8	-19.4
D(X, Y; Bichon, Mbuti) Bichon: 2,116,782 SNPs								
X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-2.4	-4.4	-16	-30.2	-24.7	-22.2	-17.1
GoyetQ116-1	2.4	NA	-2	-13.5	-28.8	-21.2	-19.7	-13.9

Vestonice16	4.4	2	NA	-12	-27.7	-19.8	-18.2	-12.7
ElMiron	16	13.5	12	NA	-15.7	-7.5	-6.8	0
Loschbour	30.2	28.8	27.7	15.7	NA	9.2	9	17
LaBrana1	24.7	21.2	19.8	7.5	-9.2	NA	0	8
Hungarian.KO1	22.2	19.7	18.2	6.8	-9	0	NA	8.2
Motala12	17.1	13.9	12.7	0	-17	-8	-8.2	NA
Malta1	-0.2	-2.5	-4.3	-15.6	-31.1	-24.5	-23	-17.1

D(X, Y; Rochedane, Mbuti) Rochedane: 237,390 SNPs

X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-2.1	-3.5	-14	-26.3	-17.9	-16.1	-12.6
GoyetQ116-1	2.1	NA	-1.5	-12.2	-23.1	-15.3	-13.5	-9.8
Vestonice16	3.5	1.5	NA	-10.7	-23.3	-14.5	-13.5	-10.2
ElMiron	14	12.2	10.7	NA	-11.2	-3.2	-2.6	1.6
Loschbour	26.3	23.1	23.3	11.2	NA	9.2	8.5	14.1
LaBrana1	17.9	15.3	14.5	3.2	-9.2	NA	0.7	5.1
Hungarian.KO1	16.1	13.5	13.5	2.6	-8.5	-0.7	NA	4.2
Motala12	12.6	9.8	10.2	-1.6	-14.1	-5.1	-4.2	NA
Malta1	-0.7	-3.3	-4.3	-13.5	-25.2	-17.9	-15.5	-13.6

D(X, Y; Ranchot88, Mbuti) Ranchot88: 414,863 SNPs

X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-3.7	-5.2	-16.8	-34.4	-23	-20.4	-14.4
GoyetQ116-1	3.7	NA	-0.9	-13.2	-29	-19	-17	-10.5
Vestonice16	5.2	0.9	NA	-12.2	-29.2	-17.8	-17.1	-10.1
ElMiron	16.8	13.2	12.2	NA	-16.8	-5.8	-4.7	2.6
Loschbour	34.4	29	29.2	16.8	NA	12.2	12.2	20.4
LaBrana1	23	19	17.8	5.8	-12.2	NA	0.9	8.2
Hungarian.KO1	20.4	17	17.1	4.7	-12.2	-0.9	NA	7.1
Motala12	14.4	10.5	10.1	-2.6	-20.4	-8.2	-7.1	NA
Malta1	0	-3.9	-5.2	-16.1	-31.5	-21.8	-20.7	-15

D(X, Y; Falkenstein, Mbuti) Falkenstein: 64,428 SNPs

X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-0.8	-2.6	-9.3	-20.4	-14.1	-11.2	-10.1
GoyetQ116-1	0.8	NA	-1.4	-8	-18.4	-11.7	-9.5	-8.1
Vestonice16	2.6	1.4	NA	-6.9	-17.1	-11.5	-9.2	-6.9
ElMiron	9.3	8	6.9	NA	-10.6	-5.1	-3.1	-0.2
Loschbour	20.4	18.4	17.1	10.6	NA	5.7	6.5	9.8
LaBrana1	14.1	11.7	11.5	5.1	-5.7	NA	1.4	4.3
Hungarian.KO1	11.2	9.5	9.2	3.1	-6.5	-1.4	NA	3.1
Motala12	10.1	8.1	6.9	0.2	-9.8	-4.3	-3.1	NA
Malta1	-0.8	-1.6	-3	-8.4	-19.6	-14.2	-10.7	-9.4

D(X, Y; Chaudardes1, Mbuti) Chaudardes1: 92,657 SNPs

X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-1.8	-4	-12	-26	-17.4	-12.4	-12.4
GoyetQ116-1	1.8	NA	-1.9	-9.7	-23.1	-14.2	-11.3	-9.9
Vestonice16	4	1.9	NA	-7.9	-20.6	-12.5	-8.7	-7.1
ElMiron	12	9.7	7.9	NA	-11.8	-3.5	-1.5	0.8
Loschbour	26	23.1	20.6	11.8	NA	8.8	10.7	13.6
LaBrana1	17.4	14.2	12.5	3.5	-8.8	NA	3.4	4.6
Hungarian.KO1	12.4	11.3	8.7	1.5	-10.7	-3.4	NA	1.6
Motala12	12.4	9.9	7.1	-0.8	-13.6	-4.6	-1.6	NA
Malta1	1.6	0.7	-1.7	-8.7	-21.7	-13.7	-9.9	-10

D(X, Y; BerryAuBac, Mbuti) BerryAuBac: 54,690 SNPs

X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-3.1	-2.6	-8.5	-19.9	-12.8	-12.4	-9.2
GoyetQ116-1	3.1	NA	0.1	-7	-16.9	-10.6	-9.8	-6.5
Vestonice16	2.6	-0.1	NA	-6	-17.6	-10.9	-9.6	-6.8
ElMiron	8.5	7	6	NA	-9.2	-3.9	-3.1	0.1
Loschbour	19.9	16.9	17.6	9.2	NA	6.8	6.6	10.6
LaBrana1	12.8	10.6	10.9	3.9	-6.8	NA	0.7	3.3
Hungarian.KO1	12.4	9.8	9.6	3.1	-6.6	-0.7	NA	2.5
Motala12	9.2	6.5	6.8	-0.1	-10.6	-3.3	-2.5	NA
Malta1	0.9	-1.7	-1.2	-7.5	-17.3	-12.5	-10.5	-8.1

D(X, Y; Continenza, Mbuti) Continenza: 11,717 SNPs

X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-0.4	-0.8	-4.4	-10.2	-6.1	-6	-3.1
GoyetQ116-1	0.4	NA	0.1	-4	-9.7	-4.8	-5.7	-2.8
Vestonice16	0.8	-0.1	NA	-4.4	-9.5	-5.5	-6.2	-2.9
ElMiron	4.4	4	4.4	NA	-4.7	-1.2	-1.8	1.7
Loschbour	10.2	9.7	9.5	4.7	NA	4.3	2.2	6.6
LaBrana1	6.1	4.8	5.5	1.2	-4.3	NA	-0.9	2.7
Hungarian.KO1	6	5.7	6.2	1.8	-2.2	0.9	NA	3.5
Motala12	3.1	2.8	2.9	-1.7	-6.6	-2.7	-3.5	NA

Malta1	-1.3	-1.2	-1.8	-5.1	-10.2	-7	-7	-4.9

D(X, Y; Bockstein, Mbuti) Bockstein: 21,977 SNPs								
X/Y	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Loschbour	LaBrana1	Hungarian.KO1	Motala12
Kostenki14	NA	-1.7	-0.8	-4.6	-11	-6.5	-6.6	-5.4
GoyetQ116-1	1.7	NA	0.5	-3.1	-9.6	-4.6	-6.2	-3.5
Vestonice16	0.8	-0.5	NA	-4.2	-10.2	-6	-6.9	-4.9
ElMiron	4.6	3.1	4.2	NA	-6.1	-1.7	-2.5	-0.5
Loschbour	11	9.6	10.2	6.1	NA	3.8	2.3	5.4
LaBrana1	6.5	4.6	6	1.7	-3.8	NA	-1.1	1
Hungarian.KO1	6.6	6.2	6.9	2.5	-2.3	1.1	NA	2.4
Motala12	5.4	3.5	4.9	0.5	-5.4	-1	-2.4	NA
Malta1	-0.2	-1.5	-0.6	-5.1	-10.6	-6	-7.4	-5.6

Table S5.7 Z of *D(Villabruna_Cluster$_1$, Villabruna_Cluster$_2$; Villabruna_Cluster$_3$, Mbuti)*

D(X, Y; Villabruna, Mbuti) Villabruna: 1,215,433 SNPs									
X/Y	Bichon	Rochedane	Ranchot88	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Bichon	NA	-0.7	-0.4	-1.3	0.9	0.3	-1.2	7.5	3.9
Rochedane	0.7	NA	0.4	-0.9	0.8	0.9	-0.1	7.8	4.5
Ranchot88	0.4	-0.4	NA	-0.2	0.2	0.7	-1	8.2	3.9
Falkenstein	1.3	0.9	0.2	NA	-1	-0.1	0.2	6.8	3.7
Chaudardes1	-0.9	-0.8	-0.2	1	NA	0.5	-2.1	4.7	2.1
BerryAuBac	-0.3	-0.9	-0.7	0.1	-0.5	NA	-1	4.2	1.7
Loschbour	1.2	0.1	1	-0.2	2.1	1	NA	9.1	5.9
LaBrana1	-7.5	-7.8	-8.2	-6.8	-4.7	-4.2	-9.1	NA	-3
Hungarian.KO1	-3.9	-4.5	-3.9	-3.7	-2.1	-1.7	-5.9	3	NA
D(X, Y; Bichon, Mbuti) Bichon: 2,116,782 SNPs									
X/Y	Villabruna	Rochedane	Ranchot88	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-3.4	-1.5	-2.8	-0.1	-1	-4.2	4.4	4.7
Rochedane	3.4	NA	2.3	0.2	2.4	0.7	0.1	8.1	7.4
Ranchot88	1.5	-2.3	NA	-1.2	0.7	-0.3	-2.6	6.6	6.2
Falkenstein	2.8	-0.2	1.2	NA	-0.3	-0.2	0.6	6.4	5.4
Chaudardes1	0.1	-2.4	-0.7	0.3	NA	0.1	-2.6	4.1	3.3
BerryAuBac	1	-0.7	0.3	0.2	-0.1	NA	-0.5	4.2	4.3
Loschbour	4.2	-0.1	2.6	-0.6	2.6	0.5	NA	9.2	9
LaBrana1	-4.4	-8.1	-6.6	-6.4	-4.1	-4.2	-9.2	NA	0
Hungarian.KO1	-4.7	-7.4	-6.2	-5.4	-3.3	-4.3	-9	0	NA
D(X, Y; Rochedane, Mbuti) Rochedane: 237,390 SNPs									
X/Y	Villabruna	Bichon	Ranchot88	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-2.7	-0.8	-1.3	0.6	2.9	-0.6	7.9	7.8
Bichon	2.7	NA	1.6	0.8	2.4	3	2.7	10.1	9.2
Ranchot88	0.8	-1.6	NA	-0.9	0.1	2.2	-0.1	7.7	5.9
Falkenstein	1.3	-0.8	0.9	NA	0.4	-0.6	2	3.6	4.1
Chaudardes1	-0.6	-2.4	-0.1	-0.4	NA	0.6	-2.4	3.2	1.5
BerryAuBac	-2.9	-3	-2.2	0.6	-0.6	NA	-2.6	1.4	1.1
Loschbour	0.6	-2.7	0.1	-2	2.4	2.6	NA	9.2	8.5
LaBrana1	-7.9	-10.1	-7.7	-3.6	-3.2	-1.4	-9.2	NA	0.7
Hungarian.KO1	-7.8	-9.2	-5.9	-4.1	-1.5	-1.1	-8.5	-0.7	NA
D(X, Y; Ranchot88, Mbuti) Ranchot88: 700833 SNPs									
X/Y	Villabruna	Bichon	Rochedane	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-1.1	-1.1	-2.2	-2.8	-0.3	-6.1	5	4.5
Bichon	1.1	NA	-0.8	-2	-1.9	-0.3	-5.2	5.9	6.2
Rochedane	1.1	0.8	NA	-2	-0.8	0	-3.9	6	4.8
Falkenstein	2.2	2	2	NA	0.4	-1	-0.5	4.4	4.5
Chaudardes1	2.8	1.9	0.8	-0.4	NA	0.5	-1.5	5.2	4
BerryAuBac	0.3	0.3	0	1	-0.5	NA	-2.1	3	3.3
Loschbour	6.1	5.2	3.9	0.5	1.5	2.1	NA	12.2	12.2
LaBrana1	-5	-5.9	-6	-4.4	-5.2	-3	-12.2	NA	0.9
Hungarian.KO1	-4.5	-6.2	-4.8	-4.5	-4	-3.3	-12.2	-0.9	NA
D(X, Y; Falkenstein, Mbuti) Falkenstein: 64,428 SNPs									
X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-1.5	-0.4	-2.1	-3.4	-0.3	-2	3.4	3.3
Bichon	1.5	NA	0.6	-0.9	-0.3	0.7	-0.9	5	4.4
Rochedane	0.4	-0.6	NA	-1.2	-0.1	-1	-1.2	1.5	2.2
Ranchot88	2.1	0.9	1.2	NA	0.4	0.8	0.5	4.7	4.4
Chaudardes1	3.4	0.3	0.1	-0.4	NA	1.1	0.1	1.8	2.2
BerryAuBac	0.3	-0.7	1	-0.8	-1.1	NA	0.2	1	1.5
Loschbour	2	0.9	1.2	-0.5	-0.1	-0.2	NA	5.7	6.5
LaBrana1	-3.4	-5	-1.5	-4.7	-1.8	-1	-5.7	NA	1.4
Hungarian.KO1	-3.3	-4.4	-2.2	-4.4	-2.2	-1.5	-6.5	-1.4	NA
D(X, Y; Chaudardes1, Mbuti) Chaudardes1: 92,657 SNPs									
X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-1	-0.2	-3	-2.3	-1	-4.7	3.4	5.4
Bichon	1	NA	0.2	-2.7	0	-0.6	-4.1	4.8	6.2
Rochedane	0.2	-0.2	NA	-0.9	-0.5	-0.8	-2.5	2.9	2.9
Ranchot88	3	2.7	0.9	NA	0.1	-1.3	-0.8	5.3	6.4
Falkenstein	2.3	0	0.5	-0.1	NA	1	0.3	2.9	3.9
BerryAuBac	1	0.6	0.8	1.3	-1	NA	0.1	2.5	2.3
Loschbour	4.7	4.1	2.5	0.8	-0.3	-0.1	NA	8.8	10.7

LaBrana1	-3.4	-4.8	-2.9	-5.3	-2.9	-2.5	-8.8	NA	3.4
Hungarian.KO1	-5.4	-6.2	-2.9	-6.4	-3.9	-2.3	-10.7	-3.4	NA

D(X, Y; BerryAuBac, Mbuti) BerryAuBac: 54,690

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Loschbour	LaBrana1	KO1
Villabruna	NA	-1.2	2.2	-1.1	-0.1	-1.5	-3.4	3.2	3.4
Bichon	1.2	NA	2.5	0	1	-0.7	-2	4.4	4.2
Rochedane	-2.2	-2.5	NA	-2.4	-0.4	-1.3	-3.7	0.8	0.7
Ranchot88	1.1	0	2.4	NA	1.9	-1.8	-1.1	3.6	3
Falkenstein	0.1	-1	0.4	-1.9	NA	-0.2	-0.5	1.5	1.2
Chaudardes1	1.5	0.7	1.3	1.8	0.2	NA	1.3	3.7	2.6
Loschbour	3.4	2	3.7	1.1	0.5	-1.3	NA	6.8	6.6
LaBrana1	-3.2	-4.4	-0.8	-3.6	-1.5	-3.7	-6.8	NA	0.7
Hungarian.KO1	-3.4	-4.2	-0.7	-3	-1.2	-2.6	-6.6	-0.7	NA

D(X, Y; Loschbour, Mbuti) Loschbour: 2,091,584 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Berry	LaBrana1	KO1
Villabruna	NA	-3	-0.6	-5.2	-2.2	-2.6	-2.4	4.2	4.6
Bichon	3	NA	2.5	-2.8	-1.5	-1.4	-1.5	7.2	7
Rochedane	0.6	-2.5	NA	-3.7	-3.2	-0.3	-1.1	5.2	5.2
Ranchot88	5.2	2.8	3.7	NA	1.1	0.7	1	10	9.8
Falkenstein	2.2	1.5	3.2	-1.1	NA	0.2	-0.6	5.7	5
Chaudardes1	2.6	1.4	0.3	-0.7	-0.2	NA	1.1	6.9	6.8
BerryAuBac	2.4	1.5	1.1	-1	0.6	-1.1	NA	6	4.8
LaBrana1	-4.2	-7.2	-5.2	-10	-5.7	-6.9	-6	NA	0.1
Hungarian.KO1	-4.6	-7	-5.2	-9.8	-5	-6.8	-4.8	-0.1	NA

D(X, Y; LaBrana1, Mbuti) LaBrana1: 1,884,745 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Berry	Loschbour	KO1
Villabruna	NA	-3.2	0.1	-3.6	-3.4	-1.4	-0.9	-5.7	3.4
Bichon	3.2	NA	2	-0.6	-1.6	0.7	0.4	-1.8	5.7
Rochedane	-0.1	-2	NA	-1.6	-2	-0.2	-0.6	-4.2	3.4
Ranchot88	3.6	0.6	1.6	NA	0.3	0.5	0.7	-2.1	6.6
Falkenstein	3.4	1.6	2	-0.3	NA	1.1	0.5	-0.4	5
Chaudardes1	1.4	-0.7	0.2	-0.5	-1.1	NA	1	-2.2	5.2
BerryAuBac	0.9	-0.4	0.6	-0.7	-0.5	-1	NA	-1.2	2.7
Loschbour	5.7	1.8	4.2	2.1	0.4	2.2	1.2	NA	7.9
Hungarian.KO1	-3.4	-5.7	-3.4	-6.6	-5	-5.2	-2.7	-7.9	NA

D(X, Y; Hungarian.KO1, Mbuti) Hungarian.KO1: 1,410,303 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Berry	Loschbour	LaBrana1
Villabruna	NA	0.7	2.9	0.4	-0.2	3.5	1.6	-1.4	6.4
Bichon	-0.7	NA	2.1	0.1	-0.6	2.9	0	-1.8	5.5
Rochedane	-2.9	-2.1	NA	-1.1	-1.7	1.4	-0.4	-3.1	2.9
Ranchot88	-0.4	-0.1	1.1	NA	-0.1	2.5	-0.2	-1.7	5.4
Falkenstein	0.2	0.6	1.7	0.1	NA	1.6	-0.3	-1.9	3.4
Chaudardes1	-3.5	-2.9	-1.4	-2.5	-1.6	NA	0.3	-4	1.8
BerryAuBac	-1.6	0	0.4	0.2	0.3	-0.3	NA	-2.1	2
Loschbour	1.4	1.8	3.1	1.7	1.9	4	2.1	NA	7.4
LaBrana1	-6.4	-5.5	-2.9	-5.4	-3.4	-1.8	-2	-7.4	NA

D(X, Y; Villabruna, Mbuti) Villabruna: 1,215,433 SNPs

X/Y	Bichon	Rochedane	Ranchot88	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Bichon	NA	-0.7	-0.4	-1.3	0.9	0.3	-1.2	7.5	3.9
Rochedane	0.7	NA	0.4	-0.9	0.8	0.9	-0.1	7.8	4.5
Ranchot88	0.4	-0.4	NA	-0.2	0.2	0.7	-1	8.2	3.9
Falkenstein	1.3	0.9	0.2	NA	-1	-0.1	0.2	6.8	3.7
Chaudardes1	-0.9	-0.8	-0.2	1	NA	0.5	-2.1	4.7	2.1
BerryAuBac	-0.3	-0.9	-0.7	0.1	-0.5	NA	-1	4.2	1.7
Loschbour	1.2	0.1	1	-0.2	2.1	1	NA	9.1	5.9
LaBrana1	-7.5	-7.8	-8.2	-6.8	-4.7	-4.2	-9.1	NA	-3
Hungarian.KO1	-3.9	-4.5	-3.9	-3.7	-2.1	-1.7	-5.9	3	NA

D(X, Y; Bichon, Mbuti) Bichon: 2,116,782 SNPs

X/Y	Villabruna	Rochedane	Ranchot88	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-3.4	-1.5	-2.8	-0.1	-1	-4.2	4.4	4.7
Rochedane	3.4	NA	2.3	0.2	2.4	0.7	0.1	8.1	7.4
Ranchot88	1.5	-2.3	NA	-1.2	0.7	-0.3	-2.6	6.6	6.2
Falkenstein	2.8	-0.2	1.2	NA	-0.3	-0.2	0.6	6.4	5.4
Chaudardes1	0.1	-2.4	-0.7	0.3	NA	0.1	-2.6	4.1	3.3
BerryAuBac	1	-0.7	0.3	0.2	-0.1	NA	-0.5	4.2	4.3
Loschbour	4.2	-0.1	2.6	-0.6	2.6	0.5	NA	9.2	9
LaBrana1	-4.4	-8.1	-6.6	-6.4	-4.1	-4.2	-9.2	NA	0
Hungarian.KO1	-4.7	-7.4	-6.2	-5.4	-3.3	-4.3	-9	0	NA

D(X, Y; Rochedane, Mbuti) Rochedane: 237,390 SNPs

X/Y	Villabruna	Bichon	Ranchot88	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-2.7	-0.8	-1.3	0.6	2.9	-0.6	7.9	7.8
Bichon	2.7	NA	1.6	0.8	2.4	3	2.7	10.1	9.2
Ranchot88	0.8	-1.6	NA	-0.9	0.1	2.2	-0.1	7.7	5.9
Falkenstein	1.3	-0.8	0.9	NA	0.4	-0.6	2	3.6	4.1
Chaudardes1	-0.6	-2.4	-0.1	-0.4	NA	0.6	-2.4	3.2	1.5
BerryAuBac	-2.9	-3	-2.2	0.6	-0.6	NA	-2.6	1.4	1.1
Loschbour	0.6	-2.7	0.1	-2	2.4	2.6	NA	9.2	8.5
LaBrana1	-7.9	-10.1	-7.7	-3.6	-3.2	-1.4	-9.2	NA	0.7
Hungarian.KO1	-7.8	-9.2	-5.9	-4.1	-1.5	-1.1	-8.5	-0.7	NA

D(X, Y; Ranchot88, Mbuti) Ranchot88: 700,833 SNPs

X/Y	Villabruna	Bichon	Rochedane	Falkenstein	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-1.1	-1.1	-2.2	-2.8	-0.3	-6.1	5	4.5
Bichon	1.1	NA	-0.8	-2	-1.9	-0.3	-5.2	5.9	6.2
Rochedane	1.1	0.8	NA	-2	-0.8	0	-3.9	6	4.8
Falkenstein	2.2	2	2	NA	0.4	-1	-0.5	4.4	4.5
Chaudardes1	2.8	1.9	0.8	-0.4	NA	0.5	-1.5	5.2	4
BerryAuBac	0.3	0.3	0	1	-0.5	NA	-2.1	3	3.3
Loschbour	6.1	5.2	3.9	0.5	1.5	2.1	NA	12.2	12.2
LaBrana1	-5	-5.9	-6	-4.4	-5.2	-3	-12.2	NA	0.9
Hungarian.KO1	-4.5	-6.2	-4.8	-4.5	-4	-3.3	-12.2	-0.9	NA

D(X, Y; Falkenstein, Mbuti) Falkenstein: 64,428 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Chaudardes	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-1.5	-0.4	-2.1	-3.4	-0.3	-2	3.4	3.3
Bichon	1.5	NA	0.6	-0.9	-0.3	0.7	-0.9	5	4.4
Rochedane	0.4	-0.6	NA	-1.2	-0.1	-1	-1.2	1.5	2.2
Ranchot88	2.1	0.9	1.2	NA	0.4	0.8	0.5	4.7	4.4
Chaudardes1	3.4	0.3	0.1	-0.4	NA	1.1	0.1	1.8	2.2
BerryAuBac	0.3	-0.7	1	-0.8	-1.1	NA	0.2	1	1.5
Loschbour	2	0.9	1.2	-0.5	-0.1	-0.2	NA	5.7	6.5
LaBrana1	-3.4	-5	-1.5	-4.7	-1.8	-1	-5.7	NA	1.4
Hungarian.KO1	-3.3	-4.4	-2.2	-4.4	-2.2	-1.5	-6.5	-1.4	NA

D(X, Y; Chaudardes1, Mbuti) Chaudardes1: 92,657 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Berry	Loschbour	LaBrana1	KO1
Villabruna	NA	-1	-0.2	-3	-2.3	-1	-4.7	3.4	5.4
Bichon	1	NA	0.2	-2.7	0	-0.6	-4.1	4.8	6.2
Rochedane	0.2	-0.2	NA	-0.9	-0.5	-0.8	-2.5	2.9	2.9
Ranchot88	3	2.7	0.9	NA	0.1	-1.3	-0.8	5.3	6.4
Falkenstein	2.3	0	0.5	-0.1	NA	1	0.3	2.9	3.9
BerryAuBac	1	0.6	0.8	1.3	-1	NA	0.1	2.5	2.3
Loschbour	4.7	4.1	2.5	0.8	-0.3	-0.1	NA	8.8	10.7
LaBrana1	-3.4	-4.8	-2.9	-5.3	-2.9	-2.5	-8.8	NA	3.4
Hungarian.KO1	-5.4	-6.2	-2.9	-6.4	-3.9	-2.3	-10.7	-3.4	NA

D(X, Y; BerryAuBac, Mbuti) BerryAuBac: 54,690 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Loschbour	LaBrana1	KO1
Villabruna	NA	-1.2	2.2	-1.1	-0.1	-1.5	-3.4	3.2	3.4
Bichon	1.2	NA	2.5	0	1	-0.7	-2	4.4	4.2
Rochedane	-2.2	-2.5	NA	-2.4	-0.4	-1.3	-3.7	0.8	0.7
Ranchot88	1.1	0	2.4	NA	1.9	-1.8	-1.1	3.6	3
Falkenstein	0.1	-1	0.4	-1.9	NA	-0.2	-0.5	1.5	1.2
Chaudardes1	1.5	0.7	1.3	1.8	0.2	NA	1.3	3.7	2.6
Loschbour	3.4	2	3.7	1.1	0.5	-1.3	NA	6.8	6.6
LaBrana1	-3.2	-4.4	-0.8	-3.6	-1.5	-3.7	-6.8	NA	0.7
Hungarian.KO1	-3.4	-4.2	-0.7	-3	-1.2	-2.6	-6.6	-0.7	NA

D(X, Y; Loschbour, Mbuti) Loschbour: 2,091,584 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Berry	LaBrana1	KO1
Villabruna	NA	-3	-0.6	-5.2	-2.2	-2.6	-2.4	4.2	4.6
Bichon	3	NA	2.5	-2.8	-1.5	-1.4	-1.5	7.2	7
Rochedane	0.6	-2.5	NA	-3.7	-3.2	-0.3	-1.1	5.2	5.2
Ranchot88	5.2	2.8	3.7	NA	1.1	0.7	1	10	9.8
Falkenstein	2.2	1.5	3.2	-1.1	NA	0.2	-0.6	5.7	5
Chaudardes1	2.6	1.4	0.3	-0.7	-0.2	NA	1.1	6.9	6.8
BerryAuBac	2.4	1.5	1.1	-1	0.6	-1.1	NA	6	4.8
LaBrana1	-4.2	-7.2	-5.2	-10	-5.7	-6.9	-6	NA	0.1
Hungarian.KO1	-4.6	-7	-5.2	-9.8	-5	-6.8	-4.8	-0.1	NA

D(X, Y; LaBrana1, Mbuti) LaBrana1: 1,884,745 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Berry	Loschbour	KO1
Villabruna	NA	-3.2	0.1	-3.6	-3.4	-1.4	-0.9	-5.7	3.4
Bichon	3.2	NA	2	-0.6	-1.6	0.7	0.4	-1.8	5.7
Rochedane	-0.1	-2	NA	-1.6	-2	-0.2	-0.6	-4.2	3.4
Ranchot88	3.6	0.6	1.6	NA	0.3	0.5	0.7	-2.1	6.6
Falkenstein	3.4	1.6	2	-0.3	NA	1.1	0.5	-0.4	5
Chaudardes1	1.4	-0.7	0.2	-0.5	-1.1	NA	1	-2.2	5.2
BerryAuBac	0.9	-0.4	0.6	-0.7	-0.5	-1	NA	-1.2	2.7
Loschbour	5.7	1.8	4.2	2.1	0.4	2.2	1.2	NA	7.9
Hungarian.KO1	-3.4	-5.7	-3.4	-6.6	-5	-5.2	-2.7	-7.9	NA

D(X, Y; Hungarian.KO1, Mbuti) Hungarian.KO1: 1,410,303 SNPs

X/Y	Villabruna	Bichon	Rochedane	Ranchot88	Falkenstein	Chau	Berry	Loschbour	LaBrana1
Villabruna	NA	0.7	2.9	0.4	-0.2	3.5	1.6	-1.4	6.4
Bichon	-0.7	NA	2.1	0.1	-0.6	2.9	0	-1.8	5.5
Rochedane	-2.9	-2.1	NA	-1.1	-1.7	1.4	-0.4	-3.1	2.9
Ranchot88	-0.4	-0.1	1.1	NA	-0.1	2.5	-0.2	-1.7	5.4
Falkenstein	0.2	0.6	1.7	0.1	NA	1.6	-0.3	-1.9	3.4
Chaudardes1	-3.5	-2.9	-1.4	-2.5	-1.6	NA	0.3	-4	1.8
BerryAuBac	-1.6	0	0.4	0.2	0.3	-0.3	NA	-2.1	2
Loschbour	1.4	1.8	3.1	1.7	1.9	4	2.1	NA	7.4
LaBrana1	-6.4	-5.5	-2.9	-5.4	-3.4	-1.8	-2	-7.4	NA

"Mal'ta cluster" (these individuals all lived 24,000-17,000 BP in Siberia)

Samples from the sites of Mal'ta (*Malta1*) and Afontova Gora (*AfontovaGora2* and *AfontovaGora3*)—both east of the Ural Mountains—descend from a common ancestral population relative to the other samples, but also show important sub-structure (Table S5.8):

- *Malta1* shares more drift with *AfontovaGora3* than with pre-Neolithic Europeans.
- *AfontovaGora3* appears to derive from a lineage of the Mal'ta Cluster that contributed more to some later human populations than did the lineage leading to *Malta1* itself, as: (i) *Karelia* shares more alleles with *AfontovaGora3* than with *Malta1*, and (ii) Native Americans share more alleles with *AfontovaGora3* than with *Malta1*.
- *AfontovaGora2* is not genetically closer to *AfontovaGora3* than it is to *Malta1*. Thus, there is no evidence of an "Afontova Gora Cluster".

Table S5.8 *D(X, Y; Z, Mbuti)* with *Malta1* and *AfontovaGora3* always in the statistic.

X	Y	Z	Mbuti	D value	Z score	Sites used
Malta1	AfontovaGora3	UstIshim	Mbuti	0.0008	1.3	509641
Malta1	AfontovaGora3	Oase1	Mbuti	0.0003	0.4	116250
Malta1	AfontovaGora3	Kostenki14	Mbuti	0.0011	1.7	486631
Malta1	AfontovaGora3	GoyetQ116-1	Mbuti	0.0009	1.4	401152
Malta1	AfontovaGora3	Vestonice16	Mbuti	0.0010	1.5	398604
Malta1	AfontovaGora3	Ostuni1	Mbuti	0.0006	0.7	189230
Malta1	AfontovaGora3	KremsWA3	Mbuti	0.0004	0.5	120508
Malta1	AfontovaGora3	ElMiron	Mbuti	0.0001	0.1	354861
Malta1	AfontovaGora3	Villabruna	Mbuti	-0.0004	-0.6	459597
Malta1	AfontovaGora3	Ranchot88	Mbuti	-0.0001	-0.2	223016
Malta1	AfontovaGora3	Loschbour	Mbuti	0.0004	0.7	505247
Malta1	AfontovaGora3	LaBrana1	Mbuti	-0.0008	-1.3	489381
Malta1	AfontovaGora3	Hungarian.KO1	Mbuti	-0.0005	-0.8	354548
Malta1	AfontovaGora3	Motala12	Mbuti	-0.0005	-0.4	55428
Malta1	AfontovaGora3	Karelia	Mbuti	-0.0055	-7.5	444390
Malta1	AfontovaGora3	AfontovaGora2	Mbuti	0.0016	1.2	45578
Malta1	AfontovaGora3	Stuttgart	Mbuti	0.0000	-0.1	502197
Malta1	AfontovaGora3	French	Mbuti	-0.0006	-1.4	510782
Malta1	AfontovaGora3	Sardinian	Mbuti	0.0001	0.2	510785
Malta1	AfontovaGora3	Han	Mbuti	0.0002	0.4	510784
Malta1	AfontovaGora3	Dai	Mbuti	0.0003	0.6	510787
Malta1	AfontovaGora3	Karitiana	Mbuti	-0.0030	-5.3	510778
Malta1	UstIshim	AfontovaGora3	Mbuti	0.0247	28.6	509641
Malta1	Oase1	AfontovaGora3	Mbuti	0.0266	22.9	116250
Malta1	Kostenki14	AfontovaGora3	Mbuti	0.0192	23.8	486631
Malta1	GoyetQ116-1	AfontovaGora3	Mbuti	0.0181	20.9	401152
Malta1	Vestonice16	AfontovaGora3	Mbuti	0.0180	21.0	398604
Malta1	Ostuni1	AfontovaGora3	Mbuti	0.0179	18.2	189230
Malta1	KremsWA3	AfontovaGora3	Mbuti	0.0180	16.0	120508
Malta1	ElMiron	AfontovaGora3	Mbuti	0.0179	20.8	354861
Malta1	Villabruna	AfontovaGora3	Mbuti	0.0165	20.0	459597
Malta1	Ranchot88	AfontovaGora3	Mbuti	0.0174	18.7	223016
Malta1	Loschbour	AfontovaGora3	Mbuti	0.0167	21.1	505247
Malta1	LaBrana1	AfontovaGora3	Mbuti	0.0162	20.0	489381
Malta1	Hungarian.KO1	AfontovaGora3	Mbuti	0.0161	19.4	354548
Malta1	Motala12	AfontovaGora3	Mbuti	0.0116	8.9	55428
Malta1	Karelia	AfontovaGora3	Mbuti	0.0037	4.2	444390
Malta1	AfontovaGora2	AfontovaGora3	Mbuti	0.0011	0.8	45578
Malta1	Stuttgart	AfontovaGora3	Mbuti	0.0212	27.7	502197
Malta1	French	AfontovaGora3	Mbuti	0.0173	25.3	510782
Malta1	Sardinian	AfontovaGora3	Mbuti	0.0202	29.1	510785
Malta1	Han	AfontovaGora3	Mbuti	0.0222	31.0	510784
Malta1	Dai	AfontovaGora3	Mbuti	0.0220	31.2	510787
Malta1	Karitiana	AfontovaGora3	Mbuti	0.0112	14.9	510778
UstIshim	AfontovaGora3	Malta1	Mbuti	-0.0239	-26.8	509641
Oase1	AfontovaGora3	Malta1	Mbuti	-0.0263	-23.1	116250
Kostenki14	AfontovaGora3	Malta1	Mbuti	-0.0180	-20.8	486631
GoyetQ116-1	AfontovaGora3	Malta1	Mbuti	-0.0172	-18.6	401152
Vestonice16	AfontovaGora3	Malta1	Mbuti	-0.0170	-19.2	398604
Ostuni1	AfontovaGora3	Malta1	Mbuti	-0.0173	-17.5	189230
KremsWA3	AfontovaGora3	Malta1	Mbuti	-0.0176	-16.4	120508

ElMiron	AfontovaGora3	Malta1	Mbuti	-0.0178	-20.1	354861
Villabruna	AfontovaGora3	Malta1	Mbuti	-0.0169	-20.0	459597
Ranchot88	AfontovaGora3	Malta1	Mbuti	-0.0175	-18.5	223016
Loschbour	AfontovaGora3	Malta1	Mbuti	-0.0163	-20.4	505247
LaBrana1	AfontovaGora3	Malta1	Mbuti	-0.0171	-21.0	489381
Hungarian.KO1	AfontovaGora3	Malta1	Mbuti	-0.0166	-19.9	354548
Motala12	AfontovaGora3	Malta1	Mbuti	-0.0121	-9.1	55428
Karelia	AfontovaGora3	Malta1	Mbuti	-0.0091	-10.9	444390
AfontovaGora2	AfontovaGora3	Malta1	Mbuti	0.0005	0.4	45578
Stuttgart	AfontovaGora3	Malta1	Mbuti	-0.0212	-26.7	502197
French	AfontovaGora3	Malta1	Mbuti	-0.0180	-25.9	510782
Sardinian	AfontovaGora3	Malta1	Mbuti	-0.0202	-28.0	510785
Han	AfontovaGora3	Malta1	Mbuti	-0.0220	-28.6	510784
Dai	AfontovaGora3	Malta1	Mbuti	-0.0218	-28.8	510787
Karitiana	AfontovaGora3	Malta1	Mbuti	-0.0142	-19.4	510778

"Satsurblia Cluster" (these individuals all lived 13,000-10,000 BP in the Caucasus)

The *Satsurblia* and *Kotias*[8] samples have several features that are distinctive from those of the other samples in this study (Table S5.9):

- *Satsurblia* shares more drift with *Kotias* than with pre-Neolithic Europeans.
- *ElMiron* shares more drift with *Kotias* (9,700 BP) than with *Satsurblia* (13,300 BP).

The relationships of these samples to the ancient Europeans who are the focus of this study is investigated in more detail in Supplementary Information section 12.

Table S5.9 *D(X, Y; Z, Mbuti)* where *Satsurblia* and *Kotias* are always in the statistic.

				All sites			Transversions only		
X	**Y**	**Z**	**Mbuti**	**D value**	**Z score**	**Sites used**	**D value**	**Z score**	**Sites used**
Satsurblia	Kotias	UstIshim	Mbuti	0.0002	0.4	1415190	0.0001	0.2	766712
Satsurblia	Kotias	Oase1	Mbuti	-0.0007	-1.1	189526	-0.0009	-1.0	82749
Satsurblia	Kotias	Kostenki14	Mbuti	-0.0007	-1.2	1152869	-0.0004	-0.8	553455
Satsurblia	Kotias	GoyetQ116-1	Mbuti	-0.0012	-1.9	558196	-0.0014	-1.6	130649
Satsurblia	Kotias	Vestonice16	Mbuti	-0.0005	-0.8	625570	-0.0002	-0.3	217427
Satsurblia	Kotias	Ostuni1	Mbuti	-0.0003	-0.4	246277	-0.0006	-0.5	71079
Satsurblia	Kotias	KremsWA3	Mbuti	-0.0010	-1.2	157117	-0.0014	-1.0	35085
Satsurblia	Kotias	ElMiron	Mbuti	-0.0019	-3.4	529583	-0.0028	-4.1	181778
Satsurblia	Kotias	HohleFels49	Mbuti	-0.0003	-0.3	43680	-0.0028	-1.1	9354
Satsurblia	Kotias	Villabruna	Mbuti	-0.0011	-2.0	801136	-0.0013	-2.2	307477
Satsurblia	Kotias	Bichon	Mbuti	-0.0002	-0.5	1407253	0.0000	0.0	759382
Satsurblia	Kotias	Ranchot88	Mbuti	-0.0017	-2.4	272886	-0.0016	-1.4	59040
Satsurblia	Kotias	Loschbour	Mbuti	-0.0007	-1.3	1407104	-0.0007	-1.4	763252
Satsurblia	Kotias	LaBrana1	Mbuti	-0.0014	-2.7	1257041	-0.0013	-2.3	659801
Satsurblia	Kotias	Hungarian.KO1	Mbuti	-0.0007	-1.2	957314	-0.0012	-2.0	511190
Satsurblia	Kotias	Motala12	Mbuti	-0.0016	-3.0	1260562	-0.0014	-2.6	679159
Satsurblia	Kotias	Karelia	Mbuti	-0.0006	-1.1	1186256	-0.0004	-0.8	627984
Satsurblia	Kotias	Malta1	Mbuti	-0.0002	-0.4	970714	-0.0002	-0.3	514145
Satsurblia	Kotias	AfontovaGora3	Mbuti	0.0003	0.3	190536	-0.0002	-0.1	42150
Satsurblia	Kotias	Stuttgart	Mbuti	-0.0005	-1.0	1397826	-0.0004	-0.8	759544
Satsurblia	Kotias	French	Mbuti	-0.0011	-2.8	1417761	-0.0011	-2.6	767709
Satsurblia	Kotias	Sardinian	Mbuti	-0.0010	-2.6	1417775	-0.0009	-2.3	767719
Satsurblia	Kotias	Han	Mbuti	-0.0003	-0.8	1417774	-0.0003	-0.7	767720
Satsurblia	Kotias	Dai	Mbuti	-0.0005	-1.4	1417780	-0.0006	-1.4	767723
Satsurblia	Kotias	Karitiana	Mbuti	-0.0006	-1.3	1417731	-0.0008	-1.7	767698
UstIshim	Kotias	Satsurblia	Mbuti	-0.0177	-24.0	1415190	-0.0162	-22.3	766712
Oase1	Kotias	Satsurblia	Mbuti	-0.0210	-23.0	189526	-0.0177	-17.0	82749
Kostenki14.sg	Kotias	Satsurblia	Mbuti	-0.0137	-19.9	1272240	-0.0122	-17.8	678547
Kostenki14	Kotias	Satsurblia	Mbuti	-0.0140	-20.2	1152869	-0.0125	-18.4	553455
GoyetQ116-1	Kotias	Satsurblia	Mbuti	-0.0152	-19.2	558196	-0.0152	-15.5	130649
Vestonice16	Kotias	Satsurblia	Mbuti	-0.0143	-19.6	625570	-0.0133	-16.4	217427
Ostuni1	Kotias	Satsurblia	Mbuti	-0.0140	-16.2	246277	-0.0128	-10.9	71079
KremsWA3	Kotias	Satsurblia	Mbuti	-0.0143	-15.1	157117	-0.0138	-9.0	35085
ElMiron	Kotias	Satsurblia	Mbuti	-0.0137	-18.1	529583	-0.0125	-14.8	181778
HohleFels49	Kotias	Satsurblia	Mbuti	-0.0143	-10.9	43680	-0.0115	-4.3	9354
Villabruna	Kotias	Satsurblia	Mbuti	-0.0116	-16.6	801136	-0.0107	-14.2	307477
Bichon	Kotias	Satsurblia	Mbuti	-0.0107	-15.8	1407253	-0.0094	-14.0	759382
Ranchot88	Kotias	Satsurblia	Mbuti	-0.0127	-14.9	272886	-0.0107	-8.3	59040
Loschbour	Kotias	Satsurblia	Mbuti	-0.0105	-16.1	1407104	-0.0096	-15.0	763252
LaBrana1	Kotias	Satsurblia	Mbuti	-0.0120	-17.9	1257041	-0.0107	-16.1	659801
Hungarian.KO1	Kotias	Satsurblia	Mbuti	-0.0107	-15.5	957314	-0.0100	-14.3	511190
Motala12	Kotias	Satsurblia	Mbuti	-0.0106	-16.4	1260562	-0.0096	-14.5	679159
Karelia	Kotias	Satsurblia	Mbuti	-0.0096	-14.0	1186256	-0.0086	-12.3	627984
Malta1	Kotias	Satsurblia	Mbuti	-0.0118	-16.6	970714	-0.0106	-14.6	514145

AfontovaGora3	Kotias	Satsurblia	Mbuti	-0.0103	-11.4	190536	-0.0093	-6.6	42150
Stuttgart	Kotias	Satsurblia	Mbuti	-0.0100	-14.9	1397826	-0.0091	-13.3	759544
French	Kotias	Satsurblia	Mbuti	-0.0093	-16.5	1417761	-0.0087	-15.3	767709
Sardinian	Kotias	Satsurblia	Mbuti	-0.0103	-18.3	1417775	-0.0096	-16.8	767719
Han	Kotias	Satsurblia	Mbuti	-0.0162	-26.6	1417774	-0.0147	-24.6	767720
Dai	Kotias	Satsurblia	Mbuti	-0.0164	-26.7	1417780	-0.0150	-24.8	767723
Karitiana	Kotias	Satsurblia	Mbuti	-0.0144	-22.5	1417731	-0.0134	-21.1	767698
UstIshim	Satsurblia	Kotias	Mbuti	-0.0179	-24.5	1415190	-0.0162	-22.3	766712
Oase1	Satsurblia	Kotias	Mbuti	-0.0203	-21.6	189526	-0.0168	-15.7	82749
Kostenki14.sg	Satsurblia	Kotias	Mbuti	-0.0133	-19.2	1272240	-0.0120	-17.3	678547
Kostenki14	Satsurblia	Kotias	Mbuti	-0.0134	-19.4	1152869	-0.0120	-17.4	553455
GoyetQ116-1	Satsurblia	Kotias	Mbuti	-0.0140	-17.3	558196	-0.0138	-13.7	130649
Vestonice16	Satsurblia	Kotias	Mbuti	-0.0138	-18.2	625570	-0.0131	-15.7	217427
Ostuni1	Satsurblia	Kotias	Mbuti	-0.0136	-16.1	246277	-0.0122	-10.2	71079
KremsWA3	Satsurblia	Kotias	Mbuti	-0.0133	-14.2	157117	-0.0124	-7.8	35085
ElMiron	Satsurblia	Kotias	Mbuti	-0.0118	-15.6	529583	-0.0097	-11.0	181778
HohleFels49	Satsurblia	Kotias	Mbuti	-0.0140	-10.3	43680	-0.0087	-3.2	9354
Villabruna	Satsurblia	Kotias	Mbuti	-0.0105	-14.9	801136	-0.0093	-12.4	307477
Bichon	Satsurblia	Kotias	Mbuti	-0.0104	-15.6	1407253	-0.0094	-14.4	759382
Ranchot88	Satsurblia	Kotias	Mbuti	-0.0110	-12.4	272886	-0.0090	-6.8	59040
Loschbour	Satsurblia	Kotias	Mbuti	-0.0098	-15.2	1407104	-0.0089	-14.1	763252
LaBrana1	Satsurblia	Kotias	Mbuti	-0.0106	-15.3	1257041	-0.0094	-13.6	659801
Hungarian.KO1	Satsurblia	Kotias	Mbuti	-0.0100	-14.6	957314	-0.0088	-12.4	511190
Motala12	Satsurblia	Kotias	Mbuti	-0.0090	-13.1	1260562	-0.0082	-12.1	679159
Karelia	Satsurblia	Kotias	Mbuti	-0.0090	-13.6	1186256	-0.0082	-12.0	627984
Malta1	Satsurblia	Kotias	Mbuti	-0.0115	-15.8	970714	-0.0104	-14.3	514145
AfontovaGora3	Satsurblia	Kotias	Mbuti	-0.0106	-11.5	190536	-0.0091	-6.1	42150
Stuttgart	Satsurblia	Kotias	Mbuti	-0.0095	-14.2	1397826	-0.0087	-13.0	759544
French	Satsurblia	Kotias	Mbuti	-0.0082	-14.1	1417761	-0.0077	-13.2	767709
Sardinian	Satsurblia	Kotias	Mbuti	-0.0093	-15.7	1417775	-0.0087	-14.6	767719
Han	Satsurblia	Kotias	Mbuti	-0.0159	-26.4	1417774	-0.0144	-24.0	767720
Dai	Satsurblia	Kotias	Mbuti	-0.0159	-26.5	1417780	-0.0144	-24.0	767723
Karitiana	Satsurblia	Kotias	Mbuti	-0.0139	-21.7	1417731	-0.0127	-19.9	767698

Population affinities of other ancient samples

We briefly discuss the population affiliations of the remaining samples that had at least 10,000 SNPs covered. When relevant, we use *D*-statistics of the form *D(X, Y; Test, Mbuti)* to explore how each *Test* sample in turn relates to other samples (*X, Y*).

- *Oase1:* This has been the subject of a published paper[3].

- *Kostenki14:* This sample has been the subject of a published paper[9]. We present a separate note analyzing this sample's affinities in Supplementary Information section 9.

- *GoyetQ116-1:* We present a separate note analyzing this sample's affinities in Supplementary Information section 10.

- *Cioclovina1*, *Kostenki12 and Muierii2:* As shown in Table S5.10, these samples all have evidence of shared genetic drift with present-day West Eurasians (unlike *Oase1*[3]), as documented by the statistic *D(UstIshim/Han, other pre-Neolithic European; Cioclovina1/Kostenki12/Muierii2, Mbuti)*, which give Z<-3 scores (Table S5.10). *Muierii2* and *Kostenki12* are possibly closer to *Kostenki14*, *Ostuni1* and *Vestonice16* than to other samples (Table S5.10). However, there is too little data to infer more refined relationships.

Table S5.10 Z-score of *D(X, Y; Cioclovina/Kostenki12/Muierii2 , Mbuti)*

D(X, Y; Muierii2, Mbuti) Muierii2: 98,618 SNPs												
X/Y	Han	Ust	Oase1	Kost14	Q116-1	Vestonice16	Ostuni1	Miron	Villabruna	Losch	Brana	Malta1
Han	NA	-0.5	0.1	-12.8	-9.3	-11.3	-8.2	-9.4	-11.1	-12.3	-10	-6.6
UstIshim	0.5	NA	0.8	-10.8	-7.9	-9.3	-8	-7.8	-8.7	-9.6	-8.3	-4.8
Oase1	-0.1	-0.8	NA	-6.3	-4.6	-6.4	-4.5	-4.6	-5.6	-5.1	-4.6	-3.6
Kostenki14	12.8	10.8	6.3	NA	2.3	0.5	0.8	2.5	2.1	2	3.3	5.5
GoyetQ116-1	9.3	7.9	4.6	-2.3	NA	-0.9	-0.9	0.6	-0.5	-0.6	0.4	2.5
Vestonice16	11.3	9.3	6.4	-0.5	0.9	NA	-0.1	2.4	1.1	1.5	2.5	3.9
Ostuni1	8.2	8	4.5	-0.8	0.9	0.1	NA	2.6	0.6	1.5	1.5	2.5
ElMiron	9.4	7.8	4.6	-2.5	-0.6	-2.4	-2.6	NA	-0.7	-0.5	0.7	1.6

Villabruna	11.1	8.7	5.6	-2.1	0.5	-1.1	-0.6	0.7	NA	0	1.7	3.3
Loschbour	12.3	9.6	5.1	-2	0.6	-1.5	-1.5	0.5	0	NA	1.6	3.4
LaBrana1	10	8.3	4.6	-3.3	-0.4	-2.5	-1.5	-0.7	-1.7	-1.6	NA	2
Hungarian.KO1	7.9	6	3.7	-4.4	-1.7	-3.6	-3.9	-2	-2.1	-2.2	-1.2	-0.4
Malta1	6.6	4.8	3.6	-5.5	-2.5	-3.9	-2.5	-1.6	-3.3	-3.4	-2	NA
D(X, Y; Cioclovina1, Mbuti) Cioclovina1: 12,784 SNPs												
X/Y	Han	Ust	Oase1	Kost14	Q116-1	Vestonice16	Ostuni1	Miron	Villabruna	Losch	Brana	Malta1
Han	NA	-0.3	0	-4	-4.1	-6.5	-5.2	-5.5	-6.1	-5.7	-4.6	-4
UstIshim	0.3	NA	-0.2	-3.6	-3.8	-5.6	-3.5	-4.4	-5.2	-4.7	-3.9	-3.3
Oase1	0	0.2	NA	-2.6	-1.8	-2.4	-1.8	-2.4	-1.5	-2.2	-1.7	-1.3
Kostenki14	4	3.6	2.6	NA	-0.5	-2	-0.5	-1.3	-1.4	-0.3	-0.5	-0.5
GoyetQ116-1	4.1	3.8	1.8	0.5	NA	-1.9	-1.7	-0.9	-0.9	-0.4	0.3	-0.4
Vestonice16	6.5	5.6	2.4	2	1.9	NA	0.4	0.8	0.9	2	1.8	0.9
Ostuni1	5.2	3.5	1.8	0.5	1.7	-0.4	NA	-0.1	0.4	1.6	1.9	0.9
ElMiron	5.5	4.4	2.4	1.3	0.9	-0.8	0.1	NA	0.9	1.6	0.9	1.6
Villabruna	6.1	5.2	1.5	1.4	0.9	-0.9	-0.4	-0.9	NA	1.6	1.5	0.8
Loschbour	5.7	4.7	2.2	0.3	0.4	-2	-1.6	-1.6	-1.6	NA	0.2	0
LaBrana1	4.6	3.9	1.7	0.5	-0.3	-1.8	-1.9	-0.9	-1.5	-0.2	NA	0
Hungarian.KO1	2.8	2.3	2.3	-0.6	-1.2	-3	-1.1	-1.6	-1.3	-0.7	-1	-1.4
Malta1	4	3.3	1.3	0.5	0.4	-0.9	-0.9	-1.6	-0.8	0	0	NA
D(X, Y; Kostenki12, Mbuti) Kostenki12: 61,228 SNPs												
X/Y	Han	Ust	Oase1	Kost14	Q116-1	Vestonice16	Ostuni1	Miron	Villabruna	Losch	Brana	Malta1
Han	NA	0	1.1	-11.1	-6.8	-10.3	-8.4	-8.2	-8	-9	-7.2	-4.6
UstIshim	0	NA	1.8	-10.2	-6.3	-9.7	-7.6	-7.2	-7.1	-7.8	-6.8	-4
Oase1	-1.1	-1.8	NA	-6.7	-3.1	-5.9	-4.9	-4.8	-4.8	-4.2	-3.3	-3.2
Kostenki14	11.1	10.2	6.7	NA	3.1	0.1	-0.1	1.9	2.3	3.3	3.5	4.6
GoyetQ116-1	6.8	6.3	3.1	-3.1	NA	-3.1	-2.9	-1.3	-0.8	-0.5	-0.3	1.5
Vestonice16	10.3	9.7	5.9	-0.1	3.1	NA	-0.4	2.2	2.5	2.9	3.3	4.7
Ostuni1	8.4	7.6	4.9	0.1	2.9	0.4	NA	1.8	1.8	2.2	3.3	3.3
ElMiron	8.2	7.2	4.8	-1.9	1.3	-2.2	-1.8	NA	0.8	1.2	1.4	2.1
Villabruna	8	7.1	4.8	-2.3	0.8	-2.5	-1.8	-0.8	NA	0.5	0.6	1.7
Loschbour	9	7.8	4.2	-3.3	0.5	-2.9	-2.2	-1.2	-0.5	NA	0.5	1.6
LaBrana1	7.2	6.8	3.3	-3.5	0.3	-3.3	-3.3	-1.4	-0.6	-0.5	NA	1.1
Hungarian.KO1	6.7	5.4	3.7	-3	0.7	-2.2	-2.1	-1	-0.3	-0.2	0	1.2
Malta1	4.6	4	3.2	-4.6	-1.5	-4.7	-3.3	-2.1	-1.7	-1.6	-1.1	NA

References

1 Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87-91, doi:10.1038/nature12736 (2014).

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

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

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

5 Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225-228, doi:10.1038/nature12960 (2014).

6 Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications* **5**, 5257, doi:10.1038/ncomms6257 (2014).

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

8 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications* **6**, 8912, doi:10.1038/ncomms9912 (2015).

9 Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, doi:10.1126/science.aaa0114 (2014).

Section 6

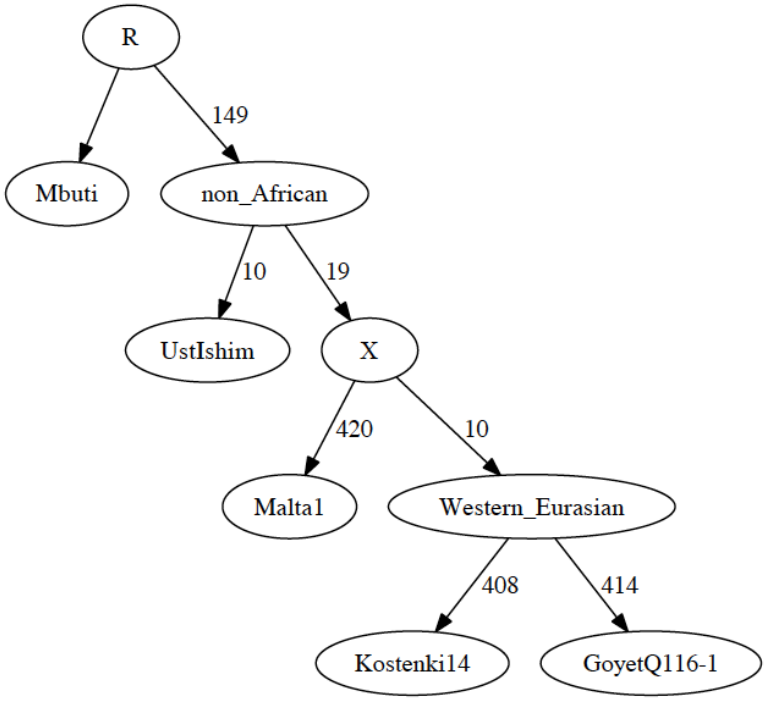
Admixture Graph Modeling of high coverage ancient genomes

Strategy

We began with the Admixture Graph (Figure S6.1) in Supplementary Information section 8, which relates the high coverage *Mbuti* (African outgroup), *UstIshim*, *Malta1*, *Kostenki14*, and *GoyetQ116-1*. This tree fits the allele frequency correlation patterns in the data to within the limits of our resolution, in the sense that the f_2-, f_3-, and f_4-statistics among all possible pairs, triples, and quadruples of populations match the observed values within three standard errors (using an empirical standard error computed with a Block Jackknife).

We proceeded by attempting to fit representative high coverage (>1×) samples—*Vestonice16*, *ElMiron*, *Villabruna*, *Loschbour*—into the Admixture Graph in turn, starting with the oldest and moving forward in time. We evaluated whether each tested model was a fit to the data, again by testing whether the predicted values of all the f_2-, f_3-, and f_4-statistics among all possible pairs, triples, and quadruples of populations matched the observed values, and assessing the significance of the difference using a Block Jackknife. In each of the Admixture Graphs shown in this note, the labels on the solid edges give the estimated genetic drift in f_2-units of squared frequency difference (parts per thousand). The labels on the dotted edges give mixture proportions.

Figure S6.1: Base model. This uses 324,336 SNPs covered in all populations.

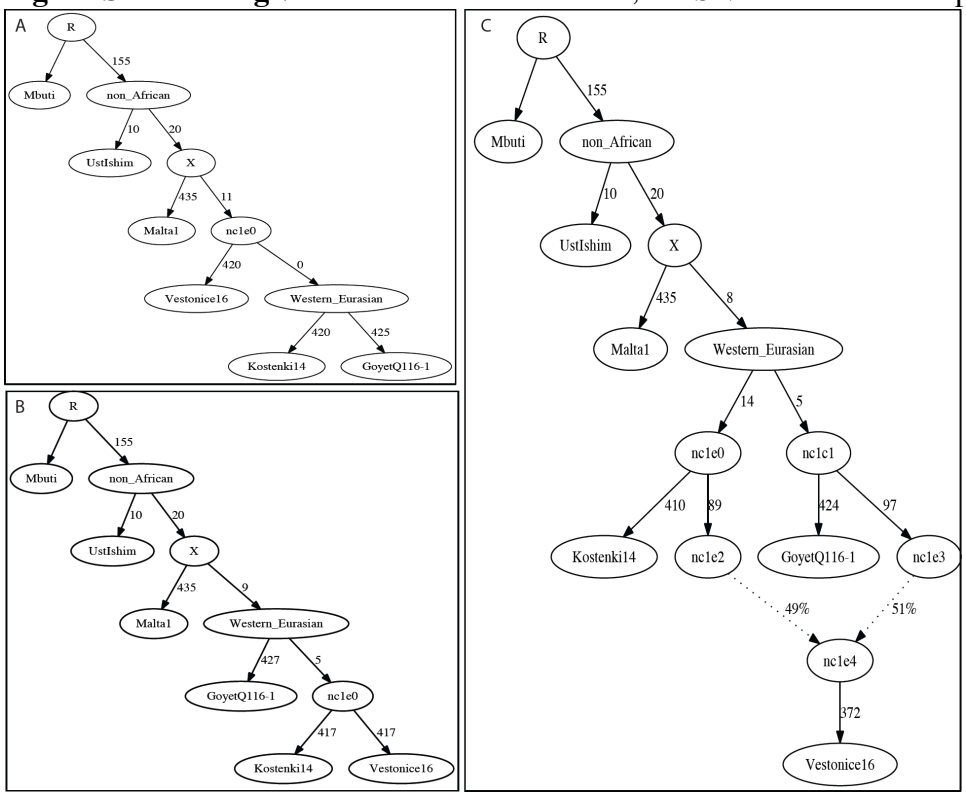


Vestonice16 (~31,000 BP)

We added *Vestonice16* to all possible nodes of Figure S6.1 either as a simple branch without mixture, or as a mixture between two branches. Altogether, we identified 3 models that fit the data (maximum |Z|<3). We show these in Figure S6.2:

- Figure S6.2A shows a fit in which *Vestonice16*, *Kostenki14* and *GoyetQ116-1* have an unresolved splitting order: a trifurcation.
- Figure S6.2B shows *Vestonice16* as a clade with *Kostenki14*, with *GoyetQ116-1* as an outgroup to both of them.
- Figure S6.2C shows *Vestonice16* mixed of lineages related to *Kostenki14* and *GoyetQ116-1*.

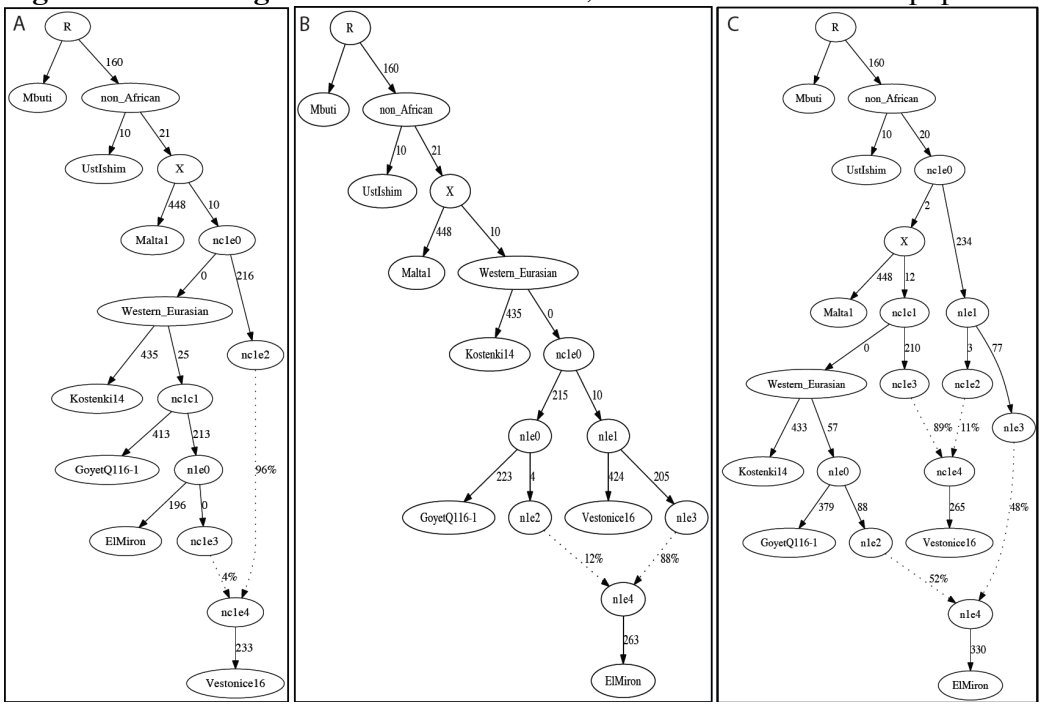
Figure S6.2 Adding *Vestonice16*. This uses 247,637 SNPs covered in all populations.



***ElMiron* (~19,000 BP)**

We added *ElMiron* to all possible nodes of the 3 models that fit the data for *Vestonice16* (Figure S6.2). We added it either as a simple branch without mixture, or as a mixture of two branches. Figure S6.3 shows the three models that fit. Figure S6.3A shows *ElMiron* as a clade with *GoyetQ116-1* and *Vestonice16* as mixed. Figure S6.3B shows *Vestonice16* as not mixed and *ElMiron* is mixed. Figure S6.3C shows both as mixed.

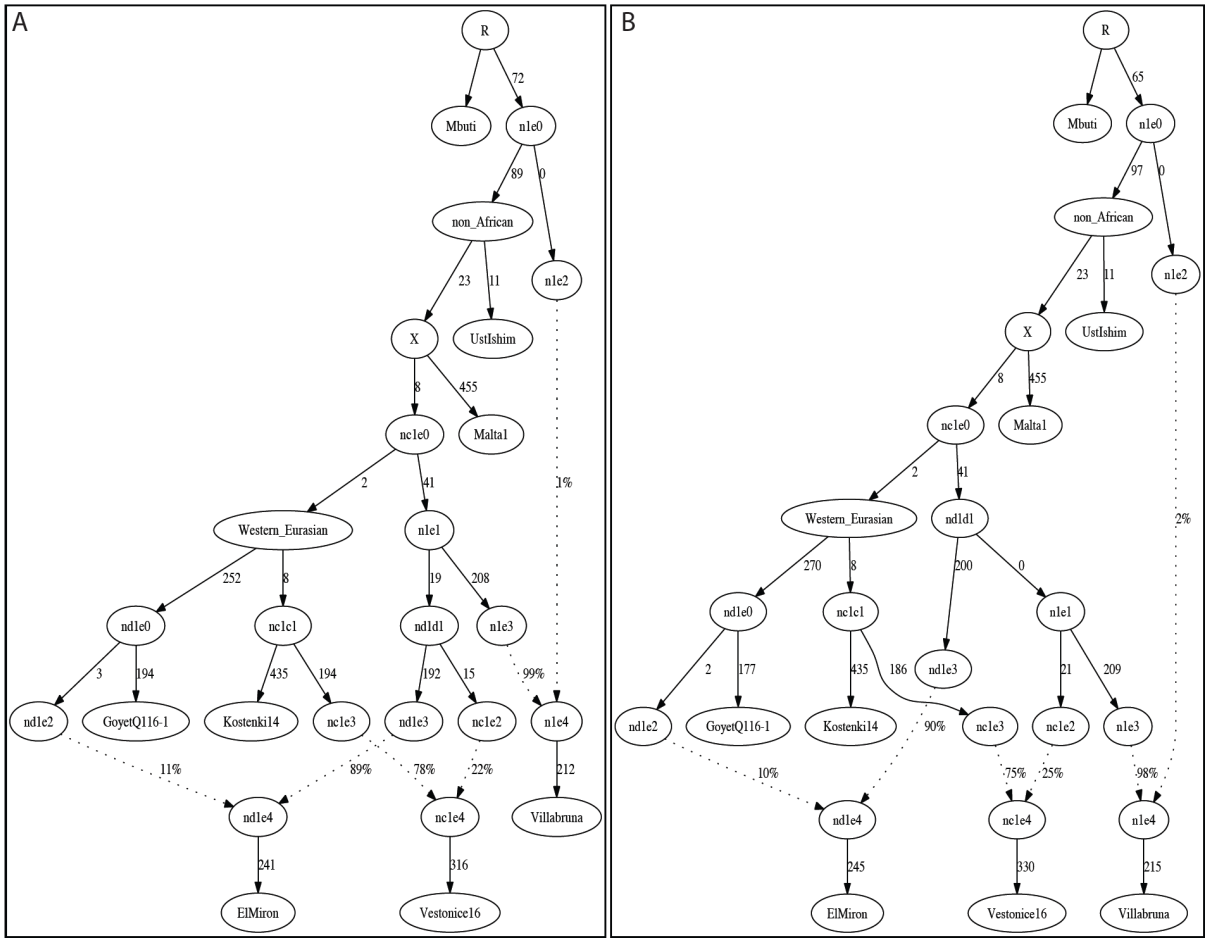
Figure S6.3 Adding *ElMiron*. This uses 186,469 SNPs covered in all populations.



***Villabruna* (~14,000 BP)**

We added *Villabruna* to all possible branches of the 3 models that fit the data for *Vestonice16* and *ElMiron* (Figure S6.3), again either as a simple branch or as a mixture between two branches. A total of nine models fit the data with at least three admixture events. We highlight two in Figure S6.4 (the other 7 models are qualitatively similar, with slight differences in the insertion points for the admixing lineages).

Figure S6.4 Adding *Villabruna*. This uses 181,563 SNPs covered in all populations.



The models that fit the data for *Villabruna* have the shared features that:

- *Vestonice16* is a mixture of lineages related to *Kostenki14* and *Villabruna*.
- *ElMiron* is a mixture of lineages related to *GoyetQ116-1* and a lineage that contributed most of the ancestry of *Villabruna*.

In both models of Figure S6.4, *Villabruna*, too, is inferred to be admixed, with 1-2% of its ancestry deriving from a deep Eurasian branch that split before the separation of *UstIshim* from all other Eurasians (drift distance ~0.09). This lineage is inferred to derive from earlier than the founder of all non-Africans, and is inferred to be more drifted than the "Basal Eurasian" lineage[1]. To investigate this unexpected signal, we removed the deep ancestry lineage, and found a single outlier: f_4*(Mbuti, UstIshim; GoyetQ116-1, Villabruna)*, which is significantly different from expectation (|Z|=3.1). When we compute the statistic *D(Mbuti, UstIshim; GoyetQ116-1, Villabruna)*, however, we obtain Z-scores of -2.2 (all SNPs) and -1.0 (transversions only). Given the number of statistics we are computing to test the fit of this model to the data, and the failure to replicate the signal for transversions only, we do not view a single outlier at |Z|>3 as a strong rejection of the model of no deep ancestry in

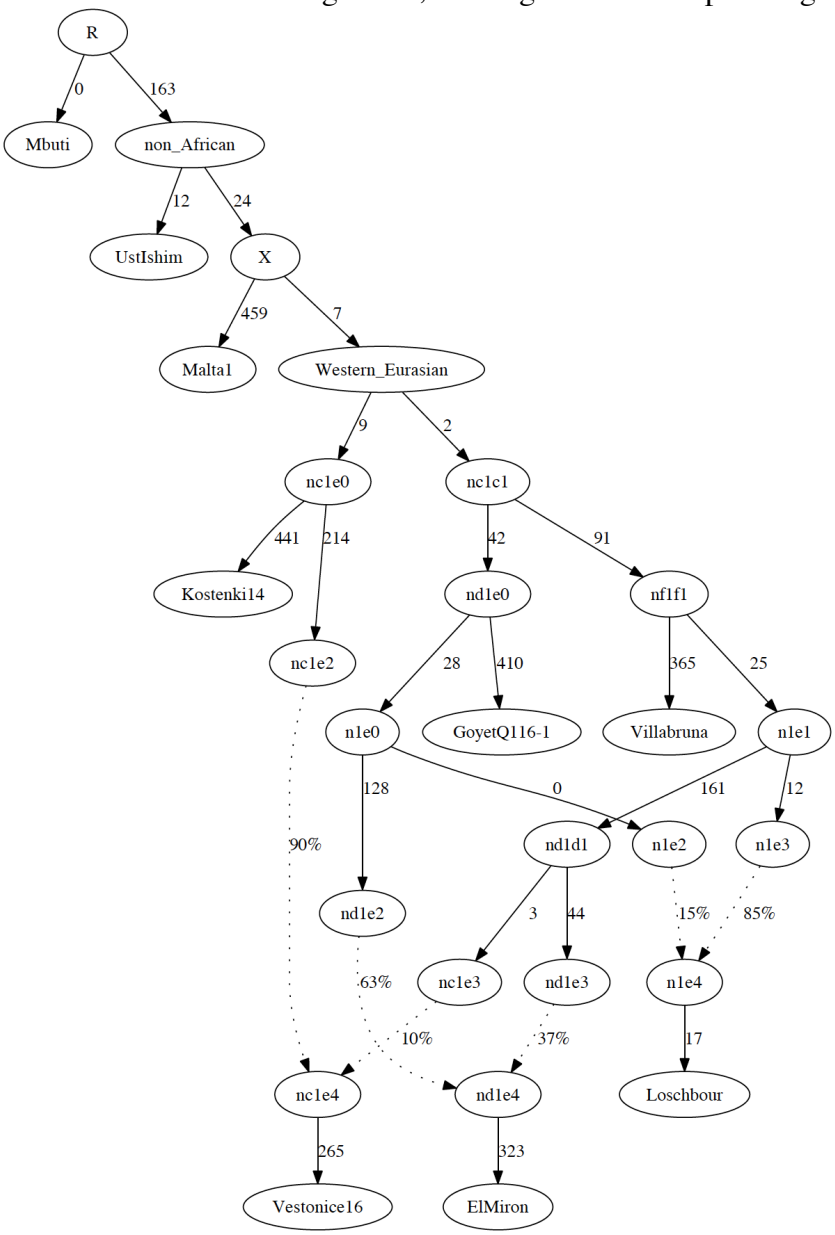
Villabruna. *Villabruna* is also artifact-prone in some ways because it is one of the few non-UDG-treated samples in the study. For subsequent analyses, we therefore do not model this deep ancestry signal in *Villabruna* (at the cost of tolerating slightly poorer fits).

***Loschbour* (~8,000 BP)**

We added *Loschbour* to all possible branches of the nine models that fit the data for *Vestonice16*, *ElMiron*, and *Villabruna*, removing in each case the deep branching lineage contributing *Villabruna*. A total of five models fit the data, after allowing for some maximum |Z|-scores slightly larger than 3 due to the deep ancestry signal in *Villabruna* (see above).

Figure S6.5 shows one passing model according to these criteria, which posits *Loschbour* as a mixture of two lineages whose closest relatives are *ElMiron* (more distantly *GoyetQ116-1*) and *Villabruna*. We use this model to explore the relationships of lower coverage samples in Supplementary Information section 7.

Figure S6.5 Adding *Loschbour*. This uses 179,232 SNPs covered in all populations. This is an alternate version of Figure 4a, which gives branch-specific genetic drift estimates.

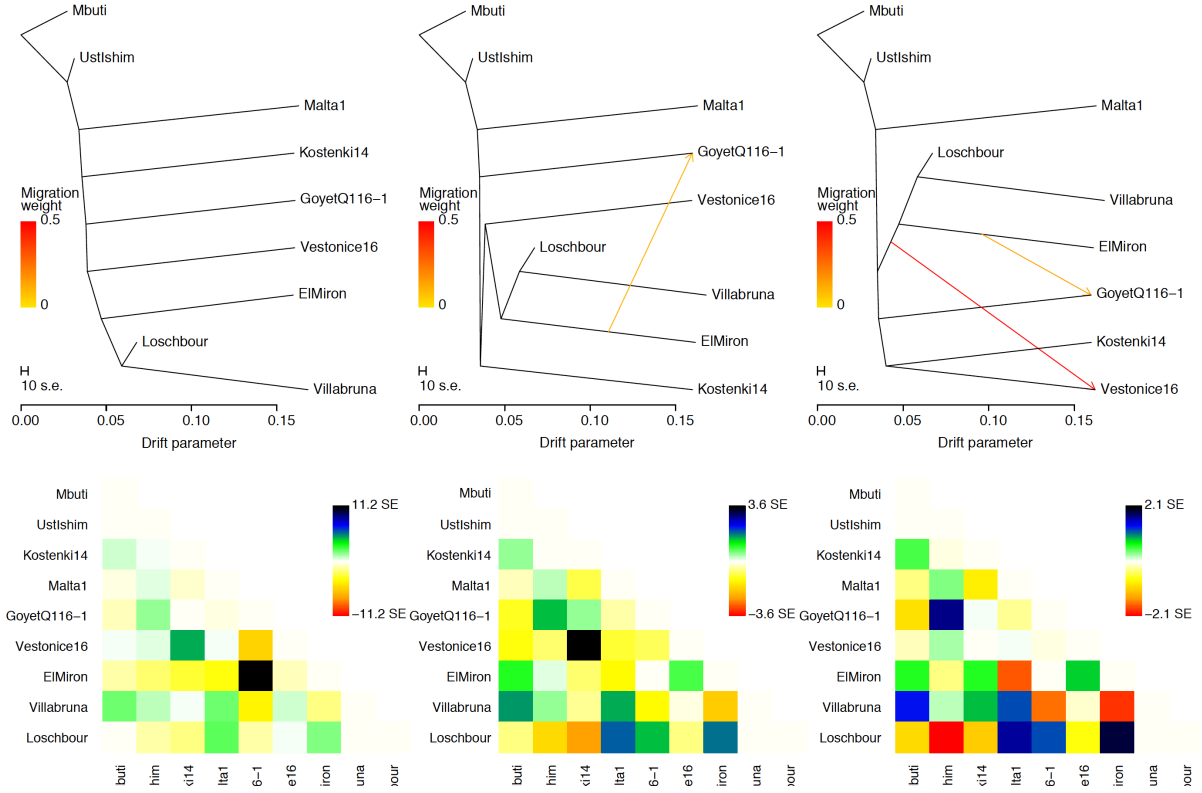


Treemix

As a complementary analysis to study the historical relationships among the samples, we used *TreeMix*[2]. Although *TreeMix* uses similar allele frequency correlation statistics as does *qpGraph*[3] to fit models to data, it takes a very different approach to exploring graph topologies. In *qpGraph*, we manually explored alternative topologies, adding in samples in a specified chronological order. In contrast, with *TreeMix*, samples are all fitted at once. We were curious to see if the two approaches gave similar solutions

We applied *TreeMix* to the same individuals shown as in Figure S6.5. The best fitting model without mixture is strongly inconsistent with the data, with some f_2-statistics among pairs of populations as high as |Z|=11.2 standard errors from expectation (Figure S6.6). When one mixture event is allowed, *Treemix* infers gene flow between the *ElMiron* and *GoyetQ116-1* lineages (Figure S6.6). With two mixture events, *TreeMix* infers gene flow between the *Vestonice16* lineage and *Kostenki14* (Figure S6.6). These two major admixture events are qualitatively similar to those that emerge from the *qpGraph* based manual analysis, although the directionalities of mixture are different.

Figure S6.6 *TreeMix* results for 0, 1 and 2 admixture events. The bottom plots shows residuals, expressed as the number of standard errors.



Summary and caveat

We have used the principle of parsimony to build these models. We make no claim that the models are exact. Indeed, we think that the models are likely to be wrong in some details. However, we hypothesize that the models may capture some important qualitative feature of the shared history of these samples:

- *Malta1* fits as an outgroup to *Vestonice16*, *Kostenki14*, *GoyetQ116-1*, and *ElMiron*, consistent with the results of Supplementary Information section 8.

- *Vestonice16* fits as a mixture of (primarily) *Kostenki14*-related ancestry, with a lesser amount of *Villabruna*-related ancestry. *GoyetQ116-1* is an outgroup.

- *ElMiron* fits as a mixture of (mostly) *GoyetQ116-1*-related ancestry, and a lesser amount of *Villabruna*-related ancestry.

- *Loschbour* fits as a mixture of (mostly) *Villabruna*-related ancestry, and a lesser amount of *GoyetQ116-1*-related ancestry.

In the rest of this study, we use the model of Figure S6.5 (which is the same as Figure 4a in the main text) as a null hypothesis that fits the allele frequency correlations. In Supplementary Information section 7, we employ this model to determine positions in the Admixture Graph for fitting lower coverage samples.

References

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

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

3 Patterson, N. J. *et al.* Ancient Admixture in Human History. *Genetics*, doi:10.1534/genetics.112.145037 (2012).

Section 7

Admixture Graph based assignment of ancestry

Overview

We used the Admixture Graph model of Figure 4a in the main text—which is a good fit to the allele frequency correlation patterns of eight ancient samples and a sub-Saharan African outgroup—as a framework to test the positioning of other samples.

Confirming the position in the Admixture Graph of samples with >0.1x coverage

For a subset of samples, we have high enough sequence coverage to be able to directly test whether they fit in the Admixture Graph of Figure 4a. For the samples that fit, we can also, in some instances, use the Admixture Graph to estimate mixture proportions.

Vestonice16, Ostuni1 and KremsWA3 are interchangeable in the Admixture Graph

In the Admixture Graph of Figure 4a, *Vestonice16* is modeled as a mixture of 90% ancestry from a lineage related to *Kostenki14* and 10% ancestry related to *Villabruna*. We were interested in whether other samples from the Věstonice Cluster (Supplementary Information section 5) fit in the same position in the Admixture Graph. We therefore replaced *Vestonice16* with *Ostuni1* and *KremsWA3*—the two other high coverage samples in the Věstonice Cluster—to test whether the Admixture Graph model fits them as well.

We began by replacing *Vestonice16* in this Admixture Graph with *Ostuni1*, and found that we obtained a reasonable fit. We observed a single outlier statistic that is |Z|=3.4 standard errors from expectation, but note that this is the same statistic discussed in Supplementary Information section 6: f_4*(Mbuti, UstIshim; GoyetQ116-1, Villabruna)*. This is not too concerning given the number of hypotheses tested, and the fact that the deep ancestry signal in Villabruna—one of the few non-UDG-treated samples in our study—becomes non-significant when restricting to transversions which are not affected by characteristic ancient DNA damage (Supplementary Information section 6).

We also replaced *Vestonice16* with *KremsWA3*. This resulted in two outliers, f_4*(Mbuti, Malta1; ElMiron, Loschbour)* (|Z|=3.2) and f_4*(UstIshim, Malta1; ElMiron, Loschbour)* (|Z|=3.1). We view this as a tolerable fit given the number of hypotheses tested. Table S7.1 summarizes the quality of each of the fits.

Table S7.1. Interchangeability of samples in the Věstonice cluster. We switch the position of *Vestonice16* with other samples and show parameters of the fits.

Sample	SNPs analysed	Coverage	Kostenki14 branch	Villabruna branch	#outliers |Z|>3	Maximum |Z|-score
Vestonice16	179,232	1.31	90%	10%	1	3.1
Ostuni1	87,879	0.25	83%	17%	1	3.4
KremsWA3	55,190	0.11	84%	16%	2	3.2

El Mirón Cluster samples are interchangeable in the Admixture Graph

In the Admixture Graph of Figure 4a, *ElMiron* is modeled as a mixture of 63% ancestry related to *GoyetQ116-1*, and 37% ancestry related (deeply) to *Villabruna*.

We replaced *ElMiron* with the pool of non-Iberian El Mirón Cluster samples (ElMiron_NI_C). This resulted in a fit, with a larger proportion of ancestry from the *GoyetQ116-1* lineage (Table S7.2). This finding is consistent with Supplementary

Information section 10, which shows formally that the proportion of *GoyetQ116-1* related ancestry is significantly higher in the ElMiron_N_CI than in *ElMiron*.

Table S7.2. Interchangeability of samples in the El Mirón Cluster. We switch the position of *ElMiron* with ElMiron_NI_C and show parameters of the fits.

Sample	SNPs analysed	Coverage	GoyetQ116-1 branch	Villabruna branch	# outliers |Z|>3	Maximum |Z|-score
ElMiron	179,232	1.01	63%	37%	1	3.1
ElMiron_NI_C	54,997	0.22	80%	20%	0	<3

Loschbour, LaBrana1, Bichon and Rochedane are interchangeable in the Admixture Graph
Loschbour fits in the Admixture Graph of Figure 4a as a mixture of 16% *GoyetQ116-1* and 84% *Villabruna*-related ancestry.

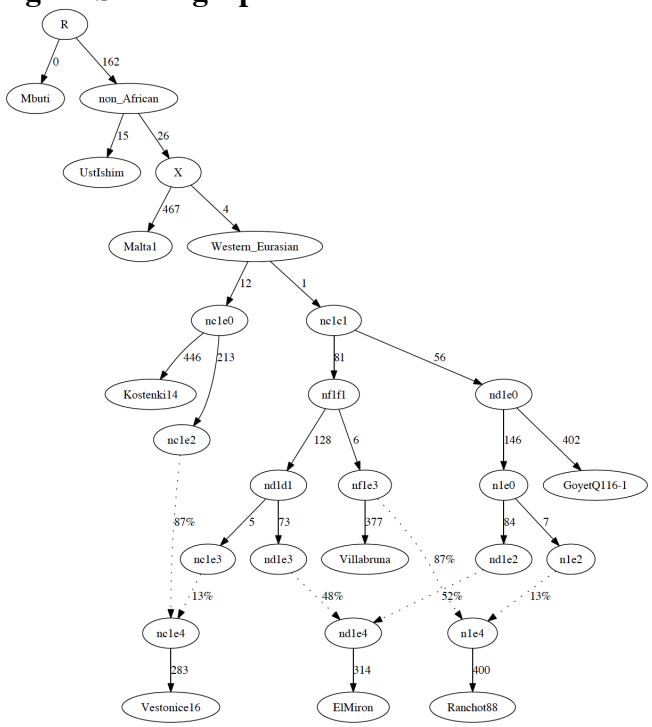
We replaced *Loschbour* with *LaBrana1*, and obtained a similar fit (Table S7.3). For *Rochedane* and *Bichon* we obtained a fit to the data without requiring any ancestry from the *GoyetQ116-1* branch, suggesting that *Rochedane* and *Bichon* may both be unadmixed descendants of the same population to which *Villabruna* belonged.

Table S7.3. Interchangeability of samples in the Villabruna Cluster. *LaBrana1, Bichon* and *Rochedane* can all be exchanged with *Loschbour* and produce a good fit.

Villabruna Cluster sample	SNPs analysed	Coverage	GoyetQ116-1 branch	Villabruna branch	# outliers |Z|>3	Maximum |Z|-score
Loschbour	179,232	20	16%	84%	1	3.1
LaBrana1	176,901	3.3	20%	80%	1	3.2
Bichon	181,406	8.1	0%	100%	2	3.1
Rochedane	50,560	0.13	0%	100%	0	<3

We finally replaced *Rochedane* with *Ranchot88*. The fit is worse in the same position in the graph (maximum |Z|-score = 4.1). However, if we change the topology so the *Villabruna*-related ancestry of *Ranchot88* is from a lineage that is a direct sister group to the *Villabruna* lineage—only a modest qualitative change in topology—we obtain a good fit (Figure S7.1).

Figure S7.1 A graph that fits *Ranchot88*. This uses 98,757 SNPs.

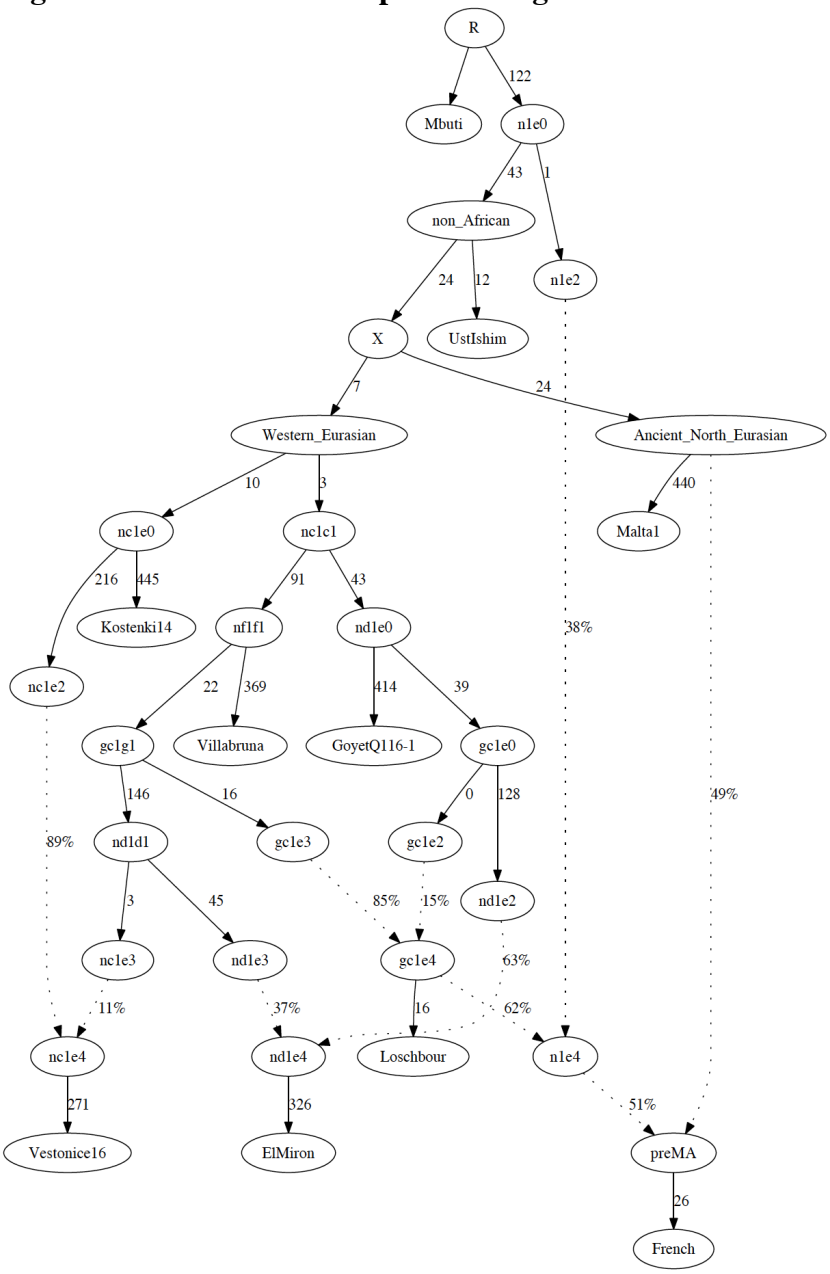


Position in the Admixture Graph of low coverage samples

For a large number of samples in this study, there are relatively few SNPs covered, reflecting the fact that we restricted to damaged sequences for our main analyses to reduce the danger of contamination. Combined with the fact that only 179,232 SNPs are covered in all nine samples included in the Admixture Graph of Figure 4a—less than 10% of the targeted SNP positions—this meant that we had very little information to use for placing low coverage samples onto the Admixture Graph.

To address the problem of limited data, we carried out Admixture Graph analysis of all samples that are not part of the skeleton admixture graph of Figure 4a, without restricting to damaged sequences. While this introduces contamination into a number of samples, we address this by assuming that the contamination is from a present-day European source, and then modeling this contamination into the Admixture Graph by using present-day *French* as a surrogate for the contamination (Figure S7.2).

Figure S7.2 Admixture Graph modeling of *French*. This uses 179,232 SNPs.



In this Admixture Graph, we model *French* as a mixture of three lineages inspired by the model of ref. [1]: (a) "Basal Eurasian" ancestry that diverged from the ancestors of *UstIshim* and *Loschbour* before they separated from each other; (b) ancestry from a sister group to *Malta1*, and (c) ancestry from a sister group to *Loschbour* (Figure S7.2). This Admixture Graph is a fit to the data in the sense that there are only two *f*-statistics for which the observed and predicted values are |Z|>3 standard errors from expectation: f_4*(Mbuti, Ust-Ishim; GoyetQ116-1, Villabruna)* (|Z|=3.1) and f_4*(UstIshim, French; GoyetQ116-1, Villabruna)* (|Z|=3.2). We make no claim that the modeling of the *French* in Figure S7.2 is correct. Indeed, it is almost certainly incorrect, as the ancestral relationships of present-day Europeans to pre-Neolithic Europeans are known to be complicated[1]. Our philosophy here is not to build an accurate model for *French*, but instead to model them into the graph as a "nuisance ancestry source" to adjust for biases from contamination. We model each ancient sample in turn as a mixture of a chosen point in the Admixture Graph and this nuisance ancestry source.

Table S7.4 gives the maximum absolute Z-score between the observed and predicted values for any f_2, f_3, or f_4-statisic relating samples in the graph, when we replace each of the non-*French* samples in the Figure S7.2, with a *Test* sample modeled as harboring contaminating ancestry from a lineage closely related to *French*. We make several observations:

- *Cioclovina1, Kostenki12* and *Muierii2* are closer to *Kostenki14* or *Vestonice16* than to other pre-Neolithic Europeans (especially the early branching *GoyetQ116-1*), consistent with the results in Supplementary Information section 5.

- *Vestonice13, Vestonice15, Vestonice14, Vestonice43, KremsWA3, Ostuni1, Paglicci133* and *Pavlov1* cluster with *Vestonice16*, consistent with the evidence from Supplementary Information section 5 that they are in the "Věstonice Cluster". They have a somewhat weaker affinity to *Kostenki14*, consistent with the Admixture Graph of Figure 4a, which has *Vestonice16* as primarily derived from a lineage related to *Kostenki14*. *Ostuni2* shows evidence of having a possibly closer affinity to *Kostenki14* than to *Vestonice16*, so we do not consider this sample to be confidently placed (Table S7.4).

- *GoyetQ376-19, GoyetQ53-1* and *GoyetQ56-16* are not possible to assign, since all three have many possible placements (Table S7.4).

- *Brillenhohle, Burkhardtshohle, GoyetQ-2, HohleFels79, HohleFels49,* and *Rigney1* are related to *ElMiron*, consistent with the finding of Supplementary Information section 5 that they are in the "El Mirón Cluster".

- *BerryAuBac, Bockstein, Continenza, Chaudardes1, Falkenstein, Hungarian.KO1, Iboussieres39, LaBrana1, LesCloseaux13, Ofnet, Ranchot88, Bichon* and *Rochedane* are closest to *Villabruna* and *Loschbour*, as expected from being in the "Villabruna Cluster". The present analysis gives us the resolution we need to conclude that *Iboussieres39*, too, is in the Villabruna Cluster.

Table S7.4. Fit of the Admixture Graph when each sample in the row is fit into the position of the sample in the column. For the samples in the rows (ordered chronologically), we use all sequences instead of restricting to damaged sequences. We model each sample as having a portion of their ancestry derived from a lineage that is a sister-group to the *French* to adjust for the contamination that we expect from undamaged sequences. Each cell reports the maximum absolute Z score for all possible *f*-statistics relating the samples in the Admixture Graph. We highlight in yellow all |Z|≤4 scores, except where there are no such |Z|-scores in which case we highlight the smallest in the row. "Min SNPs" refers to the minimum number of SNPs used in Admixture Graph fits of the sample on this row.

Sample ID	Min SNPs	UstIshim	Malta1	GoyetQ116-1	Kostenki14	Vestonice16	ElMiron	Villabruna	Loschbour
Cioclovina1	16642	5.3	3.9	5.0	3.7	4.3	5.5	5.5	5.5
Kostenki12	12996	7.2	4.8	5.2	<3	<3	7.3	7.2	7.2
Muierii2	133585	11.2	8.5	7.3	3.7	4.7	11.8	11.2	11.2
Vestonice13	150689	23.9	20.8	18.5	16.5	3.7	20.7	24.5	24.8
Vestonice15	70226	20.5	17.9	16.2	13.9	4.0	17.8	20.7	21.2
Vestonice14	31694	5.8	5.8	5.5	3.9	4.0	5.7	6.0	6.0
Vestonice43	150210	15.0	11.3	10.3	6.2	3.4	11.6	15.0	15.3
Pavlov1	105279	13.3	11.9	10.1	8.5	3.9	12.5	14.3	14.6
Paglicci133	15242	5.8	6.1	5.6	3.8	3.1	5.8	5.8	5.8
KremsWA3	49637	17.9	14.3	12.6	10.3	3.2	13.9	17.4	17.6
Ostuni2	31456	6.9	7.2	7.8	5.5	7.1	7.8	7.8	7.8
Ostuni1	169076	16.8	14.6	12.3	7.9	3.3	15.9	17.7	18.0
Paglicci108	3409	3.3	3.1	3.7	3.0	<3	3.4	3.5	3.8
GoyetQ53-1	15729	<3	<3	<3	<3	<3	<3	<3	<3
GoyetQ376-19	28263	4.7	4.8	4.0	4.4	4.1	3.9	4.8	4.8
GoyetQ56-16	14747	3.9	3.9	3.0	3.6	3.3	3.1	4.7	4.6
HohleFels79	54032	21.2	20.2	12.1	17.9	14.0	6.5	17.3	17.4
HohleFels49	69602	24.3	24.3	12.6	19.0	15.4	7.0	18.8	19.0
Rigney1	33743	19.3	19.3	10.1	15.9	13.2	3.0	14.7	14.7
GoyetQ2	74642	22.9	22.0	11.9	19.6	15.9	7.5	20.5	20.1
Brillenhohle	21200	12.4	12.4	6.9	11.0	8.2	3.5	10.9	10.9
Burkhardtshohle	36476	15.0	15.0	6.7	12.0	9.9	5.6	14.2	13.7
Bichon	179080	24.7	24.7	22.4	24.7	20.6	14.4	3.9	3.2
Rochedane	49861	20.2	20.2	16.4	20.2	15.9	11.1	3.2	<3
Continenza	26349	18.2	18.2	15.2	18.2	12.6	9.0	3.4	3.4
Iboussieres39	6487	6.9	6.9	5.6	7.1	5.4	<3	<3	<3
Ranchot88	145660	29.6	29.6	24.6	29.6	23.5	16.2	6.3	4.3
LesCloseaux13	7090	5.9	5.9	5.9	5.9	5.6	4.4	3.7	<3
Falkenstein	63695	23.9	23.9	18.5	23.9	18.5	12.0	5.4	3.6
Bockstein	43064	16.1	16.1	14.5	16.1	13.7	9.5	3.4	3.1
Ofnet	29705	8.4	8.3	8.7	8.4	8.7	5.0	3.8	4.3
CuiryLesChaudardes1	58788	21.3	21.3	20.8	21.3	17.4	12.9	4.6	3.2
LaBrana1	174613	21.3	21.3	18.0	21.3	17.6	9.8	7.4	7.0
Hungarian.KO1	124921	18.9	18.9	18.9	18.9	16.0	12.8	3.9	3.6
BerryAuBac	12882	13.2	13.2	10.5	13.2	11.4	7.5	<3	<3

References

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

Section 8

No evidence for Basal Eurasian ancestry in pre-Neolithic Europeans

Overview

Seguin-Orlando et al. analysed 2.8-fold coverage shotgun sequencing data from *Kostenki14*[1] and suggested that this individual had "Basal Eurasian" ancestry from the lineage that contributed also to Early European Farmers like *Stuttgart* from the *Linearbandkeramik* culture[2]. Here, we present compelling evidence against this hypothesis.

Replication of the statistics in Seguin-Orlando et al. in our data

The evidence for Seguin-Orlando et al.'s claim is summarized in Figure 2A of their study, in which they cite a "Middle East component for *Kostenki14* in clustering analysis" and two statistics of the form *D(Kostenki14, X; Y, Mbuti)*:

D(Kostenki14, North Eurasian Hunter Gatherer; East Asians, Mbuti) << 0
D(Kostenki14, Early European Farmers; East Asians, Mbuti) ~ 0

The first of these statistics was significantly negative when *X* = a pre-Neolithic North Eurasian (either *Loschbour* or *Malta1*), and *Y* = *Han*. This is inconsistent with the hypothesis that *Han* is an outgroup to a clade including *Kostenki14* and *X*. We replicate this statistic when *X* is any sample in the "Mal'ta Cluster" (Malta_C), or any sample in the "Villabruna Cluster," (*Villabruna, Rochedane, Loschbour, LaBrana1* and *Hungarian.KO1*) whether in capture data or shotgun data, and whether restricting analysis to transversion SNPs (Table S8.1) or analyse all sites (Table S8.2). We also replicated the second statistic.

Table S8.1 Z-score of *D(Kostenki14, X; Y, Mbuti) restricting to transversions*

X/Y	Karitiana	Han	Dai	Onge	UstIshim	Oase1	GoyetQ116-1	Vestonice_CE_C	Vestonice_I_C	ElMiron	ElMiron_NI_C	Villabruna	Loschbour	LaBrana1	Hungarian.KO1	Malta_C
Capture data																
UstIshim	2.2	-1	-0.9	-1.4	NA	1.8	8.2	11.5	11.1	10.2	7.6	11.1	10.6	10.4	10.9	7.3
Oase1	3.6	2.3	2.7	1.8	3.5	NA	6.7	7.3	7.1	7	5.3	8.5	10.6	8.8	8	6.4
GoyetQ116-1	0	-2	-2.1	-1.1	-1.3	0	NA	1.3	-1.5	-8.6	-11.5	-0.9	-3.6	-4.4	-1.5	-1.2
Vestonice_CE_C	-0.1	-0.2	0.7	0.8	0.4	0.3	-0.7	NA	-6.2	-4.2	-2.3	-4.8	-5	-2.8	-3.8	0.3
Vestonice_I_C	0.7	-0.9	-0.1	0.1	1.2	-0.8	-3.2	-5.8	NA	-4.7	-2	-5	-3.4	-2.3	-3.1	0.2
ElMiron	-0.8	-1.3	-0.6	-0.7	0.8	0	-7.6	-2.1	-3.7	NA	-18.4	-11.7	-14.3	-15.5	-9.1	-0.4
ElMiron_NI_C	-0.2	-1.4	-1.3	-0.3	0	0.2	-10.5	0.2	-1.4	-19.2	NA	-5.1	-10.2	-12.2	-5.5	-1.2
Villabruna	-1.2	-1.1	-0.4	0.4	0.4	-0.4	-0.6	-3.2	-3.6	-11.9	-5.3	NA	-27	-19.4	-22.6	-1.1
Rochedane	0.1	0.3	-0.4	0.2	1.6	-0.6	-1.8	-1.2	-1.4	-6.9	-5.8	-16	-18.3	-10.4	-10.7	-0.4
Ranchot88	-1.2	-1.7	-1.2	0.1	1	0.6	-1.8	-2.6	-4	-11.9	-7	-19.5	-24.3	-16.5	-14.8	-2.5
Loschbour	-3.2	-4.2	-3	-2.5	0.1	1.8	-2.1	-1.2	-1.3	-14	-10	-25.7	NA	-24.3	-23.4	-2.7
LaBrana1	-2.9	-3.6	-2.5	-2.1	0.5	1.5	-2.8	1	-1	-14.8	-10.8	-18.8	-23.3	NA	-19	-1.8
Hungarian.KO1	-3.3	-4	-2.9	-1.3	1.8	1.2	0.3	0	-0.6	-7.4	-4.3	-20.8	-22.6	-18.7	NA	-2.6
Malta_C	-12.3	-4.8	-5.5	-3.9	-0.3	0.1	2.3	4.7	4.3	1.9	2.2	2.8	0.3	0.2	-1.3	NA
Stuttgart	2.5	0.1	0.4	2.1	3.3	3.1	5.7	6.2	6.1	3.1	3.7	0	-1.9	-0.6	-3.6	3.7
Shotgun data																
UstIshim	2	-0.8	-0.9	-1.4	NA	1.4	8.5	11.6	11	8.9	8.3	9.7	10.4	10	11.2	6.8
Oase1	4.2	2.6	3	2.3	4.1	NA	6.6	8	6.5	7.1	5.6	7.6	11.2	8.7	8.9	6.9
GoyetQ116-1	0.4	-1.7	-1.5	-0.8	-0.8	-0.7	NA	1.1	-0.8	-9.2	-10.1	-1.5	-3.5	-4.1	-1	-0.9
Vestonice_CE_C	1.2	1.1	1.8	2.1	1.4	0.4	0.2	NA	-5.6	-3.5	-2.2	-4.2	-4.2	-2	-2.8	-0.1
Vestonice_I_C	1.1	-0.7	0.4	1.1	2	-1.2	-1.6	-5.5	NA	-4.9	-1.8	-5	-3.2	-2.7	-2.3	0.2
ElMiron	0.1	-1.1	-0.3	0	0.9	-0.4	-7	-1.7	-2.8	NA	-18.5	-11.7	-14.9	-15.5	-8.7	-0.8
ElMiron_NI_C	0.4	-0.6	0	-0.2	0.7	-1.5	-9.2	-0.8	-1.4	-19.5	NA	-5.4	-9.9	-10.9	-5.1	-0.5
Villabruna	-0.5	-0.5	0.2	1	0.9	-0.7	0.3	-3	-3	-12.1	-4.8	NA	-26.6	-19.2	-22	-1.3
Rochedane	0.2	0.3	0.2	0.5	2.2	-1.4	-0.1	-1	-1	-7.2	-4.6	-15.9	-17.3	-10.4	-9.8	0.4
Ranchot88	-0.5	-1.7	-0.9	-0.3	1.2	-0.1	-1.8	-2.6	-3.8	-12	-6.2	-20.6	-24.3	-16.2	-15.2	-1.9
Loschbour	-3.2	-4.3	-3.1	-2.2	-0.1	1	-1.8	-1	-1	-15	-9	-26.5	NA	-24.2	-23.5	-2.8
LaBrana1	-2.5	-3.2	-2.2	-1.7	0.4	1.2	-2.1	1.5	-0.8	-15.6	-10.4	-19.7	-23.5	NA	-18.6	-1.7
Hungarian.KO1	-3.2	-4.3	-3.2	-1.2	1.8	0	0.4	0.2	-0.2	-8.4	-3.5	-21.4	-22.8	-19	NA	-3.2
Malta_C	-11.2	-4.6	-5.1	-3.3	-0.4	-0.8	2.8	4.3	4.6	1	2.6	1.6	0.5	0.3	-1.2	NA
Stuttgart	2.5	0.1	0.4	2.1	2.9	2.7	6.1	6.3	6.8	1.9	4.4	-1	-2	-0.7	-3.4	3.5

Table S8.2 Z-score of *D(Kostenki14, X; Y, Mbuti)* for all SNPs

X/Y	Karitiana	Han	Dai	Onge	UstIshim	Oase1	Q116-1	Vestonice_CE_C	Vestonice_I_C	ElMiron	ElMiron_NI_C	Villabruna	Loschbour	LaBrana1	Hungarian.KO1	Malta_C
Capture data																
	2.5	-1.3	-1.4	-1.7	NA	1.1	9.1	16.3	13.4	11.2	9.9	11.6	11.5	11.5	11.8	7.9
Oase1	5.6	2.9	3.2	2.9	4.1	NA	10.5	12.1	10.6	10.6	9.3	11.5	13.1	11.5	10.2	9.6
GoyetQ116-1	0.3	-2.7	-2.8	-1.9	-1.7	-1.2	NA	0.8	-0.6	-11	-15.7	-1.8	-4.2	-5.7	-1.6	-1.3
Vestonice_CE_C	-0.5	-1.5	-0.8	0.5	0.9	0.6	-3	NA	-10.1	-5.7	-3.9	-6.9	-6.5	-5.8	-5.4	-0.8
Vestonice_I_C	0.4	-1.7	-0.6	-0.3	1.5	-0.8	-1.7	-8.6	NA	-4.7	-3.5	-5.7	-4.2	-3.8	-3.2	0.4
ElMiron	-1.4	-2.2	-1.8	-1.1	0.2	-0.4	-9.8	-2.5	-3.5	NA	-25.9	-14.4	-16.9	-19.4	-10.8	-0.6
ElMiron_NI_C	-0.7	-1.9	-1.5	-0.8	-0.2	0.5	-13.6	-0.1	-2.5	-26.8	NA	-7.6	-12.9	-16.9	-8.4	-1.3
Villabruna	-1.5	-1.3	-1.1	0.6	0.6	-0.6	-0.5	-3.2	-3.9	-13.5	-6.8	NA	-29	-21.8	-26.7	-1.8
Rochedane	-0.1	-1	-1.1	0.1	1.5	-0.2	-1.4	-0.2	-1.8	-12.5	-7.7	-23.2	-24.8	-18	-17.3	-0.9
Ranchot88	-1.4	-1.6	-1.3	-0.2	2.3	-0.4	-2.4	-1.8	-3.8	-15.4	-10.1	-25.8	-33	-23.9	-22.4	-1.6
Loschbour	-3.7	-4.3	-3.2	-2.5	0.2	1.1	-3.1	-2.5	-2.3	-16.4	-13.5	-29.5	NA	-27	-26.3	-3.1
LaBrana1	-3.3	-3.8	-2.9	-2.1	0.8	0.5	-3.7	0.2	-1.8	-18.6	-15.5	-21.4	-25.8	NA	-21	-2.2
Hungarian.KO1	-3.5	-4	-3	-1.1	2.3	0.4	1	0.3	-0.4	-9.1	-6.2	-24.2	-24.2	-19.8	NA	-3
Malta_C	-14.3	-5.3	-6.1	-3.9	-0.1	-0.7	2.6	6.8	5.1	2.5	2.5	1.9	0.7	0.6	-1.4	NA
Stuttgart	2.7	0.4	0.7	2.5	3.9	2.8	6.7	9.3	7.8	3.4	5	-0.7	-1.5	-0.2	-4	4.3
Shotgun data																
UstIshim	2.1	-1.4	-1.5	-1.8	NA	0.5	8.8	15.1	13.2	10.4	9.8	10.3	11	11	11.6	7.5
Oase1	5.8	3.1	3.4	3.1	4.6	NA	10.3	12.6	10.9	11.2	10.3	10.7	13.5	11.8	11.1	9.6
GoyetQ116-1	-0.1	-2.9	-3	-2	-1.9	-2.1	NA	-0.1	-0.6	-11.8	-15.5	-2.6	-4.7	-5.8	-2	-1.6
Vestonice_CE_C	0.6	-0.8	-0.4	0.8	1.4	-0.3	-2.8	NA	-10.2	-5.6	-3.6	-7	-6.1	-5.9	-4.8	-0.8
Vestonice_I_C	0.4	-1.7	-0.5	0.2	1.7	-1.6	-1.6	-8.7	NA	-5	-3.1	-6.1	-4.4	-4.1	-3.1	0.2
ElMiron	-0.9	-1.9	-1.5	-0.8	0.4	-0.8	-9.6	-2.9	-3.2	NA	-26.4	-14.9	-17.2	-19.8	-10.8	-0.7
ElMiron_NI_C	-0.3	-1.4	-0.7	-0.3	0.2	-0.2	-12.8	-0.4	-2.4	-26.9	NA	-7.9	-12.7	-16.5	-7.5	-1.4
Villabruna	-1.3	-1.3	-1	0.8	0.5	-1.6	-0.5	-3.7	-3.8	-13.9	-6.7	NA	-29.2	-21.9	-26.6	-1.9
Rochedane	-0.7	-1.5	-1.4	-0.3	1	-0.9	-0.5	-0.6	-1.9	-13.1	-7.5	-24.2	-24.8	-18.7	-17.6	-0.8
Ranchot88	-2.2	-2.2	-1.8	-0.4	1.9	-1.3	-2.5	-2.9	-3.6	-15.6	-9.7	-26.5	-33	-23.8	-22.7	-1.7
Loschbour	-3.9	-4.6	-3.4	-2.5	-0.2	0	-3.5	-3	-2.7	-17.3	-13.6	-30.6	NA	-26.9	-26.3	-3.2
LaBrana1	-3.4	-3.9	-3	-2.1	0.4	-0.2	-3.9	-0.5	-2.2	-19.2	-15.9	-22.4	-26.1	NA	-21	-2.4
Hungarian.KO1	-3.7	-4.4	-3.3	-1.2	2	-1	0.5	-0.4	-0.5	-10	-6.1	-24.6	-24.7	-20	NA	-3.5
Malta_C	-13.5	-5.1	-5.8	-3.6	-0.2	-1.5	2.5	5.6	5.2	2.1	2.4	0.9	0.7	0.5	-1.7	NA
Stuttgart	2.5	0.2	0.4	2.3	3.4	1.9	6.3	8.2	7.7	2.5	5.1	-1.7	-1.9	-0.5	-4.2	4

Kostenki14 patterns are general in Upper Palaeolithic Europeans

The present study allows computation of these statistics in a much larger series of Upper Palaeolithic European samples. This analysis reveals that the observed signals interpreted as Basal Eurasian ancestry are not unique to *Kostenki14*, and are also seen in later samples until around 14,000 years ago. Only after around 14,000 years ago (from Villabruna onwards) do samples with genetic affinities like those in the Loschbour cluster appear.

(1) When *X* = a European from 39,000-14,000 BP, the statistic *D(Kostenki14, X; Han/Dai/Karitiana, Mbuti)* is ~ 0 (Table S8.1 and Table S8.2).

(2) When *W* = a European from 39,000-14,000 BP including *Kostenki14*, *D(W, Early European Farmers; East Asians, Mbuti)* ~ 0 (Table S8.3).

Table S8.3 *D*(X, Stuttgart; Han, Mbuti) showing that East Asians do not always share more alleles with pre-Neolithic Europeans than with Neolithic Europeans

				Transversion sites only			All sites		
X	**Stuttgart**	**Han**	**Mbuti**	**D value**	**Z score**	**SNPs used**	**D value**	**Z score**	**SNPs used**
UstIshim	Stuttgart	Han	Mbuti	0.0005	1	1105740	0.0008	1.5	2019942
Oase1	Stuttgart	Han	Mbuti	-0.0015	-2.2	122034	-0.002	-3.2	274308
Kostenki14	Stuttgart	Han	Mbuti	0	0.1	805172	0.0002	0.4	1651428
Kostenki14.sg	Stuttgart	Han	Mbuti	0.0001	0.1	975742	0.0001	0.2	1811868
GoyetQ116-1	Stuttgart	Han	Mbuti	0.0018	2.8	189027	0.0018	3.3	797550
Vestonice_CE_	Stuttgart	Han	Mbuti	-0.0003	-0.6	108097	0.0005	1.1	321529
Vestonice_I_C	Stuttgart	Han	Mbuti	0.0011	1.8	118730	0.0009	1.8	412907
ElMiron	Stuttgart	Han	Mbuti	0.0008	1.5	268429	0.0013	2.7	764796
ElMiron _NI_C	Stuttgart	Han	Mbuti	0.0014	2.2	108687	0.0012	2.4	450103
Villabruna	Stuttgart	Han	Mbuti	0.0006	1.4	451676	0.0008	1.8	1151986
Rochedane	Stuttgart	Han	Mbuti	0.0007	0.8	50355	0.0013	2.4	225677
Ranchot88	Stuttgart	Han	Mbuti	0.0025	3.3	86407	0.0016	3	394773
Loschbour	Stuttgart	Han	Mbuti	0.0021	4.6	1100117	0.0024	5.2	2007432
LaBrana1	Stuttgart	Han	Mbuti	0.0016	3.7	954247	0.0019	4.5	1797590
Hungarian.KO1	Stuttgart	Han	Mbuti	0.002	4.6	728993	0.0021	4.9	1353157
Malta_C	Stuttgart	Han	Mbuti	0.0024	5.1	755470	0.0027	5.9	1455444

These results show that whatever history is driving the signals identified in ref. [1] is not unique to *Kostenki14*. Two scenarios can potentially explain these observations:

(1) There is Basal Eurasian ancestry in *Kostenki14* (consistent with ref . [1]) as well as in all Europeans prior to some Villabruna Cluster samples. After that, a population replacement occurred so that some European hunter-gatherers had little or no Basal Eurasian ancestry.

(2) There is no Basal Eurasian ancestry in *Kostenki14*, and instead, gene flow occurred between the ancestors of East Asians and Europeans after 14,000 years ago. (Independently, there would need to be gene flow between the ancestors of East Asians and *Malta1* to explain its affinities.) In this scenario, the observation that *D(Kostenki14, Early European Farmers; East Asians, Mbuti)* ~ 0, highlighted in ref. [1], does not reflect Basal Eurasian ancestry in *Kostenki14*, and instead a balancing of Basal Eurasian ancestry in Early European Farmers (which biases the statistic positively), and gene flow between the ancestors of Early European Farmers and East Asians (biases the statistic negatively).

The strongest evidence of Basal Eurasian ancestry comes from *UstIshim* and *Oase1*, and does not support Basal Eurasian ancestry in *Kostenki14*

The Basal Eurasian ancestry in *Kostenki14* was initially proposed when only one reference Eurasian lineage (*Malta1*) that was not West Eurasian in affinity was available. However, data from two additional Eurasian lineages distinct from West Eurasians and East Asians have since become available: *Oase1*[3] and *UstIshim*[4] (Table S8.4). This makes it possible to distinguish scenarios (1) and (2) outlined in the previous section.

The initial statistic that motivated the Basal Eurasian hypothesis[2] was of the form: *D(North Eurasian hunter-gatherer, Early European Farmer; East Asian, Outgroup)* >> 0. A key observation of Seguin-Orlando et al.[1] was that while this statistic was greater than 0 when the North Eurasian hunter-gatherer was a Mesolithic European or *Malta1*[2], it gave a qualitatively different result (consistent with 0) when the North Eurasian hunter-gatherer was *Kostenki14* (see also Table S8.1, Table S8.2). When *Oase1* or *UstIshim* is used in the statistic in place of East Asians, however, the statistic is greater than 0 through the whole period; we no longer see a qualitative change comparing early to late pre-Neolithic Europeans (Table S8.4).

Table S8.4 *D(European hunter-gatherers, Stuttgart; X, Mbuti).*

pre-Neolithic European	X = Han				X = Oase1				X = Ust-Ishim			
	Transversions		All sites		Transversions		All sites		Transversions		All sites	
	D value	Z score	D value	Z score	D value	Z score	D value	Z score	D value	Z score	D value	Z score
Kostenki14	0	0.1	0.0002	0.4	0.0024	3.1	0.0019	2.8	0.0019	3.3	0.0025	3.9
Kostenki14.sg	0.0001	0.1	0.0001	0.2	0.002	2.7	0.0013	1.9	0.0018	2.9	0.0021	3.4
GoyetQ116-1	0.0018	2.8	0.0018	3.3	0.0052	4.1	0.0034	4.2	0.0035	4.6	0.004	5.9
Vestonice_CE_C	-0.0003	-0.6	0.0005	1.1	0.0022	1.6	0.0022	2.5	0.001	1.4	0.0017	3.1
Vestonice_I_C	0.0011	1.8	0.0009	1.8	0.0046	3.7	0.003	3.5	0.0023	3.2	0.0019	3.1
ElMiron	0.0008	1.5	0.0013	2.7	0.0035	3.7	0.0026	3.6	0.0016	2.5	0.0025	4.3
ElMiron _NI_C	0.0014	2.2	0.0012	2.4	0.0053	4	0.0019	2.4	0.0034	4.5	0.0027	4.8
Villabruna	0.0006	1.4	0.0008	1.8	0.0031	4	0.0025	3.9	0.0018	3.3	0.0022	3.9
Rochedane	0.0007	0.8	0.0013	2.4	0.0054	2.5	0.0037	3.4	0.0022	2.2	0.0023	3.5
Ranchot88	0.0025	3.3	0.0016	3	0.005	3.1	0.003	3.4	0.0025	2.8	0.0014	2.3
Loschbour	0.0021	4.6	0.0024	5.2	0.0014	2.1	0.0015	2.4	0.0018	3.3	0.0023	4.2
LaBrana1	0.0016	3.7	0.0019	4.5	0.0013	1.8	0.0016	2.4	0.0015	3	0.0019	3.5
Hungarian.KO1	0.002	4.6	0.0021	4.9	0.002	2.4	0.0017	2.4	0.0005	0.9	0.0009	1.5
Malta_C	0.0024	5.1	0.0027	5.9	0.0028	3.7	0.0025	3.9	0.0021	3.8	0.0024	4.4

A second statistic informative about Basal Eurasian ancestry—which was not possible to compute at the time when the Basal Eurasian hypothesis was initially proposed[2]—also takes advantage of *Oase1* or of *UstIshim*, and is of the form:

$D(Test_1, Test_2; Oase1/UstIshim, Mbuti) \sim 0$

As shown in Table S8.5, this statistic is consistent with 0 whenever we use a pair of pre-Neolithic samples as the *Test* samples, thus providing no evidence of Basal Eurasian ancestry in these samples. In other words, there is no evidence in pre-Neolithic samples for any ancestry from a lineage that diverged before the separation of *Ust'-Ishim*. In contrast, when *Stuttgart* is used in place of one of the pre-Neolithic samples, we observe attraction of it to *Mbuti*, confirming the distinct signal of Basal Eurasian ancestry in this sample (Table S8.5).

Table S8.5 Z-score of *D(Eurasian$_1$, Eurasian$_2$, Oase1/UstIshim/Han, Mbuti).* This table is for all sites. We add *Stuttgart* to show the Basal Eurasian signal.

X/Y	Kostenki14	GoyetQ116-1	Vestonice_CE_C	Vestonice_I_C	ElMiron	ElMiron _NI_C	Villabruna	Loschbour	LaBrana1	Hungarian.KO1	Malta_C	Stuttgart
Oase1												
Kostenki14	NA	-1.2	0.6	-0.8	-0.4	0.5	-0.6	1.1	0.5	0.4	-0.7	2.8
GoyetQ116-1	1.2	NA	0.5	0.8	0.8	1.5	1	1.7	1.4	1.4	0.7	4.2
Vestonice_CE_C	-0.6	-0.5	NA	-1	0.1	0.9	-0.8	0.9	0.4	1	0.2	2.5
Vestonice_I_C	0.8	-0.8	1	NA	1.4	1.2	0.4	1.7	0.5	1.2	-0.4	3.5
ElMiron	0.4	-0.8	-0.1	-1.4	NA	-0.3	-0.1	1.5	1.6	1.1	0.5	3.6
ElMiron _NI_C	-0.5	-1.5	-0.9	-1.2	0.3	NA	-0.9	0.4	0.3	-0.3	-0.7	2.4
Villabruna	0.6	-1	0.8	-0.4	0.1	0.9	NA	1.7	1.3	0.5	0.1	3.9
Loschbour	-1.1	-1.7	-0.9	-1.7	-1.5	-0.4	-1.7	NA	-0.2	-0.8	-2	2.4
LaBrana1	-0.5	-1.4	-0.4	-0.5	-1.6	-0.3	-1.3	0.2	NA	-0.6	-1.1	2.4
Hungarian.KO1	-0.4	-1.4	-1	-1.2	-1.1	0.3	-0.5	0.8	0.6	NA	-0.8	2.4
Malta_C	0.7	-0.7	-0.2	0.4	-0.5	0.7	-0.1	2	1.1	0.8	NA	3.9
Stuttgart	-2.8	-4.2	-2.5	-3.5	-3.6	-2.4	-3.9	-2.4	-2.4	-2.4	-3.9	NA
Ust'-Ishim												
Kostenki14	NA	-1.7	0.9	1.5	0.2	-0.2	0.6	0.2	0.8	2.3	-0.1	3.9
GoyetQ116-1	1.7	NA	3.9	3.2	2.1	2	2.2	2.3	3.1	3.8	1.8	5.9
Vestonice_CE_C	-0.9	-3.9	NA	-0.6	-2.5	-1.2	-1.2	-1.1	0.3	0.9	-1.5	3.1
Vestonice_I_C	-1.5	-3.2	0.6	NA	-1.1	-1.9	-1	-1	-0.6	0.8	-1.6	3.1
ElMiron	-0.2	-2.1	2.5	1.1	NA	0	0.7	0.4	1.3	2.2	-0.3	4.3
ElMiron _NI_C	0.2	-2	1.2	1.9	0	NA	1	0.7	1.6	2.7	0.5	4.8
Villabruna	-0.6	-2.2	1.2	1	-0.7	-1	NA	-0.5	0.6	2.2	-0.5	3.9
Loschbour	-0.2	-2.3	1.1	1	-0.4	-0.7	0.5	NA	0.9	3	-0.3	4.2
LaBrana1	-0.8	-3.1	-0.3	0.6	-1.3	-1.6	-0.6	-0.9	NA	2.1	-0.8	3.5
Hungarian.KO1	-2.3	-3.8	-0.9	-0.8	-2.2	-2.7	-2.2	-3	-2.1	NA	-3.1	1.5
Malta_C	0.1	-1.8	1.5	1.6	0.3	-0.5	0.5	0.3	0.8	3.1	NA	4.4
Stuttgart	-3.9	-5.9	-3.1	-3.1	-4.3	-4.8	-3.9	-4.2	-3.5	-1.5	-4.4	NA
Han												
Kostenki14	NA	-2.7	-1.5	-1.7	-2.2	-1.9	-1.3	-4.3	-3.8	-4	-5.3	0.4
GoyetQ116-1	2.7	NA	2	1.1	0.8	0.9	1.5	-1.1	-0.5	-1.2	-2.6	3.3
Vestonice_CE_C	1.5	-2	NA	-1.8	-2.4	-1.1	-0.4	-3.8	-3	-2.6	-4.8	1.1
Vestonice_I_C	1.7	-1.1	1.8	NA	-0.8	-0.4	0.8	-2.5	-2.5	-2.3	-4.1	1.8
ElMiron	2.2	-0.8	2.4	0.8	NA	0.3	1.5	-2	-1.3	-1.8	-3.3	2.7
ElMiron _NI_C	1.9	-0.9	1.1	0.4	-0.3	NA	1	-2.7	-1.5	-1.9	-3.1	2.4
Villabruna	1.3	-1.5	0.4	-0.8	-1.5	-1	NA	-3.9	-2.9	-3.3	-4	1.8
Loschbour	4.3	1.1	3.8	2.5	2	2.7	3.9	NA	1	0.5	-0.8	5.2
LaBrana1	3.8	0.5	3	2.5	1.3	1.5	2.9	-1	NA	-0.3	-1.8	4.5
Hungarian.KO1	4	1.2	2.6	2.3	1.8	1.9	3.3	-0.5	0.3	NA	-1.3	4.9
Malta_C	5.3	2.6	4.8	4.1	3.3	3.1	4	0.8	1.8	1.3	NA	5.9
Stuttgart	-0.4	-3.3	-1.1	-1.8	-2.7	-2.4	-1.8	-5.2	-4.5	-4.9	-5.9	NA

Gene flow among eastern non-Africans, the Mal'ta Cluster & pre-Neolithic Europeans

What is driving the statistics reported in Seguin-Orlando et al.[1]? Table S8.5 shows that the two statistics highlighted in that study—*D(Kostenki14, Loschbour/LaBrana1/Malta1; Y, Mbuti)* << 0 and *D(Kostenki14, Early European Farmer; Y, Mbuti)* ~ 0—are only observed when *Y* = Eastern non-Africans. When *Y* = *Oase1* or *UstIshim*, the signals are, as expected from Basal Eurasian ancestry, only seen in Early European Farmers.

The only way to explain these patterns is a history of gene flow between the ancestors of eastern non-Africans on the one hand, and the ancestors of three groups:

(a) A subset of Villabruna Cluster samples
(b) Early European farmers
(c) Mal'ta Cluster.

Such gene flow would induce a negative bias in two key statistics highlighted in Seguin-Orlando et al.[1] Note that for this to explain the data, three separate gene flow events are not required. Supplementary Information section 11 and Figure 4b document a link between (a) and (b), so as few as two gene flow events may be needed. Understanding the exact gene flow history responsible for these patterns is difficult with the ancient DNA sample series available here, but is an important question to address in future work.

References

1 Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, doi:10.1126/science.aaa0114 (2014).

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

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

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

Section 9

Malta1 is an outgroup to Upper Palaeolithic Europeans after 37,000 years ago

Malta1 is consistent with being an outgroup to Upper Palaeolithic Europeans

We studied the relationship of the ~24,000-year-old *Malta1* individual to pre-Neolithic mainland Europeans. Table S9.1 shows statistics of the form *D(X, Y; Malta1, Mbuti)*. For all samples younger than *Oase1*, we observe that $D(Test_1, Test_2; Malta1, Mbuti) \sim 0$. This is consistent with *Malta1* being an outgroup to pre-Neolithic mainland Europeans from the date of *Kostenki14* onward. When *X* is represented by *UstIshim, Oase1* or *Han*, we observe a significantly negative value, indicating that *Malta1* is not an outgroup to a clade including them and later Europeans (red).

Table S9.1 Z-score of *D(X, Y; Malta1, Mbuti)*

X/Y	Date	Han	UstIshim	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	HohleFels49	Villabruna	Rochedane	Ranchot88	Loschbour	LaBrana1	Hungarian.KO1
Han		NA	3.1	6.3	-5.7	-7.2	-7.6	-5.9	-6.2	-2.8	-7.4	-7.2	-7.6	-9	-8.1	-9.1
UstIshim	48370-43070	-3.1	NA	3.5	-7.6	-8.7	-8.7	-6.6	-7.4	-3.6	-8.5	-8.3	-8.5	-10	-9.4	-10
Oase1	41761-37615	-6.3	-3.5	NA	-9.4	-10.2	-9.7	-6.1	-9.3	-2	-9.5	-8.5	-7.2	-11.9	-9.5	-11.3
Kostenki14	38650-37845	5.7	7.6	9.4	NA	-1.3	-1.5	-0.1	-0.1	-0.5	-1.1	-0.5	-0.9	-2.7	-1.7	-2.6
GoyetQ116-1	35160-34430	7.2	8.7	10.2	1.3	NA	-0.2	0.6	1.6	1.3	0.3	0.2	-0.4	-0.5	0.3	-0.8
Vestonice16	~31155	7.6	8.7	9.7	1.5	0.2	NA	1.2	1.4	1.5	0.7	0	-0.6	-1	0.2	-0.9
Ostuni1	27730-27530	5.9	6.6	6.1	0.1	-0.6	-1.2	NA	0.6	0.7	-0.6	-1.4	-1.1	-2.1	-1.4	-2.4
ElMiron	18735-18600	6.2	7.4	9.3	0.1	-1.6	-1.4	-0.6	NA	1.4	-0.4	-1.4	-0.7	-2.5	-1.3	-2.5
HohleFels49	16250-15568	2.8	3.6	2	0.5	-1.3	-1.5	-0.7	-1.4	NA	0	0.1	-0.1	-2	-0.7	-1.7
Villabruna	14075-13905	7.4	8.5	9.5	1.1	-0.3	-0.7	0.6	0.4	0	NA	-0.6	-0.4	-2	-0.4	-2.2
Rochedane	13090-12830	7.2	8.3	8.5	0.5	-0.2	0	1.4	1.4	-0.1	0.6	NA	0.3	-0.7	-0.2	-0.8
Ranchot88	10235-9933	7.6	8.5	7.2	0.9	0.4	0.6	1.1	0.7	0.1	0.4	-0.3	NA	-1.1	0.5	-1.6
Loschbour	8170-7940	9	10	11.9	2.7	0.5	1	2.1	2.5	2	2	0.7	1.1	NA	1.4	0
LaBrana1	7940-7690	8.1	9.4	9.5	1.7	-0.3	-0.2	1.4	1.3	0.7	0.4	0.2	-0.5	-1.4	NA	-0.9
Hungarian.KO1	7731-7596	9.1	10	11.3	2.6	0.8	0.9	2.4	2.5	1.7	2.2	0.8	1.6	0	0.9	NA

Contradiction of the *Malta1* ancestry model proposed in Seguin-Orlando et al.[1]

Statistics of the form *D(Malta1, X; Y, Mbuti)* are negative (usually highly significantly) if *(X, Y)* is any pair of pre-Neolithic mainland Europeans from *Kostenki14* until *Loschbour* and *LaBrana1* (Table S9.2). This implies that all pre-Neolithic Europeans all the way back to the time of *Kostenki14* share alleles with each other that are not shared with *Malta1*, as predicted by the hypothesis above that *Malta1* is an outgroup to Upper Palaeolithic Europeans.

These results directly contradict the findings of Seguin-Orlando et al.[1] who suggested a model in which *Kostenki14* is an outgroup to a clade that includes (*Malta1*, *Loschbour* and *LaBrana1*) (Figure 2 of that study). If that is true, then statistics of the form *D(Malta1, Loschbour/LaBrana1; Kostenki14, Mbuti)* are predicted to be consistent with 0. However, these statistics are significantly negative (Z=-3.1 for *Loschbour*, and Z=-2.2 for *LaBrana1*). In other words, there are more shared alleles between *Kostenki14* and Mesolithic Europeans, than between *Kostenki14* and *Malta1,* contradicting the model of Seguin-Orlando et al.[1]

No ancestral connection between *Malta1* and samples from the Věstonice Cluster

We were struck by the fact that the most significantly negative *D(Malta1, X; Kostenki14, Mbuti)* values were observed when *X* is any sample from the "Vestonice Cluster": *Vestonice16* gives Z=-6.1 and *Ostuni1* gives Z=-3.9 (Table S9.2). This is opposite to the positive bias in Z-scores that would be expected from shared ancestry between *Malta1* and

samples in the Věstonice Cluster. Shared ancestry between Malta1 and the Věstonice Cluster would be predicted if the assignment of both to the Gravettian culture was due to movements of people (the Mal'ta site yielded Venus figurines stylistically similar to Gravettian sites such as Dolní Věstonice in Europe, thousands of kilometers to the west). Thus, if the cultural similarity that led to the assignment of *Malta1* to the Gravettian culture is not a coincidence, it is likely to reflect communication of ideas rather than movements of people.

Table S9.2 Z-score of *D(Malta 1, X; Y, Mbuti)*

X/Y	Han	UstIshim	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	HohleFels49	Villabruna	Rochedane	Ranchot88	Loschbour	LaBrana1	KO1
Han	NA	0.5	0.5	10.9	9.4	11.8	9	9.1	4.4	10.8	10.5	10.4	9.9	9.4	10.6
UstIshim	3.6	NA	1.7	7.7	6.9	8.4	8	7.8	4.1	9.6	9.2	11.4	10.8	10.5	13.4
Oase1	7.1	4.9	NA	9.3	8.2	10.4	6.4	8.9	2.7	9.1	7.3	7.5	13.8	10.1	11.8
Kostenki14	5.3	0.4	0.3	NA	-1.9	-4.6	-3.9	-2.1	-1.3	-1.8	-0.7	0	-0.4	-0.4	1.6
GoyetQ116-1	2.6	-1.3	-1.6	-3.2	NA	-2.6	-3.3	-11.9	-8.1	-2.2	-3.3	-3.9	-4.9	-6.1	0
Vestonice16	5	0.2	0.5	-6.1	-3	NA	-13.8	-7.5	-1.9	-6.6	-4.3	-5.2	-5.6	-4.1	-3.3
Ostuni1	3.2	1.5	0	-3.9	-2.5	-11.9	NA	-7	-2.4	-6.8	-3	-5.5	-4.7	-4.7	-2
ElMiron	3.4	0.8	-0.8	-2.4	-10.6	-6.3	-6.6	NA	-14	-15	-13.5	-16.1	-17.1	-20.1	-8.6
HohleFels49	2	0.4	0.6	-1.7	-6.7	-0.5	-1.7	-12.5	NA	-3.7	-3.9	-4.8	-6	-6.3	-2
Villabruna	4.2	1	-0.2	-2.9	-1.9	-5.6	-7.3	-15	-3.6	NA	-23.6	-25.3	-29.4	-22.1	-23
Rochedane	3.4	0.7	-1	-1.3	-3.2	-4.2	-4.3	-15.1	-3.7	-24.6	NA	-21	-24.8	-17.8	-15.1
Ranchot88	3.5	2.8	-0.2	-1	-4.1	-5.4	-6.3	-15.7	-4.6	-25.3	-20.9	NA	-31.6	-23.2	-19.1
Loschbour	0.9	0.6	1.8	-3.1	-5.1	-6.1	-6.7	-18.5	-7.4	-31.1	-25.2	-31.5	NA	-27.2	-25.5
LaBrana1	1.9	1.1	0.8	-2.2	-5.5	-3.8	-6.3	-19.9	-6.9	-21.6	-17.9	-21.8	-27	NA	-19.4
Hungarian.KO1	1.4	3.3	0.4	-1	-0.8	-4	-4.6	-11	-3.5	-23.8	-15.5	-20.7	-25.9	-21.3	NA

New model: An Admixture Graph that jointly fits *Malta1*, *Kostenki14* and *GoyetQ116-1*

Figure S9.1 shows that a model with no admixture events is consistent with the joint data for *Mbuti, UstIshim, Malta1, Kostenki14* and *Goyet116-1*, in the sense that there are no significant deviations between observed and predicted *f*-statistics (maximum |Z|=2.5, not significant correcting for multiple hypothesis testing). No other simple model fits the data

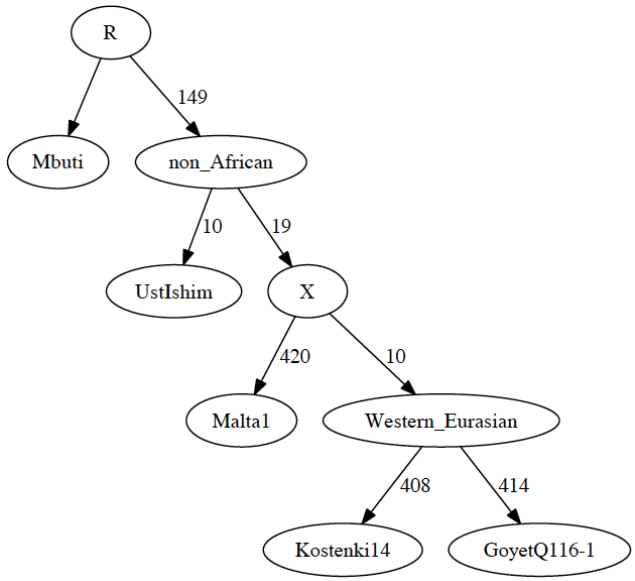


Figure S9.1 A parsimonious phylogeny with no admixture events. This is the only model that can jointly fit *Malta1, Kostenki14, UstIshim* and *GoyetQ116-1* without invoking any admixture events. (This is used as the base model for Supplementary Information section 6, and is reproduced there as Figure S6.1.)

References

1 Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, doi:10.1126/science.aaa0114 (2014).

Section 10

A genetic link between *GoyetQ116-1* and the El Mirón Cluster

Outgroup f_3-statistics show affinity of *GoyetQ116-1* to the El Mirón Cluster

In the matrix of pairwise $f_3(X, Y; Mbuti)$ (Figure 3A and Extended Data Figure 2), *GoyetQ116-1*, which dates to around 35,000 BP, shares substantial genetic drift with all the El Mirón Custer samples *ElMiron, HohleFels79, HohleFels49, Rigney1, GoyetQ-2* and *Burkhardtshohle*, which date to the period 19,000-14,000 BP. No other Upper Palaeolithic Eurasian—including *Oase1, Kostenki14, Malta1*, and multiple samples in the Věstonice Cluster—shows this affinity to the El Mirón Cluster.

D-statistics suggest a genetic link between *GoyetQ116-1* and the El Mirón Cluster

We used *D*-statistics of the form *D(X, Y; GoyetQ116-1, Mbuti)* to test formally if samples from the El Mirón Cluster are consistent with sharing more alleles with *GoyetQ116-1* as suggested by outgroup f_3 analysis (Table S10.1). We find that Z-scores of the statistic *D(El Mirón Cluster sample, other pre-Neolithic Europeans; GoyetQ116-1, Mbuti)* are all significant positive ($19.1 \geq Z \geq 3.3$), confirming the genetic affinity. The only exception to this pattern is *HohleFels79*, which is limited in its dataset size (11,211 SNPs).

Table S10.1. Z-score of *D(X, Y; Q116-1, Mbuti) for all sites*

D(X, Y; GoyetQ116-1, Mbuti): 846,983 SNPs X/Y	UstIshim	Oase1	Kostenki14	Vestonice16	Ostuni1	Villabruna	Loschbour	LaBrana	KO1	ElMiron
UstIshim	NA	3.6	-9.1	-9.7	-9.4	-9.2	-12.5	-12.9	-8.4	-18.1
Oase1	-3.6	NA	-10.5	-11.5	-9.2	-10.1	-13.2	-13.5	-9.6	-17.2
Kostenki14	9.1	10.5	NA	-1.2	-1.8	-0.5	-3.1	-3.7	1	-9.8
Vestonice13	8.7	7.6	2.5	1.6	-0.4	2.2	-0.2	-0.7	3.1	-5
Vestonice15	5.5	5.2	1.4	-0.5	-0.3	1.5	-0.6	0.5	1.5	-3.5
Vestonice43	7.9	7.2	1.7	0.9	-0.2	0.8	-1.6	-1.2	2	-6.8
Vestonice16	9.7	11.5	1.2	NA	-0.7	0.8	-2	-2.3	2.1	-8.7
Ostuni1	9.4	9.2	1.8	0.7	NA	1.4	-1.2	-1.9	1.8	-7.1
ElMiron	18.1	17.2	9.8	8.7	7.1	11.1	8.4	7.8	10.5	NA
HohleFels79	6.6	2.2	2.6	3.9	1.1	4.4	2.5	3.5	3.1	1.2
HohleFels49	12.3	9.1	6.9	7	5.3	8.9	6.7	6.6	6.8	1.7
Rigney1	10	7.3	5.9	5.6	3.4	4.5	4.2	3.3	4.7	0.6
GoyetQ-2	16.5	8.9	10.5	8.2	6.5	10.7	9.7	8.1	9.3	2.4
Burkhardtshohle	12.7	8.1	8.7	6.8	5.3	7	6.7	5.2	6.9	1.4
HohleFels79_nq	17.5	13.2	10.2	9	7.9	11.7	9.9	9.1	11.7	4.6
HohleFels49_nq	17.8	14.7	10.7	9.2	9	12.1	10.3	9.7	11.7	3.9
Rigney1_nq	15.3	11.9	9.7	8.2	5.8	8.4	7.6	6.2	8.2	1.3
GoyetQ-2_nq	21.9	18.9	14.6	13.2	11.4	14.8	13.9	12.7	15.3	4.6
Villabruna	9.2	10.1	0.5	-0.8	-1.4	NA	-3.7	-4.1	0.6	-11.1
Rochedane	9.9	8.4	1.4	0.5	-0.1	1.7	-1.6	-1.5	2.2	-8.4

To increase the power of this analysis in the face of limited data for some El Mirón Cluster individuals, we repeated the analysis using a version of the dataset that does not restrict to damaged sequences for *HohleFels49, HohleFels79, Rigney1*, or *GoyetQ-2* (suffix "_nq"). The signal becomes even stronger ($21.9 \geq Z \geq 6.2$) for each of these four individuals (Table S10.1). Contamination by present-day Europeans is expected to make this signal weaker, suggesting that the strengthening of the signal is not due to contamination.

To confirm that these patterns are not artifacts of ancient DNA damage, we restricted to transversion SNPs, while at the same time retaining as many SNPs as possible for analysis by grouping *HohleFels79, HohleFels49, Rigney1, GoyetQ-2, Burkhardtshohle* and *Brillenhohle* into an "ElMiron_NI_C" cluster (Supplementary Information section 5) (Table S10.2). Z-scores of the form *D(ElMiron_NI_C, other hunter-gatherers; GoyetQ116-1, Mbuti)* remain significantly positive ($16.5 \geq Z \geq 3.3$).

Table S10.2. Z-score of *D(X, Y; Q116-1, Mbuti) restricting to transversion SNPs*

X/Y	UstIshim	Oase1	Kostenki14	Vestonice_CE_C	Vestonice_I_C	ElMiron_NI_C	Villabruna	Loschbour	LaBrana	KO1	ElMiron
UstIshim	NA	1.2	-8.2	-6.4	-8.5	-17.6	-8.7	-10.8	-11.1	-7.6	-16.1
Oase1	-1.2	NA	-6.7	-4.9	-5.9	-11	-5.6	-8.1	-7.5	-5.4	-10.5
Kostenki14	8.2	6.7	NA	-0.7	-3.2	-10.5	-0.6	-2.1	-2.8	0.3	-7.6
Vestonice_CE_C	6.4	4.9	0.7	NA	-2	-8.2	0.1	-2.6	-1.8	0.5	-6.6
Vestonice_I_C	8.5	5.9	3.2	2	NA	-7.4	1.5	0.6	-0.1	2.3	-4.8
ElMiron	16.1	10.5	7.6	6.6	4.8	-4.8	7.7	6.9	5.6	7	NA
ElMiron_NI_C	17.6	11	10.5	8.2	7.4	NA	10.6	10.8	9.4	9.3	4.8
Villabruna	8.7	5.6	0.6	-0.1	-1.5	-10.6	NA	-2.4	-2.9	-0.3	-7.7
Rochedane	6.4	5.4	1.8	1.7	-1	-5	0.8	-0.1	-1.2	0.4	-4

A striking observation is that *GoyetQ116-1* shares more alleles with the sample pool ElMiron_NI_C than with *ElMiron*, the oldest sample in the El Mirón Cluster (last column of Table S10.1 and of Table S10.2). These results suggest that there was a gradient of *GoyetQ116-1* relatedness within the El Mirón Cluster.

Estimating mixture without a full phylogenetic model

In order to estimate the proportion of mixture from two potential reference populations in a target individual, we use the *qpAdm* method[1]. This method leverages the fact that if a *Test* population is a mixture of *N* different ancestral populations that are clades with reference populations from which we have genetic data, we can write:

$$f_4(Test, O_i; O_j, O_k) \approx \sum_{i=1}^{N} \alpha_i f_4(Ref_i, O_i; O_j, O_k)$$

Here, O_i, O_j and O_k are *Outgroup* populations. For the method to work, each of the *N* strands of ancestry in the *Test* population needs to be phylogenetically more closely related to the ancestry in one of the *Reference* populations than it is to any of the ancestry in any of the *Outgroup* populations.

We can formally test the fit of this model to the data—taking into account the covariance of these f_4-statistics—and thus produce a single P-value for fit using a Hotelling T^2 test. We can then estimate mixture proportions based on the fitted weights of the α_i coefficients.

ElMiron

We used *qpAdm*[1] to estimate the proportion of mixture in the target individual (*ElMiron*), given data from reference populations that we propose to be sister groups or descendent groups to the two mixing populations.

Briefly, we divide the populations into two sets. The *left* set *L* consists of the proposed admixed population (*ElMiron*) and *Reference* populations (*GoyetQ116-1, Villabruna,*

Ranchot88, LaBrana1, Loschbour). The *right* set *R* consists of 6 worldwide populations excluding Europeans (*Ulchi, Mbuti, Ami, Tubalar, Kinh, Onge*); these are *Outgroups* that provide leverage for discerning different components of ancestry among the *Left*. The *Outgroups* are chosen to represent a range of non-European ancestries, thus maximizing leverage for teasing apart lineages within Europe.

We used *qpAdm* to estimate the rank of the matrix of F_4-statistics $M(l, r) = F_4(l_x, l; r_x, r)$ where l and r are the left and right population sets. This matrix [2] has rank N-1 if there are N waves of migration. *qpAdm* also outputs a mixture proportion.

Rank 1 is not excluded (p=0.9) with the best model being a mixture of lineages related (perhaps anciently) to *GoyetQ116-1* and *LaBrana1*. *ElMiron* can be modeled as having 49±13% *GoyetQ116-1* and 51±13% *LaBrana1* ancestry. (In Supplementary Information section 6 using full Admixture Graph modeling, we obtain a point estimate of 63% *GoyetQ116-1* related ancestry, consistent with this model.) When we add a present-day *Iraqi_Jew* into *Outgroups* to represent a Near Eastern population, the standard error narrows to 10%. We note that *LaBrana1* and *Iraqi_Jew* post-date *ElMiron*, but this is not a problem for the method. What is important for *qpAdm* to be effective is that these populations have the phylogenetic positioning relative to *ElMiron* that is specified by the *qpAdm* model, which can be true even if they are relatively recent members of these populations.

ElMiron_NI_C
We tested *GoyetQ116-1, Villabruna, Loschbour, LaBrana1,* and *Ranchot88* as *Left* populations and *Ulchi, Mbuti, Ami, Tubalar, Kinh and Onge* as *Right* populations. The method infers 69±11% GoyetQ116-1 and 32±11% *LaBrana1*-related ancestry. (In Supplementary Information section 6 using Admixture Graph modeling, we obtain a consistent point estimate of 80% *GoyetQ116-1* related ancestry.) When we add *Iraqi_Jew* into the *Right* populations, the standard errors narrow to 7%.

Conclusion
Some populations that lived ~35,000 years ago in northwest Europe (i.e. *GoyetQ116-1*) have an unambiguous link to populations of the El Mirón Cluster that lived around 19,000-14,000 BP. Of course, the link might not involve direct descent from a population that lived near Goyet Cave. Instead, it may be mediated by populations related to *GoyetQ116-1* that lived elsewhere.

References

1 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature,* doi:doi:10.1038/nature14317 (2015).
2 Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370-374, doi:10.1038/nature11258 (2012).

Section 11

Gene flow linking the Villabruna Cluster and the Near East

We investigated the relationship between pre-Neolithic Europeans and present-day as well as ancient populations using statistics of the form:

$$D(European_1, European_2; Test, Mbuti)$$

Here, the *Test* populations are Native Americans, East Asians, Oceanians and Near Eastern populations from the Simons Genome Diversity Project (SGDP) panel.

Affinities of pre-Neolithic Europeans to the Near East

When neither of the two pre-Neolithic Europeans analysed in the statistic is in the Villabruna Cluster—that is, both are older than about 14,000 BP—they tend to be symmetrically related to populations outside Europe including present-day and ancient Near Easterners. However, when one lived prior to the Villabruna Cluster (e.g. *Vestonice16, ElMiron, Kostenki14, KremsWA3*, and *GoyetQ116-1*) and the other is in the Villabruna Cluster (e.g. *BerryAuBac, Bichon, CuiryLesChaudardes1, Falkenstein, Hungarian.KO1, LaBrana1, Loschbour, Ranchot88, Rochedane* and *Villabruna*), there is a distinct attraction of the Villabruna Cluster samples to Near Eastern populations (Figure 4b; Extended Data Figure 3). Table S11.1 shows the statistics when the Near Eastern population is *Iraqi_Jew*.

There are several possible explanations for these findings. One is gene flow between relatives of Near Easterners and pre-Neolithic Europeans after ~14,000 years ago, beginning with the Villabruna Cluster. A second is population substructure in Europe. In this scenario, after post-glacial re-peopling of Europe, the balance of ancestry could have shifted toward populations that were more closely related to Near Easterners. In either case, however, major population turnovers must have occurred.

The affinity of pre-Neolithic Europeans to Near Easterners beginning around 14,000 years ago is distinct from the affinity to East Asians in Mesolithic Europeans

Seguin-Orlando et al.[1] documented that statistics of the form *D(Kostenki14; Mesolithic Europeans; East Asians, Outgroup)* are significantly less than 0. In Supplementary Information section 8, we show that this is likely due to gene flow between the ancestors of East Asians and the ancestors of Mesolithic Europeans. These patterns are evident in Figure 4b and Extended Data Figure 3, which show that a subset of Villabruna Cluster samples including Mesolithic Europeans show a significant affinity to East Asians. This pattern does not go hand-in-hand with the affinity to the Near East that is present in all Villabruna Cluster samples, and thus the two signals must therefore reflect at least two distinct historical events.

Table S11.1 ***D(European$_1$, European$_2$; Iraqi_Jew, Mbuti).*** Europeans after around 14,000 years ago in the Villabruna Cluster began to have a significant affinity to Near Easterners.

	European$_1$=Kostenki14			European$_1$=GoyetQ116-1			European$_1$=Vestonice16			European$_1$=ElMiron		
European$_2$	**D**	**Z**	**SNPs**	**D**	**Z**	**SNPs**	**D**	**Z**	**SNPs**	**D**	**Z**	**SNPs**
GoyetQ116-1	-0.0001	-0.2	778,865	n/a	n/a	n/a	0.0003	0.8	879,269	0.0014	2.5	541,039
Kostenki14	n/a	n/a	n/a	0.0001	0.2	778,865	n/a	n/a	n/a	0.0016	3	745,568
Kostenki12	0.0008	0.9	55,817	0.0013	1.4	47,155	0.0015	0.6	615,864	0.0021	2.1	44,385
Muierii2	0.0002	0.3	89,779	-0.0003	-0.4	71,687	-0.0006	1.5	49,078	0.0013	1.6	73,066
Paglicci133	0.0002	0.2	75,733	0.0008	0.8	61,511	-0.0002	-0.8	79,159	0.0015	1.7	52,021
Vestonice13	0	-0.1	128,690	-0.001	-1.3	101,433	-0.0008	-0.3	58,743	0.001	1.4	102,285
Vestonice43	0	0	147,703	-0.0002	-0.3	116,524	0.0002	-1.3	111,268	0.0016	2.4	123,622
Pavlov1	-0.0011	-1.1	50,750	-0.0007	-0.7	47,910	-0.0006	0.4	132,591	0.0018	1.8	46,263
Vestonice16	-0.0004	-0.8	879,269	-0.0003	-0.6	615,864	n/a	-0.7	49,523	0.0012	2.5	603,758
Ostuni1	-0.0009	-1.6	338,407	-0.0011	-1.7	278,753	-0.0007	-1.3	301,276	0.0007	1.3	278,203
KremsWA3	-0.0009	-1.5	219,623	-0.0003	-0.5	189,593	-0.0001	-0.2	181,849	0.0008	1.2	161,774
ElMiron	-0.0016	-3	745,568	-0.0014	-2.5	541,039	-0.0012	-2.5	603,758	n/a	n/a	n/a
HohleFels49	-0.001	-1.1	56,741	0.0002	0.3	49,569	-0.0009	-1	55,683	0.0018	2.2	52,016
GoyetQ2	-0.0025	-2.8	67,154	-0.0013	-1.5	57,936	-0.0007	-0.8	55,614	-0.0006	-0.6	49,812
Villabruna	-0.0038	-7.8	1,125,277	-0.0038	-7	718,424	-0.0033	-6.9	797,720	-0.002	-4.4	696,121
Bichon	-0.0037	-7.9	1,669,947	-0.0037	-6.9	810,452	-0.0032	-6.9	912,465	-0.0018	-4.1	776,355
Rochedane	-0.0034	-5.4	218,932	-0.0033	-5.1	189,843	-0.0031	-5	181,626	-0.0016	-2.7	161,882
Ranchot88	-0.0036	-6.3	381,831	-0.0038	-6.4	351,382	-0.0034	-6.7	337,147	-0.0022	-4.4	306,065
Falkenstein	-0.0029	-3.3	58,169	-0.0025	-2.8	49,262	-0.0023	-2.7	55,226	-0.0021	-2.7	51,576
Chaudardes1	-0.004	-4.7	86,086	-0.0043	-5	74,202	-0.0035	-4.3	71,180	-0.0027	-3.5	63,128
Loschbour	-0.0043	-9.3	1,660,854	-0.0042	-8.2	802,734	-0.004	-8.5	904,305	-0.0027	-6.4	768,832
LaBrana1	-0.0034	-7.4	1,557,333	-0.0034	-6.3	776,310	-0.0031	-6.8	875,649	-0.0018	-4.1	748,847
Hungarian.KO1	-0.0043	-9.3	1,137,698	-0.0048	-8.7	559,581	-0.004	-8.6	629,253	-0.003	-6.5	536,668
BerryAuBac	-0.0042	-4.2	50,065	-0.0036	-3.5	46,237	-0.003	-2.9	44,315	-0.0017	-1.6	40,189

References

1 Seguin-Orlando, A. *et al.* Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113-1118, doi:10.1126/science.aaa0114 (2014).

Section 12

Population affinities of the Satsurblia Cluster

Overview

Here we describe how two individuals from the Satsurblia Cluster[1]—the Upper Palaeolithic *Satsurblia* and the Mesolithic *Kotias*—relate to other samples in this study.

Satsurblia Cluster samples have Basal Eurasian ancestry

We computed statistics of the form: *D(Satsurblia/Kotias, Pre-Neolithic Europeans; UstIshim/Oase1, Mbuti)* (Table S12.1). These statistics are significantly negative, different from the pattern seen when both of the first two samples are pre-Neolithic Europeans (Supplementary Information section 8). These results suggest that Satsurblia Cluster samples, unlike pre-Neolithic people from more western parts of Europe, harbored ancestry from a "Basal Eurasian" lineage. Specifically, they harbored ancestry from a lineage that split from European hunter-gatherers, *UstIshim* and *Oase1* before those groups separated[2].

Table S12.1. Z-score of *D(Satsurblia/Kotias, X; Y, Mbuti)*

X/Y	Han	UstIshim	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Villabruna	Bichon	Loschbour	LaBrana1	Hungarian.KO1	Malta1	Motala12	Stuttgart
D(Satsurblia, X; Y, Mbuti) Satsurblia: 1,460,368 SNPs															
Han	NA	-4.6	-4.4	3.7	0.2	3.1	1.9	7.1	6.5	5.6	4.1	5.5	2.7	6.6	11.9
UstIshim	-1.7	NA	-2	1.8	-0.3	1.8	2.7	6.2	7.4	7.1	5.8	9	4.8	8.3	11.9
Oase1	2.7	1.4	NA	5.5	2.5	5.2	4.9	7.6	8.8	10	7.1	8.8	6.7	10.1	14.5
Kostenki14	-0.3	-3.7	-3.3	NA	-9.8	-11.5	-8.9	-5.3	-3.9	-4.4	-4.6	-2.3	-2.4	-1.3	4.8
GoyetQ116-1	-3.1	-5.1	-3.7	-8.9	NA	-10	-18	-6.5	-6.9	-8.1	-10.1	-3.5	-3.9	-4.8	3.2
Vestonice16	-1.1	-4.1	-3.5	-11.6	-10.6	NA	-13.4	-9.8	-8.2	-9.2	-8.3	-7	-3.7	-5.3	2.4
ElMiron	-2.7	-3.3	-3.6	-9.1	-18.7	-13.4	NA	-17.7	-20	-21.1	-24.5	-11.6	-3	-12.2	0.3
Villabruna	-1.8	-3.5	-3.6	-9.1	-11	-14.1	-22.2	NA	-31.6	-33.2	-26.7	-26.8	-4.2	-22.8	-3.8
Bichon	-4.2	-3.7	-3.4	-9.2	-12.1	-12.6	-24.9	-31.3	NA	-32.9	-30.2	-25.5	-5.4	-21.5	-4.1
Loschbour	-5.2	-4.1	-2.5	-9.9	-13.9	-13.9	-27.6	-34.8	-34.6	NA	-32.9	-27.7	-5.6	-24.4	-5.8
LaBrana1	-4.5	-3.1	-2.4	-8.3	-13.9	-11.3	-28.9	-26.1	-29.2	-31.1	NA	-21.3	-4	-19.4	-2.7
Hungarian.KO1	-4.8	-1.2	-3.4	-7.2	-9.3	-11.6	-17.8	-27.6	-26.2	-27	-24.2	NA	-5	-21.1	-6.1
Malta1	-5.4	-4.2	-3.2	-5.7	-7.9	-6.6	-6.2	-3.7	-3.4	-3.3	-4.2	-3.2	NA	-8.1	5
Motala12	-4.3	-3.1	-2.5	-6.5	-10.3	-9.7	-16.3	-22.1	-21	-22.8	-20.2	-19.8	-10	NA	-4.5
Stuttgart	-0.3	-0.1	-0.9	-2	-4.1	-4.5	-6.6	-6.4	-5.1	-6.3	-5.9	-6.5	1.9	-5.5	NA
Kotias	-0.8	0.4	-1.1	-1.2	-1.9	-0.8	-3.4	-2	-0.5	-1.3	-2.7	-1.2	-0.4	-3	-1
D(Kotias, X; Y, Mbuti) Kotias: 2,133,968 SNPs															
Han	NA	-4.9	-3.8	4.7	2.1	4.2	5.1	8.7	7.2	6.9	6.5	6.9	3.2	9.5	12.6
UstIshim	-1.2	NA	-1.6	2.7	0.9	2.2	4.7	8.1	7.9	8.4	7.9	10.4	5	10.6	13
Oase1	3.6	1.4	NA	5.8	4.1	5.4	7	8.1	10.2	10.8	8.8	10.3	7.2	11.1	14.4
Kostenki14	-0.1	-4.4	-3	NA	-8.6	-11.1	-6.9	-4.1	-3.8	-3.4	-3.4	-1.8	-2.2	1.1	6
GoyetQ116-1	-2.8	-6.3	-3.9	-9.2	NA	-10.3	-17.1	-5.6	-6.1	-8.2	-9.6	-3.9	-3.9	-2.7	3.9
Vestonice16	-0.6	-4.8	-3.4	-11.8	-9.9	NA	-11.7	-9.7	-7.9	-8.8	-7	-6.9	-3.7	-3.6	3.3
ElMiron	-2.4	-5	-3.5	-9.3	-18.3	-13.7	NA	-18.2	-18.7	-20.7	-23.8	-12.6	-2.8	-10.4	0.8
Villabruna	-1.4	-4.4	-4	-9.1	-9.7	-13.7	-21.4	NA	-30.5	-33	-26.4	-26.5	-3.7	-20.1	-3.2
Bichon	-3.9	-4.3	-2.8	-8.9	-10.6	-12.7	-23	-31	NA	-31.3	-29.1	-24.7	-5.2	-18.2	-3.1
Loschbour	-4.7	-4.5	-2.5	-9.3	-12.7	-14.1	-24.3	-33.6	-31.7	NA	-31.4	-27.3	-5.1	-21.3	-4.7
LaBrana1	-4	-4	-2.4	-8.6	-13.3	-11.7	-27.4	-26.3	-28.8	-31.4	NA	-22.6	-4.4	-17.4	-1.9
Hungarian.KO1	-4.3	-2	-2.7	-7.7	-8.8	-11.8	-16.5	-28.1	-25.8	-27.6	-24.4	NA	-5.3	-18.8	-5.4
Malta1	-5.5	-4.9	-3.8	-5.7	-7.2	-6.9	-4	-2.4	-3.8	-2.9	-2.9	-3	NA	-6.2	5.4
Motala12	-3.8	-3.3	-2.4	-6.1	-9.1	-9.8	-14.4	-21.4	-21.3	-22.4	-19.8	-19.8	-9.6	NA	-3.5
Stuttgart	0.4	-0.6	-0.4	-1.3	-2.6	-3.9	-4	-5.2	-4.6	-5.3	-4	-5.7	2.1	-3.1	NA
Satsurblia	0.8	-0.4	1.1	1.2	1.9	0.8	3.4	2	0.5	1.3	2.7	1.2	0.4	3	1

Satsurblia Cluster samples have West Eurasian as well as Basal Eurasian ancestry

Satsurblia Cluster samples have substantial amounts of Basal Eurasian ancestry. If they had entirely Basal Eurasian ancestry, however, a prediction would be that they would be no more closely related to pre-Neolithic Europeans than they are to *Han*, *UstIshim*, or *Oase1*. This is not the case. Table S12.2 shows that statistics of the form *D(Ust-Ishim/Han/Oase1, pre-Neolithic Europe; Satsurblia/Kotias, Mbuti)* are all significantly negative (Z<<-3 scores; first

three rows of Table S12.2). The only way to explain this is if Satsurblia Cluster samples harbor a mix of Basal Eurasian and West Eurasian ancestry.

Table S12.2 Z-score of *D(X, Y; Satsurblia/Kotias, Mbuti)*

X/Y	Han	UstIshim	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	ElMiron	Villabruna	Bichon	Loschbour	LaBrana1	Hungarian.KO1	Malta1	Motala12	Stuttgart
D(X, Y; Satsurblia, Mbuti) Satsurblia: 1,460,368 SNPs															
Han	NA	2.8	6.6	-4.1	-3.1	-4.2	-4.6	-9.2	-10.6	-10.8	-8.3	-10.1	-8.1	-11.2	-12.3
UstIshim	-2.8	NA	3.5	-6.1	-5.3	-6.2	-6.1	-9.9	-11.1	-11.1	-9	-10.4	-9.4	-11.7	-12
Oase1	-6.6	-3.5	NA	-8.4	-6.2	-8.6	-8.7	-11.1	-12	-11.9	-9.3	-11.3	-10.2	-12.5	-15.1
Kostenki14	4.1	6.1	8.4	NA	0.8	-0.3	-0.4	-4.2	-5.8	-6.2	-3.6	-5.1	-3.4	-5.9	-6.6
GoyetQ116-1	3.1	5.3	6.2	-0.8	NA	-0.6	-0.8	-4.3	-5.7	-6.1	-3.7	-5.7	-3.8	-5.5	-6.9
Vestonice16	4.2	6.2	8.6	0.3	0.6	NA	0.1	-3.8	-5	-5.6	-2.9	-4.4	-3	-4.8	-6.6
ElMiron	4.6	6.1	8.7	0.4	0.8	-0.1	NA	-3.9	-5.1	-5.7	-3.3	-5.5	-3.2	-4.4	-7.1
Villabruna	9.2	9.9	11.1	4.2	4.3	3.8	3.9	NA	-1.5	-2.1	0.4	-1.8	0.4	-1.2	-2.8
Bichon	10.6	11.1	12	5.8	5.7	5	5.1	1.5	NA	-0.2	2.3	0	2.1	0.1	-1.1
Loschbour	10.8	11.1	11.9	6.2	6.1	5.6	5.7	2.1	0.2	NA	2.7	0.6	2.2	0.4	-0.9
LaBrana1	8.3	9	9.3	3.6	3.7	2.9	3.3	-0.4	-2.3	-2.7	NA	-2.1	-0.3	-2	-3.2
Hungarian.KO1	10.1	10.4	11.3	5.1	5.7	4.4	5.5	1.8	0	-0.6	2.1	NA	1.7	0.1	-1.2
Malta1	8.1	9.4	10.2	3.4	3.8	3	3.2	-0.4	-2.1	-2.2	0.3	-1.7	NA	-2.3	-2.9
Motala12	11.2	11.7	12.5	5.9	5.5	4.8	4.4	1.2	-0.1	-0.4	2	-0.1	2.3	NA	-1.1
Stuttgart	12.3	12	15.1	6.6	6.9	6.6	7.1	2.8	1.1	0.9	3.2	1.2	2.9	1.1	NA
Kotias	26.6	24	23	20.2	19.2	19.6	18.1	16.6	15.8	16.1	17.9	15.5	16.6	16.4	14.9
D(X, Y; Kotias, Mbuti) Kotias: 2,133,968 SNPs															
Han	NA	3.9	7.4	-4.9	-4.7	-5	-7.5	-10.9	-11.3	-11.7	-10.7	-11.6	-8.5	-13.9	-12.9
UstIshim	-3.9	NA	3.3	-8	-7.5	-7.7	-10.1	-13.3	-12.4	-12.9	-12.2	-13.3	-9.9	-14.4	-14.5
Oase1	-7.4	-3.3	NA	-8.3	-8	-8.7	-10.9	-12.6	-13	-13.3	-11.6	-12.3	-10.7	-13.1	-15.3
Kostenki14	4.9	8	8.3	NA	-0.6	-0.2	-2.3	-5.7	-5.3	-6.6	-5	-6.5	-3.6	-7.6	-7.2
GoyetQ116-1	4.7	7.5	8	0.6	NA	0.1	-1.7	-4.6	-4.6	-5.2	-3.7	-5.2	-3	-6.5	-6.6
Vestonice16	5	7.7	8.7	0.2	-0.1	NA	-2	-4.9	-5	-6	-4.9	-5.6	-3.3	-7.1	-7.2
ElMiron	7.5	10.1	10.9	2.3	1.7	2	NA	-2.5	-3.4	-4	-3.1	-4.1	-1.1	-4.2	-5.2
Villabruna	10.9	13.3	12.6	5.7	4.6	4.9	2.5	NA	-0.2	-1.2	0.2	-1.3	1.3	-2.1	-2
Bichon	11.3	12.4	13	5.3	4.6	5	3.4	0.2	NA	-1.2	0.2	-1.1	1.6	-2.5	-1.7
Loschbour	11.7	12.9	13.3	6.6	5.2	6	4	1.2	1.2	NA	1.5	0.2	2.5	-1.3	-0.7
LaBrana1	10.7	12.2	11.6	5	3.7	4.9	3.1	-0.2	-0.2	-1.5	NA	-1.2	1.3	-2.4	-2.2
Hungarian.KO1	11.6	13.3	12.3	6.5	5.2	5.6	4.1	1.3	1.1	-0.2	1.2	NA	2.4	-1.5	-0.7
Malta1	8.5	9.9	10.7	3.6	3	3.3	1.1	-1.3	-1.6	-2.5	-1.3	-2.4	NA	-4	-3.2
Motala12	13.9	14.4	13.1	7.6	6.5	7.1	4.2	2.1	2.5	1.3	2.4	1.5	4	NA	0.5
Stuttgart	12.9	14.5	15.3	7.2	6.6	7.2	5.2	2	1.7	0.7	2.2	0.7	3.2	-0.5	NA
Satsurblia	26.4	24.5	21.6	19.4	17.3	18.2	15.6	14.9	15.6	15.2	15.3	14.6	15.8	13.1	14.2

The Satsurblia and Villabruna Clusters are not particularly closely related

What is the nature of the West Eurasian genetic affinities in the Satsurblia Cluster samples?

We observe significantly positive statistics of the form *D(Villabruna Cluster, Early Upper Palaeolithic Europeans; Satsurblia Cluster, Mbuti)*, showing that Satsurblia Cluster samples share more alleles with Villabruna Cluster samples—for example, *Villabruna*, *Bichon*, *Loschbour*, *LaBrana1*, and *Hungarian.KO1*—than with Early Upper Palaeolithic Europeans (*Kostenki14*, *GoyetQ116-1*, *Vestonice16* and *ElMiron*) (Table S12.2). This suggests the possibility that a component of the non-Basal Eurasian ancestry in the Satsurblia Cluster may be related to the ancestry that appears in our European sample series with the Villabruna Cluster. In other words, migrations involving populations related to the Satsurblia Cluster could be responsible for the genetic link between the Near East and the Villabruna Cluster (Supplementary Information section 11).

It is important to emphasize, first of all, that Satsurblia Cluster can not be the direct source of the Near Eastern affinity that appears in our European sample series from *Villabruna* onward (Figure 4b), as Satsurblia Cluster samples have substantial Basal Eurasian ancestry, whereas Villabruna Cluster samples do not (Supplementary Information section 8).

To explore the relationship between the non-Basal Eurasian ancestry in the *Satsurblia* cluster and the Near Eastern related ancestry in the Villabruna Cluster in more detail, we fit

Satsurblia into the Admixture Graph of Supplementary Information section 6 that includes *Mbuti, UstIshim, Malta1, GoyetQ116-1, Kostenki14, Vestonice16,* and *ElMiron* (Figure S6.3). In all fitting models, *Satsurblia* is inferred to harbor ~32% ancestry from a Basal Eurasian lineage that branched before *UstIshim* (Extended Data Figure 4). These results are consistent with the finding that Satsurblia Cluster samples have Basal Eurasian ancestry[1], while also rejecting the previous model in which *Satsurblia* is of entirely Basal Eurasian ancestry[1]. Our fitted model specifies that the majority of ancestry in *Satsurblia* is from a West Eurasian, lineage providing an explanation both for why *Satsurblia* has Basal Eurasian ancestry, while at the same time sharing alleles with all Europeans beginning with *Kostenki14*.

All three of the best fitting models in Extended Data Figure 4 specify that the majority ancestry component in *Satsurblia* branched very deeply in the tree of West Eurasian populations, forming a clade with *Malta1*. Further evidence for a deep connection to *Malta1* and populations with admixture of *Malta1*-related ancestry comes from the observation in Table S12.2 that *D(Motala12/Malta1, Early Upper Palaeolithic Europeans; Satsurblia Cluster, Mbuti)* is significantly positive. In a simple model of this type, a prediction is that statistics of the form *D(Villabruna Cluster, Early Upper Palaeolithic Europeans; Malta Cluster, Mbuti)*, would be significantly positive, as *Malta1* would share more alleles with Villabruna Cluster samples than with Early Upper Palaeolithic Europeans. However, we do not detect any such a signal (Supplementary Information section 9).

Regardless of whether a population closely related to *Satsurblia* is responsible for the affinity of Villabruna Cluster samples to the Near East, there is evidence that a new lineage with affinities to present-day Near Easterners spread across Europe at this time. The evidence for this spread is that the genetic affinity of pre-Neolithic Europeans to Near Easterners abruptly increases with the appearance Villabruna Cluster, with no earlier European sample showing as strong an affinity despite sharing large amounts of genetic drift with the Villabruna Cluster (Figure 4b). An important direction for future work is to analyse more ancient DNA samples from southeastern Europe and western Asia in order to understand the history of the migratory events that our data show must have occurred around this time.

References

1 Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications* **6**, 8912, doi:10.1038/ncomms9912 (2015).

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

Section 13

Population structure in the Villabruna Cluster

Villabruna Cluster samples harbor variable proportions of at least 3 ancestral lineages

We co-analysed Villabruna Cluster samples with earlier European hunter-gatherers and present-day populations. The goal of this analysis was to begin to understand the complexity of the relationship of the Villabruna Cluster samples to other samples.

We used *qpWave*[1] to compute f_4-statistic vectors of the form:

$$f_4(Left_h, Left_i; Right_j, Right_k) \qquad \text{(Equation S13.1)}$$

Left (Villabruna Cluster samples with >0.1x coverage):
Villabruna, Bichon, Rochedane, Ranchot88, Loschbour, LaBrana1, Hungarian.KO1

Right (pre-Villabruna Cluster samples at >0.1x, and present-day humans):
UstIshim, GoyetQ116-1, Kostenki14, Malta1, Vestonice16, ElMiron, Stuttgart, Ami, Bougainville, Chukchi, Eskimo_Sireniki, Eskimo_Naukan, French, Han, Ju_hoan_North, Karitiana, Kharia, Mbuti, Onge, Papuan, She, Ulchi, Yoruba, Iraqi_Jew.

The papers that introduced *qpWave* showed that if the *Left* populations are mixtures – in various proportions – of *N* sources differently related to the *Right* populations, then all f_4-statistics of the form of Equation S13.1 will be consistent with being a linear combination of *N* vectors of statistics that correspond to the mixing populations, and the matrix will have rank N-1[1,2]. We can test whether this is the case using a Hotelling's T^2 test[1].

Applying this test to our data, we reject rank 0 at high statistical significance ($P=1.1\times10^{-13}$), and we also reject rank 1 (P=0.026). Rank 2 is consistent with the data to within the limits of our resolution (P=0.61). Thus, the Villabruna Cluster populations derive from a mixture of at least three ancestral populations that are differentially related to the *Right* set.

We obtained consistent results when we removed *Rochedane* with its modest coverage (its removal brought the total number of SNPs with coverage in all samples from 36,372 to 114,473). Again, we reject rank 0 at high significance ($P=2.4\times10^{-21}$), and we also reject rank 1 (P=0.020). Rank 2 is consistent with the data (P=0.41). This again supports the model of few as three ancestral populations, although the truth could of course be more.

Hints about the ancestral populations that contributed to the Villabruna Cluster

The *qpWave* analysis is able to document that at least three sources are necessary to account for the allele frequency correlation patterns in Villabruna Cluster samples. However, we did not succeed at convincingly modeling the ancient relationships among these sources. We nevertheless found some hints about the history relating these ancient samples, and we discuss several relevant observations.

The common thread that binds together Villabruna Cluster samples

Table S13.1 documents an attraction of the *Villabruna* genome to all other samples in the cluster, especially after ~14,000 years ago. This reflects the common thread that binds together the Villabruna Cluster samples, and is the reason for the designation of these samples as a cluster with at least one shared source of distinctive ancestry.

Table S13.1. Z-score of *D(X, Y; Villabruna, Mbuti).* *Villabruna* shares more drift with Villabruna Cluster samples post-dating it in Europe than with European samples pre-dating it.

X/Y	Han	UstIshim	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	HohleFels49	Rochedane	Ranchot88	Loschbour	LaBrana1	Hungarian.KO1
Han	NA	0.4	3.3	-13.2	-14	-18.7	-18.8	-26.9	-11.3	-38	-39.9	-45.5	-36.6	-39.8
UstIshim	-0.4	NA	2.5	-11.6	-12	-16.1	-15.4	-23.7	-10.5	-34.1	-35.5	-39.8	-31.6	-32.5
Oase1	-3.3	-2.5	NA	-11.5	-11.1	-14.4	-12.9	-20.4	-5.5	-22	-28.3	-34.9	-27.5	-28.3
Kostenki14	13.2	11.6	11.5	NA	-1.8	-5.3	-6	-14.4	-4.4	-23.2	-25.8	-29.5	-21.4	-24.2
GoyetQ116-1	14	12	11.1	1.8	NA	-3.4	-4.3	-12.4	-3.7	-23.6	-24.8	-29.4	-19.4	-22.4
Vestonice16	18.7	16.1	14.4	5.3	3.4	NA	-1.8	-9.2	-1.8	-19.7	-22	-25.1	-16.7	-18.5
Ostuni1	18.8	15.4	12.9	6	4.3	1.8	NA	-7.6	-0.3	-16.1	-17.2	-21.4	-12.4	-14.9
ElMiron	26.9	23.7	20.4	14.4	12.4	9.2	7.6	NA	4.3	-11.4	-13	-15.1	-6.6	-9.3
HohleFels49	11.3	10.5	5.5	4.4	3.7	1.8	0.3	-4.3	NA	-5.6	-9.8	-14	-6.4	-9
Bichon	41.7	37.3	32.0	27.7	24.9	21.8	18.9	12.8	11.0	-0.7	-0.4		7.5	3.9
Rochedane	38	34.1	22	23.2	23.6	19.7	16.1	11.4	5.6	NA	0.4	-0.1	7.8	4.5
Ranchot88	39.9	35.5	28.3	25.8	24.8	22	17.2	13	9.8	-0.4	NA	-1	8.2	3.9
Loschbour	45.5	39.8	34.9	29.5	29.4	25.1	21.4	15.1	14	0.1	1	NA	9.1	5.9
LaBrana1	36.6	31.6	27.5	21.4	19.4	16.7	12.4	6.6	6.4	-7.8	-8.2	-9.1	NA	-3
Hungarian.KO1	39.8	32.5	28.3	24.2	22.4	18.5	14.9	9.3	9	-4.5	-3.9	-5.9	3	NA

Variable proportions of GoyetQ116-1-related ancestry in Villabruna Cluster samples

Despite their shared thread of ancestry, a different set of *D*-statistics show directly that Villabruna Cluster samples do not form clade with respect to all later European hunter-gatherers. In particular, statistics of the form *D(Villabruna, X; Y, Mbuti)* show that El Mirón cluster samples and *GoyetQ116-1* share more alleles with a subset of Villabruna Cluster samples (*Rochedane, Ranchot88, Loschbour and LaBrana1*) than they do with *Villabruna* itself (Table S13.2). These results document a variable fraction of ancestry from the European *GoyetQ116-1* lineage within the Villabruna Cluster, as is also shown in the fitted Admixture Graphs in Supplementary Information section 6, and the variable mixture proportion estimates in Supplementary Information section 7.

We observed that *Hungarian.KO1* does not show the same affinity to *ElMiron* and *GoyetQ116-1* as do the other Villabruna Cluster (Table S13.2). We were concerned that these results might be an artifact of the fact that the libraries that were used to build *Hungarian.KO1* were not UDG-treated. However, this does not appear to explain the observations, as the pattern replicates when restricting to transversion SNPs that are not affected by ancient DNA damage (Table S13.3). These results suggest the possibility that *Hungarian.KO1* may have less *ElMiron* and *GoyetQ116-1* affinity than does *Villabruna.*

Table S13.2. Z-score of *D(Villabruna, X; Y, Mbuti)* for all sites.

X/Y	Han	UstIshim	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	HohleFels49	Rochedane	Ranchot88	Loschbour	LaBrana1	Hungarian.KO1
Bichon	-0.4	1	0.1	-1.3	0.9	1.1	-2.3	-2.4	-2.7	-1.1	-3	-3.2	0.7	-3
Rochedane	-0.8	0.5	-0.7	0.6	-1.7	0.6	-0.2	-3.1	-1	NA	-1.1	-0.6	0.1	2.9
Ranchot88	-1.2	1.1	-0.5	-0.1	-3	0.6	-0.3	-3.2	-3.6	-0.8	NA	-5.2	-3.6	0.4
Loschbour	-3.9	-0.5	1.7	-0.3	-3.7	-0.6	1.7	-4.1	-3.8	-0.6	-6.1	NA	-5.7	-1.4
LaBrana1	-2.9	0.6	1.3	0.9	-4.1	2.9	2.6	-4.1	-3	7.9	5	4.2	NA	6.4
Hungarian.KO1	-3.3	2.2	0.5	1.6	0.6	2.1	3.4	5	1.1	7.8	4.5	4.6	3.4	NA

Gene flow between relatives of Han and relatives of some Villabruna Cluster samples

Table S13.4 shows that *Han* shares more alleles with a subset of Villabruna Cluster samples (*Bichon, Loschbour, La Brana1, Hungarian.KO1*) than it does with another subset of Villabruna Cluster samples (including *Villabruna* itself) as well as non-Villabruna Cluster Europeans. The variability in the Villabruna Cluster in their degree of allele sharing with East Asians documents another complexity of the history of the Villabruna Cluster (Figure 4b).

Table S13.3. *Z*-score of *D(Villabruna, X; Y, Mbuti)* for transversion sites.

X/Y	Han	UstIshim	Oase1	Kostenki14	GoyetQ116-1	Vestonice16	Ostuni1	ElMiron	HohleFels49	Rochedane	Ranchot88	Loschbour	LaBrana	Hungarian_KO1
Bichon	-2.7	-0.6	0.8	0.4	-1.2	1.5	1.7	-1.7	-1.2	-2	0.7	-2.3	-2.3	0.1
Rochedane	-0.5	0	-0.7	-0.8	-0.8	-0.8	0.4	-3.2	-1.4	NA	-0.8	-1.1	0	0.6
Ranchot88	-1.2	0.8	-0.8	-1.1	-2	-0.4	-0.6	-2.9	-1.8	-1.1	NA	-3.3	-3.1	-0.5
Loschbour	-3.9	-0.7	1.8	0.9	-2.4	0.6	2.7	-3.2	-1.7	0.8	-2.6	NA	-4.7	-0.8
LaBrana1	-2.6	0	1.7	1.2	-2.9	2.7	2.7	-3.1	-1.1	5.7	5.1	3.5	NA	5.7
Hungarian_KO1	-3.8	1.5	0.4	1.4	-0.3	1.5	2.6	3.3	0.4	3.4	3.7	3.7	2	NA

Table S13.4. *D(X, Y; Han, Mbuti)* for all sites. Han Chinese share more alleles with a subset of Villabruna Cluster samples (Bichon, Loschbour, LaBrana1, and Hungarian.K01) than they do with early European hunter-gatherers.

X/Y	Villabruna	Loschbour	LaBrana1	Hungarian.KO1	Malta1
UstIshim	0.2	-2.9	-2.2	-2.4	-3.6
Oase1	-3.7	-6.1	-5.2	-6.1	-7.1
Kostenki14	-1.3	-4.3	-3.8	-4	-5.3
Kostenki14.sg	-1.3	-4.6	-3.9	-4.4	-5.1
GoyetQ116-1	1.5	-1.1	-0.5	-1.2	-2.6
Vestonice16	-0.9	-3.9	-3.4	-3.4	-5
Ostuni1	1	-2	-1.9	-1.9	-3.2
ElMiron	1.5	-2	-1.3	-1.8	-3.4
HohleFels49	0	-2.5	-0.8	-1.8	-2
Villabruna	NA	-3.9	-2.9	-3.3	-4.2
Bichon	3	-1.1	0	-0.5	-1.9
Rochedane	0.8	-2	-0.9	-1.1	-3.4
Ranchot88	1.2	-1.9	-1.4	-2.1	-3.5
Loschbour	3.9	NA	1	0.5	-0.9
LaBrana1	2.9	-1	NA	-0.3	-1.9
Hungarian.KO1	3.3	-0.5	0.3	NA	-1.4
Malta1	4.2	0.9	1.9	1.4	NA

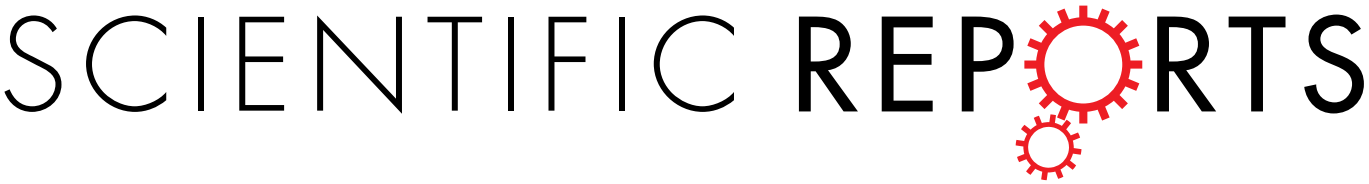
Summary

We have documented complexity in the ancestry of Villabruna Cluster samples, with variability in the degree of allele sharing with *ElMiron/GoyetQ116-1*, and independently with East Asians. We are confident that at least three sources of ancestry contribute to Villabruna Cluster samples, but at present we do not have a good model for the mixture.

References

1 Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370-374, doi:10.1038/nature11258 (2012).

2 Patterson, N. J. *et al.* Ancient Admixture in Human History. *Genetics*, doi:10.1534/genetics.112.145037 (2012).



OPEN

Neandertal cannibalism and Neandertal bones used as tools in Northern Europe

Received: 15 January 2016
Accepted: 06 June 2016
Published: 06 July 2016

Hélène Rougier[1], Isabelle Crevecoeur[2], Cédric Beauval[3], Cosimo Posth[4,5], Damien Flas[6], Christoph Wißing[7], Anja Furtwängler[4], Mietje Germonpré[8], Asier Gómez-Olivencia[9,10,11,12], Patrick Semal[8], Johannes van der Plicht[13,14], Hervé Bocherens[7,15] & Johannes Krause[4,5,15]

Almost 150 years after the first identification of Neandertal skeletal material, the cognitive and symbolic abilities of these populations remain a subject of intense debate. We present 99 new Neandertal remains from the Troisième caverne of Goyet (Belgium) dated to 40,500–45,500 calBP. The remains were identified through a multidisciplinary study that combines morphometrics, taphonomy, stable isotopes, radiocarbon dating and genetic analyses. The Goyet Neandertal bones show distinctive anthropogenic modifications, which provides clear evidence for butchery activities as well as four bones having been used for retouching stone tools. In addition to being the first site to have yielded multiple Neandertal bones used as retouchers, Goyet not only provides the first unambiguous evidence of Neandertal cannibalism in Northern Europe, but also highlights considerable diversity in mortuary behaviour among the region's late Neandertal population in the period immediately preceding their disappearance.

Neandertal funerary practices remain at the forefront of palaeoanthropological research, generating heated debates following the revision of old data and new excavations at key sites such as La Chapelle-aux-Saints[1,2], Roc de Marsal[3], Saint-Césaire[4] and La Ferrassie[5]. More generally, attention has focused on the variability of Neandertal mortuary practices to evaluate their cognitive and symbolic implications, especially as they may provide insights concerning the social systems of this fossil human group[6]. Neandertals are known to have buried their dead and are associated with mortuary behaviours that are often difficult to interpret in Palaeolithic contexts. The site of Krapina (Croatia) is an instructive example in this sense. Evidence for cannibalism was first proposed for this site as early as 1901[7] based on the fragmentation and traces of burning from a large collection of early Neandertal remains. This evidence has since been disputed by proponents of alternative explanations for the human bone modifications who argue for natural processes while others maintain that the anthropogenic manipulations are best interpreted in the context of secondary burials[8]. Several studies dedicated to cannibalism have proposed that securely identifying anthropogenic modifications related to this practice should incorporate evidence for the

[1]Department of Anthropology, California State University Northridge, 18111 Nordhoff St, Northridge, CA 91330-8244, USA. [2]Université de Bordeaux, CNRS, UMR 5199-PACEA, A3P, Allée Geoffroy Saint Hilaire, CS 50023, 33615 Pessac Cedex, France. [3]Archéosphère, 2 Rue des Noyers, 11500 Quirbajou, France. [4]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Rümelinstr. 23, 72070 Tübingen, Germany. [5]Max Planck Institute for the Science of Human History, Khalaische Straße 10, 07745 Jena, Germany. [6]Laboratoire TRACES – UMR 5608, Université Toulouse Jean Jaurès, Maison de la Recherche, 5 Allée Antonio Machado, 31058 Toulouse Cedex 9, France. [7]Department of Geosciences, University of Tübingen, Hölderlinstr. 12, 72074 Tübingen, Germany. [8]Royal Belgian Institute of Natural Sciences, 29 Vautier St, 1000 Brussels, Belgium. [9]Departmento de Estratigrafía y Paleontología, Facultad de Ciencia y Tecnología, Euskal Herriko Unibertsitatea, UPV-EHU. Apdo. 644, 48080 Bilbao, Spain. [10]IKERBASQUE, Basque Foundation for Science, María Díaz de Haro 3, 48013 Bilbao, Spain. [11]UMR 7194 CNRS, Département de Préhistoire, Muséum national d'Histoire naturelle, Musée de l'Homme, 17 Place du Trocadéro, 75016 Paris, France. [12]Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, Avda. Monforte de Lemos 5, 28029 Madrid, Spain. [13]Centre for Isotope Research, Groningen University, Nijenborgh 4, 9747 AG Groningen, Netherlands. [14]Faculty of Archaeology, Leiden University, PO Box 9514, 2300 RA Leiden, Netherlands. [15]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72072 Tübingen, Germany. Correspondence and requests for materials should be addressed to H.R. (email: helene.rougier@csun.edu)

Figure 1. Neandertal remains from the Troisième caverne of Goyet (Belgium). *Designates the specimens that have been directly dated. Scale = 3 cm.

similar treatment of both faunal and human remains in the interest of extracting nutrients[9–11]. In addition to Gran Dolina (level TD6; Early Pleistocene) in Spain, which has produced the earliest undisputed evidence for cannibalism[12], further examples have also been documented at several Western European Neandertal sites, including El Sidrón and Zafarraya[13,14] in Spain, and Moula-Guercy and Les Pradelles[15,16] in France.

Here we provide new data on the diversity of Neandertal mortuary behaviour, focusing on a small area of their known range, Northern Europe, during Marine Isotope Stage (MIS) 3 (ca. 60–30 thousand years ago), in order to identify small-scale processes during this short period that witnessed the disappearance of the Neandertals[17]. We present 99 new Neandertal remains recently identified among the collections from the Troisième caverne of Goyet (Belgium), some of which exhibit anthropogenic modifications, and discuss their implications.

The Troisième caverne (or "Third cave") of Goyet, excavated in the latter half of the 19th and beginning of the 20th century, and again at the end of the 1990s[18], is part of a large cave system located in the Mosan Basin (Supplementary Fig. S1). The most extensive excavations were carried out by Edouard Dupont in 1868, who described five "fauna-bearing levels" (FBL; ref. 19; Supplementary Note S1). The Troisième caverne yielded a rich archaeological sequence with Middle and Upper Palaeolithic deposits containing Mousterian, Lincombian-Ranisian-Jerzmanowician (LRJ), Aurignacian, Gravettian and Magdalenian artefacts as well as Neolithic and historic period material[20–23]. Whether the Mousterian material derives from a single or multiple phases of occupation is currently impossible to discern (Supplementary Note S2). Unfortunately, the excavation methods did not meet today's standards, and it appears that the levels described by Dupont actually represent a mix of material from different periods (e.g., ref. 24).

Several human remains from different levels were published by Dupont[19] and Hamy[25], although only a few figure in the Catalogue of Fossil Hominids[26], all of which were attributed to the Magdalenian. In 2004, we identified both a Neandertal mandible fragment and an isolated tooth among the human material recovered by Dupont from the Troisième caverne and currently housed at the Royal Belgian Institute of Natural Sciences (RBINS)[27], making Goyet one of the few Northern European sites north of 50° N to have yielded MIS 3 Neandertal remains (Supplementary Note S1 and Supplementary Fig. S1).

Results

Identification of new Neandertal remains at Goyet and their biogeochemical characterization. The reanalysis of the Goyet material comprised (i) the revision of the human skeletal material, (ii) systematic sorting of the faunal collections to check for unidentified human remains (Supplementary Fig. S2), and (iii) a multidisciplinary study of the human remains and their context. Two-hundred and eighty three human remains were identified from different periods, including 96 bone specimens and three isolated teeth identifiable as Neandertal (Supplementary Table S1 and Supplementary Notes S3, S4 and S5). A good number (n = 47) of the bone specimens refit, reducing the total number of isolated Neandertal remains to 64 (Fig. 1 and Supplementary Table S2), of which 10 were directly radiocarbon (^{14}C) dated, 15 were sampled for stable isotope analyses, and 10 for DNA extraction (Table 1 and Supplementary Table S3). Based on their morphology and morphometric characteristics, developmental stage and side for paired elements, as well as the successful recovery of endogenous mitochondrial DNA (mtDNA) sequences, the minimum number of individuals (MNI) represented by the Goyet sample is estimated at five (four adolescents/adults and one child represented by a single tooth; Supplementary Note S5 and Supplementary Fig. S3). Although the Neandertal sample includes cranial and postcranial elements (Fig. 1), with long bones best represented and extremities mostly absent, the minimum number of elements (MNE = 35) demonstrates a very low overall skeletal representation. The best represented elements are, in decreasing order, the tibia (six of the eight tibias expected for four adolescents/adults, 75% representation), femur and cranium (50%), humerus and mandible (25%; Supplementary Table S4).

Chemical elemental analyses performed together with stable isotope analyses were used to assess collagen preservation in preparation of ^{14}C dating (see Methods). The ecology of the Goyet Neandertals was also investigated using δ^{13}C and δ^{15}N isotope composition of bone collagen[28]. Direct ^{14}C dates obtained from the newly identified skeletal material place the Goyet Neandertals to ca. 40.5–45.5 ky calBP. However, when the youngest ages, which likely reflect undetected bone collagen contamination, are excluded (Supplementary Note S6), we cannot rule out the possibility that the Goyet Neandertals represent a single chronological group dating to ca. 44–45.5 ky calBP.

Specimen		Radiocarbon dating			Genetic analyses	Anthropogenic marks
ID	Description	Lab #	^{14}C age (BP)	Calibrated age (calBP) 95% probability		
2878–2D*	Lower lt P2 (mandible 2878–8)	GrA-54028	32,190 +200, −190	36,510–35,630	–	–
C5–1	Lt parietal frag.	–	–	–	Nean	–
Q53–4	Rt humerus diaph. frag. (humerus III)	GrA-54022	39,870 +400, −350	44,330–42,920	–	–
Q55–1	Lt clavicle frag.	GrA-54257	37,860 +350, −310	42,650–41,700	–	C
Q55–4	Rt tibia diaph. frag. (tibia IV)	–	–	–	Nean	C+N+P+R
Q56–1	Rt femur diaph. frag. (femur I)	GrA-46170	38,440 +340, −300	43,000–42,080	1	C+N+P
Q57–1	Lt tibia diaph. frag. (tibia II)	GrA-46173	41,200 +500, −410	45,630–43,910	2	C+N
Q57–2	Rt femur diaph. frag. (femur II)	GrA-54024	36,590 +300, −270	41,800–40,620	2	C+N+P
Q57–3	Rt tibia diaph. frag. (tibia VI)	GrA-60019	38,260 +350, −310	42,900–41,960	2	C+N
Q119–2	Lt rib 7? frag.	–	–	–	Nean	–
Q305–4	Lt tibia diaph. frag. (tibia I)	GrA-46176	40,690 +480, −400	45,150–43,430	3	C+N
Q305–7	Rt tibia diaph. frag. (tibia III)	–	–	–	1	C+N+P+R
Q374a–1	Rt tibia diaph. frag. (tibia V)	–	–	–	1	C+N+P+R
Q376–1	Hand prox. phalanx 2-4	GrA-46178	39,140 +390, −340	43,650–42,440	–	–
Q376–20	Rt humerus diaph. frag. (humerus II)	GrA-60018	37,250 +320, −280	42,240–41,290	–	C+N?

Table 1. Sample information and results of the ^{14}C and genetic analyses of the Neandertal remains from Goyet. *This specimen may have been varnished resulting in a young age (Supplementary Note S6). For the calibration of the ^{14}C ages, see Supplementary Note S6. Genetic analyses: 1–3 represent three distinct Neandertal mtDNA lineages, Nean: Neandertal status confirmed; Anthropogenic modifications: C: cutmarks, N: percussion notches, P: percussion pits, R: retoucher traces. All of the specimens are part of the RBINS collections and were excavated by E. Dupont in 1868.

Although this appears the most parsimonious hypothesis when individual bone associations, taphonomic aspects and similar anthropogenic modifications observed across the sample are taken into account, we retain the conservative range of ca. 40.5–45.5 ky calBP for the Goyet Neandertals in the absence of definitive evidence.

Out of the 10 samples processed for genetic analysis, seven show three distinct complete or almost complete mtDNA lineages (noted 1–3 in Table 1). The newly reconstructed mtDNAs from Goyet were compared with the mtDNA of 54 modern humans, eight previously sequenced Neandertals and one Denisovan individual[29–34]. Phylogenetic relationships were assessed using maximum parsimony and maximum likelihood trees (Fig. 2 and Supplementary Fig. S4), confirming the analysed specimens to fall within the known diversity of Neandertal mtDNA. The Goyet Neandertal mtDNAs appear most closely related to late Neandertal mtDNAs from Central and Western Europe, such as those from the Neandertal type-site (Germany), El Sidrón (Spain) and Vindija (Croatia), which all show only modest genetic variation despite large geographic distances when compared to modern humans. As previously suggested[31], this might reflect a low effective population size of Neandertals in general, and for the late Neandertals in particular.

Taphonomic analysis of the Goyet Neandertal material and anthropogenic modifications. Overall, the Neandertal remains are highly fragmented. Forty-nine percent of the bone specimens (47 out of 96) were refit to at least one other, with the number of specimens per refit set ranging from 2 to 8 (tibia I; Supplementary Fig. S5). Several examples of refits between levels 1 through 3 were also identified. None of the Neandertal bones are complete, although the proximal extremity of a hand phalanx (2878–37) is only slightly eroded (Fig. 1). Cortical surfaces are well preserved and exhibit limited post-depositional modifications. Most long bones fractures involve green breaks, as indicated by smooth margins and spiral fractures[35]. Traces of peeling may also provide evidence for the fresh bone fracture of a cranial fragment and several ribs (ref. 11; Supplementary Fig. S6). Although bears can produce such traces[36], the presence of cutmarks on several ribs (see below) suggests that the most parsimonious hypothesis is that they are anthropogenic. Traces of human chewing[37,38] are also suspected on the Neandertal phalanges but are inconclusive (Supplementary Fig. S6). The numerous unambiguous anthropogenic marks on the Goyet Neandertal remains can be attributed to three categories of bone surface modifications (Figs 3–5, Table 2, and Supplementary Figs S7 and S8):

(1) Cutmarks. Nearly a third of the Neandertal specimens bear cutmarks. The locations of the limited number of cutmarks observed on the upper limb may indicate disarticulation whereas those on the lower limb are consistent with defleshing. Several cutmarks on the internal and external surfaces of the ribs may be connected to evisceration, dismemberment of the thoracic cage and removal of the thoracic muscles. An additional cutmark on the medial side of the mandible, close to the mandibular condyle, appears consistent with dismemberment.
(2) Two types of percussion marks (notches and pits) were identified. Observed only on a single radius alongside several femurs and tibias, notches are likely connected to the fracturing of fresh diaphyses and marrow extraction. Percussion pits are common and probably indicate failed attempts at fracturing bones. Both

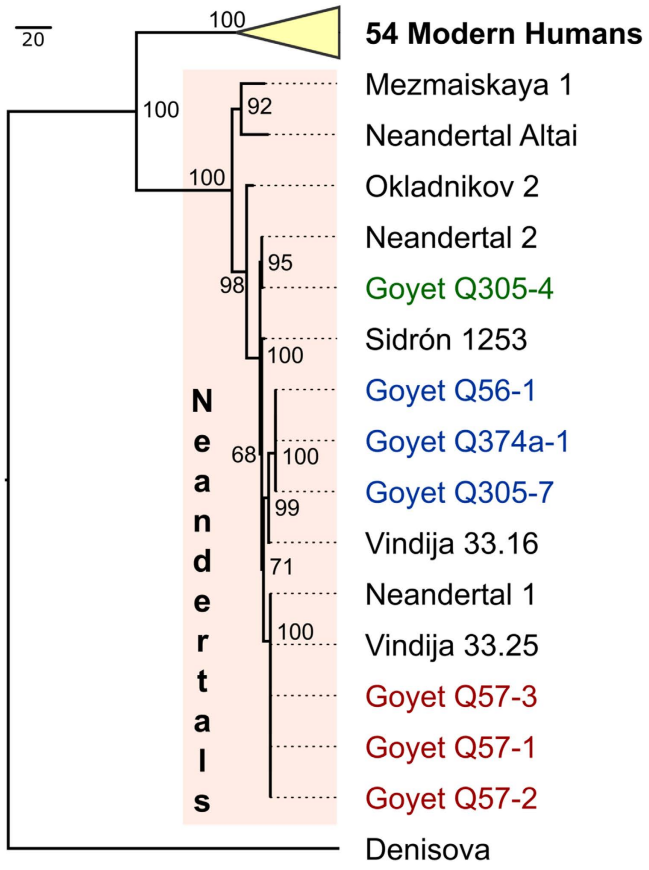


Figure 2. Maximum parsimony tree for the seven analysed Goyet samples that produced complete or almost complete mitochondrial genomes compared to 63 published modern human, Neandertal and Denisovan mtDNAs. Numbers at the main branch nodes represent bootstrap values after 1,000 iterations.

	Nean dertal	Horse	Reindeer	Carnivore
NISP Observed	*96*	*442*	*287*	*89*
NISP Cutmarks	31 (32%)	85 (19%)	126 (44%)	3 (3%)
NISP Percussion Notches	20 (21%)	107 (24%)	151 (53%)	0
NISP Percussion Pits	10 (10%)	6 (1%)	1 (0.3%)	0
NISP Retoucher Traces	5 (5%)	22 (5%)	58 (20%)	0
NISP Toothmarks	1 (1%)	27 (6%)	4 (1%)	17 (19%)

Table 2. Numbers and proportions of Neandertal, horse, reindeer and carnivore remains bearing anthropogenic modifications and toothmarks in the Goyet assemblage. Carnivores include bear (*Ursus spelaeus* or *Ursus arctos*), fox (*Vulpes vulpes* or *Vulpes lagopus*), a large canid (*Canis sp.*), hyaena (*Crocuta crocuta spelaea*), and badger (*Meles meles*). The observed faunal specimens were identified among a sample of Dupont's collection from FBL 2 and 3 (Supplementary Table S5). Note that the high percentage of retouchers made on reindeer bones is most likely related to the under-representation of fragments less than 55 mm long in our sample.

percussion notches and pits were also identified on eight bones (e.g. femur I, Fig. 5).

(3) Retouching marks. These marks, found on a femur and three tibias (Supplementary Figs S9–S12), result from retouching the edges of stone tools. The fact that none of the affected areas overlap on adjacent fragments suggests the bones to probably have first been marrow cracked. Femur III shows two retouching zones on the anterior and postero-medial surfaces, both located at mid-shaft. Interestingly, the traces found on the tibias are located in the same areas of the shaft on all three bones (posterior or postero-medial surface at mid-shaft). The retouchers are made on four different Neandertal bones that represent at least three of the four adolescent/adult Neandertal individuals (Supplementary Note S5).

While animal bone retouchers are common in European Middle Palaeolithic contexts (e.g., refs 39–41), Goyet is one of only four sites (Krapina in Croatia[42], La Quina and Les Pradelles in France[43,16]) to have yielded

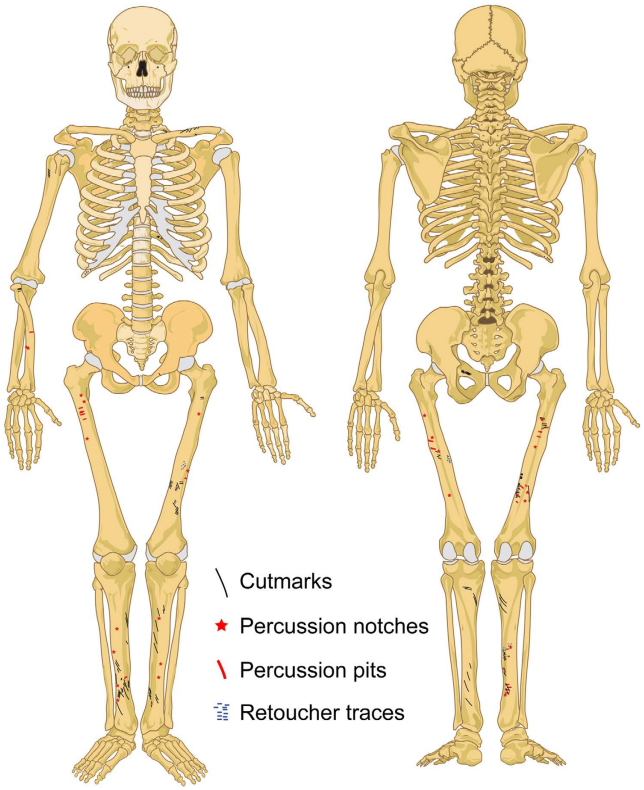


Figure 3. Overview of the anthropogenic modifications observed on the Neandertal remains from the Troisième caverne of Goyet (Belgium). See Supplementary Fig. S8 for individual Neandertal bones with anthropogenic modifications. Skeleton diagrams modified from https://en.wikipedia.org/wiki/File:Human_skeleton_front_en.svg and https://en.wikipedia.org/wiki/File:Human_skeleton_back_en.svg using Adobe Illustrator CS4 v. 14.0.0.

retouchers on Neandertal skeletal elements and the sole to have produced multiple examples (Table 3). At Krapina and Les Pradelles, femur shaft fragments were used as retouchers, whereas the La Quina example is on a parietal fragment. According to the criteria proposed by Mallye *et al.*[40], the blanks used for the Goyet retouchers made on Neandertal bones were most likely green due to the absence of scaled areas, and in addition, two of the five retoucher areas exhibit concentrated and superposed marks which imply prolonged use. The rectilinear morphology of the marks also supports the use of the bones for retouching flint flakes, the most common raw material found at Goyet.

Comparative taphonomic analysis of the fauna from the Troisième caverne. Due to the large size of the Goyet faunal collection (>30,000 specimens), only a sample from Dupont's excavation was examined (see Methods; Supplementary Fig. S2 and Supplementary Table S5). The skeletal material analysed corresponds mostly to long bone shaft fragments from various species that were mixed together within the collection and did not appear to have been previously sorted. We focused on remains from levels 3 and 2, which yielded the Neandertal remains, and on material from the same storage trays containing the human remains in order to have an overview of the associated faunal spectrum and assess food procurement and management strategies. Horse and reindeer are by far the most frequent species in the studied assemblage (86% of the 1,556 identified specimens; Supplementary Table S5). No rodent toothmarks were observed, carnivore remains are relatively sparse and carnivore damage is extremely rare on the Neandertal, horse and reindeer remains (Table 2), indicating carnivores to have had limited access to the bone material.

Anatomical profiles reveal numerous similarities between the Neandertal sample on one hand and horse and reindeer on the other (Supplementary Table S6 and Supplementary Fig. S13). The tibia is the most abundant element of all three species, whereas the axial skeleton and extremities of the forelimb and hindlimb are poorly represented. Bones of the hindlimb are better represented for all three species compared to forelimb elements, this is especially the case with the Neandertal material. The only notable difference between the faunal and Neandertal remains is the high representation of cranial elements for the latter. Unfortunately, the absence of contextual data precludes an analysis of the spatial distribution of both the faunal and Neandertal remains within the Troisième caverne.

The most intensely processed Neandertal elements are femurs and tibias (Supplementary Fig. S7), which are also the bones with the highest nutritional content (meat and marrow). The same pattern was documented for horse and reindeer bones. Overall, anthropogenic marks on the Neandertal remains match those most commonly recorded on the faunal material (Supplementary Figs S14–S16). All three taxa were intensively exploited,

		Femur III anterior area	Femur III medial area	Tibia III posterior area	Tibia IV posterior area	Tibia V medial area
Area	Length (mm)	14.6	19.2	11.4	17.0	20.9
	Width (mm)	4.4	7.5	9.0	5.8	13.2
	Preparatory scraping	no	no	no	no	no
	Morphology (if concentrated and superposed traces)	–	hatched	–	–	hatched
Marks	Orientation (to the long axis of the fragment)	oblique	transverse	transverse	transverse	transverse, slightly oblique
	Position	centered	centered?	centered	centered	centered
	Concentration	dispersed	concentrated and superposed	dispersed	dispersed	concentrated and superposed
	Morphology	rectilinear - smooth	rectilinear - rough	rectilinear - smooth	rectilinear - smooth and rough	rectilinear - smooth

Table 3. Description of the Neandertal bone retouchers from Goyet using the criteria of Mallye *et al.*[40] and Daujeard *et al.*[41].

exhibiting evidence of skinning, filleting, disarticulation and marrow extraction. However, the Neandertal remains stand out as they show a high number of percussion pits (Table 2), which may be linked to the thick cortical structure of Neandertal long bones. Although the Neandertal remains show no traces of burning, the possibility that they may have been roasted or boiled cannot be excluded. The high number of cutmarks and the fact that DNA could be successfully extracted are, however, inconsistent with this possibility[44–46]. Lastly, similar to what has been noted at other sites[40,41,47], the Neandertal retouchers are made on fragments of dense bones with comparable mechanical properties to the horse and reindeer bones. At Goyet, as at several French Middle Palaeolithic sites, large bone fragments of medium and large-sized animals were selected[40,41,48–51]. Among the Goyet Neandertal material, the largest and thickest fragments were also selected, as was the case at Les Pradelles[16] and Krapina[42]. Interestingly, a femur and tibias of cave bears were also among the retoucher blanks selected by Neandertals at Scladina[52].

The observed patterns of faunal exploitation can be interpreted as the selective transport of meat and marrow rich elements to the site that were subsequently intensively processed. However, this apparent pattern may reflect a collection bias favoring the largest and most easily identifiable fragments. Similarities in anthropogenic marks observed on the Neandertal, horse and reindeer bones do, however, suggest similar processing and consumption patterns for all three species.

Discussion

Our results show that the Neandertals from the Troisième caverne of Goyet were butchered, with the hypothesis of their exploitation as food sources the most parsimonious explanation for the observed bone surface modifications. Goyet provides the first unambiguous evidence of Neandertal cannibalism in Northern Europe and given the dates obtained on the Neandertal remains, it is most likely that they were processed by their fellow Neandertals as no modern humans are known to have been in the region at the time[17,23]. However, the available data make it impossible to determine whether the modifications observed on the Neandertal skeletal material represent symbolic practices or simply result from the processing of immediately available sources of food. In addition, Goyet is the first site to have yielded multiple Neandertal bone retouchers. It has been proposed that Middle Palaeolithic retoucher blanks were by-products of the processing of carcasses for food consumption[40,41], which may have been selected to be re-used[51]. The data at hand do not allow us to propose a different scenario for the Goyet retouchers made on Neandertal bones. However, the freshness of the blanks used suggests that Neandertals may have been aware that they were using human remains. Whether this was part of a symbolic activity or induced by a functional motivation cannot be attested, as was the case for the La Quina Neandertal retoucher[43].

Although the Goyet late Neandertals date to 40.5–45.5 ky calBP, the lack of reliable contextual information makes it impossible to associate them with any of the technocomplexes from the site. However, coeval Mousterian assemblages are known from sites in the Mosan Basin, as at unit 1A of Scladina[53], located only 5 km from Goyet, layer CI-8 of Walou Cave[54], and layer II of Trou de l'Abîme at Couvin[55] (Supplementary Note S2). While the LRJ is known from two sites in Belgium, Spy and Goyet, with its first appearance dated at other sites to around 43–44 ky calBP[23,56]), no reliable information is currently available for its regional chronology. Given the direct ^{14}C dates obtained for the Goyet Neandertals, it is impossible to securely associate them with either the Mousterian occupation(s) or the LRJ.

In terms of the region's late Neandertal mortuary practices, four sites within an approximately 250 km radius around Goyet produced Neandertal remains reliably dated to between 50–40 ky calBP (Supplementary Fig. S1). Interestingly, none of these sites produced evidence for the treatment of the corpse similar to that documented for Goyet. Two Belgian sites, Walou Cave and Trou de l'Abîme, produced, respectively, a premolar and a molar[55,57]. Although impossible to infer the behavioural signature represented by these remains, given their state of preservation it is highly unlikely that they involved funerary practices, including burial. In Germany, the Neandertal individuals from Feldhofer, including Neandertal 1, are possibly associated with the "Keilmesser group", a late

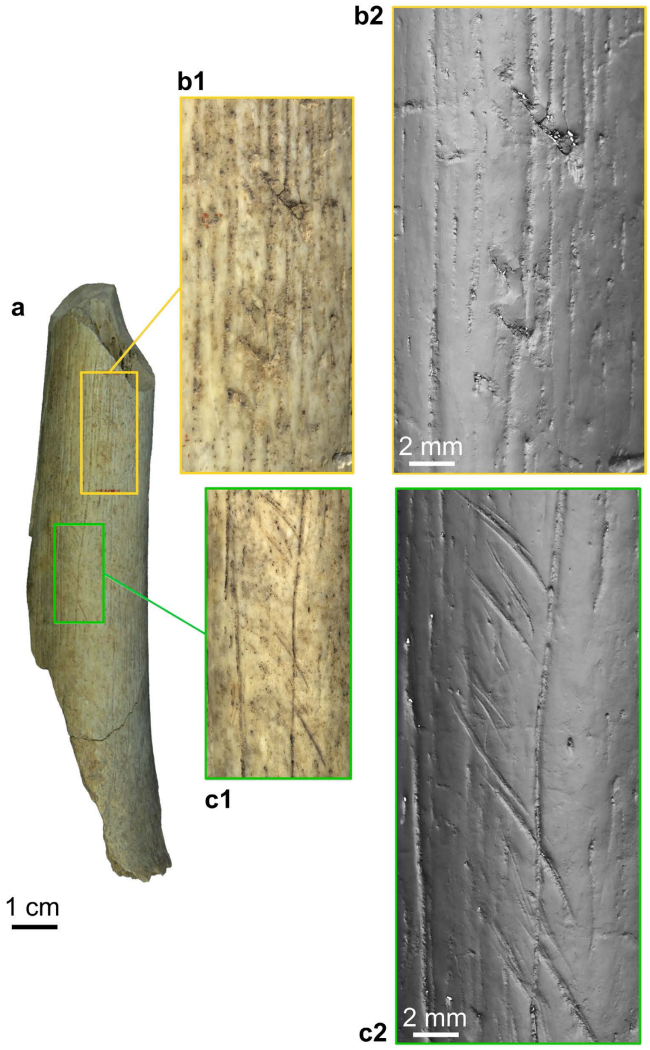


Figure 4. Retouching marks (b1,b2) and cutmarks (c1,c2) present on the Goyet Neandertal bones (example of femur III). (**a**) femur III in anterior view; (**b1**,**c1**) close-up photos; (**b2**,**c2**) images obtained using a minidome (see Methods).

Middle Palaeolithic technocomplex[58,59] unknown at Goyet (Supplementary Note S2). Neandertal 1 comprises elements of the cranial and postcranial skeleton of a single individual. Despite cutmarks on the cranium, clavicle and scapula, the long bones are intact and damage to still articulated skeletal elements during their recovery indicates that at least part of the skeleton may have originally been in anatomical connection[60,61]. Finally, at Spy, direct dates obtained on the two Neandertal adults place them within the current chronology of the LRJ[62], although the association between the human remains and this technocomplex is uncertain due to the lack of contextual information. A recent reassessment of the Spy specimens and their context suggests that both individuals were buried[63]. And, it is worth noting that the most complete individual, Spy II, was originally described as a complete skeleton found in a contracted position. Moreover, the completeness of the skeleton and the absence of post-depositional alterations suggest the body to have been rapidly protected[63].

Considerable diversity is evident in the mortuary behaviour of the late Neandertal populations of Northern Europe, possibly involving both primary and secondary deposits, alongside other types of practices, including cannibalism. Despite low genetic diversity amongst late Neandertal populations, the presence of various late Middle Palaeolithic technocomplexes, as well as the LRJ, nevertheless suggests significant behavioural variability amongst these groups in Northern Europe.

Methods

Collection assessment. The assessment of the Goyet collections included material housed at the RBINS and Royal Museums of Art and History (RMAH) in Brussels, which originate from the Troisième caverne, as well as collections from the *Grand Curtius* Museum (Liège), the *Cercle d'Histoire et d'Archéologie du Pays de Genappe* (Genappe), and the *Préhistosite de Ramioul* (Ramioul), whose origin is less secure. The Neandertal remains

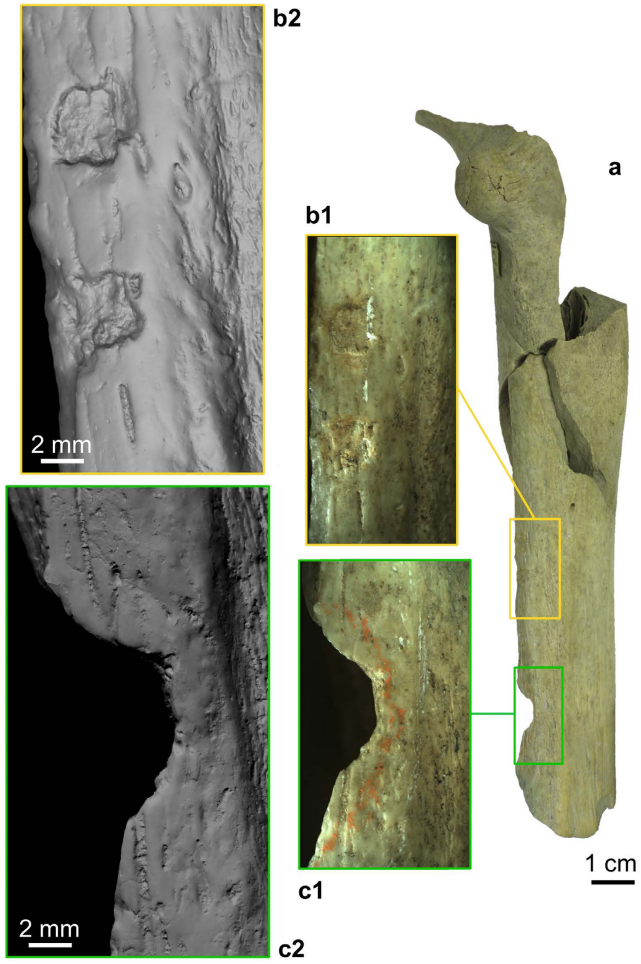


Figure 5. Percussion pits (b1,b2) and percussion notch (c1,c2) present on the Goyet Neandertal bones (example of femur I). (**a**) femur I in posterior view; (**b1**,**c1**) close-up photos; (**b2**,**c2**) images obtained using a minidome (see Methods).

presented here were found among the first two collections only. The numbering system of the specimens and their origin are discussed in Supplementary Note S4.

Taphonomic study. After determining the composition of the faunal assemblage sampled from Dupont's collection (Supplementary Table S5), a total of 442 horse and 287 reindeer remains were observed using a monocular microscope (×10), as were all of the Goyet Neandertal remains. Taphonomic and anthropogenic modifications were recorded and drawn on anatomical charts (Supplementary Figs S8 and S14–S16). Cutmarks and trampling marks were distinguished according to their morphology and placement on bones[64]. Only unambiguous notches with a negative flake scar[65,66] made on fresh bone[35] and percussion pits (left by impact events after ref. 66) were recorded as percussion marks. The identified bone retouchers are all long bone diaphysis fragments that exhibit marks as described by Mallye *et al.*[40]. Finally, toothmarks were recorded using Binford's typology[67]. Only pits and scores were observed. Some of these pits might have been produced by human chewing[38,67–69] but they are not characteristic enough to definitely distinguish them from marks left by carnivores. Following Bello *et al.*[70], the anthropogenic modifications recorded on the Neandertal remains were documented using drawings, close-up photographs and high-resolution imaging. The high-resolution images (Figs 4 and 5, and Supplementary Figs S8–S12) were obtained by using a minidome, a digital imaging device developed by VISICS at KULeuven (http://www.minidome.be). Based on the polynomial texture mapping technique, the dome consists of 260 LEDs and a single fixed camera, which captures an image with each LED individually lit. The results allow to display an object interactively under varying lighting to reveal all of the details of its surface. Additionally, 3D models of the retouchers made on Neandertal bones obtained using a white light 3D measurement system (http://www.mechscan.co.uk/) are available at http://virtualcollections.naturalsciences.be/virtual-collections/anthropology-prehistory/human-remains/goyet.

Sample selection and preparation for isotopic and genetic analyses. All sampled specimens were untreated (glued or varnished), newly identified Neandertal bones, except for tooth 2878–2D (see Table 1).

Specimens were scanned or μ-scanned and molded using DC-3481 silicone elastomer before sampling, with photos taken both before and after. Collagen was extracted at the Centre for Isotope Research of Groningen University (CIO, Netherlands) and the Biogeology working group of the Department of Geosciences of Tübingen University (Germany; see Supplementary Table S3). Radiocarbon dating was done at the CIO; stable isotope and genetic analyses were performed at Tübingen University.

Isotope analyses. Collagen extraction at the CIO followed the procedure developed by Longin[71], with additional chemical pretreatment using standard procedures[72]. Collagen extraction at Tübingen University followed a procedure modified from Longin[71] described by Bocherens *et al.*[73]. Stable isotopic measurements (^{13}C, ^{15}N) used an elemental analyser NC 2500 connected to a Thermo Quest Delta+XL mass spectrometer. The degree of chemical preservation of collagen is expressed as the atomic ratio of C_{coll}:N_{coll}, whose acceptable range of variation is 2.9–3.6[74], while the nitrogen content (N_{coll}) should be above 5%[75]. The carbon content of the extracted collagen ranges between 29.5 and 47.1% and nitrogen content between 10.1 and 17.0% (Supplementary Table S3), both of which fall in the range of fresh collagen[76]. The C_{coll}:N_{coll} atomic ratios span a narrow range (3.2–3.4), indicating exceptionally well-preserved collagen for all bone specimens (see Supplementary Note S6 for discussion of tooth 2878–2D). Subsequently, the collagen was sent to the CIO for ^{14}C dating by AMS[77].

Genetic analyses. Ten specimens were sampled (Table 1). DNA was extracted[78] from bone powder and converted to double-indexed genetic libraries[79,80]. Mitochondrial DNA (mtDNA) was enriched using a bait capture technique[81] and sequenced on a next generation sequencing platform (Illumina HiSeq). After quality filtering and merging paired-end reads[82], a modified version of the BWA mapper and the SAMtools package in combination with a custom iterative mapping assembler[30,83–85] were used to align reads to a reference Neandertal mtDNA sequence (Supplementary Table S7). Reads from the three low coverage specimens were also aligned to the modern human mtDNA reference sequence (rCRS) in order to exclude reference biases (Supplementary Table S8). The newly reconstructed complete or almost complete (i.e. at least 98% complete) mitochondrial genomes were compared with 63 other hominin mtDNA sequences in gene trees (Fig. 2 and Supplementary Fig. S4) to assess their phylogenetic placements and intergroup genetic relationships (Supplementary Note S7). The authenticity of the obtained mitochondrial sequences as endogenous ancient DNA was verified by analysing typical ancient DNA damage patterns (ref. 86 and Supplementary Fig. S17) as well as estimating the percentage of modern human DNA contamination (ref. 30 and Supplementary Table S7). Finally, damaged DNA molecules indicating an ancient origin were filtered[34] and used to build new mtDNA consensus sequences. These were co-analysed with the same 63 mtDNAs in order to validate the assigned phylogenetic placement (Supplementary Figs S18 and S19).

References

1. Rendu, W. *et al.* Evidence supporting an intentional Neandertal burial at La Chapelle-aux-Saints. *Proc. Natl. Acad. Sci. USA* **111,** 81–86 (2014).
2. Dibble, H. L. *et al.* A critical look at evidence from La Chapelle-aux-Saints supporting an intentional Neandertal burial. *J. Archaeol. Sci.* **53,** 649–657 (2015).
3. Sandgathe, D. M., Dibble, H. L., Goldberg, P. & McPherron, S. P. The Roc de Marsal Neandertal child: A reassessment of its status as a deliberate burial. *J. Hum. Evol.* **61,** 243–253 (2011).
4. Bachellerie, F. *et al.* Preliminary results from ongoing excavations at La Roche à Pierrot, Saint-Césaire. *XVII World UISPP Congress* (Burgos, Sept. 1–7, 2014).
5. Gómez-Olivencia, A., Crevecoeur, I. & Balzeau, A. La Ferrassie 8 Neandertal child reloaded: New remains and re-assessment of the original collection. *J. Hum. Evol.* **82,** 107–126 (2015).
6. Pettitt, P. B. The Neanderthal dead: exploring mortuary variability in Middle Palaeolithic Eurasia. *Before Farming* **2002/1,** 1–26 (2002).
7. Gorjanović-Kramberger, K. Der paläolithische Mensch und seine Zeitgenossen aus dem Diluvium von Krapina in Kroatien I. *Mitt. Anthropol. Gesell. Wien* **31,** 164–197 (1901).
8. Orschiedt, J. Der Fall Krapina – neue Ergebnisse zur Frage von Kannibalismus beim Neandertaler. *Quartär* **55,** 63–81 (2008).
9. Villa, P. *et al.* Cannibalism in the Neolithic. *Science* **233,** 431–437 (1986).
10. White, T. D. *Prehistoric Cannibalism at Mancos 5MTUMR-2346* (Princeton University Press, 1992).
11. Boulestin, B. Approche Taphonomique des Restes Humains. Le Cas des Mésolithiques de la Grotte des Perrats et le Problème du Cannibalisme en Préhistoire Récente Européenne (BAR IS 776, Archaeopress, 1999).
12. Fernández-Jalvo, Y., Díez, J. C., Cáceres, I. & Rosell, J. Human cannibalism in the Early Pleistocene of Europe (Gran Dolina, Sierra de Atapuerca, Burgos, Spain). *J. Hum. Evol.* **37,** 591–622 (1999).
13. Rosas, A. *et al.* Paleobiology and comparative morphology of a late Neandertal sample from El Sidrón, Asturias, Spain. *Proc. Natl. Acad. Sci. USA* **103,** 19266–19271 (2006).
14. Barroso Ruiz, C. *et al.* In *El Pleistoceno Superior de la Cueva del Boquete de Zafarraya*. (coord Barroso Ruiz, C.) 389–419 (Archeología Monografías, Junta de Andalucía, 2003).
15. Defleur, A., White, T., Valensi, P., Slimak, L. & Crégut-Bonnoure, E. Neanderthal cannibalism at Moula-Guercy, Ardèche, France. *Science* **286,** 128–131 (1999).
16. Mussini, C. Les Restes Humains Moustériens des Pradelles (Marillac-le-Franc, Charente, France): Etude Morphométrique et Réflexions sur un Aspect Comportemental des Néandertaliens. Thèse de doctorat (Université Bordeaux 1, 2011).
17. Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512,** 306–309 (2014).
18. Toussaint, M. In *Neanderthals in Europe.* (dirs Demarsin, B. & Otte, M.) 115–134 (ERAUL 117, ATVATVCA 2, 2006).
19. Dupont, E. L'Homme Pendant les Ages de la Pierre dans les Environs de Dinant-sur-Meuse, 2ème Edition (Muquardt, C. Ed., 1872).
20. Ulrix-Closset, M. *Le Paléolithique Moyen dans le Bassin Mosan en Belgique* (Universa, 1975).
21. Otte, M. Le Paléolithique supérieur ancien en Belgique. *Monogr. Archéol. Nat*. **5,** 1–684 (1979).
22. Dewez, M. *Le Paléolithique Supérieur Récent dans les Grottes de Belgique* (Université catholique de Louvain, 1987).
23. Flas, D. The Middle to Upper Paleolithic transition in Northern Europe: the Lincombian-Ranisian-Jerzmanowician and the issue of acculturation of the last Neanderthals. *World Archaeol*. **43,** 605–627 (2011).
24. Germonpré, M. A reconstruction of the spatial distribution of the faunal remains from Goyet, Belgium. *Notae Praehistoricae* **21,** 57–65 (2001).

25. Hamy, E.-T. Sur quelques ossements humains découverts dans la troisième caverne de Goyet, près Namèche (Belgique). *Bull. Soc. Anthrop. Paris* **8,** 425–435 (1873).
26. Twiesselmann, F. In *Catalogue of Fossil Hominids - Part 2: Europe*. (eds Oakley, K. P., Campbell, B. G. & Molleson, T. I.) 5–13 (Trustees of the British Museum of Natural History, 1971).
27. Rougier, H., Crevecoeur, I., Semal, P. & Toussaint, M. In *Paléolithique Moyen en Wallonie. La Collection Louis Eloy*. (eds Di Modica, K. & Jungels, C.) 173 (Collections du Patrimoine culturel de la Communauté française 2, 2009).
28. Wißing, C. *et al.* Isotopic evidence for dietary ecology of Late Neandertals in North-Western Europe. *Quat. Int.* doi: 10.1016/j.quaint.2015.09.091 (2015).
29. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708–713 (2000).
30. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134,** 416–426 (2008).
31. Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325,** 318–321 (2009).
32. Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464,** 894–897 (2010).
33. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
34. 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).
35. Villa, P. & Mahieu, E. Breakage patterns of human long bones. *J. Hum. Evol.* **21,** 27–48 (1991).
36. Arilla, M., Rosell, J., Blasco, R., Domínguez-Rodrigo, M. & Pickering, T. R. The "bear" essentials: Actualistic research on *Ursus arctos arctos* in the Spanish Pyrenees and its implications for paleontology and archaeology. *PLoS ONE* **9,** e102457 (2014).
37. Bello, S. M., Saladié, P., Cáceres, I., Rodríguez-Hidalgo, A. & Parfitt, S. A. Upper Palaeolithic ritualistic cannibalism at Gough's Cave (Somerset, UK): The human remains from head to toe. *J. Hum. Evol.* **82,** 170–189 (2015).
38. White, T. D. & Toth, N. In *Breathing Life into Fossils: Taphonomic Studies in Honor of C.K.* (*Bob*) *Brain*. (eds Pickering, T. R., Schick, K. & Toth, N.) 281–296 (Stone Age Institute Press, 2007).
39. Patou-Mathis, M. ed. In *Fiches de la Commission de Nomenclature sur l'Industrie de l'Os Préhistorique, X.* 11–19 (Société Préhistorique Française, 2002).
40. Mallye, J.-B. *et al.* The Mousterian bone retouchers of Noisetier Cave: experimentation and identification of marks. *J. Archaeol. Sci.* **39,** 1131–1142 (2012).
41. Daujeard, C. *et al.* Middle Paleolithic bone retouchers in Southeastern France: Variability and functionality. *Quat. Int.* **326–327,** 492–518 (2014).
42. Patou-Mathis, M. Analyse taphonomique et paléoethnographique du matériel osseux de Krapina (Croatie): nouvelles données sur la faune et les restes humains. *Préhistoire Européenne* **10,** 63–90 (1997).
43. Verna, C. & d'Errico, F. The earliest evidence for the use of human bone as a tool. *J. Hum. Evol.* **60,** 145–157 (2011).
44. Villa, P. *et al.* Un cas de cannibalisme au Néolithique. *Gallia préhistoire* **29,** 143–171 (1986).
45. Abe, Y. Hunting and Butchery Patterns of the Evenki in the Northern Transbaikalia, Russia. PhD thesis (Stony Brook University, 2005).
46. Thompson, T. *The Archaeology of Cremation: Burned Human Remains in Funerary Studies* (Oxbow Books, 2015).
47. Van Kolfschoten, T., Parfitt, S. A., Serangeli, J. & Bello, S. M. Lower Paleolithic bone tools from the 'Spear Horizon' at Schöningen (Germany). *J. Hum. Evol.* **89,** 226–263 (2015).
48. Beauval, C. In *Le Site Paléolithique de Chez-Pinaud à Jonzac, Charente-Maritime*. (ed Airvaux, J.) 125–156 (Préhistoire du Sud Ouest Suppl. 8, 2004).
49. Jaubert, J. *et al.* In *Les Sociétés du Paléolithique dans un Grand Sud-Ouest de la France: Nouveaux Gisements, Nouveaux Résultats, Nouvelles Méthodes. Journées SPF, Université Bordeaux 1, Talence, 24–25 nov. 2006*. (dirs Jaubert, J., Bordes, J.-G. & Ortega, I.) 203–243 (Mémoire XLVII de la Société Préhistorique Française, 2008).
50. Costamagno, S., Meignen, L., Beauval, C., Vandermeersch, B. & Maureille, B. Les Pradelles (Marillac-le-Franc, France): A mousterian reindeer hunting camp? *J. Anthropol. Archaeol.* **25,** 466–484 (2006).
51. Rosell, J. *et al.* Recycling bones in the Middle Pleistocene: Some reflections from Gran Dolina TD10-1 (Spain), Bolomor Cave (Spain) and Qesem Cave (Israel). *Quat. Int.* **361,** 297–312 (2015).
52. Abrams, G., Bello, S. M., Di Modica, K., Pirson, S. & Bonjean, D. When Neanderthals used cave bear (*Ursus spelaeus*) remains: Bone retouchers from unit 5 of Scladina Cave (Belgium). *Quat. Int.* **326–327,** 274–287 (2014).
53. Di Modica, K. Les Productions Lithiques du Paléolithique Moyen en Belgique: Variabilité des Systèmes d'Acquisition et des Technologies en Réponse à une Mosaïque d'Environnements Contrastés. Thèse de doctorat (MNHN, Université de Liège, 2010).
54. Pirson, S. *et al.* Chronostratigraphic context of the Middle to Upper Palaeolithic transition: Recent data from Belgium. *Quat. Int.* **259,** 78–94 (2012).
55. Toussaint, M. *et al.* The Neandertal lower right deciduous second molar from Trou de l'Abîme at Couvin, Belgium. *J. Hum. Evol.* **58,** 56–67 (2010).
56. Cooper, L. P. *et al.* An Early Upper Palaeolithic open-air station and Mid-Devensian hyaena den at Grange Farm, Glaston, Rutland, UK. *Proc. Prehist. Soc.* **78,** 73–93 (2012).
57. Toussaint, M. In *La Grotte Walou à Trooz* (*Belgique*). *Fouilles de 1996 à 2004. Volume 2*. (dirs Draily, C., Pirson, S. & Toussaint, M.) 148–161 (Etudes et Documents – Archéologie 21, 2011).
58. Schmitz, R. W. *et al.* The Neandertal type site revisited: Interdisciplinary investigations of skeletal remains from the Neander valley, Germany. *Proc. Natl. Acad. Sci. USA* **99,** 13342–13347 (2002).
59. Hillgruber, K. F. In *Neanderthal 1856–2006*. (ed. Schmitz, R.) 111–144 (Rheinische Ausgrabungen 58, 2006).
60. Schmitz, R. W. & Pieper, P. Schnittspuren und Kratzer. Anthropogene Veränderungen am Skelett des Urmenschenfundes aus dem Neandertal - Vorläufige Befundaufnahme. *Rheinisches Landesmuseum Bonn* **2,** 17–19 (1992).
61. Auffermann, B. & Orschiedt, J. *Die Neandertaler, eine Spurensuche* (Konrad Theiss Verlag, 2002).
62. Semal, P. *et al.* New data on the Late Neandertals: Direct dating of the Belgian Spy fossils. *Am. J. Phys. Anthrop.* **138,** 421–428 (2009).
63. Maureille, B., Toussaint, M. & Semal, P. In *Spy Cave. 125 Years of Multidisciplinary Research at the Betche aux Rotches* (*Jemeppe-sur-Sambre, Province of Namur, Belgium*), *Volume 2*. (eds Rougier, H. & Semal, P.) (Anthropologica et Praehistorica 124/2013, 2014).
64. Behrensmeyer, A. K., Gordon, K. D. & Yanagi, G. T. Trampling as a cause of bone surface damage and pseudo-cutmarks. *Nature* **319,** 768–771 (1986).
65. Blumenschine, R. J. & Selvaggio, M. M. Percussion marks on bone surfaces as a new diagnostic of hominid behaviour. *Nature* **333,** 763–765 (1988).
66. Pickering, T. R. & Egeland, C. P. Experimental patterns of hammerstone percussion damage on bones: implications for inferences of carcass processing by humans. *J. Archaeol. Sci.* **33,** 459–469 (2006).
67. Binford, L. R. *Bones: Ancient Men and Modern Myths* (Academic Press, 1981).
68. Binford, L. R. *Nunamiut Ethnoarchaeology* (Academic Press, 1978).
69. Fernández-Jalvo, Y. & Andrews, P. When humans chew bones. *J. Hum. Evol.* **60,** 117–123 (2011).
70. Bello, S. M., Parfitt, S. A., De Groote, I. & Kennaway, G. Investigating experimental knapping damage on an antler hammer: a pilot-study using high-resolution imaging and analytical techniques. *J. Archaeol. Sci.* **40,** 4528–4537 (2013).

71. Longin, R. *Extraction du Collagène des Os Fossiles pour leur Datation par la Méthode du Carbone 14*. Thèse de doctorat (Université de Lyon, 1970).
72. Mook, W. G. & Streurman, H. J. Physical and chemical aspects of radiocarbon dating. *PACT Publications* **8,** 31–55 (1983).
73. Bocherens, H. *et al.* Paleobiological implications of the isotopic signatures (13C, 15N) of fossil mammal collagen in Scladina Cave (Sclayn, Belgium). *Quaternary Res.* **48,** 370–380 (1997).
74. DeNiro, M. J. Postmortem preservation and alteration of *in vivo* bone collagen isotope ratios in relation to palaeodietary reconstruction. *Nature* **317,** 806–809 (1985).
75. Ambrose, S. H. Preparation and characterization of bone and tooth collagen for isotopic analysis. *J. Archaeol. Sci.* **17,** 431–451 (1990).
76. Rodière, E., Bocherens, H., Angibault, J. M. & Mariotti, A. Isotopic particularities of nitrogen in roe-deer (*Capreolus capreolus* L.): implications for palaeoenvironmental reconstructions. *C. R. Acad. Sci. IIa* **323,** 179–185 (1996).
77. van der Plicht, J., Wijma, S., Aerts, A. T., Pertuisot, M. H. & Meijer, H. A. J. Status report: The Groningen AMS facility. *Nucl. Instrum. Meth. B* **172,** 58–65 (2000).
78. 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).
79. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **6,** pdb.prot5448 (2010).
80. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40,** e3 (2012).
81. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5,** e14004 (2010).
82. Kircher, M. In *Ancient DNA. Methods and Protocols.* (eds Shapiro, B. & Hofreiter, M.) 197–228 (Humana Press, 2012).
83. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
84. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
85. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17,** 60 (2016).
86. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104,** 14616–14621 (2007).

Acknowledgements

We are indebted to the curators who granted us access to the collections in their care: A. Folie & I. Jadin (RBINS), N. Cauwe & L. Paquay (RMAH), J.-L. Schütz (*Grand Curtius* Museum), B. Clarys (CHAPG), and C. Jungels (*Préhistosite de Ramioul*). We thank A. & G. Jourdan, E. Dewamme & A. Mathys for the photos and casts of the Goyet remains, I. Jadin for his help with the RBINS archives, E. Morin and three anonymous reviewers for their useful comments and suggestions. This research was funded by the Wenner-Gren Foundation (Gr. 7837 to H.R.), the College of Social and Behavioral Sciences of CSUN, the CSUN Probationary Faculty Support Program, and the RBINS. A.G.-O. was supported by a Marie Curie-IEF research fellowship, the research group IT834-13 (Eusko Jaurlaritza/Gobierno Vasco), and the Spanish Ministerio de Economía y Competitividad (projects CGL2012-38434-C03-01 and CGL2015-65387-C3-2-P -MINECO/FEDER-). J.K. and C.P. were supported by the DFG grant KR 4015/1-1 and the Baden Württemberg Foundation.

Author Contributions

H.R. and I.C. designed the research. All authors performed the research. H.R., I.C., C.B., C.P., D.F., C.W., M.G., A.G.-O., P.S., J.v.d.P., H.B. and J.K. analysed the data. H.R., I.C., C.B., C.P., D.F., C.W., M.G., H.B. and J.K. wrote the paper with input from all authors.

Additional Information

Accession codes: The seven Neandertal mtDNA sequences reported in this article were deposited in GenBank and are available under accession numbers KX198082-KX198088.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

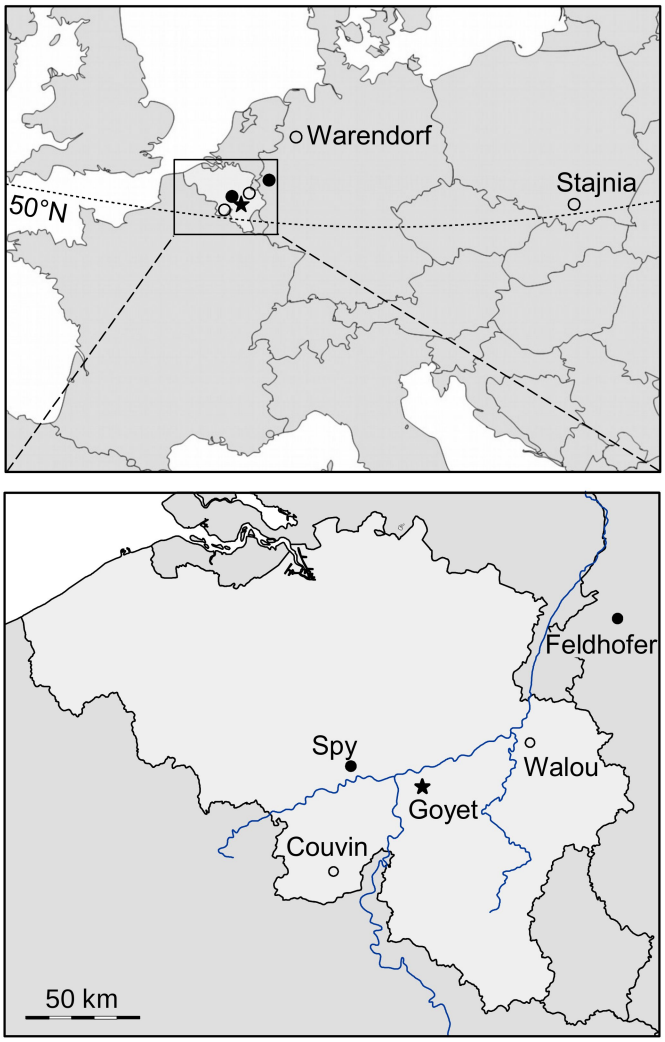
How to cite this article: Rougier, H. *et al.* Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci. Rep.* **6**, 29005; doi: 10.1038/srep29005 (2016).

Neandertal cannibalism and Neandertal bones used as tools in Northern Europe – Supplementary Information

Hélène Rougier, Isabelle Crevecoeur, Cédric Beauval, Cosimo Posth, Damien Flas, Christoph Wißing, Anja Furtwängler, Mietje Germonpré, Asier Gómez-Olivencia, Patrick Semal, Johannes van der Plicht, Hervé Bocherens & Johannes Krause

Table of Contents

Supplementary Figures



Supplementary Fig. S1. Map of Northern Europe centred on Belgium (top) and detail of the central area (bottom) showing the location of Goyet and other sites above 50° north that yielded late Neandertal remains.

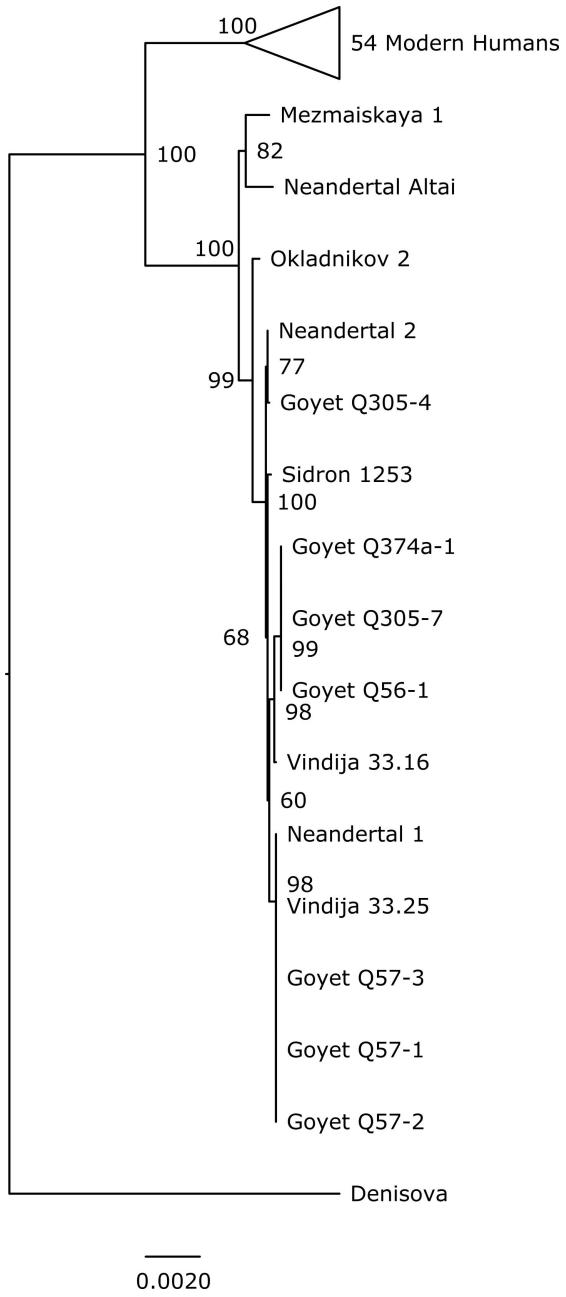
Black and white circles indicate, respectively, sites with directly- and indirectly-dated Neandertal remains. Map created using LibreOffice 5.0 Draw and exported in PDF 600 DPI (http://www.libreoffice.org/); Tiff file generated from the PDF using GIMP 2.8 (https://www.gimp.org).

Supplementary Fig. S2. The 21 drawers of fragmentary, "indeterminate" fauna from Dupont's excavations at the Troisième caverne of Goyet that were systematically sorted in order to identify any overlooked human remains.

The comparative taphonomic analysis of the Goyet Neandertal remains was conducted on the faunal remains identified in drawers Q53, Q55, Q375, and Q376 (see Supplementary Table S5).

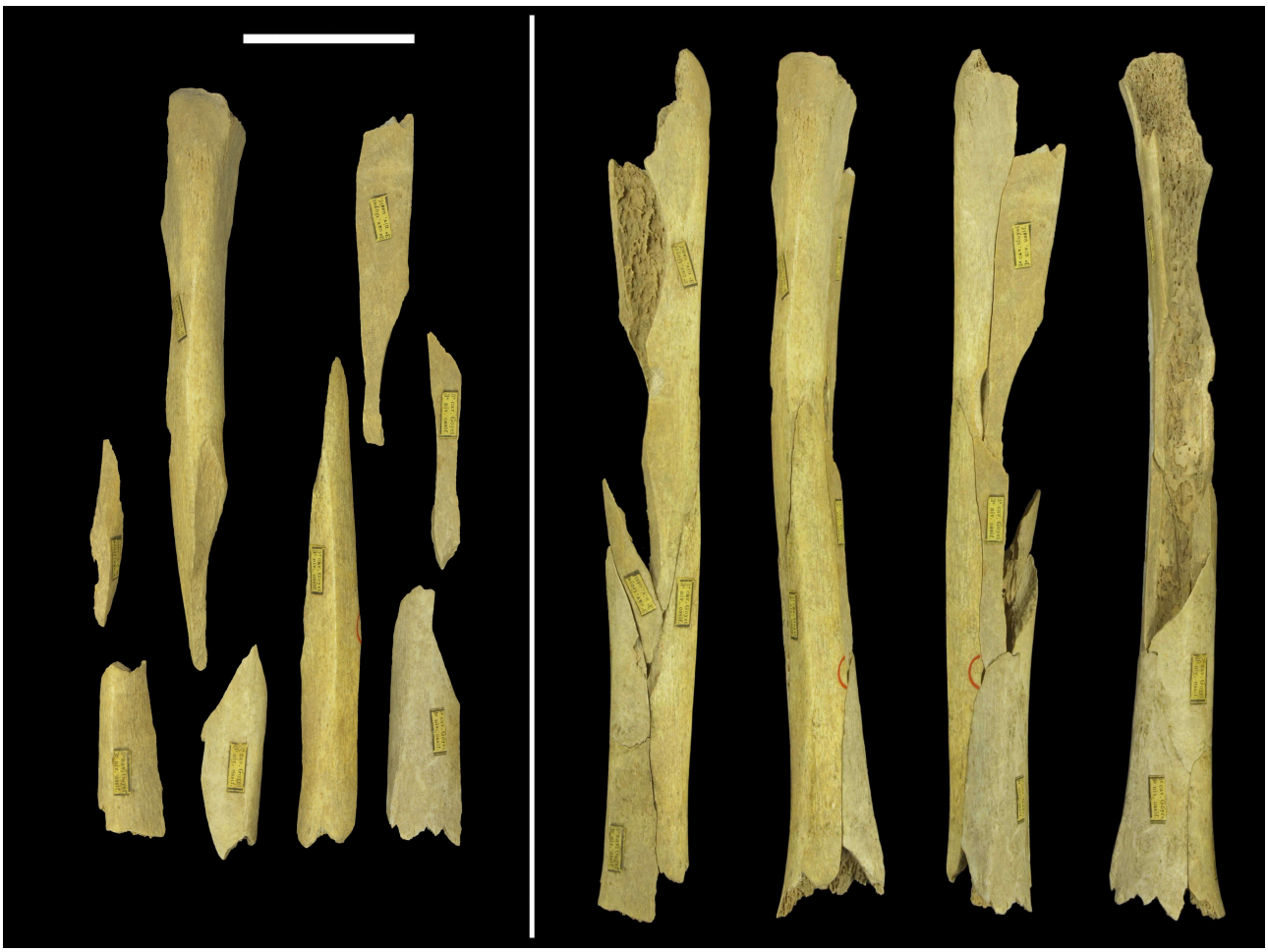
Supplementary Fig. S3. Left and right tibia pieces and lower left lateral incisor used to estimate the Neandertal MNI for Goyet.

Anterior (left) and posterior (right) view of each tibia piece; lingual (left) and distal (right) view of the lower left lateral incisor. Tibias I and II and Q375-2 are from the left side; all of the other tibia pieces are from the right side. Scale = 3 cm for the bone pieces and 1 cm for the tooth.



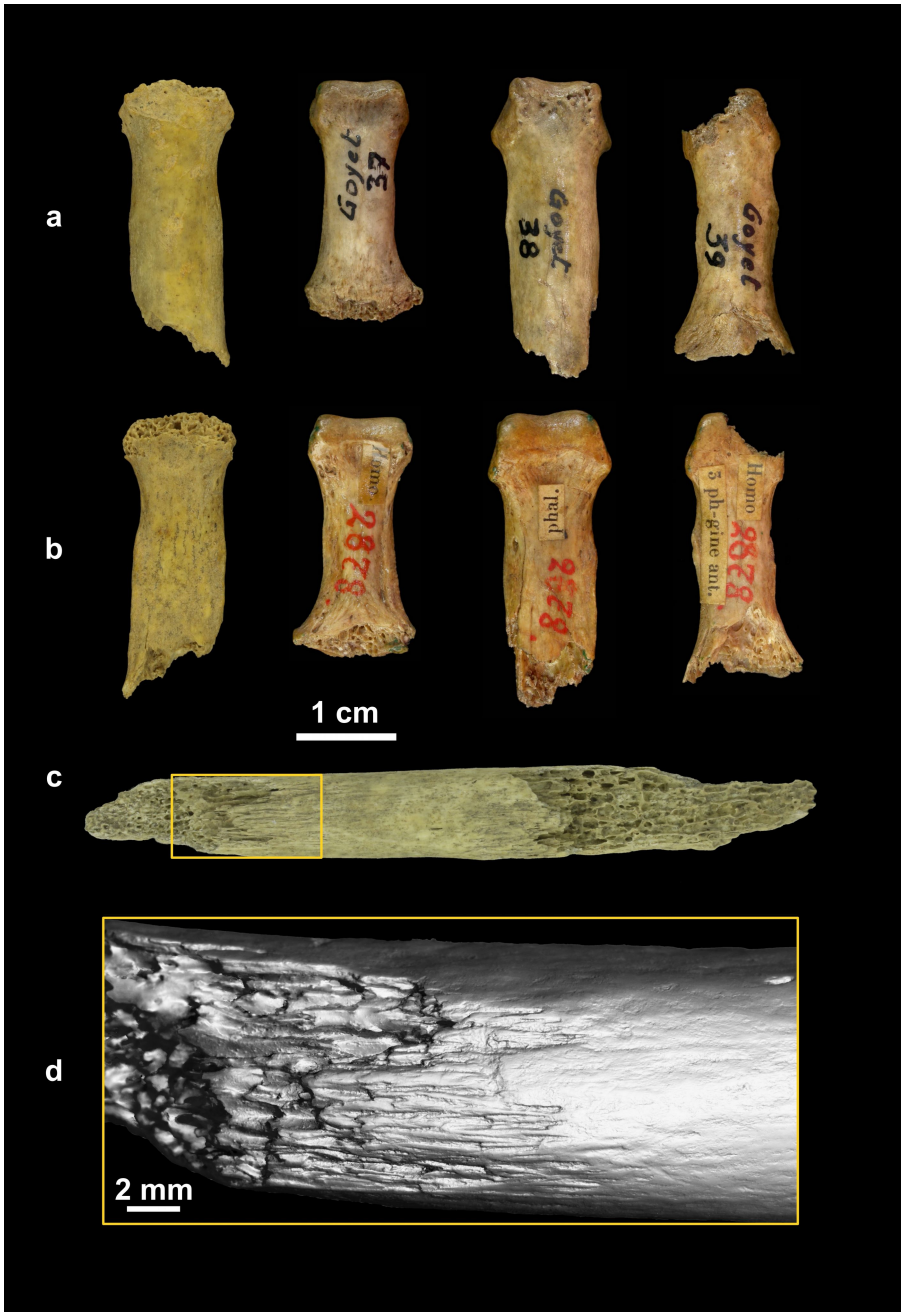
Supplementary Fig. S4. Maximum likelihood tree for the seven analysed Goyet samples that produced complete or almost complete mitochondrial genomes compared to 63 published modern human, Neandertal and Denisovan mtDNAs.

Numbers at the main branch nodes represent bootstrap values after 1,000 iterations.



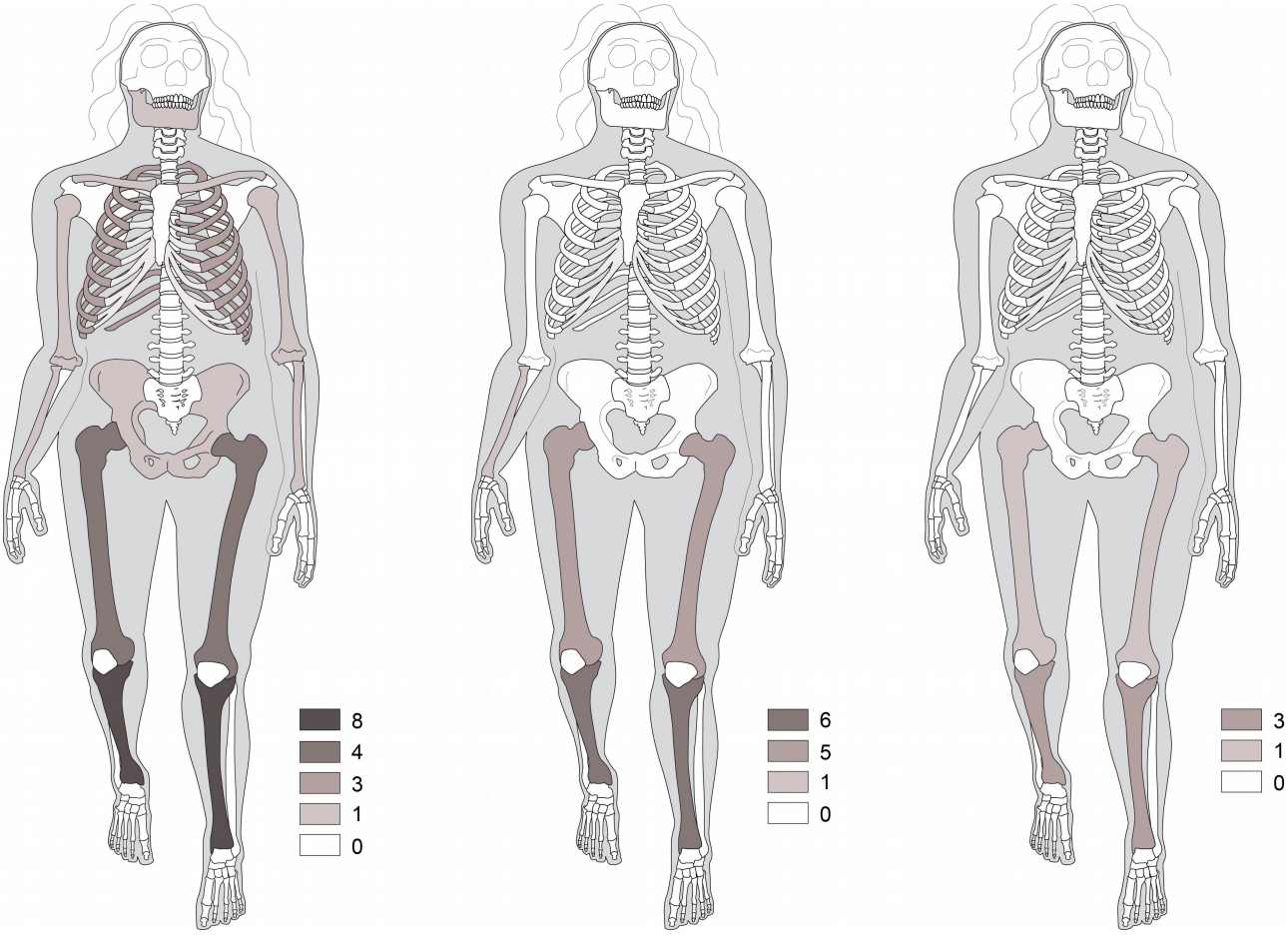
Supplementary Fig. S5. Tibia I, the most complete refit piece (8 refits) from the Neandertal assemblage.

Left: individual specimens; right: refit piece (from left to right: in medial, anterior, lateral and posterior views). Note that all of the specimens were found mixed with fauna from E. Dupont excavations and a small yellow label indicating their stratigraphic origin was glued to each at the beginning of the 20th century; red traces were also drawn on likely faunal fragments to delimit impact notches and retoucher areas. Scale = 5 cm.



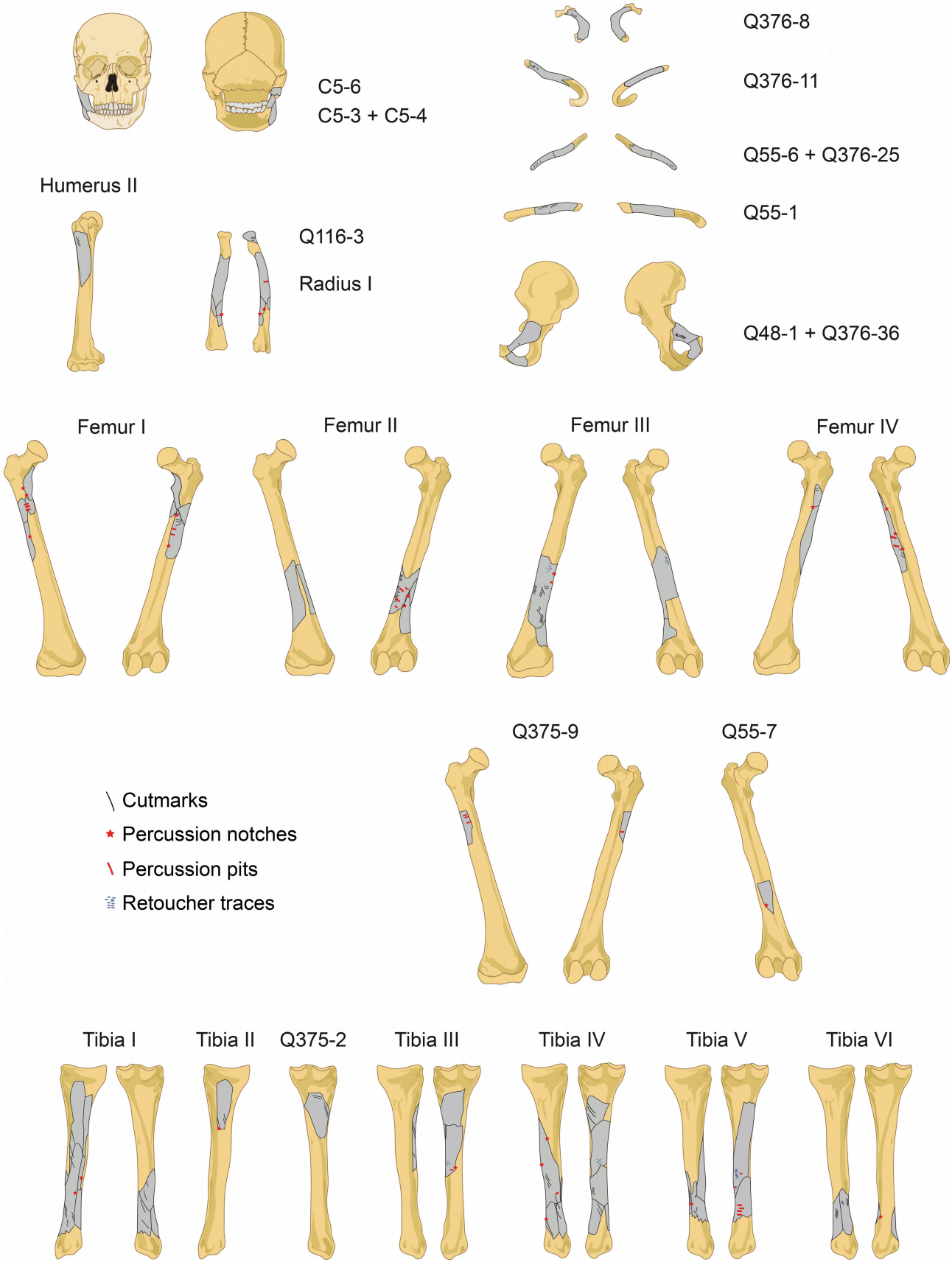
Supplementary Fig. S6. The Neandertal hand phalanges Q376-1, 2878-37, 2878-38, and 2878-39 (from left to right) in dorsal (a) and palmar (b) view and the Neandertal rib Q376-26 in posterior view (c) showing traces of peeling (d).

Image d was obtained using a minidome (see Methods).



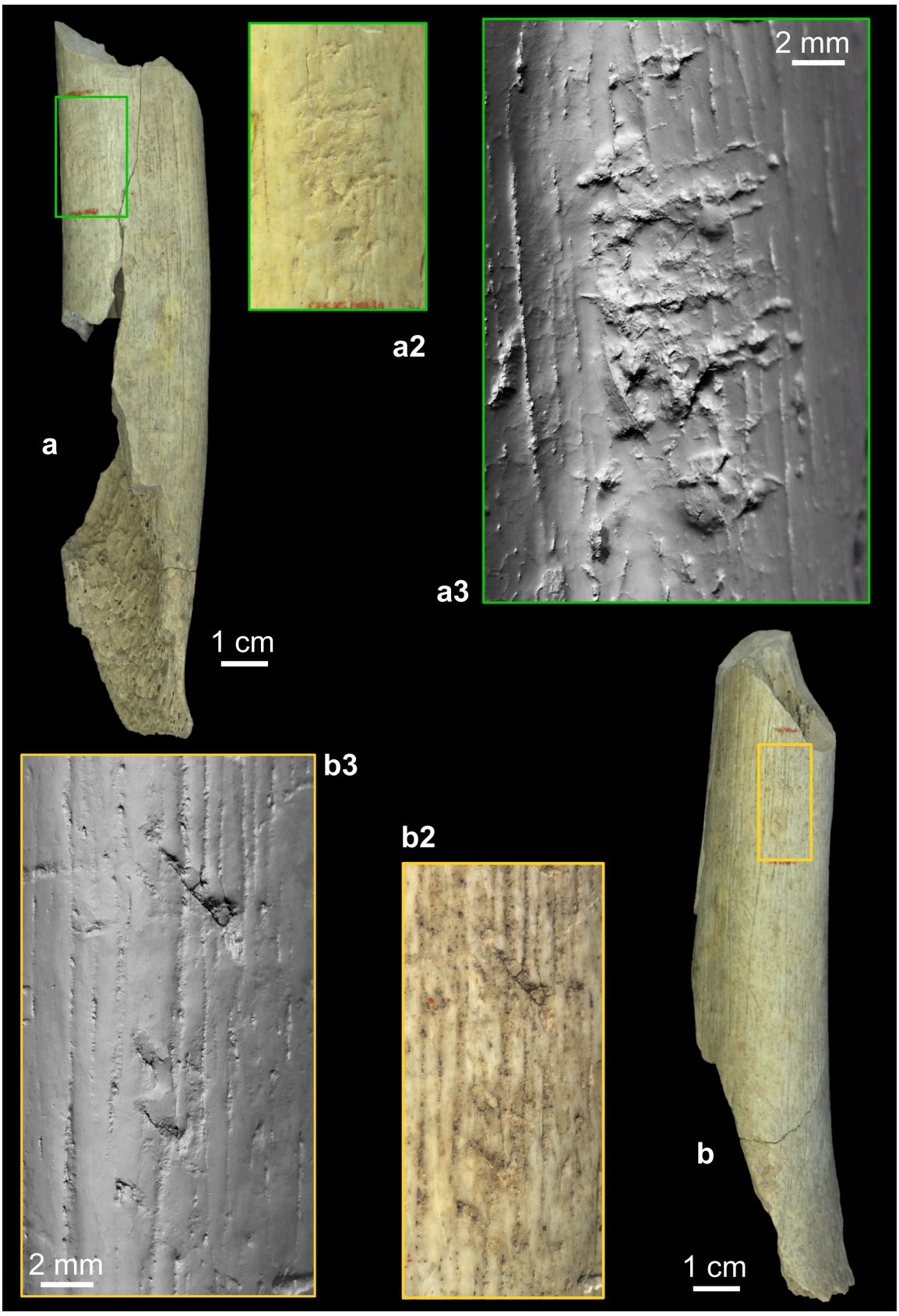
Supplementary Fig. S7. Number of Neandertal bones (after refitting) with cutmarks (left), percussion notches (centre) and retoucher traces (right).

Neandertal diagram modified from http://archeozoo.org/archeozootheque/index/category/102-hominides_langen_hominid_lang_langes_hominidos_lang_ (diagram by C.B–© 2013 ArcheoZoo.org, after ref. 43) using Adobe Illustrator CS4 v. 14.0.0.



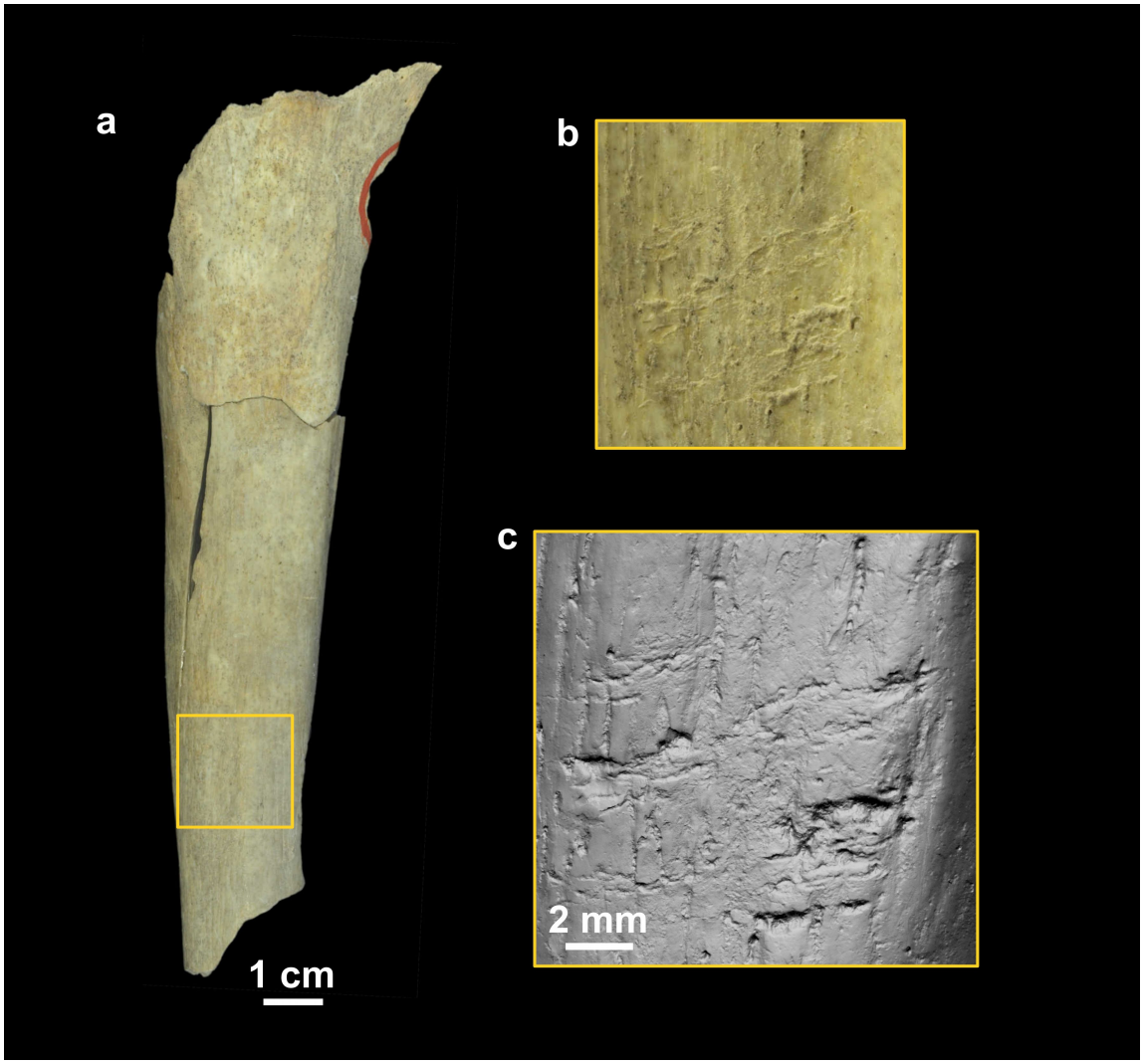
Supplementary Fig. S8. Drawings of the Goyet Neandertal remains bearing anthropogenic modifications.

Bone identifications are given in Supplementary Table S2. Grey areas indicate preserved bone portions. Sufficiently preserved elements are shown in both anterior (left) and posterior (right) views. Bone diagrams modified from https://en.wikipedia.org/wiki/File:Human_skeleton_front_en.svg and https://en.wikipedia.org/wiki/File:Human_skeleton_back_en.svg using Adobe Illustrator CS4 v. 14.0.0.



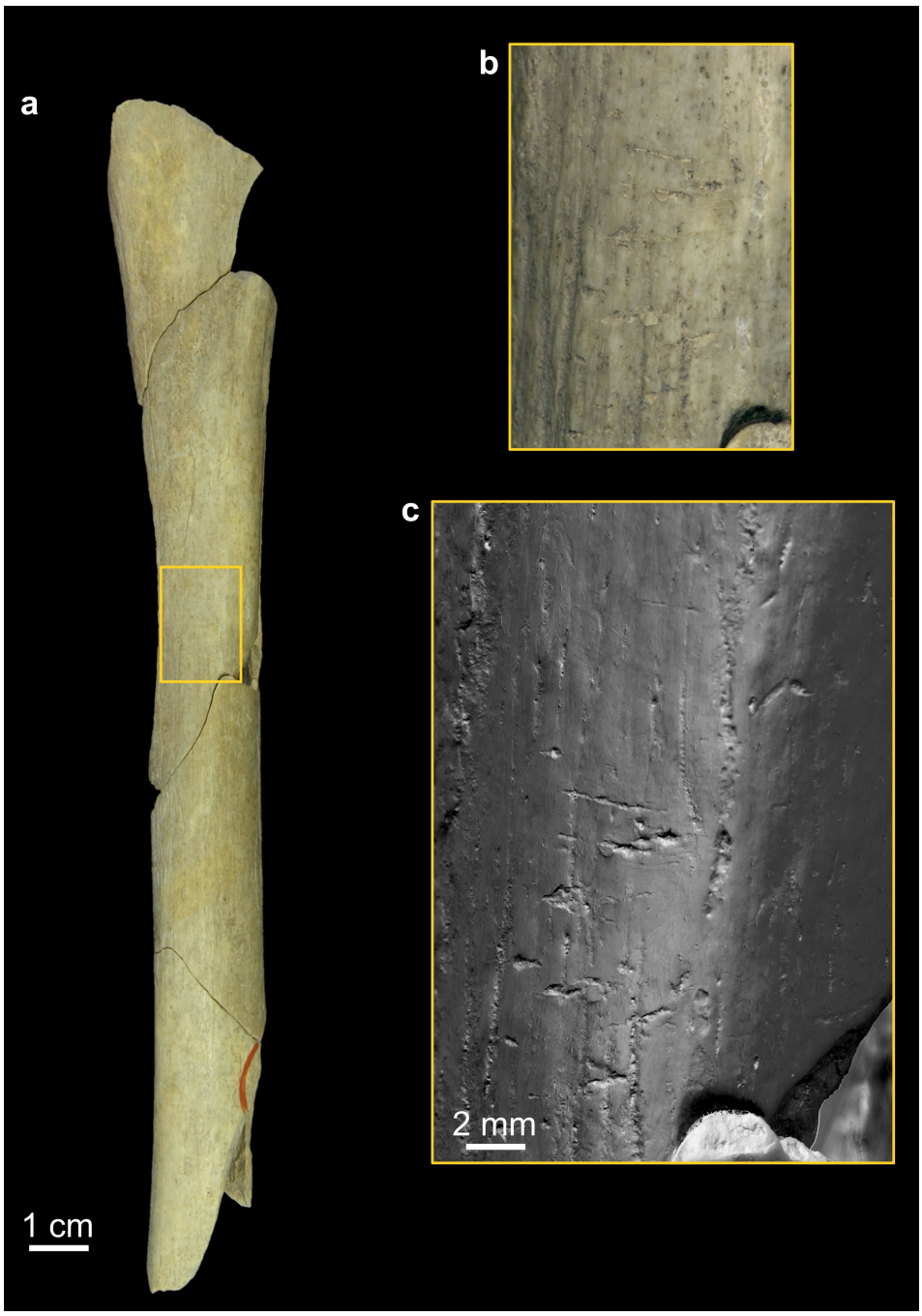
Supplementary Fig. S9. Femur III in medial (a) and anterior (b) view and details of the two areas of the bone used as retouchers.

a2 and b2: close-up photos of the areas showing retouching marks; a3 and b3: images of the areas showing retouching marks obtained using a minidome, see Methods.



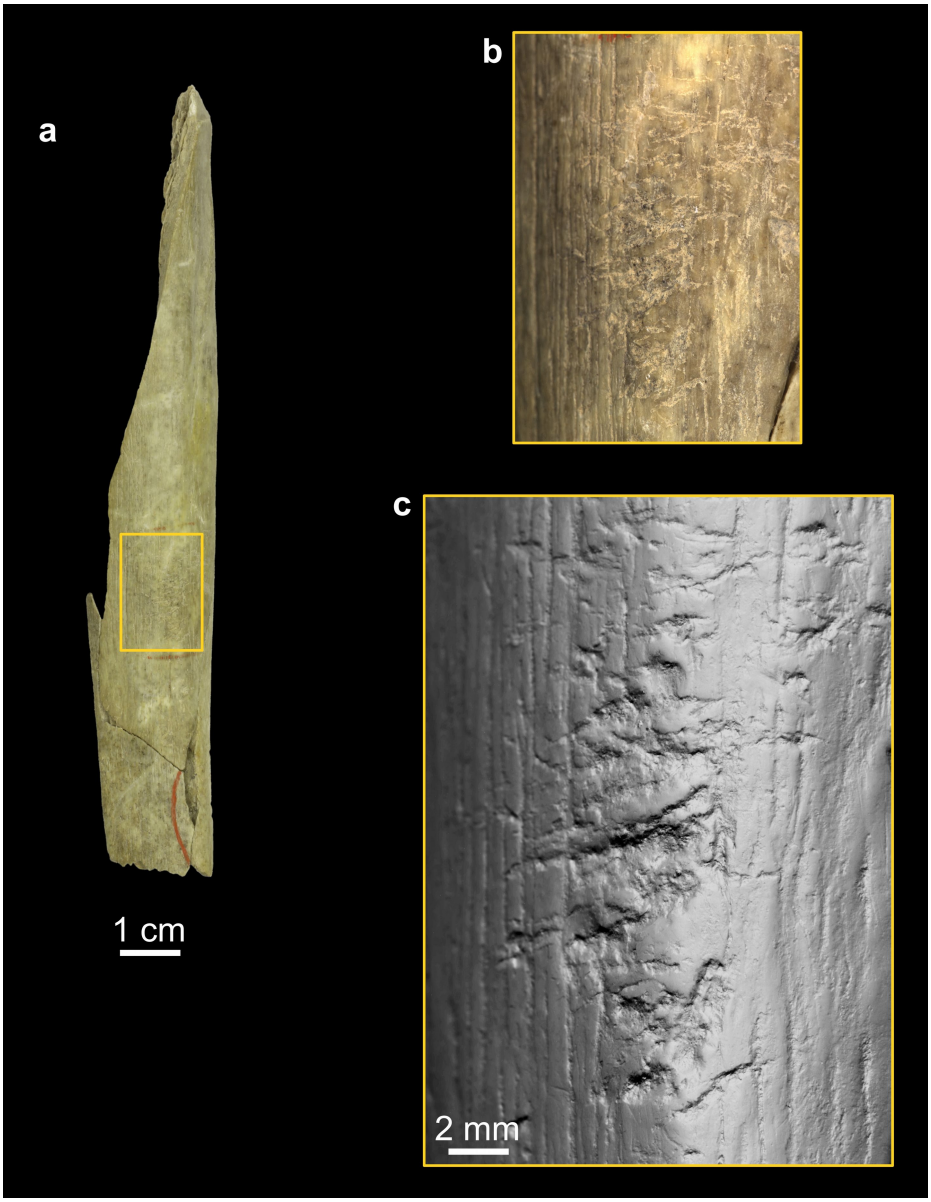
Supplementary Fig. S10. Tibia III in posterior view (a) and details of the area of the bone used as a retoucher (b and c).

b: close-up photo of the area showing retouching marks; c: image of the area showing retouching marks obtained using a minidome, see Methods.



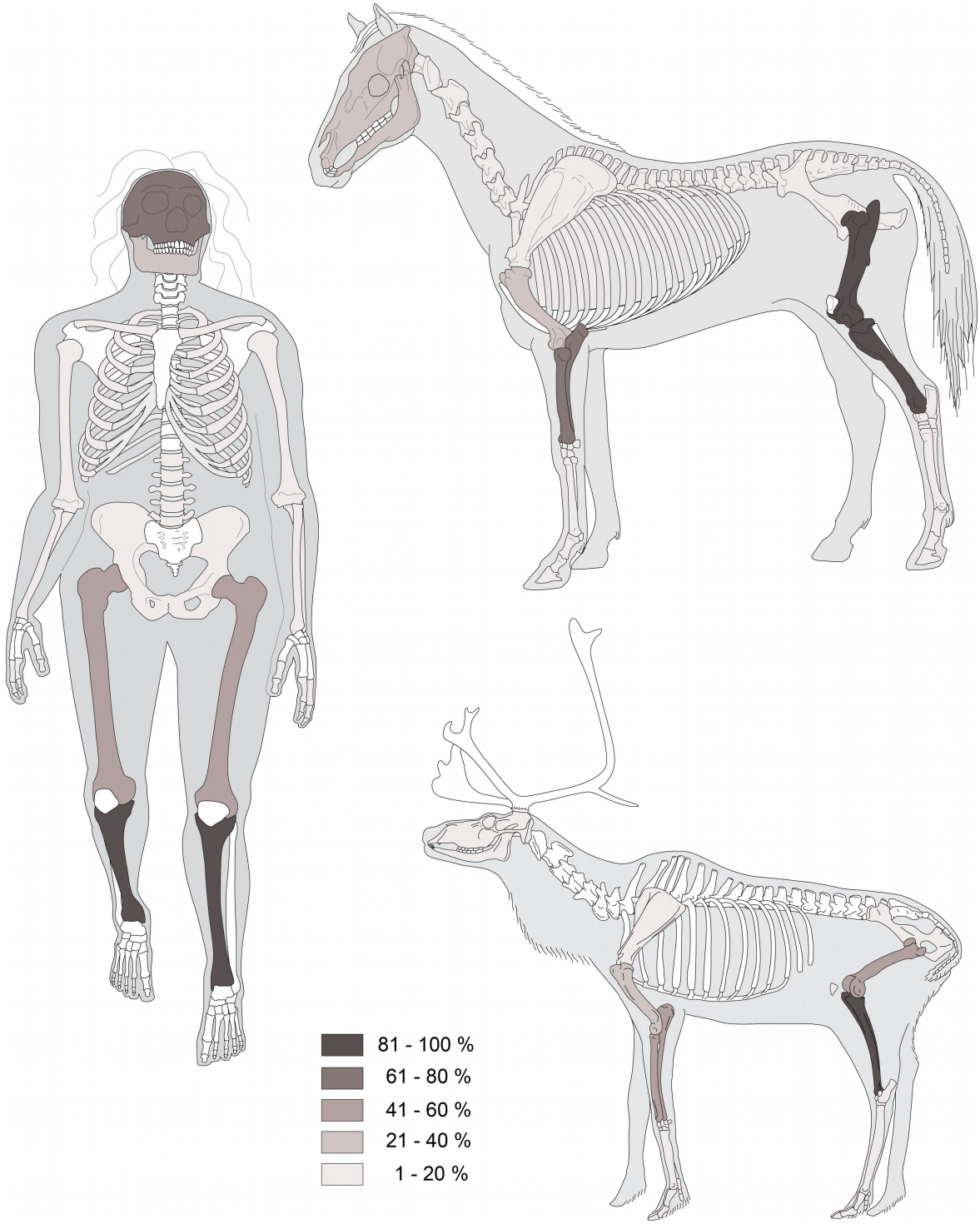
Supplementary Fig. S11. Tibia IV in posterior view (a) and details of the area of the bone used as a retoucher (b and c).

b: close-up photo of the area showing retouching marks; c: image of the area showing retouching marks obtained using a minidome, see Methods.



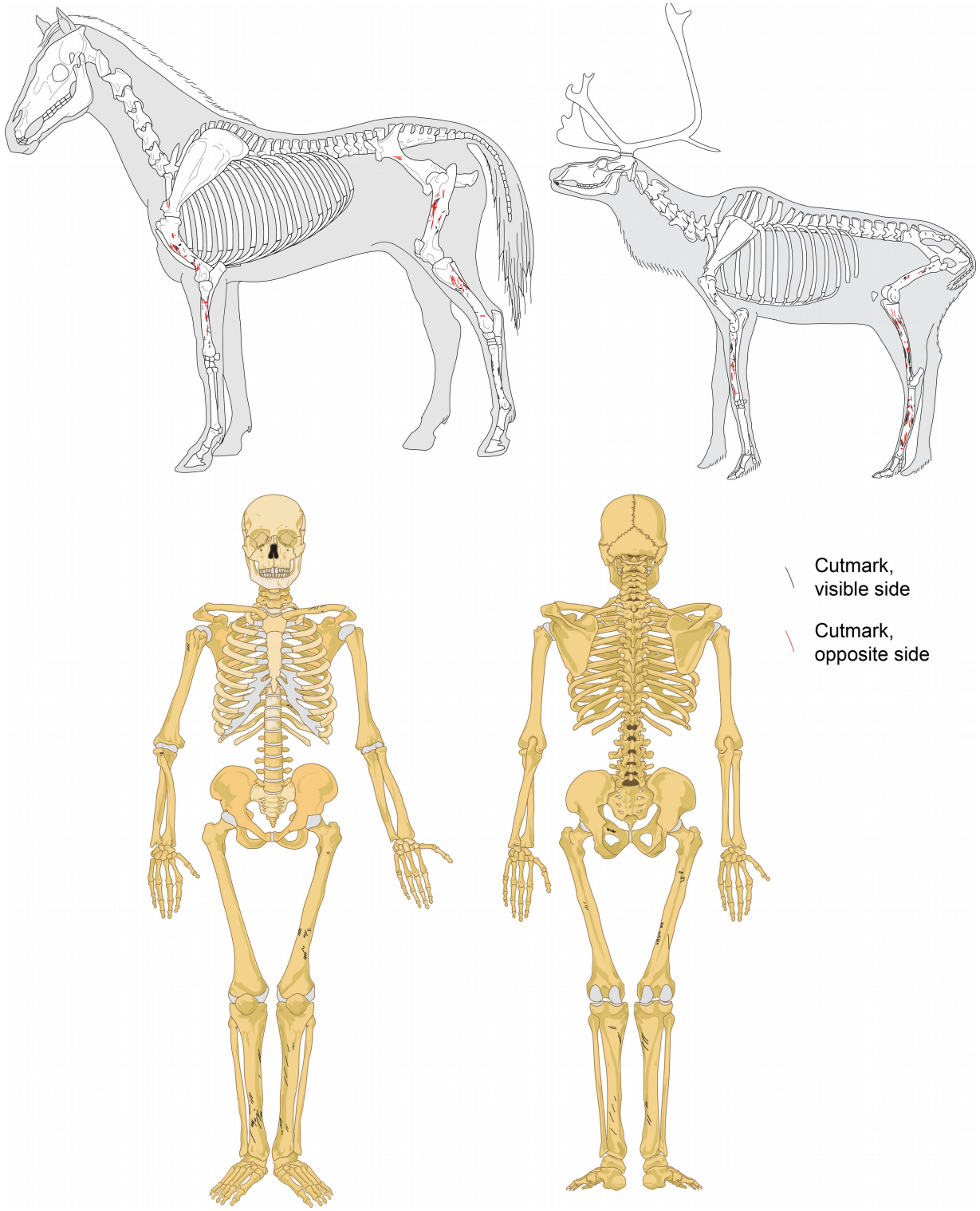
Supplementary Fig. S12. Tibia V in medial view (a) and details of the area of the bone used as a retoucher (b and c).

b: close-up photo of the area showing retouching marks; c: image of the area showing retouching marks obtained using a minidome, see Methods.



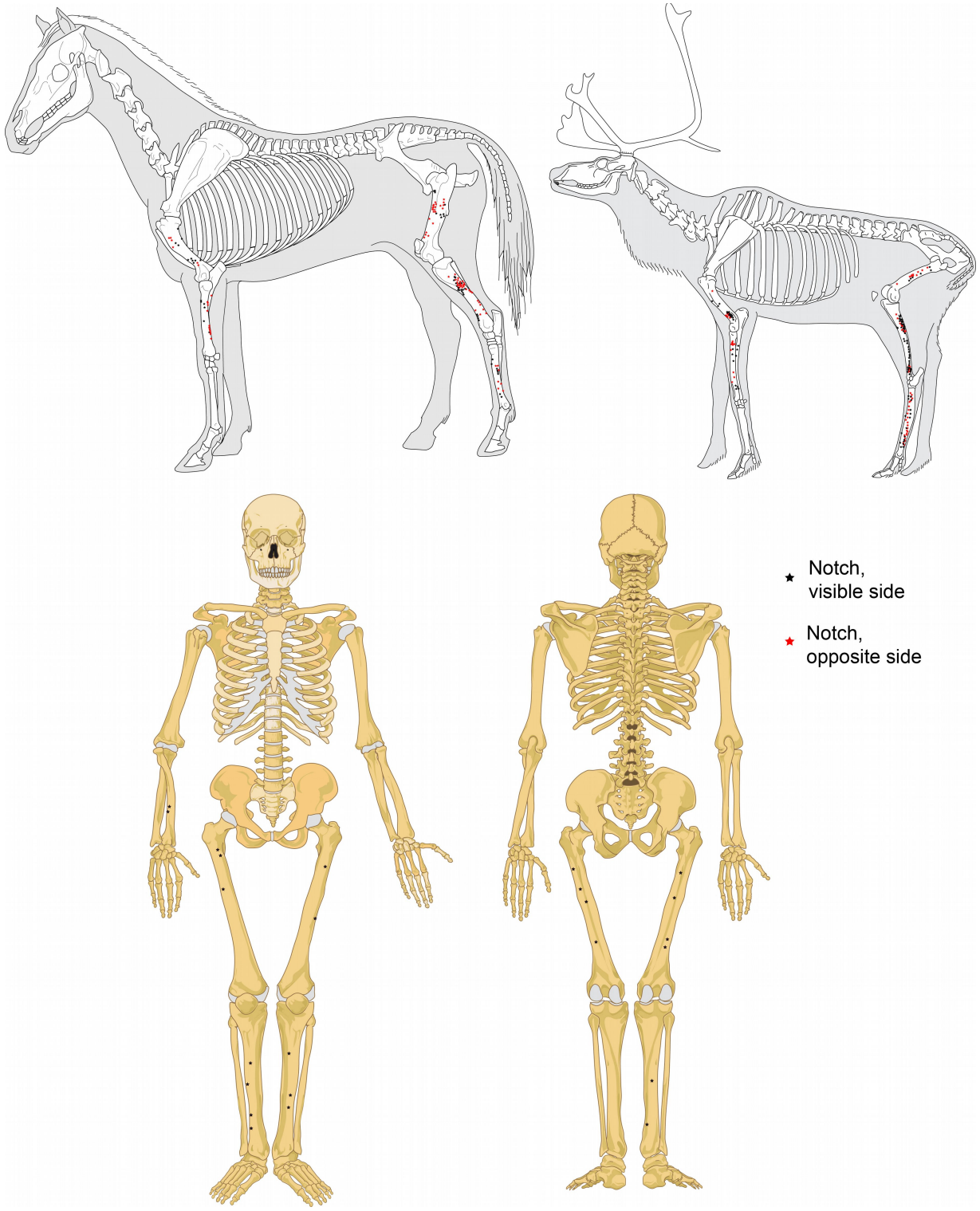
Supplementary Fig. S13. Skeletal representation of the Neandertal (left), horse (top right) and reindeer (bottom right) remains.

See Supplementary Table S6 for the values of each species. Neandertal diagram modified from http://archeozoo.org/archeozootheque/index/category/102-hominides_langen_hominid_lang_langes_hominidos_lang_ (diagram by C.B–© 2013 ArcheoZoo.org, after ref. 43), horse diagram from http://archeozoo.org/archeozootheque/picture/2595-equus_caballus/category/85-perissodactyles_langen_odd_toed_ungulate_lang_langes_perisodactilos_lang_ (diagram by M. Coutureau (Inrap), in coll. with V. Forest–© 1996 ArcheoZoo.org, after ref. 44: p. 21), and reindeer diagram from http://archeozoo.org/archeozootheque/picture/2610-rangifer_tarandus/category/92-cervides_langen_cervidae_lang_langes_cervidos_lang_ (diagram by C.B & M. Coutureau (Inrap)–© 2003 ArcheoZoo.org, after ref. 45: p. 182). All diagrams modified using Adobe Illustrator CS4 v. 14.0.0.



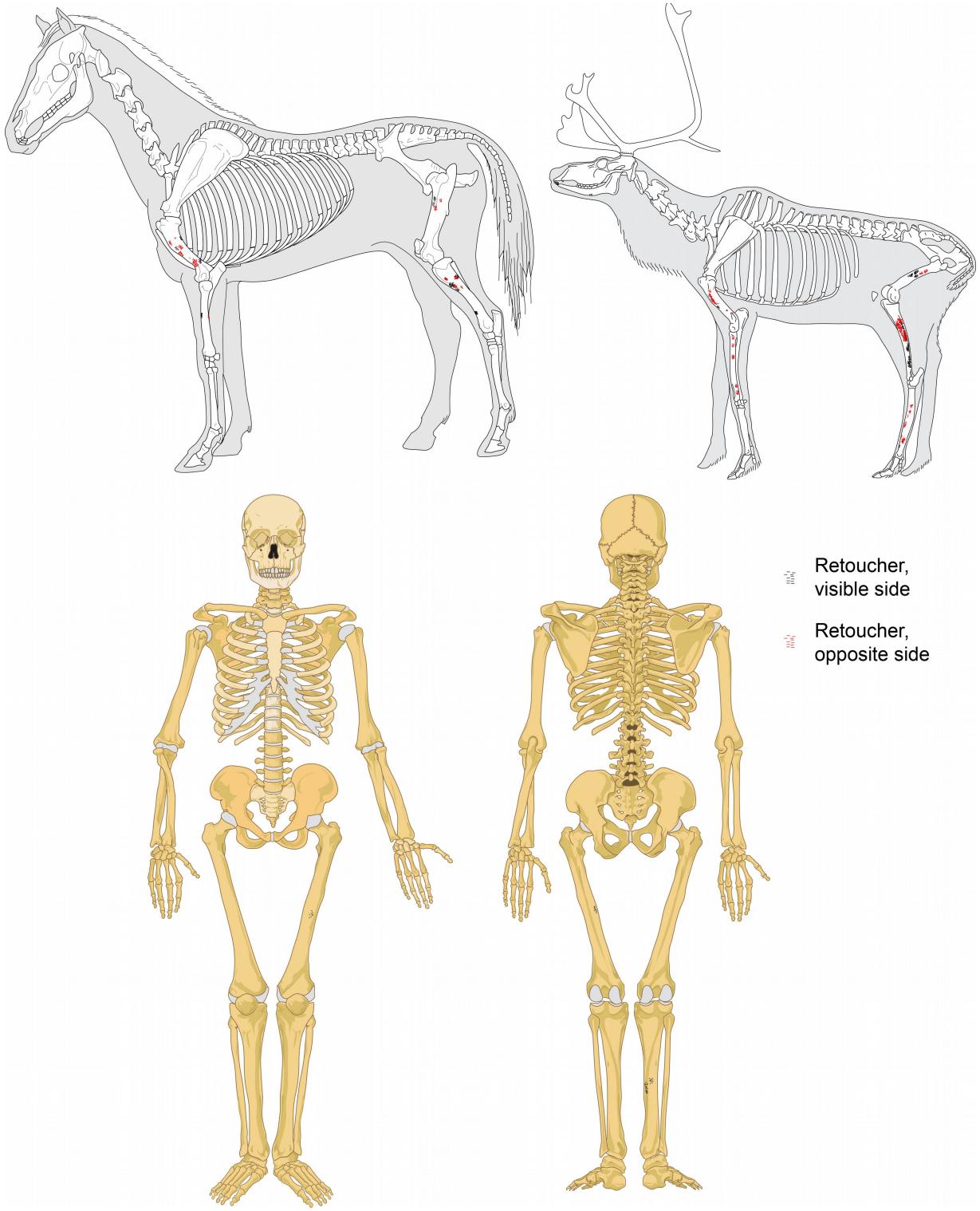
Supplementary Fig. S14. Cutmarks on horse (top left), reindeer (top right) and Neandertal (bottom) bones from Goyet.

The faunal specimens were identified among a sample of Dupont's collection from FBL 2 and 3 (horse: N = 442, reindeer: N = 287; Supplementary Table S5). Horse diagram modified from http://archeozoo.org/archeozootheque/picture/2595-equus_caballus/category/85-perissodactyles_langen_odd_toed_ungulate_lang_langes_perisodactilos_lang_ (diagram by M. Coutureau (Inrap), in coll. with V. Forest–© 1996 ArcheoZoo.org, after ref. 44: p. 21), reindeer diagram from http://archeozoo.org/archeozootheque/picture/2610-rangifer_tarandus/category/92-cervides_langen_cervidae_lang_langes_cervidos_lang_ (diagram by C.B & M. Coutureau (Inrap)–© 2003 ArcheoZoo.org, after ref. 45: p. 182), and human skeleton diagrams from https://en.wikipedia.org/wiki/File:Human_skeleton_front_en.svg and https://en.wikipedia.org/wiki/File:Human_skeleton_back_en.svg. All diagrams modified using Adobe Illustrator CS4 v. 14.0.0.



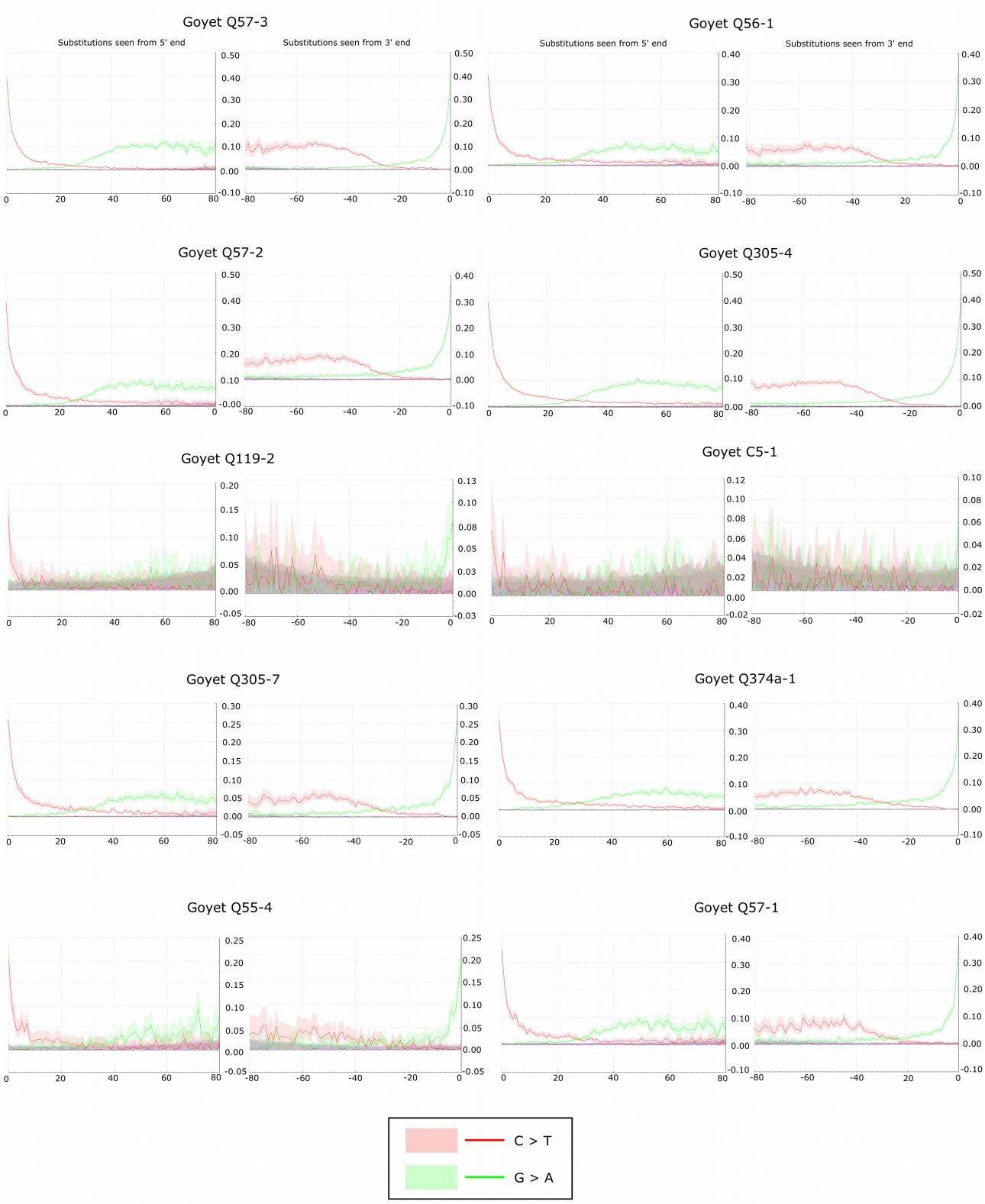
Supplementary Fig. S15. Percussion notches on horse (top left), reindeer (top right) and Neandertal bones (bottom) from Goyet.

The faunal specimens were identified among a sample of Dupont's collection from FBL 2 and 3 (horse: N = 442, reindeer: N = 287; Supplementary Table S5). Horse diagram modified from http://archeozoo.org/archeozootheque/picture/2595-equus_caballus/category/85-perissodactyles_langen_odd_toed_ungulate_lang_langes_perisodactilos_lang_ (diagram by M. Coutureau (Inrap), in coll. with V. Forest–© 1996 ArcheoZoo.org, after ref. 44: p. 21), reindeer diagram from http://archeozoo.org/archeozootheque/picture/2610-rangifer_tarandus/category/92-cervides_langen_cervidae_lang_langes_cervidos_lang_ (diagram by C.B & M. Coutureau (Inrap)–© 2003 ArcheoZoo.org, after ref. 45: p. 182), and human skeleton diagrams from https://en.wikipedia.org/wiki/File:Human_skeleton_front_en.svg and https://en.wikipedia.org/wiki/File:Human_skeleton_back_en.svg. All diagrams modified using Adobe Illustrator CS4 v. 14.0.0.

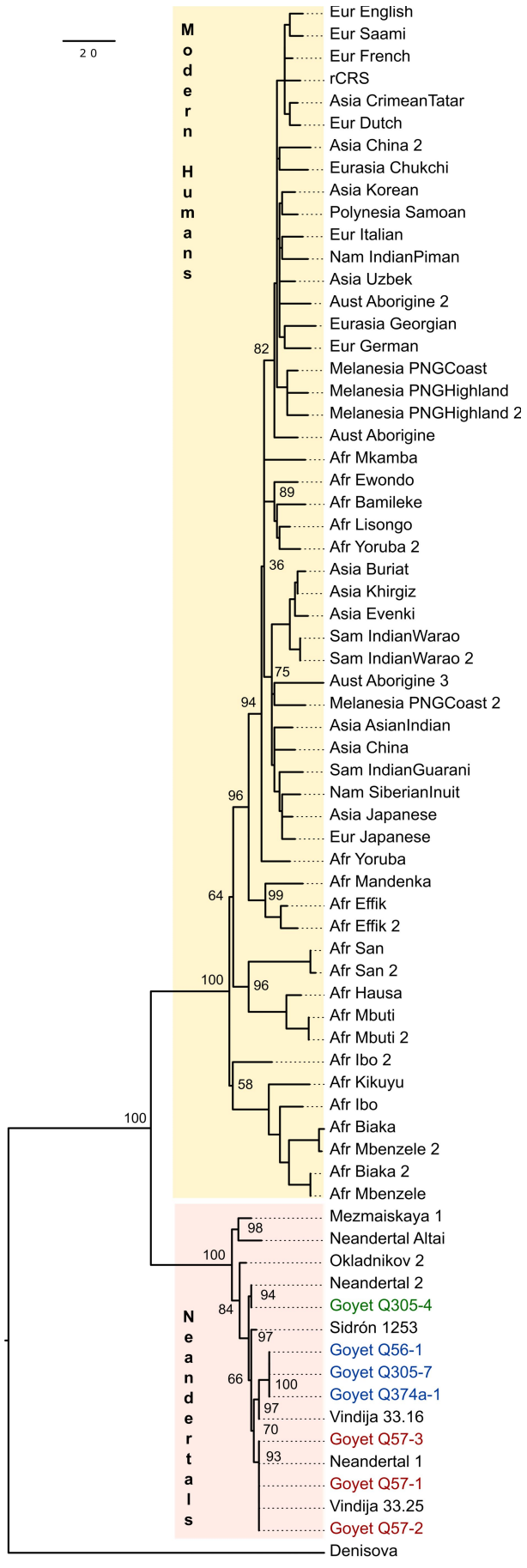


Supplementary Fig. S16. Retoucher traces on horse (top left), reindeer (top right) and Neandertal (bottom) bones from Goyet.

The faunal specimens were identified among a sample of Dupont's collection from FBL 2 and 3 (horse: N = 442, reindeer: N = 287; Supplementary Table S5). Horse diagram modified from http://archeozoo.org/archeozootheque/picture/2595-equus_caballus/category/85-perissodactyles_langen_odd_toed_ungulate_lang_langes_perisodactilos_lang_ (diagram by M. Coutureau (Inrap), in coll. with V. Forest–© 1996 ArcheoZoo.org, after ref. 44: p. 21), reindeer diagram from http://archeozoo.org/archeozootheque/picture/2610-rangifer_tarandus/category/92-cervides_langen_cervidae_lang_langes_cervidos_lang_ (diagram by C.B & M. Coutureau (Inrap)–© 2003 ArcheoZoo.org, after ref. 45: p. 182), and human skeleton diagrams from https://en.wikipedia.org/wiki/File:Human_skeleton_front_en.svg and https://en.wikipedia.org/wiki/File:Human_skeleton_back_en.svg. All diagrams modified using Adobe Illustrator CS4 v. 14.0.0.

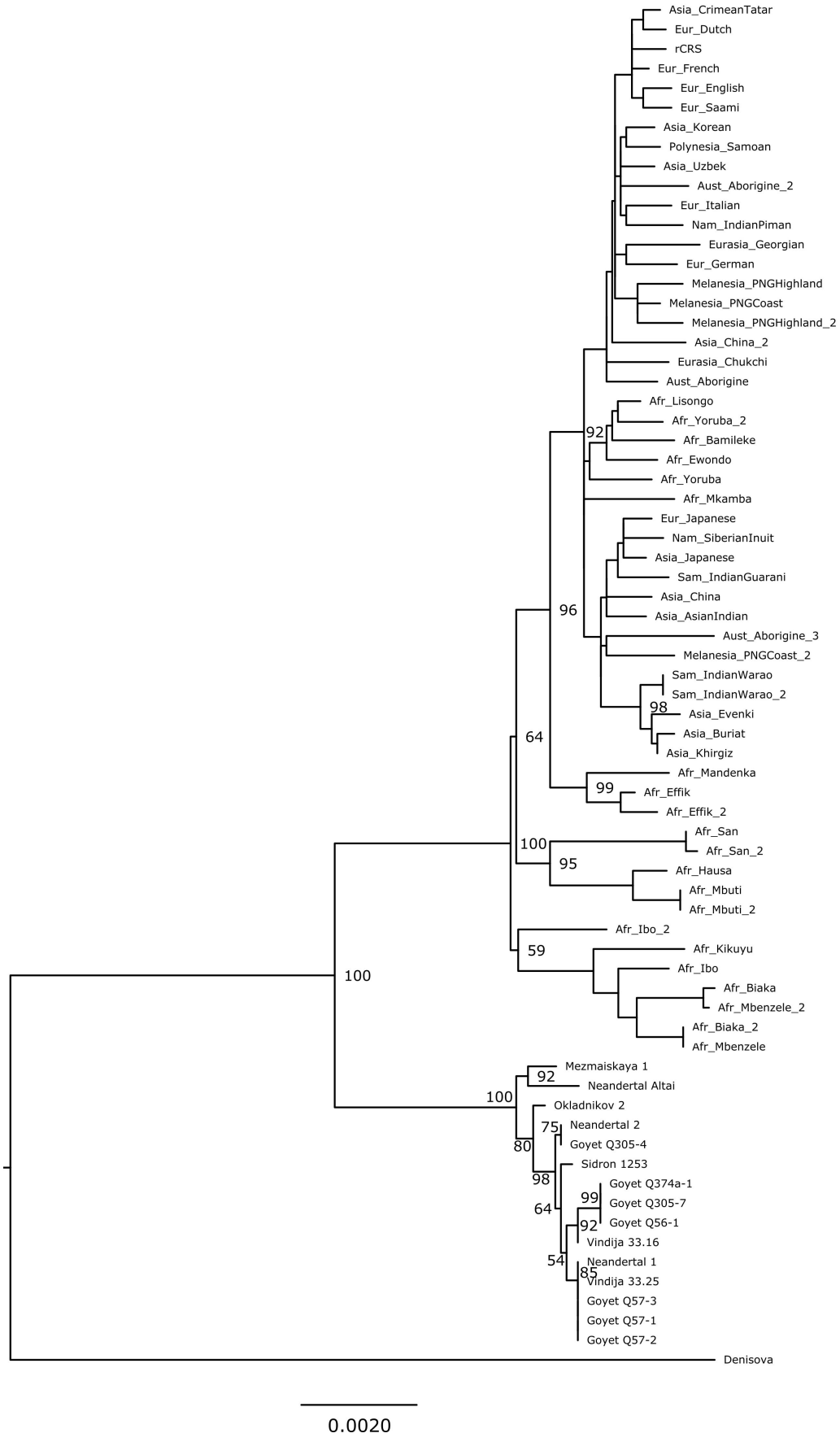


Supplementary Fig. S17. Damage plots for all of the Goyet Neandertal samples before postmortem damage (PMD) filtering.



Supplementary Fig. S18. Maximum parsimony tree of the seven analysed Goyet Neandertal mitochondrial genomes after selection of damaged reads (PMD filtering) compared to 63 published modern human, Neandertal and Denisovan mtDNAs.

Numbers at the main branch nodes represent bootstrap values after 1,000 iterations.



Supplementary Fig. S19. Maximum likelihood tree of the seven analysed Goyet Neandertal mitochondrial genomes after selection of damaged reads (PMD filtering) compared to 63 published modern human, Neandertal and Denisovan mtDNAs.

Numbers at the main branch nodes represent bootstrap values after 1,000 iterations.

Supplementary Tables

Supplementary Table S1. Human remains from the Troisième caverne of Goyet identified as Neandertal with indication of the "fauna-bearing level" (FBL) and analyses performed.

Specimen	FBL	Identification	Refits with / on	Analyses		
				^{14}C	^{13}C-^{15}N	DNA
1189-1	2	Left femur diaphysis frag.	Femur III			
1424-3D*	/	Lower left I2 (isolated)				
2861-1	4?	Right and left maxillae, alveolar and palatine processes	2878-1D			
2861-19D	4?	Root of upper left I1	on 2861-1			
2878-1D	3?	Upper left I2 (isolated)	2861-1			
2861-20D	4?	Root of upper left C	on 2861-1			
2861-21D	4?	Root of upper left P1	on 2861-1			
2861-22D	4?	Upper left P2	on 2861-1			
2861-23D	4?	Upper left M1	on 2861-1			
2861-24D	4?	Upper left M2	on 2861-1			
2861-25D	4?	Root of upper right C	on 2861-1			
2861-26D	4?	Root of upper right P1	on 2861-1	Xº		
2861-27D	4?	Root of upper right P2	on 2861-1			
2878-1	1 or 3	Right parietal, postero-superior frag.				
2878-2	1 or 3	Right parietal, anterior frag.				
2878-3	1 or 3	Right and left parietal fragments articulating along the sagittal suture	C5-1			
2878-4	1 or 3	Occipital, left nuchal plane frag.				
2878-8	3	Mandible, left body frag. with P1 and M1	2878-2D			
2878-21D	3	Lower left P1	on 2878-8			
2878-2D	1, 2 or 3	Lower left P2 (isolated)	2878-8	X	X	
2878-22D	3	Lower left M1	on 2878-8			
2878-37	3	Hand middle phalanx 2-4				
2878-38	3	Hand proximal phalanx 3-4, proximal extremity broken off				
2878-39	3	Left hand proximal phalanx 5, prox. extremity broken off, distal extremity partially broken				
C5-1	3	Left parietal frag. without sutures	2878-3		X	X
C5-2	3	Lumbar vertebra 1-4, left pedicle and left superior articular process				
C5-3	3	Mandible, right ascending ramus frag.	C5-4			
C5-4	3	Mandible, right gonial angle	C5-3			
C5-5	3	Mandible, left body inferior frag.				
C5-6	3	Right temporal, squamous frag. and complete petrous				
C5-7	3	Left temporal, mastoid portion frag.				
C5-8	3	Left zygomatic, frontal process				
Q48-1	2	Left pubis superior frag.	Q376-36			
Q53-4	3	Right humerus diaphysis frag.	Humerus III	X	X	
Q53-5	2	Ulna? diaphysis frag.				
Q54-4	1	Right tibia diaphysis frag.	Tibia III			
Q54-5	1	Right tibia diaphysis frag.	Tibia IV?			
Q55-1	3	Left clavicle, lateral half		X	X	
Q55-3	3	Right tibia diaphysis frag.	Tibia IV			
Q55-4	3	Right tibia diaphysis frag.	Tibia IV		X	X
Q55-5	3	Left femur diaphysis frag.	Femur III			
Q55-6	3	Right rib 11?, distal half	Q376-25			
Q55-7	3	Femur diaphysis frag.				
Q56-1	3	Right femur diaphysis frag.	Femur I	X	X	X
Q56-2	3	Left tibia diaphysis frag.	Tibia I			
Q56-5	3	Right tibia diaphysis frag.	Tibia III			
Q56-6	3	Left femur diaphysis frag.	Femur IV			
Q56-7	2	Right femur diaphysis frag.	Femur II			
Q56-8	3	Right tibia diaphysis frag.	Tibia IV			
Q56-9	3	Right tibia diaphysis frag.	Tibia V			
Q56-10	3	Right tibia diaphysis frag.	Tibia V			
Q56-11	2	Right tibia diaphysis frag.	Tibia IV			
Q56-12	2 or 3	Right radius diaphysis frag.	Radius I			

Q56-13	2 or 3	Right radius diaphysis frag.	Radius I			
Q56-14	1	Right humerus, diaphysis and neck frag.	Humerus III			
Q56-17	2	Tibia? diaphysis frag.				
Q57-1	1	Left tibia diaphysis frag.	Tibia II	X	X	X
Q57-2	1	Right femur diaphysis frag.	Femur II	X	X	X
Q57-3	1	Right tibia diaphysis frag.	Tibia VI	X	X	X
Q98-1	2	Right femur, prox. extremity frag. with lesser trochanter and diaphysis frag.	Femur I			
Q100-3	2	Right rib 1, distal shaft frag.				
Q115-1	3	Left femur diaphysis frag.	Femur IV			
Q115-2	3	Left femur diaphysis frag.	Femur III			
Q115-3	3	Tibia or femur diaphysis frag.				
Q116-2	3	Left tibia diaphysis frag.	Tibia I			
Q116-3	3	Radius, head and neck frag.				
Q119-2	1	Left rib 7? shaft frag.	Q376-7?		X	X
Q305-2	3	Left tibia diaphysis frag.	Tibia I			
Q305-3	3	Left tibia diaphysis frag.	Tibia I			
Q305-4	3	Left tibia diaphysis frag.	Tibia I	X	X	X
Q305-7	3	Right tibia diaphysis frag.	Tibia III		X	X
Q305-8	3	Right tibia diaphysis frag.	Tibia VI			
Q305-11	3	Tibia diaphysis frag.				
Q305-12	3	Femur diaphysis frag.				
Q374a-1	3	Right tibia diaphysis frag.	Tibia V		X	X
Q375-1	3	Right femur diaphysis frag.	Femur I			
Q375-2	3	Left tibia diaphysis frag.				
Q375-3	3	Left tibia diaphysis frag.	Tibia I			
Q375-4	3	Left tibia diaphysis frag.	Tibia I			
Q375-6	3	Right tibia diaphysis frag.	Tibia IV			
Q375-7	3	Right radius diaphysis frag.	Radius I			
Q375-8	3	Right radius diaphysis frag.	Radius I			
Q375-9	3	Femur diaphysis frag.				
Q375-10	3	Right humerus, diaphysis and neck frag.	Humerus II?			
Q376-1	3	Hand proximal phalanx 2-4, both extremities broken off		X	X	
Q376-2	3	Right femur diaphysis frag.	Femur I			
Q376-5	3	Left tibia diaphysis frag.	Tibia I			
Q376-6	3	Right tibia diaphysis frag.	Tibia IV			
Q376-7	3	Left rib 7? shaft frag.	Q119-2?			
Q376-8	3	Left rib 1, sub-complete				
Q376-9	3	Right rib 11? shaft frag.				
Q376-11	3	Left rib 10 shaft frag.				
Q376-12	3	Left rib 4?, shaft frag. with costal angle				
Q376-13	3	Right rib 2 shaft frag.				
Q376-14	3	Left rib 4-5?, distal half/third				
Q376-16	3	Right rib 3? shaft frag.				
Q376-17	3	Left? rib 4-9? shaft frag.				
Q376-18	3	Right tibia diaphysis frag.	Tibia III?			
Q376-20	3	Right humerus diaphysis frag.	Humerus II	X	X	
Q376-25	3	Right rib 11? shaft frag.	Q55-6			
Q376-26	3	Rib 3-11 shaft frag.				
Q376-27	3	Right? rib 3-7? shaft frag.				
Q376-28	3	Left rib 5-9 shaft frag.				
Q376-29	3	Right rib 3-9 shaft frag.				
Q376-30	3	Right? rib 6-8?, sternal end				
Q376-31	3	Left? rib 11?, distal half				
Q376-32	3	Left rib 3, frag. preserving the neck, costal tubercle and angle				
Q376-33	3	Rib 8-11? shaft frag.				
Q376-35	3	Humerus head frag.				
Q376-36	3	Left pubis inferior frag.	Q48-1			

For the numbering system, see Supplementary Note S4. Element numbers (Roman numerals) were given to the most complete bones and indicate refits (column 'Refits with / on'). º see Supplementary Note S6. Note that all of the specimens come from the RBINS collections and were excavated by E. Dupont, except for * that belongs to the RMAH material from A. de Loë's excavations.

Supplementary Table S2. Neandertal bones (after refitting) from the Troisième caverne of Goyet with indication of anthropogenic modifications.

Bone piece	Identification	Anthropogenic modifications			
		Cutmarks	Percussion notches	Percussion pits	Retoucher
Craniofacial skeleton					
2878-3 + C5-1	Right and left parietal fragments with portion of sagittal suture				
2878-1	Right parietal, postero-superior frag.				
2878-2	Right parietal, anterior frag.				
2878-4	Occipital, left nuchal plane frag.				
C5-6	Right temporal, squamous frag. and complete petrous	X?			
C5-7	Left temporal, mastoid portion frag.				
C5-8	Left zygomatic, frontal process				
2861-1	Right and left maxillae, alveolar and palatine processes				
2878-8	Mandible, left body frag. with P1 and M1				
C5-3 + C5-4	Mandible, right gonial angle & ascending ramus frag.	X			
C5-5	Mandible, left body inferior frag.				
Trunk					
C5-2	Lumbar vertebra 1-4, left pedicle and left superior articular process				
Q100-3	Right rib 1, distal shaft frag.				
Q376-8	Left rib 1, sub-complete	X			
Q376-13	Right rib 2 shaft frag.				
Q376-16	Right rib 3? shaft frag.				
Q376-32	Left rib 3, frag. preserving the neck, costal tubercle and angle				
Q376-27	Right? rib 3-7? shaft frag.				
Q376-29	Right rib 3-9 shaft frag.				
Q376-30	Right? rib 6-8?, sternal end				
Q376-9	Right rib 11? shaft frag.				
Q55-6 + Q376-25	Right rib 11? shaft	X			
Q376-26	Rib 3-11 shaft frag.				
Q376-12	Left rib 4?, shaft frag. with costal angle				
Q376-14	Left rib 4-5?, distal half/third				
Q376-17	Left? rib 4-9? shaft frag.				
Q376-28	Left rib 5-9 shaft frag.				
Q119-2	Left rib 7? shaft frag.				
Q376-7	Left rib 7? shaft frag.				
Q376-11	Left rib 10 shaft frag.	X			
Q376-31	Left? rib 11?, distal half				
Q376-33	Rib 8-11? shaft frag.				
Upper limb					
Q55-1	Left clavicle, lateral half	X			
Humerus II (Q376-20)	Right humerus diaphysis frag.	X			
Q375-10 (Humerus II?)	Right humerus, diaphysis and neck frag.				
Humerus III (2 spec.)	Right humerus, diaphysis and neck frag.				
Q376-35	Humerus head frag.				
Radius I (4 spec.)	Right radius diaphysis		X	X	
Q116-3	Radius, head and neck frag.	X			
Q53-5	Ulna? diaphysis frag.				
2878-37	Hand middle phalanx 2-4				
2878-38	Hand proximal phalanx 3-4, proximal extremity broken off				
2878-39	Left hand proximal phalanx 5, prox. extremity broken off, distal extremity partially broken				
Q376-1	Hand proximal phalanx 2-4, both extremities broken off				
Lower limb					
Q48-1 + Q376-36	Left pubis	X			
Femur I (4 spec.)	Right femur diaphysis & prox. extremity frag.	X	X	X	
Femur II (2 spec.)	Right femur diaphysis frag.	X	X	X	
Femur III (3 spec.)	Left femur diaphysis frag.	X	X	X	X
Femur IV (2 spec.)	Left femur diaphysis frag.	X	X	X	
Q55-7	Femur diaphysis frag.			X	
Q375-9	Femur diaphysis frag.		X		
Q305-12	Femur diaphysis frag.				

Q115-3	Tibia or femur diaphysis frag.				
Tibia I (8 spec.)	Left tibia diaphysis	X	X		
Tibia II (Q57-1)	Left tibia diaphysis frag.	X	X		
Tibia III (3 spec.)	Right tibia diaphysis frag.	X	X	X	X
Q376-18 (Tibia III?)	Right tibia diaphysis frag.	X			
Tibia IV (6 spec.)	Right tibia diaphysis	X	X	X	X
Q54-5 (Tibia IV?)	Right tibia diaphysis frag.				
Tibia V (3 spec.)	Right tibia diaphysis frag.	X	X	X	X
Tibia VI (2 spec.)	Right tibia diaphysis frag.	X	X		
Q375-2	Left tibia diaphysis frag.	X			
Q56-17	Tibia? diaphysis frag.				
Q305-11	Tibia diaphysis frag.				

See Supplementary Fig. S8 for the placement of the anthropogenic modifications on the bones.

Supplementary Table S3. Sample information and results of the elemental chemical analyses of the Neandertal remains from Goyet.

Specimen ID	^{14}C dating lab # (CIO)	$\%C_{coll}$	$\%N_{coll}$	C_{coll}:N_{coll}
2878-2D	GrA-54028	41.4	14.4	3.4
C5-1	-	43.0	14.7	3.4
Q53-4	GrA-54022	42.9	15.1	3.3
Q55-1	GrA-54257	36.9	12.9	3.3
Q55-4	-	39.6	14.0	3.3
Q56-1	GrA-46170*	45.4	15.5	3.4
Q57-1	GrA-46173*	46.0	16.8	3.2
Q57-2	GrA-54024	42.7	15.0	3.3
Q57-3	GrA-60019	43.8	15.4	3.3
Q119-2	-	38.9	13.8	3.3
Q305-4	GrA-46176*	47.1	16.7	3.3
Q305-7	-	41.9	14.9	3.3
Q374a-1	-	43.1	15.2	3.3
Q376-1	GrA-46178*	46.7	17.0	3.2
Q376-20	GrA-60018	39.8	14.0	3.3

* indicates collagens extracted at the CIO; all others were extracted at Tübingen University where elemental chemical analyses and stable isotope analyses (41) were also performed.

Supplementary Table S4. Minimum number of elements (MNE) and minimum number of individuals (MNI) represented by the Goyet Neandertal bone assemblage along with the percentage representation (PR) of the elements.

	MNE	MNI	PR (%)
Cranium	2	2	50.0
Mandible	1	1	25.0
Vertebrae	1	1	1.0
Sacrum	0	0	0.0
Ribs	11	2	11.5
Sternum	0	0	0.0
Scapula	0	0	0.0
Clavicle	1	1	12.5
Humerus	2	2	25.0
Radius	1	1	12.5
Ulna	1	1	12.5
Carpus + metacarpus	0	0	0.0
Hand phalanges	4	1	3.6
Os coxae	1	1	12.5
Femur	4	3	50.0
Patella	0	0	0.0
Tibia	6	4	75.0
Fibula	0	0	0.0
Foot	0	0	0.0
Total	*35*	*4*	*4.9*

The PR is calculated as MNE*100/(MNImax*NEind) with MNImax being the highest MNI for the whole sample and NEind the number of elements per individual.

Supplementary Table S5. Faunal sample from Dupont's excavations at the Troisième caverne of Goyet identified during the present study with indication of the storage drawers and "fauna-bearing levels" (FBL).

		Q53 (FBL 2)	Q55 (FBL 3)	Q375 (FBL 3)	Q376 (FBL 3)	*Total*
Perissodactyla	Horse (*Equus caballus*)	99*	89*	146*	108*	*442*
Artiodactyla	Reindeer (*Rangifer tarandus*)	47*	240*	354	262	*903*
	Red deer (*Cervus elaphus*)	2	3	6	3	*14*
	Roe deer (*Capreolus capreolus*)			1		*1*
	Megaceros (*Megaloceros giganteus*)		2	3		*5*
	Bovid (*Bos primigenius* or *Bison priscus*)	11	10	15	15	*51*
	Ibex (*Capra ibex*)	7	1	5	1	*14*
	Wild boar (*Sus scrofa*)	2			1	*3*
Proboscidea	Mammoth (*Mammuthus primigenius*)	2	3		1	*6*
Lagomorpha	Leporid (*Lepus timidus* or *Oryctolagus cuniculus*)	1		2	3	*6*
Ungulata	Ungulate 3/4 (horse or bovid)		5			*5*
	Ungulate 5 (rhinoceros or mammoth)			1		*1*
Carnivora	Bear (*Ursus spelaeus* or *Ursus arctos*)	16		30	10	*56*
	Fox (*Vulpes vulpes* or *Vulpes lagopus*)	6		5	1	*12*
	Large canid (*Canis sp.*)	4		5	2	*11*
	Hyaena (*Crocuta crocuta spelaea*)	1	1	4	2	*8*
	Badger (*Meles meles*)	1				*1*
	Carnivora indet			1		*1*
Mammal indet		2				*2*
Bird		5	1		8	*14*
Total		*206*	*355*	*578*	*417*	*1556*

* indicates the faunal specimens that were observed for the presence of anthropogenic modifications.

Supplementary Table S6. Comparison of the horse, reindeer and Neandertal skeletal representation at Goyet.

	NEANDERTAL				HORSE				REINDEER			
	NISP	NEind	NISP/ NEind	%NISP/NEind / (NISP/NEind)max	NISP	NEind	NISP/ NEind	%NISP/NEind / (NISP/NEind)max	NISP	NEind	NISP/ NEind	%NISP/NEind / (NISP/NEind)max
Cranium	9	1	9	64.29	24	2	12	21.43	12	2	6	4.67
Mandible	4	1	4	28.57								
Cervical vertebrae	0	7	0	0	4	86	0.05	0.08	0	64	0	0
Thoracic vertebrae	0	12	0	0								
Lumbar vertebrae	1	5	0.2	1.43								
Sacrum	0	1	0	0								
Ribs	21	24	0.88	6.25								
Scapula	0	2	0	0	14	2	7	12.5	2	2	1	0.78
Clavicle	1	2	0.5	3.57	N/A				N/A			
Humerus	5	2	2.5	17.86	39	2	19.5	34.82	101	2	50.5	39.30
Radius	5	2	2.5	17.86	74	2	37	66.07	116	2	58	45.14
Ulna	1	2	0.5	3.57								
Coxal	2	2	1	7.14	5	2	2.5	4.46	4	2	2	1.56
Femur	14	2	7	50	96	2	48	85.71	137	2	68.5	53.31
Tibia	28	2	14	100	112	2	56	100	257	2	128.5	100
Fibula	0	2	0	0	0	2	0	0	0	2	0	0
Carpus+tarsus+patella	0	32	0	0	3	28	0.11	0.19	4	24	0.17	0.13
Metacarpus+metatarsus	0	20	0	0	64	12	5.33	9.52	259	12	21.58	16.80
Phalanges	4	56	0.07	0.51	5	12	0.42	0.74	6	48	0.13	0.10
Indet	1	-	-	-	2	-	-	-	5	-	-	-
Total	*96*	*177*	*0.54*	*3.87*	*442*	*132*	*3.35*	*5.98*	*903*	*151*	*5.98*	*4.65*

NISP: Number of identified specimens, NEind: Number of elements per individual, (NISP/NEind)max: maximal value of NISP/NEind. Variable definitions are after ref. 42. The horse and reindeer specimens were identified among a sample of Dupont's collection from FBL 2 and 3 (Supplementary Table S5).

Supplementary Table S7. Palaeogenetic results of the mtDNA reads for the 10 Goyet Neandertal specimens analysed before and after postmortem damage (PMD) filtering.

Before filtering (NoPMD)

Sample ID	Library	Unique mapping reads	Average coverage of mtDNA (fold)	Nucleotides covered at 5-fold coverage (% of mtDNA)	Average read length (base pairs)	Damage at 5' end (%)	Polluting fragments	Clean fragments	Contaminaton estimate (95 % confidence interval)
C5-1	Sample_MA130m	957	5.56	9445 (57.00)	96.21	7	137	86	61.4 (54.9–67.6) %
Q55-4	Sample_MA124m	1964	9.60	14838 (89.57)	81.01	20	85	351	19.5 (16.0–23.5) %
Q56-1	Sample_GA71	11682	44.93	16555 (99.94)	63.70	32	58	1715	3.3 (2.5–4.2) %
Q57-1	Sample_GA72	6096	24.19	16257 (98.14)	65.73	35	137	819	14.3 (12.3–16.7) %
Q57-2	Sample_GA73	12905	49.91	16564 (99.99)	64.06	39	166	1886	8.1 (7.0–9.3) %
Q57-3	Sample_GA66	29758	94.92	16538 (99.84)	52.83	39	167	3452	4.6 (4.0–5.3) %
Q119-2	Sample_MA131m	1025	4.94	8465 (51.11)	79.79	14	75	138	35.2 (29.1–41.8) %
Q305-4	Sample_GA74	36399	146.25	16565 (100.00)	66.55	39	136	5739	2.3 (2.0–2.7) %
Q305-7	Sample_MA127m	10690	46.68	16561 (99.98)	72.33	26	45	2147	2.1 (1.5–2.7) %
Q374a-1	Sample_MA125m	15796	69.85	16562 (99.99)	73.24	34	12	3124	0.4 (0.2–0.7) %

After filtering (PMD)

Sample ID	Library	Unique mapping reads	Average coverage of mtDNA (fold)	Nucleotides covered at 5-fold coverage (% of mtDNA)	Average read length (base pairs)	Damage at 5' end (%)	Polluting fragments	Clean fragments	Contaminaton estimate (95 % confidence interval)
C5-1	Sample_MA130m	187	0.96	282 (1.71)	91.72	20	8	24	25.0 (13.3–42.1) %
Q55-4	Sample_MA124m	613	2.69	2740 (16.55)	75.07	46	5	90	5.3 (2.3–11.7) %
Q56-1	Sample_GA71	4096	14.53	14925 (90.11)	58.77	67	5	506	1.0 (0.4–2.3) %
Q57-1	Sample_GA72	2142	7.77	11119 (67.12)	60.09	64	9	264	3.3 (1.7–6.1) %
Q57-2	Sample_GA73	4741	16.77	16109 (97.25)	58.60	71	4	601	0.7 (0.3–1.7) %
Q57-3	Sample_GA66	12387	38.51	16011 (96.66)	51.50	71	16	1312	1.2 (0.7–1.9) %
Q119-2	Sample_MA131m	223	0.93	126 (0.77)	76.00	28	9	43	17.3 (9.4–29.7) %
Q305-4	Sample_GA74	13980	52.37	16520 (99.73)	62.04	70	9	1986	0.5 (0.2–0.9) %
Q305-7	Sample_MA127m	3784	15.60	16333 (98.60)	71.04	60	5	690	0.7 (0.3–1.7) %
Q374a-1	Sample_MA125m	6922	28.68	16522 (99.75)	71.17	65	1	1228	0.1 (0.0–0.5) %

Supplementary Table S8. Palaeogenetic results of the mtDNA reads for the three low coverage specimens mapped against modern human and Neandertal mitochondrial reference sequences before postmortem damage filtering.

Sample ID	Library	Unique mapping reads		Average coverage of mtDNA (fold)		Nucleotides covered at 5-fold coverage (% of mtDNA)		Average read length (base pairs)	
		Modern human	*Neandertal*	*Modern human*	*Neandertal*	*Modern human*	*Neandertal*	*Modern human*	*Neandertal*
C5-1	Sample_MA130m	996	957	5.87	5.56	9950 (60.05)	9445 (57.00)	97.69	96.21
Q55-4	Sample_MA124m	1899	1964	9.17	9.60	14550 (87.85)	14838 (89.57)	80.01	81.01
Q119-2	Sample_MA131m	1037	1025	4.97	4.94	8504 (51.34)	8465 (51.11)	79.45	79.79

Supplementary Notes

Supplementary Note S1. The Troisième caverne of Goyet and its regional context

The Goyet caves are located in Mozet, Belgium, some 20 km from the well-known site of Spy (Supplementary Fig. S1). The Troisième caverne of Goyet (50°26'44"N, 5°00'48"E) is part of a large karstic system developed in a Carboniferous limestone cliff of the Ardenne Massif some 130 m above sea level on the right bank of the Samson Valley, a tributary of the Meuse. The main Goyet caves open onto a large terrace about 15 m above the river. The Troisième caverne, the archaeologically richest of the Goyet cave system, is about 120 meters deep and consists of three chambers (1, 2). Chamber A lies at the entrance of the cave and is connected to Chamber B by a small gallery, with Chamber C situated at the back of the cave. Edouard Dupont, the main excavator of the site, distinguished four "fauna-bearing levels" or FBL in Chamber A (2). He recovered a significant quantity of Middle and Upper Palaeolithic artefacts, numerous Pleistocene mammal bones, especially herbivores, and human remains from the uppermost three FBL (ref. 3; see also Supplementary Note S3). Many of these bones were marrow cracked, have cutmarks, or bear traces of ochre (4, 5). At the rear of Chamber A and in Chamber B, Dupont (2) distinguished a fourth and fifth FBL containing mainly cave bear, cave lion, and cave hyaena remains. The faunal material from these levels appears unrelated to the anthropogenic assemblages from the front of Chamber A given the presence of numerous carnivore traces and comparatively less human-modified material (3, 4). The mammal assemblage from Chamber C contains remains of, amongst others, cave bear, cave hyaena, horse, reindeer, as well as human skeletal material (3, 6) that might not be of Palaeolithic age (see "Stratigraphic provenience" in Supplementary Note S4).

Interestingly, E. Dupont considered the possibility of cannibalism at Goyet in his unpublished handwritten notes from 1906 recently discovered by one of us (M.G). Concerning the human remains from Chamber A, Dupont wrote:

> [ ] all of the Caves and their archaeological levels contained some human remains mixed with those of consumed animals. The consistency of the evidence leads us to conclude cannibalism. And, indeed, in the 3rd level, adult and adolescent: cutmarks on an ulna; 2nd level, two adults and adolescent: cutmarks on a clavicle; 1st level, adult and adolescent: cutmarks on a skull fragment (our translation).

However, the two human remains we were able to match with Dupont's description (fragments of a clavicle and parietal) are not Neandertal but rather modern human remains. Nearly all of the Goyet Neandertal material was identified during our work (Supplementary Notes S3 and S4).

Very few Northern European sites north of 50° N have yielded MIS 3 Neandertal remains (Supplementary Fig. S1). In Germany, the remains of at least three individuals identified at the Neandertal type-site have been attributed to this group along with a partial parietal bone from Warendorf (7). While several other German sites have yielded Middle Palaeolithic human remains, their age is still too uncertain to be considered here (7). This is also the case with an isolated frontal bone fragment recovered from the North

Sea (8). Further east, a small number of isolated teeth have been recently identified at Stajnia Cave (Poland; 9). In Belgium, the Trou de l'Abîme at Couvin and Walou Cave at Trooz have each yielded an isolated tooth, whereas the remains of two adults and one juvenile were discovered at Spy (10).

Supplementary Note S2. Assessment of the Goyet Middle Palaeolithic material and overview of the Late Mousterian from the Mosan Basin

The lack of field data from Dupont's excavations at the Troisième caverne makes it impossible to determine whether the typo-technologically Middle Palaeolithic material corresponds to one or several occupations. The latter possibility was suggested in the only, albeit incomplete study of this material by Ulrix-Closset (11), basing her conclusion on the diversity of surface alterations and several typological arguments. Ulrix-Closset noted that Levallois technology is poorly represented, with *débitage* involving primarily "spherical cores." Various scrapers forms are the most abundant retouched tool type, followed by 45 Mousterian points and the occasional limace. She also stressed the presence of bifacial scrapers alongside 28, often 'atypical' bifacial tools, three of which resemble foliate pieces (11). The numerous denticulates and raclettes present in the assemblage are more appropriately interpreted as edge-damaged artefacts, highlighting the significant post-depositional reworking of the deposits. Ulrix-Closset (11) concluded that the majority of the assemblage could be assigned to the Quina Mousterian ("Charentian") mixed with a smaller bifacial component representing either the Mousterian of Acheulean Tradition or an "evolved Mousterian." The latter was tentatively identified as the final phase of the Mosan Mousterian marked by foliate bifacial pieces (but see ref. 12 for a critical review of the "evolved Mousterian").

Several similarly dated late Middle Palaeolithic sites in the Mosan Basin provide insights concerning the Goyet lithic assemblage. At Scladina, the material from layer 1A contains evidence for different flake production systems similar to the Levallois, Quina and discoid methods that are adapted to the local raw materials (13). The assemblage from layer CI-8 of Walou Cave comprises primarily unifacial and Levallois *débitage* accompanied by a scraper-rich retouched tool component, including one Mousterian point (14). The material from Trou de l'Abîme at Couvin is made on a fine-grained, non-local flint that was heavily reduced. Scrapers are the most well-represented tool type and occur alongside small bifacial pieces (15).

Supplementary Note S3. Reassessment of the Goyet collections

In 2008, we began the revision of the Goyet human collections and systematic sorting of the faunal material from the Troisième caverne in order to identify any overlooked human remains. Focusing on the Dupont collections, and more specifically, a series of 21 drawers each measuring approximately 75 cm x 54 cm and containing "indeterminate" fauna (Supplementary Fig. S2), all material possibly associated with the Troisième caverne for which we had access was reassessed by two biological anthropologists (H.R and I.C; see Methods). Given the fragmentary nature of the human remains identified in these drawers, as well as clear anthropogenic marks indicating non-taphonomic (i.e. intentional), post-mortem fragmentation, the faunal collections were entirely resorted in order to isolate any additional skeletal fragments with similar

taphonomic features and morphometric characteristics (e.g. cortical thickness, medullar morphology). The refitting of specimens securely identified as human during the first sorting with non-diagnostic fragments isolated during this second phase confirmed the identification of the latter as human. Several of these newly identified human specimens were then selected for direct radiocarbon (^{14}C) dating. A palaeontological and taphonomic study of a sample of fragmentary, "indeterminate" fauna from the Dupont collection was also carried out. Additionally, human remains were sampled for stable isotope and genetic analysis, and additional samples were selected for dating based on the results of the previous analyses. The "indeterminate" fauna from Dupont's excavations was sorted a third time by another biological anthropologist (A.G.-O) who identified several additional human remains. Refits were attempted again to maximize the number of Neandertal remains.

At the start of the project, the collections held by the Anthropology and Prehistory Section of the Royal Belgian Institute of Natural Sciences (RBINS) included 70 bone specimens and 33 isolated teeth from Dupont's excavations at the Troisième caverne identified as human. In addition to identifying numerous new human remains among the faunal collections, our analysis excluded six bone fragments and a single tooth erroneously identified as human. The human remains from the Troisième caverne of Goyet now comprise 244 bone specimens and 39 isolated teeth. The remains of at least 16 individuals (nine adults/adolescents and seven juveniles) can be associated with levels 1 through 4 and represent a mix of materials from different periods. A fragment of human tibia from level 3 was previously dated to 1,985 +/- 70 years BP (OxA-5678) (16) and we identified human remains from three different periods of the Upper Palaeolithic amongst the material from levels 1 to 3 (17). Moreover, we identified 96 bone specimens and three isolated teeth that we attribute to Neandertals (Supplementary Table S1 and Supplementary Notes S4 and S5) and which are the focus of the present contribution.

Supplementary Note S4. Labelling and provenience of the Goyet Neandertal remains

Labelling system. The Neandertal specimens were labelled using several codes that reflect their recent research history. Some of the Neandertal remains (nine bone specimens and two isolated teeth) were already identified as human by E. Dupont and labelled 2878 or 2861 in his catalogue. These labels were subsequently amended with an additional number to individualize each specimen, probably when Fr. Twiesselmann was head of the Anthropology and Prehistory Section of the RBINS (Jadin, pers. com.). The remains numbered 2878 were labelled in red ink, with the additional number later added in black ink. The fact that the human remains numbered 2861 bear both numbers in black ink probably reflects their not having been originally labelled at the time of Dupont. We know that the numbers from Dupont's catalogue were lost before Twiesselmann arrived at the RBINS and that he asked one of his technicians to try to find the correspondence between Dupont's numbering system and the materials housed at the RBINS (Jadin, pers. com.). The numbers in black probably date from this time.

The remains labelled Qxxx-xx and Cx-x are specimens that we identified among the fauna recovered by E. Dupont from the Troisième caverne. Qxxx and Cx indicate the fauna drawers in which the human remains were found, and each of the specimens was individualized by adding a dash and a number after the

drawer number. These drawers are currently housed in two different reserves of the Palaeontology Section of the RBINS. Finally, one human specimen (1189-1) was identified in a drawer (no. 1189) of the archaeology collection held by the Anthropology and Prehistory Section of the RBINS.

A single tooth (1424-3D) was found amongst the material held by the Royal Museums of Art and History (RMAH, Brussels) from A. de Loë's early 20th century excavations at Goyet. Its identification number (1424) follows the RMAH inventory system for the portion of the Goyet collection it comes from. This was the third human tooth (3D) we isolated from this collection. Note that the ID of each human tooth from both the RBINS and RMAH includes a "D" (*dent* being French for tooth) to clearly differentiate them from the human bones.

Stratigraphic provenience. E. Dupont stated having discovered human remains in the upper four FBL of the Troisième caverne (2). The provenience of most of the Neandertal remains was indicated on small yellow labels glued to the bone fragments. While the remains numbered 2878 lack such labels, unpublished notes from the end of the 19th century kept in the RBINS archives contain a list of the remains labelled 2878 as well as the FBL from which they were recovered (i.e. FBL 1, 2 or 3). Comparison of the brief description of the remains in these notes with those labelled 2878 allowed us to correlate most of the remains with their FBL. Only a single Neandertal bone labelled 2861 (maxilla 2861-1) proved problematic. Dupont indicated in his unpublished handwritten notes that the remains numbered 2861 come from FBL 4 in Chamber C of the Troisième caverne and were found together with the remains of cave hyaena. Chamber C lies more than 100 m from Chamber A, where all the other Neandertal remains are supposed to originate, including the upper lateral incisor 2878-1D that refits with 2861-1 (Supplementary Table S1). Moreover, the latter shows a very different aspect from the other human remains labelled 2861, which are fresh and partially eroded, pointing to a different taphonomic history. We believe that maxilla 2861-1 may have been misplaced with the 2861 material after the excavations but before the remains numbered 2878 were inventoried and labelled at the end of the 19th century.

Supplementary Note S5. The Neandertal assemblage from the Troisième caverne

Identification. The Neandertal remains from the Troisième caverne were isolated from the rest of the human sample on the basis of their morphometric characteristics combined with their taphonomic aspect, isotopic ratios, radiocarbon dating and genetic analysis (see Table 1 and Supplementary Tables S1 and S3). Distinguishing morphometric traits include, the mastoid development on temporal C5-6; evidence for several missing wormian bones on the parietal fragment 2878-1; posterior position of the mental foramen on mandible 2878-8 and presence of the "horizontal-oval" type mandibular foramen on C5-3; crown and root morphology of the teeth; development of the manual phalanx extremity relative to the diaphysis; curvature of the diaphysis of radius I and constant height of the articular edge of the head of radius fragment Q116-3 along its circumference; length of the superior ramus of pubis Q48-1; presence of a developed gluteal buttress and a well-developed lesser trochanter on femur I; short length of the diaphysis of tibia I relative to its overall dimensions.

Age-at-death. The Neandertal sample is relatively homogeneous both in terms of size and robustness. Most of the Neandertal bone pieces are too poorly preserved to securely determine whether they are fully mature. However, they are all of adult size and compatible with an age-at-death during adolescence or adulthood. No modifications associated with senescence are visible. The stage of development and attrition of the dental material associated with maxilla 2861-1 (including the isolated upper left lateral incisor 2878-1D) and mandible 2878-8 (including the isolated lower left second premolar 2878-2D) is also compatible with an age-at-death during adolescence or young adulthood. The third isolated tooth (lower left lateral incisor 1424-3D) has an open root apex (stage A½ after ref. 18) pointing to an age-at-death between ca. 6.5 and 12.5 years according to modern standards (18, 19), making it the youngest Neandertal individual of the sample.

Minimum Number of Individuals (MNI). The Neandertal sample includes a child (see above) as well as several adolescent or adult individuals based on the representation and morphometrics of the tibia combined with the mtDNA analysis. Right tibias are represented by six pieces and left tibias by three (Supplementary Fig. S3).

- The three most complete right tibias (tibias III, IV and V) each represent a different individual as they all overlap in the area of the soleal line.
- Right tibia VI, a distal portion of the anterior diaphysis, not only has a different surficial aspect compared to the other right tibias but it also produced a different mtDNA sequence to those obtained from tibias III, IV and V. It thus represents a fourth individual. Left tibia II, a proximal portion of the anterior diaphysis, shares the same surficial aspect as right tibia VI as well as an identical mtDNA sequence for all covered positions (ca. 98 % of the mtDNA). Additionally, the morphometric characteristics of these two tibias are compatible with them belonging to the same individual. With the data at hand, tibia II thus does not represent an additional individual.
- The most complete tibia of the Neandertal sample, left tibia I, cannot be the antimere of tibias III, V or VI as its mtDNA sequence differs from those obtained for these bones nor can it be associated with left tibia II as they both preserve the area of the tibial tuberosity in addition to carrying different mtDNA sequences. Unfortunately, the mtDNA of tibia IV is not sufficiently preserved to determine whether it carried the same mtDNA sequence as that of tibia I. The morphometric characteristics and taphonomic aspect of the two bones are, however, compatible with their being corresponding antimeres, and we cannot exclude that they belong to the same individual.
- The remaining tibia specimens (Q54-5 –a proximal portion of right anterior diaphysis, Q375-2 –a proximal portion of left posterior diaphysis, and Q376-18 –a long portion of right anterior crest) are all compatible with belonging to at least one of the tibial elements identified above or their antimere, and thus cannot presently be attributed to additional individuals.

The Goyet Neandertal collection represents a minimum of five individuals: at least four different adolescent or adult individuals alongside the child represented by a single tooth (the lower left lateral incisor 1424-3D).

Individual associations. The number of elements from the Neandertal collection for which individual associations can be securely proposed is limited given the fragmentary nature of the collection. Left and right

tibias II and VI may belong to the same individual based on morphometric, taphonomic and genetic similarities (see above). They also share identical mtDNAs for all covered positions (ca. 98 % of the mtDNA) with femur II whose morphometric and taphonomic characteristics are compatible with it belonging to the same individual, hence we tentatively associated them. Femur I shares an identical mtDNA sequence with right tibias III and V, and its morphometric and taphonomic characteristics are compatible with it belonging to the same individual as one of these tibias. Finally, we tentatively associated left and right tibias I and IV based on morphometric and taphonomic similarities (see above).

Supplementary Note S6. Radiocarbon dating of the Goyet Neandertals

Radiocarbon dates are reported in years before present (BP) following the convention proposed by Mook and van der Plicht (20). These dates require calibration to obtain calendar ages. The presently recommended calibration curve is IntCal13 (21), and calibration was done using the OxCal software (version 4.2; ref. 22). Calibrated ages are reported in calBP, defined as calendar age relative to 1950 AD.

The first attempt at dating the Goyet Neandertal material concerned the roots of two teeth (2861-26D, the upper right P1 of maxilla 2861-1, and 2878-2D, the lower left P2 of mandible 2878-8). However, collagen was not extracted and the dates (GrA-46009: 27,070 +160, -150 BP and GrA-46010: 18,090 +80, -70 BP) obtained from dentine powder were too young given that the morphometric characteristics of the teeth securely identify them as Neandertal. Material being still available for 2878-2D, collagen extracted from this tooth was re-dated. Unfortunately, the second date (GrA-54028) also came back too young (Table 1), possibly due to undetected contamination (Supplementary Table S3). Tooth 2878-2D was part of the sample identified as human by E. Dupont and, although unnoticed during sampling, it may have been varnished like the rest of the human bones in Dupont's collection.

The dates obtained on the Goyet Neandertal bones range from 36,590 +300, -270 to 41,200 +500, -410 BP, or from 40.6 to 45.6 ky calBP at 2 sigmas. Although the ^{14}C dates of specimens Q57-1, Q57-2 and Q57-3 span the whole range of dates obtained for the Neandertal sample (Table 1), their morphology, taphonomy and mtDNA do not preclude their belonging to the same individual. Moreover, anthropogenic modifications on the Neandertal remains reflect similar behaviours as well as being located in the same position across the assemblage (see "Taphonomic analysis of the Goyet Neandertal material and anthropogenic modifications" section). These different lines of evidence suggest undetected contamination as the most likely explanation for the youngest ^{14}C ages, in which case the Goyet Neandertal sample would represent a single chronological group dated to ca. 44–45.5 ky calBP. In the absence of definitive evidence, we propose a conservative range of ca. 40.5–45.5 ky calBP for the Goyet Neandertals.

Supplementary Note S7. Palaeogenetic analyses of the Goyet Neandertals

Each bone fragment selected for palaeogenetic analysis was first irradiated with UV light in order to reduce surface DNA contamination. A dental drill was used to remove a thin layer of bone surface and to

sample inside the bone. An aliquot of between 30 mg and 120 mg of bone powder was utilized in the DNA extraction following an optimized protocol to retrieve typical short ancient DNA (aDNA) fragments (23). 10–20 µl out of 100 µl of extract were transformed into a sequencing library using a double stranded library preparation protocol (24) and indexed with an individual double index combination (25). Different amplification cycles were used for each indexed sequencing library in order to avoid heteroduplex products formation. Mitochondrial DNA (mtDNA) was subsequently enriched through a bead-capture protocol that uses modern human mtDNA fragments as baits (26). The enriched libraries were re-amplified, quantified on a DNA 1000 chip (Agilent), pooled in equal concentration with other samples, and sequenced on an Illumina HiSeq 2500 rapid run (2x101+8+8 cycles).

The sequenced reads were quality filtered and merged using established protocols (27). Only merged reads with a length above 30 base pairs (bp) were mapped against a Neandertal mitochondrial reference sequence using BWA (28) in combination with an in-house developed circular mapping tool (29). After removal of identical reads appearing more than once, sequences with a mapping quality score lower than 30 were excluded using the SAMtools software package (30). Only unique sequences securely placed within the mtDNA (Supplementary Table S7) were used to reconstruct the mitochondrial consensus sequence of each sample for positions with at least 5-fold coverage using the custom iterative assembler MIA (31).

In order to authenticate taxonomic assignment, reads of the three low coverage samples (C5-1, Q55-4 and Q119-2) were mapped against the modern human mitochondrial reference sequence (rCRS). The mapping results are in the same range as the values obtained using the Neandertal reference (Supplementary Table S8) therefore excluding a reference bias in the taxonomic assignment. Moreover, potential contamination with modern human DNA was assessed with a contamination estimation software that considers positions where the Neandertal mtDNA reference sequence differs from at least 99 % of 311 worldwide modern human mtDNAs (31). For each of these diagnostic positions, we compared the number of sequences matching the Neandertal reference better (clean fragments) than modern human mtDNAs (polluting fragments) to calculate contamination (Supplementary Table S7). Moreover, the characteristic damage pattern of aDNA was calculated as the percentage of reads showing C to T or G to A misincorporations respectively at the 5' and 3' ends of DNA fragments (Supplementary Table S7 and Supplementary Fig. S17) using a program first used in Briggs *et al.* (32).

The phylogenetic placement of the seven newly generated complete or almost complete mitochondrial sequences (i.e. at least 98 % complete) was assessed by comparing them to modern human, Neandertal and Denisovan mitochondrial genomes. The MUSCLE software (33) was used to align mtDNA consensus sequences of the Goyet specimens, 54 modern humans belonging to different worldwide language groups (34), eight Neandertals (31, 35–37), and one Denisovan individual (38). A maximum parsimony tree and a maximum likelihood tree with complete deletions (16,110 positions considered) and 1,000 iterations as bootstrap support were built using MEGA 5.2 (39) and refined with FigTree (ref. 40; Fig. 2 and Supplementary Fig. S4).

Subsequently, reads with nucleotide misincorporations (postmortem damage or PMD score ≥ 3)

indicating authentic ancient origin (Supplementary Table S7) were selected with PMD tools (37). The percentage of filtered fragments with damaged termini increased up to 71 % whereas contamination decreased for all samples (Supplementary Table S7). Only the filtered reads of the Goyet samples represented in Fig. 2 were used to generate new mitochondrial consensus sequences (for positions covered at least 5 times) that were aligned to the same Neandertal mtDNA reference and assembled in additional maximum parsimony and maximum likelihood trees (Supplementary Figs. S18 and S19) using the same parameters mentioned above (10,234 positions considered). This confirmed the phylogenetic placement of the original reconstructed mitochondrial consensus sequences within the Neandertal mtDNA diversity and validated the intragroup matrilineal relationships.

Supplementary Information References

1. van den Broeck, E., Martel, E. A. & Rahir, E. *Les Cavernes et les Rivières Souterraines de la Belgique Etudiées Spécialement dans leurs Rapports avec l'Hydrologie des Calcaires et avec la Question des Eaux Potables. Tome 2: Les Calcaires Carbonifériens du Bassin de Dinant et Coup d'Oeil sur le Bassin de Namur* (H. Lamertin, 1910).
2. Dupont, E. *L'Homme Pendant les Ages de la Pierre dans les Environs de Dinant-sur-Meuse, 2ème Edition* (C. Muquardt Ed., 1872).
3. Germonpré, M. A reconstruction of the spatial distribution of the faunal remains from Goyet, Belgium. *Notae Praehistoricae* **21,** 57–65 (2001).
4. Germonpré, M. Preliminary results on the mammals of the Magdalenian upper horizon of Goyet (Belgium). *Notae Praehistoricae* **16,** 75–85 (1996).
5. Germonpré, M. & Hämäläinen, R. Fossil bear bones in the Belgian Upper Paleolithic: The possibility of a proto-bear-ceremonialism. *Arctic Anthropol.* **44,** 1–30 (2007).
6. Germonpré, M. & Sablin, M. V. The cave bear (*Ursus spelaeus*) from Goyet, Belgium. The bear den in Chamber B (bone horizon 4). *Bull. Inst. R. Sc. N. B.* **71,** 209–233 (2001).
7. Street, M., Terberger, T. & Orschiedt, J. A critical review of the German Paleolithic hominin record. *J. Hum. Evol.* **51,** 551–579 (2006).
8. Hublin, J.-J. *et al.* Out of the North Sea: the Zeeland Ridges Neandertal. *J. Hum. Evol.* **57,** 777–785 (2009).
9. Nowaczewska, W. *et al.* The tooth of a Neanderthal child from Stajnia Cave, Poland. *J. Hum. Evol.* **64,** 225–231 (2013).
10. Pirson, S. *et al.* Chronostratigraphic context of the Middle to Upper Palaeolithic transition: Recent data from Belgium. *Quat. Int.* **259,** 78–94 (2012).
11. Ulrix-Closset, M. *Le Paléolithique Moyen dans le Bassin Mosan en Belgique* (Universa, 1975).
12. Flas, D. *La Transition du Paléolithique Moyen au Supérieur dans la Plaine Septentrionale de l'Europe* (Anthropologica et Praehistorica 119, 2008).
13. Di Modica, K. *Les Productions Lithiques du Paléolithique Moyen en Belgique: Variabilité des Systèmes d'Acquisition et des Technologies en Réponse à une Mosaïque d'Environnements Contrastés*. Thèse de doctorat (MNHN, Université de Liège, 2010).
14. Draily, C. *La Grotte Walou à Trooz (Belgique). Fouilles de 1996 à 2004. Volume 3: L'Archéologie* (Etudes et Documents – Archéologie 22, Institut du Patrimoine Wallon, 2011).
15. Toussaint, M. *et al.* The Neandertal lower right deciduous second molar from Trou de l'Abîme at Couvin, Belgium. *J. Hum. Evol.* **58,** 56–67 (2010).
16. Bronk Ramsey, C., Higham, T. F. G., Owen, D. C., Pike, A. W. G. & Hedges, R. E. M. Radiocarbon dates from the Oxford AMS System: Datelist 31. *Archaeometry* **44(3), Suppl 1,** 1–149 (2002).
17. Rougier, H. *et al.* The first Upper Paleolithic human remains from Belgium: Aurignacian, Gravettian and Magdalenian fossils at the "Troisième caverne" of Goyet. *PaleoAnthropology* **2013,** A 33 (2013).
18. Moorrees, C. F. A., Fanning, E. A. & Hunt, E. E. Age variation of formation stages for ten permanent teeth. *J. Dent. Research* **42,** 1490–1502 (1963).
19. AlQahtani, S. J., Hector, M. P. & Liversidge, H. M. Brief communication: The London atlas of human tooth development and eruption. *Am. J. Phys. Anthrop.* **142,** 481–490 (2010).
20. Mook, W. G. & van der Plicht, J. Reporting ^{14}C activities and concentrations. *Radiocarbon* **41,** 227–239 (1999).
21. Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. *Radiocarbon* **55,** 1869–1887 (2013).
22. Bronk Ramsey, C. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51,** 337–360 (2009).
23. 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).
24. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **6,** pdb.prot5448 (2010).
25. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40,** e3 (2012).
26. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5,** e14004 (2010).
27. Kircher, M. in *Ancient DNA. Methods and Protocols*. (eds Shapiro, B. & Hofreiter, M.) 197–228 (Humana Press, 2012).
28. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
29. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17,** 60 (2016).
30. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
31. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-

throughput sequencing. *Cell* **134,** 416–426 (2008).
32. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104,** 14616–14621 (2007).
33. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
34. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708–713 (2000).
35. Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325,** 318–321 (2009).
36. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
37. 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).
38. Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464,** 894–897 (2010).
39. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731–2739 (2011).
40. Rambaut, A. *FigTree v1. 3.1: Tree figure drawing tool*. (2009) Available at: http://tree.bio.ed.ac.uk/software/figtree/.
41. Wißing, C. *et al.* Isotopic evidence for dietary ecology of Late Neandertals in North-Western Europe. *Quat. Int.* doi: 10.1016/j.quaint.2015.09.091 (2015).
42. Lyman, R. L. Quantitative units and terminology in Zooarcheology. *Am. Antiq.* **59,** 36–71 (1994).
43. Sawyer, G. J. & Maley, B. Neanderthal reconstructed. *Anat. Rec. B New Anat.* **283B,** 23–31 (2005).
44. Barone, R. *Anatomie Comparée des Mammifères Domestiques, Tome I Ostéologie – Atlas* (Vigot, 1976).
45. Fontana, L. Mobilité et subsistance au Magdalénien dans le Bassin de l'Aude. *Bull. Soc. Préhist. Fr.* **96,** 175–190 (1999).

LETTER

doi:10.1038/nature19844

Genomic insights into the peopling of the Southwest Pacific

Pontus Skoglund[1,2,3], Cosimo Posth[4,5], Kendra Sirak[6,7], Matthew Spriggs[8,9], Frederique Valentin[10], Stuart Bedford[9,11], Geoffrey R. Clark[11], Christian Reepmeyer[12], Fiona Petchey[13], Daniel Fernandes[6,14], Qiaomei Fu[1,15,16], Eadaoin Harney[1,2], Mark Lipson[1], Swapan Mallick[1,2], Mario Novak[6,17], Nadin Rohland[1], Kristin Stewardson[1,2,18], Syafiq Abdullah[19], Murray P. Cox[20], Françoise R. Friedlaender[21], Jonathan S. Friedlaender[22], Toomas Kivisild[23,24], George Koki[25], Pradiptajati Kusuma[26], D. Andrew Merriwether[27], Francois-X. Ricaut[28], Joseph T. S. Wee[29], Nick Patterson[2], Johannes Krause[5], Ron Pinhasi[6]§ & David Reich[1,2,18]§

The appearance of people associated with the Lapita culture in the South Pacific around 3,000 years ago[1] marked the beginning of the last major human dispersal to unpopulated lands. However, the relationship of these pioneers to the long-established Papuan people of the New Guinea region is unclear. Here we present genome-wide ancient DNA data from three individuals from Vanuatu (about 3,100–2,700 years before present) and one from Tonga (about 2,700–2,300 years before present), and analyse them with data from 778 present-day East Asians and Oceanians. Today, indigenous people of the South Pacific harbour a mixture of ancestry from Papuans and a population of East Asian origin that no longer exists in unmixed form, but is a match to the ancient individuals. Most analyses have interpreted the minimum of twenty-five per cent Papuan ancestry in the region today as evidence that the first humans to reach Remote Oceania, including Polynesia, were derived from population mixtures near New Guinea, before their further expansion into Remote Oceania[2–5]. However, our finding that the ancient individuals had little to no Papuan ancestry implies that later human population movements spread Papuan ancestry through the South Pacific after the first peopling of the islands.

Pacific islanders today derive from a mixture of two highly divergent ancestral populations[3]. The first ancestral modern human population arrived in island southeast Asia more than 40,000 years before present (BP), and contributed to the ancestry of both indigenous Australians and Papuans, and hence to other Pacific islanders[4]. The second ancestral population is more closely related to mainland East Asians[4], and is not found in unadmixed form today. The first humans to reach Remote Oceania—a term we use to refer to the region unoccupied before approximately 3,000 BP beyond the main Solomon Islands and, in this case, excluding Micronesia—were associated with the Lapita culture, which existed between 3,450–3,250 and 2,700–2,500 BP. These people spread into Remote Oceania using the first boats capable of long-distance sea travel and introduced new domesticated animals and plants, and their successors reached the most isolated islands of the eastern and southern Pacific by 1,000–700 BP[6]. Several hypotheses have been proposed to explain why present-day indigenous people of Near Oceania (New Guinea, the Bismarck Islands, and the Solomon Islands area) and Remote Oceania have ancestry both from Papuans and from populations of ultimate East Asian origin. In one set of models that has been favoured by recent genetic studies[3–5,7], the mixture occurred at around 3,000 BP, during the expansion of populations of East Asian origin through the New Guinea region[8]. In the other set of models, the population of ultimate East Asian origin initially mixed little with Papuans[9], and thus later gene exchanges account for the ubiquitous Papuan ancestry today[2,10].

We obtained genome-wide ancient DNA data from three individuals from Teouma, an archaeological site on Efate island, Vanuatu (Supplementary Information section 1), which were all directly radiocarbon dated to between 3,110 and 2,740 BP, an interval that is chronologically part of the Lapita period (Extended Data Table 1). We also obtained genome-wide ancient DNA data from an individual from the Talasiu site on Tongatapu island, Tonga, directly radiocarbon dated to 2,680–2,340 BP, a period spanning the late Lapita and immediately post-Lapita period (Supplementary Information section 2 and Extended Data Table 1). In dedicated clean rooms, we prepared powder from petrous bones[11], extracted DNA[12], and prepared up to four double-stranded libraries from each extract[13]. We enriched the libraries for 1.24 million targeted single nucleotide polymorphisms (SNPs)[14], sequenced the products, and represented each individual by a single randomly drawn sequence for each SNP. This procedure resulted in 139,461–231,944 SNPs that were covered at least once in each of the individuals. The low ratio of sequences aligning to Y-chromosome targets compared to targets on other chromosomes[15] reveals that all four individuals are females (Extended Data Table 1). We obtained three mitochondrial DNA sequences from Vanuatu and all were haplogroup B4a1a1a, the classic 'Polynesian motif'[16].

Multiple features of the data suggest that the DNA was authentic and minimally contaminated. First, in all individuals, around 40% of all sites that are cytosines in the human reference sequence appear as thymines in the terminal nucleotide, as expected for genuine ancient

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [3]Archaeological Research Laboratory, Department of Archaeology and Classical Studies, Stockholm University, 10691 Stockholm, Sweden. [4]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen 72070, Germany. [5]Max Planck Institute for the Science of Human History, 07745 Jena, Germany. [6]School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Dublin, Ireland. [7]Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. [8]School of Archaeology and Anthropology, College of Arts and Social Sciences, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. [9]Vanuatu National Museum, Vanuatu Cultural Centre, Port Vila, Vanuatu. [10]Maison de l'Archéologie et de l'Ethnologie, CNRS, UMR 7041, 92023 Nanterre, France. [11]Department of Archaeology and Natural History, College of Asia and the Pacific, The Australian National University, Canberra, Australian Capital Territory 2601, Australia. [12]College of Arts, Society and Education, James Cook University, Queensland 4870, Australia. [13]Radiocarbon Dating Laboratory, University of Waikato, Hamilton 3240, New Zealand. [14]CIAS, Department of Life Sciences, University of Coimbra, 3000-456 Coimbra, Portugal. [15]Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, IVPP, CAS, Beijing 100044, China. [16]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. [17]Institute for Anthropological Research, 10000 Zagreb, Croatia. [18]Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. [19]RIPAS Hospital, Bandar Seri Begawan, Brunei Darussalam. [20]Institute of Fundamental Sciences, Massey University, Palmerston North 4442, New Zealand. [21]Independent Scientist, Sharon, Connecticut 06069, USA. [22]Department of Anthropology, Temple University, Gladfelter Hall, Philadelphia, Pennsylvania 19122, USA. [23]Estonian Biocentre, Evolutionary Biology group, Tartu, 51010, Estonia. [24]Division of Archaeology, University of Cambridge, Fitzwilliam Street, Cambridge CB2 1QH, UK. [25]Papua New Guinea Institute of Medical Research, Goroka, Eastern Highlands Province 441, Papua New Guinea. [26]Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia. [27]Department of Anthropology, Binghamton University, Binghamton, New York 13902, USA. [28]Evolutionary Medicine Group, Laboratoire d'Anthropologie Moléculaire et Imagerie de Synthèse UMR 5288 CNRS, Université de Toulouse, Toulouse 31073, France. [29]National Cancer Centre Singapore, Singapore 169610, Singapore.
§These authors jointly supervised this work.

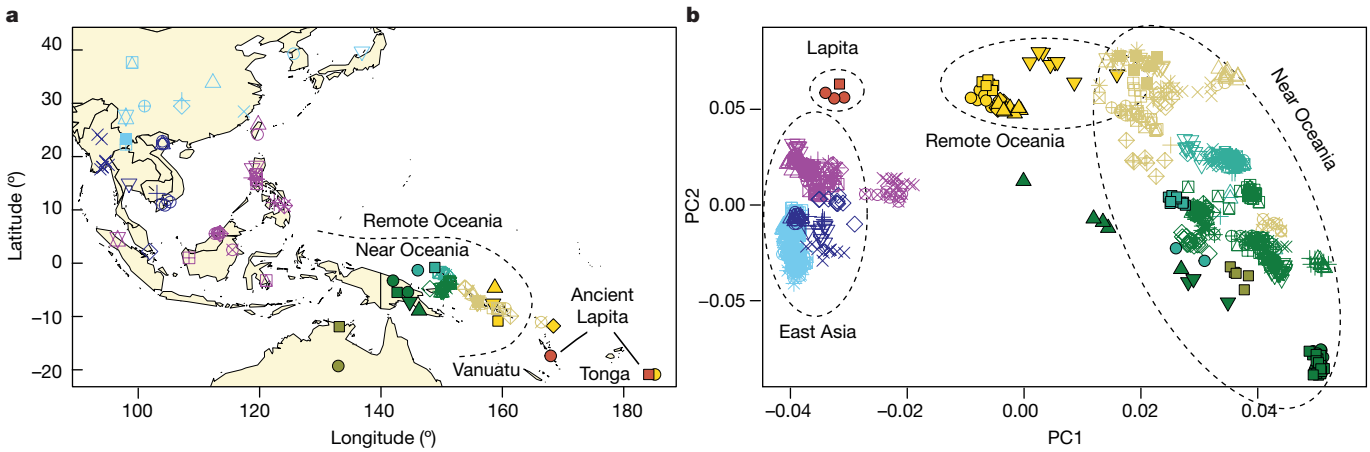


Figure 1 | Data from ancient and present-day populations. a, Locations of 778 present-day individuals genotyped on the Affymetrix Human Origins Array and 4 ancient individuals (red symbols). **b**, Ancient individuals projected onto principal components (PC) 1 and 2 computed using only present-day samples. Individual population labels are given in Extended Data Fig. 2.

DNA (Extended Data Fig. 1a). Second, when we carried out principal component analysis (PCA; Fig. 1) of 778 present-day people from 83 East Asian and Oceanian populations genotyped at 621,799 SNPs (of which 356 individuals from 38 groups were newly genotyped for this study; Extended Data Table 2) and projected the ancient individuals, we found that all clustered tightly with each other and with data from the same individuals restricting to sequences with cytosine-to-thymine changes at the terminal nucleotide (these sequences are unlikely to be contaminants[17,18]) (Extended Data Fig. 1b). Third, the cluster of ancient individuals does not overlap with present-day populations, indicating that the data are from a population that is not present in unmixed form today (Fig. 1). The distinctiveness of the ancient individuals is also highlighted by their high differentiation from all present-day groups ($0.05 < F_{ST} < 0.26$; between all modern individuals and the ancient Vanuatu individuals, using the statistic F_{ST}, which compares within- and between-group squared allele frequency differences) (Extended Data Table 3).

The ancient Vanuatu and Tongan individuals are not shifted in the PCA in the direction of Papuan ancestry, in contrast to all present-day Remote Oceanians. In this respect, they are similar to indigenous Taiwanese populations such as the Ami and Atayal as well as to populations from the Philippines such as the Kankanaey, who have no detectable Papuan ancestry (Fig. 1). To test whether the ancient individuals had any evidence of Papuan ancestry, we used the *qpWave*/*qpAdm* software (Methods) to analyse allele frequency correlation statistics[19]. The results were consistent with the ancient individuals and the Taiwanese Ami having descended from a common ancestral population to the exclusion of 14 worldwide outgroups ($P > 0.05$ for the ancient individuals from both Vanuatu and Tonga). We estimate the possible range of Papuan ancestry in the Vanuatu individuals to be 0–11% and in the Tongan individual to be 0–17% (99% confidence intervals truncated at zero), which does not overlap the point estimates of at least 25% Papuan ancestry in all present-day Oceanians (Fig. 2a). To test the hypothesis that the ancient Remote Oceanian individuals might be from the source population of the non-Papuan ancestry in Oceanians today, we computed the statistic f_4(Africa, *Test*; Australian, Polynesian), which evaluates the degree of allele sharing of a candidate *Test* population with Polynesians (at sites where Polynesians differ

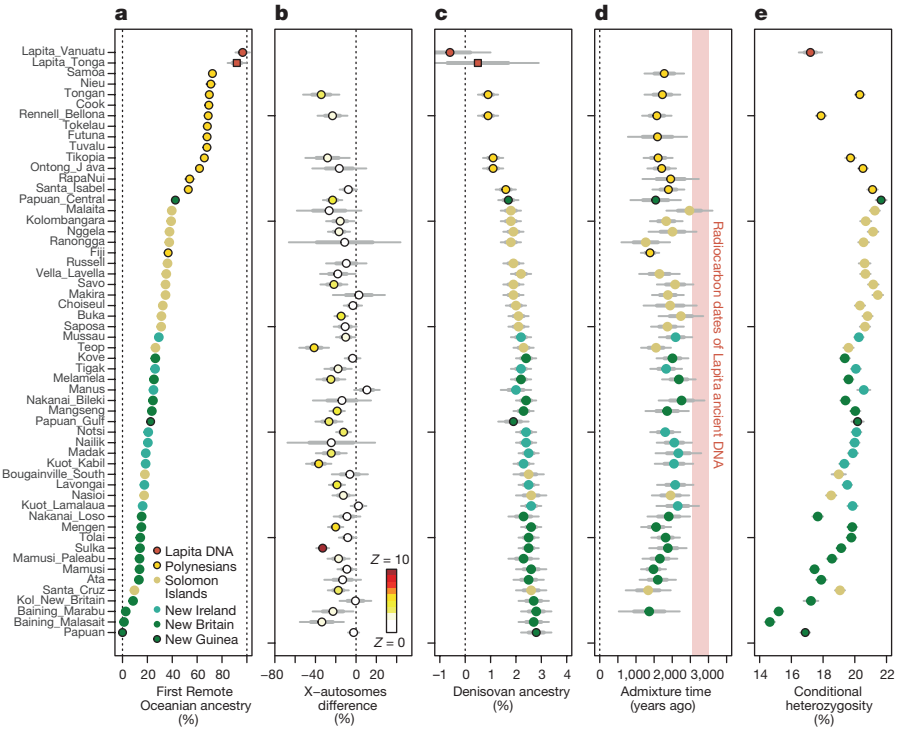


Figure 2 | Genetic characteristics of the Oceanian ancestry cline. a, Estimated proportion of First Remote Oceanian ancestry. The Papuan ancestry can be estimated as 100% minus the estimate of First Remote Oceanian ancestry. **b**, Difference between First Remote Oceanian ancestry estimates on chromosome X and the autosomes. **c**, Denisovan ancestry estimates are inversely related to First Remote Oceanian ancestry estimates. **d**, Estimated date of admixture in all populations with at least four individuals and significant evidence of decay of weighted admixture linkage disequilibrium as measured in ALDER. We used Han and New Guinean Highlanders as surrogates for the ancestral populations. We assumed a generation interval of 28.1 years, and show 95% confidence intervals (thin whiskers) incorporating uncertainty both in the ALDER date and the value of the human generation interval. We show the range of radiocarbon dates for the ancient individuals. **e**, Conditional heterozygosity (genetic diversity) estimated by drawing two random chromosomes from different individuals at each locus, using only SNPs ascertained in a single Yoruba, and restricting to transversion SNPs to avoid any concerns about inflated heterozygosity owing to ancient DNA degradation. Thick and thin error bars in all five panels correspond to 1 and 1.96 standard error of the estimate, respectively.

from Australians), and found that it was maximized when *Test* was the ancient Vanuatu or Tonga individuals (Extended Data Fig. 2b), as expected if a population related to them was the true source. We conclude that the non-Papuan ancestry that is ubiquitous in Oceania is derived from a population related to the ancient individuals we analysed, and that this ancestry reached uninhabited islands in Remote Oceania with little or possibly no mixture with Papuans. We call the population of which both the ancient Vanuatu and Tongan individuals were a part the 'First Remote Oceanians' and find that the ancestry fraction from this population is the single most important factor shaping genetic variation among Pacific islanders, accounting for most variation in measurements including genetic diversity (Pearson's $R = 0.86$, $P = 2 \times 10^{-12}$ for 42 non-Polynesian groups; Extended Data Fig. 2) and the proportion of archaic Denisovan ancestry ($R = -0.96$, $P < 10^{-16}$ for all 56 Oceanian groups; Fig. 2).

Our evidence that early and geographically diverse Remote Oceanian individuals had little or no Papuan ancestry contradicts models in which there were significant Papuan contributions to Lapita people before their dispersal into Remote Oceania[3–5]. Instead, our results show that the Papuan genetic signature appeared in many Remote Oceanian populations only subsequent to initial settlement. To gain further insight into when the Papuan ancestry may have become ubiquitous in Remote Oceanians, we leveraged the fact that chromosome segments from ancestral populations break up at a known rate owing to recombination and that the length distribution of these segments translates to a date of mixture[20]. We estimate dates of approximately 50–80 generations ago using ALDER[21], or 1,500–2,300 BP assuming 28.1 years (see Methods) per generation[22] (Fig. 2d and Extended Data Fig. 3). We combined the statistical error of the genetic estimate and the uncertainty about the generation interval (Methods), and obtained a 95% confidence interval of 1,239–1,927 BP for a pool of Polynesians, all of whom have similar Papuan ancestry proportions. This finding that Papuan–First Remote Oceanian mixture occurred long after the end of the Lapita period implies that the Polynesian ancestral population was not fully formed at that time, although we note that alternative methods for dating Papuan admixture in Remote Oceanians arrived at older dates[4,23–25]. However, our ALDER dates are supported by direct ancient DNA evidence, as the Tongan individual at 2,680–2,340 BP carried little or no Papuan ancestry, providing unambiguous confirmation that the ancestral population of Polynesians was not fully formed or widespread by the end of the Lapita period.

We used *qpGraph* to explore models of population separation and mixture that might accommodate the ancient DNA data[26] (Supplementary Information section 3). We obtained fits using models in which Polynesians today are mixtures of First Remote Oceanians and a Papuan population related to Highland New Guineans (Fig. 3a). We also obtained consistent findings using *TreeMix*[27] (Extended Data Fig. 4). In Fig. 3 we show the best fitting model, which suggests that the ancient individuals from Vanuatu and Tonga descended from an ancestral (presumably Lapita) population that separated earlier from the population that is the primary component in present-day Polynesians. This implies that not just Papuan ancestry but also deeply branching First Remote Oceanian ancestry was introduced to Remote Oceania through movement of people after the time of the ancient individuals. Thus, the minimum 25% Papuan ancestry seen in present-day Remote Oceanians is a conservative underestimate of the later population displacement. It is unlikely that there was 100% replacement, however, as we observed weak excess affinity of present-day Tongans to the ancient Tongan individuals in symmetry tests (see Methods). More deeply in time, our modelling indicates that Philippine populations (Kankanaey) are the closest outgroup to the First Remote Oceanians, indigenous Taiwanese (Atayal) second closest, and mainland southeast Asians such as the Dai most remote, consistent with models of population movement along a route from Taiwan to the Philippines to Near Oceania to Remote Oceania[28]. We were surprised that we could not fit Australians as outgroups to New Guinean Highlanders and the

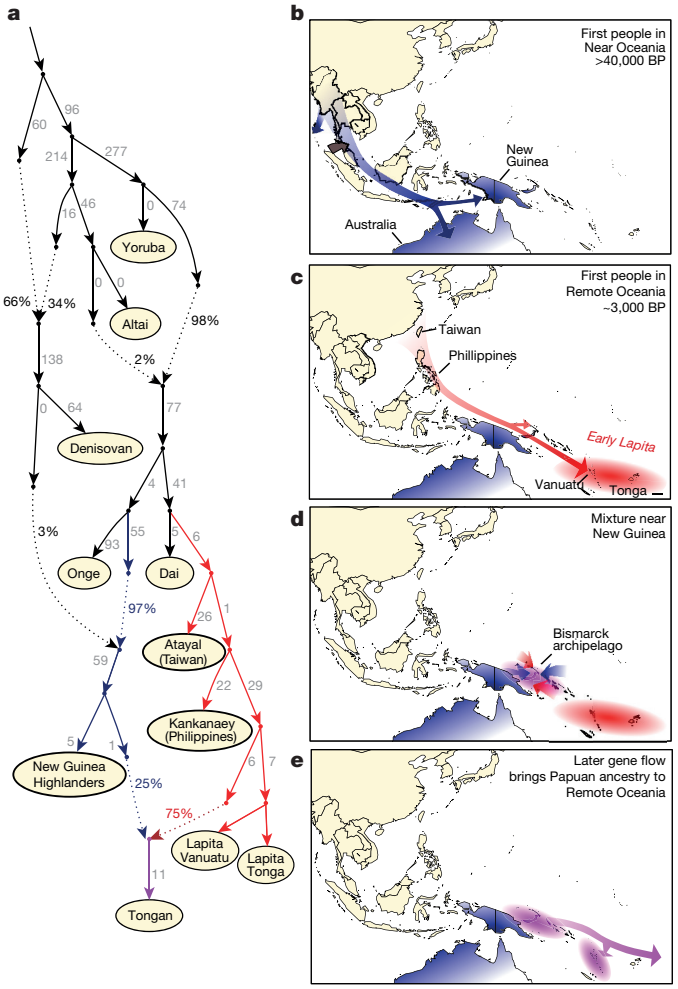


Figure 3 | A model of population history. a, A model of population relationships that fits allele frequency patterns (all empirical *f*-statistics within 3 standard errors of expectation). Branch lengths are shown in units of $F_{ST} \times 1{,}000$. Admixture edges show mixture proportions. Altai, the Altai Neandertal genome. **b**, A model of population movements more than 40,000 years ago in which modern humans arrived in the Australia–New Guinea region (blue shading) and mixed with archaic Denisovans (brown arrow). **c**, A model of events before 3,000 years ago, in which the First Remote Oceanian population formed by spread of a population of ultimate East Asian origin to a region including Vanuatu and Tonga, and experienced little or no mixture with the Papuans they encountered along the journey (red shading). Note that geographic routes are speculative. **d**, A model of populations of mixed Papuan–First Remote Oceanian ancestry in Near Oceania less than 3,000 years ago in a patchwork of islands with different proportions of First Remote Oceanian ancestry (pink shading). **e**, A model of secondary expansion of admixed populations bringing Papuan ancestry into Remote Oceania, which was still not complete in Tonga by the date of the Talasiu individual at 2,680–2,340 BP.

Papuan ancestry in Polynesians (Extended Data Fig. 5). However, we could fit Australians as deriving from a mixture of an ancient Australian lineage and a Papuan lineage from the same group that expanded into Polynesia. This is plausible if there was continuing gene flow between New Guinea and Australia. Another parsimonious model is that the ancestry in present-day Polynesians is not all Papuan, but a Papuan–Australian mix.

Previous studies of mitochondrial DNA and Y chromosomes suggested that present-day people of the South Pacific harbour more East Asian ancestry from female than from male ancestors[3]. Our genome-wide analyses confirm a significant excess of First Remote Oceanian ancestry on the X chromosome compared to the autosomes (*Z* scores up to 10) (Fig. 2b). Females carry two-thirds of the X chromosomes in

a population but only half of the autosomes (Extended Data Fig. 6), and we compared the ancestry estimates in these two parts of the genome to obtain the most accurate estimates of sex-biased admixture in diverse Oceanians to date (Extended Data Fig. 6 and Extended Data Table 4). It has been suggested that matrilocal social structure in the primarily First Remote Oceanian ancestry populations of the region is one likely factor to explain these patterns[29,30]. However, it is also possible that some of these patterns reflect a scenario in which the later movement of Papuan ancestry into Remote Oceania was largely mediated by males who then mixed with resident females.

Our study has shown that many of the first humans in Remote Oceania had little, if any, Papuan ancestry, in stark contrast to the situation today. While our findings cannot rule out the possibility that multiple groups—some of which carried substantial amounts of Papuan ancestry—settled Remote Oceania early on, the lack of such ancestry in both Vanuatu and Tonga can be more parsimoniously explained by later population movements bringing the Papuan ancestry. The scenario emerging from ancient DNA analysis is thus radically different from that suggested by previous genetic studies, which have generally posited that the first people in Remote Oceania and Polynesia[2–5] had substantial Papuan ancestry. Our finding of major post-Lapita flow of Papuan ancestry into Remote Oceania also cannot be related to the later arrival of Papuan ancestry that has been suggested for Fiji, which is estimated to have occurred at least a millennium later at 500 BP[4] or 1,100 BP[24] (Fig. 2). Systematic study of ancient DNA from throughout Remote Oceania should make it possible to provide a detailed chronicle of the population movements and sex-biased population mixtures that shaped the ancestry of present-day Oceanians.

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 20 April; accepted 13 September 2016.
Published online 3 October 2016.

1. Sheppard, P. J., Chiu, S. & Walter, R. Re-dating Lapita movement into Remote Oceania. *J. Pacific Archaeol.* **6,** 26–36 (2015).
2. Kayser, M. *et al.* Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am. J. Hum. Genet.* **82,** 194–198 (2008).
3. Kayser, M. The human genetic history of Oceania: near and remote views of dispersal. *Curr. Biol.* **20,** R194–R201 (2010).
4. Wollstein, A. *et al.* Demographic history of Oceania inferred from genome-wide data. *Curr. Biol.* **20,** 1983–1992 (2010).
5. Matisoo-Smith, E. Ancient DNA and the human settlement of the Pacific: a review. *J. Hum. Evol.* **79,** 93–104 (2015).
6. Bellwood, P. S. *First Farmers: the Origins of Agricultural Societies* (Blackwell Publishing, 2005).
7. Duggan, A. T. *et al.* Maternal history of Oceania from complete mtDNA genomes: contrasting ancient diversity with recent homogenization due to the Austronesian expansion. *Am. J. Hum. Genet.* **94,** 721–733 (2014).
8. Kayser, M. *et al.* Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* **10,** 1237–1246 (2000).
9. Blust, R. Remote Melanesia: one history or two? An addendum to Donohue and Denham. *Oceanic Linguistics* **47,** 445–459 (2008).
10. Friedlaender, J. S. *et al.* The genetic structure of Pacific Islanders. *PLoS Genet.* **4,** e19 (2008).
11. Pinhasi, R. *et al.* Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One* **10,** e0129102 (2015).
12. 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).
13. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Phil. Trans. R. Soc. Lond. B* **370,** 20130624 (2015).
14. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524,** 216–219 (2015).
15. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40,** 4477–4482 (2013).
16. Melton, T. *et al.* Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* **57,** 403–414 (1995).
17. Skoglund, P. *et al.* Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336,** 466–469 (2012).
18. 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).
19. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488,** 370–374 (2012).
20. Moorjani, P. *et al.* The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* **7,** e1001373 (2011).
21. Loh, P.-R. *et al.* Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193,** 1233–1254 (2013).
22. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **128,** 415–423 (2005).
23. Xu, S., Pugach, I., Stoneking, M., Kayser, M. & Jin, L. Genetic dating indicates that the Asian-Papuan admixture through Eastern Indonesia corresponds to the Austronesian expansion. *Proc. Natl Acad. Sci. USA* **109,** 4574–4579 (2012).
24. Pugach, I., Matveyev, R., Wollstein, A., Kayser, M. & Stoneking, M. Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biol.* **12,** R19 (2011).
25. Lipson, M. *et al.* Reconstructing Austronesian population history in Island Southeast Asia. *Nat. Commun.* **5,** 4689 (2014).
26. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192,** 1065–1093 (2012).
27. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8,** e1002967 (2012).
28. Bellwood, P. Holocene population history in the Pacific region as a model for worldwide food producer dispersals. *Curr. Anthropol.* **52,** S363–S378 (2011).
29. Jordan, F. M., Gray, R. D., Greenhill, S. J. & Mace, R. Matrilocal residence is ancestral in Austronesian societies. *Proc. R. Soc. Lond. B* **276,** 1957–1964 (2009).
30. Stephen Lansing, J. *et al.* An ongoing Austronesian expansion in Island Southeast Asia. *J. Anthropol. Archaeol.* **30,** 262–272 (2011).

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

Acknowledgements We thank the 356 volunteers who donated samples for genome-wide analysis; M. Stoneking for co-funding genotyping of the Bismarck samples; M. Brilliant, H. Norton, and L. Scheinfeldt, for help in the preparation of the Bismarck samples and establishment of a repository for them at the Marshfield Foundation; A. Wissgott for help in data generation from the ancient Tongan individual; A. Kim, I. Pugach, and M. Stoneking for comments, and I. Mathieson for critiques and advice on estimating sex-specific ancestral contributions. The maps in Figs 1a and 3b–e maps were plotted in R using the *world()* map of the 'fields' and 'maps' packages (using public domain data from the CIA World Data Bank II). P.S. was supported by the Wenner-Gren foundation, SciLifeLab, and the Swedish Research Council (VR grant 2014-453). The Teouma research by M.S. and S.B. was supported by the Australian Research Council (Discovery Grants DP0880789 and DP110101415), the National Geographic Society, and the Australia-Pacific Science Foundation. F.V. was supported by CNRS-UMR 7041. M.N. was supported by an Irish Research Council grant (GOIPD/2013/1). D.F. was supported by an Irish Research Council grant (GOIPG/2013/36). Q.F. was funded by the Key Research Program of Frontier Sciences of CAS (QYZDB-SS W-DQC003), the National Natural Science Foundation of China (L1524016) and the Chinese Academy of Sciences Discipline Development Strategy Project (2015-DX-C-03). T.K. was supported by ERC starting grant FP7-261213. C.P. and J.K. were supported by the Baden Wuerttemberg Foundation. J.K was supported by the DFG grant KR 4015/1-1 and the Max Planck Society. R.P. was supported by ERC starting grant ADNABIOARC (263441). D.R. was supported by NIH grant GM100233, by NSF HOMINID BCS-1032255, and is a Howard Hughes Medical Institute investigator.

Author Contributions N.P., J.K., R.P. and D.R. supervised the study. M.S., F.V., S.B., G.R.C., and C.R. assembled archaeological material and information. P.S., C.P., Q.F., M.L., S.M., N.R. and D.R. analysed genetic data. C.P., K.Si., F.P., D.F., E.H., M.N., N.R, and K.St. performed laboratory work. S.A., M.P.C., F.R.F., J.S.F., T.K., G.K., P.K., D.A.M., F-X.R., and J.T.S.W. assembled the sample collection from present-day populations. P.S. and D.R. wrote the manuscript with major input from C.P., M.S., F.V., G.R.C., M.P.C., J.S.F, J.K. and R.P. and additional input from all other co-authors.

Author Information The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB14728. The newly reported SNP genotyping data for the subset of individuals who provided informed consent consistent with fully public distribution are available at http://genetics.med.harvard.edu/reichlab/Reich_Lab/Datasets.html. To access data for the remaining samples, researchers should send a signed letter to D.R. containing the following text: "(a) I will not distribute the data outside my collaboration; (b) I will not post the data publicly; (c) I will make no attempt to connect the genetic data to personal identifiers for the samples; (d) I will use the data only for studies of population history; (e) I will not use the data for any selection studies; (f) I will not use the data for medical or disease-related analyses; (g) I will not use the data for commercial purposes." Extended Data Table 2 specifies which samples are consistent with which type of data distribution. 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 P.S. (skoglund@genetics.med.harvard.edu), R.P. (ron.pinhasi@ucd.ie) or D.R. (reich@genetics.med.harvard.edu).

Reviewer Information *Nature* thanks P. Bellwood, C. Capelli and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Ancient DNA sampling, extraction, library preparation, enrichment and sequencing. The Vanuatu skeletal samples B30A, B10B, B17 were analysed with permission from the Vanuatu National Museum and the excavators of the Teouma site. The Tonga skeletal sample SK10 was analysed with permission from the excavators of the Talasiu site.

All preparation of skeletal samples, DNA extraction, and library preparation was carried out in dedicated ancient DNA laboratories at University College Dublin, Ireland (sample preparation of the three Vanuatu individuals), at Harvard Medical School in Boston, USA (DNA extraction and library preparation of the three Vanuatu individuals), and at the Max Planck Institute for the Science of Human History in Jena, Germany (sample preparation, DNA extraction and library preparation of the Tonga individual). Each of these facilities is spatially separated from other molecular biology laboratories, and measures are taken to protect ancient individuals from contamination including HEPA filtered air, head-to-toe suits, face masks with visors, multiple layers of gloves, bleaching of all surfaces, ultraviolet light (UVC) decontamination of (non-sensitive) consumables and chemicals, and UVC decontamination of the facility when researchers are not in the room[31]. The final step of the library preparation (amplification) was performed outside the ancient DNA laboratory.

We prepared powder from the cochlea of petrous bones, extracted DNA[12], and prepared libraries with standard protocols (ref. 13 for the Vanuatu individuals and ref. 32 for the Tonga individual). For the three Vanuatu individuals, the first library was prepared in the presence of uracil DNA glycosylase (UDG) to cut out errors due to ancient DNA damage, whereas the remaining three libraries as well as the Tonga library were prepared without UDG as this preserves more DNA for any given sample. We performed in-solution enrichment using previously reported protocols[13,14,33,34] for a targeted set of 1,237,207 SNPs that comprises two previously separately reported sets of 394,577 SNPs[34] and 842,630 SNPs[14]. We sequenced the product on Illumina NextSeq500 instruments for 2×75 cycles. Following demultiplexing, and, for the Vanuatu samples, removal of both oligonucleotide barcodes that were used to identify the libraries and trailing adaptor sequences, we merged the forward and reverse reads of each read pair requiring a 15-base pair overlap (allowing one mismatch). We then aligned merged sequences to the human genome *hg19* using *BWA 0.6.1* (ref. 35). We removed sequences aligned to identical outer coordinates, choosing the highest quality sequence for each duplication cluster. We merged the data from the four libraries for each Vanuatu individual.

Genomic analysis. We determined sex by comparing the number of X and Y chromosome alignments[15]. We estimated damage patterns using *PMDtools* v0.60[18], separating damage patterns observed inside and outside a CpG context. Since all four individuals were female, we could not estimate contamination using X chromosome data. We investigated whether there was evidence of excess relatedness between any pair of individuals among the Vanuatu individuals, but found that the pairwise mismatch rate using panel 5 of the Affymetrix Human Origins array (see below) was $19.8\% \pm 0.4\%$ for I1368/I1369, $19.7\% \pm 0.6\%$ for I1368/I1370, and $20.5\% \pm 0.4\%$ for I1369/I1370. This suggests no atypical pair of individuals and a similar within-population mismatch rate (heterozygosity) as some present-day Polynesian populations (Fig. 2).

Genotyping of present-day humans. We genotyped 356 individuals from 38 southeast Asian and Oceanian populations on the Affymetrix Human Origins array (Extended Data Table 2). The individuals all contributed DNA samples voluntarily and provided informed consent consistent with studies of human genetic variation and history. Ethical approval of the component studies was provided by the Singapore Health IRB, the Research Ethics Committee at the Facultés de Médecine de Toulouse, the Brunei Medical and Health Research Ethics Committee, the University of Cambridge Biology Research Ethics Committee, the Government of Papua New Guinea Medical Research Advisory Committee, and the Temple University IRB. The collection of genome-wide variation data on de-identified samples was approved by the Harvard Human Research Protection Program (Protocol 11681), re-reviewed on 12 July 2016.

We restricted analysis to samples that had >95% genotyping completeness and that were not visual outliers in PCA with respect to the main cluster of samples in the group. We merged with previously reported Affymetrix Human Origins SNP array data[26,36–39]. We also co-analysed our data with samples genotyped on the Affymetrix 6.0 platform where we removed three previously published[39] Rapa Nui individuals (*5s5j*, *XB3B*, and *3p3p*), and two previously published[40] Samoan individuals (*PLY_07* and *PLY_11*), all of which appeared to have recent European ancestry based on clustering analyses. We finally compared our data to high-coverage genomes from an archaic Neanderthal and an archaic Denisovan, both from Denisova Cave in the Altai Mountains of Siberia[41–43].

Population genetic analysis. When overlapping with the Affymetrix Human Origins SNP array data set of present-day human populations, we have between 74,000 and 126,000 SNPs covered at least once for each of the four individuals (Extended Data Table 1). This is more than the minimum coverage required for high-resolution analysis using allele frequency correlation statistics, e.g. 10,000 SNPs per individual according to Supplementary Information section 6.2 of ref. 44, a study that had the same median coverage ($0.19\times$) as ours (the range in the present study is $0.14–0.26\times$). For all analyses, we called genotypes by randomly sampling a single non-duplicate sequence read at each position[45]. This procedure is standard for analysis of low-coverage ancient DNA data and is also often used for higher-coverage data to minimize reference genome biases that can be introduced when determining diploid genotypes[14,17,34,36,41,44–50]. For the *qpAdm, qpWave* and *qpGraph* analyses we excluded transition SNPs to avoid potential biases from post-mortem damage (see below).

We performed PCA using *smartpca*[51], with the option *inbreed: YES* in order to sample a single genotype from each individual randomly to match the pseudo-haploid nature of the ancient DNA genotypes from the ancient individuals[52]. We computed f_3-, f_4- and D-statistics as in ref. 26, and F_{ST} using the Hudson estimator and randomly sampled a single haploid sequence to represent each individual at each SNP position, using *popstats*[38]. We estimated the date of admixture using ALDER[21]. We tested the consistency of a matrix of f_4-statistics with one or more sources of ancestry with respect to a set of outgroups (New_Guinea, Denisova, Sardinian, English, Yakut, Chukchi, Mala, Japanese, Ju_hoan_North, Mixe, Onge, Yoruba, and Mbuti) using *qpWave*[19,34].

For the ancient individuals and all present-day populations genotyped on the Human Origins array, we used *qpAdm*[34], which estimates ancestry proportions from two or more proxy source populations assuming that the proxies are more closely related to the real source populations than they are to a set of outgroups (*qpAdm* also provides a formal statistical test for whether this is the case, which passes in the context that we use it here). We estimated First Remote Oceanian and Papuan ancestry using Denisova, Sardinian, English, Yakut, Chukchi, Mala, Japanese, Ju_hoan_North, Mixe, Onge, Yoruba, and Mbuti as outgroups and New_Guinea and Ami as proxies for the Papuan and First Remote Oceanian source populations, respectively. For the ancient individuals, we excluded all transition SNPs to avoid possible biases due to post-mortem damage, resulting in 35,194 transversion SNPs for Vanuatu (covered by at least one of the individuals) and 22,030 for Tonga. For estimating *qpAdm* ancestry proportions in the Affymetrix 6.0 Polynesian data, we used whole-genome sequences from the same populations as outgroups[53]. We estimated Denisovan ancestry using the Denisovan genome and Japanese as the two sources, and chimpanzee (*Pan troglodytes*), Ju_hoan_North, Mbuti, Yoruba, Dinka and the Altai Neanderthal genome as outgroups.

We computed conditional heterozygosity using panel 5 of the Affymetrix Human Origins array, which contains SNPs ascertained as heterozygous in a single West African Yoruba individual. This provides an unbiased estimate of relative heterozygosity since the Yoruba individual is approximately symmetrically related to all Oceanians (Denisovan ancestry violates this assumption but is not expected to change the ranking of populations). We estimated heterozygosity as the average pairwise mismatch rate when sampling 2 chromosomes from two different individuals using *popstats*[38], restricting to transversion SNPs for all populations and computing standard errors using a weighted block jack-knife.

For authentication, we used *PMDtools*[18] to extract sequences with clear evidence of post-mortem damage patterns (PMD score of at least 3), disregarding individual bases with phred-scaled base quality <30. We randomly sampled new pseudo-haploid genotypes from the resulting set of sequences and projected the ancient individuals onto the principal components inferred from the present-day populations as above. After this filtering, we retained 68,450 SNPs for I1368; 98,722 SNPs for I1369; 83,024 SNPs for I1370; and 117,023 SNPs for CP30. The ninety-nine per cent confidence intervals for *qpAdm* estimates of Papuan ancestry (see above) using the PMD score-restricted data were 0–21% for the ancient Vanuatu individuals and 0–24% for the ancient Tonga individual, consistent with the confidence intervals obtained from the full data.

To test whether the ancient Vanuatu and the ancient Tonga individuals form a clade, we used *qpWave* to test whether a model of Dai, Ami, Kankanaey and a fourth population were consistent with being outgroups to the two ancient sample groups (we used Dai, Ami and Kankanaey as these span present-day Mainland East Asia, Taiwan, and the Philippines, and lack Papuan ancestry to the limits of our resolution). The analysis used the ~12,000 SNPs that remained after excluding transition SNPs and SNPs missing in one of the two ancient sample groups. We found that the model was consistent with the data for all tested Oceanian and Asian populations shown in Fig. 1, but that the lowest P value was observed for present-day Tongans ($P = 0.09$). We also found that f_4(*Ami, Present day Tongan; Lapita_Vanuatu,Lapita_Tonga*) $= 0.006$, $Z = 3.2$, when using all SNPs. This suggests a possible affinity between present-day Tongans and the ancient Tongan individual, consistent with the hypothesis that the ancient population of Tonga with little or no Papuan ancestry may have contributed some of the ancestry of present-day Tongans.

Admixture date estimation. To estimate the date of historical admixture between First Remote Oceanians and Papuans, we used ALDER[21,25] on the full Human Origins array data, with New Guinean Highlanders and Han Chinese as the two sources. We use Han Chinese for this analysis owing to their substantial sample size compared to populations more closely related to the ancestral First Remote Oceanian population such as the ancient individuals we analysed, indigenous Taiwanese, and indigenous Philippine groups. ALDER estimates are known to be robust even when using imperfect surrogates for the ancestral populations in this way[26]. We estimate an admixture date for a pool of Polynesian populations by combining data from Tongan, Tikopia, Russell and Bellona populations, all genotyped on the Affymetrix Human Origins SNP array.

ALDER and other methods based on admixture linkage disequilibrium estimate dates in units of generations, which need to be converted to years. For this purpose, we require an estimate of the generation interval—the average age of a parent at the time their gametes were formed—weighted by the fraction of recombination events that occur in each sex (62.3% of all autosomal crossovers are estimated to occur in females, based on table 1 of ref. 54.). Using estimates from the anthropological literature, this quantity is 27.8 years for hunter–gathering societies, 28.6 years for developed nation states, and 29.6 years for less developed nation states[22]. These numbers are in the range of the point estimate we use of 28.1 years based on breakdown of admixture linkage disequilibrium in radiocarbon-dated ancient genomes[55]. To account for the substantial variability in generation intervals across human societies, we use the sample standard error of 2.15 years measured across eleven diverse hunter-gatherer groups based on Table 4 of ref. 22. The date estimates in Fig. 2 and Extended Data Fig. 4 thus use a generation interval of 28.1 years, and combine the standard error from ALDER (a) with the uncertainty in generation time, that is, $\sqrt{a^2 \times 2.15^2 + A^2 \times 2.15^2 + 28.1^2 \times a^2}$, where A is the ALDER point estimate in number of generations.

We do not subtract 66 years from the dates produced by ALDER to obtain BP dates (conventionally the date before 1950 AD, 66 years ago), because what ALDER is estimating is a number that is close to the BP date. To see this, note that ALDER estimates the date between when chromosomes of the two ancestries began crossing over (one generation after mixing began), and the date of the last crossover (when the germ cells that mixed to produce the present-day samples in our study were formed, likely one or two generations before 2016 AD). Accounting for these corrections means that ALDER is estimating a date of mixture that is likely to be within a generation of the true BP date.

Fitting models of population history. We used *qpGraph*[26,56] to assess the fit of admixture graph models to allele frequency correlation patterns as measured by f_2, f_3-, and f_4-statistics. We started with a skeleton phylogenetic tree consisting of Yoruba, New_Guinea, Dai, Atayal, Kankanaey and the pool of ancient Vanuatu individuals. We added Tongan, Mamanwa (a Philippine Negrito group), Nasioi and Kolombangara, respectively, to all possible edges in the tree, and retained only the graph solutions that provided no individual f_4 statistics with $|Z| > 3$ between empirical and predicted statistics. For the extended version of the admixture graph, we also added Australians to all possible edges of the graph that included these populations. Finally, we modelled the previously documented admixture history relating Denisovans and the Altai Neanderthal genome to the outgroup chimpanzee and the anatomically modern human populations, to which we added the Andamanese Onge and the ancient Tongan individual. The final graph visualized in Fig. 3 used 10,893 SNPs after restricting to transversion SNPs to avoid complications due to ancient DNA damage and also SNPs with coverage in all groups. For more information on the admixture graph inference procedure, see Supplementary Information section 3.

As an alternative inference method, we used *Treemix* v1.12 (ref. 27) to test models for Yoruba, Dai, Atayal, Kankanaey, Tongan, New Guinean Highlanders, the ancient Vanuatu individual and the ancient Tongan individual. The total number of SNPs after excluding transitions, SNPs with minor allele count of less than 4 in the selected data, and SNPs where one population had missing data, was 10,119, which we divided into 337 blocks of 30 consecutive SNPs each to estimate the covariance matrix. We first fitted a maximum likelihood tree of all populations, but found that several of the fitted allele frequency covariances deviated from those empirically observed by up to 16.4 standard errors. We then used the automated heuristic optimization in *Treemix* to infer a graph model with one admixture event using the same populations, and found that the optimal fit was for a model with an admixture event in the history of Tongans, where one portion of their ancestry diverged before the split of the ancestors of the ancient Vanuatu and Tonga individuals, and the other (25% ± 3%) derived from the New Guinean lineage. This maximum deviation between empirical and model covariances observed for the graph with one admixture edge was 1.6, indicating a good fit, consistent with our investigation of models using *qpGraph*.

Female and male ancestral contributions. To estimate the proportion of female ancestors (F) and male ancestors (M) for a given population, we used two different methods both based on the estimates of ancestry for the X chromosome and autosomes. Both used the same underlying model, in which the observed admixture proportion estimates that $\hat{H}_{auto}$ and $\hat{H}_X$ for the autosomes and X chromosome, respectively, depend on M and F such that:

$$\hat{H}_{auto} = \left(\frac{M}{2} + \frac{F}{2}\right) \quad (1)$$

$$\hat{H}_X = \left(\frac{M}{3} + \frac{2F}{3}\right) \quad (2)$$

The first approach obtains unbounded point estimates of M and F by rearranging equations:

$$M = 4 \times \hat{H}_{auto} - 3 \times \hat{H}_X \quad (3)$$

$$F = 3 \times \hat{H}_X - 2 \times \hat{H}_{auto} \quad (4)$$

Similarly, we obtained standard errors for M and F using the weighted block jackknife standard errors for $\hat{H}_{auto}$ and $\hat{H}_X$, SE_{auto} and SE_X, as

$$\mathrm{SE}_M = \sqrt{(9 \times \mathrm{SE}_X{}^2) + (16 \times \mathrm{SE}_{auto}{}^2)} \quad (5)$$

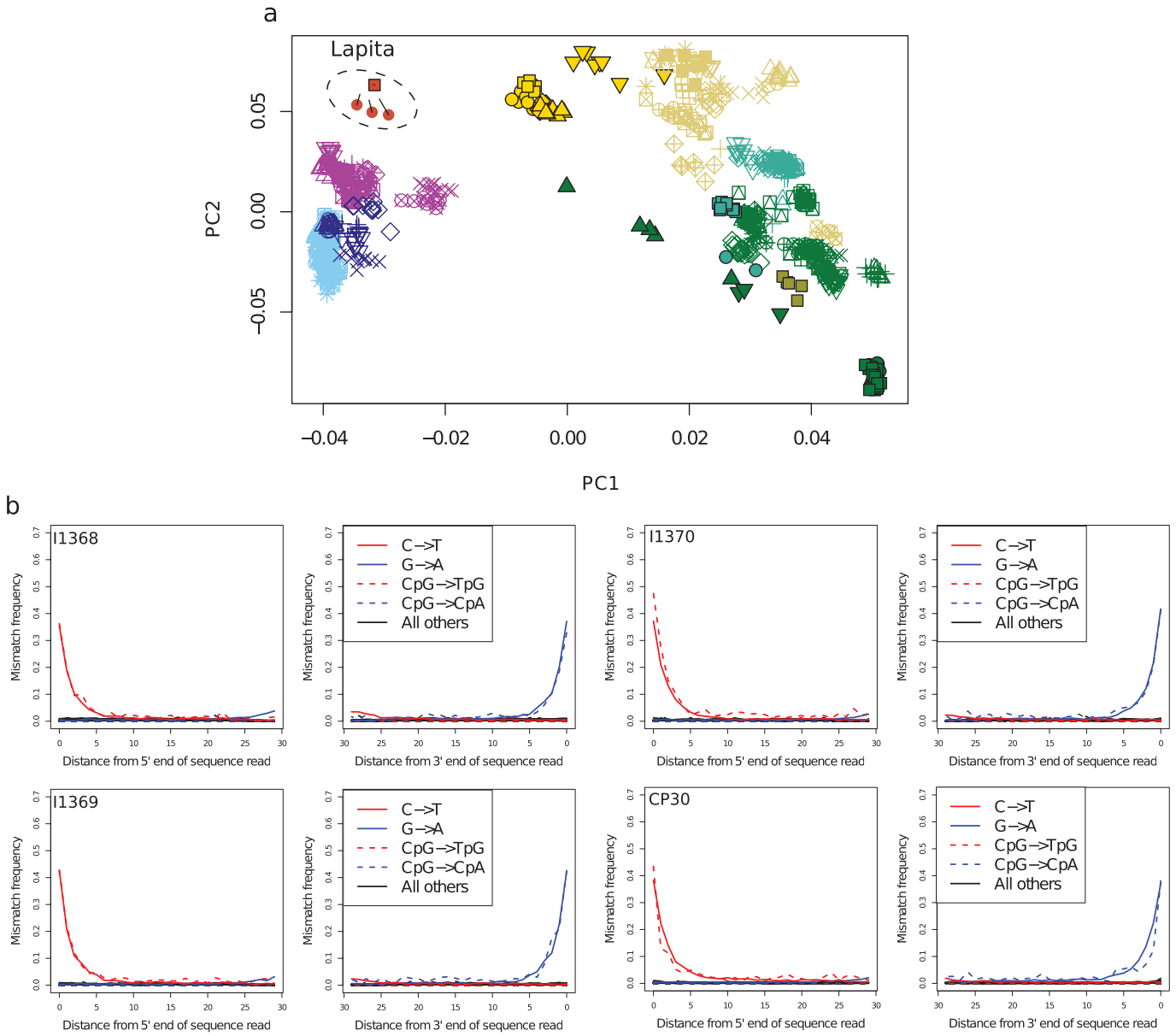
$$SE_F = \sqrt{(9 \times \mathrm{SE}_X{}^2) + (4 \times \mathrm{SE}_{auto}{}^2)} \quad (6)$$

As an alternative to estimating M and F, we took an approximate Bayesian approach by performing 1 million simulations in which M and F were sampled from a uniform prior distribution (0, 1). We then simulated ancestry estimates specifying normal distributions with means and standard errors matching the empirical values (equations 1 and 2). We used the *abc* R package[57] to run a rejection algorithm retaining the 1% of all simulation replicates with the closest Euclidean distances to the empirical $\hat{H}_{auto}$ and $\hat{H}_X$, and performed local linear regression on log-transformed summary statistics to obtain a posterior distribution. The results of the two methods are qualitatively similar. In Extended Data Fig. 6, we plot the posterior intervals of these distributions for selected populations.

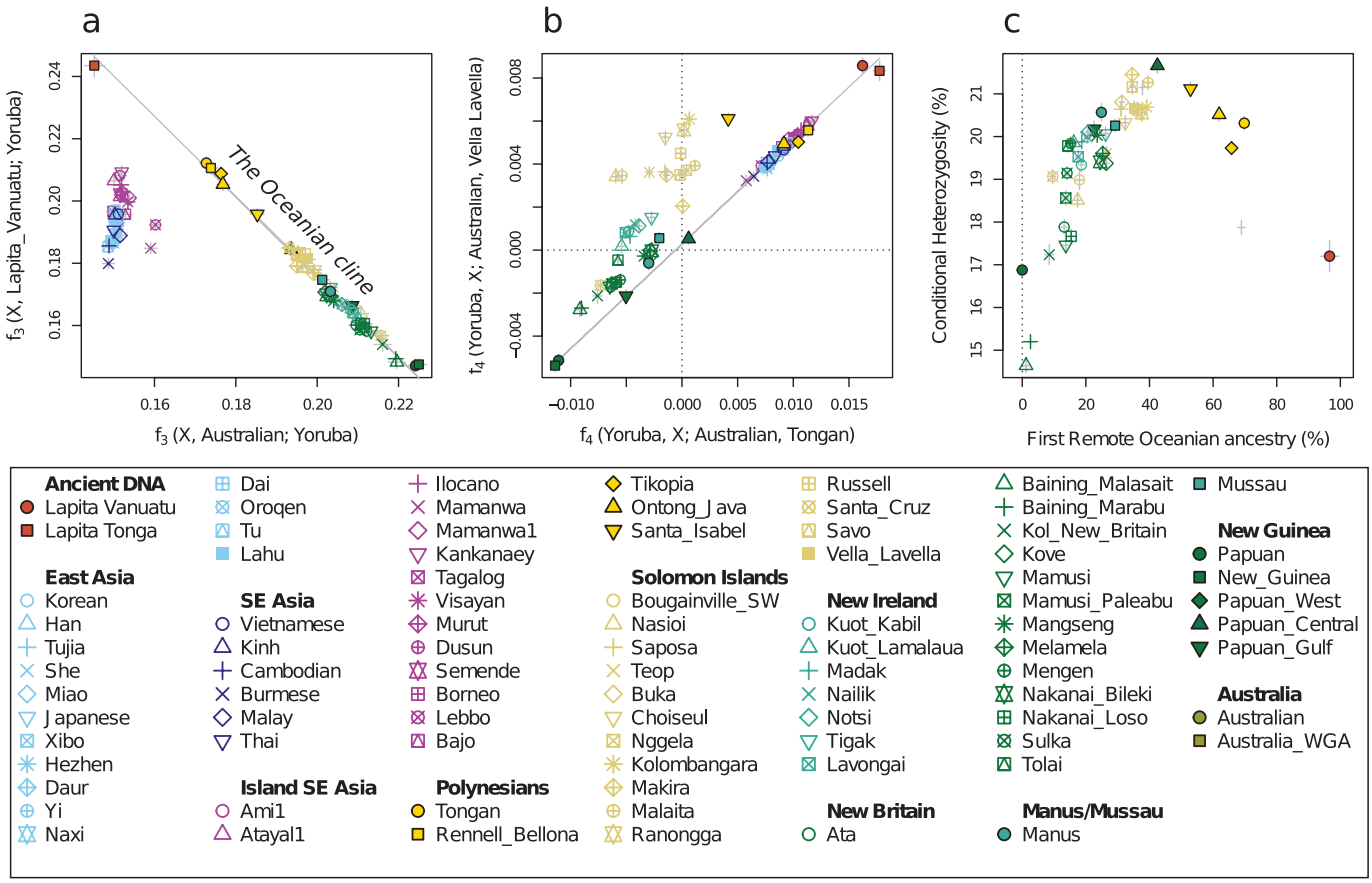
Sample size. No statistical methods were used to predetermine sample size.

31. Knapp, M., Clarke, A. C., Horsburgh, K. A. & Matisoo-Smith, E. A. Setting the stage – Building and working in an ancient DNA laboratory. *Ann. Anat. Anat. Anz.* **194,** 3–6 (2012).
32. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Prot.* http://dx.doi.org/10.1101/pdb.prot5448 (2010).
33. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110,** 2223–2227 (2013).
34. Haak, W. *et al.* Massive migration from the steppe is a source for Indo-European languages in Europe. *Nature* **522,** 207–211 (2015).
35. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
36. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513,** 409–413 (2014).
37. Qin, P. & Stoneking, M. Denisovan ancestry in East Eurasian and Native American populations. *Mol. Biol. Evol.* **32,** 2665–2674 (2015).
38. Skoglund, P. *et al.* Genetic evidence for two founding populations of the Americas. *Nature* **525,** 104–108 (2015).
39. Moreno-Mayar, J. V. *et al.* Genome-wide ancestry patterns in Rapanui suggest pre-European admixture with Native Americans. *Curr. Biol.* **24,** 2518–2525 (2014).
40. Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89,** 516–528 (2011).
41. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468,** 1053–1060 (2010).
42. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338,** 222–226 (2012).
43. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505,** 43–49 (2014).
44. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522,** 167–172 (2015).
45. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328,** 710–722 (2010).
46. Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505,** 87–91 (2014).
47. Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345,** 1255832 (2014).

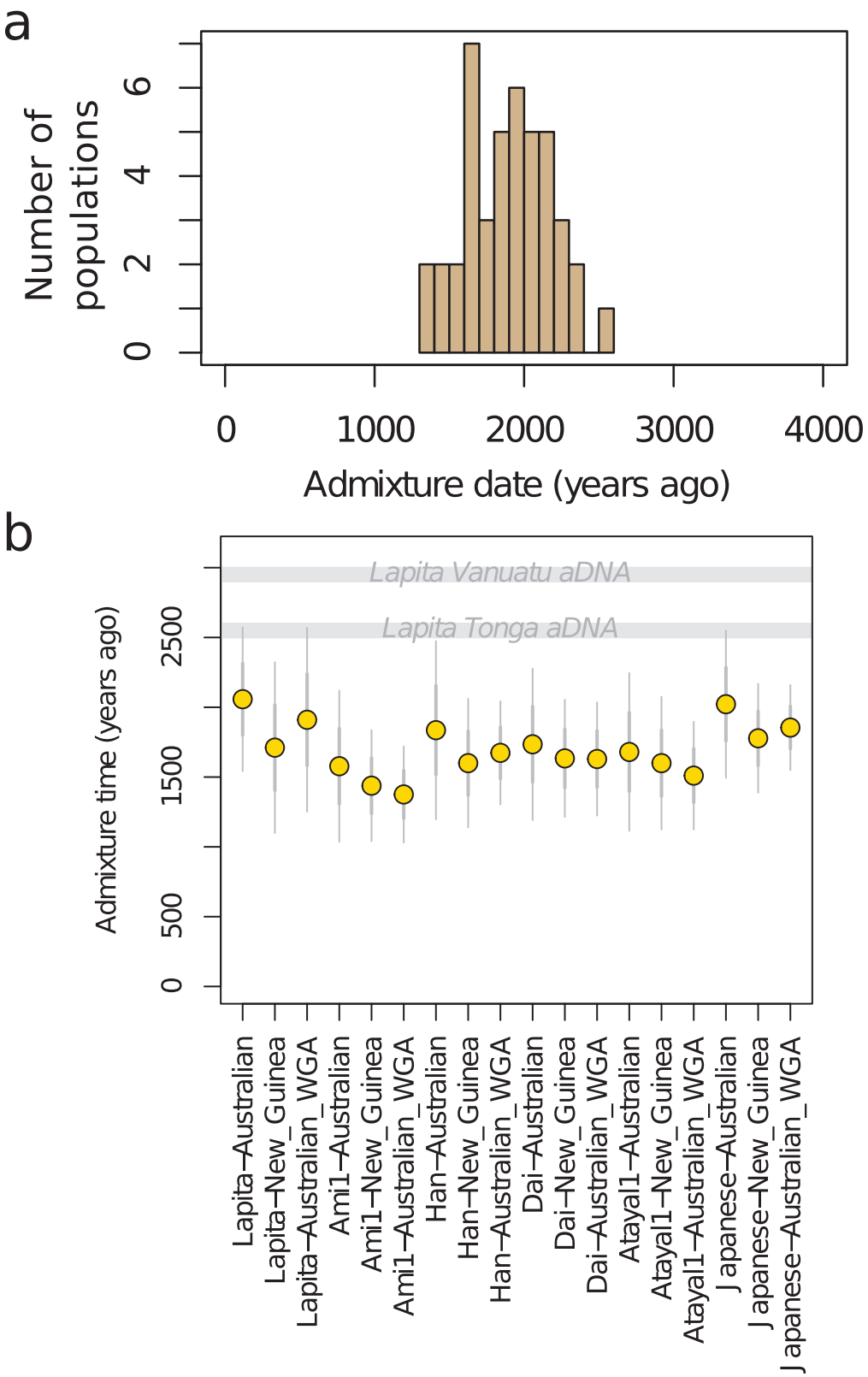
48. Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506,** 225–229 (2014).
49. Rasmussen, M. *et al.* The ancestry and affiliations of Kennewick Man. *Nature* **523,** 455–458 (2015).
50. Skoglund, P. *et al.* Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344,** 747–750 (2014).
51. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
52. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **45,** 18301–18306 (2011).
53. Sudmant, P. H. *et al.* Global diversity, population stratification, and selection of human copy-number variation. *Science* **349,** 1174–1181 (2015).
54. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nat. Genet.* **31,** 241–247 (2002).
55. 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. *Proc. Natl Acad. Sci. USA* **113,** 5652–5657 (2016).
56. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461,** 489–494 (2009).
57. Csilléry, K., Blum, M. G. B., Gaggiotti, O. E. & François, O. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* **25,** 410–418 (2010).
58. Bronk Ramsey, C. *OxCal Program v4.2.4.* (Radiocarbon Accelerator Unit, Univ. Oxford, 2016).
59. Reimer, P. J. *et al.* IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 Years cal BP. *Radiocarbon* http://dx.doi.org/10.2458/azu_js_rc.55.16947 (2013).
60. Ambrose, S. H. Isotopic analysis of paleodiets: methodological and interpretive considerations in *Food and Nutrition in History and Anthropology (USA)* http://agris.fao.org/aos/records/US9514697?output=xml (1993).
61. Petchey, F., Spriggs, M., Bedford, S., Valentin, F. & Buckley, H. Radiocarbon dating of burials from the Teouma Lapita cemetery, Efate, Vanuatu. *J. Archaeol. Sci.* **50,** 227–242 (2014).
62. Petchey, F., Anderson, A., Zondervan, A., Ulm, S. & Hogg, A. New marineΔ R values for the South Pacific subtropical gyre region. *Radiocarbon* **50,** 373–397 (2008).



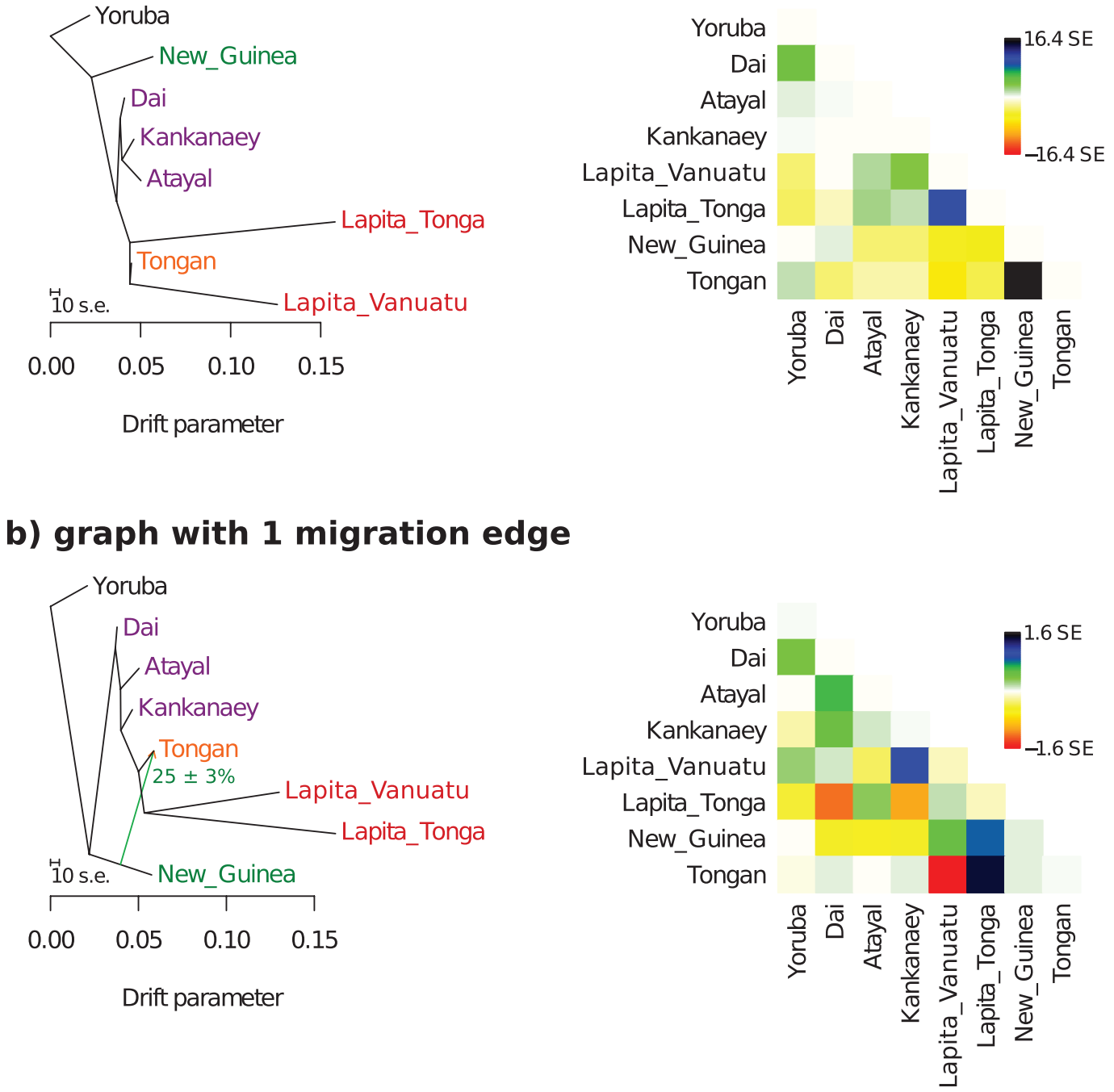
Extended Data Figure 1 | Ancient DNA authenticity. a, PCA performed as for Fig. 1, but with the four ancient individuals represented only by sequences that show clear evidence of post-mortem damage (PMD score of at least 3) to remove contaminating sequences that might be present[17,18]. The numbers of SNPs remaining after restriction to damaged sequences is 68,450 SNPs for I1368; 98,722 SNPs for I1369; 83,024 SNPs for I1370; and 117,023 SNPs for CP30. The lines indicate the projection of the samples when no damage-restriction is performed. The large number of SNPs retained, and the fact that the ancient individuals cluster tightly and have the same qualitative positioning in the plot as Fig. 1, indicates that contamination did not contribute to our findings. We also find that estimates of Papuan ancestry using PMD score restricted data are consistent with those obtained using the full data (see Methods). **b**, Post-mortem damage patterns for genome-wide in-solution enrichment data from four ancient individuals.



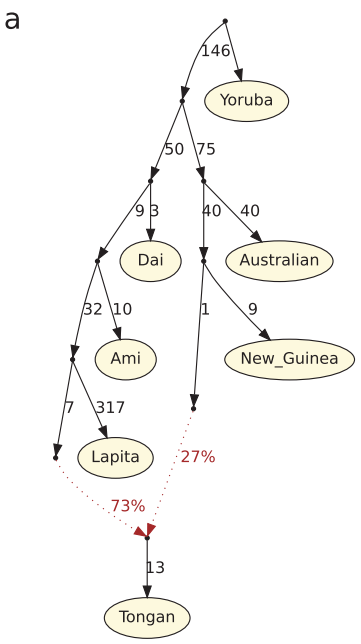
Extended Data Figure 2 | *f*-statistics document the Oceanian ancestry cline. a, Shared genetic drift with the ancient Vanuatu individuals is negatively related to shared drift with Australians. Except for the ancient Tongan individual, populations from Taiwan, the Philippines and Polynesia share the most genetic drift with the ancient Vanuatu individuals, who are not shown in the plot because they are used as reference in the computation. The trend line was fitted without the East Asian populations in the off-cline cluster. The absence of off-cline Oceanian individuals suggests the possibility that present-day Oceanians may largely be derived from a mixture of two source populations. **b**, The ancient Vanuatu individuals and the ancient Tongan individual maximize statistics of the form f_4(Yoruba, *Test*; Australian, Oceanian), suggesting that they are the most closely related to the East Asian ancestry in Oceanians of any sampled population. The trend line was fitted using populations >0.005 on the *x*-axis, together with the two populations with the lowest values on the *x*-axis (Papuan and New_Guinea). **c**, Biplot of First Remote Oceanian ancestry proportions against conditional heterozygosity. Populations with intermediate admixture proportion show the greatest genetic diversity. Thick and thin error bars in all panels are 1 and 1.96 standard errors of the estimate, respectively.

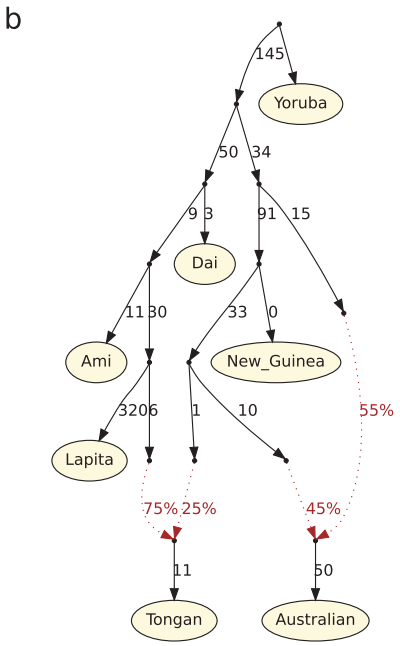


Extended Data Figure 3 | Admixture date estimates. a, Histogram of the point estimate dates in Fig. 2d. **b**, Admixture date estimates for Tongans using different pairs of source populations ('Lapita' in this figure refers to the pool of ancient Vanuatu individuals). Error bars show 1 (thick whiskers) and 1.96 (thin whiskers) standard errors. WGA, whole-genome amplified DNA.

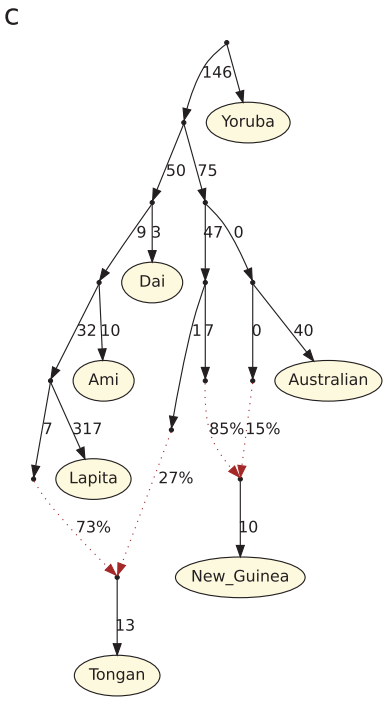


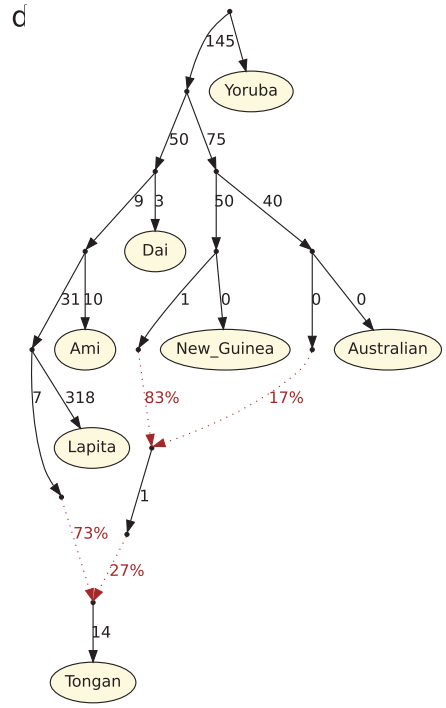
Extended Data Figure 4 | Admixture graph inferred using *Treemix*. **a**, A simple tree-like model without admixture fits the data poorly, as can be seen from the matrix of residuals between empirical and modelled allele frequency covariance on the right. **b**, The optimal placement of a single 25% admixture event is from the lineage related to New Guinean Highlanders into the lineage leading to Tongans. Tongans derive the other portion of their ancestry from the lineage leading to the two ancient groups of individuals. This graph has no significant deviations between empirical and modelled allele frequency covariances.





Outliers	fitted	empirical	z
f4(New_Guinea, Australian; Dai, Tongan)	-0.010860	-0.006083	4.768
f4(New_Guinea, Australian; Tongan, Ami)	0.010860	0.006205	-4.431
f4(New_Guinea, Australian; Tongan, Lapita)	0.010860	0.005384	-3.113
f4(Dai, Tongan; Tongan, Australian)	-0.017124	-0.012147	3.653
f4(Tongan, Ami; Tongan, Australian)	0.010509	0.005834	-3.395

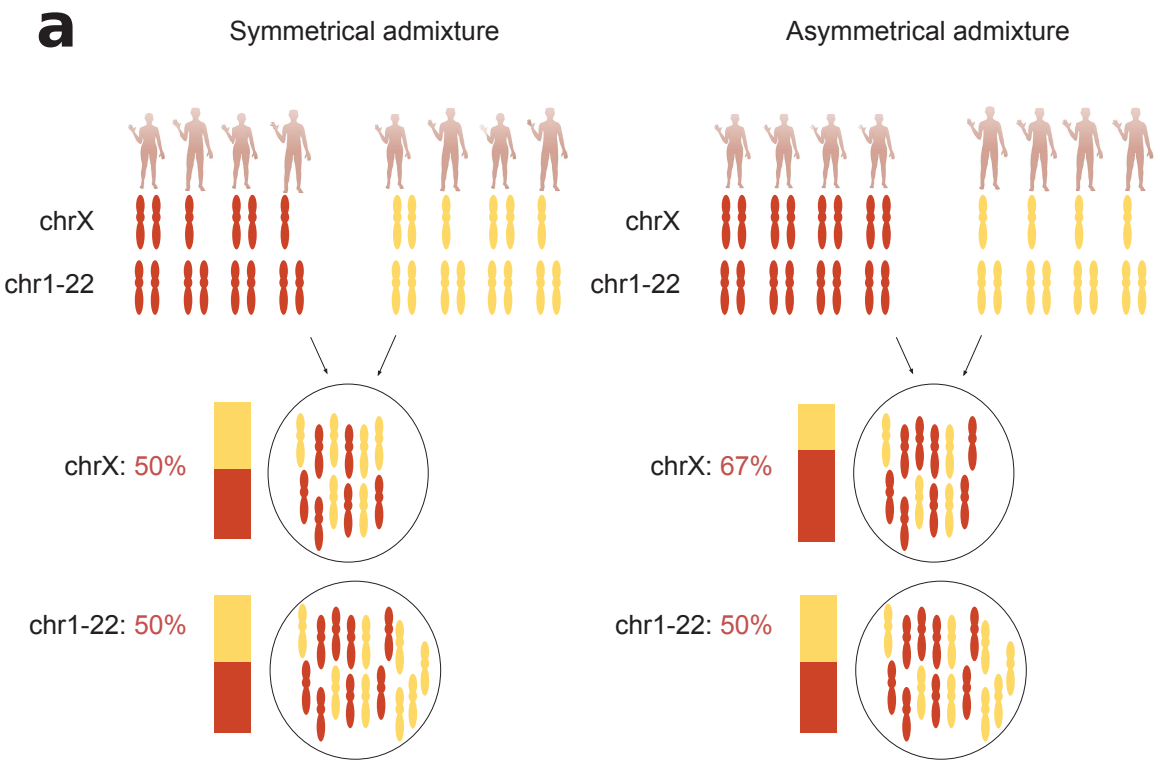


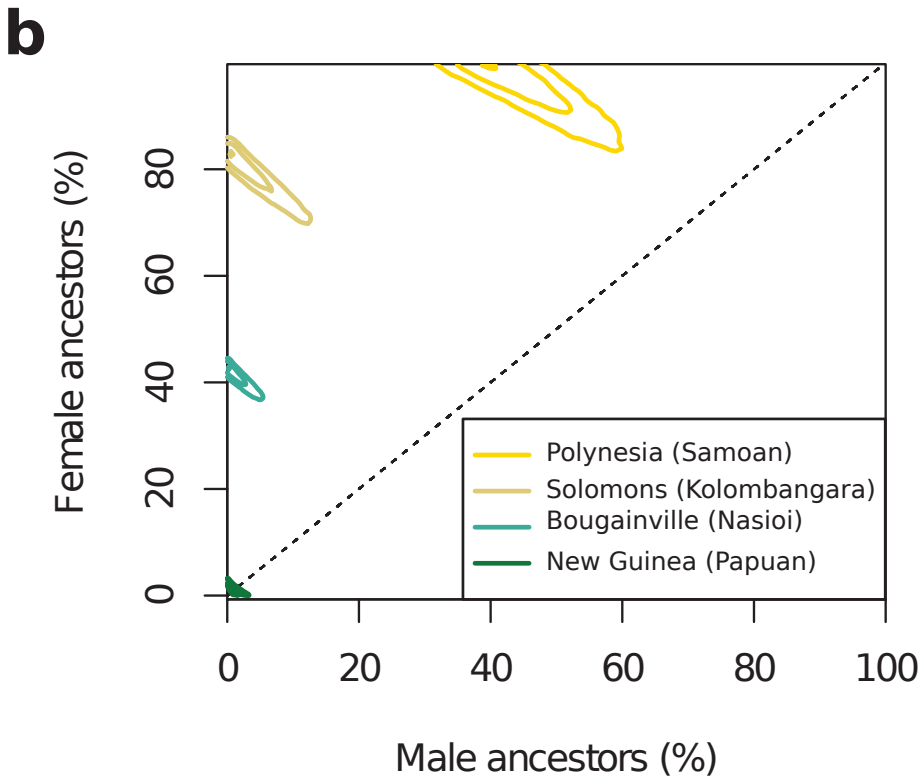


Outliers	fitted	empirical	z
f4(New_Guinea, Australian; Dai, Tongan)	-0.010850	-0.006083	4.758
f4(New_Guinea, Australian; Tongan, Ami)	0.010850	0.006205	-4.422
f4(New_Guinea, Australian; Tongan, Lapita)	0.010850	0.005384	-3.107
f4(Dai, Tongan; Tongan, Australian)	-0.017123	-0.012147	3.652
f4(Tongan, Ami; Tongan, Australian)	0.010509	0.005834	-3.395

Outlier	fitted	empirical	z
f4(New_Guinea, Australian; Dai, Tongan)	-0.009261	-0.006083	3.172

Extended Data Figure 5 | Admixture graphs modelling the population history of Australians. Outlier f_4-statistics are shown ($|Z| > 3$). **a**, A model with a single admixture edge positing that Australians are an outgroup to the Papuan ancestry in Tongans does not fit the data (5 outlier statistics). **b**, An alternative model with 2 admixture edges in which the Papuan ancestry in Tongans also contributed to Australians fits the data (no outliers). **c**, A model with 2 admixture edges in which New Guinean Highlanders are admixed from an Australian source after the divergence of the Papuan source in Tongans does not fit the data (5 outliers). **d**, A model with 2 admixture edges in which the Papuan ancestry in Tongans is intermediate between the New Guinean Highlander lineage and the Australian lineage. Branch lengths are in units of $F_{ST} \times 1{,}000$. Lapita in this figure refers only to Vanuatu, which is the only group for which we have multiple individuals.





Extended Data Figure 6 | First Remote Oceanian ancestry today comes primarily from females. a, Illustration of the rationale for using the X chromosome to study asymmetrical admixture between males and females. The example on the left illustrates admixture with equal proportion of males and females in both the red and the yellow ancestral population. The example on the right illustrates an extreme case of asymmetrical admixture where the red ancestral population only contributes females and the yellow ancestral population only contributes males to the admixed generation, demonstrating the disproportional contribution of X chromosomes by females to the admixed population. **b**, Female and male ancestral contributions based on an admixture model fitted to estimated ancestry proportions on the autosomes and X chromosome. We show the 95%, 70%, and 5% highest posterior intervals for four selected populations from Polynesia (Samoans), the Solomon Islands (Kolombangara), Bougainville (Nasioi), and mainland New Guinea (Papuans).

Extended Data Table 1 | In-solution DNA enrichment and sequencing of ancient individuals

Sample information						Coverage on chromosomes 1-22			Sex determination		
Location	ID1	ID2	Bone for aDNA	Bone for dating	^{14}C Date: Calibrated 95.4% Conf. Int. (Uncalibrated date, Lab number)	Mean depth	All SNPs	SNPs overlapping array	Y SNPs	X SNPs	Sex
Vanuatu	I1368	B30A	Petrous	Skull	2990-2740 BP (2983±32 BP, Wk-22657)	0.26	139,461	74,631	321	18,231	F
Vanuatu	I1369	B10B	Petrous	Petrous	3000-2750 BP (3045±30 BP, Poz-81126)	0.14	199,500	107,523	341	24,255	F
Vanuatu	I1370	B17	Petrous	Skull	3110-2780 BP (3083±26 BP, Wk-21026)	0.21	167,311	90,402	231	19,303	F
Tonga	CP30	SK10	Petrous	Fibula	2680-2340 BP (2594±20 BP, Wk-41883)	0.16	231,994	125,908	75	25,943	F

All dates are calibrated using OxCal v4.2.4[58] with a mixture of the Marine13 and Intcal13 curves[59] as determined by linear interpolation between dietary terrestrial and marine δ^{13}C isotopic endpoints (−21‰/−12‰) with an uncertainty of ±10% on the per cent marine carbon result following previous recommendations[60]. Two of the dates have been previously reported (for I1368/B30A and I1370/B17)[61], and in this study we add two new dates: for I1369/B10B from Vanuatu (on the same petrous bone used for ancient DNA analysis) and on CP30/SK10 from Tonga (on a fibula). Measured ^{13}C and ^{15}N values for I1369/B10B are −14.5‰ and 13.7‰ respectively, and for SK10 −16.44‰ and 10.48‰. As justified in ref. 61, we also applied a location-specific reservoir correction (ΔR) of 40 ± 44 ^{14}C years to the marine curve to adjust for regional oceanic variation in ^{14}C around Vanuatu, and 11 ± 83 ^{14}C years for Tongatapu[62].

Extended Data Table 2 | 356 individuals newly genotyped on the Human Origins Array

Population	N	Country of origin	Land mass	Language	Lat.	Long.	Co-authors for samples	Protocol Numbers for informed consent	Data distribution
Ata	8	Papua New Guinea	New Britain	Papuan	-5.7	150.9	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Baining_Malasait	5	Papua New Guinea	New Britain	Papuan	-4.47	151.9	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Baining_Marabu	10	Papua New Guinea	New Britain	Papuan	-4.63	152.3	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Bajo	10	Indonesia	Sulawesi	Austronesian	-3.97	122.59	M.P.C, P.K., F.-X.R.	4.13.2013 approval AMIS-UPS Ethics Committee	Signed Letter
Buka	8	Papua New Guinea	Bougainville	Austronesian	-5.42	154.67	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Burmese	10	Myanmar	Asia	Sino-Tibetan	16.41	95.89	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Dusun	10	Brunei	Borneo	Austronesian	4.71	114.67	T.K., S.A.	MHREC/EDU/2012/3(1) and HBREC.2011.01	Fully public
Ilocano	2	Philippines	Luzon	Austronesian	14.6	120.98	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Kankanaey	10	Philippines	Luzon	Austronesian	17.07	121.03	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Kol_New_Britain	2	Papua New Guinea	New Britain	Papuan	-5.38	151.63	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Kove	18	Papua New Guinea	New Britain	Austronesian	-5.47	148.95	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Kuot_Kabil	9	Papua New Guinea	New Ireland	Papuan	-3.07	151.7	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Kuot_Lamalaua	8	Papua New Guinea	New Ireland	Papuan	-3	151.5	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Lavongai	15	Papua New Guinea	New Ireland	Austronesian	-2.57	150.43	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Lebbo	8	Indonesia	Borneo	Austronesian	1.66	117.16	M.P.C, P.K., F.-X.R.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Madak	10	Papua New Guinea	New Ireland	Austronesian	-3.1	151.7	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Malay	9	Singapore	Asia	Austronesian	1.35	103.82	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Mamusi	20	Papua New Guinea	New Britain	Austronesian	-6	151	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Mamusi_Paleabu	6	Papua New Guinea	New Britain	Austronesian	-5.95	150.9	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Mangseng	6	Papua New Guinea	New Britain	Austronesian	-5.93	150.7	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Manus	2	Papua New Guinea	Manus	Austronesian	-2.08	147	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Melamela	10	Papua New Guinea	New Britain	Austronesian	-5	151.25	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Mengen	10	Papua New Guinea	New Britain	Austronesian	-5.1	151.4	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Murut	10	Brunei	Borneo	Austronesian	4.62	115.14	T.K., S.A.	MHREC/EDU/2012/3(1) and HBREC.2011.01	Fully public
Mussau	10	Papua New Guinea	St. Matthias	Austronesian	-1.58	149.73	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Nakanai_Bileki	10	Papua New Guinea	New Britain	Austronesian	-5.75	150.8	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Nakanai_Loso	7	Papua New Guinea	New Britain	Austronesian	-5.48	150.8	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Nailik	9	Papua New Guinea	New Ireland	Austronesian	-2.98	151.52	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Notsi	9	Papua New Guinea	New Ireland	Austronesian	-3.05	151.65	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Saposa	10	Papua New Guinea	Bougainville	Austronesian	-5.58	154.67	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Southwest_Bougainville	2	Papua New Guinea	Bougainville	Papuan	-6.6	155.5	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Sulka	20	Papua New Guinea	New Britain	Papuan	-4.5	152.3	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Tagalog	5	Philippines	Luzon	Austronesian	14.6	120.98	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Teop	10	Papua New Guinea	Bougainville	Austronesian	-5.85	155.18	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Tigak	10	Papua New Guinea	New Ireland	Austronesian	-2.57	150.83	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Tolai	24	Papua New Guinea	New Britain	Papuan	-4.31	152.14	F.R.F., J.S.F., G.K., D.A.M.	MRAC.1998.2000.2010 and 99-226.4320	Signed Letter
Vietnamese	10	Vietnam	Asia	Austroasiatic	10.82	106.64	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public
Visayan	4	Philippines	Mindanao	Austronesian	9.76	125.51	T.K., J.T.S.W.	2010/820/A and HBREC.2011.01	Fully public

Extended Data Table 3 | *f*-statistics for populations on the Oceanian cline

	F_{ST}(Lapita_Vanuatu, Test)		f_3(Lapita_Vanuatu, Australian; Test)	
Test population	**Estimate**	**SE**	**Estimate**	**Z-score**
Baining_Malasait	0.263	0.005	0.143	38.4
Baining_Marabu	0.249	0.004	0.115	40.1
Papuan	0.225	0.004	0.066	33.3
Kol_New_Britain	0.216	0.006	0.086	16.7
Mamusi	0.204	0.004	0.059	26.5
Ata	0.197	0.004	0.050	21.6
Nakanai_Loso	0.194	0.004	0.055	21.9
Mamusi_Paleabu	0.187	0.004	0.040	17.8
Santa_Cruz	0.185	0.004	0.021	10.7
Nasioi	0.178	0.004	0.028	14.7
Bougainville_South	0.176	0.006	0.034	8.1
Sulka	0.174	0.004	0.022	13.1
Mengen	0.168	0.004	0.017	9.1
Tolai	0.163	0.004	0.009	5.6
Kuot_Kabil	0.162	0.004	0.014	7.7
Lavongai	0.160	0.004	0.009	5.2
Kuot_Lamalaua	0.160	0.004	0.008	4.3
Nakanai_Bileki	0.153	0.004	0.013	7.0
Melamela	0.152	0.004	0.010	5.1
Madak	0.151	0.004	0.003	1.8
Papuan_Gulf	0.151	0.005	0.002	0.7
Kove	0.150	0.004	0.013	7.4
Mangseng	0.145	0.004	0.003	1.7
Nailik	0.145	0.004	0.000	-0.1
Teop	0.144	0.004	0.005	2.8
Notsi	0.144	0.004	-0.002	-1.4
Manus	0.141	0.006	-0.005	-1.4
Tigak	0.141	0.004	0.002	1.2
Mussau	0.132	0.004	-0.005	-3.3
Choiseul	0.123	0.004	-0.007	-3.7
Saposa	0.122	0.004	-0.011	-6.5
Buka	0.118	0.004	-0.016	-9.9
Vella_Lavella	0.112	0.004	-0.016	-8.6
Ranongga	0.110	0.004	-0.015	-8.2
Savo	0.109	0.004	-0.020	-12.1
Russell	0.108	0.005	-0.017	-6.8
Kolombangara	0.108	0.004	-0.015	-8.0
RenBel	0.106	0.004	0.035	14.0
Gela	0.103	0.004	-0.024	-14.3
Makira	0.101	0.004	-0.024	-16.1
Malaita	0.096	0.004	-0.025	-15.5
Papuan_Central	0.094	0.004	-0.031	-18.2
Bajo	0.082	0.004	0.022	12.8
Isabel	0.079	0.004	-0.024	-15.9
Tikopia	0.077	0.004	-0.003	-1.3
Ontong_Java	0.069	0.004	-0.018	-10.4
Tongan	0.053	0.004	-0.018	-9.9

Standard error (SE) is shown for F_{ST} between each *Test* population and the pool of ancient Vanuatu individuals. The *Z* score is given for the statistic f_3(Lapita_Vanuatu, Australian; *Test*), where $Z < 3$ provides significant evidence that the *Test* is admixed between sources related to the ancient Vanuatu and Australians.

Extended Data Table 4 | Ancestry estimates for populations on the Oceanian cline

	First Remote Oceanian ancestry estimate							Anc. contrib. (method of moments)				Two source model
Test	Auto.	SE	chrX	SE	Diff.	SE	Z	Male	SE	Female	SE	*P*-value
Tongan	69.8%	1.0%	104.1%	8.9%	-34.3%	9.0%	-3.8	-33.1%	27.0%	172.7%	22.0%	0.22
Rennel & Bellona	68.9%	1.2%	92.1%	7.3%	-23.2%	7.4%	-3.1	-0.7%	22.4%	138.5%	32.8%	0.89
Tikopia	65.8%	1.0%	93.9%	10.9%	-28.1%	10.9%	-2.6	-18.5%	32.9%	150.1%	39.9%	0.58
Ontong_Java	61.9%	0.9%	78.4%	13.3%	-16.5%	13.3%	-1.2	12.4%	40.1%	111.4%	12.7%	0.83
Santa Isabel	52.9%	0.9%	60.5%	4.2%	-7.6%	4.3%	-1.8	30.1%	13.1%	75.7%	14.2%	0.87
Papuan_Central	42.5%	0.9%	65.6%	4.7%	-23.1%	4.8%	-4.8	-26.8%	14.6%	111.8%	47.7%	0.14
Malaita	39.6%	0.9%	66.1%	15.9%	-26.5%	15.9%	-1.7	-39.9%	47.8%	119.1%	20.8%	0.25
Kolombangara	39.1%	1.0%	54.7%	6.9%	-15.6%	7.0%	-2.2	-7.7%	21.1%	85.9%	16.3%	0.87
Nggela	37.8%	0.9%	54.6%	5.4%	-16.8%	5.5%	-3.1	-12.6%	16.6%	88.2%	82.5%	0.59
Ranongga	37.6%	1.0%	48.8%	27.5%	-11.2%	27.5%	-0.4	4.0%	82.6%	71.2%	29.8%	0.15
Russell	36.2%	1.2%	45.8%	9.9%	-9.6%	10.0%	-1.0	7.4%	30.1%	65.0%	26.2%	0.19
Vella_Lavella	35.2%	1.0%	53.2%	8.7%	-18.0%	8.8%	-2.1	-18.8%	26.4%	89.2%	19.9%	0.10
Savo	34.6%	0.9%	56.4%	6.6%	-21.8%	6.7%	-3.3	-30.8%	20.1%	100.0%	38.4%	0.18
Makira	34.6%	0.9%	31.8%	12.8%	2.8%	12.8%	0.2	43.0%	38.6%	26.2%	13.4%	0.08
Choiseul	32.4%	1.0%	35.5%	4.4%	-3.1%	4.5%	-0.7	23.1%	13.8%	41.7%	9.8%	0.20
Buka	31.3%	0.9%	46.0%	3.2%	-14.7%	3.3%	-4.4	-12.8%	10.3%	75.4%	16.9%	0.52
Saposa	31.1%	0.9%	41.8%	5.6%	-10.7%	5.7%	-1.9	-1.0%	17.2%	63.2%	14.2%	0.25
Mussau	29.2%	0.9%	39.4%	4.7%	-10.2%	4.8%	-2.1	-1.4%	14.6%	59.8%	22.0%	0.54
Teop	26.5%	0.9%	67.8%	7.3%	-41.3%	7.4%	-5.6	-97.4%	22.2%	150.4%	11.5%	0.86
Kove	26.4%	0.9%	29.8%	3.8%	-3.4%	3.9%	-0.9	16.2%	12.0%	36.6%	20.2%	0.01
Tigak	26.2%	0.9%	43.9%	6.7%	-17.7%	6.8%	-2.6	-26.9%	20.4%	79.3%	21.4%	0.87
Melamela	25.3%	0.9%	50.1%	7.1%	-24.8%	7.2%	-3.5	-49.1%	21.6%	99.7%	18.8%	0.94
Manus	24.9%	1.2%	14.2%	6.2%	10.7%	6.3%	1.7	57.0%	19.2%	-7.2%	42.9%	0.87
Nakanai_Bileki	24.5%	0.9%	38.4%	14.3%	-13.9%	14.3%	-1.0	-17.2%	43.1%	66.2%	11.2%	0.28
Mangseng	23.6%	0.9%	42.2%	3.7%	-18.6%	3.8%	-4.9	-32.2%	11.7%	79.4%	19.9%	0.34
Papuan_Gulf	22.6%	1.1%	49.3%	6.6%	-26.7%	6.7%	-4.0	-57.5%	20.3%	102.7%	10.9%	0.07
Notsi	20.7%	0.8%	33.0%	3.6%	-12.3%	3.7%	-3.3	-16.2%	11.3%	57.6%	64.8%	0.13
Nailik	20.4%	0.8%	44.8%	21.6%	-24.4%	21.6%	-1.1	-52.8%	64.9%	93.6%	22.9%	0.08
Madak	18.7%	0.8%	43.1%	7.6%	-24.4%	7.6%	-3.2	-54.5%	23.0%	91.9%	18.7%	0.11
Kuot_Kabil	18.6%	0.9%	55.4%	6.2%	-36.8%	6.3%	-5.9	-91.8%	18.9%	129.0%	26.5%	0.04
Bougainville_Sout	18.0%	1.4%	24.1%	8.8%	-6.1%	8.9%	-0.7	-0.3%	27.0%	36.3%	11.8%	0.26
Lavongai	17.6%	0.8%	36.4%	3.9%	-18.8%	4.0%	-4.7	-38.8%	12.1%	74.0%	16.0%	0.07
Nasioi	17.4%	1.0%	29.9%	5.3%	-12.5%	5.4%	-2.3	-20.1%	16.4%	54.9%	11.5%	0.40
Kuot_Lamalaua	16.2%	0.8%	13.8%	3.8%	2.4%	3.9%	0.6	23.4%	11.8%	9.0%	19.6%	0.07
Nakanai_Loso	15.4%	1.1%	24.3%	6.5%	-8.9%	6.6%	-1.4	-11.3%	20.0%	42.1%	11.2%	0.37
Mengen	15.2%	0.8%	35.3%	3.7%	-20.1%	3.8%	-5.3	-45.1%	11.6%	75.5%	13.0%	0.41
Tolai	14.3%	0.8%	22.5%	4.3%	-8.2%	4.4%	-1.9	-10.3%	13.3%	38.9%	9.4%	0.01
Sulka	14.1%	0.8%	47.0%	3.1%	-32.9%	3.2%	-	-84.6%	9.8%	112.8%	16.3%	0.48
Mamusi_Paleabu	13.7%	1.0%	30.9%	5.4%	-17.2%	5.5%	-3.1	-37.9%	16.7%	65.3%	13.9%	0.21
Mamusi	13.7%	0.9%	22.7%	4.6%	-9.0%	4.7%	-1.9	-13.3%	14.3%	40.7%	27.1%	0.35
Ata	13.2%	1.0%	26.4%	9.0%	-13.2%	9.1%	-1.5	-26.4%	27.3%	52.8%	15.4%	0.24
Santa_Cruz	9.6%	0.9%	27.1%	5.1%	-17.5%	5.2%	-3.4	-42.9%	15.7%	62.1%	23.3%	0.17
Kol_New_Britain	8.5%	1.4%	9.1%	7.7%	-0.6%	7.8%	-0.1	6.7%	23.8%	10.3%	30.1%	0.66
Baining_Marabu	2.6%	1.0%	25.3%	10.0%	-22.7%	10.0%	-2.3	-65.5%	30.3%	70.7%	32.5%	0.24
Baining_Malasait	1.2%	1.2%	34.9%	10.8%	-33.7%	10.9%	-3.1	-99.9%	32.8%	102.3%	8.2%	0.12
Papuan	0.0%	0.5%	2.4%	2.7%	-2.4%	2.7%	-0.9	-7.2%	8.3%	7.2%	0.0%	0.58

Auto., estimate on the autosomes (chromosomes 1–22). Diff., difference between the autosome and X chromosome estimates.

doi:10.1038/nature19844

Supplementary Note 1: The Teouma site

The Teouma Lapita site is located on the south coast of Efate Island, Vanuatu, and is currently 800 meters from Teouma Bay. Tectonic uplift, volcanic ashfalls prior to and during the period of human utilisation of the site, and alluvial deposition from an adjacent stream have altered its immediately coastal location during Lapita times. Excavations by a joint ANU-Vanuatu National Museum team took place between 2004 and 2010[1-3], during which 68 burial features, including remains of just over 100 individuals were found, concentrated in a band running northeast to southwest, parallel to the former beach and some 10-15m wide. Teouma has numerous indictors of being an initial colonisation site for Efate, including extinct faunal remains, early ceramic forms and decoration, and New Britain obsidian. This combined with the extensive and previously very rare Lapita skeletal remains underline its significance to those investigating colonization in this region.

There are two key zones of the site; the Cemetery Zone (CZ) and the Midden Zone (MZ) located to the east of the Lapita cemetery, that both comprise three layers. The Lapita midden contained a number of domesticated species (i.e., *Sus scrofa*, *Gallus gallus* and the commensal *Rattus exulans*), as well as a large number of bones from indigenous birds and animals, many now extinct, including land tortoise (*Meiolania damelipi*), crocodile (*Mekosuchus kalpokasi*) and a range of bird species[2,4,5]. Layer 3 of the Midden Zone (MZ3) deposit is considered to be contemporary with Layer 3 of the cemetery (CZ3) and both are emplaced onto a pre-human tephra deposit[6]. A later midden deposit (Layer 2 Cemetery Zone [CZ2]), up to 50 cm thick, covered the cemetery and extended downslope over the former beach and within alluvial deposits from the adjacent stream, representing a shift to purely habitation use at the main site. This is associated with Post-Lapita Arapus and Early Erueti ceramics currently dated to c. 2800-2500 BP[2]. Continued tectonic uplift and alluvial deposition led to abandonment of the settlement as immediate access to the sea became problematic because of shoreline progradation. Layer 1 from both zones has not been dated but represents natural post-occupation accumulation, rich in tephra from the 2300 BP eruption of Nguna Island to the north. In Area 7C some 70m south of the main Teouma site further burials were located, dating to within the Erueti phase at about 2400 BP[7].

An extensive dating program has been undertaken at the site on a range of materials including charcoal and marine shell, as well as human, terrestrial native fauna and domesticate bones[8,9]. That analysis places the start date of the Lapita deposits at 2920-2870 cal BP with a possible use as early as 3000 cal BP, and an end date of 2870-2750 cal BP. The results are in keeping with evaluation of the burials (36 direct dating on human cranial and infracranial elements[8]), which places earliest use of the cemetery at c. 2970 cal BP with regular use underway by c. 2940-2880 cal BP and the last internment occurring c. 2770-2710 cal BP. Continued use of the area is evident by the later burials nearby dating to the end of the Erueti phase and by components of the

CZ2 midden deposits of post-Lapita age.

The exceptionally well-preserved and extensive Lapita skeletal remains have provided an opportunity to start to establish a population profile, gaining insights into the health, morphology and diet of a group of Lapita settlers who were among the first to have ventured into Remote Oceania. The 68 burials at Teouma display a wide range of mortuary practices and indicate that ritual was a multi-faceted and on-going process, rather than a one-off event. The burials were generally placed in shallow graves dug into the underlying tephra deposits amongst gaps in the uplifted reef and coral boulders on the upper part of the beach. Evidence indicates manipulation of the corpse prior to burial or at least during the early stages of decomposition and of the skeletal remains after body decomposition. All the adult burials had their skulls, and often many other bones, removed during this process and only a few of these bones were secondarily re-deposited at the site[1,3,10,11]. Nine skulls were found out of a total of 52 infracranial skeletons in articulation. Although disconnected from infracranial skeletons, the cranial elements appears to have belonged to the same group as shown by similarities in isotopic values measured in bone collagen and dental enamel. Seven of them, belonging to female and male adults, were preserved well enough for morphological examination. Five of them were useable for metrical analysis that demonstrated that each displays a lack of Australian or Papuan affinities and an alignment with ancient and more recent Asiatic populations[12]. Ancient DNA was successfully obtained from three skulls from striking mortuary contexts: a jar burial containing a single skull (B17), an alignment of three skulls lying on the chest of a skeleton without a skull (B10B), and a triangular bone arrangement, with skulls at each vertex, lying on the legs of a skeleton without a skull (B30A).

References

1 - Bedford, S., Spriggs, M. & Regenvanu, R. 2006. The Teouma Lapita Site and the Early Human Settlement of the Pacific Islands. *Antiquity* 80:812-828.

2 - Bedford, S., Spriggs, M., Buckley, H., Valentin, F. & Regenvanu, R. 2009. The Teouma Lapita Site, South Efate, Vanuatu: A Summary of Three Field Seasons (2004-2006). In P. Sheppard, T. Thomas and G. Summerhayes (eds) Lapita: *Ancestors and Descendants*. *New Zealand Archaeological Association Monograph* 28, pp. 215-234.

3 - Bedford, S., Spriggs, M., Bukley, H., Valentin, F., Regenvanu, R. & Abong, M. 2010. A cemetery of first settlement: Teouma, South Efate, Vanuatu/ Un cimetiere de premier peuplement: le site de Teouma, sud d'Efate, Vanuatu. In C. Sand and S. Bedford (eds) *Lapita: Oceanic Ancestors/Lapita: Ancetres Oceaniens*, pp. 140-161. Paris: Musee de Quai Branly/Somogy.

4 - White, A., Worthy, T., Hawkins, S., Bedford, S. & Spriggs, M. 2010. Megafaunal meiolaniid horned turtles survived until early human settlement in Vanuatu, Southwest Pacific. *Proceedings of the National Academy of Sciences* 107:15512-15516.

5 - Worthy, T., Hawkins, S., Bedford, S. & Spriggs, M. 2015. The avifauna from the Teouma Lapita site, Efate Island, Vanuatu, including a new genus and species of megapode. *Pacific Science* 69(2):205-254.

6 - Spriggs, M. & Bedford, S. 2013. Is there an incised Lapita phase after dentate-stamped pottery ends? Data from Teouma, Efate Island, Vanuatu. In G. Summerhayes and H. Buckley (eds) *Pacific Archaeology: Documenting the Past 50,000 Years*, pp. 148-156. University of Otago Studies in Archaeology 25. Dunedin: Department of Anthropology and Archaeology, University of Otago.

7 - Valentin F., Herrscher E., Bedford S., Spriggs M., Buckley H., 2014. Evidence for social and cultural change in Central Vanuatu between 3000 and 2000 BP: comparing funerary and dietary patterns of the first and later generations at Teouma, Efate. *Journal of Island and Coastal Archaeology* 9:3, 381-399.

8 - Petchey F., Spriggs M., Bedford S., Valentin F., Buckley H., 2014. Radiocarbon dating of burials from the Teouma Lapita cemetery, Efate, Vanuatu. Journal of Archaeological Science, 50: 227-242.

9 - Petchey F., Spriggs M., Bedford S., Valentin F., 2015. The chronology of occupation at Teouma, Vanuatu: Use of a modified chronometric hygiene protocol and Bayesian modeling to evaluate midden remains. Journal of Archaeological Science: Reports 4 (2015), p. 95-105.

10 - Valentin, F., Bedford, S., Buckley H., Spriggs, M., 2010. Lapita Burial Practices: Evidence for Complex Body and Bone Treatment at the Teouma Cemetery, Vanuatu, Southwest Pacific. *Journal of Island and Coastal Archaeology*, 5 (2): 212-235.

11 - Valentin F., Choi J.-I., Lin H., Bedford S., Spriggs M. 2015. Three thousand year old jar burials at the Teouma cemetery (Vanuatu): a Southeast Asian-Lapita connection?, in C. Sand, S. Chiu and N. Hogg (eds) *The Lapita Cultural Complex in Time and Space: Expansion Routes*, *Chronologies and Typologies*, Nouméa and Taipei, IANCP and Center for Archaeological Studies, Research Center for Humanities and Social Sciences, Academia Sinica. (Archaeologia Pasifika 4), p. 81-101.

12 - Valentin, F., Détroit, F., Spriggs, M. J. T. & Bedford, S. Early Lapita skeletons from Vanuatu show Polynesian craniofacial shape: Implications for Remote Oceanic settlement and Lapita origins. *Proceedings of the National Academy of Sciences*, 113 (2): 292-297 (2016).

Supplementary Note 2: The Talasiu site

The Talasiu site (TO-Mu-2) is located on the palaeoreef-limestone shoreline of the Fanga 'Uta Lagoon around 10 kilometers from Nuku'alofa and immediately north of Lapaha Village. The site consists of a dense shell midden deposit ~90 cm thick covering some 450 square meters and includes fire features and burials. The palaeoshoreline at Talasiu is fronted by reclaimed land that was deposited in the lagoon during the expansion of the Tu'i Tonga chiefdom 800-200 cal. BP. During the Lapita period in Tonga (2900-2650 cal. BP) (95.4% prob.)[1], the Talasiu site would have been an attractive location for human settlement as it overlooked an embayment holding large numbers of sessile and gregarious shellfish close to a fresh water solution channel, with gardening soils immediately inland. The site was first investigated in 1957 by Jack Golson (ANU) and in the 1980s by Dirk Spennemann (ANU) who surface collected pottery (including sherds decorated with dentate-stamping) lithics and human bone[2,3]. Site monitoring after road grading in 2008 identified a concentration of burned and partially burned human bone which was excavated and found to be a cremation mortuary context containing the incomplete remains of four individuals[4]. In 2011, new burials were found eroding from the road cut and as the area was about to be intensively gardened a rescue archaeology project to recover human remains was directed by Frederique Valentin (CNRS) and Geoffrey Clark (ANU) in 2013-2014 with the support of the Ministry of Internal Affairs (Kingdom of Tonga) and funded by the French Government (MAEDI, Commission des fouilles à l'étranger).

The site stratigraphy consists of a series of six layers distinguished by varying quantities of shellfish, ceramics, charcoal and bone resting on sterile basal clay (Layer 6). The 95% probability range of radiocarbon determinations obtained on coconut endocarp (n=5), unidentified charcoal (n=2), worked shell grave goods (n=3) and human bone sample on articulated burials (n=6) fall between 2870 and 2340 cal. BP. Calibrated charcoal and bone ^{14}C ages between 2600 and 2300 BP are influenced by curve flattening resulting in multiple curve intercepts and a substantial widening in the calibrated age range of determinations. However, a high-resolution chronology based on U-Th dating of coral files and AMS determinations on short-lived material demonstrates the Lapita period on Tongatapu spanned 2900-2650 cal. BP[1] and it is highly likely that the midden and burials at Talasiu date to ~2700-2600 cal. BP[2] and are of late Lapita/immediately post-Lapita age.

Analysis of the midden using archaeozoological and microbotanical techniques provided evidence for a broad spectrum mixed economy in Tonga[2]. Pottery was predominantly plain with dentate-stamped vessels bearing simple open designs typical of late Lapita ceramics throughout the deposit. Shell artefacts included short and long shell units, broad *Conus* spp. rings, and narrow rings made in *Conus* spp. and *Tridacna* spp. that are characteristic Lapita ornaments. Lithics included adzes, flakes, grind stones and oven stones. Thin-section and compositional study of the lithics with pXRF and LA-ICP-MS identified obsidian flakes from Tafahi in

northern Tonga[5], adzes made in volcanic material chemically sourced to Central Tonga, eastern Fiji and Samoa, and manuports and grindstones/oven stones from nearby 'Eua Island and Central Tonga[2,5,6].

The 19 late-Lapita/immediately post-Lapita burial contexts at Talasiu were placed at the base of the midden or interstratified within the midden. Intact lenses of ash and marine shell overlying several burials conclusively demonstrate that interments were made during midden formation. The burials were single or multiple interments containing the skeletal remains of one to six individuals (male and female adults and children of various ages). Field evidence indicates a wide range of mortuary practices involving primary burial in various positions, secondary deposits of burnt[4] or unburnt bone, including placement of isolated skulls, and post-decomposition removal of skulls and long bones. This diversified mortuary pattern recalls that of late Lapita and immediately post-Lapita burials from sites in Vanuatu[7], while the body treatment of some individuals is the same as at the Lapita site of Teouma[8-9]. These 19 burial contexts have yielded the skeletal remains of 62 individuals, providing an opportunity to obtain information on health, diet and morphology of a group of late Lapita/immediately post-Lapita people from Tongatapu, using macroscopic and microscopic observations and biogeochemical and palaeomolecular data.

Preliminary metrical analysis performed on one skull (BG3) shows that the individual lacks Australian or Papuan affinities and instead has affinities to mainland Asian populations, as do the Teouma Lapita skulls[10]. Ancient DNA was successfully obtained from the right petrous bone of burial SK10, a single primary interment of an adult female with a marine-adjusted calibrated age of 2680-2340 cal. BP (95.4% confidence interval; see **Extended Data Table 1**). The SK10 ^{14}C calibration is influenced by the radiocarbon plateau around 2600-2300 BP, which results in a wide spread in the calibrated age range. It is important to note that this burial was sealed by a layer of intact midden which is inferred from essentially identical radiocarbon results on carbonized coconut shell from the upper/middle and basal layers to have been deposited around 2700-2600 cal. BP[2].

1. Burley, D.V., Edinborough, K., Weisler, M., Zhao, J.-x., 2015. Bayesian modeling and chronological precision for Polynesian settlement of Tonga. PLoS One 10 (3): e0120795. http://dx.doi.org/10.1371/journal. pone.0120795.
2. Clark, G., Grono, E., Ussher, E. and Reepmeyer, C. 2015. Early settlement and subsistence on Tongatapu, Kingdom of Tonga: Insights from a 2700-2650 cal BP midden deposit. *Journal of Archaeological Science: Reports* DOI: 10.1016/j.jasrep.2015.08.005.
3. Spennemann, D.H.R. 1986. *Archaeological fieldwork in Tonga 1985-1986. Preliminary report on the archaeological activities during the 1985/86 field season of the Tongan Dark Ages Research Programme*. Department of Prehistory, The Australian National University, Canberra.

4. Valentin, F. and Clark, G. 2013. Early Polynesian mortuary behaviour at the Talasiu site, Kingdom of Tonga. *Journal of Pacific Archaeology* 4(1): 1-14.
5. Reepmeyer, C., Clark, G. and Sheppard, P. 2012. Obsidian source use in Tongan prehistory: New results and implications. *Journal of Island and Coastal Archaeology* 7(2): 255-271.
6. Clark, G., Reepmeyer, C., Melekiola, N., Woodhead, J., Dickinson, W.R. and Martinsson-Wallin, H. 2014. Stone tools from the ancient Tongan state reveal prehistoric interaction centres in the Central Pacific. *Proceedings of the National Academy of Sciences U.S.A*. 111(29): 10491-10496.
7. Bedford, S., Buckley, H., Valentin, F., Tayles, N., Longga, F. 2011. Lapita burials, a new Lapita cemetery and post-Lapita burials from Malakula, northern Vanuatu, Southwest Pacific. *Journal of Pacific Archaeology*, 2(2): 26-48.
8. Valentin, F., Bedford, S., Buckley H., Spriggs, M., 2010. Lapita Burial Practices: Evidence for Complex Body and Bone Treatment at the Teouma Cemetery, Vanuatu, Southwest Pacific. *Journal of Island and Coastal Archaeology*, 5 (2): 212-235.
9. Valentin F., Choi J.-I., Lin H., Bedford S., Spriggs M. 2015. Three thousand year old jar burials at the Teouma cemetery (Vanuatu): a Southeast Asian-Lapita connection?, in C. Sand, S. Chiu and N. Hogg (eds) *The Lapita Cultural Complex in Time and Space: Expansion Routes, Chronologies and Typologies*, Nouméa and Taipei, IANCP and Center for Archaeological Studies, Research Center for Humanities and Social Sciences, Academia Sinica. (Archaeologia Pasifika 4), p. 81-101.
10. Valentin, F., Détroit, F., Spriggs, M. J. T. & Bedford, S. 2016. Early Lapita skeletons from Vanuatu show Polynesian craniofacial shape: Implications for Remote Oceanic settlement and Lapita origins. *Proceedings of the National Academy of Sciences*, 113(2): 292-297.

Supplementary Note 3: Population history graph models

We used admixture graph models to model the population history of Oceania and Polynesia. For our main analyses, we used *qpGraph*[1,2] to test systematically a variety of historical models, and to assess their fit to the data, but see also our complementary analysis with *TreeMix* documented in the **Methods** section and **Extended Data Figure 4**. For *qpGraph*, we focused on assessing whether graphs fulfill the criterion that all or nearly all f_4-statistics predicted by the model are consistent with the empirical statistics. We refer to individual f_4-statistics with $|Z|>3$ between empirical and predicted statistics as 'outliers', providing evidence for unmodeled shared genetic drift between populations. We also assess the relative fit of models by considering a sum of squared differences-statistic (RMSE) summarizing the overall fit between predicted and empirical *f*-statistics, but note that the distribution of this statistic under arbitrary admixture graph models is not known.

We focused on a set of populations representing broad genetic and geographic diversity:

1. Outgroup: African Yoruba
2. Unadmixed Papuans: New Guinean Highlanders
3. Mainland East Asia: Dai Chinese
4. Indigenous Taiwanese: Atayal
5. Philippines population without Papuan ancestry: Kankanaey
6. Philippines population with Papuan ancestry: Mamanwa
7. Northern Solomons Papuan-speakers who are a mixture of Papuan and First Remote Oceanian ancestry: Nasioi (HGDP Bougainville)
8. Solomon Islander Austronesian-speakers who are a mixture of Papuan and First Remote Oceanian ancestry: Kolombangara
9. Polynesians: Tongan
10. Australians

A 'skeleton graph' of a tree-like population history that fits the ancient DNA data

We first added Lapita_Vanuatu to all 8 possible edges of a skeleton phylogeny of (Yoruba, (New_Guinea, (Dai, (Atayal, Kankanaey)))), which we found fitted the data without outliers either using all SNPs or only transversions. Two principal graph topologies fitted with no outliers, one of which modeled the Lapita_Vanuatu as being most closely related to the Kankanaey of the Philippines, and the other which modeled the Lapita_Vanuatu, Atayal, and Kankanaey as a trifurcation. Since the trifurcation model also had a slightly greater chi-squared statistic (RMSE=1.71) than the model in which the Lapita_Vanuatu share on the order of F_{ST} = 0.001 drift with the Kankanaey lineage (RMSE=1.58), we continued to use the latter graph, but note that the shared drift between the Kankanaey and Lapita_Vanuatu is small.

Adding populations to the skeleton graph

We proceeded by adding four genetically differentiated populations as mixtures of all 45 possible pairs of edges of the graph: the Polynesian Tongans, the Solomon Islander Kolombangara, the Northern Solomons Nasioi (HGDP Bougainville), and the Philippine Mamanwa[3].

Tongans, Kolombangara, and Nasioi

We found that Tongans, Kolombangara and Nasioi can only be fit as having derived their ancestry from the First Remote Oceanian lineage on the one hand, and the New Guinean Highlander lineage on the other hand, with no outlier statistics and $\chi^2 < 1.9$ for all three populations. All other graphs for Tongans and Kolombangara have 32 or more outlier statistics, and $\chi^2 > 15$. The second best fitting graph for Tongans posits that Tongans received their Austronesian ancestry from the Atayal lineage, but this graph has 50 outlier f_4-statistics; for example f_4(Yoruba, Tongan; Atayal, ancient Oceanians) which is predicted to be ~0 in this graph is empirically 0.021 (Z=9.6). In addition, this graph has a chi-squared statistic of $\chi^2 = 41$ compared to 1.6 for the best fitting graph. For Nasioi the second best fitting graph has 10 outlier statistics, but a slightly smaller χ^2-statistic (5.3). We conclude that there is overwhelming evidence that all three populations received their non-Papuan ancestry from a source most closely related to the Lapita_Vanuatu, which in this paper we call the First Remote Oceanian lineage.

Mamanwa

We found that there were three phylogenies that fit the Mamanwa without outliers. One portion of Mamanwa ancestry was in all cases derived from the lineage leading to New Guinean Highlanders, but the second source was either the First Remote Oceanian lineage, the Kankanaey lineage, or the lineage ancestral to Kankanaey and the Lapita_Vanuatu. In all three cases the inferred drift length separating this second source of the Mamanwa ancestry to the (Lapita_Vanuatu, Kankanaey) ancestral node was $F_{ST} = 0.0001$. We are thus not able to uniquely place the Mamanwa with respect to the First Remote Oceanians and Kankanaey, but use the placement along the Kankanaey lineage due to a smaller chi-squared statistic.

Modeling Tongans, Solomon Islanders and Bougainville populations simultaneously

We next proceeded to testing all possible models that included two populations out of the set (Tongan, Kolombangara, Nasioi). To do this we used the unique fitting graph for each population from our initial additions to the skeleton graph, and added the two others as mixture between all possible edges of that graph. We find that for all these combinations of populations there are at least 23 graph topologies without outliers, and we thus do not have enough resolution in the data to distinguish between different graphs that include these populations simultaneously. We therefore focus on graphs where only one of these populations is included at the same time.

Adding populations to a skeleton graph with Australians

We also consider a graph that models Australians as a clade with New Guineans (Yoruba, ((New_Guinea, Australian), (Dai, (Atayal, (Kankanaey, Lapita_Vanuatu))))). The Lapita_Vanuatu fit in this graph as a clade with Kankanaey without outliers, but when we add Tongan, Kolombangara, Nasioi and Mamanwa as admixed between all possible 66 pairs of graph edges, there are no solutions without outlier statistics. For the Tongan, Kolombangara, and Nasioi, the best fitting graph posits that they are mixed between the Lapita_Vanuatu and New Guinea Highlander lineages, although both the Nasioi and Tongan graph show outliers (2 and 7, respectively) suggestive of unmodeled affinity to Australians. For Tongans:

f_4(New_Guinea, Australia; Dai, Tongan): model =-0.011118, empirical=-0.006075, (Z = 4.9)
f_4(New_Guinea, Australia; Atayal, Tongan): model=-0.011118, empirical=-0.005472, (Z = 4.9)

The Mamanwa are optimally fitted as a lineage basal to Australians and New_Guineans, but a single Z = 3 outlier (f_4[Yoruba, Mamanwa; New_Guinean, Australian]) suggests that the portion of Mamanwa ancestry related to Papuans and Australians is slightly closer to Australians. The difference between this statistic and the statistic in **Table S1** (see below) is that the graph statistic is based on transversion SNPs where we have at least one Lapita_Vanuatu genotype.

Graphs that can account for the excess affinity between Oceanians and Australians

We have shown that when we model the non-First Remote Oceanian ancestry of Tongans as entirely Papuan—with the lineage leading to Australians basal to both lineages—we observe more allele sharing between Tongans and Australians than is predicted by this model. In contrast, differences between Australians and Papuans affinity are not detectable in non-Oceanians, who are approximately symmetrically related to Australians and Papuans (**Table S1**).

Table S1. Australians and New Guinean Highlanders are approximately symmetrically related to non-Oceanians based on the statistic f_4(Yoruba, *X*; Australian, New_Guinea).

X	f_4	***Z***
Primate_Chimp	0.000044	0.1
Dinka	0.000078	0.5
Sardinian	-0.000558	-1.6
Onge	0.000287	0.7
Dai	0.000214	0.6
Atayal	0.000261	0.6
Karitiana	0.000289	0.6
Mixe	0.000076	0.2
Mamanwa	0.000819	2.3
Tongan	0.002998	7.7
Kolombangara	0.002590	6.6

We tested different permutations of a graph where either Papuans, Australians, or Tongans are admixed (and excluded the Kankanaey for simplicity), and found that the only permutation tested that resulted in no outliers posited that the Papuan population that contributed ~25% of the ancestry of Tongans also contributed ~45% of the ancestry of Australians (**Extended Data Figure 3**). An alternative graph that only has a single Z=3.17 outlier posits that the non-First Remote Oceanian portion of Tongan ancestry is itself admixed between Papuan- and Australian-related sources. We cannot statistically distinguish between these two models, but conclude that there is evidence for complexity in the history of Papuans, Australians and Polynesians.

Graph model with Tongans and Mamanwa

Combining the observations from the systematic tests of all possible placements of different populations on skeleton graphs above, we find that:

1. Tongans, Nasioi and Kolombangara are generally fitted as mixed between the same two lineages: First Remote Oceanian represented by the ancient samples, and Papuan.
2. The Philippine Mamanwa can be fitted as being mixed from a lineage related to the Philippine Kankanaey, and a lineage that split off prior to the separation of the ancestors of Papuans and Australians.
3. Australians cannot be successfully fit as an outgroup to the non-First Remote Oceanian ancestry in Tongans, and will be treated separately.

We found that a graph with Mamanwa and Tongans results in an excellent fit with no outliers. We also confirmed that Tongans could be successfully replaced with Nasioi or Kolombangara without resulting in a poor fit.

Extended graph model with archaic human genomes and Onge Andaman Islanders

In the above, we do not model the Neanderthal ancestry that separates non-Africans from Africans, or the Denisovan ancestry that separates Australians and Papuans from other populations. While this ancestry is important when considering the histories of these populations, the reason that we can successfully fit Australians and Papuans without modeling their Denisovan ancestry in our graphs is that the difference in archaic ancestry between the Australians and Papuans and mainland non-Africans can be accounted for by shifting the bifurcation point with Yoruba Africans. If multiple African populations with different divergences from non-Africans had been included the Denisovan component of Australian and Papuan ancestry would be expected to share different amounts of drift with the African populations as they have different degrees of relatedness to non-Africans. This would make it impossible for a model that does not model the archaic admixtures to accommodate the real data.

To fit an extended model incorporating archaic admixture inferred by previous studies, we used chimpanzee as an outgroup, and both the archaic Denisova and Altai Neanderthal genomes[4,5]. Following previous findings, we modeled the Denisovans as being admixed between the Neanderthal lineage and a more basal ('unknown archaic') lineage[4]. We also modeled

Neanderthal admixture in the ancestral non-African population[6], and Denisovan admixture in the ancestral population of Australians and Papuans[7]. As a final expansion of our population history model, we added the Andamanese Onge on the lineage ancestral to Australians and Papuans but diverging prior to the Denisovan admixture event[3]. We found that the resulting graph (**Figure 3A**) fits well given the large number of populations included. There are only two outliers that deviate by slightly more than 3 standard errors from empirical statistics:

f_4(Yoruba, Tongan; Denisova, Altai): model = 0.001284, empirical = 0.006947, (Z = 3.1)
f_4(Atayal, Mamanwa; Onge, Tongan): model = -0.010912, empirical = -0.007444, (Z = 3.2)

The first outlier could be interpreted as suggesting unmodeled affinity between Tongans and Neanderthals, but the second one is more difficult to interpret, as it is consistent for example with unmodeled affinity between Mamanwa and Tongan or Onge and Atayal.

Adding the ancient Tonga sample

For the graph we display in **Figure 3**, we do not include the Mamanwa but added the Lapita_Tonga sample which despite being represented by only a single individual can be adequately fitted as being a clade with the Lapita_Vanuatu sample. The three minor outliers with Z-scores between 3 and 3.5 were all for population configurations that have little to do with the history of admixture and population splits in Oceania that we are most interested in:

f_4(Yoruba, New_Guinea; Yoruba, Dai): model = 0.143671, empirical = 0.133643, (Z = 3.3)
f_4(Yoruba, Dai; Yoruba, Dai): model = 0.193597, empirical = 0.183619, (Z = 3.0)
f_4(Yoruba, Dai; Dai, Atayal): model = 0.004789, empirical = -0.000185, (Z = 3.4)

We note that we can also fit modified models with the Lapita_Tonga as being a clade with the First Remote Oceanian lineage contributing to present-day Tongans with approximately the same support, but we chose to conservatively display the graph in **Figure 3**.

References:

1 Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489-494 (2009).
2 Patterson, N. *et al*. Ancient admixture in human history. *Genetics* **192**, 1065-1093 (2012).
3 Reich, D. *et al*. Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics* **89**, 516-528, (2011).
4 Prufer, K. *et al*. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).
5 Meyer, M. *et al*. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* **338**, 222-226, doi:10.1126/science.1224344 (2012).
6 Green, R. E. *et al*. A Draft Sequence of the Neanderthal Genome. *Science* **328**, 710-722 (2010).
7 Reich, D. *et al*. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).

Title

Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neandertals

Cosimo Posth [1,2,*], Christoph Wißing [3], Keiko Kitagawa [1,4], Luca Pagani [5], Laura van Holstein[6], Kurt Wehrberger [7], Nicholas J. Conard [1,8], Claus Joachim Kind [9], Hervé Bocherens [3,10] and Johannes Krause [1,2,*]

[1]*Institute for Archaeological Sciences, University of Tübingen, Rümelin Strasse 23, 72070 Tübingen, Germany*
[2]*Max Planck Institute for the Science of Human History, Khalaische Strasse 10, 07745 Jena, Germany*
[3]*Department of Geosciences, Biogeology, University of Tübingen, Hölderlin Strasse 12, 72074 Tübingen, Germany*
[4]*Department of Prehistory, National Museum of Natural History, UMR 7194 CNRS, 1 rue René Panhard, 75013 Paris, France*
[5]*Estonian Biocentre, Riia 23b, 51010 Tartu, Estonia*
[6]*Department of Archaeology and Anthropology, University of Cambridge, Fitzwilliam Street, CB2 1QH, Cambridge, UK*
[7]*Ulmer Museum, Marktplatz 9, 89073 Ulm, Germany*
[8]*Department of Early Prehistory and Quaternary Ecology, University of Tübingen, Schloss Hohentübingen, 72070 Tübingen, Germany*
[9]*State Office for Cultural Heritage Baden-Württemberg, Berliner Strasse 12, 73728 Esslingen, Germany*
[10]*Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Hölderlin Strasse 12, 72074 Tübingen, Germany*

Abstract

Ancient DNA retrieved from archaic human remains has revealed new insights into genetic relationships between Pleistocene hominins and modern humans. Nuclear DNA attested Denisovans as a sister group of Neandertals after diverging from the modern human lineage. However, the closer affinity of the Neandertal mitochondrial DNA (mtDNA) to modern humans than Denisovans has been recently interpreted as the result of gene flow from Africa into Neandertals. Here we report the complete mtDNA of a hominin femur displaying archaic features (HST) from the Hohlenstein-Stadel cave in southwestern Germany. HST represents the deepest divergent mtDNA that split from other Neandertals around 270,000 years ago, providing a lower boundary for the time of the putative mtDNA introgression event. We demonstrate that a complete mtDNA replacement in Neandertals is feasible over such a time interval even with minimal hominin introgression from Africa. Finally, the highly divergent HST branch is indicative of greater Neandertal mtDNA diversity during the Middle Pleistocene than in later periods.

Introduction

In recent years, an increasing number of mitochondrial (mtDNA) and nuclear genome (nDNA) data from archaic human remains have reshaped the understanding of evolutionary relationships among various hominin groups. Mitochondrial genomes provided evidence for at least two distinct mtDNA branches associated with Neandertals and Denisovans respectively, suggesting a sister group relationship between modern humans and Neandertals with Denisovans as a basal mtDNA outgroup[1-4]. Nuclear DNA data however revealed that Neandertal and Denisovan populations separated only after their divergence from the lineage leading to modern humans[2,5-8].

The estimate for the population split time between the two archaic hominin groups and modern humans was calculated to 765,000-550,000 years ago (765-550 ka) based on a recent estimate of the genome-wide human mutation rate[5]. Furthermore, analyses of Y chromosome data from a male Neandertal returned an age of 806-447 ka for the divergence of Neandertal and modern human Y chromosome lineages[9]. These time intervals largely overlap, suggesting that the Neandertal Y chromosome differentiated through the population split from the most recent common ancestor (MRCA) of modern humans and Neandertals. In contrast, the corresponding divergence time for mtDNA has been dated to ~400 ka (95% highest posterior density [HPD], 498-295 ka)[10,11] and was thus found to be considerably younger compared to the time estimates obtained from autosomal and Y chromosome data.

In addition, nDNA analyses of the Middle Pleistocene hominins from the Sima de los Huesos site in northern Spain confirmed their closer affinity to the Neandertal lineage[8], suggesting that at least by ~430 ka, Neandertals and Denisovans had already diverged (Fig. 1d). However, in contrast to genome-wide data, the Sima de los Huesos mtDNA was found to branch off with the deeply divergent Denisovan mtDNA lineage[3]. The phylogenetic discrepancies could be reconciled if the mtDNA of early Neandertals was indeed Denisovan-like and subsequently replaced by a more derived mtDNA lineage. Therefore, a genetic introgression event from African hominins into the early Neandertal population that gave rise to the "Late Pleistocene" Neandertal mtDNA lineage has been proposed[8]. This event must have occurred after archaic and modern human populations diverged. However, the exact timing of the proposed gene flow is unknown and merely based on possible archaeological evidence for contacts between African and Eurasian populations[8].

While genomic evidence showed that gene flow between lineages as divergent as humans and Neandertals took place[12,13] in both directions[14], it is unclear whether such small scale phenomena (all inferred introgression events account for less than 5% of the genome of either populations) were sufficient to explain the complete replacement of the initial Neandertal mtDNA pool (found in Sima de los Huesos) by a Middle Pleistocene human lineage from Africa. Moreover, the temporal placement of such admixture event into Neandertal populations is still under debate, partly due to the limited availability of additional archaic DNA. Therefore an assessment of the feasibility of such a replacement as well as the availability of more ancient specimens are required to conclude whether the African introgression hypothesis is a viable one and to refine its time boundaries.

Here we use an analytical approach to explore the possibility of the proposed mtDNA turnover and we introduce a novel Neandertal mtDNA lineage (HST) from a femur found at the Hohlenstein-Stadel cave of the Swabian Jura in southwestern Germany[15] (Fig. 1a and 1b). We use HST to further explore the mtDNA genetic diversity among

archaic humans and describe their phylogenetic relationships with modern humans to infer Neandertal demographic processes across the Middle and Late Pleistocene.

Results

Archaeology and stable isotopes

The HST specimen is a right femur shaft circa 25 cm long displaying archaic hominin morphology, affected by heavy mineralization and gnawing by a large carnivore on both sides[15] (Fig. 1a). During excavations in 1937, it was found in a black clayey layer named Black Mousterian based on the sediment color and the cultural assignment of the techno-complex retrieved in the stratigraphic unit, which is associated throughout Europe with Neandertals[16]. The femur is the sole archaic human remain originating from a Mousterian context, not only at the site, but in the entire Swabian Jura region[17] (Supplementary Note 1).

Direct radiocarbon dating attempts have resulted in inconsistent results (Supplementary Note 2) suggesting that the bone may be suffering from modern ^{14}C contamination and possibly be beyond the detection limit of this dating method. Isotopic analyses performed on the femur's collagen revealed considerably lower $\delta^{13}C$ and $\delta^{15}N$ values than those reported for late Neandertals from western and central Europe (Supplementary Table 1 and Supplementary Fig. 1)[18]. Moreover, collagen from two faunal remains recovered from the same stratigraphic unit of HST was analyzed. ZooMS analyses[19] confirmed the morphological assignment to red deer and radiocarbon dating resulted in an age range beyond radiocarbon dating (Supplementary Table 1 and Supplementary Note 2). Both deer specimens provided notably lower $\delta^{13}C$ values compared to cervids from open steppic environment[20] (Supplementary Fig. 2). The ecological background of the HST femur and deer specimens is therefore equivalent and indicates a more forested, closer environment compared to the habitat of late Neandertals in northwestern Europe[18,21].

Ancient DNA retrieval and consensus reconstruction

The femur shaft was sampled from the proximal diaphysis longitudinally to the cortical bone, at the opposite site of the previous sampling for radiocarbon dating. DNA was extracted from 130mg of bone powder[22], immortalized in a double stranded library[23] and hybridized to modern human mtDNA probes[24]. The enriched library was sequenced and between 12,750 and 12,848 DNA reads were successfully aligned to four reference sequences with the same mapping parameters (Methods section): the reconstructed MRCAs of Neandertal and *Homo sapiens* mtDNA (RNRS and RSRS)[25], the Neandertal type specimen mtDNA (Feldhofer1)[26], and the present-day human mtDNA reference (rCRS)[27], respectively. A consensus sequence for each of the four references was reconstructed with *endoCaller* implemented in the software *schmutzi*[28] followed by visual inspection to confirm the called polymorphisms (Methods section). Using the RNRS as reference sequence resulted in the highest number of mapped reads and ~35fold average mtDNA coverage. Around 50% of mtDNA fragments were damaged at the molecule termini with an average length of ~43bp, both displaying the degradation patterns typical for ancient DNA (aDNA)[29] (Supplementary Table 2 and Supplementary Fig. 3). When comparing the four consensus sequences obtained by mapping against the different references we observed the influence of reference biases in reconstructing the HST mtDNA (Supplementary Fig. 4a). After manual inspection

of the inconsistent positions we identified RNRS as the reference producing a consensus sequence closest to the endogenous mtDNA (Methods section). However, mapping bias disappeared when excluding from the alignment the highly variable D-loop region (Supplementary Note 3 and Supplementary Fig. 4b). Following a more conservative approach subsequent Bayesian and phylogenetic analyses were performed using the reconstructed HST mtDNA coding region. The phylogenetic comparison with 54 modern human, three Denisovan and an extended dataset of 17 Neandertal mtDNA sequences revealed a closer relationship of the femur's mtDNA to Neandertals. However, the HST mtDNA revealed a short phylogenetic branch length and fell basal to all other Neandertal individuals, representing the deepest diverging lineage among Neandertal mtDNAs discovered to-date (Fig. 1c, Supplementary Fig. 5 and Supplementary Fig. 6).

The same HST genetic library before mtDNA capture was also sequenced through a shotgun approach. Only 0.46% of the over half million reads were aligned to the human reference genome (hg19) despite choosing high sensitive mapping parameters to account for aDNA damage and divergence from the reference sequence (Supplementary Table 3 and Methods section).

Contamination estimates

Three measurements were performed to estimate the level of present-day human contamination in the isolated mtDNA reads. The first approach is based on the assumption that aDNA is damaged whereas contaminant DNA is less effected by this chemical modification (*contDeam*[28]). One molecule end is conditioned to exhibit damage while deamination levels are measured at the opposite end of the fragment. The discrepancy between the unfiltered and conditioned damage levels required a contamination with 9.5-11.5% of modern humans fragments (Supplementary Table 2).

This estimate is used as prior in an iterative likelihood approach in which mtDNA reads are compared to a dataset of 256 Eurasian modern mtDNA sequences to refine the level of contamination (*mtCont*[28]). According to this second method, 9-11% of the bases aligned to the rCRS turned out to be of contaminant origin.

Third, we identified mtDNA diagnostic positions, as the nucleotides where the reconstructed HST complete sequence differed from more than 99% of 311 worldwide mtDNAs[30]. From 123 differences only eleven transversions were considered in order to avoid the risk of wrongly classifying damage that is typically seen as transitions[31], as real substitutions. For each transversion we counted the total number of fragments harboring a base consistent with the HST consensus over the ones showing the almost fixed modern human variant. The mtDNA contamination was estimated to be 5.4-12.2%.

Overall the three approaches consistently returned modern human DNA contamination levels with an upper value of ~12%. This may be associated to the presence of modern collagen contamination that possibly resulted in inconsistent radiocarbon dates. While mtDNA consensus sequences can be confidently reconstructed with such contamination proportions[32], nuclear DNA analyses would be highly affected.

MtDNA Neandertal diversity

In a previous study[2], the mtDNA diversity among seven Neandertals, three Denisovans and 311 modern humans were compared through the Watterson's estimator θw, resulting in the lowest mtDNA distance within Neandertals. The value decreased even further when ten additional Neandertal mtDNAs available in the literature were included ($1.37*10^{-3}$), which confirms the small population size of late Neandertals[26] (Supplementary Table 4). However, by adding the HST mtDNA in the Neandertal group the θw estimation almost doubled to $2.50*10^{-3}$. Although the value is still below the results obtained from the three Denisovan sequences ($3.46*10^{-3}$), the HST mtDNA exhibits an average pairwise nucleotide distance to the other Neandertal mtDNAs of 104 (89-111) positions (Fig. 2 and Supplementary Table 5). These values are greater than among any Denisovan mtDNA pair and are in the upper range of modern human worldwide pairwise distance distribution (Fig. 2). This shows that HST belongs to a mtDNA branch highly divergent from the one represented in other Neandertals (Altai branch) and overall Neandertal mtDNA diversity was larger than previously assumed.

The Neandertal mtDNA effective population size (*Ne*) through time was estimated in a Bayesian statistic framework[33] under the simplified assumption they belonged to a panmictic population with a fixed mutation rate previously calculated with ancient modern human mtDNAs as calibration points[10] (Supplementary Note 4). The reconstructed skyline plot describes a *Ne* reduction through Middle and Late Pleistocene, reaching the lowest mean value at around 42 ka (Supplementary Fig. 7). Subsequently a steep population expansion appears to have occurred before the Neandertal extinction, in accordance to the reported analyses of chromosome 21 of the Vindija late Neandertal[14].

Molecular dating analyses

In order to estimate the molecular age of HST and other undated Neandertal mtDNAs as well as the temporal range of most recent common ancestors (TMRCAs) on the mtDNA tree, we performed a Bayesian dating analysis as implemented in BEAST v1.8.1[33]. A multiple genome alignment of the coding region from 54 modern humans, 18 Neandertals and one Denisovan mtDNA were tested for strict and uncorrelated lognormal relaxed clock under both a constant size and a Bayesian skyline tree prior (Methods section). As reported above, a fixed mutation rate was selected for the coding region[10] with the addition of eight dated Neandertal mtDNAs as time anchors on the Neandertal branch (Supplementary Table 6). The four model combinations were compared by stepping-stone and path sampling methods[34] attesting skyline associated to a strict rate variation among branches as the model that most adequately fits the data (Supplementary Table 7). In Table 1 we report the TMRCAs between Neandertal and modern human mtDNAs and among modern human mtDNAs itself, which largely overlap with previously published studies[10,11]. We further estimate the divergence time between HST and all other Neandertals to ~270 ka (95% HPD 316-219 ka), while the TMRCA for the Altai branch was assessed to ~160 ka (95% HPD 199-125 ka).

Based on phylogenetic branch shortening, we furthermore molecularly dated ten Neandertal sequences that had not been radiocarbon dated previously or were considered beyond the radiocarbon dating detection limit (Table 1). The two oldest mtDNAs were HST with an age of 124 ka (95% HPD 183-62 ka) and Altai

Neandertal estimated to 130 ka (95% HPD 172-88 ka). Notably, the mean value for the latter individual largely overlaps with the inferred age of 136-129 ka from its high coverage nuclear genome analyses, when applying recent estimates of the human mutation rate[5].

Analytical exploration of putative Neandertal mtDNA replacements

The probability of having the initial Denisovan-like Neandertal mtDNA present in Eurasia totally displaced by an incoming lineage[8] is highly dependent on the size of the introgressing population compared to the local one. Assuming that a complete mtDNA replacement took place we estimated under neutrality[35] (Methods section) the mean time period necessary for such an event to reach fixation given a mtDNA introgressing fraction below 20% and initial effective population size (*Ne*) up to 10,000 units (Supplementary Table 8). We molecularly dated the split of the HST lineage from other Neandertal mtDNAs to ~270 ka (Table 1) that represents the minimum time available for the Late Pleistocene branch to replace the pre-existing Denisovan-like mtDNA. From our calculations, if *Ne* was less than 5,000 units a mean temporal interval of 300 ka is sufficient for an incoming mtDNA lineage even below 0.1% in frequency to drift up to fixation.

Within the Late Pleistocene mtDNA clade, we explored if the HST mtDNA branch might have survived long after the estimated molecular age of the HST femur. All complete Neandertal mtDNAs were combined with sequences from published hypervarible regions (HVRI) of four additional Neandertal individuals. We identified the Valdegoba sequence (JQ670672) sharing three derived mutations with HST and falling on the same branch in a HVRI tree (Supplementary Fig. 8 and Methods section). This specimen was found on the Iberian peninsula and dates to 48,400 ± 3,300 ^{14}C years BP[36]. Although a complete mtDNA would be necessary to measure the total mtDNA distance between HST and Valdegoba, this finding might suggest that the HST branch was found during the Late Pleistocene as far as Western Europe. Based on geographical and temporal distributions of HVRI sequences, it was proposed that the Neandertal population in Western Europe underwent a demographic turnover followed by a subsequent recolonization[36]. Under that scenario, the HST lineage would have been largely replaced towards the end of the Neandertal temporal range from mtDNAs descendants on the Altai branch.

Discussion

Our analytical calculations (Supplementary Table 8) show that the African introgression hypothesis suggesting that Late Pleistocene Neandertal mtDNAs originated from gene flow of an African mtDNA more than ~270 ka is plausible even if the introgressing lineage represented a minimal proportion of the initial gene pool. This scenario reconciles the discrepancy in the nDNA and mtDNA phylogenies of archaic hominins and the inconsistency of the modern human-Neandertal population split time estimated from nDNA and mtDNA (Fig. 1d). Under this demographical model the Denisovan mtDNA type was common among early Neandertals in Eurasia (e.g. Sima de los Huesos) and was then largely replaced by an introgressing African mtDNA that evolved into the Late Pleistocene Neandertal mtDNA type. While the upper bound for the time of this putative gene flow event would be the divergence time between Neandertal and modern human mtDNAs, here dated to 413 ka (95% HPD 468-360 ka), the lower temporal limit was represented so far by the ~160 ka

TMRCA of all published Neandertal mtDNAs (Table 1). However, the finding of the deeply diverged HST lineage splitting from the Altai branch ~270 ka, sets an older lower boundary for the time of this admixture event. An alternative but less parsimonious scenario is that both HST and Altai mtDNA lineages reached Eurasia independently after diverging inside Africa. In that case the suggested introgression event might have occurred later but most likely before 160 ka, our estimated date for the start of the Altai branch diversification (Fig. 1c and Table 1).

The presence of modern human admixture into archaic humans has already been detected in the high coverage Neandertal genome from the Altai region but not in sequences of chromosome 21 of two Neandertals from Spain and Croatia[14]. The authors therefore suggested that a genomic contribution estimated between 0.1 and 2.1% occurred after the divergence of Altai from other late Neandertals. However, there is a high level of uncertainty around the time of the inferred gene flow event since only one high coverage Neandertal nuclear genome has been analyzed so far. Moreover, the divergence time of the introgressing African population was estimated to date before or right after the TMRCA of modern-day humans (~200 ka)[14], while the mtDNA coalescence time between Neandertals and modern humans is calculated at least twice as old (~400 ka). The evolutionary scenario responsible for providing the mtDNA to the Late Pleistocene Neandertals might have been an even earlier Middle Pleistocene gene flow from Africa, occurring in a time interval that we date between 413 ka and 268 ka (460-219 ka including upper and lower 95% HPD). It should be highlighted that this additional genomic contribution might have already been accounted for by Kuhlwilm *et al.*[14]'s estimation, which effectively measures the total amount of African introgression into Neandertals after their split from Denisovans (473-381 ka[5]).

The phylogenetic branch length of mtDNA sequences from ten non-dated Neandertal individuals was considered in BEAST, to assess individual molecular ages spanning from 130 to 40 ka. Although it is not known if the mtDNA mutation rate in modern humans is comparable to that of Neandertals (Supplementary Note 4), molecular dating can at least be used to provide relative ages when the radiocarbon absolute chronometric method is not applicable. After the Altai mtDNA, HST is estimated to be the second oldest mtDNA with an age of 124 ka (95% HPD 183-62 ka). This wide temporal interval largely overlaps with the Marine Isotope Stage 5 (MIS 5: ~130-73 ka). After its initial interglacial period (MIS 5e: 130-122 ka), central Europe was characterized by climatic fluctuations resulting in forestation cycles (MIS 5c/5a) alternated with the development of steppe-tundra biomass (MIS 5d/b)[37]. The stable isotopic $\delta^{13}C$ and $\delta^{15}N$ values of the archaic femur collagen and associated faunal remains support a more temperate, forested rather than a colder, steppe environment and is therefore consistent with an ecological context during the early warm phases of the last glaciation (MIS 5d-a: 122-73)[17].

Despite having only a single complete mtDNA on the HST lineage, the two highly differentiated Neandertal mtDNA branches suggest higher mtDNA diversity during the Middle Pleistocene, which then reduced during the Late Pleistocene (Supplementary Table 4). This observation is also supported by the steady decline in mtDNA effective population size displayed in skyline plots before a steep growth in late Neandertals (Supplementary Fig. 7). Studies focusing on the demographic patterns of late Neandertals who overlapped with the earliest modern humans in Europe are of key importance to understand population dynamics and interactions between archaic and modern humans.

In conclusion, the HST mtDNA provided insights into the mtDNA diversity of the Neandertal populations through the Middle and Late Pleistocene. Its deep divergence time allowed us to further constrain the lower boundary for the time of the proposed African mtDNA gene flow into Neandertal populations. The time range for this introgression event between 460 ka and 219 ka is compatible with the evidence of archaeological similarities between Africa and western Eurasia during the Late Middle Pleistocene[38]. Interglacial conditions across this time span might have facilitated a hominin expansion out of Africa and potentially spread cultural innovations such as the Levallois technology into Eurasia[39].

Nuclear data from the HST femur would be pivotal in assessing its genomic relationships with Neandertals, Denisovans and modern humans. However the scarce preservation of HST endogenous DNA in combination with high level of modern human contamination challenge the retrieval of its complete genome. Analyses of high quality nDNA from more than one well-preserved Neandertal individual are necessary to detect the consequences of African admixture into archaic human populations.

Methods

Ancient DNA lab work

Ancient DNA work was performed in the dedicated facilities of the Institute for Archaeological Sciences in Tübingen, Germany. The HST femur was first irradiated with UV light on the selected sampling area and then drilled with a dentist drill along the cortical bone. A total of 130 mg of bone powder went into the DNA extraction following an established protocol[22]. DNA was eluted in 100 µl of TET and 20% of the extract (GX35) was used to build a double stranded genetic library (GA87)[23]. The total copies in the resulting library were measured with qPCR ($7.53*10^9$ copies). They were split into four 100 µl indexing PCR reactions with 10 cycles where an individual index pair (8bp each) was assigned in order to create an unique double indexed library[40]. The total copies were measured again via qPCR ($2.60*10^{11}$ copies) and the reaction efficiency was calculated by dividing the number of total molecules after indexing by the number of total molecules before indexing PCR. An aliquot of two fifth of the indexed library was split into two reactions that were amplified for 7 cycles each with AccuPrime Pfx DNA polymerase. The PCR products were purified over a single MinElute spin column and the concentration after amplification was quantified to 286 ng/µl on an Agilent 2100 Bioanalyzer DNA 1000 chip. Extraction and library negative controls were carried along the workflow and treated equally.
The amplified library was enriched for mtDNA using modern human baits as reported by Maricic *et al*.[24]. This protocol has been previously used to successfully capture complete Neandertal mtDNA genomes[4]. 400 ng of the amplified library were pooled with the same amount of four other libraries for a total of 2,000 ng and captured with 500 ng of mtDNA probes. After purification the isolated molecules were quantified with qPCR ($4.84*10^6$ copies) and re-amplified for additional 20 cycles as described above. The captured pool as well as the uncaptured GA87 library was quantified with Agilent 2100 Bioanalyzer DNA 1000 chip, diluted to 10nM and sequenced with other equimolar libraries on an Illumina HiSeq2500 Rapid run via 2x100+8+8 cycles and on an Illumina NextSeq500 run via 2x75+8+8 cycles.

Sequence processing and mtDNA consensus reconstruction

Sequenced molecules were converted from bcl to fastq files and reads containing the defining library indexes were binned in an individual folder. EAGER pipeline was used for all subsequent data processing[41]. Initially adapter and index sequences were trimmed off. Only merged reads where forward and reverse reads overlapped by at least 10bp were retained. Shotgun sequences above 30bp were aligned to the complete human genome (hg19) with BWA (parameters -n 0.01 and seeding off) to calculate the percentage of human DNA. Duplicates and reads with mapping quality below 30 were discarded to estimate damage patterns and average fragment length (Supplementary Table 3). From the total of ~3Ma paired-end reads sequenced after mtDNA capture, 89.31% were successfully merged and fragments below 30bp length were further discarded for mapping. The resultant ~1.3 Ma merged reads were aligned to four reference mtDNA sequences: the reconstructed Sapiens reference sequence (RSRS)[25], the revised Cambridge reference sequence (rCRS)[27], the Neandertal Feldhofer 1 sequence[26] and the reconstructed Neandertal reference sequence (RNRS) originally proposed in Behar et. al[25] and later updated when the more basal Altai mtDNA was published[5]. The rCRS and Feldhofer 1 references are two derived

mtDNA sequences on the modern human and Neandertal branch, respectively. Instead, RSRS and RNRS represent the most recent common ancestor mtDNA for modern humans and Neandertals, respectively. Reads were mapped using BWA[42] with identical parameters (-n 5 and seeding off) for all four references, in combination with a tool able to consider the circularity of mtDNA as part of EAGER. The percentage of target DNA was calculated dividing the total number of input reads by the reads mapping to each mtDNA reference. Duplicates with the same start and end coordinates were removed and the duplication factor was measured dividing the total reads mapping before by total reads mapping after duplicate removal. All fragments with map quality below 30 were removed to estimate the average mtDNA coverage. The resulting molecules were also used to calculate average fragment length and deamination patterns[43] (Supplementary Fig. 3). Statistics for each processing step of the four reference sequences are reported in Supplementary Table 2.

Consensus reconstruction was performed in a two-step approach. First, *schmutzi*[28] was used to infer the endogenous sequence. An internal program of the software package (*contDeam*) was first run to calculate the endogenous deamination rate and a contamination prior. To each nucleotide a base likelihood value was assigned incorporating damage, base quality and mapping quality information in a Bayesian framework[28]. The endogenous consensus was then determined by *endoCaller* after the first iteration of the program. No cut-off to the nucleotide posterior probability was selected resulting in base called even in positions covered with only one fragment. This produced a consensus sequence with three unassigned positions.

Second, the four consensus sequences, one from each reference, were visually compared in Geneious 8.1.7 (http://www.geneious.com)[44]. A multiple genome alignment was produced and each of the 19 inconsistent positions between the four consensus was evaluated. We imported the bam files in Geneious and for each read covering those positions we inspected if they also overlapped neighboring confidently assigned SNPs (e.g. called in all four consensus). Fragments containing such SNPs were considered as endogenous whereas reads presenting the alternative allele were considered as contaminant. In every case the consensus sequence reconstructed after mapping against the RNRS reference was found to exhibit the endogenous base. This confirms that mapping against a reference sequence that is phylogenetically closer to the consensus sequence increases mapping accuracy (see Supplementary Note 3). Using the same criterion described above we furthermore manually screened the RNRS mapped consensus and edited the following positions according to rCRS coordinates. Two miss-mapped insertions were removed (pos. 247delT and 16184delA), two uncertain positions with low coverage were edited (A189G and A4296N) and two regions covered with only contaminant reads were masked (pos. 203-214Ns and 5486-5508Ns). We additionally masked the known troublesome regions of poly-C (pos. 303-315) and poly-AC (518-524) stretches[45]. Combining the two approaches resulted in a total of 59 unassigned positions in the final consensus sequence that was used for phylogenetic (Supplementary Fig. 5) and mtDNA diversity analyses. We furthermore generated a more conservative consensus by setting a coverage cut-off to 2-fold. The resulting mtDNA sequence exhibits 81 Ns but none of the 22 additional unassigned bases were covering polymorphic positions within the known Neandertal mtDNA diversity. Therefore the tree topology and mtDNA diversity within Neandertals was not affected.

Contamination with modern human mtDNA

We followed three different approaches to estimate modern human contamination levels in the isolated mtDNA. The first method is implemented in *contDeam*[28] and it relies on deamination patterns. This program works on two assumptions e.g. that modern human DNA contamination presents no damage and that the damage at a molecule end is independent to the one at the other end. Reads with deamination at 5' end are selected and the deamination rate is measured at the 3' end and vice versa. The calculated value is supposed to represent the true damage signal of the endogenous mtDNA fragments. Contamination estimate is then computed as the percentage of undamaged reads necessary to shift the damage rate from the endogenous value to the one initially calculated on all fragments. We obtained an estimate ranging from 9.5% to 11.5% (for all four references combined) that could be an underestimation of the real contamination level if the contaminant DNA is also damaged. However simulations have shown that this effect is marginal if the deamination rate of the endogenous DNA is over 50% at the molecule termini[28], like observed for HST mtDNA fragments (Supplementary Fig. 3).
The second method makes use of the probabilistic iterative method integrated in *schmutzi*[28]. The program was run with the fallowing parameters: "-- notusepredC – uselength". The contamination estimate is performed with the tool *mtCont* using sites where the endogenous sequence differs from a non-redundant dataset of 256 Eurasian mtDNAs. For this method we used the mtDNA reads mapping to the rCRS reference according to which base frequencies of the comparative dataset are calculated. While *contDeam* measured contamination rate on a fragment level, *mtCont* provided an estimate at a base level of 9-11%.
The third method is based on the diagnostic positions where the reconstructed HST consensus differs from present-day worldwide mtDNAs. All polymorphic positions with a frequency above 1% in a dataset with 311 worldwide mtDNAs are not considered. We then identified 123 positions where HST has a different base compared to more than 99% of the 311 mtDNAs. Of those, we restricted the analysis to only transversions (positions 2831, 6265, 7105, 9328, 9354, 11457, 13761, 13878, 14457, 14925, 16138). Of the total 262 reads covering the eleven positions, 239 reads presented the endogenous base while 23 reads the contaminant variant. This resulted in a contamination rate of 8.8% (CI 95%, 5.4-12.2%). The last approach[46] provides a direct measure for the proportion of contaminant fragments and overlaps with the two previous estimates.

Phylogenetic analyses

To further explore the maternal relationships of the HST mtDNA with other archaic and modern human mtDNAs we compared the phylogenetic placement of the HST consensus sequence with and without D-loop with 17 Neandertal, 54 modern human[47], three Denisovan[1,2] and Sima de los Huesos[3] mtDNAs, plus a chimpanzee mtDNA (GenBank: X93335.1) to root the tree. Two maximum parsimony trees with the 77 mtDNAs and 1000 iterations each were built, one including the whole molecule and 97% partial deletion (16,536 positions) and one with the coding region only and 98% partial deletion (15,417 positions) (Fig. 1c and Supplementary Fig. 5). The topology of both trees is consistent with HST diverging from the Neandertal branch more basal than any other sequence and presenting a short phylogenetic branch length.

We also tested in Modelgenerator[48] the same multiple genome alignment with only the coding region, including missing sites but not gaps. The substitution model that best fits the data (AIC1) was GTR with invariant sites and gamma-distributed correction for rate heterogeneity. These parameters were selected in MrBayes[49], used to build a Bayesian phylogenetic tree (Supplementary Fig. 6). 50 millions iterations of the Markov chain Monte Carlo (MCMC) were run with 10,000 sampling interval. From the total trees, the first 10% were removed as burn-in and a summarized tree was generated. All major branches show posterior support of 1 and confirmed the maximum parsimony trees topology.

We finally explored the diversity of the Neandertal hypervariable region I (HVRI) including four additional sequences for which only HVRI was available in GenBank (Valdegoba JQ670672, Scladina DQ464008, Teshik-Tash EU078679, Monti Lessini DQ836132). We aligned them to the HVRI of 17 complete Neandertal (excluding Denisova 11 because of several unassigned positions in the HVRI), three Denisovans and the rCRS mtDNAs using MUSCLE[50]. We then built a maximum parsimony phylogeny in MEGA6[51] with complete deletion (105 positions) and 1000 bootstrap iterations (Supplementary Fig. 8).

Mitochondrial DNA diversity

The pairwise nucleotide distance among Neandertals with (n=18) and without HST (n=17), Denisovan (n=3) and modern human (n=311) mtDNAs was calculated in MEGA6. For this analysis we used the complete mtDNAs sequences and the number of differences between them was counted with pairwise deletion where all unassigned positions were removed for each sequence pair. We plotted the pairwise nucleotide distance against their frequencies for each of the four datasets (two Neandertals, Denisovan and modern human) in Figure 2. We also reported the average distance (and minimal - maximal values) of 311 modern humans, three Denisovans, Sima de los Huesos and 17 Neandertals to the HST complete mtDNA (Supplementary Table 4). The lowest distance is with Neandertals, followed by modern humans, Sima de los Huesos and Denisovans in agreement with the phylogenetic assignment. However the nucleotide distances between HST and other Neandertals are the largest observed among Neandertals (89-111 nucleotides). These values are higher than between Denisova 3–Denisova 4 and Denisova 8 and around the uppermost edge among 311 worldwide mtDNAs (Fig. 2).

We further measured the mtDNA diversities of the enlarged Neandertal mtDNA dataset with the Watterson's estimator, as reported in Sawyer *et al*.[2]. We first prepared a multiple genome alignment of Neandertal mtDNAs both including HST (18 sequences) and excluding HST (17 sequences) using MUSCLE. Then, the number of segregating sites (K) was estimated with DNA Sequence Polymorphism (DnaSP) v.5.10.01[52]. Finally, θw was calculated as follows: $K / a_n / 16{,}595$ where a_n is $\Sigma_{i=1}^{n-1} 1/i$ to take in consideration the number of mtDNA sequences in each dataset. Adding HST to the 17 Neandertal mtDNAs the number of segregating sites almost doubled (from 78 to 145) while θw increased from $1.37*10^{-3}$ to $2.50*10^{-3}$ (Supplementary Table 4). The latter value is closer to the mtDNA diversity estimated within three Denisovan mtDNAs $(3.46*10^{-3})$[2].

BEAST analyses

We used the software package BEAST v1.8.1[33] to both estimate the divergence times between and within modern and archaic humans as well as to track the changes in the maternal effective population size (*Ne*) of Neandertal mtDNAs through time.

For the skyline analyses we first created a multiple genome alignment with only the mtDNA coding region of 18 Neandertal mtDNAs and the rCRS as outgroup. We then removed from the alignment all columns where at least one mtDNA presented a gap or missing data, resulting in 15,345 positions. We run Modelgenerator v.85[48] on our dataset to identify Tamura-Nei 93 with a fixed fraction of invariable sites as the best-supported model. We set a fixed mutation rate ($1.57*10^{-8}$ μ / site / year)[10] calculated for the coding region of modern humans with ancient mtDNAs as calibration points (Supplementary Note 4). As tree prior we selected the Bayesian Skyline coalescent with 10 as group number and piecewise-linear as the Skyline model. We tested both a strict clock and an uncorrelated lognormal-distributed relaxed clock. For both models three MCMC runs with 50,000,000 iterations were run, with 10,000 sampling frequency. We discarded 10% of the states from each run as chain burn-in and then combined the three independent runs for both models using LogCombiner v1.8.1 (included in the BEAST package), resulting in a total of 135 millions iterations. The two models were compared with a marginal likelihood estimation (MLE) using path sampling (PS) and stepping-stone sampling (SS)[34]. The Skyline tree prior in combination with a strict variation among tree branches performed better according to PS while lognormal-distributed relaxed clock was best supported according to SS (Supplementary Table 7). The strict clock provided higher ESS values because of earlier chain convergence therefore it was the chosen model to reconstruct a Skyline plot for the 18 Neandertal mtDNAs. We used Tracer v1.6 selecting linear change as Bayesian skyline variant and a default of 100 as number of bins. In Supplementary Figure 7 we report the mean *Ne* (black line) and the 95% HDP interval (purple lines) of the Neandertal mtDNAs in logarithmic value on the Y-axis and the temporal range from 350 ka to 32 ka on the X-axis. We observe a *Ne* reduction until around 42 ka followed by a rapid and short growth inversion, predating the Neandertal disappearance.

For the dating analyses instead we used a dataset composed by 18 Neandertal, 54 modern humans and 1 Denisovan mtDNAs as outgroup. As described above we removed the D-loop from the alignment and further excluded all positions containing gaps and missing data for a total of 15,334 positions. The best-supported model for this dataset was again Tamura-Nei 93 with invariable sites as tested in Modelgenerator v.85. The same fixed mutation rate was selected and tip dates were indicated for the eight dated Neandertal mtDNAs (Supplementary Table 6). The date for all 54 mtDNA was kept as zero while a range between 30 ka and 500 ka (initial value 50 ka) was given for all the undated Neandertals. We tested two models of rate variation within branches: a strict clock and an uncorrelated lognormal-distributed relaxed clock. We also investigated two different tree priors: a Bayesian Skyline coalescent and a population Constant size. As before, for each of the four model combinations MCMC was run three times with 50,000,000 iterations, sampling frequency 10,000 and 10% burn-in. The resulting 135,000,000 iterations per model were combined using LogCombiner v1.8.1. Best-supported model assessment was again implemented with MLE using PS and SS. The strict molecular clock with the Skyline tree prior provided higher likelihoods than the three other tested models (Supplementary Table 7).

Replacement inference

Under neutrality and assuming the Eurasian Neandertal effective population size (*Ne*) to be relatively small (i.e. <10,000 *Ne* units), we calculated the mean time period necessary for an introgressing mtDNA lineage below 20% in frequency to reach fixation, when conditioning for that (Supplementary Table 8). This was computed using the following formula from Kimura and Otha[53] where N is the Eurasian Neandertal *Ne* and *p* is the proportion of the introgressing mtDNA lineage: $T(p) = -2N(1-p)\ln(1-p)/p$ that was readapted from 4N for autosomes to 2N for mtDNA in Pagani *et al*.[35]. Generations were converted into years assuming a generation time of 30years.

References

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

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

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

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

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

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

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

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

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

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

11 Rieux, A. *et al.* Improved calibration of the human mitochondrial clock using ancient genomes. *Mol Biol Evol* **31**, 2780-2792, doi:10.1093/molbev/msu222 (2014).

12 Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sanchez, E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet* **16**, 359-371, doi:10.1038/nrg3936 (2015).

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

14 Kuhlwilm, M. *et al.* Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530**, 429-433, doi:10.1038/nature16544 (2016).

15 Kunter, M. & Wahl, J. Das Femurfragment eines Neandertalers aus der Stadelhöhle des Hohlensteins im Lonetal. *Fundberichte aus Baden-Württemberg* **17**, 111-124 (1992).

16 Wetzel, R. Der Hohlestein im Lonetal. Dokumente alteuropäischer Kulturen vom Eiszeitalter bis zur Völkerwanderung. *Mitteilungen des Vereins für Naturwissenschaft und Mathematik in Ulm (Donau)* **27**, 21-75 (1961).

17 Street, M., Terberger, T. & Orschiedt, J. A critical review of the German Paleolithic hominin record. *J Hum Evol* **51**, 551-579, doi:10.1016/j.jhevol.2006.04.014 (2006).
18 Wißing, C. *et al.* Isotopic evidence for dietary ecology of late Neandertals in North-Western Europe. *Quaternary International* **411**, 327-345 (2016).
19 Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **23**, 3843-3854, doi:10.1002/rcm.4316 (2009).
20 Bocherens, H., Drucker, D. G. & Madelaine, S. Evidence for a (15)N positive excursion in terrestrial foodwebs at the Middle to Upper Palaeolithic transition in south-western France: Implications for early modern human palaeodiet and palaeoenvironment. *J Hum Evol* **69**, 31-43, doi:10.1016/j.jhevol.2013.12.015 (2014).
21 Bocherens, H. in *The Scladina I-4A Juvenile Neandertal* Vol. ERAUL (ed Eds. M. Toussaint; D. Bonjean) Ch. 351-362, (2014).
22 Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A* **110**, 15758-15763, doi:10.1073/pnas.1314445110 (2013).
23 Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* **2010**, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).
24 Maricic, T., Whitten, M. & Paabo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004, doi:10.1371/journal.pone.0014004 (2010).
25 Behar, D. M. *et al.* A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* **90**, 675-684, doi:10.1016/j.ajhg.2012.03.002 (2012).
26 Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321, doi:10.1126/science.1174462 (2009).
27 Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147, doi:10.1038/13779 (1999).
28 Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol* **16**, 224, doi:10.1186/s13059-015-0776-0 (2015).
29 Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131, doi:10.1371/journal.pone.0034131 (2012).
30 Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426, doi:10.1016/j.cell.2008.06.021 (2008).
31 Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A* **104**, 14616-14621, doi:10.1073/pnas.0704665104 (2007).

32 Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr Biol* **26**, 827-833, doi:10.1016/j.cub.2016.01.037 (2016).
33 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214, doi:10.1186/1471-2148-7-214 (2007).
34 Baele, G., Lemey, P. & Vansteelandt, S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinformatics* **14**, 85, doi:10.1186/1471-2105-14-85 (2013).
35 Pagani, L. *et al.* Genomic analyses inform on migration events during the peopling of Eurasia. *Nature*, doi:10.1038/nature19792 (2016).
36 Dalen, L. *et al.* Partial genetic turnover in neandertals: continuity in the East and population replacement in the West. *Mol Biol Evol* **29**, 1893-1897, doi:10.1093/molbev/mss074 (2012).
37 Preusser, F. Towards a chronology of the Late Pleistocene in the northern Alpine Foreland. *Boreas* **33**, 195-210, doi:10.1080/03009480410001271 (2004).
38 Adler, D. S. *et al.* Early Levallois technology and the Lower to Middle Paleolithic transition in the Southern Caucasus. *Science* **345**, 1609-1613, doi:10.1126/science.1256484 (2014).
39 Lahr, M. M. & Foley, R. A. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am J Phys Anthropol* **Suppl 27**, 137-176 (1998).
40 Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**, e3, doi:10.1093/nar/gkr771 (2012).
41 Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**, 60, doi:10.1186/s13059-016-0918-z (2016).
42 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
43 Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682-1684, doi:10.1093/bioinformatics/btt193 (2013).
44 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).
45 van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**, E386-394, doi:10.1002/humu.20921 (2009).
46 Brown, S. *et al.* Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci Rep* **6**, 23559, doi:10.1038/srep23559 (2016).
47 Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713, doi:10.1038/35047064 (2000).

48 Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McLnerney, J. O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* **6**, 29, doi:10.1186/1471-2148-6-29 (2006).
49 Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).
50 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
51 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729, doi:10.1093/molbev/mst197 (2013).
52 Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452, doi:10.1093/bioinformatics/btp187 (2009).
53 Kimura, M. & Ohta, T. The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics* **61**, 763-771 (1969).

Acknowledgements
We are grateful for comments from Ilan Gronau, Adam Powell, Wolfgang Haak, Maria Spyrou and Alvise Barbieri. We thank Alexander Peltzer and Gabriel Renaud for support in running EAGER and *schmutzi*, respectively, Mannis van Oven for access to the updated version of the RNRS reference and Annette Günzel for graphical support. C.P. and J.K. were supported by the Baden Württemberg Foundation and the Max Planck Society.

Author contributions
K.W., N.J.C. and C.J.K. provided archeological material and related information. C.P., C.W. and K.K. performed laboratory work. C.W. and H.B. analyzed isotopic data. C.P., L.P., L.V.H. and J.K. analyzed genetic data. C.P., L.P. and J.K. wrote the manuscript with input from all co-authors.

Materials & correspondence
Correspondence: C.P. (posth@shh.mpg.de) and J.K. (krause@shh.mpg.de)

Additional information
Accession code: the HST mtDNA consensus sequence reported in this paper is available in GenBank: ### (released upon publication).
Competing financial interests: the authors declare no competing financial interests.

Tables

Table 1: **Divergence times and molecular ages estimated in BEAST.** Reported values derive from the skyline tree prior and strict molecular clock model that fits best the data (Methods section).

Mitochondrial lineages	Mean value	95% HPD Interval
Modern humans - Neandertals	412,930	360,230 - 467,720
HST – Altai branch Neandertals	267,770	218,980 - 316,080
Altai - rest of Altai branch Neandertals	160,480	125,410 - 198,800
San - rest of modern humans	146,730	123,650 - 169,520
Altai age	130,010	171,600 - 88,010
HST age	123,800	182,560 - 62,013
Mezmaiskaya 1 age	89,075	126,700 - 51,648
Denisova 11 age	88,244	113,760 - 63,840
Okladnikov 2 age	81,446	109,290 - 56,213
Vindija 33.17 age	48,809	57,157 - 40,532
Vindija 33.19 age	43,939	51,029 - 35,336
Vindija 33.25 age	42,996	52,305 - 34,450
Goyet Q374a_1 age	40,867	46,942 - 32,697
Goyet Q305_7 age	40,832	47,057 - 33,134

Figures

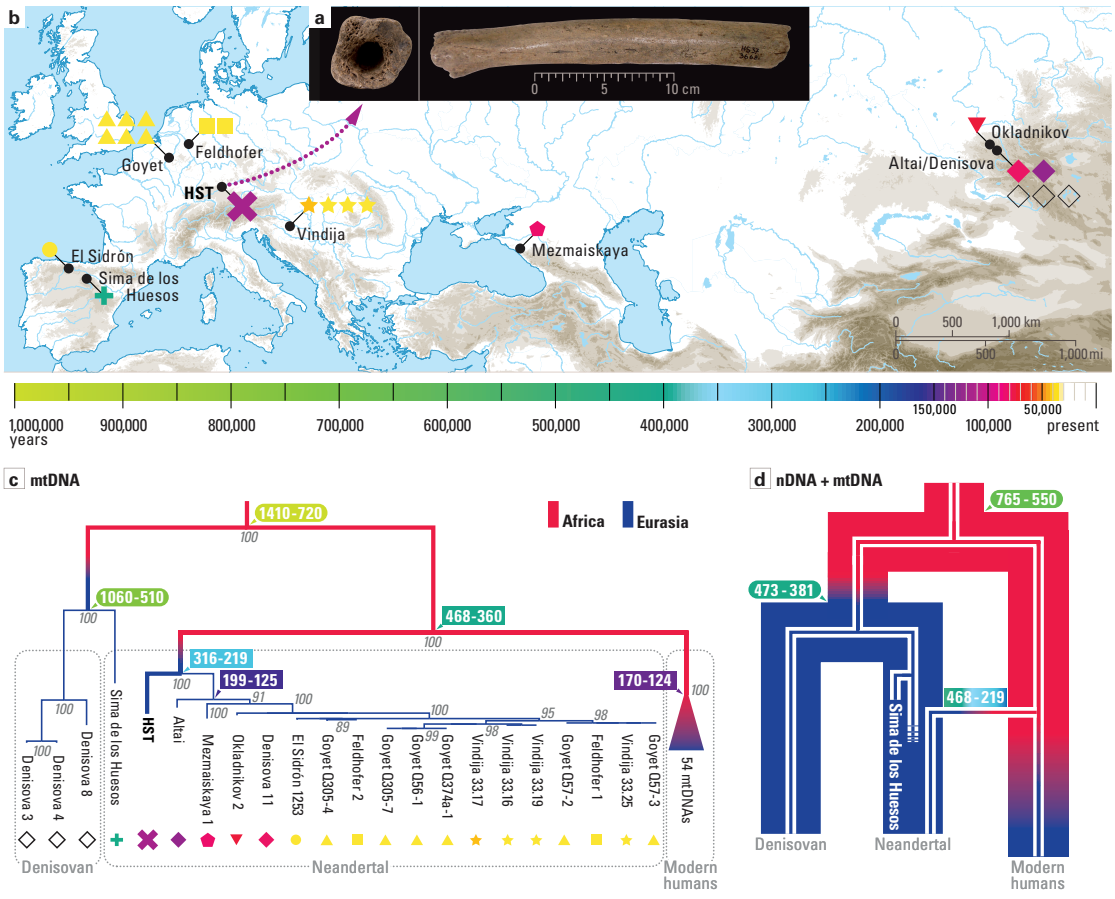


Figure 1: **Archaic and modern humans mtDNA and nDNA evolutionary scenarios.** a) Pictures of the HST femur, b) Map of archaeological sites where complete mtDNA from archaic humans were reconstructed, c) Maximum parsimony tree of 54 modern human (collapsed), 18 Neandertal, three Denisovan and one Sima de los Huesos mtDNAs build with coding region only and 98% partial deletion. Gray node numbers refer to bootstrap support after 1000 iterations. Tree rooted with a chimpanzee mtDNA (not shown). d) Schematic comparison of the nDNA (wide lines) with the mtDNA (thin lines) phylogenies of Neandertals, Denisovans and modern humans. In Fig. 1c and 1d, color legend for individual symbols and node numbers is illustrated in the horizontal time line. Node numbers in rectangular boxes are divergence times estimated in this study (Table 1) while in oval boxes are dates estimated in Prüfer *et al.*[5] and Meyer *et al.*[3] in thousand years before present. Red and blue tree branches represent supposed African and Eurasian distribution, respectively.

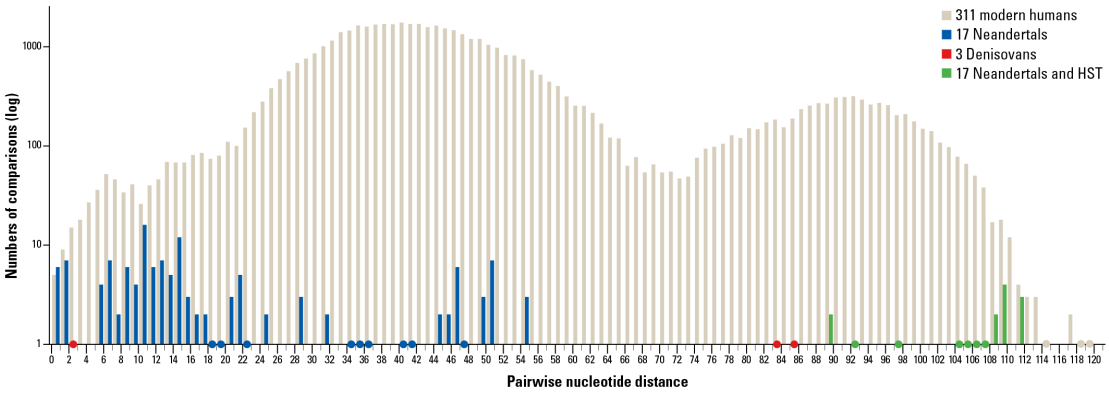
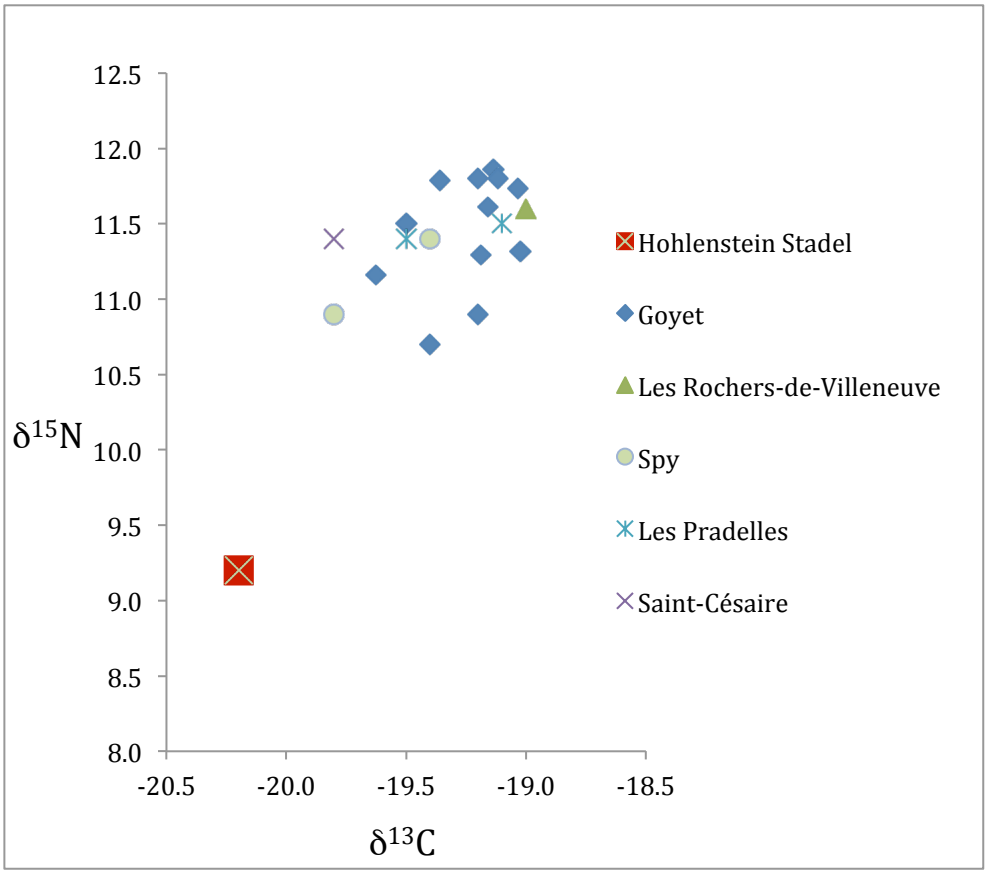
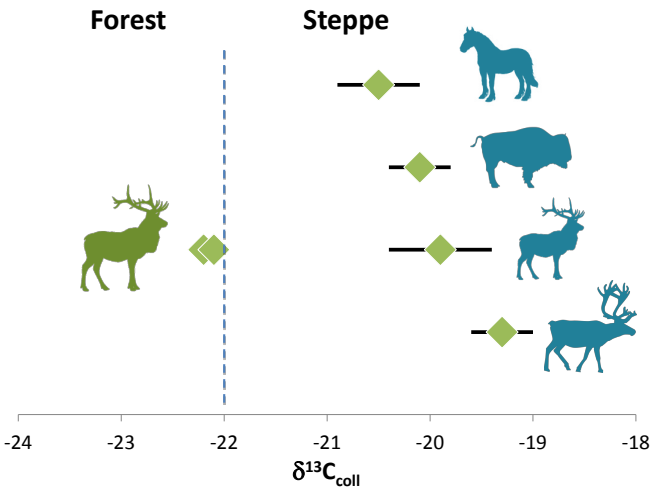


Figure 2: **Archaic and modern humans mtDNA diversity.** The pairwise nucleotide distance over its frequency (in logarithmic scale) is measured among 311 worldwide modern human, 17 Neandertal, 3 Denisovan and 18 Neandertal (including HST) mtDNAs. Points on the X-axis represent one sequence pair comparison.

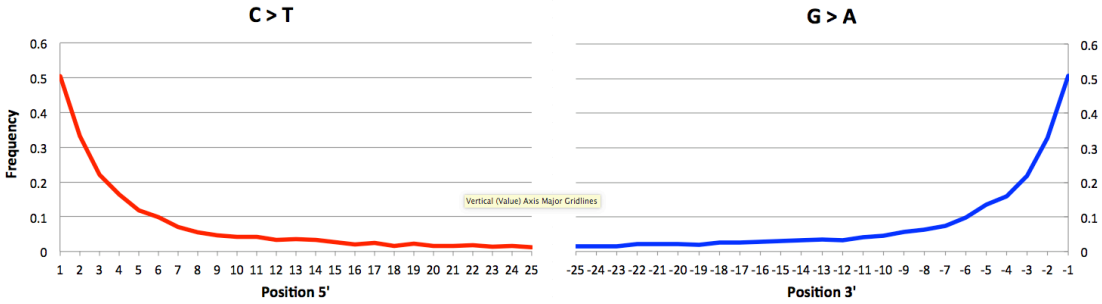
Supplementary Figures



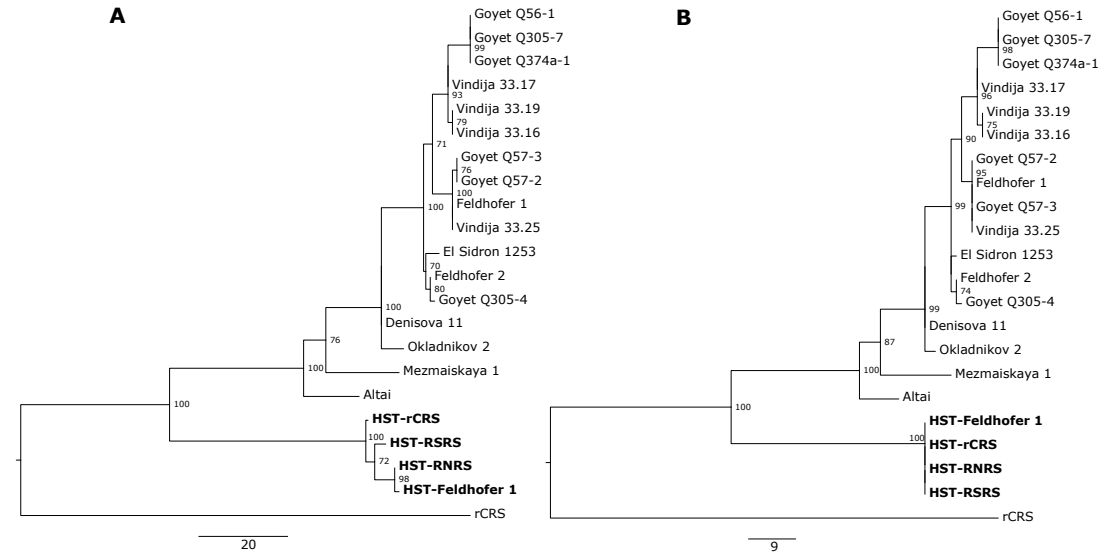
Supplementary Figure 1. Stable isotopic composition ($\delta^{13}C$ and $\delta^{15}N$) in collagen extracted from the HST femur (Supplementary Table 1) and other late Neandertals[1-3].



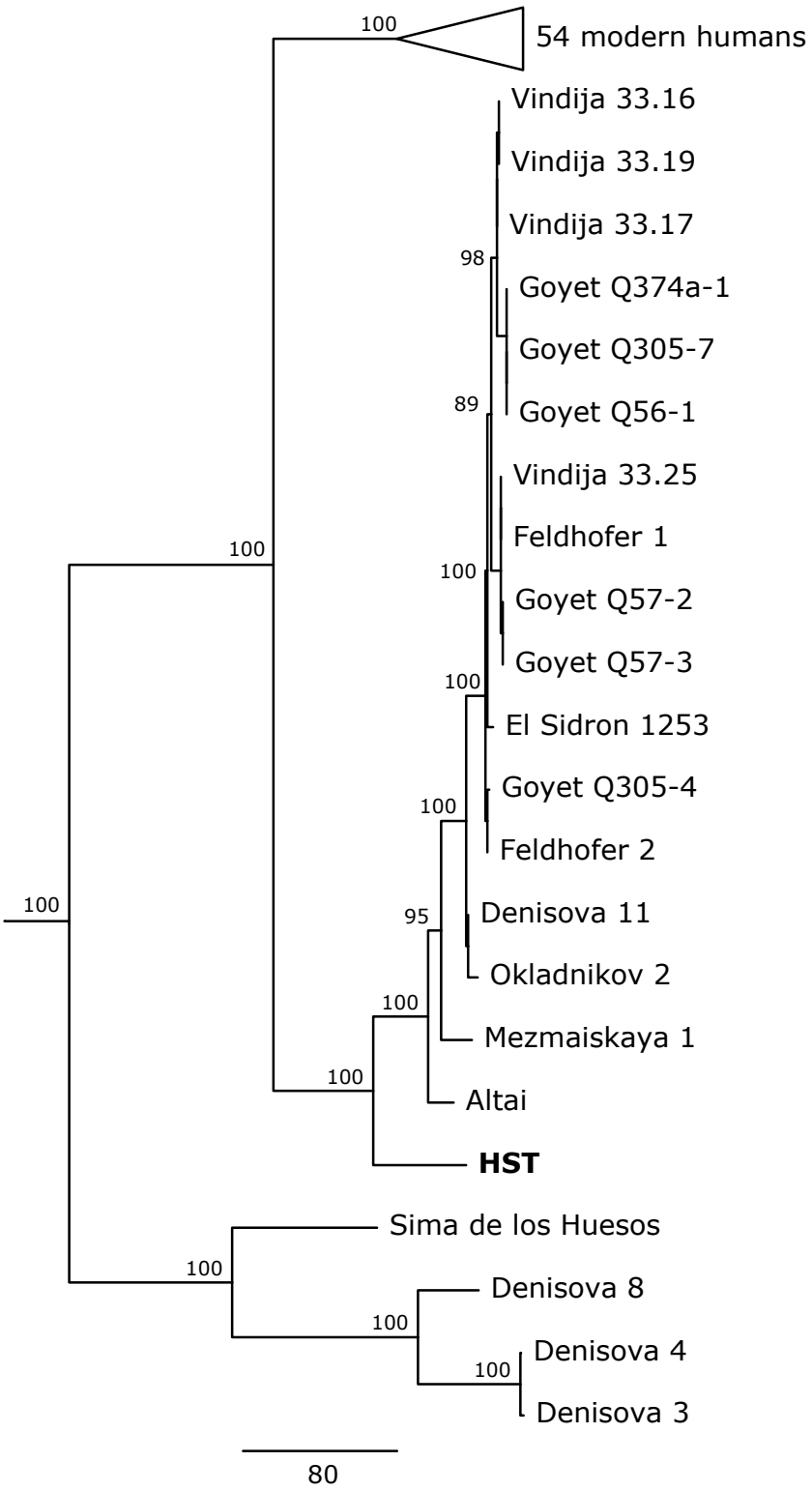
Supplementary Figure 2. Collagen $\delta^{13}C$ range of steppe herbivores (in blue from the top right: horse, bison, red deer and reindeer) compared to the $\delta^{13}C$ values of two red deer (in green on the left) from the same stratigraphic unit of HST suggesting for the latter a light forested environment (Supplementary Table 1).



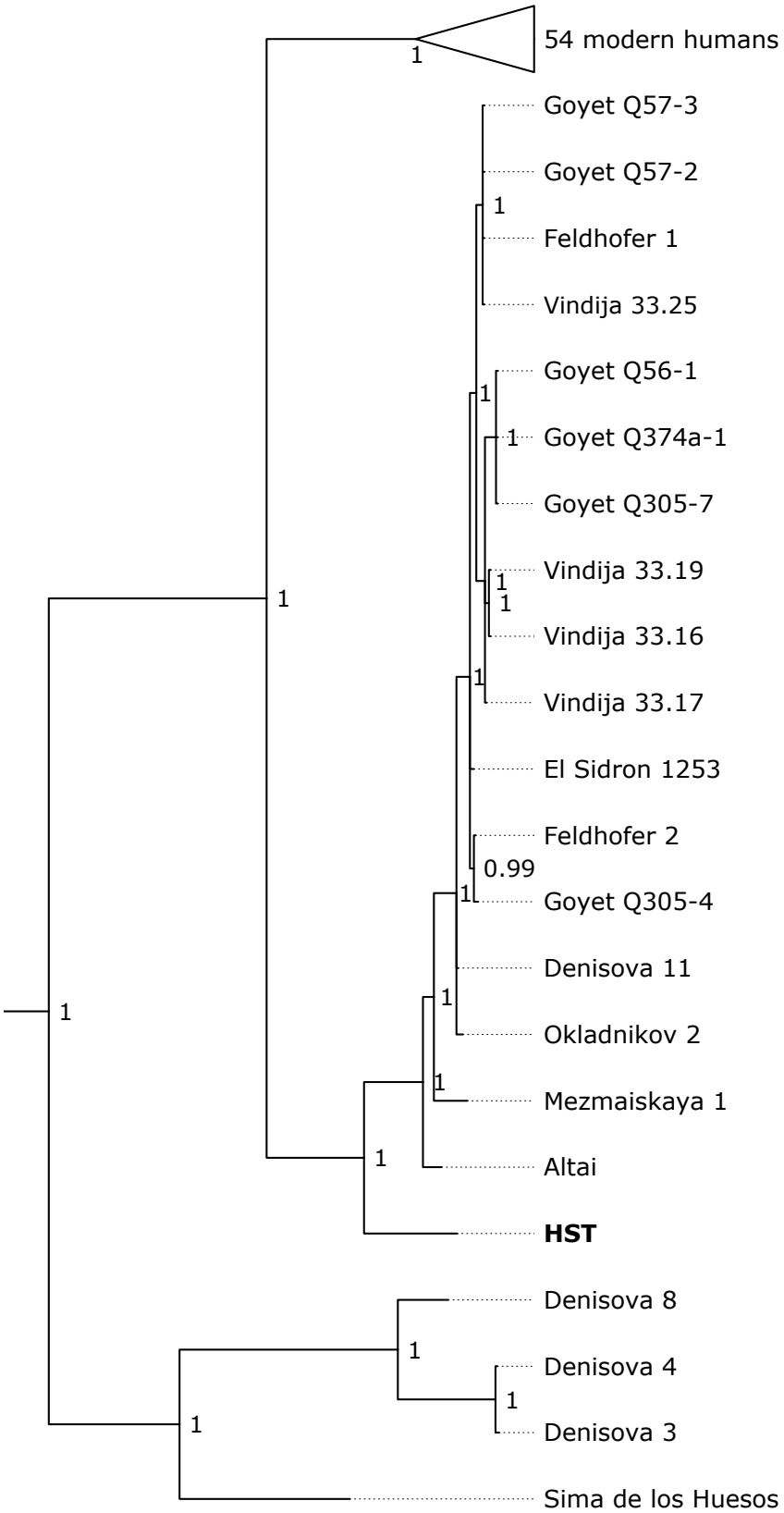
Supplementary Figure 3. Deamination patterns C to T at the 5' molecule end (in red) G to A at the 3' molecule end (in blue) in the mtDNA fragments mapped against the RNRS reference.



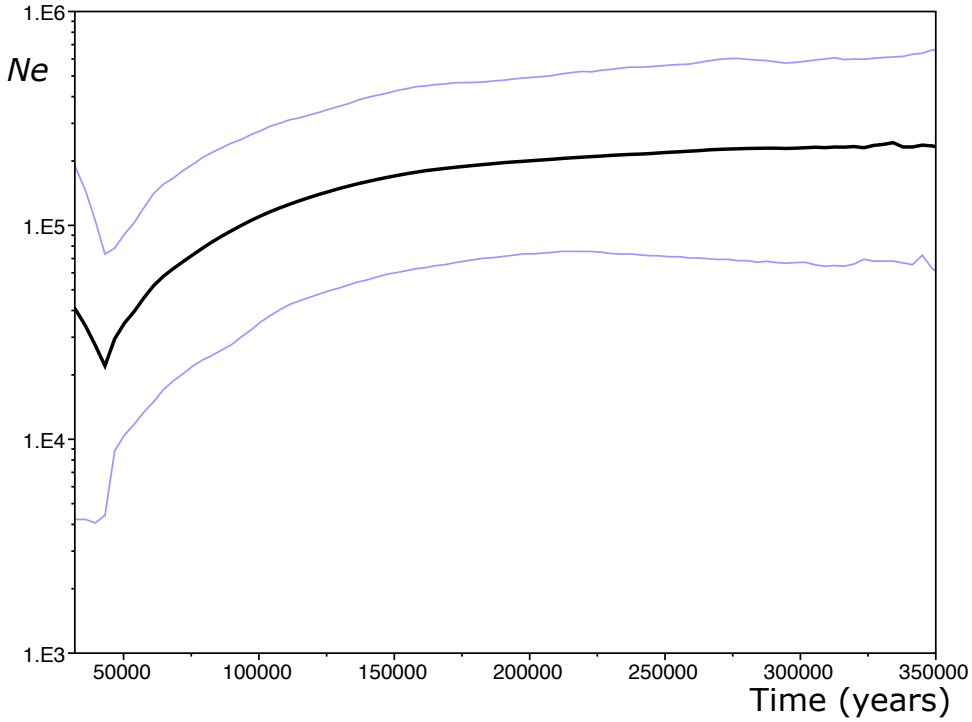
Supplementary Figure 4. Reference mapping bias. A) Maximum parsimony tree with complete mtDNA sequences using 1000 bootstrap iterations. HST consensus mtDNA show different branch lengths when mapping to four different references. B) Maximum parsimony tree with mtDNA coding region (without D-loop) and 1000 iterations as bootstrap support. Using only the HST coding region reduces reference mapping bias.



Supplementary Figure 5. Maximum parsimony tree built in MEGA6 with the complete mtDNA sequences (including D-loop) of: HST, 54 modern human, 17 Neandertals, three Denisovans and Sima de los Huesos. The tree was tested with 1000 bootstrap iterations and 97% partial deletion and rooted with a chimpanzee mtDNA (not shown).



Supplementary Figure 6. Bayesian tree built in MrBayes with the coding mtDNA region of: HST, 54 modern human, 17 Neandertals, three Denisovans and Sima de los Huesos. The tree was rooted with a chimpanzee mtDNA (not shown) and built including unassigned positions.



Supplementary Figure 7. Skyline plot depicting the Neandertal mtDNA effective population size (*Ne* in logarithmic scale) through time (from 350 ka to 32 ka). The black line represents the mean value while the purple lines the 95% HPD interval.

Supplementary Figure 8. Maximum parsimony tree built in MEGA6 with the HVRI regions of: HST, 20 Neandertals, three Denisovans and rCRS. The tree was tested with 1000 bootstrap iterations and complete deletion. The accession numbers of the HVRI published sequences are reported in graphs next to individual name.

Supplementary Tables

Supplementary Table 1. Summary isotopic and dating results

Sample	Species	Element	%N-bone	%C-bone	%Sbone	Amount collagen (mg)	Yield (mg/g)	%Ccoll	%Ncoll	C/N	d13C	d15N	ZooMS	14C date
HST17	Neandertal	Femur	n/a	n/a	n/a	20.0	63.8	41.5	14.8	3.3	-20.2	9.2	n/a	30570 ± 190 (GrA-43925)
HST21	Cervus elaphus	Tibia	2.1	7.4	0.1	11.8	38.9	41.5	14.7	3.3	-22.2	4.8	n/a	46975±1000 (MAMS-23208)
HST26	Cervus elaphus	Pelvis	2.1	7.5	0.1	19.3	58.8	42.4	15.0	3.3	-22.1	5.6	Red deer	49000±1000 (MAMS-23209)

Supplementary Table 2. Result statistics for each of the four mtDNA reference sequences (RNRS, Feldhofer 1, RSRS and rCRS) including mapped reads before and after duplicate removal and mapping quality filtering, percentage of fragments mapping the mtDNA reference, average coverage, average length, deamination at molecule termini and modern human contamination estimated with *contDeam*[4].

Reference	Merged Reads	Mapped reads	Duplicate removal	Mapping quality filter	Target DNA (%)	Duplication factor	Average Coverage (fold)	Insert size (bp)	3' deamination (%)	5' deamination (%)	*contDeam* (%) contamination (low - high)
RNRS	1,295,805	56,364	13,148	12,848	4.35	4.29	34.24	44.15	50.88	50.32	10 (9 - 11)
Feldhofer1	1,295,805	56,257	13,121	12,801	4.34	4.29	34.08	44.1	50.86	50.1	9.5 (8.5 - 10.5)
RSRS	1,295,805	56,604	13,130	12,790	4.37	4.31	34.05	44.11	50.82	50.26	10.5 (9.5 - 11.5)
rCRS	1,295,805	56,699	13,105	12,750	4.37	4.33	33.92	44.07	50.74	50.42	10.5 (9.5 - 11.5)

Supplementary Table 3. Result statistics of shotgun sequencing and mapping against the hg19 reference genome, duplicate removal and mapping quality filtering, percentage of human reads, average length and deamination at molecule termini.

Library ID	Raw Reads	Merged Reads	Mapped reads	Duplicate removal and quality filter	Endogenous DNA (%)	Insert size (bp)	3' deamination (%)	5' deamination (%)
GA87	554,766	242,521	1,110	713	0.46	41.12	32	42.48

Supplementary Table 4. Watterson's estimator among Neandertal mtDNAs with and without HST.

Group	# segregating sites	n (# individuals)	θw
Neandertals without HST	78	17	1.37E-03
Neandertals with HST	145	18	2.50E-03

Supplementary Table 5. Pairwise nucleotide distance calculated by the number of nucleotide differences of modern human, Denisovan, Sima del los Heusos and other Neandertal mtDNAs to the HST mtDNA.

	Nucleotide distance to HST	
MtDNA dataset	Average	Interval
311 modern humans	201	187-214
17 Neandertals	104	89-111
3 Denisovans	362	354-366
Sima de los Huesos	292	na

Supplementary Table 6. Eight Neandertal mtDNAs with radiocarbon dates used in BEAST analyses. Radiocarbon dates were calibrated using Oxcal 4.2[5] reporting 95,4% confidence intervals in years before present (rounded to the nearest decade).

Sample	Initial value	Lower value	Upper value	^{14}C date uncalibrated (lab-#)	Publication radiocarbon date	Publication mtDNA
Felhofer 1	43,710	42,670	44,750	39,900±620 (ETH-20981)	Schmitz 2002	Briggs 2009
Felhofer 2	43,265	42,190	44,340	39,240±670 (ETH-19660)	Schmitz 2002	Briggs 2009
Vindija 33.16	43,710	39,240	48,180	38,310±2,130 (U-n/a)	Serre 2004	Green 2008
El Sidron 1253	43,040	40,300	47,050	40,840±1,200 (Beta-192065) 37,300±830 (Beta-192066) 38,240±890 (Beta-192067)	Lalueza Fox 2005 (teeth and bone level III)	Briggs 2009
Goyet Q56-1	42,540	42,080	43,000	38,440 +340, -300 (GrA-46170)	Rougier 2016	Rougier 2016
Goyet Q57-2	41,210	40,620	41,800	36,590 +300, -270 (GrA-54024)	Rougier 2016	Rougier 2016
Goyet Q57-3	42,430	41,960	42,900	38,260 +350, -310 (GrA-60019)	Rougier 2016	Rougier 2016
Goyet Q305-4	44,290	43,430	45,150	40,690 +480, -400 (GrA-46176)	Rougier 2016	Rougier 2016

Supplementary Table 7. Log Marginal Likelihoods for Stepping stone (SS) and Path Sampling (PS) evaluation for dating analyses (four models compared) and skyline plot reconstruction (two models compared). For each model three runs each of 50Ma states were combined after 10% burn-in.

Dating analyses		**Log marginal Likelihood**	
Tree prior	Clock	Stepping-stone Sampling	Path Sampling
Constant	Strict	-28,641	-28,641
Constant	Relaxed	-28,642	-28,642
Skyline	Strict	-28,609	-28,610
Skyline	Relaxed	-28,611	-28,611
Skyline plot		**Log marginal Likelihood**	
Skyline	Strict	-22,276	-22,277
Skyline	Relaxed	-22,273	-22,286

Supplementary Table 8. Time to fixation (in 1000 years, with 1gen=30y) for various mtDNA *Ne* and introgressing proportions.

Introgressing mtDNA proportion	**Neandertal *Ne* (mtDNA)**									
	10	**50**	**100**	**250**	**500**	**1000**	**2500**	**5000**	**7500**	**10000**
1×10^{-5}	1	3	6	15	30	60	150	300	450	600
1×10^{-4}	1	3	6	15	30	60	150	300	450	600
1×10^{-3}	1	3	6	15	30	60	150	300	450	600
5×10^{-3}	1	3	6	15	30	60	150	299	449	598
0.01	1	3	6	15	30	60	149	298	448	597
0.02	1	3	6	15	30	59	148	297	445	594
0.05	1	3	6	15	29	58	146	292	439	585
0.1	1	3	6	14	28	57	142	284	427	569
0.15	1	3	6	14	28	55	138	276	414	553
0.2	1	3	5	13	27	54	134	268	402	536

Supplementary Notes

Supplementary Note 1: Morphology and archaeology of HST

The femur diaphysis (HST) was found on August 26th 1937 during a test excavation at the entrance-zone of Hohlenstein-Stadel cave, underneath the foundation of a wall, which blocked the access to the cave since the late 16th century. A museum's inventory (Inventar Hahn) lists the object under "Probegrabung unter Mauer, 4. Hieb, No. 3668" and the find-number was written directly onto the bone after the excavation. Because of its archaic morphology the excavators suggested already in their first publications that the finding was a skeletal element of a Neandertal individual[6-8]. It was not until several decades later that the specimen was published and morphologically described in detail[9]. The presence of cut mark traces from carnivores on both ends of the long bone, as well as the absence of the bone's epiphyses, has led to the suggestion that the skeletal fragment was brought inside Hohlenstein-Stadel as a post depositional process[10].

The femur was discovered at the entrance of the cave in a horizon that was correlated with an inner layer of the cave, which was called the "Black Mousterian", as described by the archaeologists during excavations in the 1930s. Inside the cave the "Black Mousterian" was found at the lowermost of the stratigraphic sequence. During recent excavations on the forecourt of the cave, a displaced layer of black sediment was discovered (layer BG)[11] and radiocarbon dated to >50,000 ^{14}C years BP (ETH-38795)[12]. This displaced horizon from the forecourt can be correlated with the layer of the "Black Mousterian" from inside the cave as well as with the horizon where the HST femur was found. In fact, layer BG was discovered only a few meters away from the spot where the hominin remain was unearthed and showed clear indication of periglacial movements of sediments from inside the cave to the forecourt. It can be supposed that the black layer containing the HST femur was also not in situ but displaced from inside the cave. A substantial amount of small mammals was discovered in the forecourt BG horizon, suggesting the presence of moderate climatic conditions that were also confirmed by the presence of a limited amount of specimens from arctic faunal species[11]. This composition of small mammals is unknown in southwestern Germany during the warm Marine Isotope Stage 3 (MIS 3), but it is also unlikely to have originated during the Eemian interglacial period (MIS 5e). Therefore it is more plausible that this layer belong to one of the moderate interstadial periods at the beginning of the last glaciation (MIS 5c or 5a)[13].

Supplementary Note 2: Isotopic results

Collagen extraction was performed at Tübingen University and followed the method from Longin[14] and described in Bocherens et al[15]. Isotopic measurements of $\delta^{15}N$ and $\delta^{13}C$ were done using an elemental analyzer NC2500 connected to a Thermo Quest Delta + XL isotopic ratio mass spectrometer. The isotopic ratios are expressed using the "δ" (delta) value as follows: $\delta^{13}C = [(^{13}C/^{12}C)_{sample}/(^{13}C/^{12}C)_{reference}-1] \times 1000‰$, $\delta^{15}N = [(^{15}N/^{14}N)_{sample}/(^{15}N/^{14}N)_{reference}-1] \times 1000‰$. The standard for $\delta^{13}C$ is the internationally defined marine carbonate V-PDB. For $\delta^{15}N$ the atmospheric nitrogen (AIR) is used. Analytical error based on laboratory standards is ±0.1‰ for $\delta^{13}C$ values and ±0.2‰ for $\delta^{15}N$. The chemical preservation of collagen is expressed through the atomic ratio of C_{coll}:N_{coll}, whose acceptable range of variation is 2.9–3.6[16],

while the nitrogen content (N_{coll}) should be above 5%[17].
The attempt to directly radiocarbon date HST femur resulted in an age of 34,130-34,880 years cal BP (GrA-43925: 30,570 ± 190 ^{14}C years). This date is inconsistent with the estimated end of the Mousterian around 40 ka[18] and with the femur's stratigraphic position (see Supplementary Note 1). Stable isotopic composition in the collagen of HST femur (HST 17) differed from that of late Neandertals[1] (Supplementary Fig. S1 and Table S1). In fact, the lower δ^{13}C and δ^{15}N values in the hominin specimen correspond to a different ecology from the one of late Neandertals in western-central Europe.
Furthermore, we analyzed two faunal remains (HST 21 and HST 26) from the same stratigraphic unit where the HST femur was discovered. Collagen from both specimens was extracted and radiocarbon dated to >49,000 ^{14}C years (HST 26) and 46,975 ± 1000 ^{14}C years (HST 21) (Table S1). In this time range, both dates can be considered beyond the limit of radiochronometric dating, where a minimal proportion of contamination with modern collagen could result in a wrongly finite date. Both specimens were initially identified as red deer on a morphological basis and ZooMS analyses[19] performed on HST 26 collagen confirmed the species assignment. Moreover, collagen of both cervid remains provided δ^{13}C isotopic values distinctively lower than individuals of the same species grazing in open steppic habitats[20] (Table S1, Supplementary Fig. S2). Isotopic evidence thus suggests that both the hominin and the two deer from Hohlenstein-Stadel Black Mousterian lived in a more forested rather than a steppic environment typical for the Late Neandertals in northwestern Europe.

Supplementary Note 3: Reference mapping bias

In order to evaluate the impact in the consensus reconstruction of the four different mapping references, we used the software MUSCLE[21] to first generate a multiple genome alignment of the four consensus with 17 Neandertal mtDNAs[22-27] and rCRS[28] as out-group. We excluded the individual Goyet Q57-1 from Rougier et al.[27] in the phylogeny and further analyses because around 2% of mtDNA positions were unassigned. We then built a phylogenetic tree with the maximum parsimony method (SPR algorithm) in MEGA6[29]. A total of 16,255 positions were considered in the phylogeny with complete deletion and 1000 replicates as bootstrap support (Supplementary Fig. S4A). Looking at the tree topology, all four consensus variants are placed on a basal mtDNA Neandertal lineage but with different branch lengths. In particular, when the two modern human mtDNA sequences (rCRS and RSRS) are used as mapping references, the resulting consensus sequences show shorter phylogenetic branches. We interpret this phenomenon as a reference bias that facilitates contaminant reads to map better than the endogenous ones, producing consensus sequences closer to the modern human references. Instead, when using Neandertal references (Feldhofer 1 and RNRS) that are phylogenetically more similar to the endogenous mtDNA we observe a greater amount of derived positions. After visual inspection (Methods section) we could confirm that all derived positions observed in the RNRS consensus are indeed of endogenous origin and thus this sequence was used for phylogenetic and mtDNA diversity analyses. We further observed that 18 of the 19 inconsistent positions across the four consensus sequences are placed in the D-loop (rCRS pos. 16023-577) where the most polymorphic regions of the mtDNA are located (HVRI and HVRII). We then removed the D-loop from the alignment and constructed an additional maximum parsimony tree with the same

parameters described above (15,345 positions) (Supplementary Fig. S4B). As expected, the reference bias highlighted previously was overcome. Thus we used the most conservative coding region consensus to perform BEAST and additional phylogenetic analyses.

Supplementary Note 4: Mutation rate mtDNA

The enlarged dataset of 18 complete Neandertal mtDNAs provide us with the opportunity to create for the first time to our knowledge a skyline plot for Neandertal population assuming panmixia. We used only the mtDNA coding region and not the whole molecule for BEAST analyses because the vast majority of unassigned nucleotides in the published Neandertal mtDNAs are located in the D-loop. This region is the most variable of the mtDNA with the highest mutation rate. Having the D-loop poorly covered but using a mutation rate for the whole molecule would effectively accelerate the mutation rate of the coding region. Therefore as molecular clock we set a fixed rate of $1.57*10^{-8}$ μ / site / year[30] calculated for the coding region of modern humans with ancient mtDNAs as calibration points. Additionally we used the eight radiocarbon Neandertal dates (Table S6) as time anchors on the Neandertal branch. When we tried to estimate the mutation rate independently with these dates as tip calibrations, BEAST runs did not converge or provided a non-realistic rate of one order of magnitude lower than observed for modern humans. Two possible reasons could explain these patterns. First, late Neandertals ages are at the limit for the radiocarbon dating method, therefore minimal contaminations with modern collagen could result in considerably younger ages, as reported for El Sidron Neandertal specimens[31]. Second, the eight radiocarbon dated Neandertals have an average age in a range of only 3,000 years (mean values from ~41 to ~44 ka) (Table S6). Therefore there might be not enough temporal depth to calibrate the molecular clock in a phylogenetic tree with divergence times in the order of several hundred thousands of years. For BEAST analyses we thus assumed that the mtDNA mutation rate of Neandertals would be similar to the one of modern humans.

Supplementary Information References

1 Wißing, C. *et al.* Isotopic evidence for dietary ecology of late Neandertals in North-Western Europe. *Quaternary International* **411**, 327-345 (2016).

2 Beauval, C., Lacrampe-Cuyaubere, F., Maureille, B. & Trinkaus, E. Direct radiocarbon dating and stable isotopes of the neandertal femur from Les Rochers-de-Villeneuve (Lussac-les-Châteaux, Vienne). *Bulletins et mémoires de la Société d'Anthropologie de Paris* **18**, 35-42 (2006).

3 Bocherens, H., Drucker, D. G., Billiou, D., Patou-Mathis, M. & Vandermeersch, B. Isotopic evidence for diet and subsistence pattern of the Saint-Cesaire I Neanderthal: review and use of a multi-source mixing model. *J Hum Evol* **49**, 71-87, doi:10.1016/j.jhevol.2005.03.003 (2005).

4 Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol* **16**, 224, doi:10.1186/s13059-015-0776-0 (2015).

5 Ramsey, C. B. & Lee, S. Recent and Planned Developments of the Program Oxcal. *Radiocarbon* **55**, 720-730 (2013).

6 Völzing, O. Die Grabungen 1937 am Hohlestein im Lonetal. *Fundberichte aus Schwaben* **NF 9**, 1-7 (1938).

7 Wetzel, R. Die Kopfbestattung und die Knochentrümmerstätte des Hohlensteins im Rahmen der Urgeschichte des Lonetals. *Verhandl. Dt. Ges. f. Rassenforschung* **9**, 193-212 (1938).

8 Wetzel, R. Der Hohlestein im Lonetal. Dokumente alteuropäischer Kulturen vom Eiszeitalter bis zur Völkerwanderung. *Mitteilungen des Vereins für Naturwissenschaft und Mathematik in Ulm (Donau)* **27**, 21-75 (1961).

9 Kunter, M. & Wahl, J. Das Femurfragment eines Neandertalers aus der Stadelhöhle des Hohlensteins im Lonetal. *Fundberichte aus Baden-Württemberg* **17**, 111-124 (1992).

10 Camarós, E., Münzel, S. C., Cueto, M., Rivals, F. & Conard, N. J. The evolution of Paleolithic hominin–carnivore interaction written in teeth: Stories from the Swabian Jura (Germany). *Journal of Archaeological Science: Reports* **6**, 798-809, doi:10.1016/j.jasrep.2015.11.010 (2016).

11 Jahnke, T. *Vor der Höhle. Zur Fundplatzgenese am Vorplatz des Hohlenstein-Stadel (Lonetal)* MA thesis thesis, Universität Tübingen, (2015).

12 Kind, C. J. *Löwenmensch und mehr. Die Ausgrabungen 2008 - 2013 in der Stadel-Höhle im Hohlenstein (Lonetal), Gemeinde Asselfingen, Alb-Donau-Kreis*. (in prep.).

13 Street, M., Terberger, T. & Orschiedt, J. A critical review of the German Paleolithic hominin record. *J Hum Evol* **51**, 551-579, doi:10.1016/j.jhevol.2006.04.014 (2006).

14 Longin, R. New method of collagen extraction for radiocarbon dating. *Nature* **230**, 241-242 (1971).

15 Bocherens, H. *et al.* Paleobiological implications of the isotopic signatures (C-13, N-15) of fossil mammal collagen in Scladina cave (Sclayn, Belgium). *Quaternary Research* **48**, 370-380, doi:DOI 10.1006/qres.1997.1927 (1997).

16 Deniro, M. J. Postmortem Preservation and Alteration of Invivo Bone-Collagen Isotope Ratios in Relation to Paleodietary Reconstruction. *Nature* **317**, 806-809, doi:DOI 10.1038/317806a0 (1985).
17 Ambrose, S. H. Preparation and Characterization of Bone and Tooth Collagen for Isotopic Analysis. *Journal of Archaeological Science* **17**, 431-451, doi:Doi 10.1016/0305-4403(90)90007-R (1990).
18 Higham, T. *et al.* The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306-309, doi:10.1038/nature13621 (2014).
19 Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J. C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **23**, 3843-3854, doi:10.1002/rcm.4316 (2009).
20 Bocherens, H., Drucker, D. G. & Madelaine, S. Evidence for a (15)N positive excursion in terrestrial foodwebs at the Middle to Upper Palaeolithic transition in south-western France: Implications for early modern human palaeodiet and palaeoenvironment. *J Hum Evol* **69**, 31-43, doi:10.1016/j.jhevol.2013.12.015 (2014).
21 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
22 Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321, doi:10.1126/science.1174462 (2009).
23 Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416-426, doi:10.1016/j.cell.2008.06.021 (2008).
24 Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc Natl Acad Sci U S A* **111**, 2229-2234, doi:10.1073/pnas.1318934111 (2014).
25 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).
26 Gansauge, M. T. & Meyer, M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. *Genome Res* **24**, 1543-1549, doi:10.1101/gr.174201.114 (2014).
27 Rougier, H. *et al.* Neandertal cannibalism and Neandertal bones used as tools in Northern Europe. *Sci Rep* **6**, 29005, doi:10.1038/srep29005 (2016).
28 Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147, doi:10.1038/13779 (1999).
29 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725-2729, doi:10.1093/molbev/mst197 (2013).
30 Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* **23**, 553-559, doi:10.1016/j.cub.2013.02.044 (2013).
31 Wood, R. E. *et al.* A New Date for the Neanderthals from El Sidron Cave (Asturias, Northern Spain)*. *Archaeometry* **55**, 148-158, doi:10.1111/j.1475-4754.2012.00671.x (2013).