Article

Current Biology

Genome of Peştera Muierii skull shows high diversity and low mutational load in pre-glacial Europe

Highlights

- Peştera Muierii woman is related to Europeans, but she is not a direct ancestor
- Reduced diversity in Europe caused by Last Glaciation, not out-of-Africa bottleneck
- Genetic load appears indifferent across 40,000 years of European history
- New DNA extraction approach recovers up to 33 times more DNA from ancient remains

Authors

Emma Svensson, Torsten Günther, Alexander Hoischen, ..., Concepción de-la-Rua, Mihai G. Netea, Mattias Jakobsson

Correspondence

mattias.jakobsson@ebc.uu.se (M.J.), torsten.guenther@ebc.uu.se (T.G.)

In brief

Svensson et al. sequence the complete genome of a woman from "Peştera Muierii," Romania, who lived 34,000 years ago. Her genome is similar to modern-day Europeans, but she is not a direct ancestor. Her genome shows high levels of diversity, revealing that much loss of diversity in non-Africans occurred after she lived rather than before her time.

Svensson et al., 2021, Current Biology *31*, 2973–2983
July 26, 2021
https://doi.org/10.1016/j.cub.2021.04.045

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

Genome of Peştera Muierii skull shows high diversity and low mutational load in pre-glacial Europe

Emma Svensson,[1,8] Torsten Günther,[1,8,*] Alexander Hoischen,[2,3,8] Montserrat Hervella,[4] Arielle R. Munters,[1] Mihai Ioana,[5] Florin Ridiche,[6] Hanna Edlund,[1] Rosanne C. van Deuren,[3] Andrei Soficaru,[7] Concepción de-la-Rua,[4] Mihai G. Netea,[3,5] and Mattias Jakobsson[1,9,*]

[1]Human Evolution, Department of Organismal Biology, Uppsala University, 752 36 Uppsala, Sweden
[2]Department of Human Genetics, Radboud University Medical Center, 6526 Nijmegen, the Netherlands
[3]Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, 6526 Nijmegen, the Netherlands
[4]Department of Genetics, Physical Anthropology and Animal Physiology, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), B° Sarriena s/n 48940 Leioa, Bizkaia, Spain
[5]Laboratory of Human Genetics, University of Medicine and Pharmacy, Craiova, Romania
[6]Oltenia Museum Craiova, Craiova, Romania
[7]"Francisc J. Rainer" Institute of Anthropology, Romanian Academy, 050474 Bucharest, Romania
[8]These authors contributed equally
[9]Lead contact
*Correspondence: torsten.guenther@ebc.uu.se (T.G.), mattias.jakobsson@ebc.uu.se (M.J.)
https://doi.org/10.1016/j.cub.2021.04.045

SUMMARY

Few complete human genomes from the European Early Upper Palaeolithic (EUP) have been sequenced. Using novel sampling and DNA extraction approaches, we sequenced the genome of a woman from "Peştera Muierii," Romania who lived ~34,000 years ago to 13.5× coverage. The genome shows similarities to modern-day Europeans, but she is not a direct ancestor. Although her cranium exhibits both modern human and Neanderthal features, the genome shows similar levels of Neanderthal admixture (~3.1%) to most EUP humans but only half compared to the ~40,000-year-old Peştera Oase 1. All EUP European hunter-gatherers display high genetic diversity, demonstrating that the severe loss of diversity occurred during and after the Last Glacial Maximum (LGM) rather than just during the out-of-Africa migration. The prevalence of genetic diseases is expected to increase with low diversity; however, pathogenic variant load was relatively constant from EUP to modern times, despite post-LGM hunter-gatherers having the lowest diversity ever observed among Europeans.

INTRODUCTION

The chronology of the Upper Palaeolithic (UP) from about 45 kya in Europe was characterized by drastic climatic change[1] and marks the time of the first appearance of anatomically modern humans (AMHs) in Europe (~45,000 BP).[2–4] It is commonly believed that the migration into Europe followed two main routes: along the Mediterranean and along the Danube fluvial corridor.[5,6] The Carpathian Mountains in present-day Romania are located close to the second suggested route, and some of the earliest AMH remains from Europe have been found in this region, cementing it as being an important region for early human occupation in Europe. Remains excavated in caves from South-West Romania, such as "Peştera cu Oase," "Peştera Muierii," or "Peştera Cioclovina Uscată," are part of the handful of European AMH individuals older than 30 kya that have been found to date (Figure 1).[7–9] In 1952, skeletal parts attributed to three AMHs were found in Peştera Muierii in current-day Romania (STAR Methods), although with some elements showing Neanderthal-like traits.[10] Even though the context is slightly unclear, the Aurignacian tool found in the vicinity make it likely that the human remains can be associated with this technological tradition.[10] During the Early Upper Palaeolithic (EUP), several shifts in material culture have been documented,[9,11,12] and genetic evidence[13–16] points to reoccurring population changes.

Placing modern human origins in Africa and the ensuing migrations out of Africa to colonize the rest of the globe is one of the success stories of molecular genetics,[17] confirming hypotheses that emerged from the fossil record. The observed difference in genetic diversity between sub-Saharan Africans and non-Africans[18] has been explained by the bottleneck associated with the migration of a relatively small group of individuals out from Africa.[18–20] The lower diversity outside of Africa has also been suggested to cause less effective purging of harmful variants, resulting in an increased "genetic load" in, e.g., Europeans compared to Africans,[21–24] although others argue that the reduction in effective population size has not been substantial enough to cause increased genetic load.[25,26]

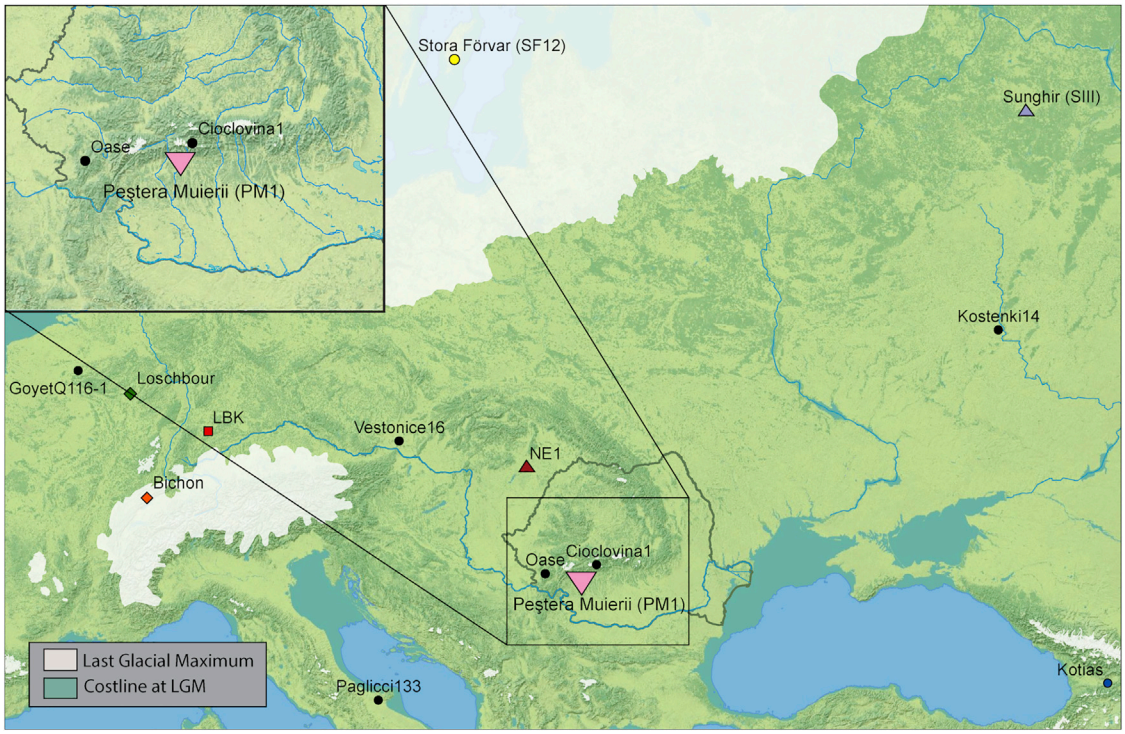

Figure 1. Map of selected EUP individuals

These individuals have been genomically investigated. The map also shows later individuals which were sequenced to a high genome coverage. See also Data S1.

We sequenced and analyzed the genome of Peştera Muierii 1 (PM1) dated to ~34 ky calBP.[27] Taken together with the data offered by the few other complete EUP genomes,[28,29] a picture of relatively high genetic diversity in EUP Europe emerges , a diversity that declines during and after the Last Glacial Maximum (LGM) (~24–19 kya) and that only recovers after the migration (and mixing) of genetically diverse Neolithic groups into Europe. We further investigated pathogen-selective pressures on immune system genes and describe the landscape of medically relevant variants in the genomes of EUP Europeans, showing deleterious variant loads that are similar to those of modern-day European populations.

RESULTS

The preservation of DNA is generally poor in specimens from the EUP, typically limiting the possible inferences from these data. To date, genomes of >1× coverage have been retrieved from AMH older than 30 ky from four sites and eight individuals: Ust'-Ishim (n = 1), Kostenki (n = 1), Yana (n = 2), and Sunghir (n = 4),[28–31] with Peştera Muierii 1 being the 9th. A few more EUP individuals have been analyzed for genome-wide SNP data, although the poor preservation of the material has limited the retrieval of data to in-solution SNP capture.[14–16]

We extracted DNA from small portions of four teeth from the Peştera Muierii 1 individual. Using state-of-the-art approaches, including silica-based DNA extraction protocols,[32,33] the DNA preservation and proportion of endogenous human DNA in Peştera Muierii 1 was at about 1% to 2% (Figure 2), similar to previous observations from material of this age and from Europe.[14,30] Based on these DNA extracts, the genome coverage that could be expected when sequencing libraries to depletion was calculated at about 0.5×. However, we used a novel DNA extraction approach, including a less invasive sampling technique coupled with improved digestion buffer. In contrast to sampling into the bone material in order to collect bone powder, a larger proportion of the bone material is protected from the heat created from the sampling equipment when a diamond cutting wheel is used to remove a piece of the bone material. Using this sampling method, only the part where the diamond cutting wheel is cutting through is affected by increased temperature from the equipment, and thereby a less invasive method is used. More recently, we have expanded this "less invasive" technique to other bone elements, including to plug-drill a small hole (~5 mm diameter) into, e.g., the petrous bone, thereby minimizing the impact on the petrous bone. This sampling approach resulted in the proportion of endogenous genomic DNA , increasing up to 33-fold in the novel DNA extracts compared to current state-of-the-art ancient DNA (aDNA) extraction techniques (Figure 2).[33–35] In these DNA extracts, the endogenous DNA content was 49% (mean across libraries; STAR Methods), allowing us to obtain a genome coverage of 13.49×. We note, however, that this improvement could be partly due to stochasticity regarding the preservation within this particular specimen.

In total, we built 42 blunt-end double-stranded libraries from 8 independent DNA extractions, which were sequenced on the

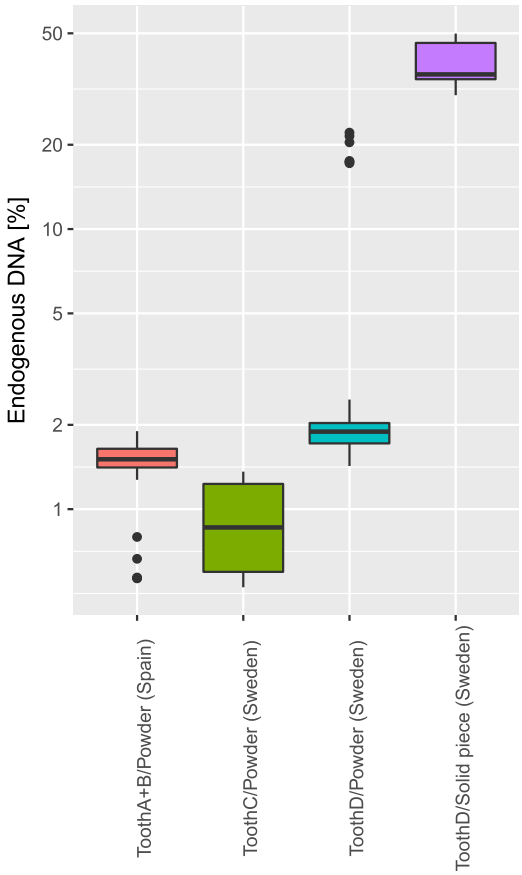

Figure 2. Proportion of endogenous human DNA

Comparison of the traditional sampling method compared to the here described DNA extraction method. See also Figure S1.

Illumina X10 and HiSeq 2500 platforms (STAR Methods) until depletion of unique sequences. The fragmentation and nucleotide mis-incorporation patterns were consistent with a pattern of DNA damage typical of aDNA, with low-moderate contamination levels. We estimate autosomal and mitochondrial contamination of Peştera Muierii 1 to be below 6%, with a point estimate of 5.7% for the nuclear genome and 1.5% after filtering potentially affected heterozygous sites (STAR Methods; Figure S1C). The sequence data of Peştera Muierii 1 confirm the previous assignment of Peştera Muierii 1 to the basal mitochondrial haplogroup U6* (STAR Methods; Figure S2A),[36] making PM1 the only occurrence of the U6 haplogroup found until now in prehistoric Europeans. Derived U6 haplotypes are primarily found among prehistoric and present-day North African populations, but not among modern-day Europeans.[36] There is ample evidence of migrations from Eurasia into North Africa during the Holocene and possibly earlier,[37,38] potentially explaining this observation.[36] f_4 statistics show no evidence of an excess of autosomal allele sharing between PM1 and Iberomaurusian Moroccan hunter-gatherers (STAR Methods; Figure S3D).[38]

Although the Peştera Muierii 1 cranium has been suggested to exhibit a mosaic of morphological features related to both AMHs and Neanderthals,[9,27] the genome shows similar levels of Neanderthal admixture (point estimate of 3.1%) compared to other EUP humans from the same time frame (Figures S3E and S3F)[14,28–30] and only modest elevation compared with modern-day Europeans (2.2%–2.7%).[39] We further find that Peştera Muierii 1 carries fewer but longer Neanderthal segments compared to post-LGM individuals and more but shorter segments than the older Ust'-Ishim individual, consistent with a single Neanderthal introgression event for these individual's ancestors. The ~40-ky-old Peştera Oase 1 mandible, also excavated in modern-day Romania, showed a different pattern, with a Neanderthal ancestor just a few generations back and 6%–9% Neanderthal admixture.[40] Hence, despite the fact that both Peştera Muierii 1 and Peştera Oase 1 have been suggested to carry archaic morphological traits,[7,9,27] they show distinctly different Neanderthal admixture levels and history. Peştera Muierii 1 carried the ancestral variants for known SNPs involved in pigmentation similar to other individuals from EUP Eurasia and likely had relatively dark skin pigmentation and brown eyes (STAR Methods).

Several genetically distinct hunter-gatherer groups were likely present in Eurasia between 45 and 30 kya,[13–16] differing in their relationship to later Stone Age groups and modern-day populations. Archaeogenomic studies have revealed a separation between European (e.g., Goyet Q116-1 from present-day Belgium and Sunghir III and Kostenki 14 from present-day Russia)[14,29,30] and East Asian hunter-gatherers (e.g., Tianyuan)[41,42] in this period, but genetic data also exist from groups that did not contribute directly to modern-day Eurasians (e.g., Ust'-Ishim and Oase 1).[28,40] Genetically, Peştera Muierii 1 falls into the genetic variation of European hunter-gatherers of similar age (Figure 3), but not with the older Ust'-Ishim or Oase 1 individuals, despite the geographic proximity to the latter. The genetic resemblance between these EUP individuals (Sunghir III, Kostenki 14, and Peştera Muierii 1) shows extensive genetic similarities across space (2,000 km separate the Romanian and Russian sites) and suggests that stratification rather follows time than geography. Modeling the relationships as admixture graph, Peştera Muierii 1 is genetically intermediate between Eastern and Western European hunter-gatherer groups and shows distant relationships to later hunter-gatherers who contributed to modern Europeans (Figure 3). PM1 shows similar affinities to all modern-day European populations (Figure S2B), but she also displays substantial private genetic drift (Figure S3H), suggesting that she represents a group that was a side branch to the ancestor of modern-day Europeans.

The availability of high-coverage ancient genomes allows us to obtain unbiased estimates of genetic diversity in these different groups and time periods. Interestingly, heterozygosity is significantly higher in pre-LGM hunter-gatherers than in post-LGM hunter-gatherers (Figure 4A; only the confidence intervals of Sunghir III and Kotias overlap). A similar loss of diversity during the LGM has been seen in mitochondrial haplogroups.[13] Although most of the reduced diversity between modern-day African and non-African populations has been attributed to bottleneck(s) associated with the out-of-African migration, the genetic diversity across time in Europe (Figure 4) illustrates the impact and importance of subsequent climatic and demographic events. First, we can conclude that it is not the migration out of Africa that solely caused the reduction in diversity; rather, it appears that the low diversity was caused by the low population

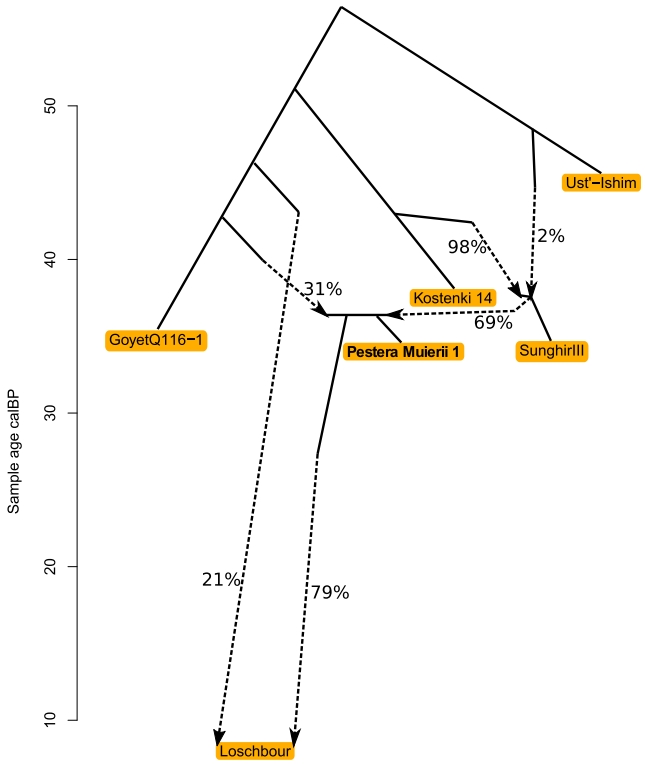

Figure 3. Admixture graph modeling of pre-Neolithic human history in Europe focusing on PM1

The positioning of individuals along the vertical axis corresponds to their radiocarbon dates; population split times and dates of admixture events are unknown; private drift not to scale. See also Figures S1–S4 and Data S1.

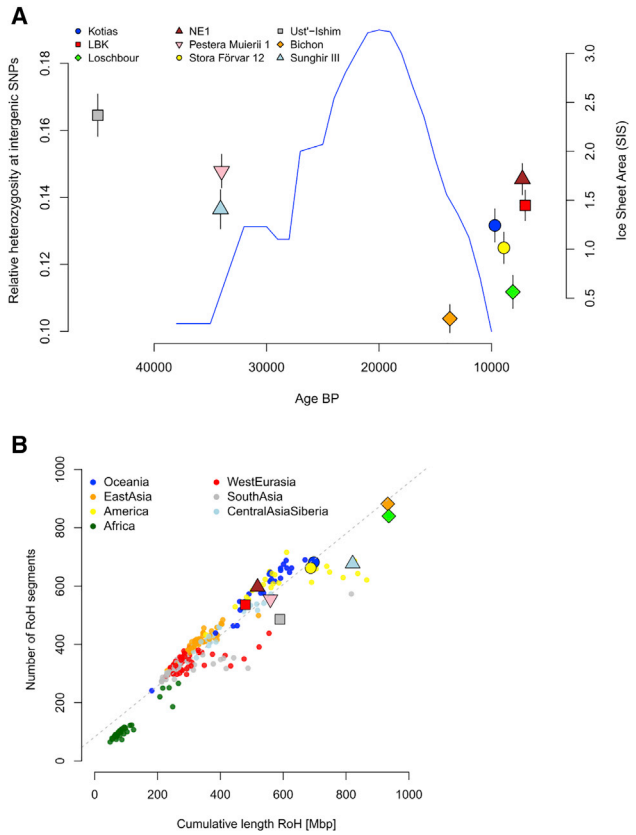

Figure 4. Diversity in the EUP

(A) Relative heterozygosity displayed for 9 ancient high-coverage samples, also showing the size of ice cover in Europe (area in million km^2) during the last glacial period.[43] Error bars indicate two standard errors estimated through a block-jacknife procedure.

(B) Runs of homozygosity for the 9 ancient individuals and modern populations.[18]

See also Figures S1 and S3 and Data S1.

density outside Africa for an extended period of time coupled with population turnovers, as seen in Europe. Second, after the LGM, Europe was likely recolonized by relatively small hunter-gatherer groups from one or very few glacial refugia, and only later large-scale migrations associated with farming practices led to an increase of genetic diversity approaching the levels before the LGM. In addition to heterozygosity, we assessed runs of homozygosity (ROHs) in ancient and modern individuals with high-coverage genomes sequenced. ROHs can be seen as evidence for identity by descent and used to reflect the recent demographic history of an individual, because the length of the ROH is affected by factors such as recent consanguinity leading to long identical tracts and long-term small population sizes indicated by many short identical tracts. ROHs have been used to show that pre-LGM hunter-gatherers probably had social behaviors avoiding mating between close relatives.[29] Consistent with per-site heterozygosity, ROHs suggest pre-LGM hunter-gatherers to be much more diverse than their post-LGM counterparts (Figure 4B). Notably, the younger pre-LGM hunter-gatherers Sunghir III and PM1 show intermediate levels of genetic diversity in both analyses. This illustrates that the loss of genetic diversity during and after the LGM was likely caused by the harsh climate conditions during those times, coupled with recolonizations and population turnovers initiated from small groups.

The complete genome sequence from Peştera Muierii 1, together with the sequence data from previously published high-coverage ancient genomes,[28,29,34,44–46] allowed us to investigate the disease-associated variant landscape among European Stone-Age individuals (from EUP to the Neolithic). We used a methodology drawn from medical genetics[47] to investigate potential pathogenic mutations and changes in their frequency.

We first explored whether there are any differences in terms of mutational burden in the coding part of the genome (exome) of ancient individuals grouped into pre-LGM hunter-gatherers, post-LGM hunter-gatherers, and Neolithic farmers, compared to each other and to modern-day healthy individuals (STAR Methods). We focused on substitution variants that were homozygous, with known dbSNP entries (rs IDs) in the coding regions, in order to avoid the potential false-positive variants in ancient genomes due to effects of *post mortem* DNA damage and slightly greater sequencing error rates in these genomes. There were no significant differences in the burden of coding protein-altering variants between the ancient genomes compared to modern-day genomes, as shown by similar numbers of non-synonymous variants, non-synonymous and synonymous variant ratio, and stop-gain variants (Data S1G; Figure 5). We used

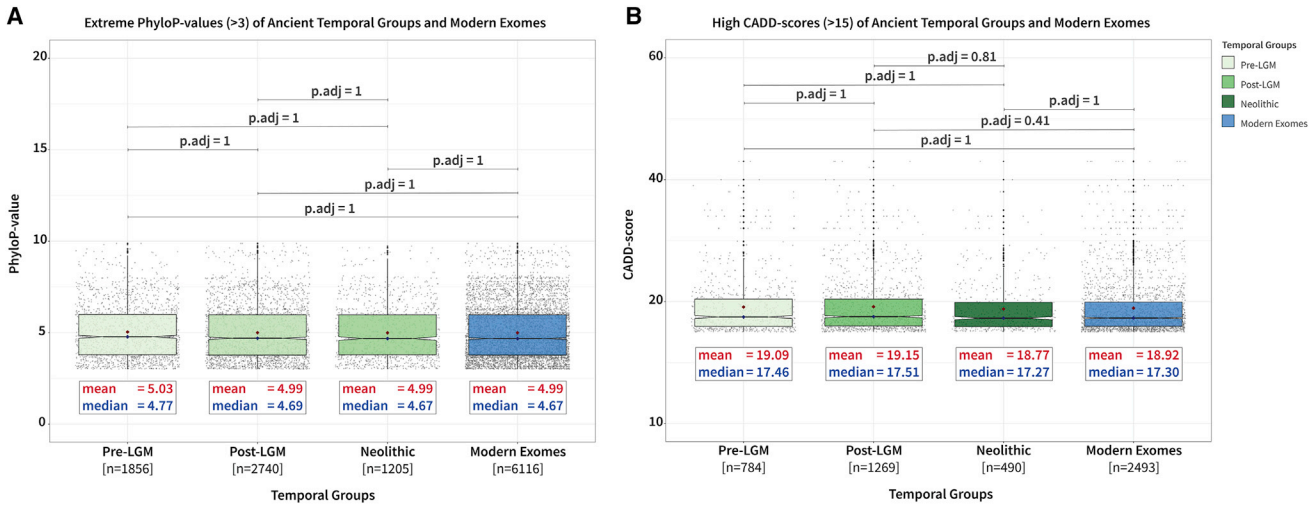

Figure 5. Comparison of predicted deleterious variants in temporal groups of ancient (pre-LGM, post-LGM, and Neolithic) and modern exomes

For (A) extreme PhyloP scores and high CADD scores (B) (see also Figure S5). P.adj., Wilcoxon test Bonferroni-adjusted p value. See also Figure S5 and Data S1.

two proxies to assess the load of possibly damaging variants among all missense variants: (1) distribution of nucleotide conservation scores (vertebrate PhyloP > 3.0)[48] and (2) distribution of CADD (combined annotation-dependent depletion) scores >15 (Figure 5).[49] For these measures, the three groups of ancient genomes (pre-LGM, post-LGM, and Neolithic) showed no significant difference from modern exomes (Figure 5). In addition, none of the individual ancient exomes showed a significant difference from modern-day exomes (Figure S5). The EUP census population size could have led to increased occurrence of possibly damaging variants due to inbreeding and bottlenecks, but the lack of difference is instead in line with the high diversity of the pre-LGM population.

Although we do not have a strong indication of a higher number of damaging variants in exomes from most time periods, the ancient individuals may carry a different set of damaging variants compared to modern-day humans. We scanned the ancient genomes and identified several variants that stood out as potentially involved in human pathology. Here, we focus on rare, likely homozygous variants in the exomes of prehistoric humans (Table 1) or variants that have a HGMD (human gene mutation database) entry (Data S1I). We identified two variants in genes that in recent literature have been suggested to be associated with severe, monogenic diseases. First, one homozygous missense variant (p.(Glu1413Lys)) was identified in *ANKRD11.* Heterozygous pathogenic variants in that gene have been described as the cause of the severe KBG syndrome;[50] patients with this disease are characterized by macrodontia, distinctive craniofacial features, short stature, skeletal anomalies, global developmental delay, seizures, and intellectual disability. However, we excluded this diagnosis based on the different inheritance pattern, the unknown effect of this missense variant, as well as the cranial characteristics of the Peştera Muierii 1 individual. Second, the exact same variant in *AIPL1* (p.(His82Tyr)) that we identified in the Bichon individual was previously described in a compound heterozygous state in a sporadic case of recessive form of Leber congenital amaurosis 4 (OMIM: 604393) and *retinitis pigmentosa*. However, the Bichon variant p.(His82Tyr) is not in a functional domain, and the same variant is described 3 times in ExAC in homozygous state (2× in EUR and 1× SA). These observations make us argue that this is very unlikely a pathogenic variant in a homozygous state. These two examples cast doubt on the idea that these variants are pathogenic and demonstrate that sequencing and in-depth analysis of ancient genomes can help draw conclusions on the pathogenicity of potential disease-causing mutations in modern-day patients: the earlier reported case of Leber congenital amaurosis 4 is thus unlikely to have been caused by this specific *AIPL1* mutation. In addition, we identified several rare, likely homozygous, non-synonymous variants in Peştera Muierii 1 that are of potential medical interest (Data S1J). Among them, we identified a homozygous nonsense variant in *IL-32* p.(Trp169*), a gene important in carcinogenesis (STAR Methods).[51]

Assessment of ancient genomes can also inform on the shaping of immune responses by different pathogen burdens during history. Cytokines are crucial immune mediators of host defense against pathogens. We assessed the presence of five gene polymorphisms known to be strongly associated with higher cytokine production capacity (STAR Methods). Interestingly, Peştera Muierii 1 genome harbors the variants associated with a strongly increased cytokine production capacity for 4 of these 5 SNPs: heterozygous carrier for C allele of *TLR1* rs4833095; homozygous carrier of G allele for *TLR6* rs5743810; heterozygous carrier for G allele of *TLR10* rs11096957; and heterozygous carrier for T allele of *IL-10* rs1800872.[53] In addition, the Peştera Muierii 1 individual was a carrier of heterozygous *IFNG* rs2069727, associated with an average cytokine production. Overall, these data suggest that Peştera Muierii 1 individual was a high responder in terms of cytokine production capacity, although less than 4% of modern-day Europeans display this combination of high-cytokine polymorphisms. Considering the protective effects of high

Table 1. Summary of rare (<1%) likely homozygous, non-synonymous variants identified in PM1 and other ancient exomes

Sample ID	R./C.	Chr.	Genomic position (hg19)	Ref.	Var.	Reads	Var. reads	SNP ID (dbSNP144)	SNP freq. (%)	In-house freq. (%)	GnomAD freq. (%)	Protein change	c.DNA variant	PhyloP	CADD PHRED score	GATK quality score
PM1	R.	chr1	243,328,858	C	T	3	3	rs572733185	0.020	0	0.0008047	p.(Val802Met)	NM_014812.2(CEP170): c.2404G > A	0.01	0.012	92.28
PM1	R.	chr2	139,316,900	G	A	27	22	rs145753116	0.439	0.276	0.3993	p.(Ser236Asn)	NM_001001664.2(SPOPL): c.707G > A	7.843	14.71	733.77
PM1	R.	chr7	2,473,130	G	C	5	5	rs201518792	NA	0.116	0.3405	p.(Ala286Pro)	NM_001243794.1(CHST12): c.856G > C	1.25	12.7	171.9
PM1	R.	chr7	77,261,719	C	T	21	20	rs201001953	NA	0.026	0.06898	p.(Ser684Leu)	NM_002835.3(PTPN12): c.2051C > T	5.381	26.8	691.81
PM1	R.	chr10	114,925,405	C	G	15	13	rs201257936	0.998	0.032	0.2812	p.(Pro495Ala)	NM_001146274.1(TCF7L2): c.1483C > G	2.527	12.79	453.89
PM1	R.	chr11	70,279,800	G	A	26	26	rs763439626	NA	0	0.001414	p.(Val461Ile)	NM_001184740.1(CTTN): c.1381G > A	0.782	3.301	1,011.77
PM1	R.	chr16	3,119,157	G	A	4	4	rs775284592	NA	0	0.0008204	p.(Trp169[a])	NM_001308078.1(IL-32): c.506G > A	−0.231	11.8	128.03
PM1	R.	chr16	89,348,713	C	T	10	9	rs140373729	0.020	0.019	0.04843	p.(Glu1413Lys)	NM_001256182.1(ANKRD11): c.4237G > A	5.366	15.28	302.18
Bichon	R.	chr17	6,337,271	G	A	35	35	rs144822294	0.059	0.274	0.1682	p.(His82Tyr)	NM_014336.3(AIPL1): c.244C > T	0.191	7.845	1,111.77
Kotias	R.	chr16	2,296,927	G	A	67	67	rs112729404	0.658	0.735	0.6972	p.(Thr76Met)	NM_001919.3(ECI1): c.227C > T	6.033	9.763	1,719.77
Kotias[a]	C.	chrX	153,595,771	C	T	2	2	rs267606816	NA	0	0	p.(Gly288Arg)	NM_001110556.1(FLNA): c.862G > A	7.809	14.76	52.74
Loschbour	R.	chr5	78,280,974	G	A	4	4	rs201168448	0.219	0.604	1.047	p.(Ala33Val)	NM_000046.4(ARSB): c.98C > T	−0.344	9.108	52.74
Ust'-Ishim	R.	chr19	15,285,063	G	T	18	18	rs141320511	NA	0.094	0.4296	p.(Leu1518Met)	NM_000435.2(NOTCH3): c.4552C > A	5.547	24.7	651.77
Ust'-Ishim	R.	chrX	105,279,368	C	T	24	24	rs2234036	NA	0.196	0.4466	p.(Ala211Thr)	NM_000354.5(SERPINA7):c.631G > A	3.867	12.14	806.77

This highlights all variants after these filters with known dbSNP entry for PM1; for all other ancient genomes, only variants with HGMD entry are shown. See also Figure S5 and Data S1. GnomAD allele frequency was added (03/2020).[52] Chr., chromosome; freq., frequency; IL-32, interleukin-32; R., rare (<1% population frequency); ref., reference nucleotide; var., variant nucleotide.

[a]This variant was only present in 2/2 reads, i.e., did not fulfill the ≥3 reads filter

immune responses in the context of high infection burden,[53] it is likely that this genetic makeup represents an adaptive state conferring protection against pathogenic microbes.

DISCUSSION

The deep sequencing of the Peștera Muierii 1 woman enabled us to identify a surprising genetic diversity in pre-LGM populations, which brings a new understanding of the early European population of AMH. These data propose a novel paradigm, in which early AMH populations after migration out of Africa were much more diverse than previously believed, and the bottlenecks associated with loss of diversity were caused by glacial climatic periods in northern latitudes. In line with the high diversity observed in EUP genomes, the burden of damaging variants in these individuals was largely the same as in modern-day individuals. This is clearly a different pattern from the high burden of deleterious variants found in small isolated populations,[26] which may help us understand the different views of genetic load in humans.[21,22,24–26] However, these ancient high-coverage genomes represent a very small sample size, and it is unclear whether the results can be extrapolated to the entire populations living during these time periods. Finally, using novel methodologies employed in medical genomics, we mined the genomes of ancient individuals for potential pathogenic variants. We have identified several interesting rare variants with medical consequences in the EUP genomes. In the case of other variants identified, such as the *AIPL1* (p. (His82Tyr)) described in a sporadic case of Leber congenital amaurosis 4, we propose that they are unlikely to be pathogenic based on insufficient literature evidence. Additionally, one could argue that living completely blind would have been very challenging in the Palaeolithic. However, we also note that care for individuals with congenital disorders or injuries is present in the archaeological record since the Middle Pleistocene,[54–56] and if the variant was verified as causing blindness, we could add another example of early human care for an individual with a severe disorder. This example shows that analysis of ancient genomes can also help in the pathogenicity assessment of rare genetic variants in modern-day patients.

This study opens the door to a new approach to the study of ancient genomes, in which classical population genetics can be combined with medical genomics to draw conclusions about demographics and disease epidemiology. Future research challenges will extend to broaden these medical genetics observations in larger populations, to study patterns of selection in ancient populations, and to directly identify the refugia and human populations from which the hunter-gatherers resurged after the LGM.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Radiocarbon data
  - Sampling and DNA extraction
  - Library building (Uppsala protocol)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Processing of next generation sequence data
  - Authentication of data and contamination estimates
  - Kinship between PM1 and PM2
  - Definition of the mitochondrial haplotype of PM1
  - Comparative Data
  - Principal Component Analysis
  - Model-based clustering
  - Outgroup f_3 statistics
  - Archaic ancestry
  - Treemix
  - Admixture graphs
  - Diversity measures
  - Phenotypic analysis
  - Medical genetics analysis
  - Damaging burden test
  - Medical genetic analysis of ancient individuals
  - Genetic variants and antimicrobial host defense

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.04.045.

ACKNOWLEDGMENTS

The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. T.G. was supported by a grant from the Swedish Research Council Vetenskapsrådet (2017-05267). M.G.N. was supported by a Spinoza grant of the Netherlands Organisation for Scientific Research and an ERC Advanced Grant (no. 833247). M.J. was supported by an ERC Starting Grant and the Knut and Alice Wallenberg Foundation. C.R. was supported by a Basque Government Grant to Research Groups of the Basque University System (IT1138-16).

AUTHOR CONTRIBUTIONS

Conceptualization, E.S., C.d.-l.-R., M.G.N., and M.J.; data curation, E.S., T.G., and A.R.M.; formal analysis, E.S., T.G., A.H., M.H., A.R.M., M.I., and R.C.v.D.; funding acquisition, T.G., A.H., C.d.-l.-R., M.G.N., and M.J.; investigation, E.S., T.G., A.H., M.H., A.R.M., M.I., F.R., H.E., R.C.v.D., A.S., C.d.-l.-R., M.G.N., and M.J.; methodology, E.S. and H.E.; project administration, E.S., T.G., M.G.N., and M.J.; resources, M.I., F.R., and A.S.; supervision, E.S., T.G., A.H., C.d.-l.-R., M.G.N., and M.J.; validation, E.S., T.G., A.H., M.H., and R.C.v.D.; visualization, E.S., T.G., A.H., A.R.M., R.C.v.D., and M.J.; writing – original draft, E.S., T.G., A.H., M.G.N., and M.J.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 3, 2020
Revised: April 9, 2021
Accepted: April 19, 2021
Published: May 18, 2021

REFERENCES

1. Morin, E. (2008). Evidence for declines in human population densities during the early Upper Paleolithic in western Europe. Proc. Natl. Acad. Sci. USA *105*, 48–53.
2. 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.
3. Higham, T. (2011). European Middle and Upper Palaeolithic radiocarbon dates are often older than they look: problems with previous dates and some remedies. Antiquity *85*, 235–249.
4. Hublin, J.-J., Sirakov, N., Aldeias, V., Bailey, S., Bard, E., Delvigne, V., Endarova, E., Fagault, Y., Fewlass, H., Hajdinjak, M., et al. (2020). Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. Nature *581*, 299–302.
5. Mellars, P. (2004). Neanderthals and the modern human colonization of Europe. Nature *432*, 461–465.
6. Mellars, P. (2006). A new radiocarbon revolution and the dispersal of modern humans in Eurasia. Nature *439*, 931–935.
7. Harvati, K., and Roksandic, M. (2017). Paleoanthropology of the Balkans and Anatolia: Human Evolution and Its Context (Springer).
8. Soficaru, A., Petrea, C., Doboş, A., and Trinkaus, E. (2007). The human cranium from the Peştera Cioclovina Uscată, Romania: context, age, taphonomy, morphology, and paleopathology. Curr. Anthropol. *48*, 611–619.
9. Trinkaus, E., Constantin, S., and Zilhão, J. (2013). Life and Death at the Peştera cu Oase: A Setting for Modern Human Emergence in Europe (Oxford University).
10. Doboş, A., Soficaru, A., and Trinkaus, E. (2010). The Prehistory and Paleontology of the Peştera Muierii (Université de Liège).
11. Hoffecker, J.F. (2009). Out of Africa: modern human origins special feature: the spread of modern humans in Europe. Proc. Natl. Acad. Sci. USA *106*, 16040–16045.
12. Svoboda, J.A. (2007). The Gravettian on the Middle Danube. Paleo *19*, 203–220.
13. Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentin, F., Thevenet, C., Furtwängler, A., Wißing, C., et al. (2016). Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a late glacial population turnover in Europe. Curr. Biol. *26*, 827–833.
14. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.
15. Prüfer, K., Posth, C., Yu, H., Stoessel, A., Spyrou, M.A., Deviese, T., et al. (2021). A genome sequence from a modern human skull over 45,000 years old from Zlatý kůň in Czechia. Nat. Ecol. Evol. Published online April 7, 2021. https://doi.org/10.1038/s41559-021-01443-x.
16. Hajdinjak, M., Mafessoni, F., Skov, L., Vernot, B., Hübner, A., Fu, Q., et al. (2021). Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry. Nature *592*, 253–257.
17. Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., and Willerslev, E. (2017). Tracing the peopling of the world through genomics. Nature *541*, 302–310.
18. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.
19. Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.-C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., et al. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. Nature *451*, 998–1003.
20. Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA *102*, 15942–15947.
21. Aris-Brosou, S. (2019). Direct evidence of an increasing mutational load in humans. Mol. Biol. Evol. *36*, 2823–2829.
22. Berens, A.J., Cooper, T.L., and Lachance, J. (2017). The genomic health of ancient hominins. Hum. Biol. *89*, 7–19.
23. Henn, B.M., Botigué, L.R., Peischl, S., Dupanloup, I., Lipatov, M., Maples, B.K., Martin, A.R., Musharoff, S., Cann, H., Snyder, M.P., et al. (2016). Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. USA *113*, E440–E449.
24. Lohmueller, K.E., Indap, A.R., Schmidt, S., Boyko, A.R., Hernandez, R.D., Hubisz, M.J., Sninsky, J.J., White, T.J., Sunyaev, S.R., Nielsen, R., et al. (2008). Proportionally more deleterious genetic variation in European than in African populations. Nature *451*, 994–997.
25. Do, R., Balick, D., Li, H., Adzhubei, I., Sunyaev, S., and Reich, D. (2015). No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nat. Genet. *47*, 126–131.
26. Simons, Y.B., and Sella, G. (2016). The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. Curr. Opin. Genet. Dev. *41*, 150–158.
27. Soficaru, A., Dobos, A., and Trinkaus, E. (2006). Early modern humans from the Pestera Muierii, Baia de Fier, Romania. Proc. Natl. Acad. Sci. USA *103*, 17196–17201.
28. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.
29. Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. Science *358*, 659–662.
30. Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.-S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., et al. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science *346*, 1113–1118.
31. Sikora, M., Pitulko, V.V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., Renaud, G., et al. (2019). The population history of northeastern Siberia since the Pleistocene. Nature *570*, 182–188.
32. Yang, D.Y., Eng, B., Waye, J.S., Dudar, J.C., and Saunders, S.R. (1998). Technical note: improved DNA extraction from ancient bones using silica-based spin columns. Am. J. Phys. Anthropol. *105*, 539–543.
33. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc. Natl. Acad. Sci. USA *110*, 15758–15763.
34. Günther, T., Malmström, H., Svensson, E.M., Omrak, A., Sánchez-Quinto, F., Kılınç, G.M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., et al. (2018). Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. PLoS Biol. *16*, e2003703.
35. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448.
36. Hervella, M., Svensson, E.M., Alberdi, A., Günther, T., Izagirre, N., Munters, A.R., Alonso, S., Ioana, M., Ridiche, F., Soficaru, A., et al. (2016). The mitogenome of a 35,000-year-old Homo sapiens from Europe supports a Palaeolithic back-migration to Africa. Sci. Rep. *6*, 25501.
37. Henn, B.M., Botigué, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J.,

et al. (2012). Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. *8*, e1002397.

38. van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., El Hajraoui, M.A., et al. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science *360*, 548–552.
39. Petr, M., Pääbo, S., Kelso, J., and Vernot, B. (2019). Limits of long-term selection against Neandertal introgression. Proc. Natl. Acad. Sci. USA *116*, 1639–1644.
40. 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.
41. Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. Proc. Natl. Acad. Sci. USA *110*, 2223–2227.
42. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., Kelso, J., and Fu, Q. (2017). 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. Curr. Biol. *27*, 3202–3208.e9.
43. Hughes, A.L.C., Gyllencreutz, R., Lohne, Ø.S., Mangerud, J., and Svendsen, J.I. (2016). The last Eurasian ice sheets–a chronological database and time-slice reconstruction, DATED-1. Boreas *45*, 1–45.
44. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.
45. 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.
46. 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.
47. Lelieveld, S.H., Reijnders, M.R.F., Pfundt, R., Yntema, H.G., Kamsteeg, E.-J., de Vries, P., de Vries, B.B.A., Willemsen, M.H., Kleefstra, T., Löhner, K., et al. (2016). Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat. Neurosci. *19*, 1194–1196.
48. Siepel, A., Pollard, K.S., and Haussler, D. (2006). New methods for detecting lineage-specific selection. In Research in Computational Molecular Biology Lecture Notes in Computer Science, A. Apostolico, C. Guerra, S. Istrail, P.A. Pevzner, and M. Waterman, eds. (Springer Berlin Heidelberg), pp. 190–205.
49. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47* (D1), D886–D894.
50. Ockeloen, C.W., Willemsen, M.H., de Munnik, S., van Bon, B.W.M., de Leeuw, N., Verrips, A., Kant, S.G., Jones, E.A., Brunner, H.G., van Loon, R.L.E., et al. (2015). Further delineation of the KBG syndrome phenotype caused by ANKRD11 aberrations. Eur. J. Hum. Genet. *23*, 1176–1185.
51. Joosten, L.A.B., Heinhuis, B., Netea, M.G., and Dinarello, C.A. (2013). Novel insights into the biology of interleukin-32. Cell. Mol. Life Sci. *70*, 3883–3892.
52. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. https://doi.org/10.1101/531210.
53. Netea, M.G., Wijmenga, C., and O'Neill, L.A.J. (2012). Genetic variation in Toll-like receptors and disease susceptibility. Nat. Immunol. *13*, 535–542.
54. Gracia, A., Arsuaga, J.L., Martínez, I., Lorenzo, C., Carretero, J.M., Bermúdez de Castro, J.M., and Carbonell, E. (2009). Craniosynostosis in the Middle Pleistocene human Cranium 14 from the Sima de los Huesos, Atapuerca, Spain. Proc. Natl. Acad. Sci. USA *106*, 6573–6578.
55. Milella, M. (2017). Subadult mortality among hunter-gatherers: implications for the reconstruction of care during prehistory. In New Developments in the Bioarchaeology of Care: Further Case Studies and Expanded Theory Bioarchaeology and Social Theory, L. Tilley, and A.A. Schrenk, eds. (Springer International), pp. 289–300.
56. Trinkaus, E., and Zimmerman, M.R. (1982). Trauma among the Shanidar Neandertals. Am. J. Phys. Anthropol. *57*, 61–76.
57. Kircher, M. (2012). Analysis of high-throughput ancient DNA sequencing data. Methods Mol. Biol. *840*, 197–228.
58. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
59. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
60. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. J. Archaeol. Sci. *40*, 4477–4482.
61. Schmidt, S. (2009). (GATK) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. In Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning 20, pp. 254–260.
62. Broad Institute (2016). Picard tools. https://broadinstitute.github.io/picard/.
63. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
64. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am. J. Hum. Genet. *91*, 839–848.
65. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.
66. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS ONE *13*, e0195491.
67. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772.
68. Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *10*, e1003537.
69. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. *67*, 901–904.
70. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
71. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
72. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

73. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

74. Behr, A.A., Liu, K.Z., Liu-Fang, G., Nakka, P., and Ramachandran, S. (2016). pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics *32*, 2817–2823.

75. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.

76. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w^{1118}; iso-2; iso-3. Fly (Austin) *6*, 80–92.

77. Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., Sijen, T., Knijff, P., Liu, F., Branicki, W., et al. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci. Int. Genet. *35*, 123–135.

78. Ramsey, C.B. (2005). OxCal Program v3.10. https://c14.arch.ox.ac.uk/oxcal.html.

79. Chirica, V., Borziac, I., and Chetraru, N. (1996). Gisements du Paléolithique Supérieur Ancien entre le Dniestr et la Tissa (Ed. Helios).

80. Păunescu, A. (1998). Paleoliticul şi Mezoliticul de pe Teritoriul Moldovei Cuprins între Siret şi Prut: Studiu Monografic (Satya Sai).

81. Păunescu, A. (1999). Paleoliticul şi Epipaleoliticul de pe Teritoriul Moldovei Cuprins între Carpaţi şi Siret: Studiu Monografic. Paleoliticul şi Mezoliticul de pe Teritoriul Moldovei Cuprins între Carpaţi şi Prut (Satya Sai).

82. Păunescu, A. (2000). Paleoliticul şi Mezoliticul din Spaţiul Cuprins între Carpaţi şi Dunăre: Studiu Monografic (Editura AGIR).

83. Păunescu, A. (2001). Paleoliticul şi Mezoliticul din Spaţiul Transilvan: Studiu Monografic (Editura Agir).

84. Cârciumaru, M. (1999). Le Paléolithique en Roumanie (Editions Jérôme Millon).

85. Chirica, V. (2001). Gisements Paleolithiques de Mitoc: le Paleolithique Superieur de Roumanie a la Lumiere des Decouvertes de Mitoc (Helios).

86. Borziac, I., Chirica, V., and Văleanu, M.C. (2006). Culture et Sociétés Pendant le Paléolithique Supérieur à Travers l'Espace Carpato-Dniestréen (Valeanu Madalin Cornel).

87. Dobrescu, R. (2008). Aurignacianul din Transilvania (Renaissance Bucharest).

88. Gibori, M. (1976). Les Civilisations du Paleolithique Moyen entre les Alpes et l'Oural (Akademiai Kiado).

89. Hahn, J. (1977). Aurignacien, das ältere Jungpaläolithikum in Mittel-und Osteuropa (Böhlau).

90. Allsworth-Jones, P. (1986). The Szeletian and the Transition from Middle to Upper Palaeolithic in Central Europe, First Edition (Oxford University).

91. Kozłowski, J.K. (2004). Early Upper Paleolithic backed blade industries in central and eastern Europe. In The Early Upper Paleolithic beyond Western Europe, P.J. Brantingham, S.L. Kuhn, and K.W. Kerry, eds. (University of California), pp. 14–29.

92. Noiret, P. (2004). Le Paléolithique supérieur de la Moldavie. Essai de synthèse d'une évolution multi-culturelle. PhD thesis (Université de Liège).

93. Noiret, P. (2005). The Aurignacian in Eastern Europe. Anadolu *29*, 39–56.

94. Sitlivy, V., and Zieba, A. (2006). Eastern and Central Europe before 30 kyr BP: Mousterian, Levallois & Blade industries. In Kabazi II: The 70000 Years since the Last Interglacial, V. Chabai, J. Richter, and T. Uthmeier, eds. (National Academy of Sciences Ukraine), pp. 361–419.

95. Douka, K., and Higham, T. (2017). The chronological factor in understanding the Middle and Upper Paleolithic of Eurasia. Curr. Anthropol. *58*, S480–S490.

96. Greenbaum, G., Friesem, D.E., Hovers, E., Feldman, M.W., and Kolodny, O. (2019). Was inter-population connectivity of Neanderthals and modern humans the driver of the Upper Paleolithic transition rather than its product? Quat. Sci. Rev. *217*, 316–329.

97. Trinkaus, E., Soficaru, A., Doboş, A., Constantin, S., Zilhão, J., and Richards, M. (2009). Stable isotope evidence for early modern human diet in southeastern Europe: Peştera cu Oase, Peştera Muierii and Peştera Cioclovina Uscată. Materiale Cercet. Arheolog. *5*, 5–14.

98. Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annu. Rev. Genet. *38*, 645–679.

99. Gilbert, M.T.P., Bandelt, H.-J., Hofreiter, M., and Barnes, I. (2005). Assessing ancient DNA studies. Trends Ecol. Evol. *20*, 541–544.

100. Hervella, M., Izagirre, N., Alonso, S., Fregel, R., Alonso, A., Cabrera, V.M., and de la Rúa, C. (2012). Ancient DNA from hunter-gatherer and farmer groups from Northern Spain supports a random dispersion model for the Neolithic expansion into Europe. PLoS ONE *7*, e34417.

101. Glocke, I., and Meyer, M. (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. Genome Res. *27*, 1230–1237.

102. Malmström, H., Svensson, E.M., Gilbert, M.T.P., Willerslev, E., Götherström, A., and Holmlund, G. (2007). More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. Mol. Biol. Evol. *24*, 998–1004.

103. Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodriguez-Varela, R., Sverrisdóttir, O.O., Daskalaki, E.A., Skoglund, P., Naidoo, T., Svensson, E.M., et al. (2015). Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proc. Natl. Acad. Sci. USA *112*, 11917–11922.

104. Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., and Allentoft, M.E. (2015). Improving access to endogenous DNA in ancient bones and teeth. Sci. Rep. *5*, 11184.

105. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.

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

107. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

108. Briggs, A.W., Stenzel, U., Johnson, P.L., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA *104*, 14616–14621.

109. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., and Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS ONE *7*, e34131.

110. Green, R.E., Malaspinas, A.-S., Krause, J., Briggs, A.W., Johnson, P.L.F., Uhler, C., Meyer, M., Good, J.M., Maricic, T., Stenzel, U., et al. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell *134*, 416–426.

111. Secher, B., Fregel, R., Larruga, J.M., Cabrera, V.M., Endicott, P., Pestano, J.J., and González, A.M. (2014). The history of the North African mitochondrial DNA haplogroup U6 gene flow into the African, Eurasian and American continents. BMC Evol. Biol. *14*, 109.

112. 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*, 1065–1093.

113. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015).

Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

114. Jones, E.R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P.R., Manica, A., Pinhasi, R., and Bradley, D.G. (2017). The neolithic transition in the Baltic was not driven by admixture with Early European farmers. Curr. Biol. *27*, 576–582.

115. Wang, C., Szpiech, Z.A., Degnan, J.H., Jakobsson, M., Pemberton, T.J., Hardy, J.A., Singleton, A.B., and Rosenberg, N.A. (2010). Comparing spatial maps of human population-genetic variation using Procrustes analysis. Stat. Appl. Genet. Mol. Biol. *9*, 13.

116. Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., and Jakobsson, M. (2012). Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science *336*, 466–469.

117. R Development Core Team (2016). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

118. Nychka, D., Furrer, R., Paige, J., and Sain, S. (2005). fields: Tools for Spatial Data (National Center for Atmospheric Research).

119. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

120. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.

121. Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science *358*, 655–658.

122. Kuhlwilm, M., and Boeckx, C. (2019). A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. Sci. Rep. *9*, 8463.

123. 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.

124. Crawford, N.G., Kelly, D.E., Hansen, M.E.B., Beltrame, M.H., Fan, S., Bowman, S.L., Jewett, E., Ranciaro, A., Thompson, S., Lo, Y., et al.; NISC Comparative Sequencing Program (2017). Loci associated with skin pigmentation identified in African populations. Science *358*, eaan8433.

125. Martin, A.R., Lin, M., Granka, J.M., Myrick, J.W., Liu, X., Sockell, A., Atkinson, E.G., Werely, C.J., Möller, M., Sandhu, M.S., et al. (2017). An unexpectedly complex architecture for skin pigmentation in Africans. Cell *171*, 1340–1353.e14.

126. Olalde, I., Allentoft, M.E., Sánchez-Quinto, F., Santpere, G., Chiang, C.W.K., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J.A., Rasmussen, S., Quilez, J., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature *507*, 225–228.

127. de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. *367*, 1921–1929.

128. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. *15*, 1034–1050.

129. Heinhuis, B., Plantinga, T.S., Semango, G., Küsters, B., Netea, M.G., Dinarello, C.A., Smit, J.W.A., Netea-Maier, R.T., and Joosten, L.A.B. (2016). Alternatively spliced isoforms of IL-32 differentially influence cell death pathways in cancer cell lines. Carcinogenesis *37*, 197–205.

130. Wang, Y., Yang, Y., Zhu, Y., Li, L., Chen, F., and Zhang, L. (2017). Polymorphisms and expression of IL-32: impact on genetic susceptibility and clinical outcome of lung cancer. Biomarkers *22*, 165–170.

131. Wang, Y.-M., Li, Z.-X., Tang, F.-B., Zhang, Y., Zhou, T., Zhang, L., Ma, J.-L., You, W.-C., and Pan, K.-F. (2016). Association of genetic polymorphisms of interleukins with gastric cancer and precancerous gastric lesions in a high-risk Chinese population. Tumour Biol. *37*, 2233–2242.

132. Yu, X., Zhou, B., Zhang, Z., Gao, Q., Wang, Y., Song, Y., Pu, Y., Chen, Y., Duan, R., Zhang, L., and Xi, M. (2015). Significant association between IL-32 gene polymorphisms and susceptibility to endometrial cancer in Chinese Han women. Tumour Biol. *36*, 5265–5272.

133. Plantinga, T.S., Costantini, I., Heinhuis, B., Huijbers, A., Semango, G., Kusters, B., Netea, M.G., Hermus, A.R., Smit, J.W., Dinarello, C.A., et al. (2013). A promoter polymorphism in human interleukin-32 modulates its expression and influences the risk and the outcome of epithelial cell-derived thyroid carcinoma. Carcinogenesis *34*, 1529–1535.

134. Li, J., Davidson, D., Martins Souza, C., Zhong, M.-C., Wu, N., Park, M., Muller, W.J., and Veillette, A. (2015). Loss of PTPN12 stimulates progression of ErbB2-dependent breast cancer by enhancing cell survival, migration, and epithelial-to-mesenchymal transition. Mol. Cell. Biol. *35*, 4069–4082.

135. Sturgeon, C.M., Duffy, M.J., Stenman, U.-H., Lilja, H., Brünner, N., Chan, D.W., Babaian, R., Bast, R.C., Jr., Dowell, B., Esteva, F.J., et al.; National Academy of Clinical Biochemistry (2008). National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. Clin. Chem. *54*, e11–e79.

136. Stone, E.M. (2007). Leber congenital amaurosis - a model for efficient genetic testing of heterogeneous disorders: LXIV Edward Jackson Memorial Lecture. Am. J. Ophthalmol. *144*, 791–811.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ancient individual	This study	PM1/Peştera Muierii 1
Chemicals, peptides, and recombinant proteins		
QIAGEN MinElute kit	QIAGEN	Cat#28004
Guanidine hydrochloride	Sigma-Aldrich	Cat#50933
EDTA, 0.5M, pH 8.0	Thermo Fisher Scientific	Cat#15575020
Agencourt AMPure XP	Beckman Coulter	Cat#A63882
Critical commercial assays		
High Sensitivity D1000 Reagents (Tapestation 2200)	Agilent	Cat#5067-5585
High Sensitivity D1000 Screentape (Tapestation 2200)	Agilent	Cat#5067-5584
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat#10616763
Deposited data		
Human sequence data (European nucleotide archive)	This study	ENA: PRJEB33172
Oligonucleotides		
IS1_adapter P5: A*C*A*C*TCTTTCCCTACACGACGCT CTTCCG*A*T*C*T (* = Phosphorothioate)	N/A	Biomers
IS2_adapter P7: G*T*G*A*CTGGAGTTCAGACGTGTG CTCTTCCG*A*T*C*T (* = Phosphorothioate)	N/A	Biomers
IS3 adaptor P5+P7: A*G*A*T*CGGAA*G*A*G*C (* = Phosphorothioate)	N/A	Biomers
IS4 PCR primer: AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTT	N/A	Biomers
P7 indexing primer: CAAGCAGAAGACGGCATACGAG ATnnnnnnnGTGACTGGAGTTCAGACGTGT	N/A	Biomers
IS7 amplification primer: ACACTCTTTCCCTACACGAC	N/A	Biomers
IS8 amplification primer: GTGACTGGAGTTCAGACGTGT	N/A	Biomers
Software and algorithms		
MergeReadsFastq_cc.py	Kircher[57]	https://bioinf.eva.mpg.de/fastqProcessing/
FilterUniqSAMCons_cc.py	Kircher[57]	https://bioinf.eva.mpg.de/fastqProcessing/
BWA	Li and Durbin[58]	http://bio-bwa.sourceforge.net/
samtools	Li et al.[59]	http://www.htslib.org/
Ry	Skoglund et al.[60]	https://github.com/pontussk/ry_compute
GATK	Schmidt[61]	https://gatk.broadinstitute.org/hc/en-us
Picard	Broad Institute[62]	https://broadinstitute.github.io/picard/
vcftools	Danecek et al.[63]	http://vcftools.sourceforge.net/
verifyBAMID	Jun et al.[64]	https://genome.sph.umich.edu/wiki/VerifyBamID
ANGSD	Korneliussen et al.[65]	http://www.popgen.dk/angsd/index.php/ANGSD
READ	Kuhn et al.[66]	https://bitbucket.org/tguenther/read/src/master/
jModeltest	Darriba et al.[67]	https://github.com/ddarriba/jmodeltest2
BEAST	Bouckaert et al.[68]	https://www.beast2.org/
Tracer 1	Rambaut et al.[69]	http://tree.bio.ed.ac.uk/software/tracer/
smartpca	Patterson et al.[70]	https://github.com/DreichLab/EIG/
ADMIXTURE	Alexander et al.[71]	https://dalexander.github.io/admixture/download.html
PLINK	Purcell et al.[72] and Chang et al.[73]	https://www.cog-genomics.org/plink/
pong	Behr et al.[74]	https://github.com/ramachandran-lab/pong

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AdmixTools	Patterson et al.[70]	https://github.com/DreichLab/AdmixTools
treemix	Pickrell and Pritchard[75]	https://bitbucket.org/nygcresearch/treemix/wiki/Home
SNPeff	Cingolani et al.[76]	https://pcingola.github.io/SnpEff/
Hirisplex-S	Chaitanya et al.[77]	https://hirisplex.erasmusmc.nl/
Other		
Proteinase K	Sigma-Aldrich	Cat#P6556
High Pure Viral Nucleic Acid Large Volume Kit (spin columns)	Roche	Cat#5114403001
Buffer Tango (10x)	Thermo Fisher Scientific	Cat#BY5
ATP Solution (100mM)	Thermo Fisher Scientific	Cat# R0441
T4 Polynucleotide Kinase (10U/μL)	Thermo Fisher Scientific	Cat#EK0032
T4 DNA polymerase (5 U/μL)	Thermo Fisher Scientific	Cat#EP0061
PEG-4000 (50%)	Thermo Fisher Scientific	Cat#EL0012
T4 DNA Ligase (5 U/μL)	Thermo Fisher Scientific	Cat#EL0012
Bst DNA polymerase, large fragment	NEB	Cat# M0275L
Maxima SYBR Green qPCR master mix (2X)	Thermo Fisher Scientific	K0252
AMpliTaqGoldDNA polymerase	Thermo Fisher Scientific	4311816
OxCal v4.3	Ramsey[78]	https://c14.arch.ox.ac.uk/oxcal/OxCal.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mattias Jakobsson (mattias.jakobsson@ebc.uu.se).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The accession number for the human sequence data reported in this paper is: European Nucleotide Archive: PRJEB33172.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

During the transition from the Middle to the Upper Palaeolithic in Europe several significant changes in both culture and climate took place. The Middle Palaeolithic Mousterian and Micoquian cultures were replaced by the Aurignacian and subsequent Gravettian cultures typical of the Early Upper Palaeolithic. How and why these changes took place between ~45 and ~40 Kyr BP is complex and still under debate.

Located at the crossroads between the Eastern, Mediterranean and Central Europe, Romania holds an obvious key geographical position in relation to all the models proposed for the origin of the European UP. Hosting some of the oldest directly dated AMH remains found thus far in Europe, Romania seems to have provided a likely early scene for various interactions between the local Neanderthal population and the dispersing Aurignacian population. In addition, the local Palaeolithic record, to which many generations of scholars have contributed, is abundant.[10,79–87] Systematic prehistoric investigations have been undertaken for more than a century, resulting in many archaeological, paleontological and anthropological discoveries, albeit unequally reported and studied. The Romanian Palaeolithic had also long attracted the scientific attention of foreign researchers aiming at incorporating this record into the general European chrono-cultural framework.[14,36,88–96]

The human bones from Peștera cu Oase were found without any lithic assemblage and it is impossible to link them to any techno-complex such Aurignacian or Gravettian. The bones were carried by water from outside of the cave along with other animal bones.[9] A similar situation is the discovery of the skull from Peștera Cioclovina Uscată, which was found in association with bears bones and three flint tools during phosphates exploitation.[8]

Peştera Muierii (The Woman's Cave, 'Muiere' is an archaism for woman.) is a multi-chambered karstic system near Baia de Fier, Gorj County, Romania. In 1952, the remains of three human individuals were found in this cave. According to the field notes the Palaeolithic bones were found in the same place, but at different depths. Initially, the archaeologist who supervised the excavation

suggested that all the bones came from the same individual. As no bone element was found more than once it is still possible that all the bones do belong to the same individual.[10,27,97]

A skull (frontal, both maxillae, left zygomatic, parietal bones – but with some missing parts – and occipital), the right side of a mandible, ten teeth, a scapula and a tibia were assigned to the individual Muierii 1. Muierii 2 is a temporal bone, and Muierii 3 is a fibula.[10] The left temporal bone (Muierii 2) cannot be articulated with the skull due to the missing areas from the left parietal and occipital. Based on the size and color (taphonomic transformation in the soil), it was considered to belong to another individual.[10] The skull is currently in Oltenia Museum, Craiova, Dolj County, comprehensive morphological description can be found in Dobos et al.[10]

METHOD DETAILS

Radiocarbon data

The dates for the zygomatic (PM1, OxA-15529) and temporal (PM2, OxA-16252) bones have been re-calibrated with OxCal 4.3 (https://c14.arch.ox.ac.uk/oxcal/OxCal.html) and the results are quite different: for OxA-15529 the determined age is 32407-31758 calBC (95.4% probability) and for OxA-16252 the determined age is 31806-30892 calBC (95.4% probability). The overlap between them is about 48 years.

Sampling and DNA extraction

The selection of the sample was made for teeth without caries or deep fissures that might extend into the pulp. Initially, two teeth were selected (Tooth A and Tooth B).[36] The processing of the samples and the DNA extraction were performed in the laboratory of the University of Medicine and Pharmacy at Craiova (Romania). The processing of the samples involved the application of a series of strict criteria detailed in Pääbo et al.[98] and Gilbert et al.[99] for the authentication of results. In our case, the extraction and preparation of the PCR was undertaken in a positive-pressure sterile chamber, physically separated from the laboratory where post-PCR processes are carried out. All the work surfaces were cleaned regularly with sodium hypochlorite and irradiated with UV light. Suitable disposable clothing was worn (lab coat, mask, gloves and cap). Contamination controls were applied in both the extraction and amplification processes. The teeth were extracted in different moments. In order to eliminate surface contamination, the teeth were subjected to a process of depurination using acids (solution of 20% acetic acid, 15% HCl and then with 70% ethanol), and the entire surface was irradiated with ultraviolet light. Each tooth was cut off with sterile jeweller saws between the crown and the root, and the pulp cavity was scraped with sterile dental tools.

Two additional teeth were sampled in 2015 due to the limited DNA preservation of the first two teeth samples (Tooth A and B). Three samples were taken: powder and a small solid piece of the tip of the root from one additional tooth (Tooth D), and powder from another tooth (Tooth C). Each sample was incubated in 5 mL of lysis buffer (0.5M EDTA; 50mM Tris HCl; SDS 0.5%; 0.01 mg/ml Proteinase K) for 2 h at 56°C with agitation. Subsequently, DNA was extracted by the phenol-chloroform method. Blank tubes were also processed as extraction controls. Extracts were purified with Centricon-30 spin columns (Amicon), obtained 60-70ul of DNA extract, add water to 300 ul, three aliquots of 100 μL each were finally stored.[36,100] These two teeth were extracted at the Institute of Anthropology, Bucharest (Romania), and DNA extracts later brought to the University of the Basque Country in Spain and Uppsala University in Sweden. DNA extraction and library preparation were performed in a dedicated ancient DNA clean room facility at the department of Organismal Biology, Human Evolution, Uppsala University, containing positive air-pressure and with incoming air passing through HEPA filter. The lab is regularly irradiated using UV (254 nm) and all work spaces are cleaned using sodium hypochlorite. Protective clothing was used and all the work was performed in safety bench class II.

DNA was extracted from powder using silica-based methods either as in Dabney et al.[33] using silica spin-column containing a volume extender from High Pre-Viral Nucleic Acid Large Volume kit (Roche)[101] or Yang et al.[32] The following modifications were implemented in the method from Yang et al.[32] The powder was incubated in 1 mL lysis buffer, (final concentrations: 0.45M EDTA, (pH 8.0 and 0.2 mg/ml Proteinase K)) and agitating the tubes for 18-24 h at 37°C. Additional 0.1 mg/ml Proteinase K was added followed by incubation at 55°C for 4-5 h. Remaining powder was pelleted by centrifugation at 2000 rpm for 5 minutes and supernatant was transferred to Amicon Ultra-4 PLTK Ultracel-PL membrane 30kDa filter (Millipore). This was followed by centrifugation for 10-15 minutes at 5000 rpm and 100 μl was transferred to silica-based membrane for binding and elution of DNA using Minelute purification kit (QIAGEN) according to manufacturer's recommendations.[32,34,102,103] The total volume of each extract was 50-110 μl (eluted in buffer EB, QIAGEN) and 1-2 extraction blanks were processed as controls.

In previous experimentation, we have noted improvements of endogenous DNA content when avoiding powerdization of the tooth material. In order to directly test that possibility, we sampled both powder and a small solid piece from the same tip of tooth (Tooth D). The solid piece was removed using a Dremel Multi Tool and a diamond cutting wheel. The powder was treated using the standard protocol, described above. For DNA isolation of the solid piece, 2mm x 1mm x 3 mm tooth piece was pre-digested in 1 mL EDTA (0.5M, pH 8) for 30 minutes at 37°C and thereafter the EDTA solution was discarded.[104] This was followed by incubation in 1 mL lysis buffer (final concentrations: 0.45M EDTA, (pH 8.0 and 0.2 mg/ml Proteinase K) agitating for 24 hours at 37°C. Thereafter, the same protocol as for powder was used.[32–34,101–103] In order to completely dissolve the small piece of the tip of the root the extraction was repeated following the same procedure as described above. In total two DNA extracts were prepared from one solid piece and a total of three DNA extracts were prepared using bone powder. The sequencing results between powder and the solid piece are compared in Figure 2.

Library building (Uppsala protocol)

Double stranded DNA libraries for Illumina sequencing were prepared from 20 μl DNA extract using protocol by Meyer and Kircher[35] with modifications as in Günther et al.[103] In total 42 DNA libraries were prepared from 8 extracts, 3 from Spain and 5 from Uppsala. In order to determine the number of cycles for library amplification, qPCR was performed (CFX Connect 96 BioRad). A total volume of 25 μl, containing 1 μl of DNA library, 1X MaximaSYBRGreen mastermix and 200 nM each of primer IS7 and IS8[35] were set up in duplicates. Libraries were amplified in quadruplicates using amplification protocol as in Günther et al.[103] The PCR reactions were pooled and purified using AMPure XP (Beckman Coulter) according to manufacturer's protocol and quantified on the Tapestation 2200 system using High Sensitivity D1000 screen tapes (Agilent Technologies). Purified libraries were pooled in equimolar concentration and deep sequenced on an Illumina HiSeq 2500 using v.4 chemistry and 125 bp paired-end reads or HiSeqX, 150bp paired-end reads using v2.5 chemistry at the SNP & SEQ Technology Platform, Uppsala University.

QUANTIFICATION AND STATISTICAL ANALYSIS

Processing of next generation sequence data

Adapters where trimmed and pair-end reads were merged using MergeReadsFastq_cc.py,[57] which requires at least 11 base pair overlap between overlapping reads. Bwa 0.7.13[58] was used to map the reads as single end reads to the human reference (build hg19). Non-default parameters for bwa were -l 16500 -n 0.01 -o 2.[46,105,106] Reads longer than 35 base pairs and with less than 10% mismatches to the reference were kept for further analysis. Biological sex was determined using the Ry script as described in Skoglund et al.[60] to female. To ensure maximal retention of reads all sequencing runs of the same library were merged (using samtools v.0.1.19[59]) before removal of PCR duplicates (per library) with a modified version of FilterUniqSAMCons_cc.py,[57] which ensures random assignment of bases in a 50/50 case. Reads with identical start and end positions are identified as PCR duplicates and collapsed. All 42 libraries were then merged, resulting in an estimated genome coverage of 13.5X and 1492X mitochondrial coverage.

The seven individuals with genome coverage of 13x or higher[28,34,44–46] were subjected to diploid genotype calling. Before genotype calling, the base qualities of all Ts in the first five base pairs of a read and all As in the last five base were set to 2. Picard[62] was then used to add read groups and indel realignment was conducted with GATK[61] using the indels identified in phase 1 of the 1000 genomes project (KGP)[107] as a reference. Genotypes were then called with GATK v3.5.0's UnifiedGenotyper and the parameters -stand_call_conf 50.0, -stand_emit_conf 50.0, -mbq 30, -contamination 0.02 and –output_mode EMIT_ALL_SITES using dbSNP version 142 as known SNPs. vcftools[63] was then used to extract SNPs from the VCF file excluding all filtered SNPs and transitions for non-UDG treated samples.

Authentication of data and contamination estimates

PM1 exhibits common characteristics of ancient DNA with short fragment length (Figure S1A) and a high frequency of cytosine to thymine transitions at the 5′ end.[108,109] Figure S1B shows that PM1 have the typical damage pattern for ancient DNA.

To investigate potential contamination different methods were applied. Mitochondrial contamination was estimated with the method described in Green et al.[110] with positions lifted over from hg18 to hg19. Private or near-private alleles (< 5% in 311 modern mtDNAs),[110] with a base and mapping quality greater than 30, as well as a coverage of at least 10x of the ancient sample were used. Sites where a transition substitution and a consensus allele of either C or G were removed to compensate for postmortem damages. To obtain a contamination estimate and counts of consensus and alternative alleles were added together across all sites[110] (Data S1A).

To estimate autosomal contamination the program VerifyBam[64] was used together with the 1000 genomes reference panel after removing all transversion sites (Data S1B). This left 5,270,838 SNPs. VerifyBamID estimated the contamination of PM1 to about 6%. To compensate for this we investigated how much this affects diploid calls. Restricting the analysis to sites with confidently called diploid genotypes with minimum coverage of 10x reduced the estimated contamination to 2.5%, probably because contamination has only an effect on diploid calls at low coverage sites. The effect of contamination would be falsely called heterozygous sites which made us apply additional filters on such sites to reduce contamination even further. We required the variant allele to be present at least to a certain proportion of the total number of reads at that site. We observed that a proportion of 0.2 results in an autosomal contamination of 1.5% (Figure S1C) which is not uncommon in aDNA studies and should not substantially affect our downstream analyses of genetic diversity. We performed identical filtering for all high coverage genomes before analyzing genetic diversity.

As a final quality control, we also estimated relative error rates using ANGSD[65] by comparison to an out-group (chimpanzee mapped against hg19) and an error free individual. To generate an ''error-free'' sample, reads with a mapping quality higher than 35 from KGP phase 1[107] CEU male, NA12342, were used to call a fasta consensus sequence (using the -doFasta command in ANGSD). By comparing the quantity of derived alleles in our samples in relation to the quantity between the error-free individual and ancestral state a relative error can be calculated. All reads were used but only sites where both ancestral, ''perfect'' sample and the sample have at least a coverage of one with a base quality higher than 30 were used. The results do not indicate that PM1 has more error than other non-damage repaired samples (Figure S1E).

Kinship between PM1 and PM2

To investigate kinship between Muierii 2[14] and our Pestera Muierii 1, READ[66] was used. READ divides the genome into non-overlapping windows and calculates the proportion of non-matching alleles inside each of these windows. The samples are then classified as

either unrelated, second-degree, first-degree or identical individuals or twins. Using the KGP SNP positions ancient samples from similar dating and geographical region were compared. It is difficult to achieve a good expectation for unrelated individuals in this case as we cannot make good population assignments for Upper Palaeolithic individuals,[14] which may lead to overestimation of the relationship between two individuals.[66] Despite Muierii 2 having a relatively low coverage (0.0046x) READ suggests that they are the same individual or identical twins (Data S1C). This is not an effect of coverage since Cioclovina 1 has lower coverage than Muierii 2 and no relationship is identified for him. We also added shotgun[30] and SNP capture data[14] for the Kostenki 14 individual as a positive control, and they were successfully identified as the same individual. To complement this analysis, we calculated pairwise mismatch rates between different sequencing libraries of PM1 and compared the values to the differences between PM1 and PM2 which falls at the lower end of the within PM1 distribution (Figure S1D). This suggests that the two remains belonged to the same individual or identical twins.

Definition of the mitochondrial haplotype of PM1

The mitogenome of PM1 was previously identified as a basal haplogroup U6*, with a private mutation at position T10517A which had not previously been found in any ancient or modern human.[36] Since then U6 has been identified in one more ancient specimen, PM2, the sample originating from the same cave as PM1,[14] now also suggested to be related to or the same individual as PM1 (Data S1D). To call a consensus sequence for the mitochondrial genome of PM1, we used mpileup and vcfutils provided by samtools[59] requiring mapping and base qualities of at least 30. We reanalyzed the mitogenome from PM1, now at 1492 X coverage, and confirm the same private mutation as previously found (Data S1D).

The Bayesian phylogenetic re-analysis was performed to infer the phylogenetic position of PM1. The best-fit model of evolution was selected using jModeltest 2[67] under AIC, BIC, and AICc criteria prior to Bayesian analyses. Bayesian analyses were carried out using BEAST 2.[68] Two simultaneous runs of 50 million generations were conducted for the datasets and trees were sampled every 1,000 generations, with the first 25% discarded as burn-in. Samples from the posterior were checked for acceptable effective sample sizes (> 200) and the adequate convergence of the MCMC chains was checked using Tracer 1.[69]

The phylogenetic reanalysis was performed using the complete mitogenomes of 245 modern *Homo sapiens*, including 61 individuals from the Upper Palaeolithic (15 individuals have been added to the previous analysis in Hervella et al.,[36] 41 humans from the Neolithic, one individual from the 15th century and 143 modern samples (Data S1E). The modern humans were selected from all the individuals with published mitogenomes and covered the whole phylogenetic diversity within the N hg lineage, with special emphasis in the U lineage. The analysis was carried out using the HKY+G+I substitution model, strict molecular clock and coalescent constant population tree prior, indicating the tip dates of the samples (Figure S2A).

We confirm the basal U6 mitochondrial haplogroup for the PM1 genome as described in Hervella et al.,[36] and increase the coverage in 1492X. Our estimates of the haplogroup U6 TMRCA that incorporate ancient genomes (including PM1) set the formation of the U6 lineage back to 49.6 ky BP (95% HPD: 42–58 ky) (using a mutation rate of $2.06^{*}\ 10^{-8}$ SD = $1.94\ ^{*}\ 10^{-9}$). Our estimates are almost identical in age to that by Secher et al.[111] (45.1 ± 6.9 ky). Moreover, our conclusion has been supported by a recent study[38] who presented genomic data from seven ~15,000-year-old modern humans from Morocco, defined six sub –haplogroup U6 (U6a1b, U6a6b and U6a7), this being the most ancient signal of the back to Africa migration from Eurasia in Africa. They estimated a divergence time of the haplogroup U6 with a value similar to Hervella et al.[36] and Secher et al.[111]

In summary, given the presence of a basal U6 mitogenome in Romania 34 ky BP, the presence of U6 haplotype derived in Northern Africa 15ky,[38] and the estimated TMRCA for U6 haplogroup, we can support the view that the PM1 mitochondrial lineage of South East Europe is an offshoot that can be traced to the Early Upper Palaeolithic back migration from Western Asia to North Africa, during which the U6 lineage diversified. However, the timing and geographical direction for the spread of the U6 haplotypes, and their association with a back-to-Africa migration, remains poorly understood.

Comparative Data

We compare the genome of PM1 to a set of published modern and ancient individuals. Comparative ancient NGS data (Data S1F) were all mapped and filtered using identical parameters as PM1. The ancient data were then merged with the Human Origins panel[112] 1.9 million transversions polymorphic in the Yorubans of the 1000 genomes phase 3 data[107] as well as 1.2 million SNPs commonly used in SNP capture.[40,113] For each individual and site a random read covering the site with a minimum mapping and base quality of 30 was drawn, and the allele was used as a pseudo haploid representation of the ancient individual. Indels were excluded and to compensate for post-mortem damages transitions were set to missing data.

Principal Component Analysis

We performed a principal component analysis (PCA) to characterize the genetic relationship of PM1 and previously published ancient individuals to modern populations. We used smartpca (v. 10210)[70] to run a PCA on a selection of the Eurasian populations from the Human Origins dataset as well as the KGP (1000 genomes) Eurasian populations represented by haplodized data. One PCA per individual was conducted for 54 ancient individuals[14,28,29,34,44–46,113,114] using overlapping SNP sites only. We then used Procrustes analysis[115] to align the coordinate systems between the different PCAs.[116] The results are shown in Figures S3A and S3B. While this analysis reproduces some similarities among ancient individuals that mainly correlate with time (Fu et al.[14]), we need to highlight that this type of comparison with modern populations may be misleading as some of the ancient individuals predate the development of different lineages in Eurasia which contribute to the pattern seen in modern groups. The location of the chronologically Upper

Palaeolithic Ust'-Ishim and Oase individuals outside of modern variation can serve as an example for this pattern, which is the reason why we caution against the interpretation of other later hunter-gatherers overlapping with modern European groups in this PC1/PC2 space (Figures S3A and S3B). In fact, we see equal affinities for most of these Upper Palaeolithic hunter-gatherers to all modern Europeans (see statistics below).

Model-based clustering

We performed model-based clustering of ancient and modern individuals using the approach implemented in ADMIXTURE.[71] To maximize the number of sites used, we employed the 1000 genomes panel as modern populations. The data were pruned for linkage disequilibrium using PLINK (v.1.90)[72,73] with parameters –indep-pairwise 200 25 0.4, and then unsupervised ADMIXTURE was run with 2 to 20 clusters (K-values), 20 iterations each. Common modes among different runs and clusters were aligned along different values of K using the greedy mode with pong (v.1.4.6).[74] The results were then plotted in R.[117] The results are shown in Figure S4.

Outgroup f_3 statistics

To estimate shared drift among ancient individuals as well as between ancient and modern populations outgroup, we calculated f_3 statistics using Admixtools version 3.0.[112] All combinations of ancient individual and modern populations were tested with Yoruban as an outgroup population. Standard errors were estimated using a block jackknife procedure.

Testing outgroup f_3 statistics of PM1 against all other ancient individuals reveals a high proportion of shared drift between PM1 and the published data for PM2 which is consistent with the results of the kinship analysis suggesting that they are either identical twins or the same individual. Comparing PM1 to other ancient individuals show the highest amounts of shared drift to European hunter-gatherers assigned to the Vestonice cluster in Fu et al.,[14] confirming PM1's position among the early European hunter-gatherers.

For eight high coverage individuals (Ust'-Ishim, Kostenki 14, Sunghir III, PM1, SF12 (Stora Förvar 12)), Loschbour, LBK, and ne1 (Polgár-Ferenci-hát 1)), we plotted the outgroup f_3 results comparing them to modern populations from the Human Origins dataset on a geographic map of Europe. We used the fields package for R[118] and the settings krig(theta = 3, m = 1) and predictSurface(extrap = TRUE, nx = 2000, ny = 2000,type = ''l''). To be able to compare ancient individuals only the top 60% of all f_3-values were included for each individual plot (Figure S2B). The amount of shared drift increases over time as expected since later individuals should share more history with modern groups. Most individuals do not show a strong geographic structure of their affinity which becomes more pronounced in post-LGM individuals (sf12, Loschbour, LBK, ne1). Notably, PM1 lacks a strong affinity to modern North African populations as suggested by her U6 mitochondrial haplogroup.[36] While U6 seems to be a marker of back-migrations from Eurasia into Africa, we do not see a stronger connection between PM1 and North Africans as opposed to other ancient Europeans. This is further supported when testing for allele sharing between PM1 and the Iberomaurusian genomes from Taforalt using f_4 statistics of the form f_4(Chimp, Taforalt; X, PM1). There is no indication that Taforalt and PM1 share alleles to the exclusion of other prehistoric Europeans. Taforalt is symmetrically related to PM1 and other pre-LGM individuals while Taforalt shows an excess of allele sharing with post-LGM individuals to the exclusion of PM1 (Figure S3D). The latter could be explained by later expansions from the Levant into both Europe and North Africa.[14,38,119] This suggests that the autosomal connection between ancient Europeans and North Africans is similar for all individuals. Admixture studies of modern North African populations have shown substantial proportions of Eurasian ancestry in these regions.[37] Other admixture components, especially from sub-Saharan populations, contribute to a very complex mixture of ancestries in North Africa.

Archaic ancestry

To estimate the proportion of Neanderthal admixture α in the ancient individuals we calculated f_4 ratios with AdmixTools version 3.0[112] using the following formula:

$$\alpha = \frac{f_4(Chimp, AltaiNeanderthal; Yoruba, X)}{f_4(Chimp, AltaiNeanderthal; Yoruba, VindijaNeanderthal)}$$

where X is the individual/population for which we want to estimate archaic ancestry. Standard errors were estimated using a block jackknife procedure. The archaic genomes used were, Altai Neanderthal[120] (44x coverage) and Neanderthal Vindija33.19[121] (24x coverage). The estimated Neanderthal ancestry in the newly sequenced PM1 is 3.1% (SE: 0.41%) which is similar to contemporary individuals. To investigate if Neanderthal ancestry decreased over time, we tested whether sample age is correlated with Neanderthal admixture. The samples were plotted based on sample age, with error bars representing 95% confidence interval (Figures S3E and S3F). Excluding the outlier with recent Neanderthal admixture, Oase, a linear regression did not show a strong slope for both all-ancient and modern samples (Figure S3F) as well as the high-coverage and modern samples (Figure S3E) (all-ancient samples: p = 0.5144). These results are largely consistent with recent studies showing no substantial decline in Neanderthal ancestry over the last 40,000 years but also sensitivity to the exact procedures used for estimation.[39]

We also investigated the distribution of Neanderthal related tracts across the genomes of prehistoric Europeans similar to the approach of Sikora et al.[29] and Seguin-Orlando et al.[30] We first investigated informative sites as sites where both high coverage Neanderthals from Altai and Vindija are homozygous derived while the derived allele is virtually absent (< 1%) among Africans in the 1000 genomes data. This information was obtained from Kuhlwilm and Boeckx.[122] To reduce the effect of sequencing depth, we conducted this analysis on the pseudohaploid data of the prehistoric Europeans. A Neanderthal tract is defined as a run of consecutive derived sites with a maximum of 200 kb and up to 13 ancestral alleles between derived alleles (corresponding to a binomial sampling

probability of < 0.0001). As expected under a model where most Neanderthal ancestry can be traced to a single admixture event, we find fewer but longer tracts in older individuals compared to post-glacial individuals (Figure S3G).

Treemix

To study the relationships among prehistoric populations, admixture graphs were constructed using TreeMix v1.12.[75] This analysis included PM1, Ust'-Ishim, Kostenki 14, Sunghir III, individuals representative of the clusters identified in Fu et al.[14] (MA1, Vestonice16, El Miron, Villabruna, Satsurblia, Goyet Q116-1), as well as Loschbour and Stuttgart to represent later populations. Modern Mbuti were used as outgroup. Only sites without missing data in any individual among the captured SNPs were included in the analysis which restricted the analysis to 37,112 SNPs. As each population is represented by a single sample, correction for sample sizes was switched off (-noss) and standard errors were calculated in blocks of 500 SNPs. TreeMix was run with ten different random seeds per setting and the run with the best likelihood was selected. Allowing for three migrations resolves all pairwise residual covariances with $|Z| > 2$ and the migration edges are similar to the population connections observed in Fu et al.[14]

The general pattern of the graph is consistent with other results (Figure S3H). PM1 clusters with European hunter-gatherers while MA1, Neolithic farmers and Caucasus hunter-gatherers form a separate clade. Migration edges mainly show connections among European hunter-gatherers but we also observe admixture from Loschbour into the common ancestor of MA1, Neolithic farmers and Caucasus hunter-gatherers.

Admixture graphs

In order to perform a phylogenetically more explicit assignment of PM1 into the major pre-LGM European hunter-gatherer gene pools,[14,29] we used qpGraph (version 6040) of ADMIXTOOLS[112] to generate admixture graphs including a number of European hunter-gatherer individuals. A user-defined model for the populations is used as input for qpGraph which then calculates all combinations of f_2, f_3 and f_4 statistics and fits drift parameters and admixture proportions. Each internal node in the graph can represent a bifurcation into two child populations and/or the recipient or source for two-way admixtures. The f statistics predicted by the model are compared to the observed f statistics and deviations are expressed in multiples of standard error (Z score) which are estimated using a block-jackknife across the genome. Low Z scores are usually interpreted as consistency between model and data. We used qpGraph with the following parameters: blgsize: 0.005, lsqmode: YES, diag: 0.001, details: YES, hires: YES, allsnps: YES, precision: 0.0001, initmix: 1000.

Previous studies on European Upper Palaeolithic hunter-gatherers have identified major several major gene pools in pre-LGM Europe.[14,29] Our goal is to understand where PM1 fits into such models. We use a single Mbuti individual[105,123] as outgroup and the pre-LGM Ust'-Ishim, Kostenki 14, Sunghir III, Goyet Q116-1 as well as the post-LGM Loschbour[14,28,30,46] as representatives of the different hunter-gather groups. Adding PM1 to this set and using the 1240K capture SNP panel left 123,560 SNPs with overlapping data. The resulting model (worst $|Z| < 2.8$) is shown in Figure S3G. It shows PM1 as a population drawing ancestry from both the Russian Sunghir III as well as the Belgian Goyet Q116-1. PM1 is also distantly related to related to a population contributing parts of the ancestry to later hunter-gatherers (Loschbour) who then serve as the connection to modern Europeans.[46,113]

Diversity measures

To obtain unbiased estimates of genetic diversity, we filtered all high coverage ancient individuals in the same way as PM1: after diploid genotype calling with GATK UnifiedGenotyper, we only accepted heterozygous genotypes if the minor allele was present in at least 20% of all reads covering that site and if the total coverage of that site was at least 0.76 times the average coverage of that individual (approximately corresponding to 10x in PM1). To have high coverage modern individuals for comparison, we applied the same genotype calling and filtering to 272 individuals from a worldwide set of populations sequenced as part of the Simons Genome Diversity Project.[18] We only used individuals published as ''FullyPublic'' and excluded two Khomani San as well as four Ju\'hoan North individuals as we ascertained SNPs in a San individual (see below).

To avoid ascertainment biases, we ascertained SNPs in a single HGDP San individual sequenced as part of a separate project.[105,123] The bam file was filtered the same way as all other bam files and we ascertained 781,487 transversion sites (to avoid post-mortem damages) heterozygous in this single individual and then used this set of sites for all analyses of genetic diversity. Additionally, we annotated these SNPs using SNPeff[76] to be able to separate the signal between different site types (e.g., intergenic, intronic, synonymous, non-synonymous). We calculated two measures of genetic diversity. First, we calculated relative heterozygosity by counting the number of heterozygous SNP sites and dividing it by the total number of SNP sites covered for the individual. Standard errors were estimated using a block jackknife and a block size of 1000 SNPs. Second, plink v1.9[72,73] was used to assess runs of homozygosity in the diploid genotype calls using the parameters–homozyg-density 50,–homozyg-gap 1000,–homozyg-kb 500,–homozyg-snp 100,–homozyg-window-het 1,–homozyg-window-missing 35,–homozyg-window-snp 100,–homozyg-window-threshold 0.02. The results of this analysis are shown in Figure 4, heterozygosity is only displayed for intergenic sites to avoid the potential impact of natural selection on diversity at functional sites.

Phenotypic analysis

We used HIrisplex-S[77] to predict pigmentation phenotypes for PM1. HIrisplex-S uses 41 SNPs to predict eye, skin and hair color and it has been shown to perform well on modern Eurasians. This prediction is associated with some uncertainty in ancient individuals as

we do not know if there were functional alleles in these populations that are unknown today. The presence of unknown functional variants is not unlikely considering the complex genetic architecture of pigmentation phenotypes.[124,125]

We did not include the insertion polymorphism rs312262906 into the prediction model. The HIrisPlexS model predicts the following phenotype for PM1: brown eyes (p = 0.995), brown (p = 0.566) to black hair (p = 0.382) with dark shade (p = 0.931) and intermediate skin color (p = 0.999). These results are similar to other European hunter-gatherers who are mainly ancestral for the known pigmentation variants.[14,126]

Medical genetics analysis

The complete genome sequence from Peştera Muierii 1, together with the genomes of a few previously published high coverage genomes[28,29,34,44–46] allows for the first analysis of ancient individuals from a medical perspective. We approached this with a methodology used in medical genetics, but not yet adopted in aDNA studies, investigating shifts in potential pathogenic variants in genomes across Eurasia from EUP to the Neolithic.

All coding substitutions were extracted from the WGS dataset. To minimize the effect of post-mortem deamination, we excluded heterozygous C/T and A/G sites with coverages below 10 and/or ratios of the two alleles more extreme than 25/75 or 75/25 across all sequencing reads.

Subsequently all variants were annotated using a clinical annotation pipeline as previously described.[47,127] This annotation included protein consequences of the variants; nucleotide conservation score (vertebrate PhyloP); frequencies from public databases (1000 genomes variant frequency data retrieved from db SNP144: https://www.ncbi.nlm.nih.gov/SNP/; ExAC [sept2016]: http://exac.broadinstitute.org/; GoNL: https://www.nlgenome.nl/) as well as exome data of > 15,000 in-house exomes mainly from individuals of Dutch descent; the annotation also checked for the presence of the same variant in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) or HGMD (http://www.hgmd.cf.ac.uk/; 2015 version); gene based annotations included OMIM entries, GO term and KEGG annotations.

In order to compare the variant burden in ancient versus modern humans, we compared the variants detected in PM1 and all other ancient exomes with 10 representative modern human exomes. The latter were exomes of healthy parents that gave birth to a child with severe intellectual disability.[47]

To identify a burden of possibly damaging variants in ancient versus modern exomes, we calculated the ratio of non-synonymous/synonymous variants; and counted most likely damaging variant categories a) stop-gains; b) variants with a CADD[49] score > 15; c) missense variants with a nucleotide conservation score (vertebrate PhyloP[128]) > 3.0.

Damaging burden test

We first explored whether there are any differences in terms of potential pathogenic variants or an overall higher burden of non-synonymous variants in the coding part of the genome (exome) of PM1 and other ancient specimen compared to modern day healthy humans. Because damage of ancient DNAs complicates a comparison with modern day human genomes, we focused on variants that were likely homozygous, i.e., > 80% variant reads, with known dbSNP entries (rs IDs) in the coding regions, in order to avoid the potential false-positive variants in ancient DNA specimen. When comparing coding variants in ancient versus modern day humans (Data S1G), several interesting observations could be made:

a) Overall, we see slightly lower NS/S-ratio in ancient versus modern humans; i.e., on average a NS/S ratio of 0.914 in ancient exomes versus 0.9471 in modern exomes. Thus, overall ancient exomes show on average lower degree of non-synonymous variants. The lowest NS/S ratio is observed for STIII; Loschbour and Ust'-Ishim i.e., we see proportionally fewer homozygous non-synonymous variants in those.
b) Despite this, among all NS variants, the percentage of possibly damaging variants (e.g CADD: > 15 or missense with high nucleotide conservation score: vertebrate PhyloP > 3.0) is somewhat higher in some ancient exomes compared to modern day exomes. The percentages of those variants are highest among post-LGM hunter-gatherers Bichon (however, this individual also has the lowest coverage and the data are not damage repaired) and Loschbour, followed by Sunghir III, sf12, Ust'-Ishim and Kotias (Figure 5). This pattern seems more pronounced for variants with a high CADD score than in PhyloP scores but statistical tests do not detect significant differences between individuals (Figure S5) or temporal groups (Figure 5).
c) Comparing the full distributions of CADD (Figure S5C) and PhyloP (Figure S5D) scores between temporal groups, shows significant differences between modern and ancient groups. The CADD scores seem highest in post-LGM hunter-gatherers while PhyloP scores seem highest in pre-LGM hunter-gatherers.

The post-LGM hunter-gatherer group has the lowest level of neutral genetic diversity (Figure 4), which would suggest that these individuals also carry a higher proportion of deleterious variants – both compared to modern populations and to pre-LGM hunter gatherers. Our analysis of genetic load is inconclusive (differences between the two measurements CADD and PhyloP scores, differences between the full distribution and outliers), which means that there is no support for strong differences of genetic load between the ancient groups. This suggests that, while the post-LGM hunter-gatherers lived in small groups and also had a reduced effective population size, their genetic load was still at a level which was not substantially higher than that of other groups.

Medical genetic analysis of ancient individuals

While we do not have a strong indication for an overall higher degree of damaging variants in ancient exomes, we identified several variants in Pestera Muierii 1 genome and the other ancient genomes that stood out as potentially involved in human pathology. Here we focus on rare, likely homozygous variants in the exome of ancient DNA specimen (Data S1I), as well as likely homozygous non-synonymous variants with a HGMD entry (http://www.hgmd.cf.ac.uk) (Data S1I).

PM1: To avoid over-interpretation of technical artifacts in the ancient DNA, we focused on variants that were known polymorphisms (dbSNP144 entries), yielding a total of 18,086 variants (Data S1I and S1J). Subsequently, we focused on likely homozygous variants, i.e., requiring > 80% variant reads, with at least 3 reads total coverage, yielding 7,320 variants that fulfilled those criteria.

In order to identify rare homozygous variants, we filtered on population frequencies < 1% (based on 1000 genomes entries of dbSNP144; ExAc, GoNL and our in-house exome database with > 15,000 exomes). This filter strategy resulted in 14 rare, likely homozygous variants, 8 of which were non-synonymous (Data S1J). This ''clinical exome'' interpretation yielded several interest variants for which additional information is available or the gene has a well understood function. These variants include rare, likely homozygous, variants in several potentially medically relevant genes, including *PTPN12 (p.(Ser684Leu)), CTTN (p.(Val461Ile)), IL32 (p.(Trp169*)),* and *ANKRD11 (p.(Glu1413Lys))*.

Among these remarkable observations is a premature stop-gain in *IL32* p.(Trp169*). As this variant affects the last exon of *IL32*, the immediate consequences on protein level are difficult to predict but an escape of nonsense-mediated-decay (NMD) and a resulting truncated protein is likely. Interestingly, this variant was never described in any modern human (dbSNP, ExAc), but other stop-gain/frameshifting variants in the same exon have been described in ExAC as rare events in the general population. *IL32* encodes the interleukin-32 protein, a member of the cytokine family, with an important role in immune responses[51] and which has been associated with cancer development.[129] Polymorphisms in this gene modulate susceptibility to several types of malignancies.[130–133] Interestingly, two additional rare, likely homozygous, non-synonymous variants (Data S1J) were identified in genes for which also a cancer association was previously shown: *PTPN12*[134,135] and *CTTN*.[50] It would be therefore tempting to speculate whether these genetic variants may have influenced susceptibility to cancer of the PM1 individual, but more data in additional ancient genomes in the future or additional functional evidence would be needed in order to be able to draw any general conclusions.

Finally, one likely homozygous missense variant was identified in *ANKRD11*. Heterozygous loss-of-function mutations or deletions of this gene have been described as the cause of KBG syndrome (CITE). KBG syndrome is characterized by macrodontia, distinctive craniofacial features, short stature, skeletal anomalies, global developmental delay, seizures and intellectual disability. We were however able to exclude this diagnosis based on the cranial characteristics: the Pestera Muierii 1 ancient individual does not resemble any phenotypic features of KBG syndrome by systematically analyzing its skull for the prevalence of the clinical hallmarks of KBG syndrome (Data S1M). We therefore consider the p.(Glu1413Lys) variant of *ANKRD11* as likely benign.

Other ancient exomes: for all other ancient exomes similar filters were applied. Of all the rare, likely homozygous variants, we only report those which have a HGMD entry, as well as more common likely homozygous variants with an HGMD entry (Data S1I). Additionally, we focused on all rare (< 1%) likely homozygous stop-gains identified in all exomes, those are summarized in Data S1K. Here we have identified likely homozygous variants in possibly medically relevant genes including *AIPL1* (p.(His82Tyr)), *NOTCH3* (p.(Leu1518Met)), *NLRP3* (p.(Gln705Lys)), and *APC* (p.(Gly2520Ser)). The exact same nucleotide changes in *AIPL1* as identified in Bichon specimen was previously described in a single, sporadic case of Leber congenital amaurosis 4 (OMIM: 604393) and retinitis pigmentosa, in compound heterozygous state together with p.(His90Asp) (the latter is a known SNP). However, the Bichon variant p.(His82Tyr) is not in a functional domain, and there is no functional evidence for pathogenicity;[136] moreover, the same variant is described 3 times in ExAC in homozygous state (2x in EUR, 1x SA). Alltogether, the evidence for pathogenicity of this variant is too limited to be able to firmly diagnose this disorder, which in itself would have been severe as it causes blindness in the affected individuals. While the *NOTCH3* variant is interpreted by ClinVar as 'likely benign', and the *APC* missense variant is most likely also benign, as it has been identified in 80 carriers in ExAC (mainly EUR, and SA/EA), the *NLRP3* p.(Gln705Lys) homozygous variant has been suggested as a recessive disease gene for cold-induced autoinflammatory syndrome, OMIM:120100). However, the high variant population frequency of > 2% also suggests a benign nature of this specific variant.

Genetic variants and antimicrobial host defense

A long-standing host-pathogen ''arm race'' has shaped the human genome, as a result of selective pressures acting upon genes involved in host resistance and immunity response. Among the genes and pathways involved in antimicrobial host defense, proinflammatory cytokines play a crucial role in the initiation of a successful response to pathogens. We assessed the presence of five genetic polymorphisms known to be strongly associated with higher cytokine production capacity. Interestingly, Peştera Muierii 1 genome harbors the variants associated with a strongly increased cytokine production capacity for 4 of these 5 SNPs: heterozygous carrier for C allele of *TLR1* rs4833095, homozygous carrier of G allele for *TLR6* rs5743810, heterozygous carrier for G allele of *TLR10* rs11096957, and heterozygous carrier for T allele of *IL10* rs1800872.[53] In addition, the Pestera Muierii 1 individuals was carrier of heterozygous *IFNG* rs2069727, associated with an average cytokine production. All in all, these data suggest that Pestera Muierii 1 individual was a high responder in terms of cytokine production capacity; this combination of high-cytokine polymorphisms is present in less than 4% of the modern European population. Considering the protective effects of high immune responses in the context of high infection burden,[53] it is likely that this represents an adaptive state conferring beneficial protective effects.