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

Emma Svensson,1,8 Torsten Günther,1,6,* 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 Jakobsson1,9,*

1Human Evolution, Department of Organismal Biology, Uppsala University, 752 36 Uppsala, Sweden
2Department of Human Genetics, Radboud University Medical Center, 6526 Nijmegen, the Netherlands
3Department of Internal Medicine and Radboud Center for Infectious Diseases (ROI), Radboud University Medical Center, 6526 Nijmegen, the Netherlands
4Department 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
5Laboratory of Human Genetics, University of Medicine and Pharmacy, Craiova, Romania
6Oltenia Museum Craiova, Craiova, Romania
7“Francisc J. Rainer” Institute of Anthropology, Romanian Academy, 050474 Bucharest, Romania
8These authors contributed equally
9Lead 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 “Peștera 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 Peștera 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 change1 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 AMHs 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 “Peștera cu Oase,” “Peștera Muierii,” or “Peștera 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 Peștera 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 evidence13–16 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-Africans18 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
We sequenced and analyzed the genome of Peștera Muierii 1 (PM1) dated to ~34 ky calBP. Taken together with the data offered by the few other complete EUP genomes, 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 >1x 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). With Peștera 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.

We extracted DNA from small portions of four teeth from the Peștera Muierii 1 individual. Using state-of-the-art approaches, including silica-based DNA extraction protocols, the DNA preservation and proportion of endogenous human DNA in Peștera Muierii 1 was at about 1% to 2% (Figure 2), similar to previous observations from material of this age and from Europe. Based on these DNA extracts, the genome coverage that could be expected when sequencing libraries to depletion was calculated at about 0.5x. 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). In these DNA extracts, the endogenous DNA content was 49% (mean across libraries; STAR Methods), allowing us to obtain a genome coverage of 13.49x. 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
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 Peștera 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 S1 C). The sequence data of Peștera Muierii 1 confirm the previous assignment of Peștera Muierii 1 to the basal mitochondrial haplogroup U6* (STAR Methods; Figure S2A), 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 f4 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 Peștera 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 Peștera 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 Peștera 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 Peștera Muierii 1 and Peștera Oase 1 have been suggested to carry archaic morphological traits,7,41,42 they show distinctly different Neanderthal admixture levels and history. Peștera 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, Peștera 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 Peștera 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, Peștera 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 2. Proportion of endogenous human DNA**
Comparison of the traditional sampling method compared to the here described DNA extraction method. See also Figure 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.\(^\text{29}\) 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 Peștera Muierii 1, together with the sequence data from previously published high-coverage ancient genomes,\(^\text{28,29,34,44–46}\) 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\(^\text{47}\) 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...
two proxies to assess the load of possibly damaging variants among all missense variants: (1) distribution of nucleotide conservation scores (vertebrate PhyloP > 3.0) and (2) distribution of CADD (combined annotation-dependent depletion) scores >15 (Figure 5). 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).

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; 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 Peștera 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 Peștera 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).

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, Peștera 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. In addition, the Peștera Muierii 1 individual was a carrier of heterozygous IFNG rs2069727, associated with an average cytokine production. Overall, these data suggest that Peștera 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

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.
| Sample ID | R.C. | Chr. | Genomic position (hg19) | Ref. Var. Reads | SNP ID (dbSNP144) | SNP freq. (%) | In-house freq. (%) | GnomAD freq. (%) | Protein change | c.DNA variant | PhyloP score | CADD PHRED score | GATK quality score |
|-----------|------|------|-------------------------|----------------|------------------|---------------|-------------------|----------------|---------------|---------------|--------------|----------------|------------------|--------------------|
| PM1       | R.   | chr1 | 243,328,858             | C T 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         | rs145753116      | 0.439         | 0.276             | 0.3993         | p.(Ser236Aln) | NM_001001664.2(SPOPL):c.707G > A | 7.843 | 14.71 | 733.77 |
| PM1       | R.   | chr7 | 2,473,130               | G C 5          | rs201518792      | 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        | rs201001953      | 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             | G C 15         | rs201257936      | 0.998         | 0.032             | 0.2812         | p.(Pro495Ala) | NM_001146274.1(TCF7L2):c.1483C > G | 2.527 | 12.7  | 453.89 |
| PM1       | R.   | chr11| 70,279,800              | G A 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          | rs775284592      | NA            | 0                 | 0.0008204      | p.(Trp169*) | NM_001308078.1(IL-32):c.506G > A | −0.231 | 11.8  | 128.03 |
| PM1       | R.   | chr16| 89,348,713              | C T 10         | rs140373729      | 0.020         | 0.019             | 0.04843        | p.(Glu141Lys) | NM_001256182.1(ANKRD11):c.4237G > A | 5.366 | 15.28 | 302.18 |
| Bichon     | R.   | chr17| 6,337,271               | G A 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         | rs112729404      | 0.658         | 0.735             | 0.6972         | p.(Thr76Met) | NM_001919.3(ECI1):c.227C > T | 6.033 | 9.763 | 1,719.77 |
| Kotias     | C.   | chrX | 153,595,771             | C T 2          | rs267060816      | NA            | 0                 | 0.00008204     | p.(Gly288Arg) | NM_001110556.1(FLNA):c.862G > A | 7.809 | 14.76 | 52.74 |
| Loschbour  | R.   | chr5 | 78,280,974              | G A 4          | rs201168448      | 0.219         | 0.604             | 1.047          | p.(Ala33Val) | NM_0000046.4(ARSB):c.98C > T | −0.344 | 9.108 | 52.74 |
| Ust'-Ishim | R.   | chr19| 15,285,063              | G T 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         | rs2234036        | NA            | 0.196             | 0.4466         | p.(Ala211Thr) | NM_000354.5(SERPIN2):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). Chr., chromosome; freq., frequency; IL-32, interleukin-32; R., rare (<1% population frequency); ref., reference nucleotide; var., variant nucleotide.

*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, 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, which may help us understand the different views of genetic load in humans. 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 APL1 (p. (His82Tyr)) described in a sporadic case of Leber congenital amaurosis, 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, 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 resurfaced after the LGM.

STAR METHODS

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

- 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 f3 statistics
  - Archica 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 (IT1138-18).

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.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.
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., Sirkov, N., Adeilas, V., Bailey, S., Bard, E., Delvigne, V., Endarova, E., Fagault, Y., Frewiass, H., Hajdinjak, M., et al. (2020). Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria. Nature 587, 299–302.

5. Mellars, P. (2004). Neanderthals and the modern human colonization of Europe. Nature 432, 461–465.

6. Mellars, P. (2008). 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., Dobos, A., and Trinkaus, E. (2007). The human cranium from the Peștera Cioclovina Uscată, Romania: context, age, taphonomy, morphology, and paleopathology. Curr. Anthrop. 48, 611–619.

9. Trinkaus, E., Constantin, S., and Zilhão, J. (2013). Life and Death at the Peștera cu Oase: A Setting for Modern Human Emergence in Europe (Oxford University).

10. Dobos, A., Soficaru, A., and Trinkaus, E. (2010). The Prehistory and Paleoanthropology of the Peștera 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. Paleontology of the Peștera Muierii (Universită de Lie`ge).

13. Posth, C., Renaud, G., Mittnik, A., Drucker, D.G., Rougier, H., Cupillard, C., Valentín, F., Thevenet, C., Furtwangler, A., Willing, C., et al. (2016). Pleistocene mitochondrial genomes suggest a major single 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., Furtwangler, A., Häak, 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., Stosssel, 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., Hüblner, 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., Maples, B.K., Martin, A.R., Musharof, 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.

19. Aris-Brosou, S. (2019). Direct evidence of an increasing mutational load in humans. Mol. Biol. Evol. 36, 2823–2829.

20. Berens, A.J., Cooper, T.L., and Lachance, J. (2017). The genomic health of ancient homins. Hum. Biol. 89, 7–19.

21. Hoffecker, J.F., Gebele, P.H., Pichler, C., Gröppel, W., Kummer, G., Lesch, K., et al. (2016). Palaeolithic back-migration to Africa. Sci. Rep. 6, 15942–15947.

22. Berens, A.J., Cooper, T.L., and Lachance, J. (2017). The genomic health of ancient homins. Hum. Biol. 89, 7–19.

23. Aris-Brosou, S. (2019). Direct evidence of an increasing mutational load in humans. Mol. Biol. Evol. 36, 2823–2829.

24. Berens, A.J., Cooper, T.L., and Lachance, J. (2017). The genomic health of ancient homins. Hum. Biol. 89, 7–19.

25. Hoffecker, J.F., Gebele, P.H., Pichler, C., Gröppel, W., Kummer, G., Lesch, K., et al. (2016). Palaeolithic back-migration to Africa. Sci. Rep. 6, 15942–15947.

26. Berens, A.J., Cooper, T.L., and Lachance, J. (2017). The genomic health of ancient homins. Hum. Biol. 89, 7–19.

27. Hoffecker, J.F., Gebele, P.H., Pichler, C., Gröppel, W., Kummer, G., Lesch, K., et al. (2016). Palaeolithic back-migration to Africa. Sci. Rep. 6, 15942–15947.

28. Berens, A.J., Cooper, T.L., and Lachance, J. (2017). The genomic health of ancient homins. Hum. Biol. 89, 7–19.

29. Hoffecker, J.F., Gebele, P.H., Pichler, C., Gröppel, W., Kummer, G., Lesch, K., et al. (2016). Palaeolithic back-migration to Africa. Sci. Rep. 6, 15942–15947.

30. Hoffecker, J.F., Gebele, P.H., Pichler, C., Gröppel, W., Kummer, G., Lesch, K., et al. (2016). Palaeolithic back-migration to Africa. Sci. Rep. 6, 15942–15947.

31. Hoffecker, J.F., Gebele, P.H., Pichler, C., Gröppel, W., Kummer, G., Lesch, K., et al. (2016). Palaeolithic back-migration to Africa. Sci. Rep. 6, 15942–15947.
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., Taibl, 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., Paabo, 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 Neandertal ancestor. Nature 524, 219–224.

41. 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.

42. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Petr, 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. Bores 45, 1–45.

44. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangelo, V., Domoróczki, L., Kóvári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennia transcript 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., and Jones, E.A. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 47 (D1), D886–D894.

46. Ockeleon, C.W., Willemsen, M.H., de Munnik, S., van Bon, B.W.M., de Leeuw, N., Verrups, A., Kant, S.G., Jones, E.A., Brunner, H.G., van Loon, R.L.E., et al. (2015). Further delineation of the KGB syndrome phenotype caused by ANKRD11 aberrations. Eur. J. Hum. Genet. 23, 1176–1185.

47. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Aföldi, J., Wang, Q., Collins, R.L., Lalicchia, 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.

48. 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.
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., 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, Å., 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. Prufer, K., Schneider, J., van Adrichem, B., de Knudsen, L.K., Albrechtsen, A., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505, 43–49.

119. de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Keesstra, 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.

120. Nychka, D., Furrer, R., Paige, J., and Sain, S. (2005). fields: Tools for spatial data analysis. R Foundation for Statistical Computing.

121. Prufer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlevi, Kuhlwilm, M., and Boeckx, C. (2019). A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. Nature 575, 207–211.

122. Crawford, N.G., Kelly, D.E., Hansen, M.E.B., Beltramé, 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 in African populations. Science 358, eaan8433.

123. 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.e114.

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

125. Crawford, N.G., Kelly, D.E., Hansen, M.E.B., Beltramé, 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 in African populations. Science 358, eaan8433.

126. 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/Pesțera Mulierii 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*TCTTTCCCTACACGACGCTCTTCCG*A*T'C'T (' = Phosphorothioate) | N/A | Biomers |
| IS2 adapter P7: G'T*G*A'CTGGAGTTCAGACGTGTGCTCTTTCCG*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: AATGATACGGCGACCACCCGAGATC TACACTCTTTCCCTACAGACGACGCTCTT | N/A | Biomers |
| P7 indexing primer: CAAGCAGAAGACGGCATACGAGATnnnnnnnGTGACTGGAGTTCAGACGTGT | N/A | Biomers |
| IS7 amplification primer: ACACTCTTTCCCTACAGAC | N/A | Biomers |
| IS8 amplification primer: GTGACTGGAGTTCAGACGTGT | N/A | Biomers |
| **Software and algorithms** | | |
| MergeReadsFastq_cc.py | Kircher | https://bioinf.eva.mpg.de/fastqProcessing/ |
| FilterUniqSAMCons_cc.py | Kircher | https://bioinf.eva.mpg.de/fastqProcessing/ |
| BWA | Li and Durbin | http://bio-bwa.sourceforge.net/ |
| samtools | Li et al. | http://www.htslib.org/ |
| Ry | Skoglund et al. | https://github.com/pontussk/ry_compute |
| GATK | Schmidt | https://gatk.broadinstitute.org/hc/en-us |
| Picard | Broad Institute | https://broadinstitute.github.io/picard/ |
| vcftools | Danecek et al. | http://vcftools.sourceforge.net/ |
| verifyBAMID | Jun et al. | https://genome.sph.umich.edu/wiki/VerifyBamID |
| ANGSD | Korneliusen et al. | http://www.popgen.dk/angsd/index.php/ANGSD |
| READ | Kuhn et al. | https://bitbucket.org/tguenther/read/src/master/ |
| jModeltest | Darriba et al. | https://github.com/ddarriba/jmodeltest2 |
| BEAST | Bouckaert et al. | https://www.beast2.org/ |
| Tracer 1 | Rambaut et al. | http://tree.bio.ed.ac.uk/software/tracer/ |
| smartpca | Patterson et al. | https://github.com/DreichLab/EIG/ |
| ADMIXTURE | Alexander et al. | https://dalexander.github.io/admixture/download.html |
| PLINK | Purcell et al. and Chang et al. | https://www.cog-genomics.org/plink/ |
| pong | Behr et al. | https://github.com/ramachandran-lab/pong |

(Continued on next page)
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. 

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. 

Systematic prehistoric investigations have been undertaken for more than a century, resulting in many archaeological, palaeontological 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.

The human bones from Peștera cu Oase were found without any lithic assemblage and it is impossible to link them to any technocomplex such Aurignacian or Gravettian. The bones were carried by water from outside of the cave along with other animal bones.

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.

Peștera Muierii (The Woman’s Cave, ‘Muier’ 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. 12 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. 38 and Gilbert et al. 95 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. 96,100 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. 35 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. 32-34,101-104 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.
Libraries were amplified in quadruplicate using amplification mastermix and 200 nM each of primer IS7 and ISB, which were set up in duplicates. Libraries were amplified in quadruplicates using amplification protocol as in Günther et al. 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 tapis (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, 150 bp paired-end reads using v2.5 chemistry at the SNP & SEQ Technology Platform, Uppsala University.

**Quantitative and Statistical Analysis**

**Processing of Next Generation Sequence Data**

Adapters where trimmed and pair-end reads were merged using MergeReadsFastq_cc.py, which requires at least 11 base pair overlap between overlapping reads. Bwa 0.7.13 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. 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. to female. To ensure maximal retention of reads all sequencing runs of the same library were merged (using samtools v.0.1.19) before removal of PCR duplicates (per library) with a modified version of FilterUniqSAMCons_cc.py, 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 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 was then used to add read groups and indel realignment was conducted with GATK using the indels identified in phase 1 of the 1000 genomes project (KGP) 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 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. 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. with positions lifted over from hg18 to hg19. Private or near-private alleles (< 5% in 311 modern mtDNAs), 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 (Data S1A).

To estimate autosomal contamination the program VerifyBam was used together with the 1000 genomes reference panel after removing all transversion sites (Data S1B). This left 5,270,838 SNPs. VerifyBam 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 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 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 that 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 and our Pester a Muierii 1, READ 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 Paleolithic individuals, which may lead to overestimation of the relationship between two individuals.65 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 shotgun60 and SNP capture data14 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.66 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 samtools59 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 267 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 1.59

The phylogenetic reanalysis was performed using the complete mitogenomes of 245 modern Homo sapiens, including 61 individuals from the Upper Paleolithic (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 study28 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,112 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 Paleolithic 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 panel112 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. 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) 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). The results were then plotted in R. 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. 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., 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 and the settings krig(theta = 3, m = 1) and predictSurface(extrap = TRUE, nx = 2000, ny = 2000, type = "I"). 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 (af12, Loschbour, LBK, ne1). Notably, PM1 lacks a strong affinity to modern North African populations as suggested by her U6 mitochondrial haplogroup. 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 \( f_3 \) statistics of the form \( f_3(\text{Chimp, Taforalt}; X, PM1) \). 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. 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. 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 \( \alpha \) in the ancient individuals we calculated \( f_4 \) ratios with AdmixTools version 3.0 using the following formula:

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

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 (44x coverage) and Neanderthal Vindija (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.

We also investigated the distribution of Neanderthal related tracts across the genomes of prehistoric Europeans similar to the approach of Sikora et al. and Seguin-Orlando et al. 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. To reduce the effect of sequencing depth, we conducted this analysis on the pseudohaploid data of the prehistoric Europeans. 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
Our goal is to understand where PM1 fits into such models. We use a single Mbuti individual as outgroup and the pre-

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.

In order to perform a phylogenetically more explicit assignment of PM1 into the major pre-LGM European hunter-gatherer gene pools, we used qpGraph (version 6040) of ADMIXTOOLS 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: bigsize: 0.005, lsqmode: YES, initmix: 1000, presicion: 0.0001, intmix: 1000.

Previous studies on European Upper Palaeolithic hunter-gatherers have identified several major gene pools in pre-LGM Europe. Our goal is to understand where PM1 fits into such models. We use a single Mbuti individual as outgroup and the pre-LGM Ust'-Ishim, Kostenki 14, Sunghir III, Goyet Q116-1 as well as the post-LGM Loschbour 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.

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. We only used individuals published as “FullyPublic” and excluded two Khomani San as well as four Ju’hoan North individuals as we ascertainment SNPs in a San individual (see below).

To avoid ascertainment biases, we ascertainment SNPs in a single HGDP San individual sequenced as part of a separate project. The bam file was filtered the same way as all other bam files and we ascertainment 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 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 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 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.\textsuperscript{124,125}

We did not include the insertion polymorphism rs312262906 into the prediction model. The HiFiPlexS 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.\textsuperscript{14,126}

**Medical genetics analysis**

The complete genome sequence from Peşteră Muierii 1, together with the genomes of a few previously published high coverage genomes\textsuperscript{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.\textsuperscript{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://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.\textsuperscript{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\textsuperscript{49} score > 15; c) missense variants with a nucleotide conservation score (vertebrate PhyloP\textsuperscript{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 SS) 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 S1J), 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 homozgyous 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 and which has been associated with cancer development. Polymorphisms in this gene modulate susceptibility to several types of malignancies. 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 and CTTN. 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 however were 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 APLF (p.(His82Tyr)), NOTCH3 (p.(Leu1518Met)), NLPR3 (p.(Gln705Lys)), and APC (p.(Gly2520Ser)). The exact same nucleotide changes in APLF 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.(His900Asp) (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; moreover, the same variant is described 3 times in ExAC in homozygous state (2x in EUR, 1x SA). All together, 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 NLPR3 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 immune 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, Peștera 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. In addition, the Peștera Muierii 1 individuals was carrier of heterozygous IFNG rs2069727, associated with an average cytokine production. All in all, these data suggest that Peștera 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, it is likely that this represents an adaptive state conferring beneficial protective effects.