



OPEN

The mitogenome portrait of Umbria in Central Italy as depicted by contemporary inhabitants and pre-Roman remains

Alessandra Modi^{1,10}, Hovirag Lancioni^{2,10}✉, Irene Cardinali^{2,10}, Marco R. Capodiferro^{3,10}, Nicola Rambaldi Migliore³, Abir Hussein³, Christina Strobl⁴, Martin Bodner⁴, Lisa Schnaller⁴, Catarina Xavier⁴, Ermanno Rizzi⁵, Laura Bonomi Ponzi⁶, Stefania Vai¹, Alessandro Raveane³, Bruno Cavadas^{7,8}, Ornella Semino³, Antonio Torroni³, Anna Olivieri³, Martina Lari¹, Luisa Pereira^{7,8}, Walther Parson^{4,9}, David Caramelli¹ & Alessandro Achilli³✉

Umbria is located in Central Italy and took the name from its ancient inhabitants, the *Umbri*, whose origins are still debated. Here, we investigated the mitochondrial DNA (mtDNA) variation of 545 present-day Umbrians (with 198 entire mitogenomes) and 28 pre-Roman individuals (obtaining 19 ancient mtDNAs) excavated from the necropolis of *Plestia*. We found a rather homogeneous distribution of western Eurasian lineages across the region, with few notable exceptions. Contemporary inhabitants of the eastern part, delimited by the Tiber River and the Apennine Mountains, manifest a peculiar mitochondrial proximity to central-eastern Europeans, mainly due to haplogroups U4 and U5a, and an overrepresentation of J (30%) similar to the pre-Roman remains, also excavated in East Umbria. Local genetic continuities are further attested to by six terminal branches (H1e1, J1c3, J2b1, U2e2a, U8b1b1 and K1a4a) shared between ancient and modern mitogenomes. Eventually, we identified multiple inputs from various population sources that likely shaped the mitochondrial gene pool of ancient *Umbri* over time, since early Neolithic, including gene flows with central-eastern Europe. This diachronic mtDNA portrait of Umbria fits well with the genome-wide population structure identified on the entire peninsula and with historical sources that list the *Umbri* among the most ancient Italic populations.

Due to its acknowledged potential, archaeogenetics is broadly applied to study ancient civilizations, demographic histories and migration events. Markedly, advances in high-throughput genotyping technology have highlighted how the present-day genetic variation of human populations is the outcome of past population movements. In prehistoric times, the Mediterranean area experienced three significant migration waves whose legacy is retrieved in the mitochondrial pool of modern and ancient populations: the Paleolithic hunter-gatherers who survived and re-expanded from glacial refuges, the Neolithic farming societies that moved from the East, and the herders from the Pontic-Caspian steppes inaugurating the Bronze Age^{1–13}.

¹Department of Biology, University of Florence, 50122 Florence, Italy. ²Department of Chemistry, Biology and Biotechnology, University of Perugia, 06123 Perugia, Italy. ³Department of Biology and Biotechnology "L. Spallanzani", University of Pavia, 27100 Pavia, Italy. ⁴Institute of Legal Medicine, Medical University of Innsbruck, 6020 Innsbruck, Austria. ⁵Istituto di Tecnologie Biomediche, CNR, Segrate, 20090 Milan, Italy. ⁶M.A.N.U. National Archeological Museum of Umbria, 06121 Perugia, Italy. ⁷IPATIMUP (Instituto de Patologia e Imunologia Molecular da Universidade do Porto), Porto, Portugal. ⁸i3S (Instituto de Investigação e Inovação em Saúde, Universidade do Porto), 4200-135 Porto, Portugal. ⁹Forensic Science Program, The Pennsylvania State University, University Park, PA 16801, USA. ¹⁰These authors contributed equally: Alessandra Modi, Hovirag Lancioni, Irene Cardinali and Marco R. Capodiferro ✉email: hovirag.lancioni@unipg.it; alessandro.achilli@unipv.it

In this scenario, the Italian Peninsula played a pivotal role in human migrations around the Mediterranean Sea, as testified by the higher degree of its current genomic variability compared with other European populations^{14–19}. This complexity is the result of multifaceted inputs that shaped its gene pool since the Upper Paleolithic. Inferring the contributions of each process is further complicated by similar (or partially overlapping) dispersal patterns from, to and even within the Italian Peninsula, often separated by short time frames. It is generally agreed that the ancestral contribution came from the ancient Italic peoples, among which Latins (also called pre-Romans) achieved a dominant position establishing Roman civilization; whereas the invasions after the fall of the Roman Empire did not significantly alter the peninsular gene pool^{18,19}.

Concerning the phylogeography of Italy, it is difficult to identify a clear genetic pattern able to discriminate southern, northern and central populations in spite of several attempts based on autosomal and uniparental markers^{19–24}. Southern populations were mostly influenced by Greek and Arab colonizations, Northern Italians might reflect admixture with French and German-speaking populations, while Central Italy occupies its own intermediate position creating a continuous cline of variation across the peninsula (with Sardinians as outliers)^{13,19,25–29}. Most of these studies were performed on a large geographic scale producing low-definition results and mainly focusing on modern populations. As for microgeographic studies on Central Italy, only Etruscans (in Tuscany) and Picentes (in Marche) were the target of specific analyses that highlighted their genetic affinity with the current inhabitants^{30–37}. However, Umbria, another crucial region in Central Italy, is still unexplored. The name derives from the ancient Umbrians (or *Umbri*), traditionally considered an indigenous and very old population³⁸. In the first century Common Era (CE) Pliny the Elder stated: “The population of Umbria is considered the oldest of Italy, and it is believed that the Umbrians had been called Ombrikòi by the Greeks because they survived the rains when the earth was flooded” (Pliny the Elder, *Naturalis Historia*, III, 112). Nevertheless, the origin and ethnic affinities of the Umbrians are still in some degree a matter of dispute.

Archaeological and historical data suggest that during the Early Iron Age (ninth/eighth centuries BCE, Before Common Era), Umbrians were among the first communities with strong and well-defined cultural identities in Central Italy, together with Etruscans (to the west), Picentes (to the east) and Samnites (to the south). They originally occupied the eastern part of the today’s Umbria region, placed on the left bank of the Tiber River, soon extending their territories in western Umbria and Tuscany. Around the sixth century BCE, the Etruscans, who had already begun to influence the Umbrian culture, took control over the western territories and the Tiber became the natural border between Umbrians and Etruscans³⁹. The degree of interaction between these ancient populations is still unclear. The Romans came into contact for the first time with the Umbrians during the fourth century BCE and established Latin colonies in the area at the beginning of the third century BCE. After 260 BCE, Umbria was already under the full control of Rome⁴⁰, while the Etruscan culture (and language) disappeared only at the time of the “Social War” (90–88 BCE) with the attribution of Roman citizenship to all Italic people⁴¹. Nowadays Umbria is somewhat smaller than ancient Umbria, but its inhabitants still preserve significant difference in the dialects spoken on the two banks of the Tiber⁴².

An important necropolis in East Umbria is placed in the so-called *Plestinam Paludem* (now Colfiorito, located at 760 m above the sea level up in the Apennines). The *Plestini* plateaus represented an obligatory way in the trans-Appennine routes, but stable settlements have not been attested before the beginning of the Iron Age^{43,44}. The geographical position, the wealth of water, the possibilities offered by the exercise of hunting and fishing, the goodness of the pastures and the abundance of timber have undoubtedly encouraged the stabilization and growth of the population during the Iron Age.

In this study, we report 198 entire mitogenomes from modern Umbrians (191 here sequenced for the first time), selected from a larger dataset of 545 samples covering the entire region, as well as the mitogenomes of 19 Iron Age *Umbri Plestini*, who were buried in *Plestinam Paludem* (Fig. 1 and Supplementary Fig. S1). This diachronic approach allowed us to study the mitochondrial DNA (mtDNA) variation (at the highest-resolution level) in a microgeographic context and to obtain new insights concerning the maternal genetic history of Umbria, a region often defined as the “Heart of Italy” because of its location.

Results and discussion

Mitochondrial variation of modern Umbrians. *Control-region data.* Through the analysis of the control-region sequence of 545 modern Umbrians (Supplementary Dataset S1), it was possible to identify a high haplotype diversity ($Hd = 0.994$) that, compared to other Eurasian and North African populations²¹, confirms the goodness of the sampling and testifies for an extensive maternal admixture (Supplementary Fig. S2). In order to verify if this variability is equally distributed within the region without any sub-population differentiation, we estimated pairwise fixation index (*Fst*) values in six sub-areas, considering geographic and historical criteria (north, south, west, center, center-east and east; Fig. 1), showing that inhabitants from eastern Umbria are genetically the most distant from the other sub-groups (Fig. 2). This high differentiation of the eastern part of Umbria suggests a distinctiveness in its ancient or recent history compared to the rest of the region.

Phylogenetic analyses were then performed. The mutational motifs of the 545 Umbrians clustered into 369 haplotypes belonging to numerous haplogroups and sub-haplogroups when using Haplogrep 2.0 and SAM 2 on EMPOP (Supplementary Dataset S1). As expected, most (97%) are members of typical western Eurasian branches. Initially, we compared macro-haplogroup distributions among the six established sub-regions identifying two significant differences in haplogroups J, which is particularly common (30%) in East Umbria, and K, with a rather high incidence (17%) in South Umbria (Supplementary Fig. S3A). In order to summarize the information embedded in these haplogroups, we performed a principal component analysis (PCA, Fig. 3) including the six Umbrian sub-regions and the Eurasian dataset previously used to analyze the neighboring Tuscany region³⁰. The relatedness of different parts of Umbria with typical Mediterranean populations can be clearly appreciated in

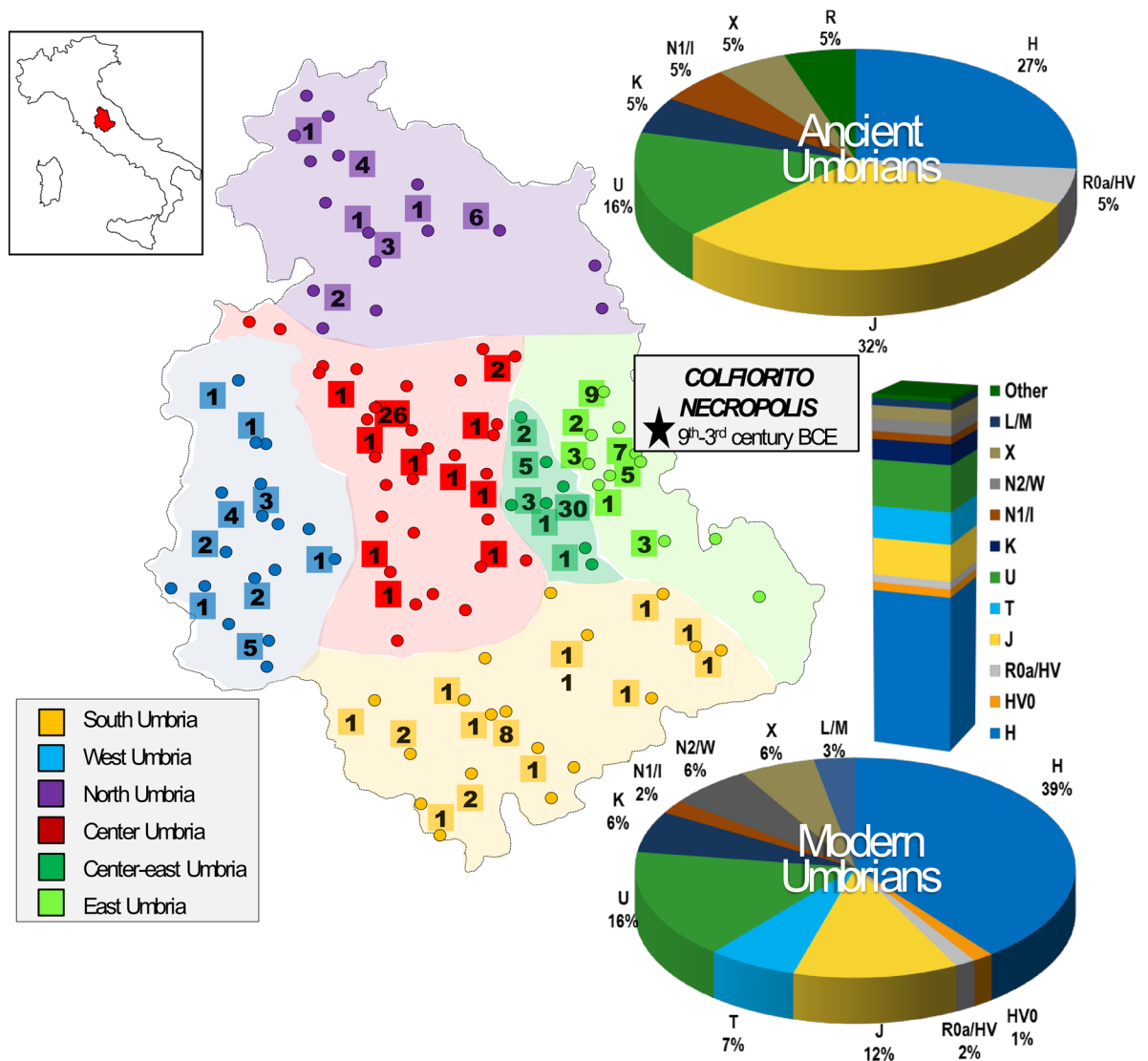


Figure 1. Geographic origin of ancient and modern Umbrians analyzed in this study. The six established sub-regions are symbolized by different colors. Dots mark the geographic origin of all modern samples (N=545, see Supplementary Dataset S1); those completely sequenced are reported in squares (N=198, see Supplementary Dataset S2). Pie charts summarize haplogroup distributions (based on Haplogrep) considering complete mitogenomes of ancient (N=19, see Supplementary Dataset S3) and modern samples, while the bar plot represents control-region data of the overall modern sample. The location of the Colfiorito necropolis is indicated by a star.

the middle portion of the plot. However, East Umbria clusters together with eastern European countries. Major contributions to this clustering come from haplogroups U4 and U5a, which show high frequencies in central-eastern Europe (inset of Fig. 3). Notably, two of their sub-branches (U4a and U5a1) have been also identified in *Yamnaya* samples² as well as in Mesolithic samples from northern and eastern Europe (Reich database V42.4; <https://reich.hms.harvard.edu>).

Complete mitogenome data. Taking the population density into account, we randomly selected samples (from 19 to 42) from each of the six regional divisions for complete mtDNA sequencing. With this approach we obtained 191 novel mitogenomes (Supplementary Dataset S2), selected considering only geographic criteria without any phylogenetic bias.

It is worth mentioning that we did not notice any difference when comparing the two NGS methodologies used to generate the complete mitogenomes. To check if any ascertainment bias was present, we performed a Site Frequency Spectrum (SFS) analysis, using the two methodologies as “artificial populations” and comparing the distributions of variant occurrences in the two datasets. As shown in Supplementary Figure S4, we observed a comparable amount of singletons and doubletons, which are used as indicators of possible inconsistencies.

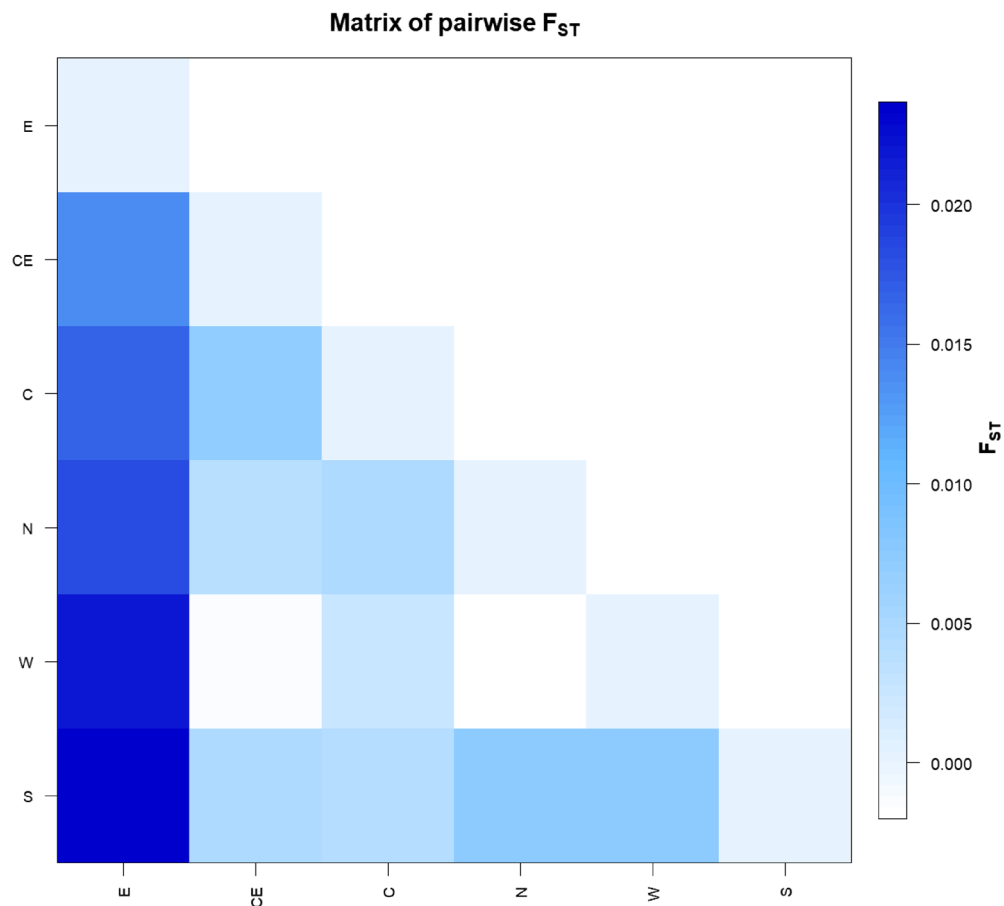


Figure 2. Pairwise population genetic distances. Plot of pairwise population genetic distances between the six established sub-regions of Umbria (E = east; CE = center-east; C = center; N = north; W = west; S = south), based on control-region data (n = 480).

Our mitogenomes, together with seven GenBank records (189 haplotypes in total), were classified into different sub-haplogroups (147 with Haplogrep and 137 with EMPOP). The frequencies of major haplogroups widely overlap with those obtained from the control-region dataset, without any significant differences (p value 0.57), thus confirming that even the 198 complete mitogenomes can be accounted as a population dataset representative of modern Umbrians. Moreover, also the macro-haplogroup distributions in the six sub-regions showed the same pattern of the control-region data, confirming significant differences only for haplogroups J and K in East and South Umbria, respectively (Supplementary Fig. S3B). On the other hand, the importance of complete mitogenome sequencing is confirmed by the increased haplotype diversity value (from 0.994 to 0.999) as well as by the accuracy of the sub-haplogroup classification, which was improved for more than 70% of haplotypes (76% for Haplogrep, 72% for EMPOP; Supplementary Dataset S2).

MtDNA variation of ancient Umbrians. Using NGS technology combined with target enrichment⁴⁵, we tried to reconstruct the mitogenomes of 28 pre-Roman samples from the necropolis of *Plestia*, located in East Umbria (Fig. 1 and Supplementary Fig. S1). Four direct radiocarbon dates confirmed the age estimated from the archaeological context placing the remains at the end of the seventh cal. century BCE (Supplementary Fig. S5). Eventually, four of the 28 samples did not amplify at all, while five produced ambiguous sequencing results that did not reach the standard quality requested to guarantee the reliability of NGS data (Supplementary Fig. S1). The final dataset of 19 ancient mitogenomes showed a depth of average coverage ranging from 5.86× to 50.98× (Supplementary Dataset S3). The damage pattern and average fragment size were used in an iterative probabilistic approach that jointly estimates modern human contaminations and reconstructs the endogenous mtDNA sequence⁴⁶. Nucleotide misincorporations and fragmentation patterns were compatible with the sample age⁴⁷, ranging between 16.7 and 42.1% at 5' molecule termini and 60.57–100.41 bps, respectively. In addition, no significant levels of contamination were detected.

The 19 mtDNA sequences were classified into 17 mitochondrial haplogroups and eight super-haplogroups. They are all typical of present-day West-Eurasian populations with the most represented lineage being J (32%), followed by H (26%) and U (16%) (Figs. 1, 4). A similar H frequency (~30%) was observed in modern samples from the eastern part of the region. Haplogroup H is the most frequent in Europe (~40%) with a declining pattern from western Europe towards the Near East and Caucasus (~10–20%), but without any conclusive scenario

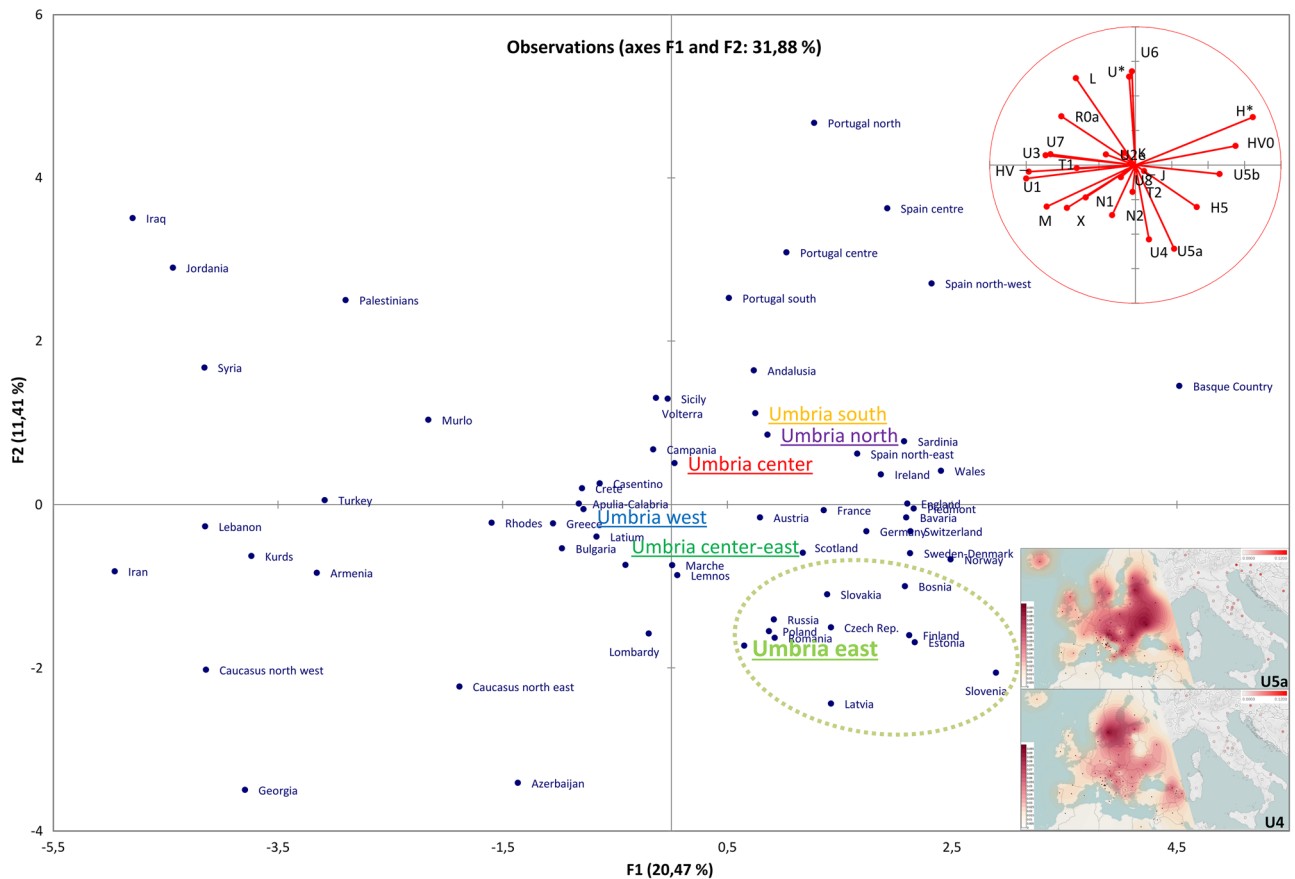


Figure 3. PCA plot. The genetic landscape of Eurasia based on haplogroup frequencies from control-region data. The inset shows the frequency distributions of U4 and U5a in western Eurasia (left side) as well as in Italy (right side) constructed with Tableau 2019.3.0 (www.tableau.com).

about its still enigmatic origin⁴⁸. Regarding the most represented haplogroup J (three mitogenomes belonging to different subsets of J1c3), it has been proposed that most of its subgroups diversified in the Near East during the Last Glacial Maximum (LGM) and spread into Europe in the Late Glacial⁴⁹. Some J1c sub-lineages have been also proposed as Early Neolithic founder lineages^{5,50}. As for super-haplogroup U, four sub-haplogroups were detected, including U4, the same lineage that pushes modern eastern Umbrians close to central-eastern Europeans in the PCA.

The incidence of each major haplogroup identified in our ancient sample is comparable with the one observed in present-day Umbrians (p value 0.33). However, the high frequency of haplogroup J in ancient Umbrians (32%) can currently be observed only in the eastern part of the region (30%). Virtually all lineages (except for the paragroups J* and R*) identified in pre-Roman remains are still recognizable nowadays in Umbria, thus suggesting a possible genetic continuity since pre-Roman times (Supplementary Dataset S3). We attempted to verify this continuity on a phylogenetic tree encompassing modern (198) and ancient (19) mitogenomes from Umbria (Fig. 4 and Supplementary Fig. S6). Firstly, the demographic change in the population size depicted by the Bayesian Skyline Plot (BSP) confirms the typical trend of European populations with two sharp increases dated to Paleolithic (from ~ 40 kya) and Neolithic (from ~ 10 kya) ages. Moreover, the age estimates of the major branches overlap with previously reported confidence intervals^{50,51}. Even if we did not pinpoint any haplotype identities between modern and ancient samples, about half of the ancient samples share terminal branches (six clades in total: H1e1, J1c3, J2b1, U2e2a, U8b1b1 and K1a4a) with modern Umbrians, all dated back to the Holocene (Figs. 4, 5). We searched public databases for ancient mtDNAs belonging to these lineages identifying 225 ancient mitogenomes from samples excavated in different western Eurasian regions and in northern Africa and dated to prehistoric and historic periods, as shown by the geographic/temporal maps of these sub-lineages (Fig. 6 and Supplementary Dataset S4). J1c3g could be considered a paradigmatic example of these heterogeneous genetic connections, as attested by its aDNA tree, which includes our sample (aUMB050) and other eight ancient mitogenomes from public databases (inset of Fig. 5). Two of these are Bronze Age samples, one from Ukraine⁶ and one from southeastern Poland⁵². Other two burials were excavated in southern Bavaria (Germany), one associated to the early Bronze Age and the other to a Bell Beaker Complex⁵³. The latter sample is at the root of the reconstructed J1c3g tree, which has been dated to 5.4 ± 0.3 kya. Four more recent J1c3g mtDNAs have been also identified in one individual from Spain dated to the sixth century CE and archaeologically interpreted as a

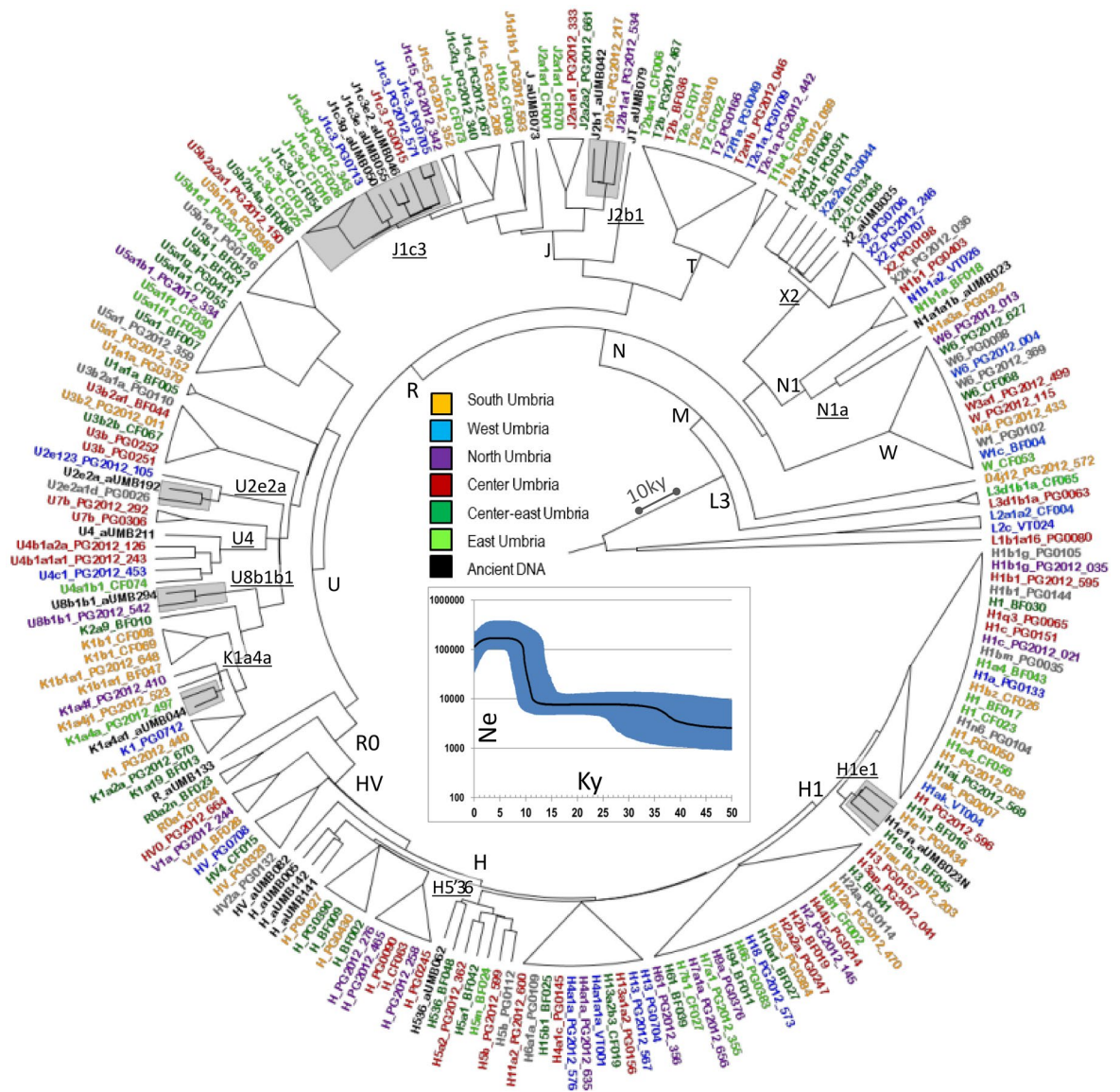


Figure 4. Schematic phylogenetic tree of modern and ancient Umbrian mitogenomes. The terminal branches shared between ancient and modern mtDNAs are shaded. Branch lengths are drawn to scale based on Bayesian time estimates. The inset shows a Bayesian Skyline Plot (BSP) analysis of Umbrian mitogenomes. See Supplementary Figure S6 for details.

Visigoth⁵⁴, one Hungarian conqueror⁵⁵ and a pre-Christian Icelander⁵⁶, both from the early tenth century CE, and a medieval sample from Denmark⁵⁷.

Conclusions

Surrounded by the Mediterranean Sea and bounded by the Alps, Italy extends over more than 1,000 km along a North–South axis and includes the two largest islands of the Mediterranean Sea, Sicily and Sardinia. The combination of this geographic complexity with a rich set of historical events and cultural dynamics had the potential to shape in a unique way the distribution of genetic variation within the Italian populations. Local peculiarities have been highlighted by analyzing the mitogenome variation of specific regions, e.g. Marche, Piedmont, Tuscany and Sardinia^{21,36,37,58}. However, a fine and exhaustive microgeographic characterization of other regions has yet to be conducted.

In this study, we describe for the first time the mtDNA variation of the current Umbrian population by analyzing 545 samples covering the entire region. Upon evaluating the genealogical information collected during the sampling campaigns, we reallocated the samples, based on their terminal maternal ancestors, into six sub-areas (north, south, west, center, center-east and east) drawn by geographic criteria and historical/cultural information. A wide range of haplotypes, mostly belonging to western Eurasian haplogroups (97%), testify for the high mtDNA diversity in Umbria. The incidence of these lineages across the region is quite homogeneous with the notable exception of haplogroup K, reaching the highest frequency (17%) in South Umbria, and haplogroup J,

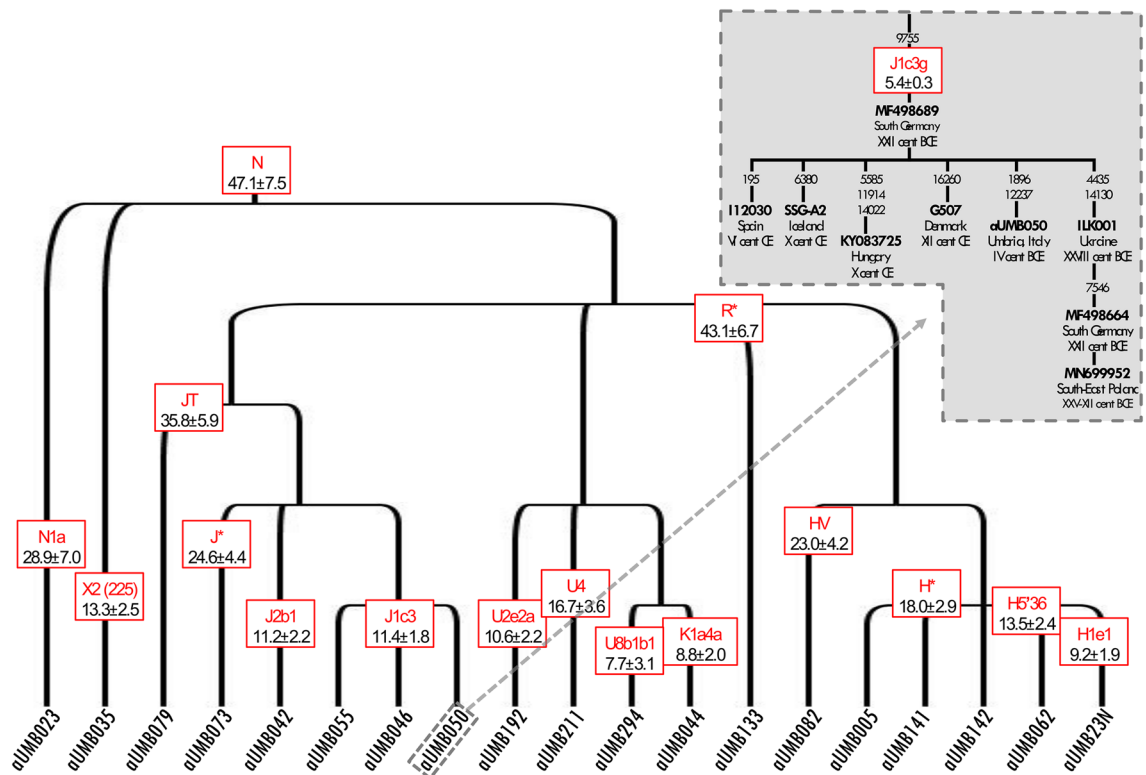


Figure 5. Schematic phylogeny of ancient Umbrians. Bayesian ages refer to the MRCA shared with modern Umbrians. The inset highlights the closeness of the Umbrian J1c3g mtDNA to other available ancient samples.

which encompasses 30% of current inhabitants of the eastern area. In the western Eurasian PCA plot, the latter sub-region is pushed close to populations from central-eastern Europe by haplogroups U4 and U5a that show high frequencies in those areas.

Then, we extended our analyses to complete mitogenomes (191 sequenced for the first time), randomly selecting the targeted samples to avoid phylogenetic biases and to maintain the population-wide characteristics of our dataset. This higher level of resolution allowed us to refine the haplogroup affiliation in more than 70% of the samples and to make a diachronic comparison with 19 ancient mitogenomes from *Umbri Plestini*. These pre-Roman samples were classified into the same haplogroups identified in contemporary inhabitants. Moreover, the six terminal branches (H1e1, J1c3, J2b1, U2e2a, U8b1b1 and K1a4a) shared between ancient and modern mitogenomes suggest a genetic continuity in the region during the Holocene. These specific lineages were also identified in a wide range of available ancient samples outside the region, including Neolithic Mediterranean remains as well as *Yamnaya*, Bell Beaker and more recent samples from central-eastern Europe. These variegated connections are summarized by the lineage geographic/temporal patterns and are specifically shown by the J1c3g ancient mtDNA tree dated between the Late Neolithic and the Early Bronze Age.

In brief, it is apparent that distinctive mtDNA variants have been brought into the region by the ancestors of *Umbri Plestini* and preserved in some, perhaps more isolated, sub-areas. These ancestors reached Umbria coming from various population sources at different times during the Holocene, from early Neolithic farmers spreading across the Mediterranean to Bronze Age and Medieval connections with central-eastern Europeans, possibly including few nomadic groups (*Yamnaya*) from the Pontic-Caspian steppes. This microgeographic and diachronic mtDNA portrait of Umbria fits well with recent genetic data on the entire peninsula. The Y-chromosome counterpart pointed to different male ancestries for the Italian populations²⁴ and the autosomal data revealed several ancient signatures and the largest degree of population structure detected so far in Europe^{19,29}. Notably, two of the three published genomic clusters (Sardinia, Northern and Southern Italy) overlap in Central Italy and precisely in Umbria, the “Heart of Italy”. In a wider multidisciplinary context, this hypothesis is also supported by historical sources that list the *Umbri* among the most ancient Italic populations^{38–40} and by the assumed Indo-European origin of their language, distinct from the Etruscan one spoken by neighboring people during the Iron Age⁵⁹.

Materials and methods

Modern Umbrians. *Sample collection.* The modern collection consisted of 538 DNA samples from healthy and unrelated subjects with an Umbrian maternal grandmother as a terminal maternal ancestor. Swab or mouth-wash rinsing samples were collected from volunteers, representing the entire Umbrian area. Written informed consents were obtained from all donors, who provided information about place of birth and geographical origins up to three generations of Umbrian maternal ancestry. Total DNA was extracted with the MagCore Automated Nucleic Acid Extractor following manufacturer’s protocols. Seven additional Umbrian samples, collected and sequenced in our labs for previous projects^{60,61}, were also included.

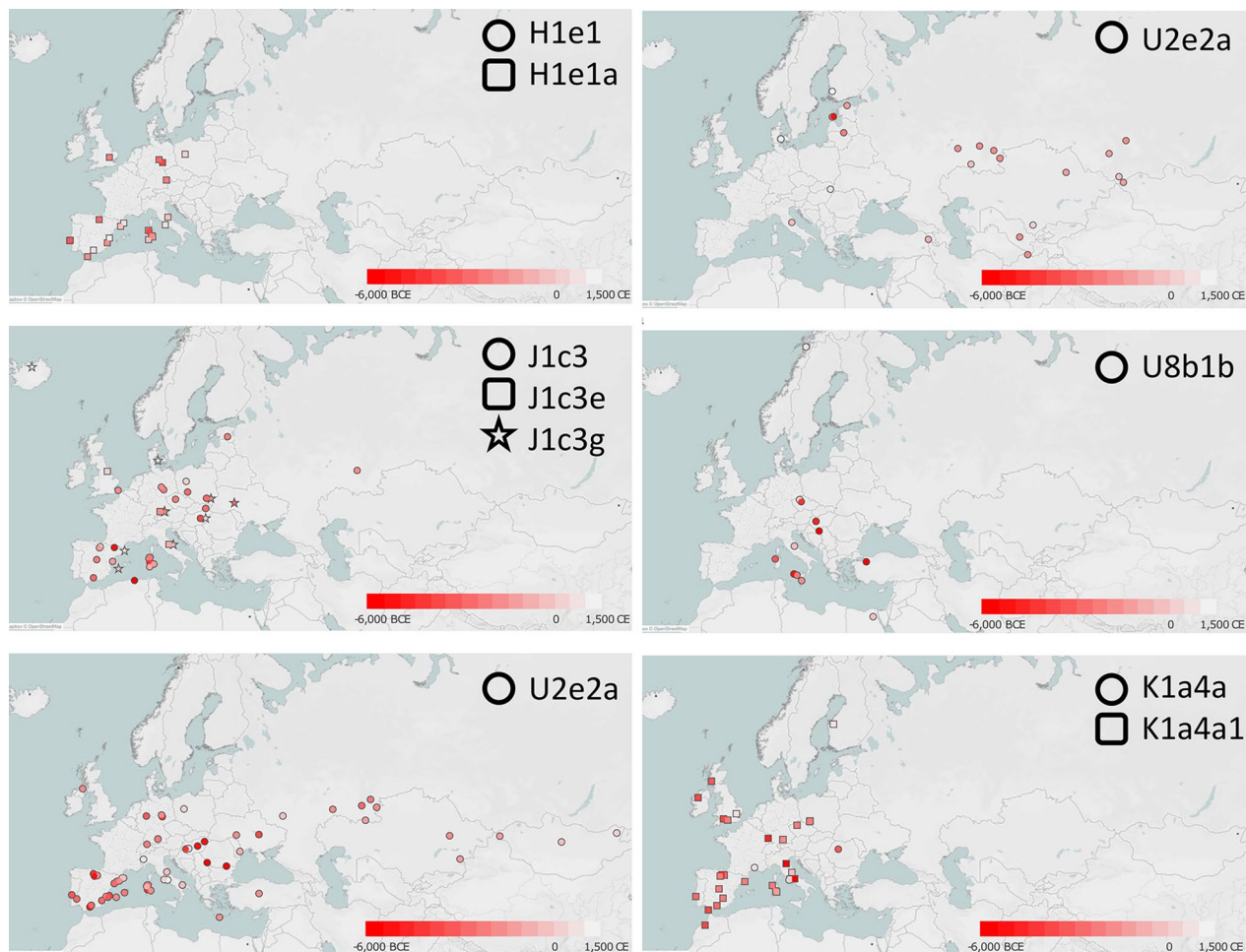


Figure 6. Maps of ancient mitogenomes from literature belonging to the six terminal branches shared between ancient and modern Umbrians. Within each of the six branches, only sub-clades identified in Ancient Umbrians are reported. See Supplementary Dataset S4 for details. Maps were constructed with Tableau 2019.3.0 (www.tableau.com).

All analyses were carried out in accordance with relevant guidelines and regulations, and all experimental protocols were approved by the Ethics Committee for Clinical Experimentation of the University of Perugia (protocol no. 2017-01).

Geographical division. Umbria was divided into six sub-areas (highlighted in different colors in Fig. 1) considering geographic criteria as well as historical and cultural information. The northern and southern areas are geographically and traditionally linked to Tuscany and Latium, respectively. The hilly lands to the west, including “Monte Peglia” and Orvieto, were part of Etruria. Eastern Umbria is characterized by high mountains (the Apennines) where ancient Umbrians settled for centuries having extensive exchanges with the neighboring Marche populations. Lastly, we decided to divide the vast and flat central area into two sub-regions, here called center and center-east, which are delimited by the Tiber and Topino rivers, respectively. The central area includes cities of known Etruscan origins, such as Bettona, Perugia and Todi. In particular, the name Todi means “border” and, even if it was founded by ancient Umbrians, the city was located at the border with the Etruscan territories and was still under their influence when it was conquered by the Romans. On the contrary, central-eastern Umbria, also known as “Valle Umbra”, includes ancient villages such as Assisi, Bevagna, Spello and the modern municipality of Foligno. Historically, these cities experienced intensive exchanges with eastern Umbria, as testified for instance by two ancient roads, Via Plestina (from Foligno) and Via della Spina (from Spoleto).

Control-region sequencing. Novel mitochondrial control-region sequences were generated through standard PCR and Sanger sequencing method³⁰, then assembled and aligned to the revised Cambridge Reference Sequence (rCRS; NC_012920.1)⁶² using Sequencher 5.10 (Gene Codes Corporation). These were analyzed together with the control-region sequences from the 191 complete genomes (see below) and seven previously published, for an overall number of 545 control regions (Supplementary Dataset S1).

Complete mitogenome sequencing. The entire mitogenome of six present-day samples was sequenced using the classic PCR-Sanger system⁶³, while 185 mitogenomes were obtained by employing two Next Generation Sequencing (NGS) techniques: 82 by the Illumina MiSeq⁶⁴ and 103 through the Ion PGM System⁶⁵ (Supplementary Dataset S2).

FASTQ files were aligned to the reference sequence (rCRS; NC_012920.1) using BWA⁶⁶, the bam files were then filtered and sorted with SAMtools⁶⁷. The variants were called employing HaplotypeCaller implemented in GATK (with ploidy flag set as 1)⁶⁸ and filtered using BCFtools to obtain the final SNP dataset. Three different in-house scripts (HeteroSeek, HaploCreate and HaploCreateBelow, developed at the IPATIMUP Institute) were used to obtain the final haplotypes (both with and without heteroplasmies). The final haplotypes were also double-checked through a manual visualization of the bam files with the Integrative Genomics Viewer (IGV) software. Common criteria used for calling mtDNA variants were adopted as reported by Olivieri and colleagues⁵⁸. In addition, some problematic fragments were replicated by Sanger sequencing and the congruence with the initial control-region data was evaluated.

Ancient Umbrians. Ancient sample collection. We analyzed the remains of 28 individuals excavated from the necropolis of *Plestia* in Colfiorito (East Umbria, Central Italy, Fig. 1), in which more than 250 tombs have been identified. According to funerary rites and grave goods, the necropolis was dated from the early ninth to the late third century BCE and provided a greater understanding of the life and culture of the ancient Umbrian civilization (see Supplementary Figure S1 and Supplementary Text for further details). Direct radiocarbon dating on the skeletal remains of four individuals was performed in outsourcing at the Curt-Engelhorn-Centre for Archaeometry (Mannheim, Germany).

Ancient mitogenome sequencing. Molecular analysis of the archaeological specimens was performed under sterile conditions in a dedicated ancient DNA (aDNA) facility at the Laboratory of Molecular Anthropology and Paleogenetics (University of Florence, Italy), following strict guidelines and standard precautions to avoid contaminations. After a silica-based DNA extraction⁶⁹ and libraries preparation⁷⁰, ancient mitogenomes were captured and sequenced on the Illumina MiSeq platform at the Institute of Biomedical Technologies, National Research Council (Segrate, Milano, Italy), as previously reported⁷¹.

After demultiplexing, raw reads were analyzed using a specific pipeline developed for aDNA. The EAGER pipeline⁷² was used for initial sequencing quality control, adapter trimming and paired-end read merging. Merged reads were filtered for a minimum length of 30 base pairs and mapped to rCRS (NC_012920.1) using CirculaMapper (BWA parameters: $-n\ 0.02, -l\ 16,500$), a tool integrated in EAGER and specifically designed for the analysis of circular reference genomes. After removing PCR duplicate, only reads with a map quality score ≥ 30 were retained and used for reconstructing mtDNA consensus sequences using *schmutzi* (parameters: $-\text{logindel}\ 1 - \text{uselength}$)⁴⁶. Bases with individual likelihood < 20 were considered as unassigned positions (Ns). Present-day human contamination was evaluated by an iterative likelihood method implemented in *schmutzi* using a non-redundant database of 197 human mitochondrial genomes available in the software package. Damage patterns at the ends of the molecules were calculated using contDeam, a program provided with the *schmutzi* package.

Phylogenetic and statistical methods. Several mtDNA sequence variation parameters were estimated using DnaSP 5.1 software⁷³. Intra- and inter-population comparisons based on the number of pairwise differences between sequences were performed using an Arlequin integrated R script⁷⁴.

Haplogroups were predicted using HaploGrep2 software⁷⁵, but the initial classification was revised and manually updated in agreement with PhyloTree build 17⁷⁶ and SAM 2⁷⁷ on EMPOP⁷⁸.

All (modern and ancient) haplotypes underwent a posteriori mtDNA sequence data quality control using EMPcheck, a tool to perform plausibility checks on a rCRS-coded data table (<https://empop.online/tools>).

In order to graphically display (and summarize) the relationships among the analyzed mtDNAs, Principal Component Analyses (PCA) were also performed using Excel software implemented by XLSTAT, as previously described³⁰. Spatial frequency distribution plots were constructed with the program Tableau 2019.3.0. Finally, after purging all positions containing gaps and ambiguous data, a maximum parsimony tree was built with mtPhyl v.5.003, while time estimates and demographic trends were evaluated using BEAST v2.6.1 (Bayesian Evolutionary Analysis of Sampling Trees), as previously reported⁵⁸.

Data availability

All novel sequences have been deposited in GenBank under accession numbers: MN686759-MN687105 for 347 mitochondrial control-region sequences from modern samples; MN687107-MN687297 for 191 complete mitochondrial sequences from modern samples; MN687298-MN687316 for 19 complete mitochondrial sequences from ancient samples. The data will be available from the EMPOP mtDNA population database (<https://empop.online/>) under accession numbers EMP00826 (control-region data) and EMP00827 (mitogenomes).

Received: 22 February 2020; Accepted: 22 May 2020

Published online: 01 July 2020

References

1. Haak, W. *et al.* Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS Biol.* **8**, e1000536. <https://doi.org/10.1371/journal.pbio.1000536> (2010).
2. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211. <https://doi.org/10.1038/nature14317> (2015).

3. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257. <https://doi.org/10.1038/ncomms6257> (2014).
4. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167. <https://doi.org/10.1038/nature14507> (2015).
5. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503. <https://doi.org/10.1038/nature16152> (2015).
6. Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197–203. <https://doi.org/10.1038/nature25778> (2018).
7. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. USA* **113**, 6886–6891. <https://doi.org/10.1073/pnas.1523951113> (2016).
8. Omrak, A. *et al.* Genomic evidence establishes anatolia as the source of the European neolithic gene pool. *Curr. Biol.* **26**, 270–275. <https://doi.org/10.1016/j.cub.2015.12.019> (2016).
9. Olalde, I. *et al.* Erratum: The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 543. <https://doi.org/10.1038/nature26164> (2018).
10. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424. <https://doi.org/10.1038/nature19310> (2016).
11. Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218. <https://doi.org/10.1038/nature23310> (2017).
12. Lazaridis, I. The evolutionary history of human populations in Europe. *Curr. Opin. Genet. Dev.* **53**, 21–27. <https://doi.org/10.1016/j.gde.2018.06.007> (2018).
13. De Angelis, F. *et al.* Mitochondrial variability in the Mediterranean area: A complex stage for human migrations. *Ann. Hum. Biol.* **45**, 5–19. <https://doi.org/10.1080/03014460.2017.1416172> (2018).
14. Di Gaetano, C. *et al.* An overview of the genetic structure within the Italian population from genome-wide data. *PLoS ONE* **7**, e43759. <https://doi.org/10.1371/journal.pone.0043759> (2012).
15. Boattini, A. *et al.* Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS ONE* **8**, e65441. <https://doi.org/10.1371/journal.pone.0065441> (2013).
16. Sarno, S. *et al.* An ancient Mediterranean melting pot: Investigating the uniparental genetic structure and population history of sicily and southern Italy. *PLoS ONE* **9**, e96074. <https://doi.org/10.1371/journal.pone.0096074> (2014).
17. Pereira, J. B. *et al.* Reconciling evidence from ancient and contemporary genomes: A major source for the European Neolithic within Mediterranean Europe. *Proc. Biol. Sci.* **284**, 1851. <https://doi.org/10.1098/rspb.2016.1976> (2017).
18. Antonio, M. L. *et al.* Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714. <https://doi.org/10.1126/science.aay6826> (2019).
19. Raveane, A. *et al.* Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *Sci Adv* **5**, eaaw3492. <https://doi.org/10.1126/sciadv.aaw3492> (2019).
20. Brisighelli, F. *et al.* Uniparental markers of contemporary Italian population reveals details on its pre-Roman heritage. *PLoS ONE* **7**, e50794. <https://doi.org/10.1371/journal.pone.0050794> (2012).
21. Vai, S. *et al.* Genealogical relationships between early medieval and modern inhabitants of Piedmont. *PLoS ONE* **10**, e0116801. <https://doi.org/10.1371/journal.pone.0116801> (2015).
22. Sazzini, M. *et al.* Complex interplay between neutral and adaptive evolution shaped differential genomic background and disease susceptibility along the Italian peninsula. *Sci. Rep.* **6**, 32513. <https://doi.org/10.1038/srep32513> (2016).
23. Parolo, S. *et al.* Characterization of the biological processes shaping the genetic structure of the Italian population. *BMC Genet.* **16**, 132. <https://doi.org/10.1186/s12863-015-0293-x> (2015).
24. Grugni, V. *et al.* Reconstructing the genetic history of Italians: New insights from a male (Y-chromosome) perspective. *Ann. Hum. Biol.* **45**, 44–56. <https://doi.org/10.1080/03014460.2017.1409801> (2018).
25. Fiorito, G. *et al.* The Italian genome reflects the history of Europe and the Mediterranean basin. *Eur. J. Hum. Genet.* **24**, 1056–1062. <https://doi.org/10.1038/ejhg.2015.233> (2016).
26. Amorim, C. E. G. *et al.* Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun.* **9**, 3547. <https://doi.org/10.1038/s41467-018-06024-4> (2018).
27. Vai, S. *et al.* A genetic perspective on Longobard-Era migrations. *Eur. J. Hum. Genet.* **27**, 647–656. <https://doi.org/10.1038/s41431-018-0319-8> (2019).
28. Tamm, E. *et al.* Genome-wide analysis of Corsican population reveals a close affinity with Northern and Central Italy. *Sci. Rep.* **9**, 13581. <https://doi.org/10.1038/s41598-019-49901-8> (2019).
29. Sazzini, M. *et al.* Genomic history of the Italian population recapitulates key evolutionary dynamics of both Continental and Southern Europeans. *BMC Biol.* **18**, 51. <https://doi.org/10.1186/s12915-020-00778-4> (2020).
30. Achilli, A. *et al.* Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans. *Am. J. Hum. Genet.* **80**, 759–768. <https://doi.org/10.1086/512822> (2007).
31. Guimaraes, S. *et al.* Genealogical discontinuities among Etruscan, Medieval, and contemporary Tuscans. *Mol. Biol. Evol.* **26**, 2157–2166. <https://doi.org/10.1093/molbev/msp126> (2009).
32. Ghirotto, S. *et al.* Origins and evolution of the Etruscans' mtDNA. *PLoS ONE* **8**, e55519. <https://doi.org/10.1371/journal.pone.0055519> (2013).
33. Tassi, F., Ghirotto, S., Caramelli, D. & Barbujani, G. Genetic evidence does not support an Etruscan origin in Anatolia. *Am. J. Phys. Anthropol.* **152**, 11–18. <https://doi.org/10.1002/ajpa.22319> (2013).
34. Pardo-Seco, J., Gómez-Carballa, A., Amigo, J., Martínón-Torres, F. & Salas, A. A genome-wide study of modern-day Tuscans: Revisiting Herodotus's theory on the origin of the Etruscans. *PLoS ONE* **9**, e105920. <https://doi.org/10.1371/journal.pone.0105920> (2014).
35. Gómez-Carballa, A., Pardo-Seco, J., Amigo, J., Martínón-Torres, F. & Salas, A. Mitogenomes from the 1000 genome project reveal new near Eastern features in present-day Tuscans. *PLoS ONE* **10**, e0119242. <https://doi.org/10.1371/journal.pone.0119242> (2015).
36. Serventi, P. *et al.* Iron Age Italic population genetics: The Piceni from Novilara (8th–7th century BC). *Ann. Hum. Biol.* **45**, 34–43. <https://doi.org/10.1080/03014460.2017.1414876> (2018).
37. Leonardi, M. *et al.* The female ancestor's tale: Long-term matrilineal continuity in a nonisolated region of Tuscany. *Am. J. Phys. Anthropol.* **167**, 497–506. <https://doi.org/10.1002/ajpa.23679> (2018).
38. Galiberti, A. In *XXIII Riunione Scientifica Il Paleolitico inferiore in Italia*. 147–163.
39. Pallottino, M. *Genti e Culture dell'Italia Preromana* (Jouvence, Rhone-Alpes, 1981).
40. Bradley, G. *Ancient Umbria: State, Culture, and Identity in Central Italy from the Iron Age to the Augustan Era* (Oxford University Press, Oxford, 2000).
41. Rasmussen, T. Urbanization in Etruria. In *Mediterranean Urbanization (600–800 BC)*. (eds. Osborne, R. & Cunliffe, B.) 91–113 (2004).
42. Mattesini, E. *I dialetti italiani. Storia, struttura, uso* (UTET, Uttarakhand, 2002).
43. Bonomi Ponzi, L. *La Necropoli Plestina di Colfiorito di Foligno* (Quattroemme, Rome, 1997).
44. Agnoletti, M. *Italian Historical Rural Landscapes: Cultural Values for the Environment and Rural Development* (Springer, New York, 2012).
45. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE* **5**, e14004. <https://doi.org/10.1371/journal.pone.0014004> (2010).

46. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224. <https://doi.org/10.1186/s13059-015-0776-0> (2015).
47. Sawyer, S., Krause, J., Guschanski, K., Savolainen, V. & Pääbo, S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* **7**, e34131. <https://doi.org/10.1371/journal.pone.0034131> (2012).
48. Richards, M. B., Soares, P. & Torroni, A. Palaeogenomics: Mitogenomes and migrations in Europe's past. *Curr. Biol.* **26**, R243–246. <https://doi.org/10.1016/j.cub.2016.01.044> (2016).
49. Pala, M. *et al.* Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am. J. Hum. Genet.* **90**, 915–924. <https://doi.org/10.1016/j.ajhg.2012.04.003> (2012).
50. Pereira, J. B. *et al.* Reconciling evidence from ancient and contemporary genomes: A major source for the European Neolithic within Mediterranean Europe. *Proc. Biol. Sci.* <https://doi.org/10.1098/rspb.2016.1976> (2017).
51. Soares, P. *et al.* Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759. <https://doi.org/10.1016/j.ajhg.2009.05.001> (2009).
52. Juras, A. *et al.* Mitochondrial genomes from Bronze Age Poland reveal genetic continuity from the Late Neolithic and additional genetic affinities with the steppe populations. *Am. J. Phys. Anthropol.* <https://doi.org/10.1002/ajpa.24057> (2020).
53. Knipper, C. *et al.* Female exogamy and gene pool diversification at the transition from the Final Neolithic to the Early Bronze Age in central Europe. *Proc. Natl. Acad. Sci. USA* **114**, 10083–10088. <https://doi.org/10.1073/pnas.1706355114> (2017).
54. Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234. <https://doi.org/10.1126/science.aav4040> (2019).
55. Neparáczki, E. *et al.* Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing. *PLoS ONE* **12**, e0174886. <https://doi.org/10.1371/journal.pone.0174886> (2017).
56. Eneaszós, S. S. *et al.* Ancient genomes from Iceland reveal the making of a human population. *Science* **360**, 1028–1032. <https://doi.org/10.1126/science.aar2625> (2018).
57. Krause-Kyora, B. *et al.* Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. *Nat. Commun.* **9**, 1569. <https://doi.org/10.1038/s41467-018-03857-x> (2018).
58. Olivieri, A. *et al.* Mitogenome diversity in sardinians: A genetic window onto an Island's past. *Mol. Biol. Evol.* **34**, 1230–1239. <https://doi.org/10.1093/molbev/msx082> (2017).
59. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World* (Princeton University Press, Princeton, 2007).
60. Cerezo, M. *et al.* Reconstructing ancient mitochondrial DNA links between Africa and Europe. *Genome Res* **22**, 821–826. <https://doi.org/10.1101/gr.134452.111> (2012).
61. Olivieri, A. *et al.* Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and neolithic dispersals within Europe. *PLoS ONE* **8**, e70492. <https://doi.org/10.1371/journal.pone.0070492> (2013).
62. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147. <https://doi.org/10.1038/13779> (1999).
63. Achilli, A. *et al.* Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proc. Natl. Acad. Sci. USA* **110**, 14308–14313. <https://doi.org/10.1073/pnas.1306290110> (2013).
64. Brandini, S. *et al.* The Paleo-Indian entry into South America according to mitogenomes. *Mol. Biol. Evol.* **35**, 299–311. <https://doi.org/10.1093/molbev/msx267> (2018).
65. Strobl, C., Eduardoff, M., Bus, M. M., Allen, M. & Parson, W. Evaluation of the precision ID whole MtDNA genome panel for forensic analyses. *Forensic Sci. Int. Genet.* **35**, 21–25. <https://doi.org/10.1016/j.fsigen.2018.03.013> (2018).
66. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595. <https://doi.org/10.1093/bioinformatics/btp698> (2010).
67. Li, H. *et al.* The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
68. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. <https://doi.org/10.1101/gr.107524.110> (2010).
69. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763. <https://doi.org/10.1073/pnas.1314445110> (2013).
70. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* <https://doi.org/10.1101/pdb.prot5448> (2010).
71. Modi, A. *et al.* Complete mitochondrial sequences from Mesolithic Sardinia. *Sci. Rep.* **7**, 42869. <https://doi.org/10.1038/srep42869> (2017).
72. Peltzer, A. *et al.* EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60. <https://doi.org/10.1186/s13059-016-0918-z> (2016).
73. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> (2009).
74. Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2007).
75. Weissensteiner, H. *et al.* HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–63. <https://doi.org/10.1093/nar/gkw233> (2016).
76. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–394. <https://doi.org/10.1002/humu.20921> (2009).
77. Huber, N., Parson, W. & Dür, A. Next generation database search algorithm for forensic mitogenome analyses. *Forensic Sci. Int. Genet.* **37**, 204–214. <https://doi.org/10.1016/j.fsigen.2018.09.001> (2018).
78. Parson, W. & Dür, A. EMPOP—A forensic mtDNA database. *Forensic Sci. Int. Genet.* **1**, 88–92. <https://doi.org/10.1016/j.fsigen.2007.01.018> (2007).

Acknowledgements

We are grateful to *Soprintendenza Archeologia, Belle Arti e Paesaggio dell'Umbria, to Istituto Comprensivo Statale Foligno 5* (Perugia) and to all the volunteers who generously participated in this survey and made this research possible. We thank our colleagues Prof. Fausto Panara and Dr. Livia Lucentini with whom we have been discussing the feasibility and the first steps of this project, and Prof. Cristina Cereda, Dr. Gaetano Grieco, Dr. Marialuisa Valente, Dr. Nicole Huber and Jannika Oeke for technical support. We would like to thank the two anonymous reviewers for their suggestions and thoughtful comments. This research received support from: the Italian Ministry of Education, University and Research projects FIR2012 RBFR126B8I (to AO and AA), PRIN2017 20174BTC4R (to AA); Dipartimenti di Eccellenza Program (2018–2022)—Department of Biology and Biotechnology “L. Spallanzani,” University of Pavia (to AA, AO, OS and AT) and Department of Biology,

University of Florence (to DC); the Fondazione Cariplo (project no. 2018–2045 to AA, AO and AT); the Fondazione Carifol (2008 to AA) and the Tiroler Wissenschaftsfonds (TWF) (UNI-404/1998) (to MB).

Author contributions

A.M., H.L., D.C. and A.A. conceived the study. A.M., H.L., I.C., M.R.C., C.S., M.B., L.S., E.R. and S.V. did the lab work. A.M., I.C., M.R.C., N.R.M., A.H., C.S., M.B., L.S., C.X., L.P., W.P. and A.A. performed analyses. H.L., L.B.P., and A.A. provided modern samples and archaeological material. A.R., B.C., O.S., A.T., A.O., M.L., L.P., W.P. and D.C. gave inputs about genomic analyses. A.M., H.L., I.C., M.R.C. and A.A. wrote the manuscript with inputs from all co-authors. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-67445-0>.

Correspondence and requests for materials should be addressed to H.L. or A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020