



Universiteit
Leiden
The Netherlands

Ancient genomics support deep divergence between Eastern and Western Mediterranean Indo-European languages

Yediay, F.E.; Kroonen, G.J.; Sabatini, S.; Frei, K.M.; Frank, A. B.; Pinotti, T.; ... ; Willerslev, E.

Citation

Yediay, F. E., Kroonen, G. J., Sabatini, S., Frei, K. M., Frank, A. B., Pinotti, T., ... Willerslev, E. (2024). Ancient genomics support deep divergence between Eastern and Western Mediterranean Indo-European languages. *Biorxiv*. doi:10.1101/2024.12.02.626332

Version: Accepted Manuscript

License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4171691>

Note: To cite this publication please use the final published version (if applicable).

Ancient genomics support deep divergence between Eastern and Western Mediterranean Indo-European languages

Fulya Eylem Yediay^{1,2*}, Guus Kroonen^{3,4*}, Serena Sabatini^{2#}, Karin Margarita Frei^{5#}, Anja B. Frank^{6,7}, Thomaz Pinotti^{1,8}, Andrew Wigman³, Rasmus Thorsø³, Tharsika Vimala¹, Hugh McColl^{1,2}, Ioanna Moutafi^{9,10}, Isin Altinkaya¹, Abigail Ramsøe¹, Charleen Gaunitz¹, Gabriel Renaud¹¹, Alfredo Mederos Martin¹², Fabrice Demeter^{1,13}, Gabriele Scorrano^{1,14}, Alessandro Canci¹⁵, Peter Fischer², Izzet Duyar¹⁶, Claude Serhal¹⁷, Alexander Varzari^{18,19}, Murat Türkteki²⁰, John O'Shea²¹, Lorenz Rahmstorf²², Gürcan Polat²³, Derya Atamtürk²⁴, Lasse Vinner¹, Sachihito Omura²⁵, Kimiyoshi Matsumura²⁵, Jialu Cao¹, Frederik Valeur Seersholm¹, Jose Miguel Morillo Leon²⁶, Sofia Voutsaki²⁷, Raphaël Orgeolet^{28,29}, Brendan Burke³⁰, Nicholas P Herrmann³¹, Giulia Recchia³², Susi Corazza¹⁵, Elisabetta Borgna¹⁵, Mirella Cipolloni Sampò³³, Flavia Trucco³⁴, Ana Pajuelo Pando³⁵, Marie Louise Schjellerup Jørkov³⁶, Patrice Courtaud³⁷, Rebecca Peake^{38,39}, Juan Francisco Gibaja Bao⁴⁰, Györgyi Parditka²¹, Jesper Stenderup¹, Karl-Göran Sjögren², Jacqueline Staring¹, Line Olsen¹, Igor V. Deyneko⁴¹, György Pálfi⁴², Pedro Manuel López Aldana³⁵, Bryan Burns⁴³, László Paja⁴², Christian Mühlenbock⁴⁴, Claudio Cavazzuti⁴⁵, Alberto Cazzella³², Anna Lagia⁴⁶, Vassilis Lambrinoudakis⁴⁷, Lazaros Kolonas⁴⁸, Jörg Rambach^{49,50}, Eugen Sava¹⁸, Sergey Agulnikov⁵¹, Vicente Castañeda Fernández⁵², Mia Broné⁵³, Victoria Peña Romo⁵⁴, Fernando Molina González⁵⁵, Juan Antonio Cámara Serrano⁵⁵, Sylvia Jiménez Brobeil⁵⁶, Trinidad Nájera Molino⁵⁵, María Oliva Rodríguez Ariza⁵⁷, Catalina Galán Saulnier⁵⁸, Armando González Martín⁵⁹, Nicolas Cauwe⁶⁰, Claude Mordant⁶¹, Mafalda Roscio⁶², Luc Staniaszek^{38,61}, Mary Anne Tafuri³², Tayfun Yıldırım⁶³, Luciano Salzani⁶⁴, Thorfinn Sand Korneliussen¹, J. Víctor Moreno-Mayar¹, Morten Erik Allentoft^{1,65}, Martin Sikora¹, Rasmus Nielsen^{1,66}, Kristian Kristiansen^{1,2*} & Eske Willerslev^{1,67*}

Affiliations

¹Lundbeck Foundation GeoGenetics Center, Globe Institute, University of Copenhagen, Copenhagen, Denmark, ²Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden, ³Leiden University Centre for Linguistics, Leiden University, Leiden, The Netherlands, ⁴Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark, ⁵Department of Environmental Archaeology and Materials Science, National Museum of Denmark, Kongens Lyngby, Denmark, ⁶Institute for Geology, University of Hamburg, Hamburg, Germany, ⁷Department of Research, Collections and Conservation, Environmental Archaeology and Materials Science, National Museum of Denmark, Kongens Lyngby, Denmark, ⁸Laboratório de Biodiversidade e Evolução Molecular (LBEM), Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁹Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig-Maximilians-Universität München, Germany, ¹⁰The M.H. Wiener Laboratory for Archaeological Science, American School of Classical Studies at Athens, ¹¹Department of Health Technology, Technical University of Denmark, Copenhagen, Denmark, ¹²Universidad Autonoma de Madrid, Spain, ¹³Eco-anthropologie (EA), Dpt ABBA, Muséum national d'Histoire naturelle, CNRS, Université Paris Cité, Musée de l'Homme 17 place du Trocadéro, 75016 Paris, France, ¹⁴Center for Molecular Anthropology for the study of ancient DNA, Department of Biology, University of Rome Tor Vergata, Rome Italy, ¹⁵University of Udine, Italy, ¹⁶Istanbul University, ¹⁷British Museum, London (UK) and University College London (UK), ¹⁸The National Museum of History of Moldova, Chişinău, Republic of Moldova, ¹⁹Laboratory of Human Genetics, Chiril Draganiuc Institute of

Phthisiopneumology, Chişinău, Republic of Moldova, ²⁰Bilecik Şeyh Edebali University, Turkey, ²¹Museum of Anthropological Archaeology, University of Michigan, USA, ²²University of Göttingen, Germany, ²³Ege University, Turkey, ²⁴Istanbul University, Turkey, ²⁵Japanese Institute of Anatolian Archaeology, Turkey, ²⁶Universidad Christian Albrecht de Kiel, Germany, ²⁷Groningen Institute of Archaeology, University of Groningen, The Netherlands, ²⁸Aix Marseille Univ, CNRS, CCJ, Aix-en-Provence, France, ²⁹École française d'Athènes, ³⁰University of Victoria, Canada, ³¹Texas State University, USA, ³²Rome University La Sapienza, Italy, ³³National Research Council, Italy, ³⁴Italian Heritage board, Italy, ³⁵Grupo de Investigación TELLUS. Departamento de Prehistoria y Arqueología. Universidad de Sevilla. España, ³⁶Laboratory of Biological Anthropology, ArchaeoScience, Globe Institute, University of Copenhagen, Denmark, ³⁷University of Bordeaux, France, ³⁸Institut national de recherches archéologiques préventives (Inrap), France, ³⁹UMR 6298, Université de Bourgogne, France, ⁴⁰Spanish National Research Council, Spain, ⁴¹Laboratory of Human Genetics, Chiril Draganiuc Institute of Phthisiopneumology, Chişinău, Republic of Moldova, ⁴²Department of Biological Anthropology, University of Szeged, Szeged, Hungary, ⁴³Classical Studies, Wellesley College, Boston, USA, ⁴⁴Lund University, Sweden, ⁴⁵University of Bologna, Italy, ⁴⁶Ghent University, Department of Archaeology, The Netherlands, ⁴⁷National and Kapodistrian University of Athens, ⁴⁸Ministry of Culture (Greece), ⁴⁹Greek Archaeological Society, ⁵⁰German Archaeological Institute, Athens, Greece, ⁵¹National Agency for Archaeology, Chişinău, Republic of Moldova, ⁵²Universidad de Cádiz, Spain, ⁵³Världskulturmuseerna, Sweden, ⁵⁴Complutense University, Madrid, Spain, ⁵⁵Department of Prehistory and Archaeology, University of Granada, Spain, ⁵⁶Department of Legal Medicine, Toxicology and Anthropology University of Granada, Spain, ⁵⁷Branch of Prehistory, Department of Historical Heritage, University of Jaen, Spain, ⁵⁸Department of Prehistory and Archaeology, Autonomous University of Madrid, Spain, ⁵⁹Department of Biology, Autonomous University of Madrid, Spain, ⁶⁰Royal Museums of Art and History Brussels, Belgium, ⁶¹UMR 6298, Université de Bourgogne, France, ⁶²UMR 8164, Université de Lille, France, ⁶³Department of Archaeology, Ankara University, Ankara, Turkey, ⁶⁴Archaeological Superintendence of Verona, Italy, ⁶⁵Trace and Environmental DNA (TrEnD) Laboratory, School of Molecular and Life Sciences, Curtin University, Perth, Australia, ⁶⁶Departments of Integrative Biology and Statistics, UC Berkeley, USA, ⁶⁷GeoGenetics Group, Department of Zoology, University of Cambridge, Cambridge, UK

*Corresponding author, #Shared second co-authors

Summary

The Indo-European languages are among the most widely spoken in the world, yet their early diversification remains contentious^{1–5}. It is widely accepted that the spread of this language family across Europe from the 5th millennium BP correlates with the expansion and diversification of steppe-related genetic ancestry from the onset of the Bronze Age^{6,7}. However, multiple steppe-derived populations co-existed in Europe during this period, and it remains unclear how these populations diverged and which provided the demographic channels for the ancestral forms of the Italic, Celtic, Greek, and Armenian languages^{8,9}. To investigate the ancestral histories of Indo-European-speaking groups in Southern Europe, we sequenced genomes from 314 ancient individuals from the Mediterranean and surrounding regions, spanning from 5,200 BP to 2,100 BP, and co-analysed these with published genome

data. We additionally conducted strontium isotope analyses on 224 of these individuals. We find a deep east-west divide of steppe ancestry in Southern Europe during the Bronze Age. Specifically, we show that the arrival of steppe ancestry in Spain, France, and Italy was mediated by Bell Beaker (BB) populations of Western Europe, likely contributing to the emergence of the Italic and Celtic languages. In contrast, Armenian and Greek populations acquired steppe ancestry directly from Yamnaya groups of Eastern Europe. These results are consistent with the linguistic Italo-Celtic^{10,11} and Graeco-Armenian^{1,12,13} hypotheses accounting for the origins of most Mediterranean Indo-European languages of Classical Antiquity. Our findings thus align with specific linguistic divergence models for the Indo-European language family while contradicting others. This underlines the power of ancient DNA in uncovering prehistoric diversifications of human populations and language communities.

Introduction

From 5,000 BP, large-scale human migrations significantly restructured the genetic makeup of human populations across Eurasia^{6,7}. Various pulses of steppe-related ancestry, associated with the mobile pastoralists of the Early Bronze Age Yamnaya culture, spread across vast distances, leaving distinct cultural and genetic signatures. These migrations likely also played a key role in the prehistoric dispersal of the Indo-European language family. However, steppe ancestry reached the various regions of Western Eurasia by different mechanisms and at different times. In various historically Indo-European-speaking regions of Europe, the arrival of steppe ancestry was mediated by populations associated with the archaeological complexes of the Corded Ware (CWC) (5,000–4,350 BP)^{6,7,14} and Bell Beakers (4,800–4,300/3,800 BP)^{15–17}. However, the extent to which similar dynamics occurred in the Eastern Mediterranean and adjacent Western Asian populations remains less well-understood. Moreover, while steppe ancestry has previously been detected in prehistoric Iberians^{18,19}, Italians^{20–23}, Greeks^{24–27}, and Caucasians²⁴, questions remain regarding the interrelatedness of the proximal source populations in the context of the Mediterranean region at large.

The spread of steppe ancestries is closely tied to the diversification of the Indo-European protolanguage into its historically attested subgroups²⁸. In the Mediterranean, important Indo-European languages from Classical Antiquity include Gaulish, Latin, Greek, and Armenian, the latter being indigenous to the South Caucasus and Eastern Anatolia. For these, multiple competing phylogenetic linguistic models have been proposed²⁹ (Linguistic Supplementary 2-4). The so-called Indo-Greek hypothesis groups Greek as well as the closely related Phrygian with Indo-Iranian³⁰, while the competing Graeco-Armenian^{12,13} hypothesis posits that Greek forms a subclade with Armenian, possibly also including Albanian (“Balkan Indo-European”). Similarly, the Italic Indo-European subgroup, which is ancestral to Latin, has been variously grouped with Celtic and Germanic, giving rise to the traditionally popular Italo-Germanic³¹ and contrastive Italo-Celtic^{10,32} hypotheses. While relative linguistic consensus exists on the Graeco-Armenian and Italo-Celtic hypotheses, these are not unchallenged^{33,34}. More fundamentally, the lack of a definitive linguistic consensus model for the Indo-European diversification constitutes a key obstacle to the interdisciplinary study of Indo-European language dispersal.

Here, we investigate the various sources of steppe ancestry along the entire northern border of the Mediterranean region, so as to establish the most parsimonious divergence model for the Indo-European languages in this area. We present new whole-genome data from 314 ancient individuals (>0.1X genomic coverage) from Southern and Central-Eastern Europe, as well as from the Eastern Mediterranean comprising Spain, France, Italy, Hungary, Moldova, Greece, Cyprus, Turkey, Syria, and Lebanon. These individuals mostly belong to the Bronze Age but span a time frame from 5,200 BP to 2,100 BP (Fig. 1) (Genetics and Strontium Supplementary S3; Supplementary Table S1). We also provide strontium isotope signatures from 224 individuals and radiocarbon-dating of 144 individuals (Supplementary Tables S8 and S9). Utilizing identity-by-descent (IBD) inferred admixture modelling^{17,35} with specific source populations, we obtained enhanced resolution of genetic ancestries, enabling us to differentiate diverse or common ancestries. Furthermore, we combined our strontium isotope data with our genetic results to deepen the understanding of mobility patterns over time.

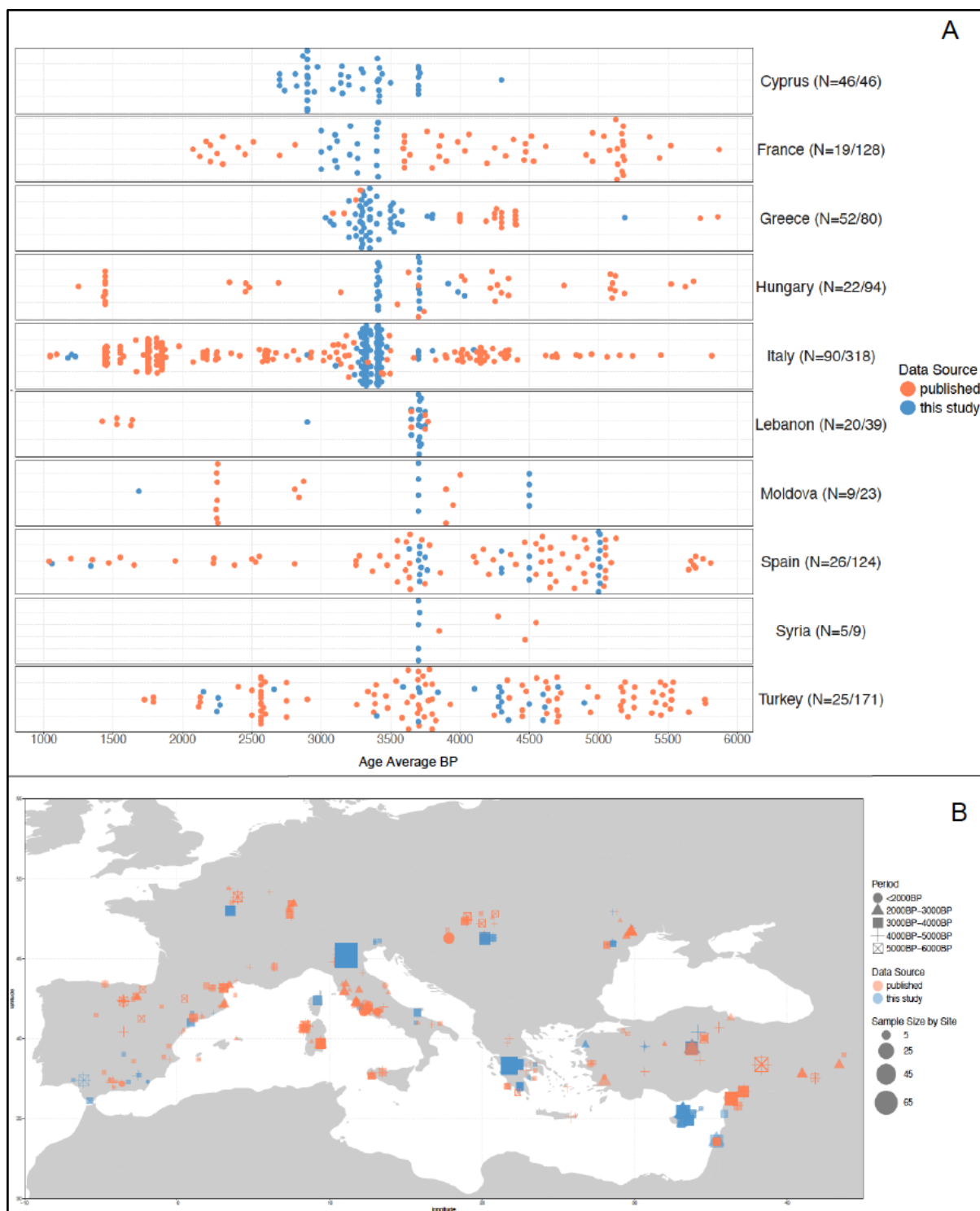


Fig. 1. Distribution of ancient individuals distributed by country (A), (N= the number of individuals in this study/total number of individuals in dataset), and locality shown on the map (B). We only demonstrate the individuals limited by time frame (6,000–1,000 cal BP) to avoid overlapping.

Population structure and overview of IBD clusters

We generated the dataset by shotgun-sequencing genomic DNA extracted from 314 ancient individuals (Genetics and Strontium Supplementary S2; S3). This was combined with 992 published shotgun-sequenced genomes and 1,097 genome-wide SNP-captured ancient

individuals (Supplementary Table S2). These were all imputed using established cut-offs^{17,35–37} of 1X capture and 0.1X shotgun data resulting in a total combined dataset of 2,403 imputed genomes across 643,430 SNP (single nucleotide polymorphism) sites (Genetics and Strontium Supplementary S3). We called IBD segments between pairs of individuals using IBDseq³⁸, and employed a network-based hierarchical approach³⁹ to obtain fine-scale population clusters based on the total length of shared IBD segments between pairs of individuals (Genetics and Strontium Supplementary S5).

Through this clustering approach, we detected six deep clusters corresponding to distinct ancestries previously related to the prehistoric formation of Eurasia: “Farmer-related (Cluster 0_1)”, “European-Hunter-Gatherers (Cluster 0_2)”, “Caucasus – Iran (Cluster 0_3)”, “Steppe-related (Cluster 0_4)”, “Central Asia – Siberia (Cluster 0_5)”, “Moroccan-Hunter-Gatherers (Cluster 0_6)” (Extended Fig. 1 and Fig. 2; Supplementary Table S3; Genetics and Strontium Supplementary Table S5.1). We deeply investigated “Farmer-related (0_1)” and “Steppe-related (0_4)” clusters, since the ancient individuals reported here fell within a number of subclusters of these main clusters. The “Farmer-related (0_1)” cluster is further divided into four subclusters, distinguishing between East (0_1_2 and 0_1_3) and West Mediterranean (0_1_1 and 0_1_4). The “Steppe-related (0_4)” cluster comprises three subclusters with individuals from Central Asia and Europe, “Russia-Central Asia (0_4_1)”, “Bell Beaker-related (0_4_2)”, and “CWC-Yamnaya-related (0_4_3)” (Fig. 2; Genetics and Strontium Supplementary Table S5.2; Extended Data Fig. 3). The occurrence of these subclusters refers to the spread of Yamnaya-related ancestry both eastward and westward, as well as the formation of new genetic profiles throughout Europe and Central Asia^{6,7,40–42}. When focusing specifically on the time frame of this study (between 5,000 BP and 3,000 BP), we observe a distinct pattern emerging between the 5th and 4th millennia in terms of the distribution of these Steppe-related clusters (Fig. 2). The distribution of the “CWC-Yamnaya-related (0_4_3)” cluster originating from the Pontic Steppe, extends into Central Eastern Europe and Northern Greece before the 4th millennium BP. This cluster persisted west of the Black Sea and in Armenia after the 4th millennium BP, whereas the “Bell Beaker-related (0_4_2)” cluster became prevalent in Europe. Further south to Greece, during this time we see a shift of the border between the “Steppe-related (0_4)” cluster and the “Farmer-related (0_1)” cluster. The first instance we detect of cross-border interactions occurs during an advanced phase of the 4th millennium BP. Here we find that two published Greece Bronze Age individuals^{43,44} (Log04 and G23) fall within the “CWC-Yamnaya-related (0_4_3)” cluster, specifically within subclusters corresponding to the “Yamnaya-related (0_4_3_1)” culture and the “Armenian-Caucasus (0_4_3_3)”, respectively (Genetics and Strontium Supplementary Fig. S5.9). At the east end of this border zone, individuals from Moldova fall within the subclusters of the “Steppe-related (0_4)” cluster, whereas all Eastern Mediterranean individuals including those from Greece fall within the subclusters of “Farmer-related (0_1)” cluster. Moreover, we detect the newly sequenced Early Bronze age individuals from Moldova within the subcluster of “Yamnaya-related (0_4_3_1)”, while the Middle Bronze Age individuals within the subcluster corresponding to “Cord Ware (0_4_3_2)” individuals like previously published Moldovan individuals²⁴ (Supplementary Table S3). Additionally, we found all individuals from Anatolia, Cyprus, and Levant within

the “Mediterranean (0_1_2)” cluster during the 3rd and 4th millennia, except one Early Bronze age individual and two European outliers from Cyprus are found within “Early farmers (0_1_3)” and “Bell Beaker-related (0_4_2)” throughout the 4th and 3rd millennia BP.

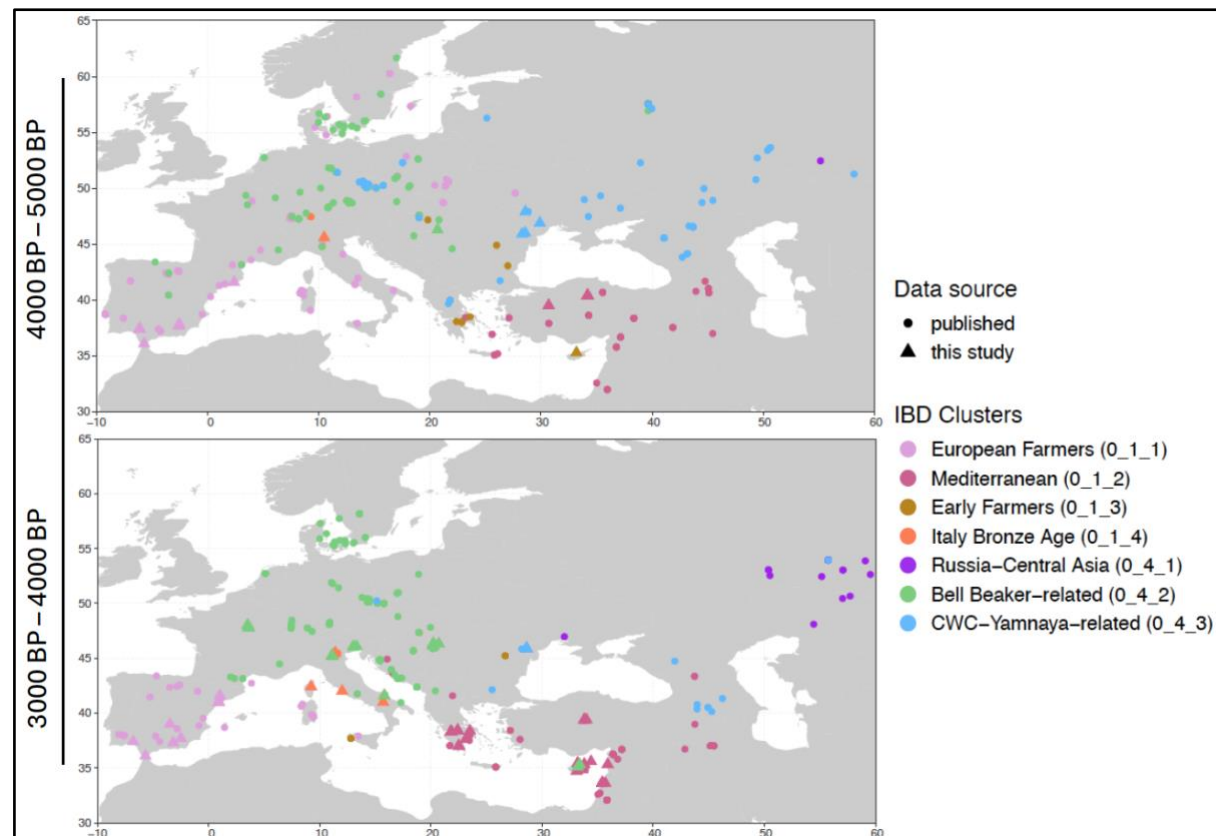


Fig. 2. Geographical distribution of the IBD clusters of individuals in the 5th and 4th millennia BP.

Southern and Central Eastern Europe

The spread of steppe-related ancestry in Europe has been traced to the Yamnaya populations and their subsequent admixture with local European populations after admixing with Globular Amphora culture-related (GAC-related) populations in the east^{6,7}. To differentiate between closely-related steppe ancestries in Southern European populations, we first generated an ancestry “palette” for every ancient individual in the database, representing the individual sharing patterns with all clusters in the dataset (Genetics and Strontium Supplementary S6.2). By designating specific clusters as sources, we are able to employ IBD Mixture Modelling^{17,35} to model the palettes of target individuals as best-fitting combinations of the source palettes. Of relevance here were various steppe-related populations from the Pontic Steppe and Europe (Genetics and Strontium Supplementary S6.2). We built on from a basic source set³⁵ that includes the Hunter-Gatherer and early Farming-related populations involved in the formation of European population structure, including a series of outgroups from Eurasia and Africa (Supplementary Table S4). We then progressively added a series of more recent populations as sources. By including two steppe sources, early Corded Ware

culture (CWC) and Yamnaya (Genetics and Strontium Supplementary S6.2; Supplementary Table S4), we revealed distinctly dissimilar patterns, contrasting Greece with France, Spain, and Italy, reflecting the expansion of two separate steppe ancestries—one mediated through Corded Ware groups and the other derived from Yamnaya populations (Extended Data Fig. 4; Genetics and Strontium Supplementary S6.2). The distinction between these two expansions is also distinguishable by the Farming-related ancestry they carry. The arrival of steppe ancestry in Europe was accompanied by GAC-related Farming ancestry: this farming ancestry is present in the Corded Ware carrying GAC-related ancestry, but absent in Greece (Genetics and Strontium Supplementary S6.2; Fig. S6.18; Supplementary Table S5). We thus detect a clear division in steppe ancestry between the Eastern and Western Mediterranean.

By introducing a third steppe-related source from the “Bell Beaker-related (0_4_2)” cluster, Southwestern Europe shows ancestry that is BB-related, while that of Greece is Yamnaya-related, particularly in the Peloponnese (Figs. 3 and 4). The Balkans, however, exhibit mixed ancestries of CWC/BB and Yamnaya, suggestive of interactions between migrants from the Pontic Steppe and with local populations associated with CWC/BB-related ancestry, or an unsampled steppe source emerging west of the Black Sea (Fig. 4). During the Mycenaean period in Greece (or Late Helladic period, c. 3,700–3,200 BP), Yamnaya ancestry became widespread, even extending into the Eastern Mediterranean (Fig. 3). We interpret this as evidence of a direct Yamnaya migration into the Peloponnese, supported also by the detection of the typical Yamnaya Y-chromosome haplogroup lineage R1b-Z2103 (Z2108)^{24,45} in two individuals from Ayios Vasileios, one from the Mycenaean cemetery Voudeni, and one from Kirrha (Fig. 5, Supplementary Table S7). However, the dominant Y-haplogroup lineage in Greece, J2a-Y7011, is also detected in individuals from Lapithos, Iron Age Cyprus, who are additionally characterized by non-local strontium isotope signatures. This haplogroup is also found in Szőreg, Hungary, suggesting a connection between the Balkans and the Eastern Mediterranean, as this lineage is present in early Balkan populations. To identify a potential connection with the Balkans, we present new data from Middle and Late Bronze Age Hungary revealing a complex population structure. We observe a mixed pattern of BB and Yamnaya components in Hungary during the Bronze Age (Fig. 4; Genetics and Strontium Supplementary S6.3). This suggests either another source population for those individuals or an admixture between two groups.

Italian Bronze Age individuals show at least three distinct admixture patterns. The first is similar to the Bronze Age Bell Beaker-related groups of France and Spain. Interestingly, that group includes all individuals from Central Italy and Corsica, a few individuals from Olmo, and one published Bell Beaker-context individual from Northern Italy¹⁵ (I2478, 4,006 BP). A second group, primarily from Olmo, shows increased Neolithic farmer ancestry with lower amounts of steppe ancestry, suggesting an ongoing local admixture process. A third limited group with increased Yamnaya ancestry, similar to Balkan and Greek Bronze Age populations, is observed in individuals from the Adriatic coast of Italy (Fig. 4; Genetics and Strontium Supplementary S6.2, CGG_2_100646, NEO806, R1)^{6,20}. Interestingly, three Early Bronze Age individuals from Northeastern Italy carry an additional CWC or Yamnaya component, and such a profile is also observed in the individuals from Hungary (Fig. 4;

Genetics and Strontium Supplementary S6.3). These individuals, found in monumental tumulus graves and a coffin burial (CGG_2_022653), likely held a special social or cultural status, as suggested by the complexity of the tumuli and the ritual activities linked to those monuments⁴⁶ (Archaeological Supplementary 2.7.4; 2.7.9–11).

To compare the BB source with local contribution, we added two Early Bronze Age individuals from Northeastern Italy as an ancestral source to model Late Bronze Age Italian individuals (Genetics and Strontium Supplementary S6.2, Fig. S6.13; Supplementary Table S5). We found this new, local source to better model the ancestry component previously modelled by the BB source, except in a few individuals who show a mixed pattern of Bell Beaker and local ancestry, rejecting the model. This local source also replaced the steppe proportion in Corsica and Sicily, suggesting a maritime spread of steppe ancestry through the Mediterranean (Genetics and Strontium Supplementary Fig. S6.15). This ancestry also increased in Late Bronze Age individuals from Migennes, France (Genetics and Strontium Supplementary S6.2; Fig. S6.15), suggesting later formation in steppe ancestry in Western Europe could also connect the Balkans and Italy.

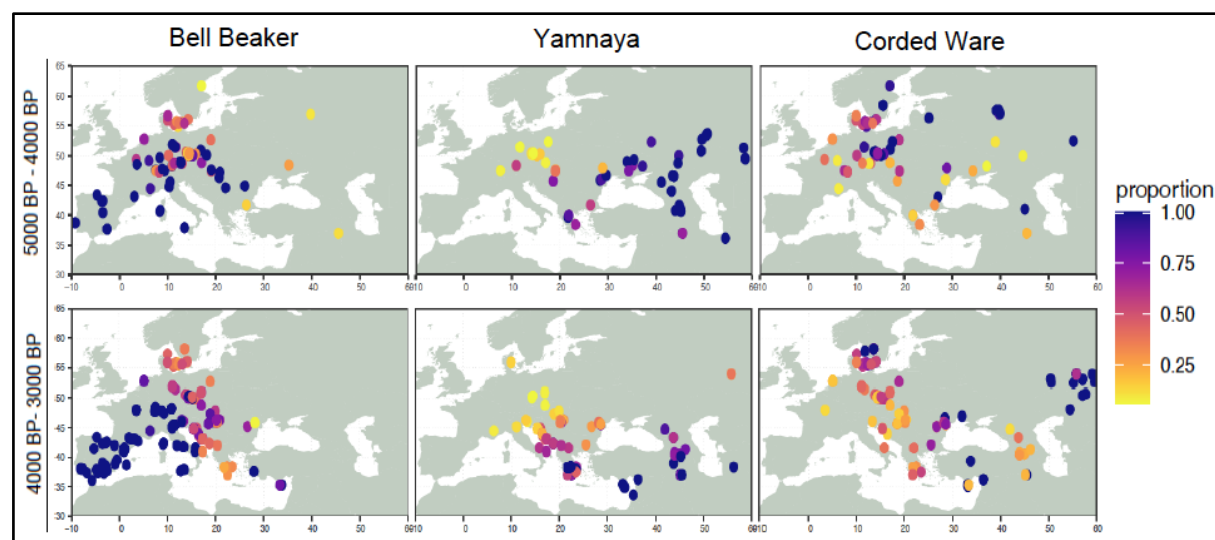


Fig. 3. Distribution of Bell Beaker-derived and Yamnaya-derived ancestry proportions obtained from the IBD admixture model. The proportion of each steppe source is standardized by the total steppe contributions, i.e. the sum of Corded Ware, Bell Beaker and Yamnaya_Samara contributions.

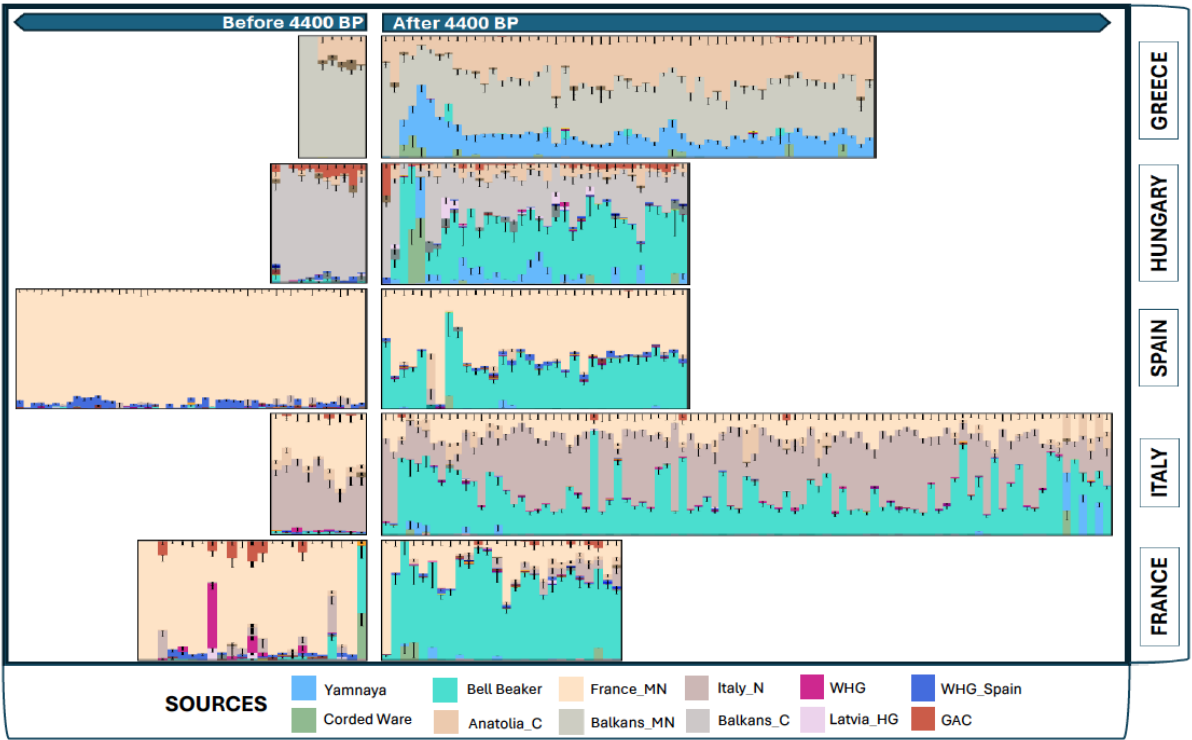


Fig. 4. Ancestry bar plots generated for each individual using source population proportions of IBD admixture modelling sorted by time BP and divided into two time series, before and after 4,400 BP, illustrating a Southern Europe split (Italy, France, and Spain vs Greece).

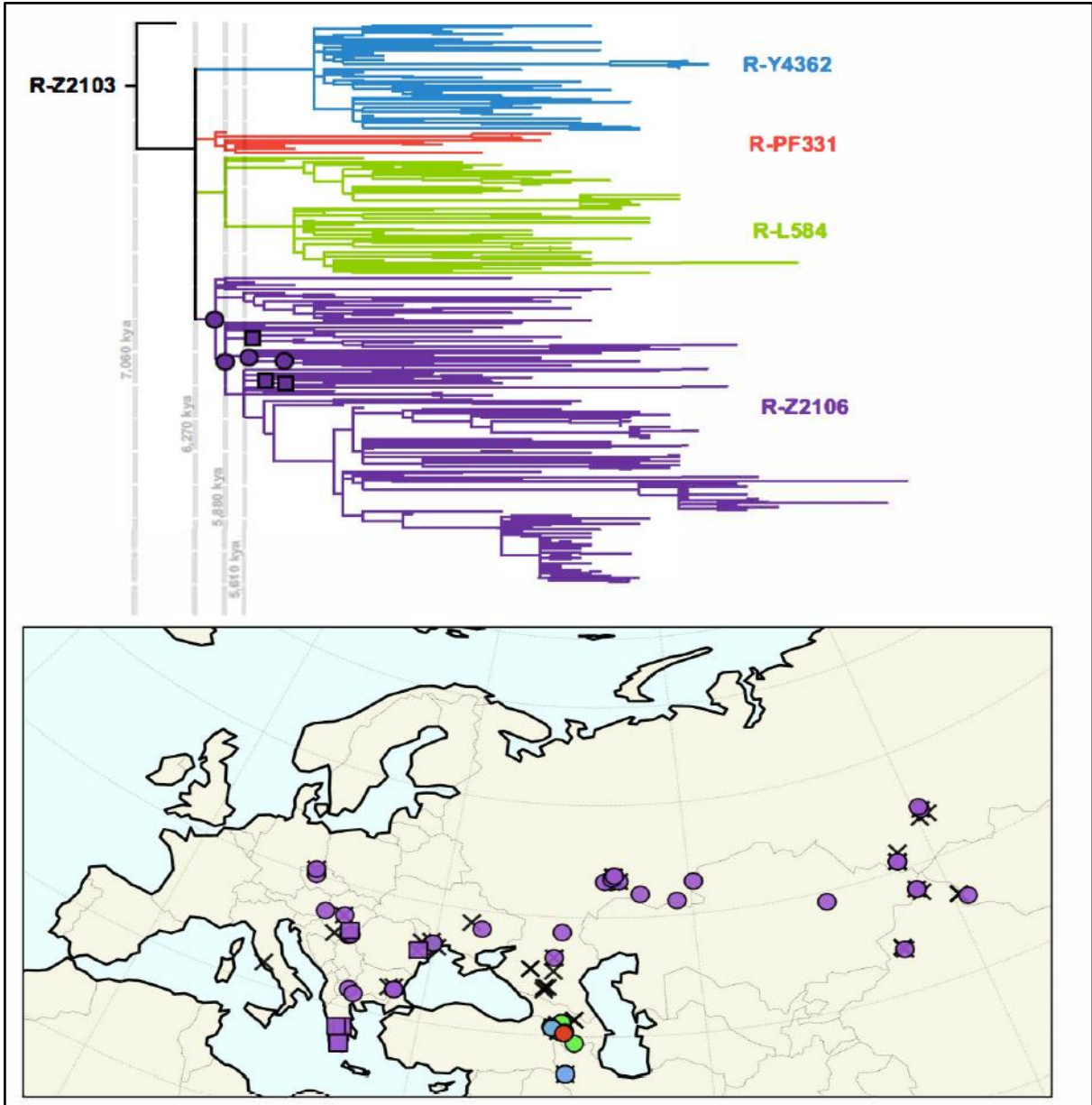


Fig. 5. Phylogeography of Y-chromosome haplogroup R-Z2103. Phylogram of haplogroup R-Z2103, with branch lengths proportional to SNP number, built from a dataset of unique variants in private datasets. The haplogroup resolves in a four-way polytomy. We plotted all occurrences older than 2,000 years ago in the ancient DNA record in Western Eurasia on an Albers Equal Area map, colour-coded by clades downstream of R-Z2103. Only haplogroup R-Z2106 extends beyond the Caucasus and Northern Iran, and we indicate phylogenetic position of individuals in the tree. Crosses mark R-Z2103 individuals with uncertain clade assignment due to low coverage, and squares indicate individuals from Greece, Moldova and Hungary generated in this study. Dotted lines denote split time estimates of key haplogroups, calculated using rho statistics.

Eastern Mediterranean

The Caucasus region has experienced a complex history of human migrations, interactions, and cultural exchanges, marked by admixture among diverse population groups. With the spread of farming to Iran and the Caucasus, populations in the region were shaped by admixture between early Neolithic farmers from the Fertile Crescent and two genetically

similar population groups, Iran Neolithic and Caucasus Hunter-Gatherers (CHG)^{24,41}. Additionally, the expansion of the Kura-Araxes culture during the 5th millennium BP connected the Caucasus with the Levant and Mesopotamia through extensive trade networks⁴⁷. Interaction between Anatolia and the Caucasus increased during the Chalcolithic and the Bronze Ages, leading to the spread of CHG ancestry. It also diffused into the Mediterranean, an early indication of which is found in Anatolian farmer groups from Tepecik-Çiftlik^{21,25,43,44}.

Here, we report new genomic data from 25 individuals from Anatolia, including individuals from Resuloğlu, an Early Bronze Age cemetery associated with Hattian culture, from Western Anatolia (Küllüoba, Keçiçayırı and Antandros), and from Central Anatolia (Kaman Kalehöyük), representing both the Bronze and Iron Ages. To assess the population structure of Anatolian individuals, we applied IBD mixture modelling, removing the CWC and BB sources, and instead progressively added a series of eastern ancestry sources from Iran and Caucasus Chalcolithic individuals, also including a Central Anatolian farmer source from Tepecik (Genetics and Strontium Supplementary S6.4; Supplementary Table S4). In our base modelling, all Chalcolithic and Bronze Age genomes from Anatolia appeared as mixtures of three components: local Neolithic farmers (from Tepecik), CHG (Caucasus Hunter-Gatherer) and a small proportion of Iranian Neolithic ancestry, with the exception of a minor proportion of EHG (Eastern Hunter-Gatherer) found in some individuals from Kalehöyük, Arslantepe, and Western Anatolia (Genetics and Strontium Supplementary Fig. S6.25; Supplementary Table S5). Together with CHG, this proportion was slightly elevated in the Iron Age. To understand the origin of the CHG and EHG contributions, we added Chalcolithic individuals from Iran and the Southern Caucasus as geographically proximal sources (Supplementary Table S4). The model revealed that Anatolian Chalcolithic and Bronze Age individuals received various proportions of ancestry from both Iran and the Southern Caucasus, with a higher amount of Iranian ancestry in Eastern than Western and Central Anatolia (Fig. 6). Moreover, the Chalcolithic Caucasus proportion increased in Central Anatolia, such as in Kaman Kalehöyük, in the Middle Bronze Age, and Iranian ancestry is minimal or absent in Western Anatolia. Admixture patterns in Eastern Anatolian individuals resemble those of Bronze Age individuals from Lebanon. Notably, Chalcolithic and Bronze Age individuals from Arslantepe, who received both Caucasus and Iranian ancestries, played a pivotal role within the Kura-Araxes region, spanning from Southern Caucasus to Levant.

To distinguish increased proportions of steppe ancestry in the Iron Age, we included multiple steppe sources (Yamnaya, CWC, BB) that revealed different signatures depending on the geographical location. In the newly sequenced Iron Age samples from Central and Northwestern Anatolia (Kalehöyük, Antandros and Keçiçayırı), we observed minor proportions of steppe ancestry with the pattern found in Balkans/Greek Late Bronze Age and probably reflects migrations from the Balkans (Genetics and Strontium Supplementary Fig. S6.37; S6.38; S6.39; Supplementary Table S5). Given that the individual from Keçiçayırı (CGG_2_022162) was unearthed from the Phrygian valley, the appearance of this ancestry may be associated with the emergence of the Phrygian state during the late 4th millennium BP⁴⁸ (Archaeology Supplementary 2.12.5; Linguistic Supplementary 3.3).

To trace the interaction between Eastern Mediterranean and Southern Europe, we sequenced Bronze Age and Iron Age individuals from Cyprus and Lebanon. Cypriot individuals make up a highly diverse group connecting all of the Eastern Mediterranean (Fig. 6). Our results suggest that Cyprus, and in particular its coastal towns, were a genetic and cultural “melting pot” during the Bronze Age. The Middle and Late Bronze Age individuals from Cyprus show a genetic pattern similar to that of individuals from Lebanon and Eastern Anatolia Bronze Age, while one (the earliest) individual from Karavas (CGG_2_022531), dated to 5,000–4,500 BP (Archaeology Supplementary 2.2.3), show extra early Anatolian farmer ancestry. A long-distance genetic connection is observed at Hala Sultan Tekke, one genetic outlier carrying Balkan/Aegean ancestry similar to that of Late Bronze Age Greece individuals. Thus, our dataset suggests close contacts of Cyprus with the Levant and the Aegean during the Bronze Age and even earlier periods. During the subsequent Iron Age, data from both Lapithos and Amathus suggest an increasingly uniform population across the island. Moreover, Iron Age genomes show a formation similar to populations of the Aegean and Western Anatolia Iron Age, carrying a small proportion of Yamnaya ancestry which reflects Greece ancestry (Fig. 6). Additionally, there is genetic evidence of long-distance interaction with Northern Europe, as seen in a Scandinavian genetic outlier (CGG_2_022535) from a rock-cut tomb at Vounous Bellapais, excavated by the Swedish-Cyprus expedition and dated to c. 4,000–3,800 BP. (Genetics and Strontium Supplementary Fig. S6.45, Archaeology Supplementary 2.2.7). This outlier clusters with Scandinavian Bronze Age individuals and, intriguingly, this origin is also supported by the Y-haplogroup I1 and by a non-local highly radiogenic strontium isotope signature compatible with some parts of Scandinavia (Genetics and Strontium Supplementary S10; Supplementary Table S8). The implications of this observation are not conclusive, since we do not have radiocarbon dating from this individual. Due to its rich copper sources, Cyprus maintained extensive trading networks with most of the Mediterranean; this situation is well mirrored by the presence of Anatolian, Levantine, Greek, and European ancestries in the Cypriot genomes.

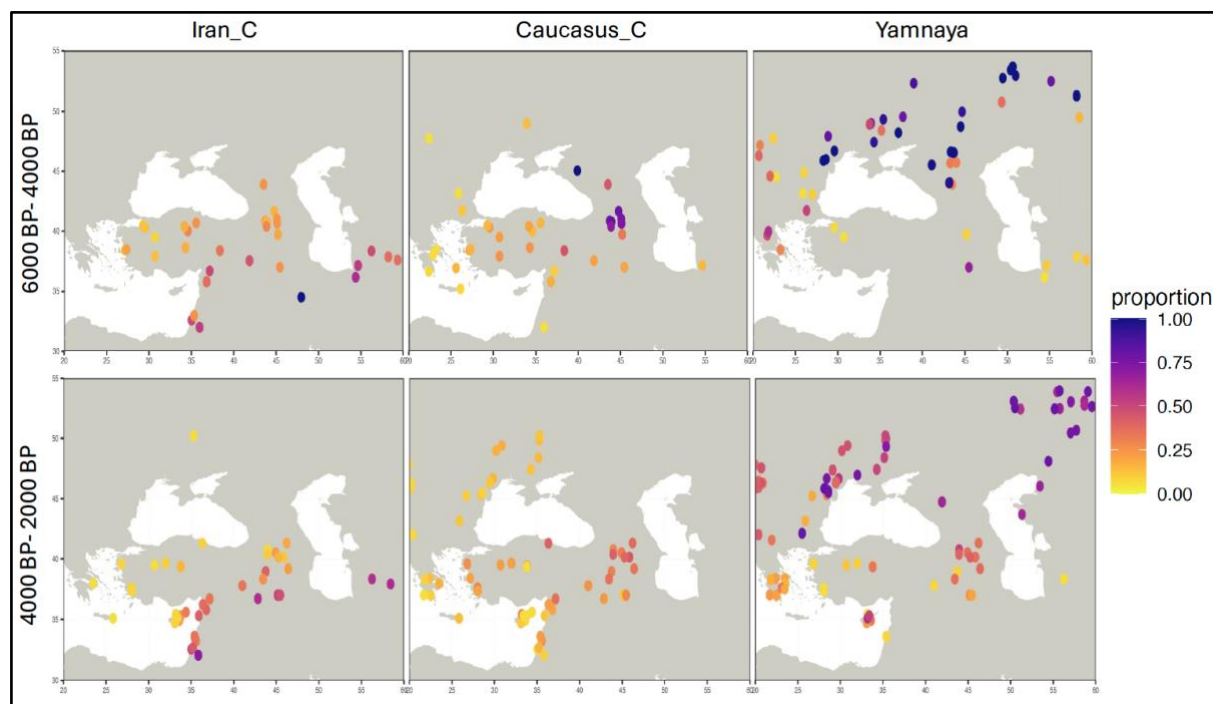


Fig. 6. Distribution of Iran Chalcolithic (Iran_C), Caucasus Chalcolithic (Caucasus_C) and Yamnaya-related ancestry proportions obtained from the IBD admixture model.

Increasing mobility towards the end of the Bronze Age

Considering the highly dynamic Bronze Age period in the Mediterranean, we analysed strontium isotope ratios from 224 ancient individuals across Cyprus, Greece, Italy, and Spain to trace the mobility of individuals (Genetics and Strontium Supplementary S10; Supplementary Table S8). We identified 56 individuals as potential non-locals and compared them to their genetic profiles.

In Greece, nine individuals from Kirrha, Voudeni, Eleon, and Apollo Maleatas returned non-local strontium isotope signatures. These might have travelled from regions within Greece where strontium isotope baselines are slightly more radiogenic. Among these nine individuals, we only have genetic data for four individuals, as several closely-related individuals were removed to avoid skewing the IBD admixture analysis. Genetically, these four individuals are similar to others from the same area, except for one individual from Eleon, who carries a small proportion of Bell Beaker ancestry reflecting a connection with Central Eastern Europe. Interestingly, two individuals (from Kirrha and Apollo Maleatas, Genetics and Strontium Supplementary S10; Supplementary Table S6) with non-local signatures have second-degree relatives who fall within the local baseline, indicating different mobility patterns within the same family. However, the relatives falling within the local baseline could also have originated from a different region given the large overlap in strontium signatures across Greece (Genetics and Strontium Supplementary S10; Supplementary Table S8). Finally, the father of a non-local individual from Apollo Maleatas (CGG_2_023933) differs from other Late Bronze Age Greece individuals by having a higher Yamnaya proportion compare to Late Bronze Age individuals, similar to the Middle Bronze Age individuals (>~3,800 BP).

In Cyprus, despite the high levels of genetic variation, only four Cypriot samples are characterized by non-local strontium isotope signatures (Supplementary Table S8). The genetic outlier from Vounous-Bellapais (CGG_2_022535) with unusual Scandinavian ancestry is also confirmed by the strontium isotope results. Another non-local Iron Age individual (CGG_2_022526) exhibits an admixture signature consistent with Greece Late Bronze Age (Genetics and Strontium Supplementary Fig. S6.45; S6.46). In contrast, the remainder of the outliers resemble populations from the Anatolia/Levant Bronze Age (Genetics and Strontium Supplementary S10). For many genetic outliers in which strontium data is also presented, long-term mobility was not detected (Genetics and Strontium Supplementary S10).

We additionally identified several individuals with non-local strontium isotope signatures within Italy, consistent with a highly dynamic society. These individuals come from Olmo, Pian Sultano, Scalvinetto, and from Coppa Nevigata. One individual (CGG_2_101264) from Pian Sultano had a bone and a tooth sample investigated for strontium isotopes, returning one signature (tooth) within their local strontium isotope baseline and another (petrous) outside of it. These results suggest that she did not originate from the area of Pian Sultano, but moved there in her early childhood/adolescence (Genetics and Strontium Supplementary S10). Yet another interesting non-local (CGG_2_022591), from Olmo, carries the highest proportion of steppe-related ancestry found among Italian Bronze Age individuals, displaying a different composition farmer and steppe proportion than other Italian Bronze Age individuals, carrying no Italian Neolithic ancestry. Another non-local from Coppa Nevigata (CGG_2_100646) genetically can be distinguished from the other Italian individuals by carrying Yamnaya-related steppe ancestry and Mediterranean farmer proportion similar to the Aegean/Balkans. Overall, non-local individuals of Bronze Age Italy cannot be distinguished from each other in terms of the steppe source obtained from IBD admixture modelling except for two cases from Coppa Nevigata and Olmo (CGG_2_100646 and CGG_2_022591; Genetics and Strontium Supplementary Fig. S6.13). Moreover, most of the non-local individuals in Olmo come from a group among Bronze Age Italians that exhibit a lower proportion of steppe-related ancestry, thus resembling individuals from Mediterranean islands (Sicily, Corsica)(Genetics and Strontium Supplementary Fig. S6.13; S6.15).

In Spain, we identified individuals with non-local strontium isotope signatures genetically belonging to two different groups: one with local farmer-related ancestry, and the other group dated to the Early Bronze Age (~4,300 BP) showing Bell Beaker-related ancestry. Since we observe only small-scale variation among farmer populations in Spain, the outliers with farmer ancestry may have originated from other farmer sites within the region.

Discussion

To elucidate the genetic formation and infer the divergence of the Indo-European language family, we investigated the contribution of potential ancestral source populations to the wider Mediterranean region, including from areas in which the Italic, Celtic, Greek, and Armenian languages are historically spoken. The genetic results obtained by IBD admixture modelling

with putative steppe-related source populations support a deep divergence between Eastern and Western Mediterranean Indo-European-speaking groups through the detection of Yamnaya-related and Bell Beaker-related steppe ancestry respectively (Fig. 4). This divergence supports the linguistic hypotheses on the existence of the so-called Graeco-Armenian and Italo-Celtic subclades. In contrast, it disqualifies the rival cladistic hypotheses known as Indo-Greek and Italo-Germanic, the steppe ancestry among the populations of historically Germanic- and Indo-Iranian-speaking areas previously having been characterized as primarily Corded Ware-related³⁵.

South-central Europe: Italic

Prior to its Romanization, Italy was characterized by marked linguistic complexity, harbouring multiple Indo-European and non-Indo-European language groups⁴⁹. The Italic languages, including Latin, Oscan, Umbrian, and possibly Venetic, are Indo-European in origin, just like the more distantly related Cisalpine Gaulish, Messapic, and Greek. Etruscan and Rhaetic, on the other hand, belong to the so-called Tyrrhenian family (Linguistic Supplementary 2). Due to this complexity, tracing the spread of the Italic group, which eventually came to dominate the Italian Peninsula, is notoriously difficult⁵⁰. Archaeological interpretations have variously linked the spread to the Remedello culture, the Rinaldone culture, the Terramare culture, and the Proto-Villanovan as well as the Villanovan cultures⁵¹. However, the oldest direct linguistic evidence for the presence of Italic is first found in Old Latin and Umbrian inscriptions from around 2,650 BP.

Previous genetic studies have dated the arrival of steppe ancestry in Northern Italy to around 4,000 BP and in Central Italy by 3,600 BP²¹. In our data, steppe ancestry appears in Central Italy a century earlier, in two newly reported individuals from Latium (Pian Sultano, CGG_2_101264, CGG_2_101266), the eventual epicentre of the Latin language. According to our IBD admixture modelling, the steppe ancestry of these individuals, along with that of published Late Bronze Age individuals from Grotta Regina Margherita and Toppo Daguzzo, is characterized as genetically Bell Beaker rather than Yamnaya-related, similar to historically Celtic-speaking populations in Western Europe. This, as well as the prevalence of this ancestry throughout the Bronze and Iron Age, is consistent with the linguistic Italo-Celtic hypothesis.

The Eastern Mediterranean: Greek

The timing and trajectory of the entry of the Greek language into Greece are traditionally much-debated topics^{2,25} (Linguistic Supplementary 4.1). In Late Bronze Age Greece, writing emerges in the form of the Linear A Minoan (c. 3,800–3,400 BP) and Linear B Mycenaean (c. 3,350–3,150 BP) syllabary scripts⁵². In Cyprus, it appears with the Cypro-Minoan syllabary (c. 3,500–3,300 BP), a local variant of Linear A, and the derived Cypriot syllabary (c. 3,000–2,300 BP)⁵³. While the Linear A text corpus remains largely undeciphered, Linear B as well as the Cypriot syllabary have been shown to represent the earliest written evidence of Greek⁵⁴.

The arrival of steppe ancestry has previously been documented in individuals from Northern Greece as early as ~4,200 BP, suggesting a connection to the Pontic Steppe^{44,55,25,43}. Using IBD admixture modelling, we establish that the Steppe source in Greece originates directly from Yamnaya-related populations, and differs from CWC populations who are formed with Yamnaya and GAC and widespread across most of Europe. Notably, the proportion of this Yamnaya-related ancestry is higher in the individuals before ~3,800 BP, including a previously unpublished male from Argolis (Apollo Maleatas, CGG_2_022928-23929), dated to around 3,800 BP. This provides the hitherto clearest evidence for the intrusion of steppe-derived, potentially Greek-speaking groups into the Peloponnese. The appearance of steppe ancestry here thus predates the oldest direct evidence of Greek in the form of the Linear B script, by which time it had stabilized at lower levels⁴³. At the same time, the preliterate interactions with local populations find a possible analogue in the absorption of so-called “Pre-Greek” vocabulary⁵⁶, reflecting contact of Greek with non-Indo-European language varieties.

In Cyprus, the Arcadocypriot Greek dialect was spoken from at least the Early Iron Age, next to Phoenician and one or more unknown languages attested through the Cypro-Minoan and Cypriot syllabaries⁵⁷. A diverse population is also suggested by genetic links between LBA and EIA Cyprus with the Levant and the Aegean. Steppe ancestry is identified in a number of unpublished individuals from Hala Sultan Tekke (CGG_2_022123, CGG_2_022924) and from Lapithos (CGG_2_022517), reflecting an affiliation with Late Helladic (i.e. Mycenaean Age) populations of the Peloponnese. This aligns with the appearance of Mycenaean pottery imports in the Late Bronze Age⁵⁸ and with the linguistic classification of Arcadocypriot as a descendant of the same South Greek dialect as Mycenaean, in contrast to other Greek dialects such as Doric and Ionic^{54,59}. However, one other individual from Lapithos (CGG_2_22488), although buried 4,100–4,000 BP, already clusters with LBA Greeks, suggesting Pre-Mycenaean connectivity with steppe-impacted populations of the Aegean.

The Caucasus and Eastern Anatolia: Armenian

The Armenian language, attested from c. 1,550 BP, is historically spoken in the South Caucasus and Eastern Anatolia (Linguistic Supplementary 4.3). Its earliest presence there has previously been estimated to date to around 3,100 BP^{51,60}. During the Late Iron Age, most of the region was part of the Urartian Kingdom. This state was culturally diverse⁶¹ and likely also contained an Armenian linguistic element, as evidenced by the exchange of vocabulary between Urartian and an early form of Armenian⁶². This is potentially supported by the presence of steppe ancestry among the published individuals from Urartian and pre-Urartian contexts²⁶.

Steppe ancestry has previously been detected in the South Caucasus from the Middle Bronze Age²⁴, coinciding with the transition from the Kura-Araxes culture to the Trialeti culture by the end of the 5th millennium BP. We can now demonstrate that these individuals, as well as those from Urartian contexts, received steppe ancestry from the same, western Yamnaya population as 4th millennium BP individuals from the Aegean (Extended Data Fig. 6, Genetics and Strontium Supplementary Fig. S6.19; S6.21). These findings support the

linguistic Graeco-Armenian hypothesis and suggest that the linguistic precursor of Armenian was introduced to the Caucasus by the end of the 5th millennium BP^{24,63}.

Archaeological implications

Bell Beaker populations exerted a pronounced genetic and cultural impact in Iberia, whereas the situation in Italy is more complex. In Northeastern Italy, Bell Beaker groups appear to have arrived in relatively small numbers, with some individuals receiving prominent tumulus burials. However, during the Early Bronze Age and at the transition to the Middle Bronze Age, evidence emerges of a connection between Central Europe and Italy in the distribution of triangular daggers, often found in hoards^{64,65,66}. These daggers are distributed across Italy, with their appearance coinciding with the widespread diffusion of Bell Beaker-related ancestry across the Italian Peninsula between 3,800 and 3,500 BP.

In the Terramare region, population growth significantly exceeded local development, suggesting an influx from surrounding areas^{67,68}. Archaeological evidence suggests connections with Hungary and regions north of the Alps^{69–72}, although these links only become evident in genetic ancestries later, during the Iron Age. This suggests that Northern Italy functioned as a cultural “melting pot”, which would later impact other Italian regions. After 3,200 BP, a noticeable exodus is archaeologically documented, particularly from the southern part of the Terramare region in the Po Valley^{67,73}.

Further south along the Adriatic coast, three genomes from the Central and Southern Italian littoral (CGG_2_100646; NEO806; R1) reveal affinities with Balkan populations, reflecting sustained contacts between both sides of the Adriatic Sea during the 5th and 4th millennia BP^{74–76}. These contacts extended beyond the Adriatic, reaching as far as the Aegean, as indicated by the presence of Mycenaean pottery and other goods in settlements along the Italian coast⁷⁷. Genetic ancestries further suggest small-scale movements of people, possibly involving specialized craftspeople and traders, or potentially driven by exogamy practices, contributing to the observed genetic admixture in the region.

Greece reveals a different picture. A distinct steppe proportion was introduced into both the mainland and the Peloponnese around 3,800 BP, subsequently becoming prevalent in burial sites over the ensuing centuries. The arrival of Mycenaean culture, along with the Greek language, has been subject to divergent interpretations, ranging from gradual, peaceful infiltration to more rapid acquisition of political control^{78–80}. Neither of these models are contradicted by the new genetic evidence: following the collapse of Early Helladic society, the brief Middle Helladic period saw new migrants from the north. These eventually consolidated political control with the advent of the Late Helladic period and the formation of Early Mycenaean Culture around 3,700 BP⁸¹. The process also entailed a cultural and linguistic encounter between newcomers and the residing population. Over time the original steppe signal diminished due to admixture with local populations with farmer ancestry. Archaeological evidence documents sustained contacts with the genetic source region in Moldova and the surrounding Carpathian region, as reflected in the introduction of steppe horses and chariots⁸², as well as the trade in silver from Carpathian mines⁸³.

Thus, a long-standing debate regarding the origins of Mycenaean culture can be resolved, at least in part. Genetic links with Anatolia and Crete persisted, mirroring the cultural influences that shaped the formation of Mycenaean society. Moreover, Mycenaean culture exhibits striking similarities with the slightly earlier Trialeti culture of the Southern Caucasus⁶⁸, which likely contributed to processes behind the subsequent Armenian ethnogenesis. In the South Caucasus, we see the rise of new Bronze Age elites in the mineral-rich regions of present-day Georgia and Armenia. These elites had links to both steppe chariot traditions and the Hittite civilization in Anatolia. Trialeti burial inventories, characterized by monumental tumulus chamber burials featuring precious imports, exhibit parallels with both Hittite city-states and early Mycenaean shaft graves. This suggests the formation of new commercial and military networks linking steppe societies and Near Eastern civilizations⁶⁸, as now also corroborated by the genetics.

The genetic evidence from Cyprus underscores its significant role as a trade hub owing to its abundant copper resources from Chalcolithic until the Iron Age. Moreover, this evidence highlights the island's strong ties to Western Anatolia and the Aegean. The Late Bronze Age marked the peak of Cypriot copper mining and trade across the Mediterranean, fostering a flourishing international culture^{84,85}. Also, these connections are mirrored in the genetic evidence, which reveals links between Cyprus and both the Eastern and Western Mediterranean regions. By the early 4th millennium BP, an archaeological connection additionally existed between the Únětice culture and the Eastern Mediterranean, including Cyprus, as exemplified by the presence of Únětice ring ingots and dress pins⁸¹. This connection is potentially supported by genetic evidence from a single individual, although its historical significance remains enigmatic.

Conclusion

For the first time, we are able to coherently combine evidence from ancient DNA, strontium isotopes, linguistics and archaeology to support a dual model for the divergence and dispersal of the Italic, Greek and Armenian languages during the Bronze Age. Specifically, we show that the arrival of steppe ancestry in Spain, France and Italy was mediated by Bell Beaker populations of Western Europe, likely contributing to the emergence of the Italic and Celtic languages. In contrast, Armenian and Greek populations acquired steppe ancestry directly from Yamnaya groups of Eastern Europe. These results are consistent with the linguistic Italo-Celtic and Graeco-Armenian hypotheses accounting for the origins of most Mediterranean Indo-European languages of Classical Antiquity. In contrast, however, our results fail to support the competing Italo-Germanic and Indo-Greek hypotheses, as the steppe ancestry among the populations of historically Germanic- and Indo-Iranian-speaking areas has been characterized as primarily Corded Ware-related. Our findings thus align with specific linguistic divergence models for the Indo-European language family while contradicting others. This underlines the power of ancient DNA in uncovering prehistoric diversifications of human populations and language communities.

References

1. Ringe, D., Warnow, T. & Taylor, A. Indo-European and computational cladistics. *Trans. Philol. Soc.* **100**, 59–129 (2002).
2. Anthony, D. W. *The Horse, the Wheel, and Language*. (Princeton University Press, 2010).
3. Chang, W., Hall, D., Cathcart, C. & Garrett, A. ANCESTRY-CONSTRAINED PHYLOGENETIC ANALYSIS SUPPORTS THE INDO-EUROPEAN STEPPE HYPOTHESIS. *Language* **91**, 194–244 (2015).
4. Bouckaert, R. *et al.* Mapping the origins and expansion of the Indo-European language family. *Science* **337**, 957–960 (2012).
5. Heggarty, P. *et al.* Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* **381**, eabg0818 (2023).
6. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
7. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
8. Blažek, V. From August Schleicher to Sergei Starostin: On the development of the tree-diagram models of the Indo-European languages. *Journal of Indo-European Studies* **35**, 82–110 (2007).
9. Goldstein, D. Indo-European phylogenetics with R: A tutorial introduction. *Indo-European Linguistics* **8**, 110–180 (2020).
10. Weiss, M. Italo-Celtic. in *The Indo-European Language Family* (ed. Olander, T.) 102–113 (Cambridge University Press, 2022).
11. Schrijver, P. Sound change, the Italo-Celtic linguistic unity, and the Italian homeland of Celtic. in *Celtic from the West 3: Atlantic Europe in the Metal Ages — questions of shared language* (eds. Koch, J. T. & Cunliffe, B.) 489–502 (Oxbow Books, 2016).
12. Pedersen, H. Armenier. B. Sprache. in *Reallexikon der Vorgeschichte* (ed. Ebert, M.) vol. 1 219–226 (1924).
13. Olsen, B. A. & Thorsø, R. Armenian. in *The Indo-European Language Family: A Phylogenetic Perspective* (ed. Olander, T.) 202–222 (Cambridge University Press, 2022).

14. Malmström, H. *et al.* The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon. *Proc. Biol. Sci.* **286**, 20191528 (2019).
15. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
16. Patterson, N. *et al.* Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature* **601**, 588–594 (2022).
17. Allentoft, M. E. *et al.* Population genomics of post-glacial western Eurasia. *Nature* **625**, 301–311 (2024).
18. Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
19. Villalba-Mouco, V. *et al.* Genomic transformation and social organization during the Copper Age–Bronze Age transition in southern Iberia. *Science Advances* **7**, eabi7038 (2021).
20. Antonio, M. L. *et al.* Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714 (2019).
21. Sauppe, T. *et al.* Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. *Curr. Biol.* **31**, 2576–2591.e12 (2021).
22. Posth, C. *et al.* The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. *Sci Adv* **7**, eabi7673 (2021).
23. Moots, H. M. *et al.* A genetic history of continuity and mobility in the Iron Age central Mediterranean. *Nat. Ecol. Evol.* **7**, 1515–1524 (2023).
24. Lazaridis, I. *et al.* The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science* **377**, eabm4247 (2022).
25. Lazaridis, I. *et al.* Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
26. Lazaridis, I. *et al.* A genetic probe into the ancient and medieval history of Southern Europe and West Asia. *Science* **377**, 940–951 (2022).
27. Skourtanioti, E. *et al.* Ancient DNA reveals admixture history and endogamy in the prehistoric Aegean. *Nat Ecol Evol* **7**, 290–303 (2023).
28. *The Indo-European Puzzle Revisited: Integrating Archaeology, Genetics, and Linguistics.*

(Cambridge University Press, Cambridge, England, 2023).

29. Olander, T. *The Indo-European Language Family*. (Cambridge University Press, 2022).
30. Euler, W. *Indoiranisch-Griechische Gemeinsamkeiten Der Nominalbildung Und Deren Indogermanische Grundlagen*. (Institut für Sprachwissenschaft der Universität Innsbruck, Innsbruck, 1979).
31. Porzig, W. *Die Gliederung Des Indogermanischen Sprachgebiets*. (Carl Winter, Heidelberg, 1954).
32. Schrijver, P. Sound change, the Italo-Celtic linguistic unity, and the Italian homeland of Celtic. in *Celtic from the West 3* (eds. Koch, J. T. & Cunliffe, B.) 489–502 (Oxbow, Oxford, 2016).
33. Isaac, G. The nature and origins of the celtic languages : Atlantic seaways, italo-celtic and other paralinguistic misapprehensions. *Studia Celtica* **38**, 49–58 (2004).
34. Kim, R. I. Greco-Armenian: The persistence of a myth. *Indogermanische Forschungen* **123**, 247–272 (2018).
35. McColl, H. *et al.* Steppe Ancestry in western Eurasia and the spread of the Germanic Languages. *bioRxiv* (2024) doi:10.1101/2024.03.13.584607.
36. Sousa da Mota, B. *et al.* Imputation of ancient human genomes. *Nat. Commun.* **14**, 3660 (2023).
37. Ringbauer, H. *et al.* Accurate detection of identity-by-descent segments in human ancient DNA. *Nat. Genet.* **56**, 143–151 (2024).
38. Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* **93**, 840–851 (2013).
39. Greenbaum, G., Rubin, A., Templeton, A. R. & Rosenberg, N. A. Network-based hierarchical population structure analysis for large genomic data sets. *Genome Res.* **29**, 2020–2033 (2019).
40. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
41. Wang, C.-C. *et al.* Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat. Commun.* **10**, 590 (2019).
42. de Barros Damgaard, P. *et al.* The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).

43. Clemente, F. *et al.* The genomic history of the Aegean palatial civilizations. *Cell* **184**, 2565–2586.e21 (2021).
44. Koptekin, D. *et al.* Spatial and temporal heterogeneity in human mobility patterns in Holocene Southwest Asia and the East Mediterranean. *Curr. Biol.* **33**, 41–57.e15 (2023).
45. Scorrano, G., Yediay, F. E., Pinotti, T., Feizabadifarahani, M. & Kristiansen, K. The genetic and cultural impact of the Steppe migration into Europe. *Ann. Hum. Biol.* **48**, 223–233 (2021).
46. Borgna, E., Simeoni, G., Vinci, F. & Nicosia, C. Origin and evolution of a bronze age funerary landscape in Friuli: The ‘Lower Context’ of the Tumulus of Mereto di Tomba (Udine) and the 3rd-2nd Millennium Transition in the Northern Adriatic. *Origini* **53**, 113–140 (2019).
47. Sagona, A. *The Archaeology of the Caucasus: From Earliest Settlements to the Iron Age*. (Cambridge University Press, 2017).
48. Bang, P. F. & Scheidel, W. *The Oxford Handbook of the State in the Ancient Near East and Mediterranean*. (Oxford University Press, 2013).
49. Farney, G. D. & Bradley, G. *The Peoples of Ancient Italy*. (Walter de Gruyter GmbH & Co KG, 2017).
50. Mallory, J. P. *In Search of the Indo-Europeans: Language, Archaeology, and Myth*. (Thames and Hudson, 1989).
51. Mallory, J. P. & Adams, D. Q. *Encyclopedia of Indo-European Culture*. (Taylor & Francis, 1997).
52. Bartoněk, A. *Handbuch Des Mykenischen Griechisch*. (C. Winter, Heidelberg, 2003).
53. Ferrara, S. *Cypro-Minoan Inscriptions: Volume 1: Analysis*. (2012).
54. Gary Miller, D. *Ancient Greek Dialects and Early Authors: Introduction to the Dialect Mixture in Homer, with Notes on Lyric and Herodotus*. (Walter de Gruyter, 2013).
55. Anthony, D. W. *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. (Princeton University Press, 2010).
56. Verhasselt, G. The Pre-Greek linguistic substratum: An overview of current research. *Les Études Classiques* **77**, 211–239 (2009).
57. Steele, P. M. *A Linguistic History of Ancient Cyprus: The Non-Greek Languages and Their*

- Relations with Greek, c.1600–300 BC.* (Cambridge University Press, 2013).
58. Fischer, P. M. Interregional trade at Hala Sultan Tekke, Cyprus: Analysis and chronology of imports. *J. Archaeol. Sci. Rep.* **47**, 103722 (2023).
59. Giannakis, G. K., Crespo, E. & Filos, P. *Studies in Ancient Greek Dialects: From Central Greece to the Black Sea.* (Walter de Gruyter GmbH & Co KG, 2017).
60. Diakonoff, I. *The Pre-History of the Armenian People.* (Caravan Books, Delmar, NY, 1984).
61. Zimansky, P. Archaeological inquiries into ethno-linguistic diversity in Urartu. in *Greater Anatolia and the Indo-Hittite Language Family: Papers Presented at the Colloquium Hosted by the University of Richmond, March 18–19, 2000* (ed. Drews, R.) 15–27 (Institute for the Study of Man, Washington, DC, 2001).
62. Thorsø, R. *Prehistoric Loanwords in Armenian: Hurro-Urartian, Kartvelian, and the Unclassified Substrate.* (Leiden University, Leiden, 2023).
63. Anthony, D. W. Ten constraints that limit the Late PIE homeland to the steppes. in *Proceedings of the 33rd Annual UCLA Indo-European Conference: Los Angeles, November 12th and 13th, 2022* (eds. Goldstein, D. M., Jamison, S. W. & Yates, A. D.) 1–25 (Helmut Buske Verlag, 2024).
64. Schwenzer, S. *Frühbronzezeitliche Vollgriffdolche: typologische, chronologische und technische Studien auf der Grundlage einer Materialaufnahme von Hans-Jürgen Hundt.* (Verlag des Römisch-Germanischen Zentralmuseums, 2004).
65. Van Rosseberg, E. *Cultural Landscapes, Social Networks and Historical Trajectories: A Data-Rich Synthesis of Early Bronze Age Networks (C. 2200-1700 BC) in Abruzzo and Lazio (Central Italy).* (Sidestone Press, 2015).
66. Peroni, V. B. *I pugnali nell'Italia continentale.* (Franz Steiner Verlag Wiesbaden GmbH, 1994).
67. Cardarelli, A. The collapse of the Terramare culture and growth of new economic and social systems during the Late Bronze Age in Italy. *Scienze dell'Antichità: 15, 2009* 449–520 (2009).
68. Kristiansen, K. The rise of bronze age peripheries and the expansion of international trade 1950–1100 BC. in *Trade and Civilisation: Economic Networks and Cultural Ties, from Prehistory to the Early Modern Era* (ed. Kristiansen, K., Lindkvist, T. & Myrdal, J.) 87–112 (Cambridge University Press, Cambridge, 2018).

69. Cremaschi, M. *et al.* Climate change versus land management in the Po Plain (Northern Italy) during the Bronze Age: New insights from the VP/VG sequence of the Terramara Santa Rosa di Poviglio. *Quat. Sci. Rev.* **136**, 153–172 (2016).
70. Nicolis, F. Northern Italy. in *The Oxford Handbook of the European Bronze Age* (eds. Harding, A. & Fokkens, H.) 692–705 (OUP Oxford, 2013).
71. Vanzetti, A. The rise of the Terramara system (Northern Italy). *Cultural change in the shadow of Thera-Eruption* (2013).
72. Cardarelli, A., Cavazzuti, C., Fritzl, M., Gavranovic, M., Hajdu, T., Kiss, V., Köhler, K., Kulcsár, G., Melis, E., Rebay-Salisbury, K., Szabó, G., Szeverényi, V. The connections between the plains of the Po and the Danube during the Bronze Age seen through the spread of the urnfield model. *Rivista di Scienze Preistoriche* **LXX**, 231–243 (2020).
73. Bettelli, M., Cardarelli, A., Damiani, I. & Others. Le ultime terramare e la penisola. Circolazione di modelli o diaspora? *STUDI DI PREISTORIA E PROTOSTORIA* **3**, 187–198 (2018).
74. Arena, A. & Barbarić, V., Cardarelli, A., Govedarica, B., Radić Rossi, I., Tunzi, A.M. The Adriatic Sea and the interactions between its two shores during the late Early and Middle Bronze Age. *Rivista di scienze preistoriche* 245–257 (2020).
75. Cavazzuti, C. *et al.* The first ‘urnfields’ in the Plains of the Danube and the Po. *J. World Prehist.* **35**, 45–86 (2022).
76. Gori, M., Recchia, G., Pacciarelli, M. & Cazzella, A. 2500-2000 BC: connectivity phenomena between the Balkans, Greece, Southern Italy, Eastern Sicily, the Aeolian Islands and Malta, RIVISTA DI SCIENZE PREISTORICHE numero speciale Italia tra Mediterraneo ed Europa: mobilità, interazioni e scambi LXX S1 2020. *Rivista di Scienze Preistoriche* 181–197 (2020).
77. Iacono, F. *et al.* Establishing the Middle Sea: The Late Bronze Age of Mediterranean Europe (1700–900 BC). *Journal of Archaeological Research* **30**, 371–445 (2022).
78. Voutsaki, S. Social place space early Mycenaean Greece. International Discussions Mycenaean Archaeology (Mykenische Studien 35). in *Social change and human agency. The Argolid at the onset of the Mycenaean era. In Eder, B. and Zavadil, M. (eds) (Social) place and space in early Mycenaean Greece. International Discussions in Mycenaean Archaeology (Mykenische Studien*

35) (Austrian Academy Sciences Press, 2021).

79. Drews, R. The coming of the Greeks: Indo-European conquests in the Aegean and the Near East. (1988).

80. Wright, J. C. Early Mycenaean Greece. in *The Cambridge Companion to the Aegean Bronze Age* (ed. Shelmerdine, C. W.) 230–257 (Cambridge University Press, Cambridge, 2008).

81. Kristiansen, K. & Larsson, T. B. *The Rise of Bronze Age Society: Travels, Transmissions and Transformations*. (Cambridge University Press, 2005).

82. Librado, P. *et al.* The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* **598**, 634–640 (2021).

83. Stos-Gale, Z. A. Silver vessels in the Mycenaean Shaft Graves and their origin in the context of the metal supply in the Bronze Age Aegean. in *Metals of power – Early gold and silver* (ed. Meller, H., Risch, R. & Pernicka, E.) 183–208 (Landesmuseum für Vorgeschichte, Halle (Saale), 2014).

84. Steel, L. *Oxford Handbook Archaeology Levant: C. 8000-332 BCE. 577–591*. 577–591 (Oxford University Press, Oxford, 2013).

85. Knapp, A. B. *Cambridge World Archaeology: The Archaeology of Cyprus: From Earliest Prehistory through the Bronze Age*. (Cambridge University Press, Cambridge, England, 2013).

86. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).

87. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

88. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).

89. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

90. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

91. Fu, Q. *et al.* Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
92. 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 (2015).
93. Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
94. Hofmanová, Z. *et al.* Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6886–6891 (2016).
95. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).
96. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
97. Poznik, G. D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562–565 (2013).
98. Jobling, M. A. & Tyler-Smith, C. Human Y-chromosome variation in the genome-sequencing era. *Nat. Rev. Genet.* **18**, 485–497 (2017).
99. Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A. & Korneliussen, T. S. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* **8**, (2019).
100. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
101. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
102. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
103. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

104. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
105. Frei, K. M. *et al.* Tracing the dynamic life story of a Bronze Age Female. *Sci. Rep.* **5**, 10431 (2015).

Methods

Data generation

All steps of generating the ancient genomes were conducted in dedicated ancient DNA lab facilities at Lundbeck Centre for GeoGenetics, University of Copenhagen, using well established protocols within ancient DNA^{17,35} (Genetics and Strontium Supplementary S1). Drilling was performed manually, if possible, on both petrous bone and tooth. DNA extractions and library builds were carried out in the ancient DNA clean lab both manually and with automatisation. Double stranded “USER” and “non-USER” libraries⁸⁶ were built and sequenced on the Illumina Hiseq 4000 and the Novaseq 6000. non-USER were evaluated for the purpose of authenticating ancient reads, to allow for the detection of post-mortem damage⁸⁷. Where possible, USER treated libraries were built from the authenticated extracts to minimise the effects of post-mortem damage on downstream analysis⁸⁸ (Supplementary Table S1). We merged the USER and non-USER libraries as described in Genetics and Strontium Supplementary S2. All reads were mapped to the human reference genomes (build hs37d5) using bwa aln (0.7.17)⁸⁹, and removed the adapters using AdapterRemoval (2.3.2)⁹⁰, duplicates by using picard MarkDuplicates (2.25.0). to ensure the authenticity of the sequenced data, we performed mapDamage2.0 (v2.2.1)⁸⁷ (Supplementary Table S1). Contaminated libraries were identified using contamix⁹¹, schmutzi⁹² and for the X chromosome contamination level, we ran ANGSD⁹³. The total 380 samples yielded DNA ranging from 0.004X to 15.5X in autosomal coverage, 314 samples out of total samples provided DNA coverage >0.1X which were used for downstream analysis (Supplementary Table S1).

Imputation

We imputed³⁶ a total 2,403 ancient samples (merged whole genome shotgun and capture), 314 newly sequenced whole genome shotgun samples by following McColl *et al.* 2024³⁵ and Allentoft *et al.* 2024¹⁷ (Genetics and Strontium Supplementary S3).

Principal component analysis (PCA)

To get an overview of the basic structure of our data, we carried out a principal component analysis (PCA) on an imputed dataset, comprising a whole set (2,228 ancient individuals) and a subset (1,837 ancient individuals) from both shotgun and capture (1,240K) data using Plink (v1.90b6.21) (Genetics and Strontium Supplementary S4). We first restricted the SNPs to capture sites, and applied MAF (0.05) filtering, resulting in a dataset with 580,130 SNPs. We

visualized the main IBD clusters by colouring for all ancient genomes used for clustering (Extended Data Fig. 2) and focused on the subclusters of “Farmer-related (0_1)” and “Steppe-related (0_4)” in the PCA plot of a subset individuals (Extended Data Fig. 3).

IBD clustering and modelling

filtered the merged VCF files INFO>0.5, MAF (minor allele frequency) 0.01 and restricted to 1,240K capture SNP sites. To run IBDseq, we included the samples with >0.1x coverage for shotgun, >1x coverage for capture, and >0.90 average genotype probability across 643,430 SNP (single nucleotide polymorphism) sites³⁸. We ran IBDseq following the implementation in McColl et al. 2024 on ancient samples within a broad geographical area and period (300 BP–45,000 BP)^{17,35,37}. After removing close relatives, we ran IBD admixture models with a total of 2,228 ancient individuals, of which 274 were newly sequenced Bronze Age Mediterranean individuals

Prior to running Leiden network-based hierarchical clustering, we applied filters by removing the IBD segments of less than 1 cM, LOD score of less than 3, and hotspot regions. We also removed one pair of the first and second-degree relatives to minimize small clusters formed only with close relatives. We then ran the clustering with 2,228 ancient individuals by setting a minimum total shared IBD of 5 cM and a permutation of 200.

IBD admixture modelling is a method combining allele-matching profiling⁹⁴ and chromopainting⁹⁵ using the shared IBD length instead of allele frequency to get a better resolution for distinguishing especially genetically similar populations. Before aggregating the IBD sharing, we filtered out IBD segments of less than 1 cM, with no limit to the upper bound of total shared segments^{17,35}. The models and the source individuals for IBD Admixture modelling were given in the Supplementary Table S4.

Genetic sex determination

We estimated the depth of coverage for the individuals with newly generated data using pysam, by counting and measuring the length of the reads (MQ > 30) and dividing the sum by the reference contig length of chromosomes 1–22, X and Y. Because the Y-chromosome presents large regions of repetitive sequence not mappable using short-read sequencing technologies^{96–98}, we restricted all analyses to the 10 Mb single-copy region defined in Poznik et al 2013⁹⁷. We called chromosomal sex for all individuals in the dataset by calculating the ratio of the depth of coverage of X to the autosomes, Y to the autosomes, and Y and X chromosomes.

Relatedness

To identify relatedness, we ran NGSRelate (v2)⁹⁹ on the imputed dataset, calculating allele frequency by using only our samples since we have a dense population structure from Eurasia (Genetics and Strontium Supplementary S7).

Y-chromosome analysis

We used bcftools¹⁰⁰ mpileup and call functions to call genotypes within the 10 Mb accessible region of the Y-chromosome⁹⁷. We excluded indels, triallelic positions, and genotypes that were not called in more than 95% of the population of non-clonal reads. To determine

haplogroups, we matched ancestral and derived calls to the ISOGG 2019–2020 database using an in-house script that generates haplogroup paths in a root-to-tip manner. Those paths are ranked by the number of supporting variants – while also distinguishing C to T in forward and G to A in reverse strands – and then manually verified.

Mitochondrial DNA analyses

We re-aligned the newly sequenced ancient DNA reads to the revised Cambridge Reference Sequence (rCRS) for the human mitochondrial DNA sequence using *bwa aln* v. 0.7.17⁸⁹ (*options: -l10000*) and filtered for reads with a mapping quality of minimum 30 using SAMtools v. 1.17¹⁰⁰. We then reconstructed consensus sequences of the mitogenomes with *bcftools*¹⁰¹ v. 1.18 using *mpileup* (*options: --no-BAQ*) to obtain read pileups along the reference sequence, which were then inputted to *bcftools call* (*options: --multiallelic-caller --ploidy 1*) for haploid genotype calling. We kept variants covered by at least five reads and a genotype quality above 25. We generated the final consensus sequences with *bcftools consensus*. The reconstructed mitogenomes were aligned with *mafft*^{102,103} v.7.490, while we restricted the phylogenetic analysis to the coding region located at 577–16,023 base pairs (rCRS coordinates). We carried out a Maximum Likelihood (ML)-based phylogenetic tree analysis with *RAxML-NG*¹⁰⁴ v. 1.2.2 under the substitution model GTR+I+G4 (*options: --all -bs-trees 100*).

Sr isotope analysis

We performed Sr isotope analysis on 232 skeletal samples (139 teeth and 93 petrous bones) from 224 ancient individuals from Cyprus, Greece, Italy, and Spain (Supplementary Table S8). A diamond-tipped dental drill was used to cut a clean enamel sample (1–2 mg) from the tooth samples and to drill 1–2 mg of sample from the densest part of the otic capsule of the petrous bones. The tooth and bone samples were dissolved using a 1:1 solution of 0.5 ml 6M HCl and 0.5 ml 30% H₂O₂. Selected samples were spiked with a ⁸⁴Sr-enriched tracer to determine Sr concentration via isotope dilution (ID).

The Sr column separation was done according to the methods of Frei et al¹⁰⁵, using disposable 1 ml pipette tips fitted with pre-cleaned filters and charged using 200 µl pre-cleaned SrSpecTM resin (50–100 mesh; Eichrome Inc./Tristchem) as disposable extraction columns. The prepared samples were dissolved, loaded onto the columns, and washed using 3M HNO₃, before the Sr was collected using mq. All Sr concentrations and isotope measurements were performed at the University of Copenhagen using A VG Sector 54 IT mass spectrometer equipped with eight Faraday detectors.

Data availability

Sequence data were deposited in the ENA under accession: xxxxxxxx

Methods References

References 17, 35–36, 86–105

Acknowledgements

The Lundbeck Foundation GeoGenetics Centre is supported by grants from the Lundbeck Foundation (R302-2018-2155, R155-2013-16338), the Novo Nordisk Foundation (NNF18SA0035006), the Wellcome Trust (WT214300), Carlsberg Foundation (CF18-0024), the Danish National Research Foundation (DNRF94, DNRF174), the University of Copenhagen (KU2016 programme) and Ferring Pharmaceuticals A/S, to E.W. This work was further supported by the Swedish Foundation for Humanities and Social Sciences grant (Riksbankens Jubileumsfond M16-0455:1) and funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement no. 95138) to K.K., G.K., K.G.S., K.M.F. and F.E.Y. A.M.W., F.E.Y., G.J.K. and R.T. received support from the European Research Council (Starting Grant no. 716732). We are grateful for the support of the Carlsberg Foundation Semper Ardens research grant (CF18-0005) "Tales of Bronze Age People" and the research grant "Tales of Bronze Age Women" (CF-15 0878) both to K.M.F. and F.E.Y. S.S. and G.S. were supported by Riksbankens Jubileumsfond, grant P22-0641 Italy before Rome. J.V.M. is supported by the European Research Council (101078151) and VILLUM FONDEN (VIL53099). R.N. received support from NIH R35 GM153400-01. We would like to thank to Sapienza Università di Roma - progetti di Ricerca e Grandi Scavi, Regione Friuli Venezia Giulia; Fondazione Friuli; Università di Udine (projects "Dai tumuli ai castellieri: 1500 anni di storia in Friuli", "Protostoria della media pianura friulana"), The M.H. Wiener Laboratory for Archaeological Science, American School of Classical Studies at Athens, Sapienza Università di Roma; Soprintendenza Archeologia Belle Arti Paesaggio per le province di Foggia, Andria – Barletta, Soprintendenza Archeologia Belle Arti Paesaggio Friuli Venezia Giulia, Japanese Institute of Anatolian Archaeology, Kaman Kalehöyük Archaeology Museum, Balıkesir Kuva-yi Milliye Archaeology Museum, Eskişehir Etnik Müzesi, Çorum Archaeology Museum, Ministry of Culture and Tourism in Republic of Turkey for providing access to archaeological samples. We thank Henrik Bjerre Hansen, Maria Madrona, Lærke Daniela Kjærsgaard Hansen, Tina-Louise Marie Mortensen, Marcus Andreas Behnke Hjorth, Andreas Bak Pørksen and Nils August Thomasen for their assistance.

Author contributions

K.K., E.W., F.E.Y., K.M.F. initiated the study. F.E.Y., K.M.F. (Strontium part), K.K. and E.W. led the study. F.E.Y., G.K., K.M.F., K.K. and E.W. conceptualised the study. G.K., L.V.I., K.M.F., L.O., R.N., M.S., K.K. and E.W. supervised the research. G.K., K.M.F., R.N., K.K. and E.W. acquired funding for research. F.E.Y., S.S., M.E.A., F.D., I.M., A.M.M., P.F., I.D., A.V., M.T., G.P., J.M.O., S.O., S.V., R.O., N.P.H., Gi.R., E.B., M.L.J., G.Pa., K.G.S., K.M.F., G.F., A.L., A.Ca., E.S., S.A., M.A.T., T.Y., A.C., C.S., L.R., J.M.M.L., Gi.R., S.C., E.B., M.C.S., F.T., A.P.P., P.C., R.P., J.F.G.B., G.Pa., P.M.L.A., B.Bu., Ch.M., C.C., A.Ca., E.S., S.A., V.C.F., A.M., C.M., M.R., L.S., M.A.T., V.P.R., F.M.G., J.A.C.S., S.J.B., T.N.M., M.O.R.A., C.G.S., A.G.M., N.C. and L.Sa. were involved in sample collection. S.S., I.M., A.M.M., P.F., I.D., A.V., M.T., D.A., G.P., S.O., K.M., S.V., R.O., N.P.H., M.L.J., K.M.F., I.V.D., G.F., A.L., V.L., J.R., L.K., L.P., T.Y., S.S., A.C., C.S., L.R., J.M.M.L., B.B., Gi.R., S.C., E.B., M.C.S., F.T., A.P.P., P.C., R.P., J.F.G.B., G.Pa., P.M.L.A., B.Bu., Ch.M., C.C., A.Ca., E.S., S.A., V.C.F., A.M., C.M., M.R., L.S., M.A.T. and L.Sa. curated bioarchaeological data. F.E.Y., H.M., G.S., L.V.I., J.C., J.S., M.E.A., F.V.S., J.T.S. and C.G. were involved in genetic data generation. F.E.Y., T.P., H.M., M.S., T.V., I.A., T.S.K., A.R., G.R. and R.N. were involved in genetic data analysis. I.M., A.B.F., K.G.S., K.M.F., S.S. were involved in other data generation and analysis (Calibrating 14C-dating, diet, Strontium etc.). F.E.Y., G.K., A.W., R.T., K.K. and E.W. drafted the main text. F.E.Y., G.K., S.S., A.B.F., A.W., R.T., I.M., S.V., R.O. and N.P.H. drafted supplementary notes and materials. F.E.Y.,

G.K., S.S., A.B.F., A.W., R.T., I.M., I.A., H.M., G.S., A.V., J.M.O., L.V.I., S.V., R.O., N.P.H., Gi.R., E.B., M.L.J., G.Pa., K.G.S., K.M.F., R.N., L.O., A.Ca., M.E.A., M.S., J.V.M., K.K. and E.W. involved in reviewing and editing drafts. All co-authors read, commented on and agreed on the submitted manuscript.

Ethics declarations

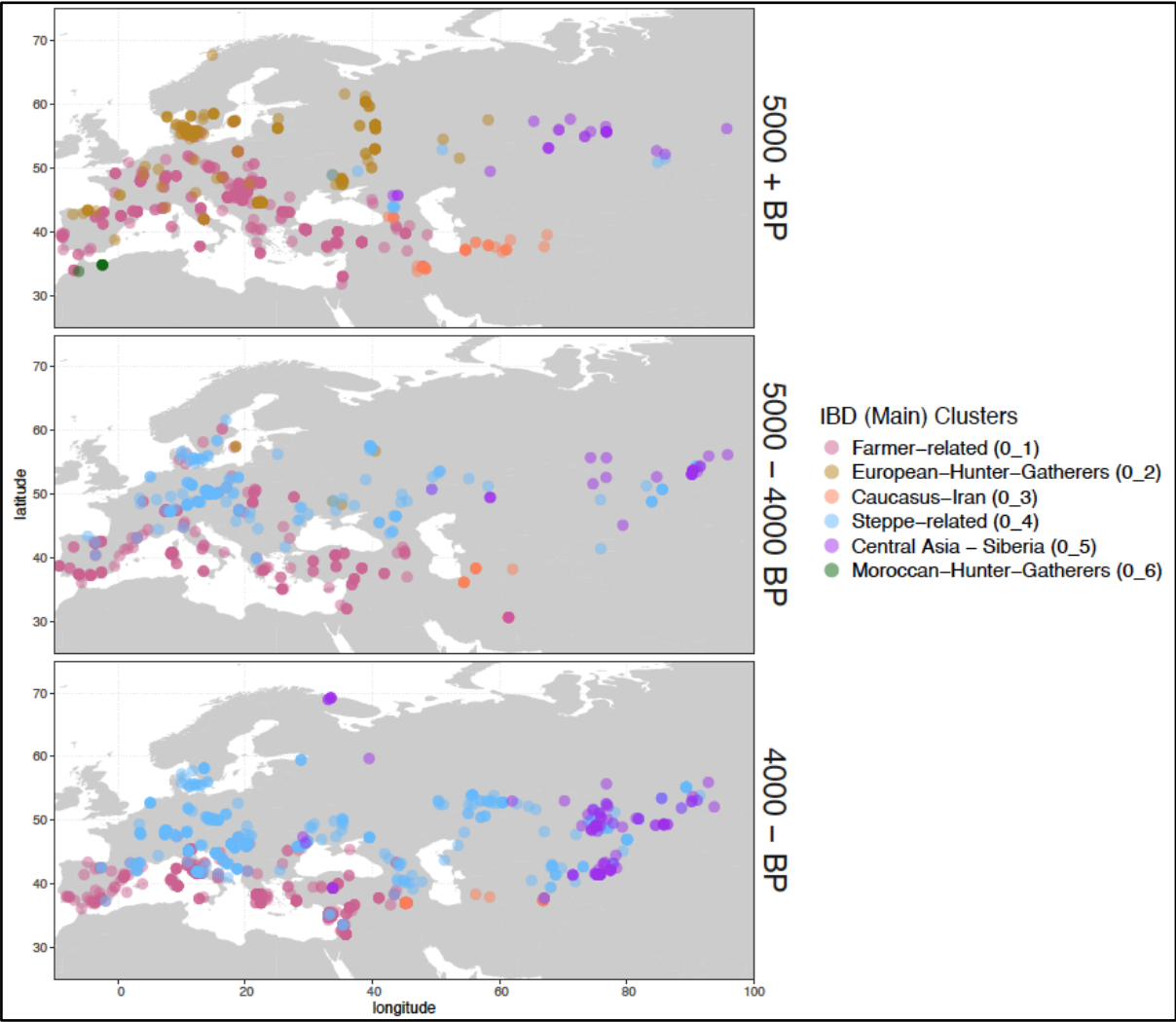
The authors declare no competing interests

Additional Information

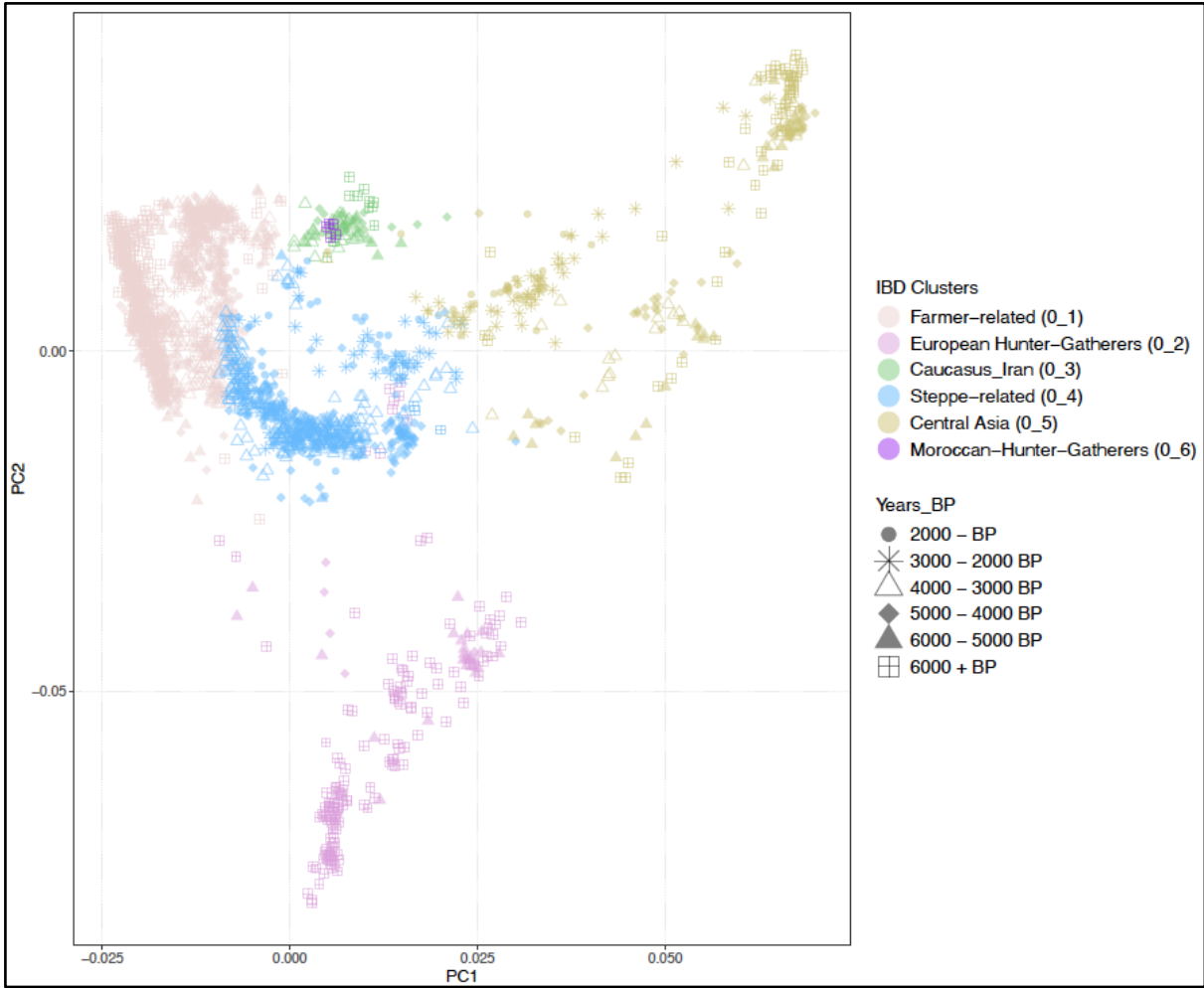
Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to: ewillerslev@sund.ku.dk, fulyay@sund.ku.dk, guus@hum.ku.dk, kristian.kristiansen@archaeology.gu.se

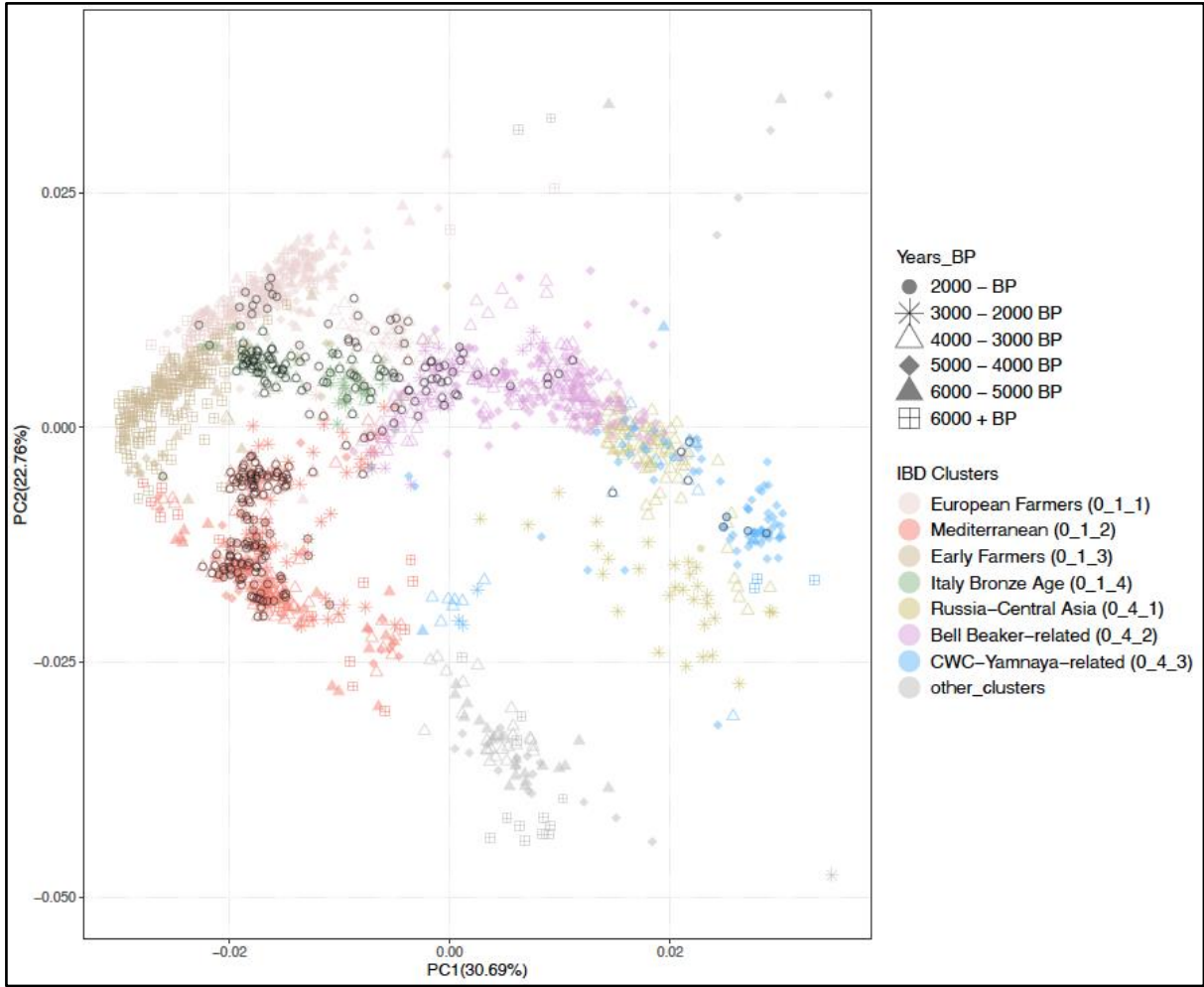
Extended Data Figures



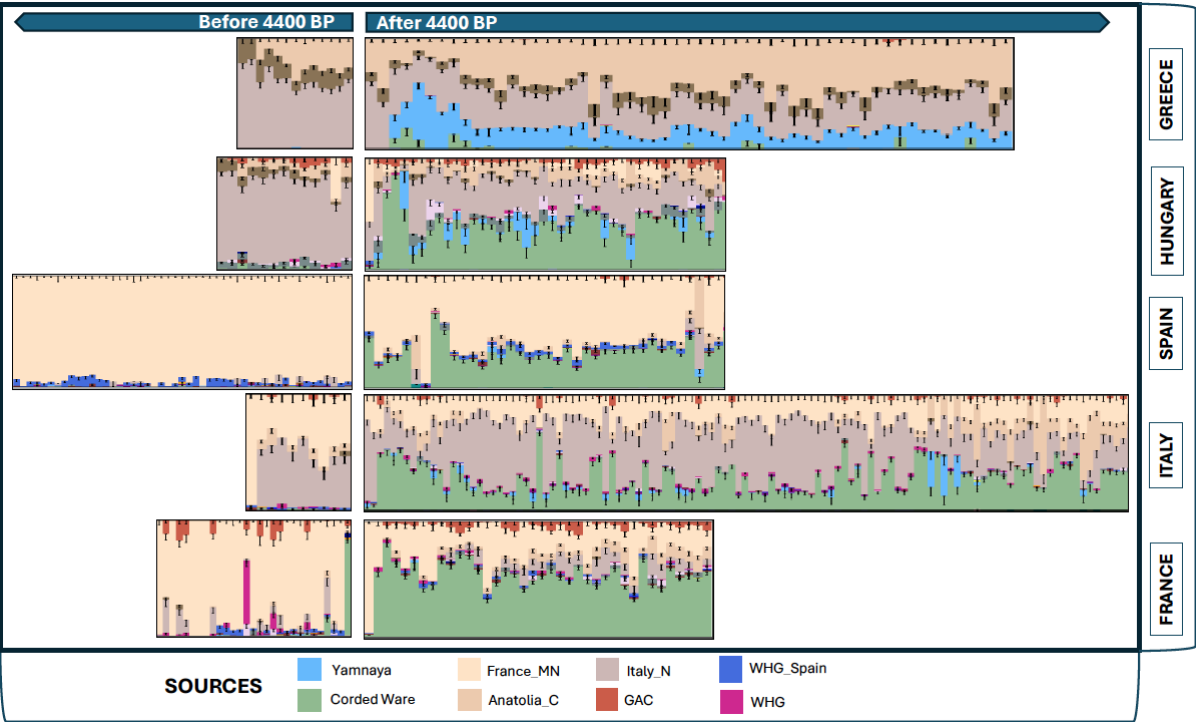
Extended Data Fig.1. Geographical distribution of the main IBD clusters, split into time ranges, pre 5,000 BP, 5,000–4,000 BP, and post 4,000 BP.



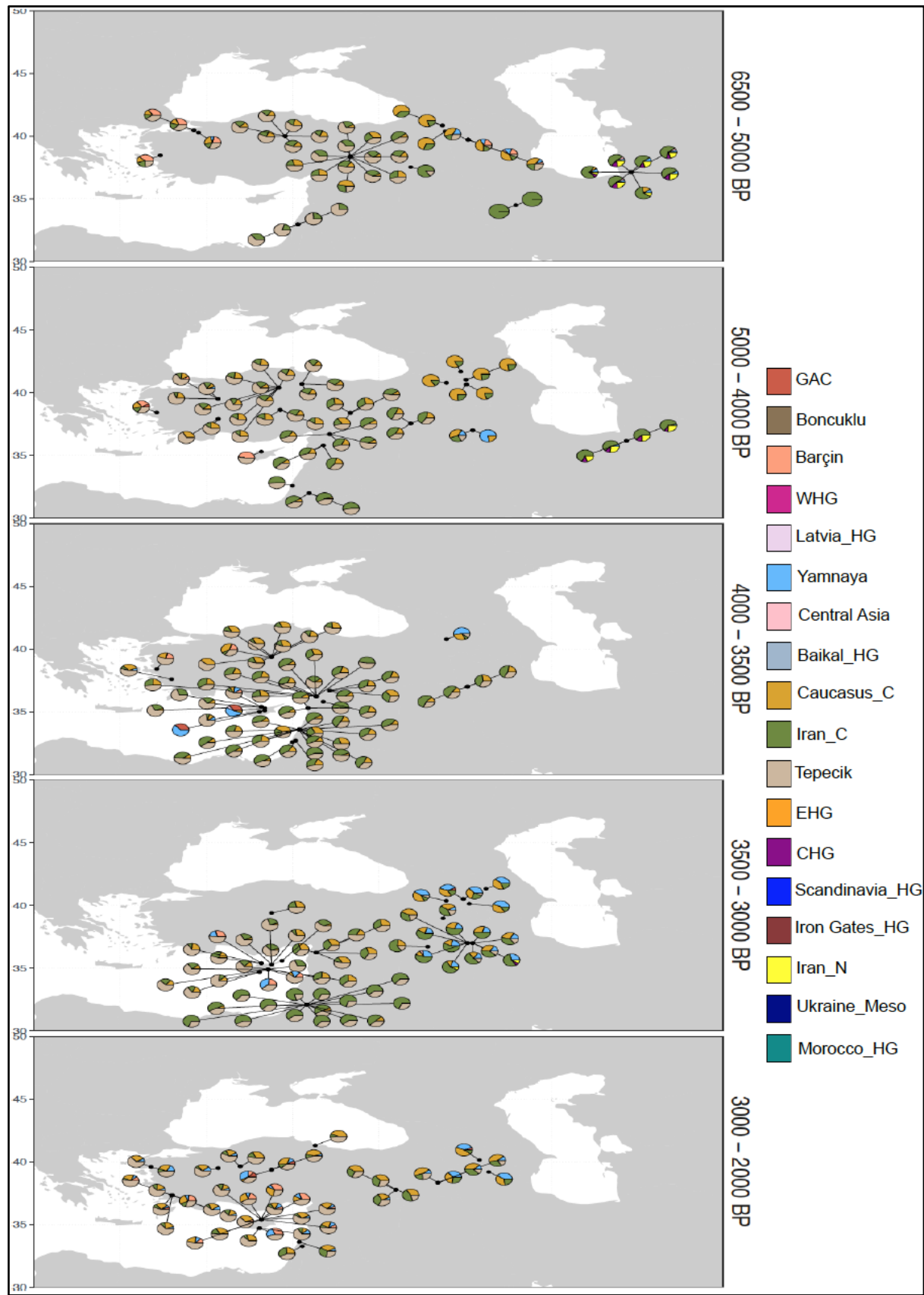
Extended Data Fig. 2. The PCA plot demonstrates the distribution of main IBD clusters on 2,228 ancient individuals.



Extended Data Fig. 3. The PCA plot demonstrates the distribution of subclusters of “Farmer-related (0_1)” and “Steppe-related (0_4)”. New genomes presented in this study were marked with black circle legends.



Extended Data Fig. 4. Ancestry bar plots generated for each individual using source population proportions of IBD admixture modelling sorted by time BP and divided into two time series, before and after 4,400 BP, illustrating a Southern and Central Eastern Europe split (Italy, France, Spain and Hungary vs Greece).



Extended Data Fig. 5. Pie charts generated by using the proportions of the applied IBD admixture model for each individual from Anatolia, Cyprus, Iran, Caucasus and Levant, divided into five time periods to avoid overlapping.

