bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ancient DNA reveals the origins of the Albanians

Leonidas-Romanos Davranoglou[1*], Aris Aristodemou[2], David Wesolowski[3], Alexandros Heraclides[4]

[1*]Oxford University Museum of Natural History, University of Oxford, UK; Leonidas-romanos.davranoglou@oum.ox.ac.uk

[2]Independent researcher; aenaristo@gmail.com

[3] Independent researcher; eurogenesblog@gmail.com

[4]School of Sciences, European University Cyprus, Nicosia, Cyprus; a.heraclides@euc.ac.cy

Abstract

The origins of the Albanian people have vexed linguists and historians for centuries, as Albanians first appear in the historical record in the 11th century CE, while their language is one of the most enigmatic branches of the Indo-European family. To identify the populations that contributed to the ancestry of Albanians, we undertake a genomic transect of the Balkans over the last 8000 years, where we analyse more than 6000 previously published ancient genomes using state-of-the-art bioinformatics tools and algorithms that quantify spatiotemporal human mobility. We find that modern Albanians descend from Roman era western Balkan populations, with additional admixture from Slavic-related groups. Remarkably, Albanian paternal ancestry shows continuity from Bronze Age Balkan populations, including those known as Illyrians. Our results provide an unprecedented understanding of the historical and demographic processes that led to the formation of modern Albanians and help locate the area where the Albanian language developed.

Introduction

During the Iron Age (1100 BCE–150 CE), the Balkans were characterised by remarkable cultural, linguistic, and genetic heterogeneity (*1–6*). In the western Balkans, "Celtic" cultures such as Hallstatt and La Tène, interacted for centuries with local groups referred to as the "Illyrians" and "Dalmatians" (*7*, *8*). Deep in the Balkan heartland, heterogenous populations named by classical authors as "Dacians", "Dardanians", "Moesians", and "Paeonians" (*1*, *7–9*) bordered nomadic cultures from the Pontic-Caspian steppe known as the "Scythians" (*10*), while the southeastern part of the peninsula was inhabited by "Thracians" and "Greeks" (*2*, *11*). Balkan peoples also expanded beyond the confines of the peninsula, with "Messapians" migrating to southeast Italy at least since 600 BCE (*12*). The linguistic and cultural diversity of the Balkans was considerably homogenised during the Hellenistic (*2*, *9*), Roman (*9*, *13–15*), and especially the Migration Period, when Germanic and Slavic-speaking groups massively settled in the region (*2*, *3*, *16*, *17*). These events ultimately led to the extinction of all palaeo-Balkan languages except Greek and Albanian. The latter is one of the most enigmatic branches of the Indo-European language family, having vexed linguists for more than two centuries (*2*, *18*, *19*).

Tracing the origins of Albanians and their language is challenging for several reasons. Only a handful of historical sources comment on the ethnic and linguistic composition of the southwest Balkans during the transition from classical antiquity to Medieval times (500-1000 CE) (*20–22*),

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and none of them mention an Albanian-speaking population from the territory of modern Albania. Speakers of Slavic languages are reported to have inhabited what is now southern Albania in the 8th century CE (*15*, *21*), where the frequency of Slavic toponyms also peaks (*23*), while the same region is characterised by the presence of Greek-speakers at least since the Medieval period (*24*). The urbanised Medieval populations of the northwest, referred to by contemporary historians as the Romani/Ῥωμᾶνοι (*20*), are thought to have spoken a variant of vulgar Latin known as West Balkan Romance (*14*, *15*, *25*) that persisted at least until the 13th century CE (*26*). The demographic and linguistic situation in the mountainous interior is unknown, and it is only in the 11th century CE that Albanians appear in the historical record (*22*), while the earliest surviving written document of their language dates to 1462 CE (*27*).

A number of linguistic hypotheses have attempted to identify the affinities of the Albanian language and to locate the region where it developed, yet no definitive conclusions have been drawn. The most prominent, mutually exclusive hypotheses can be divided into those arguing for a local west Balkan origin from an Illyrian (*28*, *29*) or Messapic background (*19*, *30*, *31*) [which may or may not have been distinct languages (*7*, *30*, *32*)], and those proposing a non-local origin from a Daco-Moesian-Thracian background (*2*, *19*, *33*) or an unattested Balkan language, whose speakers entered Albania from the central-east Balkans sometime after 400 CE (*15*, *32*, *34*, *35*). The validity of these hypotheses, although hotly debated, is hard to test, as these ancient languages are poorly recorded, being known only from fragmentary inscriptions, toponyms, and a handful of historical sources (*2*, *7*, *36*). Furthermore, all of the ethnonyms of ancient Balkan peoples, such as "Illyrian" and "Thracian", are likely artificial labels that were coined by ancient and modern authors (*37*), and may include several related languages with largely obscure geographical limits, intelligibility, and emic identities of their speakers (*8*, *9*, *32*). The most recent linguistic hypotheses propose a sister-group relationship of Albanian to Greek or to the Greek-Armenian clade (*18*, *38*, *39*), which firmly places the origin of the language in the Balkans but does not pinpoint the location of the proto-Albanian homeland within the peninsula and its potential affiliation to historically attested populations.

Archaeological data on Albania's Medieval cultures are also inconclusive, especially for the Komani-Kruja complex (*ca*. 600-800 CE), which has been interpreted as the cultural expression of either a Romanised population (local or intrusive) (*8*, *40*, *41*) or an indigenous Albanian-speaking group (*42*).

Due to the challenges associated with linking archaeological, literary, and linguistic evidence, an archaeogenetic approach may offer novel insights into the origin of the Albanians, their biological relationships to ancient people, and the affinities of their language. Although gene flow is not always accompanied by language shifts [as in the case of Basque (*43*) and Etruscan (*44*)], migration is one of the primary vectors of cultural change (*45*, *46*), of which language dissemination is a frequent outcome (*3*, *17*, *47–49*). Recent years have witnessed a surge in the palaeogenomic sampling of the Balkan peninsula (*3–6*, *11*, *16*), yet the resulting datasets have not been mined to help us understand how migration led to the emergence and spread of new material cultures, communities, and languages in the territory of modern Albania.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To gain insight into the biological and linguistic origins of modern Albanians, we undertake a palaeogenomic transect of the Balkans from the Neolithic to the modern era (Fig. 1; Tables S1-S2), using more than 6000 previously published ancient genomes from western Eurasia, which we interrogate by means of state-of-the-art statistical analyses (Tables S3-S18), large-scale algorithms that quantify human mobility (Table S19), and a meta-analysis of ancient IBD-sharing data [Tables 20-21; (*50*)]. We also mine publicly available Y-chromosome haplogroup data from more than 2500 ancient and modern Balkan samples (Tables S22-S34), which reveal the proximate ancestors of modern Albanian men.

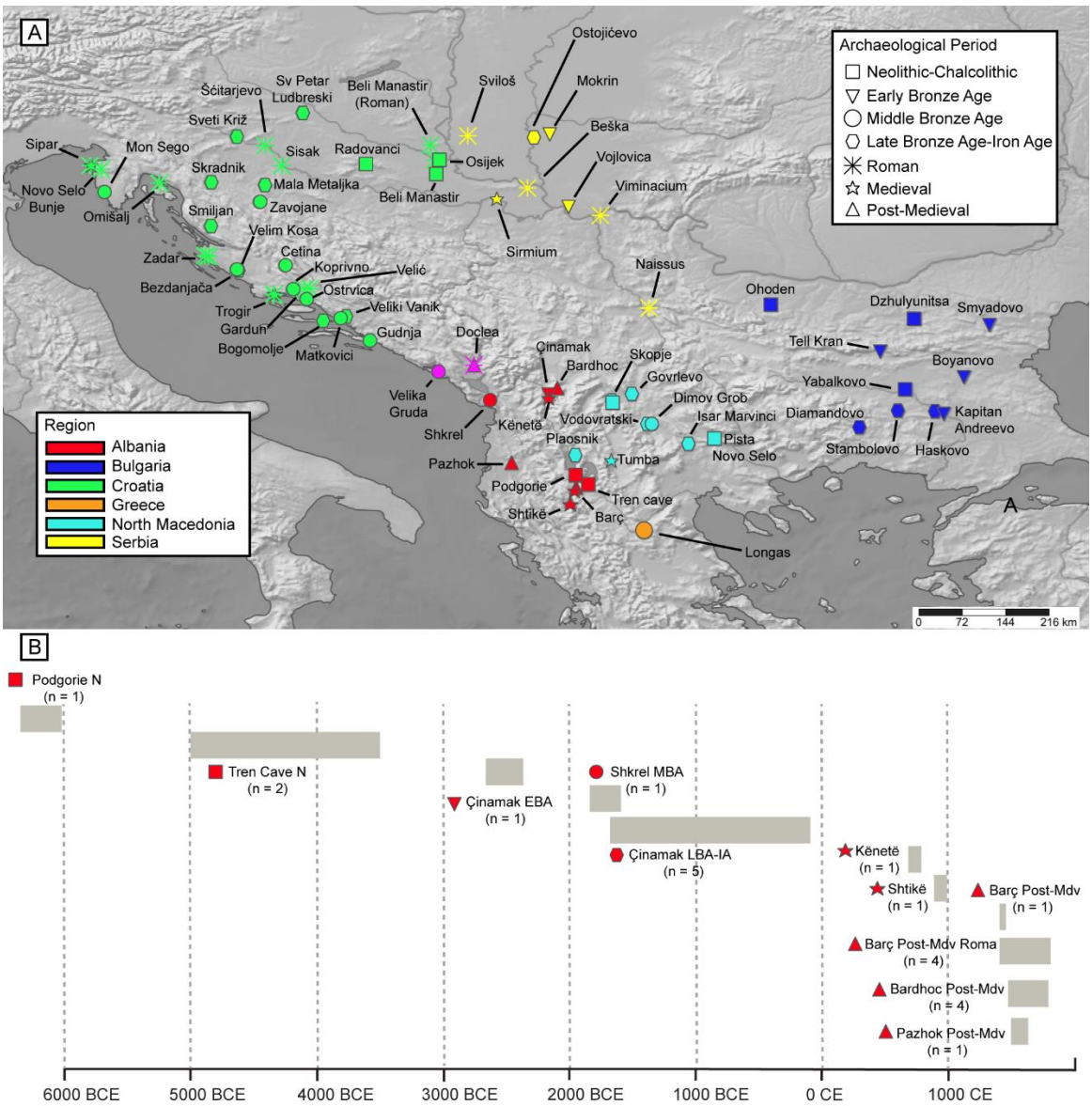


Figure 1. Geographical distribution and dating of the examined Balkan populations. A) Archaeological sites annotated by period. B) Analysed ancient individuals from the territory of modern Albania arranged by archaeological site and date (radiocarbon or archaeological chronology).

Results

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Complex genetic transformations during the Balkan Stone to Bronze Age transition

To visualise variation in genetic composition, we perform principal components analysis (PCA), where we project previously published ancient individuals from the Balkans and adjacent regions onto present day West Eurasian populations genotyped with the Human Origins array (Fig. S1). Three distinct population clusters can be seen on PC1 and PC2 of our Stone Age-Early Bronze Age PCA (Fig. 2). Hunter-Gatherer (HG) populations, which include Western Hunter Gatherers and Iron Gates Hunter-Gatherers (IGHG), occupy the right side of PC1; Early European Farmers (EEF), who introduced agriculture to Europe, cluster on the bottom left corner of PC2; and Yamnaya steppe pastoralists, who are attributed with the expansion of Indo-European languages to Eurasia and beyond, plot at the uppermost space between PC1 and PC2.

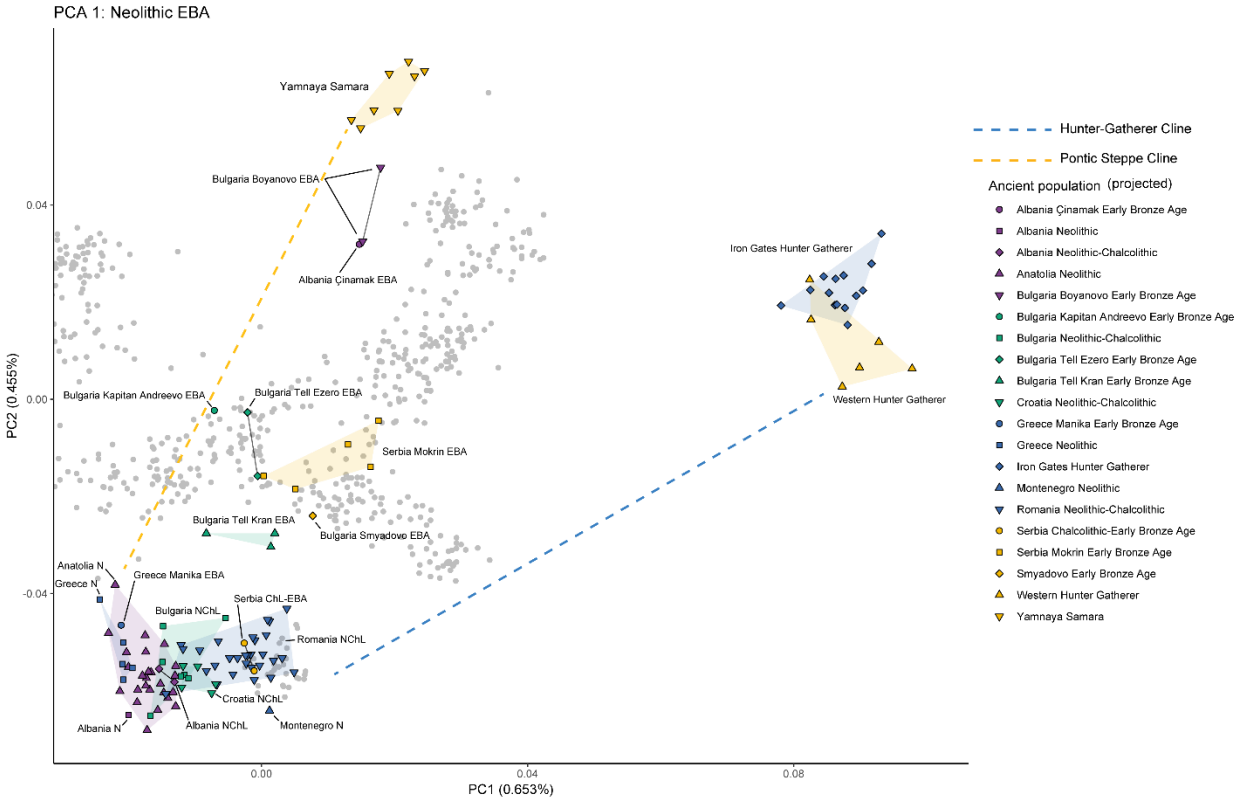
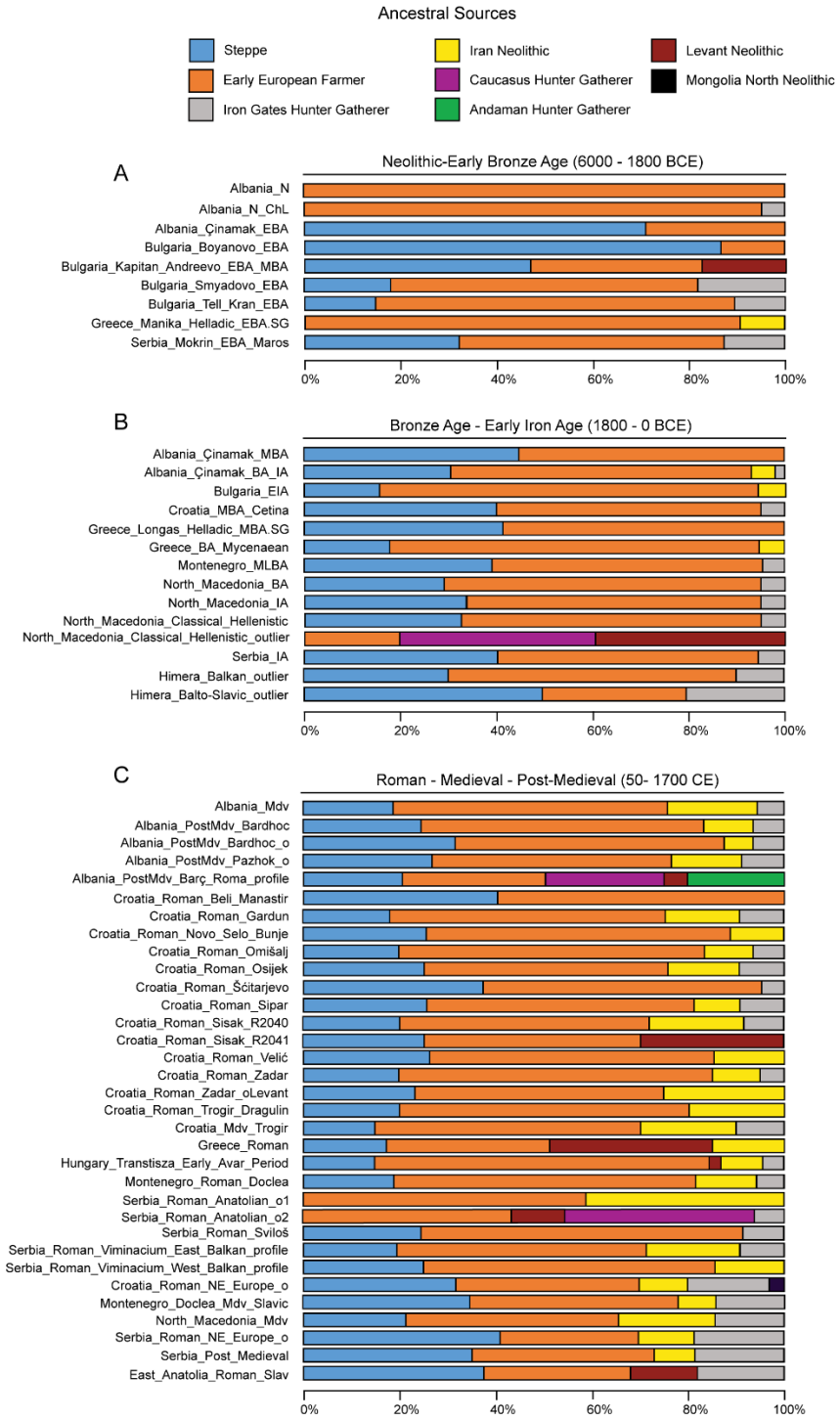


Figure 2. PCA of modern West Eurasian samples (grey points) with projection of ancient Neolithic-Early Bronze Age individuals from the Balkans. Dashed lines indicate the two poles of observed ancestry shifts from the Neolithic onwards – one generated by the assimilation of hunter-gatherers (blue line), and another by admixture with incoming Yamnaya-related populations from the Pontic-Caspian steppe (yellow line). Samples clustering within the PCA space enclosed by HG, EEF, and Yamnaya derive variable amounts of ancestry from these populations.

The PCA shows that the single Early Neolithic sample from Albania clusters with contemporary farming populations from Greece and Anatolia (Fig. 2), an affinity that is further supported by outgroup *f3*-statistics (Fig. S2A), which measure shared genetic drift between a pair of populations, as well as formal *f4*-statistics *qpAdm* tests with rotating sources (Table S3). No major genetic changes are observed in samples dating to the Neolithic-Chalcolithic transition (Fig. 2), with the

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

individuals from Albania (Alb NChl) being modelled as deriving their entire ancestry from the preceding Albanian Neolithic, possibly with an additional 5% admixture from an IGHG-like source (Figs. 3A, S3; Table S4).



bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3. Distal *qpAdm* models for Balkan populations from the Neolithic to post-Medieval period. A) Neolithic and Early Bronze Age population models. B) Bronze-Iron Age models. C) Roman-post-Roman models. Putative ancestral source populations are colour-coded and presented in the top part of the figure. The single model receiving the highest statistical support is presented (all possible models and the criteria used to select the ones shown here are provided in Tables S3-S18).

Unlike the genetic continuity observed during the Neolithic-Chalcolithic transition, the PCA indicates remarkable ancestry shifts in the Balkans during the Early Bronze Age (EBA) (*ca*. 2600-1800 BCE). Most EBA Balkan samples show a significant change in their PC position and plot towards the Yamnaya (Fig. 2), suggesting the arrival of the steppe cultural package in the region, including Indo-European speech (*4*, *11*, *43*, *47*, *49*, *51–53*). The Albania Çinamak EBA sample, dated to 2663-2472 BCE (Fig. 1B; Table S1), together with two samples from Bulgaria Boyanovo EBA (2500-2000 BCE), cluster closest to Yamnaya (Fig. 2), and *qpAdm* tests model them as having roughly 70% and 80-85% steppe ancestry, respectively, with the remainder of their genome deriving from local farming populations (Fig. 3A; Tables S5, S7). Accordingly, both male samples from Albania Çinamak EBA and Bulgaria Boyanovo EBA carry Y-chromosome haplogroup R1b-M269 (Tables S22, S26), the primary indicator of paternal ancestry from the Pontic-Caspian steppe (*6*, *49*, *53*). Importantly, steppe-derived admixture in Albania Çinamak EBA is best modelled with largely unadmixed Yamnaya groups rather than with EEF-and-HG-admixed Corded Ware populations (Fig. S2C; Table S7), suggesting that Indo-European-related ancestry arrived in the Balkans directly from the Pontic-Caspian steppe, in accordance with previous studies (*6*). This is also supported by the presence of Y-chromosomal haplogroup R1b-Z2103 in the Bulgaria Boyanovo EBA sample, which has been found in the earliest Yamnaya contexts (*6*, *49*, *53*), whereas Corded Ware samples are characterised by largely different subclades of haplogroup R1 (R1a-M417, R1b-U106, R1b-L151) (*6*, *45*, *54*). The ancestors of Albania Çinamak EBA were likely recent arrivals from the steppe, as this individual lived during the transitional period where the first kurgan burials made their appearance in Albania (*55*). Contemporary EBA archaeological finds from tumuli in Shkodër (100 km to the west of Çinamak) include artefacts associated with the Vučedol culture (*55*), revealing links with the northern Adriatic and Pannonia. It should be noted that a male with high levels of steppe ancestry discovered in a Vučedol context in Croatia also belonged to Y-chromosomal haplogroup R1b-Z2103 (*56*).

Regarding the broader Balkan EBA context, early interactions between incoming nomadic pastoralists and local farming populations were likely complex, as evidenced by the heterogenous levels of steppe admixture among geographically adjacent groups, which range from 0-7% (Greece Manika EBA), 16% (Bulgaria Tell Kran EBA), 30% (Serbia Mokrin EBA Maros) to *ca*. 45% (Bulgaria Kapetan Andreevo EBA) (Fig. 3A; Table S5). These differences in ancestry are reflected in the PCA, where the abovementioned groups occupy a middle space between largely unadmixed Yamnaya and EEF populations (Fig. 2).

The establishment of the BA-IA Balkan genetic cline

We next examined the projections of Bronze Age (BA) and Iron Age (IA) Balkan populations upon the principal components generated earlier (Fig. 4). By the Middle-Late Bronze Age (MLBA) and the Iron Age, the genetic ancestry of Balkan populations had become less heterogenous,

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

exhibiting a north-to-south cline that broadly reflects geography (Fig. 4A). Accordingly, the MLBA-IA samples from Çinamak in Albania are clinal between contemporary populations from Croatia and Montenegro on one side of PC1 and PC2, and from North Macedonia and northern Greece on the other side (Fig. 4B). This affinity is further corroborated by *f3*-statistics (Fig. S2D, E) and Mobest analysis (Fig. S4), which employs an algorithm for spatiotemporal mapping of genetic profiles using bulk aDNA data (*57*). Formal *f4*-models with *qpAdm* using ultimate sources indicate a uniform genetic profile and a resurgence of farmer ancestry across the Central-West Balkans and northern Greece compared with the preceding EBA, as most samples derive around 60% of their ancestry from EEFs, 0-5% from Iran N, and 30-40% from steppe populations (Fig. 3B; Table S5). In the southern end of the Balkan genetic cline, Bronze Age and Iron Age populations of central-southern Greece and Bulgaria have considerably higher EEF (75-80%) and Iran N (5-10%) ancestry, and a much lower proportion of steppe ancestry (15-20%) (Fig. 3B; Table S5).

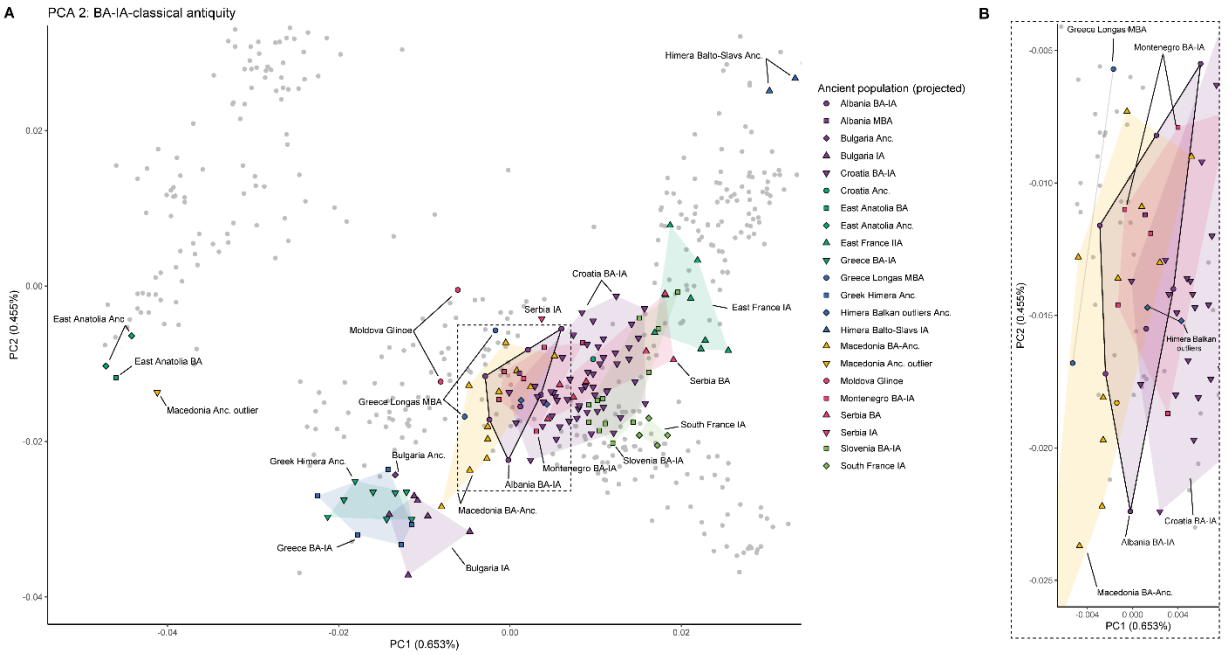


Figure 4. PCA of modern West Eurasian samples (grey circles) with projection of ancient Bronze Age and Iron Age individuals from the Balkans. A) PCA of all tested populations, showing a north-to-south cline that mirrors geography. B) Detail of region enclosed in the dotted rectangle in panel A, showing the clustering of populations from Albania (enclosed in solid line).

The close clustering of BA-IA populations from Albania, Croatia, Montenegro, North Macedonia, and northern Greece is also confirmed in proximate *qpAdm* models, as the Çinamak MLBA-IA samples derive most of their ancestry from the West Balkans (Tables S8-S9), with a possible 15-25% contribution from a southeast Balkan source (Bulgaria EIA, Greece BA Mycenaean) after the Middle Bronze Age (MBA) (Table S9). Based on the above, the MBA-IA populations of a large geographic region spanning northern Greece, North Macedonia and the entire Adriatic coast,

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

including the region of modern Albania, form a uniform genetic cluster with similar admixture proportions (Fig. 3B) that persists for at least 1.500 years and transcends the linguistic boundaries identified by classical authors (*7*, *9*). Our findings are further reinforced by IBD-sharing between certain samples from Albania and North Macedonia (Table S20) (*50*).

Intriguingly, two mercenaries from the Battle of Himera in Sicily (*58*) fall within the PCA cluster of BA-IA Albania (Fig, 4B). This position might be coincidental, as these two samples are characterised by high proportions of IGHG-related ancestry (9%), which in the Balkans can only be found in populations from EBA Serbia (Fig. 3B; Table S5). Furthermore, the proximate ancestry sources of these mercenaries are uncertain using both *f3*-and-*f4*-statistics (Fig. S5; Table S10), suggesting they might derive from a currently unsampled population, likely from the Central Balkans.

An outlier from Hellenistic North Macedonia might represent an early migrant from the Near East, as he is shifted towards populations from the Caucasus (Fig. 4A), and his ultimate ancestry is successfully modelled with proxies from the Caucasus and the Levant (Fig. 3B; Table S5).

Albania as a refugium of Iron Age ancestry during the rise and fall of the Roman Empire

It has been previously demonstrated (*3*, *16*) that during Imperial Roman times, or perhaps even the preceding Hellenistic era, as we show here (Fig. 4B), large-scale westward immigration from the Eastern provinces took place, which transformed the genetic landscape of the Empire (*59*). This demographic shift also took place in the Balkans (*3*), as can be seen in the PCA of Roman and Post-Roman samples (Fig. 5), where virtually all local populations are shifted from their Bronze Age and Iron Age position towards the Eastern Mediterranean, forming a cluster along what we term as the Anatolian cline (Fig. 5A).

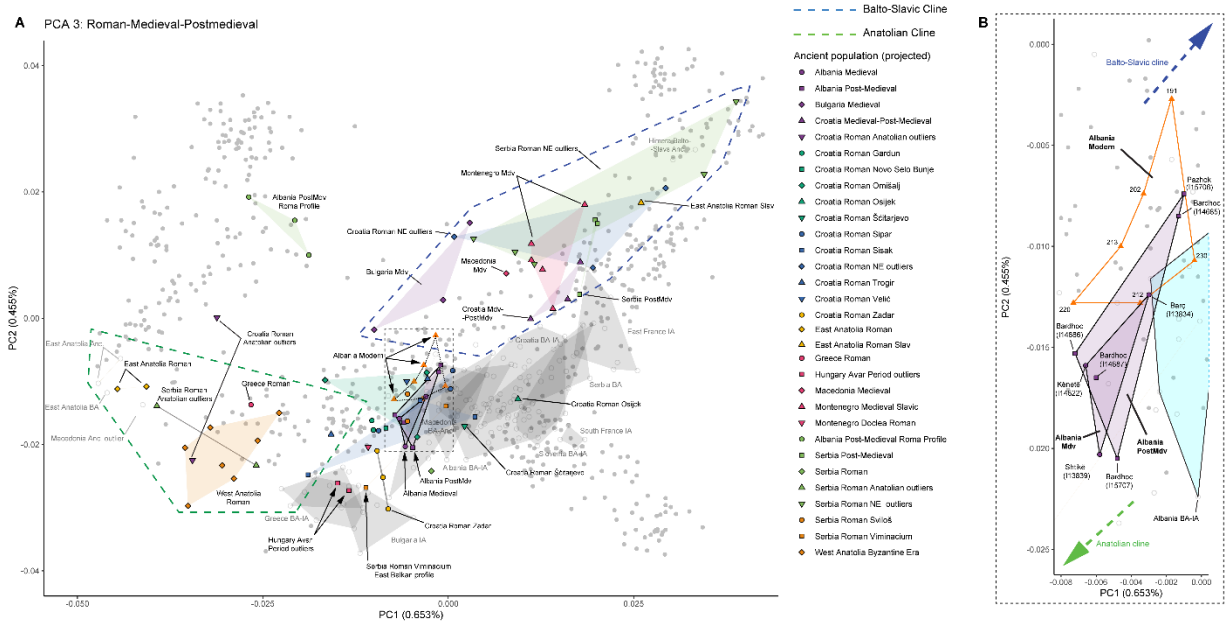


Figure 5. PCA of modern West Eurasian samples (grey circles) with projection of ancient Roman, Medieval, and post-Medieval individuals from the Balkans. A) PCA of tested populations, highlighting the

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shift in the PCA coordinates of Roman and post-Roman groups towards the Anatolian (green dashed line) and Balto-Slavic (blue dashed line) clines, compared to their counterparts from the same regions in the preceding Bronze Age and Iron Age (faded grey polygons). B) Detail of region in panel A, showing the tight clustering of populations from Albania over the past 3000 years, including modern individuals (orange polygon). For simplicity, Roman-era populations within and close to this cluster have been removed and can be seen in the previous panel.

Formal *f4*-statistical tests using *qpAdm* also support the divergence in Balkan populations between the BA/IA and the Hellenistic/Roman era, by modelling post-Neolithic Anatolian admixture with an Iran N-like reference (Fig. 3C), which became pervasive in Anatolia and the Aegean by the EBA (*60*), and was also often accompanied by Levantine ancestry (*59*). Although largely absent in the preceding Iron Age, Iran N-like admixture entered Balkan populations in a multiphased manner, with some Roman-era-Early Medieval populations being modelled as deriving 0-15% of their ancestry from this source (Croatia Roman Beli Manastir and Šćitarjevo, Montenegro Doclea Roman), which in adjacent regions can be as high as 20-30% (Croatia Novo Selo Bunje, Zadar, and Trogir Dragulin) (Fig. 3C; Table S5). These differential ancestry shifts are also corroborated using ADMIXTURE on a subset of our dataset, represented by the appearance of a West-Central Asian component that was absent prior to the Roman period (Fig. S3).

While Anatolian ancestry was becoming established in the Balkans, a second, even more significant demographic transformation began to take place. By 200 CE, individuals related to north-eastern European, Balto-Slavic, and nomadic steppe populations started appearing in the Balkans (Fig. 5A). Such movements peaked during the Migration Period (roughly 350-600 CE) with the mass migration and settlement of Avar, Germanic and Slavic-speaking groups, which transformed the cultural, linguistic, and ethnic composition of the region (*3*, *9*, *17*, *61*). Admixture with these newcomers was pervasive, as can be seen in the PCA, where most Balkan populations show a significant shift towards Balto-Slavic groups, along what we term the Balto-Slavic cline (Fig. 5A). This admixture event is mirrored in an unprecedented increase of IGHG-related ancestry in most late Roman-Early Medieval Balkan populations, which is most significant in samples located within the Balto-Slavic cline (Fig. 3C; Table S5). Both *qpAdm* and ADMIXTURE models suggest that from an average of 4-5% in the preceding BA-IA and the early Roman era, IGHG ancestry rose to 13-18% in strongly Balto-Slavic-shifted populations from late Roman, Medieval, and Post-Medieval Croatia, Macedonia, Montenegro, and Serbia (Fig. 3C; Fig. S3; Table S5).

To quantify the impact of north-eastern European migrations into the Balkans, we used *f4* rotating-source *qpAdm* tests with proximate sources dating to the Iron Age, Bronze Age and the Roman era, where Russia Ingria IA served as a proxy for Balto-Slavic-related ancestry. Our *qpAdm* models recover remarkably high Balto-Slavic-related ancestry in the late Roman and Post-Roman populations of Croatia (50-65%), Montenegro (45-65%), North Macedonia (30-50%), and Serbia (50-55%) (Table S11). These admixture proportions are almost identical to those proposed by studies on modern Slavic-speaking populations using different methods, which recover the ancestry of South Slavs as 55-70% Balto-Slavic-related, with the remainder originating from the pre-Slavic inhabitants of the Balkans (*48*), further affirming the accuracy of our models.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Considering the abovementioned Anatolian and Balto-Slavic ancestry shifts in the Roman and Medieval Balkans, we model the ancestry of samples from Medieval Albania. In contrast to neighbouring populations, the samples from Medieval South-Eastern (Shtikë, 889-989 calCE) and North-Eastern (Kënetë, 773-885 calCE) Albania (hereafter Albania Mdv) experience only a minor shift on their PC position from the BA-IA to the Migration Period (Fig. 5A, B), hinting at large-scale genetic continuity for over 2500-3000 years. This is reflected in the ancestry makeup of Albania Mdv, as in ultimate *f4 qpAdm* models they derive 17% of their ancestry from Iran N-like sources (2-5% in the BA-IA), while their IGHG ancestry increases only marginally (0-4% in BA-IA, 6% in Medieval times) (Fig. 3C; Table S5). Proximate *qpAdm* models comprising Balkan BA-IA sources and proxies for Anatolian-Levantine (East Anatolia BA IA; Syria Ebla EMBA) ancestry also replicate this ancestry shift (Table S12). The most strongly supported result was a two-way model where Albania Mdv derive 85% of their ancestry from BA-IA-Hellenistic Albania, and 15% from either the Anatolian or Levantine proxy, with equal support (Table S12).

To further resolve the observed ancestry patterns, we ran a second proximate model using Roman era West Balkan sources, together with a Slavic proxy (Russia Ingria IA), where we recovered Albania Mdv as either 100% Roman West Balkan, or 85% Roman West Balkan and 15% Slavic-related (Table S13). Accordingly, Albania Mdv cluster together with Roman era West Balkan samples that derive only a small proportion of their ancestry from Anatolian populations [Croatia (Beli Manastir, Gardun, Omišalj, Sisak, Trogir, Velić, Zadar); Montenegro (Doclea); Serbia (West-Balkan-shifted samples)] (Fig. 5B) and show little to no increase in IGHG ancestry (Fig. 3C).

Populations of largely unadmixed palaeo-Balkan ancestry persisted in pockets in other regions of the collapsing Roman Empire as well. Two samples from Early Avar Pannonia (550-650 CE) cluster with populations from Iron Age Bulgaria and Greece on the PCA (Fig. 5C), a relationship corroborated by *f3*-statistics (Fig. S6), *qpAdm* (Table S14), and IBD-sharing (Table S21) (*50*). Such outliers may affirm historical reports of the Avars undertaking mass resettlements of Roman subjects from the area of Thrace towards their Khaganate in Pannonia (*62*, *63*).

Given that some linguistic hypotheses suggest a mixed West Balkan and Thracian origin for the Medieval population of Albania (*2*, *64*), we undertake an *f4 qpAdm* test combining Roman era West Balkan, East Balkan (the two outliers from Early Avar Hungary), and Slavic-related (Russia IA Ingria) sources. Albania Mdv is once again effectively modelled as being 100% of Roman era West Balkan origin (Croatia Roman Gardun, Montenegro Roman Doclea), or as a two-way mix of a Roman period West Balkan source (85%) and a Slavic-related source (15%) with high support ($p = 0.16$; $SE = 0.04$) (Table S15). Two-way or three-way models with West Balkan + East Balkan or West Balkan + East Balkan + Slavic-related sources also passed with low statistical support ($SE = 0.12$ and $SE = 0.09$-0.14, respectively), recovering Albania Mdv as deriving 45-60% of their ancestry from the two Thracian-shifted Avar era outliers (Table S15), a finding that may also receive support from *f3*-statistics (Fig. S2F). Furthermore, models using the Avar era outliers as the sole palaeo-Balkan source for Albania Mdv receive high support ($p = 0.72$; $SE = 0.03$; Table S15). However, it is unlikely that the Avar era outliers are realistic local sources due the fact that such significant ancestry shifts would have pulled Albania Mdv samples towards the direction of

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

southeastern Balkan populations on the PCA, which is not the case (Fig. 5A, B). Overall, the PCA and *f4 qpAdm* statistical models suggest that the Medieval population of Albania was minimally affected by demographic changes during the Roman era, in stark contrast to adjacent regions such as Croatia and Serbia (Fig. 5A). However, currently unsampled urbanised areas in Albania such as Durrës and Shkodër, may have comprised populations of more complex ancestry.

Although we show that some West and East Balkan populations persisted largely unadmixed in late Roman and early post-Roman times (Fig. 5A), such ancestry profiles cannot be found post-900 CE, as Mobest analysis of Albania Mdv shows affinities only to populations from Italy, which maintained a larger proportion of Eastern Mediterranean ancestry compared to contemporary sampled locations in southeastern Europe (Fig. 6B, C). This suggests that the region of modern Albania served as a refugium of Iron Age West Balkan ancestry throughout the demographic and social upheaval that took place during the Migration Period (*9*, *17*, *61*). However, our *qpAdm* models cannot discriminate whether West palaeo-Balkan ancestry in Medieval Albania originated from indigenous populations or other incoming palaeo-Balkan groups from adjacent regions such as the northern Adriatic coast or the Balkan interior. Indeed, IBD-sharing suggests that the sample from Kënetë shares small (10.5 cM) segments with BA and Roman individuals from neighbouring Velika Gruda in Montenegro and Zadar in Croatia, respectively (Table S20) (*50*). Conversely, the individual from Shtikë shares short segments (11-14.5 cM) with geographically distant Hun and Avar-era commoners from Hungary (Table S20), suggesting that possible trade networks between Medieval Albania and the Avar Khaganate (based on archaeological artefacts) (*8*, *15*, *40*) may have also involved genetic exchange.

The formation of the modern Albanian genome

We next sought to characterise the ancestry of post-Medieval samples from Central (Pazhok, 1527-1660 calCE) and north-eastern (Bardhoc, 1400-1700 calCE) Albania. The place-name Bardhoc is of Albanian etymology (*35*), and is mentioned in Ottoman registers contemporary to the studied samples (*65*, *66*), suggesting the latter might have been Albanian-speakers. Furthermore, in the adjacent region of Has, a majority of the population is recorded as Albanian by contemporary Ottoman sources, although speakers of South-Slavic languages were also present (*65*).

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

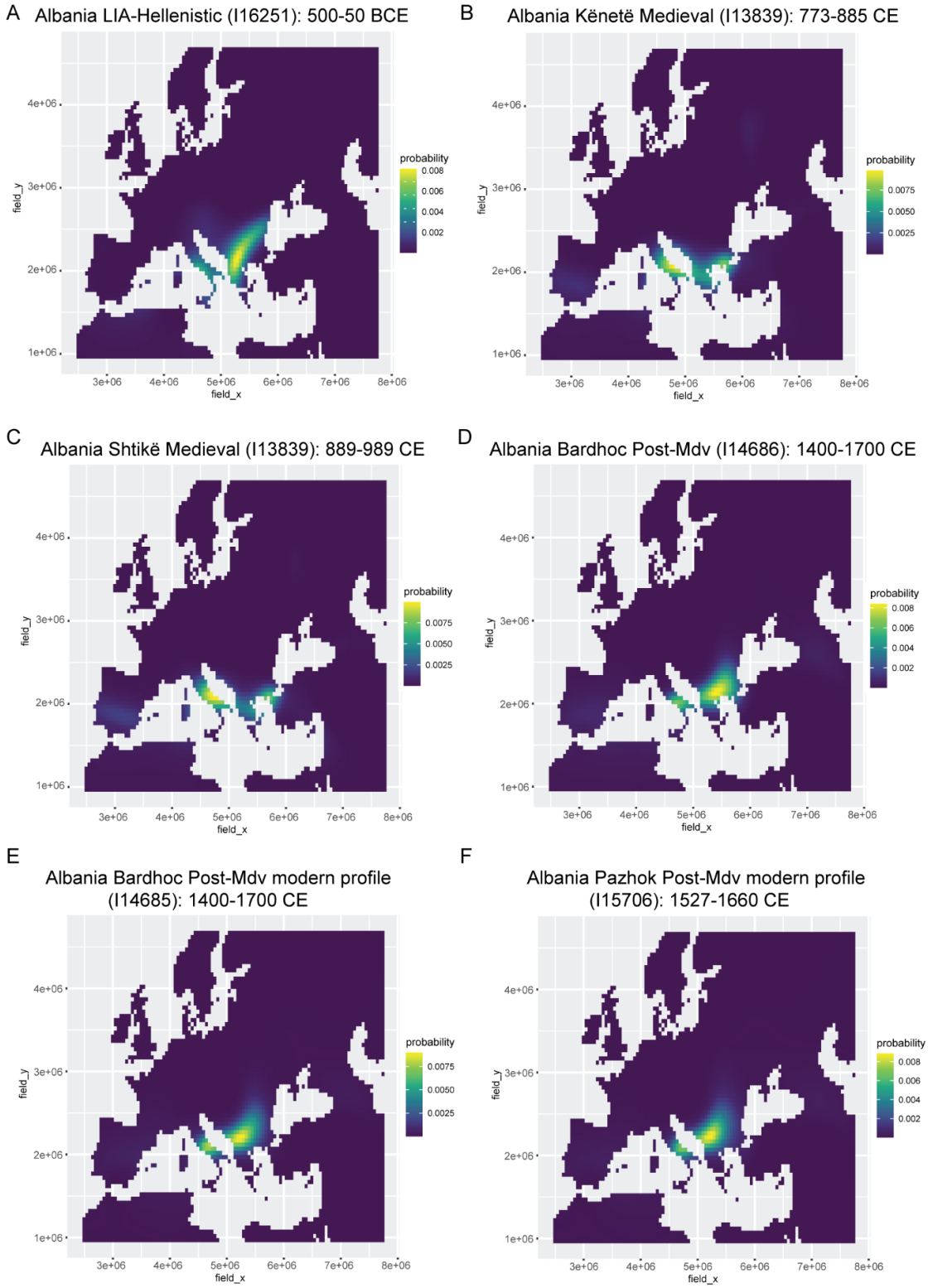


Figure 6. Mobest analysis of Late Bronze Age-Iron Age, Medieval, and post-Medieval samples from the region of Albania, plotting the probability surface that identifies the highest genetic-geographical match at the mean date of the respective individual. The higher the probability surface (light yellow-green), the closer

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the genetic match. Only a single sample from the post-Medieval population of Bardhoc is included here, as all individuals from this region display the same probability surface (Fig. S7A-C). The latitude and longitude coordinates with the best fit for the examined individuals are provided in EPSG:3035 projection in Table S19.

In the Roman-Post-Roman PCA, individuals from post-Medieval Bardhoc (hereafter Bardhoc PostMdv) cluster with the preceding Medieval samples from Kënetë and Shtikë, except for one outlier (Fig. 5B) who clusters with the Pazhok sample, both of which are pulled towards the Balto-Slavic cline (Fig. 5B), which was confirmed with *f3*-statistics (Fig. S2H, I). We employ a two-way *f4*-statistics *qpAdm* model to test whether the pattern observed in the PCA reflects shared ancestry, where Albania Mdv serves as a local source (Table S16). To test for excess north-eastern-European-related admixture, the second ancestry source is modelled with a proxy for unadmixed Balto-Slavic-related ancestry (Russia IA Ingria), as well as a suite of palaeo-Balkan-admixed populations within the Balto-Slavic cline dating to the Roman (Croatia Roman NE Europe o, Serbia Roman NE Europe o), Medieval (Macedonia Medieval, Montenegro Doclea Slavic, East Anatolia Roman Slav), and post-Medieval (Serbia Post-Medieval) periods (Tables S16-S17).

One-way models for Bardhoc PostMdv using Albania Mdv as a source receive very high statistical support, while two-way models with an additional low-level contribution from unadmixed (2%) or admixed (5-10%) Balto-Slavic-related sources are also feasible (Fig. 7; Tables S16-S17). However, a one-way *f4 qpAdm* model for the two Balto-Slavic-shifted outliers from Bardhoc and Pazhok is rejected (Tables S16-S17), suggesting more complex ancestry in these individuals. These outliers are instead recovered as deriving their ancestry from a two-way mixture between a local source (Albania Mdv) and unadmixed (ca. 15%) or admixed (ca. 25-35%) Balto-Slavic-related groups (Fig. 7; Tables S16-S17). Heterogeneity in Slavic-related admixture also characterises the modern Albanian samples from Tirana, which ranges from 18-48% (Fig. 7). It is likely that modern populations inhabiting areas in Albania that experienced little to no Slavic settlement in Medieval times, such as the south-west (Labëria) and the north (Malësi e Madhe, Rrethi i Matit) (*8*, *14*, *15*, *23*), will harbour less Balto-Slavic-related ancestry than the current modern Albanian samples.

Although models using the unadmixed Balto-Slavic-related proxy (Russia IA Ingria) receive higher statistical support (Fig. 7; Tables S16-S18), they are less likely historically, as there is no evidence that such individuals persisted in the Balkans post-600 CE based on the PCA (Fig. 5A). Likewise, Mobest analysis on the post-Medieval samples of Bardhoc and Pazhok reveals highest genetic similarity to neighbouring South-Slavic populations and not to north-eastern unadmixed Balto-Slavic groups (Fig. 6D-F). It is therefore likely that Slavic admixture entered the Albanian population via an ancestry profile related to modern South Slavs, rather than unadmixed Balto-Slavic populations. Regarding the origins of such South Slavic-related admixture in post-Medieval and modern Albanian populations, *qpAdm* models with north-western South Slavic

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Target	Source 1 (local)	Source 2 (Balto-Slavic-related)	*P* value	Coeff. 1	Coeff. 2	SE
Albania Post-Mdv Bardhoc	Albania Mdv		0.2	100%	-	-
Albania Post-Mdv Bardhoc	Albania Mdv	North Macedonia Mdv	0.16	90%	10%	0.14
Albania Post-Mdv Bardhoc	Albania Mdv	Serbia Roman NE o	0.18	95%	5%	0.04
Albania Post-Mdv Bardhoc	Albania Mdv	Russia IA Ingria	0.67	98%	2%	0.03
Albania Post-Mdv Bardhoc	Albania Mdv	Croatia Roman NE o	0.21	96%	4%	0.05
Albania Post-Mdv Bardhoc	Albania Mdv	Montenegro Doclea Mdv Slavic	0.31	93%	7%	0.07
Albania Post-Mdv Bardhoc	Albania Mdv	East Anatolia Roman Slav	0.17	94%	6%	0.06
Albania Post-Mdv Bardhoc o	Albania Mdv	Russia IA Ingria	0.19	87%	13%	0.05
Albania Post-Mdv Bardhoc o	Albania Mdv	Croatia Roman NE o	0.58	73%	27%	0.07
Albania Post-Mdv Bardhoc o	Albania Mdv	Montenegro Doclea Mdv Slavic	0.55	64%	36%	0.09
Albania Post-Mdv Bardhoc o	Albania Mdv	Serbia Roman NE o	0.55	75%	25%	0.06
Albania Post-Mdv Bardhoc o	Albania Mdv	East Anatolia Roman Slav	0.46	72%	28%	0.08
Albania Post-Mdv Pazhok o	Albania Mdv	Russia IA Ingria	0.21	86%	14%	0.04
Albania Post-Mdv Bazhok o	Albania Mdv	Croatia Roman NE o	0.09	74%	26%	0.07
Albania Post-Mdv Pazhok o	Albania Mdv	Montenegro Doclea Mdv Slavic	0.06	67%	33%	0.09
Albania Post-Mdv Pazhok o	Albania Mdv	East Anatolia Roman Slav	0.17	69%	31%	0.08
Albanian modern 191	Albania Mdv	Serbia Roman NE o	0.06	65%	35%	0.05
Albanian modern 191	Albania Mdv	Montenegro Doclea Mdv Slavic	0.48	54%	46%	0.08
Albanian modern 202	Albania Mdv	Russia IA Ingria	0.07	80%	20%	0.05
Albanian modern 202	Albania Mdv	Croatia Roman NE o	0.26	62%	38%	0.07
Albanian modern 212	Albania Mdv	Serbia Roman NE o	0.1	76%	24%	0.06
Albanian modern 212	Albania Mdv	Montenegro Doclea Mdv Slavic	0.12	70%	30%	0.09
Albanian modern 212	Albania Mdv	East Anatolia Roman Slav	0.6	73%	27%	0.09
Albanian modern 213	Albania Mdv	Montenegro Doclea Mdv Slavic	0.3	65%	35%	0.1
Albanian modern 220	Albania Mdv	Serbia Roman NE o	0.16	73%	27%	0.06
Albanian modern 220	Albania Mdv	Russia IA Ingria	0.06	82%	18%	0.05
Albanian modern 220	Albania Mdv	Croatia Roman NE o	0.18	70%	30%	0.08
Albanian modern 220	Albania Mdv	Montenegro Doclea Mdv Slavic	0.25	66%	34%	0.09
Albanian modern 230	Albania Mdv	Russia IA Ingria	0.18	72%	28%	0.05
Albanian modern 230	Albania Mdv	Serbia Roman NE o	0.21	66%	34%	0.06
Albanian modern 230	Albania Mdv	Croatia Roman NE o	0.16	63%	37%	0.08
Albanian modern 230	Albania Mdv	Montenegro Doclea Mdv Slavic	0.54	52%	48%	0.09
Albanian modern 230	Albania Mdv	East Anatolia Roman Slav	0.6	55%	45%	0.08

Figure 7. Admixture modelling of samples from post-Medieval Albania using a rotated *qpAdm* model. The samples of Albania Mdv are used as a proximate local source, while various Roman and Medieval populations along the Balto-Slavic cline serve as proxies for Balto-Slavic-related ancestry. Models with poor fit (p-value ≤ 0.05) and infeasible coefficients (≥0.15) are not shown. Note that models with modern Albanian samples are of lower resolution as they exploit 600k SNPs and not the full 1240k SNPs used for tests including solely ancient samples. However, the proportion of Slavic-related ancestry in modern Albanians is comparable with that of the post-Medieval outliers from Bardhoc, suggesting that *qpAdm* tests including modern Albanian samples are accurate. Figure based on Table S16.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sources (Croatia, Montenegro, Serbia) are highly supported, in agreement with historical and linguistic data (*33*), while a source from Medieval North Macedonia is either rejected or receives low support (SE = 0.14 and 0.15 in Tables S16-S17, respectively).

We have shown that modern Albanians from Tirana derive 25-48% of their ancestry from a South-Slavic-related source (Fig. 7; Tables S16-S17). This ancestry contribution is two to three times higher than the frequency of South-Slavic-associated Y-chromosome haplogroups (R1a-M417, I2a-M423) (*67*, *68*) in the modern Albanian population (15% combined; Fig. S8), suggesting that Slavic-related admixture may have been largely female-mediated, as has been shown in 10th century Serbia (*3*). These findings are in agreement with anthropological and historical data supporting a strongly patrilineal, kinship-focused culture among Albanians until early modern times (*69–71*).

To determine whether the similarities in autosomal ancestry and PCA position between the Medieval and post-Medieval population of Albania are the result of direct descent, we mined previously published datasets of IBD-sharing between ancient samples (*50*) (Table S20). Remarkably, all samples from post-Medieval Bardhoc share large IBD segments (10-70 cM) with the Medieval individuals from Kënetë and Shtikë, despite the latter being situated 300 km to the south and having lived 500-700 years earlier (Table S20). However, such a close relationship may be exaggerated by founder effects, especially considering that the modern Albanian population displays elevated haplotype sharing (*72*, *73*). We also observe IBD-sharing (9-13 cM; Table S20) of the Bardhoc samples with a 17th-20th century CE individual from Roopkund lake in India who clusters with modern mainland Greeks (*74*). This finding suggests that a population related to Bardhoc may have been involved in the mass migrations of Albanian-speakers into Greece in the 14th-16th centuries CE (*18*, *75*).

Based on the findings from the PCA, the *f4 qpAdm* models, and IBD-sharing, we show that the population of Bardhoc PostMdv descends from earlier medieval groups from Albania, which in turn showed continuity with IA and Roman era western Balkan individuals. However, despite its relative isolation, post-Medieval Albania represented a diverse cultural landscape. This is mirrored in three individuals from Barç, who project far from the remaining Balkan populations on the PCA (Fig. 5B). Although previously interpreted as being of Turkic origin (*6*, *16*), our *qpAdm* and ADMIXTURE analyses conclusively show that these individuals harbour ancestry from South Asia (Figs. 3C, S3; Table S6), suggesting descent from the Roma people (*76*). Furthermore, the sole male individual is assigned to Y-chromosome haplogroup J2a-Y18404 (Table S22), which is found almost exclusively among Roma populations (*77*). The same holds true for the mitochondrial haplogroup U3b1 of one of the females [especially for branch U3b1c2 (*77*)]. To our knowledge, this is the first record of Roma people in the aDNA record.

Large-scale persistence of pre-Migration Period haplogroups in modern Albanians

Given that the transmission of genetic ancestry, culture, and likely language are often male driven (*43*, *46*, *54*), we interrogated ancient and modern Balkan Y-chromosome haplogroups to further refine our understanding of demographic changes in the region. Remarkably, we find that 80% of the paternal ancestry of modern Albanians stems from pre-Migration Period populations (Fig. S8).

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We focus on the primary palaeo-Balkan lineages of modern Albanians – haplogroups E-V13 (27-35%), J2b-Z600 (15%), and R1b-BY611>Z2705 (12-14%) (Fig. S8; Table S33) and report their regional history and phylogeny in the context of the formation of modern Albanian populations.

Haplogroup J2b-Z600 branched off its parent lineage J2b-L283 around 3500-3000 BCE (*77*, *78*). Current sampling suggests that J2b-Z600 was absent from the European Neolithic-Chalcolithic (Fig. 8), as it appears abruptly on the aDNA record in the Serbian Bronze Age (2100-1800 BCE) in a Maros cultural context, alongside the parent subclade of R1b-BY611 (Fig. S9) (*5*, *79*). This places two of the most frequent paternal lineages of the Albanians (Fig. S9) in the Central-West Balkans by the EBA (Figs. 8-9, S9). Haplogroup J2b-Z600 experienced a major founder effect and diversification in the ancient populations of the Adriatic coast (Albania, Croatia, Montenegro), where it accounts for 50-70% of all paternal lineages during the BA-IA (Figs. 8-9), and has been found in samples associated with major West Balkan archaeological cultural expressions, most notably in Maros, Cetina, Japodian and Liburnian contexts (Tables S23-S24, S28) (*5*, *79*). Coupled with its remarkably local distribution in pre-Roman times (Fig. S10), J2b-Z600 may represent a reliable indicator of ultimate Bronze Age-Iron Age West Balkan paternal ancestry. The distributional expansion of J2b-Z600 in northern and western Europe in Roman and post-Roman times (Figs. 8-9, S10) is not surprising, as the West Balkans supplied the Empire with mercenaries, soldiers, and Emperors for centuries (*7*–*9*). Within an Albanian context, J2b-Z600 subclades found in BA-IA, Roman and Medieval Albania (Bardhoc, Çinamak), Montenegro (Doclea, Velika Gruda), and Southern Croatia (Gardun, Gudnja cave), have daughter or sister lineages in modern Albanians (Table S34), suggesting significant paternal continuity from ancient south-west Balkan populations identified as "Illyrians" by classical authors (Fig. S11). aDNA samples from Roman Serbia (Sviloš, Viminacium) and Late Avar-Medieval Hungary (Alattyán, Sárrétudvari) belong to J2b-Z600 lineages related to those of modern Albanians (Table S34), indicating an ultimately south-west Balkan paternal origin for these individuals, which corroborates inscriptional and historical evidence for transplantations of "Illyrian" soldiers along the Limes (*8*, *9*).

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

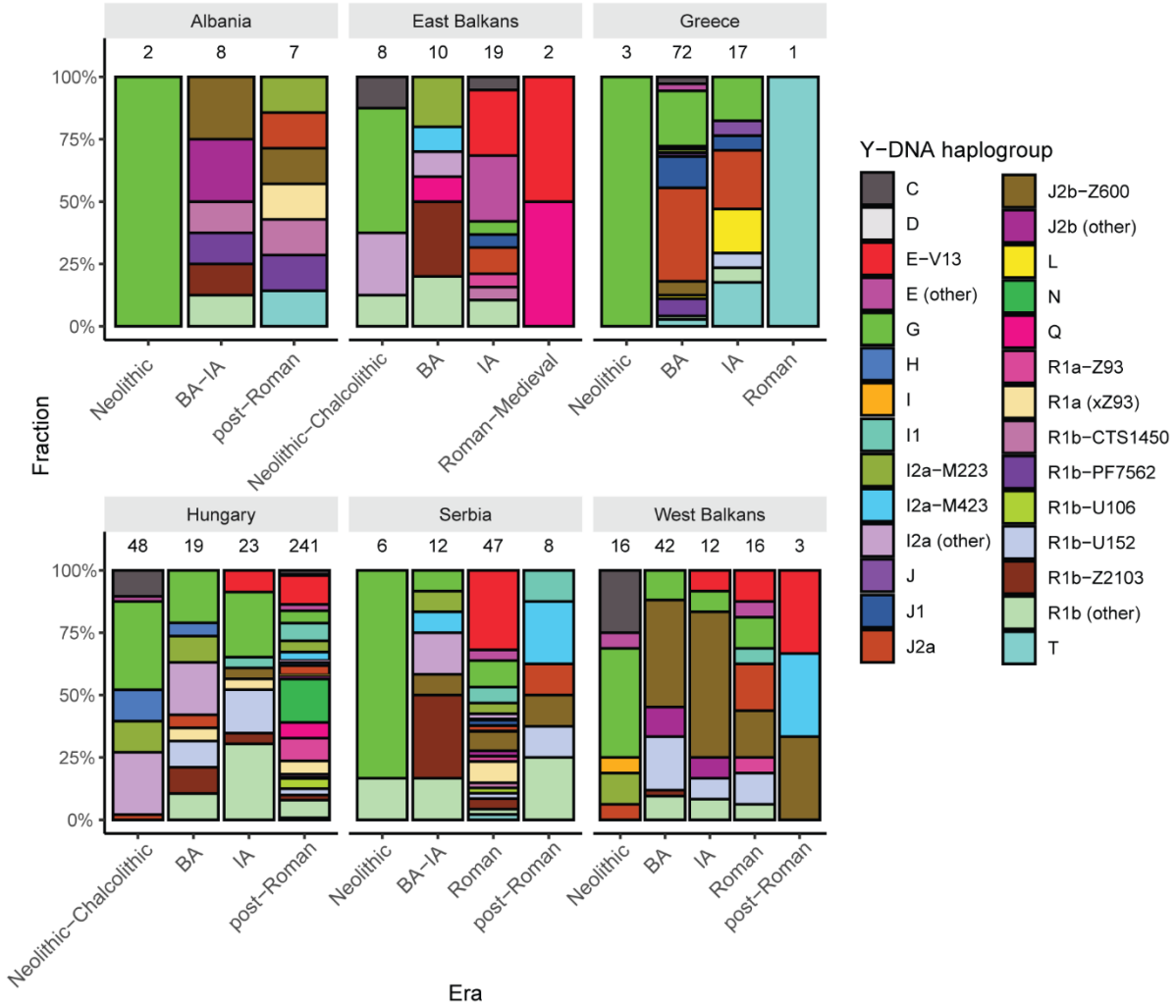


Figure 8. Y-DNA transect of the Balkans and Hungary over the past 8.000 years, using data from Tables S22-S31. Sample sizes shown above bars.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

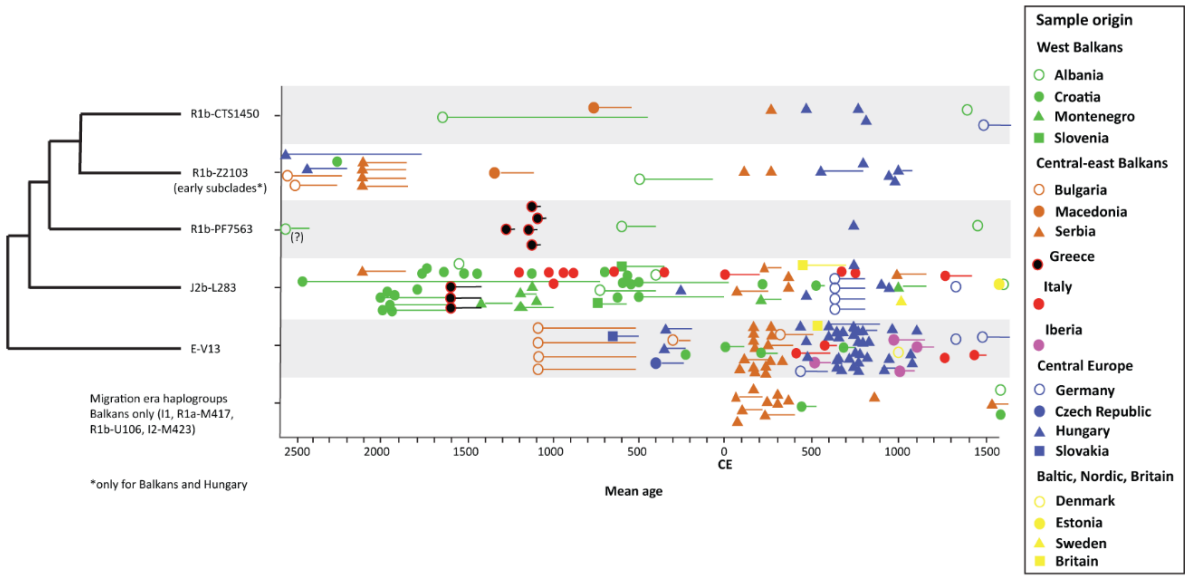


Figure 9. Temporal distribution of the principal Central-West Balkan-derived Y-chromosome haplogroups in European populations, with a schematic phylogeny on the y-axis. Horizontal lines extending from some datapoints represent the archaeologically or radio-carbon-determined time range in which a particular individual lived. Compiled from Tables S22-31.

The most frequent paternal lineage among the Yamnaya, R1b-Z2103 (*6*, *49*), is represented in the ancient Balkans by daughter haplogroup R1b-CTS7556, which was first found in a Maros cultural context in Serbia (Figs. 8-9, S9-S10) (*79*), suggesting a direct migration of Yamnaya-related groups into the Pannonian plain. In turn, the primary descendant clade of R1b-CTS7556 is R1b-CTS1450, whose various daughter lineages appear in Bronze and Iron Age populations of northeastern Albania (Çinamak) and North Macedonia (Figs. 8-9, S9-S10). Importantly, R1b-CTS1450 lineages directly ancestral to haplogroup R1b-BY611>Z2705 (which today comprises almost exclusively Albanians), were found in BA-IA Çinamak (Fig. S9C). Although these lineages ultimately did not contribute to modern Balkan populations, their presence in Çinamak suggests that the group that introduced R1b-BY611>Z2705 to the territory of modern Albania was located nearby. One of the Bardhoc PostMdv samples belongs to R1b-BY611>Z2705 (Table S22), further supporting our interpretation of that population being related to modern Albanians.

Despite being one of the most frequent haplogroups in modern Balkan populations (*67*, *80–83*), the origins of E-V13 are enigmatic. The earliest record of this haplogroup among historically attested groups is in BA-IA Bulgaria (Figs. 8-9, S10), suggesting an association with the people known as the "Thracians". By the early Roman era, E-V13 likely experienced significant demographic increase, as it appears at medium to high frequencies in areas where in the preceding Bronze and Iron Age it was either very rare (Croatia, Hungary) or entirely absent (Serbia) (Figs. 8-9). An association of the expansion of E-V13 with southeastern Balkan populations from the Thracian world is reinforced by one of the Avar-era outliers from Hungary, who is assigned to E-V13 and clusters with BA-IA populations from Bulgaria on the PCA (Fig. 5A), an affinity confirmed by *qpAdm* (Table S14), and IBD-sharing (Table S21) (*50*). A Scythian from Moldova

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Table S21) who clusters close to Balkan IA populations (Fig. 4A) and belongs to E-V13 (Fig. S10) also displays IBD-sharing with Bulgaria IA (Table 21). Our findings support late Roman historical records which mention the presence of "Thracian" groups known as the "Bessi" throughout the Balkans until the 6th century CE (*2*, *36*, *62*, *64*).

However, not all populations with E-V13 were characterised by a Bulgaria EIA-like autosomal profile, as shown by the two E-V13-bearing Himera mercenaries (Fig. 4), who were likely related to EBA-LBA populations from Serbia (Supplementary Methods; Table S10). Furthermore, several Roman-era samples with a West Balkan autosomal profile [(Croatia: Sipar R3664, Ščitarjevo R3659); Serbia (Viminacium R6756)] also harboured E-V13 (Tables S23, S28). This does not exclude a Thracian origin, as the historical region of Dardania (roughly modern Kosovo, southern Serbia, and western North Macedonia) is recorded as a zone of linguistic contact between "Illyrian" and "Daco-Thracian" groups (Fig. S7) (*1*, *7*). It is therefore likely that the population that introduced E-V13 into Albanians would plot close to Roman era West Balkan groups. Remarkably, several E-V13 subclades found in Avar and Medieval era Hungary, as well as in one of the Himera mercenaries, are characterized by sister or daughter branches that comprise almost exclusively Albanians (Table S34). Whether the individuals from Hungary had an ultimately Balkan origin is unknown, as most do not share IBD segments with Balkan populations and are significantly admixed with northern European and Central Asian populations (Table S21).

To obtain insights on the ethnogenesis of modern Albanians, we plot the mean Y-full TMRCAs of Albanian-specific subclades of E-V13, J2b-Z600, R1b-BY611 and other palaeo-Balkan haplogroups (R1b-PF7562, I-M223) (Fig. 10). Remarkably, a majority of these haplogroups (J2b-Z600, R1b-BY611, R1b-PF7562, I-M223) experience a sudden and steep increase in subclade diversity between 500-800 CE (Fig. 10), which coincides with the timing proposed by linguistic and historical hypotheses on the origins of Albanians and their language (*33–35*, *64*, *84*), as well as IBD-sharing analyses (*72*). The low number of diversifying subclades prior to 500 CE is likely caused by missing data, probably due to significant loss of diversity associated with the demographic turmoil of the Migration Period.

Unlike the abovementioned haplogroups, E-V13 exhibits continuous subclade diversification from the Bronze Age to the Roman period (Fig. 10), suggesting that populations with a high frequency of E-V13 may have followed a different demographic trajectory from those with J2b-Z600, R1b-BY611, R1b-PF7562, and I-M223. The rate of E-V13 subclade diversification increased steeply from 500 CE onwards, following the pattern of the other haplogroups found in modern Albanians (Fig. 10). Based on the above, it is possible that currently unsampled populations from the Central-West Balkan interior that were characterised by high frequencies of E-V13 may have entered the region of modern Albania around 500 CE, where they merged and co-expanded with local groups. This may also explain the absence of E-V13 from the aDNA transect of Albania, despite being the commonest haplogroup in the modern Albanian population.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

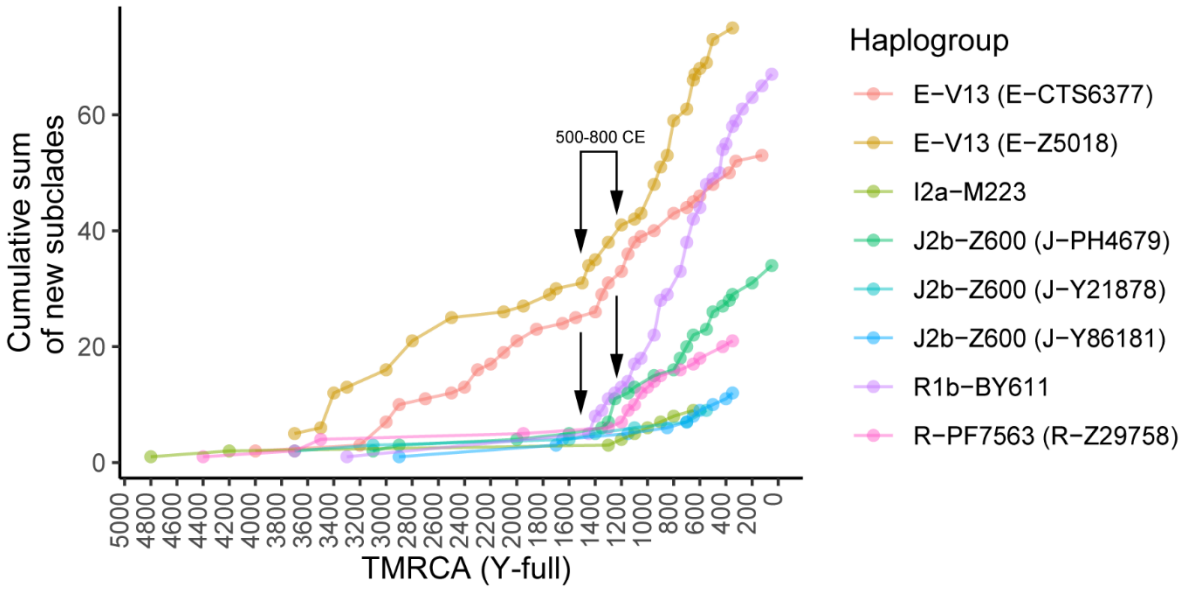


Figure 10. Graphical representation of clade formation (cumulative sum of new subclades) of the principal palaeo-Balkan haplogroups of modern Albanians, using TMRCA estimates from Y-full. Plotted using data from Table S32.

Together with *qpAdm* and IBD data, the principal Y-chromosome haplogroups add further evidence for large-scale continuity of modern Albanians from local groups, without excluding the possibility of admixture with neighbouring palaeo-Balkan populations. Continuity is also mirrored in rarer lineages, such as haplogroup T-Y206597, found in the Medieval individual from Këneta (Table S34), which today comprises almost exclusively Albanians (*77*).

Discussion

Our genomic transect of the population of Albania from the Neolithic to the modern era reveals fluctuations in genetic ancestry over a period of 8000 years. In contrast to the southeastern Balkans, where the arrival of Pontic-Caspian steppe ancestry and the associated Indo-European cultural package during the EBA did not lead to a lasting genetic turnover (*4*), we show that contemporary populations in Albania were genetically transformed both in autosomal and paternal ancestry (Fig. 8; Table S5). We find that more than a millennium later, BA-IA Balkan populations with high levels of steppe ancestry (30-40%) formed a distinct genetic cluster that extended from northwestern Greece, North Macedonia and the Adriatic coast (including Albania) and transcended archaeological and linguistic boundaries (Fig. 4A). This genetic continuum was broken down across the Balkans during the Roman and Migration period (Fig. 5A), due to mass settlement of Germanic and Slavic-speaking groups in the region.

However, in agreement with linguistic studies, we find that Albanians likely descend from a surviving West palaeo-Balkan population that experienced significant demographic increase approximately between 500-800 CE (Fig. 10), perhaps after a population bottleneck. We show that in contrast to the rest of the Balkans, the Medieval samples from both North and South Albania experienced little to no contribution from surrounding Slavic populations (Fig, 6B-C; Tables S12-

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S13, S15) and maintained high levels of BA-IA West Balkan ancestry. Remarkably, the same genetic profile persisted 500-800 years later in most of the post-Medieval samples from Bardhoc, as shown both by the PCA (Fig. 5), *qpAdm* analyses (Tables S16-S18), and IBD data (Table S20), which indicate significant genetic continuity from the Medieval populations of Albania. However, *qpAdm* models cannot exclude the possibility of additional admixture with currently unsampled neighbouring late Roman-early Medieval palaeo-Balkan groups with a similar ancestry profile. Based on linguistic data, the area of modern Kosovo and southeastern Serbia may have been such a source (*15*, *33*, *34*).

Despite being largely unaffected by the demographic changes that took place during the Migration period, the historical Albanians did not emerge in isolation. At the peak of the Migration Period, the Medieval population of Albania displayed genetic links as far as Pannonia (Tables 20-21), while in post-Medieval times we detected the presence of individuals likely related to modern Roma people (Fig. 5A). Furthermore, two of the post-Medieval samples exhibit significant admixture with South Slavic populations (Tables S16-S18), and modern Albanians display highly variable levels of Slavic ancestry (Fig. 5B, Tables S16-S18). This indicates complex historical interactions with South Slavic populations, as suggested by toponymy and linguistics (*23*, *35*).

We reveal that a significant proportion of the paternal ancestry of modern Albanians derives from groups ultimately descending from the BA-IA West Balkans (Tables S33-S34), including those traditionally known as “Illyrians” (Figs. S9, S11), which reflects our findings on autosomal ancestry. However, inferring the language spoken by the Medieval samples from Albania is challenging, as Greek, South Slavic and West Balkan Romance are the only recorded languages of the region (*14*, *15*, *25*), while there is no indication of the survival of “Illyrian” following the first centuries of Roman rule (*7*, *8*). Furthermore, Albanian displays Latin loans from both the Western and Eastern Balkans (*85*), which attests to linguistic influences beyond the confines of modern Albania. Testing the Messapic hypothesis for Albanian (*7*, *19*, *30*, *32*) was not possible due to the low coverage of said samples (*12*). Although the presence of haplogroups J2b-L283, I-M223, and R1b-Z2103 among the Messapians (Table S30) suggests a West Balkan origin, whether a related language persisted in the Balkans during Medieval times is unknown.

Even though Eastern Roman historians were unfamiliar with Albanians (*22*), we cannot exclude the possibility that proto-Albanians interacted with populations speaking Greek, Aromanian, or Slavic in what is now southern Albania during Medieval times. Given that genetic data strongly suggest a predominantly local origin for Albanians, their Medieval ancestors may have inhabited a geographically restricted area [possibly the region of Mat in central Albania (*14*)], only occasionally venturing towards the south. These movements may have increased in scale over time, finally attracting the attention of Greek-speaking historians in the 11th century (*22*).

While the quest for the origins of the Albanian language will certainly continue, we expect that the present study will shape these debates and provide the necessary framework for more extensive research on the genetic ancestry of the ancient and modern inhabitants of Albania.

Materials and Methods

Experimental Design

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The following sections describe the examined dataset, the resources providing the samples examined herein, the statistical methods used to analyse their ancestry, and the employed data visualisation software.

Dataset

A list of all the samples and naming conventions used in the *qpAdm* models and PCA of this study can be found in Tables S1-S2. On several occasions, our sample naming differs to that given in the original study, based on insights from the PCA, admixture modelling, and dating interpretation. We mention here naming differences based on a re-interpretation of the samples in question. Individuals I18723, I18721, I18719 from Bezdanjača Cave in Croatia, which were archaeologically dated to the Bronze Age (1500-800 BCE), were interpreted in previous studies as outliers compared to the contemporary population of Croatia (*5*, *6*). We argue that based on insights from the PCA and their uniparental markers, these outlying individuals likely date to post-Medieval times. Individuals I18721 and I18719 are assigned to haplogroup I2-M423>I-Y3120 (Table S23), which is associated with the Slavic expansion toward southern Europe during the Migration Period, and has experienced major founder effects in the South Slavic population of the Balkans (*48*, *68*, *82*). This haplogroup is extremely unlikely to have entered the Western Balkans in the Bronze Age, as our extensive haplogroup dataset shows that subclades downstream of I2-M423 appear in the region primarily during the Migration Period (Fig. 8), as expected. Occasional migrants from the Balto-Slavic world to southern Europe are known in the Iron Age, such as two mercenaries from the Battle of Himera in Sicily (*58*). However, the Bezdanjača Cave outliers are unlikely to represent BA migrants from northern Europe, as the mitochondrial haplogroup of individual I18719 (HV0a1a1b), has a TMRCA of 225 years before present with a person from modern Germany (*77*). Furthermore, previous archaeological studies in Bezdanjača Cave radiocarbon-dated an ancient individual to the 17th century CE, and also reported the remains of two skeletons which date to World War II (*86*). Another sample that might be misdated is I13170 from Velika Gruda in Montenegro, attributed to the Iron Age (800-400 BCE) (*6*). This individual, which was not radiocarbon dated, clustered with modern South Slavs (*6*), and was therefore excluded from our analyses.

Within the context of the samples from Albania, individual I13834 from Barç (Southeast, Korça Basin), radiocarbon dated to Post-Medieval times (1452-1619 calCE (385±15 BP, PSUAMS-8300) (*6*, *16*) is intriguing. The sample clusters with Bronze Age and Iron Age samples from Albania (Fig. 5), and also lacks any of the Iran N-related ancestry present in Medieval, Post-Medieval, and modern samples from Albania (Table S5). In contrast, sample I13839 from neighbouring Shtikë (Southeastern, Kolonja Plateau), which is radiocarbon dated to 889-989 CE (*6*, *16*) derives part of his ancestry from an Iran N-related source (Table S5), as expected. Based on the above, individual I13834 from Barç either maintained an unadmixed profile for more than 1600 years, or it might represent a case of the freshwater reservoir effect, which can cause errors in carbon dating of samples (*87*). Due to its uncertain dating, we tentatively assigned individual I13834 to the Medieval population on the PCA (Fig. 5A, B), and excluded it from all proximate admixture analyses.

Statistical Analysis

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PCA

We undertook Principal Component Analysis (PCA) on a subset of the HO (*51*) and newly genotyped Reitsema et al. (2022) (*58*) dataset using the 'smartpca' function (v16000) in EIGENSOFT (version 7.2.1) (*88*) (Fig. S1). SNP datasets in EIGENSTRAT format were combined using the mergeit function included in EIGENSOFT, with default parameters, aside from allowdups: YES; outputformat: EIGENSTRAT; strandcheck: NO; hashcheck: NO. We converted .bam files to EIGENSTRAT format using the mpileup function from the SamTools software package (v1.16), with optional flags -R -B -q30 -Q30. The resulting pileup files were processed into EIGENSTRAT using the pileupCaller function from the SequenceTools software package (v1.5.2 - https://github.com/stschiff/sequenceTools) using the default parameters and calling_method: randomHaploid.

HO dataset individuals of ancestry directly relevant to West Eurasian populations were used (Caucasus, Central Asia, Europe, Middle East), while certain Middle Eastern and North African populations with significant sub-Saharan ancestry that caused shrinkage of the PCA were removed. For compatibility with the HO dataset, SNPs within 1240K SNP datasets corresponding to those in the HO dataset were renamed prior to the datasets being combined using mergeit. We then projected ancient samples onto present-day individuals with "lsqproject:YES" and "shrinkmode:YES", in three chronologically distinct datasets (Neolithic-Early Bronze Age, Bronze Age-Iron Age, Roman-Post-Roman), resulting in Figs 2, 4, 5.

Low coverage, non-UDG-treated ancient samples were not included in our PCA, as they can cause artefactual shifts towards the direction of sub-Saharan populations, which is not only misleading but also causes shrinkage of the PCA. To avoid redundancy due to shared ancestry, we also removed close relatives from the ancient samples dataset. The final dataset of the modern and ancient samples used for the PCA is provided in Table S2.

ADMIXTURE

We used ADMIXTURE (v1.3.0) to analyze 1,044 modern and ancient human samples. Most of the individuals were sourced from the Allen Ancient DNA Resource or AADR (v54.1_1240K_public) from the David Reich Lab. Sixteen of the samples came from Antonio et al. 2023 (*89*). The initial set of 1,150,639 autosomal SNPs was pruned for linkage disequilibrium in PLINK 1.90 with parameters --indep-pairwise 50 25 0.2, which resulted in a final set of 276,155 SNPs. All individuals with a genotyping rate of less than 5% were removed from the analysis. ADMIXTURE was run with the number of ancestral populations (K) ranging from 2 to 8. The results, particularly at K7 and K8, are similar to our formal *qpAdm* statistics-based analyses. The ADMIXTURE analysis also reveals South Asian-related ancestry in the ALB_Barc_PostMdv_Roma_profile individuals. Unlike other Europeans, these individuals show ~20% membership in the K7 and K8 clusters (labeled C3 and C1, respectively) that are modal in modern samples from southern and eastern India (such as ILA.SG and BIR.SG).

f3-statistics to measure genetic drift

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We employed the ADMIXTOOLS v.2.0.0 package in R (version 4.1.1), operated using RStudio (v. 2022.07.1+554), in order to estimate outgroup *f3*-statistics of the form *f3(outgroup; population A, population B)*, where outgroup = Cameroon_SMA, population A = the tested individual or metapopulation, and population B = the ancient populations in our working dataset, respectively. Outgroup *f3*-statistics estimate pairwise genetic affinity via allele sharing (*90*). The higher the value of the *f3*-statistic the higher is the genetic affinity between population A and the tested ancient population(s) B.

qpAdm admixture modelling

We employed *qpAdm* to model the ancestry of the examined populations by using a rotating sources approach (*91*), with an emphasis on the samples from Albania. A model was accepted if its p-value was significant (≥0.05) and the standard errors (SE) were sufficiently low (≤0.15). Models with a satisfactory Z-score (≥3) were favoured if they corresponded to patterns observed on the PCA and had a significant p-value and sufficiently low standard errors, although we did accept models with lower Z-scores as well. Models with poor fit (p-value ≤0.05) and infeasible coefficients (SE>0.15) were rejected and are not shown.

We note that *qpAdm* analyses between ancient and modern individuals use around half (600k) of the SNPs compared to ancient-to-ancient models, which exploit the full 1240k SNPs available. This is due to the fact that all modern individuals have been genotyped with the HO SNP array (*90*) which tests for a much smaller number of SNPs compared to the 1240k array used for ancient samples. *qpAdm* can be used successfully even with low coverage samples (*91*), and the admixture proportions we recover for the modern Albanian samples are consistent with their position on the PCA and the ADMIXTURE analyses. However, we caution that the resolution of the models involving modern samples will be lower compared to those including only ancient metapopulations.

An extensive description of our *qpAdm* models and the rationale behind the chosen reference populations can be found in the Supplementary Material.

Mobest analysis

The Mobest analyses were run using a kernel size of 800 (corresponding to 800 km in space and 800 years in time). The predication grid was set to 50 by 50 km tiles. As genetic input we used the first two PCs of the West Eurasian PCA as well as the first two PCs of a PCA only including European individuals. Our reference dataset comprised 5664 published individuals from the Allen Ancient DNA Resource and the samples from the Antonio et al. (2022) pre-print that were used for the *qpAdm* analyses (Table S1). The reference samples date between 150 and 5000 BP. Those samples are located between 29.9° and 70° Lat. and -24° and 70° Long. (in EPSG:4326 projection). Samples with less than 15k SNPs were excluded. The relative search time was set to 0, thus, the probability surface indicates the highest genetic-geographical match at the mean date of the respective individual.

Y-chromosome analysis and interpretation

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ancient DNA Y-chromosome haplogroup data were aggregated from the literature (Tables S22-S31), albeit at low subclade resolution. In most cases, the terminal subclade we cite was assigned by yfull.com (*77*) and the free, publicly available FamilyTreeDNA Discover Beta™ (Gene by Gene Ltd), both being broadly used resources for inferring human uniparental haplogroups and their TMRCA (*6*, *92*, *93*). Additionally, publicly available raw data from Albania by Lazaridis et al. (2022) (*6*, *16*) and Central-Western Balkan samples by Antonio et al. (2022) (*89*) were manually evaluated in the present study using IGV (*94*), and were also called with SAMtools and BCFtools (*95*) by Open Genomes (Ted Kandell), and Open Genomes and the Society of Serbian Genealogists "Poreklo" (Milan Rajevac). TMRCAs are publicly available at yfull.com and FamilyTreeDNA Discover Beta™. The modern Albanian data populating the yfull.com and FamilyTreeDNA Discover Beta™ Y-chromosome public databases stem from direct-to-consumer whole genome sequencing tests (BigY (*96*), Nebula Genomics, Dante Labs, YSEQ).

To generate Figs. 8-10, S9-S11, we assembled all Y-chromosome haplogroups from Tables S22-S30 and grouped them into Table S31, which we used to produce a bar-chart in RStudio using ggplot2 (*97*). Regarding Fig. 10, which plots the TMRCAs of the principal Y-chromosome haplogroups of modern Albanians (E-V13, J2b-Z600, R1b-BY611, I-M223, and R1b-PF7562), we consulted the corresponding phylogenetic trees at yfull.com (*77*). We then assembled in Table S32 the TMRCAs of the Y-chromosome subclades associated with Albanians and their expansions into neighbouring regions (Greece, Bosnia, Montenegro, North Macedonia, Serbia). We did not include the TMRCAs of subclades lacking an association with Albanians (i.e. not having any Albanians in their daughter lineages). To study the Y-chromosome haplogroup distribution of the modern Albanian population, we used both academic samples from the scientific literature (*75*, *82*, *83*) (n = 377) and the publicly available dataset of rrenjet.com (*98*) (n = 1534) (Table S33, Supplementary Material).

IBD data

A recent study (*50*) provides a dataset with IBD-sharing between 10156 ancient Eurasian individuals. We mined said dataset for samples from Albania and their matches, which are presented in Table S20. We also provide data on IBD-based matches for Scythian and Avar era samples which were characterised by haplogroups of possible Balkan origin (E-V13, J2b-Z600, R1b-Z2103 and R1b-PF7562) in Table S21. We excluded potential false positive IBD matches appearing among shotgun-sequenced and non-UDG-treated samples of the ancIBD dataset, especially in Viking era individuals (*50*).

Data visualisation

We used R (version 4.1.1) via RStudio (v. 2022.07.1+554). All plots were created using package ggplot2. To generate vertical stacked bar plots, we additionally used package forcats. To combine plots, we used Adobe Illustrator v. 27.4. We generated maps using the QGIS Geographic Information System, QGIS Association (http://www.qgis.org) and SimpleMappr.

References

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. F. Papazoglu, *The Central-Balkan tribes in pre-Roman times. Triballi, Autariatae, Dardanians, Scordisci, and Moesians* (Hakkert, Amsterdam, 1978).

2. V. Georgiev, The Genesis of the Balkan Peoples. *Slavon E Eur Rev*. **44**, 285–297 (1966).

3. I. Olalde, P. Carrión, I. Mikić, N. Rohland, S. Mallick, I. Lazaridis, M. Korać, S. Golubović, S. Petković, N. Miladinović-Radmilović, D. Vulović, K. Stewardson, A. M. Lawson, F. Zalzala, K. Callan, Ž. Tomanović, D. Keckarević, M. Grbić, C. Lalueza-Fox, D. Reich, *bioRxiv*, in press, doi:10.1101/2021.08.30.458211.

4. I. Lazaridis, A. Mittnik, N. Patterson, S. Mallick, N. Rohland, S. Pfrengle, A. Furtwängler, A. Peltzer, C. Posth, A. Vasilakis, P. J. P. McGeorge, E. Konsolaki-Yannopoulou, G. Korres, H. Martlew, M. Michalodimitrakis, M. Özsait, N. Özsait, A. Papathanasiou, M. Richards, S. A. Roodenberg, Y. Tzedakis, R. Arnott, D. M. Fernandes, J. R. Hughey, D. M. Lotakis, P. A. Navas, Y. Maniatis, J. A. Stamatoyannopoulos, K. Stewardson, P. Stockhammer, R. Pinhasi, D. Reich, J. Krause, G. Stamatoyannopoulos, Genetic origins of the Minoans and Mycenaeans. *Nature*. **548**, 214–218 (2017).

5. N. Patterson, M. Isakov, T. Booth, L. Büster, C.-E. Fischer, I. Olalde, H. Ringbauer, A. Akbari, O. Cheronet, M. Bleasdale, N. Adamski, E. Altena, R. Bernardos, S. Brace, N. Broomandkhoshbacht, K. Callan, F. Candilio, B. Culleton, E. Curtis, L. Demetz, K. S. D. Carlson, C. J. Edwards, D. M. Fernandes, M. G. B. Foody, S. Freilich, H. Goodchild, A. Kearns, A. M. Lawson, I. Lazaridis, M. Mah, S. Mallick, K. Mandl, A. Micco, M. Michel, G. B. Morante, J. Oppenheimer, K. T. Özdoğan, L. Qiu, C. Schattke, K. Stewardson, J. N. Workman, F. Zalzala, Z. Zhang, B. Agustí, T. Allen, K. Almássy, L. Amkreutz, A. Ash, C. Baillif-Ducros, A. Barclay, L. Bartosiewicz, K. Baxter, Z. Bernert, J. Blažek, M. Bodružić, P. Boissinot, C. Bonsall, P. Bradley, M. Brittain, A. Brookes, F. Brown, L. Brown, R. Brunning, C. Budd, J. Burmaz, S. Canet, S. Carnicero-Cáceres, M. Čaušević-Bully, A. Chamberlain, S. Chauvin, S. Clough, N. Čondić, A. Coppa, O. Craig, M. Črešnar, V. Cummings, S. Czifra, A. Danielisová, R. Daniels, A. Davies, P. de Jersey, J. Deacon, C. Deminger, P. W. Ditchfield, M. Dizdar, M. Dobeš, M. Dobisíková, L. Domboróczki, G. Drinkall, A. Đukić, M. Ernée, C. Evans, J. Evans, M. Fernández-Götz, S. Filipović, A. Fitzpatrick, H. Fokkens, C. Fowler, A. Fox, Z. Gallina, M. Gamble, M. R. González Morales, B. González-Rabanal, A. Green, K. Gyenesei, D. Habermehl, T. Hajdu, D. Hamilton, J. Harris, C. Hayden, J. Hendriks, B. Hernu, G. Hey, M. Horňák, G. Ilon, E. Istvánovits, A. M. Jones, M. B. Kavur, K. Kazek, R. A. Kenyon, A. Khreisheh, V. Kiss, J. Kleijne, M. Knight, L. M. Kootker, P. F. Kovács, A. Kozubová, G. Kulcsár, V. Kulcsár, C. Le Pennec, M. Legge, M. Leivers, L. Loe, O. López-Costas, T. Lord, D. Los, J. Lyall, A. B. Marín-Arroyo, P. Mason, D. Matošević, A. Maxted, L. McIntyre, J. McKinley, K. McSweeney, B. Meijlink, B. G. Mende, M. Menđušić, M. Metlička, S. Meyer, K. Mihovilić, L. Milasinovic, S. Minnitt, J. Moore, G. Morley, G. Mullan, M. Musilová, B. Neil, R. Nicholls, M. Novak, M. Pala, M. Papworth, C. Paresys, R. Patten, D. Perkić, K. Pesti, A. Petit, K. Petriščáková, C. Pichon, C. Pickard, Z. Pilling, T. D. Price, S. Radović, R. Redfern, B. Resutík, D. T. Rhodes, M. B. Richards, A. Roberts, J. Roefstra, P. Sankot, A. Šefčáková, A. Sheridan, S. Skae, M. Šmolíková, K. Somogyi, Á. Somogyvári, M. Stephens, G. Szabó, A. Szécsényi-Nagy, T. Szeniczey, J. Tabor, K. Tankó, C. T. Maria, R. Terry, B. Teržan, M. Teschler-Nicola, J. F. Torres-Martínez, J. Trapp, R. Turle, F. Ujvári, M. van der Heiden, P. Veleminsky, B. Veselka, Z. Vytlačil, C.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Waddington, P. Ware, P. Wilkinson, L. Wilson, R. Wiseman, E. Young, J. Zaninović, A. Žitňan, C. Lalueza-Fox, P. de Knijff, I. Barnes, P. Halkon, M. G. Thomas, D. J. Kennett, B. Cunliffe, M. Lillie, N. Rohland, R. Pinhasi, I. Armit, D. Reich, Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature*. **601**, 588–594 (2022).

6. I. Lazaridis, S. Alpaslan-Roodenberg, A. Acar, A. Açıkkol, A. Agelarakis, L. Aghikyan, U. Akyüz, D. Andreeva, G. Andrijašević, D. Antonović, I. Armit, A. Atmaca, P. Avetisyan, A. İ. Aytek, K. Bacvarov, R. Badalyan, S. Bakardzhiev, J. Balen, L. Bejko, R. Bernardos, A. Bertsatos, H. Biber, A. Bilir, M. Bodružić, M. Bonogofsky, C. Bonsall, D. Borić, N. Borovinić, G. B. Morante, K. Buttinger, K. Callan, F. Candilio, M. Carić, O. Cheronet, S. Chohadzhiev, M.-E. Chovalopoulou, S. Chryssoulaki, I. Ciobanu, N. Čondić, M. Constantinescu, E. Cristiani, B. J. Culleton, E. Curtis, J. Davis, R. Davtyan, T. I. Demcenco, V. Dergachev, Z. Derin, S. Deskaj, S. Devejyan, V. Djordjević, K. S. D. Carlson, L. R. Eccles, N. Elenski, A. Engin, N. Erdoğan, S. Erir-Pazarcı, D. M. Fernandes, M. Ferry, S. Freilich, A. Frînculeasa, M. L. Galaty, B. Gamarra, B. Gasparyan, B. Gaydarska, E. Genç, T. Gültekin, S. Gündüz, T. Hajdu, V. Heyd, S. Hobosyan, N. Hovhannisyan, I. Iliev, L. Iliev, S. Iliev, İ. İvgin, I. Janković, L. Jovanova, P. Karkanas, B. Kavaz-Kındığılı, E. H. Kaya, D. Keating, D. J. Kennett, S. D. Kesici, A. Khudaverdyan, K. Kiss, S. Kılıç, P. Klostermann, S. K. B. N. Valdes, S. Kovačević, M. Krenz-Niedbała, M. K. Škrivanko, R. Kurti, P. Kuzman, A. M. Lawson, C. Lazar, K. Leshtakov, T. E. Levy, I. Liritzis, K. O. Lorentz, S. Łukasik, M. Mah, S. Mallick, K. Mandl, K. Martirosyan-Olshansky, R. Matthews, W. Matthews, K. McSweeney, V. Melikyan, A. Micco, M. Michel, L. Milašinović, A. Mittnik, J. M. Monge, G. Nekhrizov, R. Nicholls, A. G. Nikitin, V. Nikolov, M. Novak, I. Olalde, J. Oppenheimer, A. Osterholtz, C. Özdemir, K. T. Özdoğan, N. Öztürk, N. Papadimitriou, N. Papakonstantinou, A. Papathanasiou, L. Paraman, E. G. Paskary, N. Patterson, I. Petrakiev, L. Petrosyan, V. Petrova, A. Philippa-Touchais, A. Piliposyan, N. P. Kuzman, H. Potrebica, B. Preda-Bălănică, Z. Premužić, T. D. Price, L. Qiu, S. Radović, K. R. Aziz, P. R. Šikanjić, K. R. Raheem, S. Razumov, A. Richardson, J. Roodenberg, R. Ruka, V. Russeva, M. Şahin, A. Şarbak, E. Savaş, C. Schattke, L. Schepartz, T. Selçuk, A. Sevim-Erol, M. Shamoon-Pour, H. M. Shephard, A. Sideris, A. Simalcsik, H. Simonyan, V. Sinika, K. Sirak, G. Sirbu, M. Šlaus, A. Soficaru, B. Söğüt, A. Sołtysiak, Ç. Sönmez-Sözer, M. Stathi, M. Steskal, K. Stewardson, S. Stocker, F. Suata-Alpaslan, A. Suvorov, A. Szécsényi-Nagy, T. Szeniczey, N. Telnov, S. Temov, N. Todorova, U. Tota, G. Touchais, S. Triantaphyllou, A. Türker, M. Ugarković, T. Valchev, F. Veljanovska, Z. Videvski, C. Virag, A. Wagner, S. Walsh, P. Włodarczak, J. N. Workman, A. Yardumian, E. Yarovoy, A. Y. Yavuz, H. Yılmaz, F. Zalzala, A. Zettl, Z. Zhang, R. Çavuşoğlu, N. Rohland, R. Pinhasi, D. Reich, The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science (80-.).* **377**, eabm4247 (2022).

7. A. Lippert, J. Matzinger, *Die Illyrer: Geschichte, Archaologie Und Sprache (Reiche Der Alten Welt - Ethnien, Lander, Dynastien)* (Kohlhammer, Stuttgart, 2021).

8. J. Wilkes, The Illyrians. The Peoples of Europe. (1996), p. 376.

9. A. Mócsy, *Pannonia and Upper Moesia. A history of the Middle Danube provinces of the Roman Empire* (Routledge, London, 2015).

10. M. Krzewińska, G. M. Kılınç, A. Juras, D. Koptekin, M. Chyleński, A. G. Nikitin, N.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Shcherbakov, I. Shuteleva, T. Leonova, L. Kraeva, F. A. Sungatov, A. N. Sultanova, I. Potekhina, S. Łukasik, M. Krenz-Niedbała, L. Dalén, V. Sinika, M. Jakobsson, J. Storå, A. Götherström, Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. *Sci. Adv.* **4**, eaat4457 (2018).

11. F. Clemente, M. Unterländer, O. Dolgova, C. E. G. Amorim, F. Coroado-Santos, S. Neuenschwander, E. Ganiatsou, D. I. Cruz Dávalos, L. Anchieri, F. Michaud, L. Winkelbach, J. Blöcher, Y. O. Arizmendi Cárdenas, B. Sousa da Mota, E. Kalliga, A. Souleles, I. Kontopoulos, G. Karamitrou-Mentessidi, O. Philaniotou, A. Sampson, D. Theodorou, M. Tsipopoulou, I. Akamatis, P. Halstead, K. Kotsakis, D. Urem-Kotsou, D. Panagiotopoulos, C. Ziota, S. Triantaphyllou, O. Delaneau, J. D. Jensen, J. V. Moreno-Mayar, J. Burger, V. C. Sousa, O. Lao, A.-S. Malaspinas, C. Papageorgopoulou, The genomic history of the Aegean palatial civilizations. *Cell*. **184**, 2565-2586.e21 (2021).

12. S. Aneli, T. Saupe, F. Montinaro, A. Solnik, L. Molinaro, C. Scaggion, N. Carrara, A. Raveane, T. Kivisild, M. Metspalu, C. L. Scheib, L. Pagani, The Genetic Origin of Daunians and the Pan-Mediterranean Southern Italian Iron Age Context. *Mol. Biol. Evol.* **39** (2022), doi:10.1093/molbev/msac014.

13. S. Shpuza, Some aspects of the Romanization process in South Illyria. *Stud. Albanica*. **3**, 55–62 (2006).

14. E. Lafe, Toponymes latino-romans sur le territoire de l'albanais. *Iliria*. **5**, 113–117 (1976).

15. V. Popović, Byzantins , Slaves et autochtones dans les provinces de Prévalitane et Nouvelle Épire. *Publ. l'École Française Rome*. **77**, 181–243 (1984).

16. I. Lazaridis, S. Alpaslan-Roodenberg, A. Acar, A. Açıkkol, A. Agelarakis, L. Aghikyan, U. Akyüz, D. Andreeva, G. Andrijašević, D. Antonović, I. Armit, A. Atmaca, P. Avetisyan, A. İ. Aytek, K. Bacvarov, R. Badalyan, S. Bakardzhiev, J. Balen, L. Bejko, R. Bernardos, A. Bertsatos, H. Biber, A. Bilir, M. Bodružić, M. Bonogofsky, C. Bonsall, D. Borić, N. Borovinić, G. B. Morante, K. Buttinger, K. Callan, F. Candilio, M. Carić, O. Cheronet, S. Chohadzhiev, M.-E. Chovalopoulou, S. Chryssoulaki, I. Ciobanu, N. Čondić, M. Constantinescu, E. Cristiani, B. J. Culleton, E. Curtis, J. Davis, R. Davtyan, T. I. Demcenco, V. Dergachev, Z. Derin, S. Deskaj, S. Devejyan, V. Djordjević, K. S. D. Carlson, L. R. Eccles, N. Elenski, A. Engin, N. Erdoğan, S. Erir-Pazarcı, D. M. Fernandes, M. Ferry, S. Freilich, A. Frînculeasa, M. L. Galaty, B. Gamarra, B. Gasparyan, B. Gaydarska, E. Genç, T. Gültekin, S. Gündüz, T. Hajdu, V. Heyd, S. Hobosyan, N. Hovhannisyan, I. Iliev, L. Iliev, S. Iliev, İ. İvgin, I. Janković, L. Jovanova, P. Karkanas, B. Kavaz-Kındığılı, E. H. Kaya, D. Keating, D. J. Kennett, S. D. Kesici, A. Khudaverdyan, K. Kiss, S. Kılıç, P. Klostermann, S. K. B. N. Valdes, S. Kovačević, M. Krenz-Niedbała, M. K. Škrivanko, R. Kurti, P. Kuzman, A. M. Lawson, C. Lazar, K. Leshtakov, T. E. Levy, I. Liritzis, K. O. Lorentz, S. Łukasik, M. Mah, S. Mallick, K. Mandl, K. Martirosyan-Olshansky, R. Matthews, W. Matthews, K. McSweeney, V. Melikyan, A. Micco, M. Michel, L. Milašinović, A. Mittnik, J. M. Monge, G. Nekhrizov, R. Nicholls, A. G. Nikitin, V. Nikolov, M. Novak, I. Olalde, J. Oppenheimer, A. Osterholtz, C. Özdemir, K. T. Özdoğan, N. Öztürk, N. Papadimitriou, N. Papakonstantinou, A. Papathanasiou, L. Paraman, E. G. Paskary, N. Patterson, I. Petrakiev, L. Petrosyan, V. Petrova, A. Philippa-Touchais, A. Piliposyan, N. P. Kuzman, H. Potrebica, B. Preda-

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bălănică, Z. Premužić, T. D. Price, L. Qiu, S. Radović, K. R. Aziz, P. R. Šikanjić, K. R. Raheem, S. Razumov, A. Richardson, J. Roodenberg, R. Ruka, V. Russeva, M. Şahin, A. Şarbak, E. Savaş, C. Schattke, L. Schepartz, T. Selçuk, A. Sevim-Erol, M. Shamoon-Pour, H. M. Shephard, A. Sideris, A. Simalcsik, H. Simonyan, V. Sinika, K. Sirak, G. Sirbu, M. Šlaus, A. Soficaru, B. Söğüt, A. Sołtysiak, Ç. Sönmez-Sözer, M. Stathi, M. Steskal, K. Stewardson, S. Stocker, F. Suata-Alpaslan, A. Suvorov, A. Szécsényi-Nagy, T. Szeniczey, N. Telnov, S. Temov, N. Todorova, U. Tota, G. Touchais, S. Triantaphyllou, A. Türker, M. Ugarković, T. Valchev, F. Veljanovska, Z. Videvski, C. Virag, A. Wagner, S. Walsh, P. Włodarczak, J. N. Workman, A. Yardumian, E. Yarovoy, A. Y. Yavuz, H. Yılmaz, F. Zalzala, A. Zettl, Z. Zhang, R. Çavuşoğlu, N. Rohland, R. Pinhasi, D. Reich, A genetic probe into the ancient and medieval history of Southern Europe and West Asia. *Science (80-.).* **377**, 940–951 (2022).

17. J. Koder, "On the Slavic Immigration in the Byzantine Balkans" in *Migration Histories of the Medieval Afroeurasian Transition: Aspects of Mobility Between Africa, Asia and Europe, 300-1500*, J. Preiser-Kapeller, L. Reinfandt, Y. Stouraitis, Eds. (Brill, Leiden, 2020), pp. 81–100.

18. A. Hyllested, B. D. Joseph, "Albanian" in *The indo-European Language family. A phylogenetic perspective.*, T. Oleander, Ed. (Cambridge University Press, Cambridge, 2022), pp. 223–245.

19. E. P. Hamp, Albanian and Messapic. *Stud. Present. to Joshua Whatmough*, 73–89 (1957).

20. C. Porphyrogenitus, *Constantine Porphyrogenitus. De administrando imperio. Greek Text edited by Gy. Moravcsik, english translation by R. J. H. Jenkins.* (Dumbarton Oaks, Washington, D.C., 1967).

21. C. Stallman-Pacitti, *The Life of Saint Pankratios of Taormina. Edited by John B. Burke. Byzantina Australiensia, 22.* (Brill, Leiden, 2018).

22. E. L. Vranoussi, Les termes Ἀλβανοί et Ἀρβανῖται et la première mention des Albanais dans les sources du XIe siècle. *Byzantina Symmeikta*. **2**, 207–254 (1970).

23. X. Ylli, *Das slavische Lehngut im Albanischen* (Verlag Otto Sagner, München, 2018).

24. P. Malingoudis, *Slavs in Medieval Greece* (Kyriakidi Publishing, Thessaloniki, 2013).

25. Ž. Muljačić, *Das Dalmatische: Studien zu einer untergegangenen Sprache* (Bohlau, Vienna, 2000).

26. R. Elsie, *Early Albania: A Reader of Historical Texts 11th-17th Centuries.* (Otto Harrassowitz Verlag, Wiesbaden, 2003).

27. R. Elsie, The earliest references to the existence of the Albanian language. *Zeitschrift für Balk.* **27**, 101–105 (1991).

28. S. Demiraj, *The Origin of the Albanians: Linguistically Investigated* (Academy of Sciences of Albania, Tirana, 2006).

29. W. Cimochowski, Prejardhja e gjuhës shqipe. *Bul. i Univ. shtetëror të Tiranës, Ser. e Shk. shoqërore*. **12**, 37–53 (1958).

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30. K. Ackermann, M. Gavranovic, J. Matzinger, "The Indo-Europeanization of the Balkans: Some new insights at the interface of Archaeology, archaeogenetics and historical linguistics" in *The Secondary Homelands of the Indo-European Languages (IG-AT2022)* (Leiden, The Netherlands, 2022).

31. E. P. Hamp, Varia. *Etudes Celt.* **32**, 87–90 (1996).

32. J. Matzinger, Messapico e illirico. *L'Idomeneo*. **9**, 57–66 (2015).

33. E. P. Hamp, "Albanian" in *Linguistics in Western Europe. Part 2. The study of languages.*, E. Haugen, Ed. (De Gruyter Mouton, Berlin, 1972), pp. 1626–1692.

34. J. Matzinger, *Die albanische Autochthoniehypothese aus der Sicht der Sprachwissenschaft *. überarbeitete Fassung eines Thesenpapiers vom Runden Tisch in Regensburg (Südosteuropa-Institut) am 24.04.2009.* (Vienna, 2016).

35. V. E. Orel, *Albanian etymological dictionary.* (Brill, Leiden, The Netherlands, 1998).

36. E. C. Polomé, "Balkan languages (Illyrian, Thracian, and Daco-Moesian)" in *The Cambridge Ancient History. Part III. The prehistory of the Balkans, the Middle East and the Aegean world. Tenth to eighth centuries B.C.*, J. Boardman, I. E. S. Edwards, N. G. L. Hammond, E. Sollberger, Eds. (Cambridge University Press, Cambridge, 2021), pp. 866–888.

37. J. Maran, "Archaeological cultures, fabricated ethnicities and DNA research:"Minoans" and "Mycenaeans" as case examples." in *Material, Method, and Meaning Papers in Eastern Mediterranean Archaeologyin Honor of Ilan Sharon. ÄGYPTEN UND ALTES TESTAMENT Studien zu Geschichte, Kultur und Religion Ägyptens und des Alten Testaments Band 110.*, U. Davidovich, N. Yahalom-Mack, S. Matskevich, Eds. (Zaphon, Münster, 2022), pp. 7–25.

38. W. Chang, C. Bowern, R. Gray, S. Greenhill, H. Haynie, G. Holland, L. Michael, Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language (Baltim).* **91**, 194–244 (2015).

39. R. Thorsø, "Two Balkan Indo-European Loanwords" in *Dispersals and Diversification. Linguistic and Archaeological Perspectives on the Early Stages of Indo-European.*, M. Serangeli, T. Olander, Eds. (Brill, Leiden, The Netherlands, 2019), pp. 251–262.

40. W. Bowden, The construction of identies in post-Roman Albania. *Late Antiq. Archaeol.* **1**, 57–78 (2003).

41. E. Nallbani, "Early Medieval North Albania: New Discoveries, Remodeling Connections The Case of Medieval Komani" in *Adriatico altomedievale (VI-XI secolo) Scambi , porti , produzioni*, S. Gelichi, C. Negrelli, Eds. (Edizioni Ca' Foscari, Venice, 2017), pp. 311–346.

42. S. Anamali, Des Illyriens aux Albanais (Les anciens Albanais). *Iliria*. **5**, 23–40 (1976).

43. I. Olalde, S. Mallick, N. Patterson, N. Rohland, V. Villalba-Mouco, M. Silva, K. Dulias, C. J. Edwards, F. Gandini, M. Pala, P. Soares, M. Ferrando-Bernal, N. Adamski, N. Broomandkhoshbacht, O. Cheronet, B. J. Culleton, D. Fernandes, A. M. Lawson, M. Mah,

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

J. Oppenheimer, K. Stewardson, Z. Zhang, J. M. J. Arenas, I. J. T. Moyano, D. C. Salazar-García, P. Castanyer, M. Santos, J. Tremoleda, M. Lozano, P. G. Borja, J. Fernández-Eraso, J. A. Mujika-Alustiza, C. Barroso, F. J. Bermúdez, E. V. Mínguez, J. Burch, N. Coromina, D. Vivó, A. Cebrià, J. M. Fullola, O. García-Puchol, J. I. Morales, F. X. Oms, T. Majó, J. M. Vergès, A. Díaz-Carvajal, I. Ollich-Castanyer, F. J. López-Cachero, A. M. Silva, C. Alonso-Fernández, G. D. de Castro, J. J. Echevarría, A. Moreno-Márquez, G. P. Berlanga, P. Ramos-García, J. Ramos-Muñoz, E. V. Vila, G. A. Arzo, Á. E. Arroyo, K. T. Lillios, J. Mack, J. Velasco-Vázquez, A. Waterman, L. B. de Lugo Enrich, M. B. Sánchez, B. Agustí, F. Codina, G. de Prado, A. Estalrrich, Á. F. Flores, C. Finlayson, G. Finlayson, S. Finlayson, F. Giles-Guzmán, A. Rosas, V. B. González, G. G. Atiénzar, M. S. H. Pérez, A. Llanos, Y. C. Marco, I. C. Beneyto, D. López-Serrano, M. S. Tormo, A. C. Valera, C. Blasco, C. Liesau, P. Ríos, J. Daura, M. J. de Pedro Michó, A. A. Diez-Castillo, R. F. Fernández, J. F. Farré, R. Garrido-Pena, V. S. Gonçalves, E. Guerra-Doce, A. M. Herrero-Corral, J. Juan-Cabanilles, D. López-Reyes, S. B. McClure, M. M. Pérez, A. O. Foix, M. S. Borràs, A. C. Sousa, J. M. V. Encinas, D. J. Kennett, M. B. Richards, K. W. Alt, W. Haak, R. Pinhasi, C. Lalueza-Fox, D. Reich, The genomic history of the Iberian Peninsula over the past 8000 years. *Science (80-.).* **363**, 1230–1234 (2019).

44. C. Posth, V. Zaro, M. A. Spyrou, S. Vai, G. A. Gnecchi-Ruscone, A. Modi, A. Peltzer, A. Mötsch, K. Nägele, Å. J. Vågene, E. A. Nelson, R. Radzevičiūtė, C. Freund, L. M. Bondioli, L. Cappuccini, H. Frenzel, E. Pacciani, F. Boschin, G. Capecchi, I. Martini, A. Moroni, S. Ricci, A. Sperduti, M. A. Turchetti, A. Riga, M. Zavattaro, A. Zifferero, H. O. Heyne, E. Fernández-Domínguez, G. J. Kroonen, M. McCormick, W. Haak, M. Lari, G. Barbujani, L. Bondioli, K. I. Bos, D. Caramelli, J. Krause, The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. *Sci. Adv.* **7**, eabi7673 (2021).

45. L. Papac, M. Ernée, M. Dobeš, M. Langová, A. B. Rohrlach, F. Aron, G. U. Neumann, M. A. Spyrou, N. Rohland, P. Velemínský, M. Kuna, H. Brzobohatá, B. Culleton, D. Daněček, A. Danielisová, M. Dobisíková, J. Hložek, D. J. Kennett, J. Klementová, M. Kostka, P. Krištuf, M. Kuchařík, J. K. Hlavová, P. Limburský, D. Malyková, L. Mattiello, M. Pecinovská, K. Petriščáková, E. Průchová, P. Stránská, L. Smejtek, J. Špaček, R. Šumberová, O. Švejcar, M. Trefný, M. Vávra, J. Kolář, V. Heyd, J. Krause, R. Pinhasi, D. Reich, S. Schiffels, W. Haak, Dynamic changes in genomic and social structures in third millennium BCE central Europe. *Sci. Adv.* **7** (2021), doi:10.1126/sciadv.abi6941.

46. G. Scorrano, F. E. Yediay, T. Pinotti, M. Feizabadifarahani, K. Kristiansen, The genetic and cultural impact of the Steppe migration into Europe. *Ann. Hum. Biol.* **48**, 223–233 (2021).

47. B. A. Olsen, *Tracing the Indo-Europeans: New evidence from archaeology and historical linguistics* (Oxbow Books, Oxford, 2019).

48. A. Kushniarevich, O. Utevska, M. Chuhryaeva, A. Agdzhoyan, K. Dibirova, I. Uktveryte, M. Möls, L. Mulahasanovic, A. Pshenichnov, S. Frolova, A. Shanko, E. Metspalu, M. Reidla, K. Tambets, E. Tamm, S. Koshel, V. Zaporozhchenko, L. Atramentova, V. Kučinskas, O. Davydenko, O. Goncharova, I. Evseeva, M. Churnosov, E. Pocheshchova, B. Yunusbayev, E. Khusnutdinova, D. Marjanović, P. Rudan, S. Rootsi, N. Yankovsky, P. Endicott, A. Kassian, A. Dybo, T. G. Consortium, C. Tyler-Smith, E. Balanovska, M. Metspalu, T. Kivisild, R. Villems, O. Balanovsky, Genetic Heritage of the Balto-Slavic

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. *PLoS One*. **10**, 1–19 (2015).

49. W. Haak, I. Lazaridis, N. Patterson, N. Rohland, S. Mallick, B. Llamas, G. Brandt, S. Nordenfelt, E. Harney, K. Stewardson, Q. Fu, A. Mittnik, E. Bánffy, C. Economou, M. Francken, S. Friederich, R. G. Pena, F. Hallgren, V. Khartanovich, A. Khokhlov, M. Kunst, P. Kuznetsov, H. Meller, O. Mochalov, V. Moiseyev, N. Nicklisch, S. L. Pichler, R. Risch, M. A. Rojo Guerra, C. Roth, A. Szécsényi-Nagy, J. Wahl, M. Meyer, J. Krause, D. Brown, D. Anthony, A. Cooper, K. W. Alt, D. Reich, Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. **522**, 207–211 (2015).

50. H. Ringbauer, Y. Huang, A. Akbari, S. Mallick, N. Patterson, D. Reich, ancIBD - Screening for identity by descent segments in human ancient DNA. *bioRxiv Prepr. Serv. Biol.* (2023), , doi:10.1101/2023.03.08.531671.

51. I. Lazaridis, N. Patterson, A. Mittnik, G. Renaud, S. Mallick, K. Kirsanow, P. H. Sudmant, J. G. Schraiber, S. Castellano, M. Lipson, B. Berger, C. Economou, R. Bollongino, Q. Fu, K. I. Bos, S. Nordenfelt, H. Li, C. de Filippo, K. Prüfer, S. Sawyer, C. Posth, W. Haak, F. Hallgren, E. Fornander, N. Rohland, D. Delsate, M. Francken, J.-M. Guinet, J. Wahl, G. Ayodo, H. A. Babiker, G. Bailliet, E. Balanovska, O. Balanovsky, R. Barrantes, G. Bedoya, H. Ben-Ami, J. Bene, F. Berrada, C. M. Bravi, F. Brisighelli, G. B. J. Busby, F. Cali, M. Churnosov, D. E. C. Cole, D. Corach, L. Damba, G. van Driem, S. Dryomov, J.-M. Dugoujon, S. A. Fedorova, I. Gallego Romero, M. Gubina, M. Hammer, B. M. Henn, T. Hervig, U. Hodoglugil, A. R. Jha, S. Karachanak-Yankova, R. Khusainova, E. Khusnutdinova, R. Kittles, T. Kivisild, W. Klitz, V. Kučinskas, A. Kushniarevich, L. Laredj, S. Litvinov, T. Loukidis, R. W. Mahley, B. Melegh, E. Metspalu, J. Molina, J. Mountain, K. Näkkäläjärvi, D. Nesheva, T. Nyambo, L. Osipova, J. Parik, F. Platonov, O. Posukh, V. Romano, F. Rothhammer, I. Rudan, R. Ruizbakiev, H. Sahakyan, A. Sajantila, A. Salas, E. B. Starikovskaya, A. Tarekegn, D. Toncheva, S. Turdikulova, I. Uktveryte, O. Utevska, R. Vasquez, M. Villena, M. Voevoda, C. A. Winkler, L. Yepiskoposyan, P. Zalloua, T. Zemunik, A. Cooper, C. Capelli, M. G. Thomas, A. Ruiz-Linares, S. A. Tishkoff, L. Singh, K. Thangaraj, R. Villems, D. Comas, R. Sukernik, M. Metspalu, M. Meyer, E. E. Eichler, J. Burger, M. Slatkin, S. Pääbo, J. Kelso, D. Reich, J. Krause, Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. **513**, 409–413 (2014).

52. B. W. Fortson, Indo-European language and culture: an introduction. *Blackwell Textb. Linguist.* **3** (2004), pp. xviii, 468, (available at http://ccat.sas.upenn.edu/bmcr/2007/2007-05-15.html).

53. M. E. Allentoft, M. Sikora, K.-G. Sjögren, S. Rasmussen, M. Rasmussen, J. Stenderup, P. B. Damgaard, H. Schroeder, T. Ahlström, L. Vinner, A.-S. Malaspinas, A. Margaryan, T. Higham, D. Chivall, N. Lynnerup, L. Harvig, J. Baron, P. Della Casa, P. Dąbrowski, P. R. Duffy, A. V Ebel, A. Epimakhov, K. Frei, M. Furmanek, T. Gralak, A. Gromov, S. Gronkiewicz, G. Grupe, T. Hajdu, R. Jarysz, V. Khartanovich, A. Khokhlov, V. Kiss, J. Kolář, A. Kriiska, I. Lasak, C. Longhi, G. McGlynn, A. Merkevicius, I. Merkyte, M. Metspalu, R. Mkrtchyan, V. Moiseyev, L. Paja, G. Pálfi, D. Pokutta, Ł. Pospieszny, T. D. Price, L. Saag, M. Sablin, N. Shishlina, V. Smrčka, V. I. Soenov, V. Szeverényi, G. Tóth,

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S. V Trifanova, L. Varul, M. Vicze, L. Yepiskoposyan, V. Zhitenev, L. Orlando, T. Sicheritz-Pontén, S. Brunak, R. Nielsen, K. Kristiansen, E. Willerslev, Population genomics of Bronze Age Eurasia. *Nature*. **522**, 167–172 (2015).

54. I. Olalde, S. Brace, M. E. Allentoft, I. Armit, K. Kristiansen, T. Booth, N. Rohland, S. Mallick, A. Szécsényi-Nagy, A. Mittnik, E. Altena, M. Lipson, I. Lazaridis, T. K. Harper, N. Patterson, N. Broomandkhoshbacht, Y. Diekmann, Z. Faltyskova, D. Fernandes, M. Ferry, E. Harney, P. de Knijff, M. Michel, J. Oppenheimer, K. Stewardson, A. Barclay, K. W. Alt, C. Liesau, P. Ríos, C. Blasco, J. V. Miguel, R. M. García, A. A. Fernández, E. Bánffy, M. Bernabò-Brea, D. Billoin, C. Bonsall, L. Bonsall, T. Allen, L. Büster, S. Carver, L. C. Navarro, O. E. Craig, G. T. Cook, B. Cunliffe, A. Denaire, K. E. Dinwiddy, N. Dodwell, M. Ernée, C. Evans, M. Kuchařík, J. F. Farré, C. Fowler, M. Gazenbeek, R. G. Pena, M. Haber-Uriarte, E. Haduch, G. Hey, N. Jowett, T. Knowles, K. Massy, S. Pfrengle, P. Lefranc, O. Lemercier, A. Lefebvre, C. H. Martínez, V. G. Olmo, A. B. Ramírez, J. L. Maurandi, T. Majó, J. I. McKinley, K. McSweeney, B. G. Mende, A. Modi, G. Kulcsár, V. Kiss, A. Czene, R. Patay, A. Endrődi, K. Köhler, T. Hajdu, T. Szeniczey, J. Dani, Z. Bernert, M. Hoole, O. Cheronet, D. Keating, P. Velemínský, M. Dobeš, F. Candilio, F. Brown, R. F. Fernández, A.-M. Herrero-Corral, S. Tusa, E. Carnieri, L. Lentini, A. Valenti, A. Zanini, C. Waddington, G. Delibes, E. Guerra-Doce, B. Neil, M. Brittain, M. Luke, R. Mortimer, J. Desideri, M. Besse, G. Brücken, M. Furmanek, A. Hałuszko, M. Mackiewicz, A. Rapiński, S. Leach, I. Soriano, K. T. Lillios, J. L. Cardoso, M. P. Pearson, P. Włodarczak, T. D. Price, P. Prieto, P.-J. Rey, R. Risch, M. A. Rojo Guerra, A. Schmitt, J. Serralongue, A. M. Silva, V. Smrčka, L. Vergnaud, J. Zilhão, D. Caramelli, T. Higham, M. G. Thomas, D. J. Kennett, H. Fokkens, V. Heyd, A. Sheridan, K.-G. Sjögren, P. W. Stockhammer, J. Krause, R. Pinhasi, W. Haak, I. Barnes, C. Lalueza-Fox, D. Reich, The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*. **555**, 190–196 (2018).

55. M. Korkuti, *Archaeological studies on the prehistory of Albania.* (Academy of Sciences of Albania, Tirana, 2013).

56. I. Mathieson, S. Alpaslan-Roodenberg, C. Posth, A. Szécsényi-Nagy, N. Rohland, S. Mallick, I. Olalde, N. Broomandkhoshbacht, F. Candilio, O. Cheronet, D. Fernandes, M. Ferry, B. Gamarra, G. G. Fortes, W. Haak, E. Harney, E. Jones, D. Keating, B. Krause-Kyora, I. Kucukkalipci, M. Michel, A. Mittnik, K. Nägele, M. Novak, J. Oppenheimer, N. Patterson, S. Pfrengle, K. Sirak, K. Stewardson, S. Vai, S. Alexandrov, K. W. Alt, R. Andreescu, D. Antonović, A. Ash, N. Atanassova, K. Bacvarov, M. B. Gusztáv, H. Bocherens, M. Bolus, A. Boroneanţ, Y. Boyadzhiev, A. Budnik, J. Burmaz, S. Chohadzhiev, N. J. Conard, R. Cottiaux, M. Čuka, C. Cupillard, D. G. Drucker, N. Elenski, M. Francken, B. Galabova, G. Ganetsovski, B. Gély, T. Hajdu, V. Handzhyiska, K. Harvati, T. Higham, S. Iliev, I. Janković, I. Karavanić, D. J. Kennett, D. Komšo, A. Kozak, D. Labuda, M. Lari, C. Lazar, M. Leppek, K. Leshtakov, D. Lo Vetro, D. Los, I. Lozanov, M. Malina, F. Martini, K. McSweeney, H. Meller, M. Menđušić, P. Mirea, V. Moiseyev, V. Petrova, T. D. Price, A. Simalcsik, L. Sineo, M. Šlaus, V. Slavchev, P. Stanev, A. Starović, T. Szeniczey, S. Talamo, M. Teschler-Nicola, C. Thevenet, I. Valchev, F. Valentin, S. Vasilyev, F. Veljanovska, S. Venelinova, E. Veselovskaya, B. Viola, C. Virag, J. Zaninović, S. Zäuner, P. W. Stockhammer, G. Catalano, R. Krauß, D. Caramelli, G. Zariņa, B. Gaydarska, M. Lillie, A. G. Nikitin, I. Potekhina, A.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Papathanasiou, D. Borić, C. Bonsall, J. Krause, R. Pinhasi, D. Reich, The genomic history of southeastern Europe. *Nature*. **555**, 197–203 (2018).

57. C. Schmid, S. Schiffels, Estimating human mobility in Holocene Western Eurasia with large-scale ancient genomic data. *Proc. Natl. Acad. Sci.* **120**, e2218375120 (2023).

58. L. J. Reitsema, A. Mittnik, B. Kyle, G. Catalano, P. F. Fabbri, A. C. S. Kazmi, K. L. Reinberger, L. Sineo, S. Vassallo, R. Bernardos, N. Broomandkhoshbacht, K. Callan, F. Candilio, O. Cheronet, E. Curtis, D. Fernandes, M. Lari, A. M. Lawson, M. Mah, S. Mallick, K. Mandl, A. Micco, A. Modi, J. Oppenheimer, K. T. Özdogan, N. Rohland, K. Stewardson, S. Vai, C. Vergata, J. N. Workman, F. Zalzala, V. Zaro, A. Achilli, A. Anagnostopoulos, C. Capelli, V. Constantinou, H. Lancioni, A. Olivieri, A. Papadopoulou, N. Psatha, O. Semino, J. Stamatoyannopoulos, I. Valliannou, E. Yannaki, I. Lazaridis, N. Patterson, H. Ringbauer, D. Caramelli, R. Pinhasi, D. Reich, The diverse genetic origins of a Classical period Greek army. *Proc. Natl. Acad. Sci.* **119**, e2205272119 (2022).

59. M. L. Antonio, Z. Gao, H. M. Moots, M. Lucci, F. Candilio, S. Sawyer, V. Oberreiter, D. Calderon, K. Devitofranceschi, R. C. Aikens, S. Aneli, F. Bartoli, A. Bedini, O. Cheronet, D. J. Cotter, D. M. Fernandes, G. Gasperetti, R. Grifoni, A. Guidi, F. La Pastina, E. Loreti, D. Manacorda, G. Matullo, S. Morretta, A. Nava, V. F. Nicolai, F. Nomi, C. Pavolini, M. Pentiricci, P. Pergola, M. Piranomonte, R. Schmidt, G. Spinola, A. Sperduti, M. Rubini, L. Bondioli, A. Coppa, R. Pinhasi, J. K. Pritchard, Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science (80-.).* **366**, 708–714 (2019).

60. D. M. Fernandes, A. Mittnik, I. Olalde, I. Lazaridis, O. Cheronet, N. Rohland, S. Mallick, R. Bernardos, N. Broomandkhoshbacht, J. Carlsson, B. J. Culleton, M. Ferry, B. Gamarra, M. Lari, M. Mah, M. Michel, A. Modi, M. Novak, J. Oppenheimer, K. A. Sirak, K. Stewardson, K. Mandl, C. Schattke, K. T. Özdoğan, M. Lucci, G. Gasperetti, F. Candilio, G. Salis, S. Vai, E. Camarós, C. Calò, G. Catalano, M. Cueto, V. Forgia, M. Lozano, E. Marini, M. Micheletti, R. M. Miccichè, M. R. Palombo, D. Ramis, V. Schimmenti, P. Sureda, L. Teira, M. Teschler-Nicola, D. J. Kennett, C. Lalueza-Fox, N. Patterson, L. Sineo, A. Coppa, D. Caramelli, R. Pinhasi, D. Reich, The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nat. Ecol. Evol.* **4**, 334–345 (2020).

61. G. A. Gnecchi-Ruscone, A. Szécsényi-Nagy, I. Koncz, G. Csiky, Z. Rácz, A. B. Rohrlach, G. Brandt, N. Rohland, V. Csáky, O. Cheronet, B. Szeifert, T. Á. Rácz, A. Benedek, Z. Bernert, N. Berta, S. Czifra, J. Dani, Z. Farkas, T. Hága, T. Hajdu, M. Jászberényi, V. Kisjuhász, B. Kolozsi, P. Major, A. Marcsik, B. N. Kovacsóczy, C. Balogh, G. M. Lezsák, J. G. Ódor, M. Szelekovszky, T. Szeniczey, J. Tárnoki, Z. Tóth, E. K. Tutkovics, B. G. Mende, P. Geary, W. Pohl, T. Vida, R. Pinhasi, D. Reich, Z. Hofmanová, C. Jeong, J. Krause, Ancient genomes reveal origin and rapid trans-Eurasian migration of 7th century Avar elites. *Cell*. **185**, 1402-1413.e21 (2022).

62. P. Charanis, "Kouver, the chronology of his activities and their ethnic effects on the regions around Thessalonica" in *Studies on the demography of the Byzantine empire* (Variorum Collected Studies, Ashgate, Farnham, 1970), pp. 230–247.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

63. Nikephoros, *Nicephori archiepiscopi Constantinopolitani Breviarium historicum de rebus gestis post imperium Mauricii (Corpus Fontium Historiae Byzantinae 13)* (Washington, D.C., 1990).

64. A. DuNay, *The origins of the Rumanians: The early history of the Rumanian language* (Matthias Corvinus Publishing, 1996).

65. S. Pulaha, *Popullsia shqiptare e Kosovës gjatë shek: studime dhe dokumente.* (8 Nëntori, Tirana, 1984).

66. I. Rexha, *Registration of settlements and Albanian population of Kosovo (According to the Ottoman defters (records) of XV century)* (Institute of History Ali Hadri - Prishtina, Prishtina, 2016).

67. M. Mihajlovic, V. Tanasic, M. K. Markovic, M. Kecmanovic, D. Keckarevic, Distribution of Y-chromosome haplogroups in Serbian population groups originating from historically and geographically significant distinct parts of the Balkan Peninsula. *Forensic Sci. Int. Genet.* **61**, 102767 (2022).

68. D. Primorac, V. Škaro, P. Projić, S. Missoni, I. H. Zanki, S. Merkaš, J. Šarac, N. Novokmet, A. Ledić, A. Makar, G. Lauc, Š. Andelinović, Ž. Bašić, I. Kružić, M. Neuberg, M. Smolić, R. Smolić, I. Hrstić, D. Trivanović, R. Konjhodžić, L. Salihefendić, N. B. Jordamović, D. Marjanović, Croatian genetic heritage: An updated Y-chromosome story. *Croat. Med. J.* **63**, 273–286 (2022).

69. R. Elsie, *The tribes of Albania: history, society and culture* (IB Tauris, London, 2015).

70. E. Durham, *High Albania* (Edward Arnold, London, 1909).

71. G. De Rapper, "Blood and seed, trunk and hearth: kinship and common origin in southern Albania" in *Albania: Family, Society and Culture in the 20th century*, E. P. Andreas Hemming, Gentiana Kera, Ed. (LIT Verlag, Münster, 2012), pp. 79–95.

72. P. Ralph, G. Coop, The Geography of Recent Genetic Ancestry across Europe. *PLOS Biol.* **11**, 1–20 (2013).

73. E. Gilbert, A. Shanmugam, G. L. Cavalleri, Revealing the recent demographic history of Europe via haplotype sharing in the UK Biobank. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2119281119 (2022).

74. É. Harney, A. Nayak, N. Patterson, P. Joglekar, V. Mushrif-Tripathy, S. Mallick, N. Rohland, J. Sedig, N. Adamski, R. Bernardos, N. Broomandkhoshbacht, B. J. Culleton, M. Ferry, T. K. Harper, M. Michel, J. Oppenheimer, K. Stewardson, Z. Zhang, Harashawaradhana, M. S. Bartwal, S. Kumar, S. C. Diyundi, P. Roberts, N. Boivin, D. J. Kennett, K. Thangaraj, D. Reich, N. Rai, Ancient DNA from the skeletons of Roopkund Lake reveals Mediterranean migrants in India. *Nat. Commun.* **10**, 3670 (2019).

75. S. Sarno, S. Tofanelli, S. De Fanti, A. Quagliariello, E. Bortolini, G. Ferri, P. Anagnostou, F. Brisighelli, C. Capelli, G. Tagarelli, L. Sineo, D. Luiselli, A. Boattini, D. Pettener, Shared language, diverging genetic histories: high-resolution analysis of Y-chromosome variability in Calabrian and Sicilian Arbereshe. *Eur. J. Hum. Genet.* **24**, 600–606 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

76. B. Martínez-Cruz, I. Mendizabal, C. Harmant, R. de Pablo, M. Ioana, D. Angelicheva, A. Kouvatsi, H. Makukh, M. G. Netea, H. Pamjav, A. Zalán, I. Tournev, E. Marushiakova, V. Popov, J. Bertranpetit, L. Kalaydjieva, L. Quintana-Murci, D. Comas, and the G. Consortium, Origins, admixture and founder lineages in European Roma. *Eur. J. Hum. Genet.* **24**, 937–943 (2016).

77. V. Urasin, YFull. Y-Chr Sequence Interpretation Service. *yfull.com* (2023), (available at yfull.com).

78. L. Gene by Gene, FamilyTreeDNA Discover BetaTM (2023), (available at https://discover.familytreedna.com/).

79. A. Žegarac, L. Winkelbach, J. Blöcher, Y. Diekmann, M. Krečković Gavrilović, M. Porčić, B. Stojković, L. Milašinović, M. Schreiber, D. Wegmann, K. R. Veeramah, S. Stefanović, J. Burger, Ancient genomes provide insights into family structure and the heredity of social status in the early Bronze Age of southeastern Europe. *Sci. Rep.* **11**, 10072 (2021).

80. T. B. et al; Fulvio Cruciani; Roberta La Fratta, Trombetta, Beniamino; Santomalazza, Piero; Sellitto, Daniele; Eliane, Beraud Colomb; Jean-Michel, Dugoujon, Federica, Crivellaro, Tracing Past Human Male Movements in Northern/Eastern Africa and Western Eurasia: New Clues from Y-Chromosomal Haplogroups E-M78 and J-M12. *Mol. Biol. Evol.* **24**, 1300–1311 (2007).

81. M. Pericić, L. B. Lauc, I. M. Klarić, S. Rootsi, B. Janićijevic, I. Rudan, R. Terzić, I. Colak, A. Kvesić, D. Popović, A. Sijacki, I. Behluli, D. Dordevic, L. Efremovska, D. D. Bajec, B. D. Stefanović, R. Villems, P. Rudan, High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. *Mol. Biol. Evol.* **22**, 1964–1975 (2005).

82. L. Kovacevic, K. Tambets, A.-M. Ilumäe, A. Kushniarevich, B. Yunusbayev, A. Solnik, T. Bego, D. Primorac, V. Skaro, A. Leskovac, Z. Jakovski, K. Drobnic, H.-V. Tolk, S. Kovacevic, P. Rudan, E. Metspalu, D. Marjanovic, Standing at the Gateway to Europe - The Genetic Structure of Western Balkan Populations Based on Autosomal and Haploid Markers. *PLoS One*. **9**, 1–15 (2014).

83. S. Tofanelli, F. Brisighelli, P. Anagnostou, G. B. J. Busby, G. Ferri, M. G. Thomas, L. Taglioli, I. Rudan, T. Zemunik, C. Hayward, D. Bolnick, V. Romano, F. Cali, D. Luiselli, G. B. Shepherd, S. Tusa, A. Facella, C. Capelli, The Greeks in the West: genetic signatures of the Hellenic colonisation in southern Italy and Sicily. *Eur. J. Hum. Genet.* **24**, 429–436 (2016).

84. J. Matzinger, "Historical toponymy of Albania. A planned project." in *Balkan area Alatic studies general linguistics. In memoriam Albina H. Girfanova (1957–2018)*, A. V. Andronov, C. V. Lesnikov, A. N. Cobolev, Eds. (Nestor-Istoria, Saint-Petersburg, 2019), pp. 278-287.

85. W. Breu, *Romance in Contact with Albanian* (2022), Oxford Research Encyclopedia of Linguistics, doi:10.1093/acrefore/9780199384655.013.421.

86. Ivor Janković, M. Novak, The mysteries of the past - finds from the Bezdanjača Cave near

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Vrhovine. *Archaeol. Speleol.*, 65–74 (2019).

87. B. Philippsen, The freshwater reservoir effect in radiocarbon dating. *Herit. Sci.* **1**, 24 (2013).

88. N. Patterson, A. L. Price, D. Reich, Population Structure and Eigenanalysis. *PLOS Genet.* **2**, 1–20 (2006).

89. M. L. Antonio, C. L. W. Z. Gao, S. Sawyer, V. Oberreiter, H. M. Moots, J. P. Spence, O. Cheronet, B. Zagorc, E. Praxmarer, K. T. Özdo\u gan, L. Demetz, M. Lucci, T. Alihodžić, S. Amrani, P. Avetisyan, C. Baillif-Ducros, Ž. Bedić, A. Bertrand, M. Bilić, L. Bondioli, P. Borówka, E. Botte, J. Burmaz, D. Bužanić, F. Candilio, M. Cvetko, D. De Angelis, I. Drnić, K. Elschek, M. Fantar, A. Gaspari, G. Gasperetti, F. Genchi, S. Golubović, Z. Huke\v lová, R. Jankauskas, K. J. Vučković, G. Jeremić, I. Kaić, K. Kazek, H. Khachatryan, A. Khudaverdyan, S. Kirchengast, M. Korać, V. Kozlowski, M. Krošláková, D. K. Špalj, F. La Pastina, M. Laguardia, S. Legrand, T. Leleković, T. Leskovar, W. Lorkiewicz, D. Los, A. M. Silva, R. Masaryk, V. Matijević, Y. M. S. Cherifi, N. Meyer, I. Mikić, N. Miladinović-Radmilović, B. M. Zakić, L. Nacouzi, M. Natuniewicz-Sekuła, A. Nava, C. Neugebauer-Maresch, J. Nováček, A. Osterholtz, J. Paige, L. Paraman, D. Pieri, K. Pieta, S. Pop-Lazić, M. Ruttkay, M. Sanader, A. Sołtysiak, A. Sperduti, T. S. Pesterac, M. Teschler-Nicola, I. Teul, D. Tončinić, J. Trapp, D. Vulović, T. Waliszewski, D. Walter, M. Zivanovic, M. el Mostefa Filah, M. Čaušević-Bully, M. Šlaus, D. Boric, M. Novak, A. Coppa, R. Pinhasi, J. K. Pritchard, Stable population structure in Europe since the Iron Age, despite high mobility. *bioRxiv* (2022), doi:10.1101/2022.05.15.491973.

90. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient Admixture in Human History. *Genetics*. **192**, 1065–1093 (2012).

91. É. Harney, N. Patterson, D. Reich, J. Wakeley, Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics*. **217** (2021), doi:10.1093/genetics/iyaa045.

92. T. J. A. Begg, A. Schmidt, A. Kocher, M. H. D. Larmuseau, G. Runfeldt, P. A. Maier, J. D. Wilson, R. Barquera, C. Maj, A. Szolek, M. Sager, S. Clayton, A. Peltzer, R. Hui, J. Ronge, E. Reiter, C. Freund, M. Burri, F. Aron, A. Tiliakou, J. Osborn, D. M. Behar, M. Boecker, G. Brandt, I. Cleynen, C. Strassburg, K. Prüfer, D. Kühnert, W. R. Meredith, M. M. Nöthen, R. D. Attenborough, T. Kivisild, J. Krause, Genomic analyses of hair from Ludwig van Beethoven. *Curr. Biol.* (2023), doi:https://doi.org/10.1016/j.cub.2023.02.041.

93. O. Balanovsky, V. Gurianov, V. Zaporozhchenko, O. Balaganskaya, V. Urasin, M. Zhabagin, V. Grugni, R. Canada, N. Al-Zahery, A. Raveane, S.-Q. Wen, S. Yan, X. Wang, P. Zalloua, A. Marafi, S. Koshel, O. Semino, C. Tyler-Smith, E. Balanovska, Phylogeography of human Y-chromosome haplogroup Q3-L275 from an academic/citizen science collaboration. *BMC Evol. Biol.* **17**, 18 (2017).

94. J. T. Robinson, H. Thorvaldsdottir, D. Turner, J. P. Mesirov, igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*. **39** (2022), doi:10.1093/bioinformatics/btac830.

bioRxiv preprint doi: https://doi.org/10.1101/2023.06.05.543790; this version posted June 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

95. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

96. C. B. C. Davis, M. Sager, G. Runfeldt, E. Greenspan, A. Bormans, B. Greenspan, "Big Y-700 white paper: powering discovery in the field of paternal ancestry" (2019).

97. Rs. Team, Rstudio: Integrated development for R. (2020), (available at http://www.rstudio.com).

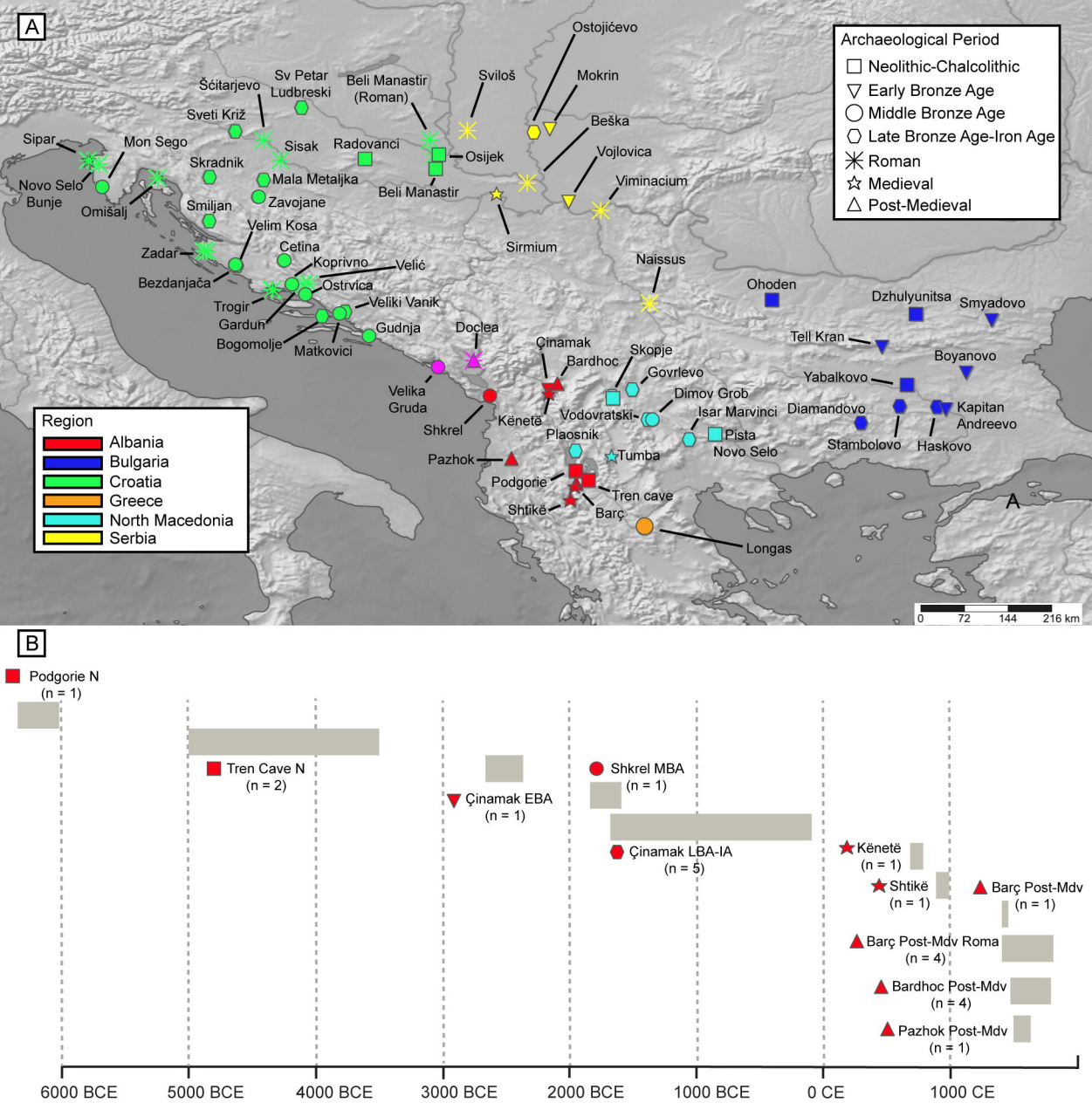
98. Alban Lauka; Gjergj Bojaxhi, RRËNJËT. Projekti Shqiptar i ADN-së. *rrenjet.com* (2023), (available at https://rrenjet.com/).

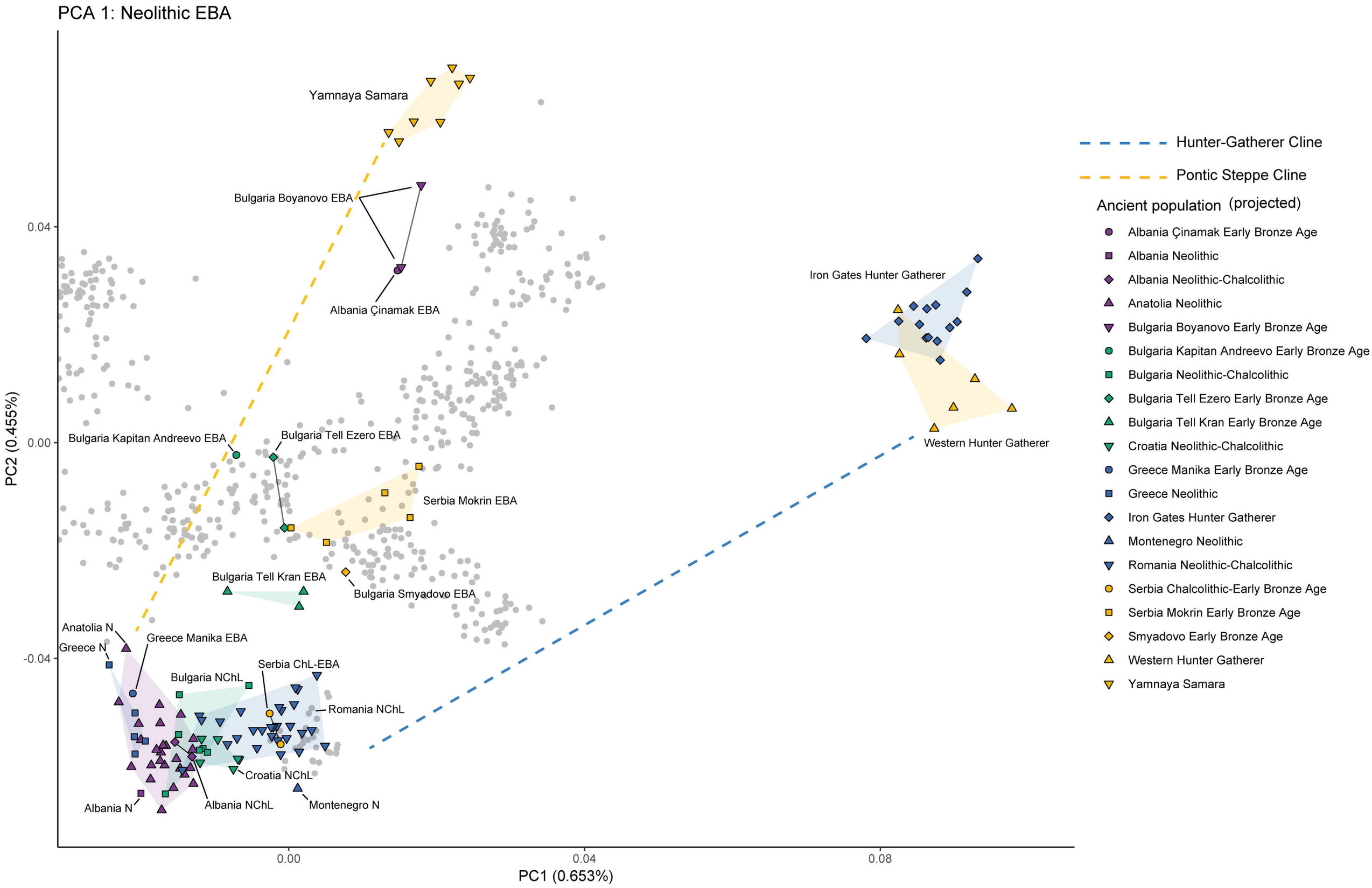
Acknowledgements

LRD acknowledges the Leverhulme Trust Early Career Fellowship grant (ECF-2021-199) for funding him during this research. The authors' contributions are as follows: Conceptualization: LRD. Data curation: LRD, AA. Formal Analysis: LRD, AA, DW, AH. Funding acquisition: LRD. Investigation: LRD. Methodology: LRD, AA, DW, AH. Project Administration: LRD. Visualization: LRD, AA, DW, AH. Writing – original draft: LRD. Writing – review and editing. LRD, AA, DW, AH.

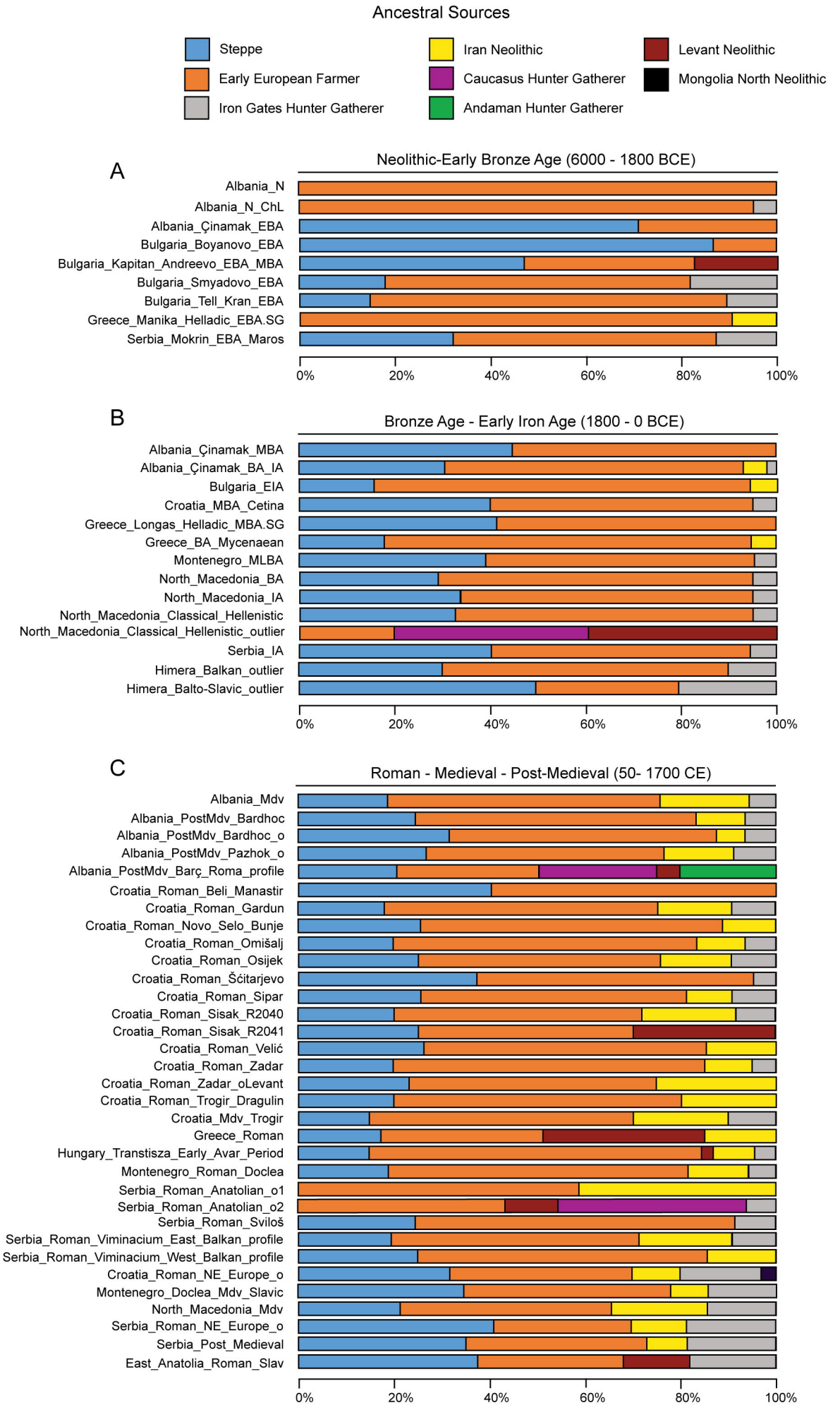
The authors are grateful to Joscha Gretzinger (Max Planck Institute for Evolutionary Anthropology, Leipzig) for undertaking the Mobest analyses. LRD thanks Andreas Kyropoulos and Leonidas Embirikos for extensive discussions on Albanian history and linguistics, and to James Kempton for assistance with QGIS mapping. LRD thanks Alexandros Spanos and Leo Cooper for assistance in compiling part of the ancient Greek Y-chromosome dataset (Table S25). The authors thank Ted Kandell (Open Genomes) and Milan Rajevac (Open Genomes, the Society of Serbian Genealogists "Poreklo") for providing detailed Y-chromosome haplogroup determinations for part of the examined dataset.

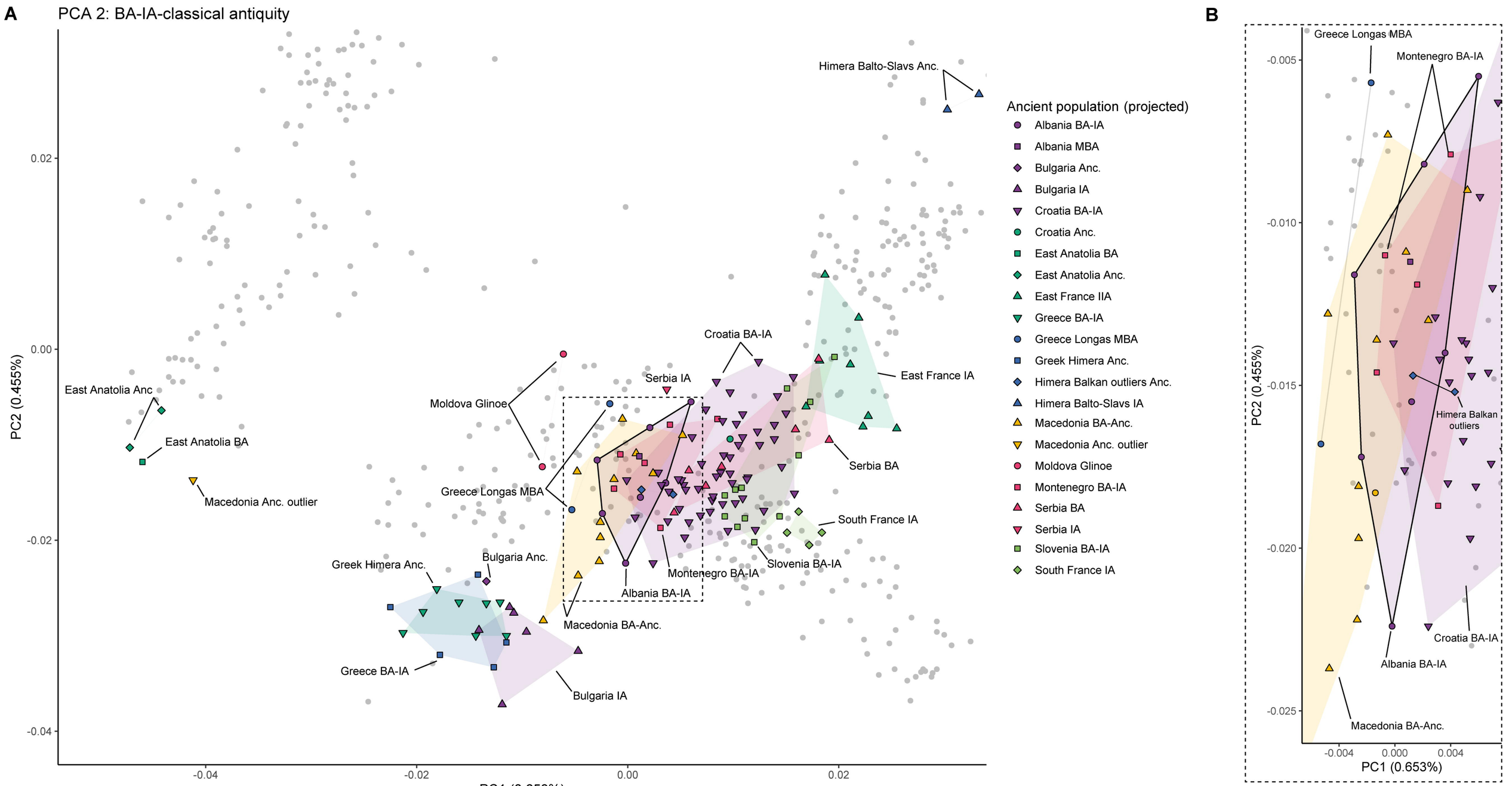
The authors declare that they have no competing interests. All data needed to evaluate the conclusions in the paper are present in the paper and the Supplementary Materials.

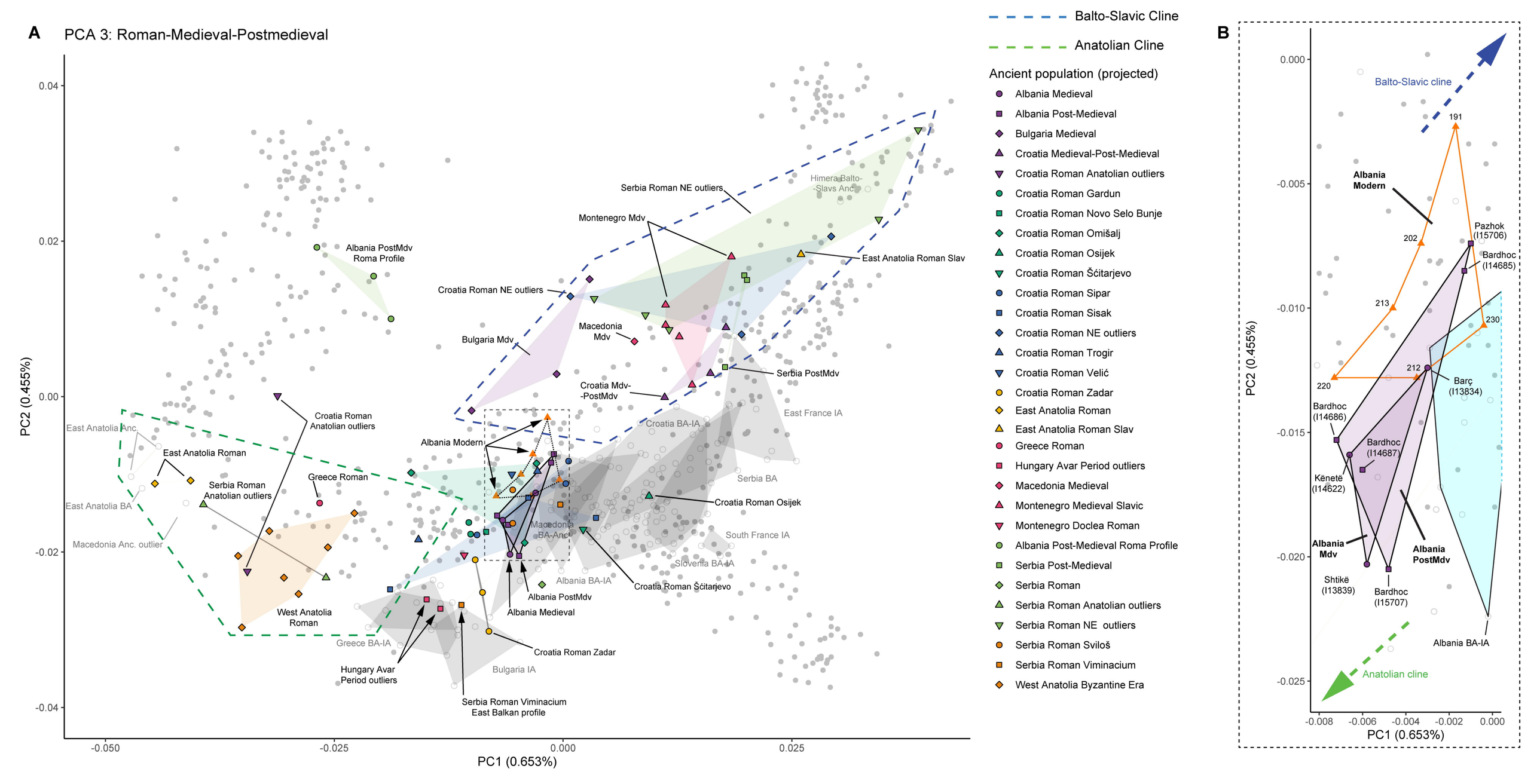

A
Archaeological Period
Neolithic-Chalcolithic
Early Bronze Age
Middle Bronze Age
Late Bronze Age-Iron Age
Roman
Medieval
Post-Medieval
Region
Albania
Bulgaria
Croatia
Greece
North Macedonia
Serbia
Ostojićevo
Mokrin
Sviloš
Beška
Vojlovica
Viminacium
Sirmium
Naissus
Beli Manastir (Roman)
Beli Manastir
Osijek
Radovanci
Šćitarjevo
Sv Petar Ludbreski
Sveti Križ
Sisak
Sipar
Mon Sego
Novo Selo
Bunje
Omišalj
Skradnik
Mala Metaljka
Zavojane
Smiljan
Velim Kosa
Zadar
Cetina
Koprivno
Velić
Bezdanjača
Ostrvica
Trogir
Garduh
Bogomolje
Matkovici
Veliki Vanik
Gudnja
Doclea
Velika Gruda
Shkrel
Kënetë
Çinamak
Bardhoc
Skopje
Govrlevo
Vodovratski
Dimov Grob
Isar Marvinci
Plaosnik
Pista
Novo Selo
Tumba
Pazhok
Podgorie
Shtikë
Barç
Tren cave
Longas
Ohoden
Dzhulyunitsa
Smyadovo
Tell Kran
Boyanovo
Yabalkovo
Diamandovo
Stambolovo
Kapitan Andreevo
Haskovo
0
72
144
216 km
B
Podgorie N (n = 1)
Tren Cave N (n = 2)
Çinamak EBA (n = 1)
Shkrel MBA (n = 1)
Çinamak LBA-IA (n = 5)
Kënetë (n = 1)
Shtikë (n = 1)
Barç Post-Mdv (n = 1)
Barç Post-Mdv Roma (n = 4)
Bardhoc Post-Mdv (n = 4)
Pazhok Post-Mdv (n = 1)
6000 BCE
5000 BCE
4000 BCE
3000 BCE
2000 BCE
1000 BCE
0 CE
1000 CE


PCA 1: Neolithic EBA
Yamnaya Samara
Bulgaria Boyanovo EBA
Albania Çinamak EBA
Iron Gates Hunter Gatherer
Western Hunter Gatherer
Bulgaria Kapitan Andreevo EBA
Bulgaria Tell Ezero EBA
Serbia Mokrin EBA
Bulgaria Tell Kran EBA
Bulgaria Smyadovo EBA
Anatolia N
Greece N
Greece Manika EBA
Serbia ChL-EBA
Bulgaria NChL
Romania NChL
Croatia NChL
Albania N
Albania NChL
Montenegro N
0.04
0.00
-0.04
PC2 (0.455%)
0.00
0.04
0.08
PC1 (0.653%)
Hunter-Gatherer Cline
Pontic Steppe Cline
Ancient population (projected)
Albania Çinamak Early Bronze Age
Albania Neolithic
Albania Neolithic-Chalcolithic
Anatolia Neolithic
Bulgaria Boyanovo Early Bronze Age
Bulgaria Kapitan Andreevo Early Bronze Age
Bulgaria Neolithic-Chalcolithic
Bulgaria Tell Ezero Early Bronze Age
Bulgaria Tell Kran Early Bronze Age
Croatia Neolithic-Chalcolithic
Greece Manika Early Bronze Age
Greece Neolithic
Iron Gates Hunter Gatherer
Montenegro Neolithic
Romania Neolithic-Chalcolithic
Serbia Chalcolithic-Early Bronze Age
Serbia Mokrin Early Bronze Age
Smyadovo Early Bronze Age
Western Hunter Gatherer
Yamnaya Samara

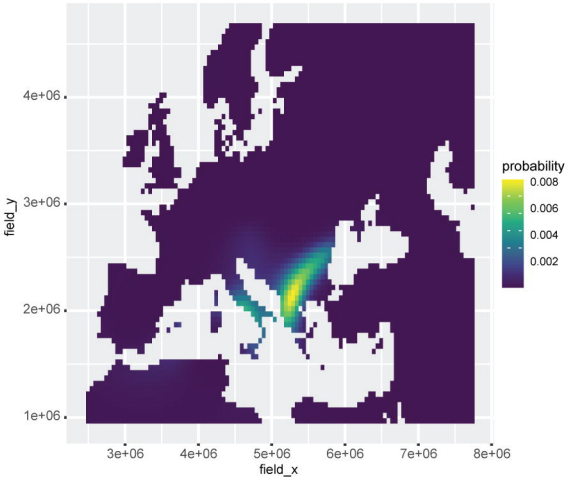
available under aCC BY-NC-ND 4.0 International license.



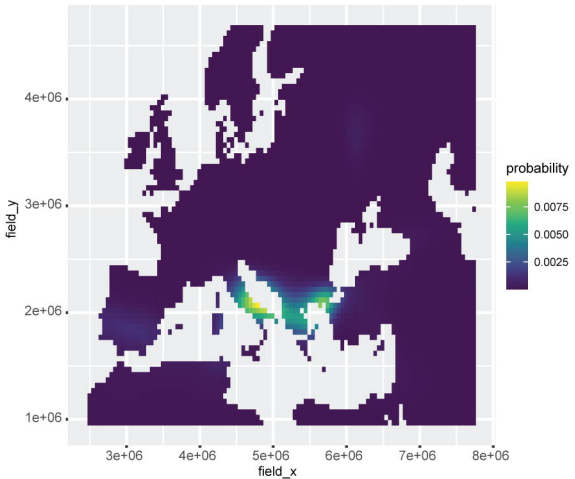

A
PCA 2: BA-IA-classical antiquity
PC2 (0.455%)
PC1 (0.653%)
0.02
0.00
-0.02
-0.04
-0.04
-0.02
0.00
0.02
Himera Balto-Slavs Anc.
East Anatolia Anc
East Anatolia BA
Macedonia Anc. outlier
Moldova Glinoe
Greece Longas MBA
Serbia IA
Croatia BA-IA
East France IA
Serbia BA
South France IA
Slovenia BA-IA
Montenegro BA-IA
Albania BA-IA
Macedonia BA-Anc.
Greek Himera Anc.
Bulgaria Anc.
Greece BA-IA
Bulgaria IA
Ancient population (projected)
Albania BA-IA
Albania MBA
Bulgaria Anc.
Bulgaria IA
Croatia BA-IA
Croatia Anc.
East Anatolia BA
East Anatolia Anc.
East France IIA
Greece BA-IA
Greece Longas MBA
Greek Himera Anc.
Himera Balkan outliers Anc.
Himera Balto-Slavs IA
Macedonia BA-Anc.
Macedonia Anc. outlier
Moldova Glinoe
Montenegro BA-IA
Serbia BA
Serbia IA
Slovenia BA-IA
South France IA
B
Greece Longas MBA
Montenegro BA-IA
Himera Balkan outliers
Croatia BA-IA
Albania BA-IA
Macedonia BA-Anc.
PC2 (0.455%)
PC1 (0.653%)
-0.005
-0.010
-0.015
-0.020
-0.025
-0.004
0.000
0.004


A
PCA 3: Roman-Medieval-Postmedieval
PC2 (0.455%)
PC1 (0.653%)
0.04
0.02
0.00
-0.02
-0.04
-0.050
-0.025
0.000
0.025
Albania PostMdv Roma Profile
Serbia Roman NE outliers
Montenegro Mdv
Himera Balto-Slavs Anc.
East Anatolia Roman Slav
Croatia Roman NE outliers
Bulgaria Mdv
Macedonia Mdv
Serbia PostMdv
Croatia Mdv-PostMdv
East France IA
Croatia BA-IA
Croatia Roman Anatolian outliers
Albania Modern
East Anatolia Anc.
East Anatolia Roman
Greece Roman
Serbia Roman Anatolian outliers
East Anatolia BA
Serbia BA
Croatia Roman Osijek
Macedonia BA-Anc.
South France IA
Macedonia Anc. outlier
Slovenia BA-IA
Albania BA-IA
Croatia Roman Šćitarjevo
Albania PostMdv
Albania Medieval
West Anatolia Roman
Greece BA-IA
Croatia Roman Zadar
Hungary Avar Period outliers
Bulgaria IA
Serbia Roman Viminacium East Balkan profile
Balto-Slavic Cline
Anatolian Cline
Ancient population (projected)
Albania Medieval
Albania Post-Medieval
Bulgaria Medieval
Croatia Medieval-Post-Medieval
Croatia Roman Anatolian outliers
Croatia Roman Gardun
Croatia Roman Novo Selo Bunje
Croatia Roman Omišalj
Croatia Roman Osijek
Croatia Roman Šćitarjevo
Croatia Roman Sipar
Croatia Roman Sisak
Croatia Roman NE outliers
Croatia Roman Trogir
Croatia Roman Velić
Croatia Roman Zadar
East Anatolia Roman
East Anatolia Roman Slav
Greece Roman
Hungary Avar Period outliers
Macedonia Medieval
Montenegro Medieval Slavic
Montenegro Doclea Roman
Albania Post-Medieval Roma Profile
Serbia Post-Medieval
Serbia Roman
Serbia Roman Anatolian outliers
Serbia Roman NE outliers
Serbia Roman Sviloš
Serbia Roman Viminacium
West Anatolia Byzantine Era
B
Balto-Slavic cline
Albania Modern
191
202
213
220
212
230
Pazhok (I15706)
Bardhoc (I14685)
Barç (I13834)
Bardhoc (I14686)
Bardhoc (I14687)
Kënetë (I14622)
Albania Mdv
Albania PostMdv
Shtikë (I13839)
Bardhoc (I15707)
Albania BA-IA
Anatolian cline
PC2 (0.455%)
PC1 (0.653%)
0.000
-0.005
-0.010
-0.015
-0.020
-0.025
-0.008
-0.006
-0.004
-0.002
0.000

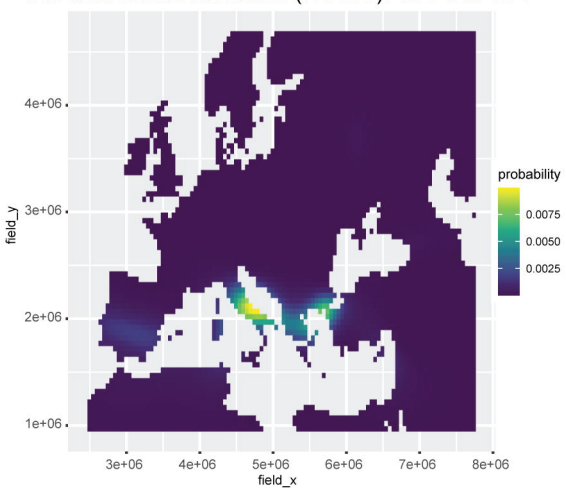
A Albania LIA-Hellenistic (I16251): 500-50 BCE



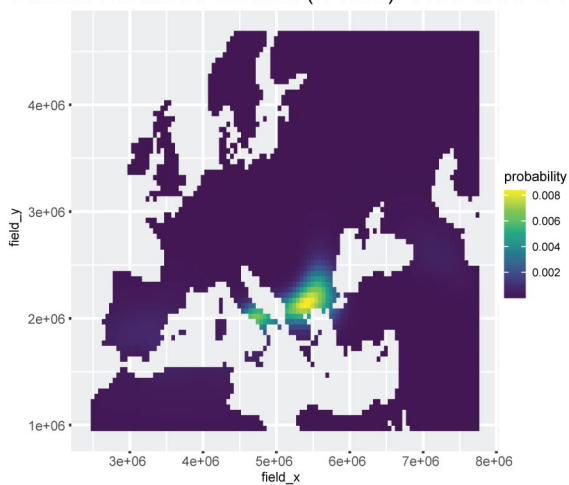
B Albania Kënetë Medieval (I13839): 773-885 CE



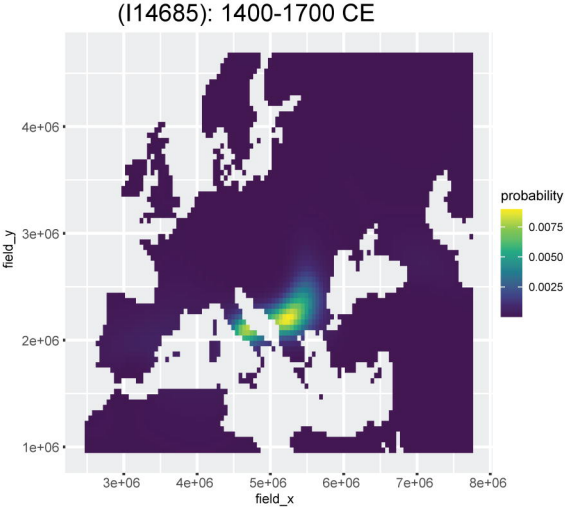
C Albania Shtikë Medieval (I13839): 889-989 CE



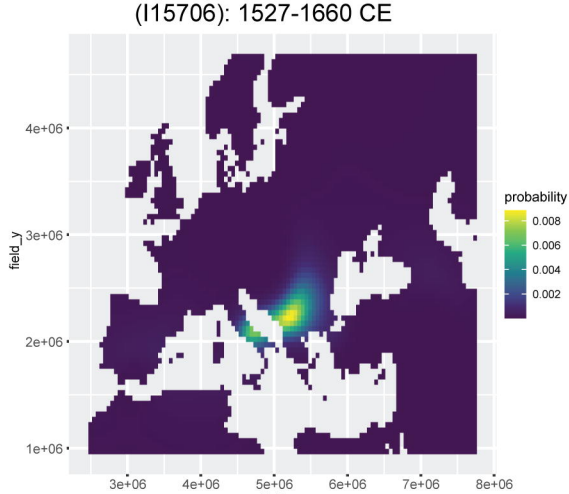
D Albania Bardhoc Post-Mdv (I14686): 1400-1700 CE



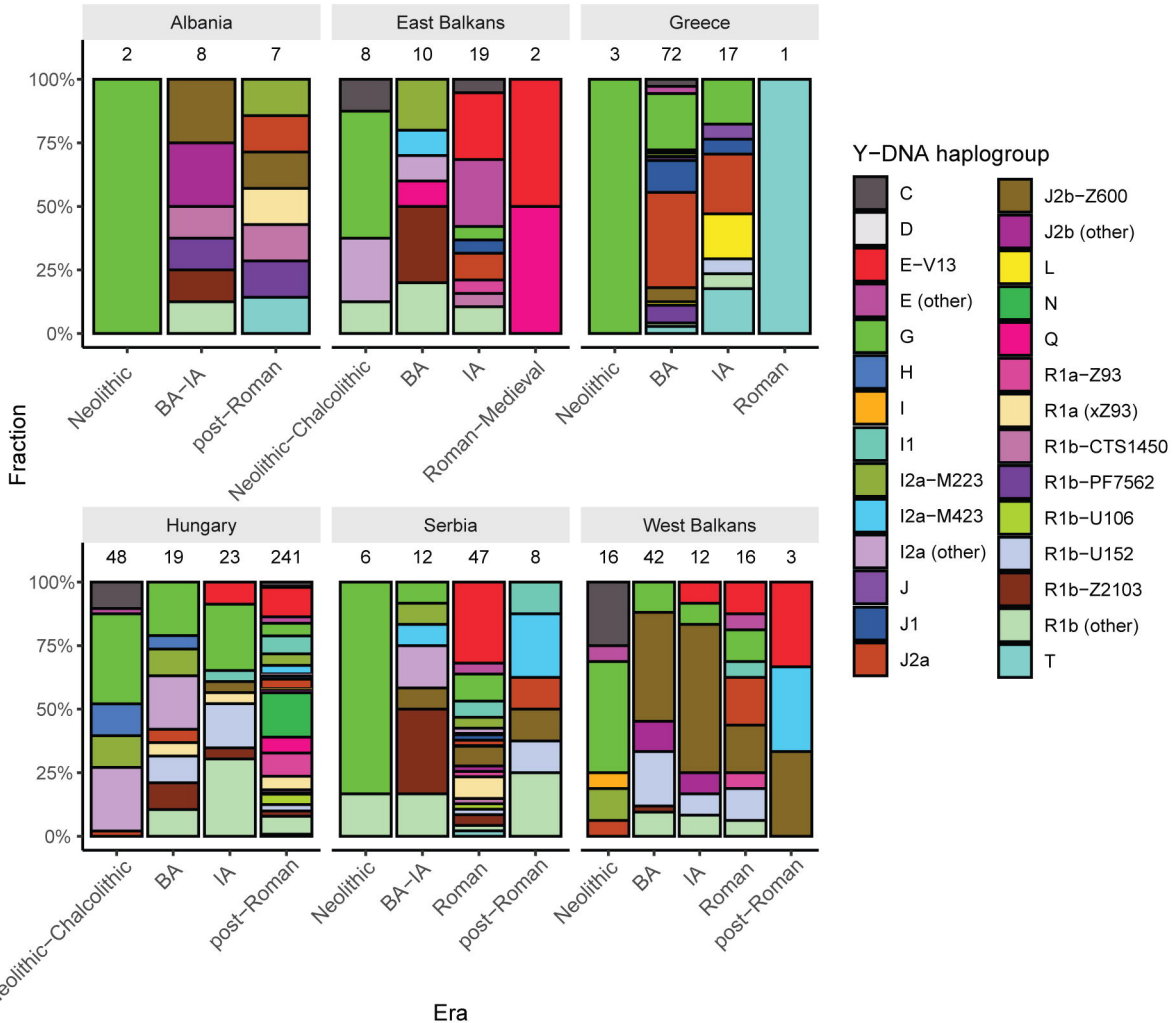
E Albania Bardhoc Post-Mdv modern profile (I14685): 1400-1700 CE

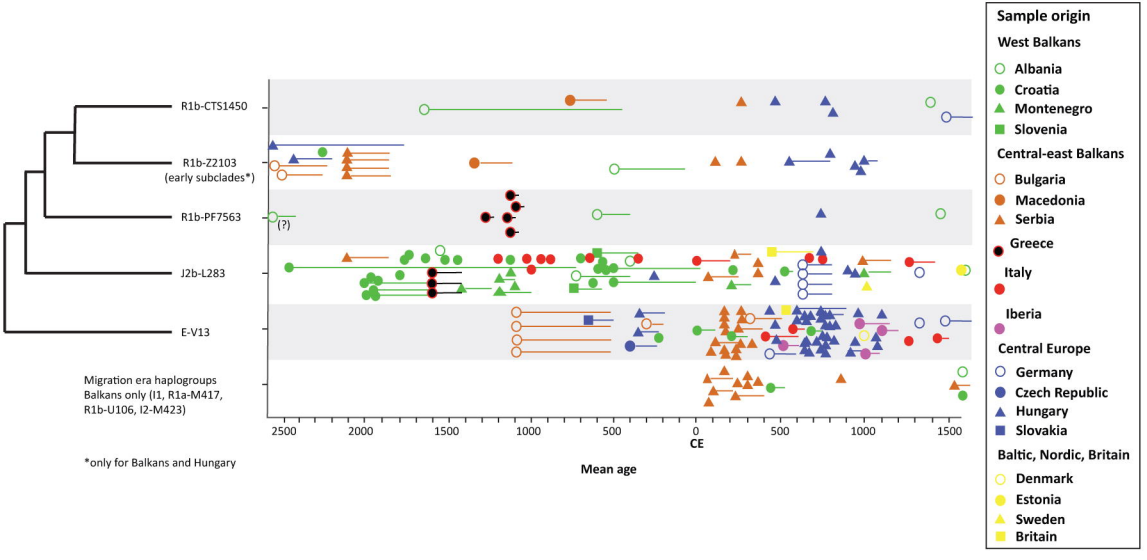


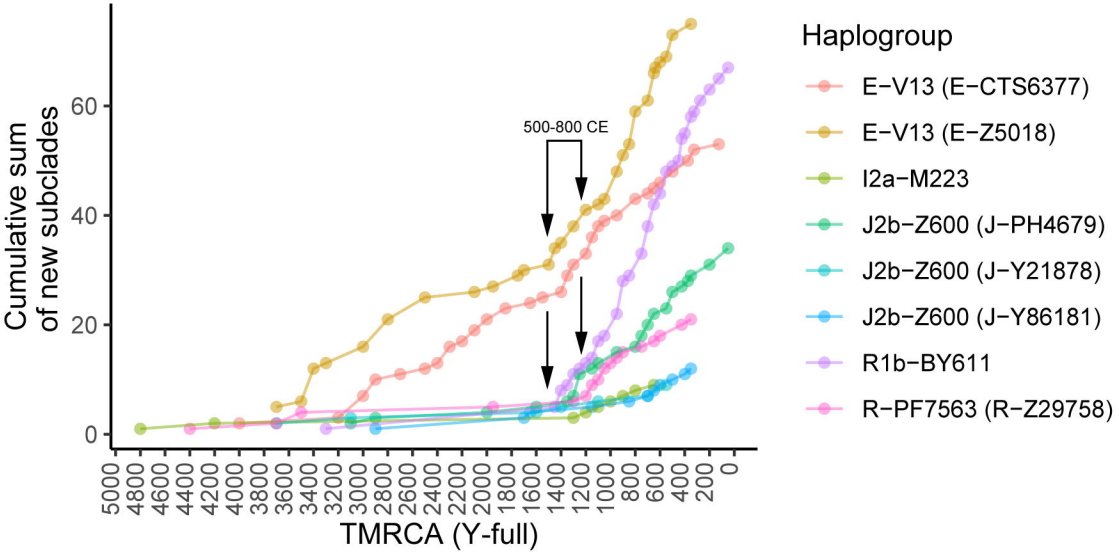
F Albania Pazhok Post-Mdv modern profile (I15706): 1527-1660 CE



Target	Source 1 (local)	Source 2 (Balto-Slavic-related)	*P* value	Coeff. 1	Coeff. 2	SE
Albania Post-Mdv Bardhoc	Albania Mdv		0.2	100%	-	-
Albania Post-Mdv Bardhoc	Albania Mdv	North Macedonia Mdv	0.16	90%	10%	0.14
Albania Post-Mdv Bardhoc	Albania Mdv	Serbia Roman NE o	0.18	95%	5%	0.04
Albania Post-Mdv Bardhoc	Albania Mdv	Russia IA Ingria	0.67	98%	2%	0.03
Albania Post-Mdv Bardhoc	Albania Mdv	Croatia Roman NE o	0.21	96%	4%	0.05
Albania Post-Mdv Bardhoc	Albania Mdv	Montenegro Doclea Mdv Slavic	0.31	93%	7%	0.07
Albania Post-Mdv Bardhoc	Albania Mdv	East Anatolia Roman Slav	0.17	94%	6%	0.06
Albania Post-Mdv Bardhoc o	Albania Mdv	Russia IA Ingria	0.19	87%	13%	0.05
Albania Post-Mdv Bardhoc o	Albania Mdv	Croatia Roman NE o	0.58	73%	27%	0.07
Albania Post-Mdv Bardhoc o	Albania Mdv	Montenegro Doclea Mdv Slavic	0.55	64%	36%	0.09
Albania Post-Mdv Bardhoc o	Albania Mdv	Serbia Roman NE o	0.55	75%	25%	0.06
Albania Post-Mdv Bardhoc o	Albania Mdv	East Anatolia Roman Slav	0.46	72%	28%	0.08
Albania Post-Mdv Pazhok o	Albania Mdv	Russia IA Ingria	0.21	86%	14%	0.04
Albania Post-Mdv Bazhok o	Albania Mdv	Croatia Roman NE o	0.09	74%	26%	0.07
Albania Post-Mdv Pazhok o	Albania Mdv	Montenegro Doclea Mdv Slavic	0.06	67%	33%	0.09
Albania Post-Mdv Pazhok o	Albania Mdv	East Anatolia Roman Slav	0.17	69%	31%	0.08
Albanian modern 191	Albania Mdv	Serbia Roman NE o	0.06	65%	35%	0.05
Albanian modern 191	Albania Mdv	Montenegro Doclea Mdv Slavic	0.48	54%	46%	0.08
Albanian modern 202	Albania Mdv	Russia IA Ingria	0.07	80%	20%	0.05
Albanian modern 202	Albania Mdv	Croatia Roman NE o	0.26	62%	38%	0.07
Albanian modern 212	Albania Mdv	Serbia Roman NE o	0.1	76%	24%	0.06
Albanian modern 212	Albania Mdv	Montenegro Doclea Mdv Slavic	0.12	70%	30%	0.09
Albanian modern 212	Albania Mdv	East Anatolia Roman Slav	0.6	73%	27%	0.09
Albanian modern 213	Albania Mdv	Montenegro Doclea Mdv Slavic	0.3	65%	35%	0.1
Albanian modern 220	Albania Mdv	Serbia Roman NE o	0.16	73%	27%	0.06
Albanian modern 220	Albania Mdv	Russia IA Ingria	0.06	82%	18%	0.05
Albanian modern 220	Albania Mdv	Croatia Roman NE o	0.18	70%	30%	0.08
Albanian modern 220	Albania Mdv	Montenegro Doclea Mdv Slavic	0.25	66%	34%	0.09
Albanian modern 230	Albania Mdv	Russia IA Ingria	0.18	72%	28%	0.05
Albanian modern 230	Albania Mdv	Serbia Roman NE o	0.21	66%	34%	0.06
Albanian modern 230	Albania Mdv	Croatia Roman NE o	0.16	63%	37%	0.08
Albanian modern 230	Albania Mdv	Montenegro Doclea Mdv Slavic	0.54	52%	48%	0.09
Albanian modern 230	Albania Mdv	East Anatolia Roman Slav	0.6	55%	45%	0.08


Albania
2
8
7
East Balkans
8
10
19
2
Greece
3
72
17
1
Hungary
48
19
23
241
Serbia
6
12
47
8
West Balkans
16
42
12
16
3
100%
75%
50%
25%
0%
Fraction
Neolithic
BA–IA
post–Roman
Neolithic–Chalcolithic
BA
IA
Roman–Medieval
Roman
Era
Y–DNA haplogroup
C
D
E–V13
E (other)
G
H
I
I1
I2a–M223
I2a–M423
I2a (other)
J
J1
J2a
J2b–Z600
J2b (other)
L
N
Q
R1a–Z93
R1a (xZ93)
R1b–CTS1450
R1b–PF7562
R1b–U106
R1b–U152
R1b–Z2103
R1b (other)
T


R1b-CTS1450
R1b-Z2103
(early subclades*)
R1b-PF7563
(?)
J2b-L283
E-V13
Migration era haplogroups
Balkans only (I1, R1a-M417,
R1b-U106, I2-M423)
*only for Balkans and Hungary
2500
2000
1500
1000
500
0
500
1000
1500
CE
Mean age
Sample origin
West Balkans
Albania
Croatia
Montenegro
Slovenia
Central-east Balkans
Bulgaria
Macedonia
Serbia
Greece
Italy
Iberia
Central Europe
Germany
Czech Republic
Hungary
Slovakia
Baltic, Nordic, Britain
Denmark
Estonia
Sweden
Britain


Cumulative sum of new subclades
0
20
40
60
500-800 CE
5000
4800
4600
4400
4200
4000
3800
3600
3400
3200
3000
2800
2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
TMRCA (Y-full)
Haplogroup
E−V13 (E−CTS6377)
E−V13 (E−Z5018)
I2a−M223
J2b−Z600 (J−PH4679)
J2b−Z600 (J−Y21878)
J2b−Z600 (J−Y86181)
R1b−BY611
R−PF7563 (R−Z29758)