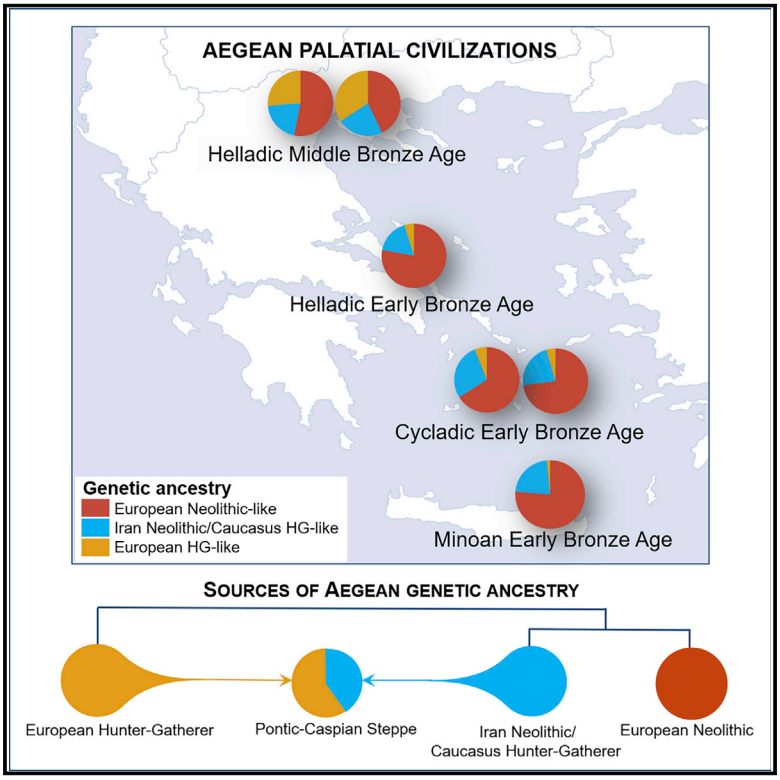
Article:
Clemente, F., Unterländer, M., Dolgova, O. et al. (2021) The genomic history of the Aegean palatial civilizations. Cell, 184 (10). 2565-2586.e21. ISSN: 0092-8674

https://doi.org/10.1016/j.cell.2021.03.039

Cell

The genomic history of the Aegean palatial civilizations

Graphical abstract



Highlights

- Bronze Age (BA) Helladic, Cycladic, and Minoan genomes from the Aegean were sequenced
- 3,000 BCE Aegeans are homogeneous and derive ancestry mainly from Neolithic farmers
- Neolithic Caucasus-like and BA Pontic-Caspian Steppe-like gene flow shaped the Aegean
- Present-day Greeks are genetically similar to 2,000 BCE Aegeans from Northern Greece

Authors

Florian Clemente, Martina Unterländer, Olga Dolgova, ..., Oscar Lao, Anna-Sapfo Malaspinas, Christina Papageorgopoulou

Correspondence

annasapfo.malaspinas@unil.ch (A.-S.M.), cpapage@he.duth.gr (C.P.)

In brief

The genomic analysis of ancient individuals from important archaeological sites across Aegean cultures suggests that the Aegean during the Bronze Age was at a genomic crossroads, and separate migration waves coincide with cultural shifts that had important impacts on Bronze Age cultures and the formation of the modern Greek population.

Clemente et al., 2021, Cell *184*, 2565–2586
May 13, 2021
https://doi.org/10.1016/j.cell.2021.03.039

CellPress

Cell

CellPress
OPEN ACCESS

Article

The genomic history of the Aegean palatial civilizations

Florian Clemente,[1,2,21] Martina Unterländer,[3,4,20,21] Olga Dolgova,[5,21] Carlos Eduardo G. Amorim,[1,2] Francisco Coroado-Santos,[6] Samuel Neuenschwander,[1,2,7] Elissavet Ganiatsou,[3] Diana I. Cruz Dávalos,[1,2] Lucas Anchieri,[1,2] Frédéric Michaud,[1,2] Laura Winkelbach,[4] Jens Blöcher,[4] Yami Ommar Arizmendi Cárdenas,[1,2] Bárbara Sousa da Mota,[1,2] Eleni Kalliga,[3] Angelos Souleles,[3] Ioannis Kontopoulos,[8] Georgia Karamitrou-Mentessidi,[9] Olga Philaniotou,[9] Adamantios Sampson,[10] Dimitra Theodorou,[11] Metaxia Tsipopoulou,[9] Ioannis Akamatis,[12] Paul Halstead,[13] Kostas Kotsakis,[12] Dushka Urem-Kotsou,[14] Diamantis Panagiotopoulos,[15] Christina Ziota,[16] Sevasti Triantaphyllou,[12] Olivier Delaneau,[1,2] Jeffrey D. Jensen,[17] J. Víctor Moreno-Mayar,[1,2,8,18] Joachim Burger,[4] Vitor C. Sousa,[6] Oscar Lao,[5,19] Anna-Sapfo Malaspinas,[1,2,22,23,*] and Christina Papageorgopoulou[3,22,*]

[1]Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[3]Laboratory of Physical Anthropology, Department of History and Ethnology, Democritus University of Thrace, 69100 Komotini, Greece
[4]Palaeogenetics Group, Institute of Organismic and Molecular Evolution, Johannes Gutenberg University of Mainz, 55099 Mainz, Germany
[5]CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4, 08028 Barcelona, Spain
[6]CE3C, Centre for Ecology, Evolution and Environmental Changes, Faculty of Sciences of the University of Lisbon, 1749-016 Lisbon, Portugal
[7]Vital-IT, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
[8]Center for GeoGenetics, GLOBE Institute, University of Copenhagen, 1350 Copenhagen, Denmark
[9]Ephor Emerita of Antiquities, Hellenic Ministry of Culture and Sports, 10682 Athens, Greece
[10]Department of Mediterranean Studies, University of the Aegean, 85132 Rhodes, Greece
[11]Ephorate of Antiquities of Kozani, Hellenic Ministry of Culture and Sports, 50004 Kozani, Greece
[12]Department of History and Archaeology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[13]Department of Archaeology, University of Sheffield, Minalloy House, 10-16 Regent St., Sheffield S1 3NJ, UK
[14]Department of History and Ethnology, Democritus University of Thrace, 69100 Komotini, Greece
[15]Institute of Classical Archaeology, University of Heidelberg, Marstallhof 4, 69117 Heidelberg, Germany
[16]Ephorate of Antiquities of Florina, Hellenic Ministry of Culture and Sports, 53100 Florina, Greece
[17]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA
[18]National Institute of Genomic Medicine (INMEGEN), 14610 Mexico City, Mexico
[19]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[20]Present address: Federal Criminal Police Office, 65173 Wiesbaden, Germany
[21]These authors contributed equally
[22]These authors contributed equally
[23]Lead contact
*Correspondence: annasapfo.malaspinas@unil.ch (A.-S.M.), cpapage@he.duth.gr (C.P.)
https://doi.org/10.1016/j.cell.2021.03.039

SUMMARY

The Cycladic, the Minoan, and the Helladic (Mycenaean) cultures define the Bronze Age (BA) of Greece. Urbanism, complex social structures, craft and agricultural specialization, and the earliest forms of writing characterize this iconic period. We sequenced six Early to Middle BA whole genomes, along with 11 mitochondrial genomes, sampled from the three BA cultures of the Aegean Sea. The Early BA (EBA) genomes are homogeneous and derive most of their ancestry from Neolithic Aegeans, contrary to earlier hypotheses that the Neolithic-EBA cultural transition was due to massive population turnover. EBA Aegeans were shaped by relatively small-scale migration from East of the Aegean, as evidenced by the Caucasus-related ancestry also detected in Anatolians. In contrast, Middle BA (MBA) individuals of northern Greece differ from EBA populations in showing ~50% Pontic-Caspian Steppe-related ancestry, dated at ca. 2,600-2,000 BCE. Such gene flow events during the MBA contributed toward shaping present-day Greek genomes.

INTRODUCTION

The Bronze Age (BA) period in Eurasia was marked by pivotal changes on the social, political, and economic levels, visible in the appearance of the first large urban centers and monumental palaces (Harding, 2000). The Aegean Sea—an embayment of the Mediterranean surrounded by mainland Greece, western Anatolia, and Crete (Figure 1A)—has played an important role in the formation of these innovations, particularly because some of the first

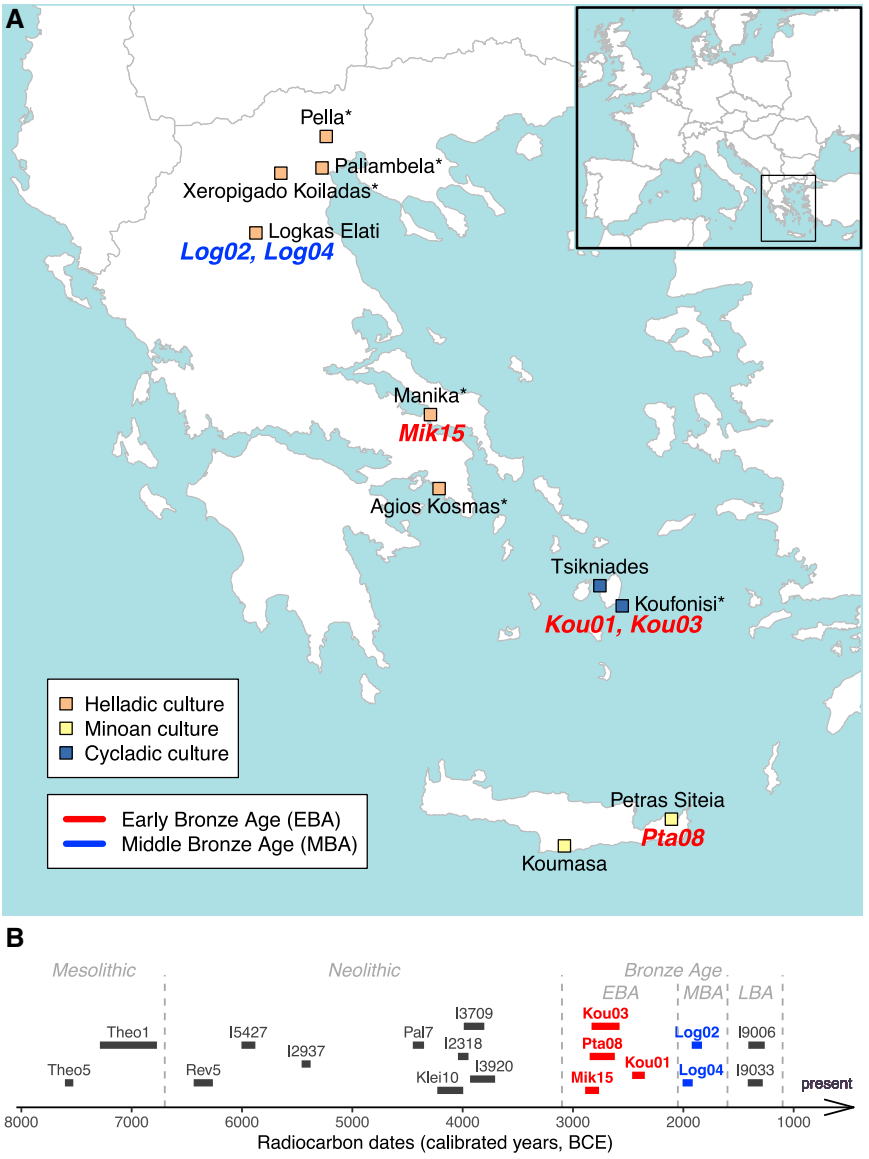


Figure 1. Geographical location of archeological sites and radiocarbon dates

(A) Archeological sites are indicated by square symbols, colored according to their corresponding Aegean cultural group. Asterisks indicate archeological sites for which mitochondrial DNA (mtDNA) capture data were generated (STAR Methods). The whole genomes of six individual samples (*Pta08*, *Kou01*, *Kou03*, *Mik15*, *Log02*, and *Log04*) sequenced are colored according to their radiocarbon dates.

(B) Radiocarbon dates of the six whole-genome sequenced individuals from this study together with 12 individuals from present-day Greece from previous studies (Lazaridis et al., 2017, Mathieson et al., 2018, Hofmanová et al., 2016). For the two Mesolithic individuals, only mtDNA data is available (Hofmanová et al., 2016). The bar indicates the range of the Cal 1-sigma OxCal calibrated date for each individual (STAR Methods).

See also Figures S1 and S3, Table 2, and Document S1.

monumental urban centers were formed around its shores (Renfrew, 2011).

The BA civilizations in the Aegean, often termed *Aegean Cultures*, include the Minoan civilization in Crete (3,200/3,000–1,100 BCE) (Wilson, 2008), the Helladic civilization in mainland Greece (3,200/3,000–1,100 BCE) (Wright, 2008) including the Mycenaean (i.e., the last phase of Helladic [1,600–1,100 BCE]), the Cycladic civilization in the Cycladic islands in the middle of the Aegean Sea (3,200/3,000–1,100 BCE) (Broodbank, 2008), and the western Anatolian cultures (3,000–1,200 BCE) (Şahoğlu, 2008; Yakar, 2012). These cultures exhibit distinct characteristics in pottery style, burial customs, architecture, and art (Cline, 2012; Shelmerdine, 2008). However, they share common innovations related to craft and agricultural (e.g., wine and oil) specializations, the creation of large storage facilities and redistribution systems as well as palaces, intensive trade, and the extensive use of metals. The increasing economic and cultural exchange that developed in the BA Aegean laid the groundwork for modern economic systems—including capitalism, long-distance political treaties, and a world trade economy (Kristiansen, 2016). In late BA (LBA), the earliest forms of writing appear—Linear A Minoan and Linear B Mycenaean scripts. Although Linear A has yet to be deciphered, Linear B (1,450 BCE) is the earliest attested form of Greek (Ventris and Chadwick, 1953a, 1953b)—one of the living languages with the longest documented history within the Indo-European family. These novelties define the early forms of urbanization, traditionally described as the *urban revolution* and the *emergence of civilization* (Childe, 1942, 1950; Renfrew, 1972), and constitute significant milestones in European history (Cline, 2012; Renfrew, 2011).

Based on extensive archaeological data, several hypotheses on the origin and development of these cultures have been proposed, including: (1) local innovation, where changes were based on genetic and cultural continuity of local Neolithic groups (Dickinson, 2016; Oakley and Renfrew, 1972; Tsountas and Manatt, 1897); (2) the immigration of new populations from Anatolia and the Caucasus during the Early BA (EBA) and the Middle BA (MBA) (Blegen and Haley, 1928; Caskey, 1971; Wace, 1957); and (3) the arrival of possible speakers of Indo-European languages from the Pontic-Caspian Steppe at the beginning of the EBA (Coleman, 2000; for review, see Pullen, 2008; Dickinson, 2016) (Document S1). In central, northern, and western Europe, most BA genomes are a mixture of local farmers, themselves descendants of Aegean Neolithic populations (Hofmanová et al.,

Table 1. Labels for key populations discussed in the text

European HG (EUHG)	Hunter-gatherers of the twelfth to the sixth millennium BCE from western, central and northern Europe (present-day Spain, Luxembourg, Hungary (*WHG*), Sweden (*SHG*), Switzerland *(Bichon)*, and Italy *(Villabruna))* (Lazaridis et al. 2014; Jones et al. 2015; Mathieson et al. 2015; Fu et al. 2016).
Eastern HG (EHG)	Hunter-gatherers of the sixth millennium BCE from eastern Europe, more specifically present-day northwestern and southwestern Russia *(EHG)* (Haak et al. 2015; Mathieson et al. 2015).
Iran Neolithic/ Caucasus HG	Neolithic (N) farmers of the eleventh to the first millennium BCE of present-day Iran and HG of the Caucasus region. These populations cluster in genomic analyses from previous studies (*Iran_HotuIIIb, Iran_N, Iran_LN, Iran_ChL, Iran_IA, CHG)* (Gamba et al. 2014; Jones et al. 2015; Broushaki et al. 2016; Lazaridis et al. 2016).
European Neolithic	European farmers of the sixth to the third millennium BCE in central, northern and western Europe *(Europe_MNChL, Europe_EN)* (Gamba et al. 2014; Allentoft et al. 2015; Günther et al. 2015; Haak et al. 2015; Mathieson et al. 2015; Olalde et al. 2015).
Aegean and Anatolian Neolithic	Neolithic farmers of the ninth to fourth millennium BCE from central Anatolia (*Anatolia_Boncuklu, Anatolia_Tepecik_Ciftlik*) as well as the Aegean (*i.e.*, northern and southern Greece (*Greece_N*) and northwestern Anatolia *(Anatolia_N)* (Mathieson et al. 2015; Hofmanová et al., 2016; Kilinç et al. 2016; Lazaridis et al. 2016)).
Balkans Neolithic	Neolithic and Chalcolithic farmers of the seventh to fourth millennium BCE from present-day Bulgaria, Croatia, Romania, Serbia, and North Macedonia (*Balkans_Neolithic, Balkans_Chalcolithic*) (Mathieson et al. 2018).
Anatolian ChL and BA	Neolithic farmers of the fifth to second millennium BCE from northwestern Anatolia (*Anatolia_Kumtepe, Anatolia_ChL, Anatolia_BA*) (Lazaridis et al. 2016; 2017; Omrak et al. 2016).
Armenia ChL_BA	Late Neolithic - Chalcolithic - farmers and Caucasus BA populations of the fourth to the first millennium BCE from present-day Armenia (*Armenia_ChL, Armenia_EBA, Armenia_MLBA*) (Allentoft et al. 2015; Lazaridis et al. 2016).
Levant/Natufian	Semi-sedentary Epipaleolithic populations, Neolithic farmers and BA populations of the twelfth to the second millennium BCE from the Levant (*Levant_N, Natufian, Levant_BA*) (Lazaridis et al. 2016).
Steppe	Bronze Age populations of the fifth to the second millennium BCE from the Pontic-Caspian Steppe region (*Steppe_EMBA, Steppe_MLBA*) (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015).
Balkans EBA	Bronze Age populations of the fourth to second millennium BCE from present-day Bulgaria and Croatia *(Balkans_EBA)* (Mathieson et al. 2018).
Balkans LBA	Bronze Age populations of the second to first millennium BCE from present-day Bulgaria *(Balkans_LBA)* (Mathieson et al. 2018).
Aegean EBA	Populations of the third millennium BCE in the Aegean (EBA) sequenced in this study: two Early Cycladic individuals from the island of Ano Koufonisi (*Kou01*, *Kou03*), one Early Helladic individual from Manika in Euboea (*Mik15*) and one Early Minoan Individual from Kephala Petras in Crete (*Pta08*).
Aegean MBA	Populations of the second millennium BCE in northern Greece (MBA) from the site of Elati-Logkas (*Log02*, *Log04*, this study).
Helladic-Manika-EBA	Early Bronze Age Helladic individuals from Manika on the island of Euboea of the third Millennium BCE (*Mik15*, this study).
Cycladic-Koufounisi-EBA	Early Bronze Age Cycladic individuals from the island of Ano Koufonisi of the third millennium BCE (*Kou01*, *Kou03*, this study).
Minoan-Petras-EBA	Early Bronze Age Minoan individuals from Kephala Petras on the island of Crete of the third millennium BCE (*Pta08*, this study).

Chronological abbreviations used throughout the text: *N-Neolithic, EN-Early Neolithic, LN-Late Neolithic, MNChL-Middle Neolithic to Chalcolithic, ChL-Chalcolithic, LNBA-Late Neolithic to Bronze Age, BA-Bronze Age, EBA-Early Bronze Age, EMBA-Early to Middle Bronze Age, MBA-Middle Bronze Age, MLBA-Middle to Late Bronze Age, IA-Iron Age.*

2016), and local hunter-gatherers (HG) (Table 1) (Allentoft et al., 2015; Lazaridis et al., 2014; Mathieson et al., 2018). Ancient DNA data have unveiled massive population movements from the East, bringing in a Caucasus HG component together with an Eastern HG component in similar proportions (de Barros Damgaard et al., 2018; Jones et al., 2015). These components may be attributed to a migration wave of Pontic-Caspian Steppe populations during the late Neolithic and EBA (~2,800 BCE) (Allentoft et al., 2015; Antonio et al., 2019; Haak et al., 2015; Olalde et al., 2019). Recently, Steppe-related ancestry was reported during the BA in the northern Balkans in Bulgaria (Mathieson et al., 2015), on the Balearic Islands, and in Sicily (Fernandes

et al., 2020), but not in Sardinia (Marcus et al., 2020). However, it remains unclear how much further this ancestry extends either temporally or geographically into southeastern Europe.

Despite their importance for understanding the rise of western civilization and the spread of Indo-European languages, no BA whole genomes from the Aegean have been sequenced to date. Hence, the genetic origins of the peoples behind the Neolithic-BA transition and their contribution to the present-day Greek population remain controversial (Coleman, 2000; Hughey et al., 2013; Lazaridis et al., 2017). Neolithic whole genomes from present-day Greece and western Anatolia are almost indistinguishable, supporting a common Aegean Neolithic population spreading across the Aegean Sea (Hofmanová et al., 2016). Caucasus HG-related ancestry is present in some of the late Neolithic Aegean individuals, Chalcolithic Anatolians (Kılınç et al., 2017; Lazaridis et al., 2017; Omrak et al., 2016), LBA Mycenaeans, and Early to Middle BA (EMBA) Minoans (Lazaridis et al., 2017), raising the possibility of gene flow from the East. LBA Mycenaeans also show evidence for an ancestry attributable to gene flow from the Pontic-Caspian Steppe, or from Armenia (Lazaridis et al., 2017). Finally, present-day Greeks were found to be quite genetically distinct from these previously reported Minoans and Mycenaeans, although the source of this difference was not investigated.

The dearth of genomic data from the Neolithic to the BA transition period in the Aegean has left key questions partially unanswered for understanding particular aspects of the BA demographic process in Europe, which we address:

(1) Were the Aegeans who triggered the BA transition related to Neolithic groups from the same area?
(2) What was the genetic affinity among the Helladic, Cycladic and Minoan EBA civilizations (i.e., did their cultural differences entail population structure, and how did they relate to LBA populations such as the Mycenaeans)?
(3) Did the Eastern (Caucasus or Iran) ancestry observed in some Neolithic and Chalcolithic Anatolians persist until the EBA in the Aegean? What was the timing of such gene flow?
(4) Did the massive migration from the Pontic-Caspian Steppe into central Europe have an influence on the Aegean BA populations? If so, what was the timing and magnitude of this gene flow?
(5) How are Aegean individuals across the BA related to present-day Greeks who inhabit the same area?

To answer these questions and to characterize the populations who were behind the sophisticated palaces and urban centers of the Aegean BA, we generated whole genomes from BA Aegeans, including four from the EBA and two representing the Cycladic culture (Figure 1). We used existing tools for phenotypic prediction on nuclear capture data and applied standard population genomic methods to characterize the relationship among ancient and present-day populations. To infer the demography of the Aegean from the Neolithic to the present-day, we capitalized on whole genome data and utilized approximate Bayesian computation (Tavaré et al., 1997) coupled with deep learning (ABC-DL) (Mondal et al., 2019), which we have extended to account for the typical low depth of coverage, damage, and modern human contamination characterizing ancient genomes.

RESULTS AND DISCUSSION

Dataset

Individual samples and radiocarbon dates

We screened 70 individual samples for the presence of human DNA. Six individual samples with a human DNA content higher than 1% were selected for whole genome sequencing (WGS) (Table S1; STAR Methods). For three of those, nuclear capture data was also generated. For these six individuals, we used sample material from the petrous bone for both radiocarbon dating and DNA sequencing (Figure 1; Table 2; STAR Methods): one EBA Helladic individual from the site of Manika on the island of Euboea (*Mik15*), one EBA Minoan from the site of Kephala Petras (the burial rock shelter) on the island of Crete (*Pta08*), two EBA Cycladic individuals from the island of Koufonisi (*Kou01* and *Kou03*), and two MBA individuals from the site of Elati-Logkas in northern Greece (*Log02* and *Log04*) (Figures 1A and S1). To improve clarity and to emphasize the archaeological site, culture, and time period of the sequenced individuals, we will refer to these individuals as: Helladic-Manika-EBA (*Mik15*), Minoan-Petras-EBA (*Pta08*), Cycladic-Koufonisi-EBA (*Kou01* and *Kou03*), and Helladic-Logkas-MBA (*Log02* and *Log04*). Moreover, four individuals (*Mik15*, *Pta08*, *Kou01*, and *Kou03*) will be jointly referred to as EBA Aegeans, distinguishing them from the more recent (by ~1,000 years) individuals, *Log02* and *Log04*, who will be referred to as MBA Aegeans. Similar labels are utilized to group published genomic data from reference individuals (Table 1). Moreover, to assist reproducibility, the labels of previously published populations are italicized throughout the text. In addition to generating data for the abovementioned six individuals, we captured the mtDNA genome for 11 individual samples (Figure 1; Table S1; STAR Methods).

Six ancient Aegean whole genomes

The resulting depth of coverage for the Aegean BA genomes ranged between 2.6× and 4.9× (average: 3.7×) (Tables 2 and S1; and STAR Methods). The number of SNPs covered by at least one read is considerably higher for the six Aegean BA genomes than for the Aegean BA SNP capture data from Lazaridis et al. (2017) when considering the "1240K" SNP capture set (Mathieson et al., 2015) but also across an "intergenic region" SNP set defined in this study (Figure S2A). The latter includes ~5,270,000 SNP sites located at least 20 kb away from annotated genes and CpG islands (STAR Methods). Note that whole genomes from the populations studied hereafter have more low frequency variants in the intergenic regions than were detected in the 1240K SNP set for the same regions (Figure S2B). This likely owes to the SNP ascertainment scheme in the latter (Clark et al., 2005).

We observed typical ancient DNA damage patterns at the 5′ and 3′ termini of the DNA fragments, as well as short sequence reads (average length between 49.9 and 74.3 bases across genomes, after adaptor removal and mapping), attesting to the authenticity of the ancient data (Figure S3; STAR Methods). Across individuals, contamination rate estimates ranged between 0.6% and 1.1%, and between 0.01% and 1.49% when

Table 2. Genomic and archaeological data for the six BA whole-genome sequenced individuals from this study

Archaeological site	Sample ID	Time period	Culture[a]	Age (cal BCE)	Shotgun DoC (all)[b]	Shotgun DoC (5 bp trim)[b]	Capture DoC[c]	Contam. mtDNA (%)[d]	Contam. X (%)[e]	Sex	mtDNA haplogroup	Y haplogroup
Manika	Mik15	EBA	Early Helladic	2890–2764	3.5	2.2	–	0.01–0.58	–	XX	J2b1	–
Petras	Pta08	EBA	Early Minoan	2849–2621	4.0	3.0	38.0	0.01–0.52	1.0–1.1	XY	H	G2-L156
Koufonisi	Kou01	EBA	Early Cycladic	2464–2349	2.6	2.0	42.9	0.02–0.97	0.6–0.8	XY	K1a2c	J2a-M410
	Kou03	EBA	Early Cycladic	2832–2578	2.8	2.2	–	0.18–1.49	–	XX	K1a	–
Logkas	Log02	MBA	Middle Helladic	1924–1831	4.3[f]	3.4	109.1	0.02–0.39	–	XX	H55a	–
	Log04	MBA	Middle Helladic	2007–1915	4.9	4.0	-	0.02–0.94	–	XX	J1c+16261	–

Details on the origin of the individuals, their cultural group, and their radiocarbon dates. ID: identifier; BCE: Before the Common Era; DoC: Depth of Coverage; 5 bp trim: 5 bp trimming; mtDNA: mitochondrial DNA; Contam.: Contaminaton;EBA: Early Bronze Age; MBA: Middle Bronze Age. See also Figures 1, S2, and S3, Tables 1 and S1, and STAR Methods.

[a]Terminology of cultural groups.

[b]Depth of Coverage (DoC) before ("all") and after trimming 5 bp from the extremities of the reads ("5 bp trim").

[c]Average number of reads covering the nuclear capture regions.

[d]95% credible interval (STAR Methods).

[e]95% confidence interval assuming a European population (HapMap CEU) as contaminant (Star Methods).

[f]USER-treated sample (STAR Methods).

estimated using the X chromosome and mtDNA, respectively (Table 2).

Population structure and demographic history

Genomic homogeneity across the Aegean during the EBA despite distinct cultural backgrounds

The overall genome-wide genetic relationship of the Aegean BA individuals was studied in the context of ancient and present-day Eurasian populations (STAR Methods). Despite their distinct cultures, the EBA Helladic, Cycladic, and Minoan genomes resemble one another in all analyses. Outgroup f_3-statistics of the form f_3(*Yoruba;Y,X*), where *X* is one of the present-day populations included in Dataset I (STAR Methods) and *Y* the ancient individuals from this study, show that the Helladic-Manika-EBA, Minoan-Petras-EBA, and Cycladic-Koufonisi-EBA have a similar profile, which contrasts with the Helladic-Logkas MBA. EBA Aegeans have higher genetic similarity with present-day southern Europeans, particularly present-day Sardinians (Figure S4). In the classical multidimensional scaling (*MDS*) analysis, the projected genetic dissimilarities between pairs of individuals estimated by an identity-by-state distance matrix in two dimensions (STAR Methods) show that the four EBA individuals (*Mik15*, *Pta08*, *Kou01*, *Kou03*) and the two MBA individuals (*Log02* and *Log04*) form two groups (Figure 2) in agreement with the f_3 profiles. In line with the results above, ancestry proportions estimated by *ADMIXTURE* for $K > 2$ using Dataset II (Table S2; STAR Methods) suggest that the EBA Aegeans are genetically similar to one another and distinct from the MBA Aegeans (Figures 3 and S5).

Compared to other ancient Eurasian populations, the EBA Aegeans are similar to other Aegean BA and Anatolian populations, but are quite distinct from all Balkan populations. For instance, in the *MDS* analysis, they fall within or near Minoan-Lasithi-MBA, Mycenaean-Peloponnese-LBA, and Anatolian populations such as *Anatolia_Tepecik_Ciftlik* (Figure 2). Similarly, in the *ADMIXTURE* analysis, the EBA Aegeans show similar ancestry proportions to other Aegean populations, such as the Minoan-EMBA and *Anatolia_Kumtepe*, as well as Anatolian populations spanning the Chalcolithic and the EMBA (e.g., *Anatolia_ChL*, *Anatolia_BA*) (Figure 3).

The genomic EBA homogeneity across cultures in the Aegean and parts of Anatolia may indicate that Aegean populations used the sea as a route to interact not only culturally but also genetically. This could have been the result of an intense network of communication in the Aegean, which has been well documented on the archaeological level and has been dubbed the "International Spirit of the Aegean" (Renfrew, 1972). Moreover, given the high similarity between Minoan-Petras-EBA and the Cycladic-Koufonisi-EBA, the genomic data also informs debates related to the formation of colonies from the Cycladic islands to Crete (Doumas, 2010; Papadatos, 2007).

Ancestry components of EBA Aegeans indicating gene flow from a source related to the CHG during the Neolithic

ADMIXTURE results indicate that the EBA Aegean population consists mostly of an ancestry component shared with Neolithic Aegeans (accounting for >65%), whereas most of the remaining ancestry can be assigned to Iran Neolithic/Caucasus HG-related populations (17%–27%) (Figure 3). These results were replicated with *qpWave*/*qpAdm* (STAR Methods) using Dataset I (Tables 3 and S3). When considering early Neolithic populations and HG populations as potential

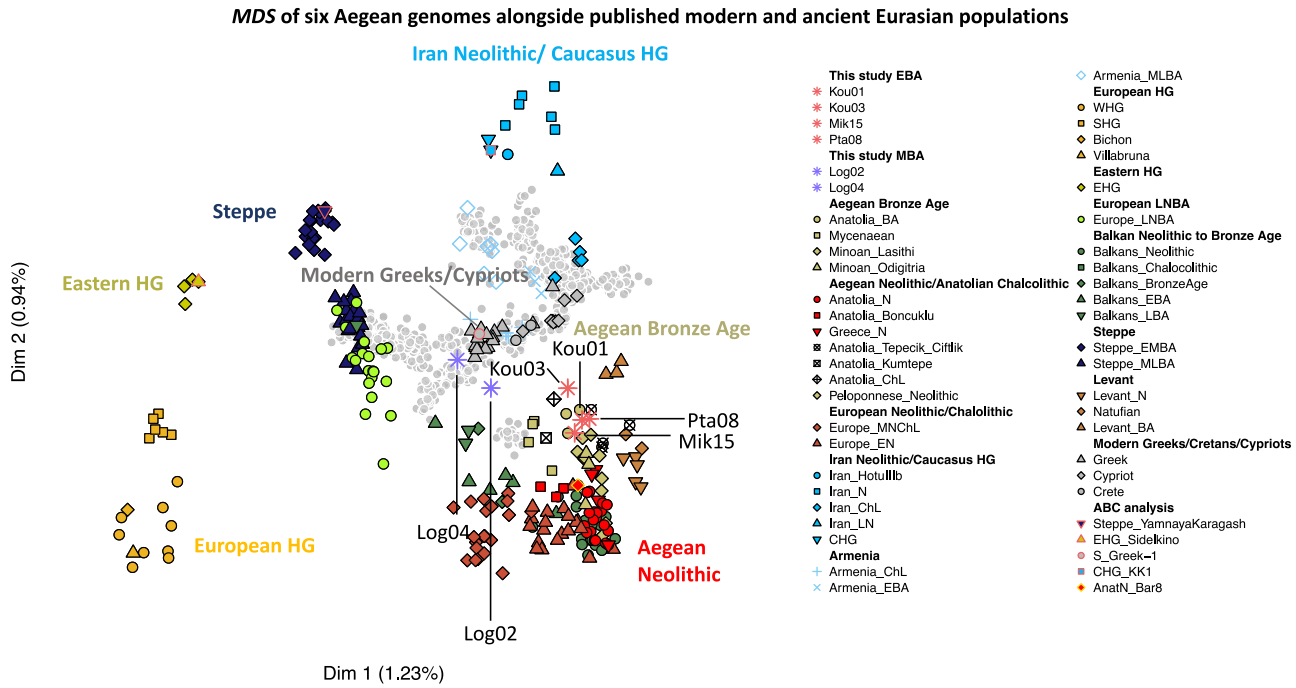


Figure 2. Multidimensional scaling analysis (*MDS*)

Included are the six Aegean Bronze Age individual samples from the present study, 259 ancient genomes, and 638 modern individuals (gray shapes) of Eurasian ancestry (Table S2; STAR Methods). Population labels are given in Table 1.
See also Table S5 and Document S1.

sources, EBA individuals were in general found to be consistent with the majority of their ancestry deriving from populations related to *Anatolia_N* (~69%–84%) (Table 3). This suggests that the people behind the Neolithic to BA transition largely had ancestors from the preceding Aegean farmers, in line with archaeological theories for the EBA transformation (Dickinson, 2016; Renfrew, 1972; Tsountas and Manatt, 1897) (Document S1). The second component in *qpWave*/*qpAdm* could be assigned to *Iran_N*/*CHG*-related populations (~16%–31%) (Table 3). In line with this result, in the *MDS* analysis (Figure 2), the Aegean EBA individuals are on an axis connecting Neolithic Aegeans to the Iran Neolithic/Caucasus HG (''Caucasus-axis'').

To further test for gene flow events from outside of the Aegean, *D-statistics* were computed. In particular, we tested whether an *H3* population (e.g., *Iran_N* or *CHG*—the blue component in *ADMIXTURE*) (Figure 3) shares more alleles with H1 = *Anatolia_N* ($D > 0$) or with Aegean/Anatolian populations from different time periods (H2 = *Greece_N*, BA Aegeans/Anatolians, present-day Greeks and Cypriots) ($D < 0$), using the ancient Ethiopian *Mota* (Gallego Llorente et al., 2015) as an outgroup *D*(*Anatolia_N*, *H2*; *H3*, *Mota*) (Figure S6). The EBA Aegean genomes were found to be similar to one another. Although EBA Aegeans carry the ''Iran Neolithic/Caucasus HG-like'' component in other analyses (e.g., Figure 3), no statistically significant evidence for gene flow from *Iran_N* or *CHG* was detected. However, a visible trend suggests that Aegeans dating to ~4,000 BCE onward (from *Anatolia_ChL* to *Mycenaean*) share more alleles with *Iran_N*/*CHG* than with *Anatolia_N* (Figure S6). This trend is replicated in the *ADMIXTURE* results (Figure 3), where small proportions of *CHG*-like components were observed from the Neolithic onward in individuals on both sides of the Aegean and in Anatolia but not in the Balkans. This *CHG*-like component increases in frequency during the early Neolithic in Anatolia (e.g., Boncuklu, Tepecik-Ciftlik) (Kılınç et al., 2016), the late Neolithic in the Aegean (e.g., *Greece_N*) (Hofmanová et al., 2016; Omrak et al., 2016), and during the BA in Anatolia (*Anatolia_BA*) (Lazaridis et al., 2017). This is not seen in the Balkans, where the transition from Neolithic to BA is mostly associated with an increase in ''European HG-like'' ancestry (Figure 3).

To compare competing scenarios, and to infer the mode and tempo of potential gene flow events into the Aegean while accounting jointly for the population history of Neolithic, BA, and present-day populations from Greece, we performed ABC-DL (Mondal et al., 2019) (Document S1; STAR Methods). To determine the relationship between HG and Aegean Neolithic, we first contrasted 3-leaf models (Figure 4A; Table S4) of the three ancestral populations: CHG, EHG, and Aegean Neolithic. In this analysis, the 3-leaf model (EHG, CHG, and Aegean Neolithic) had the greatest posterior probability ($P(M|D) = 0.999$). This result is in agreement with Jones et al. (2015), who found a closer relationship between *CHG* and ''Early Farmer'' from Stuttgart, than with *WHG*. We used this tree for the more complex 7-leaf models (Figure 4B; Table S4). In line with all of the above results, 7-leaf models without a *CHG*-like pulse of gene flow (models B1, B2, and B3) (Figure 4B) were associated with lower posterior probabilities

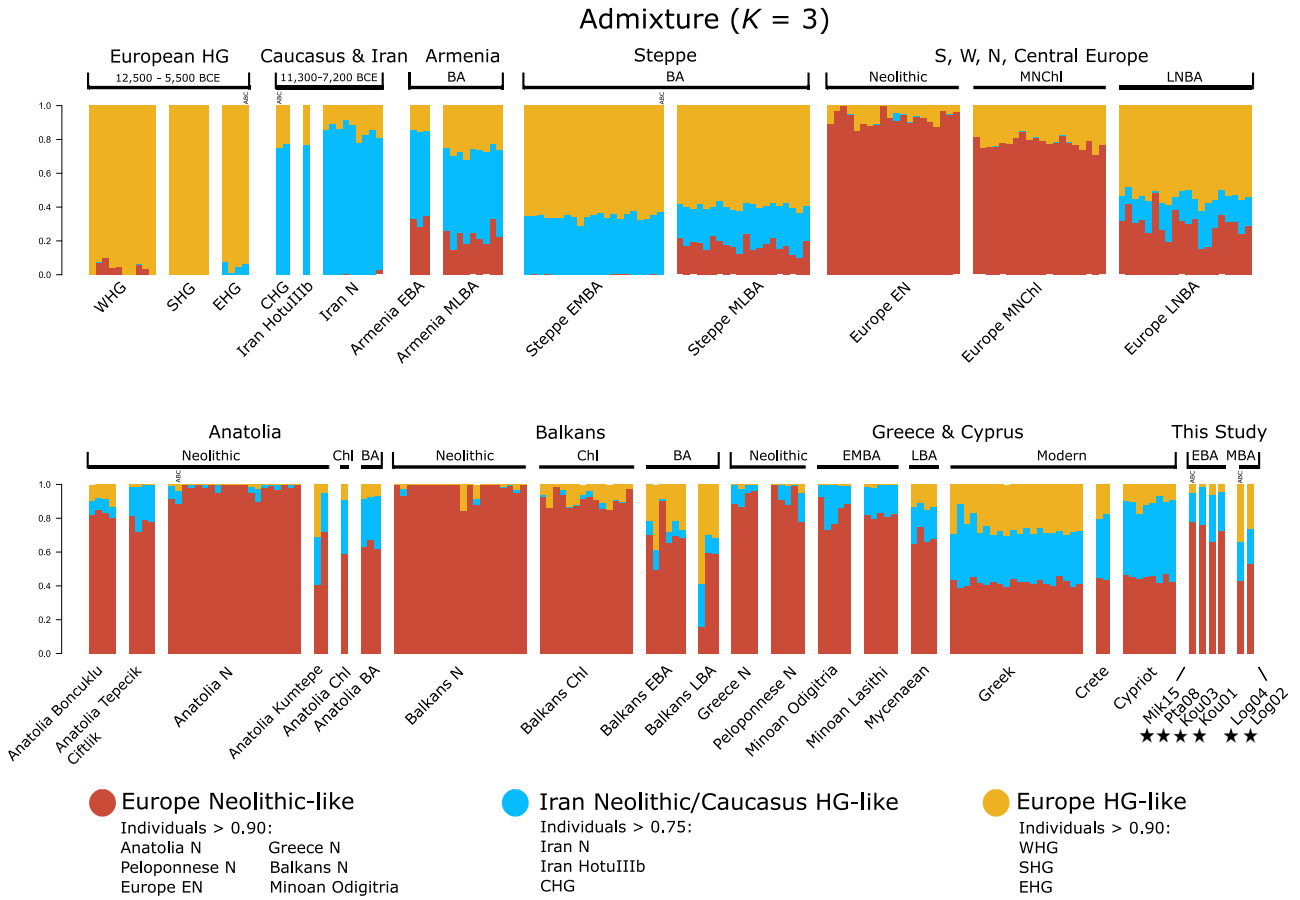


Figure 3. ADMIXTURE analysis for modern and ancient Eurasian individuals

Shown are a subset of individuals for $K = 3$, which has the lowest cross validation error (CV = 0.974). Results for the full dataset and statistical support are shown in Figure S5. The six BA individuals whole genome sequenced in this study are highlighted with an asterisk. Abbreviations for chronological periods and population names are given in Table 1.

See also Table S2, Document S1, and STAR Methods.

(0.000–0.014). In contrast, a model including such gene flow, estimated at 16% (0.2%–29%, 95% highest posterior density interval) (Table S4) at ~5,700 BCE (8,299–2,881 BCE, 95% highest posterior density interval) (Table S4) was assigned much higher support (posterior probability of 0.98 for model B4) (Document S1; STAR Methods). Taken together, these results suggest that a population related to the Caucasus HG had either directly influenced the Aegean through migration, or a *CHG*-like component was indirectly introduced through exchanges with Neolithic Anatolian populations.

Unlike for most European populations, little EHG contribution is seen during the EBA

In central, western, and northern BA Europe, the *CHG* component is generally accompanied by an *EHG* component (Allentoft et al., 2015; Haak et al., 2015; Jones et al., 2015) – which would be expected to appear in similar proportions if transmitted through Steppe-related populations (de Barros Damgaard et al., 2018). In contrast, EBA Aegeans carry little to no *EHG* ancestry. Based on the *D-statistics* analysis, we cannot reject that most EBA Aegean genomes and *Anatolia_N* are equally close to *EHG* (Figure S6). Moreover, when considering three potential sources in *qpWave/qpAdm*, EBA individuals carry only ~1%–8% *EHG* versus 24%–25% *CHG* ancestry (i.e., substantially less *EHG* ancestry) (Table 3). This is further supported by *ADMIXTURE* results (Figure 3), indicating that changes from Neolithic to EBA were mostly associated with increases in *IranN/CHG*-like ancestry in the Aegean and Anatolia, whereas the Balkans and the rest of Europe were mostly associated with increases in *EHG*-like ancestry (Figure 3). Finally, all ABC-DL models including an *EHG* pulse into the ancestor of EBA Aegeans (models B5 and B6) (Figure 4) have negligible posterior probability support (0.000–0.002). Taken together, these results suggest little influence of populations related to *EHG* during the EBA in the Aegean, further implying that the Caucasus component arrived in the Aegean independently.

Genomic heterogeneity during the Aegean MBA, likely owing to gene flow from a Steppe-like population prior to 2,000 BCE

Considerably more population structure is observed in the Aegean during the MBA compared to the EBA. MBA individuals

Table 3. *qpWave/qpAdm* admixture models

Period	Test	Ref1	Ref2	Ref3	Mixture Prop. Ref1 ± SE	Mixture Prop. Ref2 ± SE	Mixture Prop. Ref3 ± SE	p value
EBA	*Kou01*	*Anatolia_N*	*CHG*		0.75 ± 0.03	0.25 ± 0.03		0.67
	Kou01	*Anatolia_N*	*Iran_N*		0.75 ± 0.03	0.25 ± 0.03		0.90
	Kou03	*Anatolia_N*	*CHG*		0.69 ± 0.03	0.31 ± 0.03		0.10
	Mik15	*Anatolia_N*	*CHG*		0.84 ± 0.03	0.16 ± 0.03		0.08
	Mik15	*Anatolia_N*	*Iran_N*		0.84 ± 0.03	0.16 ± 0.03		0.07
	Pta08	*Mik15*	*Iran_N*		0.98 ± 0.03	0.02 ± 0.03		0.09
	Pta08	*Mik15*	*CHG*		0.99 ± 0.03	0.01 ± 0.01		0.07
	Kou01	*Anatolia_N*	*CHG*	*EHG*	0.74 ± 0.04	0.25 ± 0.03	0.01 ± 0.02	0.67
	Kou01	*Anatolia_N*	*Iran_N*	*EHG*	0.74 ± 0.03	0.24 ± 0.03	0.02 ± 0.02	0.88
	Kou03	*Anatolia_N*	*Iran_N*	*EHG*	0.67 ± 0.03	0.25 ± 0.03	0.08 ± 0.02	0.82
MBA	*Log02*	*Kou01*	*EHG*		0.81 ± 0.02	0.19 ± 0.02		0.07
	Log02	*Kou03*	*WHG*		0.91 ± 0.02	0.09 ± 0.02		0.06
	Log02	*Anatolia_N*	*Balkans_LBA*	*CHG*	0.22 ± 0.05	0.65 ± 0.06	0.12 ± 0.04	0.08
	*Log02**	*Kou01*	*Steppe_MLBA*		0.61 ± 0.03	0.39 ± 0.03		0.20
	*Log02**	*Kou01*	*Europe_LNBA*		0.56 ± 0.04	0.44 ± 0.04		0.05
	Log04	*Kou01*	*Balkans_LBA*		0.21 ± 0.06	0.79 ± 0.06		0.08
	Log04	*Kou03*	*Balkans_LBA*		0.26 ± 0.07	0.74 ± 0.07		0.07
	Log04	*Mik15*	*Balkans_LBA*		0.21 ± 0.06	0.79 ± 0.06		0.06
	Log04	*Anatolia_N*	*CHG*	*EHG*	0.58 ± 0.03	0.16 ± 0.03	0.27 ± 0.02	0.12
	*Log04**	*Anatolia_N*	*Steppe_EMBA*		0.53 ± 0.03	0.47 ± 0.03		0.35
	*Log04**	*Anatolia_N*	*Steppe_MLBA*		0.38 ± 0.03	0.62 ± 0.03		0.13
	*Log04**	*Pta08*	*Balkans_LBA*		0.15 ± 0.04	0.85 ± 0.04		0.06
	*Log04**	*Pta08*	*Steppe_MLBA*		0.44 ± 0.03	0.56 ± 0.03		0.36
LBA	*Mycenaean*	*Log04*	*Minoan_Lasithi*		0.36 ± 0.04	0.64 ± 0.04		0.35
	Mycenaean	*Log04*	*Minoan_Odigitria*		0.21 ± 0.04	0.79 ± 0.04		0.45
	Mycenaean	*Anatolia_N*	*Kou03*		0.37 ± 0.09	0.63 ± 0.09		0.40
Present-day	*Crete*	*Log02*	*Iran_N*		0.82 ± 0.04	0.18 ± 0.04		0.08
	Cypriot	*Pta08*	*CHG*	*Villabruna*	0.64 ± 0.02	0.32 ± 0.02	0.04 ± 0.01	0.31
	Cypriot	*Pta08*	*CHG*	*WHG*	0.65 ± 0.02	0.32 ± 0.02	0.03 ± 0.01	0.23
	Greek	*Log02*	*EHG*		0.93 ± 0.02	0.07 ± 0.02		0.16
	Greek	*Log02*	*MA1*		0.96 ± 0.02	0.04 ± 0.02		0.10
	Greek	*Log02*	*Kostenki14*		0.93 ± 0.02	0.07 ± 0.02		0.07

For a test population, the estimated admixture proportions (±1 standard error, SE) for n = 2 or n = 3 source populations (Ref1, Ref2, and Ref3) are shown. Ancestry was inferred from both "ultimate" sources representing the earliest populations, and "proximate" sources (row labeled with a * symbol) representing populations down to the Bronze Age (STAR Methods). Only a subset of the results with p values $\geq$0.05 are depicted. See also Tables 1 and S3.

from northern Greece are quite distinct from the EBA Aegeans, as can be seen across all analyses. For example, in the f_3 analysis, unlike the EBA Aegeans, they are equally distant to a much larger set of populations across Europe (Figure S4). In *MDS* (Figure 2) and *ADMIXTURE* (Figure 3) analyses, they form a separate group distinct from the *EBA* Aegeans, sharing the same components as the present-day Greeks. In contrast, the Minoan-Lasithi-MBA are very similar to the EBA Aegean populations (Figures 2 and 3).

The primary feature distinguishing the Helladic-Logkas-MBA from the contemporary Minoan-Lasithi-MBA, as well as from the EBA populations, is the higher proportion of "European HG-like" ancestry. For instance, in *ADMIXTURE*, the "European HG-like" component accounts for 26%–34% of the overall Logkas ancestry, more than four times greater than the 2%–6% found in the Aegean EBA individuals (Figure 3). Similarly, in *qpWave/qpAdm*, a Helladic-Logkas-MBA individual (*Log04*) was consistent with a 3-way admixture model, deriving ~58% of her ancestry from Aegean Neolithic populations; the remaining ancestry can be attributed to *CHG*-like and *EHG*-like sources (accounting for ~16% and ~27%, respectively)—that is, having a much greater contribution from *EHG* as compared to the EBA Aegeans (Table 3). Because *EHG* and *CHG* are the major components of Steppe-related populations (e.g., *Steppe_EMBA* with 66% *EHG*-like and 34% *IranN/CHG*-like0 (Figure 3), consistent with previous results (de Barros Damgaard et al., 2018), this

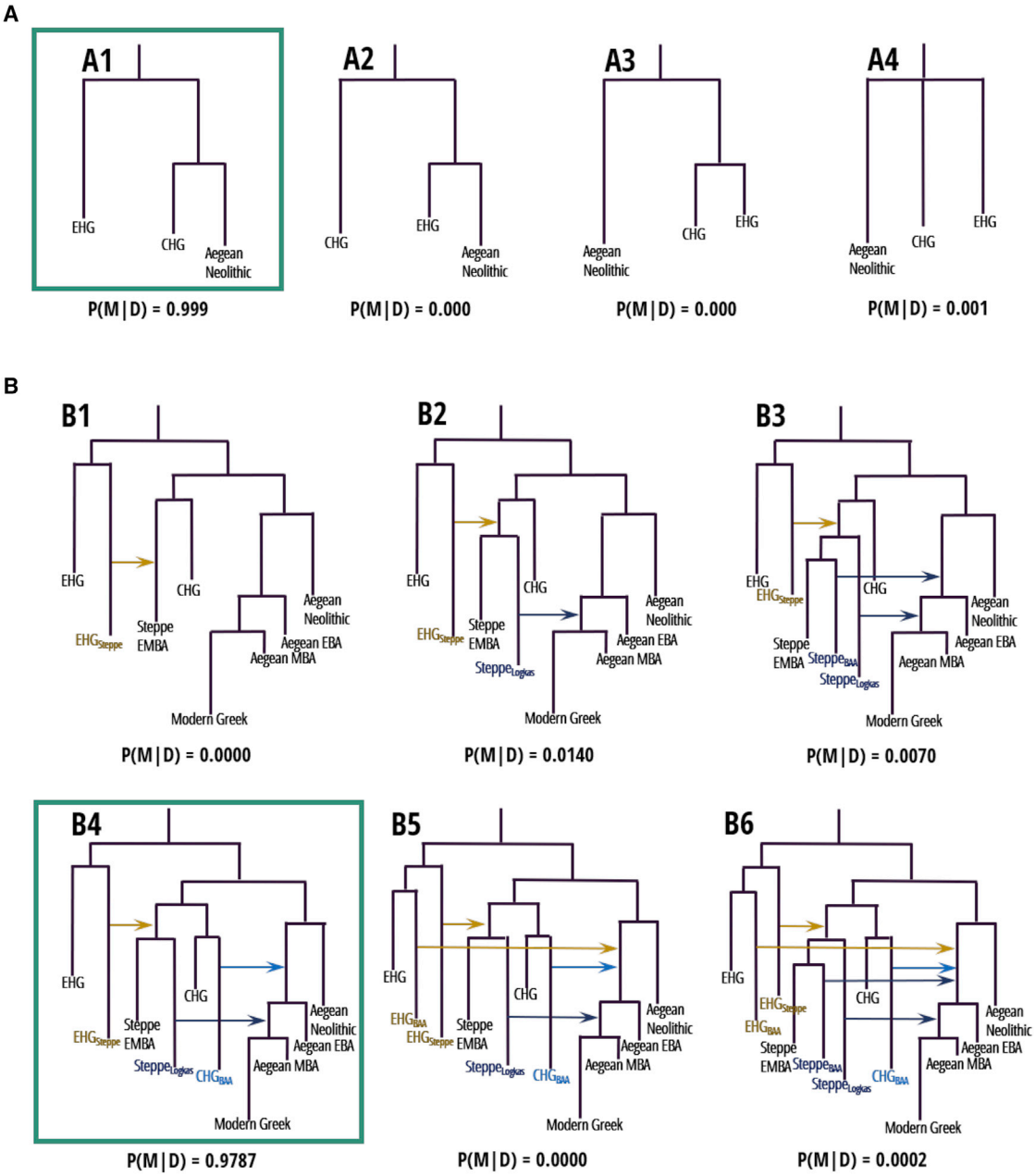


Figure 4. Model comparison in ABC-DL analysis

(A) Posterior probabilities P(M|D) of different 3-leaf models (models A1–A4) calculated with ABC-DL to establish the topology of the three ancestral populations: *CHG*, *EHG* and Aegean Neolithic.

(B) Seven-leaf demographic models (models B1–B6) extending the tree from (A) with the highest posterior probability, each including Neolithic, EBA and MBA Aegeans, present-day Greeks, *CHG*, *EHG*, and Pontic-Caspian *Steppe_EMBA* populations. Yellow (*EHG*-like), light blue (CHG-like), and dark blue (Steppe-like) arrows indicate a single pulse of gene flow from simulated "ghost" populations diverged from *EHG*, *CHG*, and *Steppe_EMBA*. Posterior probabilities are listed below each schematic topology (Document S1; STAR Methods).

See also Figure S2 and Tables 1 and S4.

supports the hypothesis that populations from the Pontic-Caspian Steppe contributed to the ancestry of the Helladic-Logkas-MBA individuals. This combined ancestry has been observed in central, western, and northern BA Europeans and interpreted as the result of a "massive" Steppe migration (Allentoft et al., 2015; Haak et al., 2015; Jones et al., 2015; Mathieson et al., 2018; Olalde et al., 2019). Our *ADMIXTURE* estimates are consistent with an increase of *EHG* components in the Late Neolithic and EBA in most regions of Europe, including in the Balkans (Figure 3; Document S1). Yet, in Anatolia, such an increase in *EHG*-like

ancestry is residual, and in the Aegean, it is only seen later in the MBA (Helladic-Logkas-MBA) and LBA (Mycenaean) individuals, suggesting a later arrival of Steppe-related ancestry in the Aegean.

Evidence for such a Steppe contribution is provided, for example, in *MDS* (Figure 2), where the Helladic-Logkas-MBA fell on a "Steppe-axis" connecting Neolithic Aegeans with Steppe populations. In *ADMIXTURE* (Figure 3), the Helladic-Logkas-MBA carries similar relative amounts of the "Iran Neolithic/Caucasus HG-like" (~1/3) and "European HG-like" (~2/3) components as *Steppe_EMBA*. Moreover, unlike the Neolithic and the EBA Aegeans and Anatolians, as well as the Minoan-Lasithi-MBA, the Helladic-Logkas-MBA share significantly more alleles with *CHG, EHG* and *Steppe_EMBA* compared to *Anatolia_N* (Figure S6). In addition, the Helladic-Logkas-MBA *Log04* individual could also be directly modeled as 2-way admixture (proximate sources) of *Anatolia_N* (~53%) and *Steppe_EMBA* (~47%), or *Anatolia_N* (~38%), and *Steppe_MLBA* (~62%), consistent with a strong genetic contribution from the Steppe (Table 3). Furthermore, demographic modeling suggests that gene flow (8%–45%, 95% highest posterior density interval) (Table S4) from a ghost population related to *Steppe_EMBA*, prior to the MBA split, considerably improves the fit of the model to the data (model B4 versus model B1) (Figure 4B). The timing of such gene flow into the ancestors of the Helladic-Logkas-MBA ought to have occurred by ~1,900 BCE, based on the radiocarbon dates of the Logkas individuals, and was estimated at ~2,300 BCE (2,616–2,003 BCE 95% highest posterior density interval) (Table S4) in the ABC-DL analysis. This suggests that a Steppe-like migration wave may have reached the Aegean by the MBA. Because Steppe-related ancestry is essentially absent in Sardinia (Fernandes et al., 2020; Marcus et al., 2020), and because we have no evidence of Steppe-like or EHG-like ancestry among Minoans, this may suggest that Steppe-related populations did not cross the sea during the BA. Supporting this hypothesis, the archaeological record does not indicate that BA populations from the Pontic-Caspian Steppe were sea-faring people (Anthony, 2010).

Note, however, that the Steppe-like ancestry observed in the Logkas individuals may have been brought directly by migrating populations originally from the Pontic-Caspian Steppe or indirectly by populations with substantial Steppe-like gene flow (e.g., *Balkans_LBA* or *Europe_LNBA*) (Table 3). Alternatively, the Steppe-like component may have been brought by an unsampled, genetically similar, population (e.g., MBA Balkans). The indirect contribution is supported by *ADMIXTURE* estimates that suggest an earlier influence of Steppe-related ancestry in the Balkans than in the Aegean (Figure 3), and by *qpWave/qpAdm* modeling of the MBA *Log04* individual as 2-way admixture involving Balkan LBA (Table 3). This finding is consistent with the suggestion of intermittent genetic contact between the Balkans and the Steppe populations during the BA (Mathieson et al., 2018) and is in line with archaeological evidence of cultural contacts between southeastern Europe and the Pontic-Caspian Steppe around 2,500 BCE (Anthony, 2010). This may further be related to previous hypotheses based on both archaeological and linguistic evidence that populations with Steppe-like ancestry contributed to the formation of the Helladic culture (Coleman, 2000) (Document S1).

Assessing sex-biased gene flow and inbreeding during the EBA and MBA

To assess sex-biased gene flow among the BA Aegeans, mtDNA, Y-, and X-chromosomes were analyzed. The 17 inferred mtDNA (Tables 2 and S2) and the two Y-chromosome haplogroups (Table 2) are common among European Neolithic individuals and do not show any clear evidence of sex-biased gene flow from outside of the Aegean (Document S1). To further investigate sex-biased gene flow, we compared the ancestry on the X chromosome versus the autosomes with a supervised *ADMIXTURE* following Goldberg et al. (2017). We found no evidence for sex-biased gene flow in EBA Aegeans, with point estimates of *Iran_N/CHG*-like ancestry on the X chromosome overlapping with those of autosomes (Figure 5A). In contrast, among MBA Aegeans, although *Log04* has similar amounts of Steppe-like ancestry on the X chromosome and the autosomes, *Log02* is inferred to harbor no Steppe-like ancestry on the X chromosome versus 25%–52% Steppe-like ancestry on the autosomes (Figure 5B). Moreover, in the mtDNA, we found no significant (STAR Methods) population structure (AMOVA p value = 0.293) between EBA and MBA Aegeans from the North of Greece (Pella, Paliambela, Xeropigado Koiladas, and Elati-Logkas) (Figure 1; Document S1). Together, these patterns on the X chromosome and mtDNA could be explained by male-biased gene flow from Steppe-like ancestry into the Aegean. Similarly, Goldberg et al. (2017) and Olalde et al. (2019) suggested that the immigration of Pontic-Caspian Steppe populations during the Late Neolithic/EBA in Europe may have involved a much larger number of males than females.

To gain further genetic clues about marital practices during the EBA and the MBA, we inferred contiguous genomic regions in homozygous states—also called runs of homozygosity (ROH)—in four present-day Greek and the six BA Aegean whole genomes (Figure S7). *Log04* had more (twenty-nine versus seven at most) and longer ROH (two ROH above 5 Mb) (Document S1) than other ancient individuals. Different evolutionary/demographic processes (Ceballos et al., 2018; Pemberton et al., 2012), including recent inbreeding (Yengo et al., 2019), could explain the *Log04* data; in any case, *Log02* does not harbor similarly long ROH, suggesting that the underlying cause may not generally characterize the Helladic-Logkas-MBA (Document S1).

LBA Myceneans: Armenia versus Steppe-like gene flow

The last phase of the BA is associated with a Late Helladic culture termed Mycenaean. Around 1,200 BCE, the Mycenaean civilization began to decline, the palaces were destroyed, the system of writing (Linear B) was abandoned, and their arts and crafts ceased. The causes of their decline are disputed (e.g., climatic change, invasions) (Middleton, 2020). Lazaridis et al. (2017) showed that Mycenaeans were quite distinct from present-day populations, but it remained unclear how they relate to EBA populations.

Despite cultural similarity with the Helladic-Logkas-MBA individuals, analyses suggest that the Mycenaean-Peloponnese-LBA were quite distinct genetically, occupying a position in-between the Logkas and the EBA Aegean and the Minoan-Lasithi-MBA in *MDS* (Figure 2). Unlike the Logkas individuals, they carry a lower European-HG-like component in *ADMIXTURE* (Figure 3) and do not share significantly more alleles with *Iran_N/CHG* or

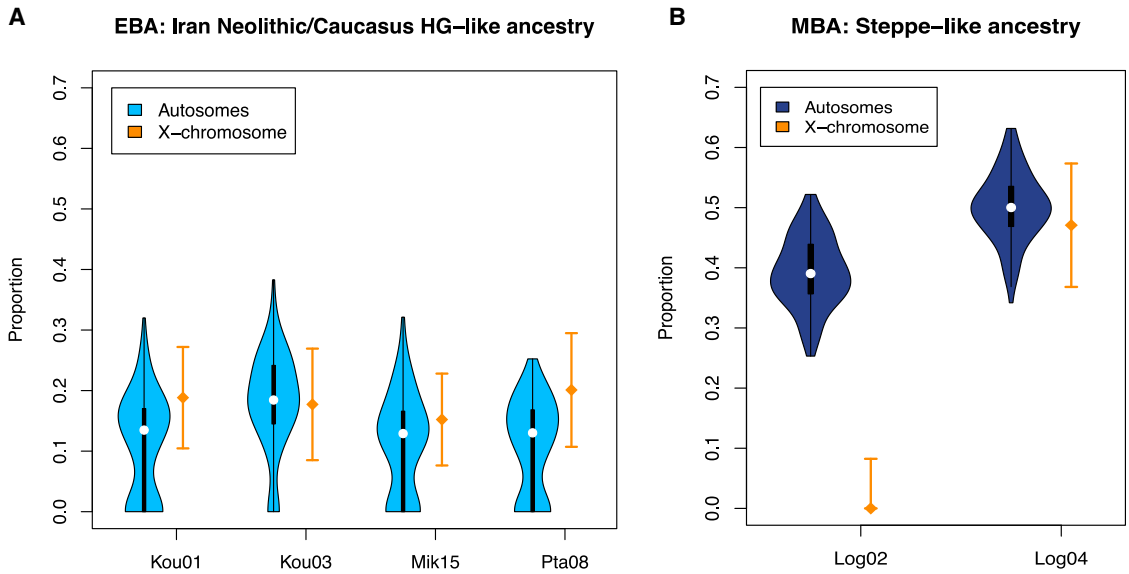


Figure 5. Sex-biased gene flow

Comparison of X-linked and autosomal genetic ancestries associated with (A) *Iran_N/CHG*-like and (B) Steppe-like components in EBA and MBA Aegean ancient genomes. Violins show the distribution of point estimates across 100 replicates for the corresponding autosomal ancestry, with median indicated by a dot and interquartile ranges indicated by boxes. For the violin plots, we considered a random set of autosomal SNPs matching X-linked SNPs in number (i.e., 8,133) (STAR Methods). Orange diamonds show the point estimates and one associated Standard Error for the same ancestries on the X chromosome.
See also Figure 3 and Table 1.

EHG compared to *Anatolia_N* in the *D*-statistics (Figure S6). However, like the Helladic-Logkas-MBA, they share more alleles with *Steppe_EMBA*. Mycenaean-Peloponnese-LBA had previously been shown to be consistent with a *qpWave/qpAdm* model that either involved BA Steppe- or Armenian-related populations (Lazaridis et al., 2017). We recapitulated this result and we additionaly found that Mycenaean-Peloponnese-LBA data are also consisent with a model involving an EBA Aegean and *Anatolia_N* as source populations (Table 3). In contrast, the Helladic-Logkas-MBA require a Steppe-like source and cannot be explained with a simple model involving an Armenian-like source (Tables 3, S3, and S5).

There are further alternative explanations consistent with the data. First, the Mycenaean-Peloponnese-LBA could be the descendants of populations closely related to the MBA Logkas population and to an EBA Aegean population—a 2-way admixture between populations related to Helladic-Logkas-MBA (~21%–36%) and the *Minoan_Odigitria_EMBA* and *Minoan_Lasithi_MBA* (~64%–79%). Similarly, a 2-way admixture between the Helladic-Logkas-MBA *Log04* individual (~34%–36%) and EBA Aegeans (~64%–66%) could not be rejected (Table S3). Second, populations related to Armenia BA may have contributed to the Aegeans in a geographically localized fashion during the LBA or earlier (Table S5). This scenario was proposed in the archaeological literature (Drews, 1988) and would imply that the Mycenaeans would not have left much trace in individuals from later generations.

Present-day Greek populations resemble MBA Logkas

Lazaridis et al., (2017) found present-day Greek populations to be quite distinct from later phases of the BA in the Aegean. In contrast, our results reveal that present-day individuals from Greece (northern Greece—Thessaloniki—and Crete) are closely related to the Helladic-Logkas-MBA individuals of northern Greece, falling near present-day Greeks in *MDS* analysis (Figure 2), sharing the same ancestry components in *ADMIXTURE* (Figure 3), and having very similar *D-statistics* (Figure S6). Moreover, in *qpWave/qpAdm* analyses (Table 3), the Thessaloniki individuals could be successfully modeled with ~93%–96% MBA Logkas-related ancestry, and a small fraction (4%–11%) of a second component (either *EHG* or Eurasian Upper Paleolithic populations such as *Kostenki14* [Fu et al., 2016] or *MA1* [Raghavan et al., 2014]). The latter are basal populations that constitute a distant outgroup to the Aegean genomes and appear to be interchangeable in this analysis across tests. This suggests that modern populations from northern Greece and Crete could be descendants of Aegean EBA populations, with subsequent admixture with populations related to the Pontic-Caspian Steppe EMBA. Interestingly, modern Cypriots carry no evidence for Steppe-like gene flow across analyses (Figures 2, 3, and S6; Table 3).

Phenotypic insights: Pigmentation and lactose intolerance

Using genotype data, we predicted that *Pta08*, *Kou01*, and *Log02* most likely had brown eyes, dark brown to black hair, and dark skin (Table S1; STAR Methods). These predictions match the visual representations of male individuals from BA wall paintings of Minoan Crete for hair and eye color. The eye and hair color predictions were similar to those from later periods of the Aegean BA (Lazaridis et al., 2017). Although the overall prediction for all three individuals was of dark skin, they also all carried alleles strongly

associated with lighter skin color (rs1426654 in the gene *SLC24A5,* and rs16891982 in *SLC45A2*) (Mathieson et al., 2015). The latter is in line with observations that skin depigmentation has been segregating since the Neolithic in southern Europe (Hofmanová et al., 2016; Mathieson et al., 2015).

Adulthood lactose tolerance was tested on two strongly associated variants, the T allele for rs4988235 under selection in ancient and modern Europeans (Enattah et al., 2008; Mathieson et al., 2015; Tishkoff et al., 2007) and the A allele for rs182549 (−22018A) (Enattah et al., 2002). All three individuals carried the ancestral state in homozygous form (including the MBA Logkas) both at −13910T and at −22018A. This is in line with other results for Neolithic Europeans and Aegeans (Allentoft et al., 2015; Hofmanová et al., 2016; Mathieson et al., 2015) and suggests that dairying may well have been practiced (Evershed et al., 2008) while individuals were lactose intolerant (Document S1). This observation supports a model of mutation-limited adaptation, as has been observed widely across species and phenotypes (Casillas and Barbadilla, 2017; Harris et al., 2018; Jensen et al., 2019).

Concluding remarks

During the EBA, the Aegean saw key innovations in trade, craft specialization, social structure, and urbanization. These changes—that mark the end of the Neolithic Period—left indelible marks on Europe and signaled the start of the urban revolution. At the beginning of this cultural transformation, the Aegean world was mostly split between three iconic palatial civilizations, the Helladic, the Cycladic, and the Minoan, each distinguishable by their artwork, pottery style, burial customs, and architecture (Cline, 2012; Shelmerdine, 2008).

To better understand the origin of the people behind this transformation, we sequenced four EBA individuals covering all three Aegean BA cultures (Helladic, Cycladic, and Minoan), two MBA individuals from northern Greece, as well as 11 mtDNA genomes from EBA Aegeans. The increased number of variants covered by the whole genomes from this study compared to previous SNP capture data from later periods in the BA Aegean (Figure S2A), as well as the inherent random variant selection characterizing whole genomes (Figure S2B), allowed us to perform demographic inference and statistically contrast population histories. Moreover, the whole genomes generated here can be easily combined with any genomic data (whole genomes, capture data—1240K or otherwise) with a limited loss of variants in future studies of human population history. Note that future work will be required to determine how representative the analyzed genomes of the Aegeans are of the BA Cycladic, Minoan, and Helladic cultures as a whole.

In summary, these genomes from the Cycladic, Minoan, and Helladic (Mycenaean) BA civilizations suggest that these culturally different populations were genetically homogeneous across the Aegean and western Anatolia at the beginning of the BA. The EBA genomes drew their ancestry mainly from local Aegean farmers and from populations related to the *CHG.* These findings are consistent with long-standing archaeological theories regarding the Neolithic-Bronze Age transformation, namely the immigration of new peoples from Anatolia and the Caucasus (Blegen and Haley, 1928; Caskey, 1971; Wace, 1957). However, because the contribution of the local Neolithic populations was significant (Dickinson, 2016; Renfrew, 1972; Tsountas and Manatt, 1897), both local and incoming elements appear to have contributed to the EBA innovations.

In contrast, the MBA Aegean population was considerably more structured. One likely reason for such structure is additional Pontic-Caspian Steppe-related gene flow into the Aegean, for which evidence was seen in the newly sequenced MBA Logkas genomes. Present-day Greeks—who also carry Steppe-related ancestry—share ~90% of their ancestry with MBA northern Aegeans, suggesting continuity between the two time periods. In contrast, LBA Aegeans (Mycenaeans) may carry either diluted Steppe- or Armenian-related ancestry (Lazaridis et al., 2017). This relative discontinuity could be explained by the general decline of the Mycenaean civilization as previously proposed in the archaeological literature (Middleton, 2019). Finally, the inferred migration waves all predate the appearance of Linear B script (1,450 BCE) (Chadwick, 2014). As a result, the genomic data could support both dominant linguistic theories explaining the emergence of Proto-Greek and the evolution of Indo-European languages (Gray et al., 2011). Namely, that these languages either originated in Anatolia (Renfrew, 1972, 1989, 2000) (correlating with the Anatolian and Caucasus-like genetic ancestries) or they originated in the Pontic-Caspian Steppe region (Anthony, 2010) (correlating with the Steppe-like ancestry). Future Mesolithic to BA genomes from Armenia and the Caucasus regions in general could help to further pinpoint the origins and the mode of gene flow into the Aegean and to better integrate the genomic data with the existing archaeological and linguistic evidence.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological material and ethics permission
- METHOD DETAILS
  - Radiocarbon dating
  - Ancient DNA Data Generation
  - Processing and Mapping of the Raw Sequencing Data
  - Authenticity of data
  - Uniparental markers
  - Reference panels
  - Multidimensional Scaling (MDS)
  - ADMIXTURE analysis: Population structure and sex-biased gene flow
  - f_3/D-statistics
  - qpWave/qpAdm analysis
  - ROH analysis
  - ABC-DL
  - Validation of the ABC-DL approach

○ Phenotype prediction

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.03.039.

ACKNOWLEDGMENTS

We would like to thank the Lausanne Genomics Technologies Facility (GTF) for performing sequencing and the former Vital-IT platform of the SIB Swiss Institute of Bioinformatics (SIB) and the DCSR for maintaining the computer infrastructure at UNIL. We thank the INCD (https://incd.pt/) for use of their computing infrastructure, which is funded by FCT and FEDER (01/SAICT/2016 022153).C.P., E.G., A.S., L.W., and J. Burger acknowledge the support of the European Union and the General Secretariat of Research and Innovation-GSRI, Ministry of Development & Investments in Greece, and the Federal Ministry of Education and Research-BMBF in Germany under the Bilateral Cooperation Program Greece – Germany 2017 (project BIOMUSE-0195). O.L. and O. Dolgova acknowledge the support of the Spanish Ministry of Science and Innovation to the EMBL partnership, Centro de Excelencia Severo Ochoa, CERCA Programme/Generalitat de Catalunya, Spanish Ministry of Science and Innovation through the Instituto de Salud Carlos III, Generalitat de Catalunya through Departament de Salut and Departament d'Empresa i Coneixement, as well as co-financing with funds from the European Regional Development Fund by the Spanish Ministry of Science and Innovation corresponding to the Programa Operativo FEDER Plurirregional de España (POPE) 2014-2020, and by the Secretaria d'Universitats i Recerca, Departament d'Empresa i Coneixement of the Generalitat de Catalunya corresponding to the Programa Operatiu FEDER de Catalunya 2014-2020. We would also like to thank David Reich and his lab for making ancient and modern data available; Iosif Lazaridis for help with *qpWave*/*qpAdm*; Aggeliki Georgiadou for help with editing the supplementary information; Iago Maceda for advice on whole genome sequence dataset building; and Julia Santiago, Alexandre Reymond, Eske Willerslev, Johannes Krause, and Morten Allentoft for helpful discussions. We thank Theodoros Pitsios (Anthropology Museum, National and Kapodistrian University of Athens) for providing samples from the osteological material of Agios Kosmas archaeological site. We thank the three anonymous reviewers for their constructive comments, which helped us to improve the manuscript. F.C., C.E.G.A., S.N., D.I.C.D., L.A., B.S.d.M., Y.O.A.C., F.M., J.V.M.-M., and A.-S.M. were supported by the Swiss National Science Foundation (SFNS) and a European Research Council (ERC) grant to A.-S.M. M.U., S.T., D.U.-K., and C.P. were co-financed by the EU Social Fund and the Greek national funds research funding program ARISTEIA II (project-3461). C.P., E.G., A.S., L.W., and J. Burger were co-financed by the Greek-German bilateral cooperation program 2017 (General Secreteriat for Research and Innovation, Ministry of Development and Investments, Greece, and Federal Ministry of Education and Research - BMBF, Germany) project BIOMUSE-0195 funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (EU Social Fund and European Regional Development Fund). J. Blöcher was funded in part by the research initiative GeneRED. E.K. was funded by the Greek State Scholarships Foundation (IKY). O. Delaneau is funded by a SNSF (project grant PP00P3_176977). V.C.S. was supported by Portuguese Foundation for Science and Technology (FCT-Fundação para a Ciência e Tecnologia) through funds granted to cE3c (UIDB/00329/2020) and individual grant CEECIND/02391/2017. O.L. was supported by a Ramón y Cajal grant from the Spanish Ministerio de Economia y Competitividad (MEIC) (RYC-2013-14797), a PGC2018-098574-B-I00 (MEIC/FEDER) grant, and the support of Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 937). O. Dolgova was supported by a PGC2018-098574-B-I00 (MEIC/FEDER) grant. J.D.J. was funded by National Institutes of Health grants R01GM135899 and R35GM139383.

AUTHOR CONTRIBUTIONS

C.P. conceived the project idea and initiated the project. A.-S.M. and C.P. headed the study. C.P. acquired corresponding samples and organized data production with critical input from J. Burger and A.-S.M. M.U. and L.W. performed the lab work. E.G., E.K., and A.S. contributed texts and archaeological and genetic background information. I.K. contributed lab work for radiocarbon dating. E.G., E.K., G.K.-M., O.P., A. Sampson, D.T., M.T., I.A., P.H., K.K., D.U.-K., D.P., C.Z., and S.T. contributed samples and background archaeological information. F.C., M.U., O. Dolgova, E.G., and C.P. compiled reference datasets. A.-S.M., O.L., and V.C.S. designed the genetic analyses with critical input from F.C., O. Dolgova, M.U., C.E.G.A., F.C.-S., S.N., D.I.C.D., L.A., F.M., J. Blöcher, B.S.d.M., Y.O.A.C., O. Delaneau, and J.V.M.-M. performed the population genetic analyses under the supervision of A.-S.M., O.L., and V.C.S. A.-S.M. wrote the manuscript together with F.C., C.P., M.U., and O. Dolgova with critical input from C.E.G.A., V.C.S., O.L., J.V.M.-M., J. Burger, and J.D.J. and from all other coauthors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 17, 2020
Revised: September 17, 2020
Accepted: March 18, 2021
Published: April 29, 2021

SUPPORTING CITATIONS

The following references appear in the Supplemental information: Aeschbacher et al., 2012); Akamatis (2009); Andreou (2012); Andreou et al. (1996); Anthony (1995); Bartonĕk (2003); Beekes (2011); Behar et al. (2008); Bellwood (2001, 2005); Bellwood and Renfrew (2003); Ben Halima et al. (2017); Bintliff and Sarri, 2018); Bogucki (1984); Boitard et al. (2016); Bouckaert et al. (2012, 2013); Brandt et al., (2013); Broodbank (2002); Brotherton et al. (2013); Brown and Brown (2011); Burger et al. (2007); Carruba (1995); Caskey (1960, 1966a, 1966b); Cassidy et al. (2016); Chadwick and Edwards (1975); Chang et al. (2015); Chapman et al. (2006); Cheer and Allen (1997); Childe (1915); Choleva (2012); Coleman and Facorellis (2018); Danecek et al. (2011); Demacopoulou (1990); Danecek et al. (2011); Demacopoulou (1990); Demoule (2017); Demoule and Perles (1993); Diamond and Bellwood (2003); Douka et al. (2017); Doumas (1977); Driessen (2008); Duhoux (1998); Durham (1991); Durham (1991); Edwards et al. (1973); Excoffier et al. (2013); Fernandes et al. (2018); Fernández et al. (2014); Fick (1905); Forsén, (1992, 2012); Fortson (2010); Foundoulakis (1985); French (1973); Furnée (1979); Gamba et al. (2014); Gamkrelidze and Ivanov (1983a, 1983b, 1995); Gerbault (2013); Gerbault et al. (2011); Giannopoulos (2012); Gilman et al. (1981); Gimbutas (1956, 1973, 1997); Goodfellow et al. (2016); Gray and Atkinson (2003); Gutenkunst et al. (2009); Haak et al. (2010); Halstead and Kotsakis (2005); Harrison (1975); Hay (2017); Heubeck (1961); Hofreiter et al. (2001); Hood (1963); Ingram et al. (2007, 2009); Itan et al. (2009); Kahveci and Singh (2006); Karamitrou-Mentessidi (2012, 2016); Karamitrou-Mentessidi and Theodorou (2013); Karamitrou-Mentessidi et al. (2015); Karnava (1999); Kirin et al. (2010); Kotsakis (2008, 2018); Kotsakis and Halstead (2004, 2016); Kouka (2002, 2011, 2013); Kretschmer (1896); Kruschke (2014); Kuchay et al. (2013); Lacan et al. (2011a, 2011b); Lander et al. (2001); Laroche (1957); Lawson et al. (2018); Lespez et al. (2016); Lifran et al., 2009); Lukić and Hey (2012); Mallory (1989); Mallory and Adams (1997); Malmström et al. (2010); Maniatis (2012); Maniatis and Ziota (2011); Manning (1995, 2012); Maran (1998, 2007); McCracken (1971); McQuillan et al. (2008); Mellaart (1973); Mendizabal et al. (2012); Meyer and Kircher (2010); Morin (2008); Mylonas (1959); Nikita et al. (2017); Mulcare et al. (2004); Olalde et al. (2018); Olivier (1986); Özdoğan (2006); Palmer (1961); Panagiotopoulos (2012, 2015, 2016); Papadatos, (2008)·Papadatos and Tomkins (2013); Papathanasopoulos (1962); Papavasileiou (1959); Parlama (1992); Perlès (2001); Parpola, (2008) Perlès et al. (2013); Peter (2016); Philaniotou, 2006, 2008, 2017a,b); Plantinga et al. (2012); Pudlo et al. (2016); Pullen (2008); Rahmstorf (2015); Rascovan et al., 2019); Rasmussen et al. (2015);

Renaud et al. (2017); Renfrew (1988, 1989, 2018); Renfrew and Bahn (2014); Ringe (2006); Rutter (1979, 1983, 1993, 2001); Şahoğlu (2005); Sampson (1981, 1985, 1988, 1991); Schachermeyr (1955); Semino et al. (2000, 2004); Shelton (2010, 2012); Sherratt (1981); Sisson et al. (2018); Swallow (2003); Szécsényi-Nagy et al. (2015); Televantou (2008); Theocharis (1959, 1973); Todorova (1995); Tomkins and Schoep (2012); Triantaphyllou (2001, 2008, 2017); Triantaphyllou et al. (2015, 2018); Tsartsidou and Kotsakis (2020); Tsipopooulou (1999, 2002, 2010, 2012a, 2012b, 2017); Tsirtsoni (2016a, 2016b); van Oven and Kayser (2009); Vavouranakis (2015); Ventris and Chadwick (1956); Voutsaki (2012); Wace and Blegen (1916); Wakeley (2008); Walsh et al. (2011, 2013); Walter and Felten (1981); Weiss (2000); Weninger et al. (2009); Whittle (1997); Wiener (2013); Windler et al. (2013); Wollstein et al. (2010); Xanthoudides (1924); Xu et al. (2010); Zapheiropoulou (2008); Ziota (2007); Ziota and Triantaphyllou (2004).

REFERENCES

Aeschbacher, S., Beaumont, M.A., and Futschik, A. (2012). A novel approach for choosing summary statistics in approximate Bayesian computation. Genetics *192*, 1027–1047.

Akamatis, I. (2009). Proistoriki Pella: Nekrotafeio Epoxhis Chalkou. In Kermatia Filias. Studies in Honour of Ioannis Touratsoglou, S. Drougou, D. Evgenidou, C. Kritzas, N. Kaltsas, B. Penna, I. Tsourti, M. Galani-Krikou, and E. Ralli, eds. (Numismatic Museum), pp. 193–213.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

Andreou, S. (2012). Northern Aegean. In The Oxford Handbook of the Bronze Age Aegean, E.H. Cline, ed. (Oxford University Press), pp. 643–659.

Andreou, S., Fotiadis, M., and Kotsakis, K. (1996). Review of Aegean prehistory V: the Neolithic and Bronze Age of northern Greece. Am. J. Archaeol. *100*, 537–597.

Andrews, S. (2010). Fastqc. A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anthony, D.W. (1995). Horse, wagon and chariot: Indo-European languages and archaeology. Antiquity *69*, 554–565.

Anthony, D.W. (2010). The horse, the wheel, and language: How Bronze-Age riders from the Eurasian steppes shaped the modern world (Princeton University Press).

Antonio, M.L., Gao, Z., Moots, H.M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R.C., et al. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. Science *366*, 708–714.

Bartoněk, A. (2003). Handbuch des mykenischen Griechisch (Universitaetsverlag Winter).

Beaumont, M.A. (2019). Approximate Bayesian Computation. Annu. Rev. Stat. Appl. *6*, 379–403.

Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025–2035.

Beekes, R.S.P. (2011). Comparative Indo European Linguistics: An Introduction (John Benjamins Publishing Company).

Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., et al.; Genographic Consortium (2008). The dawn of human matrilineal diversity. Am. J. Hum. Genet. *82*, 1130–1140.

Bellwood, P. (2001). Early Agriculturalist Population Diasporas? Farming, Languages, and Genes. Annu. Rev. Anthropol. *30*, 181–207.

Bellwood, P. (2005). First Farmers: the Origins of Agricultural Societies (Wiley-Blackwell).

Bellwood, P., and Renfrew, C. (2003). Examining the Farming/Language Dispersal Hypothesis (McDonald Institute for Archaeological Research).

Ben Halima, Y., Kefi, R., Sazzini, M., Giuliani, C., De Fanti, S., Nouali, C., Nagara, M., Mengozzi, G., Elouej, S., Abid, A., et al. (2017). Lactase persistence in Tunisia as a result of admixture with other Mediterranean populations. Genes Nutr. *12*, 20.

Bertorelle, G., Benazzo, A., and Mona, S. (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Mol. Ecol. *19*, 2609–2625.

Bintliff, J., and Sarri, K. (2018). Demographic transitions from the Earlier Neolithic stages until the first Early Bronze Age settlements in the plains and hill-country of Boeotia, Greece. Communities in Transition: The Circum-Aegean Area During the 5th and 4th Millennia BC (Oxbow Books), pp. 249–259.

Blegen, C.W., and Haley, J. (1928). The coming of the Greeks: II. The geographical distribution of prehistoric remains in Greece. Am. J. Archaeol. *32*, 146–154.

Bock, C., Walter, J., Paulsen, M., and Lengauer, T. (2007). CpG island mapping by epigenome prediction. PLoS Comput. Biol. *3*, e110.

Bogucki, P.I. (1984). Ceramic Sieves of the Linear Pottery Culture and their Economic Implications. Oxf. J. Archaeol. *3*, 15–30.

Boitard, S., Rodríguez, W., Jay, F., Mona, S., and Austerlitz, F. (2016). Inferring Population Size History from Large Samples of Genome-Wide Molecular Data - An Approximate Bayesian Computation Approach. PLoS Genet. *12*, e1005877.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S.J., Alekseyenko, A.V., Drummond, A.J., Gray, R.D., Suchard, M.A., and Atkinson, Q.D. (2012). Mapping the origins and expansion of the Indo-European language family. Science *337*, 957–960.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S.J., Alekseyenko, A.V., Drummond, A.J., Gray, R.D., Suchard, M.A., and Atkinson, Q.D. (2013). Corrections and Clarifications. Science *342*, 1446.

Brandt, G., Haak, W., Adler, C.J., Roth, C., Szécsényi-Nagy, A., Karimnia, S., Möller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., et al.; Genographic Consortium (2013). Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. Science *342*, 257–261.

Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Res. *38*, e87.

Broodbank, C. (2002). An Island Archaeology of the Early Cyclades (Cambridge University Press).

Broodbank, C. (2008). The Early Bronze Age in the Cyclades. In The Cambridge Companion to the Aegean Bronze Age, C.W. Shelmerdine, ed. (Cambridge University Press), pp. 47–76.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S.M., Der Sarkissian, C., Ganslmeier, R., Friederich, S., et al.; Genographic Consortium (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat. Commun. *4*, 1764.

Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Díez-del-Molino, D., et al. (2016). Early Neolithic genomes from the eastern Fertile Crescent. Science *353*, 499–503.

Brown, T., and Brown, K. (2011). Studying the Diets of Past People. In Biomolecular Archaeology: An Introduction (Wiley), pp. 190–209.

Burger, J., Kirchner, M., Bramanti, B., Haak, W., and Thomas, M.G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proc. Natl. Acad. Sci. USA *104*, 3736–3741.

Carruba, O. (1995). L'arrivo dei Greci, le migrazioni indoeuropee e il" ritorno" degli Eraclidi. Athenaeum *83*, 5.

Casillas, S., and Barbadilla, A. (2017). Molecular Population Genetics. Genetics *205*, 1003–1035.

Caskey, J.L. (1960). The early Helladic period in the Argolid. Hesperia *29*, 285–303.

Caskey, J.L. (1966a). Excavations in Keos 1964-1965. Hesperia *35*, 363–367.

Caskey, J.L. (1966b). Greece and the Aegean islands in the Middle Bronze Age (Cambridge University Press).

Caskey, J.L. (1971). Greece, Crete and the Aegean islands in the Early Bronze Age. In The Cambridge Ancient History, I.E.S. Edwards, C.J. Gadd, and N.G.L. Hammond, eds. (Cambridge University Press), pp. 771–807.

Cassidy, L.M., Martiniano, R., Murphy, E.M., Teasdale, M.D., Mallory, J., Hartwell, B., and Bradley, D.G. (2016). Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proc. Natl. Acad. Sci. USA *113*, 368–373.

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., and Wilson, J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet. *19*, 220–234.

Chadwick, J.L. (2014). The Decipherment of Linear B (Cambridge University Press).

Chadwick, J.L., and Edwards, I.E.S. (1975). The prehistory of the Greek language. In The Cambridge Ancient History, 2. History of the Middle East and the Aegean Region c. 1380-1000 B.C (Cambridge University Press), pp. 805–819.

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

Chang, W., Cathcart, C., Hall, D., and Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. Language *91*, 194–244.

Chapman, J., Higham, T., Slavchev, V., Gaydarska, B., and Honch, N. (2006). The social context of the emergence, development and abandonment of the Varna cemetery, Bulgaria. Eur. J. Archaeol. *9*, 159–183.

Cheer, S.M., and Allen, J.S. (1997). Lactose digestion capacity and perceived symptomatic response after dairy product consumption in Tokelau Island migrants. Am. J. Hum. Biol. *9*, 233–246.

Childe, V.G. (1915). On the Date and Origin of Minyan Ware. J. Hell. Stud. *35*, 196–207.

Childe, V.G. (1942). 59. Prehistory in the U.S.S.R. I.Palaeolithic and Mesolithic. A. Caucasus and Crimea. Man (Lond.) *42*, 98–100.

Childe, V.G. (1950). The Urban Revolution. Town Plan. Rev. *21*, 3–17.

Choleva, M. (2012). The first wheelmade pottery at Lerna: wheel-thrown or wheel-fashioned? Hesperia *81*, 343–381.

Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H., and Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. Genome Res. *15*, 1496–1502.

Cline, E.H. (2012). The Oxford Handbook of the Bronze Age Aegean (Oxford University Press).

Coleman, J. (2000). An Archaeological Scenario for the "Coming of the Greeks" ca. 3200 B.C. J. Indo-European Stud. *28*, 101–153.

Coleman, J.E., and Facorellis, Y. (2018). The shadowy "proto-Early Bronze Age" in the Aegean. In Communities in Transition: The Circum-Aegean Area during the 5th and 4th Millennia BC, S. Dietz, F. Mavridis, Ž. Tankosić, and T. Takaoğlu, eds. (Oxbow Books), pp. 33–66.

Cox, M.A.A., and Cox, T.F. (2008). Multidimensional Scaling. In Handbook of Data Visualization (Springer Berlin Heidelberg), pp. 315–347.

Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., and François, O. (2010). Approximate Bayesian Computation (ABC) in practice. Trends Ecol. Evol. *25*, 410–418.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science *360*, eaar7711.

Demacopoulou, K. (1990). Marmarino andriko eidolio ("O Arpistis tis Naxou"). In Kykladikos Politismos. I Naxos Stin 3i p.C Chilietia, L. Marangou, ed. (N.P. Goulandris Collection), p. 115.

Demoule, J.P. (2017). The transitions between Neolithic and Early Bronze Age in Greece, and the "Indo-European problem.". In Balkan Dialogues, M. Gori and M. Ivanova, eds. (Routledge), pp. 52–63.

Demoule, J.-P., and Perles, C. (1993). The Greek Neolithic: A new review. J. World Prehist. *7*, 355–416.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

Diamond, J., and Bellwood, P. (2003). Farmers and their languages: the first expansions. Science *300*, 597–603.

Dickinson, O. (1994). The Aegean Bronze Age (Cambridge University Press).

Dickinson, O. (2016). 'The Coming of the Greeks' and All That. In The Archaeology of Greece and Rome: Studies In Honour of Anthony Snodgrass, J. Bintliff and N.K. Rutter, eds. (Edinburgh University Press), pp. 3–21.

Douka, K., Efstratiou, N., Hald, M.M., Henriksen, P.S., and Karetsou, A. (2017). Dating Knossos and the arrival of the earliest Neolithic in the southern Aegean. Antiquity *91*, 304–321.

Doumas, C. (1977). Early Bronze Age burial habits in the Cyclades. In Studies in Mediterranean Archaeology (Paul Åströms), p. 144.

Doumas, C. (2010). Crete and the Cyclades in the Early Bronze Age: A view from the North. In Cretan Offerings: Studies in Honour of Peter Warren, O. Krzsyzkowska, ed. (British School at Athens Studies), pp. 101–106.

Drews, R. (1988). The Coming of the Greeks: Indo-European Conquests in the Aegean and the Near East (Princeton University Press).

Driessen, J. (2008). Chronology of the Linear B texts. In Companion to Linear B Mycenaean Greek Texts and Their World, Y. Duhoux and A.M. Davis, eds. (Peeters Louvain La Neuve), pp. 69–79.

Duhoux, Y. (1998). Pre-Hellenic language(s) of Crete. J. Indo-Eur. Stud. *26*, 1–39.

Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Mol. Biol. Evol. *28*, 2239–2252.

Durham, W.H. (1991). The relationship of genes and culture. In Coevolution: Genes, Culture and Human Diversity, W.H. Durham, ed. (Stanford University Press), pp. 154–225.

Edwards, I.E.S., Gadd, C.J., Hammond, N.G.L., and Sollberger, E. (1973). The Cambridge Ancient History (Cambridge University Press).

Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L., and Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. Nat. Genet. *30*, 233–237.

Enattah, N.S., Jensen, T.G.K., Nielsen, M., Lewinski, R., Kuokkanen, M., Rasinpera, H., El-Shanti, H., Seo, J.K., Alifrangis, M., Khalil, I.F., et al. (2008). Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. Am. J. Hum. Genet. *82*, 57–72.

Eupedia (2018). Distribution of European mitochondrial DNA (mtDNA) haplogroups by region in percentage. https://www.eupedia.com/europe/european_mtdna_haplogroups_frequency.shtml.

Evershed, R.P., Payne, S., Sherratt, A.G., Copley, M.S., Coolidge, J., Urem-Kotsu, D., Kotsakis, K., Ozdoğan, M., Ozdoğan, A.E., Nieuwenhuyse, O., et al. (2008). Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. Nature *455*, 528–531.

Ewing, G.B., and Jensen, J.D. (2016). The consequences of not accounting for background selection in demographic inference. Mol. Ecol. *25*, 135–141.

Excoffier, L., and Foll, M. (2011). fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics *27*, 1332–1334.

Excoffier, L., and Lischer, H.E.L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. *10*, 564–567.

Excoffier, L., Smouse, P.E., and Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics *131*, 479–491.

Excoffier, L., Estoup, A., and Cornuet, J.-M. (2005). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics *169*, 1727–1738.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., and Foll, M. (2013). Robust demographic inference from genomic and SNP data. PLoS Genet. *9*, e1003905.

Fenner, J.N. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. *128*, 415–423.

Fernandes, D.M., Strapagiel, D., Borówka, P., Marciniak, B., Żądzińska, E., Sirak, K., Siska, V., Grygiel, R., Carlsson, J., Manica, A., et al. (2018). A genomic Neolithic time transect of hunter-farmer admixture in central Poland. Sci. Rep. *8*, 14879.

Fernandes, D.M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., et al. (2020). The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. Nat. Ecol. Evol. *4*, 334–345.

Fernández, E., Pérez-Pérez, A., Gamba, C., Prats, E., Cuesta, P., Anfruns, J., Molist, M., Arroyo-Pardo, E., and Turbón, D. (2014). Ancient DNA analysis of 8000 B.C. near eastern farmers supports an early neolithic pioneer maritime colonization of Mainland Europe through Cyprus and the Aegean Islands. PLoS Genet. *10*, e1004401.

Fick, A. (1905). Vorgriechische Ortsnamen als Quelle für die Vorgeschichte Griechenlands verwertet (Vandenhoeck und Ruprecht).

Forsén, J. (1992). The Twilight of the Early Helladic: a Study of the Disturbances in East-Central and Southern Greece towards the End of the Early Bronze Age (Paul Åströms).

Forsén, J. (2012). Mainland Greece. In The Oxford Handbook of the Bronze Age Aegean, E.H. Cline, ed. (Oxford University Press), pp. 99–112.

Fortson, B.W. (2010). Indo-European Language and Culture. An introduction (Blackwell).

Foundoulakis, E. (1985). The skeleton remains from the cemetery of Manika, Chalkis. In An Early Helladic T. near Chalkis, Euboea, A. Sampson, ed. (Archive of Euboean Studies), pp. 249–309.

Francis, R.M. (2017). pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. *17*, 27–32.

French, D.H. (1973). Migrations and 'Minyan'pottery in Western Anatolia and Aegean. In Bronze Age Migrations in the Aegean. Archaeological and Linguistic Problems in Greek Prehistory, R. Crossland and A. Birchal, eds. (Duckworth), pp. 51–54.

Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. Curr. Biol. *23*, 553–559.

Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

Furnée, E.J. (1979). Vorgriechisch-Kartvelisches: Studien zum ostmediterranen Substrat, nebst einem Versuch zu einer neuen pelagischen Theorie (Peeters Louvain La Neuve).

Galtier, N., Gouy, M., and Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. *12*, 543–548.

Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.

Gamkrelidze, T.V., and Ivanov, V.V. (1983a). The ancient near-east and the Indo-european question-temporal and territorial characteristics of proto-Indo-european based on linguistic and historico-cultural data. J. Indo-Eur. Stud. *13*, 3–48.

Gamkrelidze, T.V., and Ivanov, V.V. (1983b). The Migrations of Tribes Speaking the Indo-European Dialects from Their Original Homeland in the Near East to Their Historical Habitations in Eurasia. Sov. Stud. Hist. *22*, 53–95.

Gamkrelidze, T.V., and Ivanov, V.V. (1995). Indo-European and the Indo-Europeans (Mouton de Gruyter).

Gerbault, P. (2013). The onset of lactase persistence in Europe. Hum. Hered. *76*, 154–161.

Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D.M., and Thomas, M.G. (2011). Evolution of lactase persistence: an example of human niche construction. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 863–877.

Giannopoulos, T.G. (2012). The Greeks: Whence and When? The Mainstream Scientific Responses and the Present State of Research on the First Beginning of the Greek Civilization (Crete University Press).

Gilman, A., Cazzella, A., Cowgill, G.L., Crumley, C.L., Earle, T., Gallay, A., Harding, A.F., Harrison, R.J., Hicks, R., Kohl, P.L., et al. (1981). The Development of Social Stratification in Bronze Age Europe. Curr. Anthropol. *22*, 1–23.

Gimbutas, M. (1956). The Prehistory of Eastern Europe (Peabody Museum).

Gimbutas, M. (1973). The Beginning of the Bronze Age in Europe and the Indo-Europeans: 3500-2500 BC (Universitätsbibliothek Johann Christian Senckenberg).

Gimbutas, M. (1997). The Kurgan Culture and the Indo-Europeanization of Europe: Selected Articles from 1952 to 1993 (Institute for the Study of Man).

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. *27*, 182–189.

Goldberg, A., Günther, T., Rosenberg, N.A., and Jakobsson, M. (2017). Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. Proc. Natl. Acad. Sci. USA *114*, 2657–2662.

Goodfellow, I.J., Bengio, Y., and Courville, A. (2016). Deep Learning (MIT Press).

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. *27*, 221–224.

Gray, R.D., and Atkinson, Q.D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature *426*, 435–439.

Gray, R.D., Atkinson, Q.D., and Greenhill, S.J. (2011). Language evolution and human history: what a difference a date makes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 1090–1100.

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

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. *5*, e1000695.

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

Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., Der Sarkissian, C.S.I., Brandt, G., Schwarz, C., Nicklisch, N., et al.; Members of the Genographic Consortium (2010). Ancient DNA from

European early neolithic farmers reveals their near eastern affinities. PLoS Biol. *8*, e1000536.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

Halstead, P., and Kotsakis, K. (2005). Paliambela. In Archaeology in Greece 2004-2005, J. Whitley, ed. (Archaeological Reports), p. 66.

Harding, A.F. (2000). European Societies in the Bronze Age (Cambridge University Press).

Harris, R.B., Sackman, A., and Jensen, J.D. (2018). On the unfounded enthusiasm for soft selective sweeps II: Examining recent evidence from humans, flies, and viruses. PLoS Genet. *14*, e1007859.

Harrison, G.G. (1975). Primary Adult Lactase Deficiency: A Problem in Anthropological Genetics. Am. Anthropol. *77*, 812–835.

Hay, M. (2017). Haplogroup H (mtDNA). https://www.eupedia.com/.

Heaton, J. (2015). Encog: Library of Interchangeable Machine Learning Models for Java and C#. J. Mach. Learn. Res. *16*, 1243–1247.

Heubeck, A. (1961). Praegraeca: sprachliche Untersuchungen zum vorgriechisch - indogermanischen Substrat (Universitätsbund Erlangen).

Hoban, S., Bertorelle, G., and Gaggiotti, O.E. (2012). Computer simulations: tools for population and evolutionary genetics. Nat. Rev. Genet. *13*, 110–122.

Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. USA *113*, 6886–6891.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., and Pääbo, S. (2001). Ancient DNA. Nat. Rev. Genet. *2*, 353–359.

Hood, M.S.F. (1963). Mylonas Aghios Kosmas. An Early Bronze Age Settlement and Cemetery in Attica. J. Hell. Stud. *83*, 199–200.

Hughey, J.R., Paschou, P., Drineas, P., Mastropaolo, D., Lotakis, D.M., Navas, P.A., Michalodimitrakis, M., Stamatoyannopoulos, J.A., and Stamatoyannopoulos, G. (2013). A European population in Minoan Bronze Age Crete. Nat. Commun. *4*, 1861.

Ingram, C.J.E., Elamin, M.F., Mulcare, C.A., Weale, M.E., Tarekegn, A., Raga, T.O., Bekele, E., Elamin, F.M., Thomas, M.G., Bradman, N., and Swallow, D.M. (2007). A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? Hum. Genet. *120*, 779–788.

Ingram, C.J.E., Mulcare, C.A., Itan, Y., Thomas, M.G., and Swallow, D.M. (2009). Lactose digestion and the evolutionary genetics of lactase persistence. Hum. Genet. *124*, 579–591.

International HapMap 3 Consortium, Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Bonnen, P.E., de Bakker, P.I., Deloukas, P., et al. (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.

Itan, Y., Powell, A., Beaumont, M.A., Burger, J., and Thomas, M.G. (2009). The origins of lactase persistence in Europe. PLoS Comput. Biol. *5*, e1000491.

Jensen, J.D., Payseur, B.A., Stephan, W., Aquadro, C.F., Lynch, M., Charlesworth, D., and Charlesworth, B. (2019). The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. Evolution *73*, 111–114.

Johri, P., Charlesworth, B., and Jensen, J.D. (2020). Toward an evolutionarily appropriate null model: Jointly inferring demography and purifying selection. Genetics *215*, 173–192.

Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. *6*, 8912.

Kahveci, T., and Singh, A.K. (2006). Index Structures for Approximate Matching in Sequence Databases. In Handbook of Computational Molecular Biology., S. Aluru, ed. (Chapman & Hall/CRC Computer and Information Science Series 9), pp. 43–6513.

Karamitrou-Mentessidi, G. (2012). Research on Mavropigi and the Ilarion dam during 2012. In AEMTh 26, P. Adam-Veleni and K. Tzanavari, eds. (Ministry of Culture Thessaloniki), pp. 47–90.

Karamitrou-Mentessidi, G. (2016). The Prefectures of Kozani and Grevena. From the Research of the Last Decade. In AEMTh 30 (Ministry of Culture Thessaloniki), In press.

Karamitrou-Mentessidi, G., and Theodorou, D. (2013). Elati, the site of Logkas: transfer and excavation of Bronze Age burial pithoi. In AEMTh 27, P. Adam-Veleni and K. Tzanavari, eds. (Ministry of Culture Thessaloniki), pp. 121–132.

Karamitrou-Mentessidi, G., Efstratiou, N., Kaczanowska, M., and Kozłowski, J.K. (2015). Early Neolithic settlement of Mavropigi in Western Greek Macedonia. Eurasian Prehistory *12*, 47–116.

Karnava, A. (1999). The Cretan Hieroglyphic Script of the Second Millennium BC: Description, Analysis, Function and Decipherment Perspectives. (Université libre de Bruxelles).

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kılınç, G.M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A.M., Bıçakçı, E., Baird, D., Dönertaş, H.M., Ghalichi, A., Yaka, R., et al. (2016). The Demographic Development of the First Farmers in Anatolia. Curr. Biol. *26*, 2659–2666.

Kılınç, G.M., Koptekin, D., Atakuman, Ç., Sümer, A.P., Dönertaş, H.M., Yaka, R., Bilgin, C.C., Büyükkarakaya, A.M., Baird, D., Altınışık, E., et al. (2017). Archaeogenomic analysis of the first steps of Neolithization in Anatolia and the Aegean. Proc. Biol. Sci. *284*, 20172064.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res. *40*, e3.

Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M., and Wilson, J.F. (2010). Genomic runs of homozygosity record population history and consanguinity. PLoS ONE *5*, e13996.

Kontopoulos, I., Penkman, K., Liritzis, I., and Collins, M.J. (2019). Bone diagenesis in a Mycenaean secondary burial (Kastrouli, Greece). Archaeol. Anthropol. Sci. *11*, 5213–5230.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.

Kotsakis, K. (2008). A sea of agency: Crete in the context of the earliest Neolithic in Greece. In Escaping the Labyrinth: The Cretan Neolithic in Context, V. Isaakidou and P.D. Tomkins, eds. (Oxbow Books), pp. 49–72.

Kotsakis, K. (2018). Transformation and changes at the end of the Neolithic. In Communities in Transition: The Circum-Aegean Area during the 5th and 4th Millennia BC, S. Dietz, F. Mavridis, Ž. Tankosić, and T. Takaoğlu, eds. (Oxbow Books), pp. 12–16.

Kotsakis, K., and Halstead, P. (2004). Anaskafi sta neolithika Paliambela Kolindrou. In AEMTh 16, P. Adam-Veleni, ed. (Ministry of Culture Thessaloniki), pp. 407–415.

Kotsakis, K., and Halstead, P. (2016). Paliambela Kolindrou: Research on the early Neolithic settlement. In AEMTh 30 (Ministry of Culture Thessaloniki), In press.

Kouka, O. (2011). Symbolism, ritual feasting and ethnicity in Early Bronze Age Cyprus and Anatolia. In On Cooking Pots, Drinking Cups, Loomweights and Ethnicity in Bronze Age Cyprus and Neighbouring Regions: An International Archaeological Symposium Held in Nicosia, V. Karageorghis and O. Kouka, eds. (AG Leventis Foundation), pp. 43–56.

Kouka, O. (2013). Prehistoric Heraion on Samos Revisited: Cultural Transformation, Interaction and Urbanism in the East Aegean. In Fritz Thyssen Stiftung Für Wissenschaftsförderung - Jahresbericht, pp. 120–122.

Kouka, O. (2002). Siedlungsorganisation in der Nord-und Ostägäis während der Frühbronzezeit (3. Jt. v. Chr.) (VML Verlag Marie Leidorf GmbH, Rahden–Westfalen).

Kretschmer, P. (1896). Einleitung in die Geschichte der griechischen Sprache (Vandenhoeck und Ruprecht).

Kristiansen, K. (2016). Interpreting Bronze Age Trade and Migration. In Human Mobility and Technological Transfer in the Prehistoric Mediterranean, C. Knappett and E. Kiriatzi, eds. (Cambridge University Press), pp. 154–180.

Kruschke, J.K. (2014). Doing Bayesian Data Analysis (Elsevier).

Kuchay, R.A.H., Anwar, M., Thapa, B.R., Mahmood, A., and Mahmood, S. (2013). Correlation of G/A -22018 single-nucleotide polymorphism with lactase activity and its usefulness in improving the diagnosis of adult-type hypolactasia among North Indian children. Genes Nutr. *8*, 145–151.

Kullback, S., and Leibler, R.A. (1951). On information and sufficiency. Ann. Math. Stat. *22*, 79–86.

Lacan, M., Keyser, C., Ricaut, F.-X., Brucato, N., Duranthon, F., Guilaine, J., Crubézy, E., and Ludes, B. (2011a). Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. Proc. Natl. Acad. Sci. USA *108*, 9788–9791.

Lacan, M., Keyser, C., Ricaut, F.-X., Brucato, N., Tarrús, J., Bosch, A., Guilaine, J., Crubézy, E., and Ludes, B. (2011b). Ancient DNA suggests the leading role played by men in the Neolithic dissemination. Proc. Natl. Acad. Sci. USA *108*, 18255–18259.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al.; International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Laroche, E. (1957). Notes de toponyme anatolienne. In MNHMHΣ XAPIN. Gedenkschrift Paul Kretschmer II, H. Kronasser, ed. (Wiener Sprachgesellschaft), pp. 1–7.

Lawson, D.J., van Dorp, L., and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat. Commun. *9*, 3258.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

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

Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., et al. (2017). Genetic origins of the Minoans and Mycenaeans. Nature *548*, 214–218.

Lespez, L., Glais, A., Lopez-Saez, J.-A., Le Drezen, Y., Tsirtsoni, Z., Davidson, R., Biree, L., and Malamidou, D. (2016). Middle Holocene rapid environmental changes and human adaptation in Greece. Quat. Res. *85*, 227–244.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, S., and Jakobsson, M. (2012). Estimating demographic parameters from large-scale population genomic data using Approximate Bayesian Computation. BMC Genet. *13*, 22.

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

Lifran, E.V., Hourigan, J.A., and Sleigh, R.W. (2009). Lactose Derivatives: Turning Waste Into Functional Foods. Aust. J. Dairy Technol. *64*, 89–93.

Lipson, M., Loh, P.-R., Sankararaman, S., Patterson, N., Berger, B., and Reich, D. (2015). Calibrating the human mutation rate via ancestral recombination density in diploid genomes. PLoS Genet. *11*, e1005550.

Gallego Llorente, M., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. Science *350*, 820–822.

Longin, R. (1971). New method of collagen extraction for radiocarbon dating. Nature *230*, 241–242.

Lorente-Galdos, B., Lao, O., Serra-Vidal, G., Santpere, G., Kuderna, L.F.K., Arauna, L.R., Fadhlaoui-Zid, K., Pimenoff, V.N., Soodyall, H., Zalloua, P., et al. (2019). Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. Genome Biol. *20*, 77.

Lukić, S., and Hey, J. (2012). Demographic inference using spectral methods on SNP data, with an analysis of the human out-of-Africa expansion. Genetics *192*, 619–639.

Malaspinas, A.-S., Tange, O., Moreno-Mayar, J.V., Rasmussen, M., DeGiorgio, M., Wang, Y., Valdiosera, C.E., Politis, G., Willerslev, E., and Nielsen, R. (2014). bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). Bioinformatics *30*, 2962–2964.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.

Mallory, J.P. (1989). Search of the Indo-Europeans (Thames and Hudson).

Mallory, J.P., and Adams, D.Q. (1997). Encyclopedia of Indo-European Culture (Taylor and Francis).

Malmström, H., Linderholm, A., Lidén, K., Storå, J., Molnar, P., Holmlund, G., Jakobsson, M., and Götherström, A. (2010). High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. BMC Evol. Biol. *10*, 89.

Maniatis, Y. (2012). Radiocarbon dating of the major cultural changes in Prehistoric Macedonia: Recent developments. In Proceedings of the International Conference on "A Century of Research in Prehistoric Macedonia, 1912-2012, E. Stefani, N. Merousis, and A. Dimoula, eds. (Archaeological Museum of Thessaloniki), pp. 22–24.

Maniatis, Y., and Ziota, C. (2011). Systematic 14C Dating of a unique Early and Middle Bronze Age cemetery at Xeropigado Koiladas, West Macedonia, Greece. Radiocarbon *53*, 461–478.

Manning, S.W. (1995). The Absolute Chronology of the Aegean Early Bronze Age: Archaeology, Radiocarbon and History (Sheffield Academy Press).

Manning, S.W. (2012). Chronology and terminology. In The Oxford Handbook of the Bronze Age Aegean, E.H. Cline, ed. (Oxford University Press), pp. 11–28.

Maran, J. (2007). Seaborne contacts between the Aegean, the Balkans and the Central Mediterranean in the 3rd millennium BC: The unfolding of the Mediterranean world. In Between the Aegean and the Baltic Seas: Prehistory across Borders Proceedings of the International Conference Bronze and Early Iron Age Interconnections and Contemporary Developments between the Aegean and the Regions of the Balkan Peninsula, Central and Northern Europe, University of Zagreb, 11-14 April 2005, I. Galanaki, H. Tomas, Y. Galanakis, and R. Laffineur, eds. (University of Zagreb), pp. 3–21.

Maran, J. (1998). Kulturwandel auf dem griechischen Festland und den Kykladen im späten 3. In Jahrtausend v. Chr (Habelt Bonn).

Marcus, J.H., Posth, C., Ringbauer, H., Lai, L., Skeates, R., Sidore, C., Beckett, J., Furtwängler, A., Olivieri, A., Chiang, C.W.K., et al. (2020). Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. Nat. Commun. *11*, 939.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.

McCracken, R.D. (1971). Origins and implications of the distribution of adult lactase deficiency in human populations. J. Trop. Pediatr. Environ. Child Health *17*, 7–10.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., et al. (2008). Runs of homozygosity in European populations. Am. J. Hum. Genet. *83*, 359–372.

Mellaart, J. (1973). The End of the Early Bronze Age in Anatolia and the Aegean. Am. J. Archaeol. *62*, 9.

Mendizabal, I., Lao, O., Marigorta, U.M., Wollstein, A., Gusmão, L., Ferak, V., Ioana, M., Jordanova, A., Kaneva, R., Kouvatsi, A., et al. (2012). Reconstructing the population history of European Romani from genome-wide data. Curr. Biol. *22*, 2342–2349.

Meyer, M., and Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448.

Middleton, G.D. (2019). Collapse of the Bronze Age Aegean. In Oxford Classical Dictionary (Oxford University Press).

Middleton, G.D. (2020). Collapse and Transformation: The Late Bronze Age to Early Iron Age in the Aegean (Oxbow Books).

Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D., and Marshall, D. (2013). Using Tablet for visual exploration of second-generation sequencing data. Brief. Bioinform. *14*, 193–202.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., et al. (2003). Natural selection shaped regional mtDNA variation in humans. Proc. Natl. Acad. Sci. USA *100*, 171–176.

Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. F1000Res. *10*, 33.

Mondal, M., Bertranpetit, J., and Lao, O. (2019). Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. Nat. Commun. *10*, 246.

Moreno-Mayar, J.V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J.P., Allentoft, M.E., Vimala, T., Racimo, F., Pinotti, T., et al. (2018). Early human dispersals within the Americas. Science *362*, eaav2621.

Moreno-Mayar, J.V., Korneliussen, T.S., Dalal, J., Renaud, G., Albrechtsen, A., Nielsen, R., and Malaspinas, A.-S. (2020). A likelihood method for estimating present-day human contamination in ancient male samples using low-depth X-chromosome data. Bioinformatics *36*, 828–841.

Morin, E. (2008). Evidence for declines in human population densities during the early Upper Paleolithic in western Europe. Proc. Natl. Acad. Sci. USA *105*, 48–53.

Mulcare, C.A., Weale, M.E., Jones, A.L., Connell, B., Zeitlyn, D., Tarekegn, A., Swallow, D.M., Bradman, N., and Thomas, M.G. (2004). The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the Lactase Gene (LCT) (C−13.9kbT) does not predict or cause the Lactase-Persistence phenotype in Africans. Am. J. Hum. Genet *74*, 1102–1110.

Mylonas, G.E. (1959). Aghios Kosmas: an Early Bronze Age settlement and cemetery in Attica (Princeton University Press).

Nikita, E., Triantaphyllou, S., Tsipopoulou, M., Panagiotopoulos, D., and Platon, L. (2017). Mobility patterns and cultural identities in pre-and proto-palatial central and eastern Crete. In Petras, Siteia: The Pre-and Proto-Palatial Cemetery in Context, M. Tsipopoulou, ed. (The Danish Institute at Athens and Aarhus University Press), pp. 325–339.

Oakley, R., and Renfrew, C. (1972). The Cyclades and the Aegean in the Third Millennium BC. Br. J. Sociol. *24*, 125.

Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., García Borja, P., Salazar-García, D.C., Talamo, S., et al. (2015). A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. Mol. Biol. Evol. msv181.

Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C.J., Gandini, F., Pala, M., et al. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. Science *363*, 1230–1234.

Olivier, J. (1986). Cretan writing in the second millennium BC. World Archaeol. *17*, 377–389.

Omrak, A., Günther, T., Valdiosera, C., Svensson, E.M., Malmström, H., Kiesewetter, H., Aylward, W., Storå, J., Jakobsson, M., and Götherström, A. (2016). Genomic evidence establishes Anatolia as the source of the European Neolithic gene pool. Curr. Biol. *26*, 270–275.

Orlando, L., Gilbert, M.T.P., and Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. Nat. Rev. Genet. *16*, 395–408.

Özdoğan, M. (2006). Neolithic cultures at the contact zone between Anatolia and the Balkans-diversity and homogeneity at the Neolithic frontier. In Aegean–Marmara–Black Sea: The Present State of Research on the Early Neolithic, I. Gatsov and H. Schwarzberg, eds. (Langenweissbach), pp. 21–28.

Palmer, L.R. (1961). Mycenaeans and Minoans: Aegean prehistory in the light of the linear B tables (Faber and Faber).

Panagiotopoulos, D. (2012). Anaskafi Koumasas. In Proceedings of the Archaeological Society at Athens (The Archaeological Society at Athens), pp. 185–216.

Panagiotopoulos, D. (2015). Anaskafi Koumasas. In Proceedings of the Archaeological Society at Athens (The Archaeological Society at Athens), pp. 525–547.

Panagiotopoulos, D. (2016). Anaskafi Koumasas. In Proceedings of the Archaeological Society at Athens (The Archaeological Society at Athens), p. 171.

Papadatos, Y. (2007). Beyond cultures and ethnicity: a new look at material culture distribution and inter-regional interaction in the Early Bronze Age Southern Aegean. In Mediterranean Crossroads, S. Antoniadou and A. Pace, eds. (Pierides Foundation), pp. 419–451.

Papadatos, Y. (2008). The Neolithic-Early Bronze Age Transition in Crete: New evidence from the Settlement at Petras Kephala, Siteia. In Escaping the Labyrinth: The Cretan Neolithic in Context, V. Isaakidou and P. Tomkins, eds. (Oxbow Books), pp. 261–275.

Papadatos, Y., and Tomkins, P. (2013). Trading, the Longboat, and Cultural Interaction in the Aegean During the Late Fourth Millennium B.C.E.: The View from Kephala Petras, East Crete. Am. J. Archaeol. *117*, 353–381.

Papathanasopoulos, G. (1962). Kykladika Naxou. Archaiologikon Delt *174*, 104–151.

Papavasileiou, G. (1959). Ancient tombs in Euboea (Monograph of Archaeological Ephemeris).

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.

Parlama, L. (1992). The prehistoric Skyros. Archaeologia *42*, 16–22.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

Parpola, A. (2008). Proto-Indo-European speakers of the Late Tripolye culture as the inventors of wheeled vehicles: Linguistic and archaeological considerations of the PIE homeland problem. In Proceedings of the 19th Annual UCLA Indo-European Conference, K. Jones-Bley, M.E. Huld, A. Della Volpe, and M.R. Dexter, eds. (Institute for the Study of Man), pp. 1–59.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., and Li, J.Z. (2012). Genomic patterns of homozygosity in worldwide human populations. Am. J. Hum. Genet. *91*, 275–292.

Perlès, C. (2001). The Early Neolithic in Greece. The First Farming Communities in Europe (Cambridge University Press).

Perlès, C., Quiles, A., and Valladas, H. (2013). Early seventh millennium AMS dates from domestic seeds in the Initial Neolithic at Franchthi Cave (Argolid, Greece). Antiquity *87*, 1001–1015.

Peter, B.M. (2016). Admixture, population structure, and f-statistics. Genetics *202*, 1485–1501.

Philaniotou, O. (2006). Tinos. Archaeology Aegean islands (Melissa Publishing House).

Philaniotou, O. (2008). Naxos, Tsikniades: an early Cycladic cemetery. In Horizon: A Colloquium on the Prehistory of the Cyclades, N. Brodie, J. Doole, G. Gavalas, and C. Renfrew, eds. (McDonald Institute Monographs), pp. 195–207.

Philaniotou, O. (2017a). Figurines from the cemetery of Tsikniades, Naxos. Early Cycladic Sculpt. Context 263–271.

Philaniotou, O. (2017b). Figurines from Potamia on Epano Koufonisi (Pandelis Tsavaris property). In Early Cycladic Sculpture in Context (Oxbow Books Limited), p. 171.

Plantinga, T.S., Alonso, S., Izagirre, N., Hervella, M., Fregel, R., van der Meer, J.W.M., Netea, M.G., and de la Rúa, C. (2012). Low prevalence of lactase persistence in Neolithic South-West Europe. Eur. J. Hum. Genet. *20*, 778–782.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv. https://doi.org/10.1101/201178.

Poznik, G.D., Xue, Y., Mendez, F.L., Willems, T.F., Massaia, A., Wilson Sayres, M.A., Ayub, Q., McCarthy, S.A., Narechania, A., Kashin, S., et al.; 1000 Genomes Project Consortium (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat. Genet. *48*, 593–599.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A., and Feldman, M.W. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. *16*, 1791–1798.

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

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., and Robert, C.P. (2016). Reliable ABC model choice via random forests. Bioinformatics *32*, 859–866.

Pullen, D. (2008). The Early Bronze Age in Greece. In The Cambridge Companion to the Aegean Bronze Age, C.W. Shelmerdine, ed. (Cambridge University Press), pp. 19–46.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

R Development Core Team (2018). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). https://www.r-project.org/.

Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. Science *345*, 1255832.

Rahmstorf, L. (2015). An introduction to the investigation of archaeological textile tools. In Tools, Textiles and Contexts: Textile Production in the Aegean and Eastern Mediterranean Bronze Age, E. Andersson-Strand and M.-L. Nosch, eds. (Oxbow Books), pp. 1–23.

Ramsey, C.B. (2017). Methods for Summarizing Radiocarbon Datasets. Radiocarbon *59*, 1809–1833.

Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., and Rasmussen, S. (2019). Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. Cell *176*, 295–305.e10.

Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M.O., et al. (2015). Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. Cell *163*, 571–582.

Reimer, P.J., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Bronk Ramsey, C., Buck, C.E., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. Radiocarbon *55*, 1869–1887.

Renaud, G., Hanghøj, K., Willerslev, E., and Orlando, L. (2017). gargammel: a sequence simulator for ancient DNA. Bioinformatics *33*, 577–579.

Renfrew, C. (1972). The Emergence of Civilisation: the Cyclades and the Aegean in the Third Millennium BC (Oxbow Books).

Renfrew, C. (1988). Archaeology and Language: the Puzzle of Indo-European origins (Cambridge University Press).

Renfrew, C. (1989). The Origins of Indo-European Languages. Sci. Am. *261*, 106–115.

Renfrew, C. (2000). At the Edge of Knowability: Towards a Prehistory of Languages. Camb. Archaeol. J. *10*, 7–34.

Renfrew, C. (2011). The Emergence of Civilisation: the Cyclades and the Aegean in the Third Millennium BC (Oxbow Books).

Renfrew, C. (2018). Inventing the Final Neolithic. In Communities in Transition: The Circum-Aegean Area during the 5th and 4th Millennia BC., S. Dietz, F. Mavridis, Ž. Tanckosic, and T. Takaoglu, eds. (Oxbow Books), pp. 3–11.

Renfrew, C., and Bahn, P.G. (2014). The Cambridge World Prehistory 3 (Cambridge University Press).

Ringe, D. (2006). From Proto-Indo-European to Proto-Germanic: A Linguistic History of English (Oxford University Press).

Rubinacci, S., Ribeiro, D.M., Hofmeister, R., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nature Genetics *53* (1), 120–126. https://doi.org/10.1038/s41588-020-00756-0.

Rutter, J.B. (1979). Ceramic Change in the Aegean Early Bronze Age: the Kastri Group, Lefkandi I, and Lerna IV: a Theory Concerning the Origin of Early Helladic II Ceramics (University of California).

Rutter, J.B. (1983). Some observations on the Cyclades in the later third and early second millennia. Am. J. Archaeol. *87*, 69–76.

Rutter, J.B. (1993). Review of Aegean prehistory II: The prepalatial Bronze Age of the southern and central Greek mainland. Am. J. Archaeol. *97*, 745–797.

Rutter, J.B. (2001). Review of Aegean Prehistory II: The Prepalatial Bronze Age of the Southern and Central Greek Mainland" and "Addendum: 1993-1999". In Aegean Prehistory, A. Review and T. Cullen, eds. (University of Michigan), pp. 148–155.

Şahoğlu, V. (2005). The Anatolian trade network and the Izmir region during the Early Bronze Age. Oxf. J. Archaeol. *24*, 339–361.

Şahoğlu, V. (2008). Liman Tepe and Bakla Tepe: New evidence for the relations between the Izmir region, the Cyclades and the Greek Mainland during the Late Fourth and Third Millennia BC. In Proceedings of the International Symposium the Aegean in the Neolithic, Chalcolithic and the Early Bronze Age, R.T.H. Erkanal, H. Hauptmann, and V. Şahoğlu, eds. (Ankara University Press), pp. 483–501.

Sampson, A. (1981). The Neolithic and Early Bronze Age I in Euboea (Archive of Euboean Studies).

Sampson, A. (1985). Manika I: an Early Helladic Town in Chalkis (Archive of Euboean Studies).

Sampson, A. (1988). Manika II. The Early Bronze Age settlement and cemetery (Archive of Euboean Studies).

Sampson, A. (1991). Some chronological problems of the end of the Neolithic and the Early Bronze Age. In Archaeometry: Proceedings of the 25th International Symposium Held in Athens, I. Maniatis, ed. (Elsevier), pp. 709–715.

Schachermeyr, F. (1955). Prähistorische Kulturen Griechenlands (Druckenmüller).

Scheu, A., Powell, A., Bollongino, R., Vigne, J.D., Tresset, A., Çakırlar, C., Benecke, N., and Burger, J. (2015). The genetic prehistory of domesticated cattle from their origin to the spread across Europe. BMC Genet. *16*, 54.

Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics *27*, 592–593.

Schroeder, H., Ávila-Arcos, M.C., Malaspinas, A.-S., Poznik, G.D., Sandoval-Velasco, M., Carpenter, M.L., Moreno-Mayar, J.V., Sikora, M., Johnson, P.L.F., Allentoft, M.E., et al. (2015). Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. Proc. Natl. Acad. Sci. USA *112*, 3669–3673.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A., Willerslev, E., Krogh, A., and Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. BMC Genomics *13*, 178.

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

Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., et al. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science *290*, 1155–1159.

Semino, O., Magri, C., Benuzzi, G., Lin, A.A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P.J., et al. (2004). Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am. J. Hum. Genet. *74*, 1023–1034.

Shelmerdine, C.W. (2008). The Cambridge Companion to the Aegean Bronze Age (Cambridge University Press).

Shelton, K. (2010). Citadel and settlement: a developing economy at Mycenae, the case of Petsas House. In Political Economies of the Aegean Bronze Age, D.J. Pullen, ed. (Oxbow Books), pp. 184–204.

Shelton, K. (2012). Mainland Greece. In The Oxford Handbook of the Bronze Age Aegean, E.H. Cline, ed. (Oxford University Press), pp. 139–148.

Sherratt, A. (1981). Plough and Pastoralism: Aspects of the Secondary Products Revolution (Cambrdge University Press).

Sisson, S.A., Yanan, F., and Beaumont, M. (2018). Handbook of Approximate Bayesian Computation (Chapman and Hall/CRC Press).

Swallow, D.M. (2003). Genetics of lactase persistence and lactose intolerance. Annu. Rev. Genet. *37*, 197–219.

Szécsényi-Nagy, A., Brandt, G., Haak, W., Keerl, V., Jakucs, J., Möller-Rieker, S., Köhler, K., Mende, B.G., Oross, K., Marton, T., et al. (2015). Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. Proc. Biol. Sci. *282*, 20150339.

Tavaré, S., Balding, D.J., Griffiths, R.C., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. Genetics *145*, 505–518.

Televantou, C.A. (2008). Strofilas: a Neolithic settlement on Andros. In Horizon: A Colloquium on the Prehistory of the Cyclades, N. Brodie, J. Doole, G. Gavalas, and C. Renfrew, eds. (McDonald Institute Monographs), pp. 43–53.

Theocharis, D. (1959). From the Prehistoric Euboea. Arch. European Stud. *6*, 279–328.

Theocharis, D. (1973). Neolithic Greece (National Bank of Greece).

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. *39*, 31–40.

Todorova, K. (1995). The Neolithic, Eneolithic and transitional period in Bulgarian prehistory. In Prehistoric Bulgaria, D.W. Bailey, I. Panayotov, and S. Alexandrov, eds. (Monographs in World Archaeology), pp. 79–98.

Tomkins, P., and Schoep, I. (2012). Crete. In The Oxford Handbook of the Bronze Age Aegean, E.H. Cline, ed. (Oxford University Press), pp. 66–76.

Treuil, R., Darcque, P., Poursat, J.-C., and Touchais, G. (1989). Les civilisations égéennes du Néolithique et de l'Age du Bronze (Presses universitaires de France).

Triantaphyllou, S. (2001). A bioarchaeological approach to prehistoric cemetery populations from western and central Greek Macedonia (British Archaeological Reports International Series).

Triantaphyllou, S. (2008). Living with the dead: a re-consideration of mortuary practices in the Greek Neolithic. In Escaping the Labyrinth: The Cretan Neolithic in Context, V. Isaakidou and P. Tomkins, eds. (Oxbow Books), pp. 139–157.

Triantaphyllou, S. (2017). Oso psila ki an anebeis lexi min peis megali, 'po choma se eftiaxe o theos ki ekeia girizeis pali. In The Pre- and Protopalatial Cemetery in Context. 2nd Petras Symposium, Athens-Greece, 14/02/2015-15/02/2015, M. Tsipopoulou, ed. (Monographs of the Danish Institute at Athens), pp. 271–290.

Triantaphyllou, S., Nikita, E., and Kador, T. (2015). Exploring mobility patterns and biological affinities in the southern Aegean: first insights from early Bronze Age eastern Crete. Annu. Br. Sch. at Athens *110*, 3–25.

Triantaphyllou, S., Tsipopoulou, M., and Betancourt, P.-P. (2018). Kephala Petras Siteias; human bones and funerary practices at the EM cave and MM cemetery. In Proceedings of the 11th International Cretan Conference, Rethymno, 21–27 October 2011, pp. 203–215.

Tsartsidou, G., and Kotsakis, K. (2020). Grinding in a hollow? Phytolith evidence for pounding cereals in bedrock mortars at Paliambela Kolindros, an Early Neolithic site in Macedonia, North Greece. Archaeol. Anthropol. Sci. *12*, 173.

Tsipopoulou, M. (1999). Before, During, After: Architectural phases of the Palatial Building at Petras, Siteia. In Meletemata: Studies in Aegean Archaeology Presented to Malcolm H. Wiener as He Enters His 65th Year, P.P. Betancourt, V. Karageorghis, R. Laffineur, and W.D. Niemeier, eds. (Aegaeum), pp. 847–859.

Tsipopoulou, M. (2002). Petras, Siteia: the Palace, the town, the hinterland and the Protopalatial Background. In Monuments of Minos, Rethinking the Minoan Palaces, Proceedings of the International Workshop "Crete of the Hundred Palaces?", J. Driessen, I. Schoep, and R. Laffineur, eds. (Aegaeum), pp. 133–144.

Tsipopoulou, M. (2010). A prepalatial burial rock-shelter at Petras, Siteia. A preliminary report. In Archaeological Work in Crete I, Proceedings of the 1st Meeting, (Rethimno, 28-30 November 2008), M. Andrianakis and I. Tzachali, eds. (University of Crete), pp. 121–133.

Tsipopoulou, M. (2012a). Burial Cave at Petras Siteias. First announcement. Archaeol. Work. Crete *I*, 121–133.

Tsipopoulou, M. (2012b). The Prepalatial-early Protopalatial cemetery at Petras, Siteia: a diachronic symbol of social coherence, Petras-Siteia. In Petras, Siteia – 25 Years of Excavations and Studies, M. Tsipopoulou, ed. (Monographs of the Danish Institute at Athens), pp. 117–131.

Tsipopoulou, M. (2017). Documenting sociopolitical changes in Pre-and Protopalatial Petras: the house tomb cemetery. In Petras, Siteia: The Pre-and Proto-Palatial Cemetery in Context, M. Tsipopoulou, ed. (Monographs of the Danish Institute at Athens), pp. 57–102.

Tsirtsoni, Z. (2016a). The chronological framework in Greece and Bulgaria between the late 6th and the early 3rd millennium BC, and the "Balkans 4000" project. In The Human Face of Radiocarbon (MOM Éditions), pp. 13–39.

Tsirtsoni, Z. (2016b). The Human Face of Radiocarbon (MOM Éditions).

Tsountas, C., and Manatt, J.I. (1897). The Mycenaean age: a study of the monuments and culture of pre-Homeric Greece (Houghton Mifflin).

van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum. Mutat. *30*, e386–e394.

Vavouranakis, G. (2015). Ikona kai Arxaiologia (Hellenic Academic Libraries).

Veeramah, K.R., Rott, A., Groß, M., van Dorp, L., López, S., Kirsanow, K., Sell, C., Blöcher, J., Wegmann, D., Link, V., et al. (2018). Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. Proc. Natl. Acad. Sci. USA *115*, 3494–3499.

Ventris, M., and Chadwick, J. (1953a). Documents in Mycenaean Greek,Three Hundred Selected Tablets from Knossos, Pylos and Mycenae with Commentary and Vocabulary (Cambridge University Press).

Ventris, M., and Chadwick, J. (1953b). Evidence for Greek Dialect in the Mycenaean Archives. J. Hell. Stud. *73*, 84–103.

Ventris, M., and Chadwick, J. (1956). Documents in Mycenaean Greek (Cambridge University Press).

Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Daly, K.G., Maisano Delser, P., Hare, A.J., Burger, J., Collins, M.J., Kehati, R., et al. (2019). Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science *365*, 173–176.

Voutsaki, S. (2012). Mainland Greece. In The Oxford Handbook of the Bronze Age Aegean, E.H. Cline, ed. (Oxford: Oxford University Press), pp. 53–65.

Wace, A.J.B. (1957). Foreword. In Documents in Mycenaean Greek, M. Ventris and J. Chadwick, eds. (Cambridge University Press), p. xvii.

Wace, A.J.B., and Blegen, C.W. (1916). The pre-Mycenaean pottery of the mainland. Annu. Br. Sch. at Athens *22*, 175–189.

Wakeley, J. (2008). Coalescent Theory: An Introduction (W.H. Freeman).

Walsh, S., Liu, F., Ballantyne, K.N., van Oven, M., Lao, O., and Kayser, M. (2011). IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci. Int. Genet. *5*, 170–180.

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., and Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci. Int. Genet. *7*, 98–115.

Walsh, S., Chaitanya, L., Clarisse, L., Wirken, L., Draus-Barini, J., Kovatsi, L., Maeda, H., Ishikawa, T., Sijen, T., de Knijff, P., et al. (2014). Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci. Int. Genet. *9*, 150–161.

Walsh, S., Chaitanya, L., Breslin, K., Muralidharan, C., Bronikowska, A., Pospiech, E., Koller, J., Kovatsi, L., Wollstein, A., Branicki, W., et al. (2017). Global skin colour prediction from DNA. Hum. Genet. *136*, 847–863.

Walter, H., and Felten, F. (1981). Die vorgeschichtliche Stadt. Befestigungen, Häuser, Funde. Alt-Ägina*III,1* (P. von Zabern).

Weiss, H. (2000). Beyond the Younger Dryas, Collapse as adaptation to abrupt climate change in ancient West Asia and the Eastern Mediterranean. In Environmental Disaster and the Archaeology of Human Response, G. Bawden and R.M. Reycraft, eds. (Maxwell Museum of Anthropology), pp. 75–98.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44*, W58–63.

Weninger, B., Clare, L., Rohling, E., Bar-Yosef, O., Böhner, U., Budja, M., Bundschuh, M., Feurdean, A., Gebe, H.G., Jöris, O., et al. (2009). The Impact of Rapid Climate Change on Prehistoric Societies during the Holocene in the Eastern Mediterranean. Doc. Praehist. *36*, 7.

Whittle, A. (1997). Moving On and Moving Around: Neolithic Settlement Mobility (Oxbow Books).

Wiener, M.H. (2013). "Minding the Gap": Gaps, Destructions, and Migrations in the Early Bronze Age Aegean. Causes and Consequences. Am. J. Archaeol. *117*, 581.

Wilson, D. (2008). Early Prepalatial Crete. In The Cambridge Companion to the Aegean Bronze Age, C.W. Shelmerdine, ed. (Cambridge University Press), pp. 77–104.

Windler, A., Thiele, R., and Müller, J. (2013). Increasing inequality in Chalcolithic Southeast Europe: the case of Durankulak. J. Archaeol. Sci. *40*, 204–210.

Wollstein, A., Lao, O., Becker, C., Brauer, S., Trent, R.J., Nürnberg, P., Stoneking, M., and Kayser, M. (2010). Demographic history of Oceania inferred from genome-wide data. Curr. Biol. *20*, 1983–1992.

Wong, W., Jiang, B., Wu, T., and Zheng, C. (2018). Learning Summary Statistic for Approximate Bayesian Computation via Deep Neural Network. Stat. Sin. *27*, 1595–1618.

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

Xanthoudides, S.A. (1924). The Vaulted Tombs of Mesará: An Account of Some Early Cemeteries of Southern Crete (University Press of Liverpool Limited).

Xu, L., Sun, H., Zhang, X., Wang, J., Sun, D., Chen, F., Bai, J., and Fu, S. (2010). The -22018A allele matches the lactase persistence phenotype in northern Chinese populations. Scand. J. Gastroenterol. *45*, 168–174.

Yakar, J. (2012). Anatolian Chronology and Terminology. In The Oxford Handbook of Ancient Anatolia: (10,000-323 BCE), S. Steadman and G. McMahon, eds. (Oxford University Press), pp. 6–18.

Yengo, L., Wray, N.R., and Visscher, P.M. (2019). Extreme inbreeding in a European ancestry sample from the contemporary UK population. Nat. Commun. *10*, 3719.

Zapheiropoulou, P. (2008). Early Bronze Age cemeteries of the Kampos Group on Ano Kouphonisi. In Horizon: A Colloquium on the Prehistory of the Cyclades, N. Brodie, J. Doole, G. Gavalas, and C. Renfrew, eds. (McDonald Institute Monographs), pp. 183–194.

Ziota, C. (2007). Tafikes Praktikes kai Koinonies tis Epochis tou Chalkou sti Ditiki Makedonia. Ta Nekrtafeia stin Koilada kai stis Goules Kozanis (Aristotle University of Thessaloniki).

Ziota, C., and Triantaphyllou, S. (2004). Early Bronze age burial practices andskeletal populations: a case study from Western Macedonia. In Spatial Analysis of Funerary Areas, L. Šmejda and J. Turek, eds. (University of West Bohemia, Department of Archaeology), pp. 38–47.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Ancient skeletal element	This study	Sample ID: Mik15
Ancient skeletal element	This study	Sample ID: Pta08
Ancient skeletal element	This study	Sample ID: Kou01
Ancient skeletal element	This study	Sample ID: Kou03
Ancient skeletal element	This study	Sample ID: Log02
Ancient skeletal element	This study	Sample ID: Log04
Ancient skeletal element	This study	Sample ID: AGI03
Ancient skeletal element	This study	Sample ID: AGI04
Ancient skeletal element	This study	Sample ID: AGI05
Ancient skeletal element	This study	Sample ID: AGI06
Ancient skeletal element	This study	Sample ID: AGI07
Ancient skeletal element	This study	Sample ID: AGI08
Ancient skeletal element	This study	Sample ID: AGI09
Ancient skeletal element	This study	Sample ID: AGI10
Ancient skeletal element	This study	Sample ID: AGI11
Ancient skeletal element	This study	Sample ID: AGI12
Ancient skeletal element	This study	Sample ID: KMA01
Ancient skeletal element	This study	Sample ID: KMA02
Ancient skeletal element	This study	Sample ID: KMA04
Ancient skeletal element	This study	Sample ID: KMA05
Ancient skeletal element	This study	Sample ID: KMA06
Ancient skeletal element	This study	Sample ID: KOU04
Ancient skeletal element	This study	Sample ID: KOU05
Ancient skeletal element	This study	Sample ID: KOU06
Ancient skeletal element	This study	Sample ID: KOU07
Ancient skeletal element	This study	Sample ID: LOG01
Ancient skeletal element	This study	Sample ID: LOG08
Ancient skeletal element	This study	Sample ID: MIK01
Ancient skeletal element	This study	Sample ID: MIK02
Ancient skeletal element	This study	Sample ID: MIK03
Ancient skeletal element	This study	Sample ID: MIK04
Ancient skeletal element	This study	Sample ID: MIK05
Ancient skeletal element	This study	Sample ID: MIK06
Ancient skeletal element	This study	Sample ID: MIK07
Ancient skeletal element	This study	Sample ID: MIK08
Ancient skeletal element	This study	Sample ID: MIK09
Ancient skeletal element	This study	Sample ID: MIK10
Ancient skeletal element	This study	Sample ID: MIK12
Ancient skeletal element	This study	Sample ID: MIK13
Ancient skeletal element	This study	Sample ID: MIK19
Ancient skeletal element	This study	Sample ID: MIK20
Ancient skeletal element	This study	Sample ID: MIK21
Ancient skeletal element	This study	Sample ID: MIK22

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ancient skeletal element	This study	Sample ID: PAL04
Ancient skeletal element	This study	Sample ID: PAL05
Ancient skeletal element	This study	Sample ID: PEL01
Ancient skeletal element	This study	Sample ID: PEL02
Ancient skeletal element	This study	Sample ID: PEL03
Ancient skeletal element	This study	Sample ID: PEL04
Ancient skeletal element	This study	Sample ID: PEL05
Ancient skeletal element	This study	Sample ID: PTA02
Ancient skeletal element	This study	Sample ID: PTA09
Ancient skeletal element	This study	Sample ID: TSI01
Ancient skeletal element	This study	Sample ID: TSI03
Ancient skeletal element	This study	Sample ID: TSI04
Ancient skeletal element	This study	Sample ID: XER01
Ancient skeletal element	This study	Sample ID: XER02
Ancient skeletal element	This study	Sample ID: XER03
Ancient skeletal element	This study	Sample ID: XER06
Ancient skeletal element	This study	Sample ID: XER07
Ancient skeletal element	This study	Sample ID: XER08
Ancient skeletal element	This study	Sample ID: XER09
Ancient skeletal element	This study	Sample ID: XER10
Ancient skeletal element	This study	Sample ID: XER11
Ancient skeletal element	This study	Sample ID: XER13
Ancient skeletal element	This study	Sample ID: XER15
Ancient skeletal element	This study	Sample ID: XER16
Chemicals, peptides, and recombinant proteins		
AccuPrime Pfx SuperMix	Thermo Scientific	Cat#12344040
AmpliTaq Gold Buffer II (10x)	Life Technologies	Cat#4311816
AmpliTaq Gold DNA Polymerase	Life Technologies	Cat#4311816
ATP Solution (100 mM)	Life Technologies	Cat#R0441
Bovine Serum Albumin (BSA) (20 mg/ml)	Roche Diagnostics	Cat#10711454001
Bradford fish gel (Gelatin from cold water Fish)	Sigma	G7041-100G
Bst Polymerase, Large Fragment (8 U/μl)	New England Biolabs GmbH	Cat#M0275S
dNTPs (each 10 mM)	QIAGEN, Hilden, Germany	Cat#201901
dNTPs (each 25 mM)	Agilent Technologies	Cat#600677
EDTA (0.5 M, pH 8.0)	Ambion/Applied Biosystems, Life Technologies	Cat#AM9262
HCl	Fisher Scientific	H/1200/PB17
Herculase II Fusion DNA Polymerase	Agilent Technologies	Cat#600677
Herculase II Reaction Buffer	Agilent Technologies	Cat#600677
IA cane	Iso Analytical	R006
IAEA 600	International Atomic Energy Agency	IAEA600
IAEA N2	International Atomic Energy Agency	IAEA N2
MgCl2 (25 mM)	Life Technologies	Cat#4311816
NEBNext End Repair Enzyme Mix	New England Biolabs GmbH	Cat#E6050L
NEBNext End Repair Reaction Buffer (10X)	New England Biolabs GmbH	Cat#E6050L
Nuclease-free H2O	Life Technologies	Cat#AM9932

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PEG-4000	Thermo Scientific	Cat#EL0011
Phenol/chloroform/ isoamylalcohol (25:24:1)	Roth, Karlsruhe, Germany	Cat#A156.1
Proteinase K	Roche Diagnostics, Mannheim, Germany	Cat#3115828001
Sodium N-lauryl sarcosinate	Merck Millipore, Merck KGaA, Darmstadt, Germany	Cat#428010
T4 DNA Ligase (5 U/μl)	Thermo Scientific	Cat#EL0011
T4 DNA Ligase Buffer (10X)	Thermo Scientific	Cat#EL0011
T4 DNA Polymerase (5 U/μl)	Thermo Scientific	Cat#EP0062
T4 Polynucleotide Kinase	Invitrogen	Cat#EK0032
Tango Buffer (10x)	Life Technologies	Cat#BY5
ThermoPol Buffer (10X)	New England Biolabs GmbH	Cat#M0275S
Trichlormethan/Chloroform	Roth, Karlsruhe, Germany	Cat#3313.1
Tris-EDTA	Sigma-Aldrich	Cat#T9285
Tris-HCl (1M, pH 8.0)	Life Technologies	Cat#15568025
USER TM enzyme	New England Biolabs GmbH	Cat#M5505L
Critical commercial assays		
Agilent 2100 Expert Bioanalyzer System and High Sensitivity DNA Analysis Kit	Agilent Technologies	Cat#5067-4626 (kit)
Qubit Fluorometric quantitation and dsDNA HS Assay Kit	Invitrogen	Cat#Q32854 (kit) Cat#Q32856 (tubes)
Deposited data		
Sequencing data (this study)	European Nucleotide Archive	https://www.ebi.ac.uk/ena/browser/view/PRJEB37782
351 individuals, whose data derive from the '1240k' SNP capture assay	Lazaridis et al., 2017	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/MinMyc.tar.gz
225 ancient '1240k' SNP captured individuals	Mathieson et al., 2018	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/Genomic_Hist_SE_Europe_Mathieson.tar.gz
2,068 worldwide modern individuals (genotyped on the Human Origins SNP array)	Lazaridis et al., 2016	https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/NearEastPublic.tar.gz
1000 Genomes	International HapMap 3 Consortium et al., 2010	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data
Bar8	Hofmanová et al., 2016	https://www.ebi.ac.uk/ena/browser/view/SAMEA3672571
KK1	Jones et al., 2015	https://www.ebi.ac.uk/ena/browser/view/SAMEA3609088
S_Greek-1, S_Greek-2, B_Crete-1, B_Crete-2 from Simons Genome Diversity Project	Mallick et al., 2016	https://www.ebi.ac.uk/ena/browser/view/PRJEB9586
Sidelkino	de Barros Damgaard et al., 2018	https://www.ebi.ac.uk/ena/browser/view/ERR2572846
YamnayaKaragash_EBA	de Barros Damgaard et al., 2018	ftp://ftp.sra.ebi.ac.uk/vol1/run/ERR257/ERR2572845/Yamnaya.realigned.calmd.readsadded.bam
Oligonucleotides		
IS4, IS5, IS6 and IS7	(Meyer and Kircher, 2010) IDT, Leuven, Belgium	N/A
MYBait kit	https://arborbiosci.com/genomics/targeted-sequencing/mybaits/mybaits-custom-dna-seq/	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
P5 and P7	(Meyer and Kircher, 2010) IDT, Leuven, Belgium	N/A
SureSelectXT in-solution target enrichment kit	Agilent Technologies (custom design); Gnirke et al., 2009	N/A
Software and algorithms		
ABC-DL	Mondal et al., 2019	https://github.com/oscarlao/ABC_DL
abc package R	Csilléry et al., 2010	https://cran.r-project.org/web/packages/abc/index.html
AdapterRemoval v2.1.7	Schubert et al., 2016	https://adapterremoval.readthedocs.io/en/latest/
Admixtools v. 5.1	Patterson et al., 2012	https://github.com/DReichLab/AdmixTools
ADMIXTURE	Alexander et al., 2009	http://dalexander.github.io/admixture/download.html
ANGSD v. 0.921	Korneliussen et al., 2014	http://www.popgen.dk/angsd/index.php/Change_log
ape v5.3	Paradis and Schliep, 2019	http://ape-package.ird.fr
ARLEQUIN version 3.5.2.2	Excoffier and Lischer, 2010	http://cmpg.unibe.ch/software/arlequin35
bwa v0.7.15	Li and Durbin, 2010	http://bio-bwa.sourceforge.net/
BamDamage v. 20140602	Malaspinas et al., 2014	https://savannah.nongnu.org/projects/bammds
cmdscale function (in R)	R Development Core Team, 2018	https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cmdscale
contaminationX	(Moreno-Mayar et al., 2020)	https://github.com/sapfo/contaminationX
contamMix v. 1.0-10	Fu et al., 2013	https://science.umd.edu/biology/plfj
Eigensoft v. 7.2.1	Patterson et al., 2006	https://github.com/DReichLab/EIG
Encog v3.4	Heaton, 2015	https://github.com/jeffheaton/encog-dotnet-core/releases/tag/v3.4
fastqc v0.11.5	Andrews, 2010	https://www.bioinformatics.babraham.ac.uk/projects/fastqc
fastSimcoal2	(Excoffier and Foll, 2011)	http://cmpg.unibe.ch/software/fastsimcoal2
fastx_trimmer v0.0.13.2	Hannon lab 2010	http://hannonlab.cshl.edu/fastx_toolkit
FigTree v1.4.4	Andrew Rambaut	https://groups.google.com/g/figtree-discuss/c/-9_1I88HPOA
GATK v3.7	DePristo et al., 2011	https://gatk.broadinstitute.org
GenotypeGVCFs	Poplin et al., 2017	https://gatk.broadinstitute.org/hc/en-us/articles/360046224151-GenotypeGVCFs
GLIMPSE v1.0.1	Rubinacci et al., 2021	https://github.com/odelaneau/GLIMPSE
haploGrep v2.1.19	Weissensteiner et al., 2016	https://haplogrep.i-med.ac.at
HaplotypeCaller	Poplin et al., 2017	https://gatk.broadinstitute.org/hc/en-us/articles/360041415292-HaplotypeCaller
HIrisPLex-S	Chaitanya et al., 2018	https://hirisplex.erasmusmc.nl
phangorn v2.5.5	Schliep, 2011	https://cran.r-project.org/web/packages/phangorn/index.html
PhyML v3.1	Guindon et al., 2010	http://www.atgc-montpellier.fr/phyml/
Picard tools v2.9.0	Broad Institute	http://broadinstitute.github.io/picard/
PLINK 1.9	Purcell et al., 2007	https://zzz.bwh.harvard.edu/plink/plink2.shtml
popHelper	Francis, 2017	http://pophelper.com/
Samtools v. 1.10	Li et al., 2009	https://github.com/samtools/samtools
SeaView v5.0.4	Galtier et al., 1996; Gouy et al., 2010	http://doua.prabi.fr/software/seaview
Snakemake v5.3.0	(Mölder et al., 2021)	https://snakemake.readthedocs.io/en/v5.3.0/
Tablet 1.19.09.03	Milne et al., 2013	https://ics.hutton.ac.uk/tablet/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Amicon Ultra-4 Centrifugal Filter Units, 30kDa	Merck Millipore, Darmstadt, Germany	Cat#UFC803096
Amicon Ultra-15 Centrifugal Filter Units, 50kDa	Merck Millipore, Darmstadt, Germany	Cat#UFC905096
MinElute PCR Purification Kit	QIAGEN, Hilden, Germany	Cat#28006
MSB Spin PCRapace	Invitek, Stratec Molecular, Berlin, Germany	Cat#1020220400
Ezee filters 9ml	Elkay Laboratory Products	Cat#127-3193-000
Tin capsules 4mm x 3.2 mm	OEA Laboratories	Cat#C11130.500P

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Anna-Sapfo Malaspinas, email: annasapfo.malaspinas@unil.ch

Materials availability

This study did not generate new unique reagents.

Data and code availability

The accession number for the bam files of the ancient genomes and nuclear and mtDNA capture data reported in this study is European Nucleotide Archive: PRJEB37782 (https://www.ebi.ac.uk/ena/browser/view/PRJEB37782). ABC-DL code with aDNA simulator is available at https://github.com/olgadolgova/ABC_DL_aDNA. Supplementary information to the present article, in addition to the supplemental tables and figures, are available in Document S1.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological material and ethics permission

Skeletal material from ten archaeological sites representing the Bronze Age civilizations were chosen (Cline, 2012; Dickinson, 1994; Treuil et al., 1989). For a detailed chronology, terminology and archaeological background on Aegean cultures see Document S1.

We produced whole genomes from four archaeological sites (Petras, Manika, Koufonisi, Elati-Logkas) and mtDNA genomes from six archaeological sites (Agios Kosmas, Koufonisi, Manika, Paliambela-Kolindrou, Pella, Xeropigado Koiladas, see Figure 1). The individual samples from two archaeological sites (Koumasa and Tsikniades) yielded no DNA. Information for the archaeological sites and individual samples is available at Document S1.

We were given permission by the Greek Ministry of Culture and Sports to sample and extract DNA as well as to radiocarbon date all human remains mentioned in this study according to Greek law for destructive sampling of archaeological material (N.3028/02).

METHOD DETAILS

Radiocarbon dating

Radiocarbon dates of the petrous bone samples used for WGS, newly reported in this study are summarized in Table S1. We report the uncalibrated ^{14}C age, the calibrated ranges (1 and 2 sigma) using OxCal v4.3.2. (Ramsey, 2017) and INTCAL13 (Reimer et al., 2013), and the C/N ratio and the % carbon as additionally quality criteria for showing collagen preservation (Table S1).

Extraction of collagen from the samples *Mik15, Kou03, Log04* and *Pta08* was performed in the BioArch facility of the University of York before being sent to the Curt-Engelhorn-Zentrum-Mannheim for ^{14}C dating. Collagen was extracted from four human petrous bones using a modified Longin (1971) method as in Kontopoulos et al. (2019). The exterior surfaces of the bone samples were mechanically cleaned using a scalpel. Bone chunks of 300-500 mg were demineralized in 8 mL 0.6 M HCl at 4°C. Samples were agitated twice daily and the acid solution was changed every two days. After demineralization, the supernatant was drained off and samples were rinsed (x3) with distilled water. Gelatinization was carried out by adding 8 mL pH3 HCl, and samples were placed in hot blocks at 80°C for 48h. The supernatant was filtered using Ezee^{TM} filters and was freeze-dried for two days in pre-weighed plastic tubes. Collagen yields (weight in % of dry bone), which are commonly used to distinguish well-preserved from poorly-preserved collagen were estimated using the formula: bone mass (mg)/collagen mass (mg) x 100, where bone mass is the weight of bone chunks after

cleaning the exterior surfaces, and collagen mass is the extracted material that remains following demineralization, gelatinization and filtering.

Samples *Kou01* and *Log02* were sent as a piece of petrous bone to the Curt-Engelhorn-Zentrum Mannheim for ^{14}C dating. Collagen was extracted following standard protocols of the Curt-Engelhorn-Zentrum Mannheim.

Ancient DNA Data Generation

DNA extraction

All pre-PCR sample preparation steps were carried out in the clean room facilities of the Palaeogenetics Group at the Johannes Gutenberg-University Mainz, which is physically separated from any post PCR laboratories as previously described (Scheu et al., 2015). Ancient DNA was extracted using two different extraction methods:

Protocol A: We used 0.5g bone powder as input for DNA extraction and followed a previously published protocol (Hofmanová et al., 2016; Scheu et al., 2015) with minor modifications. For lysis 6mL EDTA (0.5M, pH8), 250μL N-Laurylsarcosine (5%) and 30μL Proteinase K (14-22mg/mL) were added to the powder and left under constant shaking for 48h at 37°C. After a subsequent phenol/chloroform (phenol/chloroform/ isoamyl alcohol 25:24:1) extraction the aqueous phase was transferred to a 50kD Amicon filter unit (Merck Millipore) for clean-up and concentration up to a final volume of 200μL DNA extract. Optionally, an additional pre-lysis step was performed: prior to lysis, the powder was washed for 45min at 37°C using 1mL EDTA, 250μL N-Laurylsarcosine and 30μL 20 Proteinase K, the powder was then pelleted and the supernatant discarded. Extractions I and II of *Kou01*, *Log02* and *Pta08* followed Protocol A. Extraction *Kou01*_II followed Protocol A with an additional pre-lysis step. DNA extractions for the mtDNA capture experiments followed extraction protocol A.

Protocol B: DNA extraction was performed using 0.15 g bone powder. Pre-lysis was performed by adding 1mL EDTA (0.5M, pH8) to the bone powder and incubating the suspension at 37°C for 30min under constant shaking (*Kou03*_IV, *Log04*_III, *Mik15*_I and *Pta08*_III) or for 10min without shaking (*Log02*_III). Afterward, each sample was incubated with 1mL lysis buffer consisting of 950μL EDTA (0.5M, pH8), 20μL Tris-HCl (1M, pH8), 17μL N-Laurylsarcosine (5%) and 13μL Proteinase K (14-22mg/ml) at 37°C for 24h under constant shaking, followed by a concentration and washing step using 1x Tris EDTA on 30kD Amicon filter units (Merck Millipore). The extracts were then purified using silica columns (MinElute PCR Purification Kit). For extract *Log02*_III the lysis step was repeated after 24h by removing the supernatant, adding 1 mL additional lysis buffer to the bone powder pellet and incubating for another 24h, followed by merging both extracts prior to the Amicon filter wash.

Library preparation

DNA extracts were converted into double-indexed sequencing libraries according to Kircher et al. (2012) with modifications. For WGS blunt-end repair was performed using the NEBNext End Repair Module: 20 μL of DNA extract are mixed with NEBNext End Repair Reaction Buffer (10X, 7μL), NEBNext End Repair Enzyme Mix (3.5μL) and nuclease-free water (39.5μL; for a final reaction volume of 70 μL) and incubated for 15 min at 25°C followed by 5 min at 12°C. Hybridized adapters P5 and P7 were used at a concentration of 0.75 μM. The WGS libraries were amplified in 12 PCR parallels each with an individual index combination using AccuPrimeTM Pfx SuperMix and 3 μL fill-in product per parallel (final reaction volume: 25 μL; final primer concentration: 200 nM each). The PCR was performed in 10-11 cycles and the temperature profile followed the manufacturer's recommendations but using an annealing temperature of 60°C, extending for 30 s during each cycle and performing a final elongation step for 5min. The DNA extract of *Log02* was treated with USER™ enzyme prior to library preparation for WGS drastically reducing the effects of DNA deamination (Briggs et al., 2010) (Figure S3): 16,25μl of DNA extract were mixed with 5μl USER™ Enzyme and incubated for 3 h at 37°C (Verdugo et al., 2019). The blunt-end repair step followed immediately. After adaptor fill-in and before library amplification, a sample was taken for qPCR, to infer the number of DNA molecules successfully transferred into sequencing libraries (Hofmanová et al., 2016).

Sequencing

We processed a total of 70 individual bone samples; 15 samples contained no DNA that could be transferred into libraries for sequencing. The remaining 55 genomic libraries (pooled equimolar) were sequenced on an Illumina MiSeq™ platform (50 bp, SE) at StarSEQ GmbH (Mainz, Germany) to measure the human DNA content. Demultiplexing was performed by the sequencing facility. Based on the screening results, we selected six individual samples with high endogenous DNA content (ranging from 12.6% to 55.9%; Table S1) for whole genome shotgun sequencing (*Kou01*, *Kou03*, *Log02*, *Log04*, *Mik15*, and *Pta08*; all from petrous bone samples). DNA sequencing for whole genome sequence data were performed using Illumina's HiSeq2500 at the Lausanne Genomics Technologies Facility (GTF) at the University of Lausanne (*Mik15*, *Pta08*, *Kou01*, *Kou03, Log04*) as well as an Illumina HiSeq3000 at the Next Generation Sequencing Platform (Institute of Genetics) at the University of Bern (*Log02*).

For three individual samples (*Pta08*, *Kou01* and *Log02*), a capture experiment was conducted targeting 5,329 putative neutral, autosomal regions, as well as 388 autosomal SNPs associated with phenotypes of interest (*i.e.*, 5,717 regions in total). These were enriched by in-solution hybridization capture as described in Veeramah et al. (2018). The targeted regions span around 4.9Mb of the nuclear genome, each region ranges from 80 to 1,001 bp, with 79% of them spanning 1,001 bp each. For these captured libraries, a 100 bp single-end rapid run was performed on an Illumina HiSeq2000 sequencer at the Institute for molecular genetics, genetic engineering research and consulting (IMSB) (now Molecular Genetics and Genome Analysis Group) at the University Mainz, Germany.

In addition to nuclear data, the mitochondrial genome was captured for 11 samples using Agilent's SureSelect™ in-solution target enrichment kit (custom design) (Gnirke et al., 2009) as detailed in Hofmanová et al. (2016). Prior to enrichment 2-4 libraries originating

from 2-3 independent extractions (Table S1) were prepared for each sample according to Kircher et al., (2012) as described in Hofmanová et al. (2016) and pooled equimolarly. Sequencing and demultiplexing of the mtDNA capture samples were performed on an Illumina MiSeq™ platform (50 bp, SE) at StarSEQ GmbH (Mainz, Germany).

Processing and Mapping of the Raw Sequencing Data

Sequencing reads from screening, WGS, nuclear and mtDNA capture experiments, as well as those from eight published genomes (Table S1), were mapped with an in-house mapping pipeline implemented in the workflow manager snakemake (Mölder et al., 2021) and consisting of the following steps. First, sequence reads were checked for quality before and after removal of adapters using fastqc version 0.11 (Andrews, 2010). Illumina adapters were removed with AdapterRemoval version 2.1.7 (Schubert et al., 2016) filtering out ambiguous bases (N, –trimns), bases of low quality (phred score $\leq$ 2) from both ends of the reads (—trimqualities), and reads shorter than 30 base pairs (bp, –minlength 30). Reads were aligned to the human reference genome (GRCh37 for the screening, WGS and nuclear capture data; and to the revised Cambridge Reference Sequence (rCRS, NC_012920.1) for the mtDNA capture data) using BWA ALN version 0.7.15 (Li and Durbin, 2010) with disabled seeding (-l 1024) to reduce the effect of post-mortem damage-related error (Schubert et al., 2012). Alignments with a quality score below 30 were discarded with samtools version 1.4 (Li et al., 2009), PCR duplicates were removed at the library level using 'Picard tools' MarkDuplicates version 2.9.0 (http://broadinstitute.github.io/picard) and local realignment around indels was performed using GATK version 3.7 (DePristo et al., 2011). The md flag of the alignments was recomputed using calmd from samtools version 1.4 (Li et al., 2009). For the WGS data, reads were also trimmed at both ends by 5 bp after adapted removal and prior to mapping in order to reduce the impact of post-mortem damage-related error. In this case, reads shorter than 30 bp after 5 bp-trimming were discarded using fastx_trimmer version 0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit). Statistics for the screening, WGS, nuclear and mtDNA capture experiments (including the performance of the nuclear capture experiment) can be found in Table S1.

Authenticity of data

Error analysis

We used the software package ANGSD v. 0.921 (Korneliussen et al., 2014) to assess the overall and type-specific error rates of the WGS and nuclear capture data. This method uses an outgroup species to estimate the expected number of derived alleles in the human genome. Any excess of derived alleles in the sample is then interpreted as the result of errors. We obtained the expected number of derived alleles in the human genome by using a modern individual (SS6004480) from Prüfer et al. (2014) and the chimp genome (mapped to GRCh37) as outgroup species. As expected for aDNA (Orlando et al., 2015), underlying the effects of post-mortem damage, C to T and G to A substitutions accounted for most of the observed errors in the individual samples (Figure S3). For the untrimmed WGS data, we estimated overall error rates from 0.6% to 0.91% (Figure S3). Note that the *Log02* extract was USER™-treated, which drastically decreased the error rate (0.11%). After trimming 5 bp from each end of the reads, the overall error rates were reduced from 0.35% to 0.11% (excl. *Log02*, Figure S3). The trimmed dataset was used for downstream WGS population genetic analyses. For the untrimmed nuclear capture data, the overall error rates were generally higher and ranged from 0.87% to 1.18% (Figure S3), which might reflect the fact that the nuclear capture data were sequenced on an older platform (Illumina HiSeq2500) than the one used for the WGS data (Illumina HiSeq3000).

Ancient DNA damage and fragmentation

The data were inspected for characteristics of ancient DNA (e.g., being fragmented and damaged) using the bamdamage software package (Malaspinas et al., 2014) (Figure S3; Table S1). These features can also be informative about the authenticity of the DNA. The distribution of read lengths was unimodal for each sequencing lane and the reads were shorter than the number of cycles used for sequencing, as expected for ancient DNA (Figure S3; Table S1). Furthermore, the damage patterns across the reads showed increased C to T and G to A substitutions at the read termini, consistent with the degradation pattern for ancient DNA (e.g., Briggs et al., 2010). Consistent with the results from the ANGSD error analysis, the damage pattern for the USER™-treated *Log02* individual was one order of magnitude lower compared to the other individuals.

mtDNA-based contamination estimation

The software package ContamMix v. 1.0-10 (Fu et al., 2013) was used to quantify the level of contamination using the mtDNA chromosome for the WGS and mtDNA capture data. This method assumes that the coverage is high enough to call the true endogenous mtDNA consensus sequence and that the data contains less than 50% contamination. A set of 311 worldwide modern mitochondrial genomes (Green et al., 2008) serves as source of potential contamination. Sequence reads are therefore modeled as a mixture of any of these 311 genomes and the endogenous consensus sequence of unknown proportions. For this analysis, the data were mapped to the whole genome for both the WGS and the mtDNA capture data to reduce the effect of nuclear DNA of mitochondrial origin (NUMT). The data were either trimmed prior to mapping (WGS) or using Contamix (mtDNA capture). The inferred fractions of exogenous mitochondrial sequences correspond to the amount of contamination. We first built a mtDNA consensus sequence using ANGSD v. 0.921 (-doFasta 2), which was then aligned to the panel of 311 mtDNA sequences via mafft v. 7.310 (Katoh and Standley, 2013). Both the alignment and the mapped reads were then used in a MCMC framework to estimate the level of contamination. The Markov chain was run for 10,000 (for the estimates based on trimmed data, see below) to 100,000 (when restricting the analyses to transversions only, see below) iterations (after burn-in) using the default value 0.1 for the hyperparameter alpha of the Dirichlet prior distribution. For the WGS individual samples,

we estimated the contamination with modern mtDNA within the 95% credibility interval to be below 1.5% (Table S1). These findings are in line with estimates from earlier reported ancient samples (e.g., Moreno-Mayar et al., 2018) and suggest that population genetic analyses of the WGS data may not be substantially affected. As damage was high (up to 50% at the end of the reads) for the mtDNA capture data (Table S1), we report three estimates: trimming 5 bp (as for the WGS data) or 10 bp at the read termini and finally only considering transversions. Point estimates for contamination for the mtDNA capture data are below ~5% in most cases (Table S1). The exceptions are three mitochondrial genomes with relatively low coverage: AGI02 (4%–15% contamination after 5 bp trimming), PAL04 (4%–8% contamination after 5 bp trimming), and XER01 (2%–13% contamination after 5 bp trimming) (Table S1). However, in all three cases the point estimates are lower after 10 bp trimming and even lower when considering only transversions, suggesting that damage might not be accounted for properly for those estimates or that the DoC is too low to call the true consensus. In any case, the data for these three samples should be interpreted with care.

X chromosome-based contamination estimates

For this analysis we used contaminationX, an X chromosome-based method described in Moreno-Mayar et al. (2020) (https://github.com/sapfo/contaminationX) to estimate contamination for the two males (*Kou01* and *Pta08*). As described in Moreno-Mayar et al. (2020), the HapMap data (International HapMap 3 Consortium et al., 2010) was used as a reference panel. In brief, this method leverages the fact that the X chromosome (excluding the pseudoautosomal regions) is hemizygous in males. Thus, the occurrence of multiple alleles at a given site on the X chromosome can be attributed to either error or contamination. Base counts are modeled as a function of an error rate estimated from the data, the allele frequencies in the contaminant population, and the contamination fraction, which is estimated with a maximum likelihood optimization. Since estimates obtained through this method have been shown to vary slightly with the assumed ancestry of the contaminant (Moreno-Mayar et al., 2020), we considered the allele frequencies of three HapMap populations: YRI, CHB and CEU. We ran the 'two-consensus' method, considered sites with a depth of coverage $\geq$ 3X and $\leq$ 20X, and obtained a 95% confidence interval through 1,000 block-jackknife replicates following Moreno-Mayar et al., (2020). Results are summarized in Tables 1 and S2. For both individuals, contamination estimates had little variation across panels and were below 1.1%. Given the magnitude of these estimates, we assumed that contamination had little impact on subsequent analyses.

Uniparental markers

Haplogroups

Consensus sequences for the mtDNA were called from the bam files using ANGSD v. 0.921 (Korneliussen et al., 2014), with parameters "-doFasta 2" and "-doCounts 1." We used HaploGrep v2.1.19 (Weissensteiner et al., 2016) to infer the mtDNA haplogroups of the samples (Document S1). In order to determine the Y chromosome haplogroups bcftools v. 1.4 (Li et al., 2009) was used to perform haploid calls (i.e., specifying–ploidy 1) on the Y chromosome of the two males (*Pta08* and *Kou01*). The sequenced data were compared to approximately 60,000 Y chromosome variants reported for the phase 3 of the 1000 Genomes project (Poznik et al., 2016). To call haplogroups for *Kou01* and *Pta08*, the analysis was restricted to sites with a minimum depth of 5 reads and minimum base quality scores of 20. Moreover, analysis was restricted to sites with at least 80% of the reads supporting one of the two alleles at a 1000 Genomes project SNP variant. We then performed a binary tree search as in Schroeder et al. (2015) against the 1000 Genomes Y-phylogeny (Poznik et al., 2016) to determine the Y chromosome haplogroups of the two individuals. Further information about mtDNA and Y chromosome haplogroups can be found in Document S1 and Table S1.

Population structure among mtDNA sequences

To understand the structure among the mtDNA sequences, we first built a phylogenetic tree and then conducted an Analysis of Molecular Variance (AMOVA, Excoffier et al., 1992). The phylogenetic tree (Document S1) was built for the 17 mtDNA sequences in Table S1 (six and eleven from the WGS and mtDNA capture data, respectively) using an additional two San mitochondrial genomes (AY195783 and AY195789) (Mishmar et al., 2003) as outgroups. The sequences were first aligned using SeaView v5.0.4 (Gouy et al., 2010) and then converted to phylip format using the R package ape v5.3 (Paradis and Schliep, 2019). PhyML v3.1 (Guindon et al., 2010) and the R package phangorn 2100 v2.5.5 (Schliep, 2011) were used to determine the best mutational model for our data. In our case, F84 + I + Γ was the best model. The sequences were loaded on PhyML v3.1 and the tree was generated using this model with a BioNJ starting tree, SPR moves, and 10,000 bootstrap replicates. The resulting tree (with all nodes with bootstrap support below 50% collapsed) can be found in Document S1. The resulting tree topology is also mostly coherent with publicly available data (Eupedia, 2018) and haplotypes cluster together. However, samples from the same archeological site do not strictly form clades.

For the AMOVA, the individual samples were pooled into two groups according to their geographic location, culture and time period: EBA Helladic North (3 Pella, 2 Paliambela, 4 Xeropigado Koiladas); and MBA Helladic North (2 Elati-Logkas). See map on Figure 1 in the main text for grouping and locations. Using the mtDNA sequences aligned with SeaView v5.0.4 (Gouy et al., 2010), we carried an AMOVA with Arlequin v.3.5.2.2 (Excoffier and Lischer, 2010), using 10,000 permutations of individual mtDNA sequences between groups to assess significance. Pairwise F_{ST} between pairs of groups was estimated based on pairwise distances of mtDNA sequences, using 10,000 permutations to assess significance. The data type was set to DNA in haplotypic format, with a total of 16,479 variable sites. The results indicate limited differentiation between EBA and MBA in northern Greece, with an explained variance among groups of 2.96% corresponding to an estimated F_{ST} of 0.0296 (p = 0.293).

Reference panels

The six WGS ancient individuals were studied in the context of previously published modern and ancient data of two kinds: i) genotypes/variants (mostly from SNP array capture data), and ii) whole genome sequence data. To serve the purpose of different analyses, different datasets were assembled:

Dataset 0: Three reference panels made available by David Reich's laboratory were combined. These data include 2,068 modern individuals (621,799 SNPs) genotyped on the Human Origins SNP array (https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/NearEastPublic.tar.gz; Lazaridis et al., (2016), 351 ancient individuals (1,150,639 SNPs) whose data derive from the '1240K' SNP capture assay (https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/MinMyc.tar.gz; Lazaridis et al., (2017) as well as 225 ancient individuals (1,233,013 SNPs) from Mathieson et al., (2018) (https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/Genomic_Hist_SE_Europe_Mathieson.tar.gz). The data were downloaded in the PACKEDANCESTRYMAP format and we used EIGENSOFT package 7.2.1 (Patterson et al., 2006; Price et al., 2006) with default parameter settings to merge these datasets and to convert them into the Plink format. This merged dataset contained 621,272 SNPs in total, with 616,427 SNPs being autosomal. Note that the 19 ancient individuals published in Lazaridis et al. (2017) (https://www.ebi.ac.uk/ena/browser/view/PRJEB20914) as well as four additional individuals *B_Crete-1* (*SAMEA3302765*) and *B_Crete-2* (*SAMEA3302625*) (Mallick et al., 2016) (https://www.ebi.ac.uk/ena/browser/view/PRJEB9586); *YamnayaKaragash_EBA* and Sidelkino (de Barros Damgaard et al., 2018), together with the BA Aegean six samples (*Kou01*, *Kou03*, *Mik15*, *Pta08*, *Log02*, *Log04*) were joined from the BAM files after applying the following post-processing steps for these samples. Given the relative low depth of coverage for most of these individuals (Figure S2), rather than calling genotypes, one read at random for each SNP position was sampled, following the steps described below. First, we filtered the BAM files to keep only the aligned bases (minimum base quality of 20) mapping to the position of the SNPs in the reference panel using samtools 1.10 (Li et al., 2009). Second, we randomly sampled one allele at each position per individual and coded it as homozygote reference ("0/0") or homozygote alternative ("1/1"), if it matched the reference or alternative allele in the filtered reference panel, respectively, or as missing data ("./.") otherwise. Third, we merged the resulting table of genotypes with the reference panel. Fourth, given that the modern reference panel contained called genotypes, including heterozygotes, we simulated the sampling of just one read per individual (as done for the sub-set of ancient individuals mentioned above). This was done by randomly converting heterozygous sites into 0/0 or 1/1 with 50% probability. These data (Dataset 0) were further processed in two ways to fit the purpose of different analyses.

Dataset I: For the *f/D*-statistics, we tried to maximize the number of SNPs used in each analysis. Following Lazaridis et al. (2017), this dataset used the following additional filters on Dataset 0: for analyses involving both present-day and ancient data, the so-called "HO" set of 591,642 SNPs was used; for analyses involving ancient individuals only, the so-called "HOIII" set of 1,054,671 SNPs was used (see Lazaridis et al. [2017] for details on the SNP sets) of which 1,054,637 SNPs were retained for this merged dataset.

Dataset II: For this dataset, to discard rare alleles that were likely due to sequencing/mapping errors and to reduce linkage, both minor allele frequency and linkage disequilibrium filters (MAF of 0.05 and $r^2 >= 0.4$) were applied on Dataset 0 using PLINK v1.90 (Purcell et al., 2007) (–maf 0.05–indep-pairwise 200 25 0.4). To avoid biases due to the increased damage patterns for ancient DNA, these filters were only applied to the 2,068 present-day individuals from the Human Origins panel, retaining 165,447 SNPs. These data were then rejoined to the ancient individuals. Moreover, individuals marked as related and outliers were removed and we kept only present-day individuals of primarily Eurasian individuals as determined by an *ADMIXTURE* run with $K = 3$. We selected the present-day individuals with more than 90% Eurasian ancestry component, thus excluding individuals that did not cluster with those from the same continent (Table S2). A panel restricted to this set of individuals was then created by including 1346 individuals (564 ancient and 782 present-day) from Dataset 0. Finally, to allow for a more balanced sample size per population, a maximum of 20 individuals per population were retained removing the individuals with the most missing data. The resulting dataset of 969 individuals (331 ancient and 638 present-day) was used for *ADMIXTURE.* For *MDS*, we further applied a filter to exclude individuals with a proportion of missing haploid genotype calls greater than 0.95 using PLINK v1.90, removing 47 ancient low-quality samples from the panel (Table S2, "Excluded mind 0.95" tab) and removed populations of less relevance for this analysis. In Table S2, all individuals from Dataset 0 that were used in *ADMIXTURE* and *MDS* are listed.

Dataset III: This set included four ancient and four modern publicly available full-genomes that were remapped with the same procedure as the six genomes from this study (Table S1, "Additional genomes" tab): *YamnayaKaragash_EBA* (3,018-2,887 BCE) (de Barros Damgaard et al., 2018), *KK1* (*CHG*; 7,745-7,579 BCE) (Jones et al., 2015), *Bar8* (Neolithic Barçın; 6,122-6,030 BCE) (Hofmanová et al., 2016), *Sidelkino* (*EHG*; 9,386-9,231 BCE) (de Barros Damgaard et al., 2018), *S_Greek-1* and *S_Greek-2* (*SAMEA3302732* and SAMEA3302763; present-day Greeks from Thessaloniki) (Mallick et al., 2016), *B_Crete-1* and *B_Crete-2* (*SAMEA3302765I* and SAMEA3302625; modern Cretans) (Mallick et al., 2016). The modern genomes from this dataset were used for the ROH analysis. The ancient genomes selected had similar or higher mean depth of coverage than the whole genomes sequenced in this study and were therefore more suitable for demographic history reconstruction.

Dataset IV: This dataset (used for ABC-DL) consisted of the four ancient genomes described above and the present-day Greek (*S_Greek-1*) from Dataset III as well as two genomes reported in this study, *Mik15* (to represent Aegean EBA) and *Log04* (to represent Aegean MBA).

After remapping (see above), we followed the g*vcf* methods (Poplin et al., 2017) for variant calling in each of the BAM files. First, every BAM file was called for variants using HaplotypeCaller (Poplin et al., 2017), which produced a gVCF file for a single individual. We then merged all the gVCF files using GenotypeGVCFs (Poplin et al., 2017) with default parameters.

The final VCF file was re-calibrated following Genome Analysis Toolkit (GATK) Variant Quality Score Recalibration (VQSR) recommendations (DePristo et al., 2011). We converted the VCF files to PLINK format with PLINK v1.90 and added the ancestral information from the chimpanzee reference genome (panTro4) as described in Mondal et al., (2019). We marked the reference allele as Ancestral and all the alternative alleles as Derived.

An ''intergenic region SNP set'' was used for this analysis, in an effort to reduce the impact of background selection on demographic inference (Ewing and Jensen, 2016; Johri et al., 2020). All genomic regions containing CpG islands defined in Bock et al. (2007) were excluded, as well as Ensemble genes and their 20 kb upstream and downstream regions. Genomic regions of at least 10kb and separated by at least 100kb were kept. Within each genomic region we concatenate genomic fragments that were at $\leq$ 5kb from each other and that were neither in genes nor in CpG islands. Finally, only SNPs covered by at least 10 reads were included in this dataset, resulting in 7,314 regions (comprising 713 Mb) and 5,268,391 SNPs after filtering. This dataset was converted into PED format with PLINK v1.90.

The data were divided into two subsets (''training'' and ''replication'') using *–thin 0.5* and *–exclude* functions in PLINK v1.90. Each subset accounted for 50% of SNPs from each individual. The ''training'' subset was used for noise injection in the simulation and training processes, and another ''replication'' subset represented the observed data in the ABC-DL framework (Document S1).

Multidimensional Scaling (MDS)

Classical *MDS,* also called Principal Coordinate Analysis (Cox and Cox, 2008), was used to summarize in two dimensions the relationships among our six ancient samples in the context of 259 previously published ancient, and 638 primarily Eurasian genomes selected from Dataset II (Table S2). For each individual and site, the identity-by-state distances for each pair of individuals were calculated, using a minimum base quality filter of 20. These distances were then used to compute the *MDS* projection via the *cmdscale* R function with a custom script (Document S1)and to quantify how much variance is explained by each dimension (Figure 2; Document S1).

ADMIXTURE analysis: Population structure and sex-biased gene flow

We used the software *ADMIXTURE* v1.31 (Alexander et al., 2009) to infer population structure and sex-biased gene-flow.

To infer population structure, we estimated the average genomic ancestry proportions for a total of 638 modern and 331 ancient Eurasian individuals in Dataset II (Table S2) considering an unsupervised *ADMIXTURE* model with *K* ranging from 2 to 6 (Figure S5; Document S1). For the unsupervised *ADMIXTURE* analyses, ten independent replicates were run and the one with the lowest cross-validation (CV) error was selected. The ''haploid'' option was used as the data included pseudo-haploid data, generated by randomly sampling a single read for each locus. The R package popHelper (Francis, 2017) was used to visualize the estimated admixture proportions for each individual.

To assess sex-biased gene flow, *ADMIXTURE* was run considering a supervised model with two known sources ($K = 2$) for each studied period. When analyzing the EBA Aegeans, we considered *Anatolia_N* (n = 26) as proxy for the Anatolian Neolithic-like ancestry and merged *Iran_N* (n = 6) with *CHG* (n = 2) to represent the *Iran_N*/*CHG*-like ancestry. For the MBA Aegeans, *Anatolia_N* (n = 26) and *Steppe_EMBA* (n = 27) were used as proxies for the Anatolian Neolithic-like and Steppe-like ancestry respectively. For this analysis, we considered all of the SNPs on the autosomes from Dataset II (165,402 SNPs) and retrieved pseudo-haploid data for 8,133 SNPs on the X chromosome, following the same filtering criteria. As the X chromosome has fewer SNPs (8,133), 100 replicates of ancestry proportion estimates for the autosomes considering 8,133 autosomal SNPs at random were generated. The results are presented in a violin plot corresponding to the distribution (across the 100 replicates) of the autosomal ancestries for each individual (Figure 5). For this analysis, we also used the ''haploid'' option in *ADMIXTURE* and pseudo-haploid data as input.

f_3/D-statistics

Outgroup f_3-statistics (Patterson et al., 2012) were computed to explore the broad genetic affinities between individuals of interest and present-day populations in Dataset I. Specifically, f_3-statistics of the form f_3*(Yoruba; Y, X)* – where *X* represents one of the present-day populations included in Dataset I and *Y* represents an ancient or present-day individual of interest – were computed (Figure S4). In this case $Y \in$ {*Anatolia_N*, *Greece_N*, *Mik15*, *Pta08*, *Kou03*, *Anatolia_BA*, *Kou01*, *Minoan_Odigitria*, *Log04*, *Log02*, *Minoan_Lasithi*, *Mycenaean*, *Greek*, *Crete, Cypriot*}. In Figure S4, the geographic distribution of the f_3-statistics computed for each population *X* is shown.

D-statistics of the form *D(Anatolia_N, X; Y, Mota)* were computed using Dataset I (Figure S6). *Mota* is an ancient Ethiopian (Gallego Llorente et al., 2015).

For both statistics, standard errors were estimated through a weighted block jackknife approach over approximately 5-Mb blocks. For *D*-statistics, absolute Z-scores greater than 3.3 (corresponding to a *p-value* < 0.001) were regarded as statistically significant.

qpWave/qpAdm analysis

The *qpWave/qpAdm* (Haak et al., 2015; Lazaridis et al., 2017) framework was used to test for the number of migration waves of ancestry and to estimate the admixture proportion of a *Test* population based on Dataset I (Tables 3, S3, and S5; Document S1).

Here a set of "Left" populations (*Test* population and potential source populations) together with a set of "Right" populations (diverse outgroups) allowed for the testing of the number of waves of ancestry from "Right" to "Left" populations, and for estimation of the admixture proportions of the *Test* population. Following Lazaridis et al., (2017), two sets of outgroups were used, one using early Neolithic sources and HG (16 populations, ultimate sources). Note that the African individual *Mota* was used as fixed outgroup.

All: Mota, AfontovaGora3, Anatolia_N, CHG, EHG, ElMiron, GoyetQ116-1, Iran_N, Kostenki14, Levant_N, MA1, Natufian, Ust_Ishim, Vestonice16, Villabruna, WHG.

Likewise, we also defined a second set (proximate sources) by adding younger populations down to the Bronze Age into the "All" set of populations. This may allow for the identification of simpler models underlying the likely complex admixed populations.

All+: All (see above) ∪ Anatolia_ChL, Armenia_ChL, Armenia_EBA, Armenia_MLBA, Europe_LNBA, Europe_MNChL, Iberia_BA, Iran_ChL, Levant_BA, Steppe_EMBA, Steppe_MLBA.

The Left set is chosen to include the *Test* population (for which the admixture proportions are being modeled) and *N* populations from the All (or All+) set. The Right set is then All \Left, testing against the maximal set of Right outgroups. To evaluate the relatedness of the BA Greek individuals, we also added our individual samples, as well as the Minoans and Mycenaeans from Lazaridis et al. (2017), and BA Balkan individual samples (Mathieson et al., 2018) as potential sources to the Left set. We further split the BA Balkan samples into EBA and LBA since the amount of reported Steppe admixture is increased in individuals from LBA (Mathieson et al., 2018). We infer the rank = N-1 using *qpWave* and estimate the admixture proportion for the test populations using *qpAdm*. We show only feasible admixture proportions (in the interval [0,1]) and use a significance threshold of $\alpha = 0.05$ to reject models.

To make our results comparable to Lazaridis et al. (2017), we used the HO set of 591,642 SNPs for joined analyses of modern and ancient data and the HOIII set of 1,054,671 SNPs (1,054,637 SNPs retained after merging datasets) for analyses considering only ancient individuals (Dataset I).

ROH analysis

ROH were computed for the four present-day Greek genomes described in Dataset III, alongside the six BA Aegean genomes (*Kou01*, *Kou03*, *Log02*, *Log04*, *Mik15*, *Pta08*). The data were first imputed by extracting all bi-allelic SNPs from 1000 Genomes phase 3 (positions and alleles) (International HapMap 3 Consortium et al., 2010). Second, we called genotypes in the form of genotypes likelihoods at all these variant sites in the ten genomes (four present-day Greek genomes and six BA Aegean genomes). For this step, we used *bcftools v1.8* constraining the calling model to the positions and alleles extracted at the first step. Third, we imputed genotypes and estimated haplotypes using *GLIMPSE v1.0.1* (Rubinacci et al., 2021) using 1000 Genomes phase 3 as reference panel within overlapping chunks of ~2Mb. The imputation outcome takes the form of haplotype calls, which are essentially a phased version of the most likely genotypes. The imputed dataset contained 43,258,118 SNPs.

Runs of homozygosity were called on the imputed haplotype data of the 10 individuals, either using all the imputed variants (total: 43,285,118 SNPs) or a subset of the variants excluding transitions (total: 13,542,104 SNPs). ROH were identified using PLINK 1.9 (Purcell et al., 2007) with the option—homozyg printing all identified ROH of at least 500 kb (—homozyg-kb 500). The following arguments were kept with default parameters as specified in PLINK 1.9: a minimum of 100 SNPs per ROH (—homozyg-snp 100), a minimum density of one SNP per 50 kb (—homozyg-density 50), merging consecutive ROH with gaps shorter than 1 Mb (—homozyg-gap 1000), scanning window spanning 50 SNPs (—homozyg-window-snp 50), including at most one heterozygous call per window (—homozyg-window-het 1), and at most five missing calls per window (—homozyg-window-het 5). Results are shown in Figure S7.

ABC-DL

Overview

ABC is a Bayesian statistical framework that comprises a family of algorithms (Beaumont et al., 2002; Bertorelle et al., 2010; Pritchard et al., 1999; Tavaré et al., 1997) to perform model comparison and parameter estimation using simulated data. ABC has a long tradition in population genomics as - especially for complex models - it is generally straightforward to generate data but often difficult to estimate the likelihood analytically (Hoban et al., 2012). In short, ABC-based demographic inference is conducted by generating simulated data under the prior probabilities of the parameters/demographic models considered in the study; each simulated dataset is then compared with the observed data by means of a set of summary statistics that capture relevant information in the data; the parameters/models used to generate the simulations are accepted as sampled from their posterior distributions if the summary statistics are similar to the ones estimated in the observed data.

For any given dataset, several summary statistics can be defined. Nevertheless, identifying summary statistics that capture all of the information present in the data pertaining to the parameter or model of interest is usually not straightforward. For a given set of summary statistics, several computational methods - such as machine learning (Beaumont, 2019) - have been proposed to automatize the choice of informative summary statistics. ABC-DL relies on DL to define the informative summary statistics to be used in ABC (Wong et al., 2018). The joint site frequency spectrum (*jSFS*) has been previously used as input to the DL for comparing complex demographic models (Lorente-Galdos et al., 2019; Mondal et al., 2019). In this work, we have extended the ABC-DL framework to account for aDNA specificities by augmenting the simulator to include features typical of aDNA such as high error rates and low coverage prior to calling genotypes.

Simulations

We used the coalescence-based simulator *fastSimcoal2* to generate ~713 Mbp of sequenced data assuming neutrality, considering a recombination rate over the genome of 1.0e-8 (Li and Jakobsson, 2012) and a mutation rate sampled from a Normal distribution with mean 1.61e-8 and standard deviation of 0.13e-8 (Lipson et al., 2015). Since *fastSimcoal2* requires times to be specified in generations, each date was initially scaled by the average human generation time (29 years) estimated by Fenner (2005). After this step, the simulated data were first re-sampled to match the observed depth of coverage, genotypes were then recalled on the resulting alleles, noise was introduced to mimic the increased error rate in aDNA, and contamination was simulated assuming that it originated from present-day Greeks (Document S1).

Demographic models

We applied a two-step model comparison approach to define the topology of the underlying demographic model. First, we tested four three-leaf models to establish the topology of the three ancestral populations *CHG*, *EHG* and Aegean Neolithic (Figure 4A) based on the prior distributions of demographic parameters listed in Table S4. The tree with highest support was then used as a backbone for more complex 7-leaf models (Figure 4B). We compared six 7-leaf plausible models compatible with previous results for *Europe_LNBA* and with results from this study (Figure 4B), based on the prior distributions of demographic parameters listed in Table S4.

For model comparison, each network was trained with 20,000 simulations per model, setting as output for each simulation the probability of assignment to one of the models. Next, we generated an additional set of 100,000 simulations per model, and used each DL network to predict the probability of assignment to each demographic model. A combined DL prediction was obtained by averaging over the 100 DL predictions. This combined prediction was used as the summary statistic for the ABC analysis. For defining the best summary statistic of a parameter from a demographic model, we trained 10 independent DL networks with 10,000 simulations as the training dataset, and generated a combined DL prediction to be used in the ABC step by averaging over the 10 DL. We used the prediction as a summary statistic in the ABC analysis to generate the posterior distribution of the parameter.

Validation of the ABC-DL approach

Model choice

The accuracy of ABC-DL for distinguishing between the proposed models was quantified using simulated genomic data from the different models as observed data, and by inferring the posterior probability of each model with ABC-DL. The contingency table between the model that produced the simulation and the model with the largest posterior probability defines the confusion matrix (Csilléry et al., 2010). The diagonal of the confusion matrix quantifies how well the ABC-DL approach properly identifies the model used to generate the data. The confusion matrix for the four three-leaf models suggests that ABC-DL can distinguish the different models, as the P(Real Model|Inferred Model) ranged between 0.66 to 0.96 (Table S4). For the 7-leaf models, the confusion matrix suggests that ABC-DL can distinguish the models as the P(Real Model|Inferred Model) ranged between 0.69 to 0.95 (Table S4).

Parameter estimation

For each parameter, Spearman's rank correlation between the values used to simulate the data and the DL prediction was computed to quantify the amount of information that the DL prediction provided. Furthermore, we estimated how well the mean of the posterior recapitulates the true value of each parameter by means of the factor 1.25 statistic, the fraction of times that the estimated mean of the posterior was within the range of 80% to 125% of the simulated value of a parameter (see Table S4, Document S1, and Excoffier et al., [2005] for details). When ABC-DL was applied to the observed data, divergence between the shape of the prior and posterior distribution was estimated by means of the Kullback-Leibler divergence (KL-divergence) (Kullback and Leibler, 1951). Reduction in the amount of uncertainty of the posterior with regards to the prior was judged by means of the ratio of highest density interval (HDI). For the three-leaf model A1, all of the validation analyses discussed above suggest that it is possible to obtain reliable posterior distributions for the parameters (Table S4; Document S1). For the 7-leaf model B4, we observed a high variability in the performance of ABC-DL for estimating the posterior distributions of the parameters (Table S4; Document S1), which can be expected given the complexity of the model. Nevertheless, for the estimated time and amounts of *CHG*-related and *Steppe_EMBA*-related gene flow reported in the text, we have (i) a Spearman's rank correlation $\geq$ 0.5 between the DL prediction and the value used in the simulation, (ii) high ($\geq$1.5) ratio factor 1.25 posterior mean/random prior, (iii) large (> 0.5) KL-divergence between posterior and prior, and (iv) strong reduction in HDI (> 2), that is, in the uncertainty of the posterior distribution compared to the prior distribution. These results suggest that the estimations obtained by ABC-DL for these four parameters substantially reduce the amount of our prior uncertainty.

Phenotype prediction

The genotype likelihoods from the mapped reads of our nuclear capture data were calculated using ANGSD v. 0.921 (Korneliussen et al., 2014). For this, we specified the SAMtools model (-GL 1) and inferred major/minor alleles from the genotype likelihood (-doMajorMinor 1 -doMaf 1). We used a minimum read depth of 20 (-genoMinDepth 20) and trimmed 10 bases of the reads from both ends (-trim 10). Furthermore, we assumed a uniform prior for the genotypes (-doPost 2) and only considered called genotypes with posterior probabilities > 0.95 (-postCutoff 0.95) and base (-minQ 30) and mapping (-minMapQ 30) qualities of 30 in Phred score. Using these data, we extracted genotypes for *Kou01*, *Log02* and *Pta08* for two SNPs on chromosome 2 – rs4988235 and rs182549 – in which the alleles T −13,910 and A −22,018 respectively are informative of lactase persistence. Furthermore, we used the capture

data to infer eye-, hair- and skin color for the individuals *Kou01*, *Log02* and *Pta08* using the HIrisPLex-S DNA Phenotyping Webtool (https://hirisplex.erasmusmc.nl/). We compared our dataset with the 41 SNPs published as HIrisPlex-S (Chaitanya et al., 2018; Walsh et al., 2017) and found an overlap of 30 SNPs (Table S1, "Nuclear capture" tab). Note that HIrisPLex-S is an extension of the previously published HIrisPLex, which contains 24 SNPs for eye and hair color determination (Walsh et al., 2014) and that our data cover 23 of those 24 SNPs. The HIrisPLex-S webtool calculates individual prediction probabilities and associated values for the loss of prediction accuracy (AUC loss) depending on the available set of SNPs. The results for *Kou01*, *Log02* and *Pta08* are given in Table S1. The most supported eye, hair and skin color is the one with the highest prediction probability (Table S1). A discussion about lactose intolerance can be found in Document S1.

CellPress
OPEN ACCESS

Supplemental figures

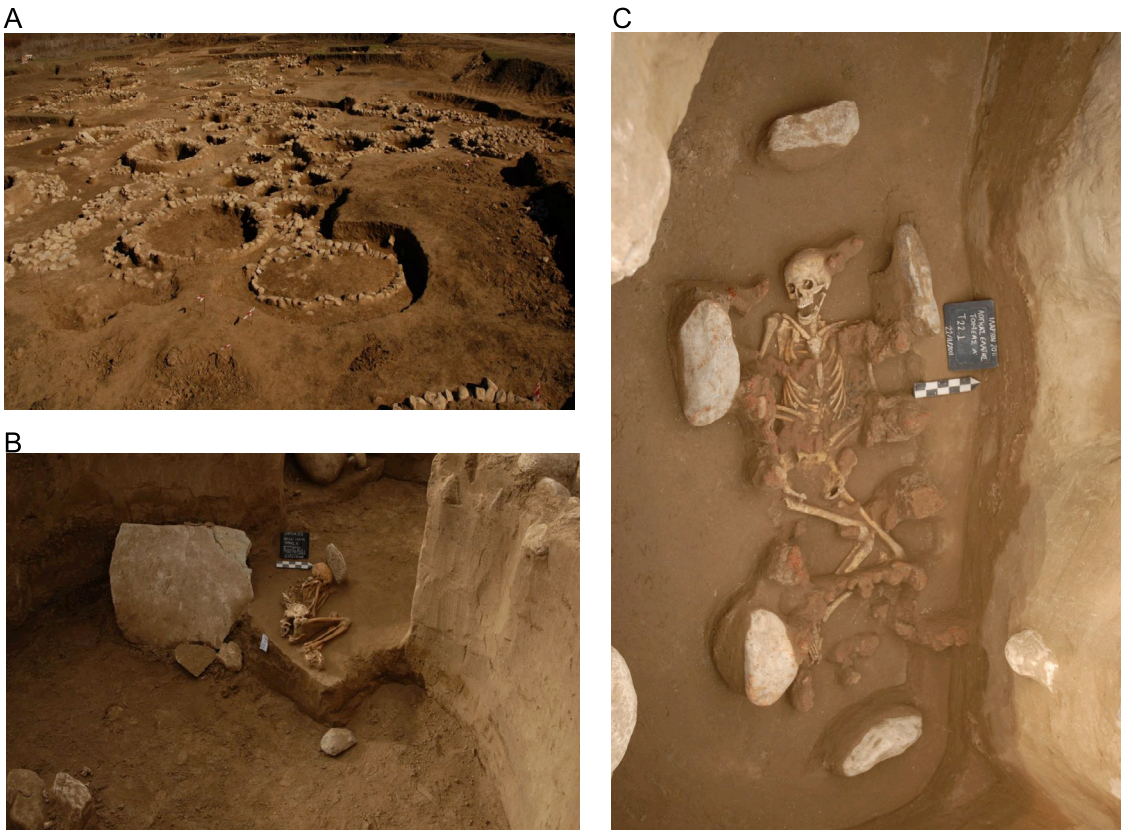


Figure S1. Images of archaeological site Elati-Logkas (*Log02*, *Log04*), related to Table 2 and Document S1

(A) Elati-Logkas, view of the cemetery with burials covered by stones known as "periboloi." (B) Elati-Logkas, Burial 80.1 (*Log04*) is a pit-grave in the circumference of an inner enclosure built from rough stones. The buried individual was in crouched position lying to the left side, with the hands bent and the palms supporting the skull. Inside the same walls, four other similar burials were excavated with no grave goods apart from only one flint stone blade in tomb 80.5. (C) Elati-Logkas, Burial 22.1 (*Log02*) is the main among three pithos-inhumations and one secondary burial inside the "peribolos 22." The grave itself is bordered by rough stones, with the buried individual laid on a ceramic "stretcher." Several vertical lines are still visible on the skeletal remains. The individual of the burial 22.1 was found in a supine position with the hands crossed on the abdomen, the legs bent in a crouched position to the left, and the skull turned to the right side. There were no grave goods found in the burial 22.1. Photo credits: Ephorate of Antiquities of Kozani, Hellenic Ministry of Culture, Greece. Courtesy of Dr. Georgia Karamitrou-Mentessidi.

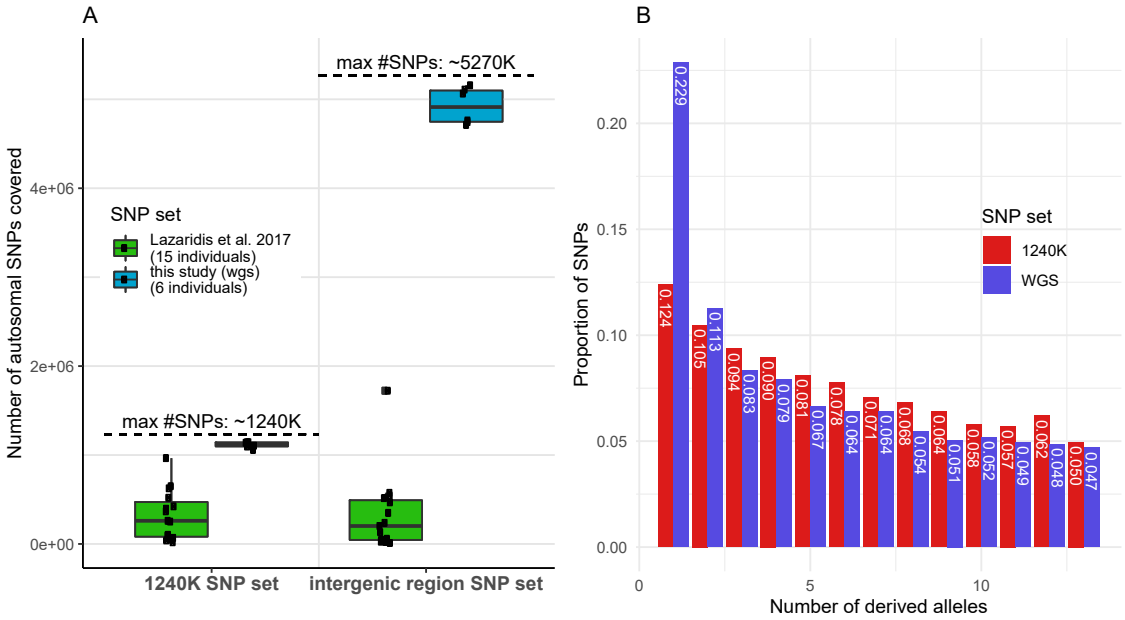


Figure S2. Comparison of SNP capture and WGS data, related to Figure 4 and Tables 2 and S1

(A) Number of single nucleotide polymorphisms (SNPs) in the nuclear genomic data from this study (WGS data) in comparison with previously published BA genomic data from the Aegean. The number of covered SNPs across BA Aegeans based on two SNP sets are shown. On the left: the number of SNPs based on the 1240K SNP set (Dataset I) defined by the array used in Lazaridis et al. (2017) to enrich the libraries. On the right: the number of SNPs based on the intergenic regions defined for the ABC-DL analysis below (Dataset IV, STAR Methods). The green box plots (median indicated by a horizontal line and interquartile range indicated by the box) correspond to the number of SNPs among the BA Aegean data from present-day Greece (Lazaridis et al., 2017); the blue box plots correspond to the number of SNPs among the whole genome sequence (WGS) data from this study. (B) One-dimensional Site Frequency Spectrum (SFS) for the seven whole genomes used for demographic analyses (ABC-DL). The seven genomes included here are: *Mik15* and *Log04* from this study, *YamnayaKaragash_EBA* (3,018-2,887 BCE) (de Barros Damgaard et al., 2018), *KK1* (CHG; 7,745-7,579 BCE) (Jones et al., 2015), Bar8 (Neolithic Barçın; 6,122-6,030 BCE) (Hofmanová et al., 2016), *Sidelkino* (*EHG*; 9,386-9,231 BCE) (de Barros Damgaard et al., 2018), and *S_Greek-1* (SAME3302732; modern Greek from Thessaloniki) (Mallick et al., 2016).STAR Methods In blue ("WGS") the SFS for the regions included in Dataset IV (STAR Methods). In red ("1240K") the SFS for the regions in Dataset IV restricted to the sites overlapping with the SNPs included in the 1240K array.

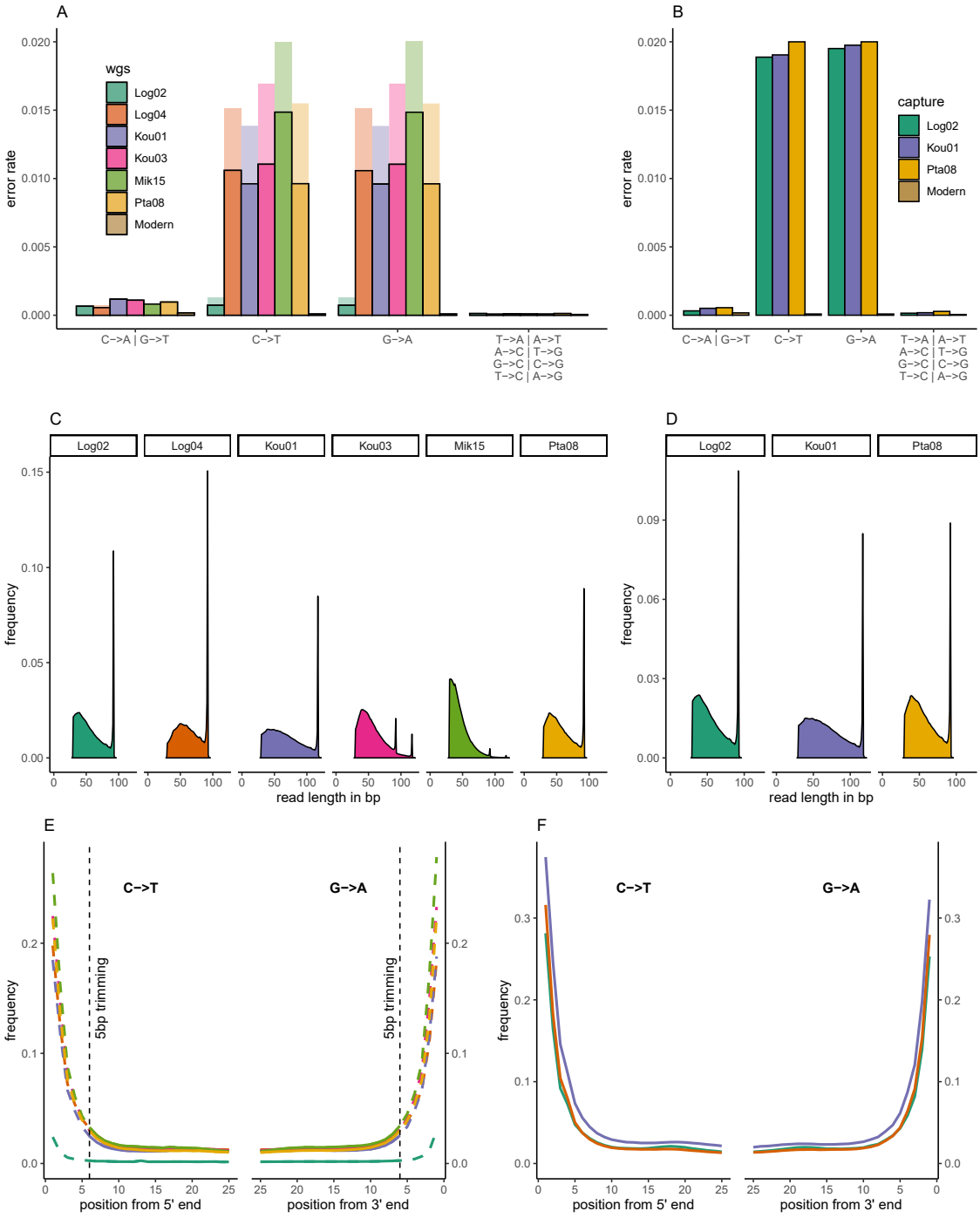


Figure S3. Error rates, damage, and read length distributions for the WGS and nuclear capture data from this study, related to Figure 1 and Tables 2 and S1

(A) Error rate for whole genome sequencing before (lighter colors) and after (darker colors) trimming 5 bp from the extremities of the reads. *Log02* was USER™-treated. (B) Error rate for nuclear capture data for different mutation types. Columns 1 and 2 show transitions and column 3 shows transversions. (C) Read length distribution for whole genome sequencing. (D) Read length distribution for nuclear capture data. (E) Post-mortem damage pattern for whole genome sequencing (C to T and G to A substitutions). Dashed lines indicate partial data removal resulting from trimming 5 bp from the extremities of the reads. The color of each curve indicates the analyzed sample according to panel A. *Log02* (dark green curve) was USER™-treated. (F) Post-mortem damage pattern for nuclear capture data. Curves are colored according to panel B. See STAR Methods for details.

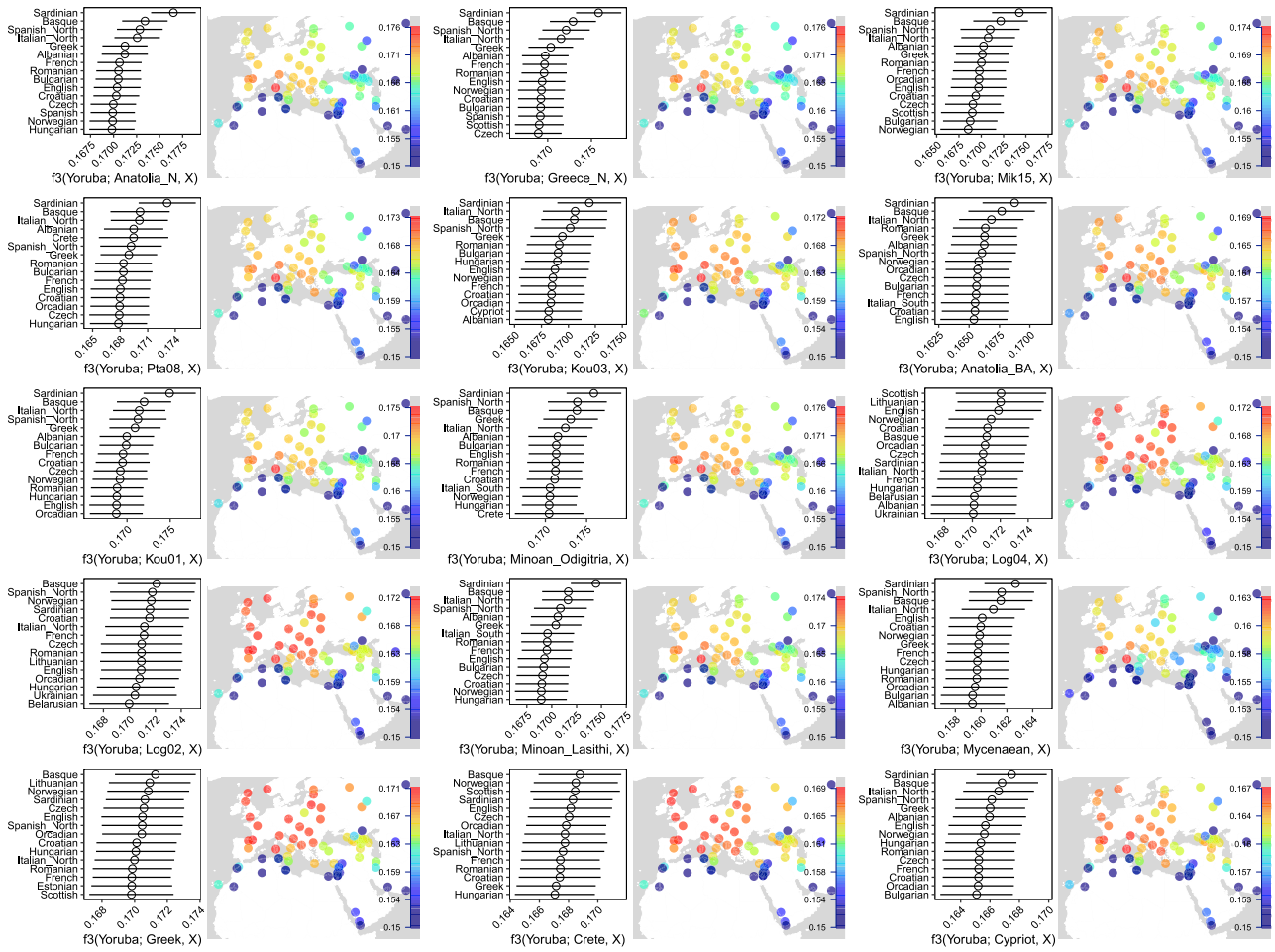
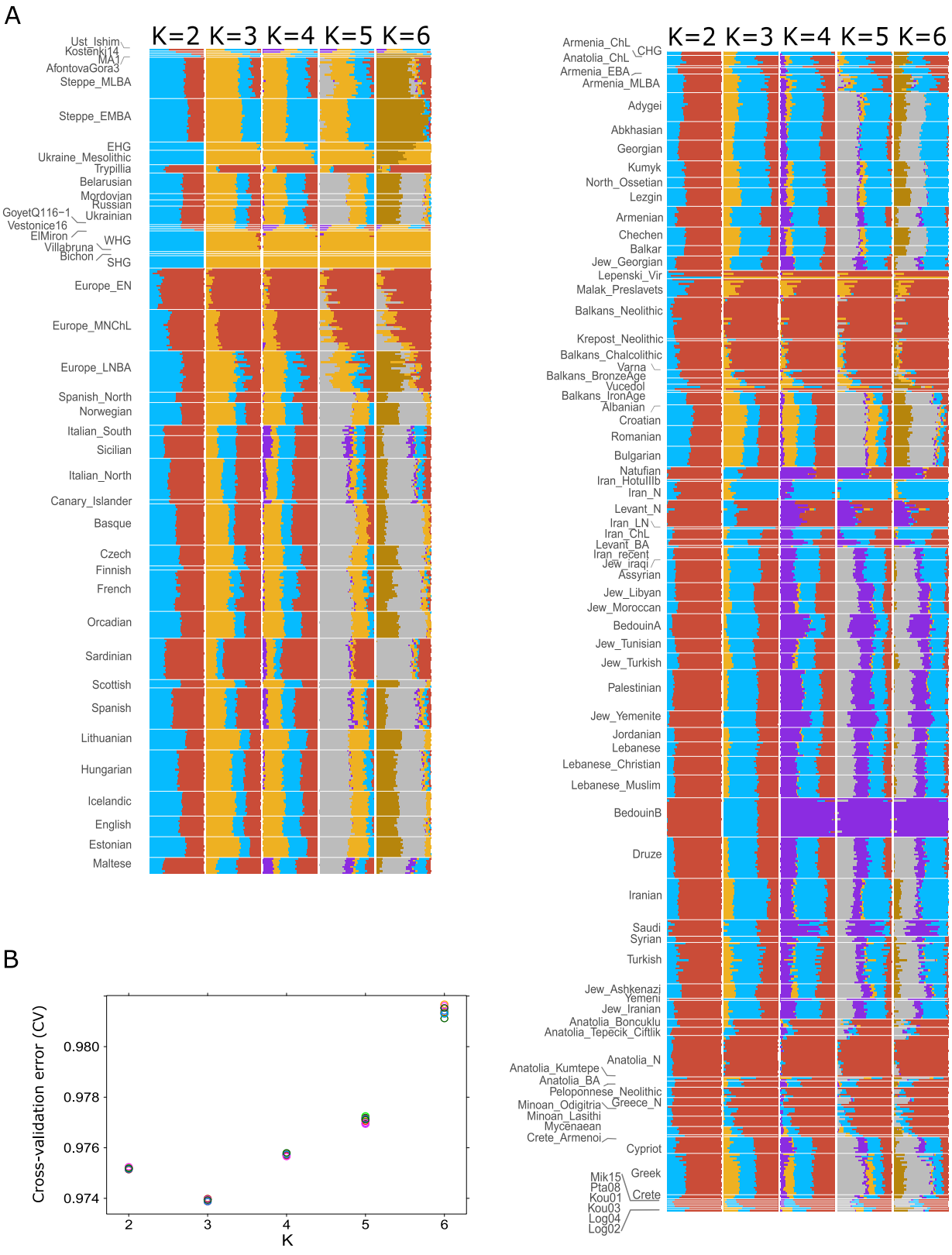


Figure S4. Genetic affinities between Neolithic, BA, and present-day Aegeans compared to other present-day Eurasian populations, related to Table 1

f_3-statistics of the form f_3(Yoruba; *Y*, *X*): *Y* corresponds to either Neolithic Anatolians or Greeks, Minoan-Petras-EBA individual from the island of Crete (*Pta08*), the Cycladic-Koufounisi-EBA individuals (*Kou01, Kou03*), the Helladic-Manika-EBA individual from the island of Euboea (*Mik15*), the Helladic-Logkas-MBA individuals from northern Greece (*Log02, Log04*), previously published BA Aegeans (Mycenaeans and Minoans), and present-day Greeks (incl. Cretans), Cypriots, while *X* are other present-day populations from Dataset I (Lazaridis et al., 2014, 2016, 2017) (STAR Methods). For clarity, we only show results for west Eurasian and north African populations and cap f_3 values below 0.15. For each case, we show the geographic distribution of f_3 (warmer colors represent greater sharing between populations *X* and *Y*). Beside each map, we plot the f_3 values for the 15 populations that are most closely related to each of the populations in *Y* (bars represent ~1.95 standard errors). In agreement with *MDS* and *ADMIXTURE* analyses, we observed that ancient and present-day Anatolians and Greeks share the most genetic drift with present-day central and southern European populations.



(legend on next page)

Figure S5. ADMIXTURE analysis using ancient and modern populations with the number of ancestry components ranging from *K* = 2–6 and cross-validation error, related to Figure 3, Table 1, and Document S1

(A) For this analysis we consider a total of 969 individuals (638 modern and 331 ancient) and 165,447 SNPs (Dataset II, STAR Methods). Each bar represents one individual. Individuals from the same population were grouped. For all K > 2, red represents the component mostly present in "European Neolithic-like," light blue in "Neolithic Iran/Caucasus HG-like" and orange for "European HG-like." (B) Cross-Validation error (CV error) for *K* ranging from 2 to 6. The CV-error is plotted for the ten runs for each value of *K* (STAR Methods).

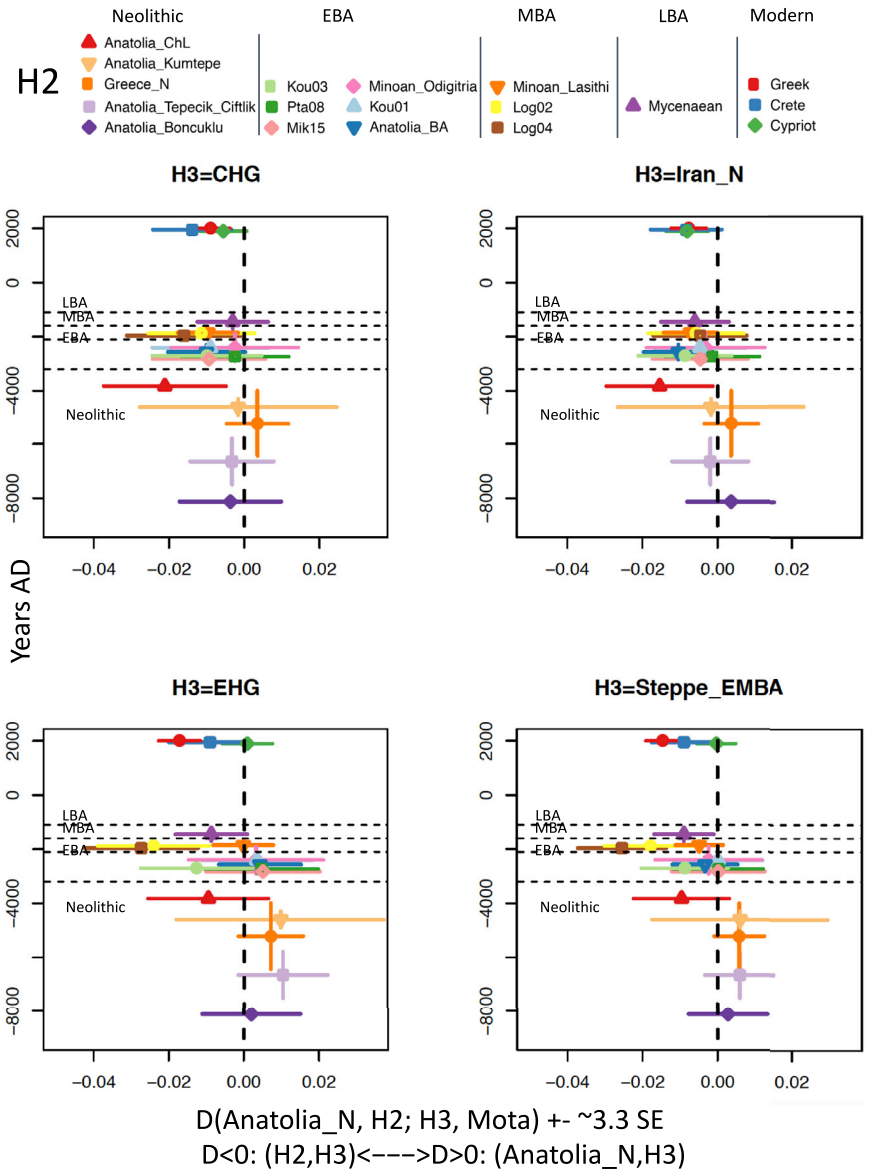


Figure S6. Exploring differential allele-sharing in Aegean/Anatolian populations through time with D-statistics, related to Table 1

D-statistics of the form D(Anatolia_N, H2; H3, Mota) were computed, testing whether Anatolia_N (H1), or ancient/modern Anatolian and Aegeans (H2) share more alleles with *CHG*, *Iran_N*, *EHG*, or *Steppe_EMBA* (H3, STAR Methods). For this analysis, the genome of an Ethiopian individual (Mota) was used as an outgroup. Points represent D-statistics, and horizontal error bars represent ~3.3 standard errors (SE corresponding to a *p-value* of ~0.001 in a Z-test). Vertical bars represent upper and lower bounds of the dates available for the populations. In this figure, the populations are ordered chronologically, using either radiocarbon dates (when available) or dated archaeological context. Horizontal dashed lines indicate time periods. Vertical dashed lines mark the zero. A value of D = 0 indicates no gene flow or ancestral population structure (Durand et al., 2011), thus H1 and H2 are symmetrically related to H3 and Mota. In this case, D < 0 would indicate potential gene flow between H3 and H2, and D > 0 would indicate potential gene flow between H3 and H1. Abbreviations for chronological periods and population names are given in Table 1.

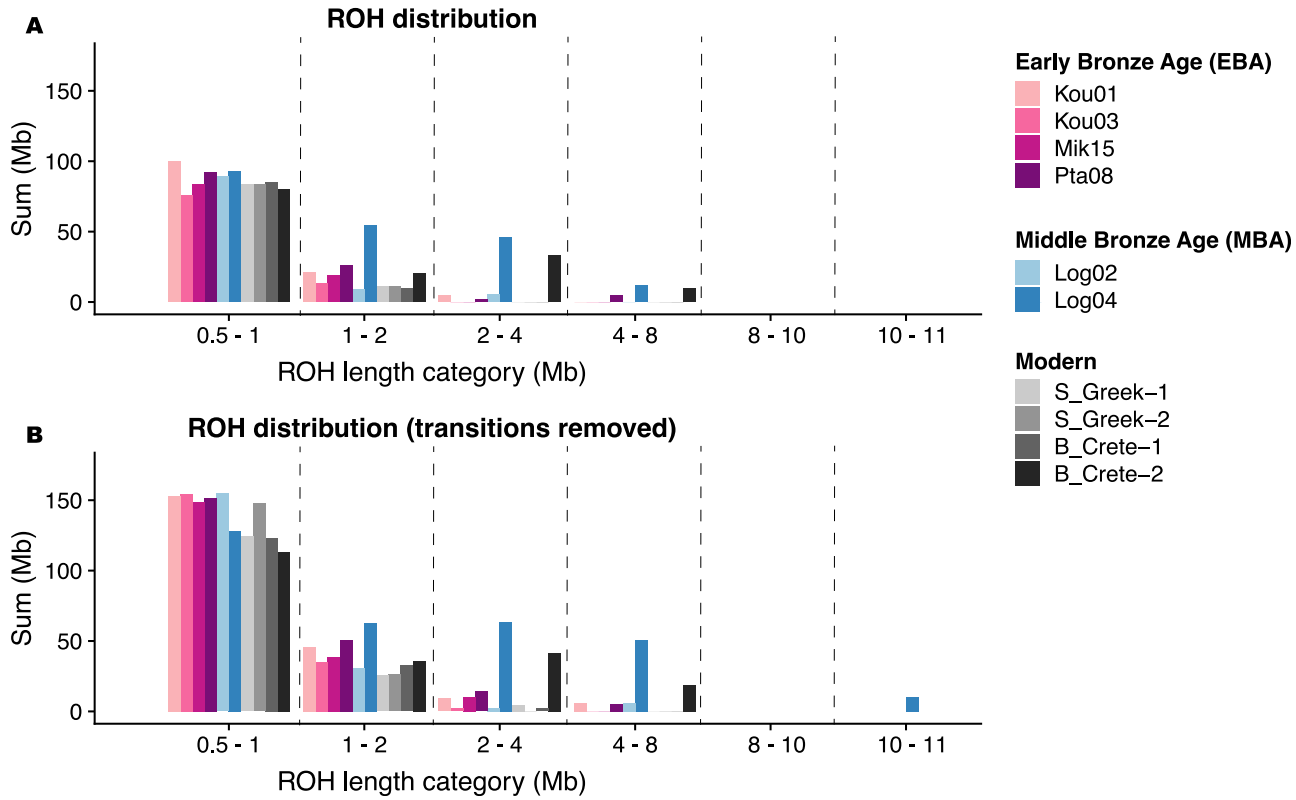


Figure S7. Estimated total ROH length by size category for six ancient and four modern genomes, related to Document S1

(A) ROHs estimated from 43 million imputed transitions and transversions, and (B) 13 million imputed transversions (STAR Methods).