Article

Punic people were genetically diverse with almost no Levantine ancestors

https://doi.org/10.1038/s41586-025-08913-3

Received: 8 May 2024

Accepted: 18 March 2025

Published online: 23 April 2025

Check for updates

Harald Ringbauer[1,2,3]✉, Ayelet Salman-Minkov[4,5], Dalit Regev[6], Iñigo Olalde[1,7,8], Tomer Peled[4], Luca Sineo[9], Gioacchino Falsone[10], Peter van Dommelen[11], Alissa Mittnik[1,2,3], Iosif Lazaridis[1,12], Davide Pettener[13], Maria Bofill[14], Ana Mezquida[14], Benjamí Costa[14], Helena Jiménez[14], Patricia Smith[15], Stefania Vai[16], Alessandra Modi[16], Arie Shaus[1,17,18], Kim Callan[12,19], Elizabeth Curtis[12,19], Aisling Kearns[12], Ann Marie Lawson[12,19], Matthew Mah[12,19,20], Adam Micco[12], Jonas Oppenheimer[12,19], Lijun Qiu[12,19], Kristin Stewardson[12,19], J. Noah Workman[12], Nicholas Márquez-Grant[21], Antonio M. Sáez Romero[22], María Luisa Lavado Florido[23], Juan Manuel Jiménez-Arenas[24], Isidro Jorge Toro Moyano[25], Enrique Viguera[26], José Suárez Padilla[26], Sonia López Chamizo[26], Tomas Marques-Bonet[27,28,29,30], Esther Lizano[31,32,33], Alicia Rodero Riaza[34], Francesca Olivieri[35], Pamela Toti[36], Valentina Giuliana[37], Alon Barash[38], Liran Carmel[39], Elisabetta Boaretto[40], Marina Faerman[41], Michaela Lucci[42], Francesco La Pastina[9,10,42], Alessia Nava[43], Francesco Genchi[44], Carla Del Vais[45], Gabriele Lauria[9], Francesca Meli[10], Paola Sconzo[10], Giulio Catalano[9], Elisabetta Cilli[46], Anna Chiara Fariselli[46], Francesco Fontani[2,3,46], Donata Luiselli[46], Brendan J. Culleton[47], Swapan Mallick[12,19,20], Nadin Rohland[1,12,20], Lorenzo Nigro[48], Alfredo Coppa[49,50,51,54], David Caramelli[16,54], Ron Pinhasi[49,52,54], Carles Lalueza-Fox[27,53,54], Ilan Gronau[4,54]✉ & David Reich[1,2,12,19,20,54]✉

The maritime Phoenician civilization from the Levant transformed the entire Mediterranean during the first millennium BCE[1–3]. However, the extent of human movement between the Levantine Phoenician homeland and Phoenician–Punic settlements in the central and western Mediterranean has been unclear in the absence of comprehensive ancient DNA studies. Here, we generated genome-wide data for 210 individuals, including 196 from 14 sites traditionally identified as Phoenician and Punic in the Levant, North Africa, Iberia, Sicily, Sardinia and Ibiza, and an early Iron Age individual from Algeria. Levantine Phoenicians made little genetic contribution to Punic settlements in the central and western Mediterranean between the sixth and second centuries BCE, despite abundant archaeological evidence of cultural, historical, linguistic and religious links[4]. Instead, these inheritors of Levantine Phoenician culture derived most of their ancestry from a genetic profile similar to that of Sicily and the Aegean. Much of the remaining ancestry originated from North Africa, reflecting the growing influence of Carthage[5]. However, this was a minority contributor of ancestry in all of the sampled sites, including in Carthage itself. Different Punic sites across the central and western Mediterranean show similar patterns of high genetic diversity. We also detect genetic relationships across the Mediterranean, reflecting shared demographic processes that shaped the Punic world.

Canaanite–Phoenician culture emerged in the city-states of the Bronze Age Levant. In the early first millennium BCE, Phoenicians established a vast network of trading posts as far as southwest Iberia[1–4], spreading their culture, religion and language to the central and western Mediterranean[6]. By the mid-sixth century BCE, Carthage, a Phoenician coastal colony in present-day Tunisia, emerged as the dominant power in the central and western Mediterranean while the influence of Levantine cities declined as they fell under neo-Assyrian and neo-Babylonian control. Carthage came into large-scale violent conflicts with Greeks in the fifth to fourth centuries BCE and with Rome in the third to second centuries BCE, culminating in its destruction in 146 BCE and annexation by the Roman Republic[7,8].

Many historians and archaeologists use the Roman term Punic to refer to culturally Phoenician communities in the central and western Mediterranean that were associated with or even ruled by Carthage after the mid-sixth century BCE[5,9–11]. While problems of such a simplifying label have been highlighted[12], we follow the recent *Handbook of the Phoenician and Punic Mediterranean*[6] and use Punic to refer to sites in the central and western Mediterranean associated with material culture with Phoenician antecedents or connotations between the late sixth and

A list of affiliations appears at the end of the paper.

second centuries BCE, a time frame of Carthaginian hegemony in this region. We also use the term to refer to the people buried at these sites. However, we caution that, by labelling sites and people as Punic, we do not make any claim of a homogeneous culture or ethnicity. Moreover, our distinction between Phoenician and Punic sites does not question the Levantine cultural origins of their inhabitants, nor does it address the question of whether the founders of Phoenician–Punic sites such as Carthage and Cádiz were Levantine[13]; we do not have the early samples from these sites needed to address these issues. We use Punic in this paper simply as a means to refer to a group of sites and their inhabitants as defined above, for which our analysis finds shared genetic signatures.

Ancient DNA (aDNA) reveals information regarding genetic ancestry, which can help resolve questions beyond the reach of conventional historical and archaeological data. Previous aDNA studies of Punic remains focused on mitochondrial DNA (mtDNA), revealing diverse maternal lineages[14–16]. However, as mtDNA traces only a single lineage, those analyses provide limited statistical resolution compared with genome-wide data. A whole-genome analysis of 5 Bronze Age individuals from Sidon[17] and 12 Iron Age individuals from Beirut[18] indicates that Phoenicians in the Levant were genetically similar to local Bronze Age Canaanite people. However, a mitochondrial genome from Carthage[14] and whole-genome data from 12 individuals from the nearby rural Punic site of Kerkouane[19] show substantial south European ancestry as well as indigenous North African ancestry. Partial North African ancestry was also found in genome-wide data from eight individuals from two Punic sites in Sardinia, combined with a broad eastern Mediterranean ancestry[20]. Together with analysis of the whole-genome sequence of an individual from Ibiza[21], which was also interpreted to harbour eastern Mediterranean ancestry, this suggested that Punic people had complex ancestry. However, these observations were based on analysing genome-wide data from a combined total of 21 individuals and did not include the key regions of Iberia or Sicily or settlements with early Phoenician foundations, such as Carthage, Motya or Cádiz. Owing to the small sample size and complex combination of ancestries, it has not been possible to study ancestry across Punic sites systematically.

Dataset assembly

We screened human skeletal remains from 398 individuals excavated from 14 sites in Iberia, Sardinia, Sicily, North Africa and the Levant, using in-solution enrichment for more than 1.2 million single-nucleotide polymorphisms (SNPs)[22] and the mitochondrial genome[23] (Methods and Supplementary Tables 1 and 2). Sampled sites cover diverse cultural, temporal and geographical contexts (Supplementary Information 1). They include the Levantine Phoenician settlement of Akhziv, large urban settlements with probable Levantine foundations, such as Cádiz, Motya and Carthage, and sites with evident cultural ties to Carthage, such as Villaricos or the rural settlement of Kerkouane (Fig. 1a). We publish results for 210 unique individuals that produced genome-wide data that met minimal standards for aDNA authenticity, of which 196 came from Phoenician and Punic contexts (Supplementary Table 3). We carried out population genetic analyses on a subset of 157 individuals with sufficient data, as described below.

We generated 111 direct accelerator-mass-spectrometry-based dates on bone for 99 of these individuals (Supplementary Table 4). These radiocarbon dates were crucial to confirming archaeological context as several sampled Punic sites were occupied in later periods and because excavations for a subset of analysed samples predate the establishment of modern archaeological methods (in some cases more than a century ago). Owing to fluctuations in the proportion of atmospheric carbon 14 in the period of interest, we could typically assign samples only to around 800–400 BCE (the Hallstatt Plateau in the calibration curve used for radiocarbon dating) or 400–200 BCE (Supplementary Information 2). Nevertheless, this enabled us to distinguish within Punic sites between people living during the earlier stages of the Carthaginian expansion (before 400 BCE), people living after this time but still in Punic contexts and people plausibly living after the transition to Roman control around 250–150 BCE. A notable limitation of our sampling is the absence of sampled individuals who predate 600 BCE (based on archaeological dating of the tombs), because cremation was the dominant form of burial in Phoenician settlements in the central and western Mediterranean before this time[24] (Discussion).

We filtered to individuals with higher-quality data (>20,000 SNPs and no evidence of substantial contamination) and with verified archaeological context, yielding 157 individuals for whom we performed population genetic analysis. Within this dataset, we identified 131 individuals deriving from a Phoenician or Punic context, out of which 108 had a firm archaeological context, indicating that they can confidently be classified as Phoenician or Punic (Fig. 1a and Supplementary Information 2). We added previously published sequences for 20 indivduals confidently classified as Phoenician or Punic (7 from 2 Sardinian sites[20], 1 from Ibiza[21] and 12 from Beirut[18]), yielding a dataset of 128 higher-quality individuals confidently classified as Phoenician or Punic (Fig. 1). A recent study reported whole-genome shotgun sequencing data for 12 individuals from the North African site of Kerkouane[19]. We used in-solution enrichment (which enabled us to convert samples with low proportions of endogenous DNA to analysable data) to generate data for 17 new individuals from Kerkouane and additional data for 10 individuals with previously reported data. We analysed the data that we generated with data from previously reported ancient individuals from other relevant times and regions, as well as 26 newly sequenced individuals from relevant contexts (which are not Phoenician or Punic). Among these are an Early Iron Age individual from the inland northeast Algerian site Khenchela, who we use in our model of North African ancestry, and 20 individuals from originally Punic sites in Sicily, Sardinia and Iberia, whose radiocarbon dates place them within the period of Roman hegemony (Supplementary Information 2).

Genetic ancestry and admixture modelling

To examine the genetic variation in our newly reported data, we projected whole genomes from the ancient individuals onto the first two principal components (PCs) computed from 1,196 present-day West Eurasian and North African individuals (Fig. 1b, Methods and Extended Data Fig. 1). For context, we also projected Bronze and Iron Age individuals from relevant archaeological cultures across the Mediterranean, who broadly cluster on the basis of their geographical distribution. We find that individuals from the Levantine Phoenician site of Akhziv in present-day Israel cluster together with previously published Bronze and Iron Age Levantine individuals, including from Megiddo in present-day Israel[25] and the Phoenician cities of Sidon and Beirut in present-day Lebanon[17,18]. By contrast, individuals from Punic sites in the central and western Mediterranean do not cluster with Levantine individuals. Instead, they are broadly distributed with a primary mode overlapping Bronze and Iron Age individuals from Sicily and the Aegean, regardless of sampled location. In addition to this cluster, a narrow cline of individuals points from this mode towards Northwest African individuals, including pre-contact Guanche individuals from the Canary Islands[26] and the newly sequenced Early Iron Age individual from Khenchela (inland Algeria). This suggests that the Phoenician people of Akhziv derived most of their ancestry from previous populations in the Levant, while the people from Punic sites in the central and western Mediterranean shared recent common ancestry with populations from Sicily and the Aegean, with additional recent admixture with North African populations.

Other methods that are useful for learning about population mixture (Methods) confirmed the main hypotheses suggested by the principal component analysis (PCA). Using ADMIXTURE, we inferred a three-component model with clusters maximized in North Africans, eastern populations (Levant/Iran), and central and western Mediterranean groups (Extended Data Fig. 2). Individuals from Akhziv exhibit similar

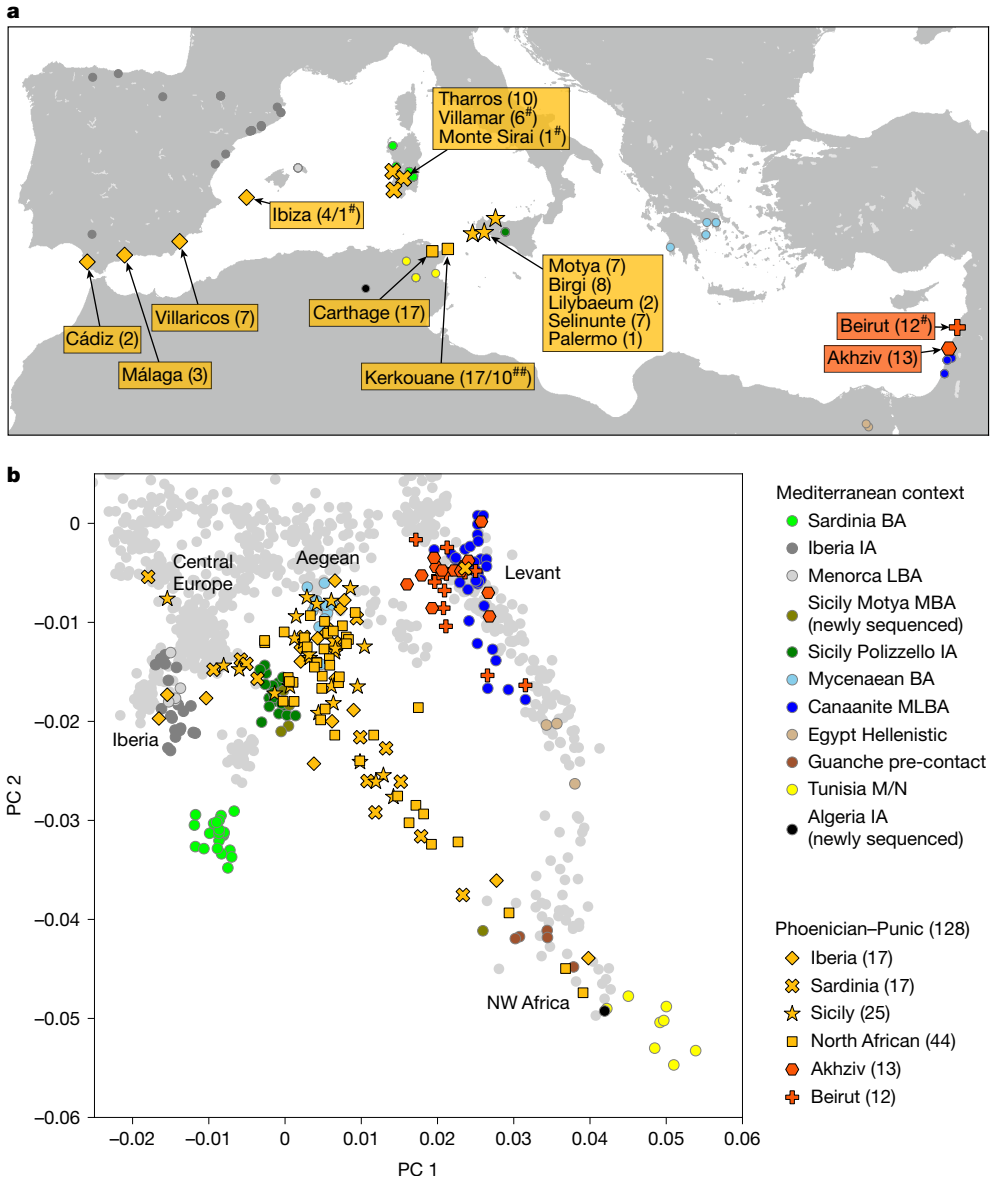


Fig. 1 | Sample overview and PCA. a, The locations of archaeological sites from which we analysed aDNA. The number of individuals with newly generated aDNA data (>20,000 SNPs) and high-confidence archaeological association (Methods and Supplementary Information 2) is indicated for each site. The single and double hash symbols indicate the numbers of previously published ancient individuals or the numbers of previously published ancient individuals for whom there are new sequence data obtained by in-solution enrichment, respectively. The locations of aDNA context are indicated by circles. **b**, PCA. We first computed PCs using 1,196 present-day individuals from Western Eurasia and North Africa (Methods). We then projected 128 individuals with more than 20,000 SNPs covered and confidently assigned to a Phoenician–Punic context (Methods and Supplementary Table 5) onto the first two PCs. For context, we also projected various Bronze and Iron Age individuals across the Mediterranean (coloured circles matching **a**; Supplementary Table 6). Extended Data Fig. 1 shows the PCA projections broken up according to archaeological sites and dates. Extended Data Fig. 10 shows a projection of the reference populations only and a magnified version of the Levantine individuals. BA, Bronze Age; IA, Iron Age; LBA, Late Bronze Age; MBA, Middle Bronze Age; MLBA, Middle–Late Bronze Age; M/N, Meso/Neolithic; NW, northwest.

ancestry patterns to those of Bronze Age individuals from the Levant. By contrast, individuals from Punic sites are assigned clusters more typical of Bronze and Iron Age populations from the central and western Mediterranean and North Africa (Supplementary Information 3). To carry out formal tests of fit for alternative models and to quantify ancestry proportion under the fitting models, we used qpAdm[27,28], with 78 ancient individuals from 8 different genetically homogeneous groups as proxies for the ancestry sources (Methods and Supplementary Table 7). While multiple ancestry models could fit most individuals (Extended Data Fig. 3), several robust observations emerged. First, all individuals from the Phoenician site Akhziv derived more than 80% of their ancestry from a Levantine Bronze Age population (Fig. 2). By contrast, only three individuals from Punic sites were assigned substantial Levantine ancestry: two from Sicily (whose radiocarbon dates overlap with the period of Roman control in Sicily) and one from Sardinia, all of whom clustered in the PCA with Levantine individuals (Extended Data Fig. 1). The primary source of ancestry in most Punic individuals is best proxied in our analysis using Bronze Age individuals from Sicily and Greece. In every region, including North Africa, and every time period, some Punic individuals derive nearly all of their ancestry from this source. However, our analysis could not more precisely identify the geographical origin of this source population as it could not confidently distinguish between Bronze Age Sicilian and Aegean sources, which had only subtle ancestry differences from each other (Discussion).

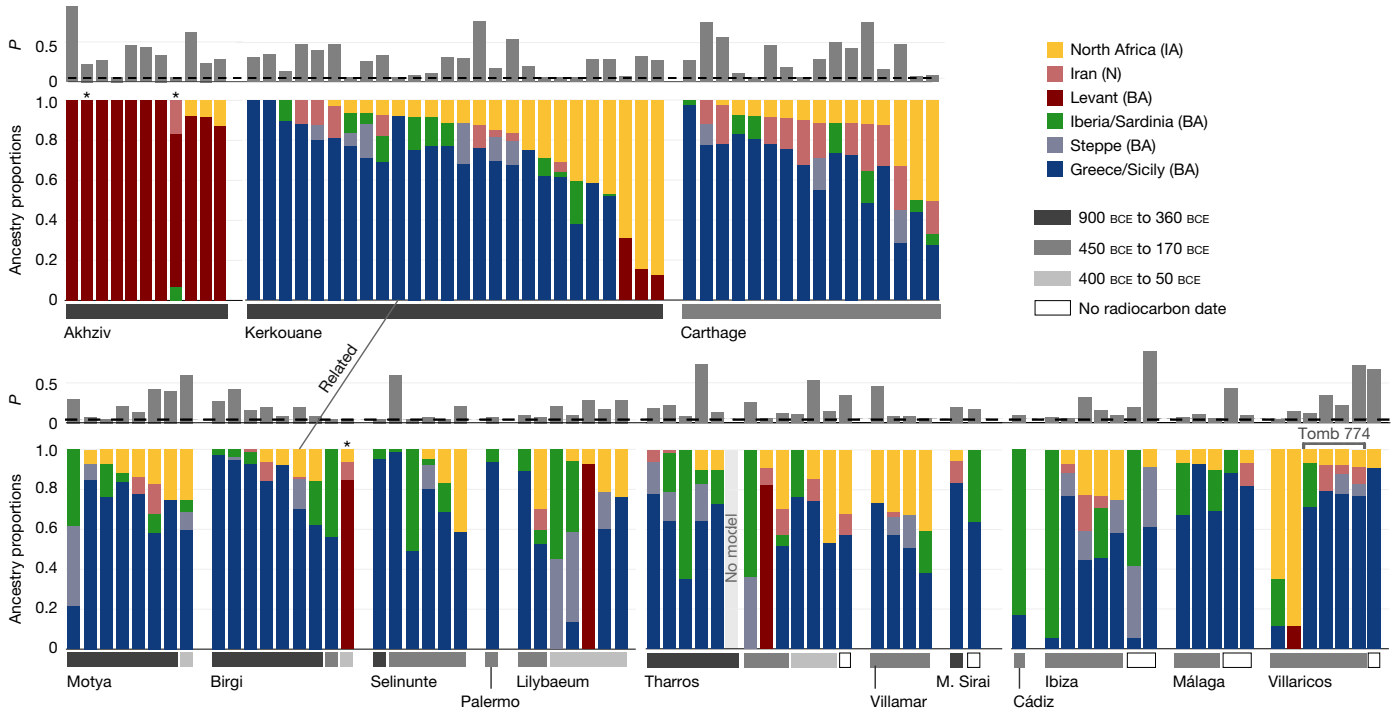


Fig. 2 | Ancestry models inferred using qpAdm. The representative admixture models for 122 Phoenician–Punic individuals in our dataset sequenced at more than 100,000 SNPs. We sorted individuals by region (Levant, North Africa, Sicily, Sardinia and Iberia), site, estimated date range (the grey horizontal bars above the site names) and, finally, by inferred North African ancestry. We combined the ancestry proportions inferred for Greece BA (Myc) and Sicily EMBA (Sicilian–Aegean ancestry), and the ancestry proportions inferred for Sardinia LBA and Iberia EBA (western Mediterranean ancestry). We combined these ancestries due to the limited ability of our qpAdm models to distinguish between them (Supplementary Information 3). As individuals typically had several valid admixture models, we selected the model that maximized Sicilian–Aegean or Levantine ancestry (Methods). For three individuals (indicated by asterisks), we could fit only a broad ancestry model (Methods), and one individual from Tharros could not be fit by any of the ancestry models that we considered. The figure highlights a pair of related individuals from Birgi and Kerkouane (Fig. 4) and related individuals buried in the same tomb in Villaricos (Fig. 5). The complete set of valid models for all 140 individuals in the dataset (including those with fewer than 100,000 SNPs) is specified in Extended Data Fig. 3. M. Sirai, Monte Sirai; N, Neolithic.

The second most common ancestry source in Punic sites is consistent with deriving from indigenous North Africans, defined as a population that is genetically similar to the Early Iron Age individual from inland northeast Algeria (Extended Data Fig. 3 and Supplementary Information 3). North African ancestry appears as early as the sixth and fifth centuries BCE in a substantial subset of individuals from Kerkouane and, sporadically, in low fractions at Sicilian sites (Motya and Birgi), and not at all in sampled Iberian and Sardinian individuals before the fourth century BCE (Fig. 2 and Extended Data Figs. 1 and 4). The rise in North African ancestry after 400 BCE outside of North Africa is particularly marked in Tharros, Sardinia (Extended Data Figs. 1 and 3). Although indigenous North African ancestry was widespread after 400 BCE in the Punic world, it remained a minority ancestry component (below 50%) in all individuals except for three from Kerkouane, two from Villaricos and one from Tharros. Even in North Africa, 10 out of the 27 individuals from Kerkouane and 5 out of the 17 individuals from Carthage can be modelled with no indigenous North African ancestry, and 84% of individuals from these sites have more than 50% Sicilian–Aegean ancestry, making it the dominant ancestry component also in North African Punic sites (Extended Data Fig. 3).

Many individuals in Punic sites in Sicily and North Africa had substantial proportions of ancestry consistent with that observed in earlier people from the same geographical regions (this ancestry not necessarily being local). By contrast, we do not observe evidence of an important genetic contribution of Bronze Age people of Sardinia and Iberia to individuals in nearby Punic sites either in clustering analysis or in formal modelling analysis (Figs. 1b and 2). Only two Iberian individuals, from Ibiza and Cádiz, had confidently high proportions of Bronze Age Iberian ancestry (Fig. 2 and Extended Data Fig. 3). Instead, Punic sites in the western Mediterranean share similar ancestry distributions of predominantly Sicilian–Aegean or North African origin (Figs. 1 and 2 and Extended Data Fig. 1; a detailed discussion is provided in Supplementary Information 3).

Genetic heterogeneity and connectedness

Five lines of population genetic analysis reveal that individuals in Punic sites were part of a genetically heterogeneous and interconnected Mediterranean population.

First, we observe a substantially higher Y haplogroup diversity per archaeological site compared with earlier people (Fig. 3a and Extended Data Fig. 5). In each sampled Punic site, most males have differing Y haplogroups (using the ISOGG classification; Methods). Overall, no single lineage dominates the sample across sites, including most commonly E1b (14 out of 58 individuals), R1b (10 out of 58), J2a (8 out of 58), G2a (7 out of 58) and J1a (5 out of 58). The high Y-chromosome diversity is notable in light of the previous proposal of J2 (J-M172) as a footprint of Phoenician demographic expansions and their genetic legacy today[29]. We observe J2 haplogroups in only 20% of the Punic male individuals, and at a similar low frequency in Levantine male individuals from the Bronze and Iron Age and at even higher frequencies in Bronze Age Aegean and Minoan populations (J2a) and Bronze and Iron Age Balkan populations (J2b) (Supplementary Information 4). This signal challenges the hypothesis that J2 is a haplogroup specifically diagnostic for Phoenician expansions from the Levant. Instead, the high Y-chromosome diversity suggests that no single Y haplogroup can serve as an effective marker for Phoenician expansion.

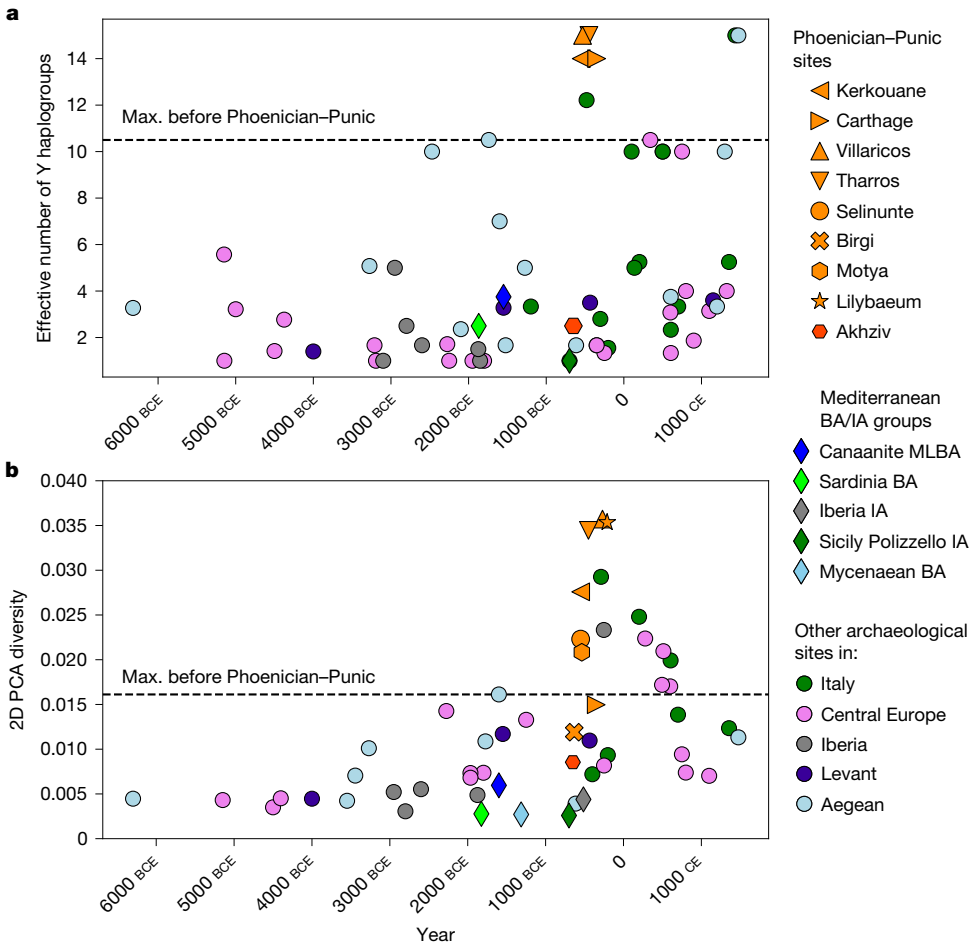


Fig. 3 | Autosomal and Y diversity per site in Phoenician–Punic contexts and the published aDNA record. a, We calculated the Y haplogroup diversity for contexts with at least five male individuals with sufficient data using Y haplogroups at the level of the first three characters of the ISOGG classification and the inverse Simpson index, also known as the effective number of types (Methods). **b**, We calculated the 2D PCA diversity for contexts with at least ten individuals with sufficient data using the first two PCs from Fig. 1 and the mean pairwise distance of those coordinates (Methods; values are listed in Supplementary Information 5). The dashed horizontal bar in both panels indicates the maximum (max.) diversity (autosomal or Y based) observed in sites before 500 BCE. The Mediterranean Bronze Age and Iron Age groups (diamonds) correspond to the ones depicted in Fig. 1. In each of these five groups, we included individuals from different archaeological sites.

Second, we observe high heterogeneity in genome-wide ancestry. The genetic variation in most Punic sites, as measured by the diversity of their positions in the PCA (Methods), is higher than in earlier sampled sites from all across the Mediterranean since the fifth millennium BCE (Fig. 3b). While varying North African ancestry is a primary driver of the high diversity we observe in Punic sites, Y haplogroup diversity levels in several sites remain higher than in earlier populations, even when this specific contribution is factored out (Extended Data Fig. 6 and Supplementary Information 5).

Third, our genetic data document that the individual ancestry variation was high within Punic sites, but this variation was overall similar between different regions, namely Sardinia, Iberia, Sicily and North Africa (Fig. 2 and Extended Data Fig. 1). The most parsimonious explanation for this overlap is that populations in different Punic sites drew from the same (or very similar) ancestry sources and had a high degree of interconnectedness, as it is unlikely that such a similar pattern of ancestry variation across individuals would be observed in geographically distant regions with differing local ancestries unless they shared a common ancestral source.

Fourth, we find evidence for mixing of people with different genetic backgrounds within extended pedigrees. We used long genomic segments that are identical by descent (IBD) among pairs of individuals to screen for biologically related individuals. In particular, multiple shared long IBD segments indicate biological relatives up to the fifth to seventh degree, and we could use this to infer a pedigree of five individuals from Kerkouane and a pedigree of three individuals from Tharros (Extended Data Fig. 7). Both pedigrees revealed extended families that integrated individuals with diverse ancestries, showing that ancestry mixing was ongoing in some Punic populations.

Fifth, we document multiple pairs of genealogical relatives separated by the Mediterranean Sea. We identify 31 pairs of individuals who share an IBD segment of length >16 centimorgan (cM), which indicates a shared ancestor probably within the last 20 generations (Methods and Supplementary Table 8). Most pairs (21 out of 31) are from within the same site (including the 2 pedigrees above) or adjacent sites, such as Birgi and Motya, but 10 pairs of individuals were from more distantly separated sites. Some of these pairs of individuals come from within the same geographical region (such as Málaga and Villaricos or Birgi and Selinunte). However, in six cases, we find IBD sharing across the Mediterranean, between an individual from Birgi or Motya (in Sicily) and an individual from Kerkouane or Carthage (North Africa), or Málaga (Iberia). This IBD sharing indicates a high enough rate of maritime mobility between Punic communities such that we could observe it multiple times in our dataset. This is particularly notable because the Mediterranean Sea is a geographical barrier that maintained high levels of genetic differentiation between populations in Sicily and North Africa

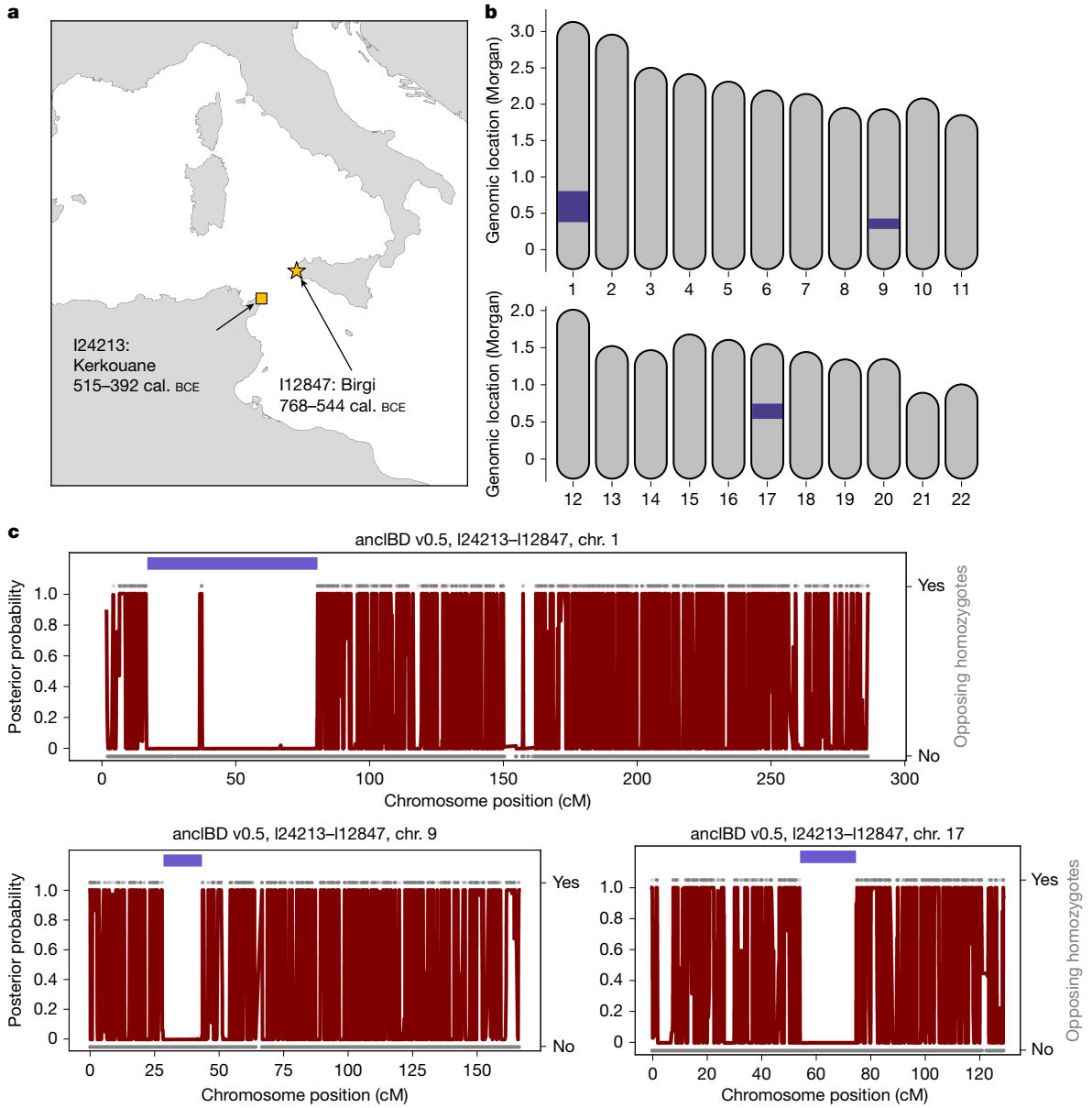


Fig. 4 | IBD segments shared between two individuals from Sicily and North Africa indicate that they were fifth-to-seventh-degree biological relatives. **a**, The sampling locations and radiocarbon date estimates of the two individuals. **b**, The genomic location of the position of three long IBD segments on the 22 autosomes. **c**, The posterior probability of not being in IBD along chromosomes 1, 9 and 17 as calculated using ancIBD (red). We also visualize opposing homozygotes (top versus bottom points, yes or no) for all SNPs of the two individuals that have an imputed genotype posterior probability of greater than 0.99. The dark blue segments indicate the inferred IBD segments. Our analysis suggests that both individuals had around 85% Sicilian–Aegean ancestry and 15% North African ancestry (highlighted in Fig. 2). We list all pairs of related Punic individuals in Supplementary Table 8, including five additional pairs linking sites separated by the Mediterranean Sea.

before Punic expansion and also in modern groups. One individual from Birgi and one individual from Kerkouane shared three long IBD segments, implying that they were fifth-to-seventh-degree relatives, such as second-to-third-degree cousins (Fig. 4 and Methods). Thus, these individuals or their immediate ancestors travelled at least once between North Africa and Sicily, making this relative pair a concrete case of population interconnectedness.

Sporadic endogamy within Punic sites

By analysing runs of homozygosity (ROHs) in an individual genome, we were able to measure the relatedness of the individual's parents and quantify the frequency of endogamy within Punic sites. Among the 90 individuals in our dataset with adequate data for such inference (Methods), we found 11 that have more than 50 cM of their genome covered by ROHs >20 cM long (Extended Data Fig. 8a), indicating parental relatedness on the level of first or second cousins[30–32]. Three individuals from Kerkouane, Carthage and Motya even had >350 cM of their genome covered by ROHs, suggesting that their parents were second-degree relatives (such as niece–uncle or half-siblings). This rate of detected close-kin unions (11 out of 86 analysed genomes) is higher than that observed in almost every other Bronze and Iron Age Mediterranean context (Extended Data Fig. 8b), except for Late Bronze Age Mycenaeans, who had similarly elevated rates of close-kin unions[33]. The fact that we find cases of close parental relatedness for individuals from six different Punic sites in all four sampled regions points to a geographically widespread practice of mating between close relatives. However, the overall low levels of ROHs in the large majority of analysed individuals

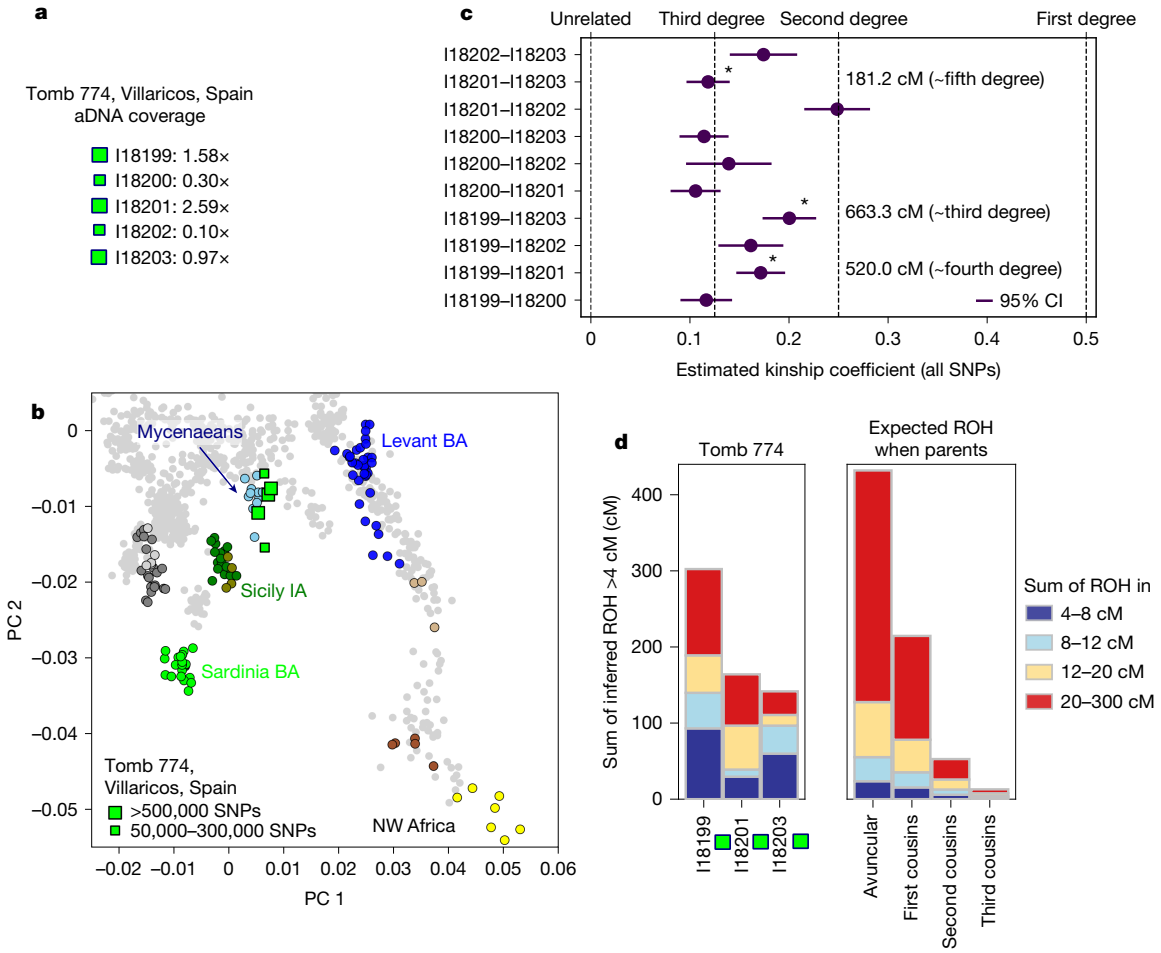


Fig. 5 | A Punic tomb from Villaricos contains individuals from an endogamous community. a, The mean sequencing depths for five individuals from Tomb 774 in Villaricos (Iberia), for which we obtained aDNA (average SNP coverage > 0.9× is indicated by large squares). **b**, PCA plot (as in Fig. 1) showing that the five individuals cluster close to Bronze Age Mycenaeans. **c**, The estimated kinship coefficients and their 95% confidence intervals (calculated using average pairwise mismatch rates) for all ten pairs of individuals (Methods) reveal a pattern of second-to-third-degree relatives. For the three individuals sequenced at sufficient coverage for IBD analysis (asterisks), we specify the total length of long IBD segments (>12 cM) shared by the three pairs, confirming that those pairs were closely related. **d**, The three individuals with high coverage exhibit exceptionally high levels of ROH, close to the level expected for the offspring of first cousins.

imply that this practice remained sporadic in most regions (with the possible exception of Iberia; Extended Data Fig. 8a), and indicates a large recent effective population size, incompatible with the hypothesis that a substantial fraction of the ancestry of Punic people derived from populations that had experienced an intense founder event[30,31].

A case study that provides insights into social practices among people in Iberian Punic sites who had offspring with their close relatives comes from Tomb 774 in Villaricos, which housed the remains of 18 individuals. We produced viable sequencing data for five individuals and dated them to the fifth to fourth centuries BCE (Supplementary Information 2). Our analysis shows that these five individuals cluster on the PCA with the primary mode of Punic ancestry (Sicilian–Aegean ancestry) and near Bronze Age Mycenaeans (Fig. 5b). The five individuals were related to each other. Every pair exhibits an average kinship coefficient typical of third-to-second-degree relatives. For the three individuals with sufficient sequence coverage, we directly observed close relationships through multiple long IBD segments (Fig. 5c). Moreover, we identified exceptionally high levels of long ROHs in the three individuals with sufficient coverage (Fig. 5d), indicating that their parents were close biological relatives. These observations suggest that the individuals in Tomb 774 were part of an intermating community of Sicilian–Aegean ancestry within the diverse Punic community of Villaricos. This tomb exhibits a unique material culture, including painted ostrich eggs typical of Punic funerary contexts and an ivory plaque depicting an Ionian-style capital associated initially with Greek themes[34]. However, there is no compelling reason to believe that the latter indicated the adoption of Greek cultural traditions, as Ionian capitals were a frequent feature of Punic iconography, and ivory working was a Phoenician–Punic specialization. This specific community probably extended beyond this particular tomb in Villaricos, as we found multiple short IBD segments shared between individuals from Tomb 774 and individuals from Málaga, who also exhibit an excess of ROHs and have similar ancestry patterns (Fig. 2, Extended Data Fig. 8a and Supplementary Table 8). These patterns contrast with the two reconstructed pedigrees in Kerkouane and Tharros (Extended Data Fig. 7), which provide examples of extended families in Punic sites incorporating diverse ancestries, highlighting the diverse social practices in the Punic world.

Discussion

Previous analyses of aDNA from Ibiza and Sardinia hypothesized a substantial proportion of eastern Mediterranean ancestry in people buried at Punic sites[20,21,35]. While a study of Punic people from Kerkouane in North Africa[19] did not find evidence of Levantine ancestry, it could not make a more general statement because it was limited to a single site. Our analysis of a sample size an order of magnitude larger and from five

times more Punic sites rules out substantial Levantine ancestry in all 21 previously published genomes and more than 80 newly sequenced individuals, including from prominent urban sites with evidence of Canaanite–Phoenician culture such as Carthage, Motya and Cádiz. Our diverse sample also enables us to trace the expansion of North African ancestry in Punic sites beyond case examples of previous studies[19,20], showing how, after around 400 BCE, North African ancestry spread to Sardinia and Iberia, and suggesting that expanding Carthaginian influence facilitated this spread.

Our results suggest that the Punic demographic expansion was primarily driven by the spread of people with Aegean-like ancestry. We conclude this because the dominant ancestry observed in our Punic dataset matches the ancestry observed in Bronze Age Aegean and Sicilian populations, and Bronze and Iron Age Sicilians in turn derived a large proportion of their ancestry from Aegean populations due to expansions in the Middle to Late Bronze Age and Iron Age[36]. The proximate origin of the Sicilian–Aegean ancestry in Punic people is difficult to resolve due to the low differentiation between Bronze Age Sicilian and Aegean populations ($F_{ST} \approx 0.01$; Supplementary Information 6) and due to sparse aDNA sampling in the Bronze and Iron Age eastern Mediterranean, particularly on the coast of Asia Minor and Cyprus. One possibility is that Aegean-like ancestry originated from early interactions of Phoenicians in Sicily with indigenous Sicilian populations harbouring such ancestry derived from earlier Bronze Age gene flow. Another possible source could be interactions between Punic people and Greek colonies established in the central Mediterranean and eastern North Africa (Cyrenaica) since the eighth century BCE. The proximity of Phoenician–Punic settlements in Sicily to some of these Greek colonies, such as Himera and Selinunte, would have provided further opportunities for gene flow.

A critical question raised by our results is how and when Canaanite–Phoenician culture and language were adopted by people without any detectable Levantine ancestry. One hypothesis is that, after Levantine Phoenicians founded settlements in the central and western Mediterranean in the early first millennium BCE, these communities continuously incorporated people with Sicilian–Aegean ancestry. As a result, most individuals living in these Punic settlements in the sixth century BCE or later would not have had detectable levels of Levantine ancestry. Cremation, the dominant funerary practice among Phoenician communities in the central and western Mediterranean before the sixth century BCE, makes obtaining viable samples for aDNA sequencing difficult, and we consequently do not have data from this period. The shift from cremation to inhumation as the preferred burial practice in the mid to late sixth century BCE is nevertheless also a cultural transition that could reflect substantial numbers of new people integrating into these communities[24].

Finally, the long-term demographic impact of the ancestry that we have documented here remains to be clarified. Our data from Roman-era individuals from Sicily (Extended Data Fig. 9c) suggest that ancestry patterns shifted toward the eastern Mediterranean, as previously observed elsewhere in the Roman Mediterranean[37–39]. Still, they only partially replaced the ancestry profile from the Punic period in Sicily. Studies that analyse Roman, Byzantine and Early Medieval material from Sicily, North Africa, Iberia and Sardinia, carried out in close collaboration between geneticists, archaeologists and historians, are critical to understanding these transitions that lastingly shaped the Mediterranean.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-08913-3.

1. Aubet, M. E. *The Phoenicians and the West: Politics, Colonies and Trade* (Cambridge Univ. Press, 2001).
2. Bondì, S. F., Garbati, G., Botto, M. & Oggiano, I. *Fenici e Cartaginesi: Una Civiltà Mediterranea* (Istituto poligrafico e Zecca dello Stato, Libreria dello Stato, 2009).
3. López-Ruiz, C. *Phoenicians and the Making of the Mediterranean* (Harvard Univ. Press, 2022).
4. Regev, D. *Painting the Mediterranean Phoenician: On Canaanite-Phoenician Trade-Nets* (Equinox, 2021).
5. Roppa, A., Botto, M. & Van Dommelen, P. *Il Mediterraneo Occidentale Dalla Fase Fenicia All'egemonia Cartaginese. Dinamiche Insediative, Forme Rituali e Cultura Materiale Nel V Secolo a. C* (Edizioni Quasar, 2021).
6. López-Ruiz, C. & Doak, B. R. *The Oxford Handbook of the Phoenician and Punic Mediterranean* (Oxford Univ. Press, 2019).
7. Hoyos, D. *Hannibal's Dynasty: Power and Politics in the Western Mediterranean, 247-183 BC* (Psychology Press, 2005).
8. Miles, R. *Carthage Must Be Destroyed: The Rise and Fall of an Ancient Civilization* (National Geographic Books, 2012).
9. Moscati, S. Fenicio o punico o cartaginese. *Rivista Studi Fenici* **16**, 3–14 (1988).
10. van Dommelen, P. A. R. & Bellard, C. G. *Rural Landscapes of the Punic World* (Equinox, 2008).
11. Doak, B. R. & López-Ruiz, C. in *The Oxford Handbook of the Phoenician and Punic Mediterranean* 3–8 (Oxford Academic, 2019); https://doi.org/10.1093/oxfordhb/9780190499341.013.1.
12. Prag, J. R. W. in *The Punic Mediterranean: Identities and Identification from Phoenician Settlement to Roman Rule* (eds Quinn, J. C. & Vella, N. C. E.) 11–23 (Cambridge Univ. Press, 2014).
13. Quinn, J. in *In Search of the Phoenicians* 25–40 (Princeton University Press, 2018).
14. Matisoo-Smith, E. et al. A European mitochondrial haplotype identified in ancient Phoenician remains from Carthage, North Africa. *PLoS ONE* **11**, e0155046 (2016).
15. Matisoo-Smith, E. et al. Ancient mitogenomes of Phoenicians from Sardinia and Lebanon: a story of settlement, integration, and female mobility. *PLoS ONE* **13**, e0190169 (2018).
16. Sarno, S. et al. Insights into Punic genetic signatures in the southern necropolis of Tharros (Sardinia). *Ann. Hum. Biol.* https://doi.org/10.1080/03014460.2021.1937699 (2021).
17. Haber, M. et al. Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. *Am. J. Hum. Genet.* **101**, 274–282 (2017).
18. Haber, M. et al. A genetic history of the Near East from an aDNA time course sampling eight points in the past 4,000 years. *Am. J. Hum. Genet.* **107**, 149–157 (2020).
19. Moots, H. M. et al. A genetic history of continuity and mobility in the Iron Age central Mediterranean. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-023-02143-4 (2023).
20. Marcus, J. H. et al. Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. *Nat. Commun.* **11**, 939 (2020).
21. Zalloua, P. et al. Ancient DNA of Phoenician remains indicates discontinuity in the settlement history of Ibiza. *Sci. Rep.* **8**, 17567 (2018).
22. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
23. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
24. López-Bertran, M. in *The Oxford Handbook of the Phoenician and Punic Mediterranean* 293–309 (Oxford Univ. Press, 2019).
25. Agranat-Tamir, L. et al. The genomic history of the Bronze Age Southern Levant. *Cell* **181**, 1146–1157 (2020).
26. Rodríguez-Varela, R. et al. Genomic analyses of pre-European conquest human remains from the Canary Islands reveal close affinity to modern North Africans. *Curr. Biol.* **28**, 1677–1679 (2018).
27. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
28. Harney, É., Patterson, N., Reich, D. & Wakeley, J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* **217**, iyaa045 (2021).
29. Zalloua, P. et al. Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *Am. J. Hum. Genet.* **83**, 633–642 (2008).
30. Ringbauer, H., Novembre, J. & Steinrücken, M. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2021).
31. Waldman, S. et al. Genome-wide data from medieval German Jews show that the Ashkenazi founder event pre-dated the 14th century. *Cell* **185**, 4703–4716 (2022).
32. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).
33. Skourtanioti, E. et al. Ancient DNA reveals admixture history and endogamy in the prehistoric Aegean. *Nat. Ecol. Evol.* **7**, 290–303 (2023).
34. Astruc, M. *La necropolis de Villaricos* (CSIC, 1951).
35. Ryan, S. E. et al. Growing up in Ancient Sardinia: infant-toddler dietary changes revealed by the novel use of hydrogen isotopes (δ^2H). *PLoS ONE* **15**, e0235080 (2020).
36. Fernandes, D. M. et al. The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nat. Ecol. Evol.* **4**, 334–345 (2020).
37. Antonio, M. L. et al. Ancient Rome: a genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714 (2019).
38. Lazaridis, I. et al. The genetic history of the Southern Arc: a bridge between West Asia and Europe. *Science* **377**, eabm4247 (2022).
39. Olalde, I. et al. A genetic history of the Balkans from Roman frontier to Slavic migrations. *Cell* **186**, 5472–5485 (2023).

[1]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. [2]Max Planck Harvard Research Center for the Archaeoscience of the Ancient Mediterranean (MHAAM), Leipzig, Germany. [3]Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [4]Efi Arazi School of Computer Science, Reichman University, Herzliya, Israel. [5]Department of Evolution and Ecology, University of California, Davis, CA, USA. [6]Israel Antiquities Authority, Jerusalem, Israel. [7]BIOMICs Research Group, Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, Vitoria-Gasteiz, Spain. [8]Ikerbasque—Basque Foundation of Science, Bilbao, Spain. [9]Dept. STEBICEF, Laboratory of Anthropology, University of Palermo, Palermo, Italy. [10]Dept. Culture e Società, University of Palermo, Palermo, Italy. [11]Joukowsky Institute for Archaeology and the Ancient World, Brown University, Providence, RI, USA. [12]Department of Genetics, Harvard Medical School, Boston, MA, USA. [13]University of Bologna, Bologna, Italy. [14]Archaeological Museum of Ibiza and Formentera, Eivissa, Spain. [15]Faculties of Medicine and Dental Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel. [16]Department of Biology, University of Florence, Florence, Italy. [17]Department of Archaeology and Ancient Near Eastern Civilizations, Tel Aviv University, Tel Aviv, Israel. [18]Department of Data Science, Mount Holyoke College, South Hadley, MA, USA. [19]Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. [20]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [21]Cranfield Forensic Institute, Cranfield University, Bedford, UK. [22]Universidad de Sevilla, Seville, Spain. [23]Archaeologist, independent researcher, Seville, Spain. [24]Universidad de Granada, Granada, Spain. [25]Museo Arqueológico de Granada, Granada, Spain. [26]Universidad de Málaga, Málaga, Spain. [27]Institute of Evolutionary Biology (UPF-CSIC), PRBB, Barcelona, Spain. [28]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain. [29]CNAG, Centro Nacional de Analisis Genomico, Barcelona, Spain. [30]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain. [31]Institut Català de Paleontologia Miquel Crusafont (ICP-CERCA), Universitat Autònoma de Barcelona, Barcelona, Spain. [32]Unidad de Paleobiología, ICP-CERCA, Unidad Asociada al CSIC por el IBE UPF-CSIC, Barcelona, Spain. [33]Departament de Medicina i Ciències de la Vida, Institut de Biologia Evolutiva (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Spain. [34]Museo Arqueológico Nacional Madrid, Madrid, Spain. [35]Ministry of Cultural Heritage, Palermo, Italy. [36]The Giuseppe Whitaker Foundation, Motya, Italy. [37]University of Palermo, Department of Archaeology, Palermo, Italy. [38]Bar Ilan University, The Azrieli Faculty of Medicine, Safed, Israel. [39]Department of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. [40]Weizmann Institute of Science, Scientific Archaeology Unit, D-REAMS Radiocarbon Dating Laboratory, Rehovot, Israel. [41]The National Natural History Collections, The Hebrew University of Jerusalem, Jerusalem, Israel. [42]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy. [43]Department of Odontostomatological and Maxillofacial Sciences, Sapienza University of Rome, Rome, Italy. [44]Italian Institute of Oriental Studies, Sapienza University of Rome, Rome, Italy. [45]Department of Literature, Languages and Cultural Heritage, University of Cagliari, Cagliari, Italy. [46]Department of Cultural Heritage, University of Bologna, Ravenna, Italy. [47]Institute of Energy and the Environment, Penn State University, University Park, PA, USA. [48]Department of Ancient World Studies, Sapienza University of Rome, Rome, Italy. [49]Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria. [50]Department of History, Anthropology, Religion, Arts and Performing Arts, Sapienza University of Rome, Rome, Italy. [51]Department of Law and Digital Society, Unitelma Sapienza, Rome, Italy. [52]Human Evolution and Archaeological Sciences - HEAS, University of Vienna, Vienna, Austria. [53]Natural Sciences Museum of Barcelona, Barcelona, Spain. [54]These authors jointly supervised this work: Alfredo Coppa, David Caramelli, Ron Pinhasi, Carles Lalueza-Fox, Ilan Gronau, David Reich. ✉e-mail: harald_ringbauer@eva.mpg.de; ilan.gronau@runi.ac.il; reich@genetics.med.harvard.edu

Methods

Laboratory work, sequencing and quality control

In clean rooms where the goal was to protect the analysed samples from contamination by environmental DNA, we sampled bones (especially petrous bones[40]) and teeth, typically aiming for 37 mg of powder per sample. We extracted DNA using a methodology optimized to retain smaller DNA fragments[41,42]. We applied uracil-DNA-glycosylase (UDG) treatment to reduce characteristic aDNA errors[43,44] and then built double- and single-stranded libraries[45]. We enriched these libraries in solution for more than 1 million SNPs[22,46] and the mitochondrial genome[23]. We sequenced the enriched products aiming for approximately 30 million paired sequences each. We also sequenced on the order of a couple of hundred thousand sequences for the unenriched libraries.

To analyse the data bioinformatically, we assigned sequences to individuals based on the associated barcodes and indices that we attached to them during library preparation, and merged sequences that overlapped by at least 15 bases (allowing up to one mismatch) using a modified version of SeqPrep v.1.1 (https://github.com/jstjohn/SeqPrep). At overlapping bases, we used the allele call from the higher-quality base. We restricted our analysis to sequences of at least 30 bp in length, which aligned with a minimum mapping quality of least 10 to either the inferred ancestral mitochondrial genome sequence[47,48] or the human reference genome sequence (hg19) (https://www.internationalgenome.org/category/grch37/), using the 'samse' command from BWA[47]. We removed duplicate sequences based on aligning to the same locations and having the same in-line barcode information. To reduce the effects of characteristic aDNA damage, we trimmed UDG-treated sequences by 2 bp on either end and non-UDG-treated sequences by 10 bp on either end. We then represented each targeted position covered by at least one sequence with a base quality of at least 20 at that location by randomly selecting a single sequence (pseudo-haploid genotyping).

We built a mtDNA consensus sequence using bcftools (https://github.com/samtools/bcftools) and SAMTools[49], analysing only sites with a minimum of twofold coverage and determining allelic status by majority rule. We used Haplogrep2[50] and the phylotree database (mtDNA tree build 17) to determine the mitochondrial haplogroup.

To assess evidence of aDNA authenticity, we estimated a 95% confidence interval for contamination on the mitochondrial genome based on the mismatch rate to the consensus sequence, using contamMix-1.0.1051[23]. To estimate a 95% confidence interval for contamination in the X chromosome in males (who should have no variation in the non-pseudoautosomal regions of the X chromosome in the absence of contamination), we used ANGSD[51]. We measured the cytosine-to-thymine mismatch rate to the consensus sequence at the terminal ends of sequences to determine whether libraries had the expected damage profile for authentic aDNA. We also used the ratio of the Y chromosome to the X and Y chromosome to determine whether the individual had a ratio consistent with being from a person with two X chromosomes (molecular female, <0.03) or a person with an X and a Y chromosome (molecular male, >0.35); libraries with a ratio in between have potential evidence of contamination.

Radiocarbon dating

We obtained 111 accelerator mass-spectrometry-based dates on bone for 99 distinct human remains in specialized C14 laboratories at Pennsylvania State University (82 measurements), the D-REAMS laboratory at the Weizmann Institute of Science (24), Oxford (4) and CIRAM (1). For the CIRAM date, the different carbon isotopes were separated in a joint venture with JSC Barnas using a 250 kV accelerator mass spectrometer from FTMC in Vilnius, Lithuania. We report laboratory codes together with raw and calibrated measurements, including conventional radiocarbon ages and their standard errors (Supplementary Table 4). We used the calibration curve IntCal20[52] to calibrate conventional radiocarbon dates using the software OxCal. Eleven individuals were analysed by two different laboratories, with date ranges from the two laboratories being concordant in all cases. We used the R_combine method to combine date ranges into a single and more precisely constrained range for these individuals.

Merging with published aDNA data

We merged the newly generated aDNA data with a collection of previously published ancient genomes compiled in version 54.1 of the Allen Ancient DNA Resource[53]. We generated a primary dataset of pseudohaploid genotype data, with one allele picked randomly from sequences overlapping that position at the targeted SNPs (1240k). We used the resulting data in Eigenstrat format as the basis for our population genetic analysis.

Dataset partitions of ancient individuals

We categorized ancient individuals in the aDNA dataset based on sequence coverage and the certainty of provenance. Our analysis in this study mainly focuses on 108 newly sequenced ancient individuals with more than 20,000 targeted autosomal SNPs covered by at least one sequence and its Phoenician or Punic association determined by the archaeological context of a site and its radiocarbon dates. We applied a series of criteria to ensure that our results did not include individuals associated with the Roman expansion following the Punic Wars (Supplementary Information 2). We identified an additional set of 23 individuals with slightly reduced confidence in context: 10 individuals from Tharros, Birgi, Motya and Lilybaeum who had radiocarbon date ranges that overlap the third and second centuries BCE, and 13 individuals from Tharros and sites in Iberia who did not have radiocarbon dates. We typically analysed them separately because some of these individuals may post-date the Roman expansion after the Punic Wars. We added nine previously published individuals to these two groups, of whom eight had radiocarbon dates[20,21,35]).

We grouped our sample into five broad geographical regions: Iberia (including the Iberian Peninsula and the island of Ibiza), Sardinia, Sicily, North Africa and Akhziv. We also grouped our sample into two primary time ranges. The radiocarbon calibration curve for the first millennium BCE has several plateaus spanning hundreds of years, particularly the 'Hallstatt plateau' 800–400 BCE and a subsequent plateau 400–200 BCE. Consequently, radiocarbon dating cannot resolve sample dates within each plateau. We therefore associate each individual with one of these plateau time ranges (Supplementary Information 2).

Principal component analysis

We computed PCs of modern individuals genotyped on the Affymetrix Human Origin SNP array with the software smartpca (v.18150) using the default settings, combining data from a standard set of West Eurasians (HO data, widely used in human aDNA analysis) with present-day North-African individuals in AADR v.54.1 (the list of 1,196 individuals is provided in Supplementary Table 6) as described previously[54]. We then projected onto the first two PCs ancient individuals with data for at least 20,000 SNPs on the autosomal targets, using least-square projection (setting, lsqproject: YES) and shrinkage correction (setting, shrinkmode: YES) and aDNA data in pseudohaploid eigenstrat format.

Ancestry modelling with ADMIXTURE

We ran ADMIXTURE v.1.3.0[55] in unsupervised mode on the 122 individuals from our Phoenician–Punic dataset who were sequenced for more than 100,000 SNPs, together with 24 individuals from related populations around the Mediterranean (Supplementary Table 12). Following a previous study[56], we pruned SNPs in linkage disequilibrium by removing one SNP from every pair of SNPs that (1) were within a genomic window that contained at most 200 SNPs and (2) had an r^2 value of association greater than 0.4. This pruning was done by applying PLINK v.1.9[57] with the following options: --indep-pairwise 200 25 0.4. This process retained 452,215 SNPs out of the original 1,233,013. We applied ADMIXTURE to

the pruned dataset with $K = 2$, 3, 4 or 5 latent ancestry components and, for each value of K, we executed 50 replicate analyses with different random seeds. We used the ΔK score described previously[58], as implemented in Clumpak[59] to find the value of K that provided the best fit of the data and used PONG (v.1.5)[60] to visualize the results.

General setup for qpAdm analysis

We estimated proportions of ancestry with qpAdm[27,28], using the R package admixr[61] and qpAdm v.1201 (from ADMIXTOOLS) with the allsnps flag turned on. We modelled background ancestry using 23 ancient individuals partitioned into 14 groups (Supplementary Table 7): El Mirón, Ust Ishim, Kostenki, GoyetQ116-1, Vestonice, MA1, Villabruna, CHG (2 individuals), EHG (3 individuals), WHG (2 individuals), Mota, Natufian, Levant_N (4 individuals) and Tunisia_N (3 individuals). These 14 groups were used as 'right populations' in all of the admixture models that we considered, with El Miron always used as the first (anchor) right population. Considering the main clusters observed in the PCA, we determined eight groups of ancient individuals to use as proxies for potential sources of ancestry (Supplementary Table 7): Greece BA (Myc), Sicily Early–Middle Bronze Age (EMBA), Sardinia LBA, Iberia Early Bronze Age (EBA), Steppe MLBA, Iran N, Levant MLBA and North Africa IA. As a proxy for North African ancestry, we use a single individual from Iron Age Algeria sequenced to high coverage (>750,000 SNPs). As three Neolithic individuals from Tunisia are used to model background ancestry, North African ancestry inferred by our qpAdm models should reflect an African source population more closely related to the Iron Age individual from Algeria than to the Neolithic individuals from Tunisia (Supplementary Information 3). Each of the other 7 groups contains 6–24 individuals from a similar archaeological context, who were sequenced with high coverage (>100,000 SNPs) and form tight clusters in the 2D PCA (Fig. 1). In each admixture model that we examined, some of these eight groups were used as proxies to source populations (left populations), others were added to the model of background ancestry (right populations) and the remaining populations were left out of the analysis. An admixture model was considered to be feasible if all proxy sources in that model were associated with non-negative admixture proportions. A feasible model was considered to be valid if the P value provided by the statistical test of qpAdm was at least 0.05 (indicating that this model cannot be rejected with high probability).

Types of qpAdm ancestry models

In preliminary analysis with qpAdm, we considered broad ancestry models, which used only the basic set of 14 groups as right populations (see above). We considered all 255 non-empty subsets of the eight potential proxy sources as left populations. In these broad ancestry models, we observed that each individual was fit by many valid admixture models with conflicting interpretations (Supplementary Information 3). We therefore decided to consider two separate and distinct types of admixture models. In the western ancestry models, we added the Levant MLBA group to the core set of 14 right populations. We considered the 127 non-empty subsets of the 7 remaining proxy sources as left populations. In the eastern ancestry models, we considered the 7 non-empty subsets of Levant MLBA, Iran N and North Africa IA as left populations, and the remaining 5 (western) proxy sources were added to the core set of 14 right populations. This approach produced interpretable valid admixture models for 135 out of the 140 individuals in our Phoenician–Punic dataset. To complete the picture, we provide admixture models inferred by the broad ancestry model described above to four of the remaining five individuals. This left only one individual (I22122 from Tharros), for whom we could not fit a valid admixture model.

Parsimonious ancestry models

The qpAdm analysis considered 134 different models for each individual (127 western models and 7 eastern models; details are provided above). As a result, many individuals obtained multiple valid ancestry models (models with non-negative ancestry proportions and P values above 0.05). We applied the following parsimonious approach to reduce the number of valid models we report for each individual. For a given model M, let pop(M) denote the set of left populations used in this model. Now, consider two models, M and M', that differ by one proxy source: pop(M) = pop(M') ∪ {p}. If both models are valid for a given individual, and the fit provided by model M is not significantly better than that provided by the more restrictive model M', then we discard model M from the set of models that we report for that individual. Formally, model M is considered to be non-parsimonious for a given individual if there is a model M' that satisfies the following three requirements: (1) pop(M') ∪ {p} = pop(M); (2) model M' is valid for the individual of interest; and (3) the P value of M is not significantly larger than the P value of M'. The difference between P values is evaluated using a standard likelihood ratio test between the two nested models. Using this approach, we removed non-parsimonious models from the set of ancestry models that we report for each individual (Extended Data Fig. 3). Ancestry contributions in the parsimonious models that we report should therefore be considered to be significant in the sense that a model without them was rejected by qpAdm.

Representative ancestry models

We applied several simplifications when presenting a representative admixture model for individuals in Fig. 2. First, we combined the ancestry proportions inferred for Greece BA (Myc) and Sicily EMBA (Sicilian–Aegean ancestry), and the ancestry proportions inferred for Sardinia LBA and Iberia EBA (western Mediterranean ancestry). These ancestries were grouped due to the limited ability of our qpAdm models to distinguish between them (Supplementary Information 3) and their genetic similarity (Supplementary Information 6). As individuals were typically inferred to have more than one valid admixture model, we selected a representative model for each individual that maximized the main source of ancestry. In particular, among eastern ancestry models, we selected the model with the largest proportion of Levantine ancestry. Among western ancestry models, we selected the model with the largest proportion of Sicilian–Aegean ancestry. For the six individuals for whom we could fit both eastern and western models, we selected the model for which this main contribution (Sicilian–Aegean or Levantine) was the largest. Thus, the selected representative model reflects the maximum amount of ancestry attributed to the main source of ancestry.

Inferring uniparental haplogroups

We used automatic Y haplogroup calling using a script described previously[38] based on the YFull YTree v.8.09 phylogeny and SNPs from ISOGG YBrowse6. For cases with low coverage or unusually shallow assignments, we double-checked the assignments by inspecting all derived SNPs—assigning haplogroups in cases of clear paths to the most derived haplogroups—allowing for mismatches due to sporadic aDNA damage (that is, C to T or G to A). We annotate all haplogroups in ISOGG19 annotation. We determined mitochondrial haplogroups using Haplogrep (v.2.1.1)[50] for sequences aligning to the mitochondrial genome (RSRS), using all sequencing data aligning to the mitochondrial genome RSRS.

Calculating autosomal and Y haplogroup diversity per site

We calculated a proxy for the autosomal diversity at each site based on the first two PCs of our primary PCA (as described above) based on the rationale that those two PCs broadly reflect a genetic map of the Mediterranean. For each site with at least ten individuals with at least 20,000 1240k SNPs covered, we calculated the pairwise Euclidean distance of all pairs of individuals and then took the mean value as a measure for the per-site diversity.

To quantify Y haplogroup diversity per archaeological site, we considered the first three characters of the ISOGG classification for each

inferred haplogroup. We used this particular cut-off for two reasons: (1) low-coverage aDNA does not consistently allow assigning Y haplogroups at more refined levels due to missing data and genotyping errors caused by aDNA damage; (2) the specific 1240k SNP set enriched in these genome sequences has little resolution for several macro-haplogroups beyond this level (for example, I1). We calculated the inverse of the probability that two randomly selected male individuals from a given site have the same Y haplogroups. This measure is known as the inverse Simpson index or the effective number of types (that is, the number of equally abundant types needed to obtain the same diversity index). We require at least five male individuals for each site, using 100,000 autosomal SNPs covered as a cut-off for sufficient genomic data for Y haplogroup calls.

For both diversity measures, in Punic sites, we included individuals with a date consistent with the Punic period (labelled Punic_Early, Punic_Late, Punic_Late2, Punic_NoRC in the column labelPCA in Supplementary Table 5). Including individuals dating to 350–50 calibrated years BCE and individuals without radiocarbon dates was necessary to obtain statistically meaningful sample sizes for quantifying diversity.

Detecting shared haplotypes (IBD and ROH segments)

To infer pairwise shared identity by descent segments (IBD), we used the software ancIBD[62], which is designed for aDNA data. We started from processed .bam files containing all aligned reads after quality control. We then imputed genotype probabilities with the software GLIMPSE, using the same imputation pipeline as described previously[62] based on the 1000 Genome reference panel. Using the recommended default settings of ancIBD, we inferred all IBD segments longer than 8 cM in all pairs of individuals with the recommended quality cut-off: at least 70% of imputed SNPs on chromosome 3 imputed with maximum genotype probability > 0.99, which corresponds broadly to at least 600,000 SNPs covered for 1240k SNP capture data (Supplementary Table 9). We report summary statistics of IBD segments per pair of individuals in Supplementary Table 8. To infer degrees of relatedness, we compared inferred IBD segments >12 cM long to IBD observed in simulated genomes with various degrees of relatedness when using the same filtering as when running ancIBD, as described previously[62]. We considered a pair of individuals to be significantly related if they shared an IBD segment of length >16 cM. Two individuals with a shared segment of 16 cM long have a common ancestor within 20 generations with high probability. Indeed, assuming a panmictic population of diploid individuals with constant size $N_e = 1{,}000$, 95% of such IBD segments originate from a common ancestor within up to 19.3 generation ago, with a median of 8.0 generations ago, calculated as described in supplementary note 4 of ref. 30. Varying N_e only has a small effect on these numbers, and longer IBD segments originate from even more recent ancestors.

We inferred ROHs using the software hapROH, which was designed and calibrated for aDNA data[30]. We applied hapROH with its recommended default settings to pseudohaploid data on 1240k SNPs using the modern 1000 Genomes haplotype panel as a reference panel. We used the recommended cut-off of at least 400,000 covered 1240k SNPs as a minimum-coverage cut-off and inferred ROH > 4 cM long. We report summary statistics of ROH segments per individual in Supplementary Table 10.

Inferring pairwise genetic relatedness

We computed pairwise mismatch rates by randomly sampling one read for each individual at the autosomal SNPs of the 1240k panel and computed 95% confidence intervals using block jackknife standard errors (Supplementary Table 11). For individuals in T774 from Villaricos, we then estimated kinship coefficients r as described previously[63]: $r = 1 - (2 \times (x - (b/2))/b)$, where x denotes the pairwise mismatch rate for that pair and b represents the mismatch rate expected for two unrelated individuals from the same ancestral background. We estimated b as the mean of the mismatch rates between individuals from T774 and Bronze–Iron Age individuals with similar ancestry from Empúries[54] and mainland Greece[38,54], who are very unlikely to be close relatives to those from T774.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Open science principles require making all data used to support the conclusions of a study fully available, and we support these principles here by making publicly available not only the digital copies of molecules (the uploaded sequences) but also the molecular copies (the aDNA libraries themselves, which constitute molecular data storage). Researchers who wish to carry out deeper sequencing of libraries published in this study should make a request to the corresponding author D.R. We commit to granting reasonable requests as long as the libraries remain preserved in our laboratories, with no requirement that we be included as collaborators or co-authors on any resulting publications. The raw DNA sequences for individuals newly sequenced in this study are deposited in the European Nucleotide Archive under accession number PRJEB86313. Their processed genotype data in pseudohaploid eigenstrat format can be obtained from the Harvard Dataverse repository (https://doi.org/10.7910/DVN/UPDESR). We include other newly reported data such as radiocarbon dates and archaeological context information in the Article and its Supplementary Information. We plotted the maps in Figs. 1 and 4 using the Python package basemap. The land-sea mask, coastline, lake, river and political boundary data are extracted from the GSHHG datasets (v.2.3.6) using GMT (5.x series). They are included under the terms of the open-source LGPLv3+ license.

Code availability

We deposited the code for analysing the data and producing the figures in this Article at GitHub (https://github.com/hringbauer/punic_aDNA).

40. Pinhasi, R. et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102 (2015).
41. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
42. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
43. Briggs, A. W. et al. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
44. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
45. Gansauge, M.-T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* **15**, 2279–2300 (2020).
46. Rohland, N. et al. Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. *Genome Res.* **32**, 2068–2078 (2022).
47. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
48. Behar, D. M. et al. A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
49. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
50. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
51. Rasmussen, M. et al. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
52. Reimer, P. J. et al. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
53. Mallick, S. et al. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci. Data* **11**, 182 (2024).
54. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
55. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

56. Lazaridis, I. et al. Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218 (2017).
57. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
58. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
59. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
60. Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P. & Ramachandran, S. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016).
61. Petr, M., Vernot, B. & Kelso, J. admixr-R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics* **35**, 3194–3195 (2019).
62. Ringbauer, H. et al. Accurate detection of identity-by-descent segments in human ancient DNA. *Nat. Genet.* **56**, 143–151 (2024).
63. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
64. van den Brink, E. C. M. et al. A Late Bronze Age II clay coffin from Tel Shaddud in the Central Jezreel Valley, Israel: context and historical implications. *Levantina* **49**, 105–135 (2017).
65. Feldman, M. et al. Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. *Sci. Adv.* **5**, eaax0061 (2019).

Acknowledgements C.L.-F. was supported by PID2021-124590NB-100 grant (MCIU/AEI/FEDER, UE). H.R. was supported by the Max Planck–Harvard Research Center for the Archaeoscience of the Ancient Mediterranean (MHAAM). I.G., A.S.-M. and D. Regev were supported by ISF grant number 1045/20. L.S., A.C.F., C.D.V. and D.L. were supported by AGED PRIN 2017 project, MUR Italy. D. Reich was supported by National Institutes of Health grant HG012287; by John Templeton Foundation grant 61220; by the Howard Hughes Medical Institute (HHMI), a gift from J.-F. Clin; by the Allen Discovery Center, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation; and by a grant from the Getty Foundation "The Classical World in Context: the Near East". We thank V. Moses and M. McCormick for their comments. The project received funding from the Italian Ministry of Foreign Affairs and International Cooperation, and ISMEO. The D-REAMS Radiocarbon analysis was supported by the Exilarch Foundation for the Dangoor Research Accelerator Mass Spectrometer. We thank the Musée de l'Homme for providing us access to the human remains from Khenchela cave.

Author contributions We annotate author contributions using the CRediT Taxonomy labels; where multiple authors serve in the same role, we specify the degree of contribution as lead, equal or support. Conceptualization (design of study): D. Reich, I.G. and D. Regev. Data curation (archaeology and bioanthropology): lead, R.P., C.L.-F., D. Regev, A.C., L.S., D.C., P.v.D. and D.P. Data curation (DNA): lead, H.R., I.G., D. Reich and S.M. Formal analysis: lead, H.R. and I.G.; support, A.S.-M., I.O., T.P., A. Mittnik, I.L., A.S. and S.M. Funding acquisition: D. Reich and I.G. Resources: G.F., M.B., A. Mezquida, B.C., H.J., P. Smith, S.V., A. Modi, K.C., E. Curtis, A.K., A.M.L., M.M., A. Micco, J.O., L.Q., K.S., J.N.W., N.M.-G., A.M.S.R., M.L.L.F., J.M.J.-A., I.J.T.M., E.V., J.S.P., S.L.C., T.M.-B., E.L., A.R.R., F.O., P.T., V.G., A.B., L.C., E.B., M.F., M.L., F.L.P., A.N., F.G., C.D.V., G.L., F.M., P. Sconzo, G.C., E. Cilli, A.C.F., F.F., D.L., B.J.C., N.R. and L.N. Supervision: lead, D. Reich and I.G. Visualization: lead, H.R., I.G. and A.S.-M. Writing (original draft preparation): H.R., I.G. and D. Reich. Writing (review and editing): all of the authors.

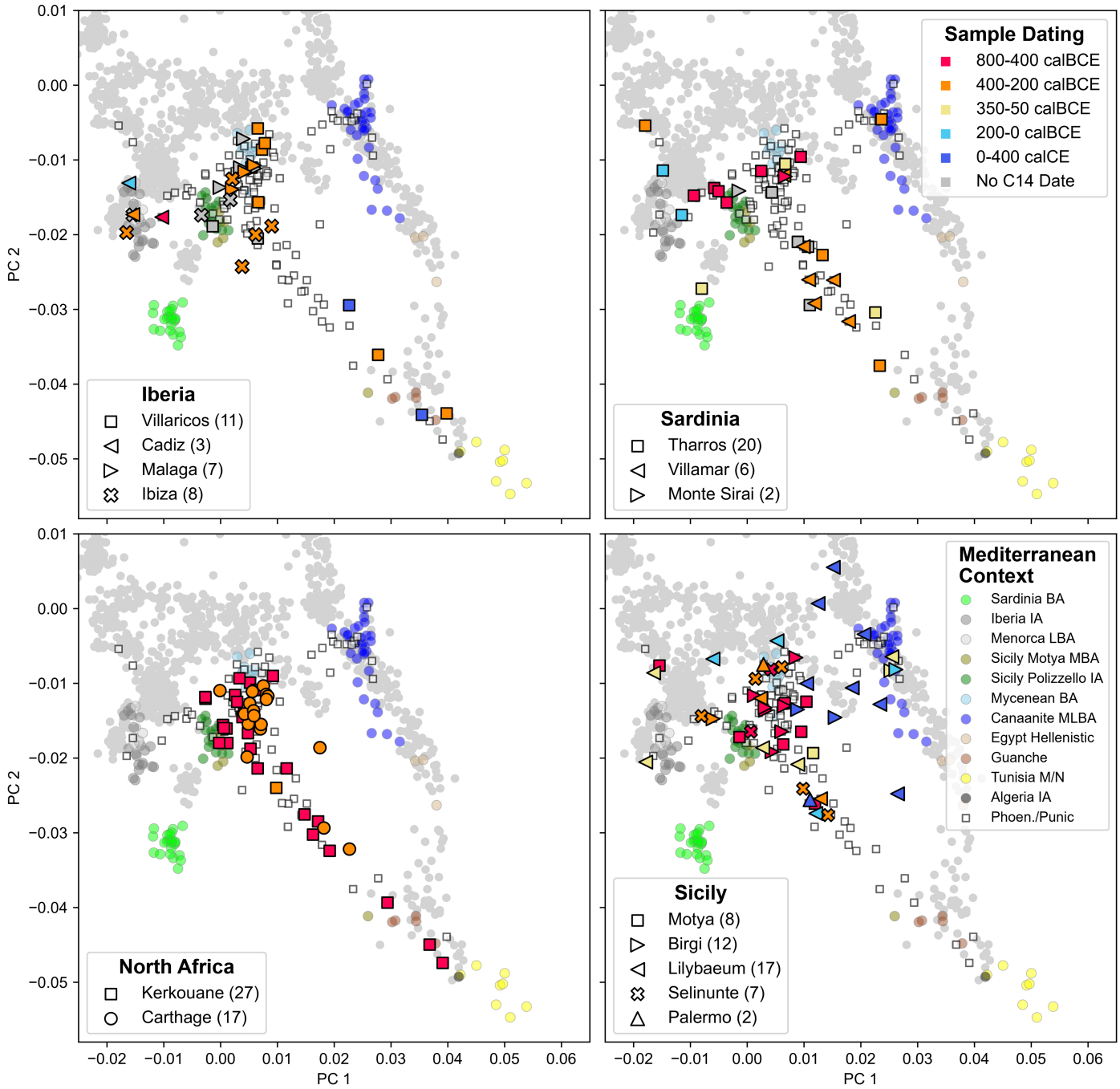
Competing interests The authors declare no competing interests.

Additional information

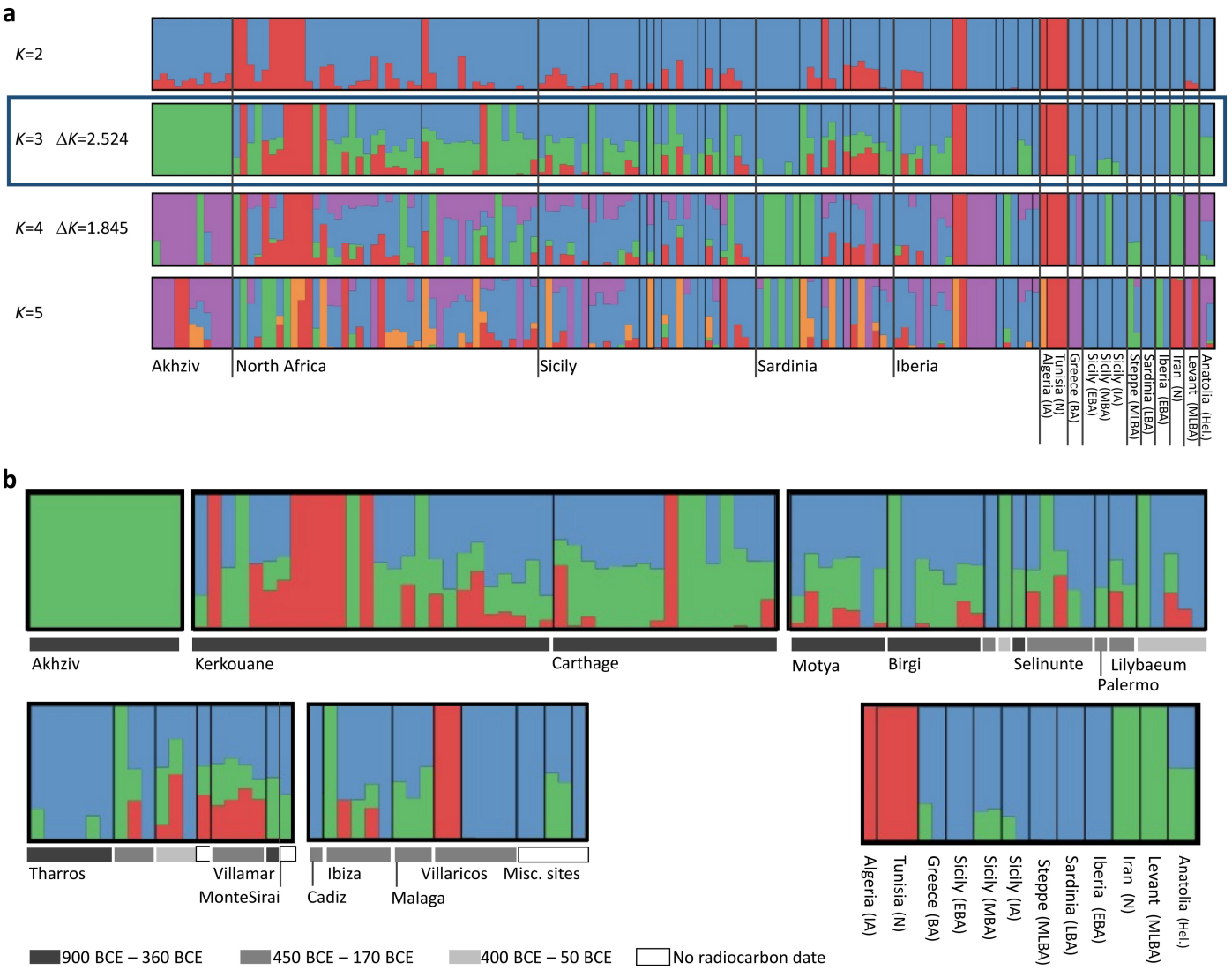
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-08913-3.

Correspondence and requests for materials should be addressed to Harald Ringbauer, Ilan Gronau or David Reich.

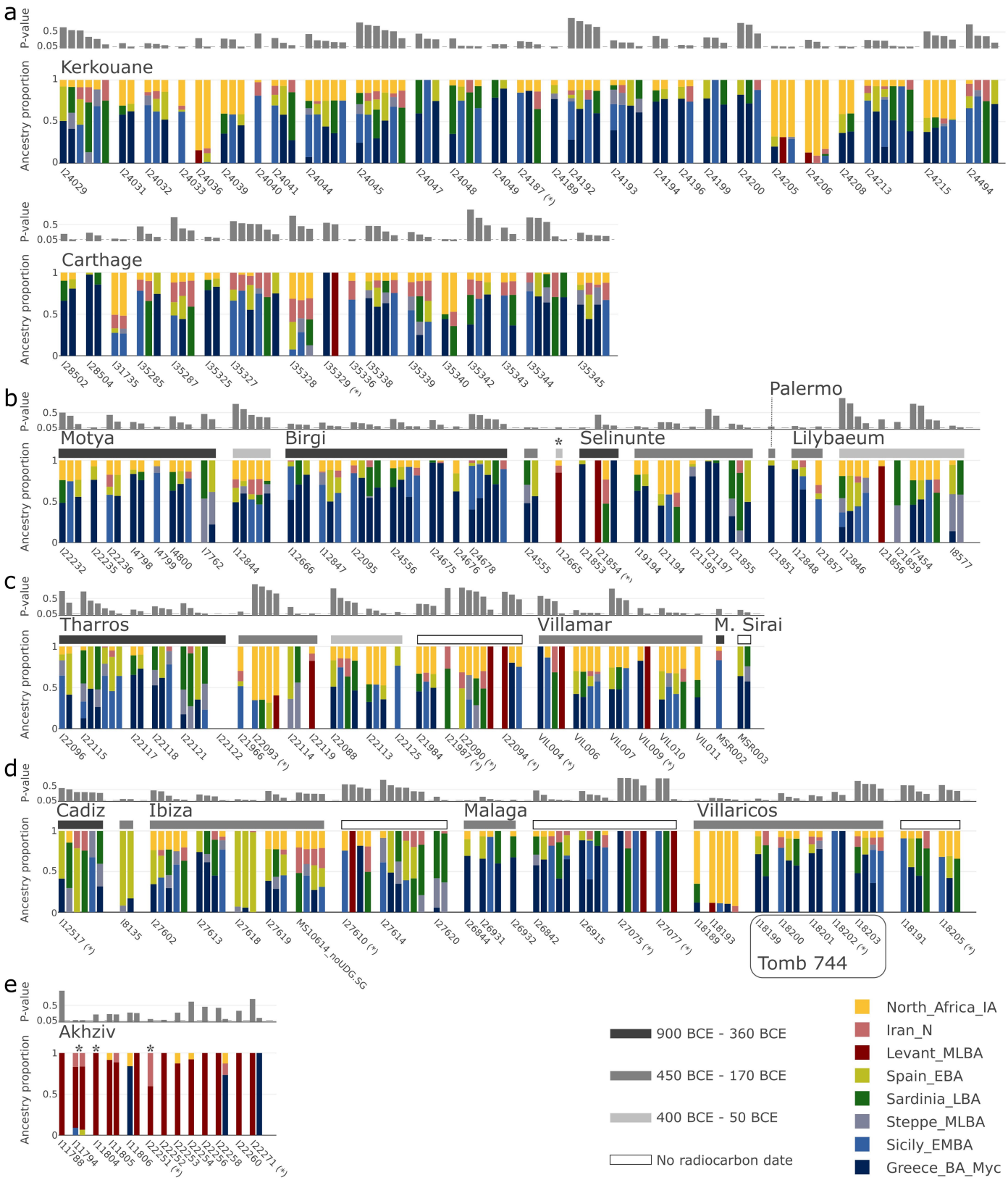
Peer review information *Nature* thanks Timothy Jull, Josephine Quinn and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



Extended Data Fig. 1 | PCA of Punic individuals grouped according to site and chronology. We project individuals sequenced at over 20,000 SNPs onto the same two PCs as in Fig. 1 (calculated from modern individuals, grey dots). Here, we plot all individuals from Punic archaeological sites, excluding the Phoenician site of Akhziv and including the 20 individuals dated to the Roman period (Supplementary Table 5). We split the sample into panels representing our four major geographic regions: Iberia, Sardinia, North Africa, and Sicily. The shape of the symbols indicates the site (lower left legend), and the colour indicates the date range of each individual (as described in the upper right legend).

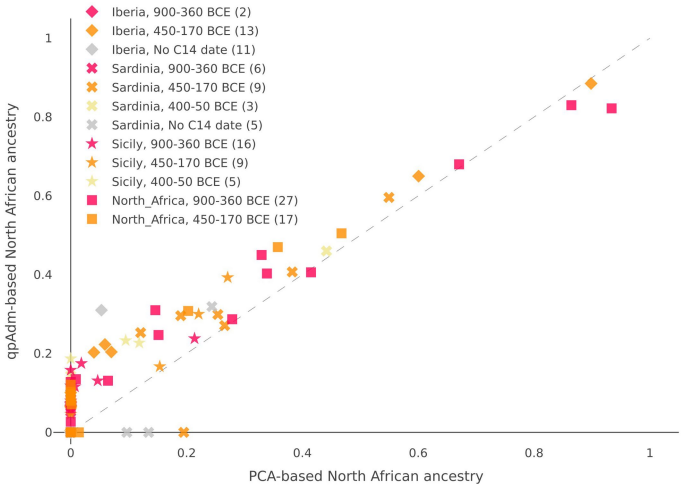


Extended Data Fig. 2 | Ancestry models inferred for 122 Phoenician-Punic individuals using unsupervised *ADMIXTURE* with $K = 2, 3, 4$ and 5 latent ancestry components. The 122 Phoenician-Punic individuals sequenced for more than 100,000 SNPs were jointly analysed with 24 individuals from related ancient populations across the Mediterranean (Supplementary Table 12). **(a)** The model with the highest likelihood was obtained for each value of K among 50 replicate runs. Individuals are grouped based on region. Values of the ΔK score of[58] are specified for $K = 3,4$, with a higher score obtained for $K = 3$, suggesting optimal fit. **(b)** A more detailed depiction of the best model obtained with $K = 3$ latent ancestry components corresponding to North African ancestry (red), eastern ancestry (green), and central/western Mediterranean ancestry (blue). Individuals are partitioned within each region according to site and time range (see legend). The unsupervised *ADMIXTURE* model does not adequately differentiate between Levantine ancestry and ancestry found in other Mediterranean locations (e.g., Anatolia and Sicily), unlike the *qpAdm* models of Extended Data Fig. 3.

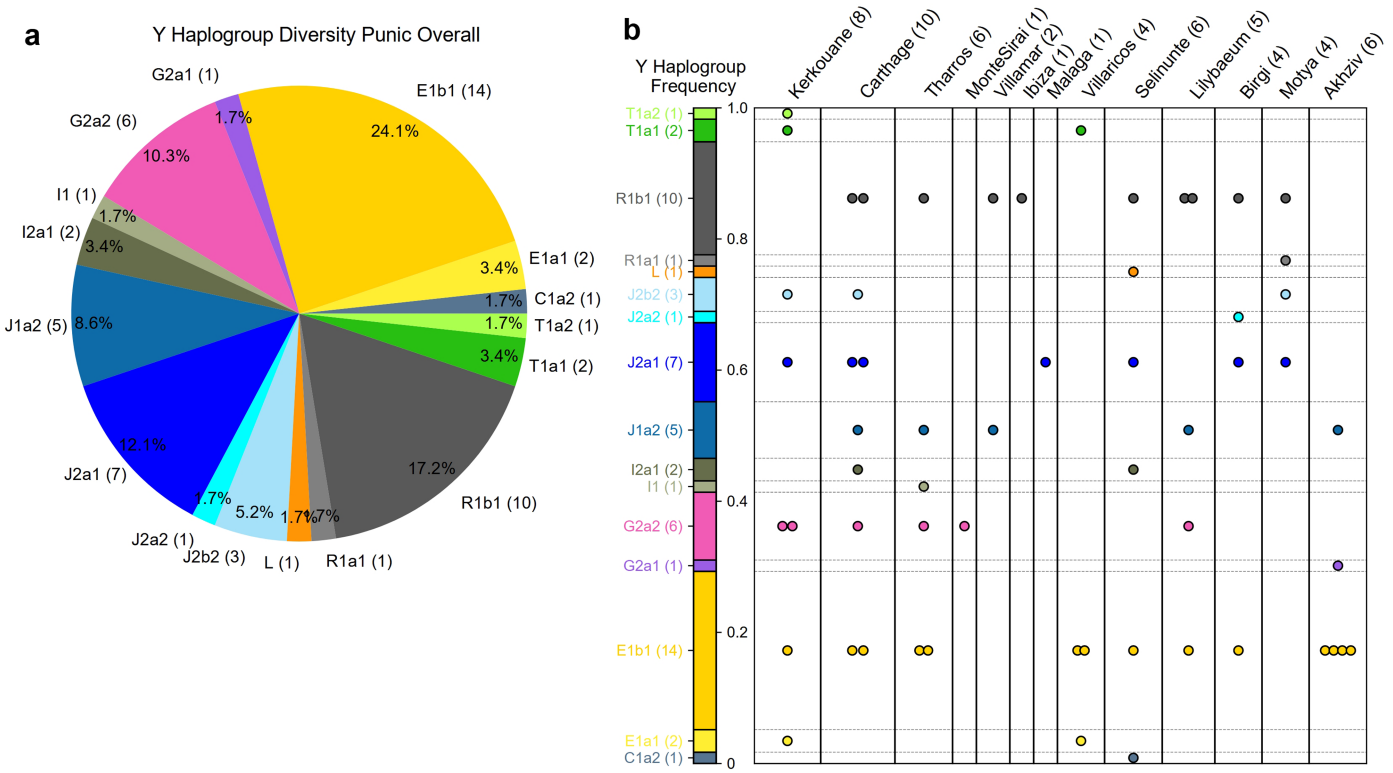


Extended Data Fig. 3 | See next page for caption.

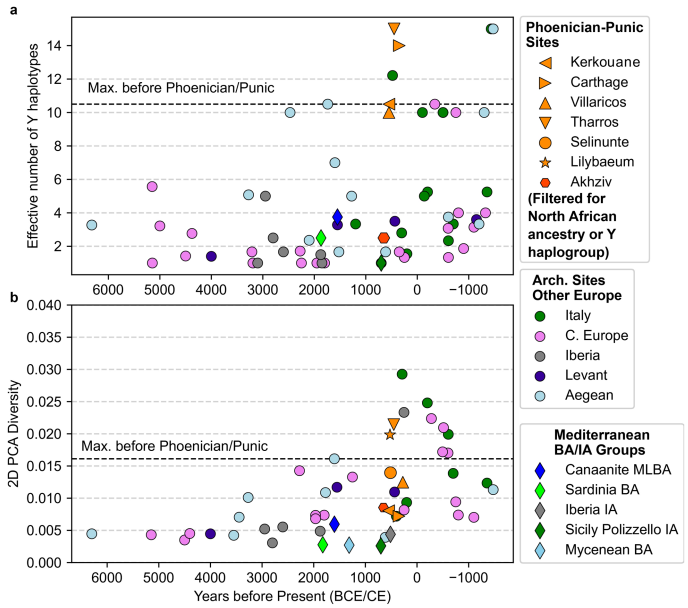
Extended Data Fig. 3 | Ancestry models inferred for the 140 Phoenician-Punic individuals in our data set by qpAdm. We partitioned individuals by region: **(a)** North Africa, **(b)** Sicily, **(c)** Sardinia, **(d)** Iberia, and **(e)** the Levant. Within each region, we grouped individuals by site, and for sites in Sicily, Sardinia, and Iberia, also by broad date ranges (see legend for colour code). We ordered the models of each individual according to their P-values (grey bar above each model). We report P-values assuming that the LRT statistic is chi-squared distributed with degrees of freedom determined by the number of populations and of contributing source populations. We did not correct these P-values for multiple testing, but this approach is conservative since we report models with comparatively high P-values (those that are not rejected by the test). Individuals with low coverage (fewer than 100,000 SNPs) are indicated by an asterisk (*) next to the sample ID. Eastern ancestry models are indicated by a contribution of the proxy sources Levant MLBA. In contrast, western ancestry models are indicated by contributions from either Greece BA (Myc), Sicily EBA, Sardinia LBA, Iberia LBA, or Steppe MLBA. There are five individuals for whom no valid eastern or western model was inferred. For four of them, we inferred valid models under the broad ancestry scheme (marked by an asterisk above the vertical bar), and for one (I22122 from Tharros, Sardinia), we could not infer any valid model.



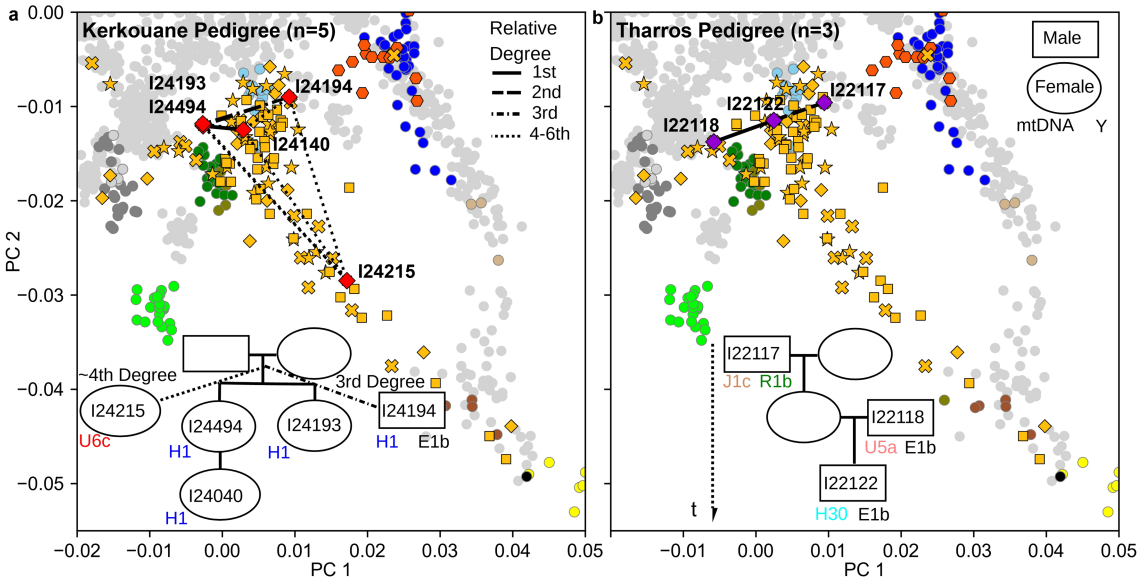
Extended Data Fig. 4 | Proportions of North African ancestry inferred using the 2D PCA and qpAdm for 123 Punic individuals. We exclude from this analysis the Akhziv sample, the three individuals from Sicily and Sardinia that cluster near Levantine individuals in the 2D PCA, and one individual for which we could not fit a qpAdm model. The qpAdm estimates are based on the smallest proportions estimated for the individual in a valid qpAdm model (see Methods). The PCA-based estimates of North African ancestry were computed by projecting the location of each sample in the PCA onto a cline from the cluster defined by Bronze Age individuals from Sicily to the cluster defined by North African individuals (see Supplementary Information S3 for more details). The two approaches yield similar estimates, with qpAdm being more sensitive to low ancestry proportions. Individuals from Kerkouane (depicted as squares) appear to have a broad range of North African ancestry (0 - 94%). Individuals from Sicily typically have lower proportions of North African ancestry (<20%), and we observe no significant shift in time. On the other hand, in Sardinia, none of the 12 individuals for which we inferred more than 10% North African ancestry (according to at least one of the two approaches) dated before 400 BCE, suggesting that North African ancestry was likely introduced around that time (Supplementary Information S3). We see a similar pattern in Iberia, but since we only have one individual from Iberia dating before 400 BCE, we cannot confidently infer the absence of North African ancestry during this time.



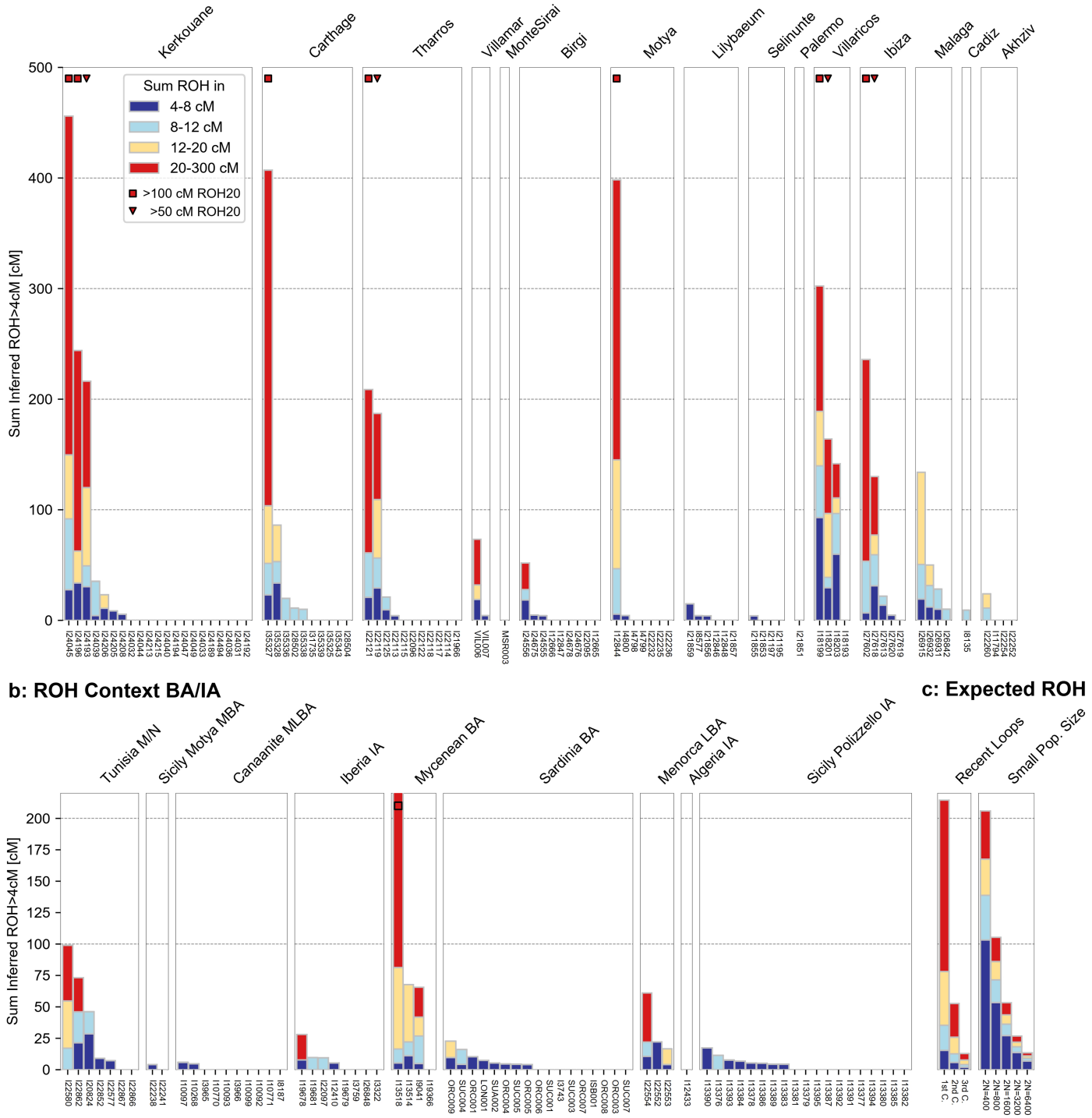
Extended Data Fig. 5 | Y Haplogroup Diversity in male individuals from Phoenician and Punic contexts. We inferred the first four characters of the ISOGG 2019 Y haplogroup classification for all Phoenician and Punic males with more than 100,000 autosomal SNPs covered (as those in almost all cases have sufficient coverage on the Y chromosome; see Methods). **(a)** Pie chart of Y haplogroup frequencies. **(b)** We visualize the Y haplogroup diversity partitioned per Phoenician or Punic site and denote each individual's haplotype by one circle. We set the height of the bar to the overall frequency (as depicted in panel **a**). The numbers in brackets indicate the total Y haplotype sample size.



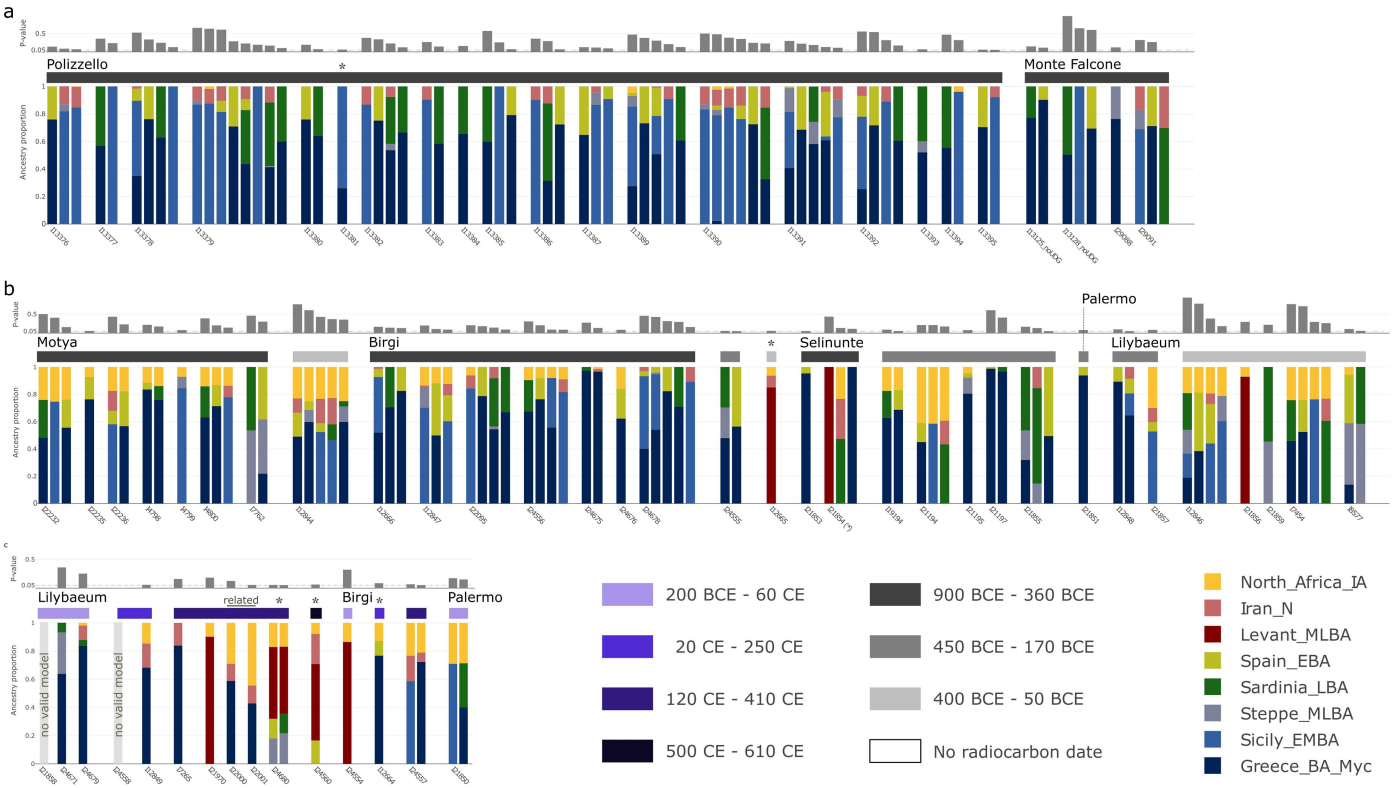
Extended Data Fig. 6 | Autosomal and Y Diversity without the contribution of filtered North African ancestry, per site in Phoenician-Punic contexts and the published aDNA record. (a) Y haplogroup diversity measured using the Inverse Simpson index. This value is computed as in Fig. 3a, excluding the three Punic individuals (from Kerkouane, Villaricos and Selinunte) with distinct North African Y haplogroups E1a and L (see Extended Data Fig. 5). **(b)** autosomal diversity measured using the first two PCs from Fig. 1 and the mean pairwise distance of those coordinates. This value is computed as in Fig. 3b, excluding individuals with more than 10% North African ancestry based on qpAdm in Phoenician-Punic sites (see Extended Data Fig. 4). Here, we combined individuals from the nearby Sicilian sites of Birgi, Motya, and Lilybaeum into one group (labelled Lilybaeum here). In both panels, the diversity measures for the context populations are as in Fig. 3 (without any additional filtering), and the dashed horizontal bar in both panels indicates the maximum diversity observed in sites dating before 500 BCE. See Supplementary Information S5 for a more detailed description of this analysis.



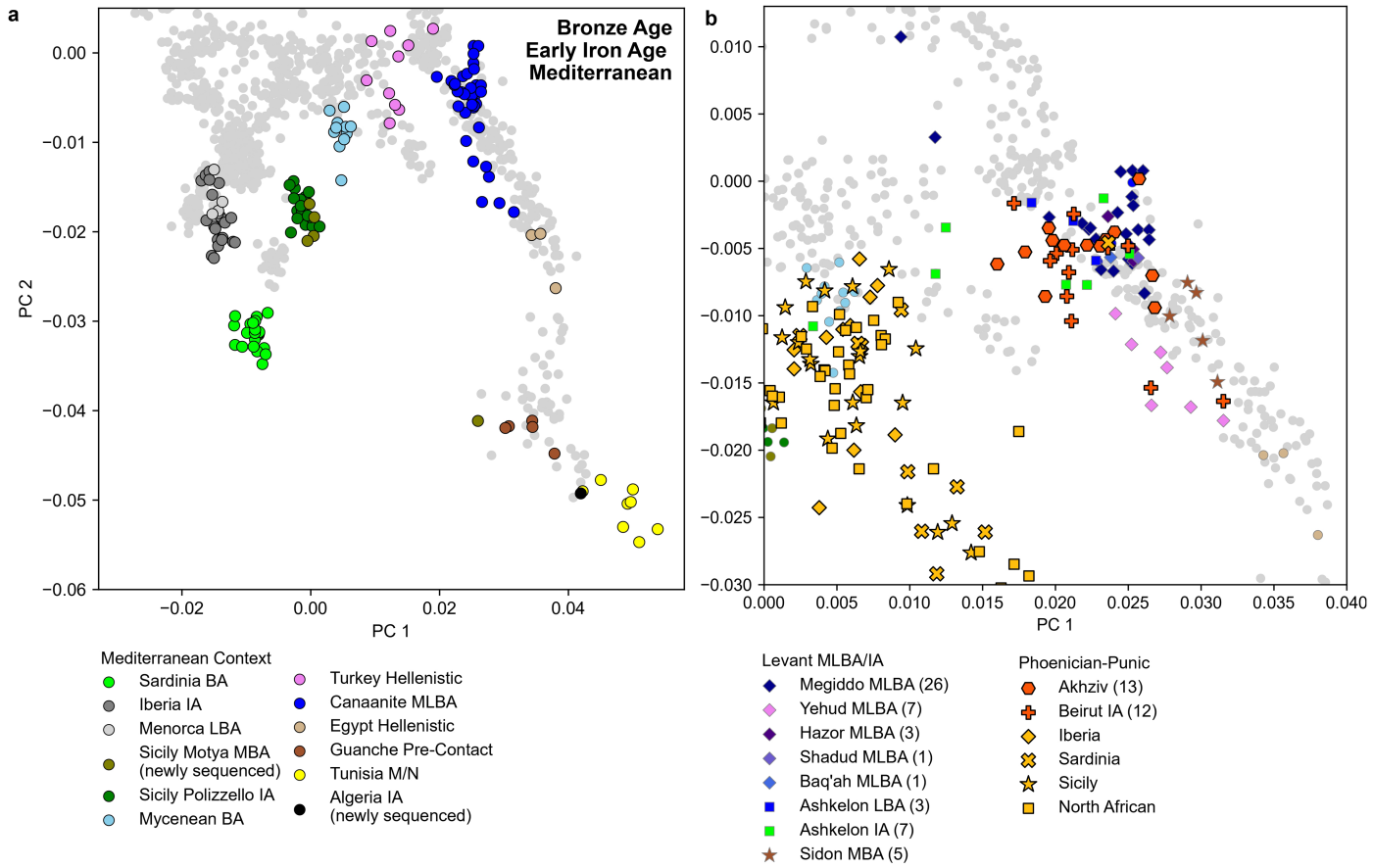
Extended Data Fig. 7 | Two reconstructed pedigrees of Punic individuals. We reconstructed two pedigrees based on inferring biological relatives with pairwise kinship (using IBD segment sharing) and uniparental haplogroups: **(a)** A pedigree linking five individuals from Kerkouane, North Africa; **(b)** A pedigree linking three individuals from Tharros, Sardinia. In the Kerkouane pedigree in (a), individuals I24215 and I24194 are inferred to be 3rd-4th degree relatives of the two siblings I24494 and I24193, but the exact pedigree relationship cannot be resolved. Each panel depicts the projection of the related individuals onto the two major PCs used in Fig. 1. Each pedigree specifies the sample IDs for all individuals, the mitochondrial (maternal) haplogroup and the Y (paternal) haplogroup for males. Both pedigrees contain individuals dating to 800–400 calBCE and link several individuals via the maternal lineage: We infer four identical maternal haplogroups in Kerkouane and a maternal grandfather in Tharros–two observations that are inconsistent with strict patrilocality.



Extended Data Fig. 8 | Runs of homozygosity inferred in Phoenician and Punic individuals and ancient individuals of relevant Bronze and Iron Age contexts. We computed runs of homozygosity (ROH) in all individuals with more than 400,000 SNPs covered and recorded the total length (in cM) of ROHs binned by length into four categories (see legend). We label individuals with at least 50 and 100 cM of their genome in long ROH (>20 cM) with triangle and square marks as in[30] - to indicate offspring of close biological parental relatives. **(a)** ROH in Phoenician and Punic individuals, grouped by site. **(b)** ROH in individuals from relevant Bronze and Iron Age contexts (as depicted in Fig. 1). **(c)** Expected ROH for offspring of various cousin matings (according to the degree of relation between parents) and for individuals sampled in populations with small effective size (calculated as described in[30]).



Extended Data Fig. 9 | Ancestry models inferred using qpAdm for individuals from Sicily from **(a)** the indigenous Iron Age sites of Polizzello and Monte Falcone, **(b)** from Phoenician sites before Roman expansion (as shown in Extended Data Fig. 3b), and **(c)** from Punic sites after Roman expansion. Colour horizontal bars indicate radiocarbon dates. The models of each individual are sorted according to their P-values (grey bar above each model). We report P-values assuming that the LRT statistic is chi-squared distributed with degrees of freedom determined by the number of populations and of contributing source populations. We did not correct these P-values for multiple testing, but this approach is conservative since we report models with comparatively high P-values (those that are not rejected by the test). Eastern ancestry models are indicated by a contribution of the proxy source Levant MLBA. In contrast, western ancestry models are indicated by contributions from either Greece BA (Myc), Sicily EBA, Sardinia LBA, Iberia LBA, or Steppe MLBA. There are seven individuals for which no valid eastern or western model was inferred. We inferred valid models under the broad ancestry scheme (marked by an asterisk above the vertical bar) for five of them. Two individuals were inferred to be related through IBD-sharing and are indicated in the figure. The analysis suggests that indigenous populations in Sicily have similar ancestry patterns as observed in the Phoenician sites but without North African ancestry. In later periods, we see the introduction of diverse ancestry sources (Levantine and western Mediterranean), likely associated with the Roman expansion into Sicily. See Supplementary Information S3 for more details.



Extended Data Fig. 10 | Additional PCA projections. (a) Bronze and Iron Age reference and Levantine populations. We show the same PCA as in Fig. 1 but focus on the ancient reference populations. **(b) Zoom in PCA projections of Levantine populations**. We show the same PCA depicted in **(a)**, but zooming into the region where Levantine individuals project. We also include additional Bronze and Iron Age Levant individuals not included in Fig. 1. Those previously published individuals originate from Sidon in present-day Lebanon[17] and various sites in present-day Israel (Megiddo, Yehud, Hazor, Baq'ah[25], Tel Shadud[64], Ashkelon[64,65]). Abbreviations: MLBA: Middle-Late Bronze Age, MBA: Middle Bronze Age, IA: Iron Age. All 13 individuals from Akhziv cluster next to other Levantine individuals, together with a single outlier individual from Tharros (I22119) inferred to have Levantine ancestry (Extended Data Fig. 3). Abbreviations: M/N: Mesolithic/Neolithic, MLBA: Middle-Late Bronze Age, MBA: Middle Bronze Age, LBA: Late Bronze Age, IA: Iron Age.

nature portfolio

Corresponding author(s): Harald Ringbauer, Ilan Gronau, David Reich

Last updated by author(s): May 7, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	BWA version 0.7.15 and other bioinformatics tools and data workflows (https://github.com/DReichLab/ADNA-Tools and https://github.com/DReichLab/adna-workflow). A detailled description of the data preprocessing steps is given in the Methods section of the manuscript.
Data analysis	We deposited the Code for analyzing the data and producing the figures in this manuscript at https://github.com/hringbauer/punic_aDNA. Relevant published software we used (including version): ancIBD version 0.5, hapROH version 0.63, SeqPrep 1.1, smartpca version 18150, ADMIXTURE version 1.3.0, qpAdm version 1201, Phylotree version 17, Yfull version 8.09, HaploGrep2 version 2.1.1, contamMix version 1.0-12, ANGSD version 0.923, OxCal version 4.4.4, SAMtools 1.3.1, pmdtools 0.60, PLINK V1.9, PONG V1.5

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw DNA sequences are deposited in the European Nucleotide Archive under the accession number XXX. Other newly reported data such as radiocarbon dates and archaeological context information are included the manuscript and supplementary files. Processed genotype data for individuals newly sequenced in this study can be obtained from the Harvard Dataverse repository through the following link (XXX).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size and geographic distribution is determined by accessibility of relevant osteological remains with sufficient aDNA preservation; it is not something we can choose a priori. We report new ancient DNA data from many contexts where ancient DNA has not previously been reported, analyzing all samples with sufficient aDNA preservation. Although even more powerful inferences could have been made if sample sizes were larger, we can make many meaningful new inferences with the available samples.
Data exclusions	We excluded samples that did not fall within the geographic scope of the study. After collecting genetic data, we excluded individuals from the analysis dataset as described in the Methods section entitled "Determination of ancient DNA authenticity." Specifically: "We determined ancient DNA authenticity based on five criteria. First, we required that the lower bound of the 95% confidence interval for contamination from ANGSD (if we were able to compute it) was <1%. Second, we required that the upper bound of the 95% confidence interval for match rate to mitochondrial consensus sequence (if we were able to compute it) was >95%. Third, we required that the average rate of cytosine-to-thymine errors at the terminal nucleotide for all sequences passing filters was >3% for double-stranded partially UDG-treated libraries39 and >10% for single-stranded USER-treated libraries and double-stranded non-UDG-treated libraries (the latter libraries are all from previously published data that we reanalysed here). Fourth, we required the ratio of sequences mapping to the Y chromosome to the sum of sequences mapping to the X and Y chromosome for the 1240K data to be less than 3% (consistent with a female) or >35% (consistent with a male). Fifth, to report an individual we required the number of SNPs covered at least once to be at least 5,000 (for most actual population genetic analyses, we required at least 20,000).
Replication	Only a single library can be made from each extract aliquot so no replication from the same extract is possible. For the individuals with more than one library (15x two libaries per sample, 3x three, 1x four, 1x five), we confirmed in all cases that the libraries were from the same individuals. Another measure of replication also derives from the fact that the ancestry distributions in individuals from the same periods tended to be very similar. As a result of this, key findings in this study are not dependent on single samples.
Randomization	Historical studies are retrospective rather than prospective -- and the actual trajectory of human history has occurred only once -- so

Randomization	randomization of the data into independent processes is not possible. The mansucript discusses a caveat about possible biases due to non-random sampling.
Blinding	Co-analysis of the genetic and archaeological data was central to the study, so we could not be blind to the sample identity.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology and Archaeology

Specimen provenance	We describe the provenance of all archaeological specifmens in Supplementary Data 1 and the "Archaeological Site Descriptions", a document that for each Archaeological site lists the 1) General Location and Chronology, 2) Excavation history, 3) Description of cemeteries and 4) Relevant references.
Specimen deposition	The bone and tooth parts that remain after analysis for ancient DNA are under the stewardship of the archaeologists and cultural institutions from which they were sampled. At present, they are either already returned to the sample stewards or they are stored on long-term loan at the ancient DNA laboratories where they were analysed. They can be re-examined upon request to the sample stewards. Researchers who wish to replicate analyses from this study or gather new data on the libraries generated for this study are welcome to make a request for aliquots of those libraries to corresponding author David Reich who will fulfill all reasonable requests
Dating methods	We obtained 111 accelerator mass spectrometry-based dates on bone for 99 distinct human remains in specialized C14 laboratories at Pennsylvania State University (82 measurements), the D-REAMS laboratory at the Weizmann Institute of Science (24), Oxford (four), and CIRAM (one). We report conventional radiocarbon ages and standard errors (Supplementary Data 1B). We used the calibration curve IntCal20 to calibrate conventional radiocarbon dates (see Methods). Eleven individuals were analyzed by two different labs, with date ranges from the two labs being highly concordant in all cases. For these individuals, we used the R_combine method to combine date ranges.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	All human skeletons analysed in this study were sampled with written permission of the stewards of the skeletons and every individual is represented by at least one co-author. Researchers who wish to obtain further information about specific individuals should write to the corresponding authors and/or the authors who provided the archaeological contextualisation for those individuals whose names are specified in the "Archaeological Site Descriptions".

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A