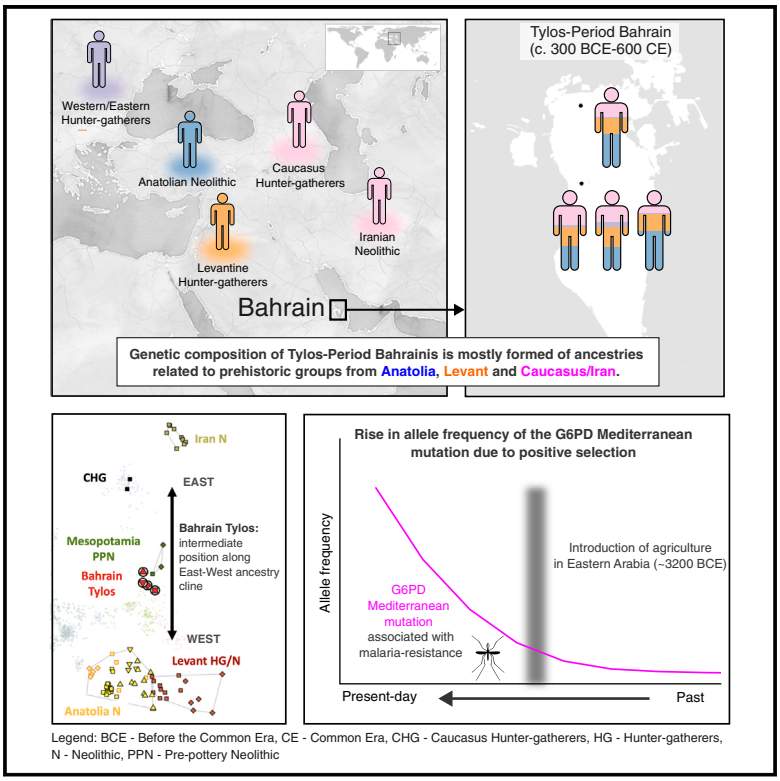
Cell Genomics

Article

Ancient genomes illuminate Eastern Arabian population history and adaptation against malaria

Graphical abstract



Legend: BCE - Before the Common Era, CE - Common Era, CHG - Caucasus Hunter-gatherers, HG - Hunter-gatherers, N - Neolithic, PPN - Pre-pottery Neolithic

Highlights

- Successful sequencing of four ancient genomes from Tylos-period Bahrain
- Genetic composition formed of ancient Anatolia, Levant, and Caucasus/Iran ancestries
- Genetic affinities to present-day inhabitants of Iraq and the Levant
- Evidence of malaria adaptation at the onset of agriculture in Eastern Arabia

Authors

Rui Martiniano, Marc Haber, Mohamed A. Almarri, ..., Daniel G. Bradley, Pierre Lombard, Richard Durbin

Correspondence

r.martiniano@ljmu.ac.uk (R.M.), rd109@cam.ac.uk (R.D.)

In brief

Due to the lack of ancient genomes from Arabia, the genetic composition of its past inhabitants was unknown. Martiniano et al. report four genomes from Tylos-period Bahrain whose ancestry is related to prehistoric groups from Anatolia, Levant, and Caucasus/Iran and provide evidence of malaria adaptation at the onset of agriculture in Eastern Arabia.

Martiniano et al., 2024, Cell Genomics *4*, 100507
March 13, 2024
https://doi.org/10.1016/j.xgen.2024.100507

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Ancient genomes illuminate Eastern Arabian population history and adaptation against malaria

Rui Martiniano,[1,14,*] Marc Haber,[2,13] Mohamed A. Almarri,[3,4,13] Valeria Mattiangeli,[5] Mirte C.M. Kuijpers,[6] Berenice Chamel,[7] Emily M. Breslin,[5] Judith Littleton,[8] Salman Almahari,[9] Fatima Aloraifi,[10] Daniel G. Bradley,[5] Pierre Lombard,[9,11] and Richard Durbin[12,*]

[1]School of Biological and Environmental Sciences, Liverpool John Moores University, L3 3AF Liverpool, UK
[2]Institute of Cancer and Genomic Sciences, University of Birmingham Dubai, Dubai, United Arab Emirates
[3]Department of Forensic Science and Criminology, Dubai Police GHQ, Dubai, United Arab Emirates
[4]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates
[5]Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland
[6]Department of Ecology, Behavior and Evolution, School of Biological Sciences, University of California, San Diego, La Jolla, CA, USA
[7]Institut Français du Proche-Orient (MEAE/CNRS), Beirut, Lebanon
[8]School of Social Sciences, University of Auckland, Auckland, New Zealand
[9]Bahrain Authority for Culture and Antiquities, Manama, Kingdom of Bahrain
[10]Mersey and West Lancashire Teaching Hospitals NHS Trust, Whiston Hospital, Warrington Road, Prescot, L35 5DR Liverpool, UK
[11]Archéorient UMR 5133, CNRS, Université Lyon 2, Maison de l'Orient et de la Méditerranée - Jean Pouilloux, Lyon, France
[12]Department of Genetics, University of Cambridge, CB2 3EH Cambridge, UK
[13]These authors contributed equally
[14]Lead contact
*Correspondence: r.martiniano@ljmu.ac.uk (R.M.), rd109@cam.ac.uk (R.D.)
https://doi.org/10.1016/j.xgen.2024.100507

SUMMARY

The harsh climate of Arabia has posed challenges in generating ancient DNA from the region, hindering the direct examination of ancient genomes for understanding the demographic processes that shaped Arabian populations. In this study, we report whole-genome sequence data obtained from four Tylos-period individuals from Bahrain. Their genetic ancestry can be modeled as a mixture of sources from ancient Anatolia, Levant, and Iran/Caucasus, with variation between individuals suggesting population heterogeneity in Bahrain before the onset of Islam. We identify the G6PD Mediterranean mutation associated with malaria resistance in three out of four ancient Bahraini samples and estimate that it rose in frequency in Eastern Arabia from 5 to 6 kya onward, around the time agriculture appeared in the region. Our study characterizes the genetic composition of ancient Arabians, shedding light on the population history of Bahrain and demonstrating the feasibility of studies of ancient DNA in the region.

INTRODUCTION

Following flint and pottery evidence of human occupation from ~5000 BCE,[1] from around 2200 BCE Bahrain can be connected with references in Mesopotamian cuneiform records to the Dilmun civilization, both as a trading center[2] and playing an important role in Sumerian mythology, including in their creation myth and in the Epic of Gilgamesh.[3,4] This period also saw the start of a funerary tradition that ultimately resulted in the highest density of burial mounds in the ancient world (on the order of 100,000).[5] The end of the Late Dilmun phase (~600 BCE) coincided with the Persian Achaemenid conquest of Mesopotamia, whose influence in Bahrain (600–300 BCE) is attested by the presence of bowls and other artifacts typical from that culture.[6]

Around 325 BCE, an expedition to the Arabian coast sent by Alexander the Great reached the shores of Bahrain, which from that time onward would be known as Tylos. The Tylos period (~325 BCE to ~600 CE) was characterized by exceptional prosperity and marked by Hellenistic and Persian influence.[7] The death of Alexander and subsequent disintegration of the Macedonian Empire led to the establishment of the Seleucid Empire (312–63 BCE), which dominated a vast area comprising Anatolia, Levant, Mesopotamia, and Iran, also controlling the eastern Arabian islands of Failaka and Bahrain. The Seleucids eventually lost authority in Mesopotamia, leading to the formation of the semi-independent Kingdom of Characene (141 BCE to 222 CE) in Southern Iraq, a vassal to the Parthian Empire that would govern Bahrain until Sasanian conquest.[8,9]

Ancient DNA (aDNA) studies from the northern regions of the Middle East revealed that the first farmers from Anatolia, the Levant, and Iran each descended from local, genetically distinct hunter-gatherer populations, with variable amounts of gene flow from neighboring groups.[10,11] For example, early Anatolians admixed with Mesopotamians and later with Levantine agriculturalists,[12] whereas in the Levant early farmers derived from Epipaleolithic Natufians and Anatolian Neolithic populations.[11]

Table 1. Ancient DNA samples from Bahrain sequenced in this study

Sample ID	Dates	Reads	Coverage	Sex	Y chromosome haplogroup	mtDNA haplogroup	mtDNA contamination (%)	X chromosome contamination (%)
AS_EMT	∼200 BCE to 300 CE	71,371,001	0.94	XX	–	J1c15a1	1.36	–
MH1_LT	432–561 cal. CE (1565 BP)	69,192,646	1.06	XY	J2a2a1a∼	R2	2.00	2.92
MH2_LT	300–600 CE	80,197,765	1.26	XX	–	T2b	0.97	–
MH3_LT	577–647 cal. CE (1456 BP)	18,586,203	0.24	XY	H2	U8b1a2a	0.64	2.00

See also Figures S1 and S2; Tables S1 and S2.

During the transition from the Chalcolithic to the Bronze Age, admixture between various Middle Eastern groups intensified, leading to increasing genetic homogenization in the region.[11] Around this time, an ancient Iranian-related component was introduced into the Levant,[13,14] followed by steppe/European-related ancestry in the Iron Age.[15,16] In the neighboring region of Mesopotamia, the aDNA record is still sparse and therefore the genetic composition of local hunter-gatherers remains unknown, but recently published Pre-Pottery Neolithic (PPN) genomes from Upper Mesopotamia are genetically intermediate along the ancestry cline extending from ancient groups from Anatolia/Levant to Iran/Caucasus.[12,17]

Due to challenges associated with the recovery of ancient genomes from hot and humid climates,[18] no ancient genomes from the Arabian Peninsula have been published so far. Surveys of present-day genomes from Arabia and the Levant suggest that they were shaped by different demographic processes: first, Arabians carry an excess of East African- and Natufian-related ancestry in comparison with Levantine populations, who in contrast bear higher proportions of European and Anatolian Neolithic ancestry.[19,20] Second, Arabian and Levantine populations present different size trajectories, with the former being affected by a pronounced bottleneck event occurring around 6,000 years ago (6 kya) that coincides with the "Dark Millennium," a period of increasing aridification in Arabia,[21] and the latter showing a more recent size reduction associated with the ∼4.2 kya climate event[20] of wider distribution across Anatolia, the Levant, and Mesopotamia.[22] Third, Arabians show increased levels of consanguinity in comparison to the Levant,[23,24] potentially leading to higher prevalence of genetic disorders including G6PD deficiency.[25]

To examine these topics directly using ancient genomes, we sequenced four individuals from the island of Bahrain dating from the Tylos period (∼300 BCE to 600 CE). Using modeling approaches, we determine that their ancestry is a mixture of sources from ancient Levant, Iran/Caucasus, and Anatolia, with one individual presenting higher affinity to Levantine groups than the others, potentially as a result of historically recorded incursions of Arabian or Levantine tribes into Bahrain,[26] and two individuals with additional CHG (Caucasus hunter-gatherer)/post-Neolithic Iranian ancestry, which can tentatively be explained by Iranian-associated influence in Bahrain during pre-Islamic times. Additionally, we detect the G6PD Mediterranean variant in three Tylos-period Bahrainis and estimate that it rose to high frequencies in Eastern Arabia due to strong positive selection exerted by malaria endemicity coinciding with the appearance of agriculture. Lastly, we detect large runs of homozygosity (ROHs) in one individual, suggesting consanguineous unions in pre-Islamic Tylos populations. The present work provides a snapshot of the genetic composition of ancient Eastern Arabia and demonstrates the feasibility of aDNA studies in the region.

RESULTS

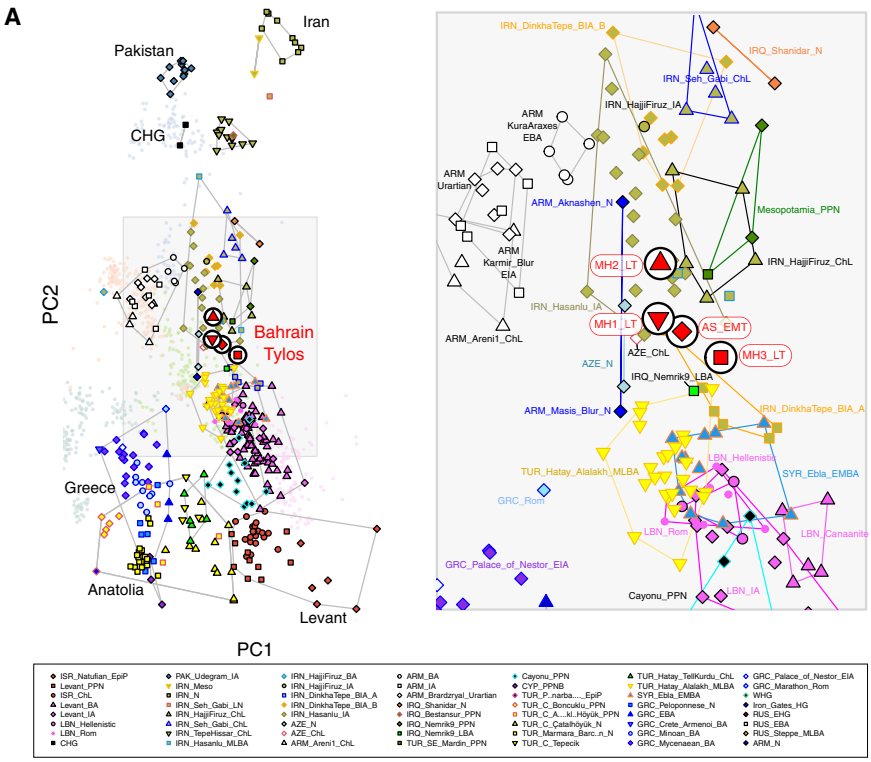
Ancient DNA sample sequencing and determination of authenticity

We extracted DNA from 25 skeletal samples from ancient burial mounds on the island of Bahrain ranging from the Dilmun period to the Tylos period and sequenced them to assess endogenous DNA preservation (Table S1). Of these, only four were sufficiently well preserved for additional sequencing, with the remaining samples presenting negligible amounts of human endogenous DNA. Here, we report shotgun sequence data obtained from these four samples, all derived from petrous bones: one from Abu Saiba (AS) and three from Madinat Hamad (MH1, MH2, MH3; Figure S1A; Tables 1 and S2). Of these samples, three were sequenced to approximately 1×, and one to 0.24× (Table 1). Due to poor collagen preservation, we could only obtain radiocarbon dates for two out of the four sequenced individuals, placing them in the Late Tylos/Sasanian period (LT, ∼300–622 CE), with MH1 being older (432–561 cal. CE) than MH3 (577–647 cal. CE) (Figure S1B). Sample MH2 was not directly dated, but its archaeological context places it in the Late Tylos period. The Abu Saiba sample was excavated from a cemetery with known occupation between 200 BCE and 300 CE,[7,27] and therefore it dates confidently within the boundaries of the Early/Middle Tylos period (EMT), more precisely during the times of Seleucid and Characene influence in Bahrain, which preceded the emergence of the Sasanian Empire.

Deamination patterns in these sequences conform with those typical of aDNA (Figure S2), supporting the presence of authentic aDNA sequences, and we observed low contamination rates on the mitochondria of all four samples (≤2%) and on the X chromosome of the two male samples (<3%, Table 1). None of the samples were close relatives of each other.

Admixed ancestry and subtle genetic differentiation within Tylos-period individuals from Bahrain

To examine the genetic affinities of the four Tylos-period samples from Bahrain, we performed a principal component analysis (PCA)[28,29] on 1,301 present-day West Eurasians,[30,31] including 117 recently reported Arabians and Levantines,[20,32] on which we projected 529 ancient individuals (Figure 1A; full PCA shown in Figure S3A). The four Bahrain Tylos samples were positioned



(legend on next page)

intermediately along the cline connecting western (Anatolian and Levantine) and eastern (CHG and IRN_Ganj_Dareh_N) sources and in the vicinity of Upper Mesopotamian PPN individuals from Southeastern Turkey and Northern Iraq (Mesopotamia_PPN) and Neolithic (N) farmers from Armenia (ARM_Masis_Blur_N and ARM_Aknashen_N) and Azerbaijan (AZE_N). Also in the proximity of the ancient Bahrain samples are various post-Neolithic groups (Chalcolithic [ChL], Iron Age [IA], Bronze Iron Age [BIA]) from Iran (IRN_HajjiFiruz_ChL/IRN_Hasanlu_IA, IRN_DinkhaTepe_A_BIA), Bronze Age (Early [E], Middle [M], Late [L] Bronze Age [BA]) groups from Eastern Turkey (TUR_Hatay_Alalakh_MLBA), Iraq (IRQ_Nemrik9_LBA), and Syria (Syria_Ebla_EMBA), and various other Levantine groups (Hellenistic/Roman), all of which were previously reported to contain a mixture of two or more sources related to Anatolian, Levantine, and Iranian Neolithic/CHG ancestries.[11–14,33,34]

The Bahrain_Tylos individuals are slightly differentiated. Compared with the older individual AS_EMT, samples MH1_LT and MH2_LT are shifted toward eastern populations from Armenia ChL-IA composed mainly of CHG- and Anatolian-related ancestry and variable amounts of steppe ancestry, with MH2_LT being further separated from the Bahrain group toward the direction of CHG and ancient groups from Iran and Pakistan. Sample MH3_LT is closer to ancient Levantine groups than the other three individuals from Bahrain.

To further investigate the ancestry composition of the Tylos individuals, we performed a temporally aware model-based clustering analysis on an expanded dataset using DyStruct (Figure 1B). At k = 9, Anatolian_N, Natufians, WHG (Western hunter-gatherers), and Iran_N/CHG define independent ancestral components that contribute to the majority of ancient and present-day West Eurasians (Figures 1B and S3B). Consistent with the PCA results, the four Tylos-period individuals are broadly similar to and intermediate between post-Neolithic groups from the Levant (BA/IA/Hellenistic/Roman) and Iran (ChL/IA), albeit with higher Natufian-related ancestry than ancient Iranians and higher amounts of Iran_N/CHG-related ancestry than ancient Levantines, a finding that is also corroborated by positive f4-statistics (Figures 2A and 2B, respectively).

We also observe some variability within the four Bahrain samples in the clustering analysis (Figure 1B), with MH3_LT carrying additional Anatolia_N and less Iran_N/CHG-related ancestry than the other samples. Here, Anatolian and Levantine ancestry were in some cases not well differentiated due to the presence of Anatolian ancestry in Neolithic Levantines. Through f4-statistics, we confirm that individual MH3_LT shares more alleles with both Anatolians and Levantines than the remaining Tylos-period samples and observe a gradient of Levantine-related ancestry which is maximized in MH3_LT and minimized in MH2_LT, with AS_EMT and MH1_LT presenting intermediate values between them (Table S3), as reflected in the PCA (Figure 1A). In Figure 1B, it is also apparent that MH1_LT and MH2_LT individuals present slight differences in comparison with the other two Bahraini samples, notably, higher Iran_N/CHG-related ancestry, and an additional small amount of Western/Eastern hunter-gatherer (WHG/EHG)-related ancestry (approximately 3.1%–6.2%), which forms around half of the genetic composition of steppe EMBA,[35,36] and can also be found in Armenia_ChL and certain post-Neolithic Iranians (HajjiFiruz_BA/IA).

To formally investigate these subtle differences, we tested for excess allele sharing between Bahrain_Tylos and ancient groups X and Y by estimating f4(Mbuti, Bahrain_Tylos; X, Y) (Figure 2C). We corroborate excess affinity between MH3_LT and Levantines, especially with LBN_Canaanites, in comparison with CHG or post-Neolithic Iranians, but we obtain no significant results for the remaining samples from Bahrain. Conversely, MH1_LT and MH2_LT present excess allele sharing with CHG in comparison with ISR_Natufians, suggesting lower Levantine and higher CHG ancestry in these samples, with AS_EMT presenting intermediate values. MH1 and MH2_LT also present increased CHG and IRN_HajjiFiruz_IA ancestry in comparison with IRN_Ganj_Dareh_N, as well as slightly higher West Siberian hunter-gatherer (WSHG) ancestry, particularly in MH2_LT.

In summary, the four Bahrain Tylos individuals are genetically intermediate along the ancestry cline that extends from ancient Anatolia/Levant to ancient Iran/Caucasus and in proximity to ancient groups from Mesopotamia, Armenia and Azerbaijan, with subtle differences in ancestry composition.

A tripartite genetic ancestry of Bahrain Tylos

We found that Bahrain Tylos is best modeled as a mixture of sources from ancient Anatolia, Levant, and Iran/Caucasus. First, we obtained significantly negative results ($z \leq -3$) for the statistic f3(Bahrain_Tylos; X, Y) when X was an ancient Levantine or an Anatolian population (Natufians, Levantine, or Anatolian Neolithic) and Y an ancient group from Caucasus or Iran (CHG or IRN_Ganj_Dareh_N) (Table S4), suggesting that these ancestries have plausibly contributed, even if distantly so, to the formation of Bahrain Tylos.

Second, nearly all the two-way qpAdm models that fit the data involve populations carrying three types of ancestry: Levantine, Anatolian, and CHG/Iranian Neolithic (Table S5). For instance, all four samples can be modeled with mixtures of Levant_PPN and CHG ($p > 0.05$; Figure 3A and Table S5). However, when replacing Levant_PPN with ISR_Natufian_EpiP, only two samples (AS_EMT and MH3_LT) were successfully modeled, although with lower p values, suggesting that an additional Anatolian-related component, which is present in Neolithic Levant, but

Figure 1. Genetic composition of Tylos-period Bahrain samples

(A) Principal component (PC) analysis with 1,830 present-day and ancient Eurasians and 579,407 SNPs. Ancient samples are indicated with larger symbols, as in the key.

(B) Pertinent results from a DyStruct analysis (k = 9) of 2,073 samples, including 658 ancient samples binned into 11 different time periods, using 85,043 transversions and with the mean ancestral components estimated for each population. Ancient samples above and present-day samples below (see also Figure S3). AKN, Aknashen; BIA, Bronze Iron Age; BMAC, Bactria-Margiana Archaeological Complex; ChL, Chalcolithic; E/M/LBA, Early/Middle/Late Bronze Age; EC, Early Chalcolithic; H, Historical; HLN, Hellenistic; IA, Iron Age; KA, Kura-Araxes; MB, Masis Blur; Min, Minoan; Myc, Mycenaean; N, Neolithic; Rom, Roman.

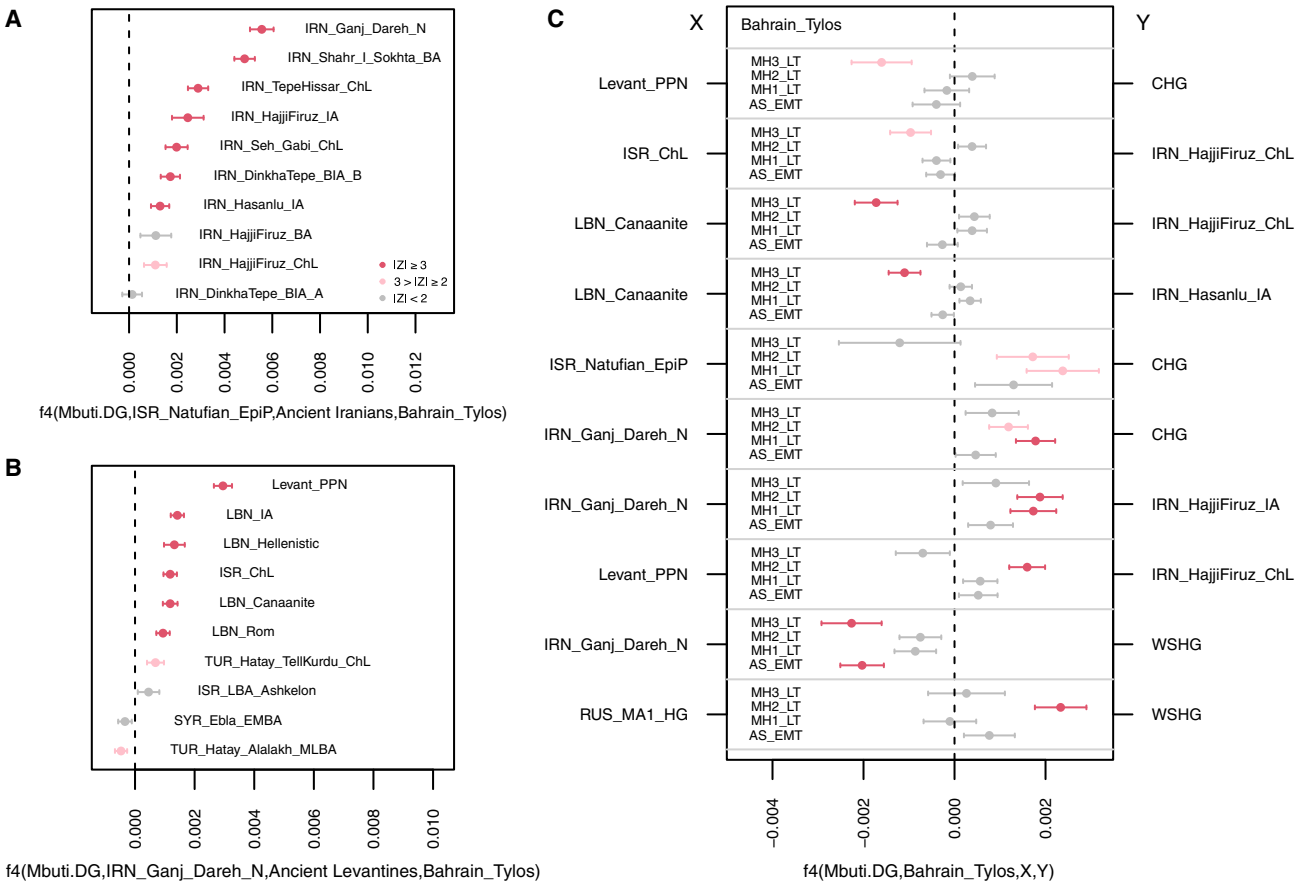


Figure 2. Allele sharing statistics

(A) f4(Mbuti, ISR_Natufians; Ancient Iranians, Bahrain_Tylos).
(B) f4(Mbuti, IRN_Ganj_Dareh; Ancient Levantines, Bahrain_Tylos).
(C) f4(Mbuti, Bahrain_Tylos; X, Y). Horizontal bars show standard errors.
See also Table S3.

absent in Natufians and CHG, is required to model the remaining individuals. We note that Anatolian Neolithic ancestry can also derive from other groups, such as ARM_Aknashen_N (formed of Anatolian and CHG ancestry), which returns feasible models when combined with ISR_Natufian_EpiP or Levant_PPN ($p > 0.05$; Table S5).

Models including Levant_PPN and IRN_GanjDareh_N sources are not as successful, given that only the lower-coverage MH3_LT individual can be modeled in this way ($p = 0.042$; Figure 3A and Table S6), suggesting that CHGs might be better representatives than Neolithic Iranians for the eastern ancestry present in the Tylos samples. Accordingly, we obtain a significantly negative ($Z = -3.81$) f4-statistic f4(Mbuti.DG, Bahrain_Tylos; CHG, IRN_Ganj_Dareh_N), suggesting excess affinity between Bahrain_Tylos and CHG in relation to IRN_Ganj_Dareh_N. This is despite the fact that, as seen in our DyStruct analysis ($k = 12$, Figure S3B), CHG and IRN_Ganj_Dareh_N ancestries are difficult to distinguish from one another. It is possible that they both contributed to the formation of Tylos-period Bahrainis. We also obtained generally less confident but feasible models when combining Iranian Neolithic or Late Neolithic with CYP_PPNB or TUR_C_Çatalhöyük_N, with the latter two populations containing both Anatolian and Levantine ancestry (Figure 3A and Table S5).

Third, we also find support for a tripartite genetic ancestry of Bahrain Tylos when including more temporally proximal groups as sources in our models (Figure 3B; Tables S5 and S6), obtaining feasible combinations of older Levantine and more recent Iranians and vice versa. The overall pattern in these models is that the most recent source contributes with a greater ancestry proportion to Bahrain Tylos samples, irrespective of being Iranian or Levantine. We interpret this as being due to the fact that later samples tend to carry the Anatolia/Levant/Iran/CHG-related ancestries required for modeling the four samples from Bahrain.

Lastly, we evaluate the ancestry of Bahrain Tylos individuals in a context of Near Eastern variation by estimating ancestry proportions using a previously published[12] three-way model with TUR_Pınarbaşı_EpiP, ISR_Natufian_EpiP, and CHG as sources in a set of relevant ancient groups which can also be modeled in this way (Figure 3C).[12] In this model, the Bahrain_Tylos samples present ancestry proportions similar

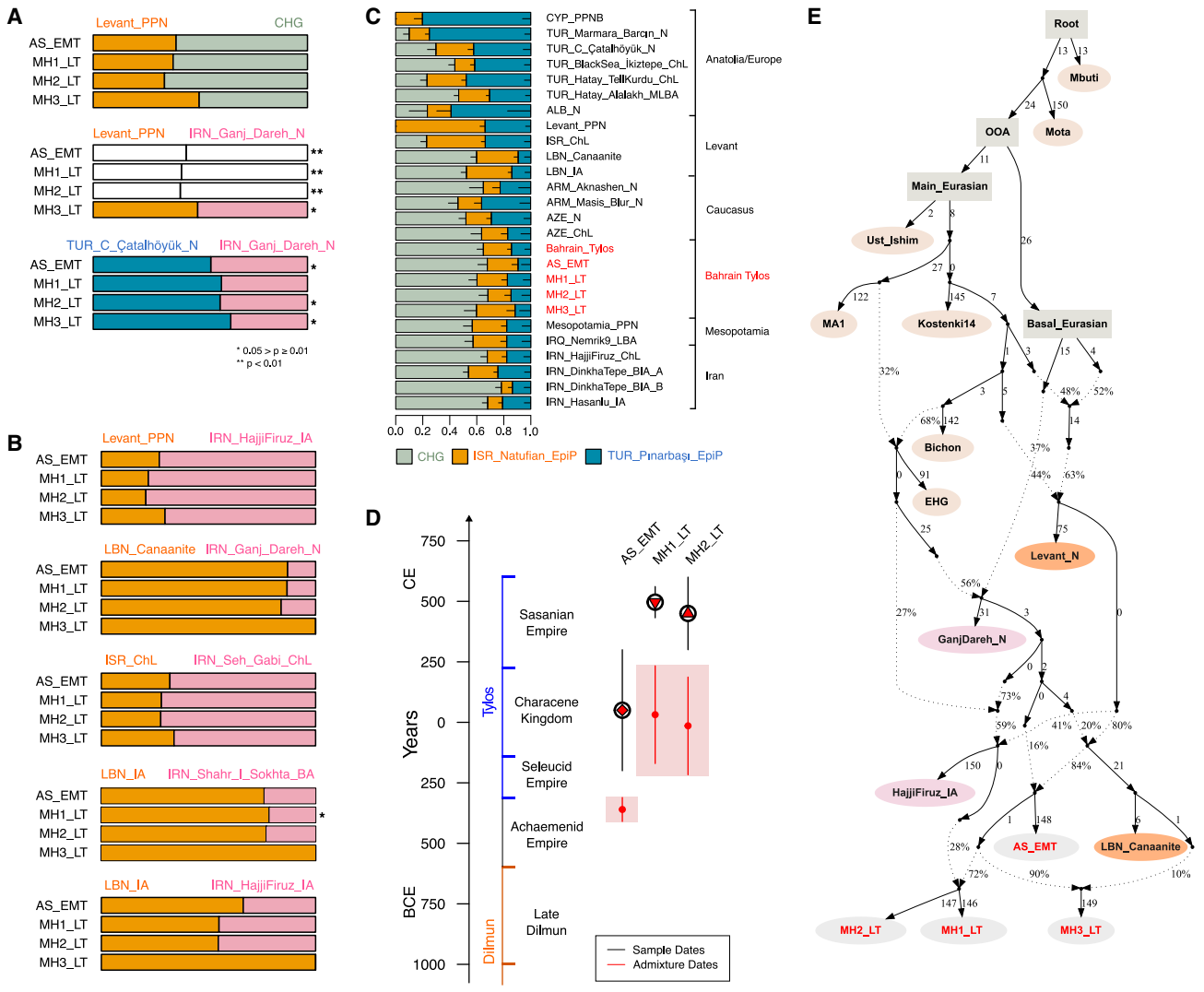


Figure 3. Models of Bahrain Tylos ancestry

(A and B) (A) Selected rank-1 qpAdm models of the four Bahrain Tylos individuals including distal sources and (B) a mixture of distal and proximal sources. Feasible rank-0 models are shown instead in the cases where the rank-1 model was rejected. Single asterisks (*) indicate $0.05 > p \geq 0.01$; double asterisks (**) and white bars indicate rejected models at $p < 0.01$.

(C) qpAdm models of ancient Near Easterners using Natufians, Pinarbasi, and CHG as sources. Horizontal bars show standard errors.

(D) Inferred time of admixture in Bahrain Tylos using DATES ancestry covariance decay curve with LBN_Canaanites and IRN_HajjiFiruz_IA as references representing ancient Levantine and ancient Iranian ancestries, respectively. Vertical bars show standard errors.

(E) Modeling Bahrain Tylos ancestry using qpGraph.

See also Figure S4 and Tables S4, S5, S6, S7, S8, and S9.

to those of various ancient samples from Mesopotamia, the Caucasus, and post-Neolithic Iranians and Levantines. Accordingly, rank-0 qpAdm models show that Tylos-period Bahrainis form a clade with several of these ancient groups ($p \geq 0.01$; Table S7), suggesting that similar sources have contributed to their ancestry.

According to our analyses, Tylos-period samples from Bahrain can be modeled using sources related to ancient groups from Anatolia, Levant, and Iran or Caucasus and are broadly similar in ancestry to various groups from neighboring regions.

Recent gene-flow events shaped the genetic composition of Tylos-period Bahrainis

We next investigated the timing of admixture between ancient Levantine and Iranian sources in Bahrain Tylos samples using DATES (Figure 3D). We observe different admixture times for AS_EMT and MH1-2_LT: the first occurred 14 ± 1.7 generations before AS_EMT (357–455 years, assuming a generation time of 29 years), coinciding with a period of Achaemenid influence in Bahrain, and the second occurred 16 ± 7 generations (261–667 years) before MH1_LT and MH2_LT, a period defined by the

incorporation of Bahrain into the Characene Kingdom, a vassal state to the Parthians.

While recognizing that there are limits to fitting complex models of population history,[37] these processes are illustrated well by an admixture graph that was semi-automatically generated to fit the data (Figure 3E). In this graph the earliest sample from Abu Saiba (AS_EMT) derives from admixture between sources related to the Levant and Iran_N/CHG. The three Late Tylos individuals descend from admixture between the earliest sample AS_EMT and other sources, with MH1_LT and MH2_LT having an additional pulse of post-Neolithic Iranian-related ancestry, here represented by IRN_Hajji_Firuz_IA, whereas the MH3_LT sample received additional ancestry associated with the Levant, represented by LBN_Canaanite from Sidon. These inferences are supported by the observation of additional IRN_IA ancestry in MH1_LT and MH2_LT relative to AS_EMT and MH3_LT, and that MH3_LT bears more Levantine ancestry than the other samples from Bahrain (Figure 3B). We show an alternative model where the post-Neolithic Iranian source is instead represented by IRN_Shahr-i-Sokhta in Figure S4. We note that the populations depicted in the graph act as representatives of different ancestries and do not necessarily correspond to the actual populations that contributed to the formation of Tylos-period Bahrainis. Further, the true population history is likely to be different in detail from, and more complex than, the admixture processes represented here.

Affinities with Mesopotamia

Considering the abundance of evidence of contacts between ancient Bahrain and Mesopotamia, we evaluated separate models using Mesopotamia_PPN (two PPN samples from Northern Iraq and one from Mardin in Turkey), an LBA sample from Iraq, and the Upper Mesopotamian Cayonu_PPN from Southeastern Turkey as a source. First, we observe that all four Tylos samples form a clade with IRQ_Nemrik9_LBA, two samples form a clade with Mesopotamia_PPN (MH1_LT and MH3_LT), and only the lower-coverage MH3_LT forms a clade with Cayonu_PPN (Table S7). When exploring two-source models with Mesopotamia_PPN, we can model all samples with Levant_PPN as a second source, except for sample MH2_LT, which requires additional EHG ancestry, or using RUS_AfontovaGora3, SRB_Iron_Gates_HG, or LBN_IA as a second source instead of Levant_PPN (Table S8). While the inability to reject various models of Bahrain Tylos ancestry containing ancient Mesopotamians as a source suggests a relationship between the two, the low *Z* scores of admixture weights attributed to non-Mesopotamian sources do not allow establishing the precise role of Mesopotamian groups in the formation of Bahrain Tylos.

Tylos-period Bahrainis are genetically closer to present-day populations from Iraq and the Levant than to present-day Arabians

Regarding affinities with present-day populations, the temporally aware model-based clustering analysis (Figure 1B) suggests that Tylos-period Bahrain samples are more similar to present-day Levantine groups than to present-day Arabians or South Asians, who show higher and lower amounts of Natufian ancestry, respectively. To investigate this in more detail, we performed a haplotype-based ChromoPainter and FineSTRUCTURE analysis with a dataset including Tylos-period Bahrainis and ancient samples from the Levant[13,38] and Iran[10] with available whole-genome shotgun sequence data, as well as present-day individuals from the Human Origins dataset[11,39] and from Arabia and the Levant[20] (Figure 4).

Three Bahrain Tylos samples (AS_EMT, MH1_LT, and MH2_LT) were included in a cluster composed of Syrians, Iraqis, and two Jewish individuals from Iran and Georgia (Figure 4A), whereas sample MH3_LT clustered with 22 ancient individuals from Lebanon ranging from the Bronze Age to the Roman period. In a PCA of the haplotype sharing matrix (Figure 4B), we observe that MH3_LT is closer to ancient Lebanese and Sardinians, corroborating our previous inferences of increased ancient Levantine and Anatolian Neolithic ancestry in this sample. The remaining three Bahrain Tylos individuals are positioned closer to present-day groups from the Levant, Iraq, and the Caucasus, including several Jewish populations from these regions, and between ancient Lebanese and an Iranian IA individual.

To gain further insights into the relationship of Tylos to present-day populations from the Arabian Peninsula and the Levant, we tested whether the Bahraini samples form a clade with any modern population in our dataset using a set of reference populations (as outgroups) that can differentiate the different ancestries in the Near East. We found that Iraqis, Assyrians, and Jewish groups from Iran, Georgia, and Iraq could derive all their ancestries from Tylos-period Bahrainis (Table S10). Arabians such as Saudis, Emiratis, and Yemenis have, in addition to ancestry from Tylos-period Bahrain, ancestry from East Africa, while Levantines such as Druze and Lebanese have additional Southeast European ancestry (Table S11). Here we should note that modeling present-day populations with Tylos-period Bahrain does not imply a direct contribution but rather that they are suitable representatives of the ancestry found in present-day populations.

Diverse uniparental lineages link ancient Bahrainis to Middle Eastern populations

We used pathPhynder[40] to examine the Y chromosome lineages of the two male individuals from Late Tylos-period Bahrain in a context of present-day and ancient samples (Figure S5 and Table S12). Individual MH3_LT carried the H2 haplogroup, which is associated with the spread of Near Eastern and Anatolian farmers into Europe.[41] MH3_LT was placed in the same clade as a PPNB Jordanian and two Anatolian samples (Alalakh_MLBA and Ilipinar_ChL), which is consistent with the genetic affinities of this individual (Figure S5A).

Sample MH1_LT, the only other male sample in our dataset, presented the J2a2a1a∼ lineage, which in a phylogeny with 2,014 individuals is carried by a present-day Brahmin and a Uighur individual. In the aDNA record, this lineage and derived haplotypes have been identified in two Turkish individuals (Gordion_Anc and Medieval), in a Canaanite, and in an Iron Age Hasanlu individual,[12,42] and in various present-day Central Asian samples from Turkmenistan and Kazakhstan. An eastern origin (Iran-Caucasus region) for haplogroup J2 is likely, given

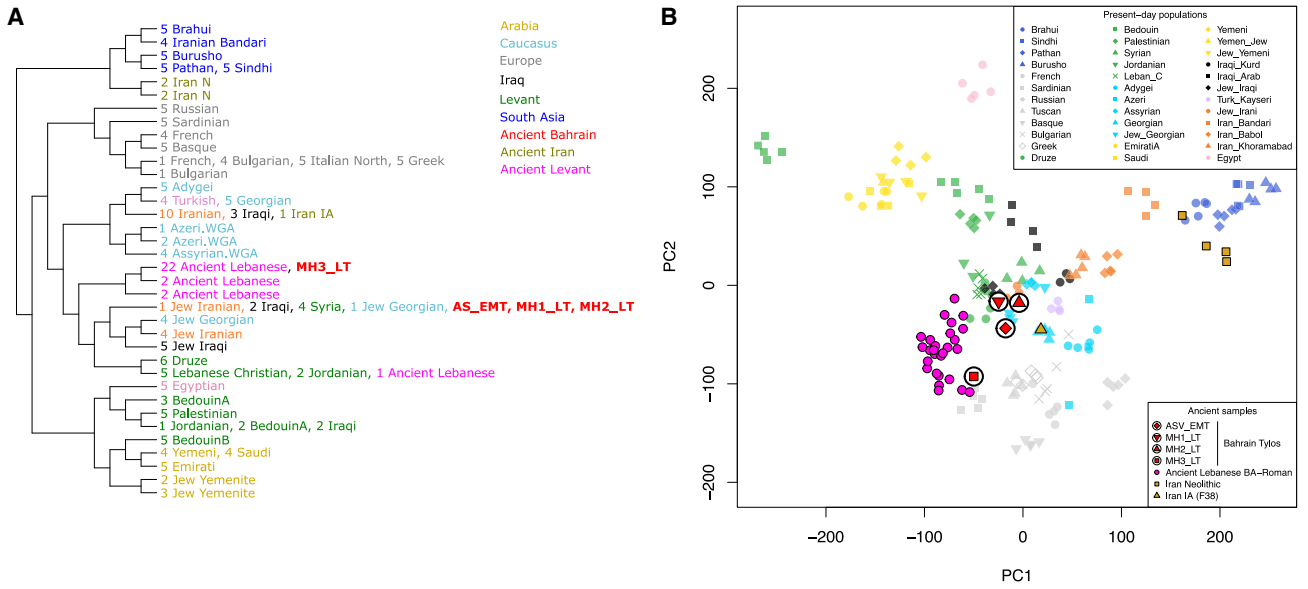


Figure 4. Haplotype-based population affinities

(A) FineSTRUCTURE phylogeny and clustering based on haplotype sharing patterns from present-day Eurasians and ancient Middle Easterners.

(B) PC analysis estimated from the ChromoPainter haplotype sharing matrix.

See also Figure S5 and Tables S10, S11, S12, and S13.

that the earliest occurrence of this lineage in the aDNA record is in hunter-gatherers from Caucasus and Iran,[11,43] with the latter being placed at the base of the J2a clade (Figure S5B). MH1_LT's inclusion in this Y chromosome clade is consistent with the subtle excess in shared autosomal ancestry with CHG/IRN_HajjiFiruz_IA individuals in this sample (Figure 2C).

Strikingly, neither of these lineages was found in present-day Arabians.[20] Their presence in both western (Anatolia/Levant) and eastern (Iran/Caucasus/Central Asia) ancient samples reflects connectivity between these ancient civilizations as early as the PPN (as attested by the presence of eastern ancestry and J2 Y chromosome lineage in Cayonu_PPN[44]) which apparently intensified during post-Neolithic times.[42]

The mitochondrial lineages of the Tylos-period samples suggest maternal ancestry sharing with various groups from the Near East, Caucasus, and South Asia. The earliest sample AS_EMT carries the mtDNA J1c15a1 haplogroup previously reported in two present-day samples from Iraq[45] and Azerbaijan,[46] groups which typically carry Iranian- and CHG-related ancestry. The Late Tylos samples from Madinat Hamad presented distinct lineages, with MH1_LT belonging to the R2 haplogroup, which is predominantly South Asian (southern Pakistan and India), but also distributed in the Middle East, Caucasus, and Central Asia.[47] In the aDNA record this lineage is most frequent in Iranian groups, including three Neolithic and two BIA/LBA individuals,[42,48] which is consistent with the additional post-Neolithic Iranian-related ancestry in this sample. Interestingly, its presence in a Bronze Age Canaanite, coinciding with the emergence of Iranian ancestry in the Levant, provides additional support for the association of R2 lineages with this source of ancestry. Sample MH2_LT carries the T2b lineage, of widespread distribution in European samples. Lastly, MH3_LT presents the U8b1a2a haplogroup, also found in two ancient LBA Armenians, two Turkish individuals (one ChL and the other dated to 750–480 BCE),[42] and a present-day Jordanian individual,[49] suggesting distribution in the Caucasus, Levant, and Anatolia and consistency with the increased Levantine and Anatolian affinities of this sample. These observations support extensive female as well as male migration in ancient West Asia, including Eastern Arabia.

Phenotypic prediction

We used HIrisPlex-S[50] to predict hair, skin, and eye color phenotypes of the Bahrain_Tylos samples. All four samples were predicted to have brown eyes (>99% probability). For hair color, the prediction was either brown (~50%) or black (~50%). Two samples (MH2_LT and MH3_LT) were predicted to have "dark" skin pigmentation (>90%), whereas the results for the remaining two samples were less certain, with AS_EMT potentially having relatively lighter skin pigmentation (Table S13). The predicted phenotypes found in ancient Bahrain are similar to those in present-day Middle Easterners and South Asians. Within these regions, subtle geographical trends occur at the level of skin pigmentation. Specifically, non-Bedouin Levantine groups from the Human Genome Diversity Project (HGDP) tend to have higher proportions of "intermediate" skin pigmentation, while Pakistanis tend to have higher proportions of "dark" skin than most Levantine populations, especially in the south.[50]

Runs of homozygosity suggest a spectrum of inbreeding similar to present-day Middle Easterners

We also examined ROHs in the Tylos samples and compared them with present-day and ancient genomes (Figure 5A). Contemporary Middle Eastern populations show long tracts of ROHs indicative of recent consanguinity, while ancient regional

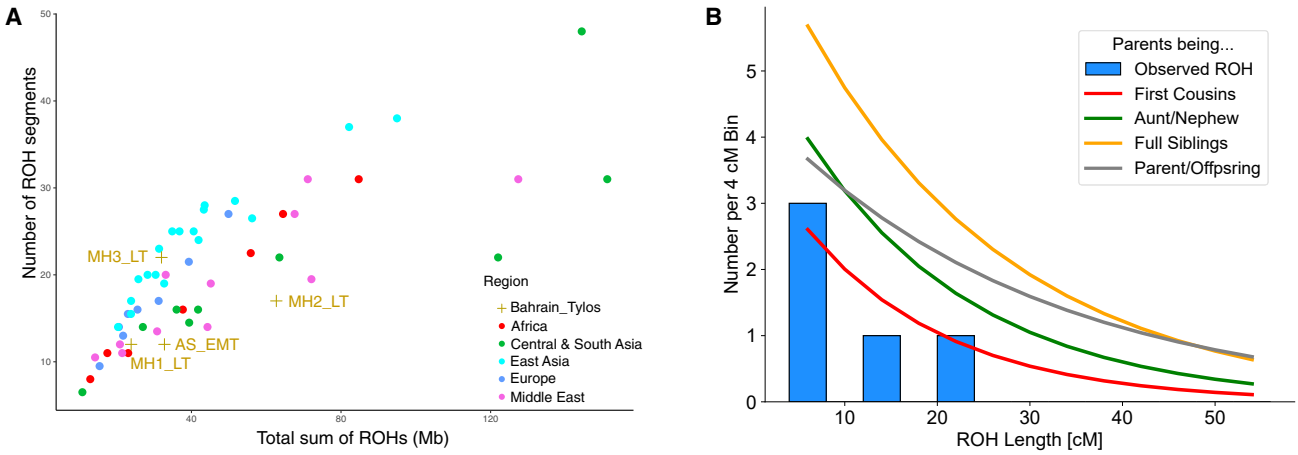


Figure 5. Runs of homozygosity

(A) Runs of homozygosity (ROHs) in Bahrain_Tylos and worldwide populations (plotting median values for each population).

(B) ROH length (>4 cM) distribution in sample MH2_LT and expected ROH distribution for the offspring of different consanguineous unions.

See also Figure S6.

hunter-gatherer groups are known to have relatively large numbers of shorter ROHs, reflecting their smaller population size.[51] We observed an excess of ROH size in sample MH2_LT which is similar to that seen in present-day Middle Eastern and South/Central Asian populations (Figure 5A). In comparison to ancient groups, we find that three individuals (AS_EMT, MH1_LT, and MH3_LT), show an ROH distribution similar to those of Bronze and Iron Age regional populations (Figure S6A). However, one sample (MH2_LT) appears to have larger ROHs, with one longer than 16 Mb, suggesting direct parental relatedness (Figure S6B).

To confirm this result, we then used an alternative method[52] for identifying ROHs in ancient samples. We observe that MH2_LT has levels of inbreeding consistent with being the offspring of related individuals (potentially second cousins; Figures 5B and S6C), including a stretch on chr7 that spans 20.47 cM (Figure S6D). This finding is evidence that consanguineous union was likely to have already been practiced in pre-Islamic Arabian societies.

High prevalence of the malaria-protective G6PD Mediterranean mutation in ancient Eastern Arabia coincided with the introduction of agriculture

G6PD deficiency is the most common enzymatic defect in humans,[53] and its distribution in worldwide populations correlates with regions currently or historically affected by malaria, including Africa, the Mediterranean, the Middle East, and Southeast Asia,[54] leading to the proposal of a link between G6PD deficiency and malaria protection.[55]

Considering that malaria was endemic in Bahrain during historical times,[56] with osteological analyses suggesting that it was present in the island at least by the Tylos period,[57] we examined the Tylos-period Bahraini samples for the presence of mutations putatively associated with malaria protection. Two ancient samples from Bahrain carried the G6PD Mediterranean mutation (rs5030868; G>A; p.S188F). MH2_LT is potentially heterozygous, given that we observed two reads overlapping this SNP in this sample, one supporting the derived allele A and the other with the ancestral allele G, whereas EMT presented two reads with the derived allele.

To corroborate our findings and obtain additional insights about the distribution of this variant in ancient populations, we imputed the X chromosome of the Bahrain_Tylos samples together with 37 ancient Lebanese and four ancient Iranians. We confirm the presence of derived alleles (genotype probability >0.95) at the rs5030868 locus in samples AS_EMT (homozygous), MH2_LT (heterozygous), and MH3_LT (male; hemizygous mutant). None of the ancient Lebanese we imputed carried this mutation and neither did five individuals from Bronze Age Greece with sequencing reads spanning this site,[58] but we observed a heterozygous genotype in a Western Iranian Neolithic sample (AH1, genotype probability >0.9). When estimating a phylogeny with the imputed X chromosome haplotypes containing the mutation of interest, we observe the inclusion of three Tylos-period Bahrainis in a clade with additional samples carrying the derived allele, including present-day Emiratis, Omanis, and Yemenis, as well as the Iranian Neolithic sample AH1 (Figure 6A).

In present-day populations, we observe relatively high frequencies of this variant in Makrani (0.19), Brahui (0.14), and Pathan (0.08) population samples, all of which bear high levels of Iranian Neolithic-related ancestry. In Europe frequencies are substantially lower, with a single Sardinian sample carrying the derived allele in the HGDP dataset (0.03), and 3 out of 53,000 European individuals in the gnomAD database.[59] In the Middle East, it can be found in Palestinians (0.06) but not in Druze, Bedouins, Jordanians, or Iraqis.[20] In Arabian populations, the derived allele reaches its highest frequencies in EmiratiC (0.38), a subgroup of the Emirati who carry substantial Iranian/South Asian-related ancestry (Figure 1B), and 6% in the rest of the Emirati population. It has also been found in the present-day populations of Yemen (0.04) and Qatar

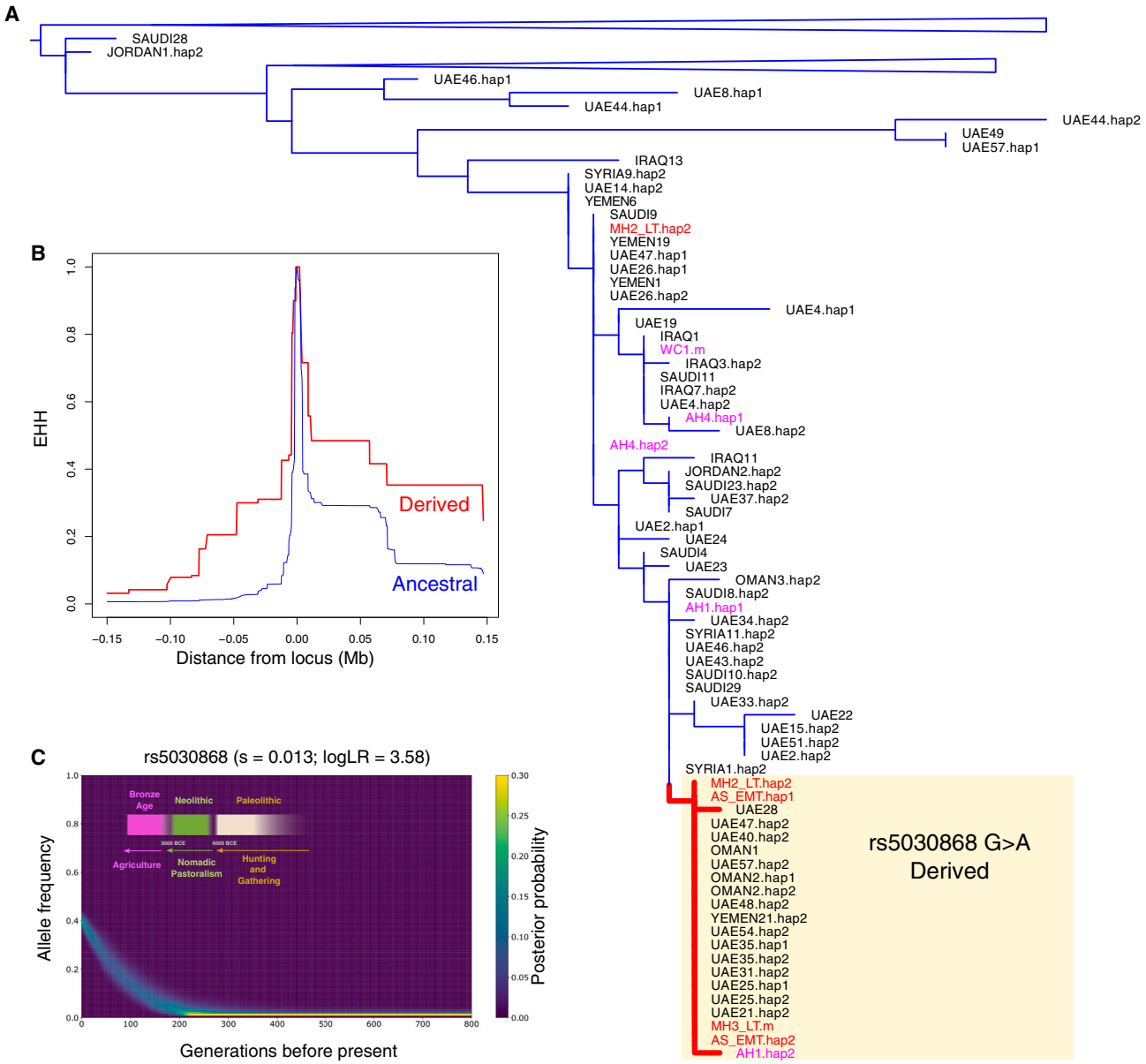


Figure 6. Malaria adaptation in Eastern Arabia

(A) Phylogeny of present-day and ancient haplotypes from Middle Easterners surrounding the G6PD Mediterranean mutation (rs5030868).

(B) Extended haplotype homozygosity (EHH) in present-day Arabians.

(C) Allele frequency trajectory of the variant in present-day Emiratis (EmiratiC) highlighting changes in modes of subsistence in Eastern Arabia.

See also Figure S7.

(0.03),[32] but not in Saudis.[20] In the Levant and Eastern Arabia, the Mediterranean mutation is responsible for the majority of G6PD-deficient cases, reaching very high conditional frequencies in Bahrainis (0.91),[60] Northern Iraqis (0.88), Kuwaitis (0.74), and Jordanians (0.54).[61]

To search for signatures of selection associated with the G6PD Mediterranean, we performed an extended haplotype homozygosity (EHH) analysis on the haplotypes surrounding this SNP in a dataset of present-day Arabian and Levantine groups. We observe that the haplotypes carrying the derived allele tend to be longer than those with the ancestral allele (Figure 6B), consistent with a signature of positive selection. We subsequently modeled the historical change in frequency of the variant and found that it has increased rapidly in the past 6,000 years in present-day Emiratis (EmiratiC) with an estimated selection coefficient of 0.013, suggesting strong selective pressure (Figure 6C). This date broadly coincides with the start of the Bronze Age in Eastern Arabia (~3200 BCE), a period of cultural transformation marked by the shift from nomadic pastoralism to agriculture.[62]

DISCUSSION

Population heterogeneity and admixture in Tylos-period Bahrain

In the present work, we address an important gap in the aDNA record by presenting whole-genome sequences from Eastern Arabia, more precisely from Tylos-period Bahrain, an era of Hellenistic, Parthian, and Sassanian influence in the region. We observe that the four individuals from the Tylos period occupy an intermediate position along the genetic ancestry cline spanning from western (Anatolian/Levantine) to eastern (Iran/CHG) sources, with some degree of heterogeneity in ancestry composition within our samples.

We detected an excess of Levantine-related ancestry in sample MH3_LT, which raises the possibility of admixture with populations from the Levant or as yet unsampled Arabian groups that may have carried this ancestry into Bahrain before or around the onset of Islam. The first hypothesis is supported by the substantial affinity of MH3_LT to Lebanese Canaanites and other Levantine groups and by shared Y chromosome and mitochondrial DNA lineages between this individual and ancient eastern Anatolians and Levantines. The second hypothesis is supported by the existence of historical records attesting to the presence of several nomadic Arab tribes, such as the Abd Al-Qays and Al-Adz, under Nasrid control[63] in Bahrain during this time. These tribes migrated from Southwestern Arabia into Bahrain and neighboring regions in pre-Islamic times, subsequently participating in the conquests of Persia and Mesopotamia in the second half of the 7th century CE.[26] Such migrations could have brought additional Levantine ancestry to Bahrain, given that Southwestern Arabians, such as the Yemeni from Maarib, carry substantial Natufian-related ancestry but virtually no African ancestry,[19] as is the case of sample MH3_LT, and are, therefore, potentially good representatives of Arabian ancestry prior to admixture with other sources including Africans.

The presence of minor steppe-related ancestry in the Late Tylos samples (MH1_LT and MH2_LT), but not in the earliest sample from Abu Saiba (AS_EMT), does not require a direct influx from steppe groups, associated, for example, with the LBA migrations that spread Indo-European languages into South Asia.[33,64] Instead, this ancestry may have been introduced to Bahrain through admixture with geographically and temporally more proximal groups from ancient Levant or Iran where it appears around the Iron Age[15,38] and Bronze Age,[33] respectively. Considering that Bahrain history is marked by Achaemenid, Parthian, and Sasanian influence, it is plausible that groups associated with these empires acted as vectors for introducing steppe ancestry into Bahrain. In support of this hypothesis, the J2a2a1 Y chromosome lineage of sample MH1_LT is also present in the Iranian IA and in other ancient groups with substantial Iranian-related ancestry (Central Asian, Anatolian, Levantine BA), but neither of the two males reported here carried the R1a-Z93 lineage associated with the spread of steppe ancestry and Indo-European languages into South Asia.[33]

We inferred admixture times between Levantine and Iranian sources occurring first, during the Achaemenid Empire, and second, during a time of Characene rule of Bahrain, suggesting that political and cultural changes were accompanied by gene flow. However, this analysis suffers from several limitations. First, it is possible that this finding may derive from a lack of power to determine exact admixture times in the case of continuous gene flow. Second, the lack of direct radiocarbon dates for two of our samples, AS_EMT and MH2_LT, limits precise estimation of admixture timing.

Given the limited sample size (n = 4) and the narrow temporal focus on a specific period in Bahrain's history, it is important to emphasize that it remains challenging to establish whether these ancestry differences derive from specific population movements occurring alongside political changes in the Near East after the Dilmun period or if they reflect pre-existing genetic diversity within the island's population. As seen in our PCA (Figure 1A), Near Eastern groups such as Iron Age Levantines and Iranians present relatively high diversity and, therefore, it is feasible that the differences we observe between our samples derive from pre-existing genetic structure within the Bahraini population not directly related to the events mentioned here. Furthermore, while we were able to identify several plausible models for describing the ancestry of Tylos-period Bahrainis, the current lack of ancient genomes from the Arabian Peninsula and sparse sampling from Mesopotamia (particularly from the south, from where no aDNA samples have been published so far) prevents us from testing models with more proximal sources relevant to Bahrain. Nevertheless, various analyses point to some degree of genetic similarity between a Late Bronze Age sample (IRQ_Nemrik9_LBA) and the Bahraini samples, which gathers support from archaeological and textual evidence of contacts between Mesopotamia and Bahrain at the time of the Dilmun civilization. However, as this observation derives from a single Mesopotamian sample, this connection should be re-examined when additional ancient genomes become available. A particularly important question remains regarding the ancestry of the Dilmun-period inhabitants, which, once characterized, will help describe more precisely how the genetic composition of Arabians has evolved through time, the relationship of Dilmun with neighboring civilizations, and their contribution to later Tylos-period groups.

Increased malaria endemicity in Eastern Arabia

We report the presence of the G6PD Mediterranean mutation in three out of four samples and four out of six alleles from ancient Bahrain. Notably, this includes a homozygous female and a hemizygous male, who would have suffered G6PD deficiency and, possibly, hemolytic anemia. This finding suggests that this mutation occurred at an appreciable frequency in Eastern Arabia during the Tylos period. The identification of this variant in an Iranian Neolithic sample, but not in ancient Levant or ancient Europe, together with its prevalence in present-day groups from Pakistan and Arabian groups with high Iran N-related ancestry (EmiratiC), challenges the hypothesis that it was introduced into the Middle East in relatively recent times alongside the Greek expansions of the first millennium BCE[65] or that its dispersal into Europe occurred alongside Neolithic migrations.[66] Our data are consistent with an alternative history in which this mutation was disseminated into Arabia through admixture with Mesopotamian/Iranian/South Asian groups,

rising to high frequency due to selective pressures, particularly along the Arabian coast, where malaria incidence was especially high. According to our estimates, the onset of selection for this variant in Eastern Arabia occurred between 5 and 6 kya, broadly coinciding with the emergence of agriculture in the region. Such changes are evidenced by the presence of domesticated cereals at Al-Hili dated to around 3000 BCE[62] and the development of oasis agriculture associated with the Umm an-Nar culture (2700–2000 BCE).[67] Increased sedentarism in areas with available water[68] would have created a propitious environment for the proliferation of malaria, causing a rise in frequency of malaria-protective variants.

Our confidence in imputing the Mediterranean variant in the three samples that carried the derived allele is supported by direct observation of the variant in a total of three reads spanning that position in two individuals. Although this SNP is a G/A change, potentially causing difficulties in distinguishing it from postmortem deamination, this is likely to be more problematic for the Iranian Neolithic sample than for our uracil-DNA-glycosylase-treated samples. Nevertheless, this Iranian individual shares the same haplotype as the other derived samples, providing additional confidence in it being a carrier. Expanding our investigation of the Mediterranean variant to a wider set of ancient groups would be desirable but currently challenging because this SNP is not included as part of the 1,240k SNP target capture array commonly used for sequencing ancient samples.[69] Future studies generating additional whole ancient genomes from the Mediterranean region, the Middle East, and South Asia should provide more insights about the spread and distribution of malaria-protective variants in the ancient world.

Lastly, we demonstrate the feasibility of aDNA studies in Arabia, paving the way for future research aimed at elucidating human population movements in the region. The sequence data and insights reported here will be of long-term importance for studying human population history in the Middle East and beyond.

Limitations of the study

One of the main limitations of our study is the small sample size, comprising only four individuals. Due to the hot climate of Eastern Arabia, DNA preservation in skeletal material tends to be poor, preventing us from obtaining a larger number of samples with sufficient endogenous content for analysis. Furthermore, the four individuals here reported belong exclusively to the Tylos period, creating challenges in determining exactly how genetic composition changed in Bahrain from the Dilmun period to the present day. Due to poor collagen preservation, we were also unable to obtain radiocarbon dates for two samples, introducing uncertainties at the level of temporal placement and estimation of admixture times. Additionally, we currently lack aDNA from mainland Arabia and Southern Mesopotamia, which could bring additional clarity to our current ancestry models. Future studies, possibly equipped with novel methodological developments, should be directed toward retrieving a larger number of samples from Arabia and Mesopotamia and covering a larger temporal range to provide further insights about the human population history of the region.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Archaeological sample processing, DNA extraction, library preparation and sequencing
  - Radiocarbon dating
  - Sequence read processing and alignment
  - Sex determination and relatedness analysis
  - Authenticity of aDNA sequences and contamination estimates
  - Population genetics analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100507.

ACKNOWLEDGMENTS

R.M. was funded by Wellcome grant 207492 and National Geographic Early Career grant EC-61437R-19, and R.D. by Wellcome grant 207492. The Abu Saiba excavation was funded by the French Ministry for Europe and Foreign Affairs and by the Bahrain Authority for Culture and Antiquities. D.G.B. was supported by Science Foundation Ireland/Health Research Board/Wellcome Trust Biomedical Research Partnership Investigator award no. 205072, "Ancient Genomics and the Atlantic Burden." M.A.A. was supported by Dubai Police GHQ. We thank the Wellcome Sanger Institute DNA Pipelines core for sequencing. We thank the President of Bahrain Authority for Culture and Antiquities, Shaikh Khalifa Bin Ahmed Al-Khalifa, for supporting the present research project and for providing access to archaeological samples for aDNA analysis. For the purpose of open access, the lead author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

AUTHOR CONTRIBUTIONS

R.M. and F.A. conceived the project. B.C., S.A., and P.L. selected and provided archaeological samples. P.L. provided information about the archaeological context with contributions from B.C. and J.L. R.M. processed bone samples. V.M. and E.M.B. carried out extractions and prepared sequencing libraries. R.M., M.H., M.A.A., and M.C.M.K. carried out analyses and interpreted them. R.M., M.H., M.A.A., P.L., and R.D. drafted the manuscript with input from D.G.B. All authors commented on and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 4, 2023
Revised: November 1, 2023
Accepted: January 31, 2024
Published: February 27, 2024

REFERENCES

1. Larsen, C.E. (1983). Life and Land Use on the Bahrain Islands: The Geoarchaeology of an Ancient Society (University of Chicago Press).
2. Crawford, H.E.W. (1998). Dilmun and its Gulf Neighbours (Cambridge University Press).
3. Nashef, K. al (1984). The deities of Dilmun. Akkadica Bruxelles *1–33*.
4. Andre-Salvini, B. (2000). The Land Where the Sun Rises. The Representation of Dilmun in Sumerian Literature. In The Civilisation of the Two Seas, from Dilmun to Tylos. An Exhibition at the Institut du Monde Arabe, P. Lombard, ed., pp. 42–48. Paris.
5. Højlund, F. (2007). The Burial Mounds of Bahrain. Social Complexity in Early Dilmun, *58* (Jutland Archaeological Society Publications).
6. Potts, D.T. (2010). In Achaemenid interests in the Persian Gulf, J. Curtis and S. John Simpson, eds., pp. 523–534, London–New York.
7. Lombard, P., Chamel, B., Cuny, J., Cotty, M., Guermont, F., Lux, R., and Noca, L. (2020). Les fouilles françaises de Abu Saiba (Mont 1). Données nouvelles sur la phase Tylos de Bahreïn (c. 200 av. J.-C. – 300 ap. J.-C.). In Proceedings of the Seminar for Arabian Studies, pp. 225–241.
8. Gatier, P.-L., Lombard, P., and Al-Sindi, K.M. (2002). Greek Inscriptions from Bahrain. Arabian Archaeol. Epigr. *13*, 223–233.
9. Kosmin, P. (2013). Rethinking the Hellenistic Gulf: The New Greek Inscription from Bahrain. J. Hellenic Stud. *133*, 61–79.
10. Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Díez-del-Molino, D., et al. (2016). Early Neolithic genomes from the eastern Fertile Crescent. Science, aaf7943.
11. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.
12. Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., et al. (2022). Ancient DNA from Mesopotamia suggests distinct Pre-Pottery and Pottery Neolithic migrations into Anatolia. Science *377*, 982–987.
13. Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S., Martiniano, R., Prado-Martinez, J., Szpak, M., et al. (2017). 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.
14. Harney, É., May, H., Shalem, D., Rohland, N., Mallick, S., Lazaridis, I., Sarig, R., Stewardson, K., Nordenfelt, S., Patterson, N., et al. (2018). Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. Nat. Commun. *9*, 3336.
15. Haber, M., Nassar, J., Almarri, M.A., Saupe, T., Saag, L., Griffith, S.J., Doumet-Serhal, C., Chanteau, J., Saghieh-Beydoun, M., Xue, Y., et al. (2020). 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.
16. Feldman, M., Master, D.M., Bianco, R.A., Burri, M., Stockhammer, P.W., Mittnik, A., Aja, A.J., Jeong, C., and Krause, J. (2019). Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. Sci. Adv. *5*, eaax0061.
17. Altınışık, N.E., Kazancı, D.D., Aydoğan, A., Gemici, H.C., Erdal, Ö.D., Sarıaltun, S., Vural, K.B., Koptekin, D., Gürün, K., Sağlıcan, E., et al. (2022). A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic Transition. Sci. Adv. *8*, eabo3609.
18. Kistler, L., Ware, R., Smith, O., Collins, M., and Allaby, R.G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. Nucleic Acids Res. *45*, 6310–6320.
19. Haber, M., Saif-Ali, R., Al-Habori, M., Chen, Y., Platt, D.E., Tyler-Smith, C., and Xue, Y. (2019). Insight into the genomic history of the Near East from whole-genome sequences and genotypes of Yemenis. Preprint at bioRxiv. https://doi.org/10.1101/749341.
20. Almarri, M.A., Haber, M., Lootah, R.A., Hallast, P., Al Turki, S., Martin, H.C., Xue, Y., and Tyler-Smith, C. (2021). The genomic history of the Middle East. Cell *184*, 4612–4625.e14.
21. Petraglia, M.D., Groucutt, H.S., Guagnin, M., Breeze, P.S., and Boivin, N. (2020). Human responses to climate and ecosystem change in ancient Arabia. Proc. Natl. Acad. Sci. USA *117*, 8263–8270.
22. Weiss, H. (2016). Global megadrought, societal collapse and resilience at 4.2-3.9 ka BP across the Mediterranean and west Asia. PAGES. Mag. *24*, 62–63.
23. Bittles, A.H., and Hamamy, H.A. (2010). Endogamy and Consanguineous Marriage in Arab Populations. In Genetic Disorders Among Arab Populations, A.S. Teebi, ed. (Springer Berlin Heidelberg), pp. 85–108.
24. Elliott, K.S., Haber, M., Daggag, H., Busby, G.B., Sarwar, R., Kennet, D., Petraglia, M., Petherbridge, L.J., Yavari, P., Heard-Bey, F.U., et al. (2022). Fine-Scale Genetic Structure in the United Arab Emirates Reflects Endogamous and Consanguineous Culture, Population History, and Geography. Mol. Biol. Evol. *39*, msac039. https://doi.org/10.1093/molbev/msac039.
25. Ferreira, J.C., Alshamali, F., Pereira, L., and Fernandes, V. (2022). Characterization of Arabian Peninsula whole exomes: exploring high inbreeding features. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.22.481461.
26. al-A'ni, A.K. (1993). The inhabitants of Bahrain at the emergence of Islam. In Bahrain Through the Ages: The History, S.A.A.-K. Rice, ed., pp. 13–24.
27. Chamel, B., and Lombard, P. (2021). Abu Saiba, une nécropole de la phase de Tylos à Bahreïn (200 BC-300 AD) : pratiques funéraires, étude du recrutement et des gestes de pillage. Bull. Mem. Soc. Anthropol. Paris *34*. S) | 2021 34 Supplément | 2022 *34*.
28. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190–e2093.
29. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
30. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.
31. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
32. Rodriguez-Flores, J.L., Fakhro, K., Agosto-Perez, F., Ramstetter, M.D., Arbiza, L., Vincent, T.L., Robay, A., Malek, J.A., Suhre, K., Chouchane, L., et al. (2016). Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. Genome Res. *26*, 151–162.
33. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487. https://doi.org/10.1126/science.aat7487.
34. Skourtanioti, E., Erdal, Y.S., Frangipane, M., Balossi Restelli, F., Yener, K.A., Pinnock, F., Matthiae, P., Özbal, R., Schoop, U.-D., Guliyev, F., et al. (2020). Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. Cell *181*, 1158–1175.e28.
35. Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

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

37. Maier, R., Flegontov, P., Flegontova, O., Işıldak, U., Changmai, P., and Reich, D. (2023). On the limits of fitting complex models of population history to f-statistics. Elife *12*, e85492. https://doi.org/10.7554/eLife.85492.

38. Haber, M., Doumet-Serhal, C., Scheib, C.L., Xue, Y., Mikulski, R., Martiniano, R., Fischer-Genz, B., Schutkowski, H., Kivisild, T., and Tyler-Smith, C. (2019). A Transient Pulse of Genetic Admixture from the Crusaders in the Near East Identified from Ancient Genome Sequences. Am. J. Hum. Genet. *104*, 977–984.

39. Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. Science *367*, eaay5012. https://doi.org/10.1126/science.aay5012.

40. Martiniano, R., De Sanctis, B., Hallast, P., and Durbin, R. (2022). Placing ancient DNA sequences into reference phylogenies. Mol. Biol. Evol. *39*, msac017. https://doi.org/10.1093/molbev/msac017.

41. Rohrlach, A.B., Papac, L., Childebayeva, A., Rivollat, M., Villalba-Mouco, V., Neumann, G.U., Penske, S., Skourtanioti, E., van de Loosdrecht, M., Akar, M., et al. (2021). Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe. Sci. Rep. *11*, 15005.

42. Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., et al. (2022). The genetic history of the Southern Arc: A bridge between West Asia and Europe. Science *377*, eabm4247.

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

44. Altınışık, N.E., Kazancı, D.D., Aydoğan, A., Gemici, H.C., Erdal, Ö.D., Sarıaltun, S., Vural, K.B., Koptekin, D., Gürün, K., Sağlıcan, E., et al. (2022). A genomic snapshot of demographic and cultural dynamism in Upper Mesopotamia during the Neolithic Transition. Sci. Adv. *8*. https://doi.org/10.1101/2022.01.31.478487.

45. Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., Tamm, E., Karmin, M., Reisberg, T., Hooshiar Kashani, B., et al. (2012). Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. Am. J. Hum. Genet. *90*, 915–924.

46. Schönberg, A., Theunert, C., Li, M., Stoneking, M., and Nasidze, I. (2011). High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. Eur. J. Hum. Genet. *19*, 988–994.

47. Quintana-Murci, L., Chaix, R., Wells, R.S., Behar, D.M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A.S., et al. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. Am. J. Hum. Genet. *74*, 827–845.

48. Ehler, E., Novotný, J., Juras, A., Chylenski, M., Moravcík, O., and Paces, J. (2019). AmtDB: a database of ancient human mitochondrial genomes. Nucleic Acids Res. *47*, D29–D32.

49. González, A.M., García, O., Larruga, J.M., and Cabrera, V.M. (2006). The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. BMC Genom. *7*, 124.

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

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

52. Ringbauer, H., Novembre, J., and Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. Nat. Commun. *12*, 5425.

53. Cappellini, M.D., and Fiorelli, G. (2008). Glucose-6-phosphate dehydrogenase deficiency. Lancet *371*, 64–74.

54. Nkhoma, E.T., Poole, C., Vannappagari, V., Hall, S.A., and Beutler, E. (2009). The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. Blood Cells Mol. Dis. *42*, 267–278.

55. Beutler, E. (1994). G6PD deficiency. Blood *84*, 3613–3636.

56. Snow, R.W., Amratia, P., Zamani, G., Mundia, C.W., Noor, A.M., Memish, Z.A., Al Zahrani, M.H., Al Jasari, A., Fikri, M., and Atta, H. (2013). The malaria transition on the Arabian Peninsula: progress toward a malaria-free region between 1960-2010. Adv. Parasitol. *82*, 205–251.

57. Littleton, J. (1998). Skeletons and Social Composition: Bahrain 300 BC-AD 250 (Archaeopress).

58. Clemente, F., Unterländer, M., Dolgova, O., Amorim, C.E.G., Coroado-Santos, F., Neuenschwander, S., Ganiatsou, E., Cruz Dávalos, D.I., Anchieri, L., Michaud, F., et al. (2021). The genomic history of the Aegean palatial civilizations. Cell *184*, 2565–2586.e21.

59. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.

60. Al Moamen, N., and Al Arrayed, S. (2004). Molecular homogeneity of G6PD deficiency. Bahrain Med. Bull., 139–142.

61. Doss, C.G.P., Alasmar, D.R., Bux, R.I., Sneha, P., Bakhsh, F.D., Al-Azwani, I., Bekay, R.E., and Zayed, H. (2016). Genetic Epidemiology of Glucose-6-Phosphate Dehydrogenase Deficiency in the Arab World. Sci. Rep. *6*, 37284.

62. Potts, D.T. (1994). Contributions to the agrarian history of Eastern Arabia II. The cultivars. Arabian Archaeol. Epigr. *5*, 236–275.

63. Fisher, G. (2011). Kingdoms or Dynasties? Arabs, History, and Identity before Islam. jla. *4*, 245–267.

64. de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science *360*, eaar7711. https://doi.org/10.1126/science.aar7711.

65. Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., et al. (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science *293*, 455–462.

66. Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. (1994). The History and Geography of Human Genes (Princeton University Press).

67. Méry, S. (2013). The first oases in Eastern Arabia: society and craft technology, in the 3rd millennium BC at Hili, United Arab Emirates. ethnoecologie. https://doi.org/10.4000/ethnoecologie.1631.

68. al-Jahwari, N.S. (2009). The agricultural basis of Umm an-Nar society in the northern Oman peninsula (2500-2000 BC). Arabian Archaeol. Epigr. *20*, 122–133.

69. Rohland, N., Mallick, S., Mah, M., Maier, R., Patterson, N., and Reich, D. (2022). Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. Genome Res. *32*, 2068–2078.

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

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

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

73. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034.

74. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. J. Archaeol. Sci. *40*, 4477–4482.

75. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS One *13*, e0195491.

76. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Paabo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. In Proceedings of the National Academy of Sciences.

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

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

79. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007.

80. Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. *25*, 918–925.

81. Joseph, T.A., and Pe'er, I. (2019). Inference of Population Structure from Time-Series Genotype Data. Am. J. Hum. Genet. *105*, 317–333.

82. Ní Leathlobhair, M., Perri, A.R., Irving-Pease, E.K., Witt, K.E., Linderholm, A., Haile, J., Lebrasseur, O., Ameen, C., Blick, J., Boyko, A.R., et al. (2018). The evolutionary history of dogs in the Americas. Science *361*, 81–85.

83. Liu, L., Bosse, M., Megens, H.-J., Frantz, L.A.F., Lee, Y.-L., Irving-Pease, E.K., Narayan, G., Groenen, M.A.M., and Madsen, O. (2019). Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. Nat. Commun. *10*, 1992.

84. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. PLoS Genet. *8*, e1002453.

85. Browning, B.L., Tian, X., Zhou, Y., and Browning, S.R. (2021). Fast two-stage phasing of large-scale sequence data. Am. J. Hum. Genet. *108*, 1880–1890.

86. Eggertsson, H.P., Jonsson, H., Kristmundsdottir, S., Hjartarson, E., Kehr, B., Masson, G., Zink, F., Hjorleifsson, K.E., Jonasdottir, A., Jonasdottir, A., et al. (2017). Graphtyper enables population-scale genotyping using pangenome graphs. Nat. Genet. *49*, 1654–1660.

87. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nat. Genet. *53*, 120–126.

88. Speidel, L., Forest, M., Shi, S., and Myers, S.R. (2019). A method for genome-wide genealogy estimation for thousands of samples. Nat. Genet. *51*, 1321–1329.

89. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.

90. Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F., and Schönherr, S. (2016). mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. Nucleic Acids Res. *44*, W64–W69.

91. Breslin, K., Wills, B., Ralf, A., Ventayol Garcia, M., Kukla-Bartoszek, M., Pospiech, E., Freire-Aradas, A., Xavier, C., Ingold, S., de La Puente, M., et al. (2019). HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms. Forensic Sci. Int. Genet. *43*, 102152.

92. Chintalapati, M., Patterson, N., and Moorjani, P. (2022). The spatiotemporal patterns of major human admixture events during the European Holocene. Elife *11*, e77625. https://doi.org/10.7554/eLife.77625.

93. Szpiech, Z.A., and Hernandez, R.D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. *31*, 2824–2827.

94. Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L.S., Cooper, A., and Haak, W. (2017). From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. Star: Science & Technology of Archaeological Research *3*, 1–14.

95. Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., and Allentoft, M.E. (2015). Improving access to endogenous DNA in ancient bones and teeth. Sci. Rep. *5*, 11184.

96. Mattiangeli, V. (2023). Multi-step ancient DNA extraction protocol for bone and teeth v1. https://doi.org/10.17504/protocols.io.6qpvr45b2gmk/v1.

97. Boessenkool, S., Hanghøj, K., Nistelberger, H.M., Der Sarkissian, C., Gondek, A.T., Orlando, L., Barrett, J.H., and Star, B. (2017). Combining bleach and mild predigestion improves ancient DNA recovery from bones. Mol. Ecol. Resour. *17*, 742–751.

98. Mattiangeli, V. (2023). Bleach Extraction Protocol: Damaged or Degraded DNA Recovery from Bone or Tooth Powder. V1. https://doi.org/10.17504/protocols.io.8epv5j88nl1b/v1.

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

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

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

102. Martiniano, R., Garrison, E., Jones, E.R., Manica, A., and Durbin, R. (2020). Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph. Genome Biol. *21*, 250.

103. Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. Science *334*, 94–98.

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

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

106. Agranat-Tamir, L., Waldman, S., Martin, M.A.S., Gokhman, D., Mishol, N., Eshel, T., Cheronet, O., Rohland, N., Mallick, S., Adamski, N., et al. (2020). The Genomic History of the Bronze Age Southern Levant. Cell *181*, 1146–1157.e11.

107. Feldman, M., Fernández-Domínguez, E., Reynolds, L., Baird, D., Pearson, J., Hershkovitz, I., May, H., Goring-Morris, N., Benz, M., Gresky,

J., et al. (2019). Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. Nat. Commun. *10*, 1218.

108. Fregel, R., Méndez, F.L., Bokbot, Y., Martín-Socas, D., Camalich-Massieu, M.D., Santana, J., Morales, J., Ávila-Arcos, M.C., Underhill, P.A., Shapiro, B., et al. (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc. Natl. Acad. Sci. USA *115*, 6774–6779.

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

110. Gokhman, D., Nissim-Rafinia, M., Agranat-Tamir, L., Housman, G., García-Pérez, R., Lizano, E., Cheronet, O., Mallick, S., Nieves-Colón, M.A., Li, H., et al. (2020). Differential DNA methylation of vocal and facial anatomy genes in modern humans. Nat. Commun. *11*, 1189.

111. González-Fortes, G., Jones, E.R., Lightfoot, E., Bonsall, C., Lazar, C., Grandal-d'Anglade, A., Garralda, M.D., Drak, L., Siska, V., Simalcsik, A., et al. (2017). Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. Curr. Biol. *27*, 1801–1810.e10.

112. González-Fortes, G., Tassi, F., Trucchi, E., Henneberger, K., Paijmans, J.L.A., Díez-Del-Molino, D., Schroeder, H., Susca, R.R., Barroso-Ruíz, C., Bermudez, F.J., et al. (2019). A western route of prehistoric human migration from Africa into the Iberian Peninsula. Proc. Biol. Sci. *286*, 20182288.

113. Kılınç, G.M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A.M., Bıçakçı, E., Baird, D., Dönertaş, H.M., Ghalichi, A., Yaka, R., et al. (2016). The Demographic Development of the First Farmers in Anatolia. Curr. Biol. *26*, 2659–2666. https://doi.org/10.1016/j.cub.2016.07.057.

114. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.

115. Martiniano, R., Cassidy, L.M., Ó'Maoldúin, R., McLaughlin, R., Silva, N.M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M.J., Serra, M., et al. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. PLoS Genet. *13*, e1006852.

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

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

118. Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., et al. (2018). The genetic prehistory of the Baltic Sea region. Nat. Commun. *9*, 442.

119. Olalde, I., Allentoft, M.E., Sánchez-Quinto, F., Santpere, G., Chiang, C.W.K., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J.A., Rasmussen, S., Quilez, J., et al. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature *507*, 225–228.

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

121. Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., García Borja, P., Salazar-García, D.C., Talamo, S., et al. (2015). A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. Mol. Biol. Evol. *32*, 3132–3142.

122. Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., et al. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. Science *358*, 655–658.

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

124. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

125. Seguin-Orlando, A., Korneliussen, T.S., Sikora, M., Malaspinas, A.-S., Manica, A., Moltke, I., Albrechtsen, A., Ko, A., Margaryan, A., Moiseyev, V., et al. (2014). Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science *346*, 1113–1118.

126. Shinde, V., Narasimhan, V.M., Rohland, N., Mallick, S., Mah, M., Lipson, M., Nakatsuka, N., Adamski, N., Broomandkhoshbacht, N., Ferry, M., et al. (2019). An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers. Cell *179*, 729–735.e10.

127. Valdiosera, C., Günther, T., Vera-Rodríguez, J.C., Ureña, I., Iriarte, E., Rodríguez-Varela, R., Simões, L.G., Martínez-Sánchez, R.M., Svensson, E.M., Malmström, H., et al. (2018). Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. Proc. Natl. Acad. Sci. USA *115*, 3428–3433.

128. van den Brink, E.C.M., Beeri, R., Kirzner, D., Bron, E., Cohen-Weinberger, A., Kamaisky, E., Gonen, T., Gershuny, L., Nagar, Y., Ben-Tor, D., et al. (2017). A Late Bronze Age II clay coffin from Tel Shaddud in the Central Jezreel Valley, Israel: context and historical implications. Levant *49*, 105–135.

129. Wang, C.-C., Reinhold, S., Kalmykov, A., Wissgott, A., Brandt, G., Jeong, C., Cheronet, O., Ferry, M., Harney, E., Keating, D., et al. (2019). Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. Nat. Commun. *10*, 590. https://doi.org/10.1038/s41467-018-08220-8.

130. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., Kelso, J., and Fu, Q. (2017). 40,000-Year-Old Individual from Asia Provides Insight into Early Population Structure in Eurasia. Curr. Biol. *27*, 3202–3208.e9.

131. van de Loosdrecht, M., Bouzouggar, A., Humphrey, L., Posth, C., Barton, N., Aximu-Petri, A., Nickel, B., Nagel, S., Talbi, E.H., El Hajraoui, M.A., et al. (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. Science *360*, 548–552.

132. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

133. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S., Davies, R., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. PubMed Abstract| Publisher Full Text.

134. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19.

135. 1000 Genomes Project Consortium; Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

136. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.

137. Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. Am. J. Hum. Genet. *103*, 338–348.

138. Hui, R., D'Atanasio, E., Cassidy, L.M., Scheib, C.L., and Kivisild, T. (2020). Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. Sci. Rep. *10*, 18542.

139. Gouy, M., Tannier, E., Comte, N., and Parsons, D.P. (2021). Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. In Multiple Sequence Alignment: Methods and Protocols, K. Katoh, ed. (Springer US), pp. 241–260.

140. Stern, A.J., Wilton, P.R., and Nielsen, R. (2019). An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. PLoS Genet. *15*, e1008384.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Four newly reported ancient individuals	This study	N/A
Chemicals, peptides, and recombinant proteins		
Proteinase K	Sigma Aldrich	Cat#3115887001
EDTA	Fisher Scientific	Cat#10628203
USER enzyme	NEB	Cat#M5505L
End repair enzyme mix	NEB	Cat#E6050L
T4 DNA ligase	Fisher Scientific	Cat#10723941
AccuPrime Pfx SuperMix	Sigma Aldrich	Cat#10043322
Deposited data		
Raw sequencing data from four newly reported ancient individuals	This study	ENA: PRJEB31781
Aligned sequencing reads from four newly reported ancient individuals	This study	ENA: PRJEB71330
Genotype data from four newly reported ancient individuals	This study	Zenodo: https://doi.org/10.5281/zenodo.10058388
Imputed SNP data from four newly reported ancient samples from Bahrain and 37 previously published ancient individuals	This study	Zenodo: https://doi.org/10.5281/zenodo.10058388
Software and algorithms		
AdapterRemoval v2	Schubert et al.[70]	N/A
bwa v0.7.17-r1188	Li and Durbin[71]	N/A
samtools v1.3.1	Li et al.[72]	N/A
sambamba v0.6.6	Tarasov et al.[73]	N/A
Ry_compute v0.4	Skoglund et al.[74]	N/A
READ	Kuhn et al.[75]	N/A
PMDtools v.0.6	Skoglund et al.[76]	N/A
ANGSD v0.940	Korneliussen et al.[77]	N/A
contamMix v. 1.0–10	Fu et al.[78]	N/A
CrossMapp v0.6.4	Zhao et al.[79]	N/A
bamUtil v1.0.15	Jun et al.[80]	N/A
sequenceTools v1.5.3.2	https://github.com/stschiff/sequenceTools	N/A
EIGENSOFT package v7.2.1	Price et al.[29]; Patterson et al.[28]	N/A
DyStruct v1.1.0	Joseph and Pe'er[81]	N/A
AdmixTools 2 R package	Maier et al.[37]	N/A
AdmixTools v7.0.2	Patterson et al.[31]	N/A
qpbrute v0.3	Ní Leathlobhair et al.[82]; Liu et al.[83]	N/A
ChromoPainter/Finestructure v4.1.0	Lawson et al.[84]	N/A
Beagle v5.2	Browning et al.[85]	N/A
GraphTyper_v2.6.2	Eggertsson et al.[86]	N/A
GLIMPSE v1.1.0	Rubinacci et al.[87]	N/A
RELATE v1.1.7	Speidel et al.[88]	N/A
PLINK v1.9	Chang et al.[89]	N/A
pathPhynder v1.a	Martiniano et al.[40]	N/A
mtDNA-Server v2	Weissensteiner et al.[90]	N/A
HIrisPlex-S	Breslin et al.[91]	N/A
DATES v4010	Chintalapati et al.[92]	N/A
selscan v1.2.0	Szpiech and Hernandez[93]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to Rui Martiniano (r.martiniano@ljmu.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw and processed sequence reads are made available at the European Nucleotide Archive (ENA) accessions ENA: PRJEB31781 and ENA: PRJEB71330, respectively.
- Pseudo-haploid genotypes for the newly reported ancient individuals and imputed variants are made available at Zenodo: https://doi.org/10.5281/zenodo.10058388.
- All code used for analysis has been previously published and is publicly available.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

A description of the archaeological context of the ancient individuals analyzed in this study can be found at Zenodo: https://doi.org/10.5281/zenodo.10058388.

METHOD DETAILS

Archaeological sample processing, DNA extraction, library preparation and sequencing

We sampled 25 petrous bones and teeth from the Bahrain National Museum and the Qal'at al-Bahrain Museum belonging to the Dilmun and Tylos periods individuals from 5 archaeological sites in Bahrain: Jari al-Shaikh (n = 2), Shakhourah (n = 3), Madinat Hamad (n = 6), Saar (n = 6) and Abu Saiba (n = 8) (Table S1). Sample processing was carried out in dedicated ancient DNA facilities at Trinity College Dublin, Ireland, where all the precautions for ancient DNA processing were followed as previously described.[94] We exposed all skeletal material to UV light for 15 min on either side to remove surface contaminants, and we cleaned the outer layer of bone with a drill before extraction. We targeted the densest section of the otic capsule region of the petrous temporal bone, or in case of teeth, the root, which we converted to powder using a homogenizer (Retsch Mixer Mill). Where possible, sample homogenization was done progressively by shaking the bone multiple times for approximately 10–15 s, keeping the resulting powder between each session. This process allows the selection of the densest part of the petrous bone. We extracted approximately ~130 mg of bone powder per sample using a silica column method[95,96] with 2 initial washing steps of 0.5M EDTA solution (EDTA1). We performed a second extraction for four samples selected from the results of the initial screening based on higher endogenous content and deamination was present. This time, we extracted both the second EDTA wash (EDTA2) and a new subset of the same bone powder treated with an initial washing step by 0.5% bleach solution (BEX1).[97,98] The samples that were sequenced to high coverage were extracted as follows: MH1_LT, MH3_LT and AS_EMT (BEX1) and MH2_LT (EDTA2).

We performed an initial screening of each sample by constructing a double-stranded DNA NGS library using the method outlined in,[99] with modifications as in.[100] Every library was first screened on a MiSeq Illumina platform (50 bp SE) at Trinseq (Ireland). DNA extracts selected for high-coverage sequencing were incubated with Uracil-DNA-glycosylase (UDG) enzyme (volume of 5 μL–16.50 μL of extract) for 3 h at 37°C to repair postmortem molecular damage prior to a second library construction with the former method. Based on initial screening results, we prioritised the four most promising libraries for additional sequencing on a HiSeq X platform at the Wellcome Sanger Institute. The list of libraries and sample IDs can be found in Table S2.

Radiocarbon dating

Radiocarbon dating was performed at the Oxford Radiocarbon Accelerator Unit (ORAU) and resulted in the following dates: MH1_LT (BAH_MTF) 1565 ± 20 BP, 432–561 cal. AD (OxA-38603); MH3_LT (BAH_N) 1456 ± 22 BP, 577–647 cal. AD (OxA-38708). It was not possible to obtain radiocarbon dates for MH2_LT (BAH_MHM) and AS_EMT (BAH_ASV) due to insufficient collagen yield in the bone samples provided for analysis.

Sequence read processing and alignment

We trimmed adapter sequences from high-throughput sequencing reads using AdapterRemoval v2[70] with the parameters (–interleaved-input –minlength 30 –trimns –trimqualities –minquality 2 –collapse), and aligned these to the human reference genome (GRCh38, obtained from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/GRCh38_reference_genome/GRCh38_full_analysis_set_plus_decoy_hla.fa) using bwa aln,[71] with the parameters -n 0.01 -o 2 and selected reads with mapping quality ≥25 using samtools,[72] as previously recommended[101,102] and removed sequence read duplicates using sambamba markdup.[73] We used the same steps to process published shotgun WGS ancient DNA from the Levant[13,15,38] and Iran.[10]

Sex determination and relatedness analysis

We determined the sex of the four ancient samples from Bahrain using a previously published method[74] (Table 1). We performed a kinship analysis on the four ancient samples from Bahrain using READ (https://bitbucket.org/tguenther/read/)[75] with default parameters on approximately 1,240k SNPs of the Allen Ancient DNA Resource, genotyped as described below, which did not reveal any close kinship relationships between them.

Authenticity of aDNA sequences and contamination estimates

To evaluate aDNA sequence authenticity, we estimated postmortem deamination patterns in aligned sequence reads using PMDtools v. 0.6 (https://github.com/pontussk/PMDtools).[76] X chromosome contamination in males was estimated using ANGSD[77,103] and mitochondrial DNA contamination for all four samples was estimated using contamMix v. 1.0–10[78] using a panel of 311 present-day mtDNA sequences.[104]

Population genetics analyses

Variant calling

For population genetics analyses, we used the following datasets: the Allen Ancient DNA Resource (AADR; Allen Ancient DNA Resource version 44.3), the Lazaridis et al. (2022) dataset,[42] present-day Middle Easterners,[20,32] present-day worldwide populations from the Human Genome Diversity Project (HGDP)[39] and from the Simons Genome Diversity Project[105] and previously published ancient individuals.[10–17,30,33–35,38,42,43,64,100,106–131]

Where necessary, we converted the coordinates of the published data to the human genome assembly GRCh38 using CrossMapp v0.6.4.[79]

For the Bahrain_Tylos samples, we trimmed 2bp at the end of each read using bamUtil (https://github.com/statgen/bamUtil),[80] while for the published Levantine and Iranian WGS we trimmed 3bp. We subsequently called variants at each position of the 1240k SNPs included in AADR using samtools mpileup, disabling base quality score recalibration and imposing a minimum base quality filter of q20. We generated pseudo-haploid genotypes by randomly sampling one allele at each SNP site using pileupCaller (https://github.com/stschiff/sequenceTools).

We genotyped present-day Middle Easterners[20,32] using bcftools v1.9[132,133] with command "bcftools mpileup -T -q30 -Q30 | bcftools call -c" on the 1240k positions and merged the datasets using the mergeit program from the EIGENSOFT package v7.2.1[28,29] with options docheck: YES and strandcheck: YES filtering sex-linked and triallelilic SNPs and sites that were outside the genomic accessibility mask,[39] leaving 1,099,436 SNPs in the final dataset.

Principal component analysis and model-based clustering

We ran principal components analysis (PCA) using smartpca v18140 from the EIGENSOFT package with parameters numoutlieriter: 0, lsqproject: YES, autoshrink: YES, projecting ancient samples on the principal components estimated with present-day Eurasian populations.

We also ran DyStruct (https://github.com/tyjo/dystruct)[81] from K = 7 to K = 12 in a dataset of 2,073 samples and 85,043 transversions with default parameters across eleven time points binned around the following times (generations ago): 450, 350, 250, 180, 150, 120, 90, 70, 50, 25, and present-day.

F-statistics, qpWave, qpAdm and qpGraph

We estimated f3-and f4-statistics and qpAdm using the AdmixTools 2 R package.[37] For modeling the ancestry of Bahrain Tylos samples with qpAdm using ancient populations as sources, we used the following right populations: Mbuti.DG, CHG, EHG3, ISR_Natufian_EpiP, MAR_Taforalt_EpiP, RUS_AfontovaGora3, SRB_Iron_Gates_HG, TUR_Pınarbaşı_EpiP and WHG4. For qpWave tests using present-day populations, we used qpWave v1520 from the AdmixTools v7.0.2 package (https://github.com/DReichLab/AdmixTools)[31] with option allsnps:YES to test if ancient Bahrain formed a clade with any other ancient or modern populations. We used qpAdm v1520 to estimate ancestry proportions in present-day populations, using the following reference population samples: Mbuti, Ust'-Ishim, Kostenki14, MA1, China_Tianyuan, WHG, WSHG, Natufian, Levant_N, Morocco_Iberomaurusian, Anatolia_N, Iran_N, CHG, EHG, and Steppe_Eneolithic. We used qpDstat v980 also from the ADMIXTOOLS package to estimate D- and f-statistics to test for admixture and differentiation in ancient samples. We generated admixture graphs using qpGraph v7580 with parameters lsqmode:YES and inbreed:NO. In Figure 3E, we present a semi-automatically fitted graph. We started with a base graph representing published ancient populations, drawing from prior knowledge.[11,13] This initial graph had an outlier Z score = 2. Subsequently, we employed qpBrute[82,83] to iteratively fit HajjiFiruz_IA, AS_EMT, MH2_LT, MH1_LT, and MH3_LT. We selected the graph with the fewest outliers and the lowest f-statistic Z score at each step. For the graph in Figure 3E, the worst f-statistic derived from an excess affinity between MH1_LT and Bichon compared to MH2_LT with a Z score of 3.3. Details about the dataset used in modeling analyses is described on Table S9.

Chromopainter analysis

We ran the ChromoPainter/Finestructure v4.1.0 inference pipeline[84] with default settings outputting a co-ancestry matrix with values depicting haplotype segments shared between individuals. We ran the pipeline on a merged dataset composed of the 41 imputed ancient samples and 168 Modern samples from the Middle East, Europe and South Asia.[11,20] We used 448K SNPs which have a minor allele frequency >5% in the reference panel and excluded sites with missingness more than 5%, and jointly phased the merged dataset using the 1000G high coverage reference panel[134] using Beagle v5.2.[85]

Imputation

As Middle Eastern populations are not found in the commonly used 1000 Genomes Imputation panel,[135] and some variants of interest are found in the Middle East but are very rare or even absent in the 1000G panel,[20] in this study we created a reference panel which includes samples from the HGDP whole-genome sequencing study[39] and present-day Middle Easterners.[20] The HGDP has several regionally-relevant populations from the Levant and South Asia, while Almarri et al. (2021) sampled populations from Arabia, the Levant and Iraq, and additionally included samples with present-day Iranian and South Asian ancestry. A total of 1064 samples were used to construct the reference panel, 929 from the HGDP and 135 samples from Almarri et al. 2021.

We first jointly called SNVs identified from both studies, excluding singletons, doubletons, and multiallelic variants, using GraphTyper_v2.6.2[86] and set variants with GQ < 20 to missing. We limited the analysis to the strict genome accessibility mask,[39] which covers ~73% of the genome. As a reference panel cannot have missing data, we used Beagle 4.1[136] with the option gtgl, which uses the genotype likelihood at sites set as missing, to identify the most likely genotype using the variation within the panel. We subsequently used a stringent filter by only including high quality sites (AR2 >0.97). The resulting reference panel comprised a total of 18.2 million SNVs.

We then used Beagle 5.1[137] to phase the reference panel using default parameters. As the samples in Almarri et al., 2021 were physically-phased using linked-read sequencing, we were able to compare the samples (statistically-phased in our reference panel, with physically-phased from Almarri et al., 2021) to evaluate phasing accuracy of the panel. Extracting variants on chromosome 1 (840,284 SNVs), we used vcftools v0.1.16 with option –diff-switch-error to calculate the switch error at heterozygous sites per sample. We find a relatively low average switch error across all samples (mean 0.44%, stdev 0.1%), although we note the switch error rate will decrease by removing singletons and doubletons.

We subsequently imputed diploid genotypes of ancient samples using GLIMPSE v1.1.0,[87] using default parameters with the exception of increasing the number of main iteration steps of the algorithm to 15. We first called genotype likelihoods for each sample at sites found in the reference panel using the command "*bcftools mpileup -f {reference_genome} -I -E -a 'FORMAT/DP' -T {variants_in_reference_panel} {sample}.bam -Ou | bcftools call -Aim -C alleles*". To assess imputation accuracy, while taking into account our mapping and filtering process, we downsampled chromosome 21 of a published regionally-relevant early Neolithic genome (WC1)[10] sampled from the Zagros region of present-day Iran from ~10x to 0.1x, 0.25x, 0.5x, 1x and 2x using samtools view -s option. Genotype consistency was measured using the GLIMPSE_concordance tool by comparing the imputed variants with the variants called using all the read information (~10x coverage) using only confidently called sites that have a minimum depth of 8 and minimum genotype probability of 0.999, a total of 85,848 SNVs. We find that for variants in the reference panel with minor allele frequency higher than 5% in the reference panel, even at 0.1x depth variants can be imputed with relatively high accuracy (r2 = 0.76), which increases with higher depth (0.25x - r2 = 0.89, 0.5x - r2 = 0.95, 1x - r2 = 0.97, 2x - r2 = 0.98). Subsequently we imputed the ancient Bahraini samples (4 samples) using the steps above, in addition to 37 published ancient Levantine and Iranian whole-genome sequences.[10,13,38] The lowest coverage of the 41 imputed ancient samples was 0.24x.

To further improve the imputation quality, we followed a two-step imputation approach as previously described.[138] We set to missing any variant that had an imputed GP < 0.99, and then re-imputed these missing variants using the larger high-coverage 1000G panel as a ref. 134 using Beagle v5.1.[137] To further check the quality of the imputed genotypes, we compared principal component analysis (PCA) and ADMIXTURE results between the imputed diploid and pseudohaploid calls across common variants (MAF >5%). For both analyses we merged our ancient samples with the HGDP. In the PCA, we projected both the imputed and pseudohaploid calls on principal components calculated using modern samples (Figure S7A). For ADMIXTURE, we ran the merged dataset using different Ks, from 3 to 8 (Figure S7B). We find similar results between the pseudohaploid calls and the imputed calls, suggesting that the imputation performed is of relatively high quality.

For rs5030868, we created a special imputation panel solely of Middle Eastern samples (9% frequency) as rs5030868 is very rare in the 1000G panel (0.08%). We used GLIMPSE as stated above to impute common SNPs (>5%) covering a 3Mb region on the X chromosome surrounding the variant (5151 SNPs). We created a phylogeny of the locus using SNPs ±50Kb around rs5030868 with Seaview v5.0.5 (PhyML - HKY85 model, default conditions).[139]

Relate

To estimate the allele frequency trajectory and selection coefficient of rs5030868 we ran RELATE v1.1.7[88] to create a sequence of genealogies across 3Mb surrounding the variant as stated above using 59 female samples from the Almarri et al., 2021 Middle East dataset setting -m 1.11e−8 and -N 20,000. We subsequently extracted the EmiratiC individuals from the tree and estimated their population size history using EstimatePopulationSize.sh and then ran 100 samples using SampleBranchLengths.sh to sample branch lengths of rs5030868 from the posterior using a mutation rate of 1.11e−8 and supplying the.coal file from the previous step. We then ran CLUES[140] (inference.py script) with the option –coal to account for population size history.

Extended haplotype homozygosity

We conducted an EHH analysis on the region surrounding SNP rs5030868 using selscan v1.2.0[93] with parameters –keep-low-freq and –ehh-win 5000.

Runs of homozygosity

We identified ROHs using PLINK v1.9[89] on the diploid imputed ancient genomes and present-day samples from the WGS HGDP[39] and Arabians and Levantines[20] using 652K common SNPs that were linkage-disequilibrium pruned (–indep-pairwise 50 5 0.9), and filtered for allele frequency (MAF >5%) and missingness (<5%). We also ran hapROH v0.64 to identify large ROHs (>4cm) and estimate parental relatedness.[52]

Uniparental lineage determination

We used pathPhynder[40] for investigating Y chromosome lineages in the two male ancient samples from Bahrain and mtDNA-Server[90] for determining mitochondrial haplogroups.

Phenotypic traits

We used the diploid imputed genotypes to explore phenotypic traits (eye color, hair color and skin color) predicted by the HIrisPlex-S software ([91] https://hirisplex.erasmusmc.nl/). Out of the 41 variants used by the software, we provided 38 variants and set the remaining 3 as "NA". Two of these variants are very rare in frequency, while one was ambiguous when flipping and correcting for strand. We checked the quality of the imputation for the 38 variants, and found that the imputation INFO score generated by GLIMPSE for these variants was high (average INFO score 0.98, minimum 0.96), suggesting they were imputed with relatively high accuracy.

Time of admixture

We used DATES v4010 (Distribution of Ancestry Tracts of Evolutionary Signals)[92] to estimate the time of admixture in the Tylos samples. We set binsize: 0.001, maxdis: 1.0, runmode: 1 and used as source populations the ancient Bronze/Iron Age Levantines and Iranians: LBN_Canaanite, LBN_IA, ISR_Canaanite_MLBA, JOR_LBA, SYR_Ebla_EMBA, IRN_DinkhaTepe_BIA_A, IRN_DinkhaTepe_BIA_B, IRN_Hasanlu_IA, IRN_Shahr_I_Sokhta_BA.