










<https://doi.org/10.1038/s42003-022-03508-4>

OPEN

## Bioarchaeological evidence of one of the earliest Islamic burials in the Levant

Megha Srigyan<sup>1,12</sup>, Héctor Bolívar <sup>2,3</sup>, Irene Ureña <sup>2</sup>, Jonathan Santana <sup>4</sup>, Andrew Petersen <sup>5</sup>, Eneko Iriarte <sup>6</sup>, Emrah Kirdök<sup>7</sup>, Nora Bergfeldt<sup>2</sup>, Alice Mora <sup>8</sup>, Mattias Jakobsson <sup>1</sup>, Khaled Abdo<sup>9</sup>, Frank Braemer<sup>10</sup>, Colin Smith<sup>6,8</sup>, Juan José Ibañez<sup>11</sup>, Anders Götherström<sup>2</sup>, Torsten Günther <sup>1,13</sup>✉ & Cristina Valdiosera <sup>6,8,13</sup>✉

The Middle East plays a central role in human history harbouring a vast diversity of ethnic, cultural and religious groups. However, much remains to be understood about past and present genomic diversity in this region. Here we present a multidisciplinary bioarchaeological analysis of two individuals dated to the late 7th and early 8th centuries, the Umayyad Era, from Tell Qarassa, an open-air site in modern-day Syria. Radiocarbon dates and burial type are consistent with one of the earliest Islamic Arab burials in the Levant. Interestingly, we found genomic similarity to a genotyped group of modern-day Bedouins and Saudi rather than to most neighbouring Levantine groups. This study represents the genomic analysis of a secondary use site with characteristics consistent with an early Islamic burial in the Levant. We discuss our findings and possible historic scenarios in the light of forces such as genetic drift and their possible interaction with religious and cultural processes (including diet and subsistence practices).

<sup>1</sup>Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden. <sup>2</sup>Centre for Palaeogenetics, 10691 Stockholm, Sweden. <sup>3</sup>Instituto del Patrimonio Cultural de España, 28040 Madrid, Spain. <sup>4</sup>Department of Historical Sciences, Universidad de Las Palmas de Gran Canaria, Las Palmas de G.C. E35001, Spain. <sup>5</sup>University of Wales Trinity Saint David, Lampeter, UK. <sup>6</sup>Laboratorio de Evolución Humana, Departamento de Historia, Geografía y Comunicación, Universidad de Burgos, 09001 Burgos, Spain. <sup>7</sup>Department of Biotechnology, Mersin University, 33343 Mersin, Turkey. <sup>8</sup>Dept. Archaeology and History, La Trobe University, Melbourne, VIC 3086, Australia. <sup>9</sup>General Directorate of Antiquities and Museums, Damascus, Syrian Arab Republic. <sup>10</sup>Université Côte d'Azur, CNRS, Culture et Environment, Préhistoire Antiquité Moyen Age, Nice, France. <sup>11</sup>Archaeology of Social Dynamics, Milà i Fontanals Institution, Spanish National Research Council (CSIC), Barcelona, Spain. <sup>12</sup>Present address: Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA. <sup>13</sup>These authors contributed equally: Torsten Günther, Cristina Valdiosera. ✉email: [torsten.guenther@ebc.uu.se](mailto:torsten.guenther@ebc.uu.se); [cevaldiosera@ubu.es](mailto:cevaldiosera@ubu.es)

The Middle East has an unparalleled place in human history. From the out-of-Africa movements of modern humans and admixture with Neandertals, to the spread of agriculture and emergence of civilisations, it has been at the crossroads of genetic as well as cultural history for millennia. While archaeological and historical sources provide instrumental insights into socio-cultural aspects, ancient DNA (aDNA) enables the recovery of past genetic information, filling crucial gaps in our understanding of population history. A number of aDNA studies from different time periods in the Middle East have provided a general overview of the genetic history in this region. These include descriptions of the earliest local farming groups from the Neolithic<sup>1</sup> and their expansions into Europe<sup>2,3</sup> as well as genetic differentiation among contemporary Neolithic groups<sup>4–6</sup>. In the later Chalcolithic period, evidence of distinctive cultural practices and associated population movements highlight the dynamic history of the region, especially in the Southern Levant<sup>7,8</sup>. Further, genomic studies from the Bronze to Iron Ages in the Levant also report admixture and population movements, suggesting some degree of continuity with modern populations<sup>6,9–11</sup>. On a more recent timescale, studies from the medieval period<sup>12</sup> and modern populations<sup>13,14</sup> describe genetic structure and the role played by culture and religion in the formation of these structures. Notably, a study on medieval individuals from Lebanon previously identified as Crusaders<sup>12</sup> demonstrated their ancestry to be either European or local as well as an admixture between Europeans and Near-Easterners. These signals cannot be detected in modern Lebanese groups, suggesting that they were only transient. This provided evidence of genetic signatures of historical religious events such as The Crusades which saw considerable movement from Europe to the Middle East, admixture with local populations and eventually, their dilution over time. Genetic analyses of modern Lebanese populations<sup>13</sup> suggest that population movements linked to the spread of religions like Islam in the past millennium led to stratification in the Levant. Further south, data from modern Yemenis combined with other Middle Eastern groups found little correlation between genetic structure and geography<sup>14</sup>. It is thus clear that the present distribution of genetic diversity in the Middle East is the result of convoluted processes with culture as an additional level of complexity.

The Late Antiquity period, roughly defined as the time between the third–eighth centuries, was a time of cultural and religious upheaval in the Middle East associated with the emerging Arab Islamic empire. Byzantine Syria–Palestine represents an instance of a region that was conquered by Islamic Arabs in the first half of the seventh century AD (630s). This area became the political centre of the empire with the founding of the Umayyad caliphate in Damascus in 661<sup>15</sup>. However, the Arabization and Islamisation of the area did not fully take place until the last decade of the seventh century led by ‘Abd al-Malik<sup>16</sup>. As such, the Aramaean, Byzantine and Christian legacy interacted with the new Arab Islamic rule and cultural values for decades. The collapse of the caliphate (750) and transfer of the political centre to Iraq caused political marginalisation and economic decline in Syria–Palestine<sup>17</sup> (see also Supplementary Note S1 – Early Islamic Southern Syria). Thus within the span of the Umayyad caliphate, this region likely mirrored some of the many political and religious shifts occurring throughout the Middle East<sup>17</sup>.

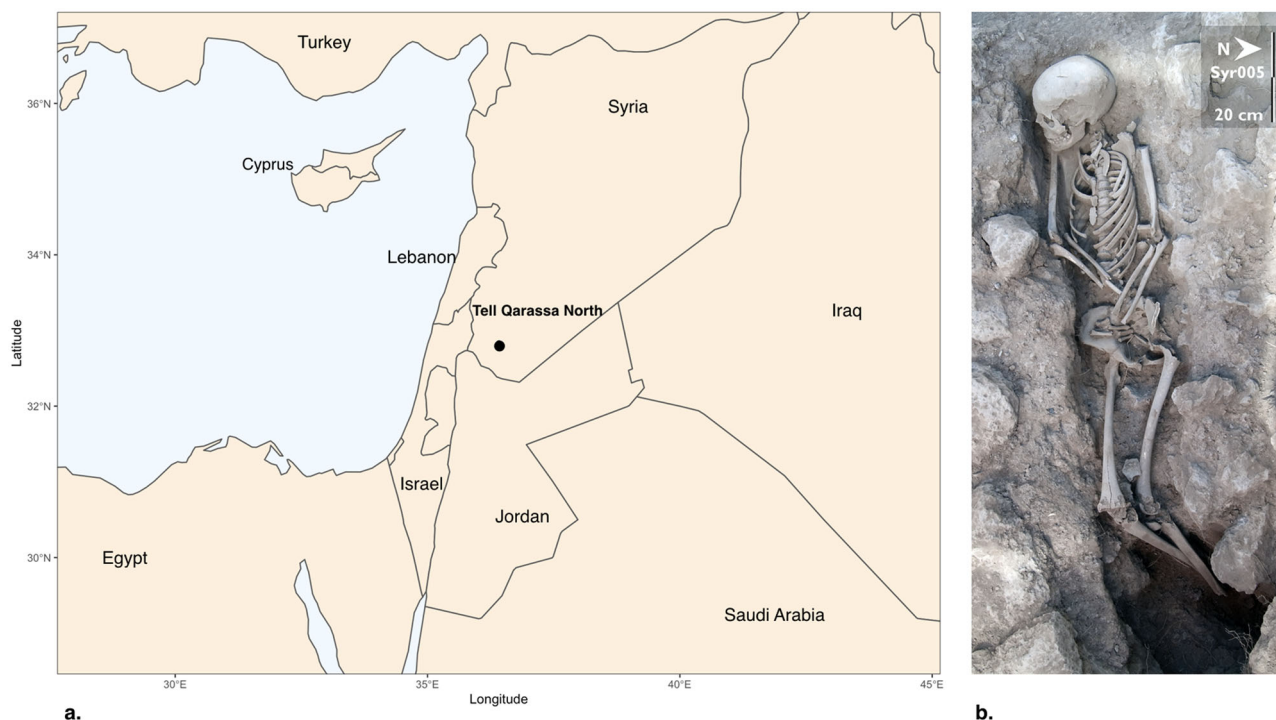
Ancient DNA analysis is a powerful tool to provide a genomic snapshot of this dynamic period, giving insight into past demographic processes of a currently conflicted and inaccessible territory of the Levant. However, while much of the focus of Near-Eastern Archaeology has been on funerary remains, few Islamic burials have been investigated as such studies might be considered as harming or disturbing the dead. We present an archaeogenomic analysis of two Umayyad Era individuals found

at a prehistoric site in modern-day Syria, with no connection to an Islamic cemetery, but buried with indications of Islamic funerary rituals (see Supplementary Note S1 – History of the excavation and subsequent bioarchaeological analysis – for the details of the circumstances of the excavation). We find genomic similarities to a group of modern-day Bedouins and Saudi rather than to most neighbouring Levantine groups. The remains are curated by the General Directorate of Antiquities and Museums (DGAM) of the Arab Republic of Syria and their genomic study is an invaluable resource to understand their ancestry and the history of this region.

## Results and discussion

**Two historic burials on top of a Neolithic site.** In this study, we perform genomic analyses of two buried individuals excavated at Tell Qarassa North, a Neolithic site in the Village of Qarassa in Syria (Fig. 1a). While Tell Qarassa North is usually known as a prehistoric site<sup>18,19</sup> (see also Supplementary Note S1), the two individuals analysed here were found in two narrow graves on surface levels of the site and directly radiocarbon dated to the Umayyad Era (seventh to eighth centuries) (Table 1). No cultural artifacts were associated with the human remains and no evidence of an Umayyad Era cemetery was documented at the site. The two Umayyad Era burials (UEB) were located very close to each other. While in the Neolithic burials the bodies are placed in a flexed position, these bodies were placed in a decubitus position, oriented east-west, with the head at the west, facing south inside pits that were intrusive in the Neolithic levels (Fig. 1b). The distribution of the skeletal elements suggests that both bodies were wrapped before burial<sup>20</sup>. Individual syr005 was a 14–15-year-old male (Fig. 1b) and individual syr013 was a female of about 15–21 years at the time of death (Table 1). Age-at-death estimation was based on the following criteria: pattern of dental eruption, synostosis of epiphyses on the long bone and closure of the sternal ends of the clavicles<sup>21–23</sup>. Together with the radiocarbon dates, the wrapping, the position and orientation of the bodies facing Mecca are concordant with Muslim funerary rituals following Early Islamic burials<sup>24</sup>. However, these individuals were not buried in a traditional Muslim cemetery. This may be explained due to special circumstances of death or cultural identity: nomadic populations, pilgrims, deviant burials or plague victims. The requirement of a Muslim burial to take place within 24 h after death might have made some compromises necessary. It is known that one of the defining features of Muslim burials is that of only one person per grave, which implies that husbands and wives are not buried together, and collective family tombs are forbidden. Nonetheless, occasionally and in extreme circumstances this can be relaxed for victims of plague or warfare (Supplementary Note S1 – Muslim Burials). Also, the close proximity of radiocarbon dates for syr005 ( $1294 \pm 18$  Cal BP) and syr013 ( $1302 \pm 15$  Cal BP) suggest that both individuals died at a similar time.

**Genome sequencing and exploratory analysis.** To investigate the genetic identity of these two individuals, their association to past, contemporaneous and present-day Middle Eastern populations as well as to shed light on past genetic variation of Syria, a conflicted region that remains currently poorly studied, we shotgun-sequenced two petrous bones (syr005 and syr013) at a depth coverage of 0.16× and 6.15×, respectively (Table 1). Sequence data from both individuals showed characteristic patterns of post-mortem damage and fragmentation expected from endogenous ancient DNA (aDNA) molecules<sup>25</sup> (Fig. S1). We used four different methods to estimate contamination at the mitochondrial<sup>25,26</sup>, autosomal<sup>27</sup> and X-chromosome<sup>28</sup> levels and



**Fig. 1** Site location and skeletal remains included in the analysis. **a** Map of the Levant indicating the location of Tell Qarassa in South Syria. **b** Skeletal remains of syr005 during the excavations at Tell Qarassa (Photo by Jonathan Santana).

**Table 1** Information about the two sequenced samples.

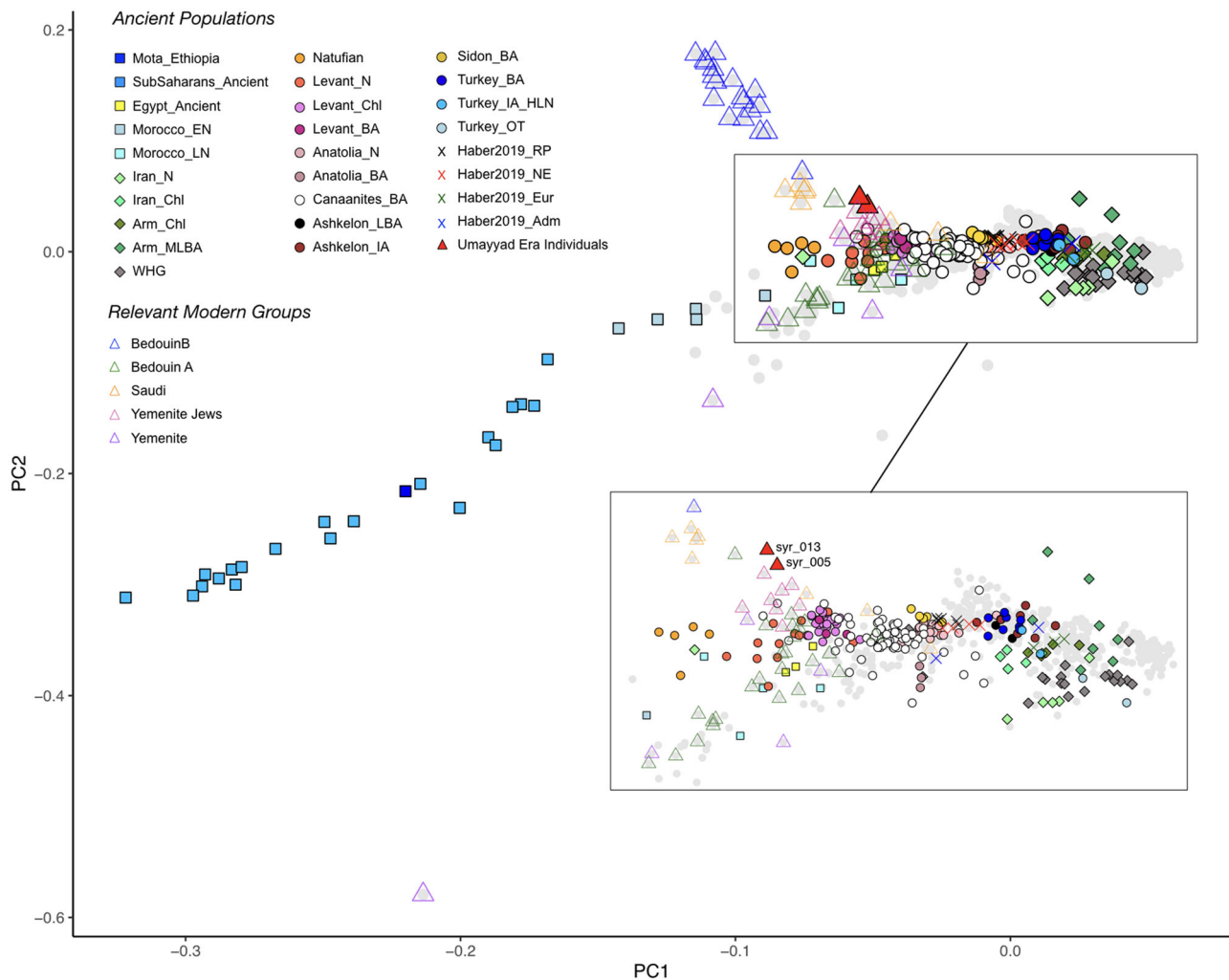
Individual	Genome coverage	mt coverage	mt haplogroup	Y haplogroup	Biological sex	Age (C14 cal yBP, 2 $\sigma$ )	Autosomal contamination estimate
syr005	0.159799	23.5158	J2a2a1a1	J	XY	666–768 cal AD	0.01475
syr013	6.15177	522.349	R0a2	-	XX	665–766 cal AD	0.0488

all four methods confirmed low levels of contamination (<5%, Table S2). Two biological sex inference methods<sup>29,30</sup> identified syr005 to be a male and syr013 a female. Individuals syr005 and syr013 were determined to carry mitochondrial haplogroups J2a2a1a1 and R0a2, respectively. Both haplogroups are common in the Arabian Peninsula, Near East and parts of Africa<sup>31,32</sup> in concordance with the broad geographical location of the samples. In addition, the Y chromosome of syr005 was determined as haplogroup J, which is the most common haplogroup across the Middle East<sup>33</sup> (Table S3).

To further explore general patterns of genetic affinity to modern populations, we performed a principal component analysis (PCA), projecting the two newly sequenced Umayyad Era individuals along with 262 published ancient individuals from the Near East, Western Europe, North and sub-Saharan Africa on a broad set of modern Middle Eastern, Arabian Peninsula, European and North African groups. The two UEB individuals fell between modern genetic variation in the Middle East and Arabian Peninsula and are shifted towards the latter (Fig. S2). Further, to obtain a better understanding of the regional variation, we conducted a second PCA, limited to 37 modern groups from the Middle East, Arabian Peninsula and Caucasus (Fig. 2). While the UEB individuals did not cluster with any published ancient Levantine individuals, the closest ancient groups were Bronze Age Canaanites<sup>8</sup> and groups from Neolithic, Chalcolithic and Bronze Age Levant<sup>1,7</sup>. Among modern human populations, the UEB individuals are positioned between groups known to originate from/inhabiting the Arabian Peninsula, i.e.

Saudi, Yemenite Jews and Bedouin A and B (Fig. 2). Hence, overall, individuals syr005 and syr013 fall between the two Bedouin groups, and show a clear genetic differentiation from other modern-day Levantines.

To gain insight into the genetic composition of ancient and modern populations, we ran an unsupervised *ADMIXTURE*<sup>34</sup> analysis (Fig. 3) with a set of 1321 individuals from Europe and the Middle East (73 modern and 28 ancient populations in total). For  $K = 2, 4, 5$  and  $6$ , all iterations with different random seeds converged to consistent results (Fig. S3). Therefore, we consider  $K = 6$  as a compromise between the resolution and robustness of the results. At  $K = 4$ , a new component appeared mostly in prehistoric Levantine groups, i.e. Natufians, Neolithic Levant, Neolithic Anatolia, Chalcolithic Levant and in the later Bronze Age Canaanites<sup>8</sup> as well as in the UEB individuals. This was also seen in moderate proportions across modern groups from the Arabian Peninsula/Middle East, and in lower proportions in some European groups. At  $K = 5$ , this component split up in two, one appeared exclusive to ancient Levantine groups (as in  $K = 4$ ) but only in low amounts in the UEB individuals and was absent in any modern population from the Middle East/Arabian Peninsula which showed high values of the second component maximised in Bedouin B. At  $K = 6$ , another component emerged in Bedouin B and was present in high to moderate proportions in groups like Saudi/Bedouin A/Yemenite Jews and Middle Eastern populations, respectively. The UEB individuals also harbour high proportions of this component along with small amounts of the Neolithic Levant component seen at  $K = 5$ .



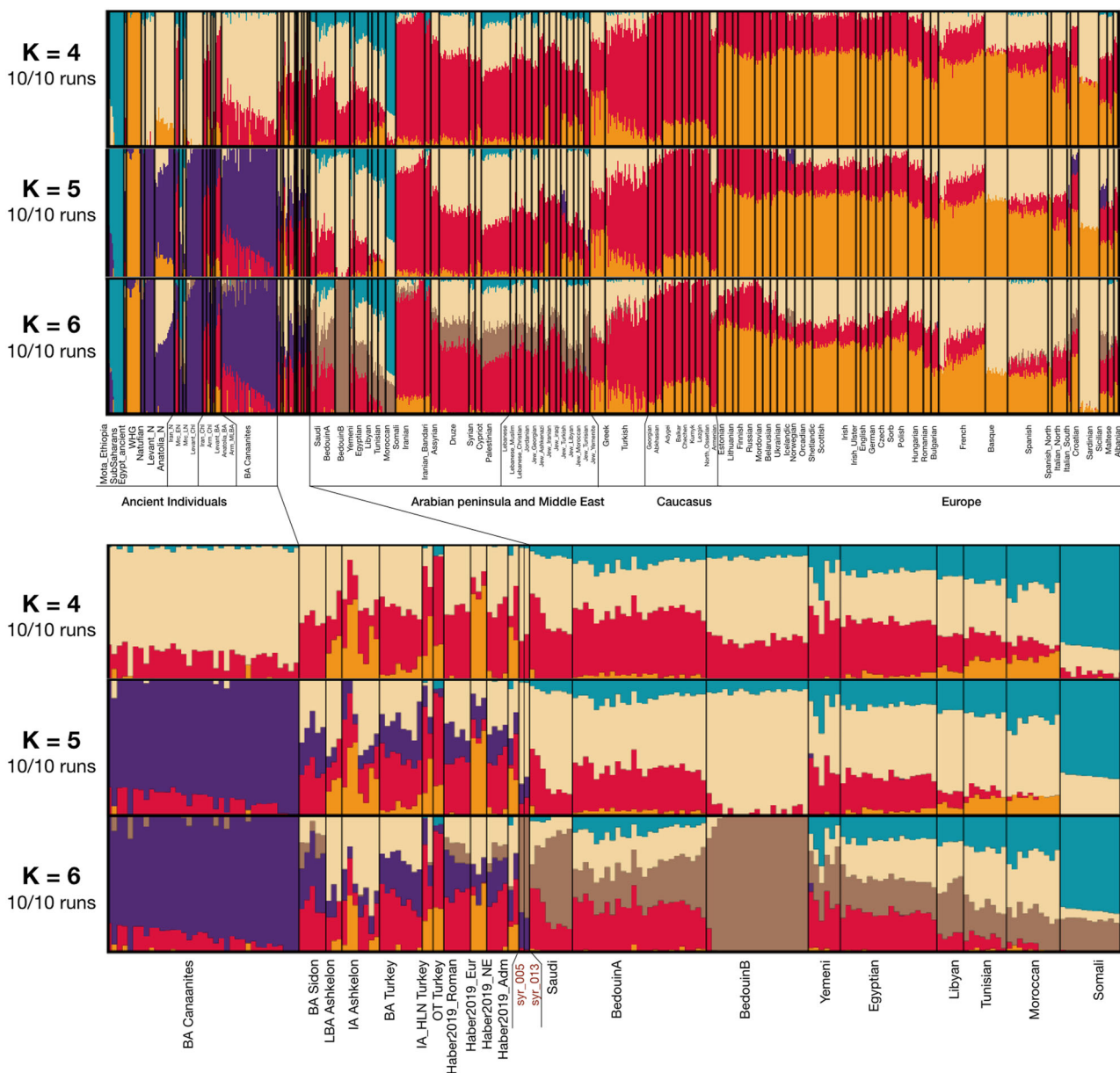
**Fig. 2 Principal component analysis of modern and ancient Middle Eastern populations.** PCA with ancient Levantine populations projected on modern Middle Eastern genetic variation (grey), with some relevant groups indicated as open triangles.

We conducted outgroup  $f_3$  statistics<sup>35</sup> to increase resolution on population affinities between the two UEB samples and modern Bedouins, Saudi and Yemenite Jews, as indicated by PCA and *ADMIXTURE* analysis. We observed high shared genetic drift with Bedouin B and Saudi (Fig. 4), but given the similar values of the statistic ( $f_3(X, \text{UEB individuals}; \text{Mbuti})$ , Table S4), establishing which modern group has the highest affinity required further analyses. To fully harness the medium coverage data for syr013, we called diploid genotypes using a genotyper designed for ancient DNA (*snpAD*<sup>36</sup>). Compared to the pseudohaploid data available for the other ancient individuals, these diploid genotypes allowed for a more fine scale analysis of the relationship between syr013 and modern populations. We used Beagle 4.1<sup>37</sup> to analyze sharing of tracts of identity by descent (IBD) between syr013 and present-day populations. Consistent with other results, the highest number and total length of IBD tracts was shared with multiple Bedouin B individuals and a single Saudi individual in the Human Origins 2.0 dataset (Fig. 5a). This confirms the connection between the UEB individuals and nomadic Levantines as well as to the Arabian peninsula.

We then tested different scenarios to investigate the most likely population affinities between the sequenced UEB samples and modern regional populations, performing D statistics of the form  $D(\text{UEB}, A; X, \text{Mbuti})$  testing whether 'X' is an outgroup with respect to the group constituted by UEB individuals and 'A' or if

there is an excess of allele sharing between 'A' and 'X' or the UEB individuals and 'X'. When we tested Bedouin B and Saudi as candidates for groups 'A' and 'X', we found that no population configuration using this topology was consistent with the data (Tables S5–S8). Bedouins B were rejected as a sister group to the UEB individuals to the exclusion of Saudi due to an excess of allele sharing between Saudi and the UEB ( $Z = 6.4$ ) while Saudi were rejected as a sister group due to an excess of allele sharing between the UEB and Bedouin B ( $Z = 13.8$ ). Testing the UEB individuals as an outgroup to Saudi and Bedouin B revealed this to be a scenario consistent with the data ( $Z = 0.19$ ) confirming that the two UEB individuals form their own group and are not directly matched by any of the modern populations in our reference data.

**Modelling the genomic ancestry of the UEB individuals.** From a genetic perspective, Levantine populations today fall into a continuum of genetic ancestry, consisting of varying proportions derived from different prehistoric populations. Recent studies suggest differences in the genomic history of the Levant and the Arabian Peninsula, with higher ancient Levantine, Iranian and Eurasian Hunter-Gatherer ancestry proportions in the Levant and excess African ancestry in the latter<sup>14,38</sup>. Within the Levant itself, groups such as Syrians, Palestinians and Jordanians have also



**Fig. 3 Model-based clustering of ancient and modern populations.** ADMIXTURE run with ancient individuals and present-day modern groups. The zoomed in figure shows ancestry proportions of the UEB individuals along with some ancient and modern populations at K = 4, 5, 6.

been suggested to have higher African ancestry relative to nearby populations<sup>13,14</sup> (see also Table 2). To obtain a deeper understanding of these different ancestries, we used *qpAdm* to explore different scenarios for the UEB individuals and relevant modern and historical Levantine populations.

We used Neolithic Levant, Neolithic Iran, Western European Hunter-Gatherers and ancient East Africans approximated by Mota (a 4500-year-old individual from Ethiopia<sup>39</sup>) as potential sources of ancestry. We found that most ancient groups do not require all four sources, but that few of them can always be modelled as a mixture of Levant\_N, Iran\_N and either WHG or Mota, (except Late Bronze Age individuals from Ashkelon<sup>10</sup> which can be modelled without Iran\_N). The UEB individuals can be modelled either as a two-way mixture of Levant\_N (61 ± 6%) and Iran\_N (39 ± 6%) (Table 2), or also with a minor contribution of ancestry from WHG (6 ± 4%) but no models work with Mota as a source (Table 2). Interestingly, Bedouin B, Saudi and Yemenite Jews could be modelled from all four sources,

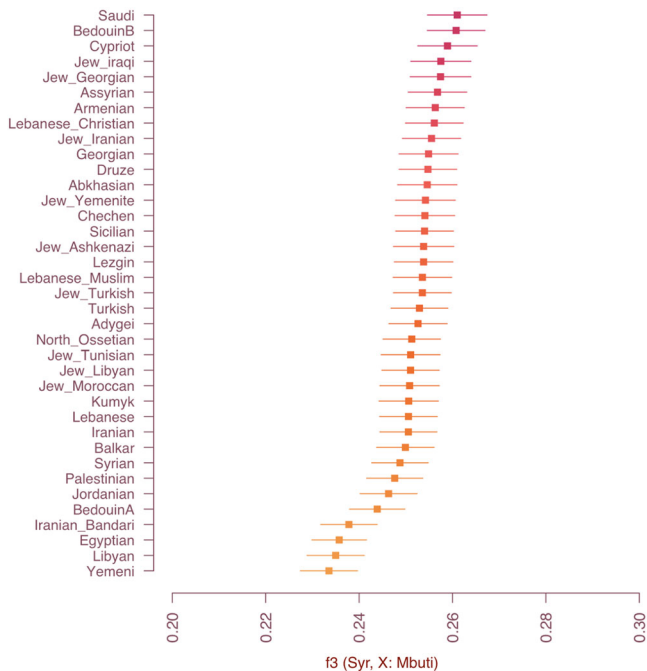
i.e. Levant\_N (56.8 ± 2.7%, 55.1 ± 2.8% and 57.9 ± 2.7%), Iran\_N (34.3 ± 4.3%, 37.6 ± 4.4% and 33.2 ± 4.4%), WHG (2.7 ± 1.6%, 3.2 ± 1.7% and 3.6 ± 1.7%) and Mota (6.2 ± 0.9%, 4.1 ± 0.9% and 5.3 ± 0.9%) as well as from only Levant\_N, Iran\_N and Mota (see Table 2). In addition, most modern Levantine populations except Lebanese Christians and Cypriots can be shown as a mixture of all four sources (Fig. 6).

Thus, despite indications from other analyses that Bedouin B and Saudi seem to be the closest modern populations to the UEB individuals, we obtained slightly different results from *qpAdm* analyses since these modern groups required East African ancestry for a fitting model. This lack of African-related ancestry in the UEB individuals in contrast to their closest modern populations might be the cause of the D statistics results (which placed the UEB individuals as an outgroup to Saudi and Bedouin B) and could be explained by different scenarios. For instance, the Neolithic Levantine ancestral background could have been diluted by the introduction of African-related ancestry later, i.e. after the

UEB individuals. This might also account for the observation that none of the other historical Levantine individuals from earlier time periods seem to produce a fitting model with Mota as an ancestral source. In fact, the only ancient individuals that could be successfully modelled with African-related ancestry are Neolithic Moroccan<sup>40</sup>. Consistent with this explanation, when we also tried to model the UEB individuals as a two-way mixture of already published Levantine individuals more proximate in time and Mota, we obtained high standard errors for Mota-related ancestry in all models and found only a few successful single-

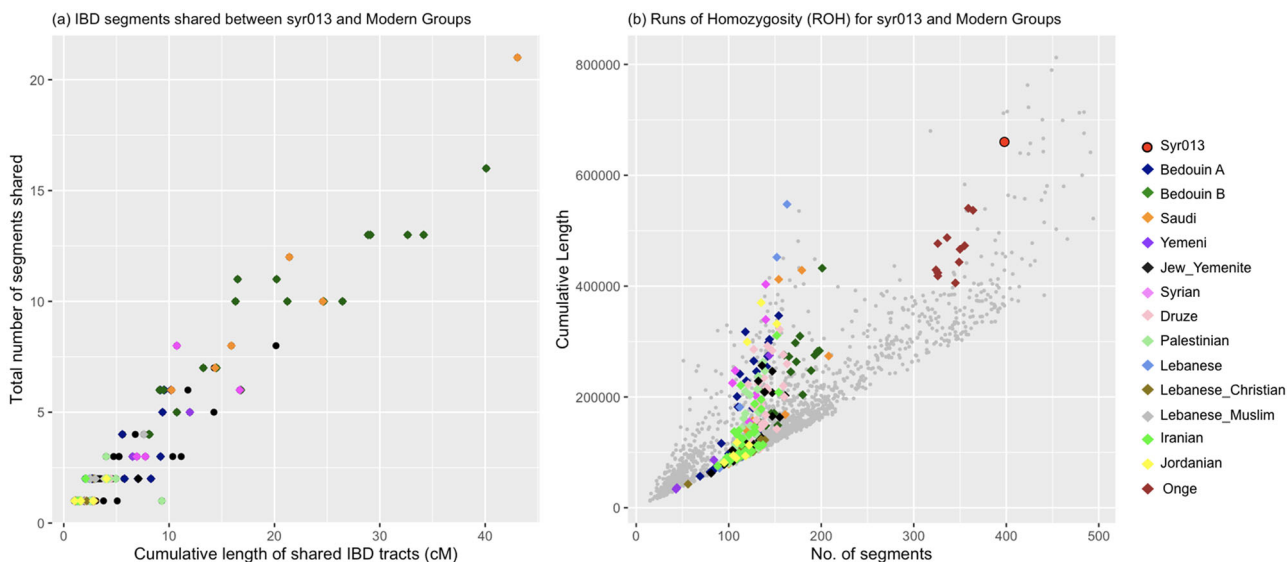
source models from published Levantine individuals (Table S9). Further, when we tried to model modern Levantine populations as a mixture of UEB individuals and Mota (Table S10), we found that all populations required a high amount of ancestry from UEB individuals (~80–99%) and variable contributions from Mota (0.1–20%).

It is also possible that the UEB individuals are representative of groups that migrated from the Arabian Peninsula to the Levant during the early years of Islam and experienced strong cultural barriers that over time prevented mixing with neighbouring populations, resulting in the highly drifted population observed in our data. This would be concordant with the fact that the ancestors of the genetically similar genotyped Negev Bedouins migrated from the Arabian Peninsula to the Negev and Sinai regions around 700 CE, i.e. shortly after the spread of Islam<sup>41</sup>. However, population-level genetic and archaeological data are required to make conclusive inferences about large-scale migrations and religious/cultural stratification especially since it is possible that the Arabian peninsula-related ancestry was present in the Levant long before the UEB individuals lived. It is worth noting that the historical UEB individuals are the population with the lowest amount of genomic data, restricting a precise estimation or even a full rejection of the presence of European or African ancestry in these individuals, thus affecting ancestry estimates. However, even if we cannot fully reject the presence of any African ancestry, all analyses consistently suggest a lower level of African ancestry in the UEB individuals. While the reduced resolution for the historical UEB individuals prevents us from making more reliable inferences at this stage, the genetic differences observed here could potentially suggest a slightly different trajectory and contacts with other groups in the long term compared to Saudi and Bedouins as well as the effect of recent history on population structure. Similar results were also obtained from models repeated with data from UEB individuals restricted to damaged sites (Tables S9, S10).



**Fig. 4 Affinity to modern populations.** Outgroup  $f_3$  statistics show shared genetic drift between the UEB individuals and modern-day populations. Standard errors (SE) were estimated using a weighted block-jackknife procedure. Error bars represent two SEs.

**Low genetic diversity in the UEB individuals.** In order to gain insight into the genetic diversity of the UEB individuals, we estimated Conditional Nucleotide diversity (CND)<sup>42</sup> to be  $0.205 \pm 0.003$ . When compared to all modern Middle East



**Fig. 5 Analysis of diploid genotype calls.** **a** Shared IBD tracts between syr013 and modern populations from the Human Origins 2.0 dataset. The X-axis indicates the cumulative length (in centiMorgans) of shared tracts between a pair consisting of syr013 and a modern-day individual, while the Y-axis shows the total number of shared segments for that pair. Relevant populations are indicated as coloured diamonds. **b** Plot showing Runs of Homozygosity (ROH) for syr013 with relevant modern populations. The X-axis indicates the number of segments with ROH and the Y-axis is their cumulative length.

**Table 2 Modelling ancient and modern Middle Eastern populations.**

**2-source models**

	<b>p value</b>	<b>Levant_N</b>	<b>Iran_N</b>	<b>WHG</b>
UEB Individuals	0.289238	0.610 ± 0.060	0.390 ± 0.060	NA
UEB Individuals_dmg	0.437987	0.666 ± 0.069	0.334 ± 0.069	NA
Turkey_IA_HLN	0.152261	0.515 ± 0.079	0.485 ± 0.079	NA
Mrc_LN	0.0880385	0.721 ± 0.096	0.279 ± 0.096	NA
Levant_BA	0.116881	0.784 ± 0.031	0.216 ± 0.031	NA
Haber2019_NE	0.0687897	0.723 ± 0.049	0.277 ± 0.049	NA
Ash_LBA	0.0636178	0.640 ± 0.131	0.360 ± 0.131	NA
Ash_LBA	0.085684	0.787 ± 0.069	NA	0.213 ± 0.069

**3-source models**

	<b>p value</b>	<b>Levant_N</b>	<b>Iran_N</b>	<b>WHG</b>	<b>Mota_Ethiopia</b>
UEB Individuals	0.396026	0.598 ± 0.058	0.348 ± 0.065	0.055 ± 0.038	NA
UEB Individuals_dmg	0.651473	0.655 ± 0.065	0.281 ± 0.074	0.064 ± 0.043	NA
Bedouin B	0.112744	0.552 ± 0.026	0.394 ± 0.030	NA	0.054 ± 0.007
Saudi	0.224519	0.534 ± 0.028	0.435 ± 0.033	NA	0.031 ± 0.008
Jew_Yemenite	0.158269	0.557 ± 0.027	0.402 ± 0.031	NA	0.041 ± 0.008
Lebanese_Christian	0.309808	0.599 ± 0.020	0.321 ± 0.024	0.080 ± 0.013	NA
Lebanese_Muslim	0.249849	0.488 ± 0.021	0.459 ± 0.026	0.052 ± 0.014	NA
Druze	0.394225	0.543 ± 0.019	0.383 ± 0.023	0.073 ± 0.013	NA
Cypriot	0.105134	0.655 ± 0.021	0.243 ± 0.025	0.103 ± 0.013	NA
Haber2019_NE	0.196697	0.704 ± 0.048	0.233 ± 0.053	0.063 ± 0.031	NA
Ash_IA	0.213846	0.531 ± 0.068	0.328 ± 0.083	0.141 ± 0.043	NA
Ash_LBA	0.164834	0.613 ± 0.119	0.241 ± 0.142	0.145 ± 0.075	NA
Levant_BA	0.866914	0.784 ± 0.030	0.169 ± 0.035	0.046 ± 0.018	NA
Turkey_IA_HLN	0.889632	0.463 ± 0.072	0.407 ± 0.076	0.130 ± 0.051	NA
Mrc_EN	0.0906708	0.327 ± 0.079	0.335 ± 0.089	NA	0.338 ± 0.023
Mrc_LN	0.604079	0.804 ± 0.100	0.120 ± 0.117	NA	0.076 ± 0.031

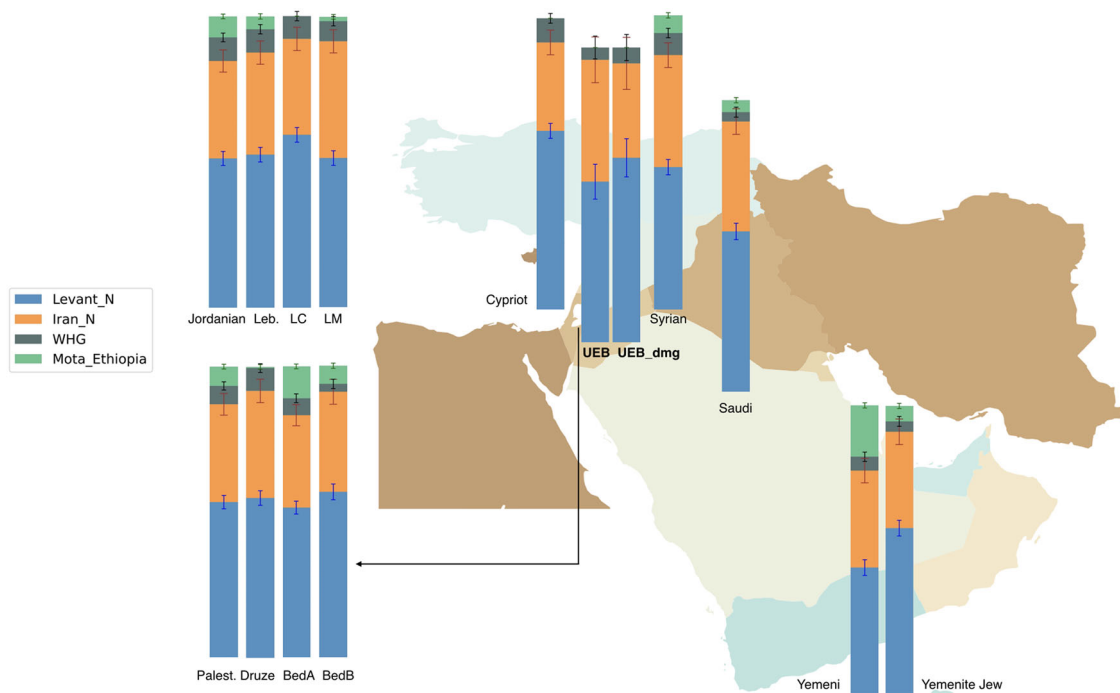
**4-source models**

	<b>p value</b>	<b>Levant_N</b>	<b>Iran_N</b>	<b>WHG</b>	<b>Mota_Ethiopia</b>
UEB Individuals (infeasible)	0.87482	0.564 ± 0.061	0.426 ± 0.080	0.043 ± 0.040	-0.032 ± 0.020
UEB Individuals_dmg (infeasible)	0.671526	0.639 ± 0.066	0.326 ± 0.090	0.055 ± 0.045	-0.021 ± 0.020
Bedouin B	0.182317	0.568 ± 0.027	0.343 ± 0.043	0.027 ± 0.016	0.062 ± 0.009
Saudi	0.561657	0.551 ± 0.028	0.376 ± 0.044	0.032 ± 0.017	0.041 ± 0.009
Jew_Yemenite	0.672415	0.579 ± 0.027	0.332 ± 0.044	0.036 ± 0.017	0.053 ± 0.009
Yemeni	0.895286	0.441 ± 0.027	0.334 ± 0.042	0.048 ± 0.016	0.177 ± 0.009
Bedouin A	0.430541	0.517 ± 0.022	0.316 ± 0.036	0.058 ± 0.015	0.109 ± 0.008
Syrian	0.589049	0.485 ± 0.026	0.380 ± 0.042	0.075 ± 0.016	0.060 ± 0.009
Lebanese	0.770664	0.527 ± 0.025	0.350 ± 0.040	0.080 ± 0.016	0.044 ± 0.009
Lebanese_Christian (infeasible)	0.172228	0.595 ± 0.025	0.329 ± 0.040	0.078 ± 0.016	-0.002 ± 0.008
Lebanese_Muslim	0.606834	0.515 ± 0.025	0.402 ± 0.040	0.069 ± 0.016	0.015 ± 0.008
Jordanian	0.858381	0.514 ± 0.024	0.334 ± 0.038	0.081 ± 0.015	0.072 ± 0.008
Druze	0.254556	0.551 ± 0.025	0.367 ± 0.040	0.078 ± 0.015	0.004 ± 0.008
Cypriot (infeasible)	0.31197	0.625 ± 0.026	0.308 ± 0.043	0.084 ± 0.017	-0.017 ± 0.009
Palestinian	0.689973	0.535 ± 0.023	0.335 ± 0.037	0.063 ± 0.014	0.066 ± 0.008
Haber2019_NE (infeasible)	0.648016	0.668 ± 0.050	0.309 ± 0.065	0.053 ± 0.032	-0.031 ± 0.016
Haber2019_RP (infeasible)	0.00555823	0.673 ± 0.052	0.298 ± 0.071	0.034 ± 0.073	-0.004 ± 0.015
Haber2019_Eur (infeasible)	0.0305447	0.390 ± 0.057	0.308 ± 0.072	0.365 ± 0.038	-0.063 ± 0.017
Haber2019_Adm (infeasible)	0.0282109	0.605 ± 0.053	0.230 ± 0.072	0.188 ± 0.035	-0.022 ± 0.017
AgTam_BACanaanites (infeasible)	0.53806	0.712 ± 0.026	0.273 ± 0.042	0.040 ± 0.016	-0.025 ± 0.008
Anatolia_BA (infeasible)	0.139901	0.801 ± 0.045	0.244 ± 0.073	0.014 ± 0.026	-0.059 ± 0.014
Arm_MLBA (infeasible)	0.0577635	0.331 ± 0.078	0.595 ± 0.110	0.120 ± 0.047	-0.046 ± 0.022
Arm_ChI (infeasible)	0.940145	0.488 ± 0.032	0.408 ± 0.050	0.152 ± 0.021	-0.048 ± 0.010
Ash_IA	0.112139	0.555 ± 0.109	0.286 ± 0.172	0.149 ± 0.052	0.010 ± 0.035
Ash_LBA	0.0944184	0.672 ± 0.199	0.105 ± 0.375	0.190 ± 0.136	0.033 ± 0.074
Iran_ChI (infeasible)	0.598333	0.429 ± 0.037	0.623 ± 0.059	-0.014 ± 0.021	-0.038 ± 0.011
Levant_ChI (infeasible)	0.239445	0.919 ± 0.032	0.117 ± 0.052	-0.004 ± 0.020	-0.032 ± 0.010
Levant_BA (infeasible)	0.738578	0.775 ± 0.040	0.189 ± 0.066	0.041 ± 0.024	-0.004 ± 0.013
Mrc_EN	0.202773	0.306 ± 0.075	0.252 ± 0.097	0.089 ± 0.049	0.352 ± 0.024
Mrc_LN (infeasible)	0.457194	0.812 ± 0.102	0.153 ± 0.133	-0.036 ± 0.070	0.071 ± 0.033
Sidon_BA (infeasible)	0.102462	0.663 ± 0.048	0.355 ± 0.066	0.014 ± 0.033	-0.032 ± 0.015
Turkey_BA (infeasible)	0.0119186	0.661 ± 0.050	0.375 ± 0.064	0.028 ± 0.030	-0.064 ± 0.014
Turkey_IA_HLN (infeasible)	0.86774	0.451 ± 0.076	0.437 ± 0.093	0.125 ± 0.052	-0.013 ± 0.023
Turkey_OT (infeasible)	0.0194304	0.107 ± 0.072	0.596 ± 0.098	0.254 ± 0.051	0.042 ± 0.022

This analysis used Neolithic Levant (Levant\_N), Neolithic Iran (Iran\_N), European Western Hunter-Gatherers (WHG) and Mota from Ethiopia as source populations. Nested two-source and three-source feasible models found in the four-source scenario are shown first, followed by the results of ancestry proportions obtained from all four sources. For three-sources, NA implies that that source was not used in the model. Values are shown along with standard errors, with infeasible models pointed out along with the target population. Outgroups used: Mbuti, Sweden Hunter-Gatherers (SHG), Kostenki14, Ust-Ishim, Central Hunter-Gatherers (CHG) and Neolithic Anatolia. UEB\_dmg indicates models run with data from the UEB individuals restricted to damaged sites.

populations in our reference panel<sup>1</sup>, syr005 and syr013 showed the lowest levels of genetic diversity. For comparison purposes, we included Onge (native hunter-gatherers from the Andaman Islands known to have extremely low genetic diversity due to their small population size and prolonged isolation<sup>43</sup>) and results showed a lower value of CND than syr005 and syr013 (Fig. S4 and Table S11). However, it should be noted that comparisons between ancient and modern populations, sequence and SNP

chip data might be subject to some biases<sup>42,44,45</sup>. Despite this low genetic difference between the two individuals, we do not see evidence of them being close relatives as the between individuals distance is almost exactly twice the within an individual distance (using a pseudo-twin representation of syr005 and syr013, see Methods) as expected for an unrelated pair of individuals. As low genetic diversity could be the result of small population size and/or inbreeding between individuals, we analysed runs of



**Fig. 6 Ancestry modelling of Levantine populations.** Ancestry of a smaller subset of Middle Eastern populations modelled from Neolithic Levant, Neolithic Iran, European Hunter-Gatherers (WHG) and a 4500-year-old individual Mota from Ethiopia using *qpAdm*. Ancient populations include the UEB individuals<sup>12</sup> (Palest. Palestinians, Leb. Lebanese, LM Lebanese Muslim, LC Lebanese Christians, BedA,B Bedouin A, B). Error bars represent one standard error.

homozygosity (ROH) to determine the extent of inbreeding. In the syr013 individual, we observe both high numbers and cumulative lengths of ROH segments in syr013 compared to other Middle Eastern populations (Fig. 5b). Consistent with CNV results, this suggests that syr013 was part of a small and/or inbred population. To ensure that the results were not affected considerably by missing sites, we also ran the analysis with the Simons Genome Diversity Panel (SGDP)<sup>46</sup> restricting the analysis to the sites covered for syr013 and found a similar pattern (Fig. S5). These findings also might suggest similar social structures as those seen in nomadic groups in the Levant, where affiliation of individuals to tribal or clan-structured groups is associated with consanguinity, strong barriers to extra-tribal marriage and results in low genetic diversity, small effective population size and a high incidence of recessive disorders<sup>47,48</sup>.

**Insights into the phenotype of the UEB individuals.** To investigate potential lactase persistence in Umayyad Era Syria we tested five SNPs known to be associated with the trait (Table S12). All five SNPs analysed here are located upstream of the gene *LCT* (encoding lactase) in introns of the gene *MCM6*, which serves as an enhancer for *LCT* transcription.<sup>49</sup> −13910C/T, −13915T/G, −14010G/C, −13907C/G, and the rare variant −14107G/A. Lactase persistence is known to be an autosomal dominant trait, thus the presence of a single derived allele is sufficient for milk digestion. Interestingly, nine reads in sample syr013 were found mapping to the SNP −13915T/G, out of which five were the derived allele and four were ancestral which indicates that syr013 was heterozygous and lactose tolerant. Despite the long history of camel herding in this region beginning with its domestication more than 6000 years ago<sup>50</sup>, syr013 being heterozygous marks the earliest observation of this variant in genomic data. Lactase persistence could not be tested for sample syr005 as no reads covered the sites of interest. Nevertheless, this finding again draws a connection to the Arabian Peninsula as this variant is common in

modern Arabian (for e.g. has a frequency of 72–88% in Saudi<sup>50</sup>) populations as well as in pastoralists that have traditionally relied on the Arabian camel (*Dromedary camelous*) for milk consumption and show high levels of lactase persistence phenotypes (>75% of individuals in Bedouins)<sup>49,50</sup>. We also tested other autosomal dominant or recessive conditions reported to occur frequently in Arabs<sup>47</sup> (including familial hypercholesterolaemia, glucose-6-phosphate dehydrogenase deficiency, sickle-cell anaemia, Bardet-Biedl syndrome, etc.) but no pathogenic alleles were found in syr013 (Table S13).

**Diet of the UEB individuals.** We attempted to infer dietary consumption patterns in these individuals using stable isotope analysis, using collagen samples extracted from the petrous bone (Table S14). The bulk stable isotope data for these individuals are typical of a C<sub>3</sub> terrestrial diet ( $\delta^{13}\text{C}$  −18.5‰ and −19.2‰;  $\delta^{15}\text{N}$  +11.5‰ and +13.1‰) with high animal protein intake. Consumption of food from an arid location or where manuring was practiced are also possible explanations for the high nitrogen values. Without comparative faunal or human data from the period and location, it is difficult to ascertain the extent of these effects. We contextualised our data with human stable isotope regional datasets from historical/proto-historical periods (e.g. from Lebanon, Jerusalem, Northern Jordan, Syria [albeit Northern], including Early and Late Roman, Parthian, Byzantine, Medieval and some modern data<sup>51–55</sup> (see Table S14 and Fig. S6). In addition, we have compared this data with those obtained from Bedouin tombs from Jordan dating between the 13th and 19th centuries<sup>56</sup>. The regional datasets demonstrate similar  $\delta^{13}\text{C}$  values to the two UEB individuals studied here, but generally lower  $\delta^{15}\text{N}$  (6.8–9.5‰, (excluding St Stephens, Jerusalem)), with the exception of a Medieval individual with European ancestry ( $\delta^{15}\text{N}$  = 11.3‰) excavated from Lebanon<sup>12</sup>. It is possible that this individual hailed from Europe and had a different diet from those in the Middle East (before being buried there). The monastic



community of St Stephens tends to have higher nitrogen values than the other datasets ( $\delta^{15}\text{N}$  7.3–12.6‰ and 6 individuals of the 68 as high or higher than 11.5‰). The high nitrogen values for this site have been attributed to animal protein consumption<sup>51</sup>. The Bedouin dataset has comparable (and higher) nitrogen isotope ratios to our two individuals ( $\delta^{15}\text{N}$  10–17.3‰), however, less negative  $\delta^{13}\text{C}$  values. The high nitrogen values of the Bedouin are attributed to 'considerable amounts of animal protein' including blood, milk (products) and meat, and the high  $\delta^{13}\text{C}$  ratios are attributed to the habitation of and resource exploitation of predominantly  $\text{C}_4$  or mixed  $\text{C}_4$ - $\text{C}_3$  ecosystems<sup>56</sup>. Compound specific amino acid carbon isotope values were obtained<sup>57,58</sup> to investigate whether the high nitrogen values were influenced by the intake of aquatic resources. The  $\delta^{13}\text{C}_{\text{GLY}}-\delta^{13}\text{C}_{\text{PHE}}$  values for the individuals are between 10 and 13‰<sup>59</sup>,  $\delta^{13}\text{C}_{\text{VAL}}-\delta^{13}\text{C}_{\text{PHE}}$  values around 0 to 1‰<sup>59,60</sup> and the  $\delta^{13}\text{C}_{\text{LEU}}-\delta^{13}\text{C}_{\text{PHE}}$  values are negative (Table S15); these values are consistent with a terrestrial diet without significant input from aquatic resources<sup>58,61</sup>. We conclude that the Tell Qarassa individual's isotope values can best be explained as diets with high levels of animal product consumption (i.e. pastoral food acquisition – like the Bedouin dataset), but obtained from an almost exclusively  $\text{C}_3$  ecosystem (like the other regional datasets).

The majority of stable isotope data discussed herein were measured from collagen samples extracted from cortical long bone and ribs, with the exception of four samples taken from molars<sup>53</sup> and the data from refs. 11,12,52, who, like us, used petrous bone. Jørkov et al.<sup>62</sup> have argued that the isotope signal from petrous bone reflects some childhood diet, reporting slightly lower  $\delta^{13}\text{C}$  values (<0.3‰) and slightly higher  $\delta^{15}\text{N}$  values (<0.8‰) in petrous bone compared to those of rib and femur samples (at a population level). Even if this offset applies to our dataset, these differences would not affect the dietary interpretation of the two individuals from Tell Qarassa.

This dietary interpretation is intriguingly consistent with the Bedouin ancestry components of the individuals and the presence of the lactase persistence variant in syr013. Dietary stable isotope studies of ancient Islamic individuals have been widely studied in Iberia (and the Balearic Islands), but less extensively elsewhere. Lopez-Costas and Alexander<sup>63</sup> conclude that there is no strong evidence that Islamic diet has a distinct isotopic signal that distinguishes it from contemporary Christian diets (e.g. ref. 64), despite some well-known cultural differences (e.g. the prohibition of pork in Islam, fasting and fish-eating in Christianity). They highlight that further research is required to investigate such distinctions<sup>63</sup> and certainly further work is required in the Middle East to expand our knowledge of dietary stable isotopes in the Umayyad Era.

**Metagenomic screening for pathogens.** A possible explanation for both individuals not being buried in a traditional Muslim cemetery could be that they represent plague victims. Several archaeogenomic studies have successfully identified different pathogen sequences in DNA extracted from archaeological remains<sup>65</sup>. We screened our sequences for DNA from different known pathogens and find potential traces of different bacterial species that can cause infections in syr005 (Supplementary Note S2 and Figs. S7–S10). The limited number of sequences does not allow us to unambiguously authenticate these findings and additional research would be required to verify whether a complex infection played a role in the death of these individuals.

## Conclusions

The continuous development and improvement in ancient DNA methodologies and molecular techniques is constantly pushing

the temporal and geographic limits of aDNA recovery, exploring older time periods<sup>66–68</sup> and hostile environments (e.g. hot and humid) for DNA preservation. The Middle East and the Arabian Peninsula are pivotal regions in the timeline of human history and an increasing number of aDNA studies have attempted to understand their genetic history. Although there have been successful studies<sup>1,3–14</sup>, given the poor conditions for DNA preservation, this process is proving to be slower than in more environmentally favourable regions of the world. Nonetheless, given its historical importance, each newly recovered DNA sequence adds an important piece to the genomic and cultural puzzle of a war-stricken region currently under difficult access.

We have been able to infer that the two Umayyad Era individuals buried on top of a prehistoric site represent individuals that were genetically close (while not identical) to a subgroup of modern-day Bedouins ('Bedouin B') from the Negev desert in Israel as well as Saudi from the Arabian peninsula in the Human Origins 2.0 dataset. Several sources document the existence of historical nomadic groups that either occupied the Tell Qarassa region and/or migrated from the Arabian Peninsula to Syria during the Umayyad Era<sup>69,70</sup>. However, the absence of contemporary genetic data from this period limits resolution on finer substructure among such groups. Additionally, while genomic methods are a powerful tool to analyse an individual's ancestry and infer past demography and population dynamics, genetic datasets often use regional, cultural and/or archaeological affiliations for genotyped individuals. Thus, while it seems that the UEB individuals are genetically similar to some nomadic groups, determining their exact cultural affiliation is a question that cannot be answered through genomic analysis.

The archaeological context proves to be slightly more informative regarding their burial characteristics, i.e. orientation towards Mecca, separate graves and wrapping practice which appears to be indicative of these individuals possibly being early adherents of Islam in a Christian-majority region<sup>71</sup>. The Tell Qarassa graves do not represent a traditional Muslim context, i.e. a traditional Muslim cemetery and – from the available evidence – do not seem to have been located near a permanent settlement of the period. This suggests that it is possible that these two individuals were transient Muslims in the region. The absence of trauma to the bones and the young age of the deceased suggests that they could have died from disease, possibly the Justinian plague which ravaged the Middle East from 541 AD to 749 AD recurring in cycles<sup>72</sup>. Specifically, the dates of the burials may be linked to the outbreak of 79 AH (698 AD) which was reported in Syria by as-Suyufi<sup>73</sup>. However, we did not find conclusive evidence of pathogens and the exact cause of their death remains difficult to pinpoint.

The genomic ancestry of the two individuals buried at Tell Qarassa in the late seventh or early eighth century offers a glimpse of early Islamic society in Syria. This study provides further insight into the possible re-use of prehistoric burial sites by Muslim groups. On a general level, this burial provides additional indications for the early adoption of specific Islamic burial rites which were followed even in remote locations. At present, there are no examples of genetic studies from the region which relate to this period, the only genetic data related to early Islamic burials is the study of two individuals from the south of France<sup>74</sup>. Our results suggest the early presence of Muslim Arabs in the Syrian countryside. Extensive additional sampling from different groups in this region is crucial to understand the extent of their genetic structure today and to potentially identify relatively genetically isolated populations, which could have implications for population genetic and clinical studies. The Middle East is a region with a complex history and a diverse ethnic and genetic composition, yet our current understanding of the genetic

structure in the past and present appears to have only scratched the surface.

## Methods

**Archaeological context.** The archaeological site of Tell Qarassa, located in modern-day Syria (Fig. 1a), is prominently known for its evidence of human occupation since the Epipaleolithic period to the Iron Age<sup>75–77</sup>. Located on the shores of an ancient dried-up lake, it is described to consist of multiple sites or so-called ‘Tells’, one of which (Northern Tell) contains remains from Pre-Pottery Neolithic B to Late Chalcolithic settlements, while the other (Southern Tell) holds evidence of Early Bronze to Iron Age remains. Evidence of a Natufian settlement has also been found in close proximity<sup>78</sup>. The village of Qarassa is a Druze community today. Pre-Pottery Neolithic mortuary practices have been described from this site, shedding light on such practices in this period<sup>18,19</sup>. The individual syr005 was found laid on his back in a decubitus supine position, although the lower limbs were slightly flexed and placed on their right side. The burial was oriented east-west, with the head at the west, facing south. Further, individual syr013 was placed on her right side in a lateral decubitus position oriented east-west, with the head at the west, facing south. The distribution of the skeletal elements suggests that both bodies were wrapped before burial<sup>20</sup>. When soft tissues of the body decay faster than the wrapping, it can create either temporary or semi-permanent spaces around the putrefied body, yielding some skeletal movements at the disarticulated joints before wrapping decay (i.e. right elbow in individual syr013, left shoulder in individual syr005). These individuals were located very close to each other. However, intensive archaeological fieldwork has not yielded evidence of further burials from this period at the site. Therefore, the archaeological record does not support that these burials belong to a cemetery for a specific community. Permits for archaeological samples processed in this study were obtained from the General Directorate of Antiquities and Museums (DGAM), the relevant authority of the Arab Republic of Syria. The remains are curated by DGAM and have been deposited at the Archaeological Museum of As-Suwayda (Syria).

**Radiocarbon dating and isotopic analysis.** Collagen was extracted from two petrous bones following a modified Longin method<sup>79,80</sup> at the Molecular Archaeology Laboratory at La Trobe University using cold 0.6 M HCl and yields are recorded in Table S14. The collagen was directly radiocarbon dated (AMS) at Waikato University in New Zealand and calibrated using the Oxcal 4.3 programme<sup>81</sup>, and the IntCal13 calibration curve<sup>82</sup>. Radiocarbon results for individuals syr005 and syr013 were 666–768 Cal AD  $2\sigma$  (1294  $\pm$  18 Cal BP, Lab code wk-46474), and 665–766 Cal AD  $2\sigma$  (1302  $\pm$  15 Cal BP, Lab code wk-46475), respectively. The collagen stable isotope values and quality parameters were also measured at Waikato and are presented in Table S14. Compound specific stable isotope analysis of amino acids (Table S15) was carried out at the Molecular Archaeology Laboratory, La Trobe University using LC IRMS following methods similar to those described in ref. <sup>57</sup> and ref. <sup>58</sup>. In brief, 1 mg of collagen was hydrolysed under vacuum for 24 h at 110 °C and the hydrolysate was dried in a rotary vacuum concentrator and frozen until required for analysis. The sample was resolved in Milli-Q water with an internal standard (2-amino isobutyric acid). Instrumental analysis was carried out using a three-phase method similar to that described in ref. <sup>58</sup> with 10–15  $\mu$ g of hydrolysate delivered to the column. Isotope values were calculated relative to standard CO<sub>2</sub> gas peaks delivered throughout the analytical run, with the CO<sub>2</sub> calibrated against USGS-40.

**Sample preparation and sequencing.** Prior to DNA extraction, the petrous bones were UV irradiated (6 J/cm<sup>2</sup> at 254 nm) and the first millimetre of bone surface abraded using a Dremel™ tool. DNA was extracted from a 100–200 mg piece of the bone using a silica binding method<sup>83</sup>, with an incubation of 24–48 h, using the MinElute column Zymo extender assembly replaced by the High Pure Extender Assembly (Roche High Pure Viral Nucleic Acid Large Vol. Kit) and performed twice for each sample. Further, blunt-end Illumina multiplex sequencing libraries were prepared<sup>84</sup> resulting in two double-stranded libraries per sample. Library amplifications were performed as in<sup>85</sup> using indexed primers<sup>84</sup> and 4–11 cycles (11 and 10 cycles for syr005; 4 and 6 cycles for syr013). All extraction and library preparation steps were conducted at the dedicated ancient DNA facility at Stockholm University. A total of four DNA libraries were shotgun sequenced on a HiSeq X10 sequencing platform (150 bp paired-end reads) at the NGI Stockholm.

**NGS data processing.** Sequenced reads were mapped to the human reference genome build 37 (hs37d5) using *BWA aln*<sup>86</sup> with non-default parameters  $-l$  16500  $-n$  0.01  $-o$  2. Data were merged at the library level using *samtools* v1.5<sup>87</sup>. PCR duplicates were collapsed using a modified version of *FilterUniqSAMCons\_cc.py*<sup>44,88</sup>. Different libraries for one individual were then merged into one bam file using *samtools* v1.5<sup>87</sup>. Reads shorter than 35 bp, showing more than 10% mismatch with the reference and/or a mapping quality score below 30 were discarded. Biological sex was inferred using two different approaches<sup>29,30</sup>.

**Contamination estimates.** Contamination for the two individuals was assessed on three different levels: mitochondria, X-chromosome and autosomes. Two methods

were used to estimate mitochondrial contamination<sup>25,26</sup>. X-chromosomal contamination was estimated for the male individual syr005 using the approach implemented in *ANGSD*<sup>28</sup>. *VerifyBAMID*<sup>27</sup> was used to estimate autosomal contamination in both samples. It uses a hypothetical ‘true genotype’ model and checks whether reads in a bam file are more likely to match a single individual or result from a mixture of other samples, such as a closely related/a different individual.

**Uniparental haplogroups.** Mitochondrial haplogroups were identified using Haplofind<sup>89</sup> and HaploGrep<sup>90</sup>, two web-based applications that use the Human Phylogenetic Tree<sup>91</sup> to identify the Y-haplotype of sample syr005, reads mapping to the Y chromosome with minimum base and mapping quality of 30 were compared to biallelic substitution SNPs from the International Society of Genetic Genealogy (ISOGG, <https://isogg.org/>). We excluded A/T and G/C SNPs to avoid strand misidentification and C/T and A/G SNPs to avoid post-mortem damages.

**Population genetic analyses.** For population genetic analyses, we merged the ancient samples with modern genotype data of the Human Origins 2.0 dataset<sup>1</sup> as well as 272 published ancient individuals from the Levant, Europe, Caucasus, North and sub-Saharan Africa<sup>1,6–10,12,39,40,92–98</sup>, resulting in a total of 2854 individuals and 541300 SNPs that passed quality control filters from the Human Origins dataset. We generated pseudohaploid representations of the ancient individuals by randomly drawing one allele from the samtools mpileup<sup>87</sup> output at each SNP site, restricting the analysis to minimum mapping and read qualities of 30 and coded transition sites as missing to avoid post-mortem damage.

**Principal components analysis (PCA).** A principal component analysis was conducted using *smartpca*<sup>99</sup> using the options *lsqproj* and *shrinkmode*. The first PCA included 73 modern populations from North, West, South, Central and East Europe, Caucasus, Turkey and the Middle East, North and North-east Africa, the Arabian Peninsula, along with 262 published ancient samples from the Levant, North and sub-Saharan Africa and the two ancient UEB individuals ( $N = 1321$ ) from this study. In order to identify the closest modern populations to the UEB individuals, we generated a second PCA with an increased resolution by reducing the original set of modern populations. For this second PCA we excluded divergent groups such as Europeans, Somalis and was limited to 37 modern groups in the Middle East, Turkey and Caucasus henceforth the Middle East, along with ancient individuals (Middle Eastern,  $N = 757$  panel). PCA plots were generated using GNU R.

**ADMIXTURE.** We used *ADMIXTURE*<sup>34</sup> for model-based clustering analysis. All data were haploidized by randomly picking one allele per individual at each SNP. Next, the dataset was thinned by pruning out SNPs in linkage disequilibrium using *PLINK* v1.9<sup>100</sup> with a window size of 200 kb, a step-size of 25 and a squared correlation ( $r^2$ ) threshold of 0.4. Following this step, 59509 of 541300 variants were removed. *ADMIXTURE* was run for ten iterations with different random starting seeds, five-fold cross validation and the number of ancestral populations ( $K$ ) was varied from 2 to 10. Results of admixture were visualised using Pong<sup>101</sup>. We ran unsupervised analyses to study the assignment of genetic clusters in the larger set of 73 modern populations from Europe, Middle East, Arabian Peninsula, North Africa and Caucasus and the ancient individuals grouped into 28 population labels determined from their respective studies ( $N = 1321$ ).

**qpAdm.** We used *qpAdm*<sup>102</sup> from the *ADMIXTOOLS*<sup>35</sup> package with the options *allsnps*: YES, *details*: YES and *summary*: YES to run various modelling scenarios for ancient and modern Levantine groups. First, we modelled the UEB individuals, other ancient and modern Levantine groups as a mixture of four basal ancient populations, namely Neolithic Levant, Neolithic Iran, Western European Hunter-Gatherers and Mota (a 4500-year-old individual from Ethiopia<sup>39</sup>) and their subsets. Secondly, we modelled the UEB individuals as a mixture from ancient Levantines more proximate in time (i.e. Iron/Bronze Age and later periods) along with Mota from Africa<sup>39</sup>. Thirdly, we modelled modern Middle Eastern groups as a mixture of the UEB individuals and Mota. Finally, all models that involve UEB individuals were also run by restricting the models to damaged reads. For this, a likelihood-based approach implemented in *PMDtools*<sup>103</sup> was used, restricting to sequences with a PMD score of at least 3. For all models, we chose 6 outgroups (‘right’ populations): Mbuti, Sweden Hunter-Gatherers<sup>92</sup> (SHG), Kostenki14<sup>104</sup>, Üst-Ishim<sup>105</sup>, CHG<sup>95</sup> and Neolithic Anatolia<sup>92</sup>. We attempted to implement suggested recommendations for *qpAdm*<sup>106</sup> and tried to minimise combinations of ancient and modern groups as references, as well as to keep the same set of sources and references for all targets.

**f and D statistics.** To detect shared genetic history between the sequenced samples and various Middle East populations, we used  $f_3$  and  $D$  statistics<sup>35,107</sup>. Population genetic summary statistics were calculated using *POPSTATS*<sup>108</sup> with sub-Saharan Mbuti rainforest hunter-gatherers as an outgroup. Standard errors were estimated using a weighted block jackknife procedure.

**Conditional nucleotide diversity.** To estimate genetic diversity in our data, we used conditional nucleotide diversity (CND)<sup>42</sup>. CND is a measure of genetic diversity within a population based on a comparison of nucleotide differences between two individuals. We used transversion sites that are polymorphic in the Human Origins 2.0 SNP array. Standard errors were estimated using a block jackknife procedure.

**Analysis of genetic kinship.** Commonly used approaches for estimating a degree of genetic relationship requires some type of reference data to set a baseline for relatedness. This is usually obtained either by using allele frequencies from modern populations or pairs of unrelated individuals from the same population<sup>109</sup>. Neither of these were an option for this study as we are faced with two individuals from a unique population. Instead, we sampled two distinct reads per individual at each SNP site with at least two reads resulting in two pseudohaploid representations for syr005 and syr013. These representations should behave as identical twins ('pseudo-twins') in a pairwise comparison, meaning that they should show half the genetic distance of an unrelated pair from the same population. Hence, we can compare individual distances to distances between syr005 and syr013 to assess their genetic relationship.

**Phenotypic analysis.** Given the observed Arabian ancestry in the sequenced samples, we looked at phenotypic variants that might be expected in this region. For instance, lactase persistence (LP) has been described in the Arabian Peninsula<sup>50</sup>. We also looked for common recessive disorders described in Arab populations from the OMIM (Online Mendelian Inheritance in Man) catalogue<sup>110</sup>. Information about SNP positions and the phenotype was taken from dbSNP<sup>111</sup> and SNPedia<sup>112</sup>. A pileup file was created in samtools from .bam files with a high base quality ( $q > 30$ ) as well as mapping quality ( $Q > 30$ ) and the occurrence of variants known to be involved in LP and other disorders was checked.

**Diploid genotype calls.** We used *snpAD* version 0.3.4<sup>36</sup>, a diploid genotype caller dedicated to work with ancient DNA to call diploid genotypes for syr013. Reads were first realigned around indels using *GATK*<sup>113</sup>. The genotype calls were restricted to reads with mapping quality of at least 30 and base quality of at least 30, calculating an error profile for the first and last 5 bp of each read. Downstream analyses were then restricted to positions with a genotype quality of at least 30 and between 5 and 22 reads sequencing depth at each position.

**ROH.** Runs of homozygosity was estimated for modern populations and the diploid calls of syr013 using *PLINK*<sup>100</sup> with parameters -homozyg-density 50, -homozyg-gap 100, -homozyg-kb 500, -homozyg-snp 100, -homozyg-window-het 1, -homozyg-window-snp 100, -homozyg-window-threshold 0.05 and -homozyg-window-missing 25 for the Human Origins dataset. For the SGDP dataset, the analysis was restricted to 6,383,589 sites with a minor allele frequency of at least 0.01 and no missing data after adding syr013. Therefore, -homozyg-window-missing 25 was set to 1 for the SGDP dataset.

**IBD.** Segments showing identity-by-descent between the UEB individual syr013 and modern human populations were determined using *Beagle* 4.1<sup>37</sup>. We used Human Origins 2.0 as well as the Simons Genome Diversity Project (SGDP)<sup>46</sup> to compare results between the HO panel and the SGDP data, which was filtered to remove missing sites from syr013. *Beagle*'s IBD was run using VCF files with the following parameters: *ibdtrim* = 30 and 150 for the Human Origins and SGDP datasets respectively; *ibd* = true; and *impute* = FALSE, for all chromosomes. We selected segments with a LOD score >3 and length >1 cM for both panels.

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

## Data availability

Generated sequence data is available at the European Nucleotide Archive (ENA) under the accession number PRJEB38008. The remains are curated by the General Directorate of Antiquities and Museums (DGAM) of the Arab Republic of Syria. Raw data for the main figures can be found in Data 1.

Received: 22 February 2021; Accepted: 20 May 2022;

Published online: 07 June 2022

## References

- Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419 (2016).
- Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409 (2014).
- Kılınc, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016).
- Gallego-Llorente, M. et al. The genetics of an early Neolithic pastoralist from the Zagros, Iran. *Sci. Rep.* **6**, 31326 (2016).
- Broushaki, F. et al. Early Neolithic genomes from the eastern fertile crescent. *Science* **353**, 499–503 (2016).
- Skourtanioti, E. et al. Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* **181**, 1158–1175.e28 (2020).
- Harney, É. et al. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun.* **9**, 1–11 (2018).
- Agranat-Tamir, L. et al. The genomic history of the Bronze Age Southern Levant. *Cell* **181**, 1146–1157.e11 (2020).
- Haber, M. et al. Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. *Am. J. Hum. Genet.* **101**, 274–282 (2017).
- Feldman, M. et al. Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. *Sci. Adv.* **5**, eaax0061 (2019).
- Haber, M. et al. A genetic history of the Near East from an aDNA time course sampling eight points in the past 4,000 Years. *Am. J. Hum. Genet.* **107**, 149–157 (2020).
- Haber, M. et al. 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 (2019).
- Haber, M. et al. Genome-wide diversity in the Levant reveals recent structuring by culture. *PLoS Genet.* **9**, e1003316 (2013).
- Haber, M. et al. 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> (2019).
- Hawting, G. R. *The First Dynasty of Islam: The Umayyad Caliphate AD 661–750* (Routledge, 2001).
- Shboul, A. & Walmsley, A. Identity and self-image in Syria-Palestine in the transition from Byzantine to early Islamic rule: Arab Christians and Muslims. *Mediterr. Archaeol.* **11**, 255–287 (1998).
- Magness, J. *The Archaeology of the Early Islamic Settlement in Palestine*. Vol. 1 (Penn State Press, 2003).
- Santana, J., Velasco, J., Ibanez, J. J. & Braemer, F. Crania with mutilated facial skeletons: a new ritual treatment in an early Pre-Pottery Neolithic B cranial cache at Tell Qarassa North (South Syria). *Am. J. Phys. Anthropol.* **149**, 205–216 (2012).
- Santana, J. et al. Interpreting a ritual funerary area at the early Neolithic site of Tell Qarassa north (south Syria, late 9th millennium BC). *J. Anthropol. Archaeol.* **37**, 112–127 (2015).
- Willis, A. & Tayles, N. Field anthropology: application to burial contexts in prehistoric Southeast Asia. *J. Archaeol. Sci.* **36**, 547–554 (2009).
- AlQahtani, S. J., Hector, M. P. & Liversidge, H. M. Brief communication: the London atlas of human tooth development and eruption. *Am. J. Phys. Anthropol.* **142**, 481–490 (2010).
- Buikstra, J. E. & Ubelaker, D. H. Standards for data collection from human skeletal remains. *Arkansas Archeological Survey Research Series* (1994).
- Coqueugnot, H. & Weaver, T. D. Brief communication: infracranial maturation in the skeletal collection from Coimbra, Portugal: new aging standards for epiphyseal union. *Am. J. Phys. Anthropol.* **134**, 424–437 (2007).
- Petersen, A. In *The Oxford Handbook of the Archaeology of Death and Burial* (eds Stutz, L. N. & Tarlow, S.) Ch. 14 (Oxford Univ. Press, 2013).
- Green, R. E. et al. The Neandertal genome and ancient DNA authenticity. *EMBO J.* **28**, 2494–2502 (2009).
- Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
- Jun, G. et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848 (2012).
- Korneliusson, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinforma.* **15**, 356 (2014).
- Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
- Mittnik, A., Wang, C.-C., Svoboda, J. & Krause, J. A molecular approach to the sexing of the triple burial at the upper Paleolithic site of Dolní Věstonice. *PLoS ONE* **11**, e0163019 (2016).
- Pala, M. et al. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am. J. Hum. Genet.* **90**, 915–924 (2012).
- Gandini, F. et al. Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Sci. Rep.* **6**, 25472 (2016).
- El-Sibai, M. et al. Geographical structure of the Y-chromosomal genetic landscape of the Levant: A coastal-inland contrast. *Ann. Hum. Genet.* **73**, 568–581 (2009).

34. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
35. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
36. Prüfer, K. snpAD: an ancient DNA genotype caller. *Bioinformatics* **34**, 4165–4171 (2018).
37. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
38. Almarri, M. A. et al. The genomic history of the Middle East. *Cell* **184**, 4612–4625.e14 (2021).
39. Llorente, M. G. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820–822 (2015).
40. Fregel, R. et al. 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 (2018).
41. Markus, B., Alshafee, I. & Birk, O. Deciphering the fine-structure of tribal admixture in the Bedouin population using genomic data. *Heredity* **112**, 182 (2014).
42. Skoglund, P. et al. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* **344**, 747–750 (2014).
43. Thangaraj, K. et al. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr. Biol.* **13**, 86–93 (2003).
44. Günther, T. et al. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.* **16**, e2003703 (2018).
45. Günther, T. & Nettelblad, C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet.* **15**, e1008302 (2019).
46. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
47. Teebi, A. S. & Teebi, S. A. Genetic diversity among the Arabs. *Public Health Genomics* **8**, 21–26 (2005).
48. Tadmouri, G. O. et al. Consanguinity and reproductive health among Arabs. *Reprod. Health* **6**, 17 (2009).
49. Ségurel, L. & Bon, C. On the evolution of lactase persistence in humans. *Annu. Rev. Genomics Hum. Genet.* **18**, 297–319 (2017).
50. Enattah, N. S. et al. Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am. J. Hum. Genet.* **82**, 57–72 (2008).
51. Gregoricka, L. A. & Sheridan, S. G. Ascetic or affluent? Byzantine diet at the monastic community of St. Stephen's, Jerusalem from stable carbon and nitrogen isotopes. *J. Anthropol. Archaeol.* **32**, 63–73 (2013).
52. Haber, M. Isotopes data from ancient Lebanon. figshare <https://doi.org/10.6084/m9.figshare.12933293.v1> (2020).
53. Al-Bashaireh, K., Al-Shorman, A., Rose, J., Jull, A. J. T. & Hodgins, G. Paleo-diet reconstruction of human remains from the archaeological site of Natfieh, Northern Jordan. *Radiocarbon* **52**, 645–652 (2010).
54. Sandias, M. & Gundula, M. Diet and herding strategies in a changing environment: stable isotope analysis of Bronze Age and Late Antique skeletal remains from Ya'amün, Jordan. *J. Archaeol. Sci.* **63**, 24–32 (2015).
55. Soltysiak, A. & Schutkowski, H. Continuity and change in subsistence at Tell Barri, NE Syria. *J. Archaeol. Sci. Rep.* **2**, 176–185 (2015).
56. Gregoricka, L. A. & Judd, M. A. Isotopic evidence for diet among historic Bedouin of Khirbat al-Mudayna Jordan. *Int. J. Osteoarchaeol.* **26**, 705–715 (2016).
57. Smith, C. I., Fuller, B. T., Choy, K. & Richards, M. P. A three-phase liquid chromatographic method for  $\delta^{13}\text{C}$  analysis of amino acids from biological protein hydrolysates using liquid chromatography–isotope ratio mass spectrometry. *Anal. Biochem.* **390**, 165–172 (2009).
58. Mora, A. et al. High-resolution palaeodietary reconstruction: amino acid  $\delta^{13}\text{C}$  analysis of keratin from single hairs of mummified human individuals. *Quat. Int.* **436**, 96–113 (2017).
59. Mora, A., Pacheco, A., Roberts, C. & Smith, C. Pica 8: refining dietary reconstruction through amino acid  $\delta^{13}\text{C}$  analysis of tendon collagen and hair keratin. *J. Archaeol. Sci.* **93**, 94–109 (2018).
60. Honch, N. V., McCullagh, J. S. O. & Hedges, R. E. M. Variation of bone collagen amino acid  $\delta^{13}\text{C}$  values in archaeological humans and fauna with different dietary regimes: developing frameworks of dietary discrimination. *Am. J. Phys. Anthropol.* **148**, 495–511 (2012).
61. Mora, A., Pacheco, A., Roberts, C. & Smith, C. Palaeopathology and amino acid  $\delta^{13}\text{C}$  analysis: Investigating pre-Columbian individuals with tuberculosis at Pica 8, northern Chile (1050–500 BP). *J. Archaeol. Sci.* **129**, 105367 (2021).
62. Jørkov, M. L. S., Heinemeier, J. & Lynnerup, N. The petrous bone—A new sampling site for identifying early dietary patterns in stable isotopic studies. *Am. J. Phys. Anthropol.* **138**, 199–209 (2009).
63. López-Costas, O. & Alexander, M. Paleo-diet in the Iberian Peninsula: exploring the connections between diet, culture, disease and environment using isotopic and osteoarchaeological evidence. *Archaeol. Anthropol. Sci.* **11**, 3653–3664 (2019).
64. Dury, G. et al. The Islamic cemetery at 33 Bartomeu Vicent Ramon, Ibiza: investigating diet and mobility through light stable isotopes in bone collagen and tooth enamel. *Archaeol. Anthropol. Sci.* **11**, 3913–3930 (2019).
65. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).
66. Meyer, M. et al. Nuclear DNA sequences from the middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016).
67. Fages, A. et al. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* **177**, 1419–1435 (2019).
68. Orlando, L. et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
69. Decker, M. J. *Encyclopedia of Ancient History* (Wiley, 2013).
70. Donner, F. M. *The Early Islamic Conquests* (Princeton Univ. Press, 2014).
71. Sartre, M. *Bostra. Des origines à l'Islam* (P. Geuthner, 1985).
72. Conrad, L. I. Ta'un and waba' Conceptions of Plague and Pestilence in Early Islam. *J. Econ. Soc. Hist. Orient/Journal Hist. Econ. Soc. Orient.* **25**, 268–307 (1982).
73. Dols, M. W. Plague in early Islamic history. *J. Am. Orient. Soc.* **94**, 371–383 (1974).
74. Gleize, Y. et al. Early medieval Muslim graves in France: first archaeological, anthropological and palaeogenomic evidence. *PLoS ONE* **11**, e0148583 (2016).
75. Ibañez, J. J. et al. The early PPNB levels of Tell Qarassa North (Sweida, southern Syria). *Antiquity* **84**, 1–5 (2010).
76. Braemer, F. et al. Qarassa, un site du Bronze ancien et moyen en Syrie du Sud. Travaux 2007–2010. *Syr. Arch.éologie Art. Hist.* **88**, 225–250 (2011).
77. Godon, M., Baldi, J. S., Ghanem, G., Ibañez, J. J. & Braemer, F. Qarassa North Tell, southern Syria: the pottery Neolithic and Chalcolithic sequence. A few lights against a dark background. *Paléorient* **41**, 153–176 (2015).
78. Terradas, X., Ibañez, J. J., Braemer, F., Lionel, G. & Teira, L. The Natufian occupations of Qarassa 3 (Sweida, southern Syria). *Natufian Foragers Levant. Termin. Pleistocene Soc. Chang. West. Asia* **19**, 45 (2013).
79. Longin, R. New method of collagen extraction for radiocarbon dating. *Nature* **230**, 241–242 (1971).
80. Richards, M. P. & Hedges, R. E. M. Stable isotope evidence for similarities in the types of marine foods used by late mesolithic humans at sites along the Atlantic Coast of Europe. *J. Archaeol. Sci.* **26**, 717–722 (1999).
81. Ramsey, C. B. Development of the radiocarbon calibration program. *Radiocarbon* **43**, 355–363 (2001).
82. Reimer, P. J. et al. Selection and treatment of data for radiocarbon calibration: an update to the International Calibration (IntCal) criteria. *Radiocarbon* **55**, 1923–1945 (2013).
83. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
84. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, pdb-prot5448 (2010).
85. Günther, T. et al. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl Acad. Sci. USA* **112**, 11917–11922 (2015).
86. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
87. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
88. Kircher, M. Analysis of high-throughput ancient DNA sequencing data. *Methods Mol. Biol.* **840**, 197–228 (2012).
89. Vianello, D. et al. HAPLOFIND: A new method for high-throughput mt DNA haplogroup assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).
90. Kloss-Brandstätter, A. et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).
91. Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).
92. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
93. Allentoft, M. E. et al. Population genomics of bronze age Eurasia. *Nature* **522**, 167 (2015).
94. Mathieson, I. et al. The genomic history of southeastern Europe. *Nature* **555**, 197–203 (2018).
95. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
96. Schuenemann, V. J. et al. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat. Commun.* **8**, 15694 (2017).

97. Wang, K. et al. Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. *Sci. Adv.* **6**, eaaz0183 (2020).
98. de Barros Damgaard, P. et al. The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).
99. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
100. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
101. Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P. & Ramachandran, S. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016).
102. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207 (2015).
103. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl Acad. Sci. USA* **111**, 2229–2234 (2014).
104. Fu, Q. et al. The genetic history of ice age Europe. *Nature* **534**, 200–205 (2016).
105. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
106. Harney, É., Patterson, N., Reich, D. & Wakeley, J. Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* **217**, iyaa045 (2021).
107. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing indian population history. *Nature* **461**, 489 (2009).
108. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104 (2015).
109. Kuhn, J. M. M., Jakobsson, M. & Günther, T. Estimating genetic kin relationships in prehistoric populations. *PLoS ONE* **13**, e0195491 (2018).
110. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: online Mendelian inheritance in man (OMIM, an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43**, D789–D798 (2015).
111. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
112. Carriaso, M. & Lennon, G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res.* **40**, D1308–D1312 (2011).
113. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

## Acknowledgements

This work was funded by a grant from the Royal Physiographic Society of Lund (Nilsson-Ehle Endowments) to T.G. and C.V. and the La Trobe Internal Research grants to C.V. M.S. was part of the Erasmus Mundus Master Programme in Evolutionary Biology (MEME). C.V. is supported by a Ramón y Cajal grant (RYC2018-025223-I). T.G. is supported by a grant from the Swedish Research Council Vetenskapsrådet (2017-05267). M.J. and A.G. were supported by grants from the Swedish Research Council Vetenskapsrådet and the Knut and Alice Wallenberg foundation. J.S. is supported by a grant from Marie Skłodowska-Curie Actions (European Commission, no. GA 750460; H2020-MSCA-IF-2016) and a Ramón y Cajal grant (RYC2019-028346-I). The archaeological research at Tell Qarassa was supported by the Spanish Ministry of Economy and Competitiveness (grant HAR2016-74999-P) and the Palarq Foundation. The authors would like to thank Nikolay Oskolkov, Claudio Mirabello and Per Unneberg from the

National Bioinformatics Infrastructure of Sweden (NBIS) for their assistance with KrakenUniq, Pathogenome databases and metagenomic workflows. We thank Arielle R. Munters for initial data processing and Carolina Bernhardsson for assistance with the diploid genotype calls. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under projects sllstore2017020 and 2018/8-150. Shotgun sequencing was performed at the National Genomics Infrastructure (NGI) Stockholm. We also thank Phillip Edwards for valuable comments on the manuscript and Christina Warinner for discussion.

## Author contributions

T.G. and C.V. conceived the study. H.B., I.U. and C.V. performed sampling and ancient DNA lab work. C.S. and A.M. performed isotopic lab work and analysis. M.S., E.K., N.B. and T.G. performed analyses. J.S., E.I., A.P., K.A., F.B., J.J.I., M.J. and A.G. provided either archaeological material and background or input about genomic analyses. M.S., T.G. and C.V. wrote the manuscript with input from all co-authors.

## Funding

Open access funding provided by Uppsala University.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03508-4>.

**Correspondence** and requests for materials should be addressed to Torsten Günther or Cristina Valdiosera.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Luke R. Grinham.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022