Cell, Volume *181*

Supplemental Information

The Genomic History of the Bronze Age Southern Levant

Lily Agranat-Tamir, Shamam Waldman, Mario A.S. Martin, David Gokhman, Nadav Mishol, Tzilla Eshel, Olivia Cheronet, Nadin Rohland, Swapan Mallick, Nicole Adamski, Ann Marie Lawson, Matthew Mah, Megan Michel, Jonas Oppenheimer, Kristin Stewardson, Francesca Candilio, Denise Keating, Beatriz Gamarra, Shay Tzur, Mario Novak, Rachel Kalisher, Shlomit Bechar, Vered Eshed, Douglas J. Kennett, Marina Faerman, Naama Yahalom-Mack, Janet M. Monge, Yehuda Govrin, Yigal Erel, Benjamin Yakir, Ron Pinhasi, Shai Carmi, Israel Finkelstein, Liran Carmel, and David Reich

Methods S1. Details of experimental and computational techniques used. Related to STAR Methods.

A. Strontium analysis of the samples.

We analyzed petrous bones and tooth enamel from twelve individuals which successfully produced DNA results, as well as eight individuals from Megiddo where DNA sequencing was not successful. The twelve individuals – one from Abel Beth Maacah, nine from Megiddo, and two from Hazor yielded ^{87}Sr/^{86}Sr values (Table SA.1). All measurements characterize the geographic region in which the samples were found, i.e., northern Hulla valley and Jezreel valley (Hartman and Richards, 2014). Both areas are characterized by sediments and soils derived from three main sources: local Mesozoic carbonate rocks, local Tertiary basaltic rocks, and atmospheric dust mostly from the Sahara, with typical ^{87}Sr/^{86}Sr values ranging between 0.706 and 0.7083, depending on the relative contribution of the three sources. The sediment and soil ^{87}Sr/^{86}Sr values determine, in turn, the ^{87}Sr/^{86}Sr values of surface water and groundwater and the ^{87}Sr/^{86}Sr values reflecting the diet of the people living in this area. For four individuals, we could measure ^{87}Sr/^{86}Sr values in both enamel and petrous bone (one from Abel and three from Megiddo). ^{87}Sr/^{86}Sr values of petrous bones were found to be almost identical to those of the enamel, indicating that Sr in petrous bones was not significantly altered, and preserve the ^{87}Sr/^{86}Sr values obtained *in-vivo*. Moreover, the male sample from Abel Beth Maacah had a small but not significant difference between the 2nd Molar, 3rd Molar and petrous, suggesting that the person did not travel far during his life-time. In addition, eight other individuals from Megiddo that did not produce DNA had similar ^{87}Sr/^{86}Sr values, suggesting that the population in Megiddo was local, with no evidence of first-generation foreigners (people that spent their first decade of life, when tooth enamel forms and obtains its ^{87}Sr/^{86}Sr value, away from this area).

Site	SampleID	Sample archa. label	Type	Sub-type	Sr ppm	87Sr/86Sr
Abel Beth Maacah	I2201 (I3813)	L2521 B25195 a	petrous	left	248	0.7077
		L2521 B25195 b	petrous	right	270	0.7078
		L2521 B25195 c	enamel	2nd molar	72	0.7080
		L2521 B25195 d	enamel	3rd molar	100	0.7079
Megiddo	I2189	2014/K/066 LB005	petrous	left	293	0.7079
	I2190	2012/K/089 LB015	petrous	right	323	0.7078
	I2195	2012/K/057 PT002 LB004	petrous	left	380	0.7079
	I2198	2014/K/049 PT007 LB029	petrous	left	371	0.7079
	I2200	2012/K/107 PT005 LB005 a	petrous	right	380	0.7079
		2012/K/107 PT005 LB005 b	enamel	unknown	82	0.7080
	I4519	2014/K/159 PT001 LB003 a	petrous	left	325	0.7082
		2014/K/159 PT001 LB003 b	enamel	1st upper molar	1393	0.7081
	I4517	1998/K/100 PT006 LB007 a	petrous	right	449	0.7083
		1998/K/100 PT006 LB007 b	enamel	1st lower molar	98	0.7081
	I4525	2010/K/106 LB007 #2	petrous	left	462	0.7082
	I4521	2006/J/008 PT023 LB003	petrous	left	363	0.7081

Hazor	I3965	L7953 B48846/1	petrous	unknown	268	0.7078
	I3966	L12-371 B78119	petrous	left	383	0.7078

Table SA.1. *Sr concentrations and isotopic values for individuals whose genome was sequenced.*

B. Sensitivity of analyses to down-sampling.

In the main text, we only used samples with a minimum autosomal SNP coverage of 30,000. To justify this threshold, we perform here down-sampling of the data, and test its effects on the different methods used in the manuscript.

For PCA, we down-sampled 16 genomes to 1K, 2.5K, 5K, 8K, 10K, 25K 50K, 250K, and 500K SNPs, with 20 replicates for each down-sampling level. All were projected onto the West Eurasian PCA, as in Figure 1B. The standard deviation of the PCA position was plotted against the down-sampling level (Figure SB.1). As expected, the standard deviation decreases with the number of SNPs and reaches ~0.005 at 30K SNPs.

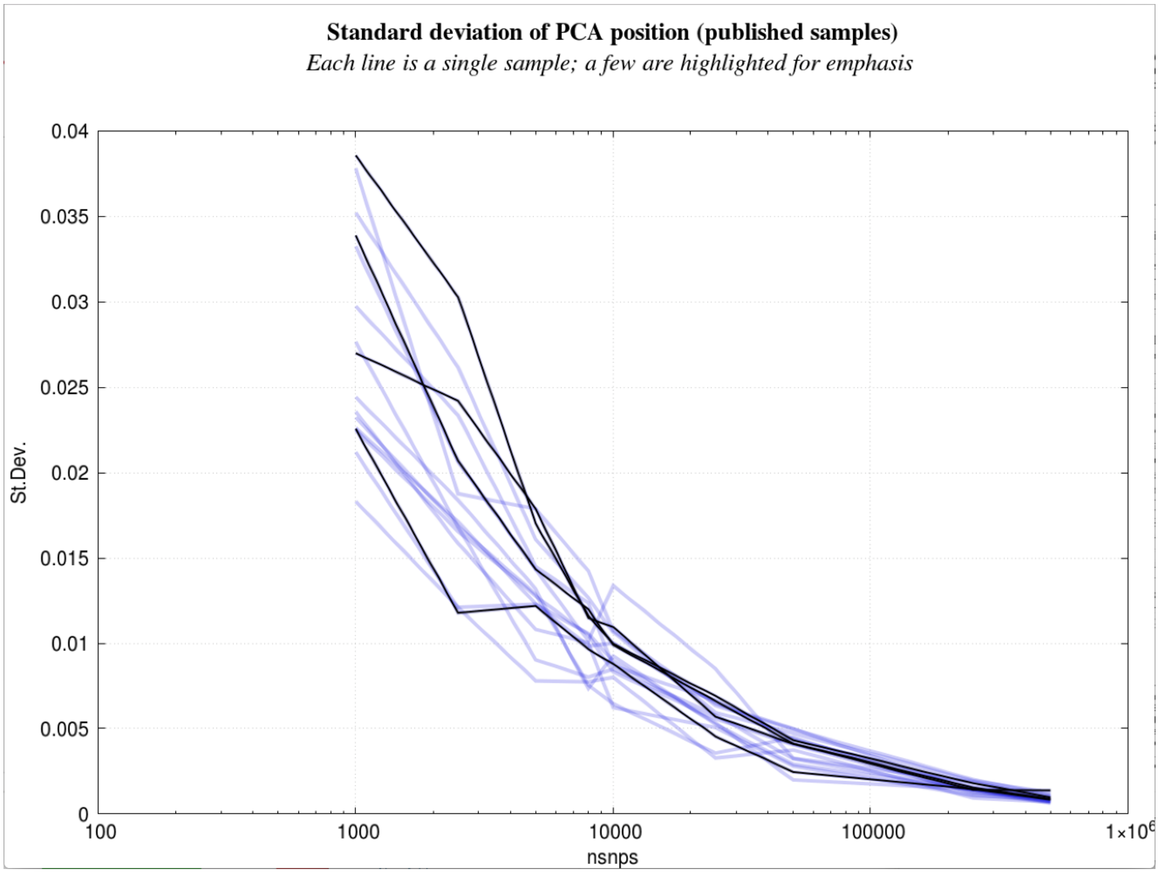

Figure SB.1. Standard deviation of PCA position. Each line is a single sample; a few are highlighted for emphasis.

For ADMIXTURE, we down-sampled the entire data set to ~50K SNPs. Due to the variability in data quality across sites, the resulting mean coverage varied across the different Levant populations, ranging from ~5,000 to ~37,000 SNPs. The qualitative results of ADMIXTURE did not change (Figure SB.2), with both outliers from Megiddo presenting an additional significant source. Populations analyzed with the reduced SNP set tend to show a small contribution of ancestor populations not inferred to be contributors in the extended SNP set. However, as we show next, this small contribution is present in all populations, and therefore hardly affects comparisons of the ADMIXTURE vectors in the different populations.

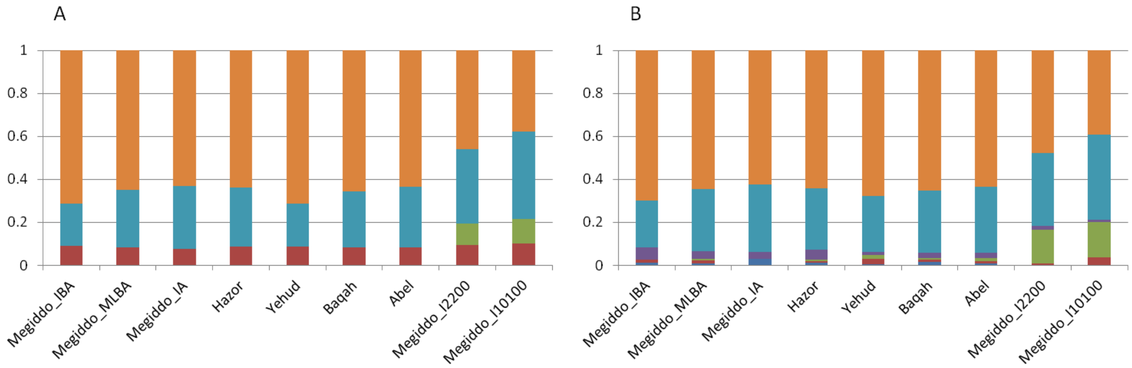

Figure SB.2*. ADMIXTURE plots for the populations first analyzed here, using either a set of 357,334 (A) or 50,165 (B) SNPs. For groups with more than one individual (Megiddo_IBA, Megiddo_MLBA, Hazor, Yehud and Baqah),* the *average across individuals is shown.*

For LINADMIX, which is a downstream analysis of ADMIXTURE, we examined the mean absolute difference in the Levant_N ancestry, in the models appearing in Table S2, before and after down-sampling. We found that when the number of SNPs exceeds 20K, the mean absolute difference in the Levant_N ancestry becomes ~5%, equal to the standard error of the model as computed by LINADMIX (Table S3, Figure SB.3). In addition, in all populations except for Tel Shadud, the change in the Levant_N ancestry was in the same direction, thus maintaining the rank of Levant_N ancestry in the different populations.

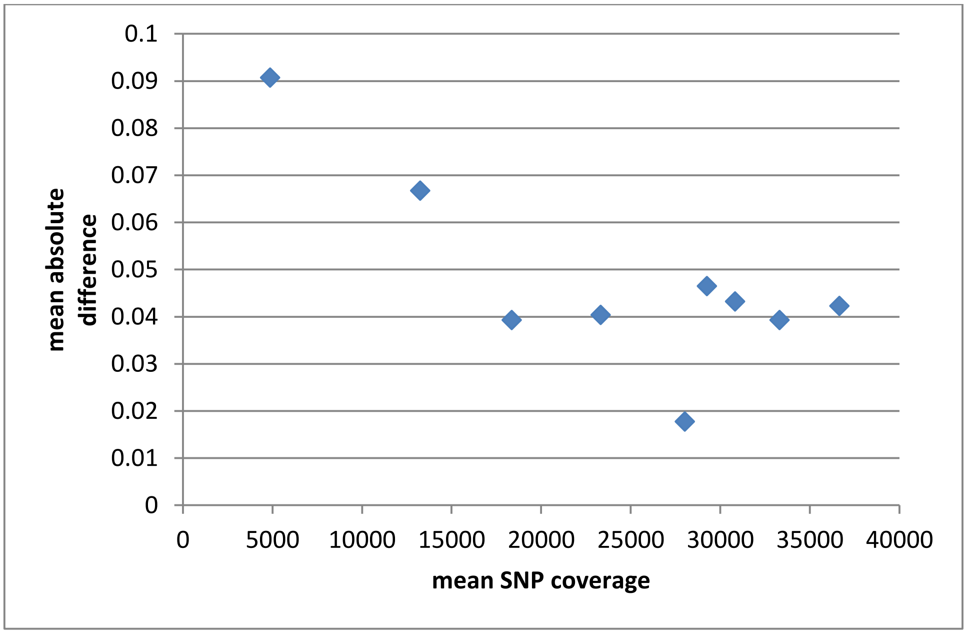

Figure SB.3*. Change in Levant_N ancestry proportions estimated by LINADMIX as a function of SNP coverage. The plot shows the mean absolute difference in Levant_N ancestry proportion calculated by LINADMIX after down-sampling, versus the mean SNP coverage of the population.*

For *qpAdm*, which takes the union of SNPs in a population, we down-sampled the entire data set to ~46K, ~33K and ~15K SNPs. We then examined the mean absolute difference of Levant_N ancestry in Megiddo_MLBA and in Baqah in the models of Table S2 before and after down-sampling (Figure SB.4). At 30K SNPs, the mean absolute difference (3.3%) decreases to approximately the standard error of the model (3%) and continues to decrease as the number of SNPs used increases.

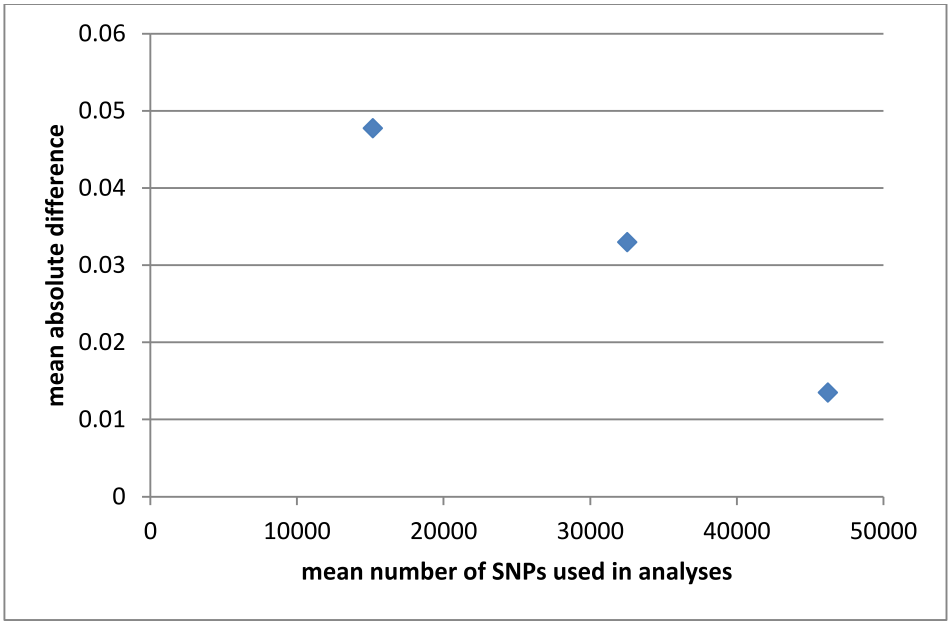

***Figure SB.4**. The mean absolute difference in Levant_N ancestry proportion calculated by qpAdm after down-sampling.*

The autosomal SNP coverage threshold for PHCP was different from the other methods, and was set to 100K SNPs throughout the manuscript. This threshold was determined using simulations, as detailed in Methods S1H.

C. Testing the robustness of the relative genomic remoteness of Sidon

Our analyses in the section "High degree of genetic homogeneity across multiple sites" showed that Sidon is relatively distinct from other contemporary populations, and resembles only Ashkelon, and to a lesser extent Abel. However, Sidon is also the only population sequenced in a different lab using a different sequencing technology. To test whether the results for Sidon are due to a batch effect, we repeated the *qpWave* analysis using SNPs that represent transversions, which are less prone to characteristic ancient DNA errors (Wang et al., 2015). Using only the 111,208 transversion SNPs, we saw that Sidon does cluster with some other Bronze and Iron Age Levant populations (Table SC.1).

Left populations	Outgroup (*right*) populations	P value	Outgroup (*right*) populations	P value
Sidon ASH IA1	o9a	0.0488	o9aamcn	0.0353
Sidon ASH IA2	o9a	0.4312	o9aamcn	0.3200
Sidon ASH LBA	o9a	0.8942	o9aamcn	0.7392
Sidon Hazor	o9a	0.1512	o9aamcn	0.1291
Sidon Abel	o9a	0.1443	o9aamcn	0.2275
Sidon Megiddo IA	o9a	0.3563	o9aamcn	0.3093
Sidon Yehud	o9a	0.0004	o9aamcn	0.0000
Sidon Megiddo I2200	o9a	0.0354	o9aamcn	0.0901
Sidon Megiddo MLBA	o9a	0.0017	o9aamcn	0.0005
Sidon Megiddo I10100	o9a	0.0290	o9aamcn	0.0495
Sidon Baqah	o9a	0.0013	o9aamcn	0.0012
Sidon ‘Ain Ghazal	o9a	0.0059	o9aamcn	0.0003
Sidon Levant N	o9a	0.0000	o9aamcn	0.0000
Sidon Megiddo IBA	o9a	0.2085	o9aamcn	0.1844
Sidon Shadud	o9a	0.2105	o9aamcn	0.1031

***Table SC.1**. qpWave results on transversion-SNPs testing Sidon and another Levant population (as left populations) using either o9a or o9aamn as the outgroups (right populations). o9a: Ust_Ishim; Kostenki14; MA1; Han; Papuan; Onge; Chukchi; Karitiana; Mbuti; Anatolia_N. o9aamcn: Ust_Ishim; Kostenki14; MA1; Han; Papuan; Onge; Chukchi;*

Karitiana; Mbuti;Anatolia_N; Armenia_MLBA; CHG; Natufian. In cyan are pairs that are compatible with being clades with each other relative to the outgroup populations.

However, these results could also be a consequence of the reduced number of SNPs used in the analysis. To test this, we randomly pruned the data to a different set of 112,392 SNPs. The qpWave results (Table SC.2) resemble those obtained for the transversion-SNPs, suggesting that transversion-SNPs do not carry a systematically different signal, and therefore the results are not showing evidence for being influenced by a batch effect.

Left populations	Outgroup (*right*) populations	P value	Outgroup (*right*) populations	P value
Sidon ASH IA1	o9a	0.0017	o9aamcn	0.0005
Sidon ASH IA2	o9a	0.4213	o9aamcn	0.3261
Sidon ASH LBA	o9a	0.5261	o9aamcn	0.7467
Sidon Hazor	o9a	0.2212	o9aamcn	0.3489
Sidon Abel	o9a	0.1726	o9aamcn	0.3277
Sidon Megiddo IA	o9a	0.1023	o9aamcn	0.0750
Sidon Yehud	o9a	0.0518	o9aamcn	0.0206
Sidon Megiddo I2200	o9a	0.0971	o9aamcn	0.2024
Sidon Megiddo MLBA	o9a	0.0063	o9aamcn	0.0000
Sidon Megiddo I10100	o9a	0.0002	o9aamcn	0.0000
Sidon Baqah	o9a	0.0000	o9aamcn	0.0000
Sidon ‘Ain Ghazal	o9a	0.0251	o9aamcn	0.0223
Sidon Levant N	o9a	0.0000	o9aamcn	0.0000
Sidon Megiddo IBA	o9a	0.0034	o9aamcn	0.0021
Sidon Shadud	o9a	0.0911	o9aamcn	0.0939

Table SC.2. qpWave *results on a random subset of 112,392 SNPs with Sidon and another Levant population (as left populations) using either o9a or o9aamn as the outgroups (right populations). o9a and o9aamcn composition is as in Table SC.1. In cyan are pairs that are compatible with being clades with each other with respect to the outgroup populations.*

To better understand the source for the genetic distinctiveness of the Sidon population, we carried out *qpWave* on each of the five Sidon individuals separately (on all SNPs). The results show that while some of the Sidon individuals form a clade with various Bronze and Iron Age Levant populations, the only populations that form a clade with all Sidon individuals are Ashkelon_LBA, Ashkelon_IA2 and Abel. These are precisely the three populations that emerged as most closely resembling the Sidon population as a whole (Table SC.3).

Population	Individuals passing p value ≥ 0.05 with o9a	Individuals passing p value ≥ 0.05 with o9aamcn
ASH IA1	None	None
ASH IA2	All	All
ASH LBA	All	All
Hazor	1,2	1,2
Abel	All	All
Megiddo IA	1,2,5	1,2,5
Yehud	1	1
Megiddo I2200	1,2	1,2
Megiddo MLBA	None	None
Megiddo I10100	None	None
Baqah	2	None
‘Ain Ghazal	None	None
Levant N	None	None
Megiddo IBA	1	None
Shadud	1,2,3,4	1,2,5

Table SC.3. qpWave *results for the different Sidon individuals and another Levant population (as left populations) compared to either o9a or o9aamn outgroup populations (right populations). Listed are the individuals that are compatible with forming a clade with the tested population. Sidon individuals: ERS1790729 – 1; ERS1790730 – 2; ERS1790731 – 3; ERS1790732 – 4; ERS1790733 – 5. o9a and o9aamcn composition is as in Table SC.1.*

We conclude that the Sidon population has evidence of heterogeneity, with some of its individuals being closer to inland Canaanites while others are less so. Nonetheless, the population as a whole forms a clade with the Ashkelon_LBA, Ashkelon_IA2 and Abel.

D. Choosing an outgroups set for *qpAdm*

When using the basic set of nine outgroups (o9) (Lazaridis et al., 2016) in the *qpAdm* analysis (Ust_Ishim, Kostenki14, MA1, Han, Papuan, Onge, Chukchi, Karitiana, Mbuti) we obtained valid models for all target populations as mixtures of Neolithic Levant (Levant_N) and Chalcolithic Iran (Iran_ChL). Similarly, most target populations showed support for being a mixture of Neolithic Levant and Early Bronze Age Armenia (Armenia_EBA; Table S2). However, analysis of Chalcolithic populations from the southern Levant found genetic evidence for influence from Anatolia (Harney et al., 2018). In order to take into account such possibility, we added Anatolia to the set of outgroups, calling this set o9a. For this set of outgroups, most target populations still show support for being a mixture of Levant_N and Iran_ChL/Armenia_EBA, with the exception of Hazor (p-values just below 0.05) and Baq'ah. However, we noticed that the addition of Anatolia to the set of outgroups significantly improved the model by reducing the standard errors of the mixing coefficients by 3-fold (Iran_ChL) and 2.1-fold (Armenia_EBA) (Figure SD.1). We therefore decided to use the o9a set for all analyses.

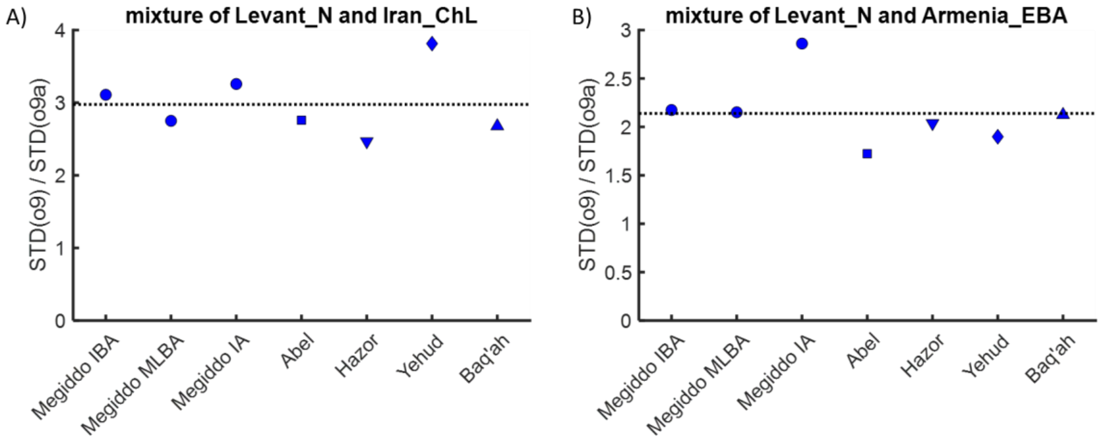

Figure SD.1. *Ratio of standard errors of the models using the o9 set of outgroups versus using the o9a set. (A) For a mixture of Levant_N and Iran_ChL. (B) For a mixture of Levant_N and Armenia_EBA. Dashed lines represent average ratios.*

E. Comparing Iran_ChL to Armenia_EBA ancestry

We modeled the Bronze and Iron Age Levant populations as a two-way admixture of Neolithic Levant (Levant_N) with either Chalcolithic Iran (Iran_ChL) or Early Bronze Age Armenia (Armenia_EBA). The *qpAdm* algorithm provides equivalent support for both models, with some populations slightly and insignificantly favoring either Iran_ChL or Armenia_EBA (Table S2).

The LINADMIX algorithm gave almost identical model norms to both models, but those of Iran_ChL were consistently lower (Table S3). To evaluate the magnitude of this difference, we repeated the computations for two additional admixture models of Levant_N with either Neolithic Iran (Iran_N) or Chalcolithic Armenia (Armenia_ChL). Whereas all four possible source populations are genetically similar, the results demonstrate that the model is sensitive enough such that the fit is not as good with Iran_N, and worst with Armenia_ChL (Table 3, Figure SE.1).

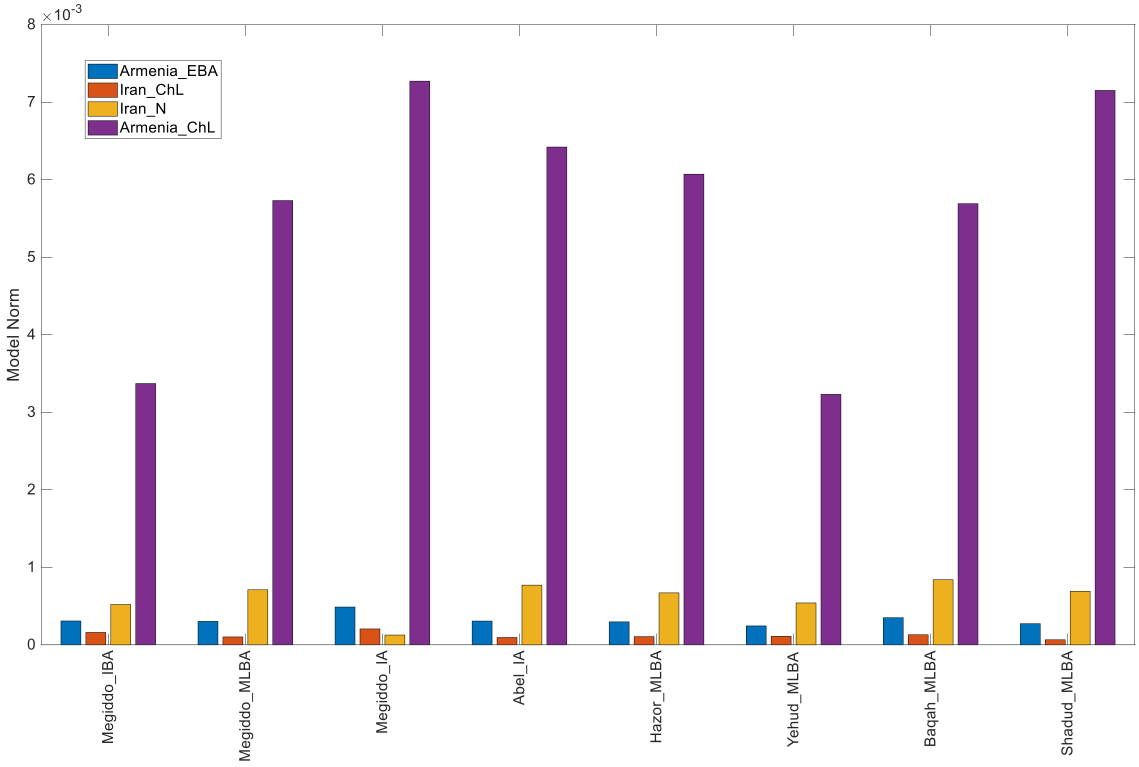

***Figure SE.1**. LINADMIX model norms when testing four models of admixture: Levant_N + either Armenia_EBA, Iran_ChL, Iran_N and Armenia_ChL.*

We therefore conclude that with the exception of the Megiddo outliers, all Bronze and Iron Age Levant populations studied here can be modeled equally well as a mixture of earlier local Neolithic populations (Levant_N) with populations related to Chalcolithic Iran or early Bronze Age Armenia.

F. PHCP parameters and evaluation

An implicit assumption in our model is that there is no substantial genetic divergence between each of the assumed source populations and the true source populations. We also assume no genetic drift in the target population since admixture. While *qpAdm* is theoretically robust to deviations from these assumptions under some scenarios, our copying profiles may be influenced. For example, recent genetic drift in the target population may increase the frequency of haplotypes coming from one source at the expense of other sources, which may affect the ancestry proportions $\beta_1, \dots, \beta_S$. However, we expect that when averaging over numerous genomic regions, changes in frequency will average out.

In an initial evaluation of the method we observed several issues. First, as we considered fewer and fewer SNPs (to simulate ancient DNA), the total length of genetic material copied from each donor population became nearly the same, making the copying vectors less informative. This also created a problem when samples with different numbers of SNPs were used in the same linear model. Second, when there were two source populations that are genetically close, even when running the method on samples with the same number of SNPs, the results from the linear model were not sufficiently accurate.

To address the dependence of the results on the number of SNPs, we down-sampled all genomes (both target and sources) to the same number of SNPs. When a target or source population was modern, we also reduced each SNP to a single random allele. An important drawback of down-sampling is that it makes inference at the individual level noisy, and thus we recommend that the results are interpreted only at the population level. Second, we noted that for high levels of missing data, the HMM tends to give nearly equal probabilities to copying from each donor group, despite the fact that the larger distances between SNPs are incorporated into the model. To guarantee that copying vectors are less uniform (and hence more informative), we reduced N_e (the effective population size) by tenfold relative to its original value inferred by ChromoPainter's EM algorithm on modern Eurasian and African populations. The reduced N_e decreased the probability of recombination, and thus forced the HMM to “choose” between each the donor populations. This, in turn, led to larger differences between the total amount of genetic material copied from each donor population (Figure SF.1).

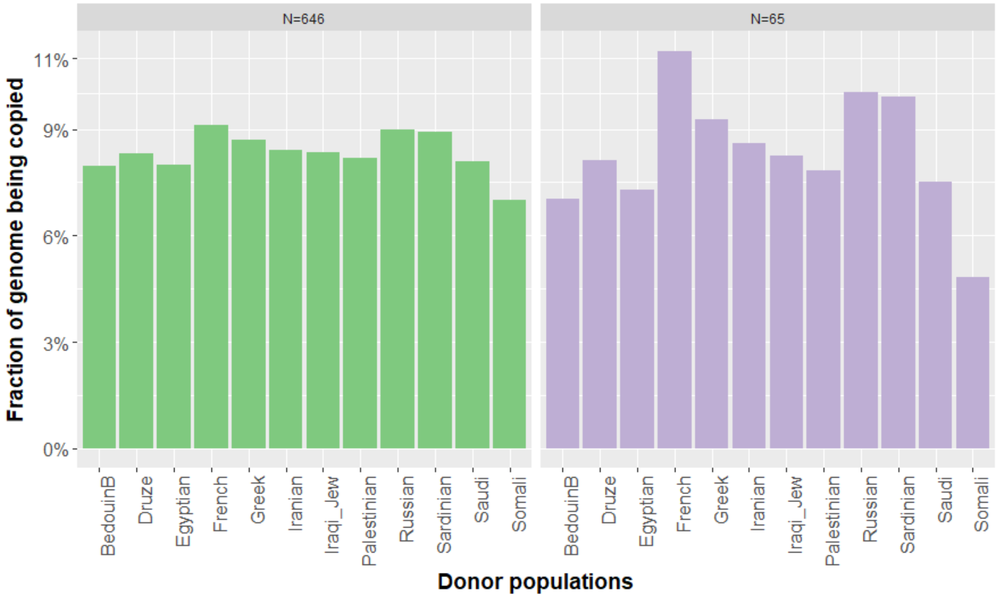

Figure SF.1. *The proportion of the genetic material copied from each of the donor populations by 3 simulated ancient Spanish genomes, with different effective population size parameter (N). Reducing N led to larger differences between the fraction of genetic material copied from each donor population.*

To improve accuracy, we noted that the ancestry estimates became more accurate when we chose the most informative donor populations instead of using a predefined set of populations. Formally, let D be a set of all donor populations and let $\vec{X}_s$, $s = 1{,}2$, be the $|D|$-element copying vector of source population s (assuming two sources), as inferred by PHCP. For each source, we ranked the donor populations according to their contribution (in the copying profile) to the source. Then, we scored each donor population based on the difference in the rank of the donor population between the two sources. A large rank difference of a donor population means that the sources are distinct in the relative amount of genetic material copied from that donor population, and thus, that the donor population is expected to be informative when estimating the contribution of each source to the target population. We then retained the twelve donor populations with the highest rank differences. Once the final set of donor populations was selected, we re-ran PHCP and the linear model with these donors, and inferred the ancestry proportions of the target population as described in Methods.

We also considered cases of three or four source populations, for two source populations that are relatively close, and other sources being more diverged. For each of the more distant source populations, we included one donor population that is closer to that source than to any other source. The remaining donor populations were chosen to differentiate between the two close source populations as described above. In all cases, the total number of donor populations was twelve.

The models that describe the contribution of ancient populations to present-day populations included four source populations - two closely related populations, Iran_ChL and Megiddo_MLBA, and two distant populations, Europe_LNBA and Somali. Hence, we chose one

European donor population, one East-African donor population, and the remaining ten donor populations were optimized to distinguish between Iran_ChL and Megiddo_MLBA. The final set of donor populations included some populations that were also target populations we wanted to model. In these cases, we dropped the target population from the donors and painted the sources and the target with the remaining eleven donor populations.

G. PHCP results

Testing PHCP on ancient target populations

Applying PHCP to the Bronze and Iron Age populations of the Southern Levant, we obtained results that are not statistically significantly different from those of LINADMIX (Figure S3C). We also noticed that when the other source population was selected as Armenia_EBA instead of Iran_ChL, the contribution of Levant_N slightly increased (Table S3).

We ran PHCP only on individuals with at least ≈95 k SNPs and only on target populations with at least three individuals. Three populations met the inclusion criteria: Megiddo_MLBA, Hazor_MLBA and Baqah (Table S3).

Testing PHCP on present-day target populations

PHCP was also applied to the 17 present-day populations that were analyzed by LINADMIX (Table S4, Figure SG.1). We used the same model as in LINADMIX, of four source populations (Megiddo_MLBA, Iran_ChL, Somali and Europe_LNBA). As in the previous section, we ran PHCP only on individuals with at least ≈95 k SNPs.

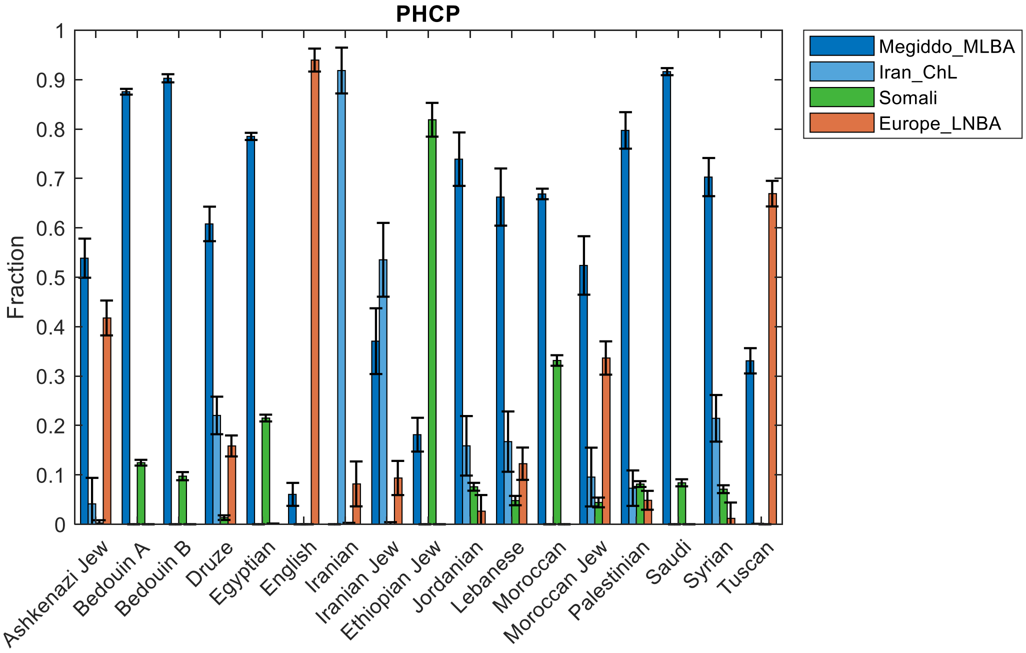

***Figure SG.1**. Genetic makeup of present-day populations, modeled by PHCP as an admixture of four source populations.*

PHCP and LINADMIX agree on the European and Somali contributions to the present-day populations (Figure S4A), and hence also on the combined contribution of Iran_ChL and Megiddo_MLBA (Figure SG.2).

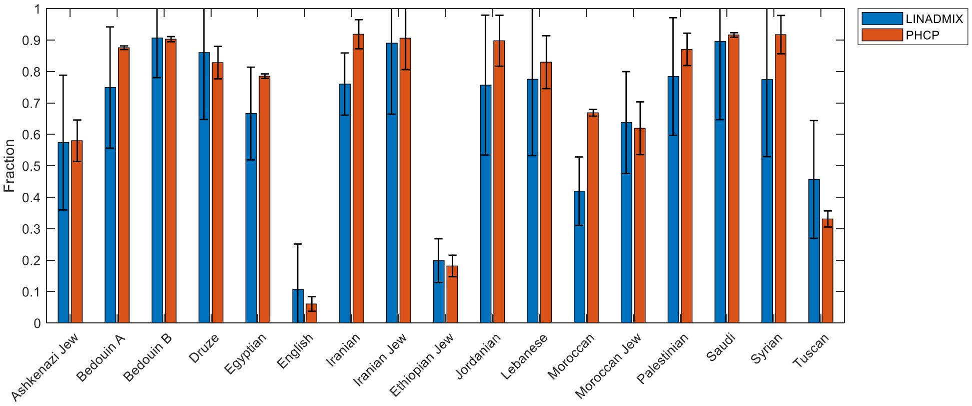

Figure SG.2. *Estimation of the combined fraction of Iran_ChL and Megiddo_MLBA by LINADMIX and PHCP*

However, the two methods differ in their estimates of the relative contributions of Megiddo and Iran, in that LINADMIX consistently gives a higher fraction of the Iranian component and a lower fraction of the Megiddo component relative to PHCP (Figures 3, S4A). In the main text we thus only report results when considering Megiddo and Iran as a single source.

H. Simulations for LINADMIX and PHCP

To test the overall accuracy of LINADMIX and PHCP in the setting of the current manuscript, we used simulations where we mixed two or three source populations. For the case of two source populations, we mixed present-day Jordanians and Iranians in order to mimic Megiddo_MLBA and Iran_ChL, which are the two most closely related source populations in our models. We used phased present-day Jordanian and Iranian genomes to simulate three admixture settings: (1) 20% Jordanian and 80% Iranian; (2) 80% Jordanian and 20% Iranian; (3) 50% Jordanian and 50% Iranian. In the simulations, the ancestry of each chromosomal segment was randomly selected based on the ancestry proportions, and the genotypes along the segment were copied from the selected source by randomly choosing a chromosome. The lengths of the segments were exponentially distributed with a rate of 0.5 per cM, representing an admixture event roughly 50 generations ago. Diploid individuals were constructed by pairing two sets of haploid chromosomes. For each admixture setting, five individuals were simulated. The resulting simulated genomes represented "present-day" genomes. To simulate the "ancient" sources of those genomes, we used other genomes (that were not used for simulating the mixed individuals) from the same source populations. In LINADMIX, those genomes were down-sampled to 60% of the SNPs (which is the average of Megiddo_MLBA in the analysis) and a random allele was selected at each SNP. In PHCP, the "source" genomes were down-sampled to 100K SNPs and a random allele was selected at each SNP. We assumed that neither the source populations nor the admixed population have undergone drift since admixture. We then reconstructed the ancestry proportions of the simulated genomes using the same pipeline as used for the real data, and compared the results to the true (known) proportions. The inferred ancestry proportions for both LINADMIX and PHCP were very close to the simulated ancestry proportions (Figure SH.1). LINADMIX had a mean deviation (of the 3 simulations) of 1%, and a maximal deviation (of the 3 simulations) of 3%. In PHCP, the mean deviation was 5%, and the maximal deviation was 8%.

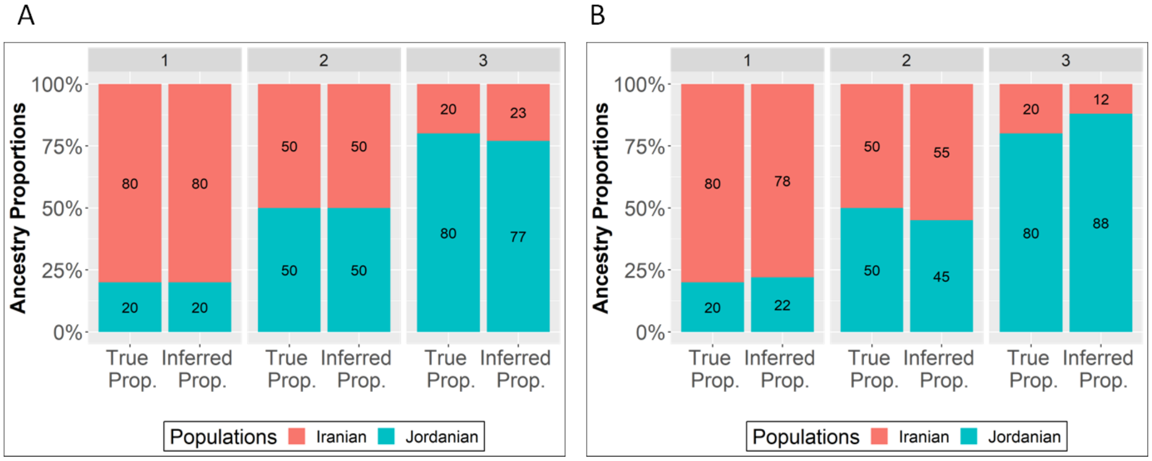

Figure SH.1. *Simulation of two source populations – Jordanians and Iranian – with varying ancestry proportions. The true and the inferred percentages are shown, averaging in each case over the five simulated admixed individuals. (A) LINADMIX. (B) PHCP.*

For the case of three source populations, we simulated a mixture of Jordanians, Iranians, and English, with three admixture configurations: (1) 60% - 30% - 10%; (2) 30% - 10% - 60%; (3) 10% - 60% - 30%. We assumed that admixture between all three populations has happened 50 generations ago. The inferred ancestry proportions are close to the simulated ancestry proportions, although, as expected, they are less accurate than in the case of two source populations (Figure SH.2). In LINADMIX, the mean deviation was 4.2% and the maximal deviation was 10%. In PHCP, the mean deviation was 5.7% and the maximal deviation was 10%.

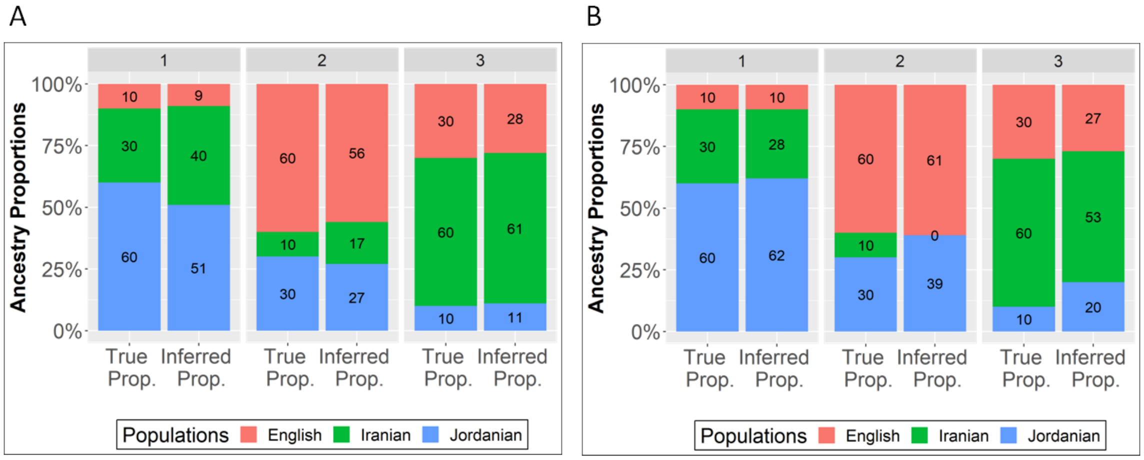

***Figure SH.2**. Simulation of three source populations – Jordanian, Iranian and English – with varying ancestry proportions. The true and the inferred percentages are shown, averaging in each case over the five simulated admixed individuals. (A) LINADMIX. (B) PHCP.*

I. LINADMIX analysis of the admixture of present-day populations

Arabic-speaking populations

We considered ten Arabic-speaking populations (Figure SI.1) – two Bedouin groupings, Druze, Egyptian, Jordanian, Lebanese, Palestinian, Saudi, Syrian and Moroccan (which served as a control). All, apart from the Druze who are also religiously distinct, have a non-negligible Somali-related component. This component seems to decrease with the distance from East Africa: it is highest in Egyptians and lowest in Lebanese and Syrians. All non-African populations we looked at – Iranians and Europeans – show little if any Somali-related component. Moroccans were not modeled well by LINADMIX, likely because of a missing source population. Nevertheless, the African-related component in Moroccans is high. Many of the Arabic-speaking populations show a significant Iranian-related component. This component could have arrived from several sources, such as deportees from the Zagros mountains who were settled in the Southern Levant by the Assyrians, the Achaemenid (Persian) Empire that ruled the region from the late 6th to the late 4th centuries BCE, and Sasanian political control in the early 7th century CE. The two conquests to which we refer took over the entire Levant, but did not extend into the heartland of the Arabian Peninsula or far beyond the northern parts of Egypt. In accord with the expectation from these scenarios, Moroccans do not have a detectable Iranian-related component, whereas the Arabic-speaking populations with the smallest Iranian-related components are Saudis, Egyptians and Bedouin B, at the fringes of these Persian empires. At the same time, as noted in the main text, the contribution of the Iranian-related component relative to Megiddo is not well modelled by our algorithms, and is subject to large standard errors. Interpretation of the magnitude of the Iranian component should therefore be taken with caution.

Notably, the two populations with the lowest European and Iranian components, Bedouin B and Saudi, seem to be the least admixed populations, possibly a result of past arid habitat, pastoral lifestyle and geographic isolation.

Culturally Jewish populations

We examined four Jewish populations that come from different regions of the world (Figure SI.1). All show a sizeable Southern Levant Bronze Age-related component, whereas the other components reflect their individual histories. Ashkenazi Jews, coming from Europe, have a substantial European-related component as already seen in previous studies (Atzmon et al., 2010; Carmi et al., 2014). Moroccan Jews, who are thought to be partially descendants of the Jews who migrated from the Iberian peninsula prior to and after the expulsion from Spain in 1492 (Schroeter, 2008), have a substantial European-related component and a minor African one. A more appropriate model should include local Berber populations as an additional source. Iranian Jews have a majority of Iranian-related ancestry, while Ethiopian Jews have a majority of Somali-related ancestry.

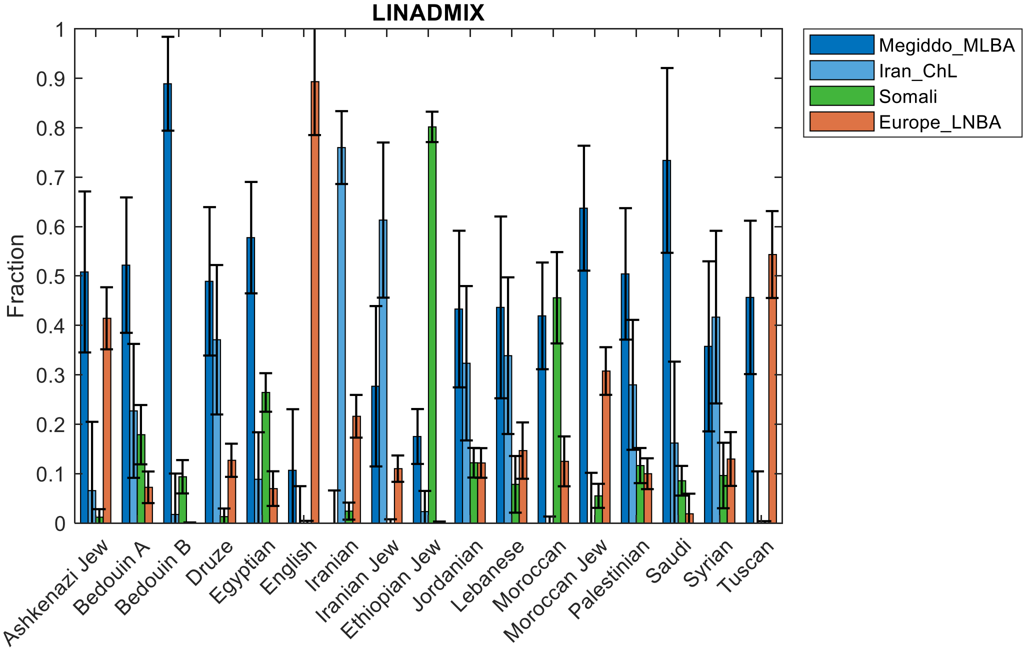

***Figure SI.1**. Genetic makeup of present-day populations, modeled by LINADMIX as an admixture of four source populations.*

J. Robustness with respect to changing the representative population of the Bronze Age Levant

When modeling present-day Levantine groups as mixtures of ancient populations, we selected Megiddo_MLBA as the Levantine ancient representative. However, as these individuals span a wide time range, we decided to repeat the analysis by selecting only a subset of individuals with low genetic variability.

To this end, we applied k-means clustering to the ADMIXTURE q-vectors of the Megiddo_MLBA individuals. We ran the algorithm for $k = 1, \dots, 12$. For each k, we measured the performance of the clustering by computing the sum of distances of all points from their clusters' centroids. To select the optimal k, we plotted this distance as a function of k and looked for an 'elbow' in the graph (Figure SJ.1).

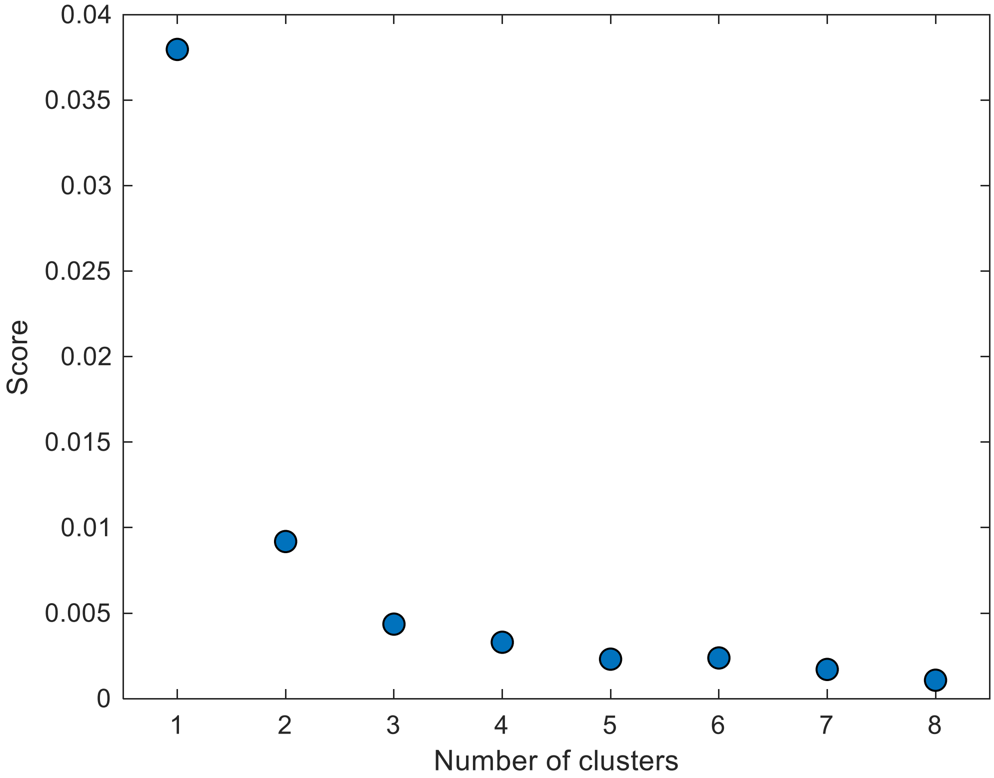

Figure SJ.1. *Sum of distances of all points from their clusters' centroids versus the number of clusters, k.*

We found that good clustering is achieved for $k = 3$, dividing the Megiddo individuals by the magnitude of their Levant_N component into three subpopulations, denoted here Megiddo_High (11 individuals), Megiddo_Medium (8 individuals) and Megiddo_Low (3 individuals; Figure SJ.2).

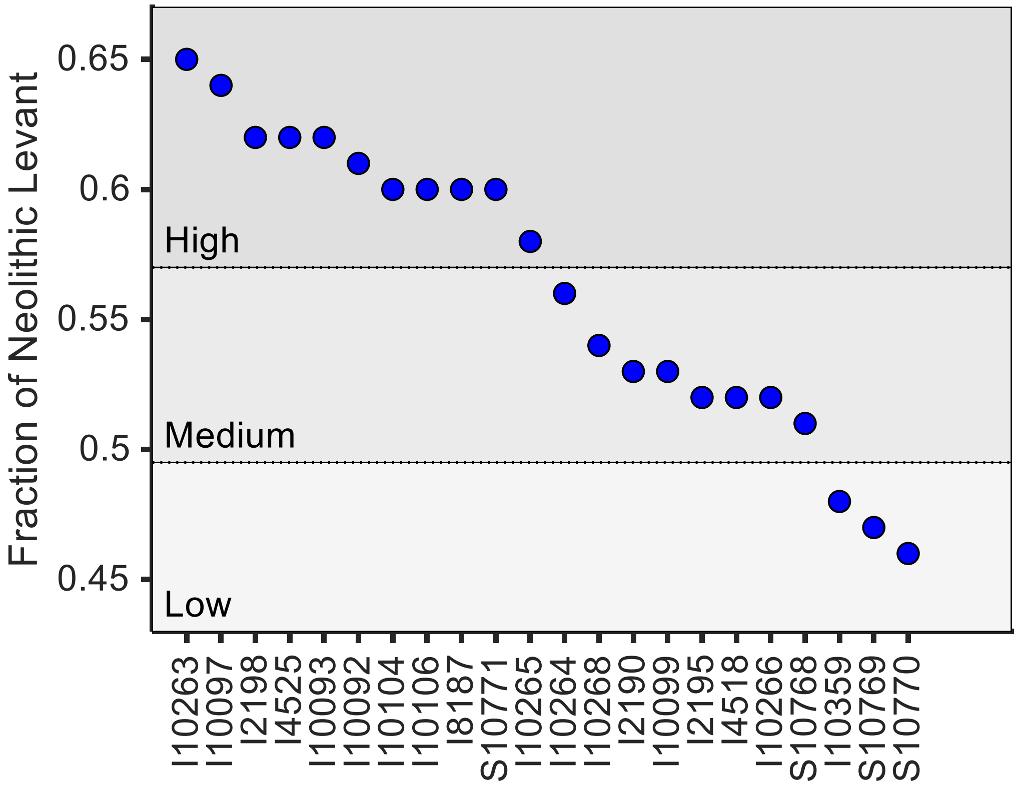

Figure SJ.2*. The fraction of Levant_N in the genomes of the Megiddo_MLBA individuals. Horizontal black lines show the clustering of the individuals to three clusters.*

Using *qpAdm* to test models of admixture between Levant_N and either Iran_ChL or Armenia EBA, and using the o9 set of outgroups, Megiddo_MLBA and the three subpopulations give comparable models. All are supported by high p-values (Table S2). When using the o9a set of outgroups, Megiddo_MLBA was weakly supported by Armenia_EBA and Iran_ChL ($P = 0.065$ and 0.043, respectively). Megiddo_High and Megiddo_Medium show support for both models ($P = 0.165$ and 0.067 for Megiddo_High, and $P = 0.277$ and 0.822 for Megiddo_Medium). Megiddo_Low has a weak support for Iran_ChL with $P = 0.036$, and a bit stronger support for Armenia_EBA model ($P = 0.125$).

In LINADMIX, the division into three subpopulations clearly improved the models, reducing the standard error from 8% to 5% (Tables S4 and S5).

Whereas Megiddo_Low has the lowest fraction of the local genetic component, we nevertheless selected Megiddo_Medium as our representative of the Bronze Age Levant source population, because of the higher number of individuals, and because two out of the three Megiddo_Low individuals were found near the royal palace and not in the residential quarters, and hence may be atypical. We repeated the analyses replacing Megiddo_MLBA with Megiddo_Medium, and obtained qualitatively identical results (Tables S4-S5, Figure SJ.3).

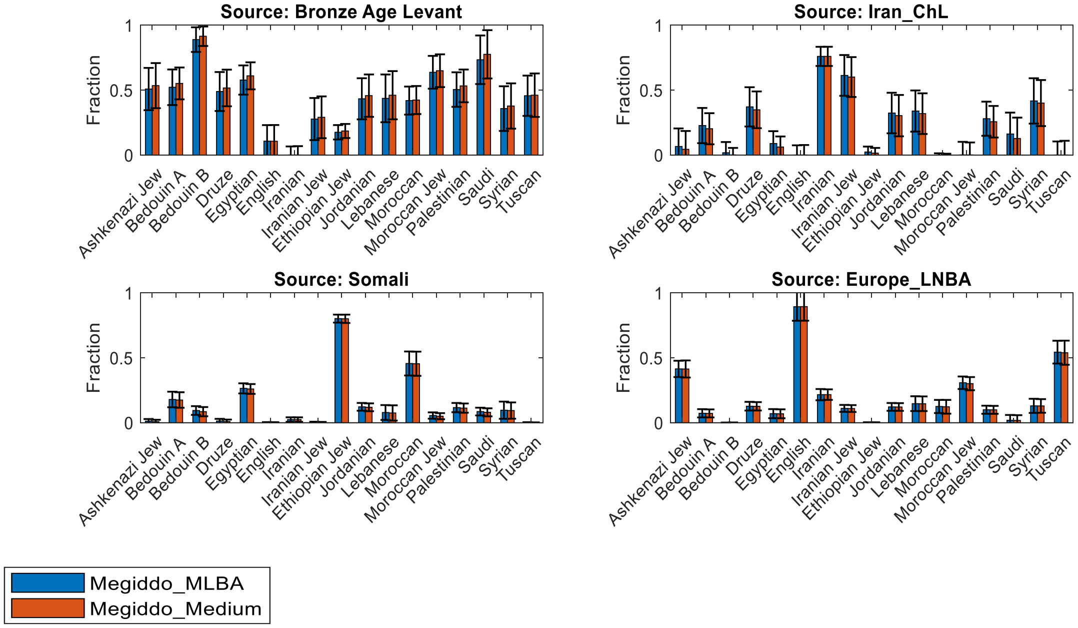

***Figure SJ.3**. The contribution of each of the source populations to the examined present-day population using LINADMIX, when the proxy for Bronze Age Levant is Megiddo_MLBA, and when changing the proxy to Megiddo_Medium.*

K. Robustness of LINADMIX results with respect to perturbations in ADMIXTURE parameters

To test the robustness of LINADMIX, we performed two perturbations to the parameters used in the main text. First, in the main text we used $K = 6$, as this was the optimal value according to ADMIXTURE's cross-validation procedure. Here, we changed this to $K = 7$, which has a slightly higher cross validation error. Second, in the main text we applied ADMIXTURE to 1,663 individuals, and here we changed it by more than two-fold to 3,515 individuals. The additional individuals came from various, apparently non relevant, populations. For this set of individuals, cross validation points at an optimal $K = 22$ (Figure S1B).

A summary of the differences between the analyses is given in Table SK.1, which compares the maximum deviation observed for each of the four source populations.

Analyses compared	Source population compared	Maximum absolute difference (target)	Minimum absolute difference (target)	Average absolute differences	Median absolute differences	Standard deviation of absolute differences
K=6 with K=7	Megiddo MLBA	0.174 (Tuscan)	$2 \cdot 10^{-10}$ (Iranian)	0.056	0.034	0.050
K=6 with K=22	Megiddo MLBA	0.097 (Lebanese)	$2 \cdot 10^{-9}$ (Iranian)	0.059	0.065	0.032
K=6 with K=7	Iran ChL	0.118 (Druze)	$2 \cdot 10^{-10}$ (Tuscan)	0.054	0.075	0.044
K=6 with K=22	Iran ChL	0.090 (Druze)	$2 \cdot 10^{-9}$ (Moroccan)	0.044	0.051	0.032
K=6 with K=7	Somali	0.023 (BedouinB)	$8 \cdot 10^{-9}$ (Tuscan)	0.010	0.011	0.006
K=6 with K=22	Somali	0.062 (BedouinB)	$1 \cdot 10^{-8}$ (Iranian Jew)	0.020	0.014	0.019
K=6 with K=7	Europe LNBA	0.317 (Tuscan)	$2 \cdot 10^{-6}$ (BedouinB)	0.126	0.135	0.074
K=6 with K=22	Europe LNBA	0.095 (Tuscan)	$7 \cdot 10^{-8}$ (Ethiopian Jew)	0.025	0.019	0.024
K=6 with K=7	All	0.317	$2 \cdot 10^{-10}$	0.061	0.032	0.064
K=6 with K=22	All	0.097	$2 \cdot 10^{-10}$	0.037	0.031	0.030

Table SK.1*. Maximum deviation of the estimators of the contribution of ancient populations to present-day ones, when ADMIXTURE parameters are perturbed.*

It appears from the table that the differences between the analyses happen mainly in the contributions of Megiddo_MLBA, Iran_ChL and Europe_LNBA. Differences between the ADMIXTURE runs for the Somali contribution are usually small, likely because of the generally smaller African component in present-day Levantine populations (Figure SK.1), and because Megiddo_MLBA, Iran_ChL and Europe_LNBA are more similar to each other than to Somali.

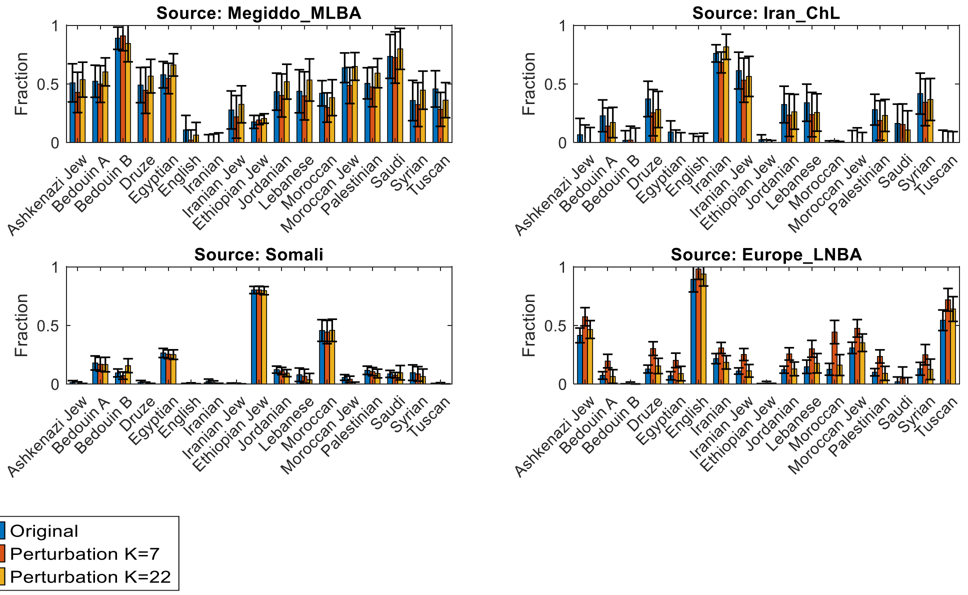

__Figure SK.1__. The contribution of each of the source populations to the examined present-day population using LINADMIX on the original data (using K=6), on ADMIXTURE on 1,663 individuals using K=7, and on ADMIXTURE with 3,515 individuals using K=22.

The norms of the residuals, measuring model quality, are shown in Figure SK.2. In both perturbations, as well as in the original analysis (Figure 5A), outgroup populations usually show worse fits (especially Tuscan and Moroccan), suggesting – as expected – that outgroup populations may have additional components poorly modeled by our analysis. It is striking that norms for $K = 7$ are at least an order of magnitude higher than in the original model, highlighting the importance of determining an optimal value of K using cross validation.

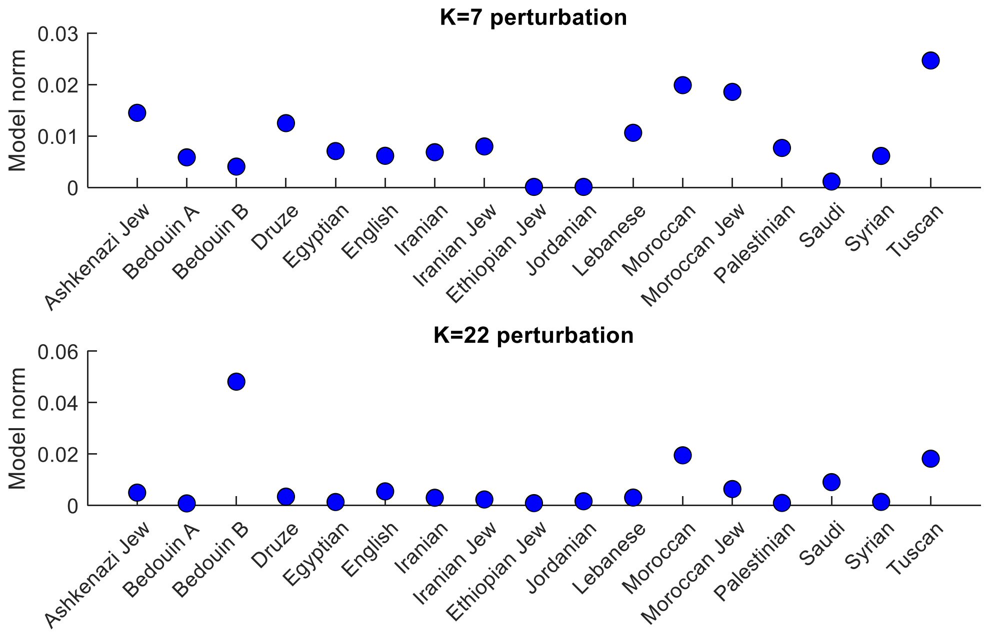

__Figure SK.2__. Residual norms of LINADMIX models, when the parameters of ADMIXTURE are perturbed.

Bedouin B, and to a lesser extent Saudi, are poorly modeled in the $K = 22$ perturbation, as well as by PHCP. Interestingly, these two populations bear the highest Megiddo_MLBA-related component, suggesting that the original analysis included in Megiddo_MLBA sources that could not be captured using $K = 22$. In the $K = 7$ perturbation, the highest norms are obtained for target populations with a pronounced European-related component, which might explain the observation that the results based on the $K = 7$ perturbation have a bias toward Europe_LNBA.

Comparison of all three ADMIXTURE runs (Table S4) shows that they agree very well with each other, but that there are roughly 15-25% of the total genetic makeup whose source is difficult to pin down and that its attribution to specific source populations fluctuates between the different ADMIXTURE runs. Generally, changing the number of ancestral populations from six to seven had a greater impact on the results than changing the background individuals, stressing the fact that addition of non-relevant populations does not have a great effect on LINADMIX. In the $K = 22$ perturbation, the fluctuating part is ≈15%, and mostly spreads between Megiddo_MLBA and Iran_ChL, reflecting their higher standard errors in the original analysis. This is likely a result of genetic similarity between Megiddo_MLBA and Iran_ChL, which makes the determination of their respective contributions more difficult. Of note, source populations that have a small contribution to the model usually also have higher confidence levels.

The models show high similarity despite the uncertainty in the respective contributions of Megiddo_MLBA and Iran_ChL. Overall, Ethiopian Jew is the population that shows the highest consistency between the analyses (average of absolute differences in ancestral components 0.0116 and 0.0144 when comparing the original model with $K = 7$ and $K = 22$, respectively). This is not surprising as Ethiopian Jews have neither Iranian nor European contributions. This also demonstrates that LINADMIX does not distribute contributions randomly between similar source populations, and that uncertainty arises only when more than one similar source population contributed to the genome of the target population. The target populations that differ the most between the original analysis and with the $K = 7$ perturbation are Moroccan and Tuscan. This is probably a result of the worse fit (higher norms) of the model to Moroccan and Tuscan in all analyses. The target population that differs the most between the original analysis and that with the $K = 22$ perturbation is Lebanese, characterized by high contributions of both Megiddo_MLBA and Iran_ChL.

In summary, we listed here some variations between the results of the different analyses. These variations are smallest for the two analyses where K has been optimized using cross validation. The highest variations amount to up to ≈15% (apart from Europe_LNBA in the poorly modeled Moroccan for the non-optimal K=7), and almost always reflect the inability to determine the relative contributions of Megiddo_MLBA and Iran_ChL.