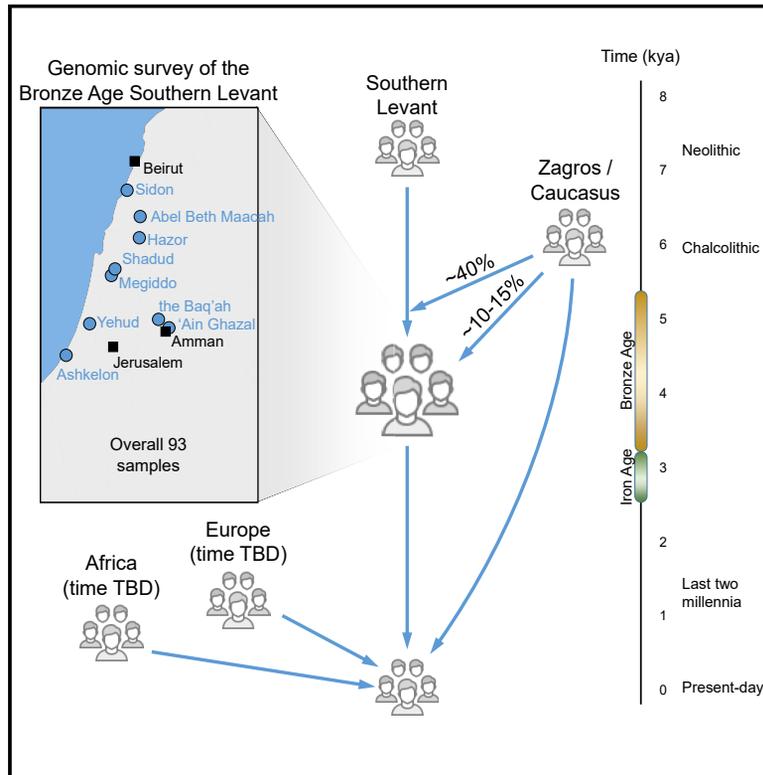


The Genomic History of the Bronze Age Southern Levant

Graphical Abstract



Authors

Lily Agranat-Tamir, Shamam Waldman, Mario A.S. Martin, ..., Israel Finkelstein, Liran Carmel, David Reich

Correspondence

mario.novak@inantro.hr (M.N.),
ron.pinhasi@univie.ac.at (R.P.),
shai.carmi@mail.huji.ac.il (S.C.),
fink2@tauex.tau.ac.il (I.F.),
liran.carmel@huji.ac.il (L.C.),
reich@genetics.med.harvard.edu (D.R.)

In Brief

Genome-wide data from Bronze Age individuals across nine sites in the Southern Levant show strong genetic resemblance, including a component from populations related to Chalcolithic Zagros and Early Bronze Age Caucasus introduced by gene flow lasting at least until the late Bronze Age and affecting modern Levantine population architecture.

Highlights

- Analysis of genome-wide data for nine sites from the Bronze Age Southern Levant
- Contemporaneous samples from multiple sites are genetically similar
- Migration from the Zagros and/or Caucasus to the Levant between 2500–1000 BCE
- People related to these individuals contributed to all present-day Levantine populations



Article

The Genomic History of the Bronze Age Southern Levant

Lily Agranat-Tamir,^{1,2} Shamam Waldman,³ Mario A.S. Martin,⁴ David Gokhman,¹ Nadav Mishol,¹ Tzilla Eshel,⁵ Olivia Cheronet,⁶ Nadin Rohland,^{7,8} Swapan Mallick,^{7,8,9,10} Nicole Adamski,^{7,9} Ann Marie Lawson,^{7,9} Matthew Mah,^{7,8,9} Megan Michel,^{7,9} Jonas Oppenheimer,^{7,9} Kristin Stewardson,^{7,9} Francesca Candilio,^{11,12} Denise Keating,¹¹ Beatriz Gamarra,^{11,13,14} Shay Tzur,³ Mario Novak,^{22,*} Rachel Kalisher,¹⁵ Shlomit Bechar,¹⁶ Vered Eshed,¹⁷ Douglas J. Kennett,¹⁸ Marina Faerman,¹⁹ Naama Yahalom-Mack,¹⁶ Janet M. Monge,²⁰ Yehuda Govrin,²¹ Yigal Erel,⁵ Benjamin Yakir,² Ron Pinhasi,^{6,24,*} Shai Carmi,^{3,24,*} Israel Finkelstein,^{4,24,*} Liran Carmel,^{1,23,24,*} and David Reich^{7,8,9,10,24,*}

¹Department of Genetics, The Alexander Silberman Institute of Life Sciences, Faculty of Science, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 9190401, Israel

²Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 9190501, Israel

³Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

⁴Institute of Archaeology, Tel Aviv University, Tel Aviv 6997801, Israel

⁵The Fredy and Nadine Herrmann Institute of Earth Sciences, the Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 9190401, Israel

⁶Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria

⁷Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

⁸Broad Institute, Cambridge, MA 02138, USA

⁹Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

¹⁰Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

¹¹School of Archaeology and Earth Institute, University College Dublin, Dublin, Ireland

¹²Soprintendenza ABAP per la città metropolitana di Cagliari e le province di Oristano e Sud Sardegna, Cagliari, Italy

¹³Institut Català de Paleocologia Humana i Evolució Social, Tarragona, Spain

¹⁴Universitat Rovira i Virgili, Àrea de Prehistòria, Tarragona, Spain

¹⁵Joukowsky Institute for Archaeology and the Ancient World, Brown University, Providence, RI 02912, USA

¹⁶The Institute of Archaeology, The Hebrew University of Jerusalem, Jerusalem 9190501, Israel

¹⁷Israel Antiquities Authority, P.O.B 586, Jerusalem 91004, Israel

¹⁸Department of Anthropology, University of California, Santa Barbara, CA 93106, USA

¹⁹Laboratory of Bioanthropology and Ancient DNA, Institute of Dental Sciences, Faculty of Dental Medicine, The Hebrew University of Jerusalem, Jerusalem 9112001, Israel

²⁰University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, PA 19104, USA

²¹The Nelson Glueck School of Biblical Archaeology, Hebrew Union College, Jerusalem 9410125, Israel

²²Institute for Anthropological Research, Zagreb, Croatia

²³Lead Contact

²⁴These authors contributed equally

*Correspondence: mario.novak@inantro.hr (M.N.), ron.pinhasi@univie.ac.at (R.P.), shai.carmi@mail.huji.ac.il (S.C.), fink2@tauex.tau.ac.il (I.F.), liran.carmel@huji.ac.il (L.C.), reich@genetics.med.harvard.edu (D.R.)

<https://doi.org/10.1016/j.cell.2020.04.024>

SUMMARY

We report genome-wide DNA data for 73 individuals from five archaeological sites across the Bronze and Iron Ages Southern Levant. These individuals, who share the “Canaanite” material culture, can be modeled as descending from two sources: (1) earlier local Neolithic populations and (2) populations related to the Chalcolithic Zagros or the Bronze Age Caucasus. The non-local contribution increased over time, as evinced by three outliers who can be modeled as descendants of recent migrants. We show evidence that different “Canaanite” groups genetically resemble each other more than other populations. We find that Levant-related modern populations typically have substantial ancestry coming from populations related to the Chalcolithic Zagros and the Bronze Age Southern Levant. These groups also harbor ancestry from sources we cannot fully model with the available data, highlighting the critical role of post-Bronze-Age migrations into the region over the past 3,000 years.

INTRODUCTION

The Bronze Age (ca. 3500–1150 BCE) was a formative period in the Southern Levant, a region that includes present-day Israel,

Jordan, Lebanon, the Palestinian Authority, and southwest Syria. This era, which ended in a large-scale civilization collapse across this region (Cline, 2014), shaped later periods both demographically and culturally. The following Iron Age (ca. 1150–586 BCE)



saw the rise of territorial kingdoms such as biblical Israel, Judah, Ammon, Moab, and Aram-Damascus, as well as the Phoenician city-states. In much of the Late Bronze Age, the region was ruled by imperial Egypt, although in later phases of the Iron Age it was controlled by the Mesopotamian-centered empires of Assyria and Babylonia. Archaeological and historical research has documented major changes during the Bronze and Iron Ages, such as the cultural influence of the northern (Caucasian) populations related to the Kura-Araxes tradition during the Early Bronze Age (Greenberg and Goren, 2009) and effects from the “Sea Peoples” (such as Philistines) from the west in the beginning of the Iron Age (Yasur-Landau, 2010).

The inhabitants of the Southern Levant in the Bronze Age are commonly described as “Canaanites,” that is, residents of the Land of Canaan. The term appears in several 2nd millennium BCE sources (e.g., Amarna, Alalakh, and Ugarit tablets) and in biblical texts dating from the 8th–7th centuries BCE and later (Binkowski, 1999; Lemche, 1991; Na’aman, 1994a). In the latter, the Canaanites are referred to as the pre-Israelite inhabitants of the land (Na’aman, 1994a). Canaan of the 2nd millennium BCE was organized in a system of city-states (Goren et al., 2004), where elites ruled from urban hubs over rural (and in some places pastoral) countryside. The material culture of these city-states was relatively uniform (Mazar, 1992), but whether this uniformity extends to their genetic ancestry is unknown. Although genetic ancestry and material culture are unlikely to ever match perfectly, past ancient DNA analyses show that they might sometimes be strongly associated. In other cases, a direct correspondence between genetics and culture cannot be established. We discuss several examples in the Discussion.

Previous ancient DNA studies published genome-scale data for thirteen individuals from four Bronze Age sites in the Southern Levant: three individuals from ‘Ain Ghazal in present-day Jordan, dated to ~2300 BCE (Intermediate Bronze Age) (Lazaridis et al., 2016); five from Sidon in present-day Lebanon, dated to ~1750 BCE (Middle Bronze Age) (Haber et al., 2017); two from Tel Shadud in present-day Israel, dated to ~1250 BCE (Late Bronze Age) (van den Brink et al., 2017); and three from Ashkelon in present-day Israel, dated to ~1650–1200 BCE (Middle and Late Bronze Age) (Feldman et al., 2019). The ancestry of these individuals could be modeled as a mixture of earlier local groups and groups related to the Chalcolithic people of the Zagros Mountains, located in present-day Iran and designated in previous studies as Iran_ChL (Haber et al., 2017; Lazaridis et al., 2016). The Bronze Age Sidon group could be modeled as a major (93% ± 2%) ancestral source for present-day groups in the region (Haber et al., 2017). A study of Chalcolithic individuals from Peqj’in cave in the Galilee (present-day Israel) showed that the ancestry of this earlier group included an additional component related to earlier Anatolian farmers, which was excluded as a substantial source for later Bronze Age groups from the Southern Levant, with the exception of the coastal groups from Sidon and Ashkelon (Feldman et al., 2019; Harney et al., 2018). These observations point to a degree of population turnover in the Chalcolithic-Bronze Age transition, consistent with archaeological evidence for a disruption between local Chalcolithic and Early Bronze cultures (de Miroschedji, 2014).

Here, we set out to address three issues. First, we sought to determine the extent of genetic homogeneity among the sites associated with Canaanite material culture. Second, we analyzed the data to gain insights into the timing, extent, and origin of gene flow that brought Zagros- and Caucasus-related ancestry to the Bronze Age Southern Levant. Third, we assessed the extent to which additional gene flow events have affected the region since that time.

To address these questions, we generated genome-wide ancient DNA data for 71 Bronze Age and 2 Iron Age individuals, spanning roughly 1,500 years, from the Intermediate Bronze Age to the Early Iron Age. Combined with previously published data on the Bronze and Iron Ages in the Southern Levant, we assembled a dataset of 93 individuals from 9 sites across present-day Israel, Jordan, and Lebanon, all demonstrating Canaanite material culture. We show that the sampled individuals from the different sites are usually genetically similar, albeit with subtle but in some cases significant differences, especially in residents of the coastal regions of Sidon and Ashkelon. Almost all individuals can be modeled as a mixture of local earlier Neolithic populations and populations from the northeastern part of the Near East. However, the mixture proportions change over time, revealing the demographic dynamics of the Southern Levant during the Bronze Age. Finally, we show that the genomes of present-day groups geographically and historically linked to the Bronze Age Levant, including the great majority of present-day Jewish groups and Levantine Arabic-speaking groups, are consistent with having 50% or more of their ancestry from people related to groups who lived in the Bronze Age Levant and the Chalcolithic Zagros. These present-day groups also show ancestries that cannot be modeled by the available ancient DNA data, highlighting the importance of additional major genetic effects on the region since the Bronze Age.

RESULTS

Dataset

We extracted DNA from the bones of 73 individuals from 5 archaeological sites in the Southern Levant (Table S1; STAR Methods; Figure 1A):

Thirty-five individuals from Tel Megiddo (northern Israel), most of whom date to the Middle-to-Late Bronze Age, except for one dating to the Intermediate Bronze Age and one dating to the Early Iron Age

Twenty-one individuals from the Baq’ah in central Jordan (northeast of Amman), mostly from the Late Bronze Age

Thirteen individuals from Yehud (central Israel), dating to the Intermediate Bronze Age

Three individuals from Tel Hazor (northern Israel) dating to the Middle-to-Late Bronze Age

One individual from Tel Abel Beth Maacah (northern Israel), dating to the Iron Age

For all analyzed samples but one, DNA was extracted from petrous bones. The DNA was converted to double-indexed half Uracil-DNA glycosylase (UDG)-treated libraries that we enriched for about 1.2 million single nucleotide polymorphism (SNPs) before sequencing (see STAR Methods). The median number

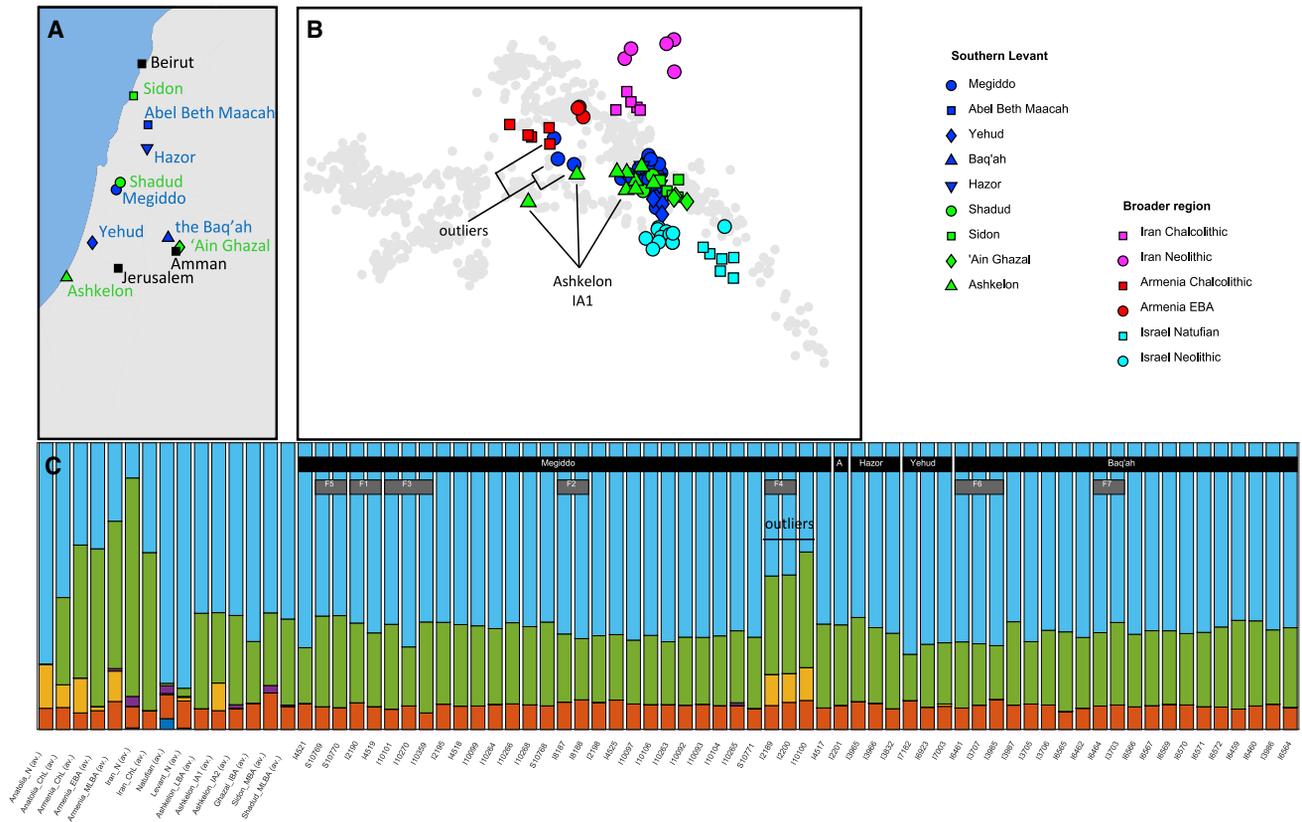


Figure 1. Bronze and Iron Age Individuals Analyzed in This Study

(A) Location of archaeological sites. Shown in blue are sites with individuals first reported in this paper. In green are sites with individuals reported in previous studies.

(B) PCA plot, showing present-day Eurasian individuals in gray (taken from Lazaridis et al., 2014) and ancient individuals in color. Only individuals with at least 30,000 autosomal SNPs were plotted. All Bronze and Iron Age individuals cluster (blue and green marks), except for the three denoted as “outliers” and for some IA1 individuals.

(C) ADMIXTURE plot with $K=6$, showing Bronze and Iron Age individuals, as well as other selected populations. Only individuals with at least 30,000 autosomal SNPs were plotted. The seven families are marked by F1–F7. “A” stands for “Abel.”

See also Figure S1 and Table S1.

of autosomal SNPs covered was 288,863 (range 4,883–945,269). In addition to genetic data, we measured values of strontium isotopes for 12 individuals (and for 8 additional individuals that did not produce DNA) (STAR Methods; Methods S1A), and generated accelerator mass spectrometry radiocarbon dates for 20 individuals (Table S1). We combined our newly generated data with published data for 13 Bronze Age Southern Levant individuals from ‘Ain-Ghazal, Sidon, Tel Shadud, and Ashkelon (van den Brink et al., 2017; Feldman et al., 2019; Haber et al., 2017; Lazaridis et al., 2016), and 7 Iron Age Southern Levant Individuals from Ashkelon (Feldman et al., 2019).

We projected the autosomal genetic data onto the plane spanned by the first two principal components of 777 present-day West Eurasian individuals genotyped for roughly 600,000 SNPs on the Affymetrix Human Origins SNP array (Lazaridis et al., 2014). We restricted the plot to 68 individuals represented by at least 30,000 autosomal SNPs (Figure 1B), a coverage threshold where the ability to infer ancestry was robust to sampling noise (Methods S1B). All Bronze and Iron Age Levant indi-

viduals (blue and green shapes) form a tight cluster, except for three outliers from Megiddo, and previously identified outliers from the Ashkelon population known as Iron Age I (IA1) (Feldman et al., 2019). We also ran ADMIXTURE on a set of 1,663 present-day and ancient individuals (see STAR Methods; Figure S1). The ADMIXTURE results are qualitatively consistent with the principal component analysis (PCA), suggesting that all individuals but the outliers from Megiddo and the Ashkelon IA1 population have similar ancestry (Figure 1C).

We used the method described in (Olaide et al., 2019) to identify 17 individuals as being first-, second-, or third-degree relatives of other individuals in the dataset. They fall within seven families: five in Tel Megiddo and two in the Baq'ah. In most families, we used only the member with the highest SNP coverage in subsequent analyses (Table S1). Two of the three Megiddo outliers are a brother and a sister (Family 4, I2189 and I2200), leaving in the final dataset two individuals marked as outliers. After removing low-coverage individuals and closely related family members, 62 individuals were left for further analysis (Table S1).

	Levant N	Megiddo I10100	Megiddo I2200	ASH IA1	Sidon	Baqah	Megiddo IBA	Megiddo IA	Hazor	Ain Ghazal	Megiddo MLBA	Shadud	Yehud	ASH LBA	ASH IA2	Abel
Levant N		7.8E-28	2.6E-31	5.4E-09	5.5E-53	2.3E-43	7.1E-05	1.0E-10	3.8E-21	1.2E-18	1.9E-31	3.8E-12	1.1E-05	3.7E-13	6.2E-22	3.2E-09
Megiddo I10100	7.8E-28		0.0111	6.3E-09	3.0E-07	2.1E-08	1.7E-13	1.0E-07	4.8E-06	9.9E-11	2.9E-12	0.0007	8.5E-05	6.4E-05	4.2E-05	1.1E-07
Megiddo I2200	2.6E-31	0.0111		1.5E-06	5.9E-06	4.8E-06	6.7E-11	0.0127	0.0005	1.8E-07	6.5E-08	8.0E-05	0.0006	0.0003	0.0036	0.0011
ASH IA1	5.4E-09	6.3E-09	1.5E-06		5.5E-14	1.0E-07	0.0189	0.1144	1.2E-05	0.0011	3.7E-07	0.0456	0.2524	0.0002	2.5E-05	0.0250
Sidon	5.5E-53	3.0E-07	5.9E-06	5.5E-14		1.5E-10	1.0E-06	0.0015	0.0002	1.2E-07	6.4E-12	0.008	1.8E-07	0.0517	0.7540	0.0438
Baqah	2.3E-43	2.1E-08	4.8E-06	1.0E-07	1.5E-10		0.0005	0.0408	0.3324	0.0039	0.0175	0.3421	0.0296	0.2916	0.2813	0.2256
Megiddo IBA	7.1E-05	1.7E-13	6.7E-11	0.0189	1.0E-06	0.0005		0.1292	0.0021	0.0054	0.0548	0.0455	0.6912	0.6578	0.0004	0.2604
Megiddo IA	1.0E-10	1.0E-07	0.0127	0.1144	0.0015	0.0408	0.1292		0.3901	0.3392	0.1486	0.006	0.2915	0.2088	0.4080	0.1435
Hazor	3.8E-21	4.8E-06	0.0005	1.2E-05	0.0002	0.3324	0.0021	0.3901		0.1028	0.1179	0.159	0.4844	0.1823	0.6114	0.1029
Ain Ghazal	1.2E-18	9.9E-11	1.8E-07	0.0011	1.2E-07	0.0039	0.0054	0.3392	0.1028		0.153	0.313	0.1087	0.1185	0.2975	0.7955
Megiddo MLBA	1.9E-31	2.9E-12	6.5E-08	3.7E-07	6.4E-12	0.0175	0.0548	0.1486	0.1179	0.153		0.8527	0.1576	0.2573	0.0631	0.4699
Shadud	3.8E-12	0.0007	8.0E-05	0.0456	0.008	0.3421	0.0455	0.006	0.159	0.313	0.8527		0.5198	0.5554	0.3117	0.7452
Yehud	1.1E-05	8.5E-05	0.0006	0.2524	1.8E-07	0.0296	0.6912	0.2915	0.4844	0.1087	0.1576	0.5198		0.0635	0.1118	0.8393
ASH LBA	3.7E-13	6.4E-05	0.0003	0.0002	0.0517	0.2916	0.6578	0.2088	0.1823	0.1185	0.2573	0.5554	0.0635		0.1997	0.3996
ASH IA2	6.2E-22	4.2E-05	0.0036	2.5E-05	0.7540	0.2813	0.0004	0.4080	0.6114	0.2975	0.0631	0.3117	0.1118	0.1997		0.3458
Abel	3.2E-09	1.1E-07	0.0011	0.0250	0.0438	0.2256	0.2604	0.1435	0.1029	0.7955	0.4699	0.7452	0.8393	0.3996	0.3458	

Figure 2. p values of qpWave for Each Pair of Populations

Values greater than 0.05 are shaded in light green, and values lower than 0.001 are shaded in light red. The rectangle shows all Levant populations excluding Sidon, the outliers from Megiddo and Ashkelon IA1. Computations were based on the o9a outgroup set (o9 + Anatolia_N).

High Degree of Genetic Affinities between Multiple Sites

We divided the 26 high-coverage individuals from Tel Megiddo into the following groups, on the basis of geographic location, archaeological period, and genetic clustering in PCA (Table S1): Intermediate Bronze Age (Megiddo_IBA, a single individual), Middle-to-Late Bronze Age (Megiddo_MLBA, 22 individuals), Iron Age (Megiddo_IA, a single individual), as well as the two outliers, Megiddo_I2200 and Megiddo_I10100, which were each treated as a separate group. We compared these groups and the other populations in our dataset to previously published data from other sites in the broader region and from earlier periods, including the Early Bronze Age Caucasus (Armenia_EBA), the Middle-to-Late Bronze Age Caucasus (Armenia_MLBA), the Chalcolithic Zagros Mountains (Iran_ChL), the Chalcolithic Caucasus (Armenia_ChL), the Neolithic of the Southern Levant (Levant_N), the Neolithic of the Zagros Mountains (Iran_N), and the Neolithic of Anatolia (Anatolia_N) (Lazaridis et al., 2016).

To test for variation in ancestry proportions among the Levant Bronze and Iron Age groups, we used qpWave. qpWave tests whether each possible pair of groups (Test_i, Test_j) is consistent with descending from a common ancestral population—that is, consistent with being a clade—since separation from the ancestors of a set of outgroup populations. qpWave works by computing symmetry test statistics of the form $f_4(\text{Test}_i, \text{Test}_j; \text{Outgroup}_k, \text{Outgroup}_l)$, which have an expected value of zero if (Test_i, Test_j) form a clade with respect to the outgroups. qpWave then generates a single p value corrected for the empirically measured correlation among the statistics (Reich et al., 2012). Using a distantly related set of outgroups, we found that with the exception of the outliers from Megiddo, Ashkelon IA1, and Sidon, all Bronze and Iron Age Levant groups are

consistent with being pairwise clades with respect to the outgroups (Figure 2).

We discuss each of qpWave’s findings of significant population substructure in turn. The Megiddo outliers not only fail to form a clade with the other populations, but also with each other. Ashkelon IA1 has previously been reported to harbor European ancestry, and so our finding that it is genetically differentiated from contemporary groups is unsurprising (Feldman et al., 2019). The significant differentiation of the Sidon individuals in qpWave—despite the fact that they roughly cluster with the other Southern Levant Bronze Age groups in PCA and ADMIXTURE—is notable, especially because we find that they are consistent with forming a clade with the two groups from coastal Ashkelon that do not have European-related admixture (the Bronze Age and later Iron Age groups ASH_LBA and ASH_IA2). Speculatively, this observation could be related to the fact that both Sidon and Ashkelon were port towns with connections to other Mediterranean coastal groups outside the Southern Levant, which could have introduced ancestry components that are absent from inland Levantine Bronze Age groups, although it is difficult to test this hypothesis in the absence of high resolution ancient DNA sampling from the eastern Mediterranean rim. The genetic distinctiveness of the Sidon individuals is also compatible with previous findings that Chalcolithic Levantine individuals from Peqi’in Cave are consistent with contributing some ancestry to the Sidon individuals, but not to the ‘Ain Ghazal ones (Harney et al., 2018). We considered the possibility that the significantly different genetic patterns we detect in the Sidon individuals could reflect their different experimental treatment compared with that of the other individuals in this study (shotgun sequencing of non-UDG-treated libraries compared

B	Levant N	Megiddo I10100	Megiddo I2200	ASH IA1	Sidon	Baq'ah	Megiddo IBA	Hazor	Ain Ghazal	Megiddo IA	Megiddo MLBA	Shadud	Yehud	ASH LBA	ASH IA2	Abel
Levant N		6.7E-40	1.1E-55	8.5E-27	6.9E-108	9.4E-106	6.1E-11	1.2E-49	4.8E-43	1.2E-22	2.4E-80	4.1E-29	1.4E-07	9.6E-31	1.4E-37	1.7E-27
Megiddo I10100	6.7E-40		4.3E-02	2.1E-11	1.5E-06	1.7E-07	1.5E-13	2.5E-05	4.6E-10	1.3E-06	1.9E-11	2.6E-04	1.2E-05	0.0004	2.1E-05	8.6E-07
Megiddo I2200	1.1E-55	4.3E-02		1.8E-06	3.3E-05	2.1E-05	6.7E-13	8.2E-04	1.1E-09	2.4E-02	1.3E-08	5.8E-05	7.0E-06	0.0012	0.0023	2.6E-03
ASH IA1	8.5E-27	2.1E-11	1.8E-06		2.7E-14	2.5E-08	0.0002	1.6E-05	8.0E-06	0.0281	8.2E-08	0.0054	0.1036	3.9E-05	2.3E-05	0.0067
Sidon	6.9E-108	1.5E-06	3.3E-05	2.7E-14		1.9E-11	2.4E-09	1.3E-04	6.2E-11	1.2E-03	1.8E-15	2.4E-03	1.4E-09	0.1060	0.1334	1.0E-01
Baq'ah	9.4E-106	1.7E-07	2.1E-05	2.5E-08	1.9E-11		2.6E-07	4.6E-01	4.3E-05	5.0E-02	2.3E-04	1.7E-01	1.9E-03	0.2744	0.2781	1.3E-01
Megiddo IBA	6.1E-11	1.5E-13	6.7E-13	0.0002	2.4E-09	2.6E-07		8.9E-06	2.3E-03	4.0E-02	7.4E-04	2.2E-02	7.9E-01	0.0495	2.0E-05	4.8E-02
Hazor	1.2E-49	2.5E-05	8.2E-04	1.6E-05	1.3E-04	4.6E-01	8.9E-06		9.4E-02	4.6E-01	2.3E-01	1.1E-01	4.5E-02	0.2173	0.7576	5.0E-02
Ain Ghazal	4.8E-43	4.6E-10	1.1E-09	8.0E-06	6.2E-11	4.3E-05	2.3E-03	9.4E-02		4.2E-01	2.1E-01	4.5E-01	2.4E-01	0.0115	0.1055	1.1E-01
Megiddo IA	1.2E-22	1.3E-06	2.4E-02	0.0281	1.2E-03	5.0E-02	4.0E-02	4.6E-01	4.2E-01		2.3E-01	2.0E-03	1.5E-01	0.2394	0.6397	1.5E-01
Megiddo MLBA	2.4E-80	1.9E-11	1.3E-08	8.2E-08	1.8E-15	2.3E-04	7.4E-04	2.3E-01	2.1E-01	2.3E-01		6.7E-01	9.8E-02	0.0440	0.0418	1.3E-01
Shadud	4.1E-29	2.6E-04	5.8E-05	0.0054	2.4E-03	1.7E-01	2.2E-02	1.1E-01	4.5E-01	2.0E-03	6.7E-01		4.3E-01	0.1817	0.0943	5.0E-01
Yehud	1.4E-07	1.2E-05	7.0E-06	0.1036	1.4E-09	1.9E-03	7.9E-01	4.5E-02	2.4E-01	1.5E-01	9.8E-02	4.3E-01		0.0375	0.0849	7.6E-01
ASH LBA	9.6E-31	0.0004	0.0012	3.9E-05	0.1060	0.2744	0.0495	0.2173	0.0115	0.2394	0.0440	0.1817	0.0375		0.1296	0.4610
ASH IA2	1.4E-37	2.1E-05	0.0023	2.3E-05	0.1334	0.2781	2.0E-05	0.7576	0.1055	0.6370	0.0418	0.0943	0.0849	0.1296		0.1013
Abel	1.7E-27	8.6E-07	2.6E-03	0.0067	1.0E-01	1.3E-01	4.8E-02	5.0E-02	1.1E-01	1.5E-01	1.3E-01	5.0E-01	7.6E-01	0.4610	0.1013	

Figure 3. p values of qpWave for Each Pair of Populations

Values greater than 0.05 are shaded in light green, and values lower than 0.001 are shaded in light red. The rectangle shows all Levant populations excluding Sidon, the outliers from Megiddo and Ashkelon IA1. Computations were based on the extended outgroup set (o9 + Anatolia_N + Armenia_MLBA + Caucasus Hunter Gatherers [CHG] + Natufians).

with enrichment of UDG-treated libraries). To test this, we repeated the analyses by using only transversion SNPs, which are less prone to characteristic ancient DNA errors, but found no indication of systematic bias (Wang et al., 2015). However, we did find evidence of substructure within the Sidon individuals, and some but not all were consistent with forming a clade with inland Southern Levant populations, a finding that could reflect substantial cosmopolitan nature of this coastal site (Methods S1C, see Discussion).

To reveal subtler population structure, we repeated the qpWave analysis adding outgroups that are genetically closer to the test groups, such as Armenia_MLBA and Natufian (Figure 3). With this more powerful set of outgroups, Baq'ah and Megiddo_IBA also provide evidence of not being pairwise clades with the remaining groups. Thus, beyond the broad observation of genetic affinities between sites, we also observe subtle ancestry heterogeneity across the region during the Bronze Age (see Discussion).

Gene Flow into the Southern Levant During the Bronze Age

Two previous studies of Bronze Age individuals from 'Ain Ghazal and Sidon modeled them as derived from a mixture of earlier local groups (Levant_N) and groups related to peoples of the Chalcolithic Zagros mountains (Iran_ChL) (Haber et al., 2017; Lazaridis et al., 2016). These groups were estimated to harbor around 56% ± 3% and 48% ± 4% Neolithic Levant-related ancestry for 'Ain Ghazal (Lazaridis et al., 2016) and Sidon (Haber et al., 2017), respectively. We used qpAdm to estimate that Bronze and Iron

Age Ashkelon (ASH_LBA and ASH_IA2) carry 54% ± 5% and 42% ± 5% Neolithic Levant-related ancestry, respectively. Next, we used qpAdm to test the same model for the data reported here and found that most Middle-to-Late Bronze Age groups fit the model, with point estimates of 48%–57% Levant_N ancestry. These ancestry proportions are statistically indistinguishable (Bonferroni-corrected z test), which corroborates the fact that they are consistent with forming pairwise clades in qpWave (Table S2; Methods S1D). The only group that failed to fit this model was Baq'ah (p = 0.0003), even when using a wide range of outgroup populations (Table S2). This might be a result of ancestry heterogeneity across the Baq'ah individuals (see below).

To obtain insight into the Zagros-related ancestry component, we focused on two questions: what is the likely origin of this ancestry component and what is its likely timing? Although people of the Chalcolithic Zagros are so far the best proxy population for this ancestry component, there is no archaeological evidence for cultural spread directly from the Zagros into the Southern Levant during the Bronze Age. In contrast, there is archaeological support for connections between Bronze Age Southern Levant groups and the Caucasus (Greenberg and Goren, 2009), a term we use to represent both present-day Caucasus, as well as neighboring regions such as eastern Anatolia (see Discussion). With regard to the timing of these events, archaeology points to cultural affinities between the Kura-Araxes (Caucasus) and Khirbet Kerak (Southern Levant) archaeological cultures in the first half of the 3rd millennium BCE (Greenberg and Goren, 2009), and textual evidence documents a number of non-Semitic, Hurrian (from the northeast of the ancient Near East)

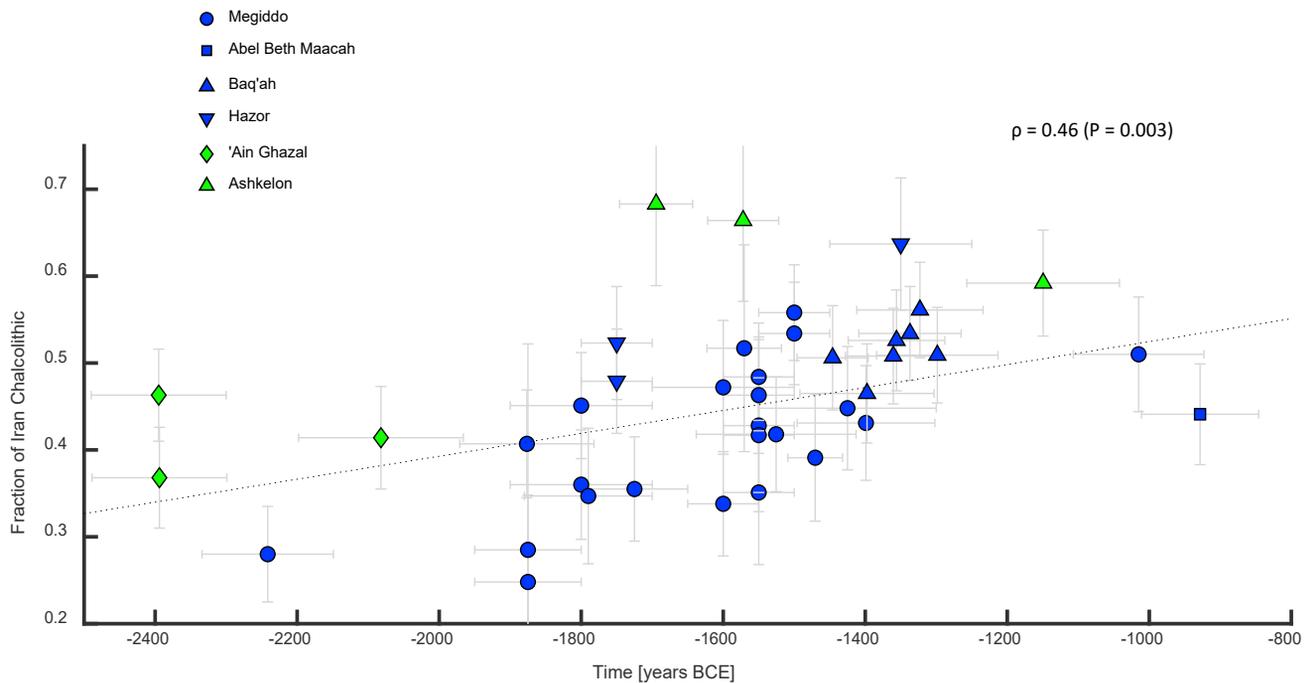


Figure 4. Temporal Changes in the Genetic Makeup of Individuals in the Bronze and Iron Age Levant

Fraction of Chalcolithic-Iran-related component in each individual as computed by *qpAdm*, modeling each individual as a mixture of Neolithic Levant and Chalcolithic Iran and using the o9a outgroup set (o9 + Anatolia_N). Vertical error bars denote one standard error in each direction. Horizontal error bars denote estimated time ranges. Dashed line describes the linear regression. Only individuals whose time range does not exceed 250 years are plotted and used in the regression. Note that the two well-dated Ash_LBA individuals happen to harbor the highest Iran_ChL component.

See also [Figures S2](#) and [S3](#) and [Tables S2](#) and [S3](#).

personal names in the 2nd millennium BCE, for example in the Amarna archive of the 14th century BCE (Na'aman, 1994b). We therefore reasoned that the Chalcolithic Zagros component might have arrived into the Southern Levant through the Caucasus (and even more proximately the northeastern areas of the ancient Near East, although we have no ancient DNA sampling from this region). This movement might not have been limited to a short pulse, and instead could have involved multiple waves throughout the Bronze Age.

To test whether the origin of the gene flow was from the Caucasus, rather than directly from the Zagros region, we ran *qpAdm*, replacing Iran_ChL with Early Bronze Age Caucasus (Armenia_EBA). We found that the Caucasus model received similar support to that of the Zagros model ([Table S2](#); [Methods S1E](#)). Next, we modeled Armenia_EBA as a mixture of an earlier Caucasus population (Chalcolithic Armenia, Armenia_ChL) and Iran_ChL and found that indeed Armenia_EBA is compatible with this model ([Table S2](#)). Altogether, we conclude that our data are also compatible with a model in which Zagros-related ancestry in the Southern Levant arrived through the Caucasus, either directly or via intermediates.

To study the timing of the admixture of Zagros-related ancestry in the Southern Levant, we leveraged the large time span of individuals in our dataset, extending across roughly 1,500 years, from the Intermediate Bronze Age to the Early Iron Age. Using *qpAdm*-based ancestry estimates for each of the individuals, we found that almost all are compatible with

being an admixture of groups related to the Neolithic Levant and Chalcolithic Zagros. One exception to this is an individual in Megiddo_MLBA that is weakly compatible with the model. Another exception is three individuals in the Baq'ah ([Table S2](#)), which suggests that the difficulty in modeling individuals from this site as a mixture of Neolithic Levant and Chalcolithic Zagros might reflect ancestry heterogeneity ([Figure 3](#)). These results do not change qualitatively when we used a larger set of outgroup populations ([Table S2](#)). We observed that the oldest individuals in our collection, from the Intermediate Bronze Age, already carried significant Zagros-related ancestry, suggesting that gene flow into the region started before ca. 2400 BCE. This is consistent with the hypothesis that people of Kura-Araxes archaeological complex of the 3rd millennium BCE might have affected the Southern Levant not only culturally, but also through some degree of movement of people. Our data also imply an increase in the proportion of Zagros-related ancestry after the Intermediate Bronze Age, as reflected in a significantly positive slope in a linear regression of the Chalcolithic-Zagros-related ancestry over the calendar year ($\beta = 1.4 \cdot 10^4 \pm 0.4 \cdot 10^4$, Jackknife), amounting to an increase of ~14% per thousand years ([Figures 4](#) and [S2A](#)). However, we caution that the number of individuals and their time span are insufficient to determine whether the increase in the Zagros-related ancestry happened continuously during the Middle and Late Bronze Ages, or whether there were multiple distinct migration events.

The two outliers from Megiddo (three including the sibling pair) provide additional evidence for the timing and origin of gene flow into the region. The three were found in close proximity to each other at Level K-10, which is radiocarbon dated to 1581–1545 BCE (domestic occupation) and 1578–1421 BCE (burials; both ± 1 s) (Martin et al., 2020; Toffolo et al., 2014), whereas the bone of one of the three (I10100) was directly dated (1688–1535 BCE, $\pm 2\sigma$). The reason these individuals are distinct from the rest is that their Caucasus- or Zagros-related genetic component is much higher, reflecting ongoing gene flow into the region from the northeast (Table S2; Figure S2B). The Neolithic Levant component is 22%–27% in I2200, and 9%–26% in I10100. These individuals are unlikely to be first generation migrants, as strontium isotope analysis on the two outlier siblings (I2189 and I2200) (Methods S1A) suggests that they were raised locally. This implies that the Megiddo outliers might be descendants of people who arrived in recent generations. Direct support for this hypothesis comes from the fact that in sensitive *qpAdm* modeling (including closely related sets of outgroups), the only working northeast source population for these two individuals is the contemporaneous Armenia_MLBA, whereas the earlier Iran_ChL and Armenia_EBA do not fit (Table S2). The addition of Iran_ChL to the set of outgroups does not change this result or cause model failure. Finally, no other Levantine group shows a similar admixture pattern (Table S2). This shows that some level of gene flow into the Levant took place during the later phases of the Bronze Age and suggests that the source of this gene flow was the Caucasus.

Altogether, our analyses show that gene flow into the Levant from people related to those in the Caucasus or Zagros was already occurring by the Intermediate Bronze Age, and that it lingered, episodically or continuously, at least in inland sites, during the Middle-to-Late Bronze Age.

Further Change in Levantine Populations Since the Bronze Age

To develop a sense of population changes in the Levant since the Bronze Age, we attempted to model groups that have a tradition of descent from ancient people in the region (Jews) as well as Levantine Arabic-speakers as mixtures of various ancient source populations. *qpAdm* assumes no admixture between groups related to the outgroups and the source populations, but almost all present-day Levantine and Mediterranean populations have significant sub-Saharan-African-related admixture that the ancient groups did not. This eliminates many key outgroups for *qpAdm* and reduces the utility of the method in this context. In particular, we were not able to apply *qpAdm* to get a single working model for the majority of present-day West Eurasian populations. As an alternative, we developed a methodology we call LINADMIX, which relies on the output of ADMIXTURE (Alexander et al., 2009) and uses constrained least-squares to estimate the contribution of given source populations to a target population (see STAR Methods). As a complementary approach, we developed a tool we call pseudo-haplotype ChromoPainter (PHCP), which is an adaptation of the haplotype-based method ChromoPainter (Lawson et al., 2012) to ancient genomes (see STAR Methods; Methods S1F). We first established that these methods provide meaningful estimates of ancestry in the context

of this study by using them to re-compute the ancestry proportions that we were able to model with *qpAdm*. Both LINADMIX and PHCP (Table S3; Figure S3; Methods S1F) produce qualitatively similar estimates as *qpAdm* (Table S2). To further establish the methods, we performed simulations that were designed to test the methods' abilities to infer ancestry proportions in present-day populations in a setup similar to the current study (Methods S1H). For this, we generated present-day populations as a mixture of two closely related ancient populations with and without a third, more distant, population. Both methods estimated the ancestry proportion of the distant source population with errors of up to 4% and the proportions of the closely related source populations with errors of up to 10%. Thus, although ADMIXTURE, the basis of LINADMIX, is known to have certain pitfalls as a tool for quantifying ancestry proportions (Lawson et al., 2018), in the case of individuals with ancestry sources similar to those we have analyzed here, our results suggest that both LINADMIX and PHCP are highly informative.

For the LINADMIX analysis of present-day populations, we used a background dataset of 1,663 present-day and ancient individuals from 239 populations genotyped by using SNP arrays and focused our analysis on 14 Jewish and Levantine present-day populations, along with modern English, Tuscan, and Moroccan populations that were used as controls. We used LINADMIX to model each of the 17 present-day populations as an admixture of four sources: (1) Megiddo_MLBA (the largest group) as a representative of the Middle-to-Late Bronze Age component; (2) Iran_ChL as a representative of the Zagros and the Caucasus; (3) Present-day Somalis as representatives of an Eastern African source (in the absence of genetic data on ancient populations from the region); and (4) Europe_LNBA as a representative of ancient Europeans from the Late Neolithic and Bronze Age (Methods S1I; Table S4; Figure S4). We also applied PHCP to these 17 present-day populations (Methods S1G; Table S4; Figure S4). Comparison of PHCP and LINADMIX shows that they agree well with respect to the Somali and Europe_LNBA component, and therefore also for the combined contribution of Iran_ChL and Megiddo_MLBA (Methods S1G; Figure S4). However, they deviate regarding the respective contributions of Iran_ChL and Megiddo_MLBA (Figure S4), likely because of the fact that the Megiddo_MLBA and Iran_ChL are already very similar populations (Table S3). To only consider results that are robust and shared by LINADMIX and PHCP, we have combined Megiddo_MLBA and Iran_ChL to a single source population representing the Middle East for our main results (Figure 5). We further verified these conclusions, as well as the robustness of the estimations, by using a different representative for the Bronze Age Levantine groups as a source (Tables S4 and S5; Methods S1J) and using perturbations to the ADMIXTURE parameters (Table S4; Methods S1K). Combined, these results suggest that modern populations related to the Levant are consistent with having a substantial ancestry component from the Bronze Age Southern Levant and the Chalcolithic Zagros. Nonetheless, other potential ancestry sources are possible, and more ancient samples might enable a refined picture (Table S4).

The results show that since the Bronze Age, an additional East-African-related component was added to the region (on

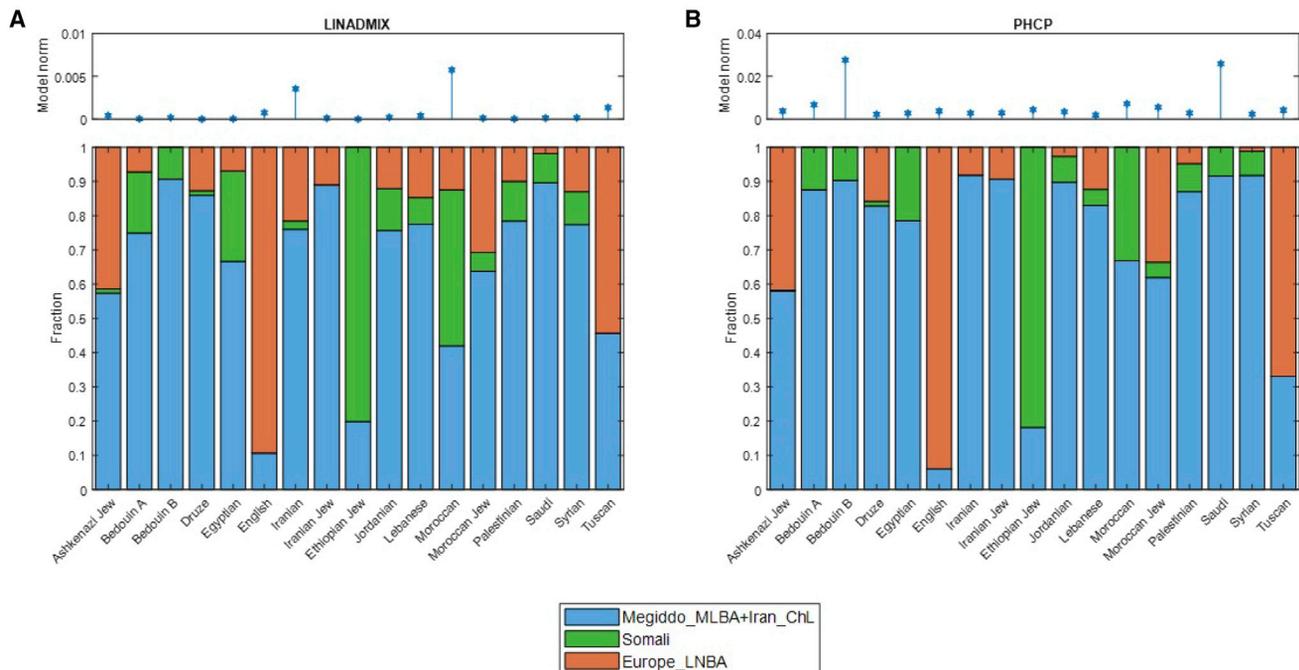


Figure 5. Estimated Fractions Contributed by Different Ancient Populations to Present-Day Groups

Seventeen present-day populations were modeled as an admixture of groups related to four source populations. The upper graphic shows the norm of residuals of the models, and the lower graphic shows the relative contribution of each of the source populations to the present-day target population listed on the x axis.

(A) LINADMIX.

(B) PHCP.

See also [Figure S4](#) and [Tables S4](#) and [S5](#).

average ~10.6%, excluding Ethiopian Jews who harbor ~80% East African component), as well as a European-related component (on average ~8.7%, excluding Ashkenazi Jews who harbor a ~41% European-related component). The East-African-related component is highest in Ethiopian Jews and North Africans (Moroccans and Egyptians). It exists in all Arabic-speaking populations (apart from the Druze). The European-related component is highest in the European control populations (English and Tuscan), as well as in Ashkenazi and Moroccan Jews, both having a history in Europe (Atzmon et al., 2010; Carmi et al., 2014; Schroeter, 2008). This component is present, although in smaller amount, in all other populations except for Bedouin B and Ethiopian Jews. As expected, the English and Tuscan populations have a very low Middle-Eastern-related component. Whereas LINADMIX and PHCP have high uncertainty in estimating the relative contributions of Megiddo_MLBA and Iran_ChL, the results and simulations nevertheless suggest that additional Zagros-related ancestry has penetrated the region since the Bronze Age (Methods S11). Except for the populations with the highest Zagros-related component, PHCP estimates lower magnitudes of this component (Figure S4A), and therefore detection by PHCP of a Zagros-related ancestry is likely an indication for the presence of this component. Indeed, examining the results of LINADMIX and PHCP on all four source populations (Figure S4), we observe a relatively large Zagros-related component in many Arabic-speaking groups, suggesting that gene flow from populations related to those of the Zagros

and Caucasus (although not necessarily from these specific regions) continued even after the Iron Age (Methods S11).

Altogether, the patterns of the present-day populations reflect demographic processes that occurred after the Bronze Age and are plausibly related to processes known from the historical literature (Methods S11). These include an Eastern-African-related component that is present in Arabic-speaking groups but is lower in non-Ethiopian Jewish groups, as well as Zagros-related contribution to Levantine populations, which is highest in the northernmost population examined, suggesting a contribution of populations related to the Zagros even after the Bronze and Iron Ages.

DISCUSSION

Our results provide a comprehensive genetic picture of the primary inhabitants of the Southern Levant during the 2nd millennium BCE, known in the historical record and based on shared material culture as “Canaanites.” We carried out a detailed analysis aimed at answering three basic questions: how genetically homogeneous were these people, what were their plausible origins with respect to earlier peoples, and how much change in ancestry has there been in the region since the Bronze Age?

Earlier genetic analyses modeled the genomes of Middle-to-Late Bronze Age people of the Southern Levant as having almost equal shares of earlier local populations (Levant_N) and populations that are related to the Chalcolithic Zagros (Feldman et al.,

2019; Haber et al., 2017; Lazaridis et al., 2016), suggesting a movement from the northeast into the Southern Levant. Here, we provide more details on this process, taking into account evidence from both archaeology and our temporally and geographically diverse genetic data. Because there is little archaeological evidence of a direct cultural connection between the Southern Levant and the Zagros region in this period, the Caucasus is a more likely source for this ancestry. We used our data to compare these two scenarios and concluded that the genetic data are compatible with both.

The Megiddo outliers, which we inferred to be relative newcomers to the region, are particularly important in demonstrating that the gene flow continued throughout the Bronze Age and that at least some of the gene flow likely came from the Caucasus rather than the Zagros. These two individuals have the highest proportions of Zagros- or Caucasian-related ancestry in our dataset. Analysis of these outliers gave significantly stronger evidence of a Caucasus source compared with a Zagros one, although this conclusion might be revised once ancient DNA data from the Middle-to-Late Bronze Age in the Zagros region become available. The two Megiddo individuals with the next lowest Neolithic Levant component (I10769 and I10770, brothers) were found near the monumental tomb that was likely related to the palace at Megiddo, raising the possibility that they might be associated with the ruling caste. Indeed, a ruler of Taanach (a town located immediately to the south of Megiddo) mentioned in a 15th century BCE cuneiform tablet found at the site and the rulers of Megiddo and Taanach mentioned in the 14th century BCE Amarna letters (found in Egypt) carry Hurrian names (a language spoken in the northeast of the ancient Near East, possibly including the Caucasus) (Na'aman, 1994b). This provides some evidence—albeit so far only suggestive—that at least some of the ruling groups in these (and other) cities might have originated from the northeast of the ancient Near East.

The Caucasus is represented in this study by ancient groups from the present-day country of Armenia, but the region known to have had cultural ties with the Southern Levant is much broader. Evidence of cultural effects on the Southern Levant is mainly focused on the Kura-Araxes culture during the Early Bronze Age (archaeology) and on the Hurrians during the Middle-to-Late Bronze Age (linguistic testimony). These two complexes were spread over the Caucasus, Eastern Anatolia, and neighboring regions. The Armenian sites we analyzed are the best representatives to date of these cultures. The Early Bronze Age individuals from Armenia (Armenia_EBA) come from an Early Bronze Age Kura-Araxes burial ground, and the later Middle-to-Late Bronze Age individuals (Armenia_MLBA) come from the Aragatsotn Province in northwestern Armenia. It is important to note that the Neolithic and Chalcolithic Anatolian individuals analyzed in this study come from the northwestern part of Anatolia, which is not part of the Caucasus. The Chalcolithic Zagros individuals come from the Kangavar Valley in Iran, which is located on the border of the Kura-Araxes influence.

The term “Canaanites” is loosely defined, referring to a collection of groups (which in the Bronze Age were organized in a city-state system) and thus in principle could lack genetic coherence. The individuals examined here cover a wide geographic span—coming from nine sites in present-day Lebanon, Israel, and Jor-

dan. Our analyses revealed that, with the exception of Sidon (and to a smaller extent the individuals of the Baq'ah), they are homogeneous in the sense of being closer to each other than to other contemporary and neighboring populations. This suggests that the archaeological and historical category of “Canaanites” correlates with shared ancestry (Eisenmann et al., 2018). This resembles the pattern observed in the Aegean basin during the 2nd millennium BCE, where the cultural categories of “Minoan” and “Mycenaean” show evidence of genetic homogeneity across multiple sites albeit with potentially subtle ancestry differences within these groupings (Lazaridis et al., 2017). Another example is the “Yamnaya” pastoralists of late 3rd and early 2nd millennium BCE in the western Eurasian Steppe (Allentoft et al., 2015; Haak et al., 2015). This contrasts with the pattern seen in other places, such as for the Bell Beaker cultural complex of the 2nd millennium BCE (Olalde et al., 2018), where people sharing similar cultural practices had widely varying ancestry. In any case, the detection of such associations—as we do here—cannot by itself prove that group identities in the past were related to genetics.

From the groups we have examined, the only one that is somewhat diverged from the rest is Sidon. We provide evidence against the possibility that this observation is a batch effect (Methods S1C). Rather, we suggest that the relative remoteness of Sidon stems from the fact that this population is genetically heterogeneous and has different individuals showing resemblance to different Southern Levantine groups (Methods S1C). During the 2nd millennium BCE, Sidon was a major port city and was connected in trading relations with the eastern Mediterranean basin, which could have led to a significant genetic inflow, making its population more heterogeneous than that of inland cities. This might also be the reason that the site that most resembles Sidon is Ashkelon, which is another coastal site. The only inland population that resembles Sidon is Abel Beth Maacah, perhaps because of its geographic proximity (Figures 1A and 2). Apart from Sidon, Baq'ah also shows some minor deviations from the rest when taking a richer set of outgroup populations (Figure 3). The Baq'ah is located on the fringe of the Syrian desert, therefore this population might be admixed with more eastern groups, which are not yet genetically sampled. This might be reflected by the fact that the individuals of the Baq'ah also show some degree of variability in their ancestry patterns (Table S2).

Although this study focuses on the Bronze Age, it also reports two new samples from the Iron Age—one from Megiddo and the other from Abel Beth Maacah. These two individuals show ancestry patterns that are very similar to those observed in the Middle and Late Bronze Age individuals (Figure 4), suggesting that the destruction at the end of the Bronze Age in the region did not necessarily lead to genetic discontinuity in each and every site. Notably, both Abel Beth Maacah and Megiddo are inland cities, and their genetic continuity throughout the transition from the Bronze Age to the Iron Age might not be representative of other sites in the region. For example, one of the two Iron Age populations in the Philistine coastal city of Ashkelon (ASH_IA1) showed evidence of mobility of populations related to southern Europe around the Bronze Age to Iron Age transition (Feldman et al., 2019).

Estimating the ancestry proportions in present-day Middle Eastern populations with substantial sub-Saharan African admixture (as well as multiple sources of admixture from different parts of the Mediterranean), is difficult. We addressed the problem by developing two statistical techniques and then testing the robustness of our inference on the basis of a comparison between these methods, simulations, and perturbations of the input (see [STAR Methods](#); [Methods S1F–S1K](#)). We examined 14 present-day populations that are historically or geographically linked to the Southern Levant and tested the contributions of East Africa, Europe, and the Middle East (combining Southern Levant Bronze Age populations and Zagros-related Chalcolithic ones) to their ancestry. We found that both Arabic-speaking and Jewish populations are compatible with having more than 50% Middle-Eastern-related ancestry. This does not mean that any these present-day groups bear direct ancestry from people who lived in the Middle-to-Late Bronze Age Levant or in Chalcolithic Zagros; rather, it indicates that they have ancestries from populations whose ancient proxy can be related to the Middle East. The Zagros- or Caucasian-related ancestry flow into the region apparently continued after the Bronze Age. We also see an Eastern-African-related ancestry entering the region after the Bronze Age with an approximate south-to-north gradient. In addition, we observe a European-related ancestry with the opposite gradient (north-to-south). Given the difficulties in separating the ancestry components arriving from the Southern Levant and the Zagros, an important direction for future work will be to reconstruct in high resolution the ancestry trajectories of each present-day group, and to understand how people from the Southern Levant Bronze Age mixed with other people in later periods in the context of processes known from the rich archaeological and historical records of the last three millennia.

STAR★METHODS

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

- [KEY RESOURCES TABLE](#)
- [LEAD CONTACT AND MATERIALS AVAILABILITY](#)
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
 - Tel Megiddo
 - Tel Abel Beth Maacah
 - Tel Hazor
 - Yehud
 - Baqāh
- [METHOD DETAILS](#)
 - Data Generation
 - Chemical and isotopic measurements
- [QUANTIFICATION AND STATISTICAL ANALYSES](#)
 - PCA
 - ADMIXTURE
 - f_4 -statistics
 - LINADMIX
- [PHCP](#)

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.04.024>.

ACKNOWLEDGMENTS

We thank Reuven Amitai, Menachem Ben-Sasson, Tom Booth, Pontus Skoglund, Philipp Stockhammer, Jason Ur, and several anonymous reviewers for critical comments. This study was funded by the Israel Science Foundation ISF grant 1009/17 to L.C. and B.Y. and ISF grant 407/17 to S.C. D.R. is an Investigator of the Howard Hughes Medical Institute and his ancient DNA laboratory work was supported by National Science Foundation HOMINID grant BCS-1032255, by National Institutes of Health grant GM100233, by an Allen Discovery Center grant, and by grant 61220 from the John Templeton Foundation. Work at Megiddo is supported by the Dan David Foundation, the Shmunis Family Foundation, Mark Weismann, and Vivian and Norman Belmonte.

AUTHOR CONTRIBUTIONS

Conceptualization, L.A.-T., R.P., Y.E., B.Y., S.C., I.F., L.C., and D.R.; Methodology, L.A.-T., S.W., B.Y., S.C., L.C., and D.R.; Formal Analysis, L.A.-T., S.W., N.M., B.Y., S.C., L.C., and D.R.; Investigation, L.A.-T., S.W., M.A.S.M., D.G., T.E., O.C., N.R., S.M., N.A., A.M.L., M. Mah, M. Michel, J.O., K.S., F.C., D.K., B.G., S.T., R.K., S.B., V.E., D.J.K., M.F., N.Y.-M., J.M.M., Y.G., Y.E., B.Y., R.P., S.C., I.F., L.C., and D.R.; Resources, M.A.S.M., S.B., V.E., N.Y.-M., J.M.M., Y.G., R.P. and I.F.; Data Curation, S.M. and D.R.; Writing, L.A.-T., S.W., T.E., Y.E., B.Y., S.C., I.F., L.C., and D.R.; Supervision, Y.E., R.P., S.C., I.F., L.C., and D.R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 25, 2019

Revised: February 10, 2020

Accepted: April 15, 2020

Published: May 28, 2020

REFERENCES

- Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664.
- Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172.
- Atzmon, G., Hao, L., Pe'er, I., Velez, C., Pearlman, A., Palamara, P.F., Morrow, B., Friedman, E., Oddoux, C., Burns, E., and Ostrer, H. (2010). Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am. J. Hum. Genet.* 86, 850–859.
- Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* 90, 675–684.
- Beherec, M.A., Levy, T.E., Tirosh, O., Najjar, M., Knabb, K.A., and Erel, Y. (2016). Iron Age Nomads and their relation to copper smelting in Faynan (Jordan): Trace metal and Pb and Sr isotopic measurements from the Wadi Fidan 40 cemetery. *J. Archaeol. Sci.* 65, 70–83.
- Bentley, R.A. (2006). Strontium Isotopes from the Earth to the Archaeological Skeleton: A Review. *J. Archaeol. Method Theory* 13, 135–187.
- Bienkowski, P. (1999). Jonathan N. Tubb. *Canaanites (Peoples of the Past)*. 160 pages, 18 colour, 106 black-and-white illustrations. 1998. London: British Museum Press; 0-7141-2089-8 hardback £20. *Antiquity* 73, 708–709.

- Carmi, S., Hui, K.Y., Kochav, E., Liu, X., Xue, J., Grady, F., Guha, S., Upadhyay, K., Ben-Avraham, D., Mukherjee, S., et al. (2014). Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat. Commun.* **5**, 4835.
- Cline, E.H. (2014). *1177 B.C.: the Year Civilization Collapsed* (Princeton: Princeton University Press).
- Dabney, J., Meyer, M., and Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harb. Perspect. Biol.* **5**, a012567, a012567.
- de Miroschedji, P. (2014). The Southern Levant (Cisjordan) during the Early Bronze Age. In *The Oxford Handbook of the Archaeology of the Levant, c. 8000–332 BCE*, M.L. Steiner and A.E. Killebrew, eds. (Oxford University Press).
- Eisenmann, S., Bánffy, E., van Dommelen, P., Hofmann, K.P., Maran, J., Lazaridis, I., Mittnik, A., McCormick, M., Krause, J., Reich, D., and Stockhammer, P.W. (2018). Reconciling material cultures in archaeology with genetic data: The nomenclature of clusters emerging from archaeogenomic analysis. *Sci. Rep.* **8**, 13003.
- Erel Y., Listovsky N., Matthews A., Ilani S., and Avni Y. (2006) Tracing end-member fluid sources in sub-surface iron mineralization and dolomitization Along a proximal fault to the Dead Sea Transform. *Geochim. Cosmochim. Acta*, <https://10.1016/j.gca-2006.08.019>.
- Ericson, J.E., Shirahata, H., and Patterson, C.C. (1979). Skeletal concentrations of lead in ancient Peruvians. *N. Engl. J. Med.* **300**, 946–951.
- Feldman, M., Master, D.M., Bianco, R.A., Burri, M., Stockhammer, P.W., Mittnik, A., Aja, A.J., Jeong, C., and Krause, J. (2019). Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. *Sci. Adv.* **5**, eaax0061.
- Finnegan, M., Husted, J.J., Rolston, S.L., and Saul, M.B. (1986). The human skeletal remains. In *The Late Bronze and Early Iron Ages of Central Transjordan: The Baq'ah Valley Project, 1977–1981*, P.E. McGovern, ed. (The University Museum Press), pp. 295–314.
- Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., and Pääbo, S. (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. USA* **110**, 2223–2227.
- Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kóvári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257.
- Goren, Y., Finkelstein, I., Na'aman, N., and Artzy, M. (2004). Makhon le-arkhe'ologyah 'a. sh. Sonyah u-Mar'ko Nadler. *Inscribed in clay : provenance study of the Amarna tablets and other ancient Near Eastern texts* (Emery and Claire Yass Publications in Archaeology).
- Greenberg, R., and Goren, Y. (2009). Transcaucasian Migrants and the Khirbet Kerak Culture in the Third Millennium BCE (Tel-Aviv).
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211.
- Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S., Martiniano, R., Prado-Martinez, J., Szpak, M., et al. (2017). Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences. *Am. J. Hum. Genet.* **101**, 274–282.
- Harney, É., May, H., Shalem, D., Rohland, N., Mallick, S., Lazaridis, I., Sarig, R., Stewardson, K., Nordenfelt, S., Patterson, N., et al. (2018). Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun.* **9**, 3336.
- Hartman, G., and Richards, M. (2014). Mapping and defining sources of variability in bioavailable strontium isotope ratios in the Eastern Mediterranean. *Geochim. Cosmochim. Acta* **126**, 250–264.
- Horstwood, M.S.A., Evans, J.A., and Montgomery, J. (2008). Determination of Sr isotopes in calcium phosphates using laser ablation inductively coupled plasma mass spectrometry and their application to archaeological tooth enamel. *Geochim. Cosmochim. Acta* **72**, 5659–5674.
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684.
- Korlević, P., Gerber, T., Gansauge, M.-T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., and Meyer, M. (2015). Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93.
- Korneliusson, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356.
- Kozlov, K., Chebotarev, D., Hassan, M., Triska, M., Triska, P., Flegontov, P., and Tatarinova, T.V. (2015). Differential Evolution approach to detect recent admixture. *BMC Genomics* **16** (Suppl 8), S9.
- Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453.
- Lawson, D.J., van Dorp, L., and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nat. Commun.* **9**, 3258.
- Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413.
- Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424.
- Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwängler, A., Peltzer, A., Posth, C., Vasilakis, A., et al. (2017). Genetic origins of the Minoans and Mycenaeans. *Nature* **548**, 214–218.
- Lemche, N.P. (1991). *The Canaanites and their land : the tradition of the Canaanites* (Sheffield: JSOT Press).
- Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Lawson, D.J., et al.; Wellcome Trust Case Control Consortium 2; International Multiple Sclerosis Genetics Consortium (2015). The fine-scale genetic structure of the British population. *Nature* **519**, 309–314.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993.
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595.
- Li, N., and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* **165**, 2213–2233.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Martin, M.A.S., Finkelstein, I., and Piasetzky, E. (2020). Radiocarbon dating the Late Bronze Age: Cultural and Historical Considerations on Megiddo and Beyond. *Bull. Am. Schools Orient. Res.*, in press.
- Mazar, A. (1992). *Archaeology of the land of the Bible: 10,000–586 B.C.E.* (Anchor Bible Reference Library).
- McGovern, P. (1986). *The late bronze and early iron ages of central Transjordan, the Baq'ah Valley project, 1977–1981* (Philadelphia, PA: University Museum, University of Pennsylvania).
- Na'aman, N. (1994a). The Canaanites and Their Land. *Ugarit-Forschungen* **26**, 397–418.
- Na'aman, N. (1994b). The Hurrians and the End of the Middle Bronze Age in Palestine. *Levant* **26**, 175–187.
- Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196.
- Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C.J., Gandini, F., Pala, M., et al. (2019). The genomic

history of the Iberian Peninsula over the past 8000 years. *Science* 363, 1230–1234.

Patterson, C., Ericson, J., Manea-Krichten, M., and Shirahata, H. (1991). Natural skeletal levels of lead in *Homo sapiens sapiens* uncontaminated by technological lead. *Sci. Total Environ.* 107, 205–236.

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

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

Perry, M.A. (2002). Life and Death in Nabataea: The North Ridge Tombs and Nabataean Burial Practices. *Near East. Archaeol.* 65, 265.

Perry, M.A., Coleman, D., and Delhopital, N. (2008). Mobility and exile at 2nd century A.D. khirbet edh-dharih: Strontium isotope analysis of human migration in western Jordan. *Geoarchaeology* 23, 528–549.

Perry, M.A., Coleman, D.S., Dettman, D.L., Grattan, J.P., and Halim al-Shiyab, A. (2011). Condemned to metallum? The origin and role of 4th–6th century A.D. Phaeno mining camps using multiple chemical techniques. *J. Archaeol. Sci.* 38, 558–569.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., et al. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* 10, e0129102.

Pinhasi, R., Fernandes, D.M., Sirak, K., and Cheronet, O. (2019). Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc.* 14, 1194–1205.

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

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. *Nature* 488, 370–374.

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2014). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20130624, 20130624.

Schroeter, D.J. (2008). The Shifting Boundaries of Moroccan Jewish Identities. *Jewish Social Studies*, pp. 145–164. Conference on Jewish Relations.

Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. USA* 111, 2229–2234.

Toffolo, M.B., Arie, E., Martin, M.A.S., Boaretto, E., and Finkelstein, I. (2014). Absolute Chronology of Megiddo, Israel, in the Late Bronze and Iron Ages: High-Resolution Radiocarbon Dating. *Radiocarbon* 56, 221–244.

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

Wang, J., Raskin, L., Samuels, D.C., Shyr, Y., and Guo, Y. (2015). Genome measures used for quality control are dependent on gene function and ancestry. *Bioinformatics* 31, 318–323.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44 (W1), W58–63.

Yasur-Landau, A. (2010). *The Philistines and Aegean Migration at the End of the Late Bronze Age* (Cambridge University Press).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and analyzed data from 73 newly reported ancient humans (see Table S1 for details)	This paper	ENA: PRJEB37057
Chemicals, Peptides, Recombinant Proteins		
Pfu Turbo Cx Hotstart DNA Polymerase	Agilent Technologies	600412
Herculase II Fusion DNA Polymerase	Agilent Technologies	600679
2x HI-RPM hybridization buffer	Agilent Technologies	5190-0403
0.5 M EDTA pH 8.0	BioExpress	E177
Sera-Mag™ Magnetic Speed-beads™ Carboxylate-Modified (1mm, 3EDAC/PA5)	GE LifeScience	65152105050250
USER enzyme	New England Biolabs	M5505
UGI	New England Biolabs	M0281
Bst DNA Polymerase2.0, large frag.	New England Biolabs	M0537
PE buffer concentrate	QIAGEN	19065
Proteinase K	Sigma Aldrich	P6556
Guanidine hydrochloride	Sigma Aldrich	G3272
3M Sodium Acetate (pH 5.2)	Sigma Aldrich	S7899
Water	Sigma Aldrich	W4502
Tween-20	Sigma Aldrich	P9416
Isopropanol	Sigma Aldrich	650447
Ethanol	Sigma Aldrich	E7023
5M NaCl	Sigma Aldrich	S5150
1M NaOH	Sigma Aldrich	71463
20% SDS	Sigma Aldrich	5030
PEG-8000	Sigma Aldrich	89510
1 M Tris-HCl pH 8.0	Sigma Aldrich	AM9856
dNTP Mix	Thermo Fisher Scientific	R1121
ATP	Thermo Fisher Scientific	R0441
10x Buffer Tango	Thermo Fisher Scientific	BY5
T4 Polynucleotide Kinase	Thermo Fisher Scientific	EK0032
T4 DNA Polymerase	Thermo Fisher Scientific	EP0062
T4 DNA Ligase	Thermo Fisher Scientific	EL0011
Maxima SYBR Green kit	Thermo Fisher Scientific	K0251
50x Denhardt's solution	Thermo Fisher Scientific	750018
SSC Buffer (20x)	Thermo Fisher Scientific	AM9770
GeneAmp 10x PCR Gold Buffer	Thermo Fisher Scientific	4379874
Dynabeads MyOne Streptavidin T1	Thermo Fisher Scientific	65602
Salmon sperm DNA	Thermo Fisher Scientific	15632-011
Human Cot-I DNA	Thermo Fisher Scientific	15279011
DyNAmo HS SYBR Green qPCR Kit	Thermo Fisher Scientific	F410L
Methanol, certified ACS	VWR	EM-MX0485-3
Acetone, certified ACS	VWR	BDH1101-4LP
Dichloromethane, certified ACS	VWR	EMD-DX0835-3
Hydrochloric acid, 6N, 0.5N & 0.01N	VWR	EMD-HX0603-3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
High Pure Extender from Viral Nucleic Acid Large Volume Kit	Roche	5114403001
MinElute PCR Purification Kit	QIAGEN	28006
NextSeq® 500/550 High Output Kit v2 (150 cycles)	Illumina	FC-404-2002
Hydrogen peroxide 30% Suprapur®	Merck	107298
Hydrochloric acid, 1N	Biolab-chemicals.com	084105
Nitric Acid 3.5N, 0.05N	Biolab-chemicals.com	147005
Software and Algorithms		
Samtools	Li, 2011; Li et al., 2009	http://samtools.sourceforge.net/
BWA	Li and Durbin, 2010	http://bio-bwa.sourceforge.net/
ADMIXTOOLS	Patterson et al., 2012	https://github.com/DReichLab/AdmixTools
SeqPrep	https://github.com/jstjohn/SeqPrep	https://github.com/jstjohn/SeqPrep
smartpca	Patterson et al., 2006	https://www.hsph.harvard.edu/alkes-price/software/
ADMIXTURE	Alexander et al., 2009	http://dalexander.github.io/admixture/download.html
PMDtools	Skoglund et al., 2014	https://github.com/pontusssk/PMDtools
Haplogrep2	Weissensteiner et al., 2016	http://haplogrep.uibk.ac.at/index.html
ContamMix	Fu et al., 2013	https://github.com/DReichLab/ADNA-Tools
ANGSD	Korneliussen et al., 2014	https://github.com/ANGSD/angsd
mapDamage2.0	Jónsson et al., 2013	https://ginolhac.github.io/mapDamage/
PHCP	This paper	https://github.com/ShamamW/PHCP
LINADMIX	This paper	liran.carmel@huji.ac.il

LEAD CONTACT AND MATERIALS AVAILABILITY

Lead Contact

Further information and requests for materials should be directed to and will be fulfilled by the Lead Contact, Liran Carmel (liran.carmel@huji.ac.il).

Materials Availability

All data generated in this study is available, see [Key Resources Table](#).

Data and Code Availability

Genomic data first reported here is available at <http://carmelab.huji.ac.il/data.html>. Raw and analyzed data from the 73 newly reported ancient samples were deposited at ENA: PRJEB37057. Codes used for LINADMIX are available upon request from Liran Carmel (liran.carmel@huji.ac.il). Code for PHCP is available from <https://github.com/ShamamW/PHCP>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The 73 Bronze and Iron Age individuals newly reported in this work come from five archaeological sites in Israel and Jordan. Below, we provide details on these sites, and the individuals they harbored and that we analyzed.

Tel Megiddo

Tel Megiddo has been excavated by four expeditions, starting in the early 20th century. The site was inhabited from Neolithic times to the Persian period. Remains of over 30 settlements have been uncovered. Megiddo is a key site for the study of the Bronze and Iron Ages in the Levant and beyond. This is due to the combination of tight control over stratigraphy, ceramic typology, and radiocarbon dating, and vast exposures of remains, including significant monuments. Megiddo was the hub of a Canaanite city-state in the Bronze and Iron I Ages and an administrative center of the biblical kingdom of Israel in the Iron II. It is mentioned in Egyptian and Assyrian

texts (possibly also in one Hittite text) and in various associations in the Hebrew Bible. This is the location of Armageddon (a Greek corruption of the Hebrew *har-Megiddo* = the mound of Megiddo) of the Book of Revelations in the New Testament.

The samples discussed here, 35 altogether (Table S1), cover a long period of time, from the Intermediate Bronze Age (ca. 2500–2000 BCE) to the late Iron I (ca. 1000 BCE). With the exception of sample I4517 (late Iron I), all were retrieved from intramural burials scattered over the mound. The majority of the samples date to the Middle Bronze III-Late Bronze I (ca. 1650–1400 BCE). The Megiddo samples come from four excavation fields:

- Area J: the cult compound of Megiddo, in the eastern sector of the mound;
- Area K: domestic buildings, burials, and fortifications in the southeastern sector of the mound;
- Area H in the north: buildings (and burials) immediately to the west of the Middle Bronze III-Late Bronze palaces of Megiddo;
- Area M: in the center of the mound, remains of a public Late Bronze building (palace?) and stone-built Late Middle Bronze/Early Late Bronze tombs.

Samples originated from the following layers (the letter designates the area and the number represents the stratum):

- Level J-8 dating to the Middle Bronze I in the early second millennium BCE.
- Level J-11/12 dating to the Middle Bronze II.
- Level K-12, Middle Bronze II, domestic architecture, burials under a house near the brick-built fortification wall.
- Level K-11, Middle Bronze III, same type of context as Level K-12.
- Level K-10, Middle Bronze III-Late Bronze I, same type of context as Level K-12. Habitation terminated at the end of the Middle Bronze III, while burials continue in the Late Bronze I.
- Level H-16, Middle Bronze III; the sample (I10771) comes from a monumental stone-built tomb, possibly associated with the palace to its east.
- Level H-15, Late Bronze I. The sampled burial, with two individuals (I10769 and I10770), was found above the monumental tomb of Level H-16; those who dug it were probably aware of this association.
- Level M-7, Late Bronze I; elaborate stone-built tombs nearby.
- Level K-4, late Iron I, large courtyard house destroyed in big fire. The individual (I4517) seems to have perished during the devastation of the city.

Two of the outliers, the brother and sister (I2189 and I2220), were buried next to each other, while the third outlier, I10100, was buried in the same grave—a stone-lined cist—as the sister, but in an earlier deposit in that grave.

Tel Abel Beth Maacah

Tel Abel Beth Maacah (Tell Abil el-Qameh) is a large site located in northern Israel, commanding several roads leading to the Lebanese inland Bekaa, the Phoenician coast and Damascus. The town of Abel is mentioned in several Egyptian 2nd millennium BCE sources and also three times in the bible, twice in relation to Aramean and Assyrian conquests in the 9th and 8th centuries BCE, as well as in the story about a rebellion against King David in the 10th century BCE.

Surveys and six seasons of excavation since 2012 have shown that the mound was occupied from the Early Bronze II-III (ca. 3000–2500 BCE) until modern times, with apparent hiatuses in the Middle Bronze I (ca. 1950–1750 BCE), Iron IIB-C (ca. 800–586 BCE) and parts of the Middle Ages. In the Middle Bronze II (ca. 1750–1650 BCE), the site was fortified and co-existed with other larger fortified sites nearby, including Dan and Hazor. During the Late Bronze Age, the town was apparently located mainly in the lower part of the mound, reusing the Middle Bronze fortifications, and probably was part of the kingdom of nearby Hazor. With the destruction of Hazor in the 13th century BCE, Abel Beth Maacah grew in size and occupation complexity, demonstrating one of the most intense sequences known in Iron Age I (12th mid–10th centuries BCE). At this time, the site included a large public complex with administrative, storage, industrial, and cultic functions. A destruction at the end of the Iron I is radiocarbon-dated to the 10th century BCE. Iron IIA remains were built directly above this destruction, demonstrating substantial architecture, including a large citadel at the top of the mound in the north, radiocarbon-dated to the 9th century BCE.

The sample described here was extracted from a complete skeleton of an adult male excavated in Area O on the western slope of the lower mound. The grave was cut into a room belonging to a courtyard building dated to the Middle Bronze II. Nearby, in topsoil, a seal dated to Iron IIA (ca. 950–800 BCE) was found. Other pits cut into the building contained pottery that can be dated to this period, along with pottery from Iron I.

Tel Hazor

Hazor is the largest Bronze and Iron Age site in Israel, covering some 200 acres. The mound is composed of an upper mound (acropolis) adjoining a huge lower mound (lower city) to its north. Occupation began in the upper mound during the Early Bronze Age II (early third millennium BCE), while the lower city was founded in the Middle Bronze Age II (approximately the 18th century BCE). Both continued to be settled until a later phase of the Late Bronze Age (13th century), when the upper and lower cities were violently destroyed or abandoned. Following this destruction, only the upper part of the mound was resettled and fortified, becoming a major city in the 10th to 8th centuries BCE, as part of the Israelite kingdom.

Canaanite Hazor is mentioned on several occasions in ancient Near Eastern texts, the earliest being the Egyptian Excretion texts of the 19th century BCE. Hazor is the only Canaanite site mentioned in the archive discovered in Mari in Syria (18th century BCE according to the middle chronology, or the 17th century BCE according to the ultra-low chronology). The Mari documents clearly demonstrate the importance, wealth and far-reaching commercial ties of Hazor. In the Amarna archive (14th century BCE) there are several references to Hazor, as well as in records of the military campaigns conducted by Egyptian pharaohs, during the 15th–14th centuries BCE.

Of the three individuals included in this study, one was found on the acropolis (I3965), in a cist grave with 2–3 individuals buried with a rich assemblage of goods. The tomb is related to Stratum pre-XVII, dated to the Middle Bronze I-II transition (ca. 1750 BCE). The two other samples were derived from burials in Area M, on the northern slope of the acropolis. Both contained young individuals or children. One (I3966) was buried beside the monumental Middle Bronze staircase; this burial too dates to the Middle Bronze I-II transition. The other was buried into a drainage channel (I3832) associated with the administrative palace of the Late Bronze Age and is dated to the Late Bronze II (ca. 14th century BCE).

Yehud

Tel Yehud (Tell el-Yehudia) is situated on the northeastern side of the Ono valley in the eastern part of the central coastal plain of Israel, ca. 12 km east of the Mediterranean Sea. Rescue excavations were carried out in 2008, followed by an excavation season in 2009 in Areas A and B, in the location of an underground parking lot. Archaeological findings at these areas include a deep shaft filled with refuse dated to the Chalcolithic Period, a cemetery from the Intermediate Bronze Age, a Late Roman-Byzantine pottery workshop, and Early Islamic cist graves.

The human remains analyzed in this study are of 13 individuals from the Intermediate Bronze Age.

Baq'ah

Human skeletal remains were excavated from the Baq'ah Valley, Jordan, approximately 20 km NW from Amman between 1977 and 1981 (McGovern, 1986). The skeletal materials spanned from the Late Bronze to the Iron Age (Caves A2, B3, and A4). Only materials from the Cave B3 context are included in these analyses. Skeletal materials from Cave A4 (Iron Age IA) were shipped to the Smithsonian Institution and the Penn Museum received the materials from A2 and B3 in 1982 from Jordan (University of Pennsylvania Museum Archival correspondence).

The tomb at Cave B3 (Late Bronze Age II) is one of a series of caves, and the only cave excavated. It is located on the lower slopes of Jebel al-Qeşir. Human remains in B3 occur in 2 units: the lower and upper burial remains. The skeletal remains are disassociated and fragmentary; some human skeletal material is also burned/charred, as a result of local depositional history (alkaline soil, percolation of subsoil groundwater). The B3 collection (Finnegan et al., 1986) includes a minimum number estimate of 64 individuals.

Sampling took place at the University of Pennsylvania Museum in March 2015. All samples were derived from Cave B3 petrous temporals. 11 males and 10 female temporal bones were sampled. The age range estimate is from late adolescent to mature adult.

METHOD DETAILS

Data Generation

We prepared powder from skeletal remains in dedicated clean rooms. Of the samples that produced working data, all but one came from petrous bones, which are known to yield up to one hundred times more DNA than other skeletal elements (Gamba et al., 2014; Pinhasi et al., 2015, 2019). We extracted DNA using a method designed to retain short and highly degraded fragments (Dabney et al., 2013; Korlević et al., 2015; Skoglund et al., 2014), and built the DNA into individual barcoded double-stranded libraries in the presence of the enzyme Uracil–DNA–glycosylase (UDG) to reduce the rate of cytosine-to-thymine errors typical of ancient DNA (Rohland et al., 2014). We enriched the libraries for sequences overlapping the mitochondrial genome as well as about 1.2 million single nucleotide polymorphisms (SNPs), and sequenced on an Illumina NextSeq500 instrument using 2x76 base pair reads and 2x7 base pair indices. We assigned reads to samples based on the match rate to expected barcodes and indices. We merged read pairs that overlapped by at least 15 nucleotides (allowing up to one mismatch) and represented each nucleotide by the read giving higher quality data. We mapped merged sequences to either the mitochondrial reference genome *rsrs* (Behar et al., 2012; Weissensteiner et al., 2016) or to the human genome reference sequence *hg19* using *bwa* (v.0.6.1) (Li and Durbin, 2010), and removed sequences that were duplicates as assessed by having the same barcode, as well as the same start and stop position when mapped to the reference genome. We determined the genotype of each SNP based on a randomly selected single read (tools used include Fu et al., 2013; Jónsson et al., 2013; Korneliusson et al., 2014; Li, 2011; Li et al., 2009).

Chemical and isotopic measurements

Tooth enamel was carefully separated from dentin, following the protocol described in Beherec et al., (2016). Bones and enamel samples were then cleaned using the methodology of Patterson and co-workers (Ericson et al., 1979; Patterson et al., 1991), which involved several steps of sample pre-cleaning and the use of clean laboratory techniques and reagents. Following the mechanical and chemical cleaning process, the samples were dissolved by concentrated, distilled HNO₃ (25–100 mg sample in 1 mL HNO₃). After sample dissolution, Sr concentrations were determined with an ICP-MS (Agilent 7500cx), following the protocol outlined in Beherec

et al., (2016). Then, all samples were subjected to Strontium isotope analysis in order to trace migration of the individuals (Beherec et al., 2016; Bentley, 2006; Hartman and Richards, 2014; Horstwood et al., 2008; Perry, 2002; Perry et al., 2008, 2011). For this, the samples went through ion exchange columns according to the protocol outlined in (Erel et al., 2006). Then, the samples were analyzed by a MC-ICP-MS (Thermo, NEPTUNE Plus) as described in Zipori et al., 2015. Replicate measurements of $^{87}\text{Sr}/^{86}\text{Sr}$ of SRM-987 standard over the course of this study yielded $^{87}\text{Sr}/^{86}\text{Sr} = 0.710273 \pm 10$ (2 s), $n = 29$.

QUANTIFICATION AND STATISTICAL ANALYSES

PCA

For PCA plots we used SMARTPCA (Patterson et al., 2006).

ADMIXTURE

We ran ADMIXTURE (Alexander et al., 2009) on a set of 1,663 individuals from Europe, Western Asia, and Africa. For analyses solely involving ancient populations, we pruned SNPs that are in linkage disequilibrium using the *indep-pairwise* flag in PLINK (Purcell et al., 2007), with parameters 200, 25 and 0.5, as in Lazaridis et al., (2014). This gave a set of 357,334 SNPs, denoted here ‘ancient’. For analyses involving modern populations, we used the stricter pruning recommended in the ADMIXTURE manual, using the parameters 50, 10, and 0.1, and only using SNPs that are missing in fewer than 10% of the individuals. This resulted in a list of 50,165 SNPs denoted here as the ‘modern’ panel. To find an optimal K we followed the default procedure of running 5-fold cross validation. For the ancient SNPs we tested K values ranging from 4 to 11, and for the modern SNPs we tested K values ranging from 4 to 14. We selected $K = 6$ for both analyses (Figure S1A).

f_4 -statistics

We computed f -statistics using the package ADMIXTOOLS (Patterson et al., 2012). To test for homogeneity we ran *qpWave* using the *allsnps:YES* parameter. To estimate the ancestry proportion for a test population given a set of source populations and a set of outgroups, we used the *qpAdm* methodology (Lazaridis et al., 2016) in ADMIXTOOLS with the *allsnps:YES* parameter.

LINADMIX

The ADMIXTURE algorithm (Alexander et al., 2009) takes as input a set of genotypes and assumes that each is a mixture of K ancestral populations. For each genotype i , it estimates the fraction contributed by ancestral population k , denoted q_{ik} . As a result, ADMIXTURE may be viewed as representing a genotype i as a vector of length K ,

$$q_i = \begin{pmatrix} q_{i1} \\ q_{i2} \\ \vdots \\ q_{iK} \end{pmatrix}.$$

Similar to Leslie et al. (Kozlov et al., 2015; Leslie et al., 2015), we sought to utilize this representation of genotypes in order to find how a *target* representation q_t might be related to a list of n *source* representations q_{s_1}, \dots, q_{s_n} . We do it by relating to a linear model, whereby we seek to model the target as a linear combination of the source populations,

$$q_t = \alpha_1 \cdot q_{s_1} + \alpha_2 \cdot q_{s_2} + \dots + \alpha_n \cdot q_{s_n},$$

Where the coefficients $\alpha_1, \dots, \alpha_n$ are called the *mixing coefficients*, and we required that they be nonnegative. Formally, the mixing coefficients are the solutions of the constrained non-negative least-squares problem

$$\min_{\alpha} \| Q_s \alpha - q_t \|_2^2$$

subject to

$$\sum_{i=1}^n \alpha_i = 1$$

$$\alpha_i \geq 0,$$

where $\alpha = (\alpha_1, \dots, \alpha_n)^T$ and Q_s is the $K \times n$ matrix $Q_s = (q_{s_1}, \dots, q_{s_n})$.

In many cases the target and source genotypes are populations rather than individuals. In these cases, we take the vectors q_i as the average across all individuals that belong to the i 'th population. More formally, if population i is made of the m_i individuals l_1, \dots, l_{m_i} , then

$$q_i = \frac{1}{m_i} (q_{l_1} + \dots + q_{l_{m_i}}).$$

In order to find the standard error of the mixing coefficients we used parametric bootstrap. We sampled $q^{(b)}$ -vectors and computed the mixing coefficients for the sampled vectors. Given that l is an individual that belongs to population i , his/her bootstrap q -vectors, $q_l^{(b)}$, are computed by the following steps:

1. Estimation of the covariance matrix of the population, $\widehat{\Sigma}_i = \text{cov}(q_{l_1}, \dots, q_{l_m})$, representing variation between individuals in the population. For populations with a single individual we use the covariance matrix of another close population.
2. Estimation of the covariance matrix of the individual, representing variation in the vector q estimated by ADMIXTURE due to the finite size of the genome. This matrix is approximated by the inverse of the empirical Fisher information matrix, $\widehat{\Sigma}_l = I_l^{-1}(q_l, F)$, where I is the empirical Fisher information matrix and F is the ADMIXTURE output for the frequencies of the minor alleles in each theoretical ancestral population (see below for details).
3. Drawing a vector $q_l^{(b)}$ from $N(q_l, \widehat{\Sigma}_i + \widehat{\Sigma}_l)$.

Finally, we used the results of all bootstrap samples to calculate the standard errors of the mixing coefficients.

The covariance matrix defined in step (2) above is computed as follows. In the ADMIXTURE model there is a large set of individuals L , a large set of markers J , and a much smaller set of K theoretical ancestral populations. The genotype of individual l at marker j is $g_{lj} \in \{0, 1, 2\}$, the number of occurrences of the minor allele at j . The proportion of the genome of individual l that comes from ancestral population k is q_{lk} , which satisfies $\sum_{k=1}^K q_{lk} = 1$. The frequency of the minor allele at marker j in ancestral population k is f_{kj} . The ADMIXTURE output comprises the matrices $Q = (q_{lk})$ and $F = (f_{kj})$.

The joint log-likelihood of the parameters given the data are (up to constant terms)

$$\ln(Q, F|G) = \sum_{l=1}^L \sum_{j=1}^J \left\{ g_{lj} \ln \left[\sum_{k=1}^K q_{lk} f_{kj} \right] + (2 - g_{lj}) \ln \left[\sum_{k=1}^K q_{lk} (1 - f_{kj}) \right] \right\}.$$

Because $q_{lK} = 1 - \sum_{k=1}^{K-1} q_{lk}$ there are only $K - 1$ free parameters for each individual l .

The empirical Fisher information matrix is calculated from the second-order partial derivatives with respect to the q parameters. The first-order derivative with respect to q_{lk} ($k = 1, \dots, K - 1$) is:

$$\frac{\partial}{\partial q_{lk}} \ln(Q, F|G) = \sum_j \left\{ \frac{g_{lj}(f_{kj} - f_{kj})}{\sum_t q_{lt} f_{tj}} + \frac{(2 - g_{lj})(f_{kj} - f_{kj})}{\sum_t q_{lt} (1 - f_{tj})} \right\}.$$

The derivative of this with respect to $q_{l'k'}$ is 0 when $l' \neq l$. Otherwise, it is equal to

$$\frac{\partial^2}{\partial q_{l'k'} \partial q_{lk}} \ln(Q, F|G) = - \sum_j \left\{ \frac{g_{lj}(f_{k'j} - f_{kj})(f_{kj} - f_{kj})}{(\sum_t q_{lt} f_{tj})^2} + \frac{(2 - g_{lj})(f_{k'j} - f_{kj})(f_{kj} - f_{kj})}{(\sum_t q_{lt} (1 - f_{tj}))^2} \right\}.$$

We calculated the empirical Fisher information matrix for single individuals. Therefore the empirical Fisher information for individual l is given by

$$\{I_l(q_l, F)\}_{k'k} = \sum_j (f_{k'j} - f_{kj})(f_{kj} - f_{kj}) \left\{ \frac{g_{lj}}{(\sum_t q_{lt} f_{tj})^2} + \frac{(2 - g_{lj})}{(\sum_t q_{lt} (1 - f_{tj}))^2} \right\}.$$

From large sample theory the $(K - 1) \times 1$ vectors q_l are asymptotically consistent estimators, are independent of each other, and are normally distributed with the covariance matrix estimated by $\widehat{\Sigma}_l = [I_l(q_l, F)]^{-1}$.

In practice, we included in the bootstrap only ADMIXTURE ancestry components greater than 1%. Also, negative simulated bootstrap values were changed to zero and the sampled vectors were normalized to sum to one.

PHCP

We used a modified version of *ChromoPainter* (Lawson et al., 2012) to “paint” an ancient genome as an imperfect mosaic of modern “donor” haplotypes. ChromoPainter is based on the hidden Markov model (HMM) of Li and Stephens (Li and Stephens, 2003), where each SNP in the target sequence is “copied” from one haplotype (the hidden state) of the donors. Transitions between hidden states are due to ancestral recombination events, and the imperfection of the copying process is due to mutations/genotyping errors.

Our modified version includes changes in the hidden states of the HMM and hence also in the transition and emission probabilities. The modifications were needed because ChromoPainter requires phased data, whereas ancient DNA is typically sequenced/genotyped to very low coverage, and thus cannot be phased using standard tools. Additionally, in many ancient genomes (such as the genomes analyzed in this study), at most one allele is reported at each SNP, based on a randomly selected read. This results in a “pseudo-haploid” sequence that represents both haplotypes. To account for this, the hidden states in our modified HMM represent pairs of haplotypes instead of a single haplotype.

Formally, let h_1, h_2, \dots, h_K denote the K donor haplotypes, and let h^* denote the ancient DNA sequence. We assume that each haplotype has L sites. Let $\bar{Y} = \{Y_1, Y_2, \dots, Y_L\}$ denote the vector of hidden states, where Y_l is an ordered pair of indices of haplotypes (y_1, y_2) from which h^* copies at site l ($y_1, y_2 = 1, \dots, K$). Each pair of haplotypes has an equal *a priori* probability, $1/K^2$, to be chosen as donors. The probability of a recombination event (in each haplotype) between sites l and $l + 1$, is $1 - e^{-\rho_l}$ where $\rho_l = N_e \cdot g_l$, g_l is the genetic distance in Morgan between sites l and $l + 1$, and N_e is a parameter related to the effective population size. Hence, the transition probability between states Y_l and Y_{l+1} is:

$$\Pr(Y_{l+1} = (y_{l+1,1}, y_{l+1,2}) | Y_l = (y_{l,1}, y_{l,2})) = \begin{cases} \frac{(1 - e^{-\rho})^2}{K^2} & \text{if } y_{l,1} \neq y_{l+1,1} ; y_{l,2} \neq y_{l+1,2} \\ \frac{e^{-\rho}(1 - e^{-\rho})}{K} + \frac{(1 - e^{-\rho})^2}{K^2} & \text{if } y_{l,1} = y_{l+1,1} ; y_{l,2} \neq y_{l+1,2} \\ \frac{e^{-\rho}(1 - e^{-\rho})}{K} + \frac{(1 - e^{-\rho})^2}{K^2} & \text{if } y_{l,1} \neq y_{l+1,1} ; y_{l,2} = y_{l+1,2} \\ e^{-2\rho} + \frac{2e^{-\rho}(1 - e^{-\rho})}{K} + \frac{(1 - e^{-\rho})^2}{K^2} & \text{if } y_{l,1} = y_{l+1,1} ; y_{l,2} = y_{l+1,2} \end{cases}$$

The emission probability is the probability to observe a certain allele, A or B, at site l of the ancient genome. It depends on the genotypes of the two haplotypes representing the hidden state and on a mutation rate parameter θ . This probability is calculated conditionally on which of the two copied haplotypes was sequenced (each with probability 0.5). Given the donor haplotype that was sequenced, we observe the same allele as in the donor haplotype with probability θ , and the other allele with probability $1 - \theta$. Thus,

$$\Pr(h_i^* = A | Y_l = y_l) = \begin{cases} 1 - \theta & \text{if } h_{y_l} = (A, A) \\ 0.5\theta + 0.5(1 - \theta) = 0.5 & \text{if } h_{y_l} = (A, B) \\ \theta & \text{if } h_{y_l} = (B, B) \end{cases}$$

Note that θ , the mutation rate, incorporates the effects of both mutations and sequencing errors. If there was no reported allele at site l , we omitted this site from the analyzed sequence.

As in the original ChromoPainter, the forward-backward algorithm was used to compute the posterior probability of each SNP of h^* to copy from each pair of donor haplotypes. The running time can be made linear in the number of states by adapting the “shortcut” offered by Li and Stephens (Li and Stephens, 2003). The expected total length of genetic material (in Morgan) copied from each possible pair of donor haplotypes was calculated based on Equation (S4) of Lawson et al., (2012). The expected length of genetic material copied from haplotype h_i was calculated as the sum of the lengths of genetic material copied from all pairs (h_i, h_j) or (h_j, h_i) , $i, j = 1, \dots, K$. The expected proportion of genetic material copied from a diploid individual was calculated as the expected length of genetic material copied from the two haplotypes of the individual, divided by the genome length. Finally, the average proportion of genetic material copied from a donor population was calculated as the average proportion over all individuals belonging to the population.

Our initial application of the modified ChromoPainter showed that differences between the proportion of genetic material copied from each donor population were small, implying that the “copying profiles” cannot be directly interpreted as ancestry proportions (Methods S1F). To reduce the noise and obtain more interpretable results, we used the regression technique based on Leslie et al., 2015, as above for LINADMIX. In this approach, the copying profile of a target population (i.e., the average copying profile over all individuals in the population) is modeled as a linear combination of the copying profiles of a set of source populations. Formally, let G be the number of distinct donor populations and let $s = 1, \dots, S$ denote the source populations. Let \bar{X}_i denote the copying vector (of length G) of source population i – the average proportion of the genome that individuals from population i copy from each of the G donor populations, as inferred by our modified ChromoPainter. The linear model is

$$\bar{X}_t = \alpha_1 \bar{X}_1 + \alpha_2 \bar{X}_2 + \dots + \alpha_S \bar{X}_S,$$

such that $\alpha_s \geq 0$ for all s , and $\sum_{s=1}^S \alpha_s = 1$ (as for LINADMIX). \bar{X}_t is the copying vector of the target population and $\bar{X}_1, \dots, \bar{X}_S$ are the copying vectors of the source populations. We interpret α_s as the average proportion of the ancestry of the target population coming from source population s . We formulated the equations as a quadratic programming problem, using the quadprog package in R. For studying the ancestry of ancient and modern Middle Eastern populations, we chose the donor populations as a set of $G = 12$ populations from Europe, Middle East, West Asia, and East Africa. We inferred the ChromoPainter parameters N_e and θ with ChromoPainter’s EM algorithm on modern populations.

For practical details on refining the PHCP parameters, and on methods for selection of SNPs and donor populations, see Methods S1F.

Supplemental Figures

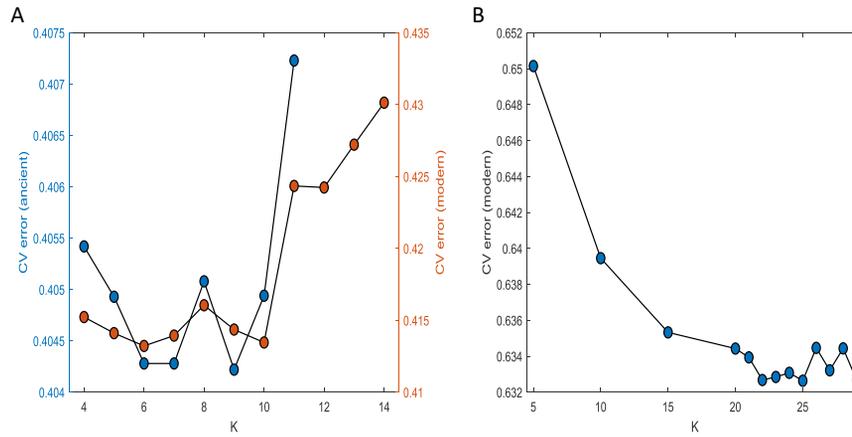


Figure S1. Cross-Validation Errors in ADMIXTURE as a Function of K, Related to Figure 1C

(A) Using 1,663 individuals. (Blue) ADMIXTURE was run on 357,334 SNPs according to the ancient samples protocol. (Orange) ADMIXTURE was run on 50,165 SNPs according to the present-day samples protocol.

(B) Using 3,515 individuals. Only the present-day samples protocol was used.

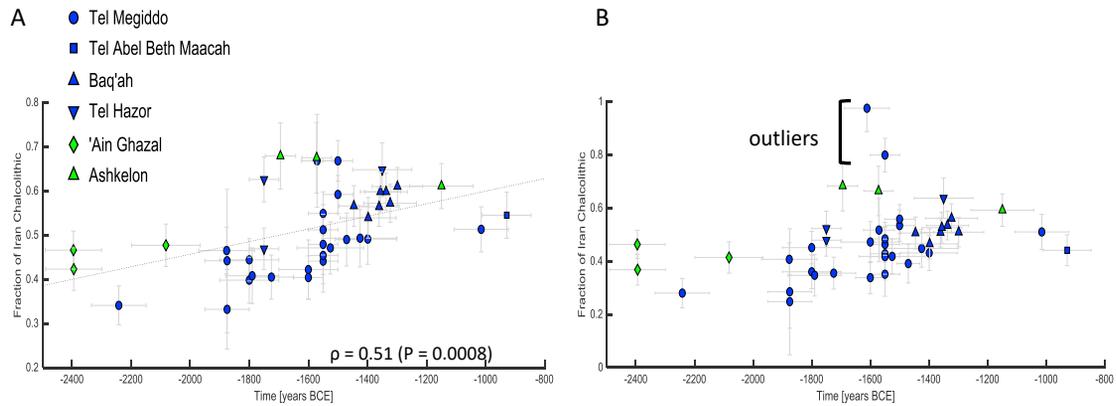


Figure S2. Fraction of Chalcolithic Iran-Related Component in Each Individual as Computed by *qpAdm*, Related to Figure 4

Modeling each individual as a mixture of Neolithic Levant and Chalcolithic Iran, and using either (A) the o9aensw outgroup set (o9 + Anatolia_N + EHG + Natufian + Switzerland_HG + WHG) or (B) the o9a outgroup set (o9 + Anatolia_N). Vertical error bars denote one standard error. Horizontal error bars denote estimated time ranges. Dashed line describes the linear regression. Only individuals whose time range does not exceed 250 years are plotted and used in the regression.

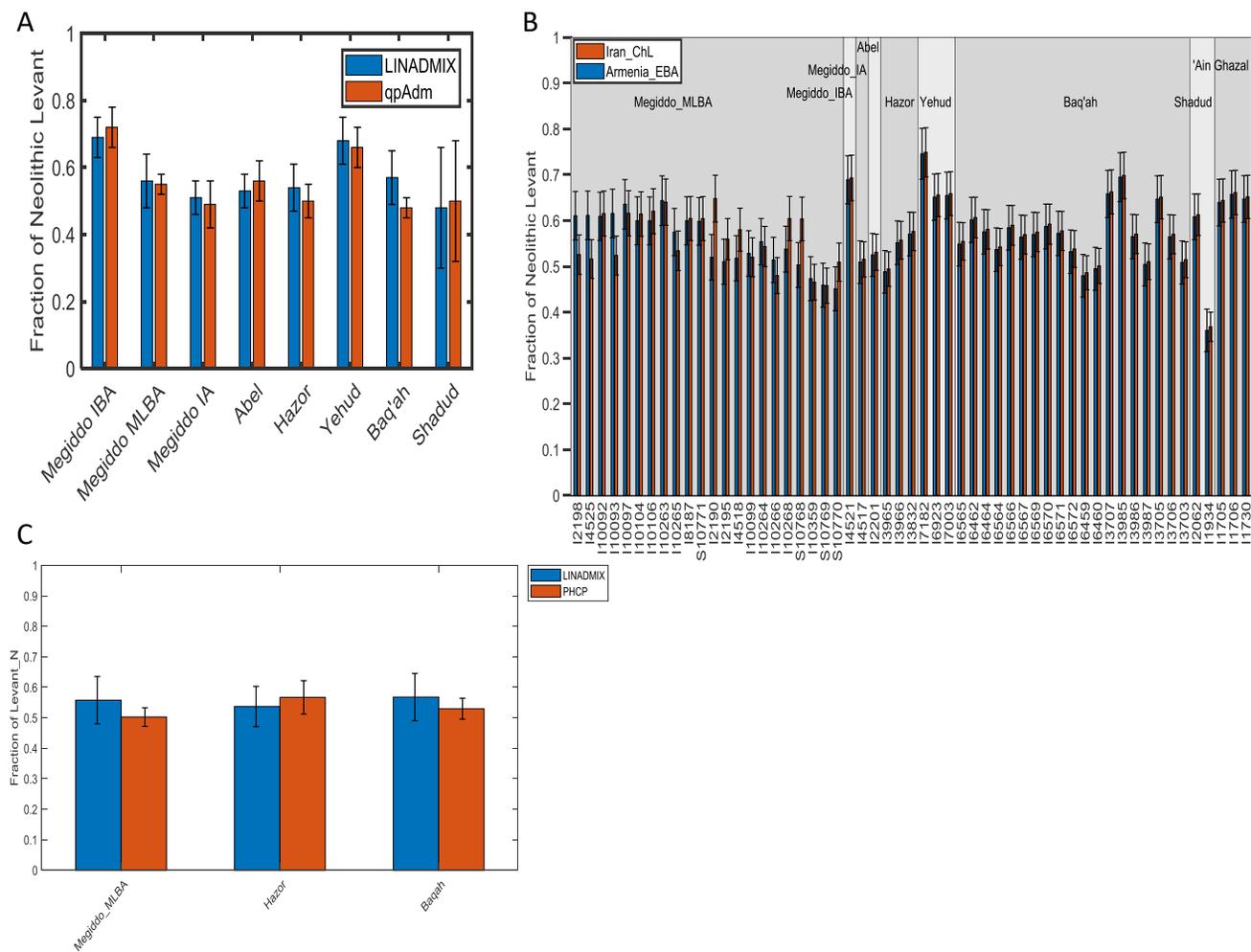


Figure S3. LINADMIX and PHCP Results on Ancient Populations, Related to Figure 4 and to STAR Methods, LINADMIX, PHCP

(A) Comparison between qpAdm and LINADMIX. Bars show the fraction of Neolithic Levant in different populations. Error bars show one standard deviation.

(B) LINADMIX of individual samples. Bars show the fraction of Neolithic Levant in different individuals, when the other source population is either Iran_ChL or Armenia_EBA.

(C) Fraction of Levant_N in different ancient populations (when the other source population is Iran_ChL) as computed by LINADMIX and PHCP.

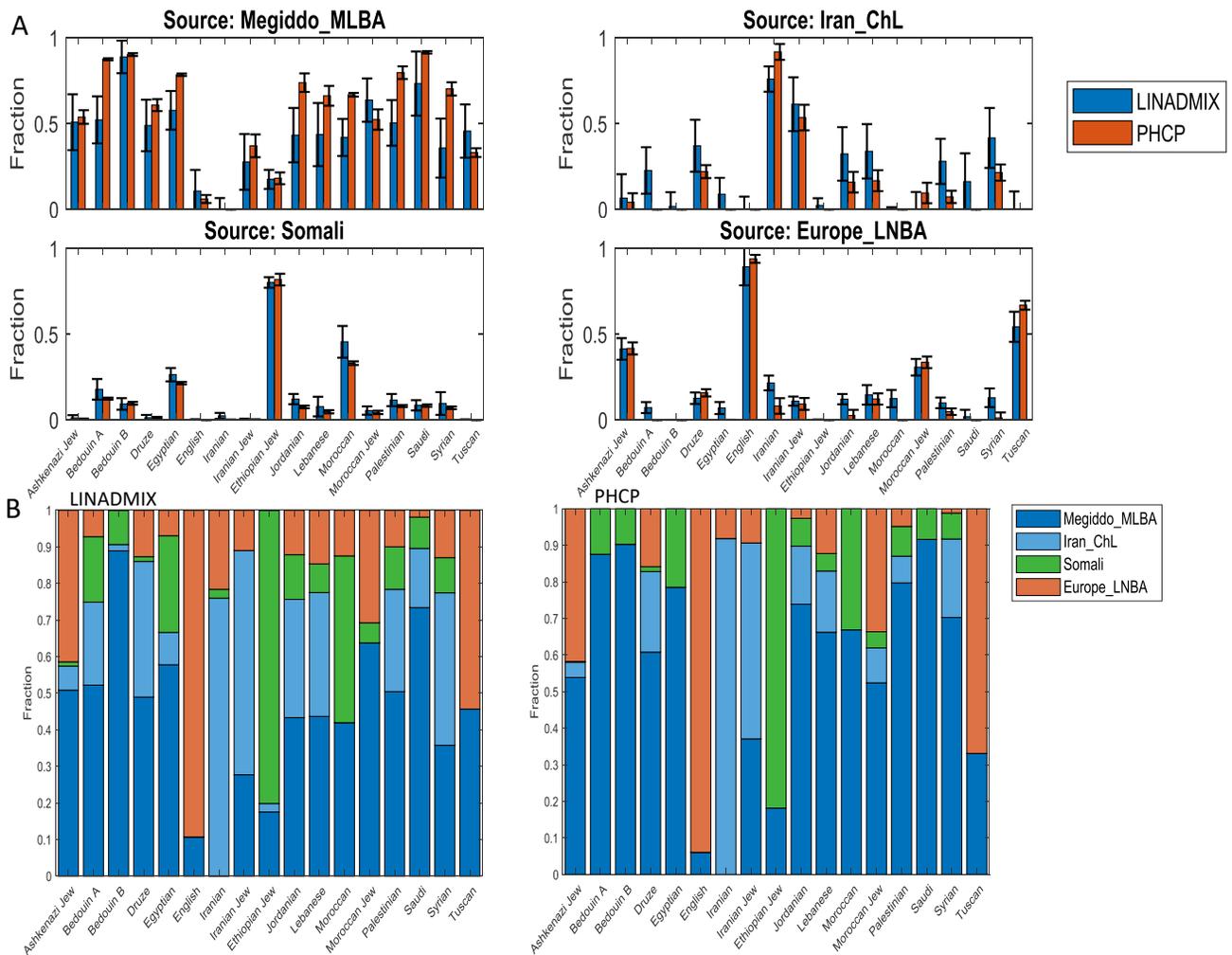


Figure S4. LINADMIX and PHCP models on modern populations, Related to Figure 5

(A) The contribution of each of the source populations to the examined present-day populations, using LINADMIX and PHCP.

(B) LINADMIX and PHCP relative contribution of each of the source populations to the present-day target population listed on the x axis.