

Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences

Marc Haber,^{1,9,*} Claude Doumet-Serhal,^{2,9} Christiana Scheib,^{3,9} Yali Xue,¹ Petr Danecek,¹ Massimo Mezzavilla,¹ Sonia Youhanna,⁴ Rui Martiniano,¹ Javier Prado-Martinez,¹ Michał Szpak,¹ Elizabeth Matisoo-Smith,⁵ Holger Schutkowski,⁶ Richard Mikulski,⁶ Pierre Zalloua,^{7,8} Toomas Kivisild,³ and Chris Tyler-Smith^{1,*}

The Canaanites inhabited the Levant region during the Bronze Age and established a culture that became influential in the Near East and beyond. However, the Canaanites, unlike most other ancient Near Easterners of this period, left few surviving textual records and thus their origin and relationship to ancient and present-day populations remain unclear. In this study, we sequenced five whole genomes from ~3,700-year-old individuals from the city of Sidon, a major Canaanite city-state on the Eastern Mediterranean coast. We also sequenced the genomes of 99 individuals from present-day Lebanon to catalog modern Levantine genetic diversity. We find that a Bronze Age Canaanite-related ancestry was widespread in the region, shared among urban populations inhabiting the coast (Sidon) and inland populations (Jordan) who likely lived in farming societies or were pastoral nomads. This Canaanite-related ancestry derived from mixture between local Neolithic populations and eastern migrants genetically related to Chalcolithic Iranians. We estimate, using linkage-disequilibrium decay patterns, that admixture occurred 6,600–3,550 years ago, coinciding with recorded massive population movements in Mesopotamia during the mid-Holocene. We show that present-day Lebanese derive most of their ancestry from a Canaanite-related population, which therefore implies substantial genetic continuity in the Levant since at least the Bronze Age. In addition, we find Eurasian ancestry in the Lebanese not present in Bronze Age or earlier Levantines. We estimate that this Eurasian ancestry arrived in the Levant around 3,750–2,170 years ago during a period of successive conquests by distant populations.

The Near East, including the Levant, has been central to human prehistory and history from the expansion out of Africa 50–60 thousand years ago (kya),¹ through post-glacial expansions² and the Neolithic transition 10 kya, to the historical period when Ancient Egyptians, Greeks, Phoenicians, Assyrians, Babylonians, Persians, Romans, and many others left their impact on the region.³ Aspects of the genetic history of the Levant have been inferred from present-day DNA,^{4,5} but the more comprehensive analyses performed in Europe^{6–11} have shown the limitations of relying on present-day information alone and highlighted the power of ancient DNA (aDNA) for addressing questions about population histories.¹² Unfortunately, although the few aDNA results from the Levant available so far are sufficient to reveal how much its history differs from that of Europe,¹³ more work is needed to establish a thorough understanding of Levantine genetic history. Such work is hindered by the hot and sometimes wet environment,^{12,13} but improved aDNA technologies including use of the petrous bone as a source of DNA¹⁴ and the rich archaeological remains available encouraged us to further explore the potential of aDNA in this region. Here, we present genome sequences from five Bronze Age Lebanese samples and show how they

improve our understanding of the Levant's history over the last five millennia.

During the Bronze Age in the Levant, around 3–4 kya, a distinctive culture emerged as a Semitic-speaking people known as the Canaanites. The Canaanites inhabited an area bounded by Anatolia to the north, Mesopotamia to the east, and Egypt to the south, with access to Cyprus and the Aegean through the Mediterranean. Thus the Canaanites were at the center of emerging Bronze Age civilizations and became politically and culturally influential.¹⁵ They were later known to the ancient Greeks as the Phoenicians who, 2.3–3.5 kya, colonized territories throughout the Mediterranean reaching as far as the Iberian Peninsula.¹⁶ However, for uncertain reasons but perhaps related to the use of papyrus instead of clay for documentation, few textual records have survived from the Canaanites themselves and most of their history known today has been reconstructed from ancient Egyptian and Greek records, the Hebrew Bible, and archaeological excavations.¹⁵ Many uncertainties still surround the origin of the Canaanites. Ancient Greek historians believed their homeland was located in the region of the Persian Gulf,^{16,17} but modern researchers tend to reject this hypothesis because of archaeological and historical evidence

¹Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK; ²The Sidon excavation, Saida, Lebanon; ³Department of Archaeology and Anthropology, University of Cambridge, Cambridge CB2 1QH, UK; ⁴Institute of Physiology, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; ⁵Department of Anatomy, University of Otago, Dunedin 9054, New Zealand; ⁶Department of Archaeology, Anthropology, and Forensic Science, Bournemouth University, Talbot Campus, Poole BH12 5BB, UK; ⁷The Lebanese American University, Chouran, Beirut 1102 2801, Lebanon; ⁸Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

^{*}These authors contributed equally to this work

*Correspondence: mh25@sanger.ac.uk (M.H.), cts@sanger.ac.uk (C.T.-S.)
<http://dx.doi.org/10.1016/j.ajhg.2017.06.013>.

© 2017 The Author(s). This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 1. Samples Analyzed in This Study

ENA Number	Burial Number	Time Ago	Mapped Reads ^a	Mapped Read %	Coverage Genomic	Coverage MT	Sex ^b	MT Haplogroup	Y Haplogroup
ERS1790733	54	3,700 ^c	69,084,826	6.24	1.19	110	M	N1a3a	J1-P58
ERS1790732	63	3,650 ^d	98,293,308	9.20	1.69	109	M	HV1b1	J2-M12
ERS1790730	65	3,650 ^d	73,701,096	7.57	1.24	124	F	K1a2	–
ERS1790731	75	3,750 ^d	128,355,897	15.48	2.32	164	F	R2	–
ERS1790729	46	3,750 ^d	23,323,399	2.64	0.40	53	F	H1bc	–

^aExcluding PCR duplicates^bGenetically determined^cRadiocarbon date^dArchaeological date

of population continuity through successive millennia in the Levant. The Canaanite culture is alternatively thought to have developed from local Chalcolithic people who were themselves derived from people who settled in farming villages 9–10 kya during the Neolithic period.¹⁵ Uncertainties also surround the fate of the Canaanites: the Bible reports the destruction of the Canaanite cities and the annihilation of its people; if true, the Canaanites could not have directly contributed genetically to present-day populations. However, no archaeological evidence has so far been found to support widespread destruction of Canaanite cities between the Bronze and Iron Ages: cities on the Levant coast such as Sidon and Tyre show continuity of occupation until the present day.

aDNA research has the potential to resolve many questions related to the history of the Canaanites, including their place of origin and fate. Here, we sampled the petrous portion of temporal bones belonging to five ancient individuals dated to between 3,750 and 3,650 years ago (ya) from Sidon, which was a major Canaanite city-state during this period (Figures S1 and S2). We extracted DNA and built double-stranded libraries according to published protocols without uracil-DNA glycosylase treatment.^{18–21} We sequenced the libraries on an Illumina HiSeq 2500 using 2× 75 bp reads and processed the sequences using the PALEOMIX pipeline.²² We retained reads ≥30 bp and collapsed pairs with minimum overlap of 15 bp, allowing a mismatch rate of 0.06 between the pairs. We mapped the merged sequences to the *hs37d5* reference sequence, removed duplicates, removed two bases from the ends of each read, and randomly sampled a single sequence with a minimum quality of ≥20 to represent each SNP. We obtained a genomic coverage of 0.4–2.3× and a mitochondrial DNA (mtDNA) genome coverage of 53–164× (Table 1). Y chromosome genotypes were jointly called across males from the 1000 Genomes Project, present-day Lebanese, and two identified Canaanite males using FreeBayes v.0.9.18.²³ A maximum likelihood phylogeny was inferred using RAxML v.8.2.10²⁴ and visualized using iTOL v.3.5.3.²⁵ In order to assess ancient DNA authenticity, we estimated mtDNA and X chromosome contamination^{26–28} (Table S1) and restricted some analyses to sequences with

aDNA damage patterns^{29,30} (Figures S3 and S4), demonstrating that the sequence data we present are endogenous and minimally contaminated.

Additionally, we sequenced whole genomes of 99 present-day Lebanese individuals with informed consent to ~8× coverage on an Illumina HiSeq 2500 using 2× 100 bp reads in a study approved by The Wellcome Trust Sanger Institute's Human Materials and Data Management Committee (13/010 and 14/072). We merged the low-coverage Lebanese data with four high-coverage (30×) Lebanese samples,³¹ 1000 Genomes Project phase 3 CEU, YRI, and CHB populations,³² and sequence data previously published from regional populations (Egyptians, Ethiopians, and Greeks).^{1,31} Raw calls were generated using bcftools (bcftools mpileup -C50 -pm3 -F0.2 -d10000 | bcftools call -mv, version 1.2-239-g8749475) and filtered to include only SNPs with the minimum of two alternate alleles in at least one population and site quality larger than ten; we excluded sites with a minimum per-population HWE and total HWE less than 0.01³³ and sites within 3 bp of an indel. The filtered calls were then pre-phased using shapeit (v.2.r790)³⁴ and their genotypes refined using beagle (v.4.1).³⁵ We have previously described the genetic structure in the Lebanese population using array data from ~1,300 individuals.⁴ A principal component analysis (PCA) using the 99 sequenced present-day individuals show that they capture the previously described genetic diversity with distinct clusters reflecting the different cultural groups in Lebanon today (Figure S5).

We combined our ancient and modern samples with previously published ancient data^{6–11,13,36–38} (Figure 1A) resulting in a dataset of 389 individuals and 1,046,317 SNPs when ancient and Lebanese samples were analyzed, and 546,891 SNPs when 2,583 modern samples from the Human Origins genotype data were included in the analysis (i.e., the small dataset was used only when a modern population other than the Lebanese was included in the test).^{9,39} A pooled Lebanese sequence dataset (99 low coverage plus 4 high coverage) was used in all analyses except for the PCA and ADMIXTURE where a subset of 15 randomly selected individuals (5 from each group described in Figure S5) was used to avoid sample size

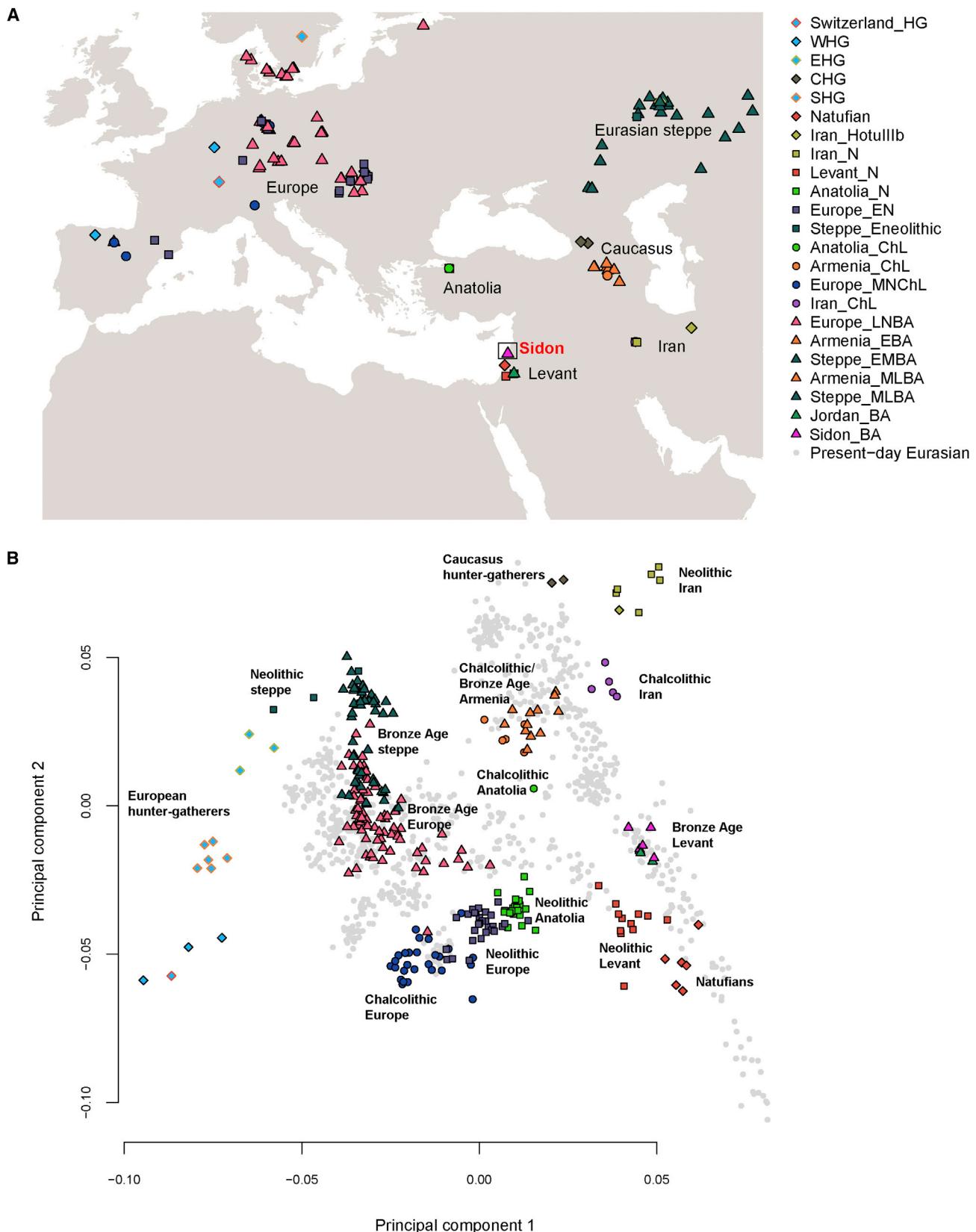


Figure 1. Population Locations and Genetic Structure

(A) The map shows the location of the newly sequenced Bronze Age Sidon samples (pink triangle labeled with red text), as well as the locations of published ancient samples used as comparative data in this study.

(B) PCA of ancient Eurasian samples (colored shapes) projected using eigenvectors from present-day Eurasian populations (gray points).

bias. The ancient samples were grouped following the labels assigned by Lazaridis et al.¹³ on the basis of archaeological culture, chronology, and genetic clustering. We used this dataset to shed light on the genetic history of the Canaanites, resolving their relationship to other ancient populations and assessing their genetic contribution to present-day populations.

We first explored our dataset using PCA⁴⁰ on present-day West Eurasian (including Levantine) populations and projected the ancient samples onto this plot (Figures 1B and S6). The Bronze Age Sidon samples (Sidon_BA) overlap with present-day Levantines and were positioned between the ancient Levantines (Natufians/Neolithic) and ancient Iranians (Neolithic/Chalcolithic). The overlap between the Bronze Age and present-day Levantines suggests a degree of genetic continuity in the region. We explored this further by computing the statistic f_4 (Lebanese, present-day Near Easterner; Sidon_BA, Chimpanzee) using $qpDstat^{39}$ (with parameter f4mode: YES) and found that Sidon_BA shared more alleles with the Lebanese than with most other present-day Levantines (Figure S7), supporting local population continuity as observed in Sidon's archaeological records. When we substituted present-day Near Easterners with a panel of 150 present-day populations available in the Human Origins dataset, we found that only Sardinians and Italian_North shared significantly more alleles with Sidon_BA compared with the Lebanese (Figure S8). Sardinians are known to have retained a large proportion of ancestry from Early European farmers (EEFs) and therefore the increased affinity to Sidon_BA could be related to a shared Neolithic ancestry. We computed f_4 (Lebanese, Sardinian/Italian_North; Sidon_BA, Levant_N) and found no evidence of increased affinity of Sardinians or Italian_North to Sidon_BA after the Neolithic (both Z-scores are positive). We next wanted to explore whether the increased affinity of Sidon_BA to the Lebanese could also be observed when analyzing functionally important regions of the genome that are less susceptible to genetic drift. Our sequence data allowed us to scan loci linked to phenotypic traits and loci previously identified as functional variants in the Lebanese and other Levantines.^{41–43} Using a list of 84 such variants (Table S2), we estimated the allele frequency (AF) in Sidon_BA using ANGSD²⁶ based on a method from Li et al.⁴⁴ and calculated Pearson pairwise correlation coefficients between AF in Sidon_BA and AF in Africans, Europeans, Asians,³² and Lebanese. We found a high significant correlation between Sidon_BA and the Lebanese ($r = 0.74$; 95% CI = 0.63–0.82; p value = 8.168×10^{-16}) and lower correlations between Sidon_BA and Europeans ($r = 0.56$), Africans ($r = 0.55$), and Asians ($r = 0.53$) (Figure S9). These results support population continuity in the region and suggest that several present-day genetic disorders might stem from risk alleles that were already present in the Bronze Age population. In addition, SNPs associated with phenotypic traits show that Sidon_BA and the Lebanese had comparable skin, hair, and eye colors (in general:

light intermediate skin pigmentation, brown eyes, and dark hair) with similar frequencies of the underlying causal variants in *SLC24A5* and *HERC2*, but with Sidon_BA probably having darker skin than Lebanese today from variants in *SLC45A2* resulting in darker pigmentation (Table S2).

The PCA shows that Sidon_BA clusters with three individuals from Early Bronze Age Jordan (Jordan_BA) found in a cave above the Neolithic site of 'Ain Ghazal and probably associated with an Early Bronze Age village close to the site.¹³ This suggests that people from the highly differentiated urban culture on the Levant coast and inland people with different modes of subsistence were nevertheless genetically similar, supporting previous reports that the different cultural groups who inhabited the Levant during the Bronze Age, such as the Ammonites, Moabites, Israelites, and Phoenicians, each achieved their own cultural identities but all shared a common genetic and ethnic root with Canaanites.¹⁵ Lazaridis et al.¹³ reported that Jordan_BA can be modeled as mixture of Neolithic Levant (Levant_N) and Chalcolithic Iran (Iran_ChL). We computed the statistic f_4 (Levant_N, Sidon_BA; Ancient Eurasian, Chimpanzee) and found that populations from the Caucasus and ancient Iran shared more alleles with Sidon_BA than with Neolithic Levant (Figure 2A and S10). We then used $qpAdm^8$ (with parameter allsnps: YES) to test whether Sidon_BA can be modeled as mixture of Levant_N and any other ancient population in the dataset and found good support for the model of Sidon_BA being a mixture of Levant_N (48.4% ± 4.2%) and Iran_ChL (51.6% ± 4.2%) (Figure 2B; Table S3).

In addition, the two Sidon_BA males carried the Y-chromosome haplogroups⁴⁵ J-P58 (J1a2b) and J-M12 (J2b) (Tables 1 and S4; Figure S11), both common male lineages in the Near East today. Haplogroup J-P58 is frequent in the Arabian peninsula with proposed origins in the Zagros/Taurus mountain region.⁴⁶ It forms the vast majority of the Y chromosomes in southwestern Mesopotamia and reaches particularly high frequencies (74.1%) in Marsh Arabs in Iraq.⁴⁷ On the other hand, haplogroup J-M12 is widespread at low frequency from the Balkans to India and the Himalayas, with Albanians having the highest proportions (14.3%).⁴⁸ We compiled frequencies of Y-chromosome haplogroups in this geographical area and their changes over time in a dataset of ancient and modern Levantine populations (Figure S12), and note, similarly to Lazaridis et al.,¹³ that haplogroup J was absent in all Natufian and Neolithic Levant male individuals examined thus far, but emerged during the Bronze Age in Lebanon and Jordan along with ancestry related to Iran_ChL. All five Sidon_BA individuals had different mitochondrial DNA haplotypes⁴⁹ (Table 1), belonging to paragroups common in present-day Lebanon and nearby regions (Table S5) but with additional derived variants not observed in our present-day Lebanese dataset.

We next sought to estimate the time when the Iran_ChL-related ancestry penetrated the Levant. Our

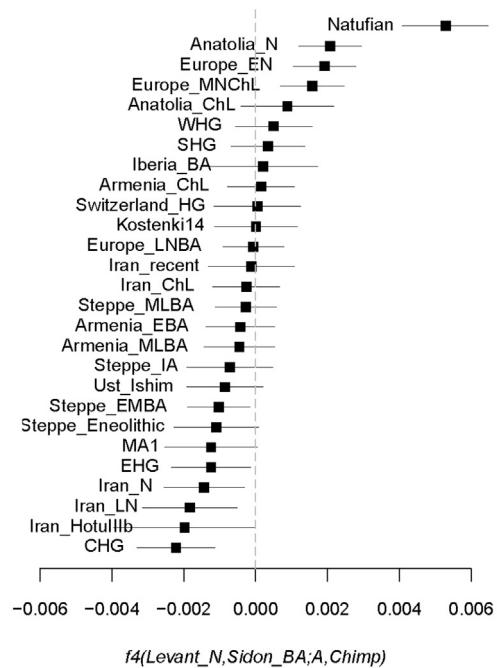
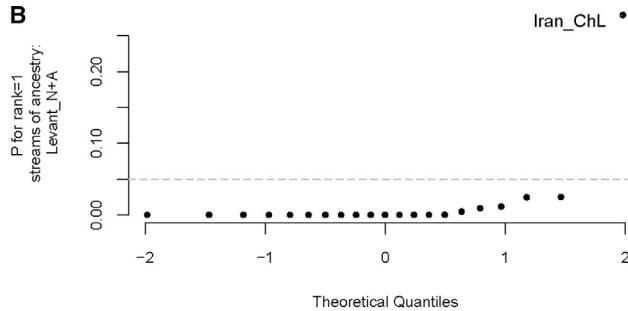
A**B**

Figure 2. Admixture in Bronze Age Levantine Populations

(A) The statistic $f_4(\text{Levant}_N, \text{Sidon}_\text{BA}; \text{Ancient Eurasian}, \text{Chimpanzee})$ is most negative for ancient populations from the Caucasus and Iran, suggesting an increase in ancestry related to these populations in Sidon after the Neolithic period. The plot shows the estimated statistic value and ± 3 standard errors.

(B) Modeling Sidon as mixture between Neolithic Levant and an ancient Eurasian population shows that Chalcolithic Iran fits the model best when using a large number of outgroups: Ust_Ishim, Kostenki14, MA1, Han, Papuan, Ami, Chukchi, Karitiana, Mbuti, Switzerland_HG, EHG, WHG, and CHG. Sidon_BA can then be modeled using $qpAdm$ as 0.484 ± 0.042 Levant_N and 0.516 ± 0.042 Iran_ChL.

results support genetic continuity since the Bronze Age and thus our large dataset of present-day Lebanese provided an opportunity to explore the admixture time using admixture-induced linkage disequilibrium (LD) decay. Using ALDER⁵⁰ (with mindis: 0.005), we set the Lebanese as the admixed test population and Natufians, Levant_N, Sidon_BA, Iran_N, and Iran_ChL as reference populations. To account for the small number of individuals in the reference populations and the limited number of SNPs in the dataset, we took a lenient minimum Z-score = 2 to be suggestive of admixture. The most significant result was for mixture of Levant_N and Iran_ChL ($p = 0.013$) around 181 ± 54 generations ago, or $\sim 5,000 \pm 1,500$ ya assuming

a generation time of 28 years (Figure S13A). This admixture time, based entirely on genetic data, fits the known ages of the samples based on archaeological data since it falls between the dates of Sidon_BA (3,650–3,750 ya) and Iran_ChL (6,500–5,500 ya). The admixture time also overlaps with the rise and fall of the Akkadian Empire which controlled the region from Iran to the Levant between ~ 4.4 and 4.2 kya. The Akkadian collapse is argued to have been the result of a widespread aridification event around 4,200 ya.^{51,52} Archaeological evidence in this period documents large-scale influxes of refugees from Northern Mesopotamia toward the south, where cities and villages became overpopulated.⁵³ Our confidence intervals for the admixture dates are wide and therefore the historical links suggested here should be considered with caution. Future sampling of ancient DNA from northern Syria and Iraq will reveal whether these populations carried the Iran_ChL-related ancestry and also provide a better understanding of the origin of the eastern migrants and the time when they arrived in the Levant.

Although f_4 tests showed that present-day Lebanese share significantly more alleles with Sidon_BA than other Near Eastern populations do, indicating genetic continuity, we failed to model the present-day Lebanese using streams of ancestry coming only from Levant_N and Iran_ChL ($qpAdm$ rank1 $p = 8.36 \times 10^{-7}$), in contrast to our success with Sidon_BA. We therefore further explored our dataset by running ADMIXTURE⁵⁴ in a supervised mode using Western hunter-gatherers (WHG), Eastern hunter-gatherers (EHG), Levant_N, and Iran_N as reference populations. These four populations have been previously¹³ found to contribute genetically to most West Eurasians. The ADMIXTURE results replicate the findings from $qpAdm$ for Sidon_BA and show mixture of Levant_N and ancient Iranian populations (Figure 3A). However, the present-day Lebanese, in addition to their Levant_N and ancient Iranian ancestry, have a component (11%–22%) related to EHG and Steppe populations not found in Bronze Age populations (Figure 3A). We confirm the presence of this ancestry in the Lebanese by testing f_4 (Sidon_BA, Lebanese; Ancient Eurasian, Chimpanzee) and find that Eurasian hunter-gatherers and Steppe populations share more alleles with the Lebanese than with Sidon_BA (Figures 3B and S14). We next tested a model of the present-day Lebanese as a mixture of Sidon_BA and any other ancient Eurasian population using $qpAdm$. We found that the Lebanese can be best modeled as Sidon_BA $93\% \pm 1.6\%$ and a Steppe Bronze Age population $7\% \pm 1.6\%$ (Figure 3C; Table S6). To estimate the time when the Steppe ancestry penetrated the Levant, we used, as above, LD-based inference and set the Lebanese as admixed test population with Natufians, Levant_N, Sidon_BA, Steppe_EMBA, and Steppe_MLBA as reference populations. We found support ($p = 0.00017$) for a mixture between Sidon_BA and Steppe_EMBA which has occurred around $2,950 \pm 790$ ya (Figure S13B). It is important to note here that Bronze Age Steppe populations used in

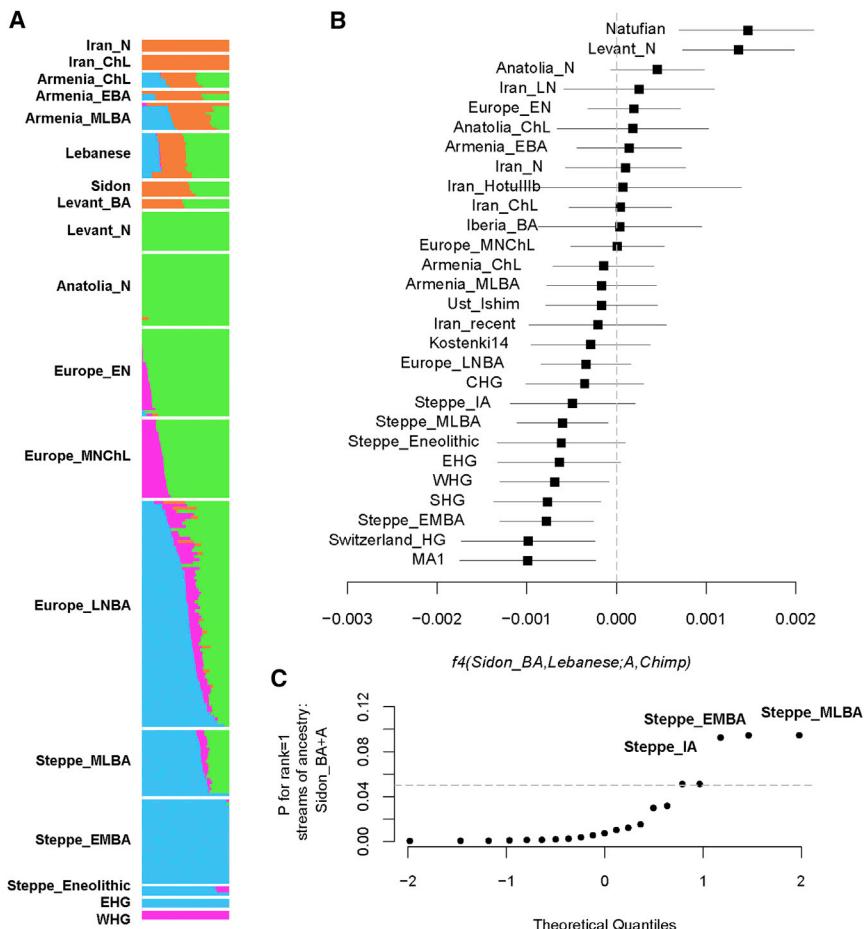


Figure 3. Admixture in Present-Day Levantine Populations

(A) Supervised ADMIXTURE using Levant_N, Iran_N, EHG, and WHG as reference populations. A Eurasian ancestry found in Eastern hunter-gatherers and the steppe Bronze Age appears in present-day Levantines after the Bronze Age.

(B) The statistic $f_4(\text{Sidon_BA}, \text{Lebanese}; \text{Ancient Eurasian, Chimpanzee})$ confirms the ADMIXTURE results and is most negative for populations from the steppe and Eurasian hunter-gatherers. We show the estimated statistic value and ± 3 standard errors.

(C) Present-day Lebanese can be modeled as mixture between Bronze Age Sidon and a steppe population. The model with mix proportions 0.932 ± 0.016 Sidon_BA and 0.068 ± 0.016 Steppe_EMBA for Lebanese is supported with the lowest SE.

the model need not be the actual ancestral mixing populations, and the admixture could have involved a population which was itself admixed with a Steppe-like ancestry population. The time period of this mixture overlaps with the decline of the Egyptian empire and its domination over the Levant, leading some of the coastal cities to thrive, including Sidon and Tyre, which established at this time a successful maritime trade network throughout the Mediterranean. The decline in Egypt's power was also followed by a succession of conquests of the region by distant populations such as the Assyrians, Persians, and Macedonians, any or all of whom could have carried the Steppe-like ancestry observed here in the Levant after the Bronze Age.

In this report we have analyzed ancient whole-genome sequence data from a Levantine civilization and provided insights into how the Bronze Age Canaanites were related to other ancient populations and how they have contributed genetically to present-day ones (Figure 4). Many of our inferences rely on the limited number of ancient samples available, and we are only just beginning to reconstruct a genetic history of the Levant or the Near East as thoroughly as that of Europeans who, in comparison, have been extensively sampled. In the future, it will be important to examine samples from the Chalcolithic/Early Bronze Age Near East to understand the events leading to admixture between local populations and the eastern mi-

grants. It will also be important to analyze samples from the Iron Age to trace back the Steppe-like ancestry we find today in present-day Levantines. Our current results show that such studies are feasible.

Accession Numbers

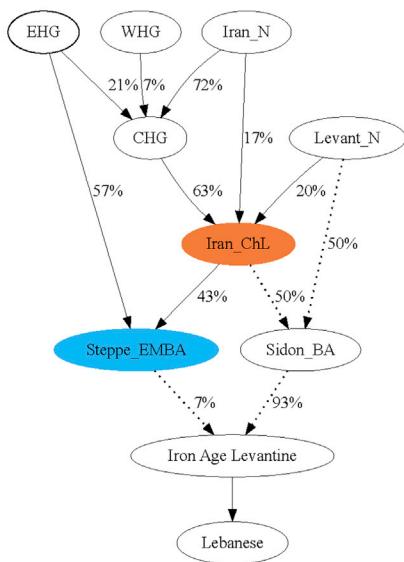
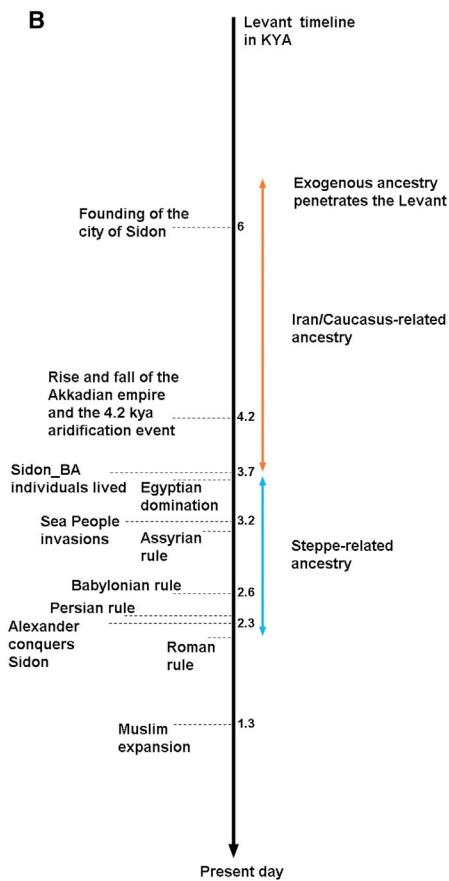
Sequencing data for 99 present-day Lebanese individuals reported in this paper are available through the European Genome-phenome Archive (EGA) under accession number EGA: EGAS00001002084. Raw sequencing reads for 5 ancient Canaanite individuals are available through the European Nucleotide Archive (ENA) under accession number ENA: PRJEB21330. Aligned sequences and genotypes can be obtained from the corresponding authors.

Supplemental Data

Supplemental Data include 14 figures and 6 tables and can be found with this article online at <http://dx.doi.org/10.116/j.ajhg.2017.06.013>.

Acknowledgments

We thank the present-day donors who contributed their samples to this study and bioRxiv readers for helpful comments. M.H., Y.X., P.D., R.M., J.P.-M., M.S., and C.T.S. were supported by The Wellcome Trust (098051).

A**B****Figure 4. Genetic History of the Levant**

(A) A model of population relationships which fits the *qpAdm* results from Lazaridis et al.¹³ (solid arrows) and this study (dotted arrows). Percentages on arrows are the inferred admixture proportions.

(B) Levant timeline of historical events with genetically inferred admixture dates shown as colored double-ended arrows with length representing the SE.

Received: May 25, 2017

Accepted: June 27, 2017

Published: July 27, 2017

Web Resources

European Genome-phenome Archive (EGA), <https://www.ebi.ac.uk/ega>

European Nucleotide Archive (ENA), <http://www.ebi.ac.uk/ena>

References

- Pagani, L., Schiffels, S., Gurdasani, D., Danecek, P., Scally, A., Chen, Y., Xue, Y., Haber, M., Ekong, R., Oljira, T., et al. (2015). Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am. J. Hum. Genet.* **96**, 986–991.
- Platt, D.E., Haber, M., Dagher-Kharrat, M.B., Douaihy, B., Khazen, G., Ashrafian Bonab, M., Salloum, A., Mouzaya, F., Luiselli, D., Tyler-Smith, C., et al. (2017). Mapping post-glacial expansions: the peopling of Southwest Asia. *Sci. Rep.* **7**, 40338.
- Hitti, P.K. (1967). Lebanon in History: From the Earliest Times to the Present (London: Macmillan).
- Haber, M., Gauguier, D., Youhanna, S., Patterson, N., Moorjani, P., Botigué, L.R., Platt, D.E., Matisoo-Smith, E., Soria-Hernanz, D.F., Wells, R.S., et al. (2013). Genome-wide diversity in the levant reveals recent structuring by culture. *PLoS Genet.* **9**, e1003316.
- Zalloua, P.A., Platt, D.E., El Sibai, M., Khalife, J., Makhoul, N., Haber, M., Xue, Y., Izaabel, H., Bosch, E., Adams, S.M., et al.; Genographic Consortium (2008). Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *Am. J. Hum. Genet.* **83**, 633–642.
- 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.
- Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodriguez-Varela, R., Svartisdóttir, O.O., Daskalaki, E.A., Skoglund, P., Naidoo, T., Svensson, E.M., et al. (2015). Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc. Natl. Acad. Sci. USA* **112**, 11917–11922.
- 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.
- 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.
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503.
- Otalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobón, I., Ramirez, O., Civit, S., García Borja, P., Salazar-García, D.C., Talamo, S., et al. (2015). A common genetic origin

- for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol. Biol. Evol.* **32**, 3132–3142.
12. Haber, M., Mezzavilla, M., Xue, Y., and Tyler-Smith, C. (2016). Ancient DNA and the rewriting of human history: be sparing with Occam's razor. *Genome Biol.* **17**, 1.
 13. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424.
 14. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257.
 15. Tubb, J.N. (1998). Canaanites (London: Published for the Trustees of the British Museum by British Museum Press).
 16. Markoe, G. (2000). Phoenicians (London: British Museum Press).
 17. Al Khalifa, H.A.S., and Rice, M. (1986). Bahrain through the Ages: The Archaeology (London: KPI).
 18. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, t5448.
 19. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763.
 20. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229.
 21. Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczkay, P., et al. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102.
 22. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., et al. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082.
 23. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv. arXiv:1207.3907 [q-bio.GN]*. <https://arxiv.org/abs/1207.3907>.
 24. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
 25. Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44** (W1), W242–W245.
 26. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356.
 27. Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98.
 28. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224.
 29. 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.
 30. Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentini, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petzsch, F., Fernandes, D., et al. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513.
 31. Haber, M., Mezzavilla, M., Bergström, A., Prado-Martinez, J., Hallast, P., Saif-Ali, R., Al-Habori, M., Dedoussis, G., Zeggini, E., Blue-Smith, J., et al. (2016). Chad genetic diversity reveals an African history marked by multiple Holocene eurasian migrations. *Am. J. Hum. Genet.* **99**, 1316–1324.
 32. The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* **526**, 68–74.
 33. Wigginton, J.E., Cutler, D.J., and Abecasis, G.R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* **76**, 887–893.
 34. Delaneau, O., Marchini, J., and Zagury, J.F. (2011). A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181.
 35. Browning, B.L., and Browning, S.R. (2016). Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126.
 36. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912.
 37. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prüfer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449.
 38. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91.
 39. 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.
 40. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* **2**, e190.
 41. Nakouzi, G., Kreidieh, K., and Yazbek, S. (2015). A review of the diverse genetic disorders in the Lebanese population: highlighting the urgency for community genetic services. *J. Community Genet.* **6**, 83–105.
 42. Hager, J., Kamatani, Y., Cazier, J.B., Youhanna, S., Ghassibe-Sabbagh, M., Platt, D.E., Abchee, A.B., Romanos, J., Khazen, G., Othman, R., et al.; FGENTCARD Consortium (2012). Genome-wide association study in a Lebanese cohort confirms PHACTR1 as a major determinant of coronary artery stenosis. *PLoS ONE* **7**, e38663.
 43. Ghassibe-Sabbagh, M., Haber, M., Salloum, A.K., Al-Sarraj, Y., Akle, Y., Hirbli, K., Romanos, J., Mouzaya, F., Gauguier, D., Platt, D.E., et al. (2014). T2DM GWAS in the Lebanese

- population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility. *Sci. Rep.* **4**, 7351.
- 44. Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sánchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliuszen, T., et al. (2010). Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat. Genet.* **42**, 969–972.
 - 45. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv*. <http://dx.doi.org/10.1101/088716>.
 - 46. Chiaroni, J., King, R.J., Myres, N.M., Henn, B.M., Ducourneau, A., Mitchell, M.J., Boetsch, G., Sheikha, I., Lin, A.A., Nik-Ahd, M., et al. (2010). The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations. *Eur. J. Hum. Genet.* **18**, 348–353.
 - 47. Al-Zahery, N., Pala, M., Battaglia, V., Grugni, V., Hamod, M.A., Hooshiar Kashani, B., Olivieri, A., Torroni, A., Santachiara-Benerecetti, A.S., and Semino, O. (2011). In search of the genetic footprints of Sumerians: a survey of Y-chromosome and mtDNA variation in the Marsh Arabs of Iraq. *BMC Evol. Biol.* **11**, 288.
 - 48. Semino, O., Magri, C., Benuzzi, G., Lin, A.A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P.J., et al. (2004). Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* **74**, 1023–1034.
 - 49. Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F., and Schönherr, S. (2016). mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res.* **44** (W1), W64–W69.
 - 50. Loh, P.R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D., and Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254.
 - 51. Cullen, H.M., deMenocal, P.B., Hemming, S., Hemming, G., Brown, F.H., Guilderson, T., and Sirocko, F. (2000). Climate change and the collapse of the Akkadian empire: Evidence from the deep sea. *Geology* **28**, 4.
 - 52. deMenocal, P.B. (2001). Cultural responses to climate change during the late Holocene. *Science* **292**, 667–673.
 - 53. Weiss, H., Courty, M.A., Wetterstrom, W., Guichard, F., Senior, L., Meadow, R., and Curnow, A. (1993). The genesis and collapse of third millennium north mesopotamian civilization. *Science* **261**, 995–1004.
 - 54. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664.