

## **Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences**

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

<sup>1</sup>The Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambs. CB10 1SA, United Kingdom.

<sup>2</sup>The Sidon excavation, Saida, Lebanon.

<sup>3</sup>Department of Archaeology and Anthropology, University of Cambridge, Cambridge, CB2 1QH, UK.

<sup>4</sup>Institute of Physiology, University of Zurich, Winterthurerstrasse 190, CH-8057, Zürich, Switzerland.

<sup>5</sup>Department of Anatomy, University of Otago, Dunedin, New Zealand, 9054.

<sup>6</sup>Department of Archaeology, Anthropology, and Forensic Science, Bournemouth University, Talbot Campus, Poole BH12 5BB, UK.

<sup>7</sup>The Lebanese American University, Chouran, Beirut 1102 2801, Lebanon; Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA.

<sup>8</sup>These authors contributed equally to this work

\*Correspondence: mh25@sanger.ac.uk (M.H.), cts@sanger.ac.uk (C.T.-S.)

Keywords: aDNA; Bronze Age; whole-genome sequences; Lebanon; Sidon; population history

The Canaanites inhabited the Levant region during the Bronze Age and established a culture which 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 catalogue 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 massive population movements in the mid-Holocene triggered by aridification ~4,200 years ago. 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 this Eurasian ancestry arrived in the Levant around 3,750-2,170 years ago during a period of successive conquests by distant populations such as the Persians and Macedonians.

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),<sup>1</sup> through post-glacial expansions<sup>2</sup> 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.<sup>3</sup> Aspects of the genetic history of the Levant have been inferred from present-day DNA,<sup>4; 5</sup> but the more comprehensive analyses performed in Europe<sup>6-11</sup> 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.<sup>12</sup> 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,<sup>13</sup> more work is needed to establish a thorough understanding of Levantine genetic history. Such work is hindered by the hot and sometimes wet environment,<sup>12; 13</sup> but improved aDNA technologies including use of the petrous bone as a source of DNA<sup>14</sup> 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 centre of emerging Bronze Age civilizations and became politically and culturally influential.<sup>15</sup> 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.<sup>16</sup> 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.<sup>15</sup> Many uncertainties still surround the origin of the Canaanites: Ancient Greek historians believed their homeland was located in the region of the Persian Gulf.<sup>16; 17</sup> However, modern researchers tend to reject this hypothesis because of archaeological and historical evidence 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 in the 9-10 kya during the Neolithic period.<sup>15</sup> 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 (Figure S1 and S2). We extracted DNA and built double-stranded libraries according to published protocols.<sup>14; 18-20</sup> We sequenced the libraries on an Illumina HiSeq 2500 using 2×75 bp reads and processed the sequences using the PALEOMIX pipeline.<sup>21</sup> We retained reads  $\geq 30$ bp and collapsed pairs with minimum overlap of 15bp, 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  $\geq 20$  to represent each SNP. We obtained a genomic coverage of 0.4-2.3x and a mitochondrial DNA (mtDNA) genome coverage of 53-164x (Table 1). In order to assess ancient DNA authenticity, we estimated X-chromosome contamination<sup>22; 23</sup> (Table S1) and restricted some analyses to sequences with aDNA damage patterns<sup>24; 25</sup> (Figure S3 and S4), and as a result demonstrate that the sequence data we present are endogenous and minimally contaminated.

Additionally, we sequenced whole-genomes of 99 present-day Lebanese individuals to ~8x coverage on an Illumina HiSeq 2500 using 2× 100 bp reads. We merged the low-coverage Lebanese data with four high-coverage (30x) Lebanese samples,<sup>26</sup> 1000 Genomes Project phase 3 CEU, YRI, and CHB populations,<sup>27</sup> and sequence data previously published from regional populations (Egyptians, Ethiopians and Greeks).<sup>1; 26</sup> 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 2 alternate alleles in at least one population and site quality larger than 10; we excluded sites with a minimum per-population HWE and total HWE less than 0.01<sup>28</sup> and sites within 3bp of an indel. The filtered calls were then pre-phased using shapeit (v2.r790)<sup>29</sup> and their genotypes refined using beagle (v4.1).<sup>30</sup>

We combined our ancient and modern samples with previously published ancient data<sup>6-11; 13; 24; 31; 32</sup> (Figure 1A) resulting in a dataset of 389 individuals and 1,046,317 SNPs when ancient and Lebanese samples were analysed, and 546,891 SNPs when 2,583 modern samples from the Human Origins

genotype data were included in the analysis.<sup>9; 33</sup> The ancient samples were grouped following the labels assigned by Lazaridis et al.<sup>13</sup> 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 principal component analysis (PCA)<sup>34</sup> on present-day West Eurasian (including Levantine) populations and projected the ancient samples onto this plot (Figure 1B and S5). 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  $f_4(\text{Lebanese, present-day Near Easterner; Sidon\_BA, Chimpanzee})$  using  $qpDstat$ <sup>33</sup> (with parameter  $f_4mode: YES$ ) and found Sidon\_BA shared more alleles with the Lebanese than with most other present-day Levantines (Figure S6), 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 only Sardinians and Italian\_North shared significantly more alleles with Sidon\_BA compared with the Lebanese (Figure S7). Sardinians are known to have retained a large proportion of ancestry from Early European farmers (EEF) and therefore the increased affinity to Sidon\_BA could be related to a shared Neolithic ancestry. We computed  $f_4(\text{Lebanese, Sardinian/Italian\_North; Sidon\_BA, Levant\_N})$  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 if the increased affinity of Sidon\_BA to the Lebanese could also be observed when analysing functionally important regions of the genome which 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.<sup>35-37</sup> Using a list of 84 such variants (Table S2), we estimated the allele frequency (AF) in Sidon\_BA using ANGSD<sup>22</sup> based on a method from Li et al.<sup>38</sup> and calculated Pearson pair-wise correlation coefficients between AF in Sidon\_BA and AF in Africans, Europeans, Asians<sup>27</sup> 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 =  $8.168e-16$ ) and lower correlations between Sidon\_BA and Europeans ( $r = 0.56$ ), Africans, ( $r = 0.55$ ) and Asians ( $r = 0.53$ ) (Figure S8). These results support substantial local population continuity and suggest that several present-day genetic disorders might stem from risk alleles which were already present in the Bronze Age population. In addition, SNPs associated with phenotypic traits show Sidon\_BA and the

Lebanese had comparable skin, hair, and eye colours (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.<sup>13</sup> 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.<sup>15</sup> Lazaridis et al.<sup>13</sup> reported that Jordan\_BA can be modelled as mixture of Neolithic Levant (Levant\_N) and Chalcolithic Iran (Iran\_ChL). We computed the statistic  $f_4(\text{Levant\_N}, \text{Sidon\_BA}; \text{Ancient Eurasian}, \text{Chimpanzee})$  and found populations from the Caucasus and Iran shared more alleles with Sidon\_BA than with Neolithic Levant (Figure 2A). We then used  $qpAdm$ <sup>8</sup> (with parameter `allsnps: YES`) to test if Sidon\_BA can be modelled 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 haplogroups<sup>39</sup> J-P58 (J1a2b) and J-M12 (J2b) (Table 1 and S4; Figure S9), both common male lineages in the Near East today. We compiled frequencies of Y-chromosomal haplogroups in this geographical area and their changes over time in a dataset of ancient and modern Levantine populations (Figure S10), and note, similarly to Lazaridis et al.,<sup>13</sup> 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. All five Sidon\_BA individuals had different mitochondrial DNA haplotypes<sup>40</sup> (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 Iranian ancestry penetrated the Levant. Our 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 ALDER<sup>41</sup> (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  $181 \pm 54$  generations ago, or  $\sim 5,000 \pm 1,500$  ya assuming a generation time of 28 years (Figure S11A). 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  $\sim 4.4$  and 4.2 kya. The Akkadian collapse is argued to have been the result of a widespread aridification event around 4,200 ya, possibly caused by a volcanic eruption.<sup>42; 43</sup> Archaeological evidence in this period documents large-scale influxes of refugees from Northern Mesopotamia towards the south, where cities and villages became overpopulated.<sup>44</sup> Future sampling of ancient DNA from Northern Syria and Iraq may reveal if these migrants carried the Iran\_ChL-related ancestry we observe in Bronze Age Sidon and Jordan.

Although *f4* 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  $p=8.36E-07$ ), in contrast to our success with Sidon\_BA. We therefore further explored our dataset by running ADMIXTURE<sup>45</sup> 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 previously<sup>13</sup> 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 Iranian populations (Figure 3A). However, the present-day Lebanese, in addition to their Levant\_N and 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 *f4*(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 (Figure 3B). 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 modelled 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±790 ya (Figure S11B). It is important to note here that Bronze Age Steppe populations used in 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 analysed the first 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 thorough 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 migrants. 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.



## **Acknowledgments**

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

**Table 1. Samples analysed in this study**

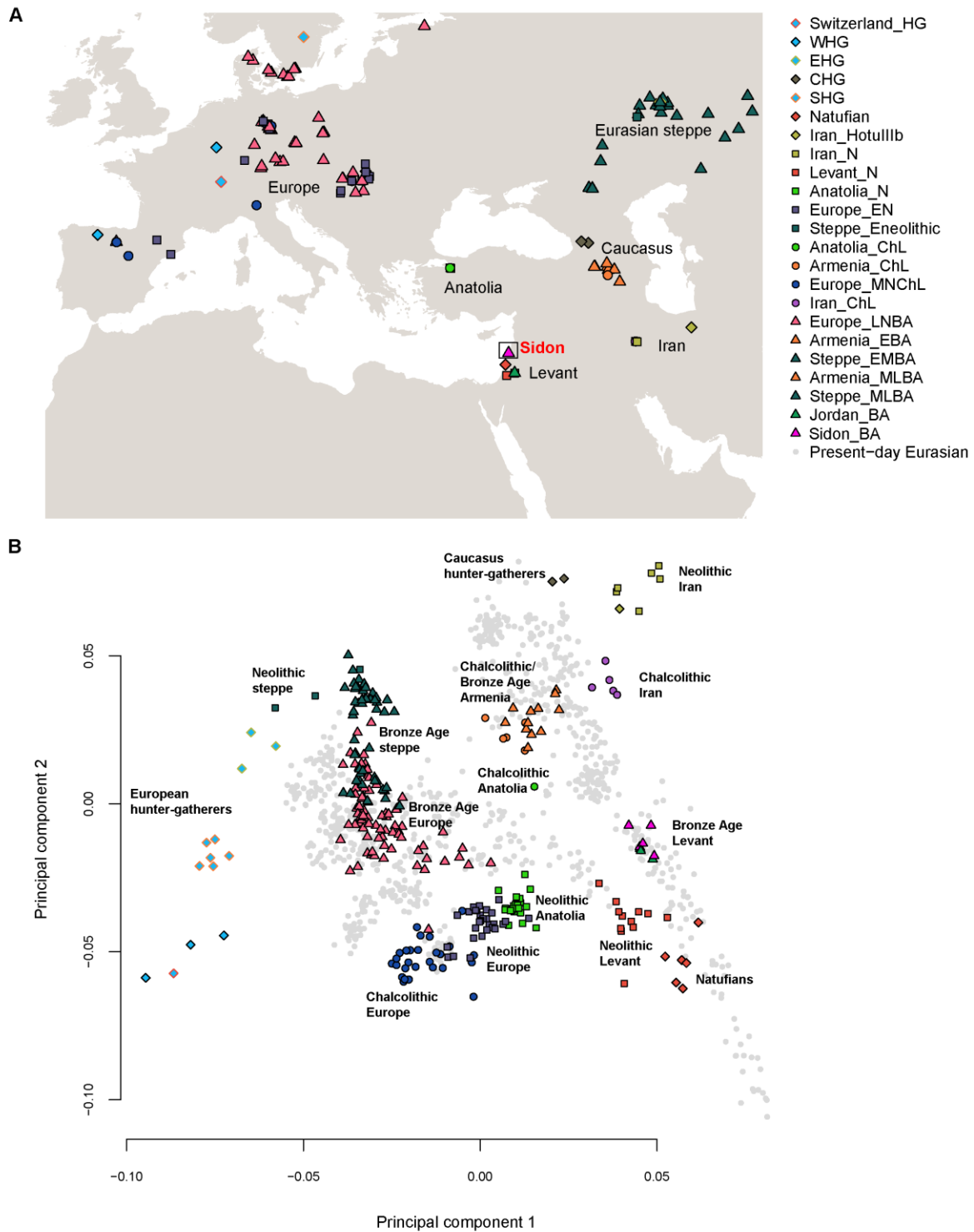
ENA number	Burial number	Time years ago	Mapped reads <sup>c</sup>	Coverage genomic	Coverage MT	Sex <sup>d</sup>	MT haplogroup	Y haplogroup
EGAN00001390967	54	3,700 <sup>a</sup>	69,084,826	1.19	110	M	N1a3a	J-P58 (J1a2b)
EGAN00001390965	63	3,650 <sup>b</sup>	98,293,308	1.69	109	M	HV1b1	J-M12 (J2b)
EGAN00001390961	65	3,650 <sup>b</sup>	73,701,096	1.24	124	F	K1a2	
EGAN00001390963	75	3,750 <sup>b</sup>	128,355,897	2.32	164	F	R2	
EGAN00001390952	46	3,750 <sup>b</sup>	23,323,399	0.40	53	F	H1bc	

<sup>a</sup>Radiocarbon date

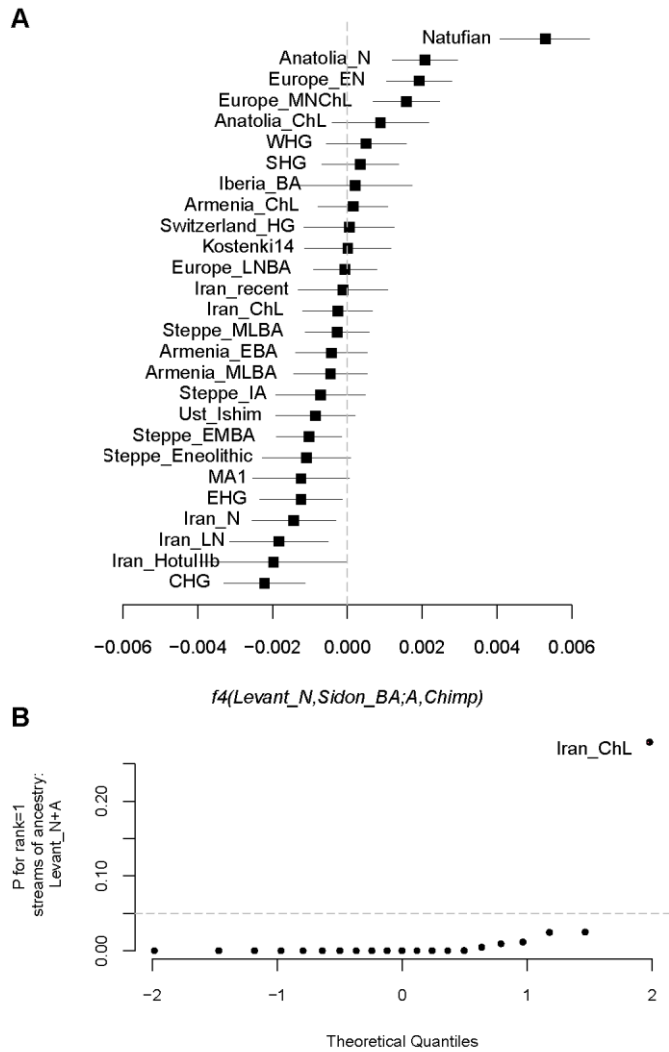
<sup>b</sup>Archaeological date

<sup>c</sup>excluding PCR duplicates

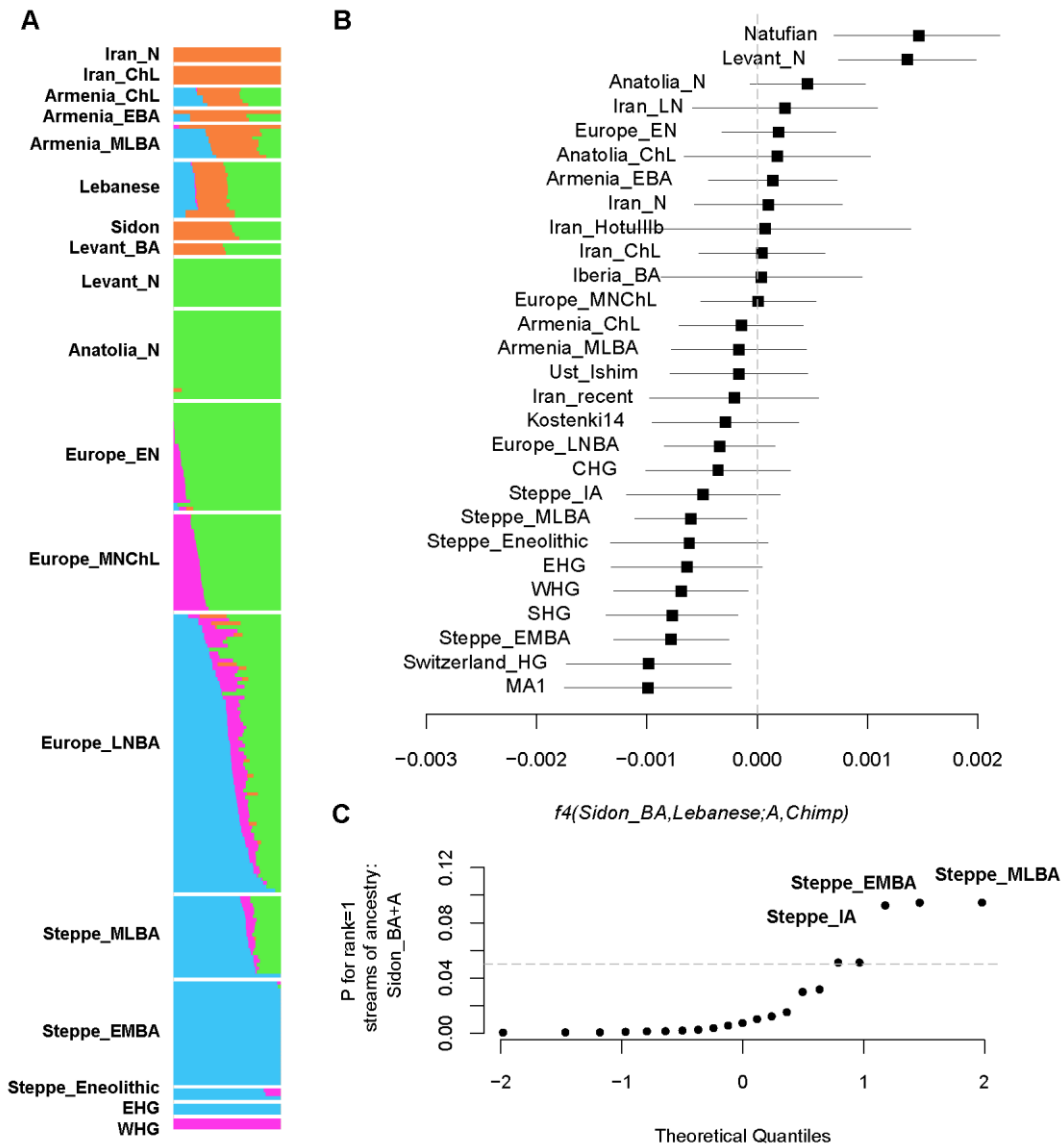
<sup>d</sup>genetically determined



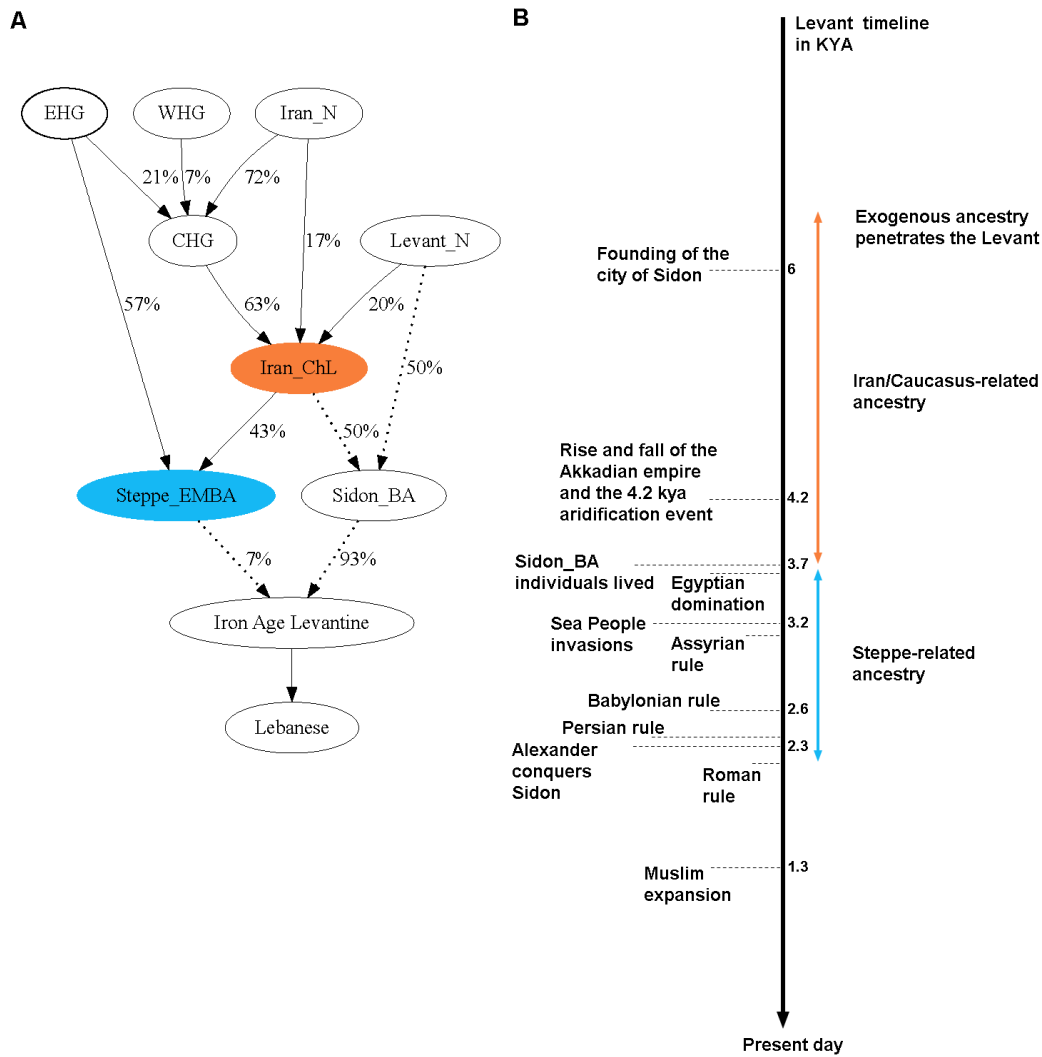
**Figure 1. Population locations and genetic structure.** (A) The map shows the location of the newly sequenced Bronze Age Sidon samples (pink triangle labelled with red text), as well as the locations of published ancient samples used as comparative data in this study. (B) PCA of ancient Eurasian samples (colored shapes) projected using eigenvectors from present-day Eurasian populations (grey points).



**Figure 2. Admixture in Bronze Age Levantine populations.** (A) The statistic  $f_4(\text{Levant}_N, \text{Sidon\_BA}; \text{Ancient Eurasian, Chimpanzee})$  is most negative for populations from the Caucasus and Iran suggesting an increase in ancestry related to these populations in Sidon after the Neolithic period. (B) Modelling 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 modelled using  $qpAdm$  as  $0.484 \pm 0.042$  Levant\_N and  $0.516 \pm 0.042$  Iran\_ChL.



**Figure 3. Admixture in present-day Levantine populations.** (A) Supervised ADMIXTURE using Levant\_N, Iran\_N, EHG and WHG as populations with fixed ancestries. 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  $f_4(\text{Sidon\_BA}, \text{Lebanese}; \text{Ancient Eurasian}, \text{Chimpanzee})$  confirms the ADMIXTURE results and is most negative for populations from the steppe and Eurasian hunter-gatherers. (C) Present-day Lebanese can be modelled as mixture between Bronze Age Sidon and a steppe population. The model with mix proportions  $0.932 \pm 0.016$  Sidon\_BA and  $0.068 \pm 0.016$  steppe\_EMBA for Lebanese is supported with the lowest SE.



**Figure 4. Genetic history of the Levant.** (A) A model of population relationships which fits the *qpAdm* results from Lazaridis et al.<sup>13</sup> (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 coloured double-ended arrows with length representing the SE.

## References

1. 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.
2. 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.
3. Hitti, P.K. (1967). *Lebanon in history: from the earliest times to the present*. (London: Macmillan).
4. Haber, M., Gauguier, D., Youhanna, S., Patterson, N., Moorjani, P., Botigue, 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.
5. 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. (2008). Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *Am J Hum Genet* 83, 633-642.
6. Allentoft, M.E., Sikora, M., Sjogren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlstrom, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167-172.
7. Gunther, T., Valdiosera, C., Malmstrom, H., Urena, I., Rodriguez-Varela, R., Sverrisdottir, 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 U S A* 112, 11917-11922.
8. 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.
9. 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.
10. 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.
11. Olalde, I., Schroeder, H., Sandoval-Velasco, M., Vinner, L., Lobon, I., Ramirez, O., Civit, S., Garcia Borja, P., Salazar-Garcia, 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. 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.
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, 5448.
19. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.L., et al. (2013). Complete mitochondrial genome sequence

- of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U S A* *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. Schubert, M., Ermini, L., Der Sarkissian, C., Jonsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernandez, 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.
  22. Korneliusson, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* *15*, 356.
  23. 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.
  24. 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.
  25. Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer, C., Petchey, F., Fernandes, D., et al. (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature* *538*, 510-513.
  26. Haber, M., Mezzavilla, M., Bergstrom, 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.
  27. The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature* *526*, 68-74.
  28. 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.
  29. Delaneau, O., Marchini, J., and Zagury, J.F. (2011). A linear complexity phasing method for thousands of genomes. *Nat Methods* *9*, 179-181.
  30. Browning, B.L., and Browning, S.R. (2016). Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet* *98*, 116-126.
  31. 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.
  32. Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prufer, K., de Filippo, C., et al. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* *514*, 445-449.
  33. 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.
  34. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet* *2*, e190.
  35. 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.
  36. 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. (2012). Genome-wide association study in a Lebanese cohort confirms PHACTR1 as a major determinant of coronary artery stenosis. *PLoS One* *7*, e38663.



37. 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.
38. Li, Y., Vinckenbosch, N., Tian, G., Huerta-Sanchez, E., Jiang, T., Jiang, H., Albrechtsen, A., Andersen, G., Cao, H., Korneliussen, T., et al. (2010). Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet* 42, 969-972.
39. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv*.
40. Weissensteiner, H., Forer, L., Fuchsberger, C., Schopf, B., Kloss-Brandstatter, A., Specht, G., Kronenberg, F., and Schonherr, S. (2016). mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Res* 44, W64-69.
41. 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.
42. 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.
43. deMenocal, P.B. (2001). Cultural responses to climate change during the late Holocene. *Science* 292, 667-673.
44. 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.
45. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655-1664.