



Autosomal genetics and Y-chromosome haplogroup L1b-M317 reveal Mount Lebanon Maronites as a persistently non-emigrating population

Daniel E. Platt¹ · Hovig Artinian² · Francis Mouzaya² · Wissam Khalil³ · Francois G. Kamar^{1,2} · Elizabeth Matisoo-Smith⁴ · Francesc Calafell⁵ · Nassim Nicolas Taleb^{2,6} · Pierre Zalloua^{1,2,7}

Received: 30 January 2020 / Revised: 16 October 2020 / Accepted: 29 October 2020 / Published online: 4 December 2020

© The Author(s), under exclusive licence to European Society of Human Genetics 2020

Abstract

Currently, there are 18 different religious communities living in Lebanon. While evolving primarily within Lebanon, these communities show a level of local isolation as demonstrated previously from their Y-haplogroup distributions. In order to trace the origins and migratory patterns that may have led to the genetic isolation and autosomal clustering in some of these communities we analyzed Y-chromosome STR and SNP sample data from 6327 individuals, in addition to whole genome autosomal sample data from 609 individuals, from Mount Lebanon and other surrounding communities. We observed Y chromosome L1b Levantine STR branching that occurred around 5000 years ago. Autosomal DNA analyses suggest that the North Lebanese Mountain Maronite community possesses an ancestral Fertile Crescent genetic component distinct from other populations in the region. We suggest that the Levantine L1b group split from the Caucasus ancestral group around 7300 years ago and migrated to the Levant. This event was distinct from the earlier expansions from the Caucasus region that contributed to the wider Levantine populations. Differential cultural adaption by populations from the North Lebanese Mountains are clearly aligned with the L1b haplotype STR haplogroup clusters, indicating pre-existing and persistent cultural barriers marked by the transmission of L1b lineages. Our findings highlight the value of uniparental haplogroups and STR haplotype data for elucidating biosocial events among these populations.

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-020-00765-x>) contains supplementary material, which is available to authorized users.

✉ Pierre Zalloua
pierre.zalloua@balamand.edu.lb

¹ Bioinformatics and Pattern Discovery, IBM T. J. Watson Research Centre, Yorktown Hgts, NY 10598, USA

² University of Balamand, East Med Res Institute, School of Medicine, P.O. Box 33 Amioun, Lebanon

³ Department of Arts and Archeology, Lebanese University, Beirut, Lebanon

⁴ Department of Anatomy, School of Biomedical Sciences University of Otago, Dunedin, New Zealand

⁵ Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain

⁶ Risk Engineering, School of Engineering, New York University, New York, NY, USA

⁷ Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA

Introduction

Lebanon is a compact country that is 217 km long and 56 km wide, located along the Eastern Mediterranean coast. Geographically, Mount Lebanon refers to the coastal mountain ridge that runs north to south and makes up about 30% of Lebanon's territory. It is a continuous mountain chain that rises sharply from the coast, peaking at a height of around 3000 m in the north. Historically, Mount Lebanon has been the most important constituent driving social and cultural development of the country. Its natural ecosystem makes it a hot spot for biodiversity that is ideal for farming and cultivation. It has also historically served as a prime refuge during political and religious conflict, due to its rich habitat and protective terrain.

Currently, there are no written historical documents describing the populations that may have occupied Mount Lebanon during the Iron or Bronze Ages or before. Pre-dynastic archeology attests to trade with Phoenicia for cedar, that only grows in mountainous regions [1]. Yet, many of the early archeologists and researchers who

undertook archeological and historical studies on Mount Lebanon believed that it was either not occupied, or very sparsely populated up until the Arab-Muslim conquest in the 7th century CE [2]. Apart from the name Amurru which, in the first millennium BCE, was used to designate a geographical area in the north of hinterland modern-day Lebanon, and its inhabitants (Amorites), there is no other reference to the inhabitants of the Lebanese mountains in the Iron Age [3].

In the first millennium AD, Roman historians, such as Strabo, Pliny and Flavius Josephus referred to the inhabitants of Mount Lebanon as Itureans [4, 5]. They were described as mountain bandits who occupied a rugged terrain that was impossible to cultivate, who lived in shelters underground or in caves along with their livestock. Many of the twentieth century archeologists and geographers who investigated the question of the inhabitants of the mountains of Lebanon and the nature of its human occupation also concluded that it was an unpopulated region during antiquity that was only visited by loggers and hunters [2, 6].

Active archeological research is currently under way investigating the question of the occupation of mountainous and rural areas of Lebanon, in an attempt to better understand the history of the region and to address what appears to be a contradiction in historic reports and other archeological remains showing evidence of active and persistent trade. Recent archeological findings of some of the isolated sites in Mount Lebanon show evidence of human occupation since at least the Bronze Age (3200 BCE). Evidence of agglomerations and fortifications from the Bronze and Iron Age are very common in Mount Lebanon [7]. Further, a considerable number of inhabited sites and sanctuaries dating to the Roman (64 BCE–391/2 CE) and the Proto-Byzantine (391/2 – 635 CE) periods dot the Lebanese mountains from north to south [8]. These sites, covering a large chronological range (from the Bronze Age to the early Arab-Muslim conquest), indicate an intense use of the mountainous territories as agricultural, trade and worship centers.

Today Mount Lebanon is mostly inhabited by the Maronite community. The Druze, a religious group of a common faith called “unitarians (Muwahhideen)” also inhabit parts of Mount Lebanon. There are an estimated one million people of the Druze faith living almost exclusively in the Chouf area of Mount Lebanon, the Houran mountain in Syria, and the mountainous areas west and north of Amman Jordan and the Galilee [9]. The Maronites represent the oldest and most dominant Christian branch within Lebanon dating to the late 6th century CE [10, 11]. While Syriac Orthodox may claim to be older, their presence in Lebanon is the result of recent diaspora. The Maronite community is spread throughout Lebanon but is the most prevalent community in the North Lebanon Mountains and Mount Lebanon in general. The Maronites draw their name from

the Syrian Monastery that was built to honor the pious Marun or Maro, who, according to Maronite church historians, lived in the early 6th century on the banks of the Orontes river, between Emesa and Apamea [11, 12]. The Maronite Church was thus formed during the tumultuous times of the development of Christian doctrine during Byzantine rule, continuing during the period of the Muslim expansion and conquest of Lebanon by the Umayyad Caliphate. The Maronites are the oldest historically attested isolated community that found refuge in the mountainous regions, which identifies it as a candidate for providing a unique genetic window into the demographic history of the region. Specifically, differentiation between mountain and coastal Maronites may reveal patterns of immigration of some specific genetic lineages.

We have previously shown, based on Y-chromosome and mitochondrial DNA analyses, that religious communities in Lebanon genetically differentiated prior to the establishment of religions [13–15]. We also showed that these communities have been influenced by local isolation, while evolving primarily within Lebanon, as evident by their Y-haplogroup distributions [14–16]. Among these, the haplogroup L1b (formerly L2), defined by the M317 SNP (L1b-M317: GRCh38 Y-chromosome: g.20592785–20592786 del GA) [17] is over-represented (p value = 2.28×10^{-12}) among the Maronites of the North Lebanon Mountains. The L1 haplogroup is divided into three sister branches, the L1a branch, defined by the M2481 SNP, found mostly in India and Pakistan; the L1b branch with representation in Central and Southwest Asia and some parts of Europe, which is defined by the L656 SNP and the L1c branch defined by the M357 SNP (https://isogg.org/tree/2011/ISOGG_HapgrpL11.html). The abundance of L1b-M317 lineages in Mount Lebanon and its near total absence from the rest of the Lebanese and surrounding populations, irrespective of their religious backgrounds including Maronites, calls for further analyses, promising to identify the factors that may have created such strong genetic stratification. Genetic analyses of the Mount Lebanon and surrounding communities focusing primarily on haplogroup L1b-M317 may provide details related to the origin and the immigration time and dispersal routes of this lineage into the mountain community and how it evolved within the region.

In this study, we selected candidate populations that may have contributed genetic admixture to the Lebanese population in general, and to Mount Lebanon in particular. Throughout ancient and modern history, Lebanon has been the site for multiple events that were associated with major human mobility. Large scale population movements have occurred in and out of Lebanon since the rule of the Babylonians and Assyrians, and have continued through the Roman, Persian, Crusader, Islamic

Expansion and Ottoman occupations. Some of these foreign conquests that involved large armies, and long-term occupations, must have certainly led to noticeable gene flow. While some of the more recent genetic admixture events, like those resulting from the Crusades and Islamic Expansion, are likely to be more readily detectable and have indeed contributed genetically, their impact alone is not significant enough to explain the observed genetic diversity and its patterns observed in the modern population that occupies Lebanon today, especially the populations of Mount Lebanon [14].

We used Y-chromosome data from Lebanon and other surrounding regions, including populations from Arabia, Armenia, the Caucasus, Greece, Italy, the Levant, Mesopotamia, and Turkey, and autosomal data primarily from communities in Mount Lebanon and from Beirut, to explore the genetic makeup of the Mount Lebanon communities and to trace the migration patterns that likely contributed to the genetic patterns observed in the mountainous regions of Lebanon. We also sought to investigate the presence or absence of genetic clustering observed within the various communities that currently occupy Mount Lebanon.

Material and methods

Populations and samples

For the Y chromosome analyses, a total of 6327 samples were used in this study and were selected based on analysis of their historic, ethnic, and geographical context. Of these, 2944 samples came from our group (511 genotyped for the current study and 2433 genotyped and published previously by our group [13, 14, 16]). These samples were actively collected in the field, and included Lebanese ($n = 1403$), Iranians ($n = 416$), Syrians ($n = 552$), Cypriotes ($n = 166$), Kuwaitis ($n = 42$), and Palestinians ($n = 365$). Data for the remaining 3383 samples came from the literature and included Turkey (452 samples), Armenia (1010 samples), Iran (104 samples), Kuwait (117 samples), Libya (175 samples) and the Caucasus (1525 samples) [14, 16, 18–23] (Supplementary Table 1).

The 178 Lebanese samples collected for autosomal analysis in the current study were obtained from the North Lebanese Mountain Maronite (NLMM) community ($n = 53$), from Other non-Marionite Lebanese Christians (OLC) living in Beirut Greek Orthodox community ($n = 43$), from Shiia community members ($n = 35$), and Sunni community members ($n = 47$). An additional 431 samples were also collected from Armenia ($n = 61$), Cyprus ($n = 16$), Iran ($n = 148$), Iraq ($n = 50$), Jordan ($n = 43$), Syria ($n = 71$), Turkey ($n = 39$), and the Ukraine ($n = 3$) making a total of 609 samples used in the autosomal analyses (data shown in Supplementary

Table 2). These samples were selected based on analysis of their historic, ethnic, and geographical context and were actively collected in the field. For genotyping, 594 samples were processed on Infinium Omni Express-24 v1.2 Bead Chip in two experiments yielding 701800 SNPs and 677830 SNPs respectively, and 15 of the samples were processed on the Human 1M-duo v3.0 array, yielding 1089836 SNPs. Additional information on DNA extraction and genotyping are presented in the Supplementary Information.

Published ancient DNA data were used representing 127 samples [24] from Anatolia_ChL ($n = 1$), Anatolia_N ($n = 24$), Armenia_ChL ($n = 5$), Armenia_EBA ($n = 3$), Armenia_MLBA ($n = 9$), Iran_ChL ($n = 5$), Iran_LN ($n = 1$), Iran_N ($n = 5$), Levant_BA ($n = 3$), Levant_N ($n = 13$), Natufian ($n = 6$), Steppe_EMBA ($n = 28$), Steppe_Eneolithic ($n = 3$), Steppe_MLBA ($n = 22$). Details on these samples are provided in Supplementary Table 3. Other samples were from Mbuti.DG ($n = 3$) [25], and 1000 Genome Project samples GBR ($n = 91$), CEU ($n = 99$), FIN ($n = 99$), GIH ($n = 103$), and IBS ($n = 107$) [26]. Data sets were lifted to GRCh38, checked against GRCh38, shared SNPs were selected, tri-alleles, rare alleles, and extreme HW deviations were removed, and phases corrected yielding 295,307 SNPs, and the datasets were merged. Additional information on these analyses can be found in the supplementary material.

The study participants recruited by our team provided three generations of paternal ancestry and detailed information on their geographical origin and religious group or religious affiliation. They all signed an informed consent form approved by the IRB of the Lebanese American University. All the Lebanese samples shown in Supplementary Table 1 were used to calculate the L1b haplogroup frequencies across the various religious communities and geographical regions within Lebanon (Table 1 & Supplementary Table 4).

For our STR analyses and in order to have a balanced representation of the L1b Haplogroup among the various regions we included, in addition to our samples ($n = 2524$), 129 L1b samples from key comparative regions (Caucasus, Turkey, Iran, Italy and Greece) obtained from published work [27] and online available data from Family Tree DNA (Supplementary Table 5).

Finally, for the L1b analyses the study samples were divided into regions (10 in total) according to their geographic distribution (Supplementary Table 6). The North Lebanese Mountain Maronites (NLMM) represent the Maronite community in the Northern Lebanon Mountains.

Analyses

NETWORK 5 (Fluxus-Engineering: <http://www.fluxus-engineering.com/>) was applied to determine a reduced median network of L1b STR haplotypes using a reduction threshold of 1 [28]. These networks were constructed with the

Table 1 Haplogroup L1b M20 frequencies in Lebanon and surrounding regions.

Region	Nb. L1b (M20) Hg	Total population	Frequency	Source
Lebanon	43	1403	3.06%	Present Study; El-Sibai et al. [16]
Turkey	13	452	2.88%	Cinnioğlu et al. [18]
Armenia	11	1010	1.09%	Hovhannisyann et al. [19], Weale et al. [20]
Palestine	3	365	0.82%	Zalloua et al. [14]
Iran	4	520	0.77%	Alshamali et al. [21], El-Sibai et al. [16]
Kuwait	1	159	0.63%	El-Sibai et al. [16], Triki-Fendri et al. [22]
Cyprus	1	166	0.60%	Zalloua et al. [14], El-Sibai et al. [16]
Libya	1	175	0.57%	Triki-Fendri et al. [23]
Syria	3	552	0.54%	Zalloua et al. [14], El-Sibai et al. [16]
Caucasus	8	1525	0.52%	Balanovsky et al. [41]

13 Y-STRs common to all samples DYS19, DYS389i, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, Y-GATA-H4, DYS456 (Supplementary Table 6) as well as the truncated set DYS19, DYS389i, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS456, to test stability. “Virtual UEPs” marking distinct L1b branches were identified in the NETWORK output, with UEP assignments identified in SR4.

Arlequin 3.5 [29] was applied to compute R_{ST} distances between populations. Euclidean distances between individual STR genotypes were clustered via UPGMA and plotted via heatmap in R 3.5.0 [30] (<https://www.r-project.org/>).

BATWING [31] was employed to compute TMRCAs for the main NETWORK-identified branches characterizing NLMM, as well as other L1b groups in Lebanon, and overall, each marked with a “virtual” UEP. Mutation rate sets were obtained from multiple published sources, with Zhivotovsky’s evolutionary rate, as well as three pedigree based [32–35] tabulations. Generation times were assumed to be 25 years for the evolutionary rates, and 30 years for the others. All were applied in separate runs to the 13 STRs described above in the NETWORK analysis. Further the Burgarella ImMR rates [32] were applied to the truncated list of 10 STRs also described in the NETWORK analysis, in order to test dependency on numbers of STRs. BATWING was first applied to the L1b samples alone tabulated in Table 2a, and then applied to a representative sample of the region drawn across haplogroups including those tabulated in Table 2b given that BATWING’s coalescence computations assume a random distribution of haplogroups and that haplogroup distributions of STRs are not random.

Results

Haplogroup distribution and frequencies

Frequencies of Y-chromosome haplogroup L1b were obtained from 10 different populations (Supplementary

Table 1). The L1b haplogroup frequencies were the highest in Lebanon, reaching 3.06%, followed by Turkey (2.88%) and Armenia (1.69%) (Supplementary Fig. 1, Table 1). Libya, Syria and the Caucasus had the lowest frequencies, at 0.57%, 0.54%, and 0.52% respectively. We note that L1b is present along the Fertile Crescent, the Levant, the Caucasus, and through the Mediterranean where Phoenicians traded, including Egypt, Cyprus, and Libya. Nowhere does it rise to the 12.5% level seen in the NLMM. While Turkey is unrepresented since the authors did not test SNPs informative of L1b, Kuwait showed no L1b. Within Lebanon (Supplementary Table 7), the frequency of the L1b haplogroup was significantly higher in the Maronite community of Northern Mount Lebanon (NLMM) compared to any other community or any other group tested in the region (Fisher exact test p value = 7.326×10^{-13}) and for L1b with Maronites compared to others in Northern Lebanon Mountains as well (Fisher exact test p value = 0.0014), consistent with our previously published results [13]. The L1b haplogroup was present in 6.4% of the Maronites compared to 3% in the general Lebanese population. Out of the 43 L1b haplotypes found in Lebanon, 34 were from Maronites from the North Lebanese Mountains.

The Y-haplogroup frequencies in our genotyped Lebanese population show gradients dominated by J2 (coastal) and J1 (inland) with J2 being the most frequent (28.7%) and J1 the second most frequent (18.3%) haplogroup. Haplogroup E1b shows strong differentiation by region irrespective of religious affiliations and was the third most frequent haplogroup (16.3%) in Lebanon. R1b reaches a frequency of 8.9% and its distribution behaves similarly to E1b but without representation in South Lebanon (Supplementary Table 4).

Network, clustering, and MDS analysis

We assessed the distribution of the L1b haplogroup through STR network analysis by observing the geographic distribution of the L1b STR haplotypes. NETWORK results

Table 2 STR derived time estimates based on L1b samples. STR derived time estimates based on all haplogroups.

Mutation model	L1b TMRCA	UEP1 TMRCA	UEP2 TMRCA	UEP3 TMRCA	UEP4 TMRCA
Zhivotovsky	36 (21–71)	27 (12–57)	16 (7.9–33)	17 (8.4–35)	17 (8.4–35)
Burgarella ImMR	11.5 (8.0–19.2)	9.0 (4.8–16.5)	5.5 (3.0–9.7)	5.7 (3.2–9.9)	5.8 (3.2–9.8)
Shi OMR	10.9 (7.5–19.2)	8.2 (4.3–16.9)	5.0 (2.8–9.4)	5.2 (2.9–9.5)	5.4 (3.0–9.9)
Burgarella OMR	11.9 (8.2–20.2)	9.1 (4.7–17.7)	5.6 (3.1–10.0)	5.8 (3.2–10.2)	5.9 (3.3–10.6)
Mutation model	L1b TMRCA UEP5	UEP1 TMRCA	UEP2 TMRCA	UEP3 TMRCA	UEP4 TMRCA
Zhivotovsky	20 (13–28)	15 (6.9–23)	9.6 (4.9–17)	10 (5.1–17)	9.9 (5.0–18)
Burgarella ImMR	10.6 (7.4–14.6)	6.4 (4.3–10.2)	4.3 (2.6–6.8)	4.3 (2.6–6.8)	4.5 (2.8–7.2)
Shi OMR	10.8 (7.4–15.1)	7.3 (4.5–10.7)	4.6 (2.8–7.8)	4.8 (3.1–7.9)	4.6 (2.8–7.5)
Burgarella OMR	9.8 (7.5–13.8)	7.5 (4.8–10.5)	5.1 (3.1–8.2)	5.2 (3.2–7.9)	5.2 (3.2–8.5)
Burgarella ImMR cut STRs	7.6 (4.7–17.3)	5.4 (3.3–9.0)	3.8 (2.0–6.5)	3.9 (2.2–7.0)	3.8 (2.1–6.7)

“OMR” = “observed mutation rates” by both Shi et al. [33] and Burgarella et al. [32], “ImMR” = “logistic model mutation rates” by Burgarella et al. – each with 30 year generation times, and Zhivotovsky “EMR” = “evolutionary mutation rates” estimated by Zhivotovsky et al. [35] with 25 years generation times. A recalibrated evolutionary mutation rate “rEMR” by Shi et al. differed from the Zhivotovsky EMR in an STR not included in these analyses, so was identical to the Zhivotovsky EMR for our purposes. Shi et al. also presented summary parameters for the other references. UEP2 is a NETWORK-determined branched cluster primarily comprised of NLMM, UEP3 is heavily NLMM but includes other Lebanese as well. UEP4 is non-NLMM Lebanese. UEP1 is the union of these, representing modern L1b in Lebanon.

UEP5 marks all the L1b’s among all the samples, and it is used to compute L1b’s TMRCA. “OMR” = “observed mutation rates” by both Shi et al. [33] and Burgarella et al. [32], “ImMR” = “logistic model mutation rates” by Burgarella et al. – each with 30 year generation times, and Zhivotovsky “EMR” = “evolutionary mutation rates” estimated by Zhivotovsky et al. [35] with 25 years generation times. A recalibrated evolutionary mutation rate “rEMR” by Shi et al. differed from the Zhivotovsky EMR in an STR not included in these analyses, so was identical to the Zhivotovsky EMR for our purposes. Shi et al. also presented summary parameters for the other references. The ‘cut STRs’ line used the Burgarella ImMR rates but with a truncated list of STRS DYS19, DYS389i, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS456. UEP5 is all the L1b sets.

are shown in Fig. 1. NETWORK offers qualitative guidance for phylogenetic relationships when applied to STR haplotypes. The results were explored by application of BATWING. This NETWORK has three heavily localized branches, primarily composed of NLMM samples, that share a common root from the L1b network in general. To these three branches, virtual UEP (unique event polymorphism) labels UEP2, UEP3, and UEP4 were assigned. UEP2 was comprised almost entirely of NLMM samples, but includes one Italian; while UEP3 contained substantial NLMM samples as well as samples from elsewhere in Lebanon, and two Italians; and UEP4 was comprised of individuals broadly distributed throughout Lebanon. Among these, the UEP4 cluster did not maintain coherence under NETWORK analysis using the 10 STR subset (Supplementary Fig. 2). UEP1 represents the union of UEP2, UEP3 and UEP4 capturing the entire Lebanese L1b. The specificity of UEP2 to NLMM p value was a significant 4.467×10^{-12} by Fisher exact test. UEP5 represents all L1b samples for the TMRCA computation in the presence of other haplogroups.

The NETWORK computation was cross-checked by application of UPGMA clustering based on Euclidean distances between individual sample STRs, and plotted on a heatmap, shown in Supplementary Fig. 3. Sidebar colors correspond to the NETWORK population colormap described

above, and show distinctive clusters marked pale and dark green indicating NLMM (containing UEP2 and UEP3) and Levantine samples (capturing UEP4), respectively. The three virtual UEP groups identified by NETWORK are largely confirmed by the Euclidean distance UPGMA clustering, showing several branches carrying L1b lineages. The UPGMA clusters agree with NETWORK in also indicating that these clusters split relatively early in the history of L1b, though they are all primarily found in Lebanon. One interesting difference is that one of the Italians that the reduced median network analysis placed in UEP2 was not associated with UEP2 haplotypes by UPGMA clustering by Euclidean distances. Turkish and Greek L1b haplotypes also tend to be marked by a distinct division among L1b STR genotypes compared to the other branches, also reflected in the NETWORK structure. The Arlequin R_{ST} population differences were also subject to UPGMA (Supplementary Fig. 4) and MDS clustering (Supplementary Fig. 5). The UPGMA clustering separates the population groups identified in the Euclidean distance UPGMA clustering between individuals in the heatmap. Individuals tend to break into 3 major sub-clusters, with some mixing between populations. The R_{ST} statistics reflect the similarity among individuals in these branches, accounting for admixture affinities. The Greek/Caucasus/Turkish cluster emerges clearly in both clusters. The R_{ST} grouping of populations included non-NLMM

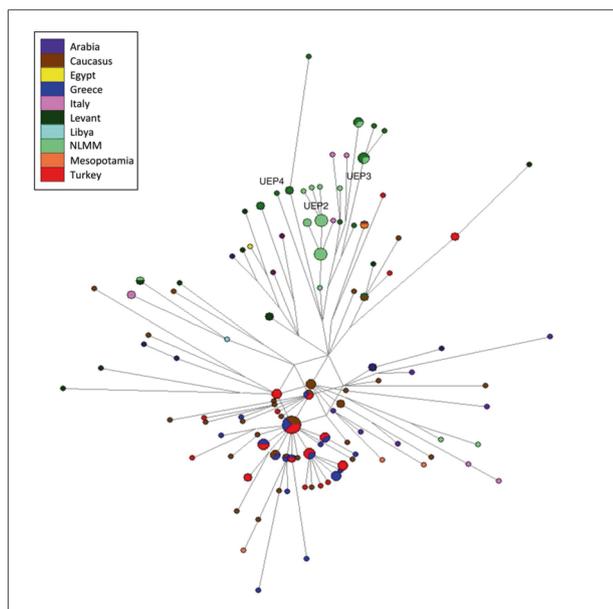


Fig. 1 Reduced median network of L1b M20 showing 13 STR haplotype distribution among populations; area is proportional to haplotype frequency, and color indicates populations. Connecting lines represent putative phylogenetic relationships between haplotypes. The three heavily localized branches that share a common root from the L1b network are assigned on the figure as UEP2, UEP3, and UEP4. Color code: Arabia - purple, Caucasus - brown, Egypt - yellow, Greece - blue, Italy - pink, Levant - dark green, Libya - light blue, NLMM - light green, Mesopotamia - orange, and Turkey - red.

Lebanon and Syria as the “Levant.” The MDS analysis in 2D showed relatively poor STRESS (0.178) and R2 (0.406) suggesting that the data are not well represented in a simple 2D plot. Nevertheless, the samples do reflect the organization of the UPGMA clustering, except that MDS appears to agree more closely with the Reduced Median network for the placement of the Egypt sample compared to the UPGMA R_{ST} clustering.

Time and place of origin of L1b branches

We used BATWING to investigate timing of isolation or group splits while taking into consideration admixture events and immigration. In estimating times of splits, BATWING will assign times based on dates of last admixture events, even when limited gene flow is involved (by immigration or admixture), thus it may significantly reduce split time estimates. For that reason, this report focused on TMRCA times of UEPs to mark expansion events.

BATWING computations were performed to estimate the TMRCA of UEP2, UEP3, and UEP4, their common ancestor in UEP1, and L1b (Table 2). UEP1 is the union of UEPs 2, 3, and 4, and represents a cluster of identifiable Lebanese L1b haplotypes. Table 2a displays estimates based only on L1b samples. The TMRCA for the whole

measures the age of L1b. Table 2b shows the results when we applied BATWING to samples drawn from regional populations across haplogroups, with UEP5 marking L1b in this analysis. The first 4 entries applied mutation rates from multiple studies yielding four distinct estimates of TMRCA, while the 5th entry applied Burgarella et al.’s ImMR rates [32] to the 10 STR truncated set. BATWING estimates for the virtual UEPs are each around 4.6–5 ky, and results suggest that they shared a common ancestor around 7.3 ky. We computed the age of L1b in two ways. First, using only L1b haplotypes, BATWING identified an overall TMRCA of around 11ky. When combined with other haplogroups, for a more representative population, UEP5 was assigned to mark L1b chromosomes (M317). This showed a TMRCA of around 10.6ky. The UPGMA on individual Euclidean distances does not suggest an oldest region, though the NETWORK reduced median plot suggests a center of radiation with the Caucasus near the center. The Caucasus haplotypes, marked brown in Fig. 1, are also distributed over a majority of the branches. Lastly, the estimates based on the 10 STR subset tended to be slightly later than the 13 STR set estimates based on the same mutation rates, with UEP2 and UEP3 being most similar. UEP4, which lost integrity as a NETWORK branch, showed a larger but insignificant difference, as did the L1b age estimate listed under UEP5.

Autosomal analyses

We applied Eigenstrat [36] for PCA, and we also applied ADMIXTURE [37], qpGraph [38], and qpAdm [38] to samples in Supplementary Tables 2 and 6 as well as to a selected 1000 genome data set with Mbuti samples.

The PCA results, computed with least-squares projections, are shown in Fig. 2. Figure 2a highlights the samples sharing the most information with the principal component. The PCA shows the North Lebanese Mountain Maronite samples strongly localized in the upper-left quadrant, Iranian and Iraqi samples in the upper-right quadrant, Jordanian, Syrian samples with some Iraqis in the lower quadrant. In the middle top, Armenians are placed just above the ancient Steppes groups and Ukrainians, and mix into Iraqi samples, then Turkish, and Iranian samples are spread from center to the right. Moving from the top center towards the left we see Cyprus and other Lebanese Christians, and the tail of the North Lebanese Mountain Maronite samples. Mutual information, augmented by the sign, between two normal factors A and B , shown in Fig. 2b, was computed as $IM_{AB} = -\text{sgn}(\rho_{AB}) \ln \sqrt{1 - \rho_{AB}^2}$ (Supplementary Information). In this case, we consider shared information between principal component as B and individual scaled and centered genotypes as A . We show in the Supplementary

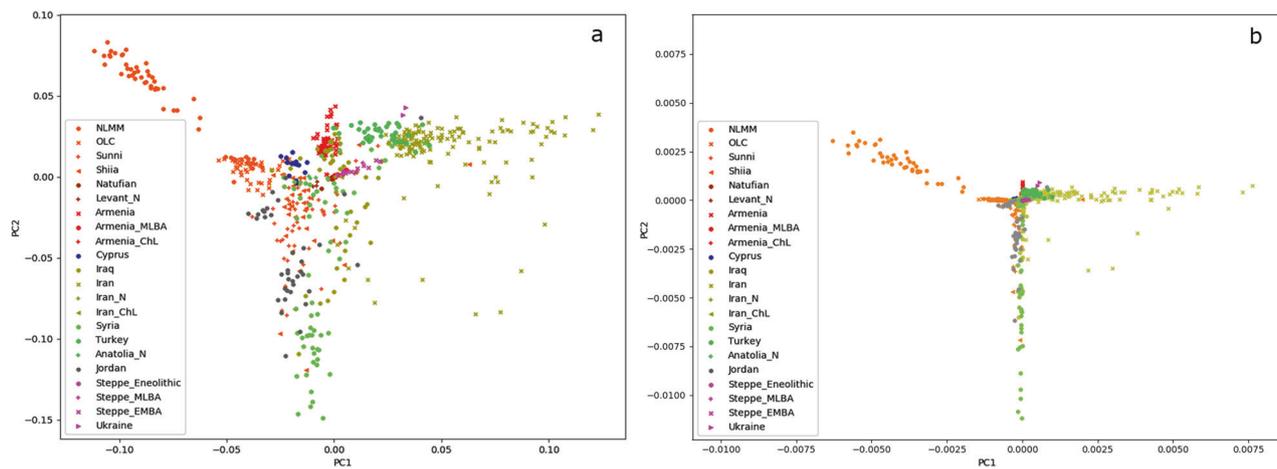


Fig. 2 PCA computed and displayed for the full list of populations indicated in Supplementary Tables 2 and 6. **a** PCA components 1 and 2 for recent and ancient regional populations, highlighting

differentiation of NLMM from all other regional populations. **b** Mutual information scaling for principal components indicating shared bitwiseness of genotypic information between samples and principal components.

Information that the correlation between samples X and singular value decomposition genotype components V is U , which comprises the plotted Principal Components. The information scaling in Fig. 2b suggests that the principal components are dominated by inter-regional genetic variations. PC1 is primarily Iranian. PC2 carries Syrian, Jordanian, Lebanese Shia and Sunni. Turkey, Steppes, and Armenia project to the origin, suggesting little shared information between their variation and PCs 1 and 2. Most of the NLMM samples share information with both PC1 and PC2, but are anti-correlated with the others, while a scattering of Iranian samples share information with PC2.

ADMIXTURE over a range of K 's, shown in Fig. 3, identified basal populations among Natufians, Early-Middle Bronze Age Steppe, and Neolithic Anatolia. North Lebanese Mountain Maronites appear to be dominated almost entirely by an ancestry component, colored pale blue in $K=5$, that is widely admixed across the Fertile Crescent populations, including those represented in ancient DNA. This suggests that this component's presence in the North Lebanese Mountains may be older than the aDNA dated populations. Another component heavily represented among Iranian samples, colored green, is also widespread throughout the more modern Fertile Crescent populations, and is represented in Armenia in most modern DNA samples. Anatolia_N is marked by an ancestral population in yellow, and shows no Iranian genetic contributions, while Turkey shows significant admixture. Lastly, Natufians are identified as a population marked by dark blue, which shows up in low contributions to modern Lebanese populations, North Lebanese Mountain Maronites least of all, though Syria and Jordan show contributions.

ADMIXTURE had identified this ancestral component as being relatively unmixed in the NLMM population, while showing admixture contributions even in very ancient aDNA samples. PCA had positioned this component in an extreme position distinct from any other regional population, and intermediary to no others.

This is suggestive either of a source ancestral to all other ancient samples in the region, including Natufians, or else it is a modern result of drift. Since NLMM showed such a distinctive position in the PCA and appears to be such an ancestral population in the ADMIXTURE analysis, we sought to understand whether this was a highly isolated population, or if this could be explained by modern admixture. To perform these tests, we constructed two candidate qpGraph topologies, working from a base in Africa, building a base topology by testing likely coalescences and admixtures, until two NLMM hypotheses could be constructed: one with NLMM deriving primarily from relatively ancient ancestral lineages similar to surrounding populations derived ultimately from Africa (Supplementary Fig. 6a), and the other candidate topology representing a relatively modern admixture of neighboring ancient populations (Supplementary Fig. 6b). The qpGraph offers Z -scores as a measure of deviation of the model from data. Without seeking to find a complete model, we sought to test whether a model of recent drift vs. more ancient expansions and admixtures mostly close corresponded to the observed populations. The Z -scores showed much less deviation for ancient admixtures vs. modern drift as an explanation for the structure of NLMM populations, suggesting that an ancient interpretation for NLMM fits the observations from ADMIXTURE more closely. The qpGraph tests yielded a z -score of 96.3 for the modern

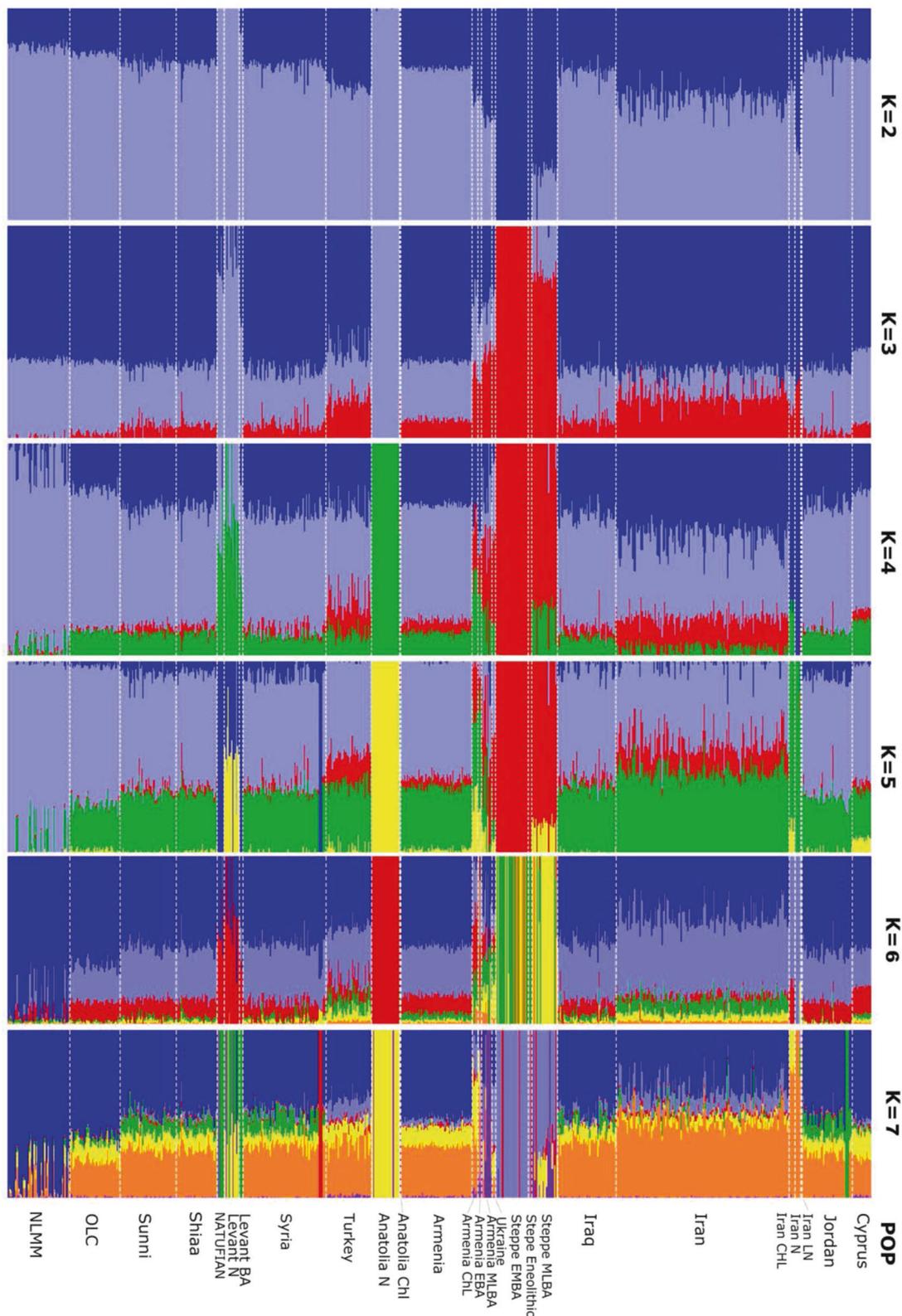


Fig. 3 ADMIXTURE plot for admixtures computed on the full population list with populations plotted listed in Supplementary Tables 2 and 6. ADMIXTURE ancestral components computed for

regional populations with $K = 4$ and 5 shows that NLMM contains the strongest regional representation of an ancestral component found distributed among Levantine and Fertile Crescent populations.

Table 3 qpAdm runs predicting NLMM from Neolithic Levant, Neolithic Iran and either Armenia or Chalcolithic Armenia (labeled 1st population). Right base population was GBR, with remaining populations being CEU, FIN, GIH, and IBS.

1st population	χ^2_2	<i>p</i> value	1st population	Levant_N	Iran_N
Armenia	2.090	0.3517	0.773 ± 0.034	0.216 ± 0.023	0.011 ± 0.013
Armenia_ChL	7.081	0.0290	0.446 ± 0.039	0.445 ± 0.026	0.109 ± 0.018

admixture topology, while the topology placing NLMM origins in the more ancient derivation yielded a z-score of 18.98, supporting the notion that the ADMIXTURE identified an ancient North Lebanese Mountain populations' ancestral component, rather than a product of more recent drift.

The basal NLMM population that ADMIXTURE and PCA identify, and that qpGraph supports does not provide information about admixtures suggested by the Y-STR analysis involving L1b admixture events. We note that qpAdm provides a way to use the “right-hand” populations to probe time slices by selecting differences primarily marking the relevant time periods and locations, performing a regression to test goodness of fit of candidate models expressed in terms of F_4 statistics. To that end, we selected GBR as the right-hand “base,” and FIN, CEU, GIH, and IBS as the remaining right-hand populations [38] to capture Neolithic through iron-age events. Then we sought to test which admixtures of either Neolithic Armenia or modern Armenia together with Neolithic Levant, Neolithic Iran, and Armenia most closely represented NLMM in the right-hand population time-slices (Table 3).

The Armenia_ChL model scored a *p* value of 0.0290, while the modern Armenia model scored 0.3510 (the better fit), suggesting that modern Armenia includes more of the admixing genetic lineages represented in modern NLMM than Chalcolithic Armenians. Therefore, either the NLMM were comprised of Chalcolithic Armenians, and both Armenians and NLMM received subsequent admixture from elsewhere, or Armenians had received post Chalcolithic admixture and the subsequent admixed Armenian group admixed into NLMM. Further tests seeking to identify a later admixture to add to the Chalcolithic Armenians among regional ancient populations (e.g., Neolithic Levantines or Natufians) did not produce improved qpAdm models. Therefore, among regional candidates, ancient Chalcolithic admixture does not account for components shared between modern Armenian and NLMM. The admixture coefficients for this model were 0.773 ± 0.034 for Armenia, 0.216 ± 0.023 for Levant_N, and 0.011 ± 0.013 for Iran_N, indicating that Iranian genetic contributions are not significant, or were adequately explained by modern Armenian contributions.

Homozygosity analysis was performed to see if the isolation of NLMM was distinct from that of other groups within Lebanon that we explored. Observations for similar

subpopulations in surrounding regions were not available for comparison. Figure 4 indicates distributions of distinct homozygosity inbreeding coefficients associated with each population. Both the NLMM and OLC are relatively isolated communities, though with different histories. The Sunni and Shia populations may have more broad regional distribution, yet show similar ranges of homozygosity, overall. However, their homozygosity distributions are distinctive, suggesting distinct origins of their homozygosity – perhaps due to consanguineous autozygosity. Both the inbreeding coefficients and runs of homozygosity present similar features to each other for all populations, except for the Shia, which includes some samples with low inbreeding coefficient estimates, not reflected in the ROH estimates. All four groups show individuals with very comparable homozygosity. The modal inbreeding coefficients for NLMM are largest across all groups, with the range of variation strongly overlapping across populations. The Christian individuals do not contain the tail of very strongly homozygous samples suggestive of consanguineous autozygosity that was observed in the Muslim samples. Autosomal data are available via EVA Study Browser at <https://www.ebi.ac.uk/eva>, Project #PRJEB39630 and Analyses #ERZ1468073.

Discussion

The NETWORK, UPGMA individual clustering and BATWING analyses indicate that the L1b Levantine STR cluster branches carry diversity that took around 5ky to generate, and that these clusters shared a common ancestor 7300 years ago. NETWORK analyses strongly suggest a common root for all the major expansions of L1b to be from the Caucasus. BATWING analysis suggests that the times of common ancestors between Greek/Turkish vs. Levantines is much older than the evolution of L1b within Lebanon, tending to refute other population expansion centers. Therefore, the Levantine STR cluster population most likely split from lineages from the Caucasus.

Migration from the Caucasus was likely triggered by the amelioration of the climate and induced migratory events driven by the search for better habitats throughout the Northern Levant. A major climatic event around 8200 years ago caused severe drought throughout the Levant and most of Europe and southwest Asia [39], after which, multiple expansion events took place across Southwest Asia.

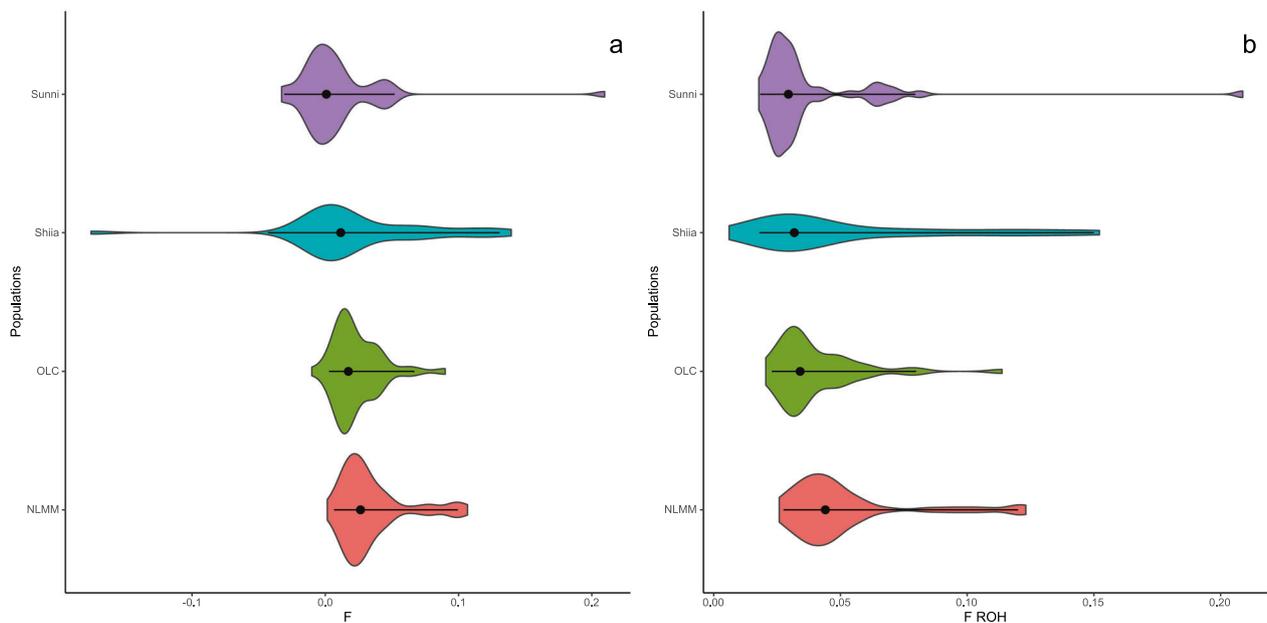


Fig. 4 Patterns of homozygosity in 4 religious groups in Lebanon. Plots of homozygosity computed **a** by co-ancestry, and **b** runs of homozygosity between strands per sample, showing a frequency chart as a fiddle diagram, and median and 95% quantiles of the fixation coefficients.

After arriving in the Levant, NETWORK-identified STR haplotype branches carrying lineages from these migrants split into different groups, with sufficient specificity to be statistically highly significant, such as the Northern Lebanese Mountains population, between 5000 and 4600 years ago. All archeological evidence recovered to date from various locations in Mount Lebanon provides dates that are consistent with our DNA findings, suggesting assimilation into community life in the Lebanese Mountains around the early Bronze Age (5000 years ago) [7, 8].

The Lebanese Northern Mountains, with their steep topography, constituted a suitable habitat for the Levantine travelers, providing shelter, and lush and verdant spaces with good precipitation. The inhabitants of Northern Lebanon have therefore been present there since at least the early Bronze Age, as attested to by archeological evidence. Interestingly, the modern genetic record indicates complete affiliation with the Maronite faith, without the conversions to Islam observed in the regions outside of the northern Lebanese mountains.

The L1b haplogroup showed different frequencies and haplotype differentiation among all regional and religious groups in Lebanon. L1b was almost exclusively present in the Maronites of Mount Lebanon and is nearly absent in all of the other religious groups tested. This haplogroup is also nearly absent from most of the Levant and the rest of Southwest Asia, while it is present in Turkey and the Caucasus. Two important observations can be stated about the haplotype differentiation of this haplogroup in the Levant. First, there is substantial geographical structure seen in the Levant. The Network analyses identified STR

branches that are strongly localized to the Mount Lebanon highlands (L1b NLMM). Second, this L1b NLMM is exclusively found in Maronites and it is present within a more diverse, but geographically restricted Levantine branch (L1b Levant). These observations indicate that the L1b NLMM marked by UEP2 has remained relatively isolated over its estimated 5000-year history with very little emigration to surrounding coastal or other inland populations that have been well established before then. Otherwise we would have observed UEP2 members in other regions. The Euclidean distance based UPGMA heat map clusters support the NETWORK results showing distinctive clusters of STRs strongly connected to the northern mountains of Lebanon. UEP3 appears to be much more broadly distributed in Lebanon with some random sample of its STR haplotypes appearing in the NLMM. It would appear that this branch evolved in the Levant, with relatively recent immigration into the NLMM. Both UEP2 and UEP3 have very similar ages, with only slightly older UEP3 dates than for UEP2. UEP4 is another non-NLMM Lebanese cluster that shares a similar age with UEP2 and 3.

We also observe that the NLMM community was not completely isolated. Other UEPs, besides UEP2, that appear to be rooted elsewhere in Lebanon include individuals now living in the NLMM community, so immigration from surrounding populations *into* the region did occur. Further, we see indications that some NLMM of the inhabitants left the region, as seen by the presence of the distinct genetic traces identified as far away as Italy. However, the numbers are quite low, perhaps due to the small population size in the

region. Such low numbers would reduce the likelihood that STR haplotypes carried by NLMM emigrants would become fixed in the population rather than being lost to drift.

Y-chromosome haplogroups show differential religious distributions in Lebanon [14], with L1b being notably found in Maronites. The NLMM UEP2 cluster offers no evidence that its members participated in the expansion of the Maronite faith into regions outside of the Northern Lebanese mountains since its haplotypes are not found outside of the mountains, though as noted above, given the relatively small NLMM population, the emigrant lineages from NLMM impacting the spread of the Maronite Church could have failed to fix in the population due to drift. Meanwhile, the other Lebanese UEP3 and UEP4 groups are largely Maronite with a few Muslim members, though they appear not to represent NLMM rooted lineages, and split from UEP2 much earlier than the establishment of the Maronite Church around the 6th Century CE. The haplogroup affinities and expansion associations further emphasize the isolation of the NLMM UEP2 group, which reflects the general very long-term resistance to successful emigration of member lineages of that population in general. While the two Sicilian individuals situated among the Lebanese UEPs may have also been the descendants of Phoenician sailors, there is another more plausible explanation. In 1584 the Maronite College of Rome was founded to form a cultural bridge between the Christians of the Levant and Rome [40]. It happened also that this College was established by a Maronite Patriarch from the mountains of North Lebanon. The College remained active for 228 years and was instrumental in instituting similar cultural practices on both sides of the Mediterranean that are still observed today. The College staff of clerical and lay people provided education to several hundreds of disciples from various age groups, not limited to Lebanon alone but also included Maronites from neighboring Syria and Cyprus. Most of these disciples would arrive in Italy from North Lebanon through Sicily. In fact, Sicily constituted a major hub of interaction between the Maronite community and Europe throughout most of the 16th and 17th Centuries. A substantial number of the names of the disciples of the Maronite College can be traced to the northern Lebanese mountains and it is not unlikely that the two Sicilian Levantine L1bs are present in Sicily as a result of the cultural exchange activities between Rome and the Christians of the Levant.

The results of the PCA (Fig. 2), ADMIXTURE (Fig. 3), and qpGraph suggest that North Lebanese Mountain Maronites carry a distinctive ancestral Fertile Crescent genetic component far above all other populations in the region. However, qpAdm explains its composition with much stronger statistical agreement with its model than qpGraph offered, but restricts these to expansions listed in the “right-

hand” population [38] list, marking admixture represented among those expansions. The qpGraph, PCA, and ADMIXTURE results all reflect the ancient population component preserved in the LNMM, while qpAdm appears to pick up more recent admixture events. Given our selection of populations, local populations that ADMIXTURE picked up will tend to be projected out by the F4 scores. However, the qpAdm admixture model results are also consistent with the L1b results identifying an admixture event into the North Lebanese Mountain populations that originated in more northern populations. This was a much more recent event than the early basal expansion suggested by qpGraph.

In general, all populations within Lebanon present similar pictures of homozygosity, with NLMM’s modal homozygosity scores being largest among groups, yet the ranges of all groups tend to overlap. Muslim populations appear to present some autozygosity. This suggests that all Lebanese populations are isolated, though with distinctive features specific to each.

The PCA also showed affinity between the “Other Christian Lebanese” (primarily Orthodox Catholic community members long present in Beirut) and Cyprus. The remaining Shiia and Sunni communities showed population similarities with other regions typical of what was observed in prior Y-chromosome studies of Lebanon [14, 16].

Conclusion

Our results indicate that the Levantine L1b group split from the Caucasus around 7300 years ago and those groups migrated to the Levant, with different branches inhabiting coastal and inland hills areas. The L1b immigration marks one of possibly several layers of immigrations into an otherwise genetically highly isolated population. Migration from the Caucasus was likely triggered by the amelioration of the climate and induced migratory events by groups in search of better habitats throughout the Northern Levant. The L1b haplogroup reached its highest frequency (12.5%) among the NLMM. The L1b haplotype STR clusters are suggestive of cultural barriers marked by the limited transmission of L1b NLMM lineages outside the Northern Lebanese mountains. These results highlight the value of investigating uniparental haplogroups and STR haplotypes to elucidate historical events among these populations.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

References

- Dreyer G. Umm el-Qaab 1. Das prädynastische Königsgrab U-j und seine frühen Schriftzeugnisse. Mainz am Rhein: Philipp von Zabern; 1997.
- Planhol XD. Les Fondements géographiques de l'histoire de l'Islam. Paris: Flammarion; 1968.
- Clay AT. The empire of the Amorites. Miami: HardPress Publishing; 2013.
- Josephus F, Henten JWV. Judean antiquities 15: translation and commentary. *Rev Hist Philos Relig.* 2015;95–3:373–4.
- Strabon, Jones HL. The geography of Strabo. 7, 7. London; Cambridge, Mass.: W. Heinemann; Harvard University press; 1966.
- Planhol XD. Les nations du prophète: manuel géographique de politique. Paris: Fayard; 1993.
- Khalil W. Routes et fortifications dans le Chouf libanais *Dossiers d'Archéologie* 22. 2012;22:70–5.
- Aliquot J. La vie religieuse au Liban sous l'Empire romain. Beyrouth: Institut Français du Proche-Orient; 2009.
- Obeid AI. The Druze & their faith in Tawhid. Syracuse: Syracuse University Press; 2006.
- Hitti PK. Lebanon in history: from the earliest times to the present. London; MacMillan; New York, St. Martin's; 1967.
- Moosa M. The Maronites in history. Piscataway, NJ: Gorgias Press; 2005.
- Henley ADM. Politics of a church at war: Maronite Catholicism in the Lebanese Civil War. *Mediterranean Politics.* 2008;13:353–69.
- Haber M, Platt DE, Badro DA, Xue Y, El-Sibai M, Bonab MA, et al. Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. *Eur J Hum Genet.* 2011;19:334–40.
- Zalloua PA, Platt DE, El Sibai M, Khalife J, Makhoul N, Haber M, et al. Identifying genetic traces of historical expansions: phoenician footprints in the Mediterranean. *Am J Hum Genet.* 2008;83:633–42.
- Badro DA, Bouchra D, Haber M, Youhanna SC, Salloum A, Ghassibe-Sabbagh M, et al. Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern middle eastern populations with European and African Populations. *PLOS ONE.* 2013;8:e54616.
- El-Sibai M, Platt DE, Haber M, Xue Y, Youhanna SC, Wells RS, et al. Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast. *Ann Hum Genet.* 2009;73:568–81.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 2008;18:830–8.
- Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, et al. Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet.* 2004;114:127–48.
- Hovhannisyann A, Khachatryan Z, Haber M, Hrechdikian P, Karafet T, Zalloua P, et al. Different waves and directions of Neolithic migrations in the Armenian Highland. *Investig Genet.* 2014;5:15.
- Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyann N, Khudoyan A, Burbage-Hall O, et al. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethnographic group. *Hum Genet.* 2001;109:659–74.
- Alshamali F, Pereira L, Budowle B, Poloni ES, Currat M. Local population structure in Arabian Peninsula revealed by Y-STR diversity. *Hum Hered.* 2009;68:45–54.
- Triki-Fendri S, Sanchez-Diz P, Rey-Gonzalez D, Alfadhli S, Ayadi I, Ben Marzoug R, et al. Genetic structure of the Kuwaiti population revealed by paternal lineages. *Am J Hum Biol.* 2016;28:203–12.
- Triki-Fendri S, Sanchez-Diz P, Rey-Gonzalez D, Ayadi I, Carra-cedo A, Rebai A. Paternal lineages in Libya inferred from Y-chromosome haplogroups. *Am J Phys Anthropol.* 2015;157: 242–51.
- Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature.* 2016;536:419–24.
- Swapan M, Heng L, Mark L, Iain M, Melissa G, Fernando R, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature.* 2016;538:201–6.
- Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature.* 2015;526:68–74.
- Boattini A, Martinez-Cruz B, Sarno S, Harmant C, Useli A, Sanz P, et al. Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS ONE.* 2013;8: e65441.
- Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999;16: 37–48.
- Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10:564–7.
- R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
- Wilson IJ, Weale ME, Balding DJ. Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *J R Stat Soc.* 2003;166:155–88.
- Burgarella C, Navascues M. Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur J Hum Genet.* 2011;19:70–5.
- Shi W, Ayub Q, Vermeulen M, Shao RG, Zuniga S, van der Gaag K, et al. A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol Biol Evol.* 2010;27:385–93.
- Wei W, Ayub Q, Xue Y, Tyler-Smith C. A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping. *Forensic Sci Int Genet.* 2013;7:568–72.
- Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, et al. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet.* 2004;74:50–61.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–9.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19: 1655–64.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. *Genetics.* 2012;192: 1065–93.
- Alley RB, Mayewski PA, Sowers T, Stuiver M, Taylor KC, Clark PU. Holocene climatic instability: a prominent, widespread event 8200 yr ago. *Geology.* 1997;25:483–6.
- Khairallah S. The Maronite College in Rome. *Theological Rev.* 1998;19:105–24.
- Balanovsky O, Dibirova K, Dybo A, Mudrak O, Frolova S, Pocheshkhova E, et al. Parallel evolution of genes and languages in the Caucasus region. *Mol Biol Evol.* 2011;28:2905–20.