

Insight into the genomic history of the Near East from whole-genome sequences and genotypes of Yemenis

Marc Haber^{1,2*}, Riyadh Saif-Ali³, Molham Al-Habori³, Yuan Chen¹, Daniel E. Platt⁴, Chris Tyler-Smith¹ and Yali Xue^{1*}

*Correspondence: ylx@sanger.ac.uk (YX) or mh25@sanger.ac.uk (MH)

¹The Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambs. CB10 1SA, United Kingdom.

²Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, B15 2TT, United Kingdom

³Department of Biochemistry and Molecular Biology, Faculty of Medicine and Health Sciences, Sana'a University, Sana'a 19065, Yemen.

⁴Computational Biology Center, IBM TJ Watson Research Centre, Yorktown Hgts, NY, USA.

Abstract

We report high-coverage whole-genome sequencing data from 46 Yemeni individuals as well as genome-wide genotyping data from 169 Yemenis from diverse locations. We use this dataset to define the genetic diversity in Yemen and how it relates to people elsewhere in the Near East. Yemen is a vast region with substantial cultural and geographic diversity, but we found little genetic structure correlating with geography among the Yemenis – probably reflecting continuous movement of people between the regions. African ancestry from admixture in the past 800 years is widespread in Yemen and is the main contributor to the country’s limited genetic structure, with some individuals in Hudayda and Hadramout having up to 20% of their genetic ancestry from Africa. In contrast, individuals from Maarib appear to have been genetically isolated from the African gene flow and thus have genomes likely to reflect Yemen’s ancestry before the admixture. This ancestry was comparable to the ancestry present during the Bronze Age in the distant Northern regions of the Near East. After the Bronze Age, the South and North of the Near East therefore followed different genetic trajectories: in the North the Levantines admixed with a Eurasian population carrying steppe ancestry whose impact never reached as far south as the Yemen, where people instead admixed with Africans leading to the genetic structure observed in the Near East today.

Text

Yemen's location and history suggest it has played an important role in early human migrations and thus is of considerable importance for understanding modern human genetic diversity.¹ But Yemen as well as the rest of the Near East have so far been poorly represented in databases that catalogue human variation: The 1000 Genomes Project did not include any Near Eastern samples among its many populations from around the globe,² and the Human Genome Diversity Project included only samples from the Northern region of the Near East.³ Here, we continue our efforts^{4;5} to catalogue the diversity of genetically under-represented populations by genotyping and whole-genome sequencing individuals from Yemen, and provide the new data as a resource for future genetic studies in the region. We present an initial exploration of the data by asking how present-day Near Easterners are related to each other and to the ancient people who inhabited the Near East before them, focusing on the historical events and processes that have contributed to the genetic structure in Yemen today.

Samples were recruited from diverse locations in Yemen (Figure 1, Table 1) and DNA extraction from 3 ml blood performed in Yemen at Sanaa University. All subjects provided an informed consent form and the study was approved by the IRB of the Lebanese American University. Samples (n=169) were genotyped at the Wellcome Sanger Institute in the UK using the Illumina HumanOmni2.5-8 BeadChip which covers 2.5 million SNPs. In addition, 46 samples were whole-genome sequenced at >30× depth using the Illumina HiSeq X Ten and the reads were mapped to build 37 and processed as described previously.⁶ The use of the samples in genetic studies was approved by the Human Materials and Data Management Committee at the Wellcome Sanger Institute (HMDMC approval number 14/072).

We merged the newly generated data with previously published data from modern and ancient populations creating two datasets for the analyses: 1) The *geno set* included the new Yemen genotype data merged with the Human Origin dataset⁷⁻⁹ and with published ancient DNA data^{4; 7; 8; 10-13} resulting in a set of 3060 samples and 211,966 SNPs. 2) The *seq set* included the new Yemen sequencing data merged with data from the Simons Genome Diversity Project¹⁴ and with sequencing data from Lebanon⁴ in addition to published ancient DNA data,^{4; 7; 8; 10-13} resulting in a set of 731 individuals and 938,848 SNPs. In this work we refer to Yemen as being the South of the Near East, while Lebanon and the surrounding Levant regions are considered the North of the Near East. This differentiation is based on geography but also captures our previous knowledge of the genetic structure in the region.⁵

The newly sequenced data contained 1,212,228 SNPs not previously reported in dbSNP build 151. Among the novel SNPs we found 46,970 SNPs that were common (>5%) in the Yemeni population; annotation on build 37 using VEP v79 as described previously¹⁵ showed these included 336 moderately

common missense SNPs, 23 splice donor/acceptor SNPs and 10 stop gained SNPs at >2% frequency (Table 2) which could be relevant to future medical studies in the region.

Investigation of the demographic history of the Yemeni populations by applying multiple sequentially Markovian coalescent (MSMC)¹⁶ analyses to the Yemen genomes showed a typical Eurasian population size change pattern characterized by a bottleneck around 50,000-60,000 years, followed by a population size expansion (Figure 2). But the Yemen population size increase after the bottleneck appears to be limited compared with the increase which occurred in the North of the Near East (Figure 3). To understand how the Yemenis were related to other populations in the Near East, we projected them and the ancient individuals in the *geno* set onto the first two principal components computed from the variations in present-day West Eurasians and East Africans. The PCA (Figure 4 and Figure 5) revealed genetic structure between the North and South of the Near East, with most of the ancient Neolithic and Bronze Age Levantines clustering at intermediate positions between the two regions. We observe on the PCA two genetic clines that affect the Near Easterners differently: 1) The African cline, which appears to be a major contributor to genetic diversity in Yemen and 2) The Eurasian cline, towards which the Northern Near Easterners appear shifted compared with the Yemenis. The Northern Near Easterners are themselves structured on the African cline with Palestinians, Jordanians, Syrians, and Lebanese Muslims having more African ancestry than Assyrians, Jews, Druze and Lebanese Christians (Figure 4) confirming our previous observation.⁵ We investigated the time when admixture has occurred in North and South of the Near East by using linkage disequilibrium,^{17; 18} using the Lebanese Muslims and the Yemenis to represent the two areas, and setting Africans, East Asians and West Eurasians in the *geno* set as references (Figure 6). We found two significant admixture events in the Lebanese Muslims; the first occurred around 600BCE-500CE ($Z=3.5$) and the second occurred around 1580CE-1750CE ($Z=3.4$), confirming our previous results on the date of admixture in North of the Near East.¹⁹ In contrast, we detect one significant admixture event in Yemen occurring 1190CE-1290CE ($Z=14.6$) and thus these results suggest that the shifts of the North and South of the Near East along the African cline could arise from independent events.

Next, we investigated genetic structure within Yemen and found it was loosely correlated with geography (Figure 7): some clustering by region is visible, but many individuals from the same region do not cluster together. We interpret this structure as resulting from continuous movement of people between the different regions. However, we note that individuals from Hudayda and from Hadramout have increased African ancestry compared with most other Yemenis, while individuals from Maarib appear to be the least admixed clustering at the base of the African cline in Yemen. We confirm these results by running ADMIXTURE²⁰ on the Yemenis, Lebanese, Africans and East Asians found in the *seq* set using $K=3$ to account for the African, Asian and Near Eastern ancestry. The ADMIXTURE results

(Figure 8) confirm the observation from the PCA showing that most Yemenis have some African ancestry which is maximized in Hudayda and Hadramout, with the later having in addition some East Asian ancestry, while the Maarib individuals appear to be the least admixed of the populations tested in Yemen. The African and Asian ancestry in Yemen can also be observed in the present Y-chromosome and mtDNA lineages (Figure 9 and Table 3) which are enriched with Y-haplogroups A, E, and R1a and mtDNA haplogroups L and M, respectively.

The African ancestry in Yemen increases the genetic distance between the Yemenis and all other Eurasians and therefore prevents us from understanding how the Yemen's Near Eastern ancestral component is related to other people in the region. But the relatively isolated Maarib individuals provide an opportunity to test Yemeni without the impact of the African ancestry. For example, using the four-population admixture test in ADMIXTOOLS,⁹ we find $f_4(\text{Lebanese, Yemen; Ancient Eurasian, Chimp})$ is always positive with all ancient Eurasians being closer to Lebanese than to Yemenis (Figure 10A). But these results are biased by Yemen's African ancestry since $f_4(\text{Lebanese, Maarib; Ancient Eurasian, Chimp})$ shows Lebanese and Maarib are equally distant from the Neolithic ($Z= 0.57$) and Bronze Age ($Z= 1.677$) Levantine populations (Figure 10B). However, we detect Eurasian gene flow to the North but not to the South of the Near East, reflected by $f_4(\text{Lebanese, Maarib; Steppe_EMBA, Chimp})$ being significantly positive ($Z=10.837$). We also note that Maarib has increased Natufian (Epipaleolithic hunter-gatherers) ancestry compared with the Lebanese, with $f_4(\text{Lebanese, Maarib; Natufian, Chimp})$ being negative ($Z= -2.95$) (Figure 10B). We propose that the increase in Natufian ancestry in Maarib is likely from structure in the ancestral population. We confirm the f_4 -statistics results by modelling the modern and ancient Near Easterners using qpGraph from the ADMIXTOOLS package and show that our data support a model where the North and South of the Near East split from a population related to the Bronze Age population which inhabited the North of the Near East, but with the South regions having an excess of Epipaleolithic hunter-gatherer ancestry (Figure 11). After the split, the Northern regions of the Near East received gene flow from a Eurasian population^{4; 19} which never reached the South, where people instead admixed with Africans resulting in the genetic structure we observe in the Near East today (Figure 11).

We thus find that the genetic structure of present-day Yemenis can be accounted for by mixture of post-glacial Levantine and other Eurasian populations with Africans within the last 1000 years. In particular, we do not find evidence for African ancestry that might have remained specifically in this region since the major expansion out of Africa 50,000-60,000 years ago.¹ Our model of the genetic relationships in the Near East relies on the available genetic data today and thus will likely be refined in the future when, most importantly, ancient DNA data from the Southern regions of the Near East

become available – work that has so far been hindered by the warm climate in the Near East negatively impacting the survival of ancient DNA.

Methods

Array genotype data were processed using PLINK v1.90b5²¹ as described previously.²² Genotypes from sequence data were called using bcftools v1.6 with the command ‘bcftools mpileup -C50 -q30 -Q30 | bcftools call’ and phased with SHAPEIT v2.r900²³ using the 1000 Genomes Project phase 3 haplotypes² as a reference panel. Genotypes were merged with data from the literature using the program mergeit from the EIGENSOFT v7.2.1 package.²⁴ Principal components were computed using smatpca v16000 also from the EIGENSOFT package with the following parameters; numoutlieriter: 0, lsqproject: YES, autoshrink: YES. Effective population size and rates of gene flow were inferred using MSMC v2.1.1¹⁶ with one (Figure 2) or two (Figure 3) high-coverage phased genomes from each population and using the following time segments -p 1*2+25*1+1*2+1*3, a generation time of 30 years and a rate of 1.25×10^{-8} mutations per nucleotide per generation. Y-chromosome haplogroups were assigned using yhaplo²⁵ and phylogenetic analysis of the whole Y-chromosome sequences was carried out using RAxML v8.2.10²⁶ with the arguments “-m ASC_GTRGAMMA” and “--ascorr=stamatakis” as described previously⁶ and plotted using iTOL v4.²⁷ mtDNA haplogroups were determined using the mtDNA-server v1.0.7.²⁸ Population mixture signals were tested using qpDstat v751 (f4mode: YES) and qpGraph v6412 (useallsnps: YES, blgsize: 0.05, forcezmode: YES, diag: .0001, bigiter: 6, hires: YES, lambdascale: 1, inbreed: YES) from the ADMIXTOOLS package.⁹ Admixture proportions were additionally estimated using ADMIXTURE v1.3.0.²⁹ MALDER v1.0^{17; 18} (mindis: 0.005, binsize: 0.0005) was used to date the time of admixture using the Africans, East Asians and West Eurasians with n>20 in the *geno* set as references.

Acknowledgements

Our work was supported by Wellcome grant number 098051 (MH, YC, YX, CTS). We thank Pierre Zalloua and Matthew Hurles for their contributions to this work.

Data availability

Sequence data are available via EGAS00001002083, and genotype data via EGAS00001001231.

References

1. Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., and Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature* 541, 302-310.
2. The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
3. Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., et al. (2002). A human genome diversity cell line panel. *Science* 296, 261-262.
4. Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S., Martiniano, R., Prado-Martinez, J., Szpak, M., et al. (2017). Continuity and Admixture in the Last Five Millennia of Levantine History from Ancient Canaanite and Present-Day Lebanese Genome Sequences. *Am J Hum Genet* 101, 274-282.
5. 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.
6. Haber, M., Jones, A.L., Connell, B.A., Asan, Arciero, E., Yang, H., Thomas, M.G., Xue, Y., and Tyler-Smith, C. (2019). A Rare Deep-Rooting D0 African Y-Chromosomal Haplogroup and Its Implications for the Expansion of Modern Humans out of Africa. *Genetics*.
7. 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.
8. 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.
9. 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.
10. Broushaki, F., Thomas, M.G., Link, V., Lopez, S., van Dorp, L., Kirsanow, K., Hofmanova, Z., Diekmann, Y., Cassidy, L.M., Diez-Del-Molino, D., et al. (2016). Early Neolithic genomes from the eastern Fertile Crescent. *Science* 353, 499-503.
11. 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.
12. 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. *Nature communications* 6, 8912.
13. 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.
14. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538, 201-206.
15. Xue, Y., Mezzavilla, M., Haber, M., McCarthy, S., Chen, Y., Narasimhan, V., Gilly, A., Ayub, Q., Colonna, V., Southam, L., et al. (2017). Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. *Nature communications* 8, 15927.
16. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat Genet* 46, 919-925.
17. 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.

18. Pickrell, J.K., Patterson, N., Loh, P.R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proc Natl Acad Sci U S A* 111, 2632-2637.
19. Haber, M., Doumet-Serhal, C., Scheib, C.L., Xue, Y., Mikulski, R., Martiniano, R., Fischer-Genz, B., Schutkowski, H., Kivisild, T., and Tyler-Smith, C. (2019). A Transient Pulse of Genetic Admixture from the Crusaders in the Near East Identified from Ancient Genome Sequences. *Am J Hum Genet* 104, 977-984.
20. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655-1664.
21. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-575.
22. Haber, M., Mezzavilla, M., Xue, Y., Comas, D., Gasparini, P., Zalloua, P., and Tyler-Smith, C. (2016). Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. *Eur J Hum Genet* 24, 931-936.
23. Delaneau, O., Zagury, J.F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 10, 5-6.
24. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet* 2, e190.
25. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv*.
26. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.
27. 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, W242-245.
28. 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.
29. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655-1664.
30. 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.

Figures and tables

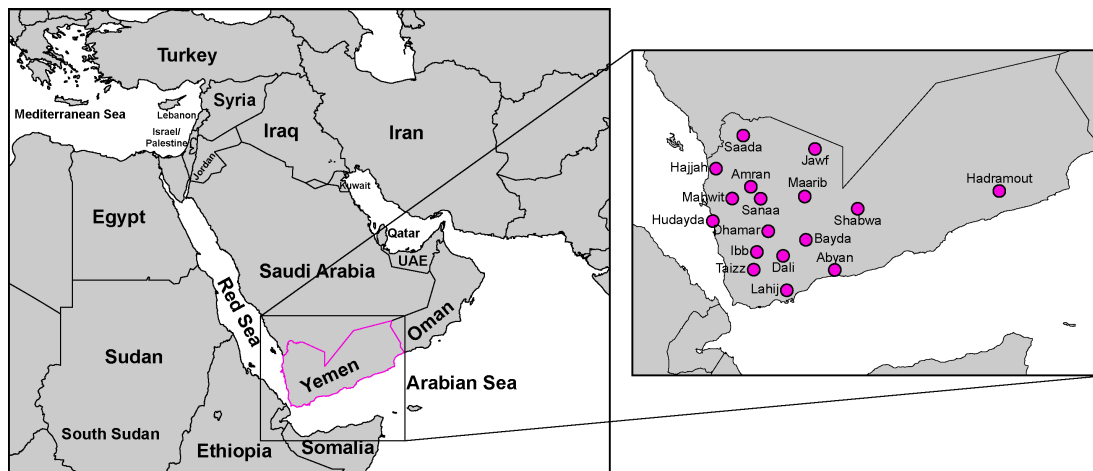


Figure 1. Map of the Near East with inset showing the location of the samples studied in this work.

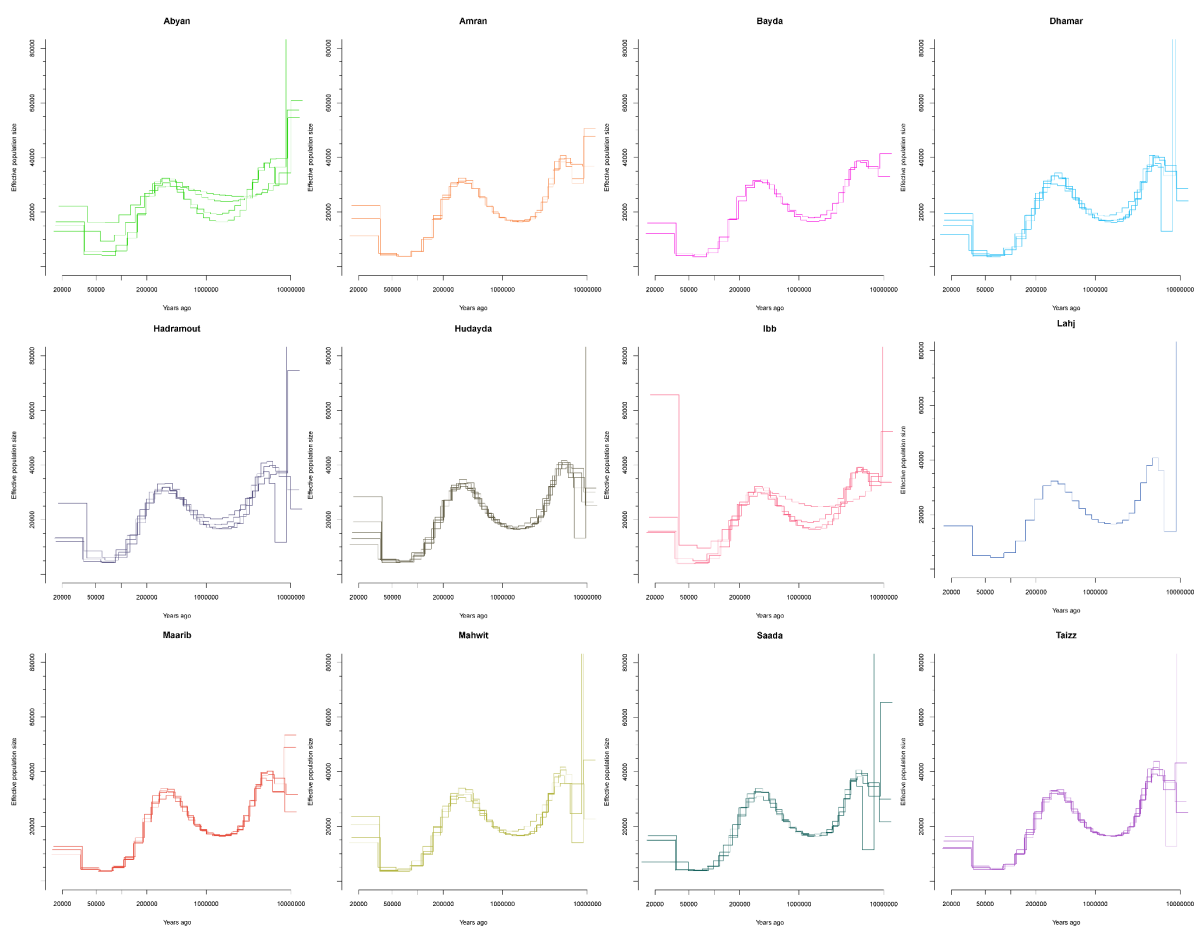


Figure 2. MSMC2 results showing effective population size change in Yemen over time. Most Yemeni genomes show a typical Eurasian population size change pattern characterized by a bottleneck around 50,000-60,000 years followed by a population size expansion.

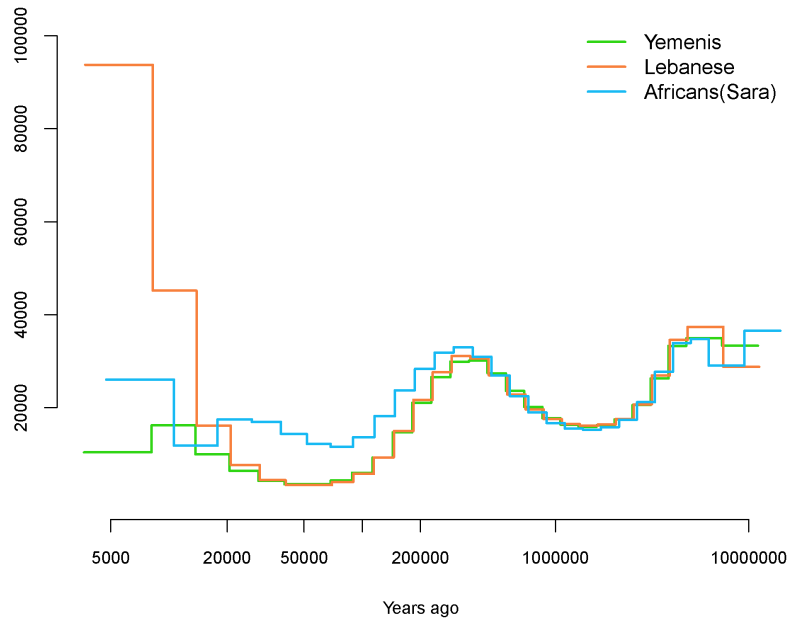


Figure 3. MSMC2 results showing population size change in Africans,³⁰ Lebanese,⁴ and Yemenis using two genomes from each population.

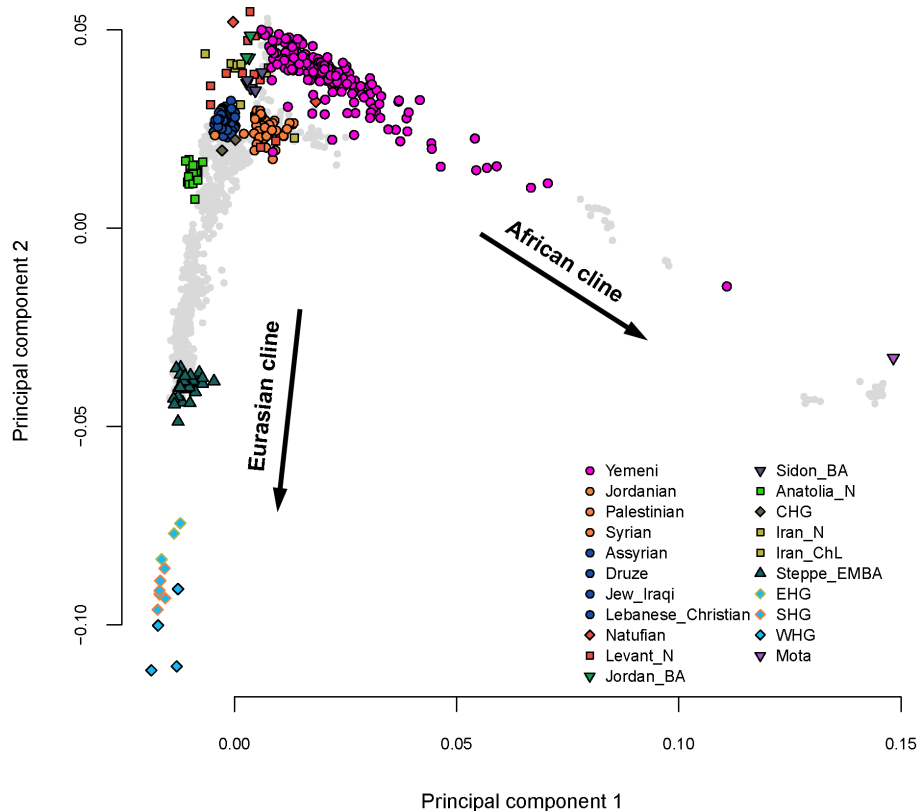


Figure 4. PCA with the Yemenis (pink circles) and the ancient samples projected into the plot using the variation from East Africans and West Eurasians (grey circles) including the Northern Near East populations (blue and orange circles).

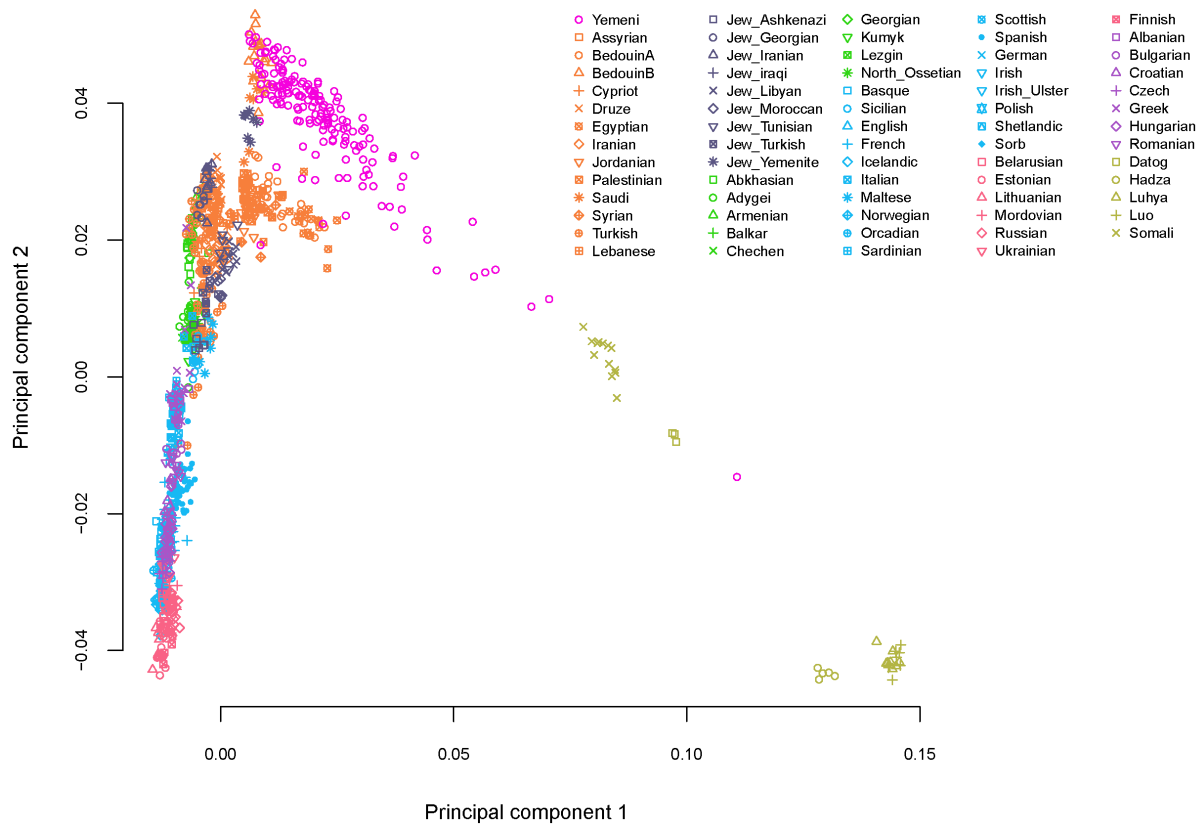


Figure 5. PCA as in Figure 4 but showing only the present-day populations.

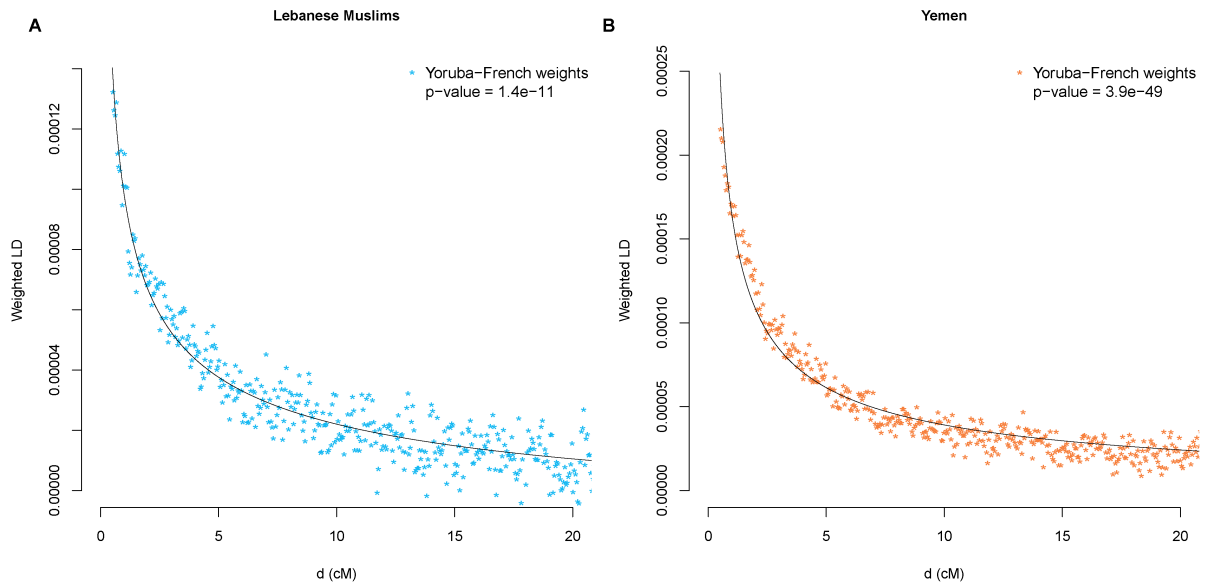


Figure 6. Weighted LD curves using Yoruba and French as reference populations for (A) Lebanese Muslims and (B) Yemenis.

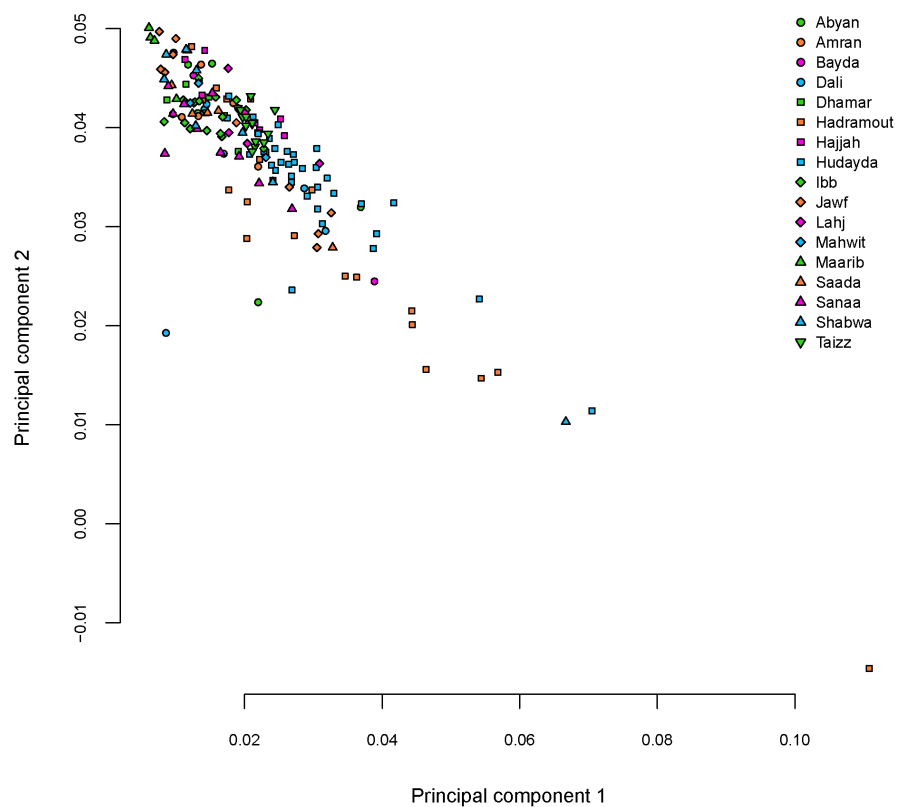


Figure 7. PCA coordinates from Figure 4 showing only samples from Yemen and their region of origin.

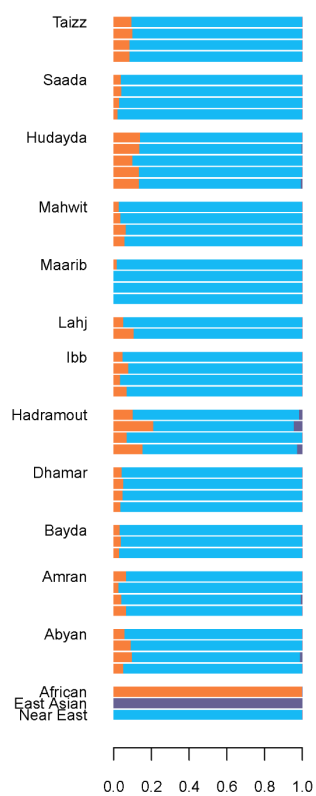


Figure 8. ADMIXTURE plot using the *geno* set and showing the Yemenis at K=3.

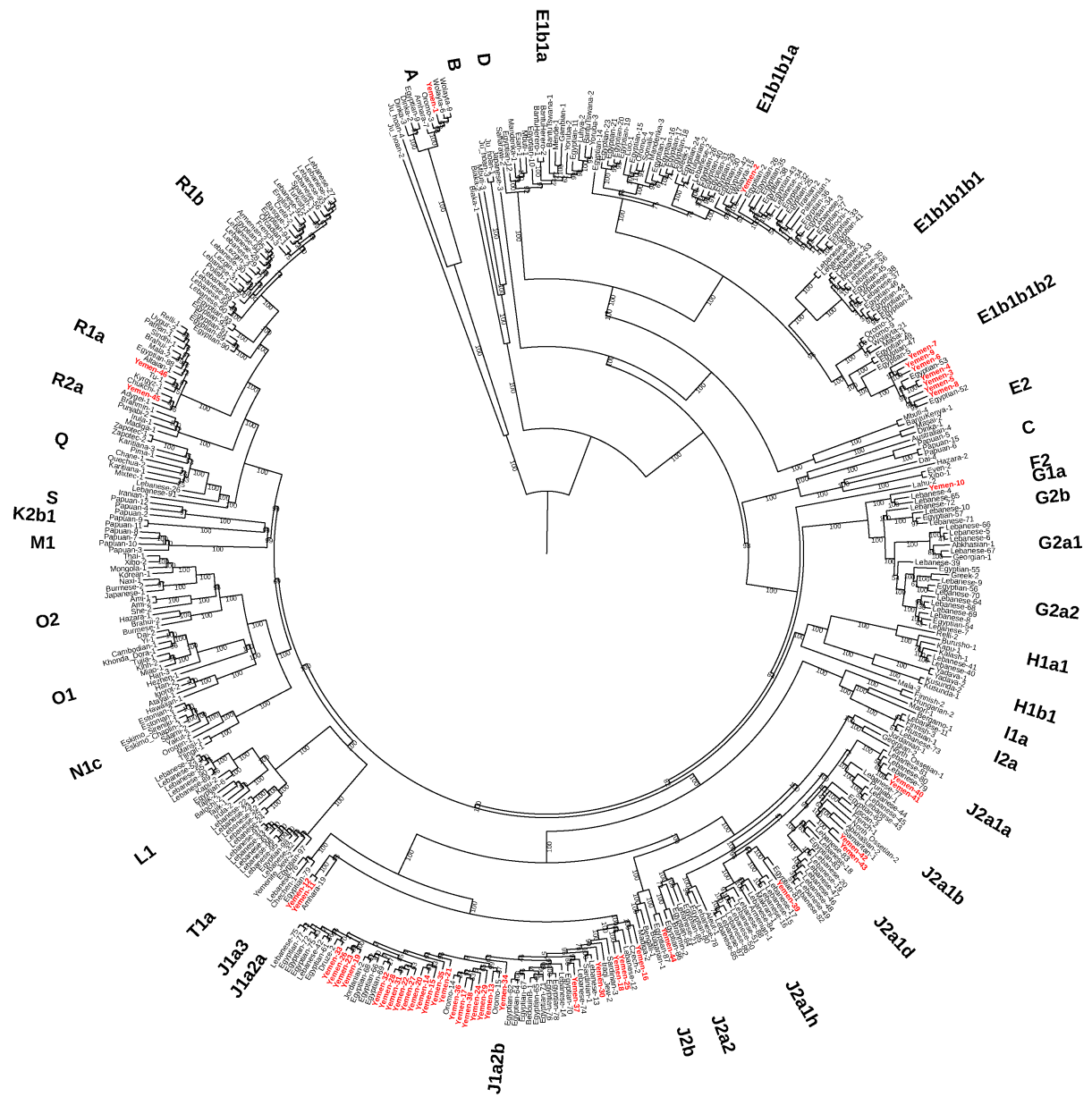


Figure 9. A maximum-likelihood tree of Y-Chromosome sequences from Yemen (red) and worldwide populations. Numbers on branches show bootstrap values from 1000 replicates.

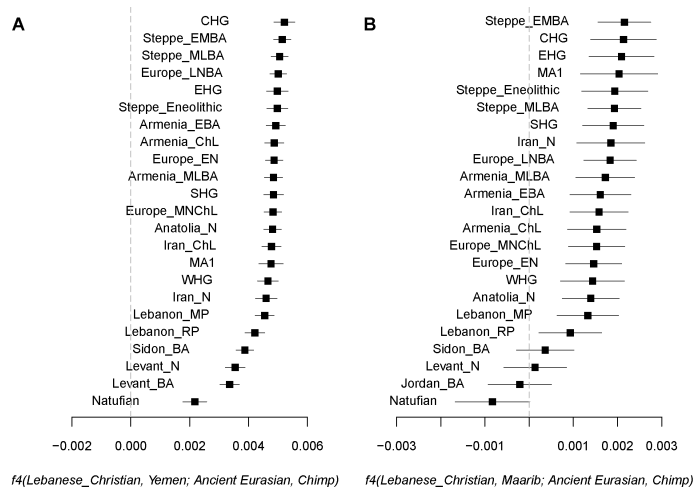


Figure 10. (A) The affinity of Yemen to ancient populations is masked by their African ancestry (B) but the less admixed Maarib population reveals the Near Eastern component in Yemen. We plot the f_4 statistic value and ± 3 standard errors.

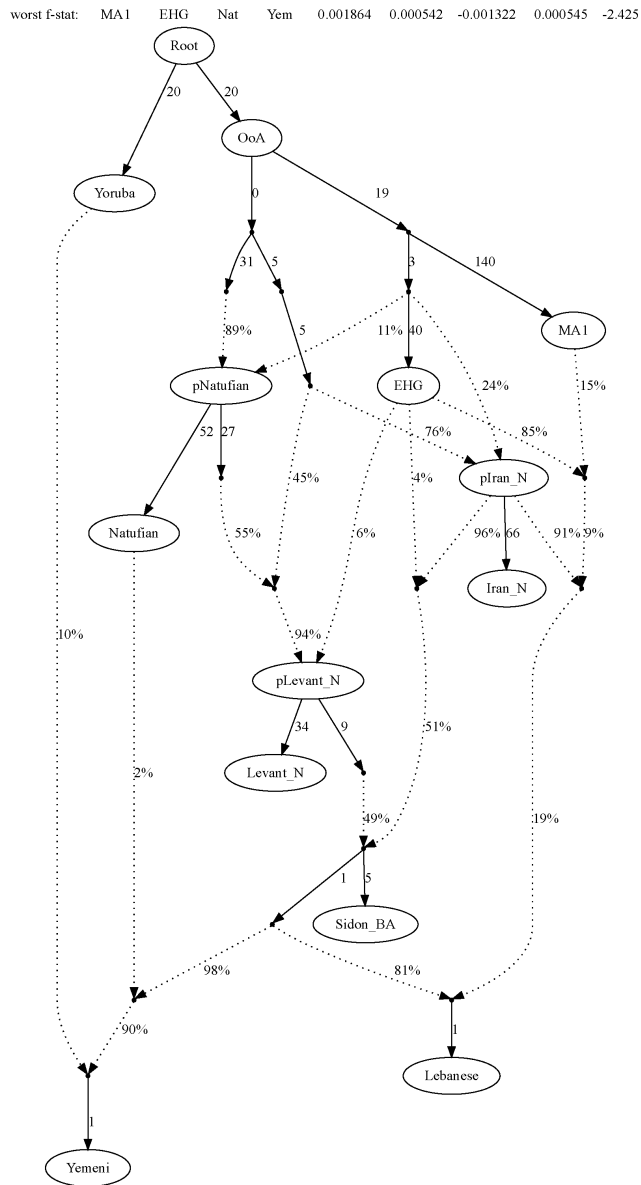


Figure 11. An admixture graph model showing the relationship of present-day Northern (Lebanese) and Southern (Yemeni) Near Easterners.

Table 1. Newly genotyped and sequenced individuals and their location in Yemen.

Location in Yemen	Genotyped Individuals	Sequenced Individuals	Latitude	Longitude
Abyan	6	4	13.568675	46.132999
Amran	6	4	15.66155	43.946018
Bayda	3	3	14.326473	45.380295
Dali	8	0	13.922642	44.786675
Dhamar	6	4	14.542997	44.400061
Hadramout	20	4	15.55487	50.418282
Hajjah	8	0	16.117894	43.03823
Hudayda	35	5	14.797868	42.954487
Ibb	17	4	14.02017	44.103111
Jawf	10	0	16.613066	45.616588
Lahij	5	2	13.05581	44.88286
Mahwit	3	4	15.365839	43.460814
Maarib	4	4	15.416376	45.35
Saada	5	4	16.950941	43.747772
Sanaa	10	0	15.36198	44.201962
Shabwa	9	0	15.108985	46.72996
Taizz	14	4	13.575224	44.021527
Total	169	46		

Table 2. Novel variants discovered in Yemen; their type and frequency.*

Type	>0--<=2%	>2--<=5%	>5--<=50%
3_prime_UTR_variant	5643	1488	194
5_prime_UTR_variant	1142	161	33
TF_binding_site_variant	242	93	24
intergenic_variant	296000	164117	19253
intron_variant	433468	220645	25712
mature_miRNA_variant	12	6	1
missense_variant	2352	259	77
non_coding_transcript_exon_variant	4855	1553	248
regulatory_region_variant	22554	8502	1358
splice_acceptor_variant	50	13	1
splice_donor_variant	46	7	2
splice_region_variant	519	81	26
start_lost	6	2	0
stop_gained	69	9	1
stop_lost	3	0	0
synonymous_variant	1196	165	40

*Annotations are from VEP v79 using build 37.

Table 3. Y-chromosome and mtDNA haplogroups in Yemen

ID	Y tree label	Region	Y haplogroup	MT Haplogroup
EGAN00001456674	Yemen-31	Abyan	J-P58	K1a27
EGAN00001456628	Yemen-45	Abyan	R-Z92	L0a2c
EGAN00001456614	Yemen-20	Abyan	J-P58	M30+16234
EGAN00001456688	Yemen-42	Abyan	J-M67	U8b1a1
EGAN00001456593	Yemen-10	Amran	G-F1761	J1b
EGAN00001456579	Yemen-16	Amran	J-P58	K1a7
EGAN00001456663	Yemen-39	Amran	J-L27	L3f1b+16292
EGAN00001456648	Yemen-6	Amran	E-CTS1096	M1a1f
EGAN00001456617	Yemen-22	Bayda	J-P58	H2a1
EGAN00001456603	Yemen-5	Bayda	E-CTS1096	L2a1j
EGAN00001456575	Yemen-15	Bayda	J-P58	M32c
EGAN00001456564	Yemen-4	Dhamar	E-CTS1096	J1b2
EGAN00001456684	Yemen-34	Dhamar	J-P58	M1a1b1b
EGAN00001456668	Yemen-30	Dhamar	J-P58	M1a1f
EGAN00001455435	Yemen-3	Dhamar	E-CTS1096	U6a+16189
EGAN00001456587	Yemen-46	Hadramout	R-Z2123	J2a2a1
EGAN00001456602	Yemen-18	Hadramout	J-P58	L2a1b1a
EGAN00001456613	Yemen-19	Hadramout	J-P58	M45a
EGAN00001456629	Yemen-25	Hadramout	J-P58	U6a2a1
EGAN00001456769	Yemen-43	Hudayda	J-M67	J1b2
EGAN00001456774	Yemen-12	Hudayda	J-CTS5368	J1d1a1
EGAN00001456561	Yemen-44	Hudayda	J-M205	L2a1
EGAN00001456745	Yemen-8	Hudayda	E-CTS1096	L2a1c+16129
EGAN00001456765	Yemen-11	Hudayda	J-CTS5368	L2a1c+16129
EGAN00001456568	Yemen-1	Ibb	A1b1-M118	K2c
EGAN00001456592	Yemen-17	Ibb	J-P58	L0a1a1
EGAN00001456560	Yemen-13	Ibb	J-P58	L0a1b1a1
EGAN00001456637	Yemen-27	Ibb	J-P58	U3b3
EGAN00001456657	Yemen-28	Lahij	J-P58	J1d1a
EGAN00001456627	Yemen-2	Lahij	E-L677	J1d1a1
EGAN00001456690	Yemen-35	Maarib	J-P58	J1b2
EGAN00001456634	Yemen-26	Maarib	J-P58	J1d1a1
EGAN00001456620	Yemen-23	Maarib	J-P58	J2a2b
EGAN00001456675	Yemen-32	Maarib	J-P58	U9a
EGAN00001456573	Yemen-14	Mahwit	J-P58	HV1b1
EGAN00001456650	Yemen-7	Mahwit	E-CTS1096	L2a1a2
EGAN00001456622	Yemen-24	Mahwit	J-P58	L5a1a
EGAN00001456680	Yemen-33	Mahwit	J-P58	T1a
EGAN00001456660	Yemen-29	Saada	J-P58	HV1a'b'c
EGAN00001456615	Yemen-21	Saada	J-P58	J1b1b1
EGAN00001456705	Yemen-41	Saada	J-M47	L2b1a3
EGAN00001456676	Yemen-40	Saada	J-M47	T1a2
EGAN00001456743	Yemen-38	Taizz	J-P58	H20
EGAN00001456732	Yemen-36	Taizz	J-P58	J2a2
EGAN00001456737	Yemen-37	Taizz	J-P58	L0a1a1
EGAN00001456748	Yemen-9	Taizz	E-CTS1096	R0a1a