Demographic history and genetic variation of the Armenian population

Authors

Anahit Hovhannisyan, Pierpaolo Maisano Delser, Anna Hakobyan, ..., Zaruhi Khachatryan, Levon Yepiskoposyan, Andrea Manica

Correspondence

hovhanna@tcd.ie (A.H.),
lepiskop@sci.am (L.Y.)

A whole-genome study of Armenians shows homogeneous genetic structure across the region, with a recent bottleneck explaining the distinctiveness of the Sasun population. Our findings reject the Balkan theory for the origin of Armenians, instead revealing genetic input from a Levantine source in the region after the Early Bronze Age.

Hovhannisyan et al., 2025, The American Journal of Human Genetics *112*, 11–27
January 2, 2025

© 2024 The Authors. Published by Elsevier Inc. on behalf of American Society of Human Genetics.
https://doi.org/10.1016/j.ajhg.2024.10.022

CellPress

Demographic history and genetic variation of the Armenian population

Anahit Hovhannisyan,[1,2,3,4,*] Pierpaolo Maisano Delser,[1,2] Anna Hakobyan,[3,5] Eppie R. Jones,[1,2] Joshua G. Schraiber,[6] Mariya Antonosyan,[3,7] Ashot Margaryan,[3,8] Zhe Xue,[2] Sungwon Jeon,[9,10] Jong Bhak,[9,10] Peter Hrechdakian,[11] Hovhannes Sahakyan,[3,4] Lehti Saag,[4] Zaruhi Khachatryan,[3] Levon Yepiskoposyan,[3,12,*] and Andrea Manica[2,12]

Summary

We introduce a sizable ($n = 34$) whole-genome dataset on Armenians, a population inhabiting the region in West Asia known as the Armenian highlands. Equipped with this genetic data, we conducted a whole-genome study of Armenians and deciphered their fine-scale population structure and complex demographic history. We demonstrated that the Armenian populations from western, central, and eastern parts of the highlands are relatively homogeneous. The Sasun, a population in the south that had been argued to have received a major genetic contribution from Assyrians, was instead shown to have derived its slightly divergent genetic profile from a bottleneck that occurred in the recent past. We also investigated the debated question on the genetic origin of Armenians and failed to find any significant support for historical suggestions by Herodotus of their Balkan-related ancestry. We checked the degree of continuity of modern Armenians with ancient inhabitants of the eastern Armenian highlands and detected a genetic input into the region from a source linked to Neolithic Levantine Farmers at some point after the Early Bronze Age. Additionally, we cataloged an abundance of new mutations unique to the population, including a missense mutation predicted to cause familial Mediterranean fever, an autoinflammatory disorder highly prevalent in Armenians. Thus, we highlight the importance of further genetic and medical studies of this population.

Introduction

Armenians are recognized as one of the ancient populations in Western Asia and have historically inhabited the area of the Armenian highlands[1] (Figure 1). Lying between Europe and Asia, the territory of the highlands has been a land bridge for major human migrations since its early settlement in the Upper Paleolithic.[2] Due to the proximity to the Fertile Crescent, the region emerged as one of the earliest centers to adopt agriculture in the Neolithic while also playing a pivotal role in disseminating technologies such as obsidian tools,[3] leather footwear,[4] and viticulture.[5] This period has been characterized by extensive genetic interactions between the Caucasus, northern Levant, Iran, and Anatolia over a long time span,[6,7] with a notable gene flow from the South-Caucasus-like source to Anatolia, starting after the end of the Neolithic.[7] Another well-documented post-Neolithic event was the onset of a population movement from the Caucasus into the steppe, which ultimately genetically contributed to the formation of Yamnaya ancestry.[7–9] The Bronze Age in the Armenian highlands is marked by the rise and fall of a number of archaeological cultures, including the Early Bronze Kura-Araxes culture, Middle Bronze Age Early Kurgan, Trialeti-Vanadzor, Sevan-Artsakh, Karmir-berd, and Karmir-vank ceramic traditions, and the Late Bronze and Early Iron Age Lchashen-Metsamor culture.[10] Approximately after the Early Bronze Age period, a migration from the steppe in a southeastern direction introduced Yamnaya-related ancestry to the region (though predominantly not reaching western and central parts of the Armenian highlands).[7] These aforementioned gene flows, backed by linguistic evidence of the formation of the proto-Armenian language during the second part of the Bronze Age (~4,000 years ago),[11] position the Armenian highlands as either near or a potential candidate for the proposed homeland of the Proto-Indo-European languages from where they subsequently radiated into western Europe, central Asia, and India.

The earliest reference to Armenia as a state is found in the Babylonian version of the Behistun inscription, dating back to 522 BC, at the start of the reign of Darius I the Great.[12,13] In the Old Persian version of the same inscription, the region was referred to as Urartu, an Iron Age kingdom that flourished in the Armenian highlands for approximately three centuries and was characterized by

[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin D02 PN40, Ireland; [2]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK; [3]Institute of Molecular Biology, National Academy of Sciences, Yerevan 0014, Armenia; [4]Estonian Biocentre, Institute of Genomics, University of Tartu, 51010 Tartu, Estonia; [5]Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, 1030 Vienna, Austria; [6]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90018, USA; [7]Max Planck Institute of Geoanthropology, 07745 Jena, Germany; [8]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark; [9]Korean Genomics Center (KOGIC), Ulsan National Institute of Science and Technology (UNIST), Uslan 44919, Republic of Korea; [10]Clinomics Inc., Ulsan 44919, Republic of Korea; [11]Armenian DNA Project at Family Tree DNA, Houston, TX 77008, USA
[12]Senior author
*Correspondence: hovhanna@tcd.ie (A.H.), lepiskop@sci.am (L.Y.)
https://doi.org/10.1016/j.ajhg.2024.10.022.
© 2024 The Authors. Published by Elsevier Inc. on behalf of American Society of Human Genetics.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

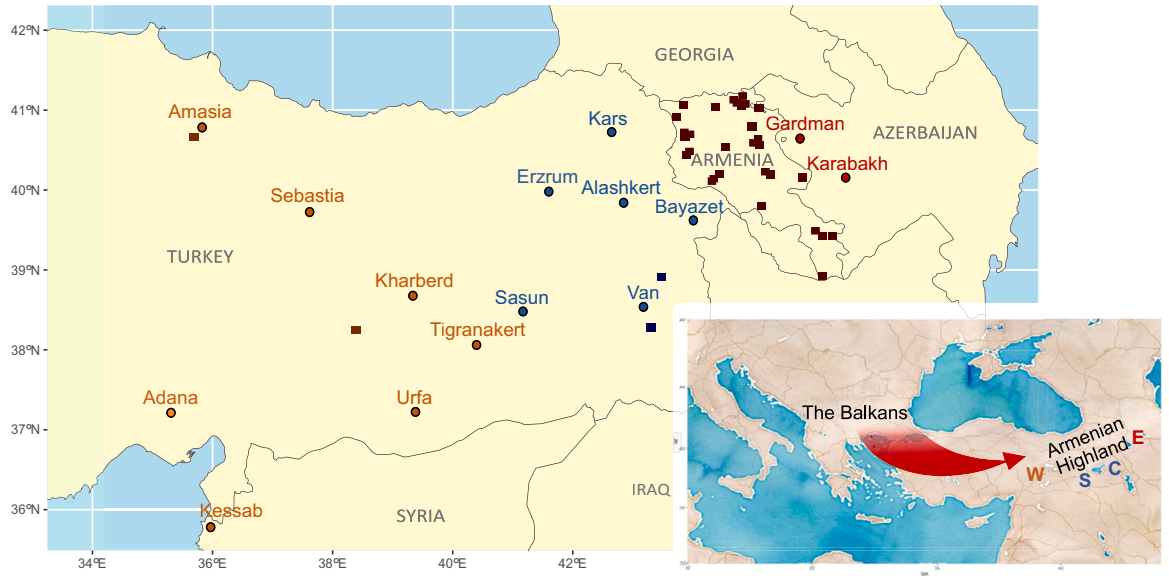


Figure 1. Maps showing Armenia with the locations of ancient and whole-genome modern individuals, along with a proposed route of Armenian migration according to the Balkan theory

The orange, blue, and red colors represent our categorization of generated whole-genome modern samples into western, central, and eastern Armenian groups, respectively (populations are denoted as W, C, and E on a smaller map). The sampling areas of ancient populations are shown in rectangles, with colors matching those of their geographically corresponding modern Armenians. Apart from geographic considerations, the division of modern Armenian groups is based on their linguistic, historical, and cultural peculiarities. While western and central regions of the Armenian highlands experienced long periods of Ottoman rule and subsequent dispersion of their populations, the eastern parts were held by the Persian and later the Russian Empire. After the displacement in the early 20th century from the western and central regions of the highlands, Armenians today are predominantly located within the territory of the Republic of Armenia, from where most of the samples in this study have been collected. As seen on the map, the Sasun population (denoted as S on a smaller map) is located within the central Armenian group. However, in some of our analyses, we treated it as a distinct group when aiming to investigate the question of its origin.

significant genetic heterogeneity.[7] Subsequent periods in the region have been marked by a series of invasions and mass migrations of the local population, which impacted various parts of the highlands differently. These included incursions by the Assyrian, Persian, and Byzantine Empires, Arabs, Seljuk Turks, Mongols, and the Ottoman Empire. Although in the past Armenians populated the regions reaching as far as the Mediterranean coast, they now primarily reside in the eastern part of the Armenian highlands, nestled in the southern Caucasus, where present-day Armenia is situated (Figure 1).

This complex historical background of the highlands makes a study of the demographic history of Armenians a matter of importance to shed light on peopling and regional migrations. At the same time, the mountainous landscape, a distinct and very old Armenian language, and a strong national and cultural identity, reinforced later by the adoption of Christianity, might have been conducive for the long-term genetic isolation of Armenians from the neighboring populations.[14–16] Indeed, a comparison of both ancient and modern mitochondrial genomes (mtDNA), across a time span of 8,000 years, revealed a remarkably high level of matrilineal genetic continuity in the region.[14] The picture of genetic isolation is further backed up by genome-wide autosomal studies of contemporary Armenians suggesting a lack of an external genetic influx at least since the Bronze Age.[15,16] Nonetheless, contemporary genomes have limited power to draw inferences on the genetic continuity through time, and a detailed study of the demographic and population histories of Armenians with the inclusion of both ancient and modern genomes has not yet been undertaken.

Another understudied and highly debated question concerns the origin of Armenians. Several theories and legends exist as to the formation of the population, albeit two of them prevail. According to the long-standing "Balkan theory" based on the ancient Greek historian Herodotus' writings, the ancestors of the Armenians were Phrygian colonists who migrated to the Armenian highlands from the Balkans.[17] The conclusion was derived mainly from the fact that Armenians were armed in the Phrygian fashion when they were part of the Persian army. Common ancestry for the Armenians and the Phrygians is further suggested by some linguists, who speculate that the proto-Armenian language belonged to a Thraco-Phrygian subgroup within the Indo-European language family.[18] A recent linguistic study has grouped the Armenian and Greek languages into a single deep branch, suggesting they diverged from the main European clade early in the Indo-European language tree.[11] However, an alternative view based on the popular legend for the origins of Armenians suggests the local formation of the population.[19] Despite extensive excavations in the area, to date no convincing archaeological evidence has been found to support either of these hypotheses.[20]

Within the question on the origin, the ethnogenesis of the Sasun, an Armenian population that inhabited the southern part of the Armenian highlands (modern-day southeastern Turkey) and speaks a distinct dialect within Armenian language,[21] stands apart as a particular controversy. According to ancient Armenian historian Movses Khorenatsi (5th century CE), the princely clans of Sasun were descendants of the sons of Sennacherib, an Assyrian king.[22] The historian's report was based on the Bible,[23] cuneiform sources, and local traditional stories. Intriguingly, the Y chromosome haplogroup composition of Sasun differs from those of other Armenian populations by a strikingly high prevalence of haplogroup T,[24] hinting at plausible population-specific demographic processes. Thus, a thorough genetic study on the relationships between modern Armenians and the ancient and modern samples from the Balkans as well as an internal substructure within Armenian populations is needed.

Finally, there is a well-known need to extend genetic studies to ethnically diverse populations, since the understanding of population-specific genetic variations has a crucial role in medical and biomedical research. The Armenians remain under-represented in current genomic databases, including the 1000 Genomes Project (1000 GP). The genetic structure of the population has been mostly scrutinized using microarray technology or uniparental genetic markers,[15,24–26] and only eight full genomes of Armenians have been reported so far (two sequenced with the Illumina platform and the other six with Complete Genomics).[27,28] Thus, a complete picture of genetic variation in Armenians remains largely unexplored.

Here, we introduce a dataset comprising 34 whole-genome sequences of Armenian individuals with all four grandparents (4GP) originating from the same region within the Armenian highlands. We divided the samples into three groups representing western (WA; $n = 11$), central (CA; $n = 17$, which includes the population of Sasun; $n = 5$), and eastern (EA; $n = 6$) regions within the Armenian highlands (Figure 1 and Table S1). We also generated genome-wide Illumina chip data of 23 Armenian individuals collected with a similar criterion for ancestry (WA = 2, CA = 7, and EA = 14). Throughout the text, we refer to these populations as modern western, central, and eastern Armenians. Armed with our modern dataset as well as densely sampled ancient genomes from the region[7,29–31] (Figure 1 and Table S2), we conducted a detailed analysis of the genetic substructure within Armenians, investigated their genetic origins and demography, and tested whether the population can indeed be considered as an isolate sheltered from the major migrations that shaped the rest of western Eurasia.[15,16] We further characterized a spectrum of genetic variants in the Armenian cohort. As part of our analyses, here we also addressed the question of the accuracy of the imputation process when the population-specific reference panel and the most updated 1000 GP haplotype reference panel for the GRCh38 assembly are included.

Material and methods

Sample collection, sequencing, and ethics

All individuals who donated samples for this project were interviewed about their origins and completed a questionnaire regarding the city of origin of their parents and grandparents. The samples were collected from the individuals both from the diaspora and the Republic of Armenia. We selected unrelated individuals whose four grandparents originated from the same region within the Armenian highlands and divided them geographically into the western ($n = 13$), central ($n = 24$), and eastern ($n = 20$) Armenian groups. Thus, in this study, we use the term "Armenian" to refer to individuals who ethnically and culturally identify themselves as part of this population and whose ancestry dates back at least four generations within a specific region of the Armenian highlands. We had no prior information about marriage practices in these communities. This study was approved by the Ethics Committee of the Institute of Molecular Biology, National Academy of Sciences of the Republic of Armenia (NAS RA) (IRB #00004079). All individuals were informed about the aim of this study and gave their consent to participate.

Blood samples were collected from our donors. DNA samples were extracted at the Institute of Molecular Biology NAS RA using a modification of the salting-out procedure.[32] DNA was normalized (~50 ng/μL), and 34 samples were sent for high coverage (~30×) paired-end whole-genome sequencing (WGS) on the Illumina HiSeq X Ten platform (Macrogen, Seoul, South Korea). The remaining samples were genotyped by Illumina 713K chip in the Estonian Biocentre, Tartu, Estonia. All the details on sample collection and sequencing are provided in Table S1.

Data processing for population genomics analyses with WGS samples

Adapter sequences were trimmed from the ends of reads using trimadap.[33] Alignment of the samples has been conducted following the pipeline described in Mallick et al.[27] In particular, sequences were aligned to the hs37d5 build of the human reference genome using Burrows-Wheeler Aligner (BWA) version 0.7.16,[34] and clonal reads were removed with samblaster.[35] Duplicate reads have been marked with Picard version 2.12.1.[36] Base calibration was conducted with the Genome Analysis Toolkit (GATK) version 4.1.9.0,[37] and mapping quality filtering was set to 20 with samtools 1.16.[38] Variant calling was performed using HaplotypeCaller in GATK to generate intermediate gVCF files, which were further merged with GenomicsDBImport and used in GenotypeGVCFs for joint genotyping of multiple samples. Variant quality score recalibration was performed by VariantRecalibrator for both SNPs and indels (insertions and deletions) separately, and filtered sites were excluded with SelectVariants.

Prior to the subsequent analyses, our newly generated WGS and Illumina chip data were checked for relatedness with KING.[39] A kinship coefficient threshold of 0.0884 was used as a cutoff to filter out second-degree or closer relatives.

Merging with ancient and modern reference datasets

We merged our 34 full genomes with two Armenian (4GP) whole genomes from the Simons project.[27] Fastq data were aligned and genotyped with the same pipeline as described above. Tidypopgen[40] was used to combine modern Armenian samples with calls from modern populations in the Human Origins (HO) dataset[41] and genome-wide data from ancient samples in Lazaridis et al.[7]

We excluded positions with multiple alleles; without any additional analysis-specific filtering, this led to the following SNP sets: 1,053,115 SNPs when the analysis was based solely on modern Armenian and ancient samples, and 591,014 SNPs when a combined dataset of ancient and other modern samples was used, from the initial 1,054,671 and 591,642 SNPs in the HO panel, respectively. To mitigate the potential impact of technological bias, we restricted our dataset to a single technology type, either shotgun or capture, whenever it was feasible.[42] Consequently, our Dstat analysis and principal component analysis (PCA) were performed solely on capture-generated genomes, while DATES analyses utilized shotgun-generated genomes only. However, for certain analyses such as ADMIXTURE, qpAdmix, and qpGraph, where we combined modern and ancient datasets, the genomes representing the ancestries required for modeling were not consistently generated using the same technology. We eliminated both the samples exhibiting close kinship and those that did not meet the contamination analysis criteria. Detailed information is presented in detail in Table S2.

Population genomic analyses

PCA was performed with a broad panel of modern and ancient samples using the smartpca software from Eigensoft 7.2.0[43] with the outlier removal option off. All ancient samples were projected onto the principal components by using the options lsqproject: YES. For the analyses, we set a cutoff of 100,000 SNPs.

D-statistics on SNP array data were computed using the qpDstat program in the ADMIXTOOLS package 7.0.1.[44] Statistics were considered significant if the Z score was greater than 3 corresponding to a p value of <0.001. To account for the potential effect of ancient DNA damage, we used transversions only.[45] Thus, D-statistics with both ancient and modern comparative samples was performed based on 109,869 SNPs, while with only ancient samples it was based on 197,022 SNPs. We applied a 35,000 SNP cutoff for the analyses.

Allele sharing between modern Armenian populations and between each Armenian population and neighboring ethnic groups was estimated for each site as the probability that two randomly drawn carriers in the pooled population are from different populations, normalized by the panmictic expectation as described in Chiang et al.[46] Taking into consideration the genetic peculiarities of the Sasun population, we did not include it in the central Armenian group in the analysis. However, due to its small population size ($n = 5$), we did not conduct allele-sharing analysis for this population separately either. To increase the number of samples, newly Illumina genotyped data of 23 Armenians were also added to the dataset with Genotype Harmonizer (Table S1).[47] Singletons, non-variant sites, and missing sites were removed. Thus, our filtering resulted in 677,047 SNPs. For the analyses between the Armenian and comparative populations, a pooled Armenian population was combined with the HO dataset. After removing two Iranian samples that had a significant number of missing sites, a total of 514,134 SNPs was included in the allele-sharing analysis.

Ancestry proportions were estimated with ADMIXTURE version 1.3.[48] Prior to the analyses, we pruned for SNPs in linkage disequilibrium (LD) with the "-indep-pairwise 200 25 0.4" command in PLINK 1.9.[49] For unsupervised ADMIXTURE with Armenian populations this resulted in 199,955 SNPs, with Armenian and comparative modern populations 203,863 SNPs, with a combined ancient and modern data 289,247 SNPs. We ran the analyses five times, and the cluster membership coefficient matrices from multiple runs were analyzed using CLUMPP.[50]

To increase our power to identify the relationship among individuals, we used a "chromosome painting" technique applied to the genome-wide haplotype data, as implemented in CHROMOPAINTER version 2.[51] The dataset included 926 samples and 482,117 SNPs after carrying out the variant filtering with –geno 0.01, –hwe 0.000001, and –maf 0.001 in PLINK. We phased our data with shapeit5 (phase_common_static tool),[52] with haplotype data from 1000 GP phase 3 provided as a reference panel.[53] First, the parameters N_e (effective population size) and μ (global mutation rate) were estimated for four chromosomes separately (chromosomes 1, 4, 15, 22) by 10 EM iterations. The convergence of the results was later confirmed, and average values across the chromosomes were calculated. These fixed values were used in a subsequent run with all chromosomes and all individuals against all others with "-a 0 0" option. ChromoCombine was applied to bin the different chromosomal outputs. We conducted PCA of haplotypic similarity based on the CHROMOPAINTER coancestry matrix.

We then used fineSTRUCTURE version 4.1.1[51] to classify the individuals into clusters based on genetic relationships. We ran 2,000,000 sample iterations of Markov chain Monte Carlo (MCMC) with 100,000 burn-in steps, keeping every 10,000th sample.

A multiple sequentially Markovian coalescent (MSMC) (version 2.1.2) approach[54] was implemented to infer historical changes in effective population size and cross-coalescence rates (CCRs). The genomes were phased with shapeit5 using 1000 GP phase 3 variant set release.[53] The custom scripts from the github repository were used to prepare input files for the analysis.[54] Missing sites were filtered out and multiallelic variants normalized prior to the analysis. For effective population size estimations, we incorporated four individuals (8 haplotypes) from each population, while for population separation we used two haplotypes per population. We assumed a mutation rate of 1.25×10^{-8} per base pair per generation and a generation time of 25 years. For the estimation of CCR, one individual (2 haplotypes) from each population was used. Comparative datasets were obtained from the Human Genome Diversity Project (HGDP)-CEPH panel[55] and previous publications.[56,57] For Syrian samples only, we proceeded with the alignment using Long Ranger[58] (version 2.2.2, using GATK v.3.7), a software designed to process chromium sequencing output. The remaining samples were aligned using the pipeline described above.

Admixture graph modeling was carried out with qpGraph in ADMIXTOOLS2[59] using Mbuti as an outgroup. We aimed to obtain graphs with modern Armenians as receiving admixture from Armenia_LBA and Lebanon_IA. For the initial exploration of the graphs analysis, we used find_graph() with 250 replications. However, we noticed that the majority of the graphs appeared incorrect in terms of the chronological order of the populations included and/or showed incorrect relationships between samples, despite existing knowledge about these relationships. We then first obtained a scaffold graph (Figure S24) with Lebanon_IA, Armenia_LBA along with the populations shown to be a source for ancient individuals from the Armenian highlands.[7] These are Caucasus hunter-gatherers (CHG), Eastern hunter-gatherers (EHG), Anatolia_N, and Levant_N. For the analyses, we used transversion only and did not allow for missing sites, which resulted in 65,370 SNPs after filtering. We acknowledge that several other graphs also fit the dataset. However, our primary approach was to align our graph with the knowledge of the relationships from previous analyses and follow the chronological order of the

samples. To model the ancestry of Armenian samples, we used qpAdm from the same package. We tested a two-way admixture from populations of Iron Age Lebanon and Late Bronze Age Armenian highlands. We used Mbuti, EHG, Kostenki14, Anatolia_N, Italy_North, and Villabruna_HG as reference populations. Prior to conducting the analysis, we assessed whether the reference populations could differentiate well between the source populations by computing qpWave. After filtering out transversions, we had 109,797 SNPs. For both qpGraph and qpAdm, we applied a 35,000 SNP cutoff for the analyses.

Estimation of F_{ST} pairwise coefficients was conducted in smartpca with default parameters and fstonly: YES.

Runs of homozygosity (ROH) were calculated with PLINK on a merged genotype data for all populations, with a minor allele frequency (MAF) filter set at 0.05. The homozygous regions were defined as those with more than 50 homozygous SNPs in 500 kb of the sliding window. We also used the scanning window hit rate of 0.05 and the options –homozyg-window-het 1 –homozyg-window-missing 5 to allow one heterozygous and five missing calls per window. This resulted in 399,509 SNPs.

Identity-by-descent (IBD) segments were detected using HapIBD[60] with segments shorter than 3 cm filtered out. We then used HapNe-IBD to infer the demographic sizes of the populations based on the remaining IBD-sharing segments.[61]

LD decay was calculated using PLINK with –r2 option, 70 kb sliding window, and no limit for r2. SNP pairs were sorted in 70-kb bins based on the distance between pairs, and mean values were calculated for each bin.

Y chromosome haplogroups were assigned with yHaplo[62] and later manually checked and refined for higher-resolution SNPs listed in International Society of Genetic Genealogy version 15.73. Mitochondrial haplogroup assignment was performed by HaploGrep2.[63]

A formal continuity test was performed using the method described in Schraiber,[64] with modern Armenian populations separately. We ran this test for captured ancient genomes only. The pysam library in python has been used to perform a pileup on bam files.[65] Prior to the analysis, we fitted alpha and beta priors to the discrete reference allele frequencies. Files of ancestral alleles were taken from 1000 GP phase 1.[53] We restricted the analyses to sites with high-confidence calls where the ancestral state is supported by all sequence comparisons. We also did not include alleles with the frequency 0 or 1 in the modern population. Thus, for tests of population continuity with capture data, we used ∼700,000 SNPs. To date the time of admixture, we used ALDER[66] (version 1.03) with mindis 0.005. We processed Fastq data on 24 Sardinian samples from the HGDP-CEPH panel[55] with a pipeline similar to that for modern Armenian genomes. We combined those with genome-wide data on 99 Armenians from Haber et al.[15] using Genotype Harmonizer, resulting in 645,992 SNPs. Bin size of 0.0001 and mindis of 0.005 were used.

We used DATES version 4010[67] to infer the dates of admixture between Iron Age Lebanon and Late Bronze Age Armenian highlands ancestry in modern Armenian individuals.

Data processing for medical genomics analyses with WGS samples

For medical purposes, we used a pipeline for alignment and genotyping similar to that described above, utilizing the GRCh38 reference genome.

Variant annotation

The variant statistics on the number of SNPs and indels were calculated with bcftools[38] for variant positions in the vcf file with 36 Armenian samples. Variant loci were defined as the ones with non-reference alleles. We defined the pathogenic state of *MEFV* mutations using supporting allele information from the dbSNP report (https://www.ncbi.nlm.nih.gov/snp/).

Vcftomaf[68] was used to run annotation by VEP[69] and convert vcf files to MAF. "Frame_Shift_Del," "Frame_Shift_Ins," "In_Frame_Del," "In_Frame_Ins," "Nonsense_Mutation," "Nonstop_Mutations," or "Splice_Site" annotated variants were grouped as loss-of-function (LoF) mutations. Maftools were used for the visualization of the gene variants.[70] The calculations involving AF refer to the normalized sites annotated by MAF and where all 36 Armenian samples have no missing sites.

Imputation

We masked SNPs from Illumina 1M chip in one of the Armenian genomes. Only biallelic non-missing variant sites have been considered. To evaluate imputation accuracy of masked genotype, we used two reference panels: (1) shapeit5 integrated phased data from 1000 GP phase 3,[71] called against GRCh38 assembly; and (2) 1000 GP reference panel merged with phased data on the rest of Armenian genomes ($n = 35$). We used the phasing_common in shapeit5 for phasing of the Armenian dataset and IMPUTE5 for imputation.[72] Squared Pearson's correlation coefficients (R^2) were calculated to evaluate imputation accuracy between the CG genotypes (0, 1, 2) and the imputed dosages (0, 2).

Results

Variants statistics for modern Armenian dataset

In total, in our dataset comprising 36 modern Armenian samples (two of which have been previously published), we identified 13,523,774 autosomal variants encompassing 11,134,554 SNPs; and 2,510,513 indels, of which 1,180,458 positions were multiallelic. We then classified the variants according to their allele frequency in the combined Armenian dataset. Most of the variants were very common (allele frequency >0.05) ($n = 8,503,355$), and only approximately one-fifth of them were singletons ($n = 3,388,182$) (Figure 2A). A total of 13,283,198 variants were already present in dbSNP, while 567,320 were novel and mostly represented by singletons ($n = 512,915$). As expected, most of the variants were identified in introns or intergenic regions (∼12 million), while LoF variants were less frequent and with a significantly larger proportion of singletons or doubletons in comparison to other variants (Figure 2B). The genomes were enriched with short indels and were mostly causing a non-frameshift change (Figure S1). Similarly, smaller indels were more abundant in protein-coding regions. In total, 12,172 and 10,442 variants were recognized as damaging or potentially damaging with various levels of confidence by SIFT and PolyPhen, respectively (with 7,472 overlapping variants between the two annotations).

The Armenian population is known to be at high risk of familial Mediterranean fever (FMF [MIM:134610]),[73] an

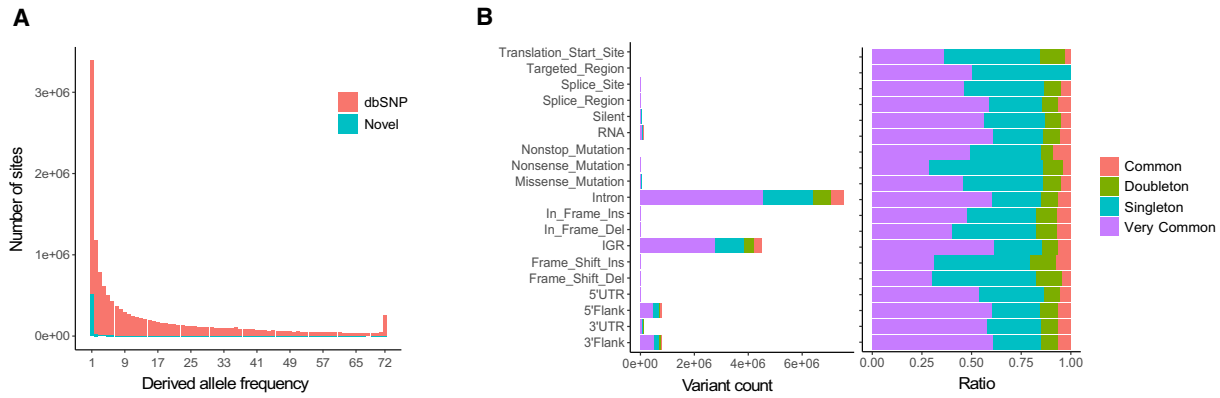


Figure 2. Autosomal variants in the modern Armenian dataset
The variants are grouped according to (A) their allele frequency and (B) their location and corresponding proportions.

autoinflammatory disorder caused by a broad mutational spectrum in the *MEFV* gene [MIM:608107] and primarily found in people of Mediterranean origin. The gene encodes a 781-amino-acid protein called pyrin, which is an important modulator of innate immunity.[74] Strikingly, within our dataset, nearly half of the individuals (16 out of 36) were found to carry pathogenic or likely pathogenic variants in the *MEFV* gene, a significantly higher frequency rate than the 20% reported in previous studies.[73] This highlights the limitations of conventional diagnostic methods such as strip assays, which do not appear to be a sensitive method for identifying pathogenic or likely pathogenic mutations in the Armenian population (Figure S2 and Table S3). Moreover, we found a novel *MEFV* missense mutation in one of our samples, which is predicted to be deleterious by PolyPhen (0.998).

Population structure within modern Armenians

We assessed the genetic diversity and substructure within the Armenian populations. Although the Sasun belongs geographically to the central part of the Armenian highlands, in most of our analyses we treat it as a separate group ($n = 5$), aiming to study its presumed genetic differences from other Armenian populations. The remainder of the Armenian individuals were grouped as western ($n = 11$), central ($n = 12$), and eastern populations ($n = 6$) of the Armenian highlands, taking into consideration their historical, geopolitical, cultural, and linguistic background (Figure 1). In general, we found no stark division between Armenian groups (Figures S3 and S4A–S4C). The allele-sharing analysis revealed a high level of genetic similarity among all groups, with particularly strong genetic links between eastern and central Armenian populations for MAF bins (Sasun is not included in the analysis due to its small sample size). Remarkably, the combined Armenian cohort displayed a substantial degree of genetic affinity with neighboring populations (Figures S4D–S4F). While the pattern in the PCA (Figure S3A) can be mostly explained by the geographic location of the populations except for a few western Armenian individuals, ADMIXTURE for the Armenian samples pointed to subtle distinctiveness in the ancestry components of Sasun (Figure S3B), which may result from admixture or elevated rates of drift in the population. ADMIXTURE analysis with other modern populations indicated close affinity between all Armenian groups and the Assyrians (Figure S5). When we reran ADMIXTURE only for a few Near Eastern populations, Sasun were inferred to have a distinct ancestral component compared to most individuals from those of other Armenian populations and Assyrians at $K \geq 3$ (Figure S6). Bearing in mind that ADMIXTURE results should be cautiously interpreted for populations with a recent bottleneck,[75] we applied additional methods, thus allowing a more robust analysis of population and demographic histories of the region.

The F_{ST} pairwise comparison between the populations revealed close genetic similarities among Armenians and their neighboring populations (Figures 3B, S7, and S8), including Assyrians, although these relationships were relatively less pronounced for the Sasun population. These results were further confirmed while exploring haplotype-sharing patterns of Armenian groups with CHROMOPAINTER (Figure S9) and further visualizing them in a form of the hierarchical clustering tree with fineSTRUCTURE (Figure S10). Even though the Sasun were located in a sister branch, all Armenian groups clustered closely together, suggesting that there are no significant genetic differences among them. At the same time, we observed a handful of Assyrian, Turkish, and Syrian individuals in the Armenian cluster, which rather suggests some genetic substructure within those populations and/or a level of admixture with Armenians. Our PCA based on haplotype sharing (Figure 3A) consistently showed Armenians clustering with neighboring populations from the Caucasus, Anatolia, Iran, and Assyria. Additionally, the results of clustering of Armenian populations support our division of them into regional groups.

To formally test whether the Sasun received additional gene flow from Assyria compared to other Armenian groups, we conducted a D-statistics in the form D(Sasun,

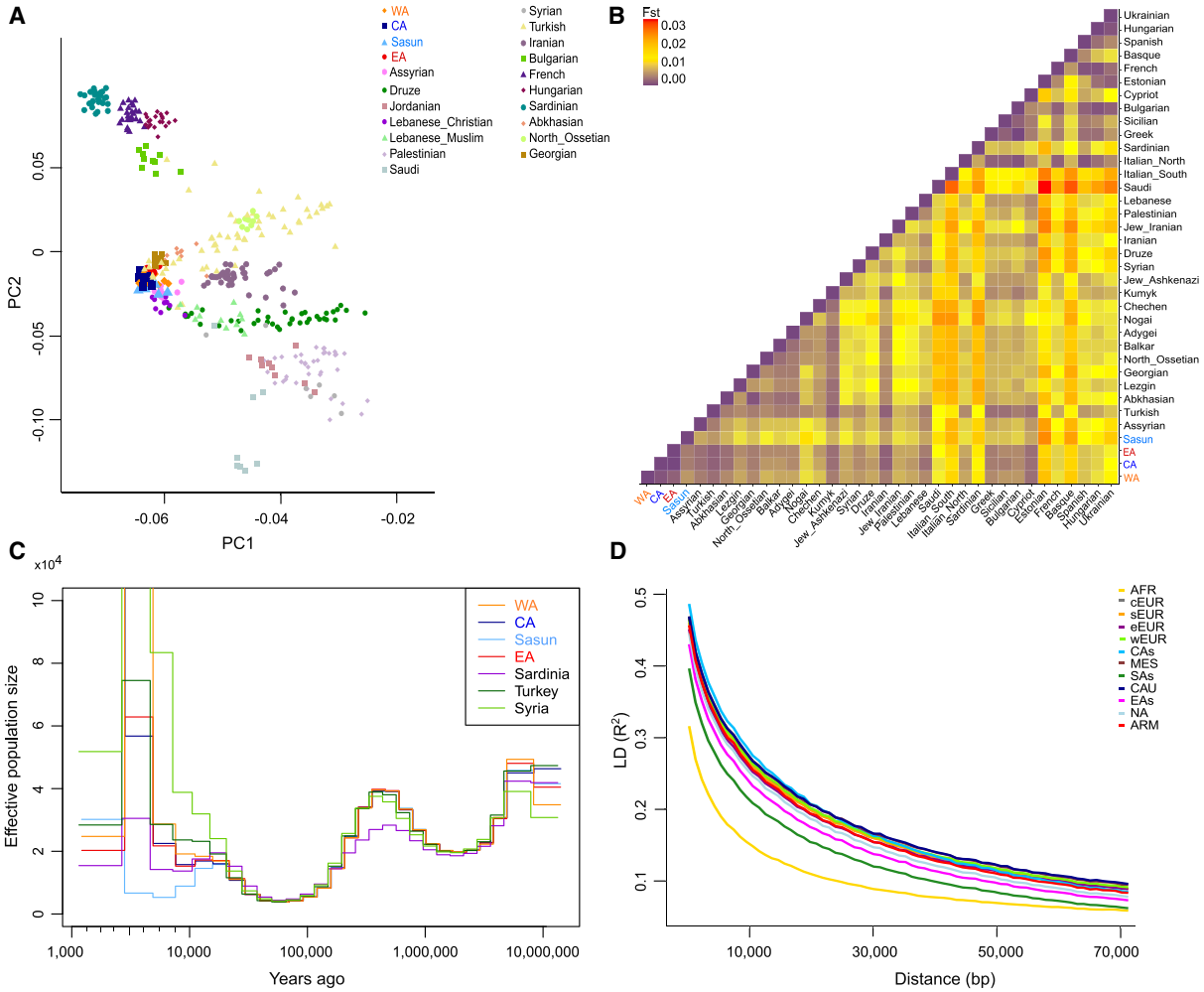


Figure 3. Analyses of population structure, genetic relationships, and population size histories of modern Armenians
(A) Principal component analysis (PCA) based on haplotype sharing. The chunkcount coancestry matrix from CHROMOPAINTER was used to perform PCA.
(B) Matrix of pairwise F_{ST} genetic distances between the populations.
(C) Effective population size estimates inferred from eight haplotypes for each population, using the MSMC model. Except for Sasun, the Armenian groups follow a pattern similar to that of the Turkish samples (a typical European shape) when compared to other populations in the analyses. WA, CA, and EA are the abbreviations for modern Armenian samples from the western, central, and eastern parts of the Armenian highlands, respectively.
(D) The rate of LD decay across the populations (ARM, Armenia; MES, Middle East; CAU, northern Caucasus; wEUR, eEUR, cEUR, and sEUR, western, eastern, central, and southern Europe, respectively; AFR, Africa; NA, SAs, CAs, and EAs, North, South, Central, and East Asia, respectively). Mean values for 1,000-bp bin over 70,000 bp are shown. When comparing LD in the Armenian population with the rest of the groups, the difference is significant with African and Middle East populations (Kolmogorov-Smirnov test, $p < 0.05$).

other_modern_Armenian; Assyrian, Mbuti). Our results showed that the Sasun formed a clade with other Armenian populations that is not broken by Assyrians (Table S4). Based on these analyses, we conclude that all modern Armenians are relatively homogeneous while sharing high similarity to the neighboring populations, and there is no evidence for the Sasun harboring an unusually high contribution from Assyrians.

Demographic history and timescale of divergence

MSMC analyses showed evidence of a population size reduction after 10,000 years ago in the case of Sasun. This finding might underlie their distinct genetic characteristics observed in previous analyses and suggests the bottleneck event and/or the lack of an external gene flow (which would exaggerate population sizes) in Sasun compared to the rest of the Armenian populations. Notably, the population size was even smaller than that of the Sardinians (as depicted in Figure 3C), for whom a substantial population size decrease around the similar time period has been previously demonstrated.[35] The other Armenian groups exhibited effective population sizes typical for most mainland Europeans, i.e., with a strong bottleneck between 50,000 and 60,000 years ago, linked to the out-of-Africa expansion, and a subsequent period of rapid population growth. Nonetheless, we

noticed that most of the cross-coalescent rate curves were quite unstable, which might be because of very recent divergence times and, consequently, insufficient resolution in the MSMC analysis (Figure S11).

We further examined a mean rate of LD decay for Armenians, which is indicative of genetic drift and population history. We found that Armenians displayed greater LD than African, Middle Eastern, South Asian, Caucasus, and western/southern European populations, and lower LD than East Asian and eastern/central European populations (Figure 3D). We also checked the inbreeding status of the population by measuring the average total number of ROH (Figure S12). The Armenians demonstrated a relatively low number of ROH in short ROH length categories in comparison with other populations. However, they exhibited a relatively elevated number of long ROH, reflecting a certain level of consanguinity marriages within the general Armenian population. When we conducted a separate analysis for Armenian groups (Figure S13), we found that the highest indication of consanguinity was observed in the western Armenian population. We also analyzed IBD sharing within the populations and found no evidence for long IBD sharing between individuals, suggesting the lack of both close and distant relatives in the populations (Figure S14). Based on the IBD-sharing patterns, we then tested the demographic scenarios over the last 100 generations, which is considered a more sensitive method for recent time periods than MSMC. Although there is little signal for population changes over the recent timescale, based on the inferred confidence interval we can conclude that the Sasun population had the lowest effective population size compared to the other groups, consistent with our previous results (Figure S15).

Distribution of mitochondrial and Y chromosome haplogroups

The analysis of the spectrum of Y chromosome and mtDNA haplogroups in Armenians showed a prevalence of region-specific lineages (Figures S16 and S17; Table S1). In particular, the most common Y chromosome haplogroups were R1b (23%), J2a (26%), and J1a (16%). In our dataset, all Armenians with haplogroup R1b belong to its Yamnaya-associated R-Z2103 lineage.[7] Interestingly, its greatest genetic diversity has been observed in the Armenian highlands.[24] The haplogroups J2a and J1a are considered Middle Eastern by their origin, but also have deep roots in the Caucasus as they were found in Paleolithic samples from Georgia.[76] The matrilineal gene pool of Armenians is mostly represented by the haplogroups H (28%), J (17%), U (14%), and N (11%). The modern mtDNA haplogroup composition was typical for ancient individuals from the Armenian highlands, except for the latter lineage that had no occurrence in the samples from Neolithic to Medieval times.[14] Overall, our observations of patrilineal and matrilineal genetic pools of Armenians were consistent with the previous results.[14,24–26]

Armenians in relation to other modern and ancient populations: Testing the Balkan theory

In our study, we confined our dataset of ancient samples from the western and central parts of the Armenian highlands to the sites located within the mountain boundaries (Figure 1 and Table S2), which naturally served as a geographic barrier to ancient population movement.[77] To assess the genetic affinities of modern Armenians to modern and ancient individuals from the Balkans, we first performed a PCA, projecting ancient samples from the region onto the first two principal components inferred from modern-day populations of Europe, Middle East, and the Caucasus (Figure 4). We observed that the modern Armenians' cluster falls in between the genetic variation of the modern Caucasus and the Middle East, which is in accordance with their geographic location. When compared with the ancient samples from the Armenian highlands, we noticed that all modern Armenian groups partly overlap with the ancient inhabitants from the eastern part of the highlands, thus suggesting a degree of regional genetic continuity since the Neolithic (the earliest samples available for this region). While the ancient samples from the central region of the Armenian highlands (only a few published at present) are located at the edge of the Armenian cluster, the majority of ancient samples from its western parts are situated further from its core. Considering that the latest ancient specimens from the western Armenian highlands belong to the Early Bronze Age (dataset spanning ~3,950–2,300 BCE), this result complements historical records on the later settlement of this region by Armenians.[78] Notably, the clustering of modern western Armenians closer to both modern and ancient central Armenian highland groups aligns with historical evidence on population movements and resettlements of Armenians from predominantly central to western territories of the highlands during the Byzantine Empire.[79] In stark contrast, both modern and ancient samples from the Balkans appear significantly distant from the Armenian cluster and are drawn mostly toward other European populations. We next conducted ADMIXTURE analysis with a comparative dataset of modern and ancient populations (Figure S18). In line with previous analyses,[7] both modern and ancient Armenian highland samples exhibit a substantial proportion of CHG ancestry. However, while modern Armenians exhibit a relatively uniform pattern of ancestry within the groups, the ancient samples from the highlands display a greater degree of regional heterogeneity in the past. In particular, there is an apparent increase of Neolithic-Anatolian-like ancestry in the samples from the western Armenian highlands (from $K > 5$). At higher K values, there are also variations in the incorporation of EHG ancestry in ancient samples from the highlands, consistent with previous observations.[7] All ancient western and Neolithic eastern samples exhibit a relative lack of this ancestry. After its appearance in the Chalcolithic period, eastern Early Bronze Age individuals demonstrate a visible drop in EHG ancestry, followed by a resurgence

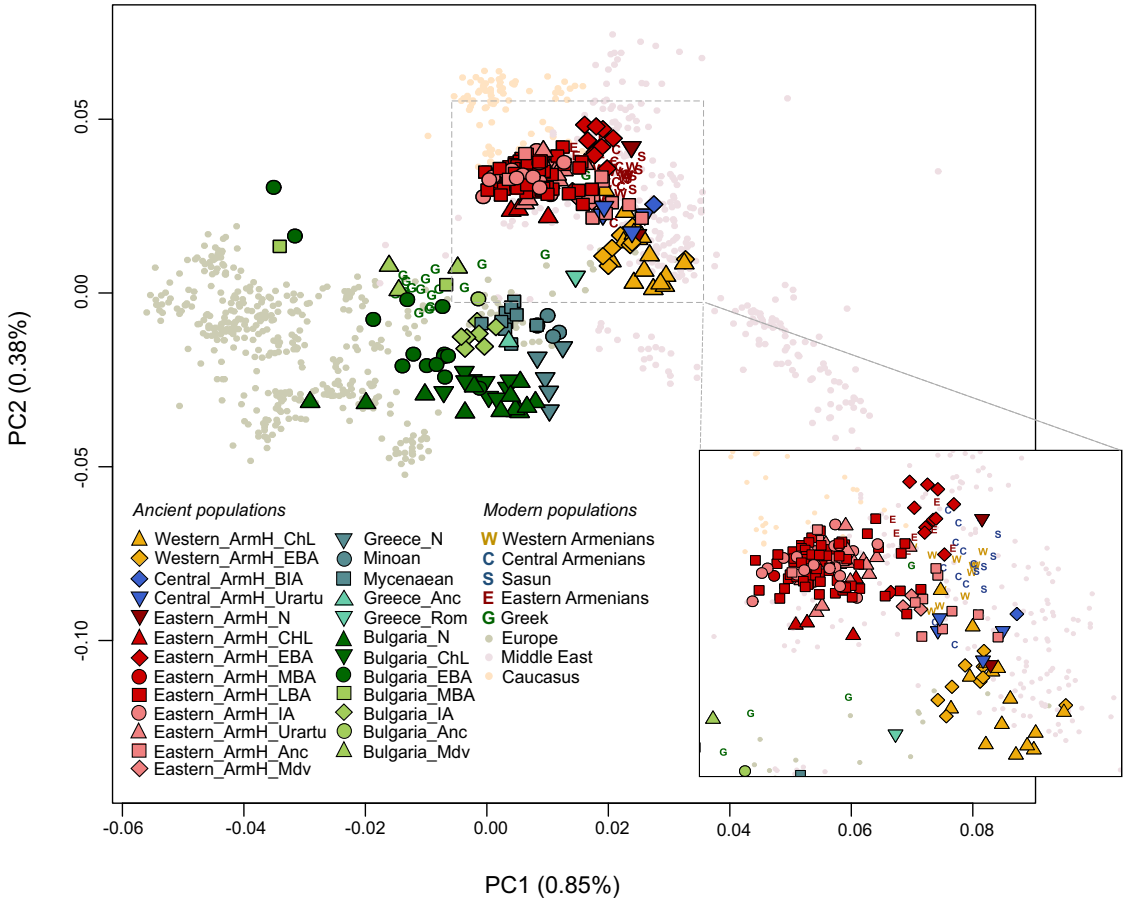


Figure 4. Principal component analysis (PC1 vs. PC2) for modern and ancient populations based on allele frequency data
Values in parentheses represent the percentage of variance explained by a given PC. According to the PCA, ancient samples from the eastern parts of the Armenian highlands cluster with all contemporary Armenian groups (all are marked with red letters), while modern (marked with a blue letter) and ancient samples from the Balkans show distinctive clustering closer to other modern European populations. On the bottom-right corner is a zoomed-in view of the analysis for the Armenian highlands cluster. S, W, C, and E are the abbreviations for modern Armenian samples from the Sasun, western, central, and eastern parts of the Armenian highlands, respectively. In the legend, ArmH refers to the Armenian highlands.

by the Late Bronze Age. Additionally, the K = 7 results suggest elevated ancient Levantine-like ancestry in both western and central ancient Armenian highland populations. This contrasts with the eastern ancient highland groups and aligns with our PCA results.

The pattern observed in previous analyses prompted us to conduct D-statistics in the form D(Modern_Armenians, Ancient_Armenian_highlands_eastern; Ancient_Armenian_highlands_western/central, Mbuti) to further examine whether all modern Armenian groups form a clade with ancient samples from the eastern part of the Armenian highlands, in comparison to the ancient representatives from its central and western regions (Table S5). To account for the plausible artifactual attraction among modern populations when compared to ancient samples affected by post-mortem damage or among two ancient samples when compared to modern populations, in all our D-statistic analyses incorporating ancient individuals we used transversions only. Additionally, in an effort to minimize the potential influence of technical biases that could emerge from the combination of capture and shotgun data, our analyses were restricted to a single technology type unless such a restriction was unfeasible. We did not detect any cases where local ancient representatives disrupted the clade between any modern Armenian populations and ancient samples from the eastern Armenian highlands (in some cases, we detected a stronger signal for attraction between ancient samples).

To better understand the population history of the western Armenian highlands, we conducted outgroup f3 statistics and found that Chalcolithic and Bronze Age inhabitants of the region share more genetic drift with Neolithic and Bronze Age samples from western Anatolia as well as Neolithic and Chalcolithic populations from Greece and Bulgaria (Figure S19). To confirm this observation and check whether any modern populations break the clade between ancient samples from the western/central and eastern parts of the Armenian highlands, we calculated D-statistics in the form D(Ancient_Armenian_highlands_western/central,

Ancient_Armenian_highlands_eastern; X_Modern, Mbuti). We detected a signal of a gene flow into ancient populations of the western Armenian highlands from the Sardinian- and Near-Eastern-like sources (Table S6). This signal is already detectable in the region during/after the Chalcolithic. Among modern-day populations in Europe, Sardinians have the highest affinity to Early European Farmers and often act as a proxy for Neolithic ancestry in population genetic analyses.[41,46] At the same time, modern Levantine populations have been shown to harbor a substantial level of ancestry from the Bronze Age Levant, which, in turn, has ancestry from local Neolithic populations.[80] We then ran a similar D-statistic test for the central Armenian highlands group (Table S6) and found only a single instance of a significantly broken clade by Near-Eastern-like source when an Urartian sample from the region and the Early Iron Age population of the eastern part of the highlands were used in the analyses. To determine whether small sample sizes combined with population heterogeneity complicate the detection of a gene flow signal, we conducted a series of D-statistic tests using ancient populations from the western Armenian highlands and those with which a significant clade was detected in previous analyses (Figures S20–S23). By drawing random subsamples of these populations for each run, we found that a signal becomes detectable ($Z > 3$) in the vast majority of runs when both tested ancient populations each have a sample size of seven (it starts improving from a sample size of five when the signal is strong). Therefore, we conclude that additional samples are necessary for the ancient central Armenian highlands dataset to ensure reliable results. We applied the same sample-size criteria and level of caution to all our D-statistic tests in our study involving ancient samples.

Given the results above, we used the eastern Armenian highland populations as the sole ancient representatives from the region in the following analyses. To further investigate the question on the genetic origin of Armenians, we used D-statistics in the form D(Modern_Armenians, Ancient_Armenian_highlands_eastern; X_Balkan, Mbuti) to formally test whether modern Armenians received any genetic input from ancient and modern samples from the Balkans (Table S7). We did not observe any significantly positive values for the D-statistics (which are expected under the Balkan theory), thus questioning the presence of the Balkan-related ancestry in modern Armenians. Furthermore, we found that in most of our comparisons ancient and modern samples from the Armenian highlands form a clade ($|Z| < 3$) to the exclusion of ancient and modern samples from the Balkans. At the same time, we detected a signal indicating that the ancestors of modern Armenians received a genetic influx from an external source presumably after the Late Bronze/Iron Age ($Z < -3$) (we acknowledge the small sample size for Middle Bronze Age samples). However, we noticed an increased derived allele sharing between modern Armenians and Greeks ($Z > 3$), which might be a consequence of the shared gene flow between the populations, and this requires further investigation. To compare the results that we obtained for the populations of the Armenian highlands, we carried out a similar set of D-statistic tests, but this time, we examined a clade involving modern and ancient Greek populations in the form D(Ancient_Greek, modern_Greek; X_ancient_populations, Mbuti) (Table S8). The findings pointed to a level of genetic discontinuity in the region following the Later Bronze Age, revealing a genetic connection between Mycenaean, Minoan, and Neolithic Greece samples with those from Anatolia, Armenia, and the Middle East (Table S8).

Taken as a whole, our results support the lack of significant genetic input from the Balkans into ancient and modern populations of the Armenian highlands.

Insights into regional continuity

We next used a maximum-likelihood approach[64] to test for continuity of ancient and modern populations of the Armenian highlands. This model assumes a scenario where changes in allele frequency from the common ancestor forward in time are explained through drift alone. The test suggests a very recent drift time in all modern Armenian populations since the common ancestor with the ancient samples, thus supporting the close relationship between ancient and modern inhabitants of this region. However, the model rejects the regional continuity hypothesis, since the ancient populations appear to have significantly larger drift times (Table S9). This result is compatible with a scenario of admixture into modern Armenians from an outside population, which diverged earlier than the split time between the modern and ancient inhabitants of the region. Consequently, this admixture would have increased heterozygosity in modern Armenians and thus would exaggerate the drift time in ancient Armenian highland populations (since no genetic influx is allowed according to the model), eventually resulting in the rejection of regional continuity. We also acknowledge that this method is highly sensitive, even to small levels of admixture, which presents a limitation given that all present-day populations have experienced some degree of admixture at various points in their history.

Identifying the source, extent, and time of gene flow

We then investigated corresponding signatures of gene flow into the Armenian highlands by using D-statistics in the form D(Modern_Armenians, Ancient_Armenian_highlands_eastern; X_ancient_population, Mbuti). We found that several ancient samples from the Middle East, Anatolia, and Eurasian Steppe are more closely associated with the ancient samples from the Armenian highlands than the latter with the modern Armenians ($Z < -3$), thus corroborating our previous results (Table S10). In agreement with the D-statistics results with ancient Balkan samples, the highest negative values have been detected in the case of the Late Bronze/Iron Age samples from the Armenian highlands. It is worth mentioning that the Sasun

population shows similar results in the D-statistics, reinforcing the finding that the population has ancestry similar to that of the other Armenian groups (Table S11). Due to the prevalence of captured ancient samples and our commitment to maintaining a consistent sequencing technology criterion for sample inclusion in our D-statistic tests, we chose not to include Iron-to-Middle-Age Lebanon and Late Bronze Age Armenian highland samples that were sequenced using the shotgun method. However, when incorporating those samples in the separate D-statistics in the form D(Modern_Armenians, Late_Bronze_Age_Armenian_highlands_eastern; ancient_Lebanon, Mbuti) using all SNPs, the Iron Age, Antiquity, and Roman-period Lebanon samples stood out as the only comparative ancient samples showing positive Z values, suggesting an affinity with modern Armenians (Table S12). When we restricted the analysis to transversions only, we had to exclude four out of five Late Bronze Age Armenian highland populations to meet the SNP cutoff of 35,000. As a result, the analysis with transversions did not yield significant results, plausibly due to the smaller sample size and the reduced number of SNPs.

To check whether any modern populations break the clade between modern Armenians and ancient samples from the region, we calculated D-statistic in the form D(Modern_Armenians, Ancient_Armenian_highlands_eastern; X_modern_population, Mbuti). We found a signal of a gene flow into modern Armenians from the Sardinian- and Near-Eastern-like sources after the Late Bronze/Iron Age (though it should be noted that our sample size for Middle Bronze Age samples is currently limited) (Table S13). It is worth noting that while all Armenian groups show a similar tendency, the central Armenian group had the highest contribution to this signal, which may be due to the relatively larger sample size of this population. The Assyrians also disrupted the clade between the Late Bronze/Iron Age populations from the eastern parts of the highlands and modern eastern and central Armenians, while showing a similar tendency for Sasun and western Armenians. This confirms our previous conclusions on the lack of any additional Assyrian ancestry in Sasun and supports the suggestion that the results of the previous D-statistic test with modern Greeks (Table S7) indicate a genetic influx into modern Armenians, which is shared with modern Greeks. Altogether, our results suggest a gene flow from a Neolithic-Levantine-Farmers-like source into the eastern parts of the Armenian highlands, presumably during/after the end of the Bronze Age, which differentiated modern Armenians from their regional ancestors. This result aligns with the findings from a prior study, which demonstrated an increase in Pre-pottery Neolithic Levant ancestry in Armenia following the Late Bronze Age (~3,200 years ago).[7]

Interestingly, we observed an affinity between Late Bronze Age Armenian samples and certain modern northern and eastern European populations. We speculate that this could be due to the later arrival of the steppe-related ancestry to that region (Table S13). We also detected that the D-statistic tests with Early Bronze Age samples from the Armenian highlands nearly reached significance (Z = 3). When randomly sampling Late Bronze Age individuals for a smaller population size, we confirmed that sample-size limitations and population heterogeneity affect the D-statistic results (Figure S24), raising caution in interpreting results for groups with limited numbers of individuals.

We then conducted a pairwise comparison among ancient Armenian highland populations in an effort to reveal the timing of the Neolithic-Levantine-Farmers-like genetic influx (Table S14). Apart from the latter, we identified the impact of gene flow associated with the steppe as a contributing factor to the genetic heterogeneity within eastern Armenian highland populations over time. Consistent with a previous study,[7] the D-statistic tests revealed the previously mentioned gene flow in the region, followed by a gradual reduction of steppe ancestry after the Iron Age.

We then attempted to date the genetic influx based on the patterns of linkage disequilibrium with ALDER (Figure S25). We had to run a one-population test (modern Armenians as the target with input from Sardinians), which has limited power, so we added previously published genotype data on 99 Armenians[15] and 24 Sardinians.[55] We were able to detect a significant signal of admixture with a Sardinian-like source in Armenians, with an estimated ancestry contribution of 41.6 ± 3.6. The timing of the admixture obtained from ALDER suggests a relatively old event (the best signal corresponds to 218.50 ± 16.39 generations ago or about 6,100 years), which is closer to the end of the Chalcolithic/Early Bronze Age in Armenia.

While ALDER has been shown to be reasonably robust with respect to having the exact source of admixture, the discrepancy in dating might be due to the Sardinian being only a distant proxy of the population that admixed with Armenians. We then used DATES, which allows one to date admixture events by incorporation of ancient DNA (Figure S26). The analysis shows that Iron Age Lebanon and Late Bronze Age Armenian highland samples were admixed 55 generations (standard error values ranging from 57.23 to 35.95), which roughly corresponds to the onset of the Late Bronze Age in Armenia.

We explicitly modeled the ancestry of Armenians by fitting the expected f-statistic values to those observed using qpGraph (Figures 5, S27, and S28). A model that fits the data suggests that, along with the major genetic contribution from the local Late Bronze Age population, modern Armenians received a gene flow from a population related to Iron Age Levant. Intriguingly, the amount of genetic contribution from the latter group varies when we consider the general Armenian population (45%, Figure 5A) or subpopulations (28%–53%), with the eastern Armenians having the smallest admixture proportion (28%, Figure 5B). The genetic differences among geographic groups are in line with PCA results (Figure S3A), where some of the western and central (including Sasun, which is the most distant) Armenian individuals appeared to be further from the core Armenian cluster.

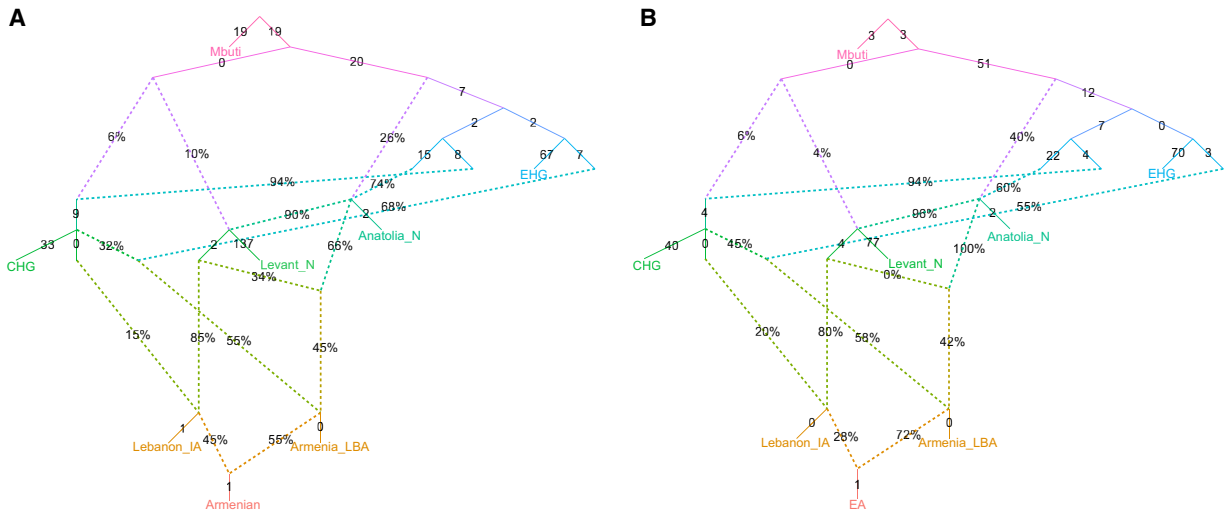


Figure 5. Admixture graphs fitting modern Armenian genomes, which can be modeled as a mixture from Armenia_LBA- and Levant_LBA-like sources

Graphs are shown for (A) pooled Armenian (LL = 25, worst-fitting f-statistic, $Z = 2.9$) and (B) eastern Armenian populations (LL = 20.17, worst-fitting f-statistic, $Z = 2.77$). EA is the abbreviation for eastern Armenians.

Finally, we used qpAdm in an attempt to model modern Armenians with the genetic contributions from the Iron Age Levant and Late Bronze Age Armenian highland populations. However, this analysis was highly inconsistent and unstable while changing the sample size in the populations or adding more populations in the reference group (despite qpWave showing their good representation). As such, we obtained a result of 0.6 and 0.4 for genetic contribution from the population of Iron Age Levant and Late Bronze Age Armenian highlands, respectively. When considering between 3 and 4 standard deviations (standard error is equal to 0.04), these proportions align with the previously mentioned finding.

Imputation

We aimed to assess the imputation accuracy when a set of population-specific genomes are added to the most updated reference panel such as the 1000 GP phase 3, called against the GRCh38 assembly. For this, we masked all positions in one Armenian genome that are not present in the Illumina Human-1M array, imputed these SNPs from the remaining variants, and compared the imputed and known genotypes. The imputation was performed with two reference panels, separately: 1000 GP haplotype data ($n = 5{,}248$) and a combined reference panel, consisting of the former and remaining Armenian genomes ($n = 35$). As a result, we found slightly improved values for rare variants when using a combined panel, whereas the distribution across the other frequency blocks remained quite similar (Figure S29).

Discussion

We conducted a comprehensive WGS analysis of Armenians using a sizable dataset of high-coverage genomes. In line with earlier studies on matrilineal genetic composition,[26] we found that Armenians are characterized by little within-population substructure. The only population that subtly differs from other Armenian groups is the Sasun, which was previously reported as distinct based on the analysis of patrilineal lineages.[24] We demonstrated that the Sasun population underwent a distinct demographic bottleneck, which is the main underlying cause of the observed genetic differences from other Armenians. Of note, all Armenian populations shared high genetic similarities with the populations in close geographic proximity, including Assyrians, and the Sasun did not receive any significantly greater gene flow from the latter, which could have explained their slight genetic divergence from other Armenian populations.

We described the variant architecture in the Armenian genomes and discovered more than 500,000 variants not previously reported in the dbSNP variant database. As a separate question, we investigated the power of imputation when population-specific reference is added to the 1000 GP phase 3 reference panel, aligned to the GRCh38 assembly. Given the predominantly rare frequency of variants in the genomes, the inclusion of a relatively small number of population-specific genomes alongside the 1000 GP dataset resulted in only marginal improvements in our imputation capability. Consequently, a substantially larger population-specific dataset is required to significantly enhance the imputation power.

We focused on solving a long-standing puzzle regarding Armenians' genetic roots. Although the Balkan hypothesis has long been considered the most plausible narrative on the origin of Armenians, our results showed that modern Armenians are genetically distinct from both the ancient and present-day populations of the Balkans. While a recent study[7] speculates that the dilution of the EHG ancestry

after the Iron Age could suggest Balkan-related gene flow, our results reveal a different source for this genetic input. At the same time, Lazaridis et al.[7] confirm the lack of a Balkan genetic component in the ancient samples from Armenia.

On the contrary, we revealed a high level of regional genetic continuity in eastern parts of the Armenian highlands for well over 6,000 years, confirming analyses in previous studies.[14,15] A recent study also suggests a similar level of stability up to the Bronze Age in the South Caucasus (there was no test for later inputs, as this study used only aDNA).[6] This pattern stands in contrast to most other western Eurasian populations, which have undergone multiple large influxes and turnovers.[9,41] A relatively similar example is the Sardinians, who have long been considered as a genetic isolate in the region since the Neolithic, but recent studies have shown that the island received numerous genetic inputs after the Bronze Age,[81] accounting for 38%–44% of their ancestry.

Although the western and central regions of the Armenian highlands are currently not densely sampled for ancient DNA, we have nonetheless been able to discern differences in the timing of Levantine-related ancestry's penetration into various parts of the highlands. While we found that the latter was already present in the western parts of the highlands during the Chalcolithic (later Levantine-Early-Farmers-like gene flow might have amplified this signal), we documented that the eastern parts received it through admixture event that likely occurred at some point after the Early Bronze Age, thereby disrupting the genetic continuity in the region. We acknowledge the sample-size limitations in pre-Late Bronze Age populations from the Armenian highlands and recognize that gene flow may have been a gradual process that reached its peak intensity by the end of the Late Bronze Age but likely started much earlier (according to the results from ALDER and DATES). Thus, the first migration to the eastern Armenian highlands after the Neolithic introduced a steppe-related ancestry in the region, while the second augmented the proportion of the Levantine-related ancestry especially in the post-Late Bronze Age inhabitants of the Armenian highlands. Our results on the source population and date for the mixture are roughly comparable with the genetic studies on Armenians conducted so far. An elevated Neolithic-European-Farmers-like ancestry in Armenians was previously detected based on modern genomes only,[15] and the gene flow rate was estimated as 29%. In our study, we noticed a cline for the Neolithic-Levantine-Farmers-like genetic influx within Armenian geographic groups; the most isolated population were eastern Armenians (28% of Iron Age Lebanon-related genetic input inferred by qpGraph), which agrees with their geography and thus shows the effect of mountains as barriers to genetic exchange. At the same time, we acknowledge that for ancestry modeling and calculations of the genetic contribution, it was not feasible to avoid mixing different sequencing technologies, which could potentially impact the results. As a note of caution, it is important to consider that in D-statistics, modern-modern or ancient-ancient affinities may result not only from postmortem damage but also from reference bias,[82] which cannot be eliminated solely by using transversions. This is why we employed additional methods in our study, drawing conclusions from all of them.

Our study including aDNA samples enhances the power of admixture tests by providing proper sources for population ancestry. As such, we could not find evidence for mixtures of multiple populations during the period 3,000–2,000 BCE.[15] In contrast, we found a signal for a single admixture event from a Levantine-Early-Farmers-like source, which continuously occurred during/after the end of the Late Bronze Age. We acknowledge that the admixture event could have occurred earlier, which would align well with our estimated dates (inferred by ALDER and DATES) and recent studies on the demographic history in the adjacent Caucasus region.[83] Interestingly, archaeological evidence indicates that Middle and Late Bronze Age periods in the Armenian highlands were marked by significant transformations.[10] The Middle Bronze Age saw the emergence of diverse cultures, a predominant nomadic lifestyle, and active relations with Anatolia and the Aegean. The Late Bronze Age, on the other hand, was characterized by intense cultural interactions with neighboring populations, including the Hurrians, Hittites, and Mesopotamians.

In summary, we conclude that there was a large-scale movement across the Middle East at some point after the Early Bronze Age. From the genetic point of view, this movement has led to a sizable input of Neolithic-like ancestry in the populations inhabiting the region. In particular, this movement reached the mountainous regions of the eastern parts of the Armenian highlands, resulting in the genetic divergence of the modern populations from their regional ancestors. The questions of exactly where and when these immigrants came from, as well as what the trigger was for such a widespread migration wave, remain unanswered. Further studies on the complex demographic processes of the region are needed along with the incorporation of additional ancient and modern data from the Armenian highlands, Anatolia, and the Levant.

Data and code availability

The generated raw Fastq files in this study are available through the European Nucleotide Archive under accession number ENA: PRJEB78867. Genotype data produced by HaplotypeCaller and Illumina genotyping data can be accessed through FigShare: https://doi.org/10.6084/m9.figshare.27277578.

Acknowledgments

L.Y., A. Hovhannisyan, M.A., and Z.K. were supported by the Science Committee of the Ministry of Education and Science of Armenia (research project no. 21AG-1F025). A. Hovhannisyan acknowledges funding by EU MSCA-IF under grant agreement 101063265, ACTIVITY 5 of the ESF DoRa PROGRAMME, Calouste Gulbenkian Foundation, and Foundation for Armenian Science

and Technology (FAST). A. Hovhannisyan is grateful to Mait Metspalu and Richard Villiems for their help and supervision of the ESF DoRa Programme research at University of Tartu. A. Manica, P.M.D., and E.R.J. were supported by the ERC Consolidator grant 647797 "LocalAdaptation." We thank the High Performance Computing Center of the University of Tartu for the provision of computational facilities. We are grateful to Hovann Simonian for supporting a project aimed at creating a Genetic Atlas of Armenia. We thank Daniel Bradley and Lara Cassidy for constructive discussions and comments.

Author contributions

A. Manica, L.Y., and A. Hovhannisyan conceived and designed the study. Sampling was performed by L.Y., M.A., and Z.K. Funding for modern whole-genome generation was acquired by L.Y. and P.H. A. Hovhannisyan performed the relevant population and medical genetic analyses under the supervision of A. Manica. P.M.D., E.R.J., A. Hakobyan, J.G.S., L.S., H.S., A. Margaryan, Z.X., S.J., and J.B. contributed to data analyses. A. Hovhannisyan and A. Manica designed the figures and wrote the manuscript. All authors contributed to the final version of the manuscript.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2024.10.022.

Received: March 1, 2024
Accepted: October 30, 2024
Published: November 25, 2024

References

1. Lang, D. (1981). The Armenians: A People in Exile (Routledge).
2. Dolukhanov, P., Aslanyan, S., Kolpakov, E., and Belyayeva, E. (2004). Prehistoric sites in northern Armenia. Antiquity *78*, 301. https://antiquity.ac.uk/projgall/dolukhanov301/.
3. Frahm, E., and Carolus, C.M. (2022). Identifying the origins of obsidian artifacts in the Deh Luran Plain (Southwestern Iran) highlights community connections in the Neolithic Zagros. Proc. Natl. Acad. Sci. USA *119*, e2109321119. https://doi.org/10.1073/pnas.2109321119.
4. Pinhasi, R., Gasparian, B., Areshian, G., Zardaryan, D., Smith, A., Bar-Oz, G., and Higham, T. (2010). First direct evidence of chalcolithic footwear from the near eastern highlands. PLoS One *5*, e10984. https://doi.org/10.1371/journal.pone.0010984.
5. Dong, Y., Duan, S., Xia, Q., Liang, Z., Dong, X., Margaryan, K., Musayev, M., Goryslavets, S., Zdunić, G., Bert, P.F., et al. (2023). Dual domestications and origin of traits in grapevine evolution. Science *379*, 892–901. https://doi.org/10.1126/science.add8655.
6. Skourtanioti, E., Erdal, Y.S., Frangipane, M., Balossi Restelli, F., Yener, K.A., Pinnock, F., Matthiae, P., Özbal, R., Schoop, U.D., Guliyev, F., et al. (2020). Genomic history of neolithic to bronze age Anatolia, northern Levant, and southern Caucasus. Cell *181*, 1158–1175.e28. https://doi.org/10.1016/j.cell.2020.04.044.
7. Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., et al. (2022). The genetic history of the Southern Arc: A bridge between West Asia and Europe. Science *377*, eabm4247. https://doi.org/10.1126/science.abm4247.
8. Wang, C.C., Reinhold, S., Kalmykov, A., Wissgott, A., Brandt, G., Jeong, C., Cheronet, O., Ferry, M., Harney, E., Keating, D., et al. (2019). Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. Nat. Commun. *10*, 590. https://doi.org/10.1038/s41467-018-08220-8.
9. 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. https://doi.org/10.1038/nature14317.
10. Bobokhyan, A., Kunze, R., Meliksetian, K., and Pernichka, E. (2017). Society and Metal in Bronze Age Armenia. In At the Northern Frontier of Near Eastern Archaeology: Recent Research on Caucasia and Anatolia in the Bronze Age. Proceedings of the International Humboldt-Kolleg Venice, E. Rova and M. Tonussi, eds. (Brepols), pp. 501–525.
11. Heggarty, P., Anderson, C., Scarborough, M., King, B., Bouckaert, R., Jocz, L., Kümmel, M.J., Jügel, T., Irslinger, B., Pooth, R., et al. (2023). Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. Science *381*, eabg0818. https://doi.org/10.1126/science.abg0818.
12. Kent, R.G. (1953). Old Persian: Grammar. Texts. Lexicon, *33* (American Oriental Society).
13. Cameron, G.G. (1951). The Old Persian text of the Bisitun inscription. J. Cuneif. Stud. *5*, 47–54.
14. Margaryan, A., Derenko, M., Hovhannisyan, H., Malyarchuk, B., Heller, R., Khachatryan, Z., Avetisyan, P., Badalyan, R., Bobokhyan, A., Melikyan, V., et al. (2017). Eight millennia of matrilineal genetic continuity in the south Caucasus. Curr. Biol. *27*, 2023–2028.e7. https://doi.org/10.1016/j.cub.2017.05.087.
15. 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. https://doi.org/10.1038/ejhg.2015.206.
16. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., and Myers, S. (2014). A genetic atlas of human admixture history. Science *343*, 747–751. https://doi.org/10.1126/science.1243518.
17. Herodotus, The Histories (1920). With an English Translation by Godley, A. D. (Harvard University Press).
18. Diakonoff, M. (1984). The Pre-History of the Armenian People (Caravan Books, Delmar, NY).
19. Khorenatsi, M. (1978). History of Armenians (Harvard University Press).
20. Redgate, A.E. (2000). The Armenians (Oxford: Blackwell).
21. Jahukian, G. (1972). Hay barbaragitutyan neratsutyun (Introduction to the Armenian dialectology) (Academy of Sciences, Yerevan).
22. Haroutyunian, S. (1997). Armenian epic tradition and Kurdish folklore. Iran Cauc. *1*, 85–92.

23. The Bible, 2 Kings, 19:37.
24. Hovhannisyan, A., Khachatryan, Z., Haber, M., Hrechdakian, P., Karafet, T., Zalloua, P., and Yepiskoposyan, L. (2014). Different waves and directions of Neolithic migrations in the Armenian highlands. Investig Genet *5*, 1–11. https://doi.org/10.1186/s13323-014-0015-6.
25. Herrera, K.J., Lowery, R.K., Hadden, L., Calderon, S., Chiou, C., Yepiskoposyan, L., Regueiro, M., Underhill, P.A., and Herrera, R.J. (2012). Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists. Eur. J. Hum. Genet. *20*, 313–320. https://doi.org/10.1038/ejhg.2011.192.
26. Derenko, M., Denisova, G., Malyarchuk, B., Hovhannisyan, A., Khachatryan, Z., Hrechdakian, P., Litvinov, A., and Yepiskoposyan, L. (2019). Insights into matrilineal genetic structure, differentiation and ancestry of Armenians based on complete mitogenome data. Mol. Genet. Genom. *294*, 1547–1559. https://doi.org/10.1007/s00438-019-01596-2.
27. 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. https://doi.org/10.1038/nature18964.
28. Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L., et al. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. Nature *538*, 238–242. https://doi.org/10.1038/nature19792.
29. 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. https://doi.org/10.1038/nature19310.
30. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172. https://doi.org/10.1038/nature14507.
31. Damgaard, P.d.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369–374. https://doi.org/10.1038/s41586-018-0094-2.
32. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. *16*, 1215. https://doi.org/10.1093/nar/16.3.1215.
33. Li, H., Trimadap. GitHub. https://github.com/lh3/trimadap.
34. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
35. Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics *30*, 2503–2505. https://doi.org/10.1093/bioinformatics/btu314.
36. Broad Institute (2019). Picard Toolkit. https://broadinstitute.github.io/picard/.
37. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
38. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
39. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.
40. Carter, E., and Manica, A. (2024). Tidypopgen: Tidy Population Genetics. R Package Version 0.0.0.9016. https://github.com/evolecolgroup/tidypopgen.
41. 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. https://doi.org/10.1038/nature13673.
42. Davidson, R., Williams, M.P., Roca-Rada, X., Kassadjikova, K., Tobler, R., Fehren-Schmitz, L., and Llamas, B. (2023). Allelic bias when performing in-solution enrichment of ancient human DNA. Mol. Ecol. Resour. *23*, 1823–1840. https://doi.org/10.1111/1755-0998.13869.
43. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190. https://doi.org/10.1371/journal.pgen.0020190.
44. 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. https://doi.org/10.1534/genetics.112.145037.
45. Molak, M., and Ho, S.Y.W. (2011). Evaluating the impact of post-mortem damage in ancient DNA: a theoretical approach. J. Mol. Evol. *73*, 244–255. https://doi.org/10.1007/s00239-011-9474-z.
46. Chiang, C.W.K., Marcus, J.H., Sidore, C., Biddanda, A., Al-Asadi, H., Zoledziewska, M., Pitzalis, M., Busonero, F., Maschio, A., Pistis, G., et al. (2018). Genomic history of the Sardinian population. Nat. Genet. *50*, 1426–1434. https://doi.org/10.1038/s41588-018-0215-8.
47. Deelen, P., Bonder, M.J., van der Velde, K.J., Westra, H.-J., Winder, E., Hendriksen, D., Franke, L., and Swertz, M.A. (2014). Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. BMC Res. Notes *7*, 901–904. https://doi.org/10.1186/1756-0500-7-901.
48. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.
49. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.
50. Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics *23*, 1801–1806.
51. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data.

PLoS Genet. *8*, e1002453. https://doi.org/10.1371/journal.pgen.1002453.
52. Hofmeister, R.J., Ribeiro, D.M., Rubinacci, S., and Delaneau, O. (2023). Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. Nat. Genet. *55*, 1243–1249. https://doi.org/10.1038/s41588-023-01415-w.
53. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
54. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925. https://doi.org/10.1038/ng.3015.
55. Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. Science *367*, eaay5012. https://doi.org/10.1126/science.aay5012.
56. Alkan, C., Kavak, P., Somel, M., Gokcumen, O., Ugurlu, S., Saygi, C., Dal, E., Bugra, K., Güngör, T., Sahinalp, S.C., et al. (2014). Whole genome sequencing of Turkish genomes reveals functional private alleles and impact of genetic interactions with Europe, Asia and Africa. BMC Genom. *15*, 963. https://doi.org/10.1186/1471-2164-15-963.
57. Almarri, M.A., Haber, M., Lootah, R.A., Hallast, P., Al Turki, S., Martin, H.C., Xue, Y., and Tyler-Smith, C. (2021). The genomic history of the Middle East. Cell *184*, 4612–4625.e14. https://doi.org/10.1016/j.cell.2021.07.013.
58. https://support.10xgenomics.com/genome-exome/software/downloads/latest.
59. Maier, R., Flegontov, P., Flegontova, O., Işıldak, U., Changmai, P., and Reich, D. (2023). On the limits of fitting complex models of population history to f-statistics. Elife *12*, e85492. https://doi.org/10.7554/eLife.85492.
60. Zhou, Y., Browning, S.R., and Browning, B.L. (2020). A fast and simple method for detecting identity-by-descent segments in large-scale data. Am. J. Hum. Genet. *106*, 426–437. https://doi.org/10.1016/j.ajhg.2020.02.010.
61. Fournier, R., Tsangalidou, Z., Reich, D., and Palamara, P.F. (2023). Haplotype-based inference of recent effective population size in modern and ancient DNA samples. Nat. Commun. *14*, 7945. https://doi.org/10.1038/s41467-023-43522-6.
62. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. Preprint at bioRxiv. https://doi.org/10.1101/088716.
63. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44*, W58–W63. https://doi.org/10.1093/nar/gkw233.
64. Schraiber, J.G. (2018). Assessing the relationship of ancient and modern populations. Genetics *208*, 383–398. https://doi.org/10.1534/genetics.117.300448.
65. https://github.com/pysam-developers/pysam.
66. 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. https://doi.org/10.1534/genetics.112.147330.
67. Chintalapati, M., Patterson, N., and Moorjani, P. (2022). The spatiotemporal patterns of major human admixture events during the European Holocene. Elife *11*, e77625. https://doi.org/10.7554/eLife.77625.
68. Kandoth, C. (2020). vcf2maf v1.6.19. https://doi.org/10.5281/zenodo.593251.
69. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.
70. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756. https://doi.org/10.1101/gr.239244.118.
71. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19. https://doi.org/10.1016/j.cell.2022.08.004.
72. Rubinacci, S., Delaneau, O., and Marchini, J. (2020). Genotype imputation using the positional burrows wheeler transform. PLoS Genet. *16*, e1009049. https://doi.org/10.1371/journal.pgen.1009049.
73. Sarkisian, T., Ajrapetian, H., Beglarian, A., Shahsuvarian, G., and Egiazarian, A. (2008). Familial Mediterranean fever in Armenian population. Georgian Med. News *156*, 105–111.
74. Shoham, N.G., Centola, M., Mansfield, E., Hull, K.M., Wood, G., Wise, C.A., and Kastner, D.L. (2003). Pyrin binds the PSTPIP1/CD2BP1 protein, defining familial Mediterranean fever and PAPA syndrome as disorders in the same pathway. Proc. Natl. Acad. Sci. USA *100*, 13501–13506. https://doi.org/10.1073/pnas.2135380100.
75. Lawson, D.J., van Dorp, L., and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat. Commun. *9*, 3258. https://doi.org/10.1038/s41467-018-05257-7.
76. 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. https://doi.org/10.1038/ncomms9912.
77. Delser, P.M., Krapp, M., Beyer, R., Jones, E.R., Miller, E.F., Hovhannisyan, A., Parker, M., Siska, V., Vizzari, M.,T., Pearmain, E.J., et al. (2021). Climate and mountains shaped human ancestral genetic lineages. Preprint at bioRxiv. https://doi.org/10.1101/2021.07.13.452067.
78. Greppin, J.A.C., and Diakonoff, I.M. (1991). Some effects of the Hurro-Urartian people and their languages upon the earliest Armenians. J. Am. Orient. Soc. *111*, 720–730.
79. Hewsen, R.H. (2008). Armenia Maritima: The historical geography of Cilicia. In Armenian Cilicia (Mazda Publishers), pp. 27–66.
80. 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. https://doi.org/10.1016/j.ajhg.2019.03.015.
81. Fernandes, D.M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., et al. (2020). The spread of steppe

and Iranian-related ancestry in the islands of the western Mediterranean. Nat. Ecol. Evol. *4*, 334–345. https://doi.org/10.1038/s41559-020-1102-0.

82. Günther, T., and Nettelblad, C. (2019). The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genet. *15*, e1008302. https://doi.org/10.1371/journal.pgen.1008302.

83. Skourtanioti, E., Jia, X., Tavartkiladze, N., Bitadze, L., Shengelia, R., Tushabramishvili, N., Neumann, G.U., Bianco, R.A., Mötsch, A., Prüfer, K., et al. (2024). The Genetic History of the South Caucasus from the Bronze to the Early Middle Ages: 5000 years of genetic continuity despite high mobility. Preprint at bioRxiv. https://doi.org/10.1101/2024.06.11.597880.