

Genetic Structure of the Armenian Population

Levon Yepiskoposyan¹ · Anahit Hovhannisyan¹ · Zaruhi Khachatryan¹

Received: 24 June 2016 / Accepted: 20 October 2016 / Published online: 12 January 2017
© L. Hirszfeld Institute of Immunology and Experimental Therapy, Wrocław, Poland 2016

Abstract Located at the crossroads of Europe and the Middle East, the Armenian Highland served as a transition corridor for major waves of prehistoric and historic migrations. The genetic history of Armenians as an indigenous population of the region attracts keen scientific interest to resolve the puzzle of ancient Middle Eastern populations' expansion and the spread of Indo-European languages. Here, we review the current state of studies on the genetic structure of both modern and ancient inhabitants of the Armenian Highland and outline further steps to be fulfilled in this regard.

Keywords Armenian Highland · Armenian population · Genetic structure

Introduction

The origin of Armenians, a population with a strong ethnic identity and speaking an Armenian language, belonging to a separate branch within the Indo-European linguistic family, is difficult to reconstruct due to several reasons (Dolukhanov et al. 2004). The vast majority of valuable historical evidence has been destroyed or lost during numerous invasions and earthquakes in the Armenian homeland. Besides, the modern political division of historical Armenia among neighbouring states has made archival, anthropological and archaeological research a sensitive and often difficult task. Moreover, for most of the

period from the 4th century AD to the end of the 20th century AD Armenia has been subjected to the hegemony of more powerful neighbors. Forced and voluntary dispersions over the years have led to large worldwide Armenian Diaspora communities with varying densities in ethnically diverse neighborhoods.

Located at the intersection of Europe and the Near East, the Armenian Highland has acted as a focal area for multicultural fusion (Lang 1980) (Fig. 1). The historical expanse of Armenia comprised roughly 235,000 square miles of land between the Caspian, Black and Mediterranean seas. In prehistoric times, it provided a fertile home for the proliferation of a diverse set of tribes throughout the Armenian Highland that included speakers of the earliest Proto-Indo-European and Caucasian languages. Linguistic analyses indicate that the Armenian language shares the greatest structural similarity with the Indo-European languages of Greece and partly the Balkans. Furthermore, of the living Indo-European languages, Armenian appears to be the oldest and, therefore, is potentially the closest living relative of the ancestral Proto-Indo-European dialects (Bouckaert et al. 2012).

The genetic study of Armenians, based on blood groups and other phenotypic traits, started about a half-century ago. Over the past few years, the genetic structure of the Armenian population investigated on the basis of DNA markers has become a subject of keen scientific interest, as it is believed that the origin of Armenians may have important implication for the theories of ancient Middle Eastern populations' expansion and the spread of Indo-European languages. Using various genetic systems, it was shown that the genetic makeup of Armenians has distinct peculiarities reflecting the complex history of their origins, migrations and contacts with other populations.

✉ Levon Yepiskoposyan
lepiskop@gmail.com

¹ Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia



Fig. 1 Geographic localization of the Republic of Armenia

Paternal Genetic Structure

A high level of Y-chromosomal genetic variability was observed in geographically different Armenian populations. These studies revealed significant genetic divergence of populations from the mountainous southern (Syunik) and eastern (Karabakh) areas of the Armenian Highland compared to those of the Ararat valley and the northern, western and central parts of the region (Weale et al. 2001). This diversity could be explained by the geographic isolation of the Karabakh and Syunik regions which led to decreased genetic contacts with neighboring populations. Moreover, the level of inter-Armenian regional stratification is even higher than that found in some comparisons between various ethno-national groups in Europe. It was also shown that the cultural tradition of patrilocality had a significant impact on the population's genetic makeup, as the Armenian gene pool is regionally more structured based on Y-chromosomal markers than for those transmitted through the maternal line (Harutyunyan et al. 2009).

In the context of genetic relationships with other ethnic groups of the region, the Y-chromosomal structure shows a closer genetic affinity of Indo-European-speaking Armenians with their geographic neighbors in the Caucasus than

with their linguistic counterparts elsewhere (Nasidze et al. 2004). Based on this, it was suggested that the Armenian language was introduced in the region via a language replacement event. However, this preliminary inference should be further tested based on biparentally inherited genetic systems and using of more powerful tools and methods.

The study of the Y-chromosomal haplogroup distribution within geographically different Armenian groups shed light on the question of possible routes of migration for the first farmers from the Near East. It was shown that all Armenian populations are characterized by the greatest prevalence of the Y-chromosomal lineages associated with the arrival of the Neolithic agriculturalists from the Near East, including the J2, R1b1a2, G2a and J1-derived haplogroups (Herrera et al. 2012; Hovhannisyan et al. 2014). Wide presence of the Neolithic genetic markers, coupled with the scarcity of Paleolithic archaeological artifacts, is thought to indicate a repopulation of the Armenian Highland by early farmers from the Near East. Based on the distribution pattern of the Neolithic genetic markers, namely the haplogroups J2, R1b1a2 and G, it was shown that the Armenian Highland is the “homeland” for the bifurcation of divergently directed paths of agriculturalists’

movement from the Near East towards Europe and the Caucasus (Hovhannisyan et al. 2014). The mountainous plateau had served as a transitional corridor for at least two distinct pathways of migration for the first farmers from the Levant. Specifically, the westward movement took place mainly via the western region of the Armenian Highland and Anatolia, alongside the coastline of the Mediterranean Sea, while the northward migration occurred predominantly across the eastern regions of the plateau. Moreover, the Armenian Highland served as a connection bridge for another distinct Neolithic wave of bidirectional movement from the Near East to Europe and the North Caucasus.

To elucidate the possible impact of host populations, a few Armenian communities living beyond the state borders of present-day Armenia were genetically studied. Among them are the Don Armenians of the Rostov-on-Don region of southwestern Russia, with complicated migration background and about a millennium-long separation from their historical homeland. Though this population could preserve a very strong national and linguistic identity, the Don Armenians' paternal genetic composition consists of signals of recent genetic influence from the neighboring Slavic populations (Chukhryaeva et al. 2015).

The origin of another Armenian Diaspora community, Hamshenis, has been a controversial subject for scholars of various disciplines throughout the years. Three regions were considered as a putative homeland for this isolated ethnic group, namely Eastern Armenia, Western Armenia and Central Asia. However, the results of the genetic survey based on the Y-chromosomal markers indicated the central part of historical Armenia as a plausible place of origin for the Hamsheni population (Margaryan et al. 2011).

Maternal Genetic Structure

In contrast with the Y-chromosomal studies, genetic surveys of Armenians based on mitochondrial DNA (mtDNA) markers are scarce. In agreement with the Y-chromosomal diversity pattern, mtDNA variation in Armenians exhibits greater impact of linguistic rather than geographic barriers as the limitation factor for genetic interactions (Nasidze et al. 2004; Schönberg et al. 2011). In particular, close genetic similarities were observed between Indo-European speaking Armenians, Turkic-speaking Azeri and Caucasian-speaking Georgians. Furthermore, an extraordinary high genetic diversity of Armenians was detected based on high-throughput sequencing of complete mtDNA genomes (Schönberg et al. 2011). Most of the mtDNA haplogroups observed in the general Armenian population (H, HV, J, K, N, T, U and X) are typical for the West Eurasia, while others considered as Central and Eastern Asian ones (A, C,

D and F) are absent in the population's maternal gene pool. Specifically, the most encountered Armenian mtDNA haplogroups are the lineages U and H, which display relatively high frequencies in the Caucasus. The lack of Central and Eastern Asian genetic influence on the Armenians was further supported by the results of the Y-chromosomal haplogroup distribution analysis (Herrera et al. 2012; Hovhannisyan et al. 2014).

Autosomal Composition

In the early stages, the autosomal structure of the Armenian populations was studied based only on a limited number of markers, while recently the genome-wide approach is becoming more applicable tool in population genetic surveys. Based on the Alu insertion polymorphisms variation, Armenians and other populations of the Caucasus have closer genetic affinity with European populations rather than with Western Asian ones (Nasidze et al. 2001). The same pattern was found while studying mtDNA diversity (Nasidze et al. 2004). However, the results of further genetic surveys based on the Y-chromosomal markers sharply contrast with this pattern, showing higher genetic similarities of Armenians with the Near Eastern than with the European populations (Herrera et al. 2012; Hovhannisyan et al. 2014; Nasidze et al. 2004). It is worth noting that in the aforementioned Alu insertion polymorphisms study there was a limited number of specimens from the Middle East populations, which potentially might influence, via sampling bias, the results obtained.

Marked regional stratification of different Armenian populations was shown based on HLA data, being consistent with the spatial distribution of the Y-chromosomal lineages (Margaryan et al. 2011). Thus, it was revealed that Iranian Armenians, characterized by a specific cultural and ethnic background, appeared to be genetically separated from the neighboring populations of the country.

Currently, a few genome-wide autosomal studies of modern Armenian populations have been performed (Behar et al. 2010; Haber et al. 2016; Yunusbayev et al. 2012). It was shown that Armenians have an excess of a Near/Middle Eastern genetic proportion compared with other studied Caucasus populations; due to this peculiarity the modern Armenians occupy an intermediate position between the Levant and Caucasus clusters on the genetic landscape of the region (Haber et al. 2016; Yunusbayev et al. 2012). Statistical analysis of modern genetic data indicated the absence of any significant admixture events for Armenians over the past 4000 years (Haber et al. 2016; Hellenthal et al. 2014), and thus justifies using the population as a reference group for addressing the issue of prehistorical migrations from the Near East. Recent genetic

survey revealed that Armenians and other genetic isolates in the Near East (Cypriots, Near Eastern Jews and Arabic-speaking Christians) have higher genetic affinity to Neolithic European farmers than other present-day Near Eastern ethnic groups do (Haber et al. 2016). Thus, in comparison to other present-day Near Eastern populations, the genetic isolates preserve the features of the ancient regional genetic landscape. More specifically, it was assessed that Armenians have about 30% of their ancestry derived from a source population represented by Neolithic Europeans. In general, these results point to the importance for genetic studies of reproductively isolated populations, such as Armenians.

At present, the first study of ancient DNA from Armenia was performed by Allentoft et al. (2015), who sequenced eight Bronze Age samples from the eastern region of the Armenian Highland. It was shown that the admixture pattern in the population of Yamnaya, pastoral nomads who migrated into Europe from the Eurasian Steppe ~4500 years ago, is detectable only when using Bronze Age Armenians and the samples from the Upper Paleolithic site in Central Siberia (Mal'ta) as source populations. Moreover, there was a strong indication of persistent genetic continuity between the ancient inhabitants of the Armenian Highland and modern populations of the region, at least since the Bronze Age.

Conclusion

Though the Armenian population is genetically studied more thoroughly than the majority of its neighbors, numerous issues remain unclear. First, only a limited number of distinct geographic groups originally inhabiting the whole stretch of the Armenian Highland have been studied: these include mainly the populations representing the eastern and central parts of the region. This drawback especially refers to the ancient DNA samples, since our archaeological excavations are confined to current state borders of Armenia. Second, the potential genetic impact of hosting populations on the Armenian Diaspora communities is not comprehensively investigated. These studies are expected to clarify to what extent and how the cultural, linguistic and religious features spur or hamper genetic contacts of Armenians with the neighboring populations. And finally, the issue of possible genetic continuity between the ancient inhabitants of the Armenian Highland and modern Armenians should be explored using

aDNA samples from varying prehistorical and historical strata.

References

- Allentoft ME, Sikora M, Sjögren KG et al (2015) Population genomics of Bronze Age Eurasia. *Nature* 522:167–172
- Behar DM, Yunusbayev B, Metspalu M et al (2010) The genome-wide structure of the Jewish people. *Nature* 466:238–242
- Bouckaert R, Lemey P, Dunn M et al (2012) Mapping the origins and expansion of the Indo-European language family. *Science* 337:957–960
- Chukhryaeva M, Dibirova Kh, Yepiskoposyan L et al (2015) Genofond donskikh armyan: stoletiya v diaspore. Proceedings of the II International Conference “Armyane yuga Rossii: Istoriya, kul'tura, obschee budushee”. Rostov-on-Don, 26–28 May 2015, pp. 81–84 (in Russian)
- Dolukhanov P, Aslanyan S, Kolpakov E et al (2004) Prehistoric sites in Northern Armenia. *Antiquity* (<http://antiquity.ac.uk/projgall/dolukhanov301/>). Accessed 2 June 2016
- Haber M, Mezzavilla M, Xue Y et al (2016) Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. *Eur J Hum Genet* 24:931–936
- Harutyunyan A, Khudoyan A, Yepiskoposyan L (2009) Patrilocal and recent migrations have little impact on shaping patterns of genetic structure of the Armenian population. *Genetica* 125:1125–1131
- Hellenthal G, Busby GB, Band G et al (2014) A genetic atlas of human admixture history. *Science* 343:747–751
- Herrera KJ, Lowery RK, Hadden L et al (2012) Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists. *Eur J Hum Genet* 20:313–320
- Hovhannisyan A, Khachatryan Z, Haber M et al (2014) Different waves and directions of Neolithic migrations in the Armenian Highland. *Investig Genet* 5:15
- Lang DM (1980) Armenia: cradle of civilization. Allen & Unwin, London, p 320
- Margaryan A, Harutyunyan A, Khachatryan Z et al (2011) HLA allele and haplotype frequencies in the Armenian population: evidence of mediterranean origin. *Int J Antropol* 26:195–210
- Nasidze I, Risch GM, Robichaux M et al (2001) Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *Eur J Hum Genet* 9:267–272
- Nasidze I, Ling EY, Quinque D et al (2004) Mitochondrial DNA and Y-chromosome variation in the Caucasus. *Ann Hum Genet* 68(Pt 1):205–221
- Schönberg A, Theunert C, Li M et al (2011) High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur J Hum Genet* 19:988–994
- Weale ME, Yepiskoposyan L, Jager RF et al (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. *Hum Genet* 109:659–674
- Yunusbayev B, Metspalu M, Järve M et al (2012) The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol* 29:359–365