

8-1-2012

Paternal Lineage Analysis Supports An Armenian Rather Than A Central Asian Genetic Origin Of The Hamshenis

Ashot Margaryan

Institute of Molecular Biology, National Academy of Sciences

Ashot Harutyunyan

Institute of Molecular Biology, National Academy of Sciences

Zaruhi Khachatryan

Institute of Molecular Biology, National Academy of Sciences

Armine Khudoyan

Institute of Man, Yerevan, Armenia

Levon Yepiskoposyan

Institute of Man, Yerevan, Armenia, lepiskop@yahoo.com

Recommended Citation

Open access pre-print, subsequently published as Margaryan, Ashot; Harutyunyan, Ashot; Khachatryan, Zaruhi; Khudoyan, Armine; and Yepiskoposyan, Levon (2012) "Paternal Lineage Analysis Supports an Armenian Rather Than a Central Asian Genetic Origin of the Hamshenis," *Human Biology*: Vol. 84: Iss. 4, Article 8.

Available at: http://digitalcommons.wayne.edu/humbiol_preprints/2

Paternal lineage analysis supports an Armenian rather than a Central Asian genetic origin of the Hamshenis

Ashot Margaryan¹, Ashot Harutyunyan¹, Zaruhi Khachatryan¹, Armine Khudoyan² and Levon Yepiskoposyan^{1*}

¹Institute of Molecular Biology, National Academy of Sciences

²Institute of Man, Yerevan, Armenia

*Corresponding author (7 Hasratyan Str., 0014, Yerevan, Armenia;

Tel: +374 10 281626, fax: +374 10 282061; email: lepiskop@yahoo.com)

Key words: Y chromosome, microsatellite haplotypes, Hamshenis, genetic anthropology

Abstract

The Hamshenis are an isolated geographic group of Armenians with a strong ethnic identity who, until the early decades of the twentieth century, inhabited the Pontus area on the southern coast of the Black Sea. Scholars hold alternative views on their origin, proposing eastern Armenia, western Armenia and Central Asia, respectively, as their most likely homeland. To ascertain whether genetic data from the non-recombining portion of the Y chromosome is supportive any of these suggestions, we screened 82 Armenian males of the Hamsheni descent for 12 biallelic and 6 microsatellite Y-chromosomal markers. These data were compared with the corresponding datasets from the representative populations of the three candidate regions. Genetic difference between the Hamshenis and other groups is significant and backs up the hypothesis of the Armenian origin of the Hamshenis, indicating the central historical Armenia as a homeland of the ancestral population. This inference is further strengthened by the results of admixture analysis, which does not support the Central Asian hypothesis of the Hamshenis origin. Genetic diversity values and patterns of genetic distances suggest a high degree of genetic isolation of the Hamshenis consistent with their retention of a distinct and ancient dialect of the Armenian language.

Introduction

The Hamshenis represent an Armenian-speaking group, which is distinguished by the use of a particular dialect (Homshetsma) of the Western Armenian language. Presently they mainly inhabit the Black Sea coastal areas of Turkey, Russia, and Georgia and also have smaller communities in Armenia and Central Asia (the latter being the consequence of forced migration from Soviet Georgia in 1944). Together the Russian and Georgian populations consist of approximately 150,000 individuals (Kuznetsov 1995). The Hamshenis of Turkey and Central Asia practice Islam, while those in Russia, Georgia, and Armenia are members of the Armenian Apostolic Church. The Christian Hamshenis studied in this paper display a strong and distinct ethnic-cultural identity (Simonian 2006).

Opinions concerning the geographic origin of the Hamshenis are contradictory and entail three alternative hypotheses: a) eastern Armenia (an area roughly covered by the present state of Armenia), b) western Armenia (presently eastern Turkey), and c) Central Asia.

The hypothesis of the eastern Armenian origin is based on medieval historical records. In 791 AD Prince Shapuh Amatuni and his son Hamam, together with 12,000 Armenian families, left Oshagan (Eastern Armenia) in order to escape the oppressive regime of the Abbasids (Khachikyan 1969). The Byzantine emperor settled them around the town of Tampour in the Pontus region, between the Black Sea and the Pontic Mountains. The name Tampour was changed to Hamamshen (built by Hamam), then to Hamshen and, finally, to the current Hemshin (Torlakian 1981). The province of Hamshen flourished and was virtually separated from other Armenian populations developing its unique dialect of the Armenian language (Khachikyan 1969).

No historical records covering the period between the late eight-century and the 13th century have been preserved. The province retained its sovereignty until the late 1480s when it fell under the Ottoman rule (Simonian 2006). As a consequence, in the 16th century and, on a larger scale, in the 18th century, a significant number of the Hamshenis were forced to convert to Islam (Khachikyan 1969). The majority of those who refused to convert fled and settled in the eastern coast of the Black Sea belonging to the Russian Empire, mainly in the districts of Batum, Abkhazia, Sochi and even Crimea. The settlers preserved their Christian lifestyle and other aspects of their cultural identity (Kuznetsov 1995).

Backers of the western Armenian origin argue that the founding of Hamshen was subsequent to the destruction of the medieval Armenian capital Ani by the Seljuk Turks in 1064 AD. A group of fugitives from Ani is said to have found refuge in the forests of Hamshen, ‘which until then had never seen any human face’ (Torlakian 1981). An alternative account of the western Armenian origin claims that Hamshen was founded by the migrants from Ispir (Figure 1) (Khachikyan 1969). Similarities between the dialects of the Khodorchur Valley of Pertakrag and Hamshen are supportive of this point of view (Vaux 2007). On the other hand, such similarities could be a reflection of contacts between the two areas throughout centuries rather than an indication of a common origin (Simonian 2007). Based on ancient historical records, Vardanian (1999) suggests that the history of Armenian population of Hamshen began much earlier, during the reigns of kings Menua, Arguishti and Sardur II, rulers of Urartu in the eighth century BC. Thus the establishment of Hamshen was a consequence of their successive invasions leading to the conquest of the northern coastal regions of Anatolia and Pontus (Redgate 2000).

Supporters of the Central Asian origin hypothesis, modern Turkish historians (see Benninghaus 2007), suggest that Muslim Hamshenis (Hemshinli) are the descendants of Oghuz Turks from Central Asia. According to this account, the ancestors of the Hemshinli originated in Khorasan and settled in the region of Hamadan in the mid 3rd century BC. During the rule of the Armenian king Ardashes II (Arsacid in Turkish sources) they moved from Hamadan in Persia to the areas of Oshagan and Aparan in Armenia. The name of the ruling family of this Turkish tribe 'Amad-Uni' is said to reflect their long stay in Hamadan. The Amatunis (derived from Amad-Uni) migrated to the valley of the Çoruh River in 604 AD (Figure 1) and later, in 620 AD, to the region of Hamshen. The leader of the migrants the 'İlbeği Hamam Beg' rebuilt the destroyed city of Dampur/Tambur and renamed it Hamamashen in his own honour.

To clarify the genetic origin of the Hamshenis we investigated their Y-chromosomal markers. This approach has been widely used as a powerful tool in population genetics to analyze genetic relationships in various ethnic groups. In many communities the paternal line of biological continuity is an important element for preservation of ethnic and cultural identity (Haber et al. 2011; Balanovsky et al. 2011; Lowery et al. 2011; Yunusbayev et al. 2012; Weale et al. 2001; Thomas et al. 2008; Bowden et al. 2008; King and Jobling 2009). Hence the paternally inherited non-recombining portion of the Y chromosome (NRY) is frequently used in studies of geographical structuring of different human populations. The main goal of this study was to test the current hypotheses of the Hamshenis origin using genetic tools such as the paternally inherited Y-chromosomal markers.

Materials and Methods

Subjects and collection of DNA

The study subjects were selected to represent the Hamshenis and the three putative areas of interest. After obtaining informed consents, the mouth swabs were collected anonymously from adult males, unrelated at the paternal-grandparental level.

Hamsheni samples were collected from 82 residents of two villages, Novomichailovskiy (n=19 out of a total population of 10,600) and Tenguinka (n=63 out of a total population of 2,500), located in the coastal area of the Black Sea in the Krasnodar region of Russia. The subjects identified themselves as direct descendents of the Hamshenis from the town Janik in northeastern Turkey. Most of the ancestors of this group moved to Russia in 1915 from their settlements in the Samsun province (westward from the original Hamshen area) (Khachikyan 1969). According to the historical accounts, earlier they had migrated to Samsun in the 17th century AD to escape the forced conversion to Islam.

Eastern Armenian samples: Data used were from a previously published paper by Weale et al. (2001) as well as from the newly collected samples. The latter were collected from the Armenian males residing in the isolated mountainous regions of Syunik (n=296) in Republic of Armenia.

Western Armenian samples: We used two groups as possible source populations. The first group, SWA (n=96), was from western Armenia (currently in Turkey), which represents southwestern historical Armenia. The second group, CA (n=145), consisted of the patrilineal descendants of Armenians from the central regions of historical Armenia (cities of Erzurum and Kars).

Central Asia: Y-chromosomal typing results of 39 Uzbek males (general population - students of Tashkent University, of Uzbek ethnicity), as a representative population of Central Asian origin, were kindly provided by Dr. Neil Bradman from TCGA, UCL.

Typing of Y chromosome polymorphisms

DNA was isolated using a standard phenol-chloroform extraction. The strategy adopted for typing of samples was designed to ensure informative comparison with the existing published data. NRY were characterised by 12 binary Y chromosome polymorphisms including 92R7, M9, M13, M20, sY81, SRY+465, SRY4064, SRY10831, Tat, M17, *Alu* insert – YAP, and p12f2 (Weale et al. 2001; Rosser et al. 2000). NRYs were also screened for six microsatellite markers, DYS19, DYS388, DYS390, DYS391, DYS392, and DYS393 according to Thomas et al. (1999). Haplogroups (hg) were defined by the UEP markers according to the updated Y Chromosome Consortium nomenclature (Karafet et al. 2008). Microsatellite repeat numbers were assigned according to the nomenclature of Kayser et al. (1997).

Statistical analysis

Pairwise F_{ST} genetic distances were estimated from analysis of molecular variance (AMOVA) Φ_{ST} values with the aid of Arlequin program v. 3.5 (Excoffier and Lischer 2010). A bootstrapping method was used to calculate unbiased values of gene diversity index, h , as well as its confidence intervals, with the help of Octave package. Tests for significant population differentiation were carried out using the exact test for population differentiation (Raymond and Rousset 1995). Testing for differences in h between two populations was performed by using the Bayesian approach (Thomas et al. 2002)

carried out with statistical package 'R' (URL: <http://www.R-project.org/>). Principal Coordinates Analysis was conducted on similarity matrices based on F_{ST} genetic distances with the use of the GenStat program.

Signature haplotype analysis (high frequency modal haplotypes and modal clusters (Thomas et al. 1998; Thomas et al. 2000; Wilson et al. 2001) was performed with Microsoft Excel.

The admixture proportions of the three hypothesized populations (4 groups) in the Hamshenis gene pool were assessed using the ADMIX 2.0 program (Dupanloup and Bertorelle 2001). Estimates of the admixture coefficients and their SEs were calculated based on microsatellite data using 10,000 bootstrap replicates with the mutation rate of 2.7×10^{-5} per year as described by Zhivotovsky et al. (2004).

The Network 4.6.1.0 was used for median networks construction (available at <http://www.fluxus-engineering.com>) considering the 6 Y-STR loci (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393). The weight of each locus was assigned to be inversely proportional to its molecular variance. The reduced median procedure ($r = 2$) was followed by a median joining analysis ($\epsilon = 0$) to reduce the complexity of the network. The star contraction and maximum parsimony options were used for further simplification of the networks.

Results

NRV haplotypes

The 12 UEP markers typed for all samples identified nine haplogroups when considering all data sets and seven haplogroups in the Hamshenis only (Table 1). The most frequent (i.e. modal) haplogroup in the Hamshenis is BT*(xDE,JK) with a

frequency of about 44%, which is two- to four-fold higher than in comparative data sets. In south-western Armenian and central Armenian groups the modal haplogroup is J, which is also highly represented in the Hamshen (31.7%) and Syunik (37.17%) groups, whereas in the Uzbeks it has a frequency of only about 7.7%. The P*(xR1a) is the most frequently encountered haplogroup in the Syunik and Uzbek groups while having only about 8.5% frequency in the Hamshenis. According to the exact test values at the haplogroup level all the populations studied significantly differ from each other ($p < 0.05$).

The six microsatellites define a total of 231 compound UEP+microsatellite haplotypes the most frequent of which are shown in Table 2. The Hamshenis share their “modal haplotype” (15.9%) only with the group of central historical Armenia (CA) (2.8%), while it is completely absent in other groups. The modal haplotypes of the three Armenian groups of Syunik, SWA and CA are shared at various degrees among them but not detected in the Hamshen and Uzbek populations. The most frequent haplotype of the Uzbeks is present at negligible levels in Hamshen, Syunik, and SWA.

The Hamshenis as well as the other four groups also significantly differ from each other ($p < 0.01$) based on the exact test of population differentiation at haplotype level.

A similar pattern of distribution is observed when analysing modal clusters defined as a modal haplotype plus its one-step neighbours (Table 3).

Bootstrap gene diversity (h) value is the lowest in the Hamshenis (Figure 2) and is significantly different from the other four groups ($p < 0.05$) when using the Bayesian approach (data not shown). The highest h value is observed in SWA and it is

significantly different from the other groups ($p < 0.01$) when relying on the same approach.

Genetic distances

No significant genetic differences are detected between the two villages according to the exact test ($p > 0.05$) for the microsatellite markers. The Hamshenis modal haplotypes (MH) and modal clusters (MC) are represented in both communities (MH: Novomichailovskiy = 15.8%, Tenguinka = 15.9%; MC: Novomichailovskiy = 36.8%, Tenguinka = 28.6%). As expected, the inhabitants of the two villages display the lowest levels of genetic diversity (h) and are not significantly different ($p > 0.05$) thus implicating the long-term reproductive isolation of the Hamshenis in these two locations. The marriage pattern analysis shows that in both communities the rate of intra village marriages is very high reaching up to 98.5% in Tenguinka and 75% in Novomikhalovskiy. In addition, more detailed examination of birthplaces among the couples indicates that all the marriages related to our sampling (either intra or inter village) can be traced back to the descendents of the Janik Hamshenis.

The pattern of genetic proximity among the Hamshenis and four other groups was visualised using the results of Principal Coordinate Analysis based on F_{ST} values obtained from the UEP data (Figure 3a) or combined UEP and microsatellite haplotype frequency data (Figure 3b). All the populations studied are significantly different from each other based on either datasets. The UEP-based genetic distance (F_{ST}) is the smallest between the Hamsheni and SWA groups while the larger F_{ST} separates the Hamsheni and Uzbek populations. The same pattern is observed when assessing the F_{ST}

values among the groups based on combined UEP and microsatellite haplotype frequency data.

Admixture and Network analysis

According to the results of admixture analysis, the Uzbeks and south-western Armenians (SWA) had virtually no genetic contribution to the Hamsheni gene pool while contemplating the time-span of 1,000 years (postulated for the western Armenian origin of the Hamshenis), 1,300 years (foreastern Armenian origin), or 2,300 years (for Central-Asian origin) in the admixture calculations. The measurable genetic inflow from the Uzbek and SWA groups can only be perceived if the time of admixture event is set to 60,000 years for the Uzbeks and 8,000 years for the south-western Armenians. These results (i.e., much longer time periods than the hypothesised ones) clearly rule out the ideas of the Central-Asian or south-western Armenian origin of the Hamshenis.

Calculation of the admixture coefficients for two other possible scenarios with the involvement of the eastern and central-western Armenians as parental populations indicated that their relative contribution to the Hamsheni gene pool is estimated at 35% and 65%, respectively, if the admixture event is set at 2300 years (Figure 4). However, only the central-western Armenian genetic influence, supported by a 100% bootstrap admixture coefficient, could be detected in the Hamsheni population if the time of the admixture event is set at 1,300 years or less.

These implications are further supported by the results of the median network analysis for the two most frequently encountered Hamsheni haplogroups, BT*(xDE,JK) and J (Figure 5). The diagrams demonstrate that the Hamsheni haplotypes cluster

predominantly with the groups of the central historical Armenians and eastern Armenians while the Uzbek haplotypes are quite distant from the modal counterparts.

Discussion

This work is the first attempt to shed light on the genetic history of one of the enigmatic Armenian geographic groups, the Hamshenis, based on the analysis of the paternally inherited Y-chromosomal markers. The current hypotheses concerning the origin of this group were tested with different methods. The distribution of the frequently encountered (modal) haplotypes and clusters provided some support only for the western Armenian hypothesis (central historical Armenia), while the analysis of genetic distances between the populations did not give an unequivocal information to make any final conclusions about the origin of the Hamshenis. At the same time, the latter analysis suggested that genetically the Hamshenis are closest to the western Armenians and very distant from the Uzbeks. The observed patterns of genetic relationship were further strengthened by the results of admixture analysis, which clearly excluded any noticeable contribution from the Central Asian group to the origin of Hamshenis. It is worth mentioning here that this result was obtained using a 2,300 years of the supposed admixture point as suggested by the proponents of the Central Asian origin of the Hamshenis.

The conclusion that the inhabitants of the Hamsheni villages are the representative of the pre-1915 settlers of Janik prior to their forced migration is supported by the distribution of NRY haplotypes in these two villages, which are not statistically significant, and by the proportion of the HMH and HMC, which are similar in both villages. The low values of genetic diversity in the Hamshenis are reflective of

the matrimonial traditions in this sub-ethnic group. In particular, the rate of intra village marriages is quite high in both villages, as a result of the tradition to marry within the frames of the same sub-ethnic group.

We understand that although an important element of the cultural identity, the paternal line of biological continuity does not address the cultural continuity, maternal continuity or genome-wide similarity. It allows, however, reaching the conclusions concerning the origin of the Hamshenis that are broadly in agreement with the interdisciplinary evidences. In particular, the unique Hamshen dialect (Homshetsma), which belongs to the northeastern subgroup (Vaux 2007) of the western branch of the Armenian language (Acharian 1947) and which is spoken in the two villages, suggests both a western connection and an extended period of isolation. The Homshetsma dialect is one of the most archaic varieties of the modern Armenian language due to its extended isolation from the rest of the Armenian-speaking world as well as due to the limited exposure to the literary dialects (Vaux 2007). Another evidence from cultural anthropologists suggests the western Armenian origin of the traditional Hamshen garments, in addition to the presence of an ancient Pontic substrate (Kuznetsov 1995).

The admixture analysis allowed not only assessing the level of genetic contribution from other possible source populations to the Hamsheni genetic pool, but also considering different time frames thus making an important link with historical records. In summary, the results suggest that within a 1,000-2,300 years time-frame of the admixture event no detectable traces of Central Asian and south-western Armenian genetic contribution could be detected. Only the central and, to some extent, eastern Armenian groups might be considered as possible source populations for the Hamshenis. The similar pattern of genetic affinity between the groups studied is

observed in median network diagrams when assessing the relationship of STR haplotypes in the two most frequently encountered haplogroups (BT*(xDE,JK) and J) in the Hamshenis. This conclusion was somewhat expected considering the geographic proximity between the Hamshen area and the central region of historical Armenia. It is worth mentioning here, that we realize that the data set of Uzbek population reflects only a small portion of the Y chromosome diversity in Central Asia and further studies of different ethnic groups living in the region will provide more insights into the topic of this paper.

According to the 6 microsatellite data, the modal haplogroup BT*(xDE,JK) in the Hamshenis is most likely to be the haplogroup G based on the results of higher resolution Y-chromosomal typing (Cinnioğlu et al. 2004; Rootsi et al. 2012). The Hamshenis modal haplotype seems to be a quite unique one, as it was encountered only in 2 samples from Turkey (Cinnioğlu et al. 2004, with a frequency of 0.44%) and 2 Armenian samples (Rootsi et al 2012, with a frequency of 0.47%) all the four belonging to the haplogroup G1a. This haplotype is completely absent in any of Caucasian (Balanovsky et al. 2011; Nasidze et al. 2003; Bertoncini et al. 2012), Near-Eastern (El-Sibai et al. 2009) and other Armenian (Herrera et al. 2012) populations. It is quite interesting to note, that the two samples from Turkey bearing Hamsheni modal haplotypes were collected in the “region 3” (Cinnioğlu et al. 2004), which coincides with the Hamshenis homeland area near the city of Trabzon. High frequency rate (about 16%) of this “unique” haplotype in the Hamsheni population and its virtual absence in the Near-East and Caucasus, along with the overall low genetic diversity of Hamshenis, may indicate a well expressed founder effect in this specific Armenian subpopulation.

Our results do not exclude other chains of events involving, for example, several waves of migration to the Hamshen area from various regions of historical Armenia. To examine these scenarios, more geographically stratified possible source populations must be included in future analyses. A higher resolution typing with additional genetic markers such as mitochondrial and autosomal DNA would be also helpful to test different scenarios.

Undertaken within the context of alternative hypotheses drawn from other disciplines, this study demonstrates the importance of genetic analysis, which can make a significant contribution to our understanding of origin of ethnic groups, in particular, in the Caucasus area.

Acknowledgments

The authors thank Prof. Mark Thomas and Dr. Neil Bradman (both at University College of London) for their invaluable help in typing the DNA samples, helpful discussion and constant support of L.Y. and A.H. visits to the UK for carrying out the experimental part of the project.

References

1. Acharian, H. 1947. Study of the Hamshen dialect. Yerevan: Academy of Science Press, p.272.
2. Balanovsky, O., K. Dibirova, A. Dybo et al. 2011. Parallel evolution of genes and languages in the Caucasus region. *Mol. Biol. Evol.* 28(10):2905–2920.
3. Benninghaus, R. 2007. Turks and Hemshinli: manipulating ethnic origin and identity. In: Simonian H.H. (ed) The Hemshin: History, society and identity in the Highlands of Northeast Turkey. Routledge, London & NY, pp.353–388.
4. Bertoncini, S., K. Bulayeva, G. Ferri et al. 2012. The dual origin of tati-speakers from dagestan as written in the genealogy of uniparental variants. *Am. J. Hum. Biol.* 24(4):391–399.
5. Bowden, G.R., P. Balaesque, T.E. King et al. 2008. Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England. *Mol. Biol. Evol.* 25:301–309.
6. Cinnioğlu C., R. King, T. Kivisild et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* 114:127–148.
7. Dupanloup I., and G. Bertorelle. 2001. Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.* 18(4):672–675.
8. El-Sibai M., D.E. Platt, M. Haber et al. 2009. Geographical structure of the Y-chromosomal genetic landscape of the Levant: a coastal-inland contrast. *Ann. Hum. Genet.* 73:568–581.

9. Excoffier, L., and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10(3): 564–567.
10. Haber, M., D.E. Platt, D.A. Badro et al. 2011. Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. *Eur. J. Hum. Genet.* 19(3):334–340.
11. Herrera K.J., R.K. Lowery, L. Hadden et al. 2012. Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists. *Eur. J. Hum. Genet.* 20(3):313–320.
12. Karafet, T.M., F.L. Mendez, M.B. Meilerman et al. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18(5):830–838.
13. Kayser, M., A. Caglià, D. Corach et al. 1997. Evaluation of Y-chromosomal STRs - A multicenter study. *Int. J. Legal. Med.* 110(3):125–133, 141–149.
14. Khachikyan, L. 1969. Pages from the history of Hamshen Armenians. *Bulletin of Yerevan State University* 2:115–144.
15. King, T.E., and M.A. Jobling. 2009. Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. *Mol. Biol. Evol.* 26:1093–1102.
16. Kuznetsov, I.V. 1995. The clothes of the Armenians in the Pontus. The semiotics of material culture. Moscow: Publish House “Vostochnaya Literatura”, p.294.

17. Lowery, R.K., K.J. Herrera, D.A. Barrett et al. 2011. Regionalized autosomal STR profiles among Armenian groups suggest disparate genetic influences. *Am. J. Phys. Anthropol.* 146(2):171–178.
18. Nasidze, I, H. Schädlich and M. Stoneking 2003. Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. *Forensic Sci. Int.* 137(1):85–93.
19. Raymond, M., and F. Rousset 1995. An exact test for population differentiation. *Evolution* 49:1280–1283.
20. Redgate, A.E. 2000. The Armenians. Blackwell Publishers Ltd, Oxford.
21. Rootsi S., N.M. Myres, A.A. Lin et al. 2012. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur. J. Hum. Genet.* May 16. doi: 10.1038/ejhg.2012.86. [Epub ahead of print].
22. Rosser, Z.H., T. Zerjal, M.E. Hurles et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* 67:1526–1543.
23. Simonian, H. 2006. History and identity among the Hemshin. *Central Asian Survey* 25:157–178.
24. Simonian, H. 2007. Hamshen before Hemshin: the prelude to Islamicization. In: Simonian H.H. (ed) *The Hemshin: History, society and identity in the Highlands of Northeast Turkey*. Routledge, London & NY, pp. 19–41.
25. Thomas, G.M., N. Bradman, and H.M. Flinn. 1999. High throughput analysis of 10 microsatellite and 11 binary polymorphisms on the human Y-chromosome. *Hum. Genet.* 105:577–581.

26. Thomas, G.M., T. Parfitt, D.A. Weiss et al. 2000. Y chromosomes travelling south: the Cohen Modal Haplotype and the origin of the Lemba - the "Black Jews of Southern Africa". *Am. J. Hum. Genet.* 66:674–686.
27. Thomas, M.G., I. Barnes, M.E. Weale et al. 2008. New genetic evidence supports isolation and drift in the Ladin communities of the South Tyrolean Alps but not an ancient origin in the Middle East. *Eur. J. Hum. Genet.* 16:124–134.
28. Thomas, M.G., K. Skorecki, H. Ben-Ami et al. 1998. A genetic date for the origin of Old Testament Priests. *Nature.* 394:138–140.
29. Thomas, M.G., M.E. Weale, A.L. Jones et al. 2002. Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am. J. Hum. Genet.* 70:1411–1420.
30. Torlakian, B.G. 1981. Ethnography of Hamshen Armenians. In: Ethnography and Folklore, Yerevan, N13, pp.24–111.
31. Vardanyan S. 1999. Historical notes. In: Galstyan S., Karapetyan L. (eds.) Voice of Hamshen Armenians. Vernatun, Yerevan, pp.7–15.
32. Vaux, B. 2007. Homshetsma: the language of the Armenians of Hamshen. In: Simonian H.H. (ed.) The Hemshin: History, society and identity in the Highlands of Northeast Turkey. Routledge, London & NY, pp.257–278.
33. Weale, M.E., L. Yepiskoposyan, R.F. Jager et al. 2001. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethn-national group. *Hum. Genet.* 109:659–674.
34. Wilson, J.F., D.A. Weiss, M. Richards et al. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl. Acad. Sci. USA* 98:5078–5083.

35. Yunusbayev, B., M. Metspalu, M. Järve et al. 2012. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol. Biol. Evol.* 29(1):359–365.
36. Zhivotovsky, L.A., P.A. Underhill, C. Cinnioğlu et al. 2004. The Effective Mutation Rate at Y Chromosome Short Tandem Repeats, with Application to Human Population-Divergence Time. *Am. J. Hum. Genet.* 74(1):50–61.

Table 1. Y chromosome haplogroup frequencies in Hamshenis and four other groups

Haplogroup	Hamshen (n=82)	Syunik (n=296)	SWA (n=96)	CA (n=145)	Uzbek (n=39)
R1a1	0.0366	0.0473	0.0729	0.0138	0.2564
P*(xR1a)	0.0854	0.402	0.1667	0.3034	0.2564
BT*(xDE,JK)	0.439	0.0912	0.1979	0.2	0.1797
K*(xL,N1c,O2b,P)	0.0122	0.0405	0.1146	0.0483	0.1538
J	0.317	0.3717	0.375	0.3931	0.0769
L	0.061	0.0068	0.0208	0.0138	0.0256
N1c	–	–	–	0.0069	0.0256
DE*(xE)	–	–	–	–	0.0256
E*(xE1b1a)	0.0488	0.0405	0.0521	0.0207	–

Table 2. Modal haplotype frequencies in Hamshenis and four other groups

Haplogroup	Haplotype	Hamshen (n=82)	Syunik (n=296)	SWA (n=96)	CA (n=145)	Uzbek (n=39)
BT*(xDE,JK)	16 12 23 10 12 13	0.1585	–	–	0.0276	–
P*(xR1a)	14 12 24 11 13 12	–	0.1216	0.0208	0.0483	–
J	14 16 23 10 11 12	–	0.0135	0.0521	0.0069	–
J	14 15 22 10 11 12	–	0.0101	0.0313	0.1034	–
R1a1	16 12 25 11 11 13	0.0122	0.0068	0.0104	–	0.1026

Table 3. Modal cluster frequencies for Hamshenis and four other groups

Haplogroup	Haplotype	Hamshen (n=82)	Syunik (n=296)	SWA (n=96)	CA (n=145)	Uzbek (n=39)
BT*(xDE,JK)	16 12 23 10 12 13	0.3049	0.0034	–	0.0414	–
P*(xR1a)	14 12 24 11 13 12	0.0366	0.2568	0.0417	0.1931	0.0513
J	14 16 23 10 11 12	0.0732	0.0912	0.1146	0.1448	0.0256
R1a1	16 12 25 11 11 13	0.0244	0.0473	0.0417	0.0138	0.1282

Table 4. Absolute frequencies of Y chromosome UEP+microsatellite haplotypes in Hamshenis and comparative data sets

Hap. no.	Microsatellite haplotype	SWA (n=96)	CA (n=145)	Hamshen (n=82)	Syunik (n=296)	Uzbek (n=39)
Haplogroup P*(xR1a)						
1	14 12 24 11 13 12	2	7	0	36	0
2	14 12 24 10 13 12	0	12	1	3	0
3	14 12 24 11 14 12	1	5	2	5	0
4	14 12 24 11 13 13	0	0	0	11	1
5	14 12 23 11 13 12	0	0	0	15	0
6	14 12 24 10 14 12	1	2	0	4	0
7	14 12 25 11 14 12	0	4	0	2	0
8	14 13 23 10 10 14	1	0	1	3	0
9	15 12 24 11 13 12	0	3	0	3	0
10	13 12 22 11 15 13	0	0	0	0	1
11	13 12 23 10 12 12	0	0	0	0	1
12	14 12 19 10 13 13	0	0	0	0	1
13	14 12 19 11 13 13	0	0	0	0	1
14	14 12 22 11 13 13	0	0	0	0	1
15	14 12 23 10 10 14	0	0	0	0	1
16	14 13 24 10 10 15	0	0	0	0	1
17	14 13 24 11 13 12	0	0	0	0	1
18	15 12 25 11 12 11	0	0	0	0	1
19	13 12 25 10 13 13	1	0	1	0	0
20	14 12 24 10 13 13	1	1	0	1	0
21	14 12 22 11 13 12	1	1	0	0	0
22	14 12 23 10 14 12	1	1	0	0	0
23	14 12 25 11 13 12	1	1	0	0	0
24	15 12 25 11 13 12	0	0	0	5	0
25	13 12 25 11 13 12	0	2	0	0	0
26	14 12 23 10 13 12	0	1	0	2	0
27	14 13 23 10 09 14	0	0	1	0	0
28	17 12 23 10 10 13	0	0	1	0	0
29	13 12 22 10 15 13	1	0	0	0	0
30	13 14 22 10 16 13	1	0	0	0	0
31	14 12 22 11 14 13	1	0	0	0	0
32	14 12 23 11 14 13	1	0	0	0	0
33	15 12 24 10 14 12	1	0	0	0	0
34	15 12 24 11 14 12	1	0	0	0	0
35	13 12 25 10 13 12	0	0	0	3	0
36	14 12 24 12 14 12	0	0	0	3	0

37	12 12 24 10 14 12	0	1	0	0	0
38	15 12 24 10 13 12	0	1	0	0	0
39	15 12 24 10 13 13	0	1	0	0	0
40	15 13 24 10 13 13	0	1	0	0	0
41	14 12 23 10 13 13	0	0	0	2	0
42	14 12 24 10 13 11	0	0	0	2	0
43	14 12 25 10 13 12	0	0	0	2	0
44	14 14 24 11 13 12	0	0	0	2	0
45	12 12 25 11 14 12	0	0	0	1	0
46	13 12 24 10 14 12	0	0	0	1	0
47	13 12 24 11 13 12	0	0	0	1	0
48	14 11 23 11 13 12	0	0	0	1	0
49	14 12 23 08 13 12	0	0	0	1	0
50	14 12 23 12 13 12	0	0	0	1	0
51	14 12 24 10 13 14	0	0	0	1	0
52	14 12 24 11 13 11	0	0	0	1	0
53	14 12 24 12 13 12	0	0	0	1	0
54	14 12 24 12 13 13	0	0	0	1	0
55	14 12 25 10 13 11	0	0	0	1	0
56	14 12 25 11 13 13	0	0	0	1	0
57	15 13 23 10 10 14	0	0	0	1	0
58	15 13 23 10 13 13	0	0	0	1	0
59	16 12 22 10 11 13	0	0	0	1	0
Total (hg P*(xR1a))		16	44	7	119	10

Haplogroup N1c

60	14 12 23 11 16 14	0	0	0	0	1
61	15 12 23 11 14 14	0	1	0	0	0
Total (hg N1c)		0	1	0	0	1

Haplogroup BT*(xDE,JK)

62	16 12 23 10 12 13	0	4	13	0	0
63	15 12 22 10 11 14	2	10	0	1	0
64	15 12 23 10 12 13	0	0	7	0	0
65	16 12 23 11 12 13	0	2	5	0	0
66	15 12 21 10 11 15	0	6	0	0	1
67	16 14 25 10 11 13	0	0	0	0	2
68	15 12 21 10 11 14	1	1	0	7	0
69	15 12 23 10 11 14	0	1	2	2	0
70	15 13 24 10 11 13	2	0	0	3	0
71	14 12 23 10 11 15	0	0	0	0	1
72	15 14 25 10 11 13	0	0	0	0	1
73	15 15 23 10 11 13	0	0	0	0	1
74	16 13 25 09 11 13	0	0	0	0	1

75	15 13 21 10 11 14	1	0	0	4	0
76	15 12 22 09 11 14	2	0	0	0	0
77	15 12 24 10 11 13	2	0	0	0	0
78	15 12 23 10 11 13	0	3	0	0	0
79	15 12 21 10 11 13	1	0	0	2	0
80	15 12 22 10 11 13	1	0	0	1	0
81	15 13 25 10 11 14	1	0	0	1	0
82	16 13 24 11 11 14	0	2	0	0	0
83	14 12 23 10 10 13	0	0	1	0	0
84	14 13 24 11 11 13	0	0	1	0	0
85	15 12 23 11 12 13	0	0	1	0	0
86	15 12 24 10 12 13	0	0	1	0	0
87	15 13 24 10 11 14	0	0	1	0	0
88	15 13 25 11 11 14	0	0	1	0	0
89	16 12 23 11 12 14	0	0	1	0	0
90	17 12 22 10 10 14	0	0	1	0	0
91	17 12 24 10 12 13	0	0	1	0	0
92	14 12 23 10 14 12	1	0	0	0	0
93	15 12 21 11 11 14	1	0	0	0	0
94	15 12 22 09 11 15	1	0	0	0	0
95	15 13 24 11 11 13	1	0	0	0	0
96	16 12 22 10 10 13	1	0	0	0	0
97	17 12 22 10 10 13	1	0	0	0	0
98	14 14 23 10 11 13	0	0	0	2	0
99	12 12 22 10 11 14	0	0	0	1	0
100	16 12 23 10 10 13	0	0	0	1	0
101	16 12 23 10 12 14	0	0	0	1	0
102	17 12 22 10 11 13	0	0	0	1	0
Total (hg BT*(xDE,JK))		19	29	36	27	7

		<u>Haplogroup E*(xE1b1a)</u>				
103	14 12 24 09 11 13	0	0	3	0	0
104	13 12 24 10 11 13	1	2	0	3	0
105	13 12 23 10 11 13	1	0	0	1	0
106	14 12 25 09 11 13	0	0	1	0	0
107	13 12 24 09 11 13	1	0	0	0	0
108	13 12 24 10 11 14	1	0	0	0	0
109	14 12 25 10 11 13	1	0	0	0	0
110	13 12 26 10 12 13	0	0	0	3	0
111	15 12 25 07 11 13	0	0	0	3	0
112	13 12 24 10 12 14	0	1	0	0	0
113	12 12 24 10 11 13	0	0	0	1	0
114	13 12 24 10 13 13	0	0	0	1	0
Total (hg E*(xE1b1a))		5	3	4	12	0

Haplogroup K*(xL,N1c,O2b,P)

115	14 12 23 10 14 13	0	0	0	2	2
116	14 12 23 10 13 13	3	0	0	4	0
117	14 12 24 10 14 13	0	0	0	0	1
118	15 12 22 10 14 13	0	0	0	0	1
119	15 14 23 10 15 13	0	0	0	0	1
120	16 13 23 11 14 14	0	0	0	0	1
121	14 12 23 10 13 14	2	0	0	0	0
122	14 12 24 10 13 13	2	0	0	0	0
123	15 12 23 11 15 14	0	0	1	0	0
124	13 12 23 10 13 13	1	0	0	0	0
125	14 12 23 11 13 13	1	0	0	0	0
126	15 12 23 09 13 14	1	0	0	0	0
127	15 12 23 10 16 13	1	0	0	0	0
128	14 12 23 10 15 14	0	0	0	3	0
129	14 12 23 10 11 13	0	1	0	0	0
130	14 12 24 11 13 13	0	1	0	0	0
131	14 12 24 11 13 14	0	1	0	0	0
132	15 12 23 10 11 13	0	1	0	0	0
133	15 12 23 10 13 13	0	1	0	0	0
134	15 12 23 10 14 13	0	1	0	0	0
135	15 12 23 10 15 13	0	1	0	0	0
136	14 12 25 10 13 13	0	0	0	1	0
137	15 12 23 10 15 14	0	0	0	1	0
138	16 12 23 10 13 13	0	0	0	1	0
Total (hg K*(xL,N1c,O2b,P))		11	7	1	12	6

Haplogroup L

139	15 12 23 10 13 11	1	0	4	0	0
140	15 12 22 10 15 12	0	0	0	0	1
141	14 12 24 10 11 13	0	0	1	0	0
142	15 12 24 10 13 11	1	0	0	0	0
143	15 12 22 10 14 12	0	1	0	0	0
144	16 12 23 10 13 11	0	1	0	0	0
145	14 12 24 10 14 12	0	0	0	1	0
146	15 12 24 10 15 11	0	0	0	1	0
Total (hg L)		2	2	5	2	1

Haplogroup R1a1

147	16 12 25 11 11 13	1	0	1	2	4
148	15 12 25 11 11 13	1	0	0	12	1
149	15 12 24 11 11 13	0	0	0	0	1
150	15 12 27 11 12 13	0	0	0	0	1

151	16 09 25 10 11 13	0	0	0	0	1
152	16 12 25 10 11 12	0	0	0	0	1
153	17 12 25 10 11 13	0	0	0	0	1
154	17 12 25 11 11 13	1	0	1	0	0
155	16 12 25 10 11 13	0	2	0	0	0
156	14 12 25 11 11 13	0	0	1	0	0
157	15 12 25 11 11 14	1	0	0	0	0
158	15 12 27 11 11 12	1	0	0	0	0
159	16 12 24 10 11 13	1	0	0	0	0
160	16 12 24 11 11 13	1	0	0	0	0
Total (hg R1a1)		7	2	3	14	10

Haplogroup DE*(xE)

161	15 12 25 10 07 13	0	0	0	0	1
Total (hg DE*(xE))		0	0	0	0	1

Haplogroup J

162	14 15 23 10 11 12	3	9	6	16	0
163	14 15 22 10 11 12	3	15	0	3	0
164	15 15 23 11 11 12	0	0	7	1	0
165	14 15 23 11 11 12	2	2	3	3	0
166	14 16 23 10 11 12	5	1	0	4	0
167	15 16 23 10 11 12	1	3	0	0	1
168	14 14 24 10 11 12	1	2	0	1	1
169	14 15 23 11 11 13	0	0	3	0	0
170	13 15 23 11 11 12	3	0	0	1	0
171	14 17 23 11 11 12	3	0	0	1	0
172	15 15 23 10 11 12	2	0	0	4	0
173	14 13 23 10 11 12	1	2	0	3	0
174	14 13 22 10 11 12	0	4	0	2	0
175	14 16 23 09 11 12	0	4	0	2	0
176	14 16 24 10 11 12	1	3	0	0	0
177	14 15 24 10 11 12	2	0	0	3	0
178	14 15 23 09 11 12	0	3	0	2	0
179	14 14 24 11 11 12	0	0	0	0	1
180	14 17 23 10 11 12	1	0	0	4	0
181	14 17 24 10 11 12	0	0	0	7	0
182	15 14 23 10 11 12	2	0	0	0	0
183	14 14 23 10 11 12	0	0	0	5	0
184	14 14 22 10 11 12	0	2	0	0	0
185	15 16 22 09 11 12	0	2	0	0	0
186	14 14 25 10 11 12	0	0	0	4	0
187	14 16 23 11 11 13	0	0	0	4	0
188	15 15 25 10 11 12	0	0	0	4	0

189	14 13 23 11 11 12	0	0	1	0	0
190	14 14 23 11 11 12	0	0	1	0	0
191	14 15 22 10 11 13	0	0	1	0	0
192	14 15 24 11 11 12	0	0	1	0	0
193	14 17 23 10 11 13	0	0	1	0	0
194	14 18 23 11 11 13	0	0	1	0	0
195	15 15 24 12 11 12	0	0	1	0	0
196	14 13 23 10 11 13	1	0	0	0	0
197	14 15 22 11 11 13	1	0	0	0	0
198	14 15 26 10 11 12	1	0	0	0	0
199	14 18 24 10 11 11	1	0	0	0	0
200	15 16 24 09 11 12	1	0	0	0	0
201	15 16 24 11 11 12	1	0	0	0	0
202	14 15 24 10 11 13	0	0	0	3	0
203	15 15 22 10 11 12	0	0	0	3	0
204	15 15 24 10 11 12	0	0	0	3	0
205	13 17 23 10 11 12	0	1	0	0	0
206	14 15 23 10 11 13	0	1	0	0	0
207	14 16 22 10 11 12	0	1	0	0	0
208	15 16 23 09 11 12	0	1	0	0	0
209	16 16 23 10 11 12	0	1	0	0	0
210	13 15 23 10 11 12	0	0	0	2	0
211	13 18 24 10 11 12	0	0	0	2	0
212	15 14 25 11 11 12	0	0	0	2	0
213	15 15 23 10 11 13	0	0	0	2	0
214	15 16 22 10 11 12	0	0	0	2	0
215	14 12 23 10 11 12	0	0	0	1	0
216	14 13 24 11 11 13	0	0	0	1	0
217	14 14 23 10 11 13	0	0	0	1	0
218	14 14 24 10 11 13	0	0	0	1	0
219	14 15 25 10 11 12	0	0	0	1	0
220	14 15 25 11 11 12	0	0	0	1	0
221	14 16 22 10 12 12	0	0	0	1	0
222	14 16 23 11 11 12	0	0	0	1	0
223	14 16 24 08 11 12	0	0	0	1	0
224	14 17 24 10 11 13	0	0	0	1	0
225	14 17 24 10 12 12	0	0	0	1	0
226	15 13 23 10 11 12	0	0	0	1	0
227	15 15 23 09 11 12	0	0	0	1	0
228	15 15 24 09 11 12	0	0	0	1	0
229	15 15 24 11 11 13	0	0	0	1	0
230	16 14 22 10 11 12	0	0	0	1	0
231	16 15 23 10 11 12	0	0	0	1	0
Total (hg J)		36	57	26	110	3

Figure 1. Map of Armenia and neighboring regions. SWA –south-western historical Armenia; CA –central historical Armenia

Figure 2. Bootstrap gene diversity, h , indexes with \pm SEs across Hamshen and four comparative data sets (notions as on Figure 1)

Figure 3. Principal coordinates plots (first two axes) based on F_{ST} values calculated from (a) haplogroup data only and (b) UEP + microsatellite data. $PCO1$ first axis, $PCO2$ second axis. *Numbers in brackets* indicate the percentage of total variation explained by each axis (notions as on Figure 1)

Figure 4. Relative genetic contribution of four hypothesized populations to the Hamsheni gene pool. The values represent the admixture coefficients while considering 2300 years to the supposed admixture event (notions as on Figure 1)

Figure 5. Median network diagrams of Y-STR haplotypes of haplogroups BT*(xDE,JK) (a) and J (b). Circle sizes are proportional to the haplotype frequencies with the smallest circles representing singletons. The branch lengths are proportional to the number of STRs separating the nodes (notions as on Figure 1)