Journal of Cell and Molecular Research (2014) 6 (1), 29-37

Research Article

Paternal genetic affinity between Iranian Azeris and neighboring populations

Ardeshir Bahmanimehr[1,3*] and Fatemeh Nikmanesh[2]

1. *Biotechnology Department, Marvdasht branch, Islamic Azad University, Marvdasht, Iran*
2. *Biotechnology and Bioengineering Department, School of Advanced Medical Science, Tabriz University of Medical Sciences, Tabriz, Iran.*
3. *Institute of Molecular Biology, National Academy of Sciences of Armenia, 07 Hasratian St.,Yerevan, Armenia*

Received 06 Feb 2014 Accepted 18 March 2014

Abstract

In certain environments such as Iran highlands, major innovations in lifestyle, as the emergence of agriculture and domestication of animals, are thought to have led to population expansions. Historical studies showed that at some point in history (from the third to the first millennium BC) dramatic changes have been taken place on the Iranian plateau. To trace the genetic affinity between the Iranian Azeris and neighboring populations, 297 samples were collected from northwest of Iran. The non-recombining portion of the Y chromosome (NRY) was genotyped at the unique event polymorphism (UEP) levels, using 48 single-nucleotide polymorphism (SNP) markers, based on the human NRY tree. According to our results, like other Iranian ethnic groups, Iranian Azeris showed a heterogeneous paternal genetic structure. Low genetic distances were also found between Iranian Azeris and their contemporary geographical neighbors. They also have preserved minor share of Y-haplogroup of central Asian ancestry tracts in their genomes, which is in agreement with the historical period of major Turkic migrations. The strategic feature for northwest of Iran to transfer and carry important ancient migratory events and gene flow across the Asia and the Europe also conducive conditions for sedentary habitation leading to sharp demographic growth in the area is supported by all molecular and statistical analysis of this study.

***Keywords*:** Iranians; Y-chromosome genomes; phylogeny; population expansions; genetic history

Introduction

The modern and interdisciplinary science, Molecular anthropology, is defined as the use of molecular genetics techniques encompasses the analysis and interpretation of; molecular genetic variation in various patterns contemporary human populations that anthropologists are interested in concerning human evolution and diversity. By examining molecular genetic structure in different populations, molecular anthropologists can figure out how closely related those populations are. Certain similarities in genetic makeup let to determine whether or not different groups of people belong to the same haplogroup, and thus if they share a common geographical origin. This is significant because it allows anthropologists to trace patterns of migration and settlement, which gives helpful insight as to how contemporary populations have formed and progressed over time.

Y-chromosome DNA documents the paternal lineage and becoming a useful tool for tracing human evolution through male lineages. Since the Y-chromosome is passed down from father to son without any recombination, can provide unique insights into the human past. It's long no recombining segment carries the most informative stable haplotypes in the genome, whereas its permanent location in the male genome links these to male specific history.

The Iranian gene pool at different times has been an important source of the Near Eastern and Eurasian Y-chromosome variability as well as a recipient of variation entered with different migratory events (Grugni et al., 2012). Y-chromosomal studies of the modern Iranian populations are indicative of the past settlements and migrations in the Middle East overall, shaping its contemporary patrilineal genetic landscape (Quintana-Murci et al., 2001). The complexity of the Iranian male gene pool is described by previous studies where some of the Iranian groups fall within the Near East and South Asian clusters. Different factors could have contributed to the observed Iranian population heterogeneity, in particular, the presence of important geographic barriers such as the Zagros and Alborz Mountain ranges and the two arid areas, the Dasht-e Kavir and the Dash-e Lut deserts.

∗Corresponding author E-mail:
Bah_ard@yahoo.com

Both types of barriers, running from North-West to South-East, have limited gene flow from neighbouring regions and free movements of internal peoples, starting from the first peopling of this area (Grugni et al., 2012; Nasidze et al., 2008; Yunusbayev et al., 2011; Derenko et al., 2013). Among various geographic features of the territory, Northwest of Iran is a unique and strategic geography due to the proximity to the Fertile Crescent, Mesopotamia, ancient Silk Road passages and its conductive positions of Neolithic agricultural diffusion.

Iranian Azeris are an indigenous population in the Northwest of Iran. The heterogeneous paternal genetic structure in this population also debates regarding the ethnic origins of the Azeris in the modern Iran, requires a reasonable comparative genetic study of Iranian Azeri and other ethnic groups of Iran as well as neighbouring population around the country. Furthermore the current analysis for this paper provides increased resolution based on a larger sample numbers, a collection on deep area and native peoples, additional control populations, higher levels of haplogroup resolution to trace the genetic affinity between the Iranian Azeris and studied populations.

Materials and Methods

The Y chromosome consortium has established a system of defining Y-DNA haplogroups by letters A through to T, with further subdivisions using numbers and lower case letters. Each of these haplogroups originated to certain geography and population and due to human community events such as migrations, nowadays haplogroups are frequent in different geography and ethnic groups. In this method, by finding out frequency of each haplogroup in the population, it is possible to discuss its origin and genetic relation with neighbors.

A total of 297 samples was collected from unrelated healthy male Azeris from three provinces (17 cities) of northwestern Iran (113 samples from the East Azerbaijan province, 140 from the Ardebil province and 44 samples from the West Azerbaijan province). Genomic DNA was extracted from whole blood by using the QIAamp DNA Mini Kit (Qiagene, Hilden, Germany).

UEP and STR genotyping were performed using 48 SNPs based on the human NRY tree published by the Y - chromosome consortium (ISOGG 2011) by PCR -RFLP analysis, Taqman assay (Applied Biosystems, Foster City, CA, USA) or direct sequencing. A set of 48 relevant bi-allelic markers that represent Near and Middle Eastern populations was selected and done base on published conditions (Hammer et al., 2001). Description of main and new Y-chromosome binary markers used in this study are in the Table 1.

A hierarchical strategy for genotyping of the NRY-UEPs was followed using deep-rooting markers of the NRY phylogeny. All samples were genotyped using markers M74, M89 and M9 to define the superclades P, F and K. Then, each sample was systematically genotyped according to the different clades within F, K or P for its final haplogroup designation. Specification for most markers was reported on ISOGG 2011 (http://www.isogg.org/tree).

Table 2. Characteristics of the studied population for Y chromosome dataset.

Region (Population)		COD	Size	Ref
IRAN	Iran- Assyrian	IR-ASY	48	11
	Iran-Persian	IR-PER	160	
	Iran-Gilak	IR-G	64	
	Iran-Turkmen	IR-TU	68	
	Iran-Hormozgan	IR-HBA	143	
	Iran-Qeshmi	IR-HQ	49	
	Iran-Khuzestan(Arab)	IR-ARB	57	
	Iran- Kurd	IR-KUR	59	
	Iran-Lur	IR-LU	50	
	Iran-Mazandarani	IR-MAZ	72	
	Iran-Baluch	IR-BAL	24	
	Iran-Zoroastrian	IR-ZO	47	
	Iranian Azeries	**IR-AZ**	**297**	*
Afghanistan	Afghanistan-Total	AF	190	14
	Afghanistan- Hazara	AF-HA	60	13
	Afghanistan-Pashtun	AF-PA	49	
	Afghanistan-Tajik	AF-TAJ	56	
	Afghanistan-Uzbek	AF-UZ	17	
Iraq	Iraq (Marsh Arab)	IQ-MAR	143	1
	Iraq/Baghdad	IQ-BA	154	
	Pakistan	PAK	176	26
	Armenia-Syunik	ARM-SY	105	**
	Turky Total	TUR	523	5

* This study
** Not published data

A subset of 90 samples belonging to Haplogroups R1b, J1, G and C was also typed for short tandem repeats of Y-chromosome (Y-STRs) using DYS388 (Kayser et al., 1997) and DYS461 (White et al., 1999) as well as 17 STR markers including DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439, DYS385a, DYS385b, DYS437, DYS438, DYS448, DYS456, DYS458, DYS635, Y_GATA_H4 by Yfiler Kit (Applied Bios-AmpFlSTR, Life technologies, California, USA) based on manufacturer's recommendations. The results were analyzed using the ABI PRISM program Gene Mapper 4.0 (Applied Biosystems).

Same population data from 2314 samples of neighbouring area (Table. 2) was used as a dataset. Genetic Fst distances and gene diversity indices were estimated by Arlequin version 3.5 (Excoffier et al., 2010). Tests for significant population differentiation were carried out using the exact test for population differentiation (Raymond et al., 1995). Similarity matrices based on Fst genetic distances were used to perform principal coordinates analysis (PCO) by GenStat version 14.2. PHYLIP version 3.6 (Felsenstein, 2004) was used to reconstruct neighbor-joining (NJ) tree. Y-STR haplotypes were used to compare populations in R1b, J1e, G and C haplogroups. Primer sequences and concentrations of STR markers presented in Table.3.

Table 3. Primer sequences and concentrations for STR markers genotyping kit.

Primer name	Primer sequence	Dye label	Final concentration (µM)
DYS19-L	CTA CTG AGT TTC TGT TAT AGT	TET	0.236
DYS19-R	ATG GCA TGT AGT GAG GAC A		0.236
DYS388-L	GTG AGT TAG CCG TTT AGC GA	TET	0.318
DYS388-R	CAG ATC GCA ACC ACT GCG		0.318
DYS390-L	TAT ATT TTA CAC ATT TTT GGG CC		0.127
DYS390-R	TGA CAG TAA AAT GAA CAC ATT GC	FAM	0.127
DYS391-L-N[a]	CTA TTC ATT CAA TCA TAC ACC CAT AT	FAM	0.384
DYS391-R-N[a]	ACA TAG CCA AAT ATC TCC TGG G		0.384
DYS392-L-N[a]	AAA AGC CAA GAA GGA AAA CAA A		0.155
DYS392-R-N[a]	CAG TCA AAG TGG AAA GTA GTC TGG	HEX	0.155
DYS393-L	GTG GTC TTC TAC TTG TGT CAA TAC		0.180
DYS393-R	AAC TCA AGT CCA AAA AAT GAG G	HEX	0.088

Results and discussions

We observed an almost similar distribution of the main Y-haplogroups (J2, R1b, R1a and G) in our studied ethnic group and neighbouring populations which is in agreement with previous findings (Grugni et al., 2012; Çinnioğlu et al., 2004; Rootsi et al., 2012) on a westward diffusion of J2-M410*, J2-PAGE55*, J2- M530, G-M201* and R1b-M269* haplogroups and pre-agricultural expansions from the Iranian plateau toward Europe via Caucasus and Turkey.

In this regard all Iranian ethnic groups, including Azeris have shared J2, R1a and G haplogroups with highest frequencies (Supplementary Table 2) while the most frequent haplogroups were J2 and R1in Afghanistan and Pakistan; J2, E1b1 and J1e in Iraq; and J2, G and R1b in Armenia and Turkey.

Distributions of Y-haplogroups

J2-M172 was the most prevalent modal haplogroup in our studied ethnic group. This haplogroup was also reported as the most common haplogroup in the Caucasus, the Fertile Crescent, Anatolia, the Balkans, Italy, the Mediterranean littoral and the Iranian plateau (Semino et al., 2004). The concordance of Iranian Azeris with neighboring populations in this haplogroup frequency, considering the origin of haplogroup J2 from the Middle East (more than 30 KYA) (Nasidze et al., 2008; Di Cristofaro et al., 2013) might suggest geographic distribution of J2-M172 in this part of the world. This issue also agrees and has strong correlation with the diffusion of agriculture from northern Mesopotamia also supported and well documented in the Neolithic archaeological record (Cauvin, 2000; Simone et al., 2013).

The paragroup J2a-PAGE55* was the most frequent of J2 lineage in Iranian Azeris. This paragroup was estimated to be distributed 10.4 KYA in Northwest of Iran, 14.5 KYA in South of Iran, Hormozgan, and 15.5 KYA in the center of Iran (Grugni et al., 2012; Kushniarevich et al., 2013). It is indicative of the distribution of this haplogroup in the Northwest of Iran after other parts of the country and it represents the signature of ancient migratory events in this area which might consider as westward diffusion of this haplogroup and pre-agricultural expansions from the Iranian plateau.

Inverse scattering of R1a and R1b haplogroups has been observed in all studied populations except Iranian Azeris. R1a was more frequent than R1b in Iran, Afghanistan and Pakistan while R1b was more frequent than R1a in Armenia, Turkey and Iraq. South Asian component, R1a, and Asian lineage, R1b, are distributed equally in Iranian Azeris. It is indicative of the merit territory in the Northwest of Iran to transfer and carry important different ancient migratory events such as the recolonizations of Indo-European nomads in the North of Iran, which possibly, linking the spread of R1a to the movement of Kurgan people from North of the Caspian Sea and farmers migration from Near East to Europe during the Paleolithic and the Last Glacial Maximum or in the dispersal process (Karafet et al., 2008; Sikora et al., 2013; Underhill et al., 2014) also varying degrees of demic diffusion and cultural diffusion of R1b

lineage. In aggregate, it emphasizes the conspicuous role of a fertile region in the North of Iran as a gateway for gene flow, of different haplogroups, through the geographical barriers in the West and East of Iran.

The issue also is in aggregation of autochthonous Middle Eastern haplogroup J1-M267 branches, J1e (Page08) and J1-M267*, that display opposite distribution in almost all studied populations except Iranian Azeris.

The haplogroup J1e (Page08), reported likely originated in the border between southeastern Turkey and North parts of Iraq (Grugni et al., 2012; Al-Zahery et al., 2011; Chiaroni et al., 2009), underwent an important Neolithic expansion in the southern countries of the Middle East and represents one of the principal haplogroups in the modern populations of the Arabian Peninsula and North Africa. Nonuniform distribution of this lineage in different geographical locations of Iran, particularly in both sides of the Zagros Mountains, indicates a possible barrier role of geographical boundaries as Zagros Mountains which hamper the flow of this linage to the other sides of Iran.

Haplogroup J1-M267* shows high variance in the Middle Eastern region including Eastern Turkey, North-West Iraq (Rootsi et al., 2012; Semino et al., 2004; Haber et al., 2013) and North-West Iran, where probably originated and then migrated westwards to the Balkans and the Italian Peninsula and southwards as far as in Saudi Arabia and Ethiopia (Grugni et al., 2012). The proportion of these two sub-lineages is highly variable in Iran. J1-M267* is almost restricted among northwestern Iranian ethnic groups and J1-Page08 is mainly observed in populations living below the Dasht-e Kavir and Dasht-e Lut deserts. It reaches a frequency of 32% among Arabs of Khuzestan on the southern border of Iraq (Wells et al., 2001; Di Giacomo et al., 2004), while in Iranian Azeris both lineages are in close portion which is suggested the origin of these lineages.

Despite the general similarity according to the dispersal of major haplogroups in studied populations, some minor haplogroups in Iranian Azeris seems entered from the eastern or northern neighbors: Central Asian components (Q, C and O) which are frequent in Eastern countries as Afghanistan and Pakistan, also European lineage, I, which is more frequent in Armenia and Turkey, comparing to Iranian Azeris, and haplogroup J1e-Page55 that frequently reported in western neighbors of Iran, as Iraq, in addition south Asian component (H, R2, L and N2) that is frequent in Afghanistan and Pakistan.

Detected gene diversity values, *h*, in most of the examined population, lies within the similar high range (0.9-0.8) except the low diversity exhibited by the Iranian Turkmen, Iraqi Marsh and Afghani Pashto (**Figure1**).

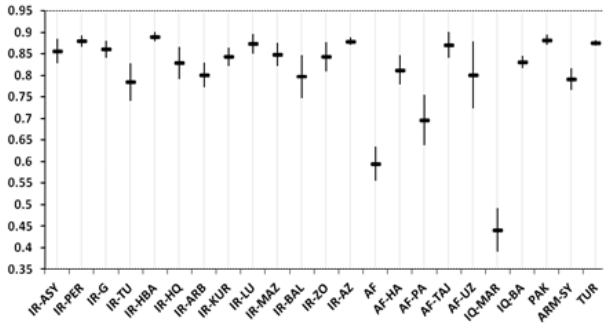

Figure 1. Genetic diversity (*h*) values, based on the Y-chromosome haplogroup frequency, with standard deviations in the studied populations.
X axis: estimated genetic diversity by Arlequin version 3.5 .
Y axis: studied Populations. Abbreviations used are described in the Table 2.

The similar range of diversity among most Iranian ethnic groups is in agreement with previously observed studies (Wells et al., 2001; Semino et al., 2004; Quintana-Murci et al., 2004; Rafiee et al., 2009) and is in concordance with Population Pairwise differences in *h* values (Supplementary Table 3).

No significant difference has been observed among Iranian ethnic groups in Pairwise differences in *h* values, except Turkmens that significantly differ from Azeris, Persians and Iran-Hormozgan. In this regard, Iranian Azeris showed a high level of gene diversity (0.8796) which was significantly different in comparison with the Arab and Turkmen ethnic groups of Iran, Afghan Pashto, both Iraqi sub-populations and Armenian Syunic ($P<0.05$). The high gene diversity values, *h*, detected throughout Iran as a multi-ethnic and multi-linguistic region points to its central role as a strategic multidirectional gateway intersecting three continents and four major linguistic families (Quintana-Murci et al., 2001). This pattern is particularly seen in the west part of the Zagros Mountains, inhabited by Iranian Azeris, Lurs, Arabs and Kurds. It is consistent with an early settlement of the fertile region by modern humans followed by subsequent migration routes across the world. It also possibly explains the peculiarity of the Northwest of Iran as a corridor for ancient human migration. The lowest level of genetic diversity is encountered in the Iranian Turkmen, Iraqi Marsh and Afghan Pashto, which reflects their long centuries of reproductive isolation due to their language, religion and other cultural features.

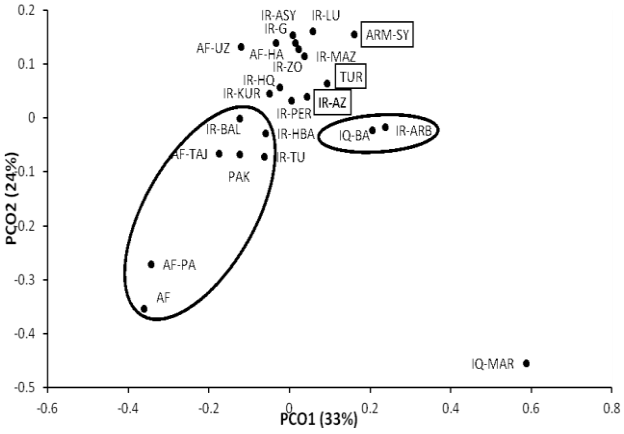

Figure 2. Principal coordinates (PCO) plot based on the F_{ST} values of studied populations using Y chromosome genotyping data. Numbers in brackets show the percentage of total variation explained by each axis. Populations close together in the circles have less genetic distance. Abbreviations used are described in the Table 2.

Genetic distance analysis

Using Y-chromosomal haplogroup frequencies we have assessed the genetic relationship between all populations under consideration by calculating the conventional F_{ST} genetic distances. The results are displayed in **Supplementary Table 4**, showing that based on the F_{ST} genetic distances Azeris, Lurs and Parsees of Iran are closer to each other and significantly ($p< 0.05$) differ from other ethnic groups of Iran and other studied populations based on the pairwise *P* value results. Interestingly, the Iranian populations are significantly far from all other neighbouring populations.

In order to depict the pattern of genetic affinities between Iranian Azeris and comparative data sets in a two-dimensional space, PCO analysis has conducted based on F_{ST} genetic distances (**Figure 2**). As shown, Iranian Arabs are clustered to other Arab populations (IQ-Mar and IQ-BA) which geographically are, also close to the west borders of Iran. Iranian Balochis and Hormozgan population are clustered to Eastern neighbors like Afghans and Pakistani population. In agreement to PCO plot, Pairwise F_{ST} values showed similarity between the Iranian Baloch and Afghan-TAJ. Iranian Azeris is located near the northwestern neighbors like Turks and Armenians and other Iranian ethnic groups in the north (Gilak and Mazandarani) as well as Lurs in the northwest of Iran. Paternal heterogeneity of Iranians also could infer geographical boundaries such as Zagros Mountains, Dasht-e-kavir and Dasht-e-Lut serve as important geographic barriers to gene flow from/to neighbouring regions.

To visualize the genetic relationship among Iranian Azeris and other populations of the area NJ tree was constructed based on FST genetic **(Figure 3).** These results, also demonstrates the correlation between the geographic proximity and genetic structures of these populations. As shown in the tree, all Iranian ethnic groups are located in a separate branch. Iranian Arabs show a common ancestor with Arab populations of Iraq, Iranian Baloch, Turkmen and the people of Hormozgan are located on the same branch of eastern neighbors, Afghanistan and Pakistan, and Iranian Azeris are located close to northwest neighbors, Turkey and Armenian populations.

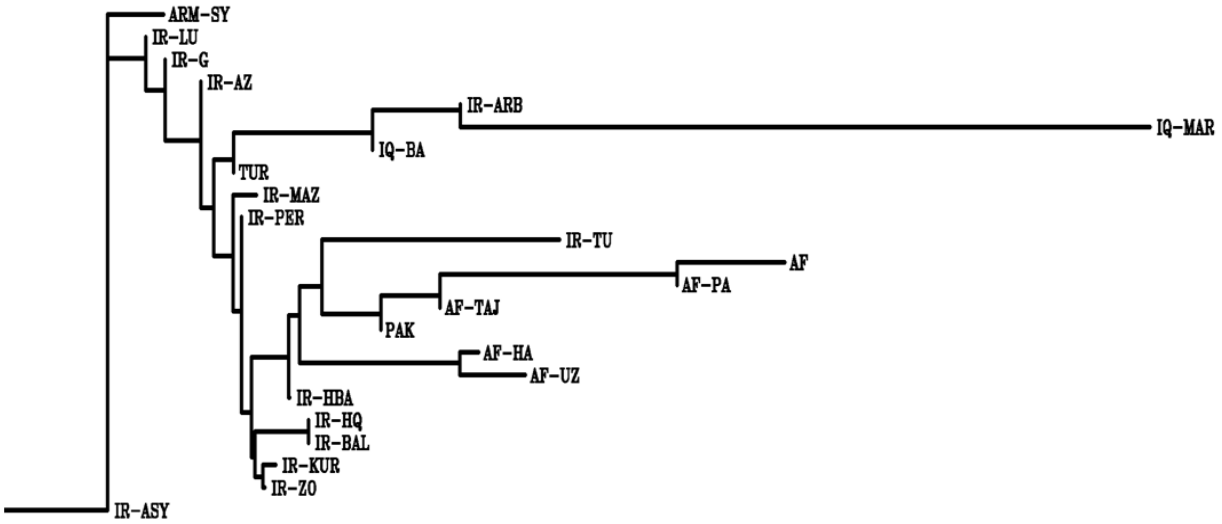

Figure 3. Rooted Neighbor-Joining dendrogram of the studied populations based on FST genetic distances of Y chromosome markers. Clustering of the distance between each pair of populations and root them according to their genetic distance. Populations on the same root have less genetic distance. Abbreviations used are described in the Table 2.

Microsatellite haplotype analysis

STR markers provide another level of differentiation of the populations. Following the aim of the study to find out the genetic affinity between the Iranian Azeris and other populations, in four major haplogroups (G-M201, J1, R1b-M343 and C-M130) results were analyzed at the haplotype level, as defined by STRs; DYS19, DYS389a, DYS389b, DYS390, DYS391, DYS392, DYS393 and DYS439. Only those haplotypes within the same haplogroup (i.e. Lineages) were analyzed together (Supplementary Table 5). For each haplogroup, haplotypes in the different population were compared with corresponding haplotype considering modal and shared haplotypes. Therefore, we present here the detailed table of all observed SNP+MS haplotypes and their frequencies in five populations (Supplementary Table 5) also encountered modal haplotypes for each population presented in Table 4.

Table 4. Frequently encountered Microsatellite (STR) haplotypes in the studied populations. STR markers used in these lotypes are DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439 respectively from left to right.

Haplogroup	**Microsatellite haplotype**	**Ir-Az** (*n=88*)	**Iranian** (*n=77*)	**Afghan** (*n=44*)	**Turkey** (n=195)	**Arm-Syu** (*n=68*)
R1b-M343	14 16 13 24 11 13 12 12	-	-	-	**0.036**	-
R1b-M343	14 13 15 24 11 14 12 13	**0.045**	-	-	-	-
R1b-M343	14 13 16 24 11 13 12 11	-	0.039	-	-	0.063
J1e-Pag09	14 13 16 23 10 11 12 11	**0.045**	0.026	-	-	0.016
J1e-Pag08	14 13 16 23 11 11 13 11	-	-	-	-	**0.078**
J1e-Pag08	14 13 17 23 10 11 12 11	-	**0.052**	-	-	0.016
G-M201	15 12 16 22 10 11 14 12	**0.045**	0.013	-	-	-
C-M130	16 13 16 25 10 11 13 10	-	-	**0.182**	-	-

The microsatellites together with mentioned SNP haplogroups defined 367 haplotypes in total; 69 haplotypes in Iranian Azeris, 64 in Iranians, 26 in Afghans, 143 in Turks and 42 in Armenians, which totally 23 haplotypes shared between populations (Supplementary Table 5).

The modal haplotypes in Iranian Azeris interestingly belong to haplogroups G-M201, R1b-M343 and J1e-Pag08 (equally 0.045) which for J1e, modal haplotype overlapped (also shared) by Iranian and Armenians and for G-M201, modal haplotype overlapped by Iranians. These results reveal a heterogeneous paternal genetic structure in Iranian Azeris. The modal haplotype of Afghanistan belongs to haplogroup C-M130 which originates from central Asia, did not observe in other populations.

Haplogroup J1e in haplotype level is modal and frequently observed in Iranian Azeris, Iranians and Armenian Syunic population. It is generally in agreement with previous reports that the haplogroup was dispersed by the westward movement of the people from the Middle East (Çinnioğlu et al., 2004; Renfrew, 1996; Simone et al., 2013; Behar et al., 2013) and might inference origin of J1e nearby eastern Anatolia or sough Caucasia.

Haplogroup R1b in haplotype level is modal in the Iranian Azeris and Turkey also frequently observed in the Iranians and Armenian Syunic population. Therefore, this issue provides a genetic signature of the Eurasia paternal gene pool, and bear witness to the expansion of this lineage across the continent after the Last Glacial Maximum and agreed previous reports that the haplogroup was originated in Asia and lies in Eurasia, most likely in Western Asia (Sikora et al., 2013).

Conclusion

In this work, we have shown the major mechanisms responsible for shaping the genetic structure of the modern Iranian Azeri population using Y-chromosomal markers which are sensitive tools in population genetics studies. Specifically, we have unmasked the major factors that have assigned strategic feature for Northwest of Iran.

The heterogeneous paternal genetic structure for the Iranian Azeris was statistically supported by different tests, particularly at the STR marker haplotype level. The Central Asian haplogroups have a notable contribution (6.4%) to the Iranian Azeri paternal gene pool. Hence the geographic location of northwest of Iran may have facilitated Mongol and other Turkic-speaking tribes from the Central Asian steppes in the thirteenth century CE therewith elite dominance model for Turkic

language dispersal have shaped the population structure of Northwest of Iran.

The high level of genetic diversity detected in Iranian Azeris is an evidence of the peculiarity of the region as a constituent part of the ancient Silk Road, as well as a settlement area for pre-Islamic Iranian people of the Central Asian origin.

In summary, our collection of samples and dataset, that cover the full extent of ethnic groups of Iran and neighboring population, shows that most Iranians proportionally contribute the majority of their genome with together and share varying minor proportion with their geographic neighbors, suggesting and emphasizing the importance of the Iranian Plateau as a source and recipient of gene flow between culturally and genetically distinct populations. Hence minor differences in genetic structure of the Iranian ethnic groups can be explained by taking into account their geographical locations in the territory.

References

1. Al-Zahery N., Semino O., Benuzzi G., Magri C., Passarino G., Torroni A. and Santachiara-Benerecetti AS. (2011)Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phylogenet Evol 28:458-472.
2. Behar D., Metspalu M., Yael B., Naama KM., Bayazit Y., Ariella G., Shay T., Sahakyan H and Bahmanimehr A.(2013) No Evidence from Genome-Wide Data of a Khazar Origin for the Ashkenazi Jews. Human Biology Open Access Pre-Prints Paper 41. 2013; http://digitalcommons.wayne.edu/humbiol_preprints/41.
3. Cauvin J. (2000) The Origins of Agriculture in the Near East: a Symbolic Interpretation (English edition). Cambridge University Press. Cambridge, 125pp.
4. Chiaroni J., King R.J., Myres N.M., Henn B.M., Ducourneau A., Mitchell M.J., Boetsch G., Sheikha I., Lin A.A., Nik-Ahd M., Ahmad J., Lattanzi F., Herrera R.J., Ibrahim M.E., Brody A., Semino O., Kivisild T. and Underhill P.A. (2009)The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations European. Journal of Human Genetics 18: 348–353.
5. Çinnioğlu C., King R., Kivisild T., Kalfoglu E., Atasoy S., Cavalleri G.L. and Lillie A.S. (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127-148.
6. Derenko M., Malyarchuk B., Bahmanimehr A., Denisova G and Perkova M. (2013) Complete Mitochondrial DNA Diversity in Iranians. PLoS ONE 8(11): e80673. doi:10.1371/journal.pone.0080673.
7. Di Cristofaro J., Pennarun E., Mazie`res S., Myres NM and Lin AA. (2013) Afghan Hindu Kush: Where Eurasian Sub-Continent Gene Flows Converge. PLoS ONE 8(10): e76748 doi:10.1371/journal.pone.0076748.
8. Di Giacomo F., Luca L. O., Popa N., Akar N., Anagnou J. and Banyko R. (2004)Y-chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Hum Genet 115:357–371.
9. Excoffier L. and Lischer H.E. (2010) Arlequin suite Ver 3.5: a new series of Programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564-567.
10. Felsenstein J. (2004) PHYLIP (Phylogeny Inference Package) version 3.6, Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.
11. Grugni V., Battaglia V., Hooshiar Kashani B., Parolo S., Al-Zahery N., Achilli A., Olivieri A., Gandini F., Houshmand M., Sanati M., Torroni A. and Semino O. (2012) Ancient Migratory Events in the Middle East: New Clues from the Y-Chromosome Variation of Modern Iranians. PLoS ONE7:7,:e41252.
12. Haber M., Gauguier D., Youhanna S., Patterson N and Moorjani P. (2013) Genome-Wide Diversity in the Levant Reveals Recent Structuring by Culture. PLoS Genet. 9:2; e1003316. doi:10.1371/journal.pgen.1003316.
13. Haber M., Platt D.E., Ashrafian Bonab M., Youhanna S.C and Soria-Hernanz D.F. (2012) Afghanistan's Ethnic Groups Share a Y-Chromosomal Heritage Structured by Historical Events. PLoS ONE 7:3. e34288. doi:10.1371/journal.pone.0034288.
14. Harlette L., Tenzin G., Regueiro M., Chennakrishnaiah S., Bukhari A., Underhill P.A. Garcia-Bertrand R.L. and Herrera R.J. (2012) Afghanistan from a Y-chromosome perspective. European Journal of Human Genetics 1–8. doi:10.1038/ejhg.2012.59.
15. Karafet T.M., Mendez F.L., Meilerman M.B., Underhill P.A., Zegura S.L. and Hammer M.F. (2008) New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. Genome Res 18:830-838.
16. Kayser M., Caglia A. and Corach D. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:141-149.
17. Kushniarevich A., Sivitskaya L., Danilenko N., Novogrodskii T and Tsybovsky I. (2013) Uniparental Genetic Heritage of Belarusians: Encounter of Rare Middle Eastern Matrilineages with a Central European Mitochondrial DNA Pool. PLoS ONE. 8(6): e66499. doi:10.1371/journal.pone.0066499.
18. Nasidze I., Quinque D., Rahmani M., Alemohamad S. and Stoneking M. (2008) Close genetic relationship between Semitic-speaking and Indo-European-speaking groups in Iran. Ann Hum Genet 72:241-252.
19. Quintana-Murci L., Chaix R., Wells S., Behar D., Sayar H., Scozzari R., Rengo C., Al-Zahery N., Semino O. and Santachiara-Benerecetti A. (2004) Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. Am J Hum Genet 74:827-845.
20. Quintana-Murci L., Krausz C., Zerjal T., Sayar S.H. and Hammer M.F. (2001) Y-chromosome lineages

trace diffusion of people and languages in southwestern Asia. Am J Hum Genet 68:537-42.

21. Rafiee M.R., Sokhansanj A., Naghizadeh M.A. and Farazmand A. (2009) Analysis of Y-Chromosomal Short Tandem Repeat (STR) Polymorphism in an Iranian Sadat Population. Russian Journal of Genetics 45:969-973.

22. Raymond M. and Rousset F. (1995) An exact test for population differentiation. Evolution 49:1280-1283.

23. Renfrew C. (1996) Languages families and the spread of farming. In: Harris DR (ed) The origins and spread of agriculture and pastoralism in Eurasia. Smithsonian Institution Press. Washington DC, USA, 217-224pp.

24. Rootsi S., Myres N.M., ALin A., Jarve M., King R.J., Kutuev I., Cabrera V.M., Khusnutdinova E.K., Varendi K., Sahakyan H., Behar D.M., Khusainova R., Balanovsky O., Balanovska E., Rudan P., Yepiskoposyan L., Bahmanimehr A., Farjadian S. and Villems R. (2012) Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. European Journal of Human Genetics 1:8, doi:10.1038/ejhg.2012.86.

25. Semino O., Magri C., Benuzzi G., Lin A.A., Al-Zahery N., Battaglia V., Maccioni L., Triantaphyllidis C., Shen P., Oefner P.J., Zhivotovsky L.A., King R., Torroni A., Cavalli-Sforza L.L., Underhill P.A. and Santachiara-Benerecetti A.S. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:5, 1023-34.

26. Sengupta S., Zhivotovsky L.A., King R., Mehdi S. Q., Edmonds C.A., Rani U., Thakur C.M., Cavalli-Sforza L. and Underhill P.A. (2006) Polarity and Temporality of High-Resolution Y-Chromosome Distributions in India Identify Both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists. Am. J. Hum. Genet 78:202–221.

27. Sikora M. J., Colonna V., Xue Y. and Tyler-Smith C. (2013) Modeling the contrasting Neolithic male lineage expansions in Europe and Africa. Investigative Genetics 4:25. doi:10.1186/2041-2223-4-25.

28. Simone R., Mohsen Z. and Nicholas J. C. (2013) Emergence of Agriculture in the Foothills of the Zagros Mountains of Iran. Science. 341: 65; DOI:10.1126/science.1236743

29. Simone R., Zeidi M. and Conard N.J. (2013) Emergence of Agriculture in the Foothills of the Zagros Mountains of Iran. Science 341:65, 64-67.DOI: 10.1126/science.1236743.

30. Underhill P.A., Poznik G.D., Rootsi S., Jarve M., Lin A.A., Wang J., Sahakyan H., Behar D.M., Kushniarevich A., Sarac J., Chaubey G., Grugni V., Semino O., Yepiskoposyan L., Bahmanimehr A., Farjadian S., Balanovsky O., Chiaroni J. and Villems R.(2014)The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. European Journal of Human Genetics. 1:8, 1018-4813/14

31. Wells R.S., Yuldasheva N., Ruzibakiev R., Underhill P.A. and Evseeva I. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244-9.

32. White P.S., Tatum O.L., Deaven L.L. and Longmire J.L. (1999) New male-specific microsatellite markers from the human Y-chromosome. Genomics 57:433-437.

33. Yunusbayev B., Metspalu M., Jarve M., Kutuev I., Rootsi S., Metspalu E., Behar D.M., Varendi K., Sahakyan H., Khusainova R., Yeppiskoposyan L., Khusnutdinova E.K., Underhill P.A., Kivisild T. and Villems R. (2011)The Caucasus as an asymmetric semi permeable barrier to ancient human migrations. Mol Biol Evol. 29:1,359–365.

Table 1. Description of some main and new Y-chromosome binary markers used in this study.

Common Name Marker	YCC nomenclature Haplogroup	Nucleotide change	Amplicon size (bp) reference sequence	Polymorphism position from 5' end of + strand	Restriction enzyme	Primer Forward 5'-3'	Primer reverse 5'-3'
M130	C*	C to T	205	41	BseL1	CTGCCCAGGGGAAAGGGCAT	CCACAAGGGGGAAAAAACAC
M78	E1b1b1a	C to T	301	197	AciI	GGGGTAACATTGGACATTCATTGCA	ATAGTGTTCCTTCACCTTTCCTT
M285	G1*	G to C	287	70	HphI	TTATCCTGAGCCGTTGTCCCTG	TGTAGAGACACGGTTGTACCCT
M485	G2a3*	C to T	312	150	Mva I	CTCATTTCCTCACATGTATGC	TTTAGGAATTACTATGTAGCGTC
M547	G2a3*	T to G	423	284	HpaII	AGAGATGGGTTTTCACCGTG	GCATAAATGTCAAGCCCACTAG
M461	G2a3a	C to T	329	114	BseYI	GCAGAAATGAAAGATGGCTG	TGAATCACACTACTCCCACG
M527	G2a3b1a1	C to T	327	128	Af1II	GTTTCATGGGAATAAACACTGG	AGTATCAAAGCACATGTGTTGC
M426	G2a3b1b	T to G	337	221	AlwI	ACTTAAACCTAAGTCATTTTGGGTG	GATCATCGGAAGTGACAGCC
M69	H	T to C	257	222	HpaII	AGCTTCAGGAGGCTGTTTACAC	AAAATATATTTCAGCAAGACAAAGG
M253	I1	C to T	400	283	Hinc II	GCAACAATGAGGGTTTTTTTG	CAGCTCCACCTCTATGCAGTTT
M267	J1*	T to G	287	148	Mva I	TTATCCTGAGCCGTTGTCCCTG	TGTAGAGACACGGTTGTACCCT
Page8	J1e	T to C	306	189	HpyF3I	ACGTCACCCATCTCAACATC	AAAGAATGTCTCCCCATGAGG
M67	J2a2*	A to T	409	377	AluI	GTGATGACAAACTCCCCTGC	GTCTTTTCACTTGTTCGTGGAC
M92	J2a2a	T to C	470	340	Eco1051	TTCAGAAACTGGTTTTGTGTCC	TTCAGAAACTGGTTTTGTGTCC
M242	Q	C toT	337	180	HphI	AACTCTTGATAAACCGTGCTG	TCCAATCTCAATTCATGCCTC
Z282	R1a1a1a*	T to C	297	155	AluI	GTTCTACAGGTTACAGGTTAGC	GGGAAACAAAAACATTCC
Z284	R1a1a1a1	C to G	275	176	Bg1II	GAGAATTTCAAAAATCATCC	GGGAAACAAAAACATTCC
M458	R1a1a1b1a1	A to G	380	87	BsrI	AGAAGAGATTTCTAGCCAGAGT	GGGGTAGAAAATTATTGGTC
Z280	R1a1a1b1a2	C to T	120	64	AlwI	GCATAATTACTGCTGTCATCTTCC	CAAAGGTCTTTACTTGTGCAATATC
M558	R1a1a1c*	T to C	281	211	AvaII	TGTTGGCTGGCCTCTCTC	GAACAAGGCAGTTGTAGGATAG
M582	R1a1a1c*	T to G	273	111	Tsp509I	GAGGCTGCAGTGAGCTATGAC	GTCACCTGCTTGGTAAAGATGAC
Z93	R1a1a2*	G to A	338	172	AluI	AACAAAGCATCATCAAAGGC	CATGATTCGTTATGACCTGC
Z95	R1a1b2a*	C to T	429	153	BsrI	TCTTTTCTGACTGGCCAGG	GGCTTATCTTTCTGTTTCCTGAAG
Z2125	R1a1b2a2*	C to T	895	284	HpyCH4 III	CCAAACCCCAGTGCCAGC	CCTAAGGCCAGGGAAGGCTC
M204	R1a1b2a2a	T to G	486	234	SfcI	AAGGGGCGAAGTATTCCAGAG	TGAAGAGGAGTCTGTTAGCCTG
M434	R1a1b2a2b	G to A	320	213	BseYI	CCAAAATTAGTGGGGAATAGT	GATCACCCAGGGTCTGGAGTT
M560	R1a1b2a3*	G to T	305	151	HphI	TGTAGATGATGGGTTAATGGGTG	GCACATAATATGTTTGAGAAGGC
M780	R1a1b2a4*	C to T	386	130	HpaII	GAAGATCCAAAACCTAAGAGAAC	GCTCAATGAGGAAGGCGATC
M70	T	A to C	257	45	SfcI		ATCTTTATTCCCTTTGTCTTGCT