

## The congruence between matrilineal genetic (mtDNA) and geographic diversity of Iranians and the territorial populations

Ardeshir Bahmanimehr <sup>1\*</sup>, Ghafar Eskandari <sup>2</sup>, Fatemeh Nikmanesh <sup>3</sup>

<sup>1</sup> Biotechnology Department, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran

<sup>2</sup> Molecular Biology, Razi Vaccine and Serum Research Institute, Ahvaz, Iran

<sup>3</sup> Biotechnology and Bioengineering Department, School of Advanced Medical Science, Tabriz University of Medical Sciences, Tabriz, Iran

### ARTICLE INFO

#### Article type:

Original article

#### Article history:

Received: Feb 20, 2014

Accepted: Jun 1, 2014

#### Keywords:

Iranian gene pool

Maternal lineage

Molecular anthropology

mtDNA

### ABSTRACT

**Objective(s):** From the ancient era, emergence of Agriculture in the connecting region of Mesopotamia and the Iranian plateau at the foothills of the Zagros Mountains, made Iranian gene pool as an important source of populating the region. It has differentiated the population spread and different language groups. In order to trace the maternal genetic affinity between Iranians and other populations of the area and to establish the place of Iranians in a broad framework of ethnically and linguistically diverse groups of Middle Eastern and South Asian populations, a comparative study of territorial groups was designed and used in the population statistical analysis.

**Materials and Methods:** Mix of 616 samples was sequenced for complete mtDNA or hyper variable regions in this study. A published dataset of neighboring populations was used as a comparison in the Iranian matrilineal lineage study based on mtDNA haplogroups.

**Results:** Statistical analyses data, demonstrate a close genetic structure of all Iranian populations, thus suggesting their origin from a common maternal ancestral gene pool and show that the diverse maternal genetic structure does not reflect population differentiation in the region in their language.

**Conclusion:** In the aggregate of the eastward spreads of proto-Elamo-Dravidian language from the Southwest region of Iran, the Elam province, a reasonable degree of homogeneity has been observed among Iranians in this study. The approach will facilitate our perception of the more detailed relationship of the ethnic groups living in Iran with the other ancient peoples of the area, testing linguistic hypothesis and population movements.

#### ► Please cite this paper as:

Bahmanimehr A, Eskandari Gh, Nikmanesh F. The congruence between matrilineal genetic (mtDNA) and geographic diversity of Iranians and the territorial populations. Iran J Basic Med Sci 2015; 18:30-37.

## Introduction

Molecular anthropology as a modern and interdisciplinary science, is trying to employ genetic tools to design genetic borders under geographical borders and reconstruct the history of modern populations also test different linguistic and historical hypothesis about population movement. In this regard by using the DNA in molecular genetic techniques and interpretation of genetic variation in different populations, could trace the paternal lineage of human evolution and understanding genes interaction with environment and climates. Due to maternally inheritance treat, mitochondrial DNA is a useful tool for studying maternal line history. mtDNA diversity in populations is good evidence to trace modern human's origin and historical events.

The people inhabiting the Western regions of Central Asia and the Iranian upland played a key role in populating East European territories in the ancient eras. It is suggested that one of the routes of the *Homo sapiens* penetration to Europe passed through the Caspian regions. Later, during the Mesolithic period,

the Caspian regions and shelters in the South of Eastern Europe were the starting points for the recolonization of European territories (1).

Conspicuous agricultural developments occurred in the eastern horn of the Fertile Crescent, notably in Elam (Southwest of Iran), connecting Mesopotamia and the Iranian plateau (1, 2). The highly urban Elamite civilization had close contacts with Mesopotamians but exhibited an extensive differentiation from the rest of the Fertile Crescent population, including the use of a language that is thought to belong to the Dravidian family (2, 3). It is presumed that the proto-Elamo-Dravidian language plausibly originated from the Southwest region of Iran, the Elam province, and spread eastwards with the movement of farmers to the Indus Valley and the Indian subcontinent (4, 5).

The Iranian plateau, in many periods, became a target to the incursions of pastoral nomads from the Central Asian steppes (6). It is emphasized that their arrivals in the Iranian plateau brought the Indo-Iranian language of the Indo-European family that

\* Corresponding author: Ardeshir Bahmanimehr. Biotechnology Department, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran. Tel: +98-917-3116097; email: Bah\_ard@yahoo.com

eventually replaced the Dravidian languages, probably by an elite-dominance model (7).

In this regard, here we ask whether can identify explicit matrilineal genetic structure shared by all Iranians or in what extend linguistic differences in the region may reflect on the patterns of mtDNA diversity. Also, in order to study the geographic diversity of the Iranian population as well as on the expansion patterns which may affect this region, in this study it has presented a large-scale complete mtDNA also combined sequence of hyper variable region (HVS1) analysis of Iranians, representative of the majority of the provinces and ethnic groups of Iran.

## Materials and Methods

The complete mtDNA sequence of 137 Iranians (Mixed collection from all over the country), 103 Kerman providence from East of Iran, 108 Qashqai Turkic speaking nomad from South of Iran also combined sequence of hyper variable regions (HVS1 and HVS2) of 294 Iranian Azeri from Northwest of Iran were analyzed. Donors were provided information on their parental ancestry for two generations. The demographic information includes the paternal grandfather's and the maternal grandmother's birthplace/language for the mtDNA analysis.

Blood samples were collected in EDTA containing vacuoliner tubes and stored in cold boxes during the collection and transferred to laboratory with maintaining the conditions. Saliva samples were collected in Oragene DNA Collection Kits.

Genomic DNA was extracted from whole blood for samples of Northwest or Iran (294 samples) by using standard procedures and protocol of Qiagene kit (QIAamp DNA Mini Kit). The mtDNA of each genomic DNA sample was genotyped for both RFLP polymorphisms of mtDNA coding regions, and was directly sequenced for the complete mtDNA D-loop.

Genotyping began with deep-rooting markers of the mtDNA phylogeny proceeding with the more derived markers. In some cases, the final designation of a sub-haplogroup mandates the combined use of a RFLP site and a stable mutation in the mtDNA D-loop. The sequence of the entire D-loop domain including hyper variable segment 1 (HVS1) (from positions 15999 to 16400) and hyper variable segment 2 (HVS2) (from positions 30 to 407) was determined in the samples followed by RFLP screening to resolve haplogroup status, by using the ABI Prism Dye Terminator cycle-sequencing protocols developed by Applied Biosystems (Perkin-Elmer) in the samples.

The sequences with uncertain phylogenetic status in Azeri samples were selected for complete mtDNA sequencing. The mtDNA genomes were amplified and sequenced by means of the procedures described by Torroni, *et al* (8). Sequencing reactions

were run on Applied Biosystems 3130 size reader. The Sequences were edited and aligned by Bioedit software, version 7.1.8, and mutations were scored relative to the revised Cambridge reference sequence (rCRS).

RFLP typing of coding-region sites that are diagnostic of all major mtDNA haplogroups and that were previously reported for Eurasian populations were conducted hierarchically. Following early reports (9) and updated novel Iranian complete mtDNA sequences from our previous publication (10) in the GenBank (accession numbers are KC911275-KC911629), we hereby set a nomenclature with some new modifications. In this regard, the exact haplogroup genotype of each sample clarified and used in the population statics.

Genetic Fst distances and gene diversity indices were estimated by Arlequin version 3.5. Tests for significant population differentiation were carried out using the exact test for population differentiation. Similarity matrices based on Fst genetic distances were used to perform principal coordinates analysis (PCO) by GenStat version 14.2. PHYLIP version 3.6 was used to reconstruct neighbor-joining (NJ) tree.

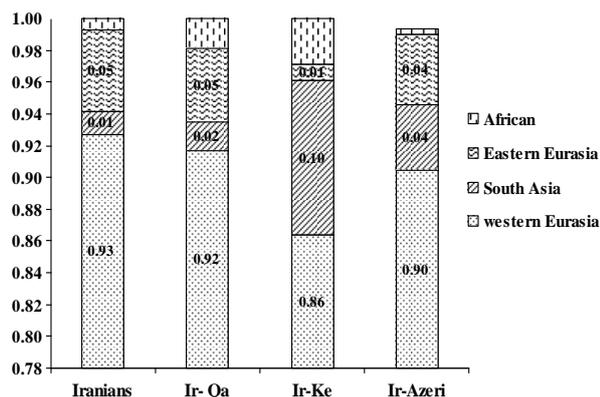
In order to trace the genetic affinity between Iranians and other populations of the area and to establish the place of Iranians in a broad framework of ethnically and linguistically diverse groups of Middle Eastern and South Asian populations a population sample (n=763) from the neighboring countries/regions and 12 different ethnic groups of Iran, from previous published data, were used as a dataset for comparison in the Iranian matrilineal lineage study based on mtDNA haplogroups (Table 1).

**Table 1.** Characteristics of the studied population by mitochondrial DNA markers as dataset

Country/Region (Population)	COD	Size	Ref	
Iran	Iranian-Arab	IR-Arb	42	
	Iranian- Baluch	IR-B	58	
	Iranian-Gilan	IR-G	21	
	Iranian-Kurd	IR-Kur	51	
	Iranian-Lur	IR-Lu	51	15
	Iranian-Mazandaran	IR-Maz	46	
	Iranian-Persian	IR-Per	54	
	Iranian-Turkmen	IR-Tur	53	
	Iranian-Zoroastrain	IR-Zo	49	
	Iran-Kerman	IR-Ke	103	
	Iran-Combined <sup>1</sup>	IR-Mix	108	*
	Iran-Qashqai	IR-Qa	111	
Iranian Azeris	IR-AZ	294		
Pakistan	Pakistani-Parsis	Pak-Pa	44	
	Pakistani-Brahui	Pak-Br	38	
	Pakistan Mixed	Pak	100	
Afghanistan-Hazara	AF-Haz	23	4	
	Uzbekistan	UZ	42	
	Turkmenistan	Turm	41	
	Turkey	Tur	50	

<sup>1</sup> Mixed collection of samples from all Iranian ethnic groups. Donors chose randomly in a period of time from Iranian students in Armenia.

\* This study (n=616)



**Figure 1.** Matrilineal genetic structure of Iranians based on the mitochondrial DNA haplogroup frequency of different origin

Abbreviations are described in the Table 1

Haplogroup composed components:

Western Eurasian: N1, N2, X, R2, JT, U, R0

African: L2a, L3d, L3f

Southern Asian: R8, M4, M5, M18, M42

Eastern Eurasian: A4, B4, C4, C5, D4, F1, G2a

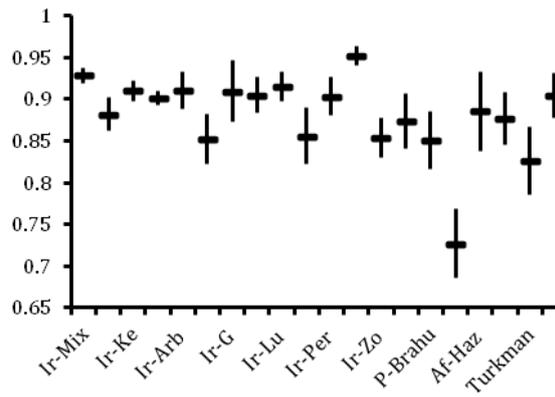
## Results

### *mtDNA haplogroup distribution and matrilineal genetic tree structure*

The haplogroup genotyping results showed that all Iranians possess the most common West Eurasian haplogroups, particularly the three most prominent mtDNA haplogroups, H, J and U; however, the frequency distribution of these lineages varies between them. Haplogroup H and J are the most frequent haplogroups and modal lineages in almost all Iranian populations except Iranian Baluch and Iranian Zoroastrian; these populations possess respectively M and R as their modal haplogroups. In contrast, all Eastern neighboring populations, Afghanistan, Pakistan, Uzbekistan and Turkmenistan, possess the South Asian haplogroup, M, as the modal and frequent lineage. This lineage is also frequent in Iranian Baluch, Turkmen and Kerman samples which are consistent with their geographically close location to their Eastern neighbours.

The majority of studied population's mtDNAs (93% for Iranians, 92% for Qashqai, 86% for Kerman and 90% for Azeris) belongs to Western Eurasian component though the impact of African, Southern Asian and Eastern Eurasian lineages is also perceptible being found at less frequencies (Figure 1).

Within West-Eurasian lineages, H and U macro-haplogroups turned out to be by far the predominant lineages: 16.8% and 23.4% for Iranians, 27.8% and 22.2% for Qashqai, 17.5% and 16.5% for Kerman, 22.8% and 20.1% for Azeris in H and U, respectively. A pattern consistent with that presented by previous studies (2, 11-14). In contrast to Europe and other Iranian studied ethnic groups (Qashqai, Kerman, Azeris) where H is predominant among the mtDNA haplogroups, in Iran, combined sample collection, the frequency of haplogroup U (23.4%) is



**Figure 2.** Genetic diversity (h) values, based on the mitochondrial DNA haplogroup frequency, with standard deviations in the studied populations

X axis: estimated genetic diversity by Arlequin version 3.5.

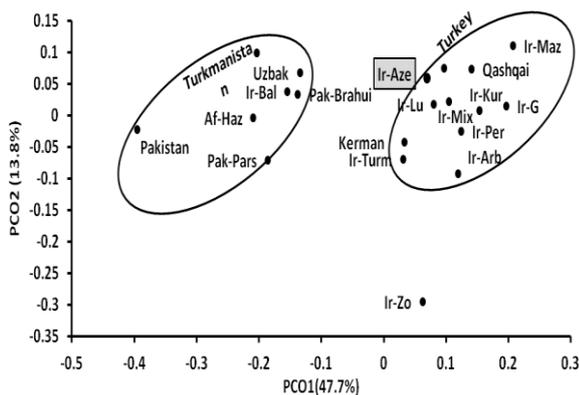
Y axis: studied Populations. Abbreviations used are described in the Table 1

higher than that of haplogroup H (16.8%). This difference accounts, at least partly for the presence in Iran of U sub-groups, such as U7 (7.3%), that are virtually absent in the Europe (13).

The South Asian components in Kerman samples are observed higher than in other Iranians. It is in agreement with the geographical position of Kerman province in the Southeast of Iran and neighboring borders of South Asian countries as Afghanistan and Pakistan. In contrast, the East Eurasian components in Qashqai and Azeris have a considerable portion. Its presence in Iran is ascribable to recent gene flow and is completely accordance with history of entrance of the Turkic invaders (Seljuk Empire 1037 - 1194 AD) of central Asia to Iran particularly among these two big Turkic speaking ethnic groups. African component, regardless of the low portion, in Qashqai and Kerman samples are more than others. Due to the close geographical position of these two populations to the Persian Gulf in the South of Iran also Hormoz strain, as gateway of African countries to Iran, it seems gene flow of African components was done through this corridor.

Middle Eastern and lower Arabian Peninsula-specific haplogroup T (6, 14) is frequent in Kerman (0.165) and Azeri populations that might interpreted as inheritance of the early groups of farmers involved in the eastwards expansion of agricultural techniques from the Fertile Crescent and Elam. This result, however, in agreement of the data published by Quintana-Murci *et al*, (4) and as a contradiction in results of small sample size collection of the North Iranian by Terreros *et al* in 2011(14).

Haplogroup U3 is frequent in Qashqai (0.083) and Azeris. Indeed it seems this haplogroup expanded from the Svaneti region (Georgia, Caucasus) (6) across the Northwest of Iran. Due to the nomadic life



**Figure 3.** Principal coordinates (PCO) plot based on the  $F_{ST}$  values of studied populations (by mtDNA 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 1

of Qashqai in the Zagros Mountains, it seems this lineage distributed in parallel of Mountains from North to South of Iran and cannot consider boundary role of the Zagros Mountains for it. However, other main geographical barriers such as Dasht-e Kavir and Dasht-e Lut have boundary roles in the gene flow for this lineage due to the complete absence of U3 in Kerman samples.

### Population statistics analysis and genetic relationships

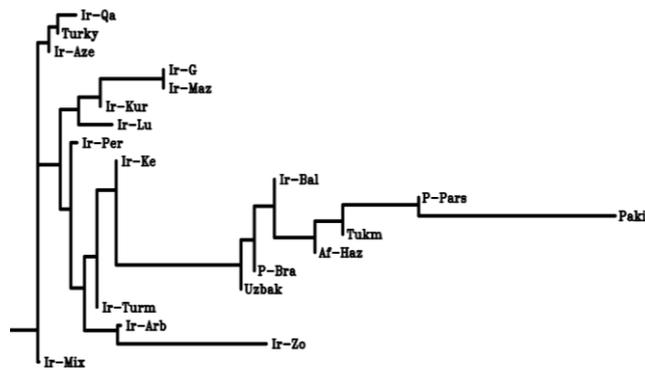
Lack of a sharp ethno-linguistic structure for mtDNA sequence diversity in Iranians was statistically supported by different tests. Using mtDNA haplogroup frequencies we have assessed the genetic relationship between all populations under consideration by calculating the conventional  $F_{ST}$  genetic distances. In order to identify whether genetic distances are statistically significant, the  $P$ -value test of population differentiation for all datasets under consideration was conducted.

mtDNA diversity values,  $h$ , detected in the examined populations lies within the similar highest range (0.95-0.9) for all Iranian and Turkish populations and likewise the high range (0.9-0.8) for other Eastern populations. The Zoroastrian, Baloch and Mazandarani among the Iranian population, and the Brahui and Turkmenistan people among the Eastern populations exhibited low diversity (Figure 2).

The similar range of diversity among most

Iranian ethnic groups is confirmed by previously observed data (13, 15-17) and is also in concordance with the Population Pairwise differences in  $h$  values, which shows no meaningful ( $P < 0.05$ ) difference among Iranian ethnic groups, except for the

Turkmen and Zoroastrians that significantly differ from some of the other ethnic groups of the country. Interestingly, Iranian Turkmen exhibited the highest diversity in the studied populations and Iranian Zoroastrians rank among the lowest in terms of



**Figure 4.** Rooted Neighbor-Joining dendrogram of the studied populations based on  $F_{ST}$  genetic distances of mtDNA 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 1

mtDNA diversity. Iranian Azeris have rather high level of gene diversity (0.9), which just significantly differs in Population Pairwise differences, in  $h$  values ( $P < 0.05$ ) by Turkmen ethnic group of Iran and combined population of Pakistan.

The genetic distances test results show that based on the  $F_{ST}$  genetic distances the Iranian Azeris and Iranian Qashqai and Turkey are closer to each other and significantly differ from other ethnic groups of Iran and other studied populations based on the pairwise  $P$  value results. Interestingly, the Iranian populations are significantly distance by all other Eastern neighboring populations except Iranian Baloch which close to Pakistani Brahui, Afghani Hazara, Uzbekistan and Turkmenistan also Kerman sample observed close to Brahui people. Turkey samples observed close to the Iranians and far from all Eastern populations.

In order to depict the patterns of genetic affinities between Iranians and comparative data sets in a two-dimensional space it also have conducted PCO plot based on values of  $F_{ST}$  genetic distances (Figure 3). In general, from the positions of the populations in the plot, it is possible to distinguish an explicit separate cluster of Iranians by other Eastern populations. Considering  $F_{ST}$  genetic distances of studied populations and their position on the PCO plot, it might be emphasized on the geographical structure for matrilineal lineage of Iranians which, despite of homogeneous structure and position of all Iranians in the same cluster, different geographically close ethnic groups located neat together. For instance, Azeris, Gilakis and Mazandarani are close to their nearby neighbor, Turkey, to the north of Iran. Also Kerman samples located close to Turkmen, other eastern ethnic groups in Iran, and together located intermediate of two separate clusters of Iranians and Eastern populations as well as Brahui position on the plot. This sharp geographical structure for the Iranian

mtDNA sequence is in contrast of previous studies (13-15). In the other sub-cluster, Iranian Azeris and Qashqai are located close to Turkey and far from Iranian Turkmens, other Turkic speaking ethnic group of Iran, which coordinates to pairwise  $F_{ST}$  values of these populations. Iranian Baloch directly located in the Eastern population cluster which is the agreement of their ethnicity. PCO plot also showed the reproductive isolated feature for Iranian Zoroastrians due to its far location on the plot. It is exactly in the contrast to its paternal feature which demonstrated location among other Iranian ethnic groups. Once again,  $F_{ST}$  genetic distances used to visualize the genetic relatedness of the Iranians and other populations of the area by constructing a phylogenetic tree using the Neighbor-Joining method (Figure 4). It displays a similar pattern of genetic relationship between the studied populations at the time of comparing the principal coordinate plots of the  $F_{ST}$  genetic distances.

It also demonstrates the correlation between the geographic proximity and the genetic structures of Iranians. As could be seen in this tree, all Eastern populations are located in the same root and the Iranian Baloch is routed directly with them. Iranian Turkmen and Kermani samples are rooted in an intermediate stage between Iranians and their neighboring Eastern populations. Turkic speaking Qashqai and Azeris are rooted analogous in Turkey, suggesting their origin from a common maternal ancestral gene pool, while being located in the Iranian cluster in the PCO plot and rooted, in the NJ tree, close to their geographically nearby neighbors the Gilakis and Mazandarans. These are other important signs to take into consideration regarding the geographical structure for the matrilineal lineage of Iranians.

## Discussion

The genetic diversity of the people that populate the area is often affected by the regional geography; certain geographic features, such as the deserts and Mountains of Iran, have served as barriers, and others, such as the Strait of Hormuz, have acted as conduits for population movements. Furthermore, environmental fluctuations occurring over time have changed the area's features and influenced human life and migrations; for example the uniquely harsh geography of the Northwest of Iran, probably had the capability to facilitate the establishment of social organizations and thus possibly caused population density in this area, a density that extended to the west foothills of the Zagros mountains; this is in contrast to the central and eastern parts of Iran which, due to the big deserts were not suitable for dense populations. Overall, this pivotal geographic position, most likely served present day Iran as a gateway of reciprocal human movements and the genetic diversity that exists throughout this vast

region point to its central role in the dispersal and expansion of human populations throughout the tri-continental area.

Overall, the various statistical tests in this study supported a striking degree of homogeneity in the mtDNA of different Iranian populations. This discordant pattern of high ethno-linguistic and low mtDNA heterogeneity observed among the examined Iranian populations is in agreement with previous studies (8, 13, 16, 18) and suggests either a common maternal origin for modern Iranians and extensive levels of gene flow amongst them.

Also, from another perspective, restricted genetic flow through geographical barriers within Iran and between Iran and neighboring regions in maternal gene flow (mtDNA) was observed only for a small number of ethnic groups and for most of the Iranian population, it is assumed that it might have preserved ancestral lineages that can be representative of Proto-Indo-Iranian or prehistoric mitochondrial profiles which survived relatively recent external contributions to the Iranian gene pool. In contrast, this barrier role in the paternal gene flow is supported by Y-chromosome data for most of all Iranians (19-22).

The presence of haplogroup H as the most frequent and modal lineage in almost all Iranian populations also highest frequency of this haplogroup and its subclades in Europe and Caucasia (9, 13), may reflect the gene flow from neighboring Southwest Asia where they are present at moderate frequencies. As in Europe, most of the present-day Near Eastern-Caucasus area variants of haplogroup H started to expand after the last glacial maximum (LGM) and presumably before the Holocene (13).

Together, the whole H haplogroup samples of Iranian, Qashqai and Kerman have clustered among 40 sub-haplogroups which just 2 of them (H\* and H13a2a2) shared among these three populations. This diverse distribution of sub-haplogroup is indicative of gene flow from a different time and origin to Iran.

We found that haplogroup M frequency drops abruptly from about 49% in Pakistan and 34% in Turkmenistan to about 1.5% in the combined samples of Iran and 4% in Turkey, marking the western border of haplogroup M distribution. It seems the two uninhabited Great Salt Deserts in the middle of the Iranian plateau, stretching from the Alborz Mountain range in the Northwest to the Southeast of the country, as particular sharp borders, cut the distribution of the South Asian specific mtDNA haplogroup to the east and the Iranian plateau.

Among Eastern populations, two marginal Pakistani ethnic groups, the Brahui and Parsis, pass mtDNA lineages differently. The West Eurasian component is more prevalent in the Parsi (70.5%) and Brahui (71%) populations of Pakistan than in other Eastern populations (50% in Pakistan, 52% in Af-Haz,

64% in Uzbekistan and 58% in Turkmenistan); in contrast, the South Asian component of these two ethnic groups is less significant than it is in the other populations and the East Eurasian component is completely missed.

Brahui population, now a day is settled in some area in the Southeast of Iran, Sistan and Balochestan province, Pakistan and India. Their language is Dravidi which also spoken in the South of India, Sri Lanka and Afghanistan. There were different linguistic hypothesis about their origin as; proto-Elamo-Dravidian hypothesis (23) which considered their origin in the Iranian ancient province Elam. The Brahui population passes the U7 haplogroup (0.105) similar to Iranian Lurs (0.118) from the west of Iran (ancient Elam province) and haplogroup H as one of the modal mtDNA haplogroups as do most Iranians in contrast to other Eastern populations.

As shown in the PCO plot (Figure 3), the Brahui lies in an intermediate position between the Iranian and the Eastern neighboring populations, far from the Pakistanis. These observations are in agreement with previous studies (13). Our results show that the maternal gene pool of the Brahui is similar to that of Indo-Iranian speakers from the Southwestern Asian corridor.

The Parsis are a group of Iranian Zoroastrians who most of them live in the Indian subcontinent particularly Southeastern Pakistan. Their Iranian origin, has supported by different historical documents (Nanavutty, 1997). Their ancestors migrated from Iran to India (province of Gujarat) on the second century AD, on the Arab governments of Iran, hence might consider them the first emigrant society of Iranians. The Parsis parental genetic structure is so interesting in the molecular anthropology due to a strong contrast in their maternal (mtDNA) and paternal (Y-chromosome) inheritance gene pool. Published Y-chromosome data take them in an Iranian population category rather than their now a day neighbors in Pakistan or India (15, 18), which is in accordance of historical records. Meanwhile, our comparison in mtDNA data shows their affinity to Eastern populations. About 29.6% of their maternal gene pool belonging to South Asian haplogroups, which make up only 1.5% of the combined Iranian samples. However, interestingly, the close similarity observed between Parsis and Iranian Zoroastrians is due to the same rate of haplogroups U2 and T in both of these ethnic groups. All these findings lead us to assume their reproductive admixture with local female or directional mating between Iranian Parsis and local women.

The similarity of Parsis to the Zoroastrians of Iran cannot be ignored. One might conjecture that this similarity, in addition to being the result of a portion of the West Eurasian component, stems from the diversity of mtDNA lineages in Parsis which remains high (0.8732). Despite the similarity in the Parsis and

Zoroastrians, the Parsis are placed between Iranians and Pakistanis (in the PCO plot) in this study. In the Parsi population, in addition to mtDNA diversity, a high Y-chromosome diversity due to a strong genetic drift has been reported (15). This kind of discordant pattern of paternal and maternal gene flow has also been reported in some simultaneous populations with the same language in different geographical places, as in the case of the two Indo-Iranian-speaking Talysh groups from Iran and Azerbaijan, which were reported similar in mtDNA while the Y-chromosomal marker compositions were considerably different (19).

## Conclusion

Given the contrasting views of the previous studies in which Iranians are mainly considered as territorially homogenous, particularly for matrilineal lineages, it should be pointed out that the number of analyzing samples and the resolution of studies were not enough to explore the geographic stratification of the Iranian population. Thus, considering the extremely high level of ethnic and linguistic diversity in Iran and based on the need for further investigation of this issue, this project aimed to study the matrilineal genetic and geographic/ language diversity of Iranians and the territorial populations according to the mtDNA markers.

Described the matrilineal genetic structure of Iranians demonstrate a close genetic structure of all Iranian populations (combined ethnic group samples of Iran and particular populations of Qashqai, Azeri and Kerman, which are studied here), thus suggesting their origin from a common maternal ancestral gene pool and show that the diverse maternal genetic structure does not reflect population differentiation in the region in their language.

However, despite this similarity, among Iranians, the high frequency of South Asian components in Kermani samples is agrees with the geographical position of the Kerman province in the southeast of Iran and the neighboring borders of South Asian countries such as Afghanistan and Pakistan. Also the considerable portion of East Eurasian components in the Qashqai and Azeri people is ascribable to the recent gene flow and is completely in accordance with the history of the entrance of the Turkic invaders (Seljuk Empire (1037 - 1194) AD) of central Asia into Iran particularly among these two big Turkic speaking ethnic groups. This study also has considered a particularly sharp geographical border role for two uninhabited Great Salt deserts in the middle of the Iranian plateau, which stretch from the Alborz Mountain range in the Northwest to the Southeast of the country in the distribution of the South Asian specific mtDNA component, haplogroup M, which cut distribution of this haplogroup to the east and to the Iranian plateau.

These issues in addition to genetic diversity that exists throughout this vast region demonstrated that for Iranian plateau; maternally transmitted genetic structure reflects the important role, to make it as a crossroad for human migrations since prehistoric times. While the significant genetic structure has not generated among this territory due to strong geographic and linguistic barriers. Cultural, political and geographical factors seem to have acted as an obstacle to maternal gene flow mostly in some ethnic groups as Balochis, Lurs, Qashqaies, Zoroastrians and Jews due to their long centuries of reproductive isolation of language, religion and other cultural features.

Generally, despite the striking degree of homogeneity in the mtDNA of different Iranian populations and the lack of a sharp ethno-linguistic structure observed among Iranians, there is a clear geographical structure for the matrilineal lineage that is statistically supported by different tests between them.

### Conflict of interest

The authors declare no conflict of interest.

### Acknowledgment

We are grateful to all individuals who donated their DNA for research and share their history and demographic information also thanks all friends for help with sample collection. Especial thanks to Estonian Biocenter, Tartu University for laboratory facility and technical assistance.

### References

- Riehl S, Zeidi M, Conard NJ. Emergence of agriculture in the foothills of the Zagros Mountains of Iran. *Science* 2013; 341:65, 64-67.
- Alizadeh A. The Rise of the highland elamite State in Southwestern Iran. *Curr Anthropol* 2010; 51:353-383.
- Sikora MJ, Colonna V, Xue Y, Tyler-Smith C. Modeling the contrasting Neolithic male lineage expansions in Europe and Africa. *Invest Genet* 2013; 4:1-18.
- Quintana-Murci L, Chaix R, Wells S, Behar D, Sayar H, Scozzari R, et al. Where west meets east: the complex mtDNA landscape of the Southwest and central asian corridor. *Am J Hum Genet* 2004; 74:827-845.
- Renfrew C. Languages families and the spread of farming. In: Harris DR. editor. The origins and spread of agriculture and pastoralism in Eurasia. Smithsonian Institution Press; Quintana-Murci L; 1996.
- Cavalli-Sforza LL. The spread of agriculture and nomadic pastoralism: insights from the genetics, linguistics and archaeology. In: Harris DR. editor. The origins and spread of Agriculture and Pastoralism in Eurasia. Washington DC: Smithsonian Institution Press; 1994.
- Renfrew C. The Emerging Synthesis: the Archaeogenetics of Farming/Language Dispersals and other Spread Zones. In: Bellwood & Renfrew C. editors. Examining the farming/language dispersal hypothesis. Mc Donald Institute for Archaeological Research; 2002.p. 3-16.
- Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, et al. Signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 2001; 69:844-852.
- Oven V, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA Variation. *Hum Mutat* 2009; 30: 386-394.
- Derenko M, Malyarchuk B, Bahmanimehr A, Denisova G, Perkova M, Farjadian S, et al. Complete mitochondrial DNA diversity in Iranians. *PLoS One* 2013; 8:1-14.
- Houshmand M, Sanati M H, Vakilian M, Akuchekian M, Babrzadeh F, Teimori M, et al. Investigation of the mitochondrial haplogroups M, BM,N, J, K and their frequencies in five regions in Iran. *Iran J Biotech* 2004; 2:44-48.
- Roostalu U, Kutuev I, Loogva EL, Metspalu E, Tambets K, Reidla M, et al. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* 2007; 24: 436-448.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, et al. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasian by anatomically modern humans. *BMC Genet* 2004; 5:26-37.
- Terreros MC. Mitochondrial DNA and Y-chromosomal stratification in Iran: relationship between Iran and the Arabian Peninsula. *J Hum Genet* 2011; 8:1-12.
- Farjadian S, Sazzini M, Tofanelli S, Castr L, Taglioli L, Pettener D, et al. Discordant patterns of mtDNA and ethno-linguistic variation in 14 Iranian Ethnic Groups. *Hum Hered* 2011; 72:73-84.
- Nasidze I, Quinque D, Rahmani M, Alemohamad SA, Stoneking M. Concomitant replacement of language and mtDNA in South Caspian populations of Iran. *Curr Biol* 2006; 16:668-673.
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, et al. Y chromosomal DNA variation in Pakistan. *Am J Hum Genet* 2002; 70:1107-1124.
- Nasidze I, Quinque D, Rahmani M, Alemohamad A, Asadova P, Zhukova O, et al. mtDNA and Y-chromosome variation in the Talysh of Iran and Azerbaijan. *Am Phys Anthropol* 2009; 138:82-89.
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, et al. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 2001; 98:10244-10249.
- Rootsi S, Myres NM, Alin A, Jarve M, King R J, Kutuev I, et al. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur. J Hum Genet* 2012; 20: 1275-1282.
- Yunusbayev B, Metspalu M, Jarve M, Kutuev I, Rootsi S, Metspalu E, et al. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol Biol Evol* 2011; 29:359-365.

22. Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ. Iran: Tricontinental nexus for Y-chromosome driven migration. *Hum Hered* 2006; 61:132-143.

23. McAlpin DW. Toward proto-elamo-dravidian. *Language* 1974; 50:89-101.