# Iranian Azeri's Y-Chromosomal Diversity in the Context of Turkish-Speaking Populations of the Middle East

*\*L Andonian[1], S Rezaie[2], A Margaryan[3], DD Farhud[4], K Mohammad[1], K Holakouie Naieni[1, 5], MR Khorramizadeh[6], M H Sanati[7], M Jamali[8], P Bayatian[1], L Yepiskoposyan[3]*

[1]*Dept. of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran Iran*
[2]*Division of Molecular Biology, Dept. of Medical Parasitology and Mycology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran*
[3]*Human Genetics Group, Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia,*
[4]*School of Public Health, Tehran University of Medical Sciences, Tehran Iran*
[5]*Iranian Epidemiological Association*
[6]*Dept. of Medical Biotechnology, School of Advanced Medical Technologies, Tehran University of Medical Sciences, Tehran Iran*
[7]*National Institute for Genetic Engineering and Biotechnology, Pajoohesh Blvd., 17 Km Karaj HWY, Tehran, Iran,*
[8]*Academic Member of Ministry of Health and Education, Tehran Iran*

## Abstract

**Background:** The main goal of this study was to conduct a comparative population genetic study of Turkish speaking Iranian Azeris as being the biggest ethno-linguistic community, based on the polymorph markers on Y chromosome.

**Methods:** One hundred Turkish-speaking Azeri males from north-west Iran (Tabriz, 2008-2009) were selected based on living 3 generations paternally in the same region and not having any relationship with each other. Samples were collected by mouth swabs, DNA extracted and multiplex PCR done, then 12 Single Nucleotide Polymorphisms (SNPs) and 6 Microsatellites (MS) were sequenced. Obtained data were statistically analyzed by Arlequin software.

**Results:** SNPs and Microsatellites typing were compared with neighboring Turkish-speaking populations (from Turkey and Azerbaijan) and Turkmens representing a possible source group who imposed the Turkish language during 11-15$^{th}$ centuries AD. Azeris demonstrated high level of gene diversity compatible with patterns registered in the neighboring Turkish-speaking populations, whereas the Turkmens displayed significantly lower level of genetic variation. This rate of genetic affiliation depends primarily on the geographic proximity.

**Conclusion:** The imposition of Turkish language to this region was realized predominantly by the process of elite dominance, i.e. by the limited number of invaders who left only weak patrilineal genetic trace in modern populations of the region.

**Keywords:** *Iranian Azeris, Y chromosome diversity, Microsatellites, SNPs, Iran*

## Introduction

Due to its geo-strategic location in the Middle East, the Iranian plateau has served as a key crossroad for human disseminations and played a critical role in the migratory waves between the populations of the Middle East and beyond (1-5). The most important long-term factor in this process was human adaptation to the Iranian plateau and its geographical, topographical, and climatic conditions with the subsequent development of agriculture, pastoralism, and pastoral nomadism. The spread of these technological innovations, along with a series of major demographic and historical events, has resulted in a large diversity and dispersal of ethnic groups and languages (1, 6, 7).

**\*Corresponding author:** E-mail: andonian@tums.ac.ir

Between the third and second millennia BC, the Iranian plateau became exposed to incursions of pastoral nomads from the Central Asian steppes (1), which were a difficult environment for agriculture but ideally suited to animal husbandry and pastoral nomadism.

Presumably, via an elite-dominance process, existing Dravidian language across the regeon was substituted by Indo-Iranian language, which is a branch of Indo-European language (8-10). Also their genetic impacts were as significant as the imposition of their language, which is clearly observed in Iran (11), Pakistan (12) and northern India (13).

In the period of the eleventh to thirteenth centuries AD the Arab-Muslim, Seljuk and subsequent Turkic-Mongol invasions signaled the arrival of a new people with flocks and culture. Specifically, in a series of rapid Arab-Muslim conquests in the seventh century, the Arab armies swept through most of the Middle East, completely engulfing the Persian lands (7).

The dominance of the Arabs came to a sudden end in the mid-eleventh century with the arrival of Seljuk Turks, a clan of the Oguz Turks (8). The expanding waves of these Altaic-speaking nomads from Central Asia involved regions farther to the west, such as Iran, Iraq, Anatolia, and the Caucasus, where they imposed Altaic (Turkish) languages. In these western regions, however, the genetic contribution is low or undetectable (14), even though the power of these invaders was sometimes strong enough to impose a language replacement, as in Turkey and Azerbaijan (1).

Later, the Mongol armies also moved westward and, by the early thirteenth century, established their rule over a vast region, including Iran and advancing as far west as the Caucasus and Turkey (1, 7). These waves of various invasions and subsequent migrations resulted in major demographic expansions in the region, which added new languages and culture to the mix of peoples that had pre-existed in Iran.

In general, a considerable genetic diversity is observed in Iranian populations, which resembles to that of Middle East patterns as a whole and

strengthens the idea of Persia, being the main crossroad for human dissemination (3, 11, 15-18). This area is remarkable for its high level of ethnic and linguistic diversity, comprising the major language families (Indo-European, Altaic, and Afro-Asiatic) currently spoken by more than seventy ethnically different populations (http://www. ethnologue.org/). This demonstrates the role of Iran, which played in population dispersal across the latitudinal belt spanning from western Anatolia to the Indus Valley. However, there have been gaps in high-resolution genetic analyses for this region to uncover population history at a fine scale, for example, for particular ethnic and linguistic groups.

In this project, we intended to get relatively comprehensive information about the Y chromosome diversity in Azeris living in Iran. Subsequently, the principal aim of this paper was to identify the place of Azeris in the frame of Turkic-speaking populations of the Middle East and to test the extent of gene flow from Central Asia. We used both SNP and STR genetic markers on the non-recombining portion of Y chromosome, which provide high level of genetic resolution and are consistent with other sets of markers applied to the studies on patrilineal genetic history of various populations (19, 20).

## Material and Methods

Buccal swab specimens were collected from 100 ethnical Azeri men currently living in Tabriz, Iran in 2008. One sample was later discarded, as the Y-chromosome typing was unsuccessful.

All donors were selected only if their paternal grandfathers were from the same region and they were unrelated to other donors at the grandfather level. The samples were stored in a DNA preservative solution consisting of 0.5% sodium dodecyl sulphate and 0.05M ethylenediaminetetraacetic acid for transport purposes. Samples were collected anonymously and informed consent was obtained from all individuals before samples were taken.

The comparative data sets have been taken from previously published papers (19, 21, 22); the re-

sults for the Turkmen population are available only for microsatellite loci, therefore the comparison with this data set was conducted only at STR markers.

Standard phenol-chloroform DNA extractions were performed. The strategy adopted for typing samples was designed to ensure informative comparison with existing published data. NRY were characterized by 12 binary Y chromosome polymorphisms: 92R7, M9, M13, M20, sY81, SRY+465, SRY4064, SRY10831, Tat, M17, *Alu* insert-YAP, and p12f2, (19, 23) and screened for six microsatellite (MS) markers: DYS19, DYS388, DYS390, DYS391, DYS392, and DYS393, as described by Thomas et al. (24). Haplogroups (hg) were defined by single nucleotide polymorphism (SNP) markers according to the Y Chromosome Consortium nomenclature (25). Microsatellite repeat numbers were assigned according to Kayser et al. (26).

Either the microsatellite PCR products and UEP digestion products were run on an ABI-310 capillary-based genetic analyzer or a gel based system such as an ABI-377 automated sequencer. For the ABI-310 genetic analyzer, 1.2 ul aliquots of the microsatellite PCR products or UEP digestion products were mixed with 0.5 ul size standard labeled with the fluorescent dye TAMRA (PE-Applied Biosystems) and 12 ul of de-ionized formamide. Samples were denatured at 96° C for 3 min and chilled on ice for 5 min before being run, using POP-4 polymer and a 36 cm POP-4 capillary.  For the ABI-377 automated sequencer, 1.0 ul aliquots of the microsatellite PCR products or UEP digestion products were mixed with 2.0 ul of a loading buffer (formamide: dextran blue: TAMRA-labeled size standard, in the ratio of 23: 4: 2).

Unbiased genetic diversity index, *h*, and its standard error were calculated using the formulae of Nei (27). Nei's Genetic Identity, I, was calculated in accordance with Nei (27). Pairwise genetic distances ($F_{ST}$) were estimated from analysis of molecular variance (AMOVA) $\Phi_{ST}$ values with the aid of Arlequin software (28). Tests for significant population differentiation were carried out using the exact test for population differentiation (27, 29).

Testing for differences in *h* between populations was performed by bootstrapping method (30). Principal Coordinates Analysis was conducted on similarity matrices calculated as one minus Genetic distance ($F_{ST}$, $R_{ST}$) or based on Nei's Genetic Identity values. Figures along the main diagonal, representing the similarity of each population sample to itself, were calculated from the estimated genetic distance between two copies of the same sample.

Signature haplotype analysis (high frequency modal haplotypes and modal clusters) (19, 31-33) was performed by hand.

## Results

The number of haplogroups detected in Azeris is the highest, nine, whereas in Eastern Turks, it is seven, and in Azerbaijanis, it is only five. The latter might be explained by the limited number of specimens in this data set. Nevertheless, we tried to make some inferences about the genetic structure of these groups based on haplogroup frequencies.

The most common haplogroup in all data sets is haplogroup J that is present at almost equal rate in the three groups: 39.39% in Azeris, 40.00% in Azerbaijanis and 39.02% in Eastern Turks. The next frequently encountered haplogroup is hg BR* (xDE,JR) which is registered at 23.23%, 40.00%, and 9.76%, respectively. The haplogroups P*(xR1a), E*(xE3a) and R1a1 are also rather frequent in Eastern Turks (19.51%, 14.63% and 10.98%, respectively). It is necessary to add that haplogourp N3 defined by Tat mutation which presumably originated in Central Asia (34) is detected only in Azeris and Azerbaijanis.

The overall comparison of population structures between the groups shows that Azerbaijanis differ significantly from Eastern Turks according to exact test of population differentiation ($P < 0.0001$); in the same time, there is no significant disparity between the Iranian Azeris and the two other (Azerbaijanis and Eastern Turks) comparative data sets ($P > 0.05$). This pattern of genetic relatedness is supported also by F*st* values, which indicate closer genetic affinity between the Azeris and Eastern Turks, as well as Azeris and Azerbai-

janis; the biggest genetic distance was detected between Azerbaijanis and Eastern Turks.

At STR level, we had possibility to add Turkmens in comparative data sets. Microsatellite markers identified 137 haplotypes in total while considering the four groups. The most diverse pattern of haplotypes is observed in Azerbaijanis and Azeris (32 haplotypes in 40 samples, and 76 in 99, respectively). In contrast, Eastern Turks and Turkmens display much lower level of variability (49 in 82 and 18 in 51, respectively). It is worth mentioning that Azeris bear 52 unique haplotypes, i.e. they are not encountered in other three data sets; only two haplotypes were shared by all groups considered.

The above-mentioned pattern of genetic variability is reflected in the actual values of gene diversity, *h* (Fig. 1). The Azeris show the highest level of genetic diversity (h= 0.9934, bootstrapped value - 0.9834). The dramatically lower rate of this parameter was registered for Turkmens (actual value -0.8267, bootstrapped value -0.8068). Azerbaijanis and Eastern Turks also have rather high level of gene diversity, although they are still lower than in Azeris. While comparing differences in h values we found two not significant differences-between Azerbaijanis and Azeris, as well as between Azerbaijanis and Eastern Turks- using bootstrap method. In case of Bayesian approach, all possible comparisons show significant level of differences (*P*< 0.0001, Table 1).

The AMOVA analysis revealed that the bulk of observed genetic diversity is explained by within-population differentiation (94.16%), and only about 5.84% reflects inter-population variability. Once again, this result supports the general rule that within-group variability is the main source of human genetic diversity.

The Azeri modal haplotypes, ht 14-15-23-10-11-12, is detected at 5.05% rate and is modal in Azerbaijani data set (7.50%). The comparable level of this haplotype is found in Eastern Turks (6.10%), while in Turkmens it is at 1.96%. The modal haplotypes in Eastern Turks data set, each at 8.54%, ht 13-12-24-10-11-13 and ht 14-12-

24-11-13-12, are detected also in Azerbaijanis (2.50% each) and in Azeris (3.03% and 2.02%, respectively), being absent in Turkmens. Two modal haplotypes of Turkmens, ht 15-12-24-10-11-13 (35.29%) and 16-12-24-10-11-13 (21.57%) display very low frequency only in Azeris (2.02% and 1.01%, respectively).

Actually, these two haplotypes are one-step neighbors and therefore might be considered as one modal cluster. The modal cluster in Turkmens, the most pronounced one, accounting for 60.78%; was found at low frequency only in Azeris (3.03%) and is totally absent in Azerbaijanis and Eastern Turks (Table 2). The modal cluster of Azeris and Azerbaijanis, being the same, accounts, respectively, for 20.00% and 14.14%; it is at comparable rate in Eastern Turkey (9.76%) and at much lower level in Turkmens (1.96%). Eastern Turks' modal clusters, accounting for 10.98% and 9.76%, are present only in neighboring populations (Azerbaijanis and Azeris) and are absent in Turkmens.

It displays some important patterns of relationships between the groups. First, it is rather evident that Turkmens are almost equally distant from the rest of Turkic-speaking populations studied. In the same time, Azeris, Azerbaijanis, and Eastern Turks form some sort of dense cluster, possibly reflecting the close genetic contacts between these groups compared to Central Asian Turkic-speaking peoples. While constructing PCO plots based on SNP+MS haplotypes, the relationship between Middle Eastern populations becomes more refined and it is in a full accordance with the results of genetic distance comparison.

Once again, the plot proves that Azeris occupy intermediate position between their close neighbors that might witness to some extent a common origin and/or intense genetic contacts since ancient times. The mentioned relationships between the populations are fully supported by the exact test of population differentiation (Table 3 based on MS data only).

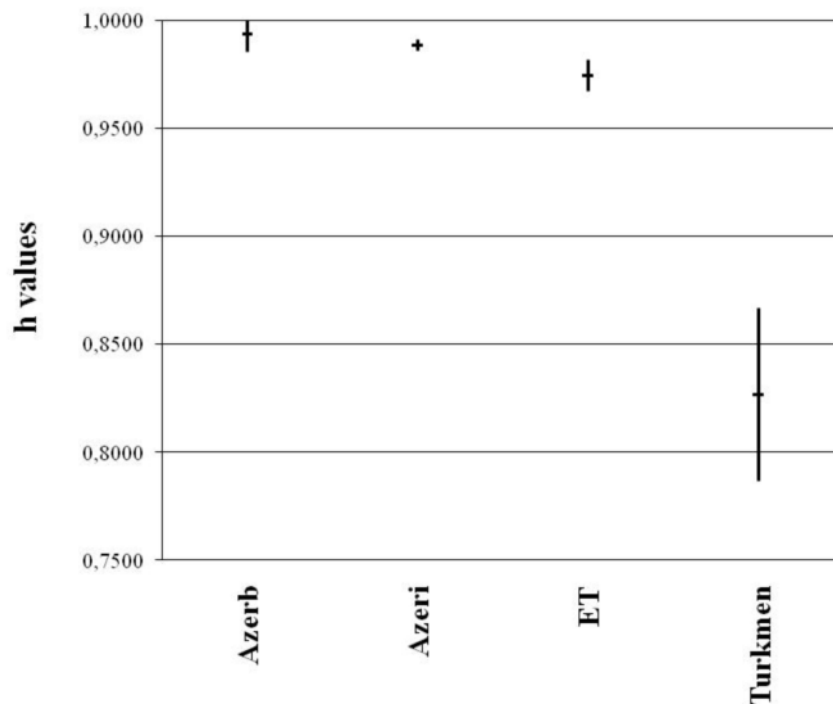Genetic distance data (Rst based on MS only) were used to visualize the spatial relationships of the groups (Fig. 2).

**Fig. 1:** Genetic diversity, h, with bootstrap 95% confidence intervals

**Table 1:** Pairwise differences in h values based on SNP+MS haplotype (lower left table, based on bootstrap approach; upper right table, based on Bayesian method)

|  | Azeri (n=99) | Azerb (n=40) | ET (n=82) | Turkmen (n=51) | h |
|---|---|---|---|---|---|
| Azeri | - | 0.0000 | 0.0000 | 0.0000 | 0.9934 |
| Azerb | 0.5721 | - | 0.0000 | 0.0000 | 0.9885 |
| ET | 0.0122 | 0.9004 | - | 0.0000 | 0.9744 |
| Turkmen | 0.0000 | 0.0001 | 0.0003 | - | 0.8267 |

Significant values in underlined

**Table 2:** Frequently Encountered Clusters

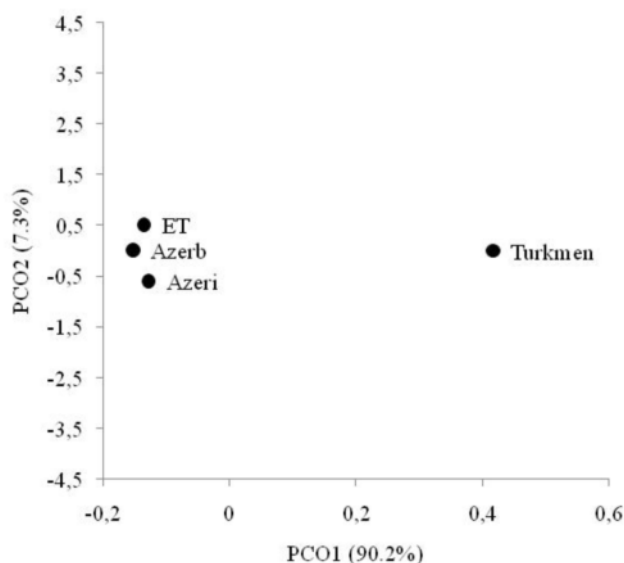| MS | Azerb (n=40) | Azeri (n=99) | ET (n=82) | Turkmen (n=51) |
|---|---|---|---|---|
| 14 15 23 10 11 12 | 0.2000 | 0.1414 | 0.0976 | 0.0196 |
| 13 12 24 10 11 13 | 0.0250 | 0.0606 | 0.1098 | 0.0000 |
| 14 12 24 11 13 12 | 0.1000 | 0.0303 | 0.0976 | 0.0000 |
| 15 12 24 10 11 13 | 0.0000 | 0.0303 | 0.0000 | 0.6078 |

Significant values in underlined

**Fig. 2:** Principal coordinates plot based on $R_{ST}$ values

**Table 3:** *P* values for the exact test of population differentiation based on MS data

|  | Azerb (n=40) | Azeri (n=99) | ET (n=82) | Turkmen (n=51) |
|---|---|---|---|---|
| Azerb | - |  |  |  |
| Azeri | **0.363** | - |  |  |
| ET | 0.009 | **0.060** | - |  |
| Turkmen | 0.000 | 0.000 | 0.000 | - |

## Discussion

The results obtained show that the rate of genetic relatedness between the populations considered depends in the first instance on the spatial proximity than on the belonging to the same linguistic group. In this context, these outcomes were highly expected taking into consideration the actual geographic location of the three populations. Azeris, being situated in between Azerbaijanis and Eastern Turks, had more possibilities of genetic contacts with both groups as the closest neighbors, while gene flow between Azerbaijan and Eastern Turkey could have been rather limited.

In general, we can make rather strong inferences about the genetic relatedness between the populations under consideration. The principal one is that Iranian Azeris have much weaker genetic affinity with Turkmens than with their immediate neigh-

bors. The same statement could be attributed to the Azerbaijanis and Eastern Turks. It seems that Turkmens had no marked input in the gene pool of Azeris, Azerbaijanis, and Eastern Turks, despite very close linguistic affinity of these groups belonging to Turkic-speaking populations. We have all grounds to suggest that language replacement took place through elite dominance phenomenon rather than demic diffusion model (35).

A weak genetic affinity between Middle Eastern Turkic-speaking populations and Turkmens is possibly explained by the fact that Central Asian populations had not any essential gene flow to the origin of Turkic speaking peoples of South Caucasus and Asia Minor, which is supported also by the results of Cinnioğlu et al. (21). Therefore, the imposition of Turkic language to this region was realized predominantly by limited number of invaders who left only weak genetic signal in modern populations of the region. The same pattern of geographic vs. genetic relatedness was revealed while comparing Indo-European speaking Bakhtiari, and Semitic-speaking Arabs (36). Both mtDNA and the Y chromosome, showed a close relationship of these groups with each other and with neighboring geographic groups, irrespective of the language spoken. Moreover, Semitic-speaking North African groups are more distant genetically from Semitic-speaking groups from the Near East and Iran. Similar results were recently obtained in the region of north-west Iran: the Uromian people (Iranian Muslim group) display a particularly close genetic relationship to the Armenians living in the same area (4). Thus, geographical proximity better explains genetic relatedness between populations than linguistic relatedness in this part of the world.

As it was shown in our recently published results (37), based on multivariate classification Iranian Azeris and their close neighbors, Persians and Armenians, form a rather distinct cluster of Middle East origin. This pattern was obtained both while using Principal Coordinate Analysis and Neighbor-Joining method for phylogenetic inferences.

As a whole, the results obtained indicate that the used set of genetic markers is an appropriate tool

for population genetics study of such an ethnically and linguistically complex area as the Middle East. The methods applied allowed distinguishing fine specific features of each population and making inferences about their origin and possible ancient genetic contacts.

We also realize that the Y chromosome represents only one locus in the human genome and describes only one, patrilineal, facet of the genetic history of human populations. The more comprehensive results on the origin, ancient relationships, and migrations of the populations of the Middle East can be achieved while using other complementary genetic systems, i.e. mtDNA and autosomal markers. Nevertheless, the Y chromosome markers provide rather strong information on the genetic history of the populations studied in the frame of this project. In addition, the main outcome of the project is that all the three populations could be considered as indigenous representatives of the area of their inhabitance with very limited genetic influence from East.

## Ethical Considerations

Ethical issues including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc. have been completely observed by the authors.

## Acknowledgements

## References

1. Cavalli-Sforza LL, Menozzi P, Piazza A (1994). *The history and geography of human genes*. Princeton University Press, Princeton, NJ.
2. Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004). The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet*, 74: pp. 532-44.
3. Regueiro M, Cadenas AM, Gayden T, Underhill PA, Herrera RJ (2006). Iran: Tricontinental nexus for Y-chromosome driven migration. *Hum Hered*, 61: 132-43.
4. Lashgary Z, Khodadadi A, Singh Y, Houshmand SM, Mahjoubi F, Sharma P, et al. (2011). Y chromosome diversity among the Iranian religious groups: A reservoir of genetic variation. *Ann of Hum Biol*, Early Online: 1-8, Doi: 10.3109/03014460. 2010.535562.
5. Terreros MC, Rowold DJ, Mirabal S, Herrera RJ (2011). Mitochondrial DNA and Y-chromosomal stratification in Iran: relationship between Iran and the Arabian Peninsula. *J of Hum Genet*, advance online publication, 17 February 2011; doi: 10.1038/ jhg. 2010.p. 174.
6. Edwards IS, Gadd CJ, Hammond NL (1971). *The Cambridge Ancient History*. Vol. 1, Part 2: Early History of the Middle East, Cambridge University Press.
7. Garthwaite GR (2005). *The Persians*. Blackwell's, Oxford. 1 Ed
8. Zvelebil M (1995). At the interface of archaeology, linguistics and genetics: Indo-European dispersals and the agricultural transition in Europe. *J Eur Archaeol*, 3: pp. 33-70.
9. Renfrew C (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. Jonathan Cape, London
10. 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.

11. Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, et al. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet*, 74: 827-45.

12. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ (2002). Y-chromosomal DNA variation in Pakistan. *Am J of Hum Genet,* 70: 1107-124.

13. Zerjal T, Pandya A, Santos FR, Adhikari R, Tarazona-Santos E, Kayser M, et al. (1999). The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. In: Papiha SS, Deka R, Chakraborty R (eds) Genomic diversity: applications in human population genetics. *Plenum,* New York, USA.

14. Johanson L (1998). *The history of Turkic*. In: Johanson L, Crato E (eds) The Turkic languages. London: Routledge.

15. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Bing S, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci*, USA, 98: 10244-249.

16. Shepard EM, Herrera RJ (2005). Iranian STR variation at the fringes of biogeographical demarcation. *For Sci Int*, 158: 140-48.

17. Nasidze I, Quinque D, Rahmani M, Ale-mohamad SA, Stoneking M (2006). Concomitant replacement of language and mtDNA in South Caspian populations of Iran. *Curr Biol*, 16: 668-73

18. Rafiee MR, Sokhansanj A, Naghizadeh MA, Farazmand A (2009). Analysis of Y-chromosomal short tandem repeat (STR) polymorphism in an Iranian Sadat population. *Russ J of Genet,* 45: 969-73.

19. Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas M (2001**)**. Armenian Y chromosome haplotypes reveal strong regional structure within a single ethnonational group. *Hum Genet,* 109: 659-74.

20. Thomas MG, Barnes I, Weale ME, Jones AL, Forster P, Bradman N, Pramstaller PP (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: 124-34.

21. Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, et al. (2004). Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet*, 114: 127-48.

22. Heyer E, Balaresque P, Jobling MA, Quintana-Murci L, Chaix R, Segurel L, Aldashev A, and Hegay T (2009). Genetic diversity and the emergence of ethnic groups in Central Asia. *BMC Genetics*, 10: 49 doi: 10. pp. 1186/1471-2156-10-49.

23. Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, et al. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet*, 67: 1526-43.

24. Thomas MG, Bradman N, Flinn HM (1999). High throughput analysis of 10 microsatellite and 11 binary polymorphisms on the human Y-chromosome. *Hum Genet*, 105: 577-81.

25. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res,* 18: 830-38.

26. Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, et al. (1997). Evaluation of Y-chromosomal STRs- A multicenter study. *Int J Legal Med,* 110: 125-41.

27. Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.

28. Excoffier L, Laval G, Schneider S (2005)**.** Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinfo Online*, 1: 47-50.

29. Raymond M, Rousset F (1995). An exact test for population differentiation. *Evolution*, 49: 1280-83.

30. Thomas MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB (2002). Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet*, 70(6): 1411-20.

31. Thomas MG, Skorecki, K, Ben-Ami, H, Parfitt, T, Bradman, N, Goldstein DB (1998). A genetic date for the origin of Old Testament Priests. *Nature,* 394: 138-40.

32. Thomas M.G, Parfitt T, Weiss DA, Skorecki K, le Roux M, Bradman N, Goldstein, DB (2000). Y chromosomes traveling south: The Cohen Modal Haplotype and the Origins of The Lemba- The "Black Jews of Southern Africa". *Am J Hum Genet,* 66: 674-86.

33. Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci*, USA, 98(9): 5078-83.

34. Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimzu K, Semjidmaa D, Sajantila A., Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997). Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J of Hum Genet,* 60(5): 1174-83.

35. Renfrew C (2002). *The Emerging Synthesis*: the Archaeogenetics of Farming/Language Dispersals and other Spread Zones. In: Bellwood & Renfrew C, (eds) Examining the farming/language dispersal hypothesis. Mc Donald Institute for Archaeological Research, pp. 3-16.

36. Nasidze I, Quinque D, Rahmani M, Alemohamad SA, and Stoneking M (2008). Close genetic relationship between Semitic-speaking and Indo-European-speaking groups in Iran. *Ann Hum Genet*, 72: 241-52.

37. Margaryan A, Andonian L, Harutyunyan A (2010). Patrilineal genetic legacy of Persians, Azeris and Armenians living in Iran in the Middle East context. *Biol J of Armenia*, Supplement 1(62): 55-8.