International Journal of Modern Anthropology
Int. J. Mod. Anthrop. (2017) 10: 115 - 138
Available online at: www.ata.org.tn ; doi: http://dx.doi.org/10.4314/ijma.v1i10.5

Original Synthesis Article

Origin of Azeris (Iran) according to HLA genes

Antonio Arnaiz-Villena[1*], Jose Palacio-Gruber[1*], Ester Muñiz[1], Diego Rey[1] Behrouz Nikbin[2], Hosein Nickman[2], Cristina Campos[1], José Manuel Martín-Villa[1], Ali Amirzargar[2]

*These authors contributed equally for this work and the order of authorship is arbitrary.

[1]*Departament of Immunology, University Complutense, School of Medicine, Madrid Regional Blood Center, Madrid, Spain.* [2]*Molecular Immunology Research Center, Medical School, Tehran University of Medical Sciences, Tehran, Iran*
Corresponding author: Antonio Arnaiz-Villena. Departamento de Inmunología, Facultad de Medicina, Universidad Complutense, Pabellón 5, planta 4. Avda. Complutense s/n, 28040 Madrid, Spain. e-mail: aarnaiz@med.ucm.es , arnaizantonio@gmail.com Web page: http://chopo.pntic.mec.es/biolmol/

(Received 10 September 2017; Accepted 17 October 2017)

Abstract - Azeris from Iran North West provinces (Tabriz city) have been studied for HLA alleles. A total of 8.902 HLA-bearing chromosomes (Chr 6) have been used for comparing their relatedness with other Middle East, Caucasus, Mediterranean and Central Asia populations. Mediterranean, Central Asian and Caucasus extended HLA haplotypes were found, i.e.: A*24:02-B*35:01-DBR1*11:01-DBQ1*03:01 and A*01:02-B*08:01-DRB1*03:01-DQB1*02:01. Genetic distances, Neighbour Joining and Correspondence analyses also showed that Azeris were close to Kurds, who have shown a closer Mediterranean/Caucasus HLA profile, and Gorgan (Turkmen) who have shown a closer Central Asia profile, as expected. It is shown that three different Iranian populations according to Language, History and Geography: Gorgans, Kurds and Azeris are genetically close. In fact, old Azeri language (Adari) was an Iranian language and not a Turkic one, which they nowadays speak. Also, present study does not support "Aryan" invasion from the East in accordance with many other previous studies. Finally, our results are useful for establishing Preventive Medicine programs in Transplantation and HLA and Pharmacogenomics/Disease linkage.

Keywords: HLA, Pharmacogenomics, Disease, Transplantation, Iran, Irak, Georgia, Turkey, Azerbajan, Azeris, Gorgan, Kurds, Yazd, Baloch, Aryan, Dene Caucasian, Zoroastro,Persia.

Introduction

The HLA genetic system consists of several closely linked loci encoding cell surface glycoproteins whose main function is activating immune system response through antigenic presentation that consists of autosomic genes inherited both through maternal and through paternal lineages that are useful as mtDNA and Y chromosome (Y Chr) maternal and paternal markers in Anthropology. It is the most polymorphic genetic system described in humans. New alleles in every single HLA locus have been described since the discovery of this genetic system and the presently available DNA typing and sequencing of these new alleles have increased the variety of HLA allelism (e.g.: 1977 HLA-DRB1 alleles have been described up to December 2016 (Robinson et al. 2015)). Due to the fact that HLA gene frequencies have both a large degree of variability and a remarkable geographical correlation, HLA genes are used as a powerful genetic markers to infer genetic background, ethnical composition of modern human populations and also for tracing migration of ancient ones: certain alleles are very frequent only in specific populations (some specific alleles and haplotypes in Amerindians (Arnaiz-Villena et al. 2010; Arnaiz-Villena et al. 2014; Arnaiz-Villena et al. 2016a; Arnaiz-Villena et al. 2016b)); and also, the strong linkage disequilibrium between HLA neighboring loci demonstrates that certain combinations of contiguous alleles (HLA haplotypes) show a characteristic frequency or are distinctive in certain living populations (Martinez-Laso et al. 2001; Gonzalez-Galarza et al. 2011). Thus, HLA genetic system is a unique tool for studying the origins of relatively isolated groups, like Azeris (also called Azerbaijanis) living not only in Azerbaijan but also in North Iran (Fig 1). In addition, HLA polymorphism is crucial for the compatibility between donor and receptor in organ transplantation and several HLA alleles have been linked to diseases and to response to drug treatments, which accomplishes relationships of certain variants with at least fifteen different pathologies treatment including AIDS (Becquemont, 2010). This is important in personalized treatments design.

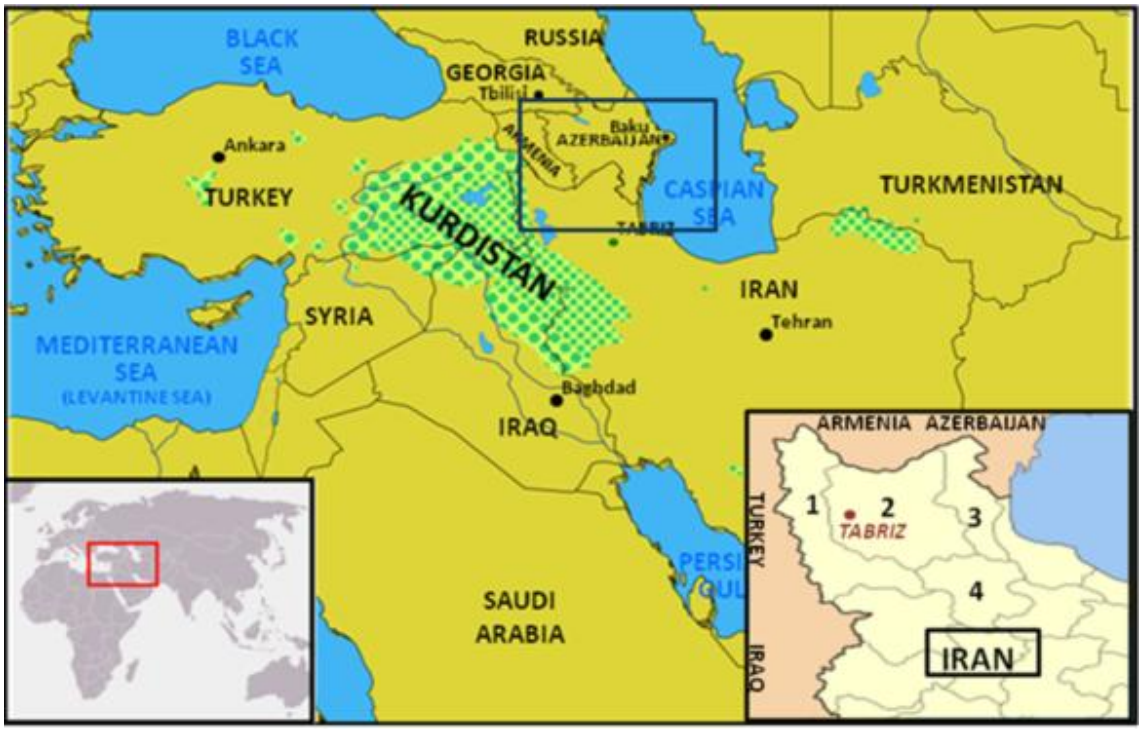

Fig. 1 Map depicting Iranian Azeri provinces and place from where samples were taken (Tabriz city -red dot-) and other relevant Middle East areas.
1: Western Azerbaijan Province; 2: Eastern Azerbaijan Province; 3: Ardabil Province and 4: Zanjan Province

HLA genetic studies in several Iranian ethnic groups, including Azeris, suggest a common genetic substratum with Eastern Mediterraneans, since these groups cluster together and close to Mediterraneans when relatedness analyses are carried out comparing with European and Asian (Siberian) populations (Farjadian et al. 2009). Genetic studies based on mitochondrial DNA (mtDNA) and Y chromosome (Y-chr), have been also carried out on Azeris from Iran and Azerbaijan, but there is not much concordance between these two genetic markers studies in order to infer the Azeri people origins, probably Middle East and/or Central Asia (Quintana-Murci et al. 2004; Grugni et al. 2012).

Azeri people are believed to be of mixed origin, whose more ancient element come from Caucasian indigenous autochthonous populations, i.e.: Mannai, who lived in northwest of modern Iran during the 1st millennium BC and underwent an assimilation

cultural process and also a certain degree of gene input by people of Andronovo culture from Central Asia (Indo-Iranian -Medians and Persians-) (Enciclopaedia Britannica, 2014). Also, historic, linguistic, archaeological and cultural data support the presence of an Iranian substratum in this ethnic group, mainly because of the existence of Old Azeri language (Adari) belonging to Iranian language family that persists in some Azeri dialects and in the regional toponymy (Yarshater, 1988; Frey, 2004). Later, the arrival of Oghuz Turkic tribes coming from Altaic region since 10^{th} century AD that conquered Southwest Asia up to Anatolia Peninsula, led to an uninterrupted "turkification" process in next centuries in Azerbaijan region that caused the substitution of Iranian Old Azeri language for the modern Azeri Turkish, belonging to Altaic language family (Yarshater, 1988).

Most of Azeris live nowadays in Republic of Azerbaijan and Iran. In this second country, this people live mainly in Iranian Azerbaijan (northwestern portion of Iran), formed by four provinces: West Azerbaijan, East Azerbaijan, Ardabil and Zanjan (Fig 1). They amount about 15 or 16 million people, being the first Azeri community in the world and the second ethnic group in Iran by population, after Persians (Central Intelligence Agency, 2013). This community is both rural and urban, living in cities in which Azeris are a majority, like Tabriz or Urmia (Frey, 2004; International Business Publication, 2013). Azeri population of Iran can be also found in other parts of the country, such as Hamadan, Kurdistan and Gilan provinces. Most of Azeri people are Shia Muslims.

We have undertaken the present study in order to: 1) Determine the HLA class I (A and B) and class II (DRB1 and DQB1) allelic Azeri lineages (hereafter "alleles" for simplicity) and specific HLA haplotypes by using PCR-SSOP-Luminex and DNA sequencing, 2) Compare the Azeri HLA allele frequencies with those Central Asian, Siberian, Mediterranean and other worldwide populations (Table 1) with computer programs in order to study the HLA relatedness with peoples most likely to be candidates for Azeris ancestors; this would clarify the still unclear origin of this people, i.e.: groups of genes frequencies comparisons by using genetic distances, Neighbor Joining (NJ) dendrograms and correspondence analyses and finally, 3) Establish the Azeri HLA profile that will be useful for HLA genetic epidemiology (HLA linked diseases, Preventive Medicine), HLA Pharmacogenomics and a virtual future transplant waiting list.

Table 1. Worldwide populations included in the analysis. A total of 8,902 chromosomes were analyzed. (N: number of individuals).

Population	N	Reference	Population	N	Reference
Algerians	51	(Arnaiz-Villena et al. 1995)	Kurds	60	(Amirzargar et al. 2015)
Ashkenazi Jews	80	(Martinez-Laso et al. 1996)	Lebanese	59	(Clayton and Lonjou, 1997)
Armenians	100	(Matevosyan et al. 2011)	Macedonians	178	(Arnaiz-Villena et al. 2001a)
Azeris	97	This study	Manchu	50	(Imanishi et al. 1992c)
Baloch	100	(Farjadian et al. 2004)	Mansi	68	(Uinuk-Ool et al. 2002)
Berbers (Souss)	98	(Izaabel et al. 1998)	Moroccans	98	(Gomez-Casado et al. 2000)
Buryat	25	(Uinuk-Ool et al. 2002)	Moroccan Jews	113	(Roitberg-Tambur et al. 1995)
Chuvash	82	(Arnaiz-Villena et al. 2003)	Negidal	35	(Uinuk-Ool et al. 2002)
Cretans	144	(Arnaiz-Villena et al. 1999)	Non Ashkenazi Jews	80	(Martinez-Laso et al. 1996)
Croatians	105	(Clayton and Lonjou, 1997)	Palestinians	165	(Arnaiz-Villena et al. 2001b)
Evenks	35	(Grahovac et al. 1998)	Russians	200	(Kapustin et al. 1999)
French	179	(Imanishi et al. 1992c)	Sardinians	91	(Imanishi et al. 1992c)
Georgians	119	(Rey et al. 2013)	Spaniards	88	(Martinez-Laso et al. 1995)
Gorgan	69	(Rey et al. 2014)	Spanish Basques	83	(Martinez-Laso et al. 1995)
Iranians [a]	65	(Gonzalez-Galarza et al. 2011)	Svan	80	(Sanchez-Velasco and Leiba-Cobian, 2001)
Iranians [b]	73	(Gonzalez-Galarza et al. 2011)	Turks	245	(Mack and Erlich, 2006)
Italians	284	(Imanishi et al. 1992c)	Todja	22	(Uinuk-Ool et al. 2002)
Japanese	493	(Imanishi et al. 1992c)	Tofalar	43	(Uinuk-Ool et al. 2002)
Khalk Mongolian	202	(Munkhbat et al. 1997)	Tuvinians	197	(Martinez-Laso et al. 2001)
Kets	22	(Grahovac et al. 1998)	Ulchi	73	(Uinuk-Ool et al. 2002)

[a]Iranians from Yazd province
[b]Iranians from Fars province

Material and Methods

- ***Population sample***

97 healthy unrelated volunteers from the city of Tabriz, Iran (Fig 1) were HLA class I and class II typed. The city of Tabriz is the capital of East Azerbaijan province in the Northwest of Iran (38°04′N 46°18′E, see Map, Fig 1) and Azeris constitute the majority of the city and the region population. A written consent to participate in the present study was signed by each individual. All subjects in the study were born in this city and their four grandparents had been born in the same area (International Business Publication, 2013). We compare our data with those of Caucasian European, Mediterranean, Siberian and Oriental populations (these populations are detailed in Table 1), obtaining the genetic distances (comparison was done with 8,902 chromosomes), relatedness dendrograms and correspondence analyses.

- ***HLA genotyping***

HLA class I (A and B) and high resolution HLA class II analysis (DRB1 and DQB1) was performed by PCR-SSOP-Luminex technique (Itoh et al. 2005). This methodology consists of: a) PCR using specific primer pairs as provided by manufacturers (Luminex Corporation, Austin, TX, USA). All of these primers are 5'-biotined and they are specific to determinate sequences of exons 2 and 3 (or only exon 2 for HLA class II) of HLA genes; b) hybridization: products of PCR biotin-labeled were denaturalized at 97 ºC and then were hybridized to complementary DNA probes associated to microbeads; and c) assignation of HLA alleles: complex resulting of hybridization was introduced into a Luminex platform; this system identifies fluorescent intensity of fluorophores on each oligobead group that has hybridized with the biotin-labeled PCR product. Luminex Software assigns HLA alleles for each DNA sample (Itoh et al. 2005). HLA-A, -B, -DRB1, and -DQB1 allele DNA automated sequencing (ABI PRISM 3700/ ABI PRISM 3730. Applied Biosystems; California) was only done when DNA typing yielded ambiguous results.

- ***Statistical analysis***

Statistical analysis was performed with Arlequin v3.0 (Schneider et al. 2000). In summary, this program calculated HLA-A, -B, -DRB1, and -DQB1 allele frequencies,

Hardy-Weinberg equilibrium and the linkage disequilibrium between *n* alleles at *n* different loci. Their level of significance (p) for 2 x 2 comparisons was determined as previously described (Imanishi et al. 1992a; Imanishi et al. 1992b). In addition, the most frequent complete extended haplotypes were deduced from: 1) the 2, 3, and 4 HLA loci haplotype frequencies (Imanishi et al. 1992a; Imanishi et al. 1992b); 2) the previously described haplotypes in other populations (Imanishi et al. 1992a; Imanishi et al. 1992b); and 3) haplotypes if they appeared in two or more individuals and the alternative haplotype was well defined (Imanishi et al. 1992a; Imanishi et al. 1992b). In order to compare phenotype and haplotype HLA frequencies with other populations, the reference tables of the 11^{th} and 12^{th} International HLA Workshops were used (Imanishi et al. 1992c; Clayton and Lonjou, 1997). Phylogenetic trees (dendrograms) were constructed with the allelic frequencies using the Neighbor-Joining (NJ) method (Saitou and Nei, 1987) with the genetic distances between populations (DA) (Nei, 1972), using DISPAN software comprising the programs GNKDST and TREEVIEW (Nei, 1973; Nei et al. 1983). Correspondence analysis in three dimensions and its bidimensional representation was carried out using the VISTA v5.05 computer program (Young and Bann, 1996) (http:/forrest.psych.unc.edu). Correspondence analysis consists of a geometric technique that may be used for displaying a global view of the relationships among populations according to HLA (or other) allele frequencies. This methodology is based on the genetic distances (DA) variance among populations (similar to the classical principal components methodology) and of a statistical visualization of the differences.

Results

- ***HLA allele and class II haplotype frequencies found in Azeri population: comparisons with other populations***

The expected and observed gene frequency values for HLA-A, -B, -DRB1, and -DQB1 loci do not differ significantly and the population is found in Hardy–Weinberg equilibrium (data not shown). HLA allele frequencies found in the sampled population are shown in Table 2. Twenty-two different HLA-A and thirty-eight different HLA-B alleles were found in our Azeri sample. Only ten HLA-A alleles and six HLA-B alleles had significant frequencies, higher than 5% (A*01:01, A*01:02, A*02:01, A*03:01, A*11:01, A*24:02, A*30:01, A*32:01, A*33:01, A*68:01, B*18:01, B*35:01, B*41:01, B*44:02, B*51:01, and B*52:01). Twenty different HLA-DRB1 and fourteen different HLA-DQB1 alleles were found (Table 2). Six HLA-DRB1 and seven HLA-

DQB1 alleles had frequencies higher than 5% (DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*11:01, DRB1*14:01, DRB1*15:01, DQB1*02:01, DQB1*03:01, DQB1*03:02, DQB1*04:01, DQB1*05:01, DQB1*06:01 and DQB1*06:02).

Table 2. HLA-A, -B, -DRB1, and -DQB1 allele frequencies in Azeris population.

Alleles	Allele frequencies%	Alleles	Allele frequencies %	Alleles	Allele frequencies %
HLA-A		18:09	0.5	03:08	0.5
01:01	11.9	27:01	1.6	04:01	15.5
01:02	6.2	35:01	15.0	07:01	10.8
02:01	13.4	35:05	0.5	08:01	3.0
02:50	4.1	35:10	0.5	09:01	1.6
02:52	0.4	35:19	0.5	10:01	3.6
03:01	9.8	37:01	1.6	11:01	15.0
03:02	1.6	38:01	4.6	12:01	1.6
11:01	6.2	39:01	1.0	13:01	4.1
23:01	0.4	40:01	1.6	13:02	1.6
24:02	12.4	41:01	5.2	13:03	1.0
24:03	1.6	44:02	5.2	13:05	0.5
26:01	2.6	44:18	0.5	13:08	0.5
26:19	0.5	46:01	0.5	14:01	6.7
29:01	3.6	49:01	4.1	14:02	1.0
30:01	5.2	50:01	4.1	15:01	15.0
31:01	1.6	51:01	8.8	16:01	4.1
32:01	5.7	51:06	1.0		
33:01	5.7	51:07	0.5	**HLA-DQB1**	
34:02	0.4	51:08	1.0	02:01	16.5
66:01	0.4	52:01	8.3	03:01	20.1
68:01	5.7	53:01	0.5	03:02	8.8
69:01	0.6	54:01	0.5	03:03	4.6
		55:01	3.6	03:05	2.1
HLA-B		56:01	0.5	03:07	2.6
07:02	4.6	57:01	1.6	03:12	0.4
08:01	3.0	58:01	2.6	04:01	5.7
13:01	3.6	58:08	0.5	05:01	16.5
14:01	3.1	78:01	0.5	05:02	3.6
15:01	2.1			06:01	8.3
15:10	0.5	**HLA-DRB1**		06:02	5.2
15:17	0.5	01:01	4.6	06:03	4.1
15:24	0.5	03:01	8.8	06:04	1.5
18:01	5.2	03:02	0.5		

Two types of analyses were done in order to compare Azeri HLA frequencies with other World population frequencies: 1) by using DRB1 allele frequencies data, and 2) by using DRB1-DQB1 haplotype frequencies (not shown). It was not possible to carry out a study comparing HLA class I allele frequencies or HLA class I and II conjointly due to the lack of class I studies in many worldwide populations (see Table 1). Thus, a single DRB1 study was also carried out in order to compare the Azeri HLA population frequencies with those of as many as possible populations. DA (plain) genetic distances based on these data between Azeri people and populations included in the analysis have been calculated and are depicted in Table 3: it is shown that Kurds and Gorgan (both from Iran) present the closest values and the closest value is that of Azeri-Kurd distance. This is in accordance with geography since Azeri sample (Tabriz) and Kurd sample (Kurdistan Province, Iran) are adjacent provinces. Gorgan sample is place far at East Iranian/Turkmenistan border. More distant Azeri values are followed by Europeans-Mediterraneans, such as French, Russians, Croatians, Italians and Turks, populations from Caucasus, like Georgians and Svan, and western Siberians, such as Tuvinians, Chuvash, Todja and Tofalar.

Table 3. Genetic distances (DA) between Azeris and other populations ($\times 10^{-2}$) obtained by using HLA-DRB1 allele frequencies.

Population	DA
Kurds	6.12
Gorgan	12.02
French	16.10
Russians	16.23
Croatians	17.74
Georgians	18.93
Tuvinians	19.28
Italians	19.45
Chuvash	19.75
Todja	21.12
Turks	21.75
Tofalar	22.51
Svan	22.74
Spaniards	22.98
Mansi	23.21
Khalk Mongolian	24.02
Macedonians	24.13
Spanish Basques	24.38
Iranians 2	26.25
Buryat	27.04
Armenians	27.35
Cretans	28.25
Berbers	29.62
Manchu	31.84
Algerians	31.92
Palestinians	32.22
Iranians 1	34.59
Moroccans	35.59
Sardinians	36.91
Ashkenazi Jews	37.38
Moroccans Jews	38.26
Non Ashkenazi Jews	38.39
Baloch	38.59
Evenks	39.54
Lebanese	39.72
Ulchi	39.83
Negidal	44.06
Kets	46.18
Japanese	48.76

Fig. 2 Neighbor-Joining (NJ) dendrogram showing relatedness between Azeris and other World populations.

Genetic distances between populations (DA) were calculated by using HLA-DRB1 (high resolution). Data from World populations were taken from references stated in Table 1.

The NJ relatedness dendrogram based on HLA-DRB1 allele frequencies (Fig 2) separates populations in two well-differentiated clusters. One of them groups North and South Mediterraneans (Europeans and Africans), Middle Easterners, Caucasians and western Siberians. The second cluster grouped the rest of analyzed populations (central and eastern Siberians and Orientals). Azeris are integrated in the first cluster, together with Gorgan (Iranian Turkmen population (Rey et al. 2014)) and Kurds (Armirzargar et al. 2015), and in intermediate position between Iranian populations (Gonzalez-Galarza et al. 2011), and western Siberians: Russian Chuvash (who live near lower Volga River,

North Caspian Sea (Arnaiz-Villena et al. 2003)), Russian Siberian Mansi (from western Siberia (Uinuk-Ool et al. 2002)), Russian-Mongols Buryat (from Baikal Lake region (Uinuk-Ool et al. 2002)) and Russian Siberian Todja (from western Siberia, inhabiting in the northeastern part of Tuva Republic (Uinuk-Ool et al. 2002)). Correspondence analyses based on HLA-DRB1 allele frequencies (Fig 3) show similar results. Also, two groups are clearly defined according to first dimension that explains most of the variability among populations. The first one includes Oriental and Siberian populations and the second one is divided according to the second dimension into two subgroups. First subgroup clusters Europeans, northern Mediterranean, Caucasus, Iranian populations (near to Kurds and Chuvash), while second subgroup clusters southern and eastern Mediterraneans; Azeris are located close to Gorgan Turkmen in an intermediate situation between these two major groups in this analysis but closer to the first one (Fig 3).

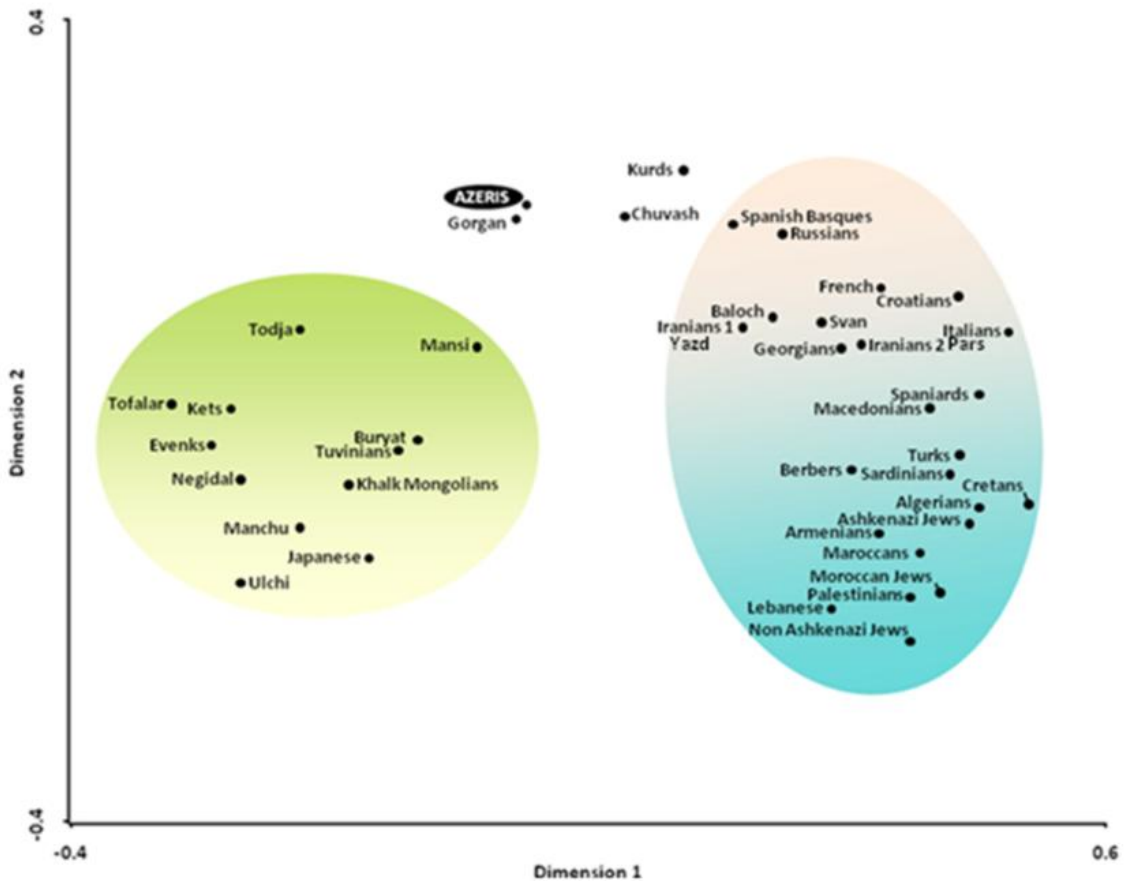

Fig. 3 Correspondence analysis showing a global view of the relationship between Azeris and Mediterranean, Siberians, and other World populations according to HLA-DRB1 allele frequencies in three dimensions (bidimensional representation).

Finally, HLA class II haplotype frequencies were used to develop NJ dendrogram and correspondence analysis too. The NJ relatedness analysis also groups Azeris with Mediterraneans, Middle Easterners, and Caucasians, close to Kurds and Gorgan, and separated from Siberians (second cluster), except Chuvash and Tuvinians. Correspondence analysis gives identical results, since our Azeri sample is located in the cluster that groups Europeans, Mediterraneans and Caucasians, defined according to first dimension, close to the other North Iranian populations (Kurds and Gorgan), although Azeris tends to be in an intermediate situation between this main group and Siberian cluster.

- ***HLA-A, -B, -DRB1 and -DQB1 extended haplotype analysis in Azeris: comparison with other populations***

The nine most frequent four HLA loci haplotype combinations (A-B-DRB1-DQB1) were calculated (Table 4); they represent 14.4 % of all haplotypes. Class I haplotype A*24:02-B*35:01 is present in association with DRB1*16:01-DQB1*05:01 and DRB1*11:01-DQB1*03:01 (the most frequent extended haplotype). This Class II haplotype DRB1*11:01-DQB1*03:01 is also associated with A*03:01-B*18:01 and A*02:01-B*35:01, and other Class II haplotype, DRB1*07:01-DQB1*02:01, is present in association with A*02:01-B*50:01, and A*30:01-B*13:01. The Azeri extended HLA haplotypes obtained allow their comparison with previously reported ones in other populations (Table 4 and its footnote). HLA-A-B-DRB1-DQB1 extended haplotypes are either from Central Asian-Siberian or Middle East-Mediterranean origin.

Table 4. The nine most frequent HLA-A, -B, -DRB1 and -DQB1 extended haplotypes in Azeris.

Haplotypes	HF (%)	Possible origin
A*24:02-B*35:01-DRB1*11:01-DQB1*03:01a	3.1	Central Asian/Mediterranean
A*01:02-B*08:01-DRB1*03:01-DQB1*02:01b	2.1	Middle East/Caucasus
A*03:01-B*07:02-DRB1*15:01-DQB1*06:02c	1.6	North Eurasian
A*03:01-B*18:01-DRB1*11:01-DQB1*03:01d	1.6	Central Asian
A*33:01-B*14:01-DRB1*01:01-DQB1*05:01e	1.6	Central Asian/Mediterranean
A*02:01-B*35:01-DRB1*11:01-DQB1*03:01f	1.1	Asian/Mediterranean
A*02:01-B*50:01-DRB1*07:01-DQB1*02:01g	1.1	Asian/Mediterranean
A*24:02-B*35:01-DRB1*16:01-DQB1*05:01h	1.1	Central Asians/Mediterranean
A*30:01-B*13:01-DRB1*07:01-DQB1*02:01i	1.1	Central Asians/Mediterranean

HF: haplotipic frequency.

[a]Found in Gorgan (HF: 2.2) and with HLA-Class I in low resolution in Chuvash (HF: 1.2), Palestinians (HF: 1.2) and in other Azeri population (HF: 6.0)

[b]Found with HLA Class I in low resolution in Chuvash (HF: 1.2) and in other Azeri population (HF: 6.0)

[c]Found in population of South Ireland (HF: 4.3), Northwest England (HF: 4.2), with HLA Class I in low resolution in Chuvash (HF: 4.9) and in other Azeri population (HF: 4.5)

[d]HLA Class II split found in Georgians (HF: 12.6), Saami (HF: 11.7), Turks (HF: 9.8), Tofalar (HF: 9.3), Todja (HF: 9.1), Cretans (HF: 8.9), Kets (HF: 8.8), Evenks (HF: 8.6) and Chuvash (HF: 4.8)

[e]HLA Class I split found in Tunisians (HF: 3.0). HLA Class II split found in Chuvash (HF: 12.1), Saami (HF: 9.2), Tuvinians (HF: 9.0), Georgians (HF: 8.8), Khoton Mongolians (HF: 8.1), Mansi (HF: 6.6), Cretans (HF: 6.3), Aleutians (HF: 5.6) and Iranians from Yazd (HF: 5.4)

[f]Found in Kurds (HF: 1.7). HLA Class I split found in population of Oman (HF: 4.9). HLA Class II split like "d"

[g]Found in Kurds (HF: 1.7) and in Armenians without HLA-DQB1 (HF: 1.0). HLA Class I split found in Moroccans (HF: 3.6). HLA Class II split found in Buryat (HF: 22.0), Berbers (HF: 20.1), Mansi (HF: 16.9), Iranians from Yazd (HF: 15.4), Kets (HF: 11.8), Moroccans (HF: 12.1), Cretans (HF: 10.4), Turks (HF: 8.9), Chuvash (HF: 5.4) and in other Azeri population (HF: 5.5)

[h]HLA Class I split found in Yupik from Alaska (HF: 11.7) and Lakota Sioux (HF: 4.3). HLA Class II split found in Tunisians (HF: 2.0)

[i]HLA Class II split like "h"

References: (Martinez-Laso et al. 2001; Gonzalez-Galarza et al. 2011; Matevosyan et al. 2011; Isabel et al. 1998; Uinuk-Ool et al. 2002; Arnaiz-Villena et al. 2003; Arnaiz-Villena et al. 1999; Grahovac et al. 1998; Rey et al. 2013; Rey et al. 2014; Munkhbat et al. 1997; Gomez-Casado et al. 2000; Arnaiz-Villena et al. 2016; Mack and Erlich, 2006; Moscoso et al. 2008; Farjadian and Ghaderi, 2007).

Discussion

- ***HLA genetics study in Azeris: relation to disease and pharmacogenomics***

The genetic distances studies based on both HLA allele and class II haplotype frequencies (i.e.: NJ relatedness dendrogram and correspondence analysis, Fig 2 and Fig 3), place Azeri sample in the Mediterranean cluster close to Kurds, Gorgan, Chuvash (South Russia, towards North Caucasus), Iranians and Caucasus populations (Svan and Georgians) (Fig 2 and Fig 3). Furthermore, HLA-DRB1 correspondence analysis shows Azeris close to Iranian populations like Baloch and Iranians from Yazd, Gorgan Turkmen and Kurds (the closest population according to plain genetic distances), but in a half-way position between Mediterraneans and Western and Central Siberians, such as Mansi or Todja, together with Gorgan, Kurds and Chuvash (South Russian towards North Caucasus).

Extended HLA haplotypes are found in Azeris in addition to the first three ones, see Table 4: A*03:01-B*18:01-DRB1*11:01-DQB1*03:01, A*33:01-B*14:01-DRB1*01:01-DQB1*05:01, A*02:01-B*35:01-DRB1*11:01-DQB1*03:01, A*02:01-B*50:01-DRB1*07:01-DQB1*02:01, A*24:02-B*35:01-DRB1*16:01-DQB1*05:01 and A*30:01-B*13:01-DRB1*07:01-DQB1*02:01. Some of the HLA class I or class II splits of these extended haplotypes are shared with other populations (mainly Mediterraneans, Caucasians and Siberians, see Table 4 footnote), but none of them with Iranian populations from Yazd and Shiraz. This fact suggests that some genetic differences exist with old Iranian surrounding populations and make this study worthwhile for establishing future transplant programs in Azeri population; HLA-DRB1-DQB1 haplotypes have been found common to Azeris and Kurds (Farjadian and Ghaderi, 2007).

Also HLA and disease epidemiology studies (Shiina et al. 2009) are addressed with the present study results. Regarding autoimmune diseases, the presence of certain HLA-DRB1 alleles has been confirmed as a risk factor of type I diabetes (Vicario et al. 1992), such as DRB1*03:01 and DRB1*04:01 (this allele was specifically found linked concretely in Azeri population (Ahmedov et al. 2006)). DRB1*04:01 is one of the most HLA-DRB1 alleles in our sample; hence the relevant presence of diabetogenic alleles in

this population could be useful for establishing preventive medicine programs. In regards to infectious diseases, better prognosis of malaria caused by *Plasmodium vivax* is linked to the presence of A*01:02 (Arevalo-Herrera et al. 2002). This malaria is endemic in North area of Zagros Mountains, corresponding to East Azerbaijan and Ardabil provinces, and A*01:02 is the most frequent HLA-A allele in our Azeri sample. Therefore, a relation between this pathology and the relevant presence of this allele in the main ethnic group of this region could exist. Finally, the study of HLA genetic profiles in this particular population could be useful for Pharmacogenomics and drug treatments. Different therapeutic drugs, such as carbamazepine (anticonvulsant) and co-amoxiclav (combination of antibiotics), may differentially affect Azeris, with possible adverse reactions, according to the presence of frequent HLA alleles A*24:02 and DRB1*15:01, respectively (Becquemont, 2010).

- ***Azeris and Iranians***

The origin of Iranian peoples as a hypothetical eastern-people (Indo-Iranian) invasion has been very much debated. The invasion by these tribes is only based on linguistics, so the fact that this invasion existed as an "indoiranization" process through an "elite" of rulers that imposed their culture, including language is possible. The concept "Aryan invasion" is neither genetically nor archaeologically based. At the beginning of the 1st millennium B.C. (1000-900 B.C.) the Medes became an empire conquering the Zagros Mountains and lower Mesopotamia (von Soden, 1994). The Persian Achaemenid dynasty that conquered the Median Empire, also with an Indo-Iranian origin, proclaimed Zoroastrianism as the official religion, expanded westwards and took over most of Middle East, including Anatolia and Egypt. Alexander the Great brought it to an end in the 4th century B.C establishing the Seleucid dynasty. Later, Arabs also brought to an end the Persian Empire by 638 A.D. The official religion of Zoroastrian was replaced by Islam but the Old Persian language survived (Paul, 2014). The concept "Aryan invasion" is not neither genetically or archaeologically based (Arnaiz-Villena et al. 2002).

Despite the fact that Azeris do not share any of the most frequent extended haplotypes with Iranians, frequent HLA class II haplotypes in our sample are also common with Iranians from Yazd (Table 4) and relatedness analyses present these populations to be close (Fig 2 and Fig 3). Previous HLA studies show that modern Iranians are close to other Middle East-Mediterranean populations Macedonians,

Cretans and Turkish (Farjadian et al. 2009; Arnaiz-Villena et al. 2002). Thus, the genetic data support the hypothesis that present day Iranian main genetic stock comes from the ancient autochthonous people and a genetic input from eastern people would be a minor one (Arnaiz-Villena et al. 2001c; Arnaiz-Villena et al. 2002).

- ***Azeris and Anatolian and Caucasus populations***

Genetic substratum of Anatolian people (Turkish) has been defined as Mediterranean with a little genetic but a high cultural influence of Oghuz Turks (Arnaiz-Villena et al. 2001c; Arnaiz-Villena et al. 2002). Despite several Anatolian invasions (Sellier and Sellier, 1993), the present day Turkish HLA profile reflects an older Mediterranean substratum not far from Cretans, Sardinians, Macedonians and Croatians (Fig 2 and Fig 3). Oghuz last invaders carried out a so-called "elite" invasion: a relatively small number of people with higher cultural and military abilities imposed a foreign culture and language, Turkish (Arnaiz-Villena et al. 2001c; Arnaiz-Villena et al. 2002).

Georgians are related with Mesopotamian and Anatolian peoples since Assyrians and Hittites took shelter in the region by 1200 B.C. and mixed with autochthonous Caucasus population. Peoples from North, such as Scythians and Cimmerians, invaded Georgia by 730 A.D. and many people fled to highlands. After that, the main Caucasian Iberian tribes Tibal and Mushki formed Georgian population nucleus, in spite of war with southern Urartian tribes, Armenian probable ancestral population (Arnaiz-Villena et al. 2002; Rey et al. 2013). Georgians speak Georgian, a member of Kartvelian family language (also called South Caucasian), which may be included into Dene-Caucasian languages (Rey et al. 2013; Arnaiz-Villena et al. 2001d). Present-day Armenians have their own alphabet and speak a doubtful Indo-European language probably imposed by the Medes (whose language is completely unknown) but their original language (Hurro-Urartian) probably was related to Kartvelian and also belonged to the Dene Caucasian group.

HLA genetic distances observed in Azeris put them into Mediterranean stock together with Turkish and Caucasian populations (Table 3, Fig 2 and Fig 3). This is consistent with data reported by previous HLA studies (Matevosyan et al. 2011; Arnaiz-Villena et al. 2001c; Arnaiz-Villena et al. 2002) and Y-chr studies in Azeris and Kurds from Iran that identify the relevant presence of haplogroups originated in Middle-East (Anatolia or Mesopotamia) showing a close association with Jews, Lebanese, Turkish

and Armenians (Wells et al. 2001; Nasidze et al. 2005; Andonian et al. 2011; Hennerbichler, 2012). Present day Azeris and Turkish people speak close related languages belonging to Altaic language family (Oghuz Turkic) that substituted old non-Turkic languages.

- ***Azeris and Central Asian populations***

Central Asian corridor is a wide geographical area that extends from Caspian Sea to West China and from South Siberia to North India and have been postulated an important role as a communication way for human groups, allowing several waves of migration of different peoples at different times. Tuvinians, Todja and Tofalar ethnic groups live in Russian Republic of Tuva, which borders South with Mongolia and is bounded in the northwest by Altai Mountains. The first ones form a relatively isolated group (Martinez-Laso et al. 2001). Todja and Tofalar are morphologically and anthropologically similar but distinct from neighboring Tuvinians, and are believed to be of mixed origin composed of Turkic populations and North Siberian Tribes, like Kets or Samoyeds (Uinuk-Ool et al. 2002). These populations are nomadic cattle breeders and speak languages belonging to Turkic language family. Mansi people are a human group related to Altaic Central Asian populations, but inhabiting the Region of the Khanty-Mansi, in Western Siberia and speak Mansi, a Finno-Ugric-group language (Uinuk-Ool et al. 2002).

Genetic studies based on mtDNA and Y-chr confirms a high degree of haplogroups diversity in Central Asian populations, suggesting that these groups are between the most ancient ones of the continent (Wells et al. 2001; Quintana-Murci et al. 2004). Previous HLA studies are consistent with these conclusions and show Siberian and Central Asian population as a monophyletic cluster, in which Altaic groups are close to European populations. HLA genetic distances observed in our study present low values in Altaic populations and Mansi with respect to Azeris, which are shown in a half-way position between Mediterranean and Central Asian, but much close to Todja, Tuvinians and Mansi, in correspondence analyses (Fig 3). These results suggest that "turkification" process caused by Oghuz Turkic tribes could also contribute to the genetic background of Azeri people, as other genetic and historic data argue (Yarshater, 1988; Schonberg et al. 2011).

- ***Iranian Azeri, Gorgan and Kurds populations***

Azeries, Kurds (Fig 1) and Gorgan from Iran are geneticallty close (Fig 2 and Fig 3). Recently analyzed Kurds from Iraq, Georgia and Iran (Arnaiz-Villena et al. 2017) seem to keep a high degree of genetic relatedness probably due to relative isolation due to language factor. Plain genetic distances (Table 3) shows that Iran Kurds have the closest genetic distance to Azeries; this is in accordance with geography (Fig 1), since Kurds from Iran sample were taken from Iran province of Kurdistan which is quite below or Iran province of West Azerbaijan. These three populations when analyzed together with Asian and Mediterraneans are close to each other and show relatedness with both Mediterraneans and Asians (Fig 2, Fig 3, Table 1 and Table 4). However, Kurds seem to be closer to a Mediterranean ancient substratum (Arnaiz-Villena et al. 2001c; Shiina et al. 2009; Arnaiz-Villena et al. 2017). Correspondence analysis (Fig 3) clearly shows a certain degree of admixture of Iranian Azeries, Kurds and Gorgan with Mediterranean and eastern genes.

Acknowledgments

This work was supported in part by Grants from the Spanish Ministry of Health and Economy PI14/01067 and European FEDER funds.

References

Ahmedov G, Ahmedova L, Sedlakova P and Cinek, O. (2006) Genetic association of type 1 diabetes in an Azerbaijanian population: the HLA-DQ, -DRB1*04, the insulin gene, and CTLA4. *Pediatr. Diabetes.* **7**, 88-93.

Andonian L, Rezaie S, Margaryan A, Farhud D, Mohammad K, Naieni KH. et al. (2011) Iranian Azeri's Y-Chromosomal Diversity in the Context of Turkish-Speaking Populations of the Middle East. *Iran. J. Public Health.* **40**, 119-123.

Arevalo-Herrera M, Valencia AZ, Vergara J, Bonelo A, Fleischhauer K, Gonzalez JM. et al. (2002) Identification of HLA-A2 restricted CD8(+) T-lymphocyte responses to Plasmodium vivax circumsporozoite protein in individuals naturally exposed to malaria. *Parasite Immunol.* **24**, 161-169.

Armirzargar A, Rey D, Muñiz E, Palacio-Gruber J, Nikbin B, Nicknam H. et al. (2015) Kurds HLA Genes: Its Implications in Transplantation and Pharmacogenomics. *Open Med. J.* ***2,*** 43-47.

Arnaiz-Villena A, Benmamar D, Alvarez M, Diaz-Campos N, Varela P, Gomez-Casado E. et al. (1995) HLA allele and haplotype frequencies in Algerians. Relatedness to Spaniards and Basques. *Hum. Immunol.* **43**, 259-268.

Arnaiz-Villena A, Iliakis P, Gonzalez-Hevilla M, Longas J, Gomez-Casado E, Sfyridaki K. et al. (1999) The origin of Cretan populations as determined by characterization of HLA alleles. *Tissue Antigens.* **53**, 213-226.

Arnaiz-Villena A, Dimitroski K, Pacho A, Moscoso J and Gomez-Casado, E. (2001a) HLA genes in Macedonians and the sub-Saharan origin of the Greeks. *Tissue Antigens.* **57,** 118-127.

Arnaiz-Villena A, Elaiwa N, Silvera C, Rostom A, Moscoso J, Gomez-Casado E. et al. (2001b) The origin of Palestinians and their genetic relatedness with other Mediterranean populations. stml. http://commons.wikimedia.org/wiki/File:Palestinians_hla.pdf; Accessed October 2016.

Arnaiz-Villena A, Karin M, Bendikuze N, Gomez-Casado E, Moscoso J, Silvera C. et al. (2001c) HLA alleles and haplotypes in the Turkish population: relatedness to Kurds, Armenians and other Mediterraneans. *Tissue Antigens.* **57**, 308-317.

Arnaiz-Villena A, Martinez-Laso J and Alonso-Garcia, J. (2001d) The correlation between languages and genes: the Usko-Mediterranean peoples. *Hum. Immunol.* **62**, 1051-1061.

Arnaiz-Villena A, Gomez-Casado E and Martinez-Laso, J. (2002) Population genetic relationships between Mediterranean populations determined by HLA allele distribution and a historic perspective. *Tissue Antigens.* **60**, 111-121.

Arnaiz-Villena A, Martinez-Laso J, Moscoso J, Livshits G, Zamora J, Gomez-Casado E. et al. (2003) HLA genes in the Chuvashian population from European Russia: admixture of Central European and Mediterranean populations. *Hum. Biol.* **75**, 375-392.

Arnaiz-Villena A, Areces C, Gomez-Prieto P, Parga-Lozano C, Moreno E, Abd-El-Fatah-Khalil S. et al. (2010) The peopling of the Americas: a complex issue for Amerindian, Na-dene, Aleut and Eskimo first inhabitants. *Int. J. Mod. Anthrop.* **3,** 65-79

Arnaiz-Villena A, Areces C, Enriquez-de-Salamanca M, Abd-El-Fatah-Khalil S, Marco J, Muñiz E. et al. (2014) Pacific Islanders and Amerindian relatedness according to HLA autosomal genes. *Int. J. Mod. Anthrop.* **7,** 44-67

Arnaiz-Villena A, Palacio-Gruber J, Muñiz E, Campos C, Alonso-Rubio J, Gomez-Casado E. et al. (2016a) HLA genes in Chimila Amerindians (Colombia), the Peopling of America and Medical implications. *Int. J. Mod. Anthrop.* **9**, 91-116

Arnaiz-Villena A, Muñiz E, Palacio-Gruber J, Campos C, Alonso-Rubio J, Gomez-Casado E. et al. (2016b) Ancestry of Amerindians and its impact in anthropology, transplantation, HLA pharmacogenomics and epidemiology by HLA study in Wiwa Colombian population. *Open Med. J.* **3**, 269-285.

Arnaiz-Villena A, Palacio-Grüber J, Muñiz E, Campos C, Alonso-Rubio J, Gomez-Casado E. et al. (2017) Genetic HLA Study of Kurds in Iraq, Iran and Tbilisi (Caucasus, Georgia): Relatedness and Medical Implications. *PLoSONE.* **12**, e0169929. doi:10.1371/journal.pone.0169929.

Becquemont L. (2010) HLA: a pharmacogenomics success story. *Pharmacogenomics.* **11**, 277-281.

Central Intelligence Agency. (2013) The World Factbook. Washington DC, USA: Central Intelligence Agency.

Clayton J and Lonjou, C. (1997) Allele and Haplotype frequencies for HLA loci in various ethnic groups. In: *Genetic diversity of HLA. Functional and medical implications* (ed. By D. Charron), pp. 665-820. EDK, Paris, France.

Encyclopaedia Britannica. Mannai. Encyclopaedia Britannica, Inc. (2014) http://global.britannica.com/EBchecked/topic/362506/Mannai; Accessed 15 April 2014.

Farjadian S, Naruse T, Kawata H, Ghaderi A, Bahram S and Inoko, H. (2004) Molecular analysis of HLA allele frequencies and haplotypes in Baloch of Iran compared with related populations of Pakistan. *Tissue Antigens.* **64**, 581-587.

Farjadian S and Ghaderi, A. (2007) HLA class II similarities in Iranian Kurds and Azeris. *Int. J. Immunogenet.* **34**, 457-463

Farjadian S, Ota M, Inoko H and Ghaderi, A. (2009) The genetic relationship among Iranian ethnic groups: an anthropological view based on HLA class II gene polymorphism. *Mol. Biol. Rep.* **36**, 1943-1950.

Frey RN. (2004) Peoples of Iran (1) A general survey. In: Yarshater, E. (ed. Encyclopaedia Iranica). Routledge and Kegan Paul, pp. 321-326. London, UK.

Gomez-Casado E, del Moral P, Martinez-Laso J, Garcia-Gomez A, Allende L, Silvera-Redondo C. et al. (2000) HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens.* **55**, 239-249.

Gonzalez-Galarza FF, Christmas S, Middleton D and Jones, AR. (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res.* 39 (Database issue), D913-D919.

Grugni V, Battaglia V, Hooshiar KB, Parolo S, Al Zahery N, Achilli A. et al. (2012) Ancient migratory events in the Middle East: new clues from the Y-chromosome variation of modern Iranians. *PLoS. One.* **7**, e41252. https://doi.org/10.1371/journal.pone.0041252

Grahovac B, Sukernik RI, O'hUigin C, Zaleska-Rutczynska Z, Blagitko N, Raldugina O. et al. (1998) Polymorphism of the HLA class II loci in Siberian populations. *Hum. Genet.* **102**, 27-43.

Hennerbichler F. (2012) The origin of Kurds. *Advances in Anthropology.* **2**, 64-79.

Imanishi T, Akaza T, Kimura A, Tokunaga K and Gojobori, T. (1992a) Estimation of allele and haplotype frequencies for HLA and complement loci. In: *HLA 1991*. (ed. By K.Tsuji et al.), pp. 76-79. Oxford University Press, Oxford, UK.

Imanishi T, Akaza T, Kimura A, Tokunaga K and Gojobori, T. (1992b) Genetic relationships among various human populations indicated by MHC polymorphisms. In: *HLA 1991*. (ed. By K.Tsuji et al.), pp. 627-632. Oxford University Press, Oxford, UK.

Imanishi T, Akaza T, Kimura A, Tokunaga K and Gojobori, T. (1992c) Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: *HLA 1991*. (ed. By K.Tsuji et al.), pp. 1065-1220. Oxford University Press, Oxford, UK.

International Business Publication. (2009) Iran: Country Study Guide. International Business Publication USA. Washington DC, USA.

Itoh Y, Mizuki N, Shimada T, Azuma F, Itakura M, Kashiwase K. et al. (2005) High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics.* **57**, 717-729.

Izaabel H, Garchon HJ, Caillat-Zucman S, Beaurain G, Akhayat O, Bach JF. et al. (1998) HLA class II DNA polymorphism in a Moroccan population from the Souss, Agadir area. *Tissue Antigens.* **51**, 106-110.

Kapustin S, Lyshchov A, Alexandrova J, Imyanitov E and Blinov, M. (1999) HLA class II molecular polymorphisms in healthy Slavic individuals from North-Western Russia. *Tissue Antigens.* **54**, 517-520.

Mack SJ and Erlich, HA. (2006) 13th International Histocompatibility Workshop Anthropology/Human Genetic Diversity Joint Report. In: Immunology of the Human MHC. Proceedings of the 13th International Histocompatibility Workshop and Conference. International Histocompatibility (ed. by Hansen, J. A.), pp. 567-766; Working Group Press, Seattle, USA.

Martinez-Laso J, de Juan D, Martinez-Quiles N, Gomez-Casado E, Cuadrado E and Arnaiz-Villena, A. (1995) The contribution of the HLA-A, -B, -C and -DR, -DQ DNA typing to the study of the origins of Spaniards and Basques. *Tissue Antigens.* **45**, 237-245.

Martinez-Laso J, Gazit E, Gomez-Casado E, Morales P, Martinez-Quiles N, Alvarez M. et al. (1996) HLA DR and DQ polymorphism in Ashkenazi and non-Ashkenazi Jews: comparison with other Mediterraneans. *Tissue Antigens.* **47**, 63-71.

Martinez-Laso J, Sartakova M, Allende L, Konenkov V, Moscoso J, Silvera-Redondo C. et al. (2001) HLA molecular markers in Tuvinians: a population with both Oriental and Caucasoid characteristics. *Ann. Hum. Genet.* **65**, 245-261.

Matevosyan L, Chattopadhyay S, Madelian V, Avagyan S, Nazaretyan M, Hyussian A. et al. (2011) HLA-A, HLA-B, and HLA-DRB1 allele distribution in a large Armenian population sample. *Tissue Antigens.* **78**, 21-30.

Moscoso J, Crawford MH, Vicario JL, Zlojutro M, Serrano-Vela JI, Reguera R. et al. (2008) HLA genes of Aleutian Islanders living between Alaska (USA) and Kamchatka (Russia) suggest a possible southern Siberia origin. *Mol. Immunol.* **45**, 1018-1026.

Munkhbat B, Sato T, Hagihara M, Sato K, Kimura A, Munkhtuvshin N. et al. (1997) Molecular analysis of HLA polymorphism in Khoton-Mongolians. *Tissue Antigens.* **50**, 124-134.

Nei M. (1972) Genetic distances between populations. *Am. Nat.* **106**: 283-292.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA.* **70**, 3321-3323.

Nei M, Tajima F and Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* **19,** 153-170.

Nasidze I, Quinque D, Ozturk M, Bendukidze N and Stoneking, M. (2005) MtDNA and Y-chromosome variation in Kurdish groups. *Ann. Hum. Genet.* **69**, 401-412.

Paul L. (2013) Persian Language I. Early New Persian. Encyclopaedia Iranica. http://www.iranicaonline.org/articles/persian-language-1-early-new-persian; Accessed 24 September 2014.

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**, 827-845.

Rey D, Areces C, Alonso-Rubio J, Enriquez-de-Salamanca M, Abd-El-Fatah-Khalil S, Bendikuze N. et al. (2013) HLA in Georgians (Caucasus) and their relationship with Eastern Mediterraneans. *Mol. Biol. Rep.* **40**, 5523-30.

Rey D, Amirzargar A, Areces C, Enriquez-de-Salamanca M, Marco J, Abd-El-Fatah-Khalil S. et al. (2014) Gorgan (Turkmen in Iran) HLA genetics: transplantation, pharmacogenomics and anthropology. *Immunol. Invest.* doi:10.3109/08820139.2014.936938.

Robinson J, Halliwell JA, Hayhurst JH, Flicek P, Parham P and Marsh, SGE. (2015) The IPD and IPD-IMGT/HLA Database: allele variant databases. *Nucleic Acids Res.* **43,** D423-431.

Roitberg-Tambur A, Witt CS, Friedmann A, Safirman C, Sherman L, Battat S. et al. (1995) Comparative analysis of HLA polymorphism at the serologic and molecular level in Moroccan and Ashkenazi Jews. *Tissue Antigens.* **46**, 104-110.

Saitou N and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425.

Sanchez-Velasco P and Leyva-Cobian, F. (2001) The HLA class I and class II allele frequencies studied at the DNA level in the Svanetian population (Upper Caucasus) and their relationships to Western European populations. *Tissue Antigens.* **58**, 223-233.

Schneider S, Roessli D and Excoffier, L. (2000) ARLEQUIN: a software for population genetics database analysis 2.0. Geneva, Switzerland: Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Switzerland.

Schonberg A, Theunert C, Li M, Stoneking M and Nasidze, I. (2011) High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *Eur. J. Hum. Genet.* **19**, 988-994.

Sellier J and Sellier, A. (1993) Atlas des peuples d´Orient. Editions La Decouverte, Paris, France.

Shiina T, Hosomichi K, Inoko H and Kulski, JK. (2009) The HLA genomic loci map: expression, interaction, diversity and disease. *J. Hum. Genet.* **54**, 15-39.

Uinuk-Ool TS, Takezaki N, Sukernik RI, Nagl S and Klein, J. (2002) Origin and affinities of indigenous Siberian populations as revealed by HLA class II gene frequencies. *Hum. Genet.* **110**, 209-226.

Vicario JL, Martinez-Laso J, Corell A, Martin-Villa JM, Morales P, Lledo G. et al. (1992) Comparison between HLA-DRB and DQ DNA sequences and classic serological markers as type 1 (insulin-dependent) diabetes mellitus predictive risk markers in the Spanish population. *Diabetologia.* **35**, 475-481.

von Soden W. (1994) The Ancient Orient. Eerdmans Publishing, Grand Rapids, USA.

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J. et al. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA.* **98**, 10244-10249.

Yarshater E. (1988) The Iranian language of Azerb. In: Yarshater, E. (ed. Encyclopaedia Iranica). Routledge and Kegan Paul, pp. 238-245. London, UK.

Young FW and Bann, CM. (1996) A visual statistics system. In: Statistical Computing Environments for Social Researches (ed. By R.A. Stine & J. Fox), pp. 207-236. Sage Publications, London, UK.