HLA Class I and Class II Genes Distribution of the Sistanis in Iran

Mahdi Aminikhah[1], Mir-Saeed Yekaninejad[1],Mohammad Hosein Nicknam[2,3], Farideh Khosravi[3,4], Mehrnaz Naroeinejad[5], Bita Ansaripour[3], Batol Moradi[3], Behrouz Nikbin[2,3], Ali Akbar Amirzargar[2,3*]

[1]Department of Epidemiology and Biostatistics, School of Public Health, [2]Molecular Immunology Research Center, [3]Department of Immunology, School of Medicine, Tehran University of Medical Sciences, [4]Noor Pathobiology Laboratory, Tehran, [5]University of Sistan and Baluchestan, Zahedan, Sistan and Baluchestan, Iran

ABSTRACT

Background: The high polymorphism in the human leukocyte antigen (HLA) genes can be used as an identity of individuals to compare with other populations. This extreme polymorphism in the HLA system is accountable for the differences in alleles and haplotypes among ethnic groups, populations, and the inhabitants of many regions. **Objective:** To define the frequency of HLA alleles and haplotypes among the Sistanis, Sistani/Zaboli population in Iran. **Methods:** In this study, genotyping of class I (A, B, C) and class II HLA (DRB1, DQA1, DQB1) loci were determined in 90 unrelated Iraninan Sistani people and the results were compared with 474,892 HLA chromosomes from a diverse worldwide population. **Results:** The highest frequently observed alleles in this study were A*02:01, B*35:01, C*12:03, C*06:02, DRB1*11, DQA1*05:05, and DQB1*03:01. Furthermore, the most frequent 3-locus haplotypes were A*02:01-B*50:01*C*06:02, DRB1*11-DQB1*03:01-DQA1*05:05, and A*02:01-B*50:01-DRB1*07. The most occurring 4-locus haplotypes were A*02:01-B*50:01-C*06:02-DRB1*07 and A*02:01-B*50:01-DRB1*07-DQB1*02:01. A*02:01-B*50:01-C*06:02-DRB1*07-DQB1*02:01 and A*02:01-B*50:01-C*06:02-DRB1*07-DQB1*02:01-DQA1*02:01 were determined to be the predominant 5- and 6-locus haplotypes, respectively. The heat maps and multiple correspondence analyses based on the frequency of HLA alleles showed that Sistanis share a common genetic inheritance with other Iranian ethnic groups such as the people from Yazd and Fars except some differences with Baluchis, Iranian Jews, Lurs of Kohgiluyeh/Buyerahmad, and Arabs of Fars, which may arise from the admixture of these groups or with foreign subgroups over centuries, and also a close relatedness with some European populations. **Conclusion:** These data could be useful for finding better donor matches for organ transplantation among Sistanis or other related Iranian ethnic groups, epidemiological studies of HLA-associated diseases, handling HLA genomics and mapping the migration pattern of different ethnic group.

Aminikhah M, et al. Iran J Immunol. 2018; 15(2): 97-111.

Keywords: Allele Frequency, Haplotype Frequency, Human Leukocyte Antigen, Population Relationship

---

*Corresponding authors: Dr. Ali Akbar Amirzargar, Molecular Immunology Research Center, Department of Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran, e-mail: amirzara@tums.ac.ir

INTRODUCTION

HLA is the most polymorphic system in the human genome and more than 17,000 of its alleles have been recognized until now (1). However, a precise and statistically sufficient fingerprinting on the dispersion of this variation at the rank of the population has not been given much attention (2).This extreme polymorphism in the HLA system is accountable for the differences in alleles and haplotypes among ethnic groups, populations, and inhabitants of many regions (3). Regarding the origin of people, the HLA system has had a significant advantage over other polymorphic genes and blood groups. Therefore, finding new alleles by DNA sequencing of HLA region have been extensively increased (4,5). Some alleles exclusively belong to special ethnic groups like A*29:11 and A*36:01, which are seen only in African Blacks (6). The robust linkage among alleles in neighboring HLA loci depicts certain haplotypes as characteristics for specific populations (7). Analysis of HLA allele frequencies is useful for finding unrelated HLA matched donors. Moreover, studying the frequency of different HLA alleles in the target population as a data source for HLA-associated diseases is necessary. High presentation of allele frequencies in some ethnic groups or inhabitants of some specific area may be indicative of the predominance of certain autoimmune disease in those ethnic groups. Scrutinizing dissimilarities in HLA allele provides good information for understanding the evolution of lineages and the origin of some ethnic groups. The Sistani ethnic group is among the principal outnumbered groups in Iran. This ethnic group lives in Zabul, a city located at the southeast of Iran, 31°1′56”N, 61°29′24”E (Figure. 1), the north of the Sistan and Baluchistan province, close to the border of Pakistan and Afghanistan in the southeast region of Iran. It is bordered in the north by Khorasan province and to the west by Kerman and Hormozgan provinces. The geographical region of the Sistani ethnic group in present-day spans from eastern Iran, southern Afghanistan in Nimruz, Kandahar, and Zabul provinces, to the western Pakistani Nok-Kundi district of Baluchistan. The typical culture of this minority group has led to their isolation and limited relationship with other Iranian populations. Hence, the HLA genetic system of this population is an ideal tool for studying the origins of partly isolated groups. Over the past couple of years, several of these people have migrated to other places like Tehran and Golestan provinces. The majority of Sistani people live in Zabul. Sistani people speak with a Persian accent called Zabuli or Sistani. The language of the Sistani people, who live in Iran, Afghanistan, Tajikistan, Pakistan, and Turkmenistan, is generally considered a dialect of Persian (Dari) (8). In this research, we aimed to define the frequency of HLA alleles and haplotypes among the Sistanis in order to 1) conduct a pilot study for determining the allele and haplotype frequencies for DNA registry bank establishment to find out better-unrelated HLA matched donor for the recipients who have not identical siblings; 2) identify HLA pharmacogenomics, adverse drug reactions affecting the skin, liver and other organs, which show strong associations with particular HLA alleles. For example, in the Eastern populations, carbamazepine adverse drug reactions like Stevens-Johnson syndrome and toxic epidermal necrosis has a meaningful relationship with HLA-B*15: 02 in far east countries and HLA-A3101 in European populations and should be preceded by HLA typing prior to administration in the patient; 3) HLA disease association and predicting the occurrence of some autoimmune diseases in this population; and 4) find the origin of the Sistani and compare it with other worldwide ethnic groups like the Asian, Mediterranean, American, and African.

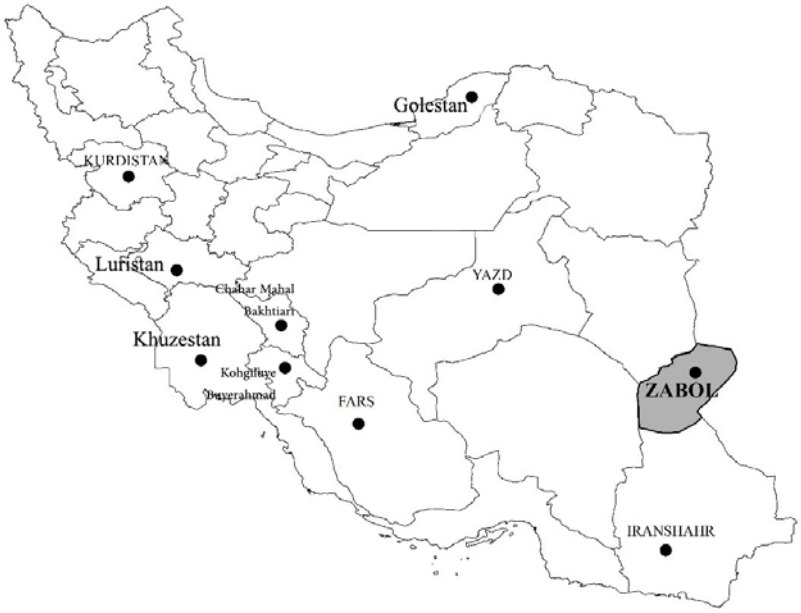

Figure 1. Map showing the area of the Iranian population used in this study. Gray area represent location of study sample (Zabol city) and other Iranian population samples cities which are included in the present paper relatedness analyses.

MATERIALS AND METHODS

Population Sample. In this study, 90 unrelated healthy volunteers from Zabul city with an equal share of both genders were genotyped for HLA classes I and II. An informed consent form was signed by each individual. All the participants were born in Zabul (Sistan region) and were Zabuli up to at least their fourth generation of ancestors. The data of our investigation of the Sistani ethnic group were compared with those other Iranian ethnic groupsas well as Africans, African Americans, Latin Americans, Europeans, Mediterranean, and Orientals (Table 1).

HLA Genotyping. In this study, 10 mL of whole blood was collected from each participant in EDTA. DNA was extracted using proteinase K and a modified salting-out method (9). HLA-A, -B, -C, -DRB1, -DQA-A1, and -DQB1 genotyping was done by commercial kits (CTS, Heidelberg, Germany) based on polymerase chain reaction (PCR) using sequence-specific primers (PCR-SSP) method according to manufacturer's instruction. Next, PCR products were electrophoresed on a 2% agarose gel including ethidium bromide and the alleles were determined under UV light.

Statistical Analysis. Allele and Haplotype frequencies, standardized disequilibrium values ($r2$) between alleles at different loci, linkage disequilibrium (LD) for genotypic data, deviations from Hardy-Weinberg equilibrium (HWE), and Ewens-Watterson homozygosity of neutrality were computed by Arlequin v3.5 software (10). To estimate the standard deviation (SD), bootstrap sampling with 3,000 bootstrap replicates and 10 starting points for SD estimations were performed. Finally, LD was estimated for all these haplotypes combinations.

Table 1. Population data in analysis.

Population	Abbreviation	n	Reference
Baloch	Sistan/Balochestan	100	[14]
Gorgan	Golestan	69	[15]
Yazd	Yazd	90	[16]
Parsi	Fars	100	[17]
Kurd	Kurdistan	60	[18]
Bakhtiari	ChaharMahall/Bakhtiari	50	[19]
Iranian Jews	Fars	91	[20]
Sistani	Sistan/Balochestan	90	Present study
Arab-1	Khuzestan	50	[20]
Arab-2	Fars	84	[20]
Lur-1	Luristan	50	[19]
Lur-2	Kohgiluyeh/Buyerahmad	50	[19]
Greece pop 7	Greece	11250	[21, 22]
Italy		159311	[21]
Argentina Buenos Aires pop 2	Argentina	1216	[21]
USA Caucasian pop 4	USA-European	1070	[21]
USA African		2411	[21]
Germany		11407	[21]
Russia Moscow	Russia	2650	[21]
Saudi Arabia pop 4	Saudi	499	[21]
Turkey pop 2	Turkey	228	[21]
Japan pop 16	Japan	18604	[21]
South Korea pop 8	Korea	7096	[21]
China Jiangsu Province	China	20248	[21]
South Africa Black	South Africa	200	[21]
Austria		200	
Macedonia pop 3	Macedonia	172	[21]

Using statistical software SPSS ver. 24, multiple correspondence analyses depending on the allele frequencies of HLA-A, -B, and -DRB1 loci were used to show a general overview of the relationships between these populations (11). This technique is controlled by the variance of allele frequencies among the populations (analogous to the classical principal constituents method) and of a statistical illustration of the differences. Heat maps analyses were implemented using the R 3.2.3 software package (14). The FST index among the different population was calculated by the aid of POPTREE software (12,13).

RESULTS

HLA Class I and Class II Allele Frequencies.

Allele frequencies of HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 loci in 90 individuals from Sistani population are illustrated in Tables 2 and 3. Altogether, 26 alleles were characterized at HLA-A locus, 40 alleles at -B locus, 27 alleles at -C locus,

17 alleles at -DRB1 locus, 10 alleles DQA1, and 10 alleles at -DQB1 locus. In the **Table 2. Class I HLA-A, -B and HLA-C allele frequencies, Hardy-Weinberg equilibrium test, in Iranian Sistani (Zabol) population (2n =180).**

HLA-A Alleles 0.1425*	Frequency	HLA-B Alleles 0.4221*	Frequency	HLA-C Alleles 0.1307*	Frequency
01:01	0.08889	07:02	0.01111	01:01	0.00556
02:01	**0.18889**	07:05	0.01667	01:02	0.01667
03:01	0.08334	08:01	0.03889	01:04	0.01111
11:01	**0.13889**	08:04	0.00556	02:01	0.02123
23:01	0.03889	13:01	0.03889	02:02	0.01667
24:02	**0.10000**	14:02	0.01667	02:03	0.01111
24:06	0.00556	15:01	0.01111	03:02	0.02222
24:07	0.00556	15:24	0.00556	03:03	0.00556
24:08	0.00556	18:01	0.01111	03:04	0.00556
24:19	0.01667	18:09	0.02222	03:08	0.02222
24:24	0.00556	27:01	0.02222	04:01	0.08889
25:01	0.00556	35:01	**0.13889**	05:01	0.00556
26:01	0.09444	37:01	0.02222	06:02	**0.15000**
26:08	0.01111	38:01	0.05556	06:04	0.03889
26:13	0.00556	38:08	0.00556	07:01	0.09444
29:01	0.00556	39:01	0.01667	07:08	0.01667
30:01	0.04444	40:01	0.02778	08:01	0.01667
31:01	0.01667	40:02	0.01111	12:02	0.05556
31:02	0.00556	40:07	0.00556	12:03	**0.15000**
31:07	0.00556	40:08	0.00556	12:04	0.00556
32:01	0.03333	41:01	0.02222	13:01	0.01667
32:04	0.01111	44:01	0.01667	14:02	0.07778
33:01	0.01111	44:02	0.02222	16:01	0.00556
68:01	0.05556	44:05	0.01111	16:02	0.03889
68:02	0.01111	44:10	0.00556	17:01	0.02778
69:01	0.00556	49:02	0.00556	18:01	0.06667
		50:01	0.08889	18:02	0.00556
		50:02	0.00556		
		51:01	**0.15000**		
		51:06	0.01667		
		51:08	0.01667		
		52:01	0.04444		
		55:01	0.02778		
		55:08	0.00556		
		57:01	0.03333		
		57:05	0.01111		
		58:01	0.00556		
		67:01	0.01111		
		73	0.00556		
		78:01	0.00556		

n: Number of individuals. The most frequent alleles were marked in bold. * *P* value for Hardy-Weinberg equilibrium test.

HLA-I for this population, 7 HLA-A, 4 HLA-B, and 7 HLA-C alleles had frequencies higher than 5% (A*01:01, A*02:01, A*03:01, A*11:01, A*24:02, A*26:01, A*68:01, B*35:01, B*38:01, B*50:01, B*51:01, C*04:01, C*06:02, C*07:01, C*12:02, C*12:03, C*14:02, C*18:01), and their cumulative frequencies were 75%, 43.3%, and 68.3%, respectively. In the HLA-II loci, 7 -DRB1, 9 -DQA1, and 6 -DQB1 alleles had frequencies higher than 5% (DRB1*01:01, DRB1*03:01, DRB1*04, DRB1*07, DRB1*11, DRB1*13:01, DRB1*15, DQA1*01:01, DQA1*01:02, DQA1*01:03, DQA1*01:04, DQA1*02:01, DQA1*03:01, DQA1*05:01, DQA1*05:05, DQA1*06:01, DQB1*02:01, DQB1*03:01, DQB1*03:02, DQB1*05:01, DQB1*06:01, and DQB1*06:02) with cumulative frequencies of 80.5%, 96.7%, and 93.3%, respectively(Tables 2 and 3).

Table 3. Class II HLA-DRB1, -DQB1 and HLA-DQA1 allele frequencies, Hardy-Weinberg equilibrium test, in Iranian Sistani (Zabol) population (2n =180).

HLA-DRB1 Alleles 0.6205*	Frequency	HLA-DQA1 Alleles 0.8064*	Frequency	HLA-DQB1 Alleles 0.6875*	Frequency
01:01	0.08333	01:01	0.08889	02:01	**0.20000**
03:01	0.07778	01:02	**0.11667**	03:01	**0.25000**
04	0.07222	01:03	**0.13333**	03:02	0.06111
04:01	0.01111	01:04	0.07222	03:03	0.02778
07	**0.12778**	02:01	**0.13889**	03:05	0.01667
07:01	0.01111	03:01	0.08889	04:01	0.03333
08	0.03889	04:01	0.02778	05:01	**0.20556**
09:01	0.00556	04:05	0.00556	06:01	0.07222
10	0.01667	05:01	**0.10000**	06:02	**0.11111**
10:01	0.01667	05:05	**0.22778**	06:04	0.02222
11	**0.25000**				
13:01	0.06111				
13:02	0.02222				
14:01	0.04444				
15	**0.13333**				
15:01	0.00556				
16	0.02222				

n: Number of individuals. The most frequent alleles were marked in bold. * *P* value for Hardy-Weinberg equilibrium test.

HLA Class I and Class II Haplotype Frequencies.

We identified 224 HLA-A-B, 205 HLA-B-DRB1, 610 HLA-A-B-C, 305 HLA-DRB1-DQA1- DQB1, 531 HLA-A-B-DRB1, 1077 HLA-A-B-C-DRB1, 1314 HLA-A-B-DRB1-DQB1, and 2089 HLA-A-B-C-DRB1-DQB1 haplotypes. The most frequently found 2-, 3-, 4-, and 5-locus haplotypes (frequency>0.01) in LD in the Sistani population are listed in Tables 4, 5, and 6. In terms of 2-locus haplotypes, 16 HLA-A-B

and the same 16 HLA-B-DRB1 were found statistically non-significant with LD frequencies exceeding 1%. The most frequent HLA-A-B haplotype was A*02:01-B*50:01 (5.55%) followed by A*24:02-B*35:01 (3.79%) and A*11:01-B*52:01 (3.33%). The most frequent HLA-B-DRB1 haplotypes were B*51:01-DRB1*11 (6.92%), B*50:01-DRB1*07 (4.97%), B*35:01-DRB1*01:01 (4.36), and B*08:01-DRB1*03:01 (2.77). In terms of 3-locus haplotypes, frequent HLA-A-B-C haplotypes were A*02:01-B*50:01-C*06:02 (4.39), A*11:01-B*52:01-C*12:02 (2.77), A*02:01-B*51:01-C*14:02 (2.22%) and A*26:01-B*38:01-C*12:03 (2.22%). The most common HLA-DRB1-DQA1-DQB1 haplotypes were DRB1*11-DQB1*03:01-DQA1*05:05 with a frequency of 22.2% followed by DRB1*07-DQB1*02:01-DQA*02:01 (11.1%), DRB1*01:01-DQB1*05:01-DQA1*01:01 (8.3%), DRB1*03:01-DQB1*02:01-DQA1*05:01 (7.2%), DRB1*15-DQB1*06:01-DQB1*01:03 (6.6%), DRB1*04-DQB1*03:02-DQA1*03:01 (5.5%), DRB1*13:01-DQB1*06:02-DQB1*03:01(5%), and DRB1*14:01-DQB1*05:01-DQA1*01:04 (4.4%).

Table 4. The most frequent 2-locus haplotypes (frequency >0.01) in linkage disequilibrium in Sistani population (2n=180).

A-B	Frequency	St.Dev.	r^2	*P*	B-DRB1	Frequency	St.Dev	r^2	*P*
02:01-50:01	0.055556	0.0178	0.0937	0.0000	51:01-11	0.069220	0.0211	0.0678	0.0004
24:02-35:01	0.037953	0.0158	0.0179	0.0225	50:01-07	0.049744	0.0165	0.1654	0.0000
11:01-52:01	0.033333	0.0135	0.1810	0.0000	35:01-01:01	0.043604	0.0164	0.1182	0.0000
26:01-38:01	0.030364	0.0132	0.1131	0.0000	08:01-03:01	0.027778	0.0123	0.2744	0.0000
03:01-51:01	0.027778	0.0142	0.0284	0.0238	51:01-08	0.027522	0.0132	0.0246	0.0352
02:01-41:01	0.022222	0.0112	0.0487	0.0031	35:01-14:01	0.022222	0.0124	0.02168	0.0482
02:01-40:01	0.020637	0.0112	0.0331	0.0146	38:01-1301	0.022211	0.0119	0.1177	0.0000
02:01-35:01	0.016667	0.0121	0.0221	0.0459	40:01-11	0.019364	0.0114	0.0186	0.0466
30:01-13:01	0.013333	0.0086	0.2076	0.0000	18:09-11	0.016667	0.0101	0.0681	0.0004
11-13	0.011111	0.0079	0.4944	0.0000	52:01-15	0.016667	0.0107	0.0235	0.0396
11-51:01	0.011111	0.0081	0.0218	0.0474	13:01-04	0.011111	0.0082	0.0458	0.0040
23:01-44:02	0.011111	0.0080	0.1293	0.0000	35:01-10:01	0.011111	0.0081	0.0394	0.0076
23:01-50:01	0.011111	0.0082	0.0194	0.0420	37:01-10	0.011111	0.0079	0.0755	0.0002
24:02-15:01	0.011111	0.0080	0.0200	0.0479	37:01-15	0.011111	0.0082	0.0264	0.0291
24:02-40:02	0.011111	0.0080	0.0200	0.0479	40:02-04	0.011111	0.0079	0.0306	0.0187
01:01-57:01	0.010176	0.0082	0.0720	0.0003	44:01-07	0.011111	0.0078	0.0441	0.0048

n: Number of individuals.r^2: standardized disequilibrium values.St.Dev.: Standard Deviation.

The most common A-B-DRB1 haplotypes were HLA-A*02:01-B*50:01- DRB1*07 with a frequency of 4.4% followed by A*01:01-B*51:01-DRB*11 (2.2%). The most frequent 4-locus (HLA-A-B-C-DRB1) haplotype was HLA-A*02:01-B*50:01-C*06:02-DRB1*07 (3.88%). The most frequent HLA-A-B-DRB1-DQB1 haplotypes were A*02:01-B*50:01-DRB1*07-DQB1*02:01 (3.88%) and A*02:01-B-51:01-

DRB1*11-DQB1*03:01 (2.22%). The common HLA-A-B-C-DRB1-DQB1 haplotypes were A*02:01-B*50:01-C*06:02-DRB1*07-DQB1*02:01 (3.33%) and A*02:01-B*51:01-C*14:02-DRB1*11-DQB1*03:01 (2.22%). The most common HLA-A-B-C-DRB1-DQB1-DQA1 haplotypes were A*02:01-B*51:01-C*14:02-DRB1*11-DQB1*03:01-DQA1*05:05 (2.77%) and A*02:01-B*50:01-C*06:02-DRB1*07-DQB1*02:01-DQA1*02:01 (2.5%).

Table 5. The most frequent 3-locus haplotypes (frequency >0.01) in linkage disequilibrium in Sistani population (2n=180).

A-B-C	Frequency	St.Dev	DRB1-DQB1-DQA1	Frequency	St.Dev.	A-B-DRB1	Frequency	St.Dev
02:01-50:01-06:02	0.043915	0.0163	11-03:01-05:05	0.22222	0.0313	02:01-50:01-07	0.044444	0.0161
11:01-52:01-12:02	0.027778	0.0126	07-02:01-02:01	0.11111	0.0236	01:01-51:01-11	0.022222	0.0126
02:01-51:01-14:02	0.022222	0.0124	01:01-05:01-01:01	0.08333	0.0203	26:01-38:01-13:01	0.016666	0.0166
26:01-38:01-12:03	0.022222	0.0121	03:01-03:05-05:01	0.07222	0.0056	02:01-40:01-11	0.016666	0.0105
01:01-35:01-04:01	0.016667	0.0092	15-06:01-01:03	0.06667	0.0187	03:01-51:01-08	0.016666	0.0103
01:01-57:01-06:02	0.016667	0.0097	04-03:02-03:01	0.05556	0.0175	11:01-52:01-15	0.016666	0.0103
02:01-51:01-04:01	0.016667	0.0107	13:01-06:02-01:03	0.05000	0.0163	24:02-40:02-04	0.011111	0.0082
02:01-35:01-18:01	0.016667	0.0111	14:01-05:01-01:04	0.04444	0.0152	24:02-35:01-14:01	0.011111	0.0085
02:01-51:01-04:01	0.016667	0.0107	15-06:02-01:02	0.03888	0.0149			
11:01-3501:1801	0.016667	0.0107	08-04:01-04:01	0.02776	0.0124			

n: Number of individuals. St.Dev.: Standard Deviation.

Hardy-Weinberg Equilibrium and Linkage Disequilibrium.

Table 2 and Table 3 show that gene frequency figures for HLA-A, -B, -C, -DRB1, -DQA1, and -DQB1 loci do not differ significantly and the population is found in Hardy-Weinberg equilibrium. Additionally, we listed all 2-, 3-, 4-, 5- and 6-locus haplotypes that were in linkage disequilibrium.

Test of Neutrality.

The Ewens-Watterson (EW) Homozygosity test of neutrality was used to determine whether the observed allele frequency distribution at each locus deviates from the null hypothesis of neutrality. The results of the EW test of neutrality for the Class I and Class II loci are presented in Table 7. The observed homozygosity values (Fobs) were lower than the expected ones (Fexp) in all loci except for HLA-B (p=0.733). HLA-DQB1 and -DAQ1 had significantly lower homozygosity than that expected for these loci in this population.

Table 6. The most frequent 4-locus haplotypes (frequency >0.01) in linkage disequilibrium in Sistani population (2n=180).

A-B-C-DRB1	Frequency	St.Dev.	A-B-DRB1-DQB1	Frequency	St.Dev.
02:01-50:01-06:02-07	0.038889	0.0149	**02:01-50:01-07-02:01**	0.038889	0.0145
02:01-51:01-14:02-11	0.011111	0.0102	**01:01-51:01-11-03:01**	0.022222	0.0124
26:01-38:01-12:03-14:01	0.011111	0.0086	**26:01-38:01-13:01-06:02**	0.016666	0.0103
01:01-57:01-06:02-07	0.011111	0.0087	**02:01-40:01-11-03:01**	0.016666	0.0107
02:01-51:01-04:01-08	0.011111	0.0085	**03:01-51:01-08-04:01**	0.016666	0.0104
02:01-51:01-14:02-15	0.011111	0.0092	**11:01-52:01-15-06:01**	0.016666	0.0103
			24:02-35:01-14:01-05:01	0.011111	0.0088
A-B-C-DRB1-DQB1					
02:01-50:01-06:02-07-02:01	0.033333	0.0139			
02:01-51:01-14:02-11-03:01	0.022222	0.0127			
11:01-52:01-12:02-04-03:02	0.011111	0.0085			
01:01-57:01-06:02-07-02:01	0.011111	0.0087			
02:01-51:01-04:01-08-04:01	0.011111	0.0088			

n: Number of individuals. St.Dev.: Standard Deviation.

Comparison of Sistani Population with other Populations.

Heat maps based on FST index among Sistani and other Iranian populations as well as other worldwide populations are depicted in Figures. 2 and 3.

Table 7. Homozygosity (F) values (Ewens-Watterson test of Neutrslity) for HLA Class I and II loci.

HLA	k	F_{obs}	F_{exp}	*P*
HLA-A	29	0.08698	0.09707	0.4990
HLA-B	42	0.06841	0.05964	0.7330
HLA-C	30	0.07957	0.09263	0.3530
HLA-DRB1	17	0.12389	0.18074	0.1380
HLA-DQA1	10	0.13438	0.30516	**0.0010**
HLA-DQB1	10	0.16870	0.30838	**0.0300**

k: Number of alleles. Fobs: observed value. Fexp: expected value. P: P values were calculated for Ewens-Watterson test using homozygosity as a test statistic.

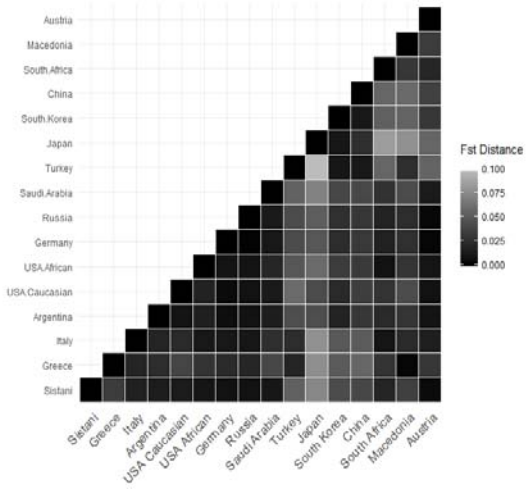

Figure 2. The heat map of F_{ST} index between Sistani and other 15 worldwide populations.

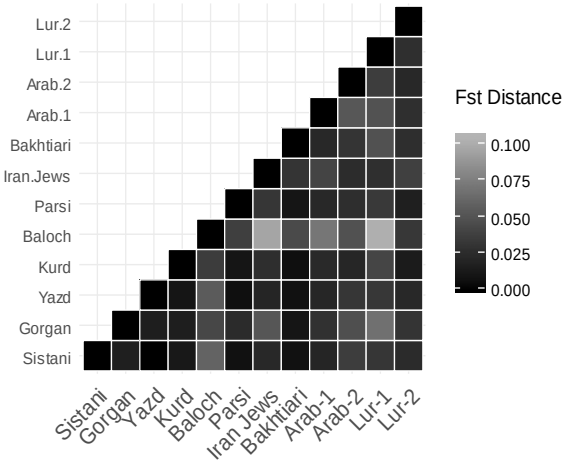

Figure 3. The heat map of F_{ST} index between Sistani and other 11 Iranian populations.

This comparison showed that the Sistani has common inheritance with other Iranian ethnic groups. Despite some differences, Sistani population has also a close relation with Yazdi (FST<0.001), Parsi (FST=0.007), and Bakhtiari (FST=0.007) populations. Additionally, we found that Baluchi population is far from the Sistani ethnic group (FST=0.06). In terms of worldwide populations, the closest people to Sistani population was Austrians (FST=0.004), followed by Germans and Russians (FST=0.007), also the farthest one was Japan (FST=0.073). Genetic distances between Sistani and other populations based on multiple correspondence analyses are shown in Figures. 4 and 5.

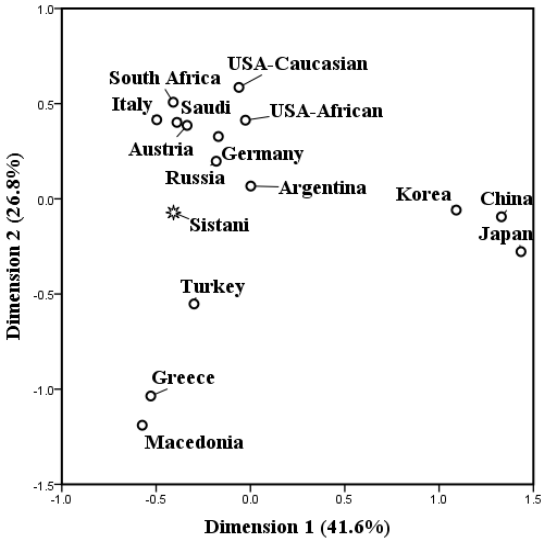

Figure 4. Multiple correspondence analysis illustrating the genetic relationships between Sistani and other 15 worldwide populations.

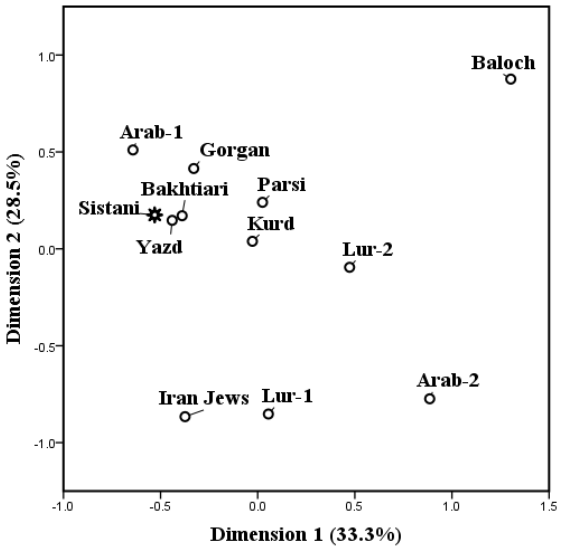

Figure 5. Multiple correspondence analysis illustrating the genetic relationships between Sistani and other 11 Iranian populations.

Based on these results, Sistanis are close to Yazdi, Bakhtiari, Parsi and Kurd populations. With respect to worldwide populations, Sistanis are near to Russian, German, Austrian, and Turk populations. Figures. 6 and 7 show alleles making the most diversity among HLA Class I and Class II. HLA-A and -C showed the highest cumulative frequency while HLA-B showed the lowest cumulative frequency. Moreover, HLA-A and -C curves attained 100% cumulative frequency first; however, the HLA-B curve was the last one to reach 100%. This result confirms that HLA-A and -C have the lowest diversity; in contrast, the HLA-B has the highest allelic diversity. As can be seen from Figure. 3, the HLA- DQA1 and -DQB1 curves had the highest cumulative frequency, while the lowest cumulative belonged to HLA-DRB1. Moreover, HLA-DRB1 curve reached 100% cumulative frequency last, suggesting that HLA-DRB1 has the highest diversity among HLA-Class II alleles.

DISCUSSION

The application of genetics data especially those of HLA and determining the genetic distance between populations is among the useful tools which can properly reveal the measure of similarity in the evolution of various races and populations (23). Therefore, recognizing and finding the HLA allele frequency distribution in populations and individuals in order to understand and identify autoimmune diseases and whole evolutionary pattern between populations is a matter of great importance. In this research, we intended to investigate the allocation of HLA alleles for the Sistani population. The Sistanis are among the paramount minority ethnic groups in the Iranian population. The cultural, moral, lifestyle and other characteristics of this ethnic minority are unique. We observed the existence of the 26, 40, 27, 17, 10, and 10 alleles at the HLA Class I (HLA-A, - B, -C, and HLA) and Class II (HLA-DRB1, -DQA1, and DQB1) loci respectively in Sistani population.

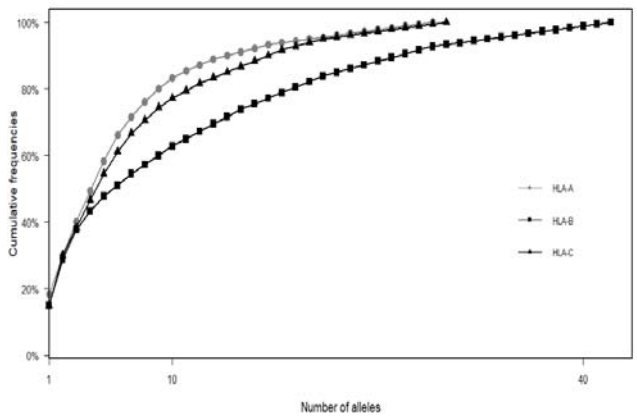

Figure 6. Cumulative Frequencies for sorted HLA class I –A, -B and – C alleles according to their frequencies in descending order.

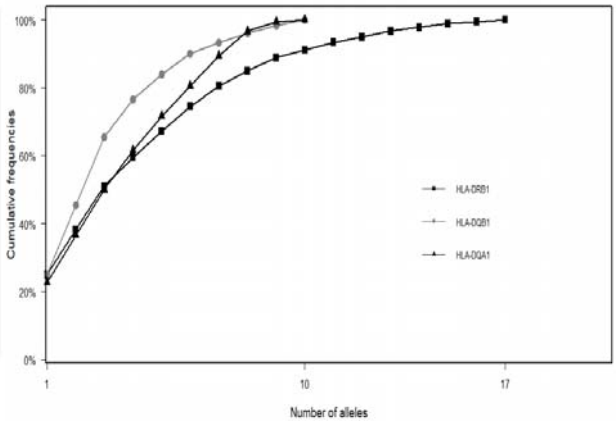

Figure 7. Cumulative Frequencies for sorted HLA class II –DRB1, -DQB1 and –DQA1 alleles according to their frequencies in descending order.

The FST distances between this population and other ethnic groups were computed. Heat map and multiple corresponding analyses showed the differences between Sistani and other populations.By FST index, we mean the genetic divergence between the populations within a species. Here, a small FST index indicates that two populations are close pertaining to the genetic relatedness and vice versa (24). According to the results achieved by both FST index and multiple correspondence analysis, the Sistanis stand among the Mediterranean and Europeans and far from the Eastern Asians. Besides, the Sistanis were in the position near the Yazdi and Bakhtiari populations, while they were far from Iranian Jews and Baluchi.

The comparison of 3-locus haplotype frequencies DRB1-DQB1-DQA1 (Figure. 8) between the Sistani ethnic population and two Iranian populations (Parsi and Yazdi) clearly showed that the trend of haplotype frequencies is generally constant. The first part (Figure. 8A) compares the 17 highest occurring frequent haplotypes common among the Sistanis and Iranian Parsis. The second part (Figure. 8B) compares the Sistanis and Iranian Yazdi for the common 15 highest occurring frequent haplotypes. In accordance with the results of the heat map, it can be defined that the Sistanis, Yazdi, Bakhtiari, and Parsis have a close genetic relationship.Based on FST index and multiple correspondence analysis, Sistanis are grouped near the Russians and Germans (Figures. 2, 3, 4, and 5).

Our analyses of HLA Classes I and II (A, B, C and DRB1, DQA1, and DAB1) polymorphism in the Sistani population revealed that the frequent alleles do not show much difference among most Iranian ethnic groups except for Baluchi population (25,26).In the current study, the most frequent 3-locus haplotypes of HLA-DRB1-DQB1-DQA1 were DRB1*11-DQB1*03:01-DQA1*05:05 (22.2% vs 19.7% in Greece and 2.56% in Japanese) followed by DRB1*07-DQB1*02:01-DQA1*02:01 (11.1% vs 4.9% in Greece) and DRB1*01:01-DQB1*05:01-DQA1*01:01 (8.33% vs 9% in Russians and 6.3 % in Greeks and 0.21% in Japanese) and DRB1*03:01-DQB1*03:05-DQA1*05:01 (7.22%). Based on previous studies on Iranian ethnic groups DRB1*11-DQB1*01:03-DQA1*05:01 with a frequency of 25% and DRB1*03:01-DQB1*02:01-DQA1*05:01 with a frequency of 10.0% were the predominant haplotypes in the Parsi population (17). Also, was DRB1*11-DQB1*03:01-DQA1*05:01 was the most common haplotype in the Lorestan (31.1%), Kohgiluyeh/Boyerahmad (23.1%), Famoori-Arabs (18.5%), and Jews of Iran (29.7%). In addition, they were frequent haplotypes in Kurds and Azeris populations (26,27).

This haplotype is the most shared haplotype in almost all Iranians. Iranian populations shared some frequent alleles HLA-A*2 (18.8% vs 22.56%), HLA-B*35:01 (13.88% vs 14.8%), HLA-B*51 (18.33% vs 14.8%), and HLA-DRB1*11 (25% vs 24.8) in the Macedonian and haplotypes HLA-A*26:01-B*38:01-C*12:03 (2.22%) and HLA-A*02:01-B*51:01-C*14:02 (2.22%) with European Caucasians in the southern and parts of Europe (28). In case of 5- and 6-locus haplotypes, we found newly extended haplotypes in Sistani that are A*02:01- B*51:01-C*14:02-DRB1*11-DQB1*03:01 (2.22%), A*02:01-B*51:01-C*14:02-DRB1*11-DQB1*03:01-DQA1*05:05 (2.77%), and A*02:01-B*50:01-C*06:02-DRB1*07-DQB1*02:01-DQA1*02:01(2.5%). Additionally, the HLA-A*02:01-B*50:01-C*06:02-DRB1*07-DQB1*02:01 (3.33%) in Sistani shared with some of European ethnics and one American population (29-31).

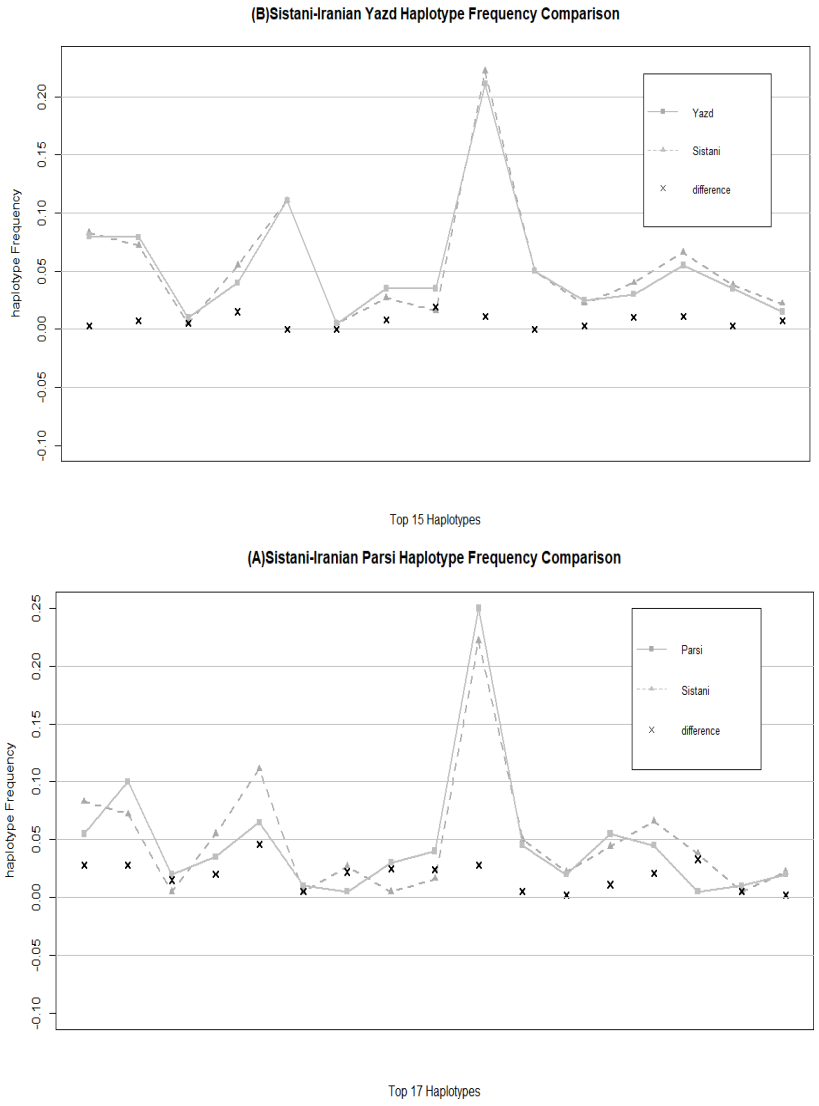

Figure 8. Complete registry low resolution-typed specimen comparison of DRB1-DQB1-DQA1 haplotype frequencies the present investigation for the initial haplotypes per ethnic group.

Sistan derived its appellation from Sakastan; the home of the Saka. The Sakas were a Scythian tribe that emigrated from the second to the first century BC to the Iranian Plateau and India and formed a kingdom called the Indo-Scythian Kingdom. In Bundahishn, the province was named “Seyansih”. After the Arabian subjugation of Iran in the 7th century BC, the province was renamed Sijistan/Sistan. As a result, there existed two dynasties in the region of Sistan, which were generally referred to as the two distinct dynasties of “Indo-Scythians” and “Indo-Parthians” (32,33).

In conclusion, to the best of our knowledge, this is the first analyses of the distribution of HLA Classes I and II allele and haplotype frequencies in the Sistani population. We believe that this work is significant for providing an outline of the HLA characteristics on the healthy population in our region and help us to precisely establish DNA registry bank in this ethnic group to find out suitable unrelated HLA-matched donors for

patients without HLA identical siblings in bone marrow transplantation.The results of this study need to be verified using larger samples size and high-resolution HLA typing methods. Another limitation of this study was that the opportunity to acquire samples was difficult due to people's reluctance to give blood samples for research purposes.

ACKNOWLEDGEMENTS

This project has been financially supported by research affair of Tehran University of medical sciences (project 4 from grant 31303379).

REFERENCES

1. Marsh, SG, et al. Nomenclature for factors of the HLA system, 2006. Human immunology. 2006. 67: p. 557.
2. Shankarkumar U, B Sridharan, Pitchappan R. HLA diversity among Nadars, a primitive Dravidian caste of South India. Tissue Antigens, 2003; 62:542-547.
3. Ivanova M, et al. HLA polymorphism in Bulgarians defined by high□resolution typing methods in comparison with other populations. Tissue antigens. 2002; 60:496-504.
4. Marsh SG. Nomenclature for factors of the HLA system, update May 2011. Int J Immunogenet. 2011; 38:449-455.
5. Robinson J, et al. The imgt/hla database. Nucleic acids research, 2013. Nucleic Acids Res. 2013; 41:D1222-7.
6. Paximadis M, et al. Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population. Hum Immunol. 2012; 73:80-92.
7. Martinez□Laso J, et al. HLA molecular markers in Tuvinians: a population with both Oriental and Caucasoid characteristics. Ann Hum Genet. 2001; 65:245-61.
8. Charles Melville, Cambridge University, Gabrielle van den Berg, Leiden University, Sunil Sharma Boston University BRILL publication. 201 ,ISSN: 22103554.
9. Olerup O, AldenerA, FogdellA. HLA‐DQB1 and‐DQA1 typing by PCR amplification with sequence□specific primers (PCR‐SSP) in 2 hours. HLA. 1993; 41:119-134.
10. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis.EvolBioinform Online. 2007;1:47-50.
11. Greenacre M,DegosL. Correspondence analysis of HLA gene frequency data from 124 population samples. Am J Hum Genet. 1977; 29:60-75.
12. Takezaki N, NeiM, TamuraK. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. MolBiolEvol. 2010; 27:747-52.
13. Slatkin M,Barton NH. A comparison of three indirect methods for estimating average levels of gene flow. Evolution. 1989; 43:1349-1368.
14. Farjadian S, et al. Molecular analysis of HLA allele frequencies and haplotypes in Baloch of Iran compared with related populations of Pakistan. Tissue antigens. 2004; 64:581-587.
15. Rey D, et al.Gorgan (Turkmen in Iran) HLA genetics: transplantation, pharmacogenomics and anthropology. Immunol Invest. 2015; 44:88-100.
16. Nikbin B, et al. Human Leukocyte Antigen (HLA) Class I and II Polymorphism in Iranian Healthy Population from Yazd Province. Iran J Allergy Asthma Immunol. 2017; 16:1-13.
17. Amirzargar A, et al. Human leukocyte antigen class II allele frequencies and haplotype association in Iranian normal population. Hum Immunol. 2001; 62:1234-8.
18. Amirzargar A, et al. Kurds HLA Genes: Its Implications in Transplantation and Pharmacogenomics. Open Medicine Journal, 2015; 2:43-47.
19. Farjadian S,GhaderiA. Iranian Lurs genetic diversity: an anthropological view based on HLA class II profiles. Iran J Immunol. 2006; 3:106-113.
20. Farjadian S, and GhaderiA. HLA class II genetic diversity in Arabs and Jews of Iran. Iran J Immunol. 2007; 4:85-93.

21. González-Galarza FF, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. Nucleic Acids Res. 2015; 43:D784-8.
22. Gonzalez-Galarza FF, et al. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res. 2011; 39:D913-D919.
23. Zhang L, et al. Distribution of HLA-A, -B and -DRB1 genes and haplotypes in the Tujia population living in the Wufeng Region of Hubei Province, China. PLoS One. 2012; 7:e38774.
24. Qiu X, et al. HLA-B* 07 is a high risk allele for familial cervical cancer. Asian Pac J Cancer Prev. 2011; 12:2597-600.
25. Khazaei HA, et al. Human leukocyte antigen profile of two ethnic groups in southeast of Iran. Iran J Allergy Asthma Immunol. 2007; 6:223-4.
26. Farjadian S, MoqadamF, GhaderiA. HLA class II gene polymorphism in Parsees and Zoroastrians of Iran. Int J Immunogenet. 2006; 33:185-91.
27. Farjadian S,GhaderiA. HLA class II similarities in Iranian Kurds and Azeris. Int J Immunogenet. 2007; 34:457-63.
28. Arnaiz□Villena A, et al. HLA genes in Macedonians and the sub□Saharan origin of the Greeks. Tissue Antigens. 2001; 57:118-127.
29. Pingel J, et al. High-resolution HLA haplotype frequencies of stem cell donors in Germany with foreign parentage: how can they be used to improve unrelated donor searches? Hum Immunol. 2013; 74:330-40.
30. Alfirevic A, et al. In silico analysis of HLA associations with drug-induced liver injury: use of a HLA-genotyped DNA archive from healthy volunteers. Genome Med. 2012; 4:51.
31. Maiers M, GragertL,KlitzW.High-resolution HLA alleles and haplotypes in the United States population. Hum Immunol. 2007; 68:779-88.
32. Bosworth CE. The Encyclopaedia of Islam, New Edition: Supplement. 1980: Brill Archive.
33. Frye RN, Handbuch der Altertumswissenschaft: Alter Orient-Griechische Geschichte-Römische Geschichte. Band III, 7: The History of Ancient Iran. Vol. 7. 1984: CH Beck.