

## BRIEF COMMUNICATION

**HLA analysis of the Parsi (Zoroastrian) population in Pakistan**

A. Mohyuddin &amp; S. Q. Mehdi

Biomedical and Genetic Engineering Laboratories, GPO Box 2891, Islamabad-44000, Pakistan

**Key words**

HLA polymorphisms; Pakistan; Parsi; population genetics

**Correspondence**

Aisha Mohyuddin  
Biomedical and Genetic Engineering  
Laboratories  
GPO Box 2891,  
Islamabad-44000, Pakistan  
e-mail: sqmehdi@comsats.net.pk

Received 7 July 2005; revised 5 August 2005;  
re-revised 24 August 2005; accepted for publication 1 September 2005

doi: 10.1111/j.1399-0039.2005.00507.x

**Abstract**

The Parsis of Pakistan are descendants of Zoroastrians from Iran who fled to Gujarat in India after the Arab invasion in 900 AD. A small group eventually migrated from India to Karachi in Pakistan. In this study, the Parsis from Pakistan were analyzed at the HLA-B, -C, -DRB1 and -DQB1 loci using the polymerase chain reaction with sequence-specific primers (PCR-SSP). The most common alleles at the HLA loci were HLA-B\*35 (15.9%), HLA-Cw\*0602 (21.4%), HLA-DRB1\*11 (23.0%), and HLA-DQB1\*02 (24.7%). Data analysis suggests that the Parsis of Pakistan and India descended from the same stock and may have the closest ancestry with Jewish and Italian populations.

Parsis are descendants of the Zoroastrians of Iran, who were the followers of the prophet 'Zarathushtra' or Zoroaster. The Achemenian (550–330 BC), Parthian (312–216 BC), and Sassanian (211–641 BC) rulers of Persia were Zoroastrians. In the year 634 AD, the Arabs attacked and defeated the Sassanians and captured the Sassanian capital Ctesiphon (present day Iraq). By 641 AD, the Zoroastrian rule had come to an end. A small group of Zoroastrians migrated to Gujarat in India in 900 AD, where they were called the Parsi (meaning from *Pars*, i.e., Iran) (1). Eventually the Parsis moved from Gujarat to Mumbai in India and Karachi in Pakistan where, according to recent estimates, they number approximately five thousand. For this study, 91 Parsi samples were collected from Karachi. Informed consent was obtained from all individuals. HLA polymorphisms in the Parsi population were characterized at the HLA-B, -C, -DRB1, and -DQB1 loci using polymerase chain reaction with sequence-specific primers (PCR-SSP) (2). Their genetic relationship with other Pakistani ethnic groups, Parsis from India, and various world populations was ascertained. The HLA-A alleles of the same Parsi samples have been reported elsewhere (3).

The HLA allele frequencies found in the Parsis are given in Table 1. Significant variation in allele frequencies between the Parsis and other Pakistani ethnic groups was estimated by a one-way parametric ANOVA ( $P < 0.05$ ;

Tukey HSD,  $\alpha = 0.05$ ) using the spss v. 10.0 (SPSS, Chicago, IL) software. Previously published HLA allele data of the Pakistani ethnic groups were used for comparison (4). The approximate geographic locations, linguistic affiliations, and proposed origins of these Pakistani ethnic groups are given in Figure 1. The alleles exhibiting significant differences between the Parsis and the Pakistani populations are listed in Table 2.

Maximum-likelihood estimates of three- and two-locus haplotype frequencies were calculated from allele frequencies for HLA-A-Cw-B, HLA-A, -B, -DRB1, and HLA-DRB1-DQB1 using ARLEQUIN v. 1.1 software (<http://anthropologie.unige.ch/arlequin>) (5) (Table 3). Pairwise linkage disequilibrium (LD) and the statistical significance of LD using a  $2 \times 2$  contingency Chi-square test were calculated as described earlier (6). To evaluate the degree of LD, relative linkage disequilibrium (RLD) defined as the ratio of LD to the possible maximum value of LD was also calculated (6). All the haplotypes showed significant ( $P < 0.0001$ ) LD values. A\*30-Cw\*0602-B\*13, found in 11% of the Parsis, is present in some European [Basque (2.0%), Austrian (2.0%), Poles (2.0%), and Spaniards (1.3%)] and Oriental populations [Hui (6.1%), Manchu (4.0%), Korean (3.8%), and Mongolian (2.9%)] (7). A\*3201-Cw\*14-B\*51 present in 4.9% of the Parsi was found in the Kalash population of Pakistan (4) and has not been reported previously in any other population.

**Table 1** HLA percent allele frequencies in the Parsis

HLA-B*		HLA-C*		DRB1*		DQB1*	
Alleles	Percent frequencies	Alleles	Percent frequencies	Alleles	Percent frequencies	Alleles	Percent frequencies
07	3.8	01	2.2	01	19.5	02	24.7
08	3.3	0302/04	1.1	03	6.3	0301/04	23.0
13	14.3	0303	0.5	04/1122	9.8	0302	8.6
1401	5.5	04	18.1	07	18.4	030302	1.1
1402	9.9	0501	0.5	08	0.6	0305	1.1
15	0.5	0602	21.4	1001	1.1	04	0.6
18	3.8	0701	5.5	11	23.0	05	25.3
35	15.9	0702	6.0	13	1.1	06	6.3
38	3.3	0802	13.2	1302	3.4	0601	8.6
39	1.1	1202	4.4	14	3.4	0602	0.6
40	1.6	1203	8.8	15	11.5		
41	1.6	14	4.9	16	1.7		
42	0.5	15	8.2				
44	5.5	1601	2.7				
45	6.0	1701	2.2				
50	2.2						
51	8.8						
52	4.9						
56	1.1						
57	0.5						
58	0.5						
73	4.9						

A\*33-Cw\*0802-B\*14 has previously been detected in Cypriots (4.0%), Italians (3.5%), Sardinians (2.9%),



**Figure 1** Map of Pakistan showing the geographical locations and origins of the Pakistani ethnic groups used for comparison in this study. With the exception of the Brahui, who speak a Dravidian language, and the Burusho, who speak a language isolate Burushaski, the remaining ethnic groups speak Indo-European languages.

Ashkenazi Jews (1.8%), and the Baloch (1.6%) (4, 8, 9). A\*33-Cw\*0701-B\*44 has been detected in populations from the Indian subcontinent and some East Asian populations (4, 7). A\*30-B\*13-DRB1\*07, found in 8.3% Parsis, was considered an Arabic and Middle Eastern haplotype (10). It has also been detected in Middle Mainland China (5.6%), North Mainland China (5.2%), Greece (1.9%), Cyprus (1.9%), Korea (1.7%), Italy (1.1%), and Portugal (1.0%) (7, 11, 12). A\*33-B\*1402-DRB1\*01 found in 6.0% of the Parsi is found in Armenians (3.6%), Cypriots (3.2%), Italians (3.0%), Sardinians (2.7%), Indians (2.2%), French (2.0%), Germans (1.2%), and Greek (1.1%) (7, 8). The haplotype DRB1\*11-DQB1\*0301/04 was present in 18.8% Parsis. This haplotype has previously been reported in the Moroccan (6.3%), Greek (4.6%), Timorese (2.7%), and Mainland Chinese (7, 11, 13). The haplotype DRB1\*01-DQB1\*05 was found at a frequency of 17.2% in the Parsi and is common in all of Pakistan (unpublished data). It is a common haplotype particularly in European populations [Finns (16.5%), Basque (14.7%); Spaniards (13.5%); Sardinians (10.6%); Poles (12.4%), Greek (7.8%)] (7). It has also been detected in the Ashkenazi Jews (15.0%), Ethiopian Jews (10.8%) Algerians (7.0%), Lebanese (7, 8), Moroccan Jews (4.4%), and Libyan Jews (2.5%) (14–16).

HLA frequencies were used for phylogenetic and correspondence analyses in order to compare the Parsis with

**Table 2** HLA alleles exhibiting significant variation ( $P < 0.05$ ) in frequency between the Parsis and other Pakistani ethnic groups

HLA allele	Significant variation in comparison to
<b>Frequency significantly increased in Parsis</b>	
B*13, *1402, *56, *73, Cw*0802	All Pakistani ethnic groups
B*1401	Brahui, Kalash, and Pathan
B*45	Baloch, Brahui, Burusho, Kalash, and Pathan
Cw*0602, DRB1*07	Baloch, Brahui, and Burusho
Cw*1601	Brahui, Burusho, Kalash, and Pathan
DRB1*01	Kalash, Pathan, and Sindhi
DRB1*11, *1302	Baloch, Brahui, Burusho, Kalash, and Sindhi
DQB1*0301/04	Baloch, Brahui, Burusho, and Sindhi
<b>Frequency significantly decreased in Parsis</b>	
B*08, DRB1*07	Baloch, Brahui, Burusho, Pathan, and Sindhi
B*15	Burusho
B*27, Cw*02	Kalash
B*40, Cw*15	Baloch, Brahui, and Sindhi
B*51	Burusho, Kalash, Pathan, and Sindhi
B*58	Baloch
Cw*0302	Baloch, Brahui, and Pathan
Cw*0702	Burusho and Sindhi
Cw*1602	Pathan
DRB1*13	Burusho, Kalash, and Sindhi
DQB1*0603	Sindhi

other Pakistani ethnic groups and other world populations. The computer software DISPAN was used to construct phylogenetic trees by the neighbor-joining method (<http://www.bio.psu.edu/people/faculty/nei/lab/software.html>).

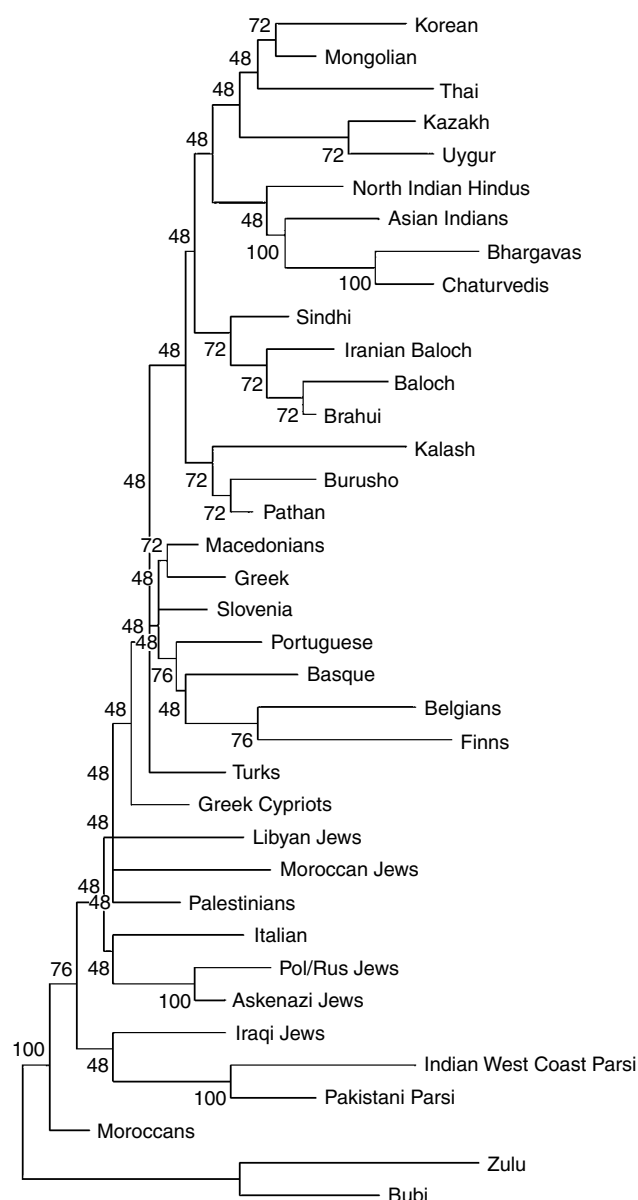
The neighbor-joining trees were based on Nei's DA genetic distances (17) (bootstrap replications = 1000). Correspondence analysis was carried out using the program VISTA v. 5.0.2 (<http://forrest.psych.unc.edu>) (18). The Parsis form an

**Table 3** Two- and three-locus haplotype frequencies in the Parsis

Haplotypes			HF (%)	LD (100)	RLD
<b>HLA-A-Cw-B</b>					
A*30	Cw*0602	B*13	11.0	10.5	0.76
A*3201	Cw*14	B*51	4.9	4.9	1.0
A*33	Cw*0802	B*1401	4.4	4.3	0.79
A*33	Cw*0701	B*44	4.4	4.3	0.80
A*2901	Cw*04	B*35	4.3	4.1	0.48
A*01	Cw*1203	B*18	3.3	3.3	0.86
A*03	Cw*04	B*35	3.3	3.3	0.85
A*11	Cw*04	B*35	3.3	2.9	0.24
A*11	Cw*12	B*52	3.3	3.3	0.75
<b>HLA-A-B-DRB1</b>					
A*30	B*13	DRB1*07	8.3	7.9	0.57
A*33	B*1402	DRB1*01	6.0	5.7	0.60
A*33	B*44	DRB1*07	3.9	3.7	0.7
A*2901	B*35	DRB1*11	3.8	3.5	0.42
A*11	B*52	DRB1*15	3.3	3.3	0.67
<b>HLA-DRB1-DQB1</b>					
DRB1*11	DQB1*0301/04		18.8	13.5	0.76
DRB1*01	DQB1*05		17.2	12.3	0.84
DRB1*07	DQB1*02		12.6	8.0	0.58
DRB1*15	DQB1*0601		8.6	7.6	1.0
DRB1*03	DQB1*02		6.3	4.8	1.0

HF, haplotype frequency; LD, linkage disequilibrium; RLD, relative linkage disequilibrium.

allele frequency out-group in the phylogenetic and correspondence analyses, with the other ethnic groups from Pakistan (figures not shown). A neighbor-joining tree comparing the Parsis with other world populations is given in Figure 2. The populations included are as described previously (4), with the addition of the Indian West Coast Parsis (19), North Indian Hindus (20), and the Iranian Baloch (21). The analysis is based upon HLA-A and HLA-B allele frequencies since the Indian Parsi population used for comparison lacked data for the remaining HLA loci. With the exception of the Parsis, all the



**Figure 2** Neighbour joining tree based on DAS genetic distances for HLA-A and -B showing the relationship between the Parsis and various world populations (16, 17). Bootstrap values shown at each fork indicate the degree of support for each branch (Bootstrap replications = 1000).

populations from Pakistan group together in a cluster. The Iranian Baloch cluster with the Pakistani Baloch as reported earlier (21). The Parsi populations from Pakistan and the Indian West Coast lie together in a cluster with the Iraqi Jews. The neighbor-joining tree resolved the populations into their regional clusters, and despite the variable range of bootstrap values, except for the Parsis, no individual population is separate from its geographic cluster. Although the bootstrap values at some of the nodes are low, correspondence analysis of this data (figure not shown) corroborates the results of the phylogenetic analysis.

These results can be explained in a historical context. The Parsis fled their homeland in present day Iran after the defeat of the Sassanians by the Arabs. Traditionally, they marry within their community and do not accept conversions. It is therefore expected that they would show little or no admixture with their current neighbors in Pakistan. Similar observations have been made for Y chromosomal markers of the Parsis. Their Y-chromosomes resemble those found in Iran more than their Pakistani neighbors (22, 23). During the Achaemenian period, at its zenith, the Zoroastrian empire extended as far west as Macedonia and Libya to as far east as the Beas River in India and from the Caucasus Mountains and the Aral Sea in the north to the Persian Gulf and the Arabian Desert in the south. The clustering of the Parsis with the Jews and Italians may therefore be the result of shared ancestry in ancient times.

### Acknowledgments

This work was supported by a core grant from the Government of Pakistan. We are grateful to Dr Qasim Ayub for his helpful comments on the manuscript. We thank Dr F. Sethna for his help in blood sample collection and the blood donors who made this study possible.

### References

1. Nanavutty P. *The Parsis*. New Delhi: National Book Trust, 1997.
2. Bunce M, O'Neill CM, Barnardo MCNM et al. Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence specific primers (PCR-SSP). *Tissue Antigens* 1995; **46**: 355–67.
3. Mohyuddin A, Williams F, Mansoor A, Mehdi SQ, Middleton D. Distribution of HLA-A alleles in eight ethnic groups from Pakistan. *Tissue Antigens* 2003; **61**: 286–91.
4. Mohyuddin A, Ayub Q, Khaliq S et al. HLA polymorphism in six ethnic groups from Pakistan. *Tissue Antigens* 2002; **59**: 492–501.
5. Schneider S, Roessli D, Excoffier L. Arlequin ver 2.0: a software for population genetic data analysis. *Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva*, 2000.

6. Imanishi T, Akaza T, Kimura A et al. Estimation of allele and haplotype frequencies for HLA and complement loci. In: Tsuji K, Aizawa M, Sasazuki T, eds. *HLA 1991. Proceedings of the Eleventh International Histocompatibility Workshop and Conference, Vol. I. Oxford: Oxford University Press, 1992, 76–9.*
7. Imanishi T, Akaza T, Kimura A et al. Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: Tsuji K, Aizawa M, Sasazuki T, eds. *HLA 1991. Proceedings of the Eleventh International Histocompatibility Workshop and Conference, Vol. I. Oxford: Oxford University Press, 1992, 1065–220.*
8. Piazza A, Lonjou C. HLA in Europe and the Mediterranean countries. In: Charron D, ed. *HLA Genetic Diversity of HLA Functional and Medical Implications, Vol. II. Paris: EDK Medical and Scientific International Publishers, 1997, 374–84.*
9. Cox ST, Marsh SGE, Scott I et al. HLA-A-B-C polymorphism in a UK Ashkenazi Jewish potential bone marrow donor population. *Tissue Antigens* 1999; **53**: 41–50.
10. Martínez-Laso J, De Juan D, Martínez-Quiles N, Gómez-Casado E, Cuadrado E, Arnaiz-Villena A. 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* 1995; **45**: 237–45.
11. Shaw C-K, Chen L-L, Lee A, Lee TD. Distribution of HLA gene and haplotype frequencies in Taiwan: a comparative study among Min-nan, Hakka, Aborigine and Mainland Chinese. *Tissue Antigens* 1999; **53**: 51–64.
12. Park MH, Kim HS, Kang SJ. HLA-A-B-DRB1 allele and haplotype frequencies in 510 Koreans. *Tissue Antigens* 1999; **53**: 386–90.
13. Gómez-Casado E, del Moral P, Martínez-Laso J et al. HLA genes in Arabic-speaking Moroccans: close relatedness to Berbers and Iberians. *Tissue Antigens* 2000; **55**: 239–49.
14. Roitberg-Tambur A, Witt CS, Friedmann A et al. Comparative analysis of HLA polymorphism at the serological level in Moroccan and Ashkenazi Jews. *Tissue Antigens* 1995; **46**: 104–10.
15. Clayton J, Lonjou C. Allele and haplotype frequencies for HLA loci in various ethnic groups. In: Charron D, ed. *HLA Genetic Diversity of HLA Functional and Medical Implications, Vol. II. Paris: EDK Medical and Scientific International Publishers 1997, 665–820.*
16. Amar A, Kwon OJ, Motro U et al. Molecular analysis of HLA class II polymorphisms among different ethnic groups in Israel. *Hum Immunol* 1999; **60**: 723–30.
17. Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* 1983; **19**: 153–70.
18. Young F, Bann CM. A visual statistics system. In: Stine, RA, Fox, J, eds. *Statistical Computing Environments for Social Researchers. New York: Sage Publications, 1996, 207–36.*
19. Shankarkumar U. HLA19 and B14 frequencies increased among Parsis from Mumbai. *MHC* 2002; **9**: 71.
20. Mehra NK, Taneja V, Kailash S, Raizada N, Vaidya MC. Distribution of HLA antigens in a sample of the North Indian Hindu population. *Tissue Antigens* 1986; **27**: 64–74.
21. Farjadian S, Naruse T, Kawata H, Ghaderi A, Bahram S, Inoko H. Molecular analysis of HLA allele frequencies and haplotypes in Baloch of Iran compared with related populations of Pakistan. *Tissue Antigens* 2004; **64**: 582–7.
22. Quintana-Murci L, Krausz C, Zerjal T et al. Y-chromosome lineages trace diffusion of people and languages in Southwestern Asia. *Am J Hum Genet* 2001; **68**: 537–42.
23. Qamar R, Ayub Q, Mohyuddin A et al. Y chromosomal DNA variation in Pakistan. *Am J Hum Genet* 2002; **70**: 1107–24.