Available online at www.blood-genomics.com

Blood and Genomics

2021, 5(1): 21−28

B & G

Original Article

Distribution of HLA class Ⅰ and class Ⅱ haplotypes in Chinese Han population based on family segregation

Lijun Wang, Wei Li, Dongmei Wang, Na Liu, Yanjun Jia*

Department of HLA, Beijing Red Cross Blood Center, Beijing 100088, China.

ABSTRACT

HLA haplotype analysis has important application value in human population genetics, anthropological research and HLA matching transplantation. Based on HLA-A, -B, -C, -DRB1 and -DQB1 genotyping data from 663 families including 663 leukemia patients and 991 related donors, the allele frequency (AF) and haplotype frequency (HF) of two-, three- and five-locus haplotype distribution patterns in the Chinese Han population were determined by family segregation. A total of 38 alleles at A locus, 75 alleles at B locus, 35 alleles at C locus, 53 alleles at DRB1 locus and 22 alleles at DQB1 locus were discovered in this population. The frequencies of these alleles were basically consistent with those of previous reports except for some tiny differences. The study found 11 A-C, 15 C-B, 4 B-DRB1 and 11 DRB1-DQB1 two-locus haplotypes with a frequency over 2%. The number of A-C-B and A-B-DRB1 three-locus haplotype with a frequency over 1% were 11 and 3 respectively. The most common HLA-A-C-B-DRB1-DQB1 haplotype (HF>1%) were A*3001-C*0602-B*1302-DR*0701-DQ*0202 (4.30%), A*0207-C*0102-B*4601-DR*0901-DQ*0303 (3.07%), A*3303-C*0302-B*5801-DR*0301-DQ*0201 (1.49%) and A*1101-C*0102-B*4601-DR*0901-DQ*0303 (1.01%). The results are helpful for finding matching donors for hematopoietic stem cell transplant patients and also contribute to transplant immunology, HLA-related diseases, research of human genetics and other fields.

Keywords: HLA polymorphism, allele frequency, haplotype frequency, family segregation

INTRODUCTION

Haplotype is a term that describes a group of specific or allelic genes inherited from the same parents. For human leucocyte antigen (HLA) genes, HLA haplotype is a tightly linked set of HLA alleles located on chromosome 6p21.31 that can be passed down from generation to generation. HLA haplotype analysis has important application value in human population genetics, anthropological research and HLA matching transplantation. For hematopoietic stem cell transplant patients, haplotype data can be used to estimate the chances of finding a fully matched donor from the HSCT unrelated donor pool and to help doctors develop improved donor selection strategies. There are two ways to determine a patient's HLA haplotype: family segregation or advanced prediction algorithm such as the expectation-maximization algorithm. In population genetic studies, the haplotype frequency (HF) obtained by family separation is closer to the real situation, however these studies are often difficult to carry out. In the past, most studies on Chinese group HLA haplotype were determined by the expectation

*Corresponding to: Yanjun Jia, Department of HLA, Beijing Red Cross Blood Center, No. 37 Beisanhuan Zhong Road, Beijing 100088, China. E-mail: jiayj321@hotmail.com.

Conflict of interests: The authors declared no conflict of interests.

https://doi.org/10.46701/BG.2021012021109

maximization algorithm[1–4]. Through the investigation of 167 families, the Chinese three-locus and two-locus haplotype frequencies of HLA-A-B-DRB1, A-B, B-DRB1 and A-DRB1 were analyzed by the First Affiliated Hospital of Nanjing Medical University[5], but without five-locus HLA haplotype frequency.

This laboratory has collaborated with Data Bank of the Chinese Hematopoietic Stem Cell Donor (also known as China Marrow Donor Program, CMDP) since 2001, and has so far carried out HLA high-resolution genotyping for more than 14 000 patients and donors, and accumulated a large amount of HLA genotyping data from different families. Based on these data, this study analyzed 663 families, including 663 leukemia patients and 991 donors, to determine the most common HLA alleles and two-, three- and five-locus haplotype distribution patterns in the Chinese Han population.

MATERIALS AND METHODS

Subjects

This study enrolled 663 families, including 663 patients awaiting hematopoietic stem cell transplants, and 991 related donors from January 1, 2010 to December 31, 2018. The patients were mainly acute myeloid leukemia, chronic myeloid leukemia, acute lymphoblastic leukemia, myelodysplastic syndromes, aplastic anemia, etc., as shown in ***Table 1***. The family relationships were 53.24% parent-child and 46.76% siblings. HLA high-resolution genotyping of these patients and donors were conducted. The two-locus haplotype included HLA-A-C, C-B, B-DRB1 and DRB1-DQB1, the three-locus haplotype included HLA-A-C-B and A-B-DRB1, and the five-locus haplotype was HLA-A-C-B-DRB1-DQB1. The determination method of haplotype referred to a previous report[6]. In short, 4 haplotypes were obtained if data from the patient and both parents were available, and 3 haplotypes were determined if data from one patient was available. Failing to obtain data from parents in a family, 3 or 4 haplotypes were provided from the patient's brother or sister. The haplotype data would be discarded if recombination occurred or if it was not possible to determine the haplotype because the locus was homozygote. The study was approved by the Beijing Red Cross Blood Center's Ethics Committee.

HLA genotyping

The whole blood genomic DNA of the patients or donors was extracted using DNA extraction kit (Qiagen Inc., Valencia, CA, USA) on the Hamilton Microlab Starlet Fully Automated DNA Extraction Platform (Hamilton Robotics, Switzerland), then genotyping of HLA-A, -B, -C, -DRB1 and -DQB1 loci was performed using the SBT method with AlleleSEQR HLA typing kit (Celera Corporation, Alameda, CA, USA). At the same time, sequence-specific oligonucleotide sequencing kit (One Lambda, Canoga Park, CA, USA) was used to further confirm the correctness of the sequences by polymerase chain reaction reverse sequence-specific oligonucleotide probe method (PCR-SSO).

Table 1 **The general characterization of patients and donors**

Characterization	No.	%
Total number of recipients	663	100
Recipient's gender		
Male	415	62.59
Female	248	37.41
Diseases		
ALL	167	25.12
AML	238	35.97
CML	77	11.67
MDS	47	7.08
SAA	65	9.79
Others	69	10.38
Number of donors	991	-
Donor's gender		
Male	396	39.96
Female	595	60.04
Relation between recipient and donor		
Parent and child	353	53.24
Brother and sister	310	46.76

AML=acute myeloid leukemia; CML=chronic myeloid leukemia; ALL=acute lymphocytic leukemia; SAA=Severe aplastic anemia; MDS=myelodysplastic syndrome.

Statistical analysis

The allele frequency (AF) of HLA-A, -B, -C, -DRB1, -DQB1 and haplotype frequency (HF) were calculated directly by EXCEL software, and the frequency of each locus was compared with the data published by CMDP. Statistical analysis was performed by Chi-square test with $P<0.05$ as statistical difference. The control was selected from HLA alleles from the "Common and Well-documented HLA Alleles List (CWD)" published in 2015 (Known as CWD 2015) by CMDP. For statistical balance, the number of cases in the control group was set to 3 500, which was consistent with the number of cases in this study. In addition, the alleles with an occurrence number of less than 3 were excluded from analysis, as a large margin of error would occur if statistical analysis was performed. The frequencies of haplotypes were compared with data

reported in previous reports.

RESULTS

HLA class Ⅰ and Ⅱ allele frequencies

The number of HLA-A, -B, -C, -DRB1, -DQB1 alleles and two-locus, three-locus, five-locus haplotypes was discovered in this population as shown in ***Table 2***. The frequency of different alleles at five loci was listed in ***Table 3***. Of the 38 HLA-A alleles, the most common alleles (AF>5%) were A*1101 (19.14%), A*2402 (18.18%), A*0201 (12.87%), and A*0207 (9.87%), A*3001 (6.90%), A*3303 (6.27%) and A*0206 (6.12%), accounting for 79.35% of the total HLA-A alleles. Of the 75 HLA-B alleles, 5 alleles with frequency greater than 5% were B*4601 (11.95%), B*4001 (10.00%), B*1302 (7.08%), B*5101 (5.16%) and B*1501 (5.09%), accounting for 39.28% of the total B alleles. Thirty-five alleles were found at the HLA-C locus, the most common alleles (AF>5%) were C*0102 (16.73%), C*0702 (15.81%), C*0602 (10.22%), C*0304 (9.54%), C*0303 (8.67%), C*0801 (8.15%) and C*0401 (5.63%), accounting for 74.75% of the total C alleles. There were 52 HLA-DRB1 alleles in this population including 6 alleles with AF over 5%, such as DRB1*0901 (16.11%), DRB1*1501 (12.71%), DRB1*0701 (9.64%), DRB1*1202 (8.21%), DRB1*1101 (6.34%) and DRB1*0803 (6.18%), accounting for 59.19% of total DRB1 alleles. Of 22 HLA-DQB1 alleles, those with AF over 5% were DQB1*0301 (22.93%), DQB1*0303 (17.96%), DQB1*0601 (10.30%), DQB1*0602 (9.18%), DQB1*0202 (7.63%) and DQB1*0502 (5.94%), accounting for 73.94% of the total DQB1 alleles.

Table 2 **Summary of HLA alleles and haplotypes in 663 patients and 991 donors**

Locus	No.	Two-locus haplotypes	No.	Three-locus haplotypes	No.	Five-locus haplotypes	No.
HLA-A	38	HLA-A-C	261	HLA-A-C-B	585	HLA-A-C-B-DRB1-DQB1	1 081
HLA-B	75	HLA-C-B	211	HLA-A-B-DRB1	1 063		
HLA-C	35	HLA-B-DRB1	529				
HLA-DRB1	53	HLA-DRB1-DQB1	141				
HLA-DQB1	22						

HLA = human leucocyte antigen.

Table 4 listed HLA alleles showing significant differences between AF in this population and AF published in the CWD 2015. Some had a greater AF than the control, such as A*0207, A*2402, B*1402, B*1511, B*3501, B*4040, B*4601, C*0303, DRB1*1201, DRB1*1401, DQB1*0303 and DQB1*0602. Others had a lower AF than the control including A*0101, A*1102, A*3303, A*6801, B*1502, B*5801, C*0302, C*0403, DRB1*0301, DRB1*1602, DQB1*0201 and DQB1*0502.

HF of two-locus haplotypes

In this population, a total of 261 A-C, 211 C-B, 529 B-DRB1 and 141 DRB1-DQB1 haplotypes were found, as shown in ***Table 5***. The all alleles with HF greater than 2% were listed in ***Table 5***. 11 A-C haplotypes with frequency greater than 2% accounted for 40.43% of the total A-C haplotypes. The number of C-B haplotypes with frequency greater than 2% were 15, accounting for 53.14% of the total C-B haplotypes. The number of B-DRB1 haplotypes with frequency greater than 2% were 4, accounting for 16.12% of the total B-DRB1 haplotypes. There were 11 DRB1-DQB1 haplotypes with HF over 2%, accounting for 64.28% of the total DRB1-DQB1 haplotypes.

HF of three-locus haplotypes

There were 585 HLA-A-C-B and 1 063 HLA-A-B-DRB1 haplotypes found in 663 families in this study, as shown in ***Table 6***. Eleven (11) HLA-A-C-B haplotypes with HF over 1% comprised 27.82% of the total HLA-A-C-B haplotypes. Only 3 HLA-A-B-DRB1 haplotypes had HF greater than 1%, accounting for 11.85% of the total HLA-A-B-DRB1 haplotypes. The HF of these three-locus haplotypes was shown in ***Table 5***.

HF of five-locus haplotypes

Table 7 listed the top 20 of five-locus HLA-A-C-B-DRB1-DQB1 haplotypes, which together comprised 18.53% of the total five-locus haplotype. The most common HLA-A-C-B-DRB1-DQB1 haplotypes (HF>1%) were A*3001-C*0602-B*1302-DR*0701-DQ*0202 (4.30%), A*0207-C*0102-B*4601-DR*0901-DQ*0303 (3.07%), A*3303-C*0302-B*5801-DR*0301-DQ*0201 (1.49%) and A*1101-C*0102-B*4601-DR*0901-DQ*0303 (1.01%).

DISCUSSION

HLA genotyping for patients and donors is a

Table 3 **Allele frequencies of HLA-A, B, C, DRB1 and DQB1 in 663 families**

Alleles	AF (%)	Alleles	AF (%)	Alleles	AF (%)	Alleles	AF (%)	Alleles	AF (%)
A*0101	2.45	B*1302	7.08	B*4501	0.32	C*1203	1.84	DRB1*1210	0.06
A*0201	12.87	B*1402	0.29	B*4601	11.95	C*1402	4.03	DRB1*1301	1.71
A*0203	2.74	B*1501	5.09	B*4701	0.07	C*1403	0.88	DRB1*1302	2.65
A*0205	0.45	B*1502	2.42	B*4801	3.03	C*1425	0.04	DRB1*1303	0.13
A*0206	6.12	B*1505	0.22	B*4803	0.14	C*1501	0.08	DRB1*1305	0.03
A*0207	9.87	B*1507	0.58	B*4901	0.32	C*1502	3.51	DRB1*1307	0.06
A*0210	0.52	B*1510	0.04	B*5001	0.51	C*1504	0.04	DRB1*1312	0.55
A*0253	0.11	B*1511	2.64	B*5101	5.16	C*1505	0.80	DRB1*1401	0.32
A*0285	0.04	B*1512	0.14	B*5102	1.16	C*1512	0.04	DRB1*1402	0.03
A*0301	3.71	B*1513	0.04	B*5201	2.96	C*1604	0.08	DRB1*1404	0.49
A*0302	0.19	B*1517	0.36	B*5301	0.07	C*1701	0.12	DRB1*1405	2.75
A*1101	19.14	B*1518	1.66	B*5401	2.64	C*5102	0.04	DRB1*1406	0.06
A*1102	0.96	B*1521	0.04	B*5501	0.11			DRB1*1407	0.42
A*1153	0.11	B*1525	0.25	B*5502	2.20	DRB1*0101	1.78	DRB1*1410	0.03
A*2301	0.30	B*1527	0.76	B*5504	0.04	DRB1*0102	0.32	DRB1*1411	0.06
A*2402	18.18	B*1529	0.07	B*5601	0.25	DRB1*0201	0.04	DRB1*1418	0.03
A*2403	0.04	B*1532	0.14	B*5701	1.05	DRB1*0202	0.03	DRB1*1454	3.14
A*2404	0.04	B*1535	0.11	B*5801	4.48	DRB1*0301	3.88	DRB1*1501	12.71
A*2407	0.04	B*1558	0.07	B*5901	0.04	DRB1*0401	1.20	DRB1*1502	2.94
A*2408	0.11	B*1601	0.04	B*6701	0.61	DRB1*0402	0.10	DRB1*1504	0.16
A*2410	0.04	B*1801	0.18	B*7301	0.04	DRB1*0403	1.36	DRB1*1601	0.10
A*2420	0.30	B*2702	0.04	B*8102	0.07	DRB1*0404	0.45	DRB1*1602	2.17
A*2601	2.63	B*2704	0.54			DRB1*0405	3.82		
A*2901	0.78	B*2705	0.51	C*0102	16.73	DRB1*0406	2.20	DQB1*0201	3.82
A*3001	6.90	B*2707	0.07	C*0103	0.68	DRB1*0407	0.19	DQB1*0202	7.63
A*3002	0.07	B*2724	0.11	C*0106	0.04	DRB1*0408	0.06	DQB1*0301	22.93
A*3004	0.04	B*3501	3.29	C*0202	0.48	DRB1*0410	0.23	DQB1*0302	4.97
A*3101	3.08	B*3502	0.29	C*0302	4.55	DRB1*0601	0.03	DQB1*0303	17.96
A*3122	0.04	B*3505	0.04	C*0303	8.67	DRB1*0604	0.18	DQB1*0317	0.07
A*3201	1.11	B*3508	0.07	C*0304	9.54	DRB1*0701	9.64	DQB1*0401	3.53
A*3301	0.19	B*3701	1.12	C*0356	0.08	DRB1*0702	0.03	DQB1*0402	1.04
A*3303	6.27	B*3801	0.47	C*0401	5.63	DRB1*0801	0.03	DQB1*0501	3.74
A*3401	0.04	B*3802	3.21	C*0403	0.36	DRB1*0802	0.65	DQB1*0502	5.94
A*3901	0.04	B*3901	1.70	C*0501	1.00	DRB1*0803	6.18	DQB1*0503	4.39
A*6801	0.33	B*3905	0.14	C*0602	10.22	DRB1*0809	0.13	DQB1*0504	0.07
A*7402	0.07	B*3924	0.04	C*0701	0.88	DRB1*0901	16.11	DQB1*0601	10.30
A*7402	0.04	B*4001	10.00	C*0702	15.81	DRB1*0904	0.10	DQB1*0602	9.18
A*7413	0.07	B*4002	1.95	C*0704	1.16	DRB1*0910	0.06	DQB1*0603	1.51
		B*4003	0.40	C*0706	0.64	DRB1*1001	1.23	DQB1*0604	1.12
B*0302	0.04	B*4006	3.72	C*0726	0.04	DRB1*1101	6.34	DQB1*0605	0.04
B*0702	2.64	B*4040	0.29	C*0766	0.04	DRB1*1104	0.74	DQB1*0609	1.40
B*0705	0.79	B*4078	0.04	C*0801	8.15	DRB1*1106	0.03	DQB1*0701	0.07
B*0706	0.04	B*4101	0.14	C*0802	0.24	DRB1*1106	0.03	DQB1*1501	0.07
B*0801	0.83	B*4130	0.07	C*0803	0.80	DRB1*1119	0.03		
B*1001	0.04	B*4402	1.16	C*0806	0.12	DRB1*1201	4.17		
B*1301	4.77	B*4403	2.06	C*1202	2.64	DRB1*1202	8.21		

HLA = human leucocyte antigen; AF = allele frequency.

necessary step before hematopoietic stem cell transplantation. The current most commonly used HLA genotyping methods are SBT and PCR-SSO, wherein the SBT method is the golden standard of HLA genotyping technology[6], from which high-resolution HLA genotyping results can be obtained. The data in this study came from both SBT and PCR-SSO methods, as each patient and donor was sequenced by SBT method and then confirmed by PCR-SSO method to ensure accuracy.

To date, polymorphism data for the Chinese population's HLA class Ⅰ and class Ⅱ genes have

Table 4 **HLA alleles showing significant differences in frequency comparing with AF of control in CWD**

HLA alleles	Allele count in this population (AF, %)	Allele count in control (AF, %)	*P*
A*0101	81 (2.45)	81 (3.60)	0.0057
A*0207	326 (9.87)	326 (8.44)	0.0454
A*1102	32 (0.96)	32 (1.74)	0.0080
A*2402	601 (18.18)	601 (15.54)	0.0042
A*3303	207 (6.27)	207 (8.19)	0.0024
A*6801	11 (0.33)	11 (0.71)	0.0451
B*1402	10 (0.29)	10 (0.05)	0.0339
B*1502	80 (2.42)	80 (3.59)	0.0197
B*1511	87 (2.64)	87 (1.83)	0.0153
B*3501	109 (3.29)	109 (2.45)	0.0172
B*4040	10 (0.29)	10 (0.05)	0.0339
B*4601	395 (11.95)	395 (10.26)	0.0031
B*5801	148 (4.48)	148 (6.14)	0.0156
C*0302	151 (4.55)	151 (5.91)	0.0147
C*0303	287 (8.67)	287 (6.96)	0.0100
C*0403	12 (0.36)	12 (1.00)	0.0025
DRB1*0301	128 (3.88)	128 (5.12)	0.0157
DRB1*1201	138 (4.17)	138 (2.47)	0.0001
DRB1*1401	11 (0.32)	11 (0.01)	0.0019
DRB1*1602	72 (2.17)	72 (3.10)	0.0237
DQB1*0201	126 (3.82)	126 (4.93)	0.0225
DQB1*0303	594 (17.96)	594 (15.84)	0.0208
DQB1*0502	196 (5.94)	196 (7.22)	0.0343
DQB1*0602	304 (9.18)	304 (7.73)	0.0318

HLA = human leucocyte antigen; AF = allele frequency; CWD = Common and Well-documented HLA Alleles List.

Table 5 **Common (HF>2%) HLA two-locus haplotypes in 663 families**

Haplotypes	HF (%)	Previously reported HF (%)[3]	Haplotypes	HF (%)	Previously reported HF (%)[3]	Haplotypes	HF (%)	Previously reported HF (%)[3]
A*0207-C*0102	6.77	NA	C*0302-B*5801	4.30	5.29	B*1301-DR*1202	2.28	NA
A*3001-C*0602	6.17	NA	C*0304-B*1301	4.06	5.36			
A*1101-C*0702	4.58	NA	C*1402-B*5101	3.74	4.04	DR*0901-DQ*0303	15.10	14.27
A*0201-C*0303	3.62	NA	C*0702-B*3802	2.91	3.34	DR*1501-DQ*0602	8.55	7.94
A*3303-C*0302	3.27	NA	C*0801-B*4006	2.71	3.41	DR*1202-DQ*0301	7.55	8.57
A*1101-C*0102	3.23	NA	C*0304-B*4001	2.67	3.34	DR*0701-DQ*0202	7.23	8.57
A*2402-C*0702	3.15	NA	C*0702-B*0702	2.51	NA	DR*1101-DQ*0301	6.47	5.2
A*1101-C*0304	2.75	NA	C*0303-B*1511	2.47	NA	DR*0803-DQ*0601	5.79	6.48
A*2402-C*0102	2.39	NA	C*0102-B*5401	2.15	NA	DR*1201-DQ*0301	3.70	3.44
A*0201-C*0702	2.27	NA	C*0303-B*1501	2.15	3.27	DR*0301-DQ*0201	3.63	3.69
A*2402-C*0304	2.23	NA	C*0801-B*1502	2.15	NA	DR*0405-DQ*0401	3.31	4.32
						DR*1501-DQ*0601	2.34	NA
C*0102-B*4601	10.99	8.71	B*1302-DR*0701	5.82	6.19	DR*1405-DQ*0503	2.05	3.06
C*0602-B*1302	7.21	6.96	B*4601-DR*0901	5.63	5.16			
C*0702-B*4001	5.10	5.65	B*5801-DR*0301	2.39	NA			

HLA = human leucocyte antigen; HF = haplotype frequency.

been reported in previous studies[1–4], as well as HLA genotyping data for populations from different regions and ethnics[7–11]. CMDP also regularly publishes the frequency of HLA's different alleles. All of these data provide references for this study. The data in this study were all from the Han population and were not distinguished according to region. This is because it is very difficult to clearly define a person's historical origin, given the large-scale population migrations that have taken place in recent decades in China. From the results of this study, the allele frequencies of the HLA loci in the population enrolled in this study were fundamentally the same as previously reported. For example, in this study, the alleles with frequencies of more than 10% at HLA-A locus were A*1101 (19.14%), A*2402 (18.18%), and A*0201 (12.87%), which was similar to Hei AL *et al*[3] and Deng YJ *et al*[12], and was also consistent with the frequencies

Table 6 **Common (HF>1%) HLA three-locus haplotypes in 663 families**

Haplotypes	HF (%)	Previously reported HF (%)[3]	Haplotypes	HF (%)	Previously reported HF (%)[3]
A*0207-C*0102-B*4601	6.49	7.06	A*1101-C*0801-B*1502	1.31	2.25
A*3001-C*0602-B*1302	5.81	5.99	A*0203-C*0702-B*3802	1.19	2.15
A*3303-C*0302-B*5801	3.07	4.59	A*1101-C*1402-B*5101	1.08	1.19
A*1101-C*0702-B*4001	2.43	2.71			
A*0201-C*0303-B*1511	1.95	1.18	A*3001-B*1302-DR*0701	4.55	5.36
A*1101-C*0102-B*4601	1.67		A*0207-B*4601-DR*0901	3.62	4.12
A*1101-C*0304-B*1301	1.47	3.05	A*3303-B*5801-DR*0301	1.49	2.37
A*2402-C*0702-B*4001	1.35	1.33			

HLA = human leucocyte antigen; HF = haplotype frequency.

Table 7 **Top 20 HLA five-locus haplotypes in 663 families**

Haplotypes	HF (%)	Order	Previously reported HF (%)[4]	Order
A*3001-C*0602-B*1302-DR*0701-DQ*0202	4.3	1	3.7	1
A*0207-C*0102-B*4601-DR*0901-DQ*0303	3.07	2	2.46	2
A*3303-C*0302-B*5801-DR*0301-DQ*0201	1.49	3	2.4	3
A*1101-C*0102-B*4601-DR*0901-DQ*0303	1.01	4	0.58	11
A*0207-C*0102-B*4601-DR*0803-DQ*0601	0.92	5	0.93	6
A*0201-C*0304-B*1301-DR*1202-DQ*0301	0.83	6	0.58	10
A*3303-C*0302-B*5801-DR*1302-DQ*0609	0.7	7	1.06	5
A*0101-C*0602-B*5701-DR*0701-DQ*0303	0.61	8	0.45	14
A*1101-C*0702-B*4001-DR*0803-DQ*0601	0.61	9	0.42	17
A*1101-C*0801-B*1502-DR*1202-DQ*0301	0.57	10	1.13	4
A*2402-C*0304-B*1301-DR*1202-DQ*0301	0.53	11	NA	NA
A*0203-C*0702-B*3802-DR*1602-DQ*0502	0.48	12	0.36	20
A*2402-C*0602-B*1302-DR*0701-DQ*0202	0.48	13	NA	NA
A*1101-C*0304-B*1301-DR*1202-DQ*0301	0.44	14	0.37	19
A*1101-C*0304-B*1301-DR*1501-DQ*0601	0.44	15	0.64	9
A*1101-C*0702-B*4001-DR*0901-DQ*0303	0.44	16	NA	NA
A*1101-C*0801-B*1502-DR*1501-DQ*0601	0.44	17	0.44	15
A*0206-C*0801-B*4006-DR*0901-DQ*0303	0.39	18	NA	NA
A*0207-C*0102-B*4601-DR*1454-DQ*0502	0.39	19	0.38	18
A*3303-C*1403-B*4403-DR*1302-DQ*0604	0.39	20	0.74	7

HLA = human leucocyte antigen; HF = haplotype frequency.

shown in CMDP's 2015 CWD list, such as A*1101 (20.95%), A*2402 (15.53%) and A*0201 (11.92%). The polymorphism of HLA-B allele was the highest. A total of 75 HLA-B alleles were found in this study. The most common alleles were B*4601 (11.95%), B*4001 (10.00%), B*1302 (7.08%), B*5101 (5.16%) and B*1501 (5.09%). Hei AL *et al*[3] and the CWD-2015 list showed the same frequency, but these alleles only accounted for 39.28% of the total B alleles, which also showed their complex polymorphism. The most common alleles (AF > 5%) on HLA-C locus were C*0102 (16.73%), C*0702 (15.81%), C*0602 (10.22%), C*0304 (9.54%), C*0303 (8.67%), C*0801 (8.15%) and C*0401 (5.63%). Although there were only 7 alleles, they accounted for 74.75% of the total, indicating that their polymorphism was far less than that of HLA-B alleles. In the CWD-2015 list, these alleles were highly polymorphic with AF of 15.78%, 15.15%, 8.92%, 9.94%, 6.96% and 8.54% respectively. Although there was a slight difference with results of this study in the order of alleles, the AF of them was similar. Among HLA-Ⅱ alleles, HLA-DRB1 polymorphism was relatively high. A total of 52 alleles were found, among which 6 alleles with frequencies greater than 5% were DRB1*0901 (16.11%), DRB1*1501 (12.71%), DRB1*0701 (9.64%), DRB1*1202 (8.21%), DRB1*1101 (6.34%) and DRB1*0803 (6.18%), respectively. The results were consistent with those reported previously, such as their frequencies of 74%, 11.67%, 9.66%, 8.65%, 5.60% and 6.28% respectively in the CWD-2015 list. The most common alleles (AF > 5%) at HLA-DQB1 locus were DQB1*0301 (22.93%), DQB1*0303 (17.96%), DQB1*0601 (10.30%), DQB1*0602 (9.18%), DQB1*0202 (7.63%) and DQB1*0502 (5.94%), which accounted for 73.94% of the total HLA-DQB1 allele. This was consistent with the order and frequency shown in the CWD-2015 list, as

21.02%, 15.84%, 10.22%, 7.73%, 7.70% and 7.22%, respectively. Relatively speaking, HLA-DQB1 polymorphism was low. Only 22 alleles were found in this study.

By comparing our results with AF data published in the CWD in 2015, some HLA alleles showed statistical differences in this population. The factors that caused the differences may include the variance in cohort sizes or subjects enrolled. For example, although the case number of control group was set to 3 500, the HLA allele frequencies in the 2015 CWD were calculated from 706 841 subjects for HLA-A, -B and-DRB1, 470 328 subjects for HLA-C and 439 617 subjects for HLA-DQB1. The number of cases in this study was limited to 1 654, so deviations in the frequency of some alleles are unavoidable. Moreover, most subjects in this study were patients with different hematological diseases and their related family members, but the AF of control was calculated from healthy unrelated hematopoietic stem cell transplant donors. Many researches have reported the difference in HLA allele frequency between hematological disease patients and healthy volunteers. For example, Fu RT *et al*[13] found that HLA-B*1502, -B*1511, -DQB1*0303 and -DQB1*0602 had correlation with the effects of immunosuppressive therapy in severe pediatric aplastic anemia patients. Al-Tonbary Y *et al*[14] reported that HLA-DQB1*0201 might be the susceptible gene in Non-Hodgkin's lymphoma, while HLA-DQB1*0502 and –DQB1*0602 might be protective genes for it. In northern Chinese Han AA patients, the frequencies of HLA-B*3501, -B*4601, -DRB1*15:01 and -DQB1*06:02 were significantly elevated as compared with healthy controls[15−16]. Qi J *et al*[17] found that the frequencies of A*0207 in Northern Han patients with AML (*n*=189) were higher than in the control group (*n*=1 241), while the frequency of B*4601 in CML group (*n*=70) was lower than that in the control group. However, none of the above alleles showed statistical differences when corrected by Bonferroni method. In an unpublished study by our team, some HLA alleles showed relevance with the occurrence of different hematological diseases, which also had significant differences in AF between experiment group of this study and control, such as HLA-A*3303, -B1511, -B4601, -B*5801, -C*0302, -DRB1*0301, -DRB1*1201, -DRB1*1602, -DQB1*0201, -DQB1*0303, -DQB1*0502, and –DQB1*0602. All of these results indicated that the different subjects enrolled in study led to different HLA allele frequencies in various populations.

There have been very few reports of HLA haplotype distribution characteristics in the Chinese population through the method of family segregation, to date we have found only one by Pan QQ *et al*[5]. This article only reported on the distribution of two-locus haplotype A-B, B-DRB1, A-DRB1 and three-locus haplotype HLA-A-B-DRB1 in 167 families, which may have also been limited by experimental data available at that time. Based on HLA genotyping data from 663 families, distribution characteristics of two- locus haplotype HLA-A-C, HLA-C-B, HLA-B-DRB1, HLA-DRB1-DQB1, three-locus haplotype HLA-A-C-B, HLA-A-B-DRB1, and five-locus haplotype HLA-A-C-B-DRB1-DQB1 were analyzed in this study. The two-, three-, four- and five-locus HLA haplotypes reported by Hei AL *et al*[3] were based on the expectation maximization algorithm, and were theoretically different from the actual distribution. By comparing the characteristics of the haplotype distribution shown in this paper with previous results, it can be seen that the distribution of the two-locus haplotype was basically consistent with the report of the two above-mentioned reports[3,5]. For the three-locus haplotype HLA-A-C-B, the data in this paper and reports from Hei AL *et al*[3], both showed the top three haplotypes were A*0207-C*0102-B*4601, A*3001-C*0602-B*1302, A*3303-C*0302-B*5801, and the HF difference was not significant. However in this population, the frequency of A*1101-C*0304-B*1301 was 1.47% (No. 7), while Hei AL *et al*.'s data was 3.05% (No. 4)[3], showing significant difference. For HLA-A-B-DRB1 haplotype, the data in this paper was consistent with those of Hei AL *et al*, but significantly different from those reported by Pan QQ *et al*[5], especially for A*3001-B*1302-DRB1*0701. Although the frequency of this haplotype was the highest in both studies, the frequency in this study was 4.55%, while the frequency reported by Pan QQ *et al*. was 8.54%[5]. In addition, A*0207-B*4601-DRB1*0901 and A*3303-B*5801-DRB1*0301 were the 2nd and 3rd high three-locus haplotype in this study respectively, while they were the 3rd and 2nd in Pan QQ *et al*[5]. This may be due to geographical factors. The study by Pan QQ *et al*[5] included more Han people from the south, while this study and the study by Hei AL *et al*[3] enrolled more northern Han populations. For the five-locus haplotypes, the top 3 were consistent with reports from Zhou XY *et al*, but there were also some differences in several five-locus haplotypes between our results and those of Zhou XY *et al*[4], such as A*1101-C*0102-B*4601-DRB1*0901-DQB1*0303 (1.01%, No.4 *vs* 0.58%, No.11), A*1101-C*0801-B*1502-DRB1*1202-DQB1*0301 (0.57%, No.10 *vs*

1.13%, No.4), A*1101-C*0304-B*1301-DRB1*1501-DQB1*0601 (0.44%, No.15 *vs* 0.64%, No.9) and A*3303-C*1403-B*4403-DRB1*1302-DQB1*0604 (0.39%, No.20 *vs* 0.74%, No.7). Some of top 20 five-locus haplotypes in this study, such as A*2402-C*0304-B*1301-DRB1*1202-DQB1*0301, A*2402-C*0602-B*1302-DRB1*0701-DQB1*0202, A*1101-C*0702-B*4001-DRB1*0901-DQB1*0303 and A*0206-C*0801-B*4006-DRB1*0901-DQB1*0303, were not included in the top 20 five-locus haplotypes in the study by Zhou XY *et al*[4]. The difference in haplotype frequency between different researches could be resulted from different calculation methods, such as family segregation or expectation –maximization algorithm. Other reasons for HF difference may also be the contribution of HLA ethnic and geographic diversity as well as various enrolled subjects. As mentioned above, there were many differences in allele frequencies between patient group and the healthy population, which also could lead to the difference in haplotype frequency. Of course, the cohort size of this study could be one of factors influencing the haplotype frequency.

In summary, the allele and haplotype frequencies of patients and donors from 663 Han families by means of family segregation were analyzed in this study. By comparing the data with previous reports, we found that the data in this paper was consistent with the HLA allele distribution characteristics of Han Chinese population in both AF and HF. The haplotype data here might be closer to the true characteristics of HLA polymorphism in the Han Chinese population. This study should be helpful for matching hematopoietic stem cell transplant patients and donors, and may also contribute to transplant immunology, HLA-related diseases, research of human genetics and other fields.

References

[1] Shen CM, Zhu BF, Deng YJ, et al. Allele polymorphism and haplotype diversity of HLA-A, -B and -DRB1 loci in sequence-based typing for Chinese Uyghur ethnic group[J]. *PLoS One*, 2010, 5(11): e13458.

[2] Li XF, Zhang X, Chen Y, et al. An analysis of HLA-A, -B, and -DRB1 allele and haplotype frequencies of 21, 918 residents living in Liaoning, China[J]. *PLoS One*, 2014, 9(4): e93082.

[3] Hei AL, Li W, Deng ZH, et al. Analysis of high-resolution HLA-A, -B, -Cw, -DRB1, and -DQB1 alleles and haplotypes in 718 Chinese marrow donors based on donor –recipient confirmatory typings[J]. *Int J Immunogenet*, 2009, 36(5): 275−282.

[4] Zhou XY, Zhu FM, Li JP, et al. High-resolution analyses of human leukocyte antigens allele and haplotype frequencies based on 169, 995 volunteers from the China bone marrow donor registry program[J]. *PLoS One*, 2015, 10(9): e0139485.

[5] Pan QQ, Fan S, Wang XY, et al. The distribution of human leukocyte antigen-A, -B, and -DRB1 alleles and haplotypes based on high-resolution genotyping of 167 families from Jiangsu Province, China[J]. *Hum Immunol*, 2011, 72(10): 872−876.

[6] Pedron B, Yakouben K, Adjaoud D, et al. Listing of common HLA alleles and haplotypes based on the study of 356 families residing in the Paris, France, area: implications for unrelated hematopoietic stem cell donor selection[J]. *Hum Immunol*, 2005, 66(6): 720−730.

[7] Danzer M, Polin H, Pröll J, et al. High-throughput sequence-based typing strategy for HLA-DRB1 based on real-time polymerase chain reaction[J]. *Hum Immunol*, 2007, 68(11): 915−917.

[8] Zou HY, Jin SZ, Li Z, et al. Study on alleleic and haplotypic polymorphism of human leukocyte antigen -A, -B, -DRB1 genes at high-resolution level in Northern Chinese Han hematopoietic stem cell donor[J]. *Int J Blood Transfus Hematol* (in Chinese), 2010, 33(2): 103−108.

[9] Shi L, Xu SB, Ohashi J, et al. HLA-A, HLA-B, and HLA-DRB1 alleles and haplotypes in Naxi and Han populations in southwestern China (Yunnan province)[J]. *Tissue Antigens*, 2006, 67(1): 38−44.

[10] Shi L, Ogata S, Yu JK, et al. Distribution of HLA alleles and haplotypes in Jinuo and Wa populations in Southwest China[J]. *Hum Immunol*, 2008, 69(1): 58−65.

[11] Yao Y, Shi L, Tao Y, et al. Distinct HLA allele and haplotype distributions in four ethnic groups of China[J]. *Tissue Antigens*, 2012, 80(5): 452−461.

[12] Deng YJ, Yang G, Wu DY, et al. HLA-A, B, DRB1 gene polymorphism of Beijing population was studied by high-resolution polymerase chain reaction sequence-based typing[J]. *Chin J Med Genet* (in Chinese), 2006, 23(1): 103−106.

[13] Fu RT, Xue HM, Zhang BH, et al. Correlation analysis of severe aplastic anemia immunosuppressive therapy and human leukocyte antigen alleles in pediatric patients[J]. *Exp Ther Med*, 2015, 10(6): 2396−2402.

[14] Al-Tonbary Y, Abdel-Razek N, Zaghloul H, et al. HLA class Ⅱ polymorphism in Egyptian children with lymphomas[J]. *Hematology*, 2004, 9(2): 139−145.

[15] Wang BJ, Wu YM, Li XH, et al. Relationship between HLA gene polymorphism and aplastic anemia in Northern Chinese Han patients[J]. *J Exp Hematol* (in Chinese), 2018, 26(6): 1731−1737.

[16] Yang K, Guo XY, Han X, et al. Relationship between gene polymorphism of HLA-A/-B-DRB1 and aplastic anemia in Chinese Han population of North Western Plateau[J]. *J Exp Hematol* (in Chinese), 2016, 24(3): 795−800.

[17] Qi J, Chen LP, Wang TJ, et al. Association of HLA-A, -B, -DRB1 alleles and haplotypes with acute lymphoblastic leukemia among ethnic Hans from northern China[J]. *Chin J Med Genet* (in Chinese), 2017, 34(5): 737−742.

Received 19 March 2021, Revised 26 April 2021, Accepted 26 May 2021