The distribution of HLA- A, -B, and -DRB1 alleles and haplotypes in inhabitants of Guizhou Province of China

Qinquin Pan, Su Fan, Xiaoyan Wang, Xing Zhao, Meng Pan, Chengya Wang*, Jie Shen*

HLA Laboratory, the First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China.

Received 27 August 2010, Revised 30 October 2010, Accepted 23 March 2011

Abstract

The present study was aimed to analyze the frequencies of human leukocyte antigen (HLA)- A, -B, and -DRB1 alleles and A-B-DRB1, A-B, A-DRB1 and B-DRB1 haplotypes in inhabitants of Guizhou province, China. All samples were typed in the HLA-A, -B, and -DRB1 loci using the polymerase chain reaction-reverse sequence specific oligonucleotide probe (PCR-rSSOP) method and HLA polymorphisms were analyzed. A total of 18 HLA-A, 31 HLA-B, and 13 HLA-DRB1 alleles were found in the Guizhou population. The first two frequent alleles in the HLA-A, -B, and -DRB1 loci were A*11(30.72%) and A*02(30.65%), B*40(16.27%) and B*46(16.27%), and DRB1*09(15.91%) and DRB1*15(13.51%), respectively. The most common haplotype was A*02-B*46-DRB1*09(5.59%) in A-B-DRB1, A*02-B*46(11.73%) in A-B, B*46-DRB1*09(7.49%) in B-DRB1, and A*02-DRB1*09(8.08%) in A-DRB1. Some haplotypes with strong linkage disequilibrium (LD) were found not only in the common haplotypes, such as A*33-B*58, B*30-DRB1*07, and B*33-DRB1*03, but also in the rare haplotypes, such as A*01-B*37, B*37-DRB1*10, and A*01-DRB1*10. Guizhou inhabitants shared some characteristics of the Southern Chinese population but also had their own unique features. Overall, HLA polymorphism in Guizhou population was more consistent with that of Chengdu population than that of other populations in China.

Keywords: human leukocyte antigen, allele, haplotype, linkage disequilibrium, Guizhou

INTRODUCTION

Guizhou province is located in the southwest of China. It adjoins Sichuan province and Chongqing municipality to the north, Yunnan province to the west, Guangxi province to the south and Hunan province to the east. Guizhou is a mountainous province; however, while it is mountainous in the west, the eastern and southern regions are relatively flat. Guizhou covers an area of over 176,000 square kilometers with a total population of more than 35,245,000. Guizhou is one of the provinces that contain the greatest number of minority groups. There are 49 ethnic groups living there, with minorities making up about 38% of the population and their compositions rank third in China after Yunnan Province and Xinjiang Autonomous Region.

Human leukocyte antigen (HLA) genes are located at the short arm of chromosome 6 within a region of a few million base pairs. HLA is an extremely polymorphic genetic system and its constituent gene products play important roles in the immune response for unrelated hematopoietic stem cell transplantation[1-2].
HLA haplotype analysis is important for identifying appropriate donors, and the most important clinical application of HLA haplotype has been the selection of suitable donors in transplantation\[3\]. HLA matching at the haplotype level may have a higher likelihood of matching at other loci than matching merely at the allele level\[10\]. On the other hand, an accurate and adequate characterization of the distribution of HLA alleles and haplotypes at the population level may have been lagging. Hence, determination of the distribution of HLA alleles and haplotypes in different populations is necessary for selecting acceptable unrelated donors for patients. With the development of the Chinese Marrow Donor Program (CMDP), more and more HLA typing data have become available, which provides us a good chance for analyzing HLA polymorphism. In addition, HLA typing technology has developed rapidly with the development of CMDP, and PCR technology has been applied in the DNA-based HLA typing method. Techniques available for DNA typing include sequence specific oligonucleotide probes (SSOP), sequence-specific primers (SSP) and sequence-based typing (SBT). However SBT technology requires expensive equipment, and the first two techniques give rise to flexibility with respect to the desired level of resolution depending on the number of oligonucleotide probes or primers used\[5,6\].

In this paper, we examined the frequencies of HLA-A, -B, and -DRB1 alleles in a total of 2,879 persons residing in the Guizhou province of China. Furthermore, we estimated the frequencies of two or three locus haplotypes and the linkage disequilibrium test between two pairs of loci.

**MATERIALS AND METHODS**

**Subjects**

Analysis included 2,879 donors recruited into the CMDP Guizhou Branch from August 2006 to December 2007. All donors, regardless of ethnic groups, were included in this study (Han 85%; Miao, Dong, and Buji etc 15%; aged from 20-45 years) and were typed for HLA-A, HLA-B and HLA-DRB1 in our laboratory. The experiment protocol was approved by the Institutional Review Board of the First Affiliated Hospital of Nanjing Medical University, and all subjects signed informed consent.

**HLA typing**

All donors were typed using PCR-reverse SSOP (PCR-rSSOP) method for HLA-A, -B and -DRB1 using commercial kits (LABtype rSSO Typing Test, lot# A007, B009, DRB0010, OLI, CA, USA). LABType SSO is a reverse SSO (rSSO) DNA typing method using SSOP and color-coded microspheres to identify HLA alleles. First, genomic DNA was isolated from whole blood using the salting-out procedure with commercial kits (DNA Isolation Kit, Dynal Biotech, Brown Deer, Wisconsin, USA). The appropriate DNA concentration was 20-40 ng/μL and the relatively good purity of A_{260}/A_{280} was 1.6-1.8. Then, the sample DNA was subjected to PCR amplification (PE9700, Thermo cycler Life technologies, USA) in a 10 μL reaction volume, with the PCR run at 96°C for 3 min, 96°C for 20 s, 60°C for 20 s, and 72°C for 20 s, for 5 cycles, and 96°C for 10 s, 60°C for 15 s, and 72°C for 20 s for 30 cycles followed by 72°C for 10 min and stored at 4°C forever. After amplification, the PCR products were denatured and neutralized with acids and bases, and then the PCR products were hybridized with the corresponding locus beads at 60°C for 15 min, which were washed three times using the washing buffer. Then, streptavidin conjugated phycoerythrin (SAPE) was reacted with the products for 5 min at 60°C, and following washing, the products were detected using the Luminex 200 after being suspended with 60 μL washing buffer. Fluorescence signals were identified by the laser Luminex 200 (Luminex, USA), and lastly the HLA typing was obtained from the software HLAtools.

**Statistical analysis**

HLA allele frequencies (AF) were determined for each allele in donors using the formula: AF (%) = (n/2N)×100%, where n indicates the sum of a particular allele and N indicates the total number of individuals.

The maximum-likelihood haplotype frequencies, the Hardy-Weinberg equilibrium, and the linkage disequilibrium (LD) test were computed by the software Arlequin 3.01 using the expectation-maximization (EM) algorithm. Hardy-Weinberg exact tests were performed on all samples for each of the three HLA loci. The EM algorithm is a very general principle for handling missing data in statistical analysis. This algorithm has been described in detail somewhere as applied to estimation of multilocus haplotype frequencies. EM is an iterative method which alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood evaluated using the current estimate for the latent variables, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step\[7,8\].

The parameters reflecting LD intensity of D, D’, and r^2, and chi-square value given by Arlequin were also shown and the mathematic definitions of D, D’ and r^2 were given in detail elsewhere\[9\].
RESULTS

Hardy-Weinberg equilibrium examination

Hardy-Weinberg exact tests were performed on the three HLA loci. The observed, expected homozygosities and the statistical P value are given in Table 1. The results showed that the P values at the three loci were all more than 0.05. The P value was used to measure the magnitude of the deviation in a population sample, if a P value greater than 0.05, indicated that the population were consistent with Hardy-Weinberg equilibrium [10], which meant that the population was random and the sample size was adequately large [11,12].

| Locus | Genot | Exp.Heter. | P       | SD   |
|-------|-------|------------|---------|------|
| A     | 2.879 | 0.774      | 0.131   | 0.001|
| B     | 2.879 | 0.911      | 0.263   | 0.000|
| DRB1  | 2.879 | 0.898      | 0.699   | 0.001|

Allele frequencies

A total of 18 HLA-A, 31 HLA-B, and 13 HLA-DRB1 alleles were found in Guizhou population. In the HLA-A locus, A*11 was the most frequent allele in the present study with a frequency of 30.72%, followed by A*02(30.65%), A*24(17.07%), and A*33(7.43%). In the HLA-B locus, B*40 and B*46 were ranked as the first two frequent alleles with the same frequency of 16.27%, followed by B*15 (13.89%), B*13(9.66%), B*51(6.34%) and B*58(6.32%). In the HLA-DRB1 locus, DRB1*09 was the most common one (15.91%), followed by DRB1*15(13.51%), DRB1*12 (13.06%), DRB1*04 (10.44%) and DRB1*14 (9.34%). In addition, some HLA alleles were found to be very rare in the Guizhou population. For example, A*25(0.02%) and A*36(0.02%) in the HLA-A locus, and B*53(0.02%) and B*59(0.02%) in the HLA-B locus. Besides, some HLA alleles were not detected at all, such as A*43, B*82 and B*83. The frequencies of HLA-A, -B, and -DRB1 alleles are described in Table 2. The HLA allele distribution (Table 3) showed that the majority of Guizhou population harbored the most common alleles. There were three alleles in the HLA-A locus (over 10%) with a cumulative frequency of 78.44%, three alleles in the HLA-B locus with a cumulative frequency of 46.43%, and four alleles in the HLA-DRB1 locus with a cumulative frequency of 52.92%. Overall, the alleles with frequencies more than 1% in the HLA-A, -B, and -DRB1 loci made up 90% of the total population.

Haplotype frequencies and linkage disequilibrium (LD)

The haplotypes of HLA-A*-B, and -DRB1 occurring at frequency over 1% are sum-
Table 3 Distribution of HLA-A, -B, and -DRB1 genes

| Frequency range (%) | HLA-A | HLA-B | HLA-DRB1 |
|---------------------|-------|-------|----------|
| Total number        | 18    | 31    | 13       |
| >10%                | 3     | 3     | 4        |
| 1%-10%              | 6     | 13    | 9        |
| 0.1%-0.1%           | 4     | 8     | 0        |
| <0.1%               | 5     | 7     | 0        |

Table 4 Common haplotypes in Guizhou population occurring at a frequency over 1%

| A-B-DRB1 | ML-HF (%) | A-B | ML-HF (%) | B-DRB1 | ML-HF (%) | A-DRB1 | ML-HF (%) |
|----------|-----------|-----|-----------|--------|-----------|--------|-----------|
| 02XX 46XX 09XX | 5.59 | 02XX 46XX | 11.73 | 46XX 09XX | 7.49 | 02XX 09XX | 8.08 |
| 33XX 50XX 03XX | 2.80 | 11XX 15XX | 6.89 | 15XX 12XX | 3.82 | 11XX 12XX | 6.07 |
| 30XX 13XX07XX | 2.23 | 11XX 40XX | 6.19 | 50XX 00XX | 3.47 | 11XX 15XX | 5.68 |
| 02XX 46XX 14XX | 2.10 | 33XX 50XX | 5.10 | 46XX 14XX | 2.94 | 11XX 04XX | 3.57 |
| 11XX 15XX 15XX | 1.58 | 24XX 40XX | 5.01 | 15XX 15XX | 2.93 | 02XX 12XX | 3.57 |
| 02XX 46XX 08XX | 1.47 | 11XX 13XX | 3.92 | 15XX 04XX | 2.68 | 02XX 15XX | 3.53 |
| 33XX 50XX 13XX | 1.40 | 02XX 40XX | 3.48 | 13XX 07XX | 2.85 | 02XX 04XX | 3.08 |
| 11XX 15XX 04XX | 1.35 | 02XX 15XX | 3.18 | 13XX 15XX | 2.76 | 33XX 03XX | 2.85 |
| 11XX 46XX 09XX | 1.27 | 11XX 46XX | 2.80 | 40XX 12XX | 2.51 | 02XX 08XX | 2.76 |
| 11XX 40XX 12XX | 1.22 | 30XX 13XX | 2.73 | 40XX 15XX | 2.45 | 02XX 14XX | 2.53 |
| 11XX 40XX 08XX | 1.16 | 24XX 15XX | 2.73 | 40XX 09XX | 2.30 | 24XX 09XX | 2.51 |
| 24XX 40XX 15XX | 1.00 | 02XX 38XX | 2.04 | 40XX 06XX | 2.16 | 11XX 08XX | 2.51 |
| 11XX 51XX | 1.98 | 46XX 06XX | 2.09 | 24XX 15XX | 2.48 |
| 02XX 13XX | 1.94 | 40XX 11XX | 1.93 | 24XX 12XX | 2.47 |
| 02XX 51XX | 1.67 | 50XX 13XX | 1.77 | 11XX 09XX | 2.46 |
| 11XX 55XX | 1.52 | 40XX 04XX | 1.67 | 11XX 14XX | 2.36 |
| 02XX 55XX | 1.40 | 15XX 09XX | 1.56 | 30XX 07XX | 2.30 |
| 24XX 51XX | 1.39 | 13XX 12XX | 1.54 | 24XX 14XX | 2.25 |
| 24XX 35XX | 1.17 | 51XX 09XX | 1.41 | 24XX 04XX | 2.19 |
| 24XX 46XX | 1.13 | 51XX 14XX | 1.27 | 02XX 16XX | 2.03 |
| 11XX 39XX | 1.10 | 40XX 16XX | 1.17 | 11XX 11XX | 1.80 |
| 24XX 54XX | 1.07 | 40XX 14XX | 1.16 | 02XX 11XX | 1.73 |
| 46XX 12XX | 1.07 | 11XX 13XX | 1.66 |
| 33XX 13XX | 1.66 |
| 11XX 16XX | 1.53 |
| 24XX 11XX | 1.24 |
| 24XX 08XX | 1.20 |

ML-HF: Maximum-likelihood haplotype frequencies.

Linkage disequilibrium

The results of linkage disequilibrium (LD) test between two pairs loci are summarized in Table 5-7 ranked by the LD parameter, $r^2$ value. Some strong LD haplotypes were detected between two loci, including the common haplotypes and the rare haplotypes. For example, in the A-B haplotype, the haplotype with the strongest LD were A*33-B*58 with a frequency of 5.1%, while the haplotype with the second strongest LD was a rare haplotype (A*01-B*37) only with a frequency of 0.68%. In the B-DRB1 haplotype, the first two strongest LD haplotypes were common haplotypes including A*30-DRB1*07 (2.30%) and A*33-DRB1*03 (2.85%). However, the third strongest one was a rare haplotype A*29-DRB1*10 with a frequency of 0.38%. While in the B-DRB1 haplotype, B*37-
Table 5 The relative strongest linkage equilibrium between HLA-A and -B

| Haplotype | Observed frequencies (%) | Expected frequencies (%) | D | D^2 | r^2 | chi-square |
|-----------|--------------------------|--------------------------|---|-----|-----|------------|
| A*33-B*58 | 5.10                     | 0.47                     | 0.046 | 0.794 | 0.530 | 3043.778   |
| A*01-B*37 | 0.68                     | 0.02                     | 0.007 | 0.760 | 0.256 | 1468.645   |
| A*29-B*07 | 0.64                     | 0.02                     | 0.006 | 0.657 | 0.217 | 1246.909   |
| A*30-B*13 | 2.73                     | 0.32                     | 0.024 | 0.823 | 0.214 | 1230.311   |
| A*02-B*46 | 11.73                    | 4.99                     | 0.075 | 0.660 | 0.192 | 1101.331   |
| A*01-B*57 | 0.51                     | 0.02                     | 0.005 | 0.637 | 0.158 | 910.244    |
| A*23-B*49 | 0.05                     | 0                       | 0     | 0.499 | 0.136 | 780.135    |
| A*23-B*45 | 0.17                     | 0                       | 0     | 0.499 | 0.045 | 259.864    |
| A*24-B*40 | 5.01                     | 2.77                     | 0.027 | 0.198 | 0.037 | 213.080    |
| A*23-B*41 | 0.02                     | 0                       | 0     | 0.332 | 0.030 | 172.608    |
| A*11-B*15 | 6.89                     | 4.26                     | 0.027 | 0.280 | 0.029 | 164.227    |
| A*32-B*44 | 0.16                     | 0.55                     | 0.002 | 0.300 | 0.027 | 156.994    |
| A*02-B*38 | 0.23                     | 1.93                     | 0.017 | -0.092 | 0.024 | 133.708    |
| A*11-B*46 | 2.80                     | 5.00                     | -0.026 | -0.513 | 0.023 | 130.330    |
| A*36-B*57 | 0.02                     | 0                       | 0     | 1.000 | 0.022 | 47.006     |
| A*33-B*44 | 0.66                     | 0.13                     | 0.005 | 0.304 | 0.021 | 122.448    |
| A*02-B*38 | 2.04                     | 1.01                     | 0.011 | 0.504 | 0.020 | 112.472    |
| A*03-B*44 | 0.29                     | 0.03                     | 0.002 | 0.146 | 0.020 | 116.708    |

Table 6 The relative strongest linkage equilibrium between HLA-A and -DRB1

| Haplotype | Observed frequencies (%) | Expected frequencies (%) | D | D^2 | r^2 | chi-square |
|-----------|--------------------------|--------------------------|---|-----|-----|------------|
| A*30-DRB1*B07 | 2.30                 | 0.17                     | 0 | 0.697 | 0.303 | 1739.976   |
| A*33-DRB1*B03 | 2.65                 | 0.37                     | 0.025 | 0.337 | 0.190 | 1094.590   |
| A*29-DRB1*B10 | 0.38                 | 0.01                     | 0.040 | 0.504 | 0.118 | 675.721    |
| A*01-DRB1*B10 | 0.58                 | 0.03                     | 0.006 | 0.357 | 0.102 | 581.727    |
| A*02-DRB1*B09 | 8.08                 | 4.87                     | 0.038 | 0.349 | 0.052 | 298.950    |
| A*33-DRB1*B13 | 1.67                 | 0.42                     | 0.013 | 0.537 | 0.049 | 279.805    |
| A*11-DRB1*B12 | 6.07                 | 4.00                     | 0.025 | 0.276 | 0.026 | 147.557    |
| A*02-DRB1*B03 | 0.36                 | 1.54                     | 0.012 | -0.008 | 0.015 | 83.747     |
| A*11-DRB1*B15 | 5.68                 | 4.16                     | 0.018 | 0.191 | 0.013 | 74.271     |
| A*11-DRB1*B07 | 0.38                 | 1.58                     | 0.012 | -0.736 | 0.013 | 74.900     |
| A*02-DRB1*B13 | 0.56                 | 1.74                     | -0.010 | -0.000 | 0.010 | 54.862     |

Table 7 The relative strongest linkage equilibrium between HLA-B and -DRB1

| Haplotype | Observed frequencies (%) | Expected frequencies (%) | D | D^2 | r^2 | chi-square |
|-----------|--------------------------|--------------------------|---|-----|-----|------------|
| B*37-DRB1*B10 | 0.76                 | 0.01                     | 0.010 | 0.860 | 0.410 | 2342.480   |
| B*58-DRB1*B03 | 3.46                 | 0.32                     | 0.030 | 0.670 | 0.360 | 2040.560   |
| B*46-DRB1*B09 | 7.39                 | 2.59                     | 0.050 | 0.400 | 0.150 | 877.110    |
| B*13-DRB1*B07 | 2.85                 | 0.50                     | 0.020 | 0.520 | 0.140 | 791.810    |
| B*07-DRB1*B01 | 0.04                 | 0.05                     | 0.010 | 0.390 | 0.130 | 766.340    |
| B*06-DRB1*B03 | 0.76                 | 0.05                     | 0.010 | 0.540 | 0.120 | 696.090    |
| B*57-DRB1*B07 | 0.53                 | 0.04                     | 0 | 0.670 | 0.070 | 377.250    |
| B*58-DRB1*B13 | 1.77                 | 0.36                     | 0.010 | 0.260 | 0.060 | 356.830    |
| B*15-DRB1*B12 | 3.82                 | 1.81                     | 0.020 | 0.200 | 0.040 | 207.150    |
| B*07-DRB1*B10 | 0.45                 | 0.04                     | 0 | 0.250 | 0.040 | 234.870    |
| B*15-DRB1*B04 | 2.88                 | 1.45                     | 0.020 | 0.190 | 0.030 | 154.070    |
| B*13-DRB1*B15 | 2.76                 | 1.31                     | 0.020 | 0.190 | 0.030 | 146.350    |

DRB1*10 with a frequency of 0.76% was ranked as the first strongest LD haplotype, followed by three common haplotypes, which were B*58-DRB1*B03 (3.46%), B*46-DRB1*B09 (7.49%) and B*13-DRB1*B07 (2.85%).

Comparison of the frequent alleles between Guizhou province and other populations

The first three frequent alleles in the HLA-A,-B, and -DRB1 loci in Guizhou province and other populations were obtained from the previous studies: the provinces of Shanxi[13], Henan[14], Jiangsu[15], Hunan[16], and Hainan[17]. As it can be seen from Table 8, A*11, A*02, and A*24 were more frequent in Guizhou than in other provinces in China, regardless of whether the
population was the northern Chinese or southern Chinese. A*11 was more frequent in southern Chinese than the northern Chinese and A*02 was less frequent in Hunan and Hainan provinces than in the northern Chinese. In the HLA-B locus, B*40 and B*46 were more frequent in Guizhou than in other provinces in China. In the HLA-DRB1 locus, DRB1*12 in Guizhou was the most frequent allele among the six provinces, and the frequency of DRB1*15 was between that of the northern Chinese and southern Chinese.

Table 2 The first three frequent alleles in HLA-A, B, and -DRB1 loci in Guizhou province and other populations in China (%)

| Allel | Guizhou | Shanxi | Henan | Jiangsu | Hunan | Hainan |
|-------|---------|--------|-------|---------|-------|--------|
| A*11  | 30.72   | 17.93  | 16.66 | 18.56   | 20.88 | 18.35  |
| A*02  | 30.65   | 29.26  | 29.09 | 29.55   | 19.64 | 15.45  |
| A*24  | 17.07   | 16.12  | 15.70 | 16.79   | 12.01 | 11.03  |
| B*40  | 16.27   | 12.98  | 13.15 | 13.58   | 13.31 | 10.78  |
| B*46  | 16.27   | 7.15   | 6.24  | 9.41    | 11.86 | 7.55   |
| B*15  | 13.89   | 13.84  | 13.70 | 14.40   | 9.54  | 8.55   |
| DRB1*09 | 15.91   | 12.85  | 13.02 | 16.15   | 13.01 | 7.56   |
| DRB1*15 | 13.51   | 16.77  | 17.86 | 13.09   | 9.55  | 9.50   |
| DRB1*12 | 13.06   | 10.67  | 9.95  | 12.22   | 8.89  | 8.25   |

DISCUSSION

Guizhou province is in the southwest of China. One would expect that Guizhou presents some of the characteristics of the Southern Chinese population. The common alleles in the Chinese population, such as A*02, A*11, A*24, A*33, B*40, B*58, B*15, B*46, DRB1*09, DRB1*15, DRB1*12, and DRB1*04, are also found frequently in Guizhou population. Meanwhile, it was shown that inhabitants of Guizhou province exhibit some differences from those of the other provinces of China in some alleles, especially differences from some minority groups, such as Hui, Wa, and Drung. For example, in locus A, A*02 (30.65%) and A*11 (30.72%) were the first two most frequent alleles in Guizhou province, and there were nearly no differences in the frequencies of two alleles. In Chengdu population, A*11 was the most frequent allele with a frequency of 31.50%, and A*02 was the second most frequent allele with a frequency of 31.03%, which is nearly the same as that reported for Guizhou province. The frequency of A*02 in Guizhou province is consistent with that of Southern Chinese, such as that in Jiangsu (A*02, 29.55%) and Shanghai (A*02, 31.34%). The first four most frequent alleles in Guizhou were in the order of A*11, A*02, A*24 and A*33, and this order is the same as that of Chengdu. In contrast, the first four most frequent alleles in the A locus in Jiangsu and Shanghai are all in the order of A*02, A*11, A*24 and A*33. The first four most frequent alleles in Yunnan province for Han Chinese are in the order of A*24, A*02, A*11 and A*33, which differ more from those in Guizhou population than Jiangsu and Shanghai, although Yunnan province is adjacent to Guizhou province. In the B locus, B*46 (16.27%) is the most common allele in Guizhou population, which is also the most common one in Chengdu (16.3%) and Yunnan (17.9%). While in Jiangsu and Shanghai, B*15 is the most common one. The first four most frequent alleles in the B locus in Guizhou are in the order of B*40, B*46, B*15 and B*13, for Han Chinese in Yunnan are B*46, B*15, B*40 and B*13, in Jiangsu are B*15, B*40, B*13 and B*46. The four alleles are all the same but in different order. In the DRB1 locus, DRB1*09 (15.91%) is the most predominant allele, which is also the most predominant one in Jiangsu, Shanghai and Chengdu and the first four most frequent alleles in DRB1 locus in Guizhou are in the same order as that in Chengdu, with the same order of DRB1*09, DRB1*12, DRB1*15, and DRB1*04. By comparison of the frequencies of the HLA alleles in several provinces, Guizhou is more consistent with Chengdu than other provinces.

HLA haplotypes of two and/or three loci have been reported in various worldwide populations, which show great difference from the population of Guizhou in haplotype distribution. The most common haplotype in Guizhou is A*02-B*46-DRB1*09, which is the same as that in Chengdu. While in Jiangsu, Shanxi province and Shanghai, the most common haplotype is A*30-B*13-DRB1*07. While in Yunnan province, A*24-B*15-DRB1*15 is the most common haplotype with a frequency of 4.1%, followed by A*24-B*46-DRB1*08 (3.5%) and A*24-B*15-DRB1*12 (2.9%). However, these three haplotypes are relatively rare in Guizhou province.

The distribution of the HLA-A, -B, and -DRB1 alleles showed that the majority of Guizhou population harbor the common alleles, which means that most patients for homologous stem cell transplantation (HSCT) would readily find HLA-matched donors in CMDP Guizhou registry if they carry those common alleles.

HLA haplotype estimate is a valuable tool in the management of donor registries. It has been used to project how many donors would be needed to achieve a certain probability of finding an HLA-matched donor. Another useful application of the haplotype frequency is to predict the probability that a donor typed at low or intermediate-resolution would match a specific patient at high resolution. Besides, HLA
haplotype provides valuable information in tracing the source of historical genetic inputs.

In summary, the present study reported HLA-A, -B, and -DRB1 allele frequencies and haplotype frequencies in Guizhou population. The results would be useful as baseline data for donor selection of hematopoietic stem cell or solid organ transplantation, anthropology studies and HLA disease association analysis.

Acknowledgements
We are grateful to each volunteer who has registered in the CMDP to offer a stranger a second chance at life.

References
[1] Hughes AL, Nei M. Maintenance of MHC polymorphism. Nature 1992;355:402.
[2] Robinson J, Waller MJ, Parham P. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Res 2003;31:311.
[3] Robinson J, Waller MJ, Parham P, de Groot N, Bonnert R, Kennedy LJ, et al. World Marrow Donor Association guidelines for use of HLA Nomenclature and its validation in the data exchange among hematopoietic stem cell donor registries and cord blood banks. Bone Marrow Transplant 2007;39:737.
[4] Gourraud PA, Lamiraux P, El-Kadhi N, Raffoux C, Cambon-Thomsen A. Inferred HLA haplotype information for donors from hematopoietic stem cells donor registries. Hum Immunol 2005;66:563-70.
[5] Hurley CK, Maiers M, Ng J, Wagage D, Hegland J, Baisch J, et al. Large-scale DNA-based typing of HLA-A and HLA-B at low resolution is highly accurate, specific, and reliable. Tissue Antigens 2000;55:352.
[6] Middleton D, Williams F, Hamill MA, Meenagh A. Frequency of HLA-B alleles in a caucasoid population determined by a two-stage PCR-SSOP typing strategy. Hum Immunol 2006;61:1285-97.
[7] Radford NF, Geoffrey H. A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Michael Irwin Jordan, Editor. Learning in Graphical Models. Toronto: Kluwer Academic Publishers,1999;355-68.
[8] Trevor H, Robert T, Friedman JH. Linear methods for regression. In: Trevor H, Robert T, Friedman JH, Editors. The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd edition. California: Springer 2008;43-92.
[9] Excoffier L, Laval G, Schneider S, Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 2007;1:47-50.
[10] Guo S, Thompson E. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48:361-72.
[11] Edwards A W, G.H. Hardy(1908) and Hardy-Weinberg equilibrium. Genetics 2008;179:1143-50.
[12] Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:887-93.
[13] Wu QJ, Liu ML, Qi J, Liu S, Zhang Y, Wei XQ. Gene and haplotype frequencies for the loci HLA-A B and DRB1 in 1755 North Chinese Han bone marrow registry donors. Zhongguo Shi Yan Xue Ye Xue Za Zhi (in Chinese) 2007;15:357-63.
[14] Yang R, Zhang BW, Cheng SG, Du J, Ma R. Gene frequency and haplotype of HLA-A, B, DRB 1 in donors from Henan branch of China marrow donor program. J Zhengzhou Univ (Medical Sciences) (in Chinese) 2007;42:930-2.
[15] Miao KR, Pan QQ, Tang RC, Zhou XP, Fan S, Wang XY, et al. The polymorphism and haplotype analysis of HLA-A, -B and -DRB1 genes of population in Jiangsu province of China. Int J Immunogenet 2007;34:419-24.
[16] Xie YB, Wang CL, LS Xie YX. Study on polymorphism of HLA-A, B and DRB1 alleles in Hunan population. Prac Prev Med (in Chinese) 2006;1:41-11.
[17] Tang QP, Yang XP, Zhao JL, Cao YX, Li Q, Liu YH, et al. Study on the application of PCR-SSP technique in the HLA-A, B, DRB1 gene polymorphism in Han people of Hainan. Mod Prev Med (in Chinese) 2008;35:4020-2.
[18] Hong W, Chen S, Shao H, Fu Y, Hu Z, Xu A. HLA class I polymorphism in Mongolian and Hui ethnic groups from Northern China. Hum Immunol 2007;68:439-48.
[19] Shi L, Ogata S, Yu JK, Ohashi J, Yu L, Shi L, et al. Distribution of HLA alleles and haplotypes in Jinuao and Wa populations in Southwest China. Hum Immunol 2008;58:65.
[20] Chen S, Fu Q, Liu Z, Fu Y, Lin J, Tao H, et al. The distribution of HLA alleles revealed a founder effect in the geographically isolated Chinese population, Drung. Mol Immunol 2007;44:2017-22.
[21] Chen Q, Zou H, Xu XH. HLA-A, -B and -DRB1 allele and haplotype frequencies in a Chinese population from southwestern China. Hum Immunol 2005;66(Suppl 1):S45.
[22] Du KM, Ji Y, Xie JH, Fu M, Sun Y, Jin Y, et al. HLA-A,-B,-DR haplotype frequencies from DNA typing data of 26,266 Chinese bone marrow donors. Hum Immunol 2007;68:454-66.
[23] Shi L, Xu SB, Ohashi J, Sun H, Yu JK, Huang XQ, et al. HLA-A, HLA-B, and HLA-DRB1 alleles and haplotypes in Naxi and Han populations in southwestern China (Yunnan province). Tissue Antigens 2006;67:38-44.
[24] Cao K, Hollenbach J, Shi X, Shi W, Chopik M, Fernandez-Vina MA. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. Hum Immunol 2001;62:1009.
[25] Williams F, Meenagh A, Darke C, Acosta A, Daar AS, Gorodezky C, et al. Analysis of the distribution of HLA-B alleles in populations from five continents. Hum Immunol 2001;62:645.