Nine polymorphic STR loci in the HLA region in the Shaanxi Han population of China

H.B. Zhang[1,2], S.G. Wei[1,2], B. Yu[1,2], L. Li[1,2] and J.H. Lai[1,2]

[1]Key Laboratory of the National Ministry of Health for Forensic Sciences,
School of Medicine, Xi'an Jiaotong University, Xi'an, Shaanxi, P.R. China
[2]Key Laboratory of Environment and Gene Related Diseases,
Ministry of Education, School of Medicine, Xi'an Jiaotong University,
Xi'an, Shaanxi, P.R. China

Corresponding author: J.H. Lai
E-mail: laijh1011@mail.xjtu.edu.cn

ABSTRACT. A large number of microsatellite genetic markers have been identified in the human leukocyte antigen (HLA) region. We investigated genetic polymorphism of the nine short tandem repeat (STR) loci (D6S276, MOGCA, D6S265, MIB, D6S273, G51152, TAP1CA, RING3CA, and D6S291) in the HLA region in the Shaanxi Han population. Using a fluorescence-labeled multiplex-PCR STR typing method, 6-13 alleles were detected in these nine STR loci in 150 unrelated Han Chinese from the region of Shaanxi, China. The distributions of the genotypes at these nine loci were in Hardy-Weinberg equilibrium. We conclude that these nine STR loci have a high level of genetic polymorphism; they would be useful for population genetic studies, pre-transplantation HLA typing, forensic and paternity testing, etc.

Key words: HLA; STR; Polymorphisms; Han population genetics

INTRODUCTION

Approximately every 2000 bases in the human genome contain a microsatellite, also called short tandem repeats (STRs) or simple sequence repeats. These STRs are widely distributed throughout the human genome and abundant in various functional regions and non-functional regions of the genome (Kashi and King, 2006). Because of their high polymorphism and comprehensive distribution in the genome, STRs have already become significant genetic markers for gene mapping and disease association analysis (Foissac et al., 2000; Cullen et al., 2003). Due to the ir extensive genetic polymorphism and heterozygosity and Mendelian codominant inheritance, STRs have also become a useful tool in forensic identification of individuals and paternity testing (Urquhart et al., 1994; Schoske et al., 2003). As one of the most intensely active gene regions in the human genome, human leukocyte antigen (HLA) is a hot spot in disease association studies. The HLA region has a full length of 3.6 Mb, including a large number of STR loci, most of which are (CA)n repeats (Fiorentino et al., 1997). As obvious characteristics of linkage disequilibrium exist in the HLA region, it was deemed important to explore the polymorphism of the short tandem repeats in the HLA region for disease correlation studies. The present study revealed that these STRs loci had high heterozygosity with highly polymorphism information content. These STRs could be applied to predict the supplement of using the specific HLA haplotypes to screen the HLA matching donor and in the forensic identification of individuals (Cullen et al., 2003; Gourraud et al., 2004).

Microsatellite typing is rapid, accurate and reproducible, providing additional incentive for their use as polymorphic markers. In this study, 9 STRs loci were evenly distributed in the human HLA region, where D6S276, MOGCA, D6S265, and MIB were located in the HLA I region, G51152, TAP1CA, RING3CA, and D6S291 in the HLA II region, and D6S273 in the HLA III region. All 9 STRs loci were the polymorphic repeat of the dinucleotide (CA)n and the length of the amplified fragments was within 63-295 base pairs. Fluorescent labeling was performed at the 5' end of the upper primer by 6-FAM or TET, and typing was carried out by using the ABI 377 DNA sequencer (Applied Biosystems). The polymorphism of 9 STR genetic markers had not been previously reported in the Han population in China. There is very little information about their polymorphism for Chinese populations.

MATERIAL AND METHODS

Sampling

After obtaining informed consent, peripheral blood samples were collected from 150 healthy Han individuals from Shaanxi Province of China. Genomic DNA was extracted from EDTA-treated peripheral blood samples by the Chelex-100 method (Walsh et al., 1991).

PCR and genotyping

Primers used for STR genotypings are published elsewhere (Cullen et al., 2003), and specific for the amplification of chromosome nine STR markers. D6S276, D6S273, G51152, TAP1CA, RING3CA, and D6S291 of the upper primers of the polymorphic markers were la-

beled with 6-FAM, and TET was used to label MOGCA, D6S265 and MIB. All primers were prepared by Shanghai Sangon Biological Engineering Technology & Service Co., Ltd. The PCR amplifications were carried out as follows: 30 cycles for 15 s at 94°C, 15 s at 55°C and 30 s at 72°C, and a final extension for 30 min at 72°C. The PCR products were electrophoresed on an ABI 377 DNA Sequencer (Applied Biosystems). The data were collected and analyzed by the GenScan 3.1 software. GeneScan and Genotyper DNA fragment analysis software were used to interpret the data. The output and molecular weight of each product were shown by curve peak and then processed, and the genotyping data were transferred into the allele code. According to the genotyping results, homozygous of each locus were selected for sequencing. Using an ABI 3730 Genetic Analyzer, the reverse unlabeled fluorescence primers were sequenced to determine the size and composition of the fragments. Based on the principle that the allele of the STR locus was named according to the repeat times of its core unit proposed by the DNA Commission of the International Society of Forensic Haemogenetics, standardized naming was performed (Gill et al., 1997; Olaisen et al., 1998).

Statistical analysis

The allele frequency and genotype frequency of the unrelated individual samples were calculated by the direct-counting method. The Arlequin3.01 software was used for the Hardy-Weinberg equilibrium test (Guo and Thompson, 1992). The PowerStats 12 software was used to calculate the application indicators of genetics, including heterozygosity, matching probability, power of discrimination, polymorphic information content, and power of exclusion.

RESULTS

In the 150 unrelated individual samples, a total of 95 alleles were detected in 9 loci with the frequency of 0.003-0.483 (Table 1). The highest polymorphism was found in locus MIB with 13 alleles and 43 kinds of genotypes, and the lowest was in locus RING3CA with 6 kinds of alleles and 15 kinds of genotypes in the nine STRs loci. Using the Arlequin 3.01 software, the observed heterozygosity and expected heterozygosity of each locus were calculated. Exact tests of Hardy-Weinberg equilibrium were performed for the nine STR loci. Based on the exact test results for the distribution of genotype frequency of the loci, all loci were in Hardy-Weinberg equilibrium.

DISCUSSION

Among the 9 STR loci, D6S276, MOGCA, D6S265, and MIB were located in the HLA I region, G51152, TAP1CA, RING3CA, and D6S291 in the HLA II region, and D6S273 in the HLA III region. For the 9 loci, the heterozygosity was greater than 0.6 in all cases, and the highest was found in locus D6S273, which was 0.8533. The polymorphic information content was higher than 0.65 in all cases, and the highest, 0.8356, was in locus MOGCA. The power of discrimination was consistently higher than 0.8, with the highest being 0.955 in MOGCA. The probability of paternity exclusion was in all cases above 0.5 and locus D6S273 had the greatest, namely 0.701. These indices are usually applicable in forensic science and population genetic studies. The 9 loci had high genetic polymorphism, heterozygosity, and

Table 1. Distribution of observed allele frequencies of nine short tandem repeat loci located on chromosome 6 (N = 150).

Alleles	MOGCA	D6S265	MIB	D6S273	D6S291	TAP1CA	RING3CA	D6S276	G51152
5						0.0067			
7								0.3667	
8						0.0033	0.1833		0.0200
9	0.0300				0.0033	0.4333	0.1700		0.4833
10			0.3033		0.0167	0.3533	0.4033		0.0800
11		0.0300			0.3267	0.0033	0.1967	0.0500	0.0167
12		0.0033			0.2900		0.0433	0.0333	0.0367
13	0.1500	0.2767	0.0433	0.0033	0.0600			0.1500	0.0967
14	0.2700	0.2233	0.0067	0.0100	0.0167	0.0700	0.0033	0.0067	0.0433
15	0.0933	0.2667	0.1300	0.0833	0.1367	0.0200			0.1800
16	0.1433	0.0567	0.0033	0.1267	0.1400	0.0400		0.0633	0.0033
17	0.0767	0.0767	0.0233	0.1433	0.0067	0.0700		0.0333	
18	0.0067	0.0033	0.0833	0.2067	0.0033			0.0600	
19	0.0033	0.0033	0.0833	0.3367				0.2200	
20	0.0800	0.0567	0.1467	0.0800				0.0133	0.0067
21	0.0033	0.0000	0.1000	0.0067					
22	0.1367	0.0033	0.0633					0.0033	
23	0.0067		0.0100						
24		0.0033		0.0033					0.0300
25									0.0033
Matching probability	0.0454	0.0798	0.0466	0.0814	0.0939	0.1967	0.117	0.0869	0.1102
Power of discrimination	0.9546	0.9202	0.9534	0.9186	0.9061	0.8033	0.883	0.9131	0.8898
Polymorphism index system	0.8256	0.7583	0.8226	0.7668	0.7315	0.6211	0.6935	0.7556	0.687
Power of exclusion	0.5505	0.5041	0.5625	0.7014	0.5867	0.5155	0.4817	0.6494	0.4283
Observed heterozygosity	0.7733	0.7533	0.78	0.8533	0.7933	0.6667	0.7333	0.8267	0.7000
Expected heterozygosity	0.848	0.7919	0.8419	0.7965	0.7693	0.6807	0.7367	0.7847	0.717

broad distribution of alleles in the Shaanxi Han population, and could have strong amplification efficiency combined. The application of the 9 loci may be more significant in the population genetics, human origin and evolution studies and forensic identification of individuals or paternity tests.

Preimplantation genetic diagnosis (PGD) is a new kind of method to avoid transferring genetic diseases to the newborn. At present, the significance of PGD has been considered for several indications beyond its original purpose, and it may not only detect fetal diseases but also select for non-disease characteristics, such as in HLA genotyping and immune compatibility before hematopoietic stem cell transplantation. The study on highly polymorphic STR genetic markers in the HLA region is an elegant method for linkage analysis of genes associated with the HLA region and disease correlation research. Studying the short tandem repeat sequences within pathogenic genes and analyzing the genetic linkage within a family has been useful for detecting the disease linked with the microsatellite markers. The STR with high polymorphism in the HLA region could be used indirectly in HLA typing in PGD and it has already become a kind of effective tool (Rechitsky et al., 2004; Fiorentino et al., 2005). To improve the accuracy of HLA typing in PGD, it has been necessary to strengthen the analysis of STRs of the HLA region, especially the polymorphic STR loci with high heterozygosity and power of discrimination. In this study, the genetic markers were all highly polymorphic in the Han population of China and they were distributed in the HLA I, HLA II and HLA III regions. These loci could be used in HLA typing in PGD in the Han population of China to find a potential donor. Although the tying of polymorphic STR loci with high polymorphism cannot replace HLA matching for organ transplantation, it could supplement the low resolution of HLA typing before the transplantation of bone marrow or hematopoietic stem cells

(Korzebor et al., 2007; Malkki et al., 2007; Sens-Abuazar et al., 2009). The reliable and highly polymorphic STR typing of the HLA region would provide an effective method for pretransplant HLA typing.

ACKNOWLEDGMENTS

REFERENCES

Cullen M, Malasky M, Harding A and Carrington M (2003). High-density map of short tandem repeats across the human major histocompatibility complex. *Immunogenetics* 54: 900-910.

Fiorentino F, Kahraman S, Karadayi H, Biricik A, et al. (2005). Short tandem repeats haplotyping of the HLA region in preimplantation HLA matching. *Eur. J. Hum. Genet.* 13: 953-958.

Foissac A, Crouau-Roy B, Faure S, Thomsen M, et al. (1997). Microsatellites in the HLA region: on overview. *Tissue Antigens* 49: 197-214.

Foissac A, Salhi M and Cambon-Thomsen A (2000). Microsatellites in the HLA region: 1999 update. *Tissue Antigens* 55: 477-509.

Gill P, Brinkmann B, d'Aloja E, Andersen J, et al. (1997). Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. *Forensic Sci. Int.* 87: 185-192.

Gourraud PA, Mano S, Barnetche T, Carrington M, et al. (2004). Integration of microsatellite characteristics in the MHC region: a literature and sequence based analysis. *Tissue Antigens* 64: 543-555.

Guo SW and Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.

Kashi Y and King DG (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 22: 253-259.

Korzebor A, Zamani M, Nouri K and Modarressi MH (2007). Statistical analysis of six STR loci located in MHC region in Iranian population for preimplantation genetic diagnosis. *Int. J. Immunogenet.* 34: 441-443.

Malkki M, Gooley T, Horowitz M and Petersdorf EW (2007). MHC class I, II, and III microsatellite marker matching and survival in unrelated donor hematopoietic cell transplantation. *Tissue Antigens* 69 (Suppl 1): 46-49.

Olaisen B, Bar W, Brinkmann B, Budowle B, et al. (1998). DNA recommendations 1997 of the international society for forensic genetics. *Vox Sang.* 74: 61-63.

Rechitsky S, Kuliev A, Tur-Kaspa I, Morris R, et al. (2004). Preimplantation genetic diagnosis with HLA matching. *Reprod. Biomed. Online* 9: 210-221.

Schoske R, Vallone PM, Ruitberg CM and Butler JM (2003). Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Anal. Bioanal. Chem.* 375: 333-343.

Sens-Abuazar C, Santos PS, Bicalho MG, Petzl-Erler ML, et al. (2009). MHC microsatellites in a Southern Brazilian population. *Int. J. Immunogenet.* 36: 269-274.

Urquhart A, Kimpton CP, Downes TJ and Gill P (1994). Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. Legal Med.* 107: 13-20.

Walsh PS, Metzger DA and Higuchi R (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotecchniqu*06-513.