HLA-A Gene Polymorphism Defined by High-Resolution Sequence-Based Typing in 161 Northern Chinese Han People

Chunxia Yan[1,2,3,4*], Ruilin Wang[3,4,5*], Jingxiang Li[3,5*], Yajun Deng[1,2,3,4*], Dongying Wu[3], Hongbo Zhang[1,2], Hongxing Zhang[6], Lidong Wang[3,4], Chunrong Zhang[5], Haiyan Sun[3], Xiuqing Zhang[3], Jian Wang[3], Huanming Yang[3], and Shengbin Li[1,2,3,4#]

[1] The State Laboratory of the Ministry of Health for Forensic Sciences, Xi'an 710061, China; [2] Forensic Department, Medical College of Xi'an Jiaotong University, Xi'an 710061, China; [3] Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 101300, China; [4] Center of Forensic Sciences, Beijing Genomics Institute, Beijing 101300, China; [5] BGI Life Tech. Co., Beijing Genomics Institute, Beijing 101300, China; [6] Technical Section of Xi'an People's Procuratorate, Xi'an 710002, China.

Human leukocyte antigen (HLA) system is the most polymorphic region known in the human genome. In the present study, we analyzed for the first time the HLA-A gene polymorphisms defined by the high-resolution typing methods—sequence-based typing (SBT) in 161 Northern Chinese Han people. A total of 74 different HLA-A gene types and 36 alleles were detected. The most frequent alleles were A*110101 (GF=0.2360), A*24020101 (GF=0.1646), and A*020101 (GF=0.1553); followed by A*3303 (GF=0.1180), A*3001 (GF=0.0590), and A*310102 (GF=0.0404). The frequencies of following alleles, A*0203, A*0205, A*0206, A*0207, A*030101, A*2423, A*2601, A*3201, and A*3301, are all higher than 0.0093. The homozygous alleles include A*020101, A*110101, A*24020101 and A*310102. Heterozygosity (H), polymorphism information content (PIC), discrimination power (DP) and probability of paternity exclusion (PPE) of HLA-A in the samples were calculated and their values were 0.8705, 0.8491, 0.6014, and 0.9475, respectively. These results by SBT analysis of HLA-A polymorphism in Northern Chinese Han population, especially the allele subtypes character, will be of great interest for clinical transplantation, disease-associated study and forensic identification. Implementation of high-resolution typing methods allows a significantly wider spectrum of HLA variation including rare alleles. This spectrum will further be extensively utilized in many fields.

Key words: HLA-A, polymorphism, sequence-based typing (SBT), Han population

Introduction

The major histocompatibility complex (MHC) is a multigene family that encodes cell-surface glycoprotein on antigen-presenting cells. Its function is to bind and present peptides to the immune system. The human MHC is known as human leukocyte antigen (HLA) system, which is located on the 6p21 (*1*). The length of the HLA complex is approximately 3.6 Mb (*1*). It contains three gene regions: Class I, Class II and Class III. Classes I and II genes have been among the most polymorphic systems known in the human genome by now. The Class I region is the most telomeric part of the HLA complex and contains three classical Class I gene loci, HLA-A, B, and C, as well as four non-classical Class I genes, HLA-E, F, G and H (*1–3*). More than 130 Classes I and II gene products have been detected by serologic techniques. In the past years, by the extensive use of the molecular biology methods, the number of the HLA-polymorphism has increased by 8 to 9 times. About 1,700 alleles in HLA Classes I and II region have been identified around the world. As of November 2003, HLA-A, B, C, E, F, and G of Class I have 303, 559, 150, 6, 2, and 15 alleles, respectively; and among them, HLA-A,

*** These authors contributed equally to this work.**
[#] Corresponding author.
E-mail: lishengbin@genomics.org.cn

B, and C have 24, 48, and 8 serogroups, respectively (http://www.ihwg.org). Much of the polymorphisms in HLA Classes I and II exons cannot be detected by serologic typing methods. Only molecular typing methods based on PCR (polymerase chain reaction) can accurately distinguish many allelic sequence variants at all these loci (*3–5*).

PCR-based HLA typing methods include: DNA amplification with sequence specific primers (SSP; ref. *3*, *4*), DNA amplification followed by hybridization with sequence specific oligonucleotide probes (SSOP; ref. *4*, *5*), sequence-based typing (SBT; ref. *6–8*) and DNA chip technology (*9*, *10*). Now, SBT is of more importance as the number of HLA alleles is increasing, and it is a high-resolution and high-throughput typing method compared with SSOP or SSP. Meanwhile, it is more practicable at present than DNA chip technology, which uses thousands of probes attached to silicon chips.

Knowledge of HLA polymorphism in different ethnic populations and determination of HLA alleles have been extensively utilized in tissue typing for organ transplantation, popular genetics, disease association, forensic paternity determination and other fields (*1, 3, 4, 5, 9*). The aim of this study is to investigate the A locus polymorphism in HLA Class I region and the wider spectrum of HLA-A variation in Northern Chinese Han population by the high-resolution typing methods—SBT.

Results

HLA-A gene polymorphism detected by SBT and their serological specificities in Northern Chinese Han population

The genotypes of HLA-A were examined in 161 Northern Chinese Han individuals by SBT and 74 gene types were identified in this population. By direct counting method, 36 alleles were detected in the gene pool, and consequently 14 serological specificities were defined (Table 1). The most frequently detected allele was A*110101 and A11 is the most common antigen type (Figure 1).

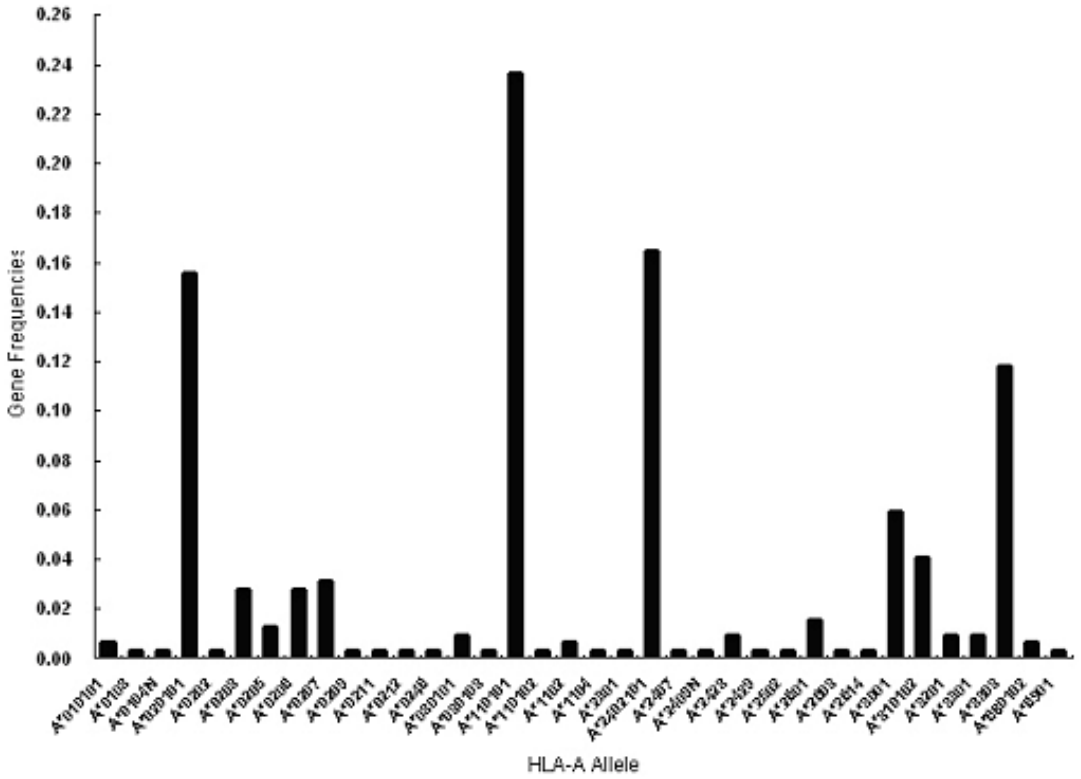

Fig. 1 HLA-A gene polymorphism detected by the high-resolution typing SBT in Northern Chinese Han population.

Table 1 HLA-A Alleles and Frequencies Detected by SBT and Serological Specificities in Northern Chinese Han Population

HLA-A allele	Gene frequencies (N=2×161)	Serological specificities	Antigen frequencies (N=2×161)
A*010101	0.0062	A1	0.0124
A*0103	0.0031		
A*0104N	0.0031		
A*020101	0.1553	A2	0.2702
A*0202	0.0031		
A*0203	0.0280		
A*0205	0.0124		
A*0206	0.0280		
A*0207	0.0311		
A*0209	0.0031		
A*0211	0.0031		
A*0212	0.0031		
A*0246	0.0031		
A*030101	0.0093	A3	0.0124
A*030103	0.0031		
A*110101	0.2360	A11	0.2484
A*110102	0.0031		
A*1102	0.0062		
A*1104	0.0031		
A*2301	0.0031	A23	0.0031
A*24020101	0.1646	A24	0.1832
A*2407	0.0031		
A*2409N	0.0031		
A*2423	0.0093		
A*2429	0.0031		
A*2502	0.0031	A25	0.0031
A*2601	0.0156	A26	0.0217
A*2603	0.0031		
A*2614	0.0031		
A*3001	0.0590	A30	0.0594
A*310102	0.0404	A31	0.0404
A*3201	0.0093	A32	0.0093
A*3301	0.0093	A33	0.1273
A*3303	0.1180		
A*680102	0.0062	A68	0.0062
A*6901	0.0031	A69	0.0031
Total: 36	**1.000**	**14**	**1.0002**

IHWG reference cell panel

The reference cell panel used in our research was supplied by International Histocompatibility Working Group (IHWG; http://www.ihwg.org). The SBT panel is a panel of well-characterized cells used as positive primer controls in HLA sequence-based typing. By using the primer in our experiment, the genotypic results of the IHWG SBT panel (Table 2) exactly matched the IHWG standard types.

Hardy-Weiburg equilibrium test

The exact test showed that Northern Chinese Han population is in the Hardy-Weinberg equilibrium for HLA-A gene ($P>0.05$).

Table 2 The Genotypes of HLA-A SBT Reference Cell Panel

IHW No.	Local designation	Consan-guineous	Locus A	IHW No.	Local designation	Consan-guineous	Locus A
IHW09016	RML	Y	A*0204	IHW09367	LCK	N	A*0203/1102
IHW09021	RSH	N	A*3001/6802	IHW09368	280599	N	A*2604/2402101
IHW09024	KY17	N	A*02206/1101	IHW09369	ISH3	Y	A*2402101(2402)
IHW09035	JBUSH	Y	A*3201	IHW09370	230699	N	A*2402101/0206
IHW09045	TUBO	N	A*0216/03011	IHW09373	FH1	N	A*0205/6802
IHW09056	KOSE	Y	A*02011	IHW09374	FH2	N	A*3402/02011
IHW09077	T7527	N	A*0207/02011	IHW09375	FH3	N	A*31012/3301
IHW09092	MB92	Y	A*2501	IHW09376	FH4	N	A*0101(01011)
IHW09103	KT14	N	A*2402/2602	IHW09380	FH6	N	A*2402101/2901
IHW09215	M7	N	A*0202/03011	IHW09381	FH7	N	A*0206/3002
IHW09220	XLI-ND	N	A*0210/3001	IHW09382	FH8	N	A*1101/3402
IHW09253	YHA1742	N	A*2403/3302	IHW09388	FH14	N	A*03011/11011
IHW09263	G085	N	A*0101/2901	IHW09389	FH15	N	A*0101/2402101
IHW09267	LEO23	N	A*2609/3201	IHW09390	FH16	N	A*0101/02011
IHW09273	LADA	N	A*0201/8001	IHW09391	FH17	N	A*0101/6801
IHW09364	GRC-212	N	A*0211/68012	IHW09397	DUG150	N	A*02011/68011

HLA polymorphism parameters in Northern Chinese Han population

Heterozygosity (H), polymorphism information content (PIC), discrimination power (DP) and probability of paternity exclusion (PPE) of HLA-A in Northern Chinese Han population were calculated with pertinent data. H, PIC, DP, and PPE values are 0.8705, 0.8491, 0.6014, and 0.9475, respectively.

Discussion

HLA-A belongs to the HLA Class I heavy chain. Its molecule is a heterodimer consisting of a heavy chain and a light chain (β-2 microglobulin). The heavy chain is anchored in the membrane. Class I molecules play a central role in the immune system by presenting peptides derived from the endoplasmic reticulum lumen. They are expressed in nearly all cells of human body. HLA-A gene contains eight exons (*1*, *3*, *5*). Exon 1 encodes the leader peptide; exons 2 and 3 encode the α1 and α2 domains, which both bind the peptide; exon 4 encodes the α3 domain; exon 5 encodes the transmembrane region, and exons 6 and 7 encode the cytoplasmic tail. Polymorphisms within exons 2 and 3 are responsible for the peptide binding specificity of each molecule in Class I. Some alleles need to sequence exon 4 variations. Typing for these polymorphisms is routinely done for clinical organ transplantation; and it is a useful DNA marker in immunogenetics and anthropology study. We analyzed HLA-A polymorphism in Han and Uygur by sequence specific oligonucleotide probe (SSOP) in the past (*5*). With the development of HLA typing technology, the SBT method results in the most possible level of resolution of HLA genotypes (*6–8*). The basic theory of SBT is to get the nucleotide sequence of the most polymorphic regions in the HLA genes and compare the sequence data to those of all possible theoretical allele combinations by software analysis (MatchTool and MT Navigator), which might ensure a highly accurate identification of the actual genotypes of samples (*11*). Meanwhile, it will permit the direct identification of new alleles by sequencing HLA conserved and polymorphic regions.

This study examined the HLA-A alleles in Northern Chinese Han population at SBT resolution for the first time. The number of alleles detected at each locus indicates the existence of high polymorphism in Northern Chinese Han population. Thirty-six different HLA-A alleles were present (Table 1) in 161 individuals. The most frequently detected alleles were A*110101 (GF=0.2360), A*24020101 (GF=0.1646), and A*020101 (GF=0.1553); followed by A*3303 (GF=0.1180), A*3001 (GF=0.0590), and A*310102 (GF=0.0404; Figure1). The frequencies of following alleles, A*0203, A*0205, A*0206, A*0207, A*030101, A*2423, A*2601, A*3201, and A*3301, are higher

than 0.0093. Other 21 alleles are in low frequency. More alleles were detected in the HLA-A locus by SBT than by SSOP method (36 vs. 22; ref. *5*), but the most frequently detected alleles are almost the same.

Fifty-six individuals were homozygous at the HLA-A locus. The homozygous alleles include A*020101, A*110101, A*24020101, and A*310102. In Berber population from North Morocco, A*0201, A*0101, A*2301, A*24020101, A*3002, and A*2902 were homozygous (*13*).

The high-resolution polymorphism profile of HLA-A gene in Northern Chinese Han population shows differences with other population. By using SBT method, only 22 different HLA-A alleles were found in Berber population from North Morocco by Piancatelli D. (*13*). Alleles A*020101 (GF=0.178) and A*010101 (GF=0.137) were most frequently detected in Berber.

All 36 HLA-A alleles detected in the Northern Chinese Han population clustered to 14 serological specificities. A2 (27.02%) and A11 (24.84%) are most common. Serological type A2 is molecular A*02 subtype and is common among many populations (*2*, *5*, *12*). During the last few years HLA-A*02 family has rapidly grown to more than 70 members (http://www.anthonyloan.com/HIG/index.html). In this study, a total of 10 different A*02 alleles were identified in 161 Han individuals, namely A*020101 (57%), A*0207 (12%), A*0203 (10%), A*0206 (10%), A*0205 (5%), A*0202 (1%), A*0209 (1%), A*0211 (1%), A*0212 (1%), and A*0246 (1%). This A*02 polymorphism profile is quite different from Shankarkumar's research in Western Indians (*12*). In 204 Western Indians individuals, totally seven different A*02 alleles were identified and the most frequently detected allele was A*0211 (52.9%). Allele A*0207 is more common in Chinese, but was not detected in Indians. The comparison with other populations evidenced a different allelic composition of the A*30 group. A*3002 is much more common among many populations (GF=0.089; ref. *13*), especially Caucasians, but it is a rarity in Chinese.

The nature of polymorphism and molecular sequence variation in HLA gene is a consequent selection pressure due to environmental factors, mutation, ethnic groups admixture and other factors (*1*, *4*).

Though more than 300 HLA-A alleles have been found, HLA alleles often occur quite differently within human populations (*1*, *2*, *5*, *13*). This means that several alleles occur at intermediate frequencies, with a few at low frequencies or as single copies, and a few with very high frequencies. This study highlights some unique genetic traits of HLA-A gene in Northern Chinese Han population. These results by high-resolution SBT analysis of HLA-A polymorphism in Northern Han, especially the particular allele subtypes, will be of great interest for clinical transplantation, disease-associated study and forensic identification.

Materials and Methods

Population samples and genomic DNA extraction

161 blood samples from healthy unrelated Han individuals living in Shanxi, Beijing and Shenyang, North China, were collected for the HLA typing. Unclotted EDTA blood samples were stored frozen at -20 °C until DNA extraction. Genomic DNA was isolated from peripheral leukocytes treated with proteinase-K by the rapid mini-scale salting-out method. Concentration and purity of DNA samples were quantified by agarose gel electrophoresis (*14*).

HLA-A sequence-based typing

SBT of the HLA-A loci was performed on a 1 Kb DNA fragment amplified by PCR, using two locus-specific primers annealing in exons 2 and 3. This fragment was sequenced with primers specific for exons 2 and 3, using Big Dye terminator chemistry. Exon 4 was amplified and sequenced separately. Sequencing was performed both in forward and reverse directions. DNA from sequencing reactions was electrophoresed on an ABI 377 or 3730XL DNA sequencer. Amplifications were accomplished on a PTC-100 thermocycler using the following cycling conditions: 30 cycles of 95 °C for 1 min, 56 °C for 1 min, 72 °C for 1 min, and a final cycle of 72 °C for 5 min. Amplification primers are AF 5′-CTCTGCGGGGAGAAGCAA-3′ and AR 5′-CCAGCAAGGATGCCCACGAT-3′. Sequencing primers are 5′-TCG.GGC.AGG.TCT.CAG.CC-3′ and 5′-TCGGACCCGGAGACTGTA-3′ for exon 2, 5′-GGG.CTC.GGG.GGA.CCG.GT-3′ and 5′-GAG.GCG.CCC.CGT.GGC-3′ for exon 3, 5′-GTGCTGGGGAGTGTCCA-3′ and 5′-TAGGGAAA GCAGGAGCCTCT-3′ for exon 4.

Allele assignment

Sequence data were processed by using dedicated software: Sequencing Analysis^{TM}, MatchTools^{TM}, MT Navigator^{TM} (Applied Biosystems, Foster City, USA). MatchTools^{TM} and MT Navigator^{TM} programs work together to assign alleles and allow manual review or editing of the sequence data. They detect the heterozygous positions within each electropherogram and assess the typing based on an alignment of the processed sequence with an updated library of HLA sequences and alleles.

Statistical analysis

The HLA-A allele frequencies were assessed by direct count; samples containing one allele were considered homozygous and that allele was counted twice in the analysis. To test the hypothesis that the observed diploid genotypes were the product of a random union of gametes, the Hardy-Weinberg equilibrium was evaluated using the exact test.

Heterozygosity (H), polymorphic information content (PIC), discrimination power (DP) and probability of paternity exclusion (PPE) of HLA-A were calculated according to following formulae (*4*, *5*):

$$
\begin{aligned}
H &= 1 - \sum_{i=1}^{n} Pi^2 \\
PIC &= 1 - \sum_{i=1}^{n} Pi^2 - \sum_{I=1}^{n} \sum_{j=i+1}^{n} 2p^2 pj^2 \\
DP &= 1 - \sum_{i=1}^{n} Pi^2 \\
PPE &= \sum_{i=1}^{n} Pi\,(1 - Pi)^4
\end{aligned}
$$

Acknowledgements

This work was supported by grants from Beijing Nature Science Foundation (No. 7032022). Special gratitude is expressed here to the volunteers for their devotion. We also thank Jianfei Hu, Beijing Genomics Institute, Chinese Academy of Sciences, for his help in statistical analysis.

References

1. Jin, P. and Wang, E. 2003. Polymorphism in clinical immunology—From HLA typing to immunogenetic profiling. *J. Transl. Med.* 1: 1-11.
2. Imanishi, T., *et al.* 1992. Genetic relationships among various human populations indicated by MHC polymorphisms. In *HLA in 1991* (eds. Tsuji, K., *et al.*), vol. 1, pp 627-632. Oxford University Press, Oxford, UK.
3. Petersdorf, E.W., *et al.* 1999. Effect of HLA mismatches on the outcome of hematopoietic transplants. *Curr. Opin. Immunol.* 11: 521-526.
4. Lee, J. (ed.) 1990. *The HLA system: A new approach*, pp. 89-154. Springer-Verlag, New York, USA.
5. Yan, C.X., *et al.* 2002. Analysis of DNA polymorphism at HLA-A locus by PCR amplification with sequence specific oligonucleotide probe in Chinese Han and Uygur populations. *Acta Genetica Sinica* 29: 384-389.
6. Adams, S.D., *et al.* 2001. High throughput HLA sequence-based typing utilizing the ABI prism 3700 analyzer. *Tumori.* 87: s40-s43.
7. Adams, S.D., *et al.* 2001. Practicality of high throughput HLA sequence-based typing. *ASHI Quarterly* 25: 54-57.
8. Wang, L. and Marincola, F.M. 2003. Applaying pyrosequencing to HLA-typing. *ASHI Quarterly* 27: 16-18.
9. Erlich, H.A., *et al.* 2001. HLA DNA typing and transplantation. *Immunity* 14: 347-356.
10. Wang, E., *et al.* 2003. A strategy for detection of known and unknown SNP using a minimum number of oligonucleotides chip applicable in the clinical settings. *J. Transl. Med.* 1: 4.
11. Duquesnoy, R.J. 2002. HLA Matchmaker: a molecularly based algorithm for histocompatibility determination. I. Description of the algorithm. *Hum. Immunol.* 63: 339-352.
12. Shankarkumar, U., *et al.* 2003. HLA A*02 allele frequencies and B haplotype associations in Western Indians. *Hum. Immunol.* 64: 562-556.
13. Piancatelli, D. 2004. Human leukocyte antigen-A, -B, and -Cw polymorphism in a Berber population from North Morocco using sequence-based typing. *Tissue Antigens.* In press.
14. Miller, S.A., *et al.* 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16: 12-15.