J Hum Genet (2004) 49:73–79
DOI 10.1007/s10038-003-0110-7

ORIGINAL ARTICLE

Xiaoyi Liu · Zehuan Liu · Bin Lin · Yuanyuan Liu
Zhuoxun Chen · Weicong He · Dong Zhong · Anlong Xu

Catalog of 162 single nucleotide polymorphisms (SNPs) in a 4.7-kb region of the HLA-DP loci in southern Chinese ethnic groups

Received: 25 August 2003 / Accepted: 28 October 2003 / Published online: 15 January 2004
© The Japan Society of Human Genetics and Springer-Verlag 2004

Abstract HLA class-II proteins are cell-surface molecules that present antigens to T cells, and their expressional regulation is crucial to the immune reaction. Sequence variation at the regulatory region can directly affect the gene expression level. We cloned and sequenced a 4.7-kb region containing the regulatory region, exon1, and partial intron1 of both *HLA-DPA1* and *DPB1* genes in 25 variable sequences from southern Chinese ethnic groups and got a high-density map of 162 single nucleotide polymorphisms (SNPs): seven in 5′-flanking regions, four in 5′-untranslated regions, and four in the coding regions. By comparing these data with SNPs in dbSNP database in the NCBI, 145 SNPs (89.5%) were novel. In addition, eight genetic variations of insertion-deletion polymorphisms (INDELs) were discovered within the 4.7-kb region. These high-resolution maps can be used as resources of markers for association studies of complex diseases, assessment of individuals' predisposition to diseases, and tailoring of therapies, as well as research markers for population genetics and evolution.

Keywords Single nucleotide polymorphisms (SNPs) · Insertion-deletion polymorphisms (INDELs) · High-density SNP map · HLA-II genes · Regulatory region · Southern Chinese populations

X. Liu · Z. Liu · B. Lin · Y. Liu · Z. Chen · W. He
D. Zhong · A. Xu (✉)
The Key Laboratory of Genetic Engineering of MOE,
Department of Biochemistry,
College of Life Sciences,
Sun Yat-Sen (Zhongshan) University,
510275 Guangzhou,
P.R. China
E-mail: ls36@zsu.edu.cn
Tel.: +86-20-84113655
Fax: +86-20-84038377

Introduction

The completion of a high-quality sequence of the human genome is a landmark event in this century, symbolizing the beginning of the postgenomic era. In this era, much interest has turned to genome variation (Collins et al. 2003), that is, to an understanding of how genomes change and take on new functional roles. Comparison of genome sequences from evolutionarily diverse species provides insight into the evolution of genes (Fu and Li 1999; Verrelli et al. 2002; Wooding et al. 2002) and a more comprehensive understanding of the function of important genomic elements. The study of sequence variation within species will also be important in defining the relationships between genotype and biological function, such as individual differences at health, susceptibility to diseases, drug response, and so on.

Besides the protein-coding sequences, a large amount of the noncoding portion of the human genome is also under active selection, suggesting that it is functionally important. It probably contains the bulk of the regulatory information controlling the expression of protein-coding genes as well as nonprotein-coding genes (Bamshad et al. 2002). It may contain sequence determinants of chromosome dynamics such as methylation and chromatin remodeling (Collins et al. 2003). Therefore, the noncoding portion of the human genome also becomes a focal point in the study of genetic variations.

Major histocompatibility complex (MHC) class-II antigens of human (HLA) are cell-surface molecules regulating a specific immune response to a pathogen by presenting antigens to T-cell receptors so as to mediate the activation of T lymphocytes. There are three isotypes of class-II molecules—DR, DQ, and DP—each consisting of two subunits, one α and one β chain encoded by separate genes *DR (DQ, DP) A* and *B*, respectively. The abnormal expression of HLA-II genes causes certain diseases. For example, the expression of class-II molecules on inappropriate cells may change the ability

of antigen-presenting cells to present antigen (both foreign and self) to T lymphocytes, triggering an autoimmune disease (Laurie et al. 1992). A deficient expression of the MHC-II gene results in a hereditary immunodeficiency disease called bare lymphocyte syndrome (BLS) (Mach et al. 1996). Thus, the expressional regulation of HLA-II genes is crucial in the control of the immune response.

Four sequence motifs within promoter proximal regions of all class-II genes have been identified as *cis*-acting regulatory elements, termed W, X_1, X_2, and Y boxes, respectively (van den Elsen et al. 1998). These four boxes are highly conserved with respect to their sequences, relative positions, orientation, and spacing. Variation within these boxes could affect the gene expression level and the nuclear protein-binding affinities, which have been confirmed on the *DRB1*, *DRB3*, *DQA1*, *DQB1*, and *DPB1* genes (Emery et al. 1993; Morzycka-Wroblewska et al. 1997; Andersen et al. 1991; Varney et al. 1999). Therefore, it is very important to study the polymorphism of the regulatory regions of HLA-II genes, which is helpful to better understand the expressional regulation in association with the immune response in humans. Our study is the first attempt to study the variation of the regulatory region in the MHC-II genes of Chinese populations. The SNPs in the promoter region of HLA-II genes found in this study can be used as resources of markers for association studies of complex diseases, assessment of individuals' predisposition to diseases, and therapy tailoring, as well as markers for population genetics and evolution research.

HLA-DPA1 and *HLA-DPB1* genes are classical HLA-II genes, and they are organized in head-to-head fashion with their 5' ends pointing toward each other resulting in the sequence between them functioning as promoters of both genes. Therefore, we selected an approximately 5-kb-region at these two loci containing the promoter region, the exon1, and partial intron1 of both *DPA1* and *DPB1* genes. To cover as much polymorphism as possible, sequence data were obtained from seven different ethnic populations, including both ethnic populations of southern origin and those of northern origin, in China (Yao et al. 2002). They are Jing, Lahu, Yao, Pumi, Naxi, Li, and Guangdong Han, mainly from southwest China.

Materials and methods

Fourteen healthy and unrelated peripheral blood samples with different *HLA-DPB1* alleles (including 02012, 0202, 03011, 0401, 0402, 0501, 1401, 2201, 6201, 2801, 6301, 5101, 5601, 8001) based on our works before were collected from southwest China populations for studying the 5-kb region. Genomic DNA was extracted from whole blood containing ACD anticoagulant by the modified salting-out method, as indicated by the International Histocompatibility Work Group (IHWG) (http://www.ihwg.org/protocols).

Based on the contig NT_033951 (gi: 27498326) in GenBank containing the complete sequence of the HLA region, two 28-nt primers were designed to amplify the 5-kb target fragment: 5'-AGGGCTTGAGGGGCTGTATTCAGGAGAT-3' and 5'-AGCTGGGTCTGGACTTCAAACTTGGCTC-3'. PCR amplification was performed in a 20-μl reaction volume containing 0.75 mmol/l each dNTP, 0.25 μmol/l each primer, 1U Extaq polymerase (Takara) and 50 ng genomic DNA. A two-step PCR program of 35 cycles in total was carried out: 95°C for 3 min; 10 cycles of 94°C for 40 s, 68°C for 4 min; and 25 cycles of 94°C for 40 s, 68°C for 4 min (increasing 5 s each cycle) followed by 72°C for 10 min at the end. The products were cloned into the pGEM-T Easy Vector (Promega, USA). Six positive plasmids for each consensus sequence were sequenced from both directions on an ABI 3700 sequencer using Bigdye reagent (Applied Biosystems, USA).

All segment sequences were assembled automatically using SeqMan in DNASTAR software package and then were carefully checked manually using the same program. All sequences were aligned with the Clustalx program (Thomson et al. 1997). Singletons and doubletons were verified by reamplifying and resequencing in both directions.

Results and discussion

From the 14 samples, 25 cloned sequences of 4751to 4759-bp-long fragments were obtained. There were 170 polymorphisms found, eight of which were insertions and deletions (INDELs) and all of which were shorter than 12 bp, with three INDELs in the intron 1 of the *HLA-DPB1* gene and five in the region between *HLA-DPB1* and *DPA1* genes. The detailed sequence information about the polymorphisms is listed in Table 1.

After exclusion of all INDELs, 4735 bp remained and were used to position the polymorphic sites (Table 1). Within the 4735-bp region, we identified 162 SNPs: 49 in intron1 of the *HLA-DPB1* gene, three in exon1 of the *HLA-DPB1* gene, two in the 5'-untranslated region of the *HLA-DPB1* gene, five in the 5'-flanking region of the *HLA-DPB1* gene, 83 in the region between the *HLA-DPB1* and *DPA1* genes, two in the 5'-flanking region of the *HLA-DPA1* gene, two in the 5'-untranslated region of the *HLA-DPA1* gene, one in exon1 of the *HLA-DPA1* gene, and 15 in intron 1 of the *HLA-DPA1* gene. The distribution was one SNP per 29 bp on average, and much denser than the average level of the human genome ($0.1 \backsim 1\%$). Frequencies of substitutions by types were 38.9% for A/G (63), 32.7% for C/T (53), 9.3% for A/T (15), 7.4% for C/G (12), 5.6% for A/C (9), 4.9% for G/T (8), and 0.6% for both C/T/G (1) and C/T/A (1). The ratio of transition and transversion was 2.6, being close to the 2.3.

Regarding the four cSNPs (SNP in coding regions), two were synonymous and the other two were nonsynonymous. One synonymous substitution in exon1 of the *HLA-DPB1* gene was C/T at +117 and encodes alanine; another synonymous substitution was A/T at +40 in exon1 of the *HLA-DPA1* gene and encodes the same proline. Both nonsynonymous substitutions were in exon 1 of the *DPB1* gene: one was a C/T transition at the +140 position, leading to Thr/Met change at 16; another was T/C transition at +152, leading to Met/Thr change at 20. None were reported in either the dbSNP database (http://www.ncbi.nlm.nih.gov/entrez/

Table 1 Characterization of variations in the 4.7-kb region in southern Chinese ethnic populations. Variation is shown by capital letter. The number of nucleotides in the coding sequences, 5'-untranslated region, and 5'-flanking regions is according to the sequence information of NT_033951 (gi: 27498326) from NCBI. *DPA1* exon 1 and *DPB1* exon 1, 100 bp each; 5'-untranslated region of *DPA1*, 31 bp; 5'-untranslated region of *DPB1*, 59 bp; 5'-flanking region of *DPA1*, 41 bp; 5'-flanking region of *DPB1*, 34 bp. The position of variations in the region between two genes is according to the distance to the *DPA1* gene's transcription initiation code. *Frequency* the number of the major single nucleotide polymorphism (SNP) allele/the number of the minor SNP allele: e.g., the variation with ID170 has 20 Gs and 5 Ts in our samples. *INDEL* insertion and deletion polymorphism

ID	Region	Position	Flanking sequence	Variation (5' to 3')	Flanking sequence	Substitution	Identity to dbSNP	Frequency
1	DPB1-intron 1	1830	tgacatgatgtcagtacctt	C/T	taggtgtgtcctaggagggg		dbSNP ID: rs3135021	22/3
2	DPB1-intron 1	1802	gtcctaggaggggtgtggag	A/G	ggtctaggaggaagggtgta			17/8
3	DPB1-intron 1	1772	ggaagggtgtaagcaggaaa	A/G	gtggaagtcagtggA/Gaaact		dbSNP ID: rs2856816	14/11
4	DPB1-intron 1	1757	ggaaaA/Ggtggaagtcagtgg	A/G	aaactaaggtacctactttg			23/2
5	DPB1-intron 1	1646	gtttacagtcccaccaacag	A/T	gtaaaagtgttcctatttct			21/4
6	DPB1-intron 1	1544	tgagatggtatctcaagaca	T/G	tgtggcA/GattcctcaaG/Agat			15/10
7	DPB1-intron 1	1537	gtatctcaagacaT/Gtgtggc	A/G	attcctcaaG/Agatctagaac			21/4
8	DPB1-intron 1	1527	acaT/GtgtggcA/Gattcctcaa	G/A	gatctagaactagaaatacc		dbSNP ID: rs2213309	17/8
9	DPB1-intron 1	1480	cccagccatcccattactgg	G/A	tatatacccaaaggattata		dbSNP ID: rs2213308	17/8
10	DPB1-intron 1	1427	ataaagacacatgcacacat	A/G	tgtttatT/Agcggcactattc			18/7
11	DPB1-intron 1	1419	acatgcacacatA/Gtgtttat	T/A	gcggcactattcacaatagc		dbSNP ID: rs2213307	22/3
12	DPB1-intron 1	1377	aagacttggaaccaacccaa	C/A	tgtcC/Tatcagtgatagactg			21/4
13	DPB1-intron 1	1372	ttggaaccaacccaaC/Atgtc	C/T	atcagtgatagactggatta			21/4
14	DPB1-intron 1	1343	tagactggattaagaaaatg	C/T	A/Gcacatatacaccatggaat			21/4
15	DPB1-intron 1	1342	agactggattaagaaaatgC/T	A/G	cacatatacaccatggaata			21/4
16	DPB1-intron 1	1259	gacatggaggaagctggaaa	C/T	catcattcacagcaaactat			17/8
17	DPB1-intron 1	1228	agcaaactattgcaaggaca	A/G	aaaaccaaagaccacatgtt			24/1
18	DPB1-intron 1	1156	acacatggacacaggaaggg	G/C	aacatcaT/CacaccG/Agggcct			23/2
19	DPB1-intron 1	1148	acacaggaagggG/Caacatca	T/C	acaccG/Agggcctgttgtggg			21/4
20	DPB1-intron 1	1142	gaagggG/CaacatcaT/Cacacc	G/A	gggcctgttgtggggtgggg			21/4
21	DPB1-intron 1	1061	aatgacgagttaatgggtgc	A/G	gcacaccaacatggcacatg			24/1
22	DPB1-intron 1	1029	tggcacatgtatacatatgt	A/G	acaaacctG/AcacG/AttgtgC/Ga			24/1
23	DPB1-intron 1	1020	tatacatatgtG/Aacaaacct	G/A	cacG/AttgtgC/Gacatgtaccc			21/4
24	DPB1-intron 1	1016	catatgtG/AacaaacctG/Acac	G/A	catatgtG/AacaaacctG/Acac			21/4
25	DPB1-intron 1	1010	tG/AacaaacctG/AcacG/Attgtg	C/G	acatgtaccctagaacttaa			23/2
26	DPB1-intron 1	987	atgtaccctagaacttaaag	(TA) 6–10	aaaagaaatatcctgagagt			
27	DPB1-intron 1	952	agaaatatcctgagagtatc	C/T	tgagagaaaaacaatggatc			24/1
28	DPB1-intron 1	890	cagccacagggaggcagcag	G/A	aagaagagatttttcacttg			24/1
29	DPB1-intron 1	827	gataacctacactcaatccc	T/G	gT/AcatgcaccC/Gacacaacct			21/4
30	DPB1-intron 1	825	taacctacactcaatcccT/Gg	T/A	catgcaccC/Gacacaacctca			23/2
31	DPB1-intron 1	816	ctcaatcccT/GgT/Acatgcacc	C/G	acacaacctcatctcattct			21/4
32	DPB1-intron 1	785	atctcattcttcctacaagt	C/T	gcaacccactctagtacC/Tct			21/4
33	DPB1-intron 1	767	gtC/Tgcaacccactctagtac	C/T	ctaaccaagttctggggagg			21/4
34	DPB1-intron 1	674	ctccctcccttctctcctta	A/G	ttgtccaagggA/Gtgtgggct			24/1
35	DPB1-intron 1	662	tctccttaG/Attgtccaaggg	A/G	tgtgggctcctgctccaccT/C		dbSNP ID: rs2071354	22/3
36	DPB1-intron 1	642	G/Atgtgggctcctgctccacc	C/T	tggggaatG/Aagaggcattct			24/1
37	DPB1-intron 1	633	cctgctccaccT/Ctggggaat	A/G	agaggcattcttcgtgagca			21/4

Table 1 (Continued)

ID	Region	Position	Flanking sequence	Variation (5' to 3')	Flanking sequence	Substitution	Identity to dbSNP	Frequency
38	DPB1-intron 1	532	ccctccaggctggatgcaga	C/T	G/Atgaacacgcccaggagcct			18/7
39	DPB1-intron 1	531	cctccaggctggatgcagaT/C	G/A	tgaacacgcccaggagcctg		dbSNP ID: rs2071353	15/10
40	DPB1-intron 1	509	gaacacgcccaggagcctgc	A/G	cttgccaacttctctctctg			24/1
41	DPB1-intron 1	462	ttgtctgtccaaggttatcc	A/T	ggacctcttggtcccatttt		dbSNP ID: rs2071352	18/7
42	DPB1-intron 1	432	gtcccattttccccaaagac	AAGCAT/ INDEL	actgaccactgggcacctct			
43	DPB1-intron 1	394	tctctctgtttactgtaccc	A/C	tgtccttggagagagaatag			23/2
44	DPB1-intron 1	378	cccC/Atgtccttggagagaga	A/INDEL	taggccT/Agtagggtagcaag			
45	DPB1-intron 1	371	ccttggagagagaataggcc	A/T	gtagggtagcaagG/Ctatcct			21/4
46	DPB1-intron 1	357	taggccT/Agtagggtagcaag	C/G	tatcctataggggatG/Aaacc			21/4
47	DPB1-intron 1	341	caagG/Ctatcctataggggat	G/A	aacctctttctctagctggg			21/4
48	DPB1-intron 1	317	ctctttctctagctgggaag	A/G	gaggacactgG/Ccttagggca			21/4
49	DPB1-intron 1	306	gctgggaagG/Agaggacactg	C/G	cttagggcaggagcccccgc			21/4
50	DPB1-intron 1	215	tccctgagccagaccctcca	A/G	gaatggcagT/CtcgG/Actctta			21/4
51	DPB1-intron 1	205	agaccctccaG/Agaatggcag	C/T	tcgG/Actcttacctggagtgg		dbSNP ID: rs2071351	15/10
52	DPB1-intron 1	201	cctccaG/AgaatggcagT/Ctcg	G/A	ctcttacctggagtggccct			21/4
53	DPB1-exon1	152	ccacagatgtgagcagcacc	A/G	tcagtaacgccG/Atcagagcc	Met20Thr		24/1
54	DPB1-exon 1	140	gcagcaccG/Atcagtaacgcc	G/A	tcagagccactgtccggggg	Thr16Met		21/4
55	DPB1-exon 1	117	agagccactgtccgggggc	C/T	gcagaaacctgcagaaccat	Ala8Ala		21/4
56	DPB1–5'untranslated region	45	catcatggagctggaaaagg	A/G	tggcaaaaT/Cgaaaagagctg			20/5
57	DPB1–5'-untranslated region	36	gctggaaaaggG/Atggcaaaa	T/C	gaaaagagctgcagtcagga			23/2
58	DPB1–5'-flanking region	−16	ccatgaacccaagtagtctt	C/T	tgtgaccctgggattggaca			24/1
59	DPB1–5'-flanking region Y box	−58	agtctgagaaaagaaccaat	G/A	gacactgagctttgtatgag			21/4
60	DPB1–5'-flanking region X_1 box	−91	tgtatgagtcattgctcact	A/G	ggcagaaagttG/Agtatgaaa			21/4
61	DPB1–5'-flanking region	−103	tgctcactG/Aggcagaaagtt	A/G	gtatgaaaggtctgaaaata			24/1
62	DPB1–5'-flanking region-W' box	−2175	cacttaagatgacggaggaa	A/G	gacagtgatactcattttaa			21/4
63	Region between DPB1 and DPA1	−2134	ccagtcagataagtcatgat	G/A	tttggggG/Agattatgcgttt		dbSNP ID: rs2071350	22/3
64	Region between DPB1 and DPA1	−2126	ataagtcatgatA/Gtttgggg	A/G	gattatgcgttttctttgct			23/2
65	Region between DPB1 and DPA1	−2040	tcatttccccaattctgcca	C/T	acC/Atcacacacccacaggac			21/4
66	Region between DPB1 and DPA1	−2037	tttccccaattctgccaT/Cac	C/A	tcacacacccacaggacatg			21/4
67	Region between DPB1 and DPA1	−2036	tccccaattctgccaT/CacA/Ct	CACACA-CCCACA/ INDEL	ggacatggtctgttgtggaa			
68	Region between DPB1 and DPA1	−2015	ggacatggtctgttgtggaa	A/G	aagtgctatcttagC/Tgtgta			21/4
69	Region between DPB1 and DPA1	−2000	tggaaG/Aaagtgctatcttag	T/C	gtgtaaaaggtcattcagtg			21/4
70	Region between DPB1 and DPA1	−1951	tagagggattagagtaccca	T/C	ctcagaactcaaatgaggC/Tc			24/1
71	Region between DPB1 and DPA1	−1932	aC/Tctcagaactcaaatgagg	T/C	ctgagtctgtctgtcttgcc			21/4
72	Region between DPB1 and DPA1	−1880	gtgtgtttaagattagcacc	C/T	attcatatttactttctccc			21/4
73	Region between DPB1 and DPA1	−1851	ttactttctcccagaggtct	G/A	tgagtcctgcgatgtgcagg			21/4
74	Region between DPB1 and DPA1	−1828	agtcctgcgatgtgcaggag	T/A	taccaggttcttT/Ccacagga			21/4
75	Region between DPB1and DPA1	−1815	gcaggagA/Ttaccaggttctt	C/T	cacaggactgtcatcaggA/Gt			24/1
76	Region between DPB1 and DPA1	−1796	tT/Ccacaggactgtcatcagg	G/A	tcaggagggT/Ctcagtctagg			21/4
77	Region between DPB1 and DPA1	−1786	tgtcatcaggA/Gtcaggaggg	C/T	tcagtctagggaccttacac			21/4
78	Region between DPB1 and DPA1	−1765	tcagtctagggaccttacac	T/C	gggagcgtggacacaccacc			21/4
79	Region between DPB1 and DPA1	−1723	accctaccatgtaaatatgc	A/T	gctttaacgacactgT/CctG/Ct			23/2

80	Region between DPB1 and DPA1	−1707	atgcT/Agctttaacgacactg	T/C	ctG/CtcttggacttT/CA/Gattcc		21/4
81	Region between DPB1 and DPA1	−1704	cT/AgctttaacgacactgT/Cct	G/C	tcttggacttT/CA/Gattccttg		21/4
82	Region between DPB1 and DPA1	−1693	acactgT/CctG/Ctcttggactt	T/C	acactgT/CctG/Ctcttggactt		24/1
83	Region between DPB1 and DPA1	−1692	cactgT/CctG/CtcttggacttT/C	A/G	attccttgtcttcaatatgg		21/4
84	Region between DPB1 and DPA1	−1634	ttatagcaaagctcttaaga	G/T	caaatgagatcaataggtct		21/4
85	Region between DPB1 and DPA1	−1598	ggtctgaaattgctttggaa	T/A	aataaagtccatagagaata		21/4
86	Region between DPB1 and DPA1	−1568	catagagaatataaattgac	C/T	ggtaaataaggaagaaagtg		21/4
87	Region between DPB1 and DPA1	−1547	ggtaaataaggaagaaagtg	(A)8–10	tcatcaaagaccccaccA/TcT/C		
88	Region between DPB1 and DPA1	−1523	gtgtcatcaaagaccccacc	A/T	cT/CttcaaagcG/Atctgattct		21/4
89	Region between DPB1 and DPA1	−1521	gtcatcaaagaccccaccA/Tc	T/C	ttcaaagcG/AtctgattctG/AC/T		21/4
90	Region between DPB1 and DPA1	−1512	gaccccaccA/TcT/Cttcaaagc	G/A	tctgattctG/AC/TtttgtttaT/G		24/1
91	Region between DPB1 and DPA1	−1502	cT/CttcaaagcG/Atctgattct	G/A	C/TttgtttaT/Gagtgaatatt		21/4
92	Region between DPB1 and DPA1	−1501	T/CttcaaagcG/AtctgattctG/A	C/T	tttgtttaT/Gagtgaatattt	dbSNP ID: rs987870	22/3
93	Region between DPB1 and DPA1	−1492	G/AtctgattctG/AC/Ttttgttta	T/G	agtgaatatttttacaaact		21/4
94	Region between DPB1 and DPA1	−1410	gatcctgttggaccaccaat	A/T	ctaatttcaaaaaaaatcaa		24/1
95	Region between DPB1 and DPA1	−1381	aaaaaaaatcaaattggcat	A/C	atttctttaaatT/CG/Aatatta		21/4
96	Region between DPB1 and DPA1	−1368	ttggcatA/Catttctttaaat	T/C	G/AatattaaaatgC/Taaataga		21/4
97	Region between DPB1 and DPA1	−1367	tggcatA/CatttctttaaatT/C	G/A	atattaaaatgC/TaaatagaG/A		21/4
98	Region between DPB1 and DPA1	−1355	tttaaatT/CG/Aatattaaaatg	C/T	aaatagaG/AggctagcaC/Tgta		24/1
99	Region between DPB1 and DPA1	−1347	G/AatattaaaatgC/Taaataga	G/A	ggctagcaC/Tgtaaaatgcaa		24/1
100	Region between DPB1 and DPA1	−1338	atgC/TaaatagaG/Aggctagca	C/T	gtaaaatgcaagagaagaaa		24/1
101	Region between DPB1 and DPA1	−1294	attatgtagtagaaggtgtc	C/T	agtaggattG/AaggggccT/Agg		21/4
102	Region between DPB1 and DPA1	−1284	agaaggtgtcC/Tagtaggatt	G/A	aggggccT/AggaC/Gaggattgt		21/4
103	Region between DPB1 and DPA1	−1276	tcC/TagtaggattG/Aaggggcc	T/A	ggaC/Gaggattgtggagaaaa		24/1
104	Region between DPB1 and DPA1	−1272	gtaggattG/AaggggccT/Agga	C/G	aggattgtggagaaaaG/Actc		21/4
105	Region between DPB1 and DPA1	−1255	ggaC/Gaggattgtggagaaaa	G/A	ctcagttttttagtaactg		20/5
106	Region between DPB1 and DPA1	−1232	cagttttttagtaactgcc	C/T	gggtgggactC/TtC/Tctccttg		24/1
107	Region between DPB1 and DPA1	−1221	gtaactgccC/Tgggtgggact	C/T	tC/Tctccttgtcatttacctg		21/4
108	Region between DPB1 and DPA1	−1219	aactgccC/TgggtgggactC/Tt	C/T	ctccttgtcatttacctgcc		17/8
109	Region between DPB1 and DPA1	−1198	ctccttgtcatttacctgcc	G/A	tagataattagacaaatact		24/1
110	Region between DPB1 and DPA1	−1177	agataattagacaaatactt	AATT/ INDEL	cctcA/Gatgtctcaggtaatt		
111	Region between DPB1 and DPA1	−1172	aattagacaaatacttcctc	A/G	atgtctcaggtaatttattt		17/8
112	Region between DPB1 and DPA1	−1137	ttatttttacagtagacatt	G/A	ataG/Atacctccctcaggttt		23/2
113	Region between DPB1 and DPA1	−1133	ttttacagtagacattG/Aata	G/A	tacctccctcaggtttaatG/A		24/1
114	Region between DPB1 and DPA1	−1113	G/Atacctccctcaggtttaat	G/A	tgaggtttaaatgaggtaat		18/7
115	Region between DPB1 and DPA1	−1084	aaatgaggtaatgcatgcaa	G/A	gcactaataacagC/Aatctct		24/1
116	Region between DPB1 and DPA1	−1070	atgcaaG/Agcactaataacag	C/A	atctctcatgtA/CctaatcC/TA/G		21/4
117	Region between DPB1 and DPA1	−1058	aataacagC/Aatctctcatgt	A/C	ctaatcC/TA/Gttagtaattatt		21/4
118	Region between DPB1 and DPA1	−1051	gC/Aatctctcatgt A/Cctaatc	C/T	A/Gttagtaattattatattta		21/4
119	Region between DPB1 and DPA1	−1050	C/AatctctcatgtA/CctaatcC/T	A/G	ttagtaattattatatttat		21/4
120	Region between DPB1 and DPA1	−1027	agtaattattatttatgcaa	C/G	gtaaggctaccaT/Caaataag		22/3
121	Region between DPB1 and DPA1	−1014	tatgcaaC/Ggtaaggctacca	T/C	aaataagataaaatattaaT/C		17/8
122	Region between DPB1 and DPA1	−994	T/Caaataagataaaatattaa	T/C	gtaaC/TatatgtaT/Cttcatca		21/4
123	Region between DPB1 and DPA1	−989	aagataaaatattaaT/Cgtaa	C/T	atatgtaT/Cttcatcattaca		22/3
124	Region between DPB1 and DPA1	−981	atattaaT/CgtaaC/Tatatgta	T/C	ttcatcattacattaaatG/Ag		21/4
125	Region between DPB1 and DPA1	−962	aT/Cttcatcattacattaaat	G/A	gtaaatgttC/G/TtgtgaT/Ccttc		24/1
126	Region between DPB1 and DPA1	−952	tacattaaatG/Agtaaatgtt	C/G/T	tgtgaT/CcttcC/Tctatgagta		1/4/20
127	Region between DPB1 and DPA1	−946	aaatG/AgtaaatgttC/G/Ttgtga	T/C	cttcC/Tctatgagtatcaaat		21/4
128	Region between DPB1 and DPA1	−941	gtaaatgttC/G/TtgtgaT/Ccttc	C/T	ctatgagtatcaaataaata		23/2
129	Region between DPB1 and DPA1	−920	ctatgagtatcaaataaata	C/G	ctcaaaaC/Ttgaaatatttaa		21/4
130	Region between DPB1 and DPA1	−912	atcaaataaataC/Gctcaaaa	C/T	tgaaatatttaaaatgagac		17/8

Table 1 (Continued)

ID	Region	Position	Flanking sequence	Variation (5' to 3')	Flanking sequence	Substitution	Identity to dbSNP	Frequency
131	Region between DPB1 and DPA1	−877	tgagactagaaaagtatttc	G/T	taaatatggtaaagagaata			23/2
132	Region between DPB1 and DPA1	−845	agagaatataaaataaaaaa	A/INDEL	ttaacattacctgtaacaga			
133	Region between DPB1 and DPA1	−821	acattacctgtaacagaaaa	G/A	cattagagaaT/Acccttctta			21/4
134	Region between DPB1 and DPA1	−810	aacagaaaaG/Acattagagaa	T/A	cccttcttaaaaacaagatc			24/1
135	Region between DPB1 and DPA1	−787	cttcttaaaaacaagatcaa	G/A	atatgggcgcacattgttac			24/1
136	Region between DPB1 and DPA1	−719	tatatgaaaatgaaataaca	T/A	gactagtaatttaaagataa			21/4
137	Region between DPB1 and DPA1	−699	gactagtaatttaaagataa	TCGAA-GATC/ INDEL	aC/Aatgttattacactgttta			
138	Region between DPB1 and DPA1	−697	actagtaatttaaagataaa	C/A	atgttattacactgtttaaa			21/4
139	Region between DPB1 and DPA1	−643	caaatactaagaaatctgaa	T/A	aataaaaattatctG/Agaaat			21/4
140	Region between DPB1 and DPA1	−628	ctgaaT/Aaataaaaattatct	G/A	gaaatttcactaacaaaggt			21/4
141	Region between DPB1 and DPA1	−590	ggtaaccaggaatagataaa	T/C	aattaaagatcaacttcatt			21/4
142	Region between DPB1 and DPA1	−503	aatgattaaataaaaaggag	A/G	tatctcagggaggtggacct			21/4
143	Region between DPB1 and DPA1	−447	tatatattttattagatagc	G/A	cattcattcttaattatcag			21/4
144	Region between DPB1 and DPA1	−398	ttttgtaaaatattgacaac	G/A	taacataaaacatgatttaa			21/4
145	Region between DPB1 and DPA1	−367	catgatttaaaatatagtga	T/C	caG/AaaaacatcG/Aaagatcat			21/4
146	Region between DPB1 and DPA1	−364	gatttaaaatatagtgaT/Cca	G/A	aaaacatcG/Aaagatcatctg			21/4
147	Region between DPB1 and DPA1	−245	ctttttacatctctttctct	G/A	actctgaaaatgaactgtga			21/4
148	Region between DPB1 and DPA1	−210	ctgtgaactggagctctctt	A/G	accacgctggtacT/Ctaaaat		dbSNP ID: rs2051548	22/3
149	Region between DPB1 and DPA1	−196	tctcttA/Gaccacgctggtac	T/C	taaaattctcccatctcttc			21/4
150	Region between DPB1 and DPA1	−71	tgattcctctccacccattt	G/C	cagtgctagaggcccacagt			20/5
151	DPA1–5'-flanking region	−14	tccactcggcctcagttcct	T/C	atcactgC/Ttcctgtgctcac			20/5
152	DPA1–5'-flanking region	−6	gcctcagttcctT/Catcactg	C/T	tcctgtgctcacagtcatca			20/5
153	DPA1–5'-untranslated region	26	cagtcatcaattatagaccc	T/C	acaG/AcatgcgcccA/Tgaagac			20/5
154	DPA1–5'-untranslated region	30	catcaattatagacccT/Caca	G/A	catgcgcccA/Tgaagacagaa			20/5
155	DPA1-exon 1	40	agacccT/CacaG/Acatgcgccc	A/T	gaagacagaatgttccatat	Pro3Pro		20/5
156	DPA1-intron 1	139	ggggccatcaagggtgagtg	G/C	tcaggaggaT/CgcaA/Ggagcgt			20/5
157	DPA1-intron 1	149	agggtgagtgG/Ctcaggagga	T/C	gcaA/GgagcgtT/Cggggtgagt			23/2
158	DPA1-intron 1	153	tgagtgG/CtcaggaggaT/Cgca	A/G	gagcgtT/Cggggtgagtgatg		dbSNP ID: rs4247257	15/10
159	DPA1-intron 1	160	tcaggaggaT/CgcaA/Ggagcgt	T/C	ggggtgagtgatggggtggt			20/5
160	DPA1-intron 1	193	ggggtggttcacatcaattg	T/C	tgcttcaggA/GatcaG/Cagatt		dbSNP ID: rs1431401	15/10
161	DPA1-intron 1	203	acatcaattgT/Ctgcttcagg	A/G	atcaG/Cagattttaggggctc		dbSNP ID: rs1431400	15/10
162	DPA1-intron 1	208	aattgT/CtgcttcaggA/Gatca	G/C	agattttaggggctcattga			20/5
163	DPA1-intron 1	287	cataataataacagcaataa	T/C	agccagaatttatgagacG/Tc			20/5
164	DPA1-intron 1	306	aT/Cagccagaatttatgagac	G/T	cctgcataT/Gtttctttcccc			13/12
165	DPA1-intron 1	315	atttatgagacG/Tcctgcata	T/G	tttctttccccatttacatc		dbSNP ID: rs3135020	23/2
166	DPA1-intron 1	345	ccatttacatctcacaggaa	C/A/T	cttcaatgaagataatattc		dbSNP ID: rs1431399	10/5/10
167	DPA1-intron 1	367	ttcaatgaagataatattcc	C/A	ttcatttagaaattG/AttcT/Ct			20/5
168	DPA1-intron 1	382	attccC/Attcatttagaaatt	G/A	ttcT/CttttatG/Ttagaaatta			23/2
169	DPA1-intron 1	386	cC/AttcatttagaaattG/Attc	T/C	ttttatG/Ttagaaattatttt			20/5
170	DPA1-intron 1	393	ttagaaatG/AttcT/Cttttat	G/T	tagaaattattttgaaaaaa			20/5

query.fcgi?db = snp) or the IGMT/HLA sequence database (http://www.ebi.ac.uk/imgt/hla/) until September 2003.

In the highly conserved X_1, Y, and W' box within the promoter of the *DPB1* gene (van den Elsen et al. 1998), there was one substitution per box, respectively. These three substitutions were all G/A transitions and had been reported before by Varney et al. (1999) who named the allele containing these three G/A substitutions as DP-PRO4. More interestingly, Varney et al. found this allele in seven individuals with eastern Asian origin. All these data suggest that DP-PRO4 allele containing these 3 G/A substitutions in the X_1, Y, and W' box may originate from China. Their competitive binding assay (Varney et al. 1999) showed that the substitutions in the W' and X_1 boxes had no effect on binding affinity, while a single substitution at the site immediately adjacent to the inverted CCAAT motif in the Y box reduced binding affinity. However, whether this substitution can influence the transcription of the *DPB1* gene in vivo should be further studied by experiments in vivo, since the Y box has not the same importance as the X_1 box in regulating gene expression.

By comparing our data with SNPs deposited in the dbSNP database in the NCBI, we found that 145 (89.5%) of 162 SNPs were novel as of August 2003. However, three SNPs found in the dbSNP database (rs2071349, rs2856830, and rs4279481) in GenBank within this region have not been found in our 25 sequences. In short, these high-resolution genome variation maps with an unusually high density of SNPs can be used as resources of markers for association studies of complex diseases, assessment of individuals' predisposition to diseases, and therapy tailoring, as well as research markers for population genetics and evolution.

Acknowledgements This research was supported by projects (Nos. 30178073 and 30100275) and the key project (No. 69935020) of the National Natural Science Foundation of China, the key project (No. 021691) of the Guangdong Natural Science Foundation, and the projects (No. 2001AA224021-04) of the State High-Tech Development Project of the Ministry of Science & Technology, Guangdong Provincial Department of Science and Technology, and the Bureau of Science and Technology of Guangzhou City.

References

Andersen LC, Beaty JS, Nettles JW, Seyfried CE, Nepom GT, Nepom BS (1991) Allelic polymorphism in transcriptional regulatory regions of *HLA-DQB1* genes. J Exp Med 173:181–192

Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Watkins WS, Wooding S, Stone AC, Jorde LB, Weiss RB, Ahuja SK (2002) A strong signature of balancing selection in the 5'*cis*-regulatory region of *CCR5*. Proc Natl Acad Sci USA 99:10539–10544

Collins FS, Green ED, Guttmacher AE, Guyer MS (2003) A vision for the future of genomics research: A blueprint for the genomic era. Nature 422:835–847

Emery P, Mach B, Reith W (1993) The different level of expression of *HLA-DRB1* and *-DRB3* genes is controlled by conserved isotypic differences in promoter sequence. Hum Immunol 38:137–147

Fu YX, Li WH (1999) Coalescing into the 21st century: an overview and prospects of coalescent theory. Theor Popul Biol 56:1–10

Laurie HG, Catherine JK (1992) Sequences and factors: a guide to MHC class-II transcription. Annu Rev Immunol 10:13–49

Mach B, Steimle V, Martinez-Soria E, Reith W (1996) Regulation of MHC class II genes: lessons from a disease. Annu Rev Immunol 14:301–331

Morzycka-Wroblewska E, Munshi A, Ostermayer M, Harwood JI, Kagnoff MF (1997) Differential expression of *HLA-DQA1* alleles associated with promoter polymorphism. Immunogenetics 45:163–170

Thomson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Varney MD, Gavrilidis A, Tait BD (1999) Polymorphism in the regulatory regions of the *HLA-DPB1* gene. Hum Immunol 60:955–961

Van den Elsen PJ, Peijnenburg A, Van Eggermond MCJA, Gobin SJP (1998) Shared regulatory elements in the promoters of MHC class I and II genes. Immunol Today 19:309–312

Verrelli BC, McDonald JH, Argyropoulos G, Destro-Bisol G, Froment A, Drousiotou A, Lefranc G, Helal AN, Loiselet J, Tishkoff SA (2002) Evidence for balancing selection from nucleotide sequence analyses of human *G6PD*. Am J Hum Genet 71:1112–1128

Wooding SP, Watkins WS, Bamshad MJ, Dunn DM, Weiss RB, Jorde LB (2002) DNA sequence variation in a 3.7-kb noncoding sequence 5' of the *CYP1A2* gene: implications for human population history and natural selection. Am J Hum Genet 71:528–542

Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118:63–76