

The Implications of Intergenic Polymorphism for Major Histocompatibility Complex Evolution

Colm O'hUigin,* Yoko Satta,† Anja Hausmann,* Roger L. Dawkins‡ and Jan Klein*

*Max-Planck-Institut für Biologie, Abteilung Immunogenetik, D-72076 Tübingen, Germany, †Department for Biosystems Science, Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan and ‡The Centre for Molecular Immunology and Instrumentation, The University of Western Australia, Perth, Australia

Manuscript received March 21, 2000
Accepted for publication May 30, 2000

ABSTRACT

A systematic survey of six intergenic regions flanking the human *HLA-B* locus in eight haplotypes reveals the regions to be up to 20 times more polymorphic than the reported average degree of human neutral polymorphism. Furthermore, the extent of polymorphism is directly related to the proximity to the *HLA-B* locus. Apparently linkage to *HLA-B* locus alleles, which are under balancing selection, maintains the neutral polymorphism of adjacent regions. For these linked polymorphisms to persist, recombination in the 200-kb interval from *HLA-B* to *TNF* must occur at a low frequency. The high degree of polymorphism found distal to *HLA-B* suggests that recombination is uncommon on both sides of the *HLA-B* locus. The least-squares estimate is 0.15% per megabase with an estimated range from 0.02 to 0.54%. These findings place strong restrictions on possible recombinational mechanisms for the generation of diversity at the *HLA-B*.

HUMANS are among the species for which sequence polymorphisms have been best characterized. A compilation of genetic data by LI and SADLER (1991) revealed that the genetic diversity in human populations averaged 0.08%, a value subsequently confirmed by TAKAHATA and SATTA (1997). There is considerable variation in the degree of polymorphism from locus to locus, however, presumably because of differences in underlying mutational mechanisms and selective influences that may promote or reduce polymorphism.

Selection can directly affect the amount of polymorphism at loci under its influence. Most commonly, negative selection reduces diversity by eliminating deleterious variants (KIMURA 1983). Selective sweeps resulting from positive selection also reduce the degree of polymorphism, due to the fixation of favored alleles. In contrast to these two mechanisms, balancing selection promotes polymorphism through the maintenance of variant allelic forms. In each case, the influence of selection on polymorphism should not be restricted to the locus under selection. Regions adjacent to such a locus should show a similar influence of selection, even if they themselves are selectively neutral (FELSENSTEIN 1974). The effect is predicted and modeled for positive (MAYNARD SMITH and HAIGH 1974; KAPLAN *et al.* 1989; BEGUN and AQUADRO 1992), negative or background (CHARLESWORTH *et al.* 1993; NORDBORG *et al.* 1996),

and balancing (HUDSON and KAPLAN 1988; TAKAHATA and SATTA 1998a) selection. Because the degree of influence depends on the strength of the linkage, a comparison of polymorphism in the linked regions to that in regions under selection gives a measure of the recombination frequency.

A spectacular example of the effects of balancing selection is provided by the major histocompatibility complex (*Mhc*) loci, the most polymorphic loci in humans (KLEIN *et al.* 1993). [There are ~200 allelic variants known at each of the *HLA-B* and *HLA-DRB1* loci (BODMER *et al.* 1999).] An unusual feature of the *Mhc* polymorphism is the sharing of sequence motifs by otherwise distinct alleles. Many mechanisms that involve genetic exchanges have been proposed for the diversification of *Mhc* loci (ANDERSSON *et al.* 1991; KUHNER and PETERSON 1992; ANDERSSON and MIKKO 1995; McDEVITT 1995). A common theme of several of the proposed mechanisms is the implication of recombination as a vehicle for promoting the dispersal of diversity into various alleles (SHE *et al.* 1991; McADAM *et al.* 1994; SATTA *et al.* 1998). Putative recombinogenic sequences, related to the bacteriophage λ chi sequence, have been identified at class II loci (GYLLENSTEN *et al.* 1991). It has been proposed that such chi-like elements promote recombination to higher than normal levels (GYLLENSTEN *et al.* 1991). As an alternative to recombinogenic mechanisms, it has been proposed that the gradual accumulation of mutations over long evolutionary periods, coupled with convergence of certain key residues in the various alleles (GUSTAFSSON *et al.* 1991; GUSTAFSSON and ANDERSSON 1994; KLEIN and O'HUIGIN 1995;

Corresponding author: Colm O'hUigin, Max-Planck-Institut für Biologie, Abt. Immunogenetik, Corrensstr. 42, 72076 Tübingen, Germany. E-mail: colm@tuebingen.mpg.de

O'HUIGIN 1995; KRIENER *et al.* 2000), could generate the observed diversity.

It is to be expected that high recombination rates will decouple the evolutionary history of loci under selection from that of adjacent regions. Two lines of evidence have suggested that such decoupling does not occur. The first of these involves conservation. Strong similarities between *DRB* haplotypes in humans and chimpanzees extending over several hundred kilobases have been reported by KLEIN *et al.* (1991). Conserved human class I region haplotype blocks extending again over several hundred kilobases have been described by DAWKINS *et al.* (1991). The results of both of these studies have been interpreted as providing evidence for the maintenance of ancient ancestral haplotypes under recombinationally frozen conditions. The second line of evidence was provided by the study of *Mhc* polymorphism. Analysis of an intergenic region called *CLI*, at least 100 kb proximal to *HLA-B*, showed a high degree of sequence variability in three haplotypes, apparently exceeding the expected neutral polymorphic levels (ABRAHAM *et al.* 1992). Neither of these lines of evidence appears easily compatible with enhanced recombination in the *Mhc* region. In the first case, excessive recombinogenic mechanisms would be expected to scramble the haplotypes during the 5 million years since the divergence of humans and chimpanzees. In the second case, the adjacent regions should not hitchhike with *HLA-B* and should therefore show no more polymorphism than any unlinked region. Nevertheless, neither of these lines of evidence is conclusive. Common haplotypes found in chimpanzees and humans may be the result of selection favoring particular combinations that can, however, recombine freely. The high degree of polymorphism in the *CLI* region may be caused, not by *HLA-B*, but by any nearby locus under balancing selection. Furthermore, it is not clear whether the three selected haplotypes are representative of polymorphisms in *CLI* and neighboring regions and whether different polymorphisms are associated with different *HLA-B* alleles. Although some studies have provided further evidence of polymorphism in noncoding regions of class I (GAUDIERY *et al.* 1997) and class II (HORTON *et al.* 1998) loci, no systematic survey relating polymorphism and linkage has been undertaken.

To decide whether the high polymorphic levels in the *CLI* locus are due to linkage to *HLA-B*, a closer examination of polymorphisms in the *HLA-B* and *TNF* interval is necessary in a wider variety of haplotypes. A higher degree of linkage to a locus under balancing selection would be expected to be associated with a higher polymorphic level, whereas lower linkage would lead to lower genetic diversity. By examining the degree of polymorphism in this interval, it should therefore be possible to determine whether it is selection at the *HLA-B* or some other locus that is maintaining the adjacent neutral polymorphisms. It should also be possible to

estimate the degree of recombination in the proximity of *HLA-B*. The recombination frequency thus estimated can be used to set limits on the amount of recombination that can be involved in generating *HLA-B* diversity. We have therefore undertaken a survey of polymorphisms in eight human *HLA-B* haplotypes at five loci in the *TNF* to *HLA-B* interval, as well as one locus distal to *HLA-B*.

MATERIALS AND METHODS

Source of DNA: Human DNA was extracted from eight *HLA-B* homozygous EBV cell lines obtained from the tissue typing service of the United Kingdom Transplantation Service, Bristol, England. The cell lines were Boleth (*HLA-B*1501*), Brip (*HLA-B*1517*), DBB (*HLA-B*5701*), DKB (*HLA-B*4001*), EA (*HLA-B*0702*), EHM (*HLA-B*3501*), Madura (*HLA-B*4001*), and Renchr. In addition, DNA from the chimpanzee cell line Yvonne (provided by Dr. R. Bontrop, TNO Health Institute, Rijswijk, The Netherlands) and from the peripheral blood leukocytes of the gorilla, Jimmy (Miami Zoo, Miami) was used. All DNA was extracted by using the standard phenol-chloroform procedure.

Polymerase chain reaction (PCR): PCR amplification was used to obtain *Mhc* segments from genomic DNA. Table 1 shows the primers used and the sizes and locations of the amplified fragments. Unless otherwise indicated, all PCRs were carried out in a volume of 50 μ l containing 100 μ M each of dATP, dCTP, dGTP, and dTTP, 5 μ l of 10 \times reaction buffer, and 2 units of *Taq* polymerase enzyme from Amersham Pharmacia Biotech, as well as 1 μ M of each primer and 200 ng of genomic DNA. The PCR program consisted of 35 cycles of 1 min denaturation followed by 30 sec primer annealing at 55 $^{\circ}$ and 1 min extension at 72 $^{\circ}$. Immediately preceding and following these 35 cycles, a 1-min denaturation step at 94 $^{\circ}$ and a 10-min extension step at 72 $^{\circ}$ were included, respectively. The amplified fragments were separated from primers on 1% low-melting-point agarose gels, cut out, purified by using the QIAEX II procedure, and cloned into pUC18 vectors by using the Sureclone kit (Amersham Pharmacia Biotech). Following transformation into competent *Escherichia coli* cells, DNA was isolated from resultant white colonies and sequenced on the A.L.F. sequencer (Amersham Pharmacia Biotech) using universal and reverse primers of the A.L.F. sequencing kit. At least two independent clones were used to verify each sequence.

Data analysis: Sequences were aligned by eye. Additional sequences for the *TNF* and *HLA-B* regions (MIZUKI *et al.* 1997; GUILLAUMEUX *et al.* 1998; SHIINA *et al.* 1998) and of *HLA-B* introns (CEREB *et al.* 1996) were obtained from database entries. Aligned sequences were used to construct phylogenetic trees by the neighbor-joining subroutine of the MEGA package (KUMAR *et al.* 1993) following distance estimation by Kimura's two-parameter method (KIMURA 1980). Polymorphism was analyzed and nucleotide diversities were estimated by using the DNASP package (version 2.2) of ROZAS and ROZAS (1997).

RESULTS

General: To simplify the interpretation, we identified nonfunctional intergenic regions for amplification. Presumably such regions evolve without the confounding influence of phenotypic selection. Our initial amplifications were made in the absence of genomic sequence data by using linker amplification to extend the se-

TABLE 1

Locations of amplified segments on large genomic sequences taken from the DNA database entries of SHIINA *et al.* (1998), MIZUKI *et al.* (1997), and GUILLAUMEUX *et al.* (1998)

Segment	Location (accession no.)	PCR fragment size (bp)	Primers ^a
<i>TNF</i>	15439–16103 (Z15026)	665	CCCCTTTGAGGCACTTGTTT TAGGACTACAGGCGCATACC
<i>CL1</i>	71612–73049 (AB000882)	1445	TTACAAAAGGGGTTGGCTGTAA AAGGACACTCAGGGAATAAATGG
<i>CL2</i>	17862–18311 (D83543)	449	CCTTTTGGAGATGTCTTTTCAGGT TAGCGGAGGAAAGGCAAGGACACT
<i>FGF</i>	26712–27620 (D83543)	888	TAACAGAAACAAGCTAAAGGTCAC CCATGAGGGAATATCAAGAGTTT
<i>DHFR</i>	46150–46733 (D83543)	583	CCCAGTTTTCCAACCTCTATTTT TTCTGCCAGTTCAGTTGAATGG
<i>HLA-B</i>	58164–60850 (D83543)	805	TGTTGGTCCCAATTGTCTCCCTC GAACSGCCTCTGYGGGGAGAAGCA
<i>DIS</i>	71670–72266 (D83543)	596	GGGGACGGTGCTAAAGGATT GGGCTTCATTCAATTCCTGA

^a Forward primer sequence given first; primers given in 5'–3' orientation.

quence from known intergenic segments (GERAGHTY *et al.* 1992; LEELAYUWAT *et al.* 1992) and by amplifying it into adjacent DNA segments. The subsequent sequencing of the *TNF* to *HLA-B* region (MIZUKI *et al.* 1997; GUILLAUMEUX *et al.* 1998; SHIINA *et al.* 1998) simplified the task of identifying and choosing regions and of evaluating those already chosen.

Several duplications in the *HLA* class I region have been revealed by large-scale sequencing and they include some of the initially chosen segments for amplifications (notably, *CL1*, *FGF*, and *DIST*). Two factors enabled us to identify paralogous PCR amplification products when they occurred. First, where two or more different sequences were obtained from a putatively homozygous source, it is likely that paralogy is involved. Second, phylogenetic analysis revealed the clustering of paralogous segments into a clade separate from the targeted product. Since the genomic sequences are known in the Boleth haplotype (MIZUKI *et al.* 1997), it was possible to assign definitive locations to the fragments amplified.

Much of the intergenic region is nearly identical in the fragments we sequenced and so the alignment is straightforward. There are many single substitutions, however, leading to a moderate degree of polymorphism. When an orthologous region was amplifiable, we used the chimpanzee and gorilla sequences to determine the likely direction of substitutions for each re-

gion. Figure 1 shows the approximate positions of the segments we amplified, as well as the positions of genes in the *TNF* to *HLA-B* interval. The number of nucleotides compared excluding gaps was 5664 bp, of which 263 sites are segregating. Of the segregating sites, 132 are phylogenetically informative and the remaining 131 are singletons. The description of the individual regions follows.

***TNF*:** The amplified segment of 665 bp is ~1.5 kb telomeric from the stop codon of the *TNF* gene and some 200 kb distant from *HLA-B*. It contains part of an *Alu* element but shows no evidence of divergent duplicate copies. Single substitutions were found in EHM (A to T at site 284 and T to C at 461), in Renchr (T to A at 436), in EA (C to T at 440), and in DKB (A to G at 672). There are two published sequences encompassing this region. One of them (GUILLAUMEUX *et al.* 1998) is identical with our Boleth, Brip, DBB, and Madura sequences. The other (IRIS *et al.* 1993, accession no. Z15026), reportedly from Boleth, shows several differences from the Boleth sequence reported here and the sequence of GUILLAUMEUX *et al.* (1998). It differs from all sequences presented here by deletion of C at site 330, insertion of G at 386, and single substitutions at sites 373, 374, 375, 430, 433, 436, 448, and 453. None of the substitutions or indels were found in any of the sequences we examined, including Boleth. It was therefore excluded from our polymorphism estimates. In the

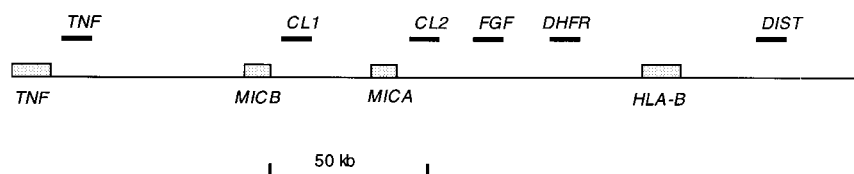


FIGURE 1.—Map of the *TNF* to *HLA-B* interval showing the approximate locations of the amplified fragments used in this study. Open boxes indicate the location of known genes. Thick lines indicate the location of PCR fragments.

set used, there are a total of five singleton polymorphic sites (Figure 2), giving a nucleotide diversity of 0.18% (SD 0.06%) at this locus.

CL1: This region was originally identified as part of a duplicated segment on a restriction enzyme fragment of 6 kb by LEELAYUWAT *et al.* (1992). It lies ~75 kb distal from *TNF* and 130 kb proximal to *HLA-B*. Several primers were designed to amplify a segment of *CL1* of up to 1.46 kb encompassing a single *Alu* element (sites 696–1034). None of a variety of PCR primers could amplify the *CL1* segment from chimpanzee or gorilla, *CL1* being presumably deleted in these two species. An *Alu*-less *CL2* fragment could be obtained instead. Since *CL1* and *CL2* are related by duplication that occurred some 30 mya (LEELAYUWAT *et al.* 1992), the *CL2* sequences can be used as outgroups to determine the substitutions that occurred in the *CL1* sequences. Both *CL2* and *CL1* contain sequences related to the LTR8 repetitive element from chromosomes 7 and 14 (accession nos. AC002508 and AC005857).

Although only twice the size of the *TNF* segment, *CL1* contains 50 polymorphic sites (Figure 2) compared with 5 found at *TNF*. The estimated nucleotide diversity of 1.24% (SD 0.19%) is more than six times greater than that estimated for *TNF*. Nineteen of the substitutions at the locus are informative. To determine the extent of incompatibility among the informative nucleotide sites in this region, an incompatibility matrix was produced (TAKAHATA and SATTA 1998b). Of the 1225 pairs of informative sites, 103 pairs in two phylogenetically distinct blocks were incompatible. In the first 700 bp, substitutions were scattered and no pattern of relationships between sequences emerged. By contrast, at the 3' end of the *CL1* fragment, the Brip and Madura sequences shared six sites and most sites were compatible (Figure 2).

CL2: The *CL1* region and some 10 kb of its flanks were duplicated in an event estimated to have occurred some 30 mya (LEELAYUWAT *et al.* 1992), giving rise to a related segment called *CL2*. This segment lies 150 kb distal to *TNF* and ~50 kb proximal to *HLA-B*. The amplified fragment is 449 bp in length and is homologous to part of the *CL1* region. By positioning one primer across the site of the *Alu* insertion in *CL1*, we were able to preferentially amplify *CL2*, which does not contain the *Alu* element. A total of 15 polymorphic sites were found in humans, giving a nucleotide diversity of 1.52% (SD 0.27%). Of these, two sites occurred in single haplotypes and the remaining 13 were shared by at least two haplotypes (Figure 2). At the 5' end (~290 bp), the sequences form at least two distinct groups on phylogenetic trees, with EHM, EA, Madura, and Boleth falling in one group (with a total nucleotide diversity of 0.48%; SD 0.25%). Three other sequences (DBB, DKB and Brip) are identical and share 10 substitutions that distinguish them from the former group. However, at the 3' end of the segment, the Boleth sequence clusters with

Brip, DKB, and DBB, suggesting at least one recombination breakpoint in the region.

Despite the higher diversity at *CL2* compared to *CL1*, no human sequence is more closely related to either the gorilla or the chimpanzee sequences than to other human sequences; in other words, we found no evidence for *trans-species* evolution of *CL2*. The high level of nucleotide diversity (1.5%) could indicate that the comparisons made are of a paralogous type. However, database searches could not identify any paralogs of *CL2*.

FGF: The *FGF* region is adjacent to a fibroblast growth factor-like pseudogene. We amplified a segment of 888 bp that lies 175 kb distal from *TNF* and 31 kb proximal to *HLA-B*. The segment was obtained as follows: a primer based on the sequence of LEELAYUWAT *et al.* (1996) was used in combination with a primer specific for a linker added to a restriction fragment that contained the LEELAYUWAT *et al.* (1996) sequence. The amplification product was sequenced and primers were then designed for the amplification of the *FGF* segment on the basis of this sequence. The primers were, however, not specific for this region; they also amplified at least one paralogous segment 50 kb proximal to the site described by LEELAYUWAT *et al.* (1996). The paralogous segment, labeled D83543-P in Figure 2, could be identified on the basis of sequences from SHIINA *et al.* (1998). A second paralogous segment occurred at an unidentified position and only this segment could be obtained from the DBB cell line (Figure 2). The fixed sequence differences enabled us, however, to distinguish both paralogous amplification products, which group in distinct clades on phylogenetic trees. By these means we could identify the sequences derived specifically from the *FGF* region. After the exclusion of the DBB paralog, nine polymorphic sites were found in this segment, eight of which were singletons. There was no incompatibility among the substitutions. The nucleotide diversity of the seven orthologous sequences was 0.32% (SD 0.07%).

DHFR: The *DHFR* region lies adjacent to a *DHFR* pseudogene identified by MIZUKI *et al.* (1997) and is 12 kb proximal to *HLA-B*. The amplified segment is 583 bp in length. In the sequences we analyzed, 15 polymorphic sites were found, giving a nucleotide diversity of 1.02% (SD 0.15%) for the eight human DNAs. Nine sites were phylogenetically informative and all sites are mutually compatible.

HLA-B: This is the locus under balancing selection that may be driving the polymorphism of adjacent regions. The *HLA-B* cDNA sequences for alleles found in seven of the eight cell lines have been characterized (PRASAD and YANG 1996) and the *HLA-B* intron sequence is available for Boleth (MIZUKI *et al.* 1997), DBB, and Madura (CEREB *et al.* 1996). To supplement the available data, we sequenced a segment encompassing exon 2 to exon 3 from each of the eight human sources. This collection enabled us to compare both the intronic and synonymous variation. In the intron 2 segments,

TNF		1	11					
CONSENSUS ==>		TGAGGTTCT	AATGCA					
Boleth		-----	-----					
Brip		-----	-----					
DBB		-----	-----					
DKB		-----	-----G					
EA		-----T-	-----					
EHM		--T-----C	-----					
Madura		-----	-----					
Renchr		-----A-	-----					
Gogo		CA-AA-----	GCCA--					
Patr		----AC--G-	-C--G-					
CL1		1	11	21	31	41		
CONSENSUS ==>		GAAGAGTCGC	GCTCCAGCGC	CTCTGAGTCC	CATGCAATFC	ATACGTAAGT		
Boleth		C-----	---T-TC---	-C-----	-----CT	-----		
Brip		-G-AG--A-A	---TC---	-C-----	---C-----G	--TGCC-GA-		
DBB		-GGAG-----	---GT-AT-	---AG---G	--G-----	-A-----		
DKB		-----	C-----C-	-C-----C-	-----T	-----G-		
EA		---A---T-	--C---CA-T	T-T---C-	-----	-A-----		
EHM		-----AA-	--C--T-A-	---C-----	-----T	T-----		
Madura		-----A-	-A---C---	---C---G-	AG-CTGGC-G	--TGCC-GAG		
Renchr		---AG-A---	-----	-----C-	-----	-A-----GA-		
CL2		1	11	21	31			
CONSENSUS ==>		GGCACGTTCA	TCGCACCTAT	GCTCAGTGGC	GTCTGA			
Boleth		-----*-TG	-----*-C	-T-----A-	---CT-			
Brip		A---T**--	CT-----**	C--T-A--A-	---CT-			
DBB		A---T**--	CT-----**	C--T-A--A-	---CT-			
DKB		A---T**--	CT-----**	C--T-A--A-	---CT-			
EA		-----*-TG	-----*	-T-----	-----			
EHM		-----*-TG	-----*	-T-----	-----			
Madura		-----*-TG	-----*	-T-----	-----			
Renchr		-----*-G-	-T-----	-----	-----			
Gogo		--A*****	--ATCG---	C-A-G-C-*	**T--G			
Patr		-*-----	---CG---	C-A-G-CT-*	**T--G			
FGF		1	11	21	31	41	51	61
CONSENSUS ==>		TCTGTGTTAGC	ATCGCGTTCG	ATTTCGTAAC	TTCTTGAACC	GAAAAGCGGG	GCATACCTGC	AA
D83543-P		C--CA---A-	--T-AAG-GT	GC*TACGC-T	CAT-----T-	TG****AAC-	A-G---CT-	-G
Boleth		-----	-----	-----	-----*	-----	-----	---
Brip		-----	T-----	-----	-----*	-----	-----	---
DBB		---AC---	-C-A-----	-----G-	---CCAC*-T	-----CA	A-G---CTT	G-
DKB		-----	-----	-----	-----*	-----	-A-----	---
EA		-----C-G	-----C-	-----	-----*	-----	-----	---
EHM		-T---C-	-----	-----	-----*	-----	-----	---
Madura		-----	-----	-----	-----*	-----	-A-----	---
Renchr		--C---G--	-----	-----	-----*	-----	-----	---
DHFR		1	11					
CONSENSUS ==>		AACCACCCTC	AGAACAC					
Boleth		-----	--G-ACT					
Brip		-----T-	--GGACT					
DBB		-----	--G-ACT					
DKB		---T-T--CT	-A-----					
EA		**G-GTT--	-----					
EHM		-----	G-----					
Madura		---T-T--CT	-A-----					
Renchr		-----	-----					
DIST		1	11	21	31	41	51	61
CONSENSUS ==>		TGTATCTGCA	CAAGAACTC	CGCTTCTACC	AGTAGTGCTA	CTTCCGTCG	ACTTACCCTA	GGTCA
Boleth		-----	-----	T-T-T---	-----*	-----	-----*	-----
Brip		-----	-----	T-T-C--A	-----T-	-----TT-	---C-----*	A-----
DBB		--C*****	***GG-T-	-----T-	-----T-C-	GCC-G--GTA	TTG-T-G--G	-A-TG
DKB		-----	-----	T-T-T---	-----	-----	-----*	A-----
EA		C-----	-----	T-T-C--A	-----T-	-----TT-	-----G-T*	-----
EHM		-----	-----	T-T-T---	-----	-----C	-----*	A-----
Renchr		-----	-----	*C--T-	-----T-C-	GCC-G--GTA	TTG-T-G-*G	-A-TG
Gogo		-A-----	-----C-	-AT---G-	CA-----	-----	-----*	--C--
Patr		-A-----	-----C-	-A-----	-----*	-----	-----*	--C--

FIGURE 2.—Variant sites found in the complete alignments of six amplified segments adjacent to *HLA-B*. The label in boldface type gives the segment name. Paralogous human sequences (indicated by the accession code and the extension -P), as well as sequences of chimpanzee (*Patr*) and gorilla (*Gogo*), are shown where available. A consensus for each segment was determined by simple majority. A dash indicates identity with the consensus; an asterisk indicates a deletion. When available, known paralogs are shown immediately below the consensus sequences and other human sequences are grouped according to their similarities.

the nucleotide diversity reaches 2.10% (SD 0.40%) for the haplotypes surveyed here. The variation at synonymous sites is higher at 3.36% (SD 0.36%) in the same haplotypes.

DIST: This segment lies 11 kb distal from *HLA-B* and 80 kb proximal to *HLA-C*. The primers we used were not specific for this site and amplified at least one other segment that presumably arose with the duplication of the ancestral *HLA-B* and *HLA-C* loci. The amplified fragment was 596 bp long and contained 41 variant sites (Figure 2). Only paralogous fragments could be obtained from the Renchr and DBB haplotypes. After the exclusion of the paralogous segments, the estimated nucleotide diversity in the remaining six haplotypes was 1.56% (SD 0.38%).

Other loci: In addition to the data we have obtained for these six loci from the eight human cell lines, other surveys have been conducted of coding sequences in the *HLA-B* to *TNF* interval. Most prominent among these is the search for polymorphism at the class I-related *MICA* and *MICB* loci (BAHRAM *et al.* 1994; FODIL *et al.* 1996; ANDO *et al.* 1997; PELLET *et al.* 1997; VISSER *et al.* 1998; YAO *et al.* 1999). These surveys have concentrated on determining whether the sites of polymorphisms can be related to molecular function. However, the patterns found are quite unlike those of class I genes in which polymorphism is concentrated on sites specifying the peptide-binding region (PBR). In both *MICA* and *MICB*, little variation is found at putative PBR-specifying sites; it is instead concentrated in the fourth exon (YAO *et al.* 1999), which is not variable in classical class I genes. The estimated nucleotide diversities of 20 near-complete *MICA* and 13 *MICB* coding sequences are 1.02 and 0.23%, respectively, but comparison with other diversity estimates is difficult since the haplotypes have not been systematically selected. The patterns of variability in *MIC* genes are consistent with the hypothesis that these genes represent nonclassical class I genes. There is evidence that *MICA* may be involved in induction of antitumor NK and T cell responses (BAUER *et al.* 1999). On the other hand, some *MICA* alleles contain in-frame stop codons, suggesting that functionality of this locus may

not be critical. Rates of synonymous and nonsynonymous substitution per site have been approximately equal since the divergence of *MICA* and *MICB* (~7.5% synonymous substitutions per site *vs.* 5.5% nonsynonymous substitutions). In contrast to *HLA-B*, *MIC* polymorphisms are not *trans*-specific between apes and humans (CATTLEY *et al.* 1999). There is evidence of a linkage disequilibrium between the *MIC* and *HLA-B* genes (YAO *et al.* 1999) and thus their polymorphism fits a picture of being generated, as in noncoding regions, by virtue of their linkage to *HLA-B*. In this view the *MIC* genes may be functional and polymorphic, but selection at *MIC* loci does not produce polymorphism in the manner found for *HLA-B*.

Table 2 illustrates the extent of nucleotide diversity found for the regions we sequenced. The lowest levels of diversity are found in segments most distant from *HLA-B*, close to the *TNF* locus. The diversity found at *TNF* is nevertheless approximately double the average levels found by LI and SADLER (1991) for unselected sites of human gene sequences. The nucleotide diversity increases with increasing proximity to *HLA-B*, although the increase is not monotonic. It is unclear what effect any functional restrictions on *MICA* and *MICB* may have on the polymorphism in regions adjacent to those loci and whether these loci may be a cause of the nonmonotonic relationship. The absence of a simple relationship between diversity and distance from a locus under selection is also seen in the analysis of two haplotypes in the class I region (SATTA and TAKAHATA 2000). The highest nucleotide diversity is found at *HLA-B* itself, whether measured at the synonymous sites or in introns.

Recombination frequency: The most likely factor driving the high levels of polymorphism in the six adjacent regions reported here is their linkage to *HLA-B*. Since the degree of polymorphism can be related to recombination frequency, it is of some interest to determine what recombination frequency can account for the observed degree of polymorphism. By using the population genetics models of TAKAHATA and SATTA (1998a) and SATTA and TAKAHATA (2000), we can attempt to fit a recombination rate to the observed data (Figure

TABLE 2

Nucleotide diversities (π) \pm standard errors (SE) of orthologous segments and the divergences (d) from gorilla or chimpanzee estimated by the Kimura two-parameter method

Segment	π (%) \pm SE	d from chimpanzee	d from gorilla	d between chimpanzee and gorilla
<i>TNF</i>	0.18 \pm 0.06	0.81 \pm 0.25	1.25 \pm 0.46	1.31 \pm 0.44
<i>CL1</i>	1.24 \pm 0.19	ND	ND	ND
<i>CL2</i>	1.52 \pm 0.27	3.47 \pm 1.01	3.98 \pm 1.08	0.98 \pm 0.49
<i>FGF</i>	0.32 \pm 0.07	ND	ND	ND
<i>DHFR</i>	1.02 \pm 0.15	ND	ND	ND
<i>HLA-B</i>	3.36 \pm 0.36	ND	ND	ND
<i>DIS</i>	1.56 \pm 0.38	1.76 \pm 0.24	2.25 \pm 0.72	0.68 \pm 0.34

ND, no orthologous sequence was obtained.

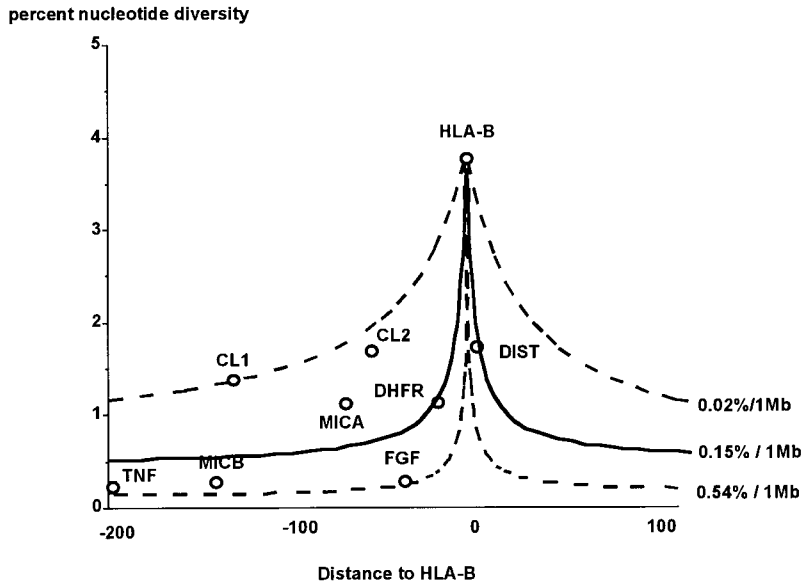


FIGURE 3.—The relationship between the percentage of nucleotide diversity (ordinate) found for six amplified intergenic segments, as well as exons 1–7 of *HLA-B* in the same haplotype, plotted as a function of distance from *HLA-B*. The nucleotide diversity found in a range of *MICA* and *MICB* haplotypes is also shown. The distances are given in kilobases with negative and positive numbers used for locations proximal to and distal from *HLA-B*, respectively. Lines show estimates of expected diversity at particular recombination frequencies following the model of TAKAHATA and SATTA (1998a). The solid lines show the least-squares estimate of recombination rate (0.15% per megabase per generation) and the dashed lines show the rates required to incorporate the maxima (0.02% per megabase) and minima (0.54% per megabase). The mutation rate is assumed to be 10^{-9} per site per year and the selection intensity is 1%.

3). The least c-squares estimate gives a rate of 0.15% per megabase (Figure 3), which is one-sixth of the average rate found over the entire human genome (in humans estimated at 1% per megabase per generation: see VOGEL and MOTULSKY 1997; TAKAHATA and SATTA 1998a). The maximum and minimum diversity values found can be encompassed by recombination frequencies of 0.02 and 0.54%, respectively. The curve of the estimate is monotonic and does not account for variations found in the vicinity of the *MIC* loci. However, these variations affect primarily the shape of the recombination curve rather than its magnitude. The present estimate agrees with another estimate for this region from SATTA and TAKAHATA (2000) that is based on a comparison of the sequences of GUILLAUMEUX *et al.* (1998) and SHIINA *et al.* (1998) extending throughout the *HLA-B* and *HLA-C* regions. It therefore appears that recombination frequency in proximity to *HLA-B* is not elevated compared to other genomic regions, but rather reduced.

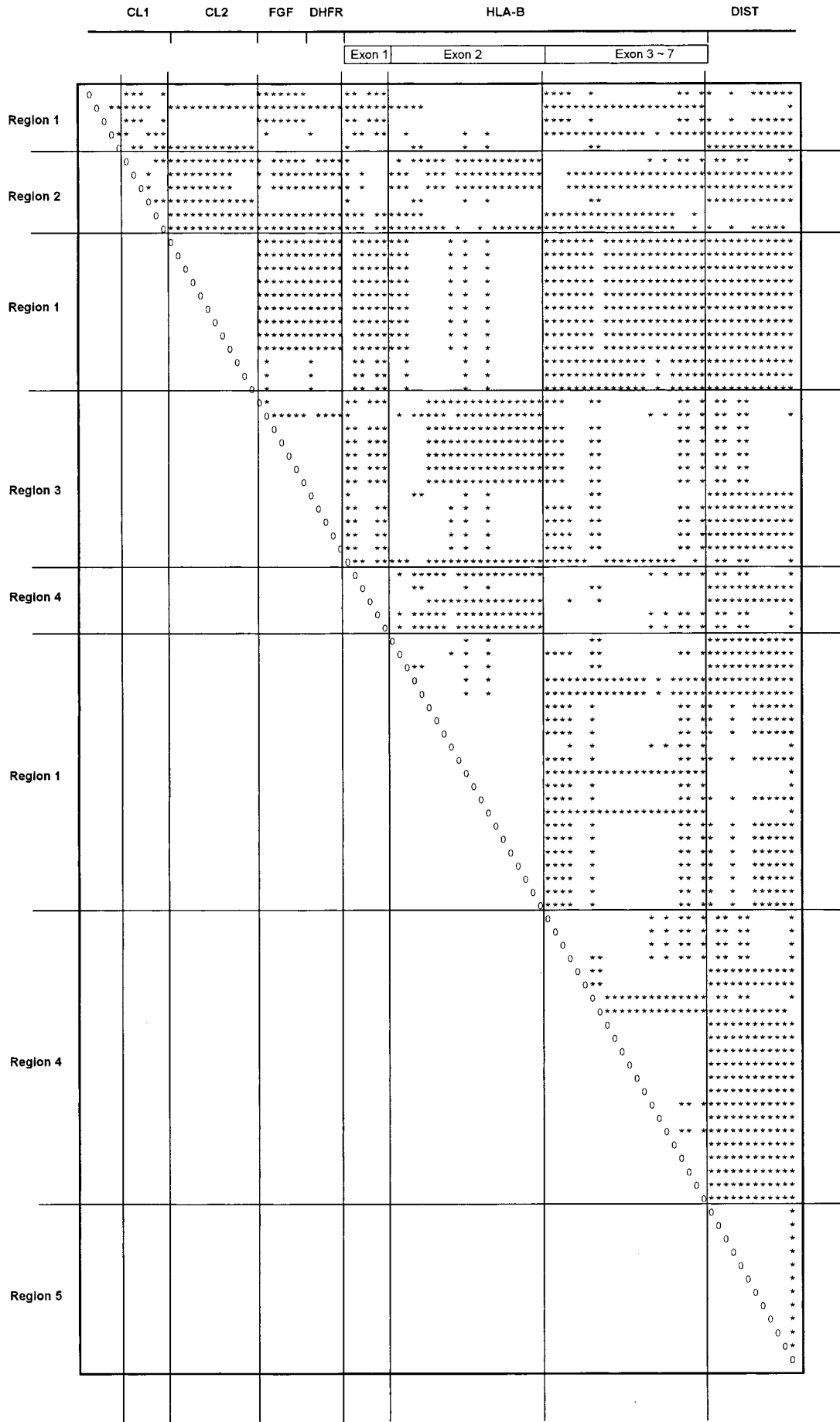
To examine the extent of incompatibility in this region, an incompatibility matrix was constructed (Figure 4). Among all the pairs of informative sites (9870 pairs), including those in *HLA-B*, 3769 pairs are incompatible. From the diagram, there appear to be at least five phylogenetically distinct blocks (Figure 4). Block 1 is composed of the 5'-most segment (5' region of *CL1*, the entirety of *CL2*, and exon 2 of *HLA-B*). These three regions are mutually compatible and support a single phylogeny. Block 2 contains the only the 3' part of *CL1*. Block 3 is composed of *FGF* and *DHFR*. Block 4 is formed by exon 1 as well as exons 3–7 of *HLA-B* and block 5 by *DIST*. We reconstructed a phylogeny for each block. Although the number of informative sites for each block is not large, most clusters in the phylogenies are statistically supported. Then we counted the minimum num-

ber of branch switching by a modified method of ROBINSON and FOULDS (1981) for all blocks to support a single phylogeny (not shown). The minimum number was 18. Taking this number as a minimum number of recombination events in the 200-kb region during the past 17 million years (this date is based on the average synonymous divergence of eight sampled *HLA-B* exons, 0.034/site, and the synonymous substitution rate of 10^{-9} per site per year), the recombination frequency is calculated as 1.1/200 kb/million years. Assuming 20 years per generation in humans, the recombination rate per megabase per generation is 0.01%. This possible minimum rate, although roughly estimated, is not much different from the lowest limit of the least-squares estimate.

The polymorphisms at the *HLA-B* locus are thought to be *trans*-specific between humans and chimpanzees, having arisen at least 5 million years before present ($\sim 250,000$ generations ago, taking 20 years as generation time). For *TNF*, *CL2*, and *DIST*, orthologous segments from the chimpanzee and gorilla were obtained. In every case the nucleotide diversities found in humans were smaller than the nucleotide divergences between humans and apes. This observation suggests that, unlike at *HLA-B*, the polymorphisms at the adjacent linked loci have not evolved *trans*-specifically between humans and apes.

DISCUSSION

For all the loci we examined we were able to find sequence variation in pairwise comparisons of segments generally <1 kb in length. This is in contrast to findings from surveys of genomic variation, in which fewer than one polymorphism per 1 kb is found in humans (LI and SADLER 1991). At one extreme, the survey indicates that each of the eight haplotypes shows some unique



variation at the *DIST* or *FGF* loci. At the other extreme, the *TNF* locus shows variation that is only slightly above that expected from the Li and Sadler survey. These findings are in agreement with isolated observations of high levels of polymorphism in the neighborhood of the *HLA-B* locus (LEELAYUWAT *et al.* 1992).

The survey of polymorphisms proximate to *HLA-B* is limited by the choice of sample. The eight human cell lines were chosen because of their ready availability, their serologically determined *HLA* homozygosity, and the variety of *HLA-B* haplotypes they represent. There is an indication that the *HLA-B* genes of the EHM and Brip cell lines differ in a minor way only and that those of the Madura and DKB cell lines are identical (PRASAD and YANG 1996). It may be relevant to ask how representative are the cell lines used. This question can reasonably be answered only for the *HLA-B* locus, at which a large sample of variants is available. The nucleotide diversity of 23 *HLA-B* intron 2 sequences obtained by CEREB *et al.* (1996) is 1.68% (SD 0.28%). This value is slightly lower than that found in the eight chosen haplotypes (2.1%, SD 0.4%), suggesting that the alleles sampled here are more variable than those sequenced by CEREB *et al.* (1996). Nevertheless, the diversity at the *HLA-B* synonymous sites of these eight haplotypes ($3.4 \pm 0.4\%$) is not different from that ($3.4 \pm 1.1\%$) of 82 available *HLA-B* allele sequences (TAKAHATA and SATTA 1998a). It appears, therefore, that the present samples are representative of a large number of *HLA-B* alleles and the sampling cannot account for the 20-fold increase in diversity of the present haplotypes compared to that in the LI and SADLER (1991) survey.

A high degree of nucleotide diversity may be caused by high underlying mutation rates. However, studies by SATTA *et al.* (1993) on substitutions in the *Mhc* indicate that the mutation rate at *HLA-B* may be even lower than the human average. On the basis of the synonymous substitution rate of 10^{-9} (per site per year), the divergence time at the six intergenic segments adjacent to *HLA-B* ranges from 0.9 to 7.8 million years. Although chimpanzee or gorilla orthologs at *TNF*, *DHFR*, and *CL2* show no evidence of the *trans*-specific mode of evolution, the diversity of several pairs of alleles at these loci as well as other loci exceeds the average nucleotide divergence between humans and chimpanzees (1.75% averaged over 37 loci; TAKAHATA and SATTA 1997). Thus the polymorphisms observed in the intergenic segments adjacent to *HLA-B* have taken several millions of years to accumulate. When we explain this high extent of nucleotide diversity by mutation and random genetic drift, the required mutation rate or the effective popula-

tion size is too high or too large to be compatible with the nucleotide diversity at other neutral regions. It is, therefore, likely that this long persistence of the intergenic segments is due to their linkage to a locus under balancing selection.

The highest nucleotide diversities are observed at the *DIST* locus that is the closest to *HLA-B* (Figure 1, Table 2). In addition, the variation at the most polymorphic loci does not exceed that found at the *HLA-B* locus itself in the eight haplotypes; the locus has an intron 2 diversity of 2.1% (SD 0.4%) and a synonymous nucleotide diversity of 3.4% (SD 0.4%). The balancing selection driving polymorphism at *HLA-B* is of sufficient magnitude to explain the nucleotide diversity found at the adjacent neutral regions. Increasing levels of polymorphism correlate with proximity to *HLA-B*, although the relationship is not simple. The *MIC* genes fall into the pattern of high polymorphism levels in proximity to *HLA-B*, although selection at these loci may modify the pattern. Our evidence suggests that linkage to *HLA-B* is the main factor that maintains polymorphisms in adjacent regions. We find little or no evidence for any strong influence of another locus under balancing selection in the interval.

The extent of nucleotide diversity at these regions does not appear to be compatible with a high level of recombination at *HLA-B*. The long time (>5 million years) available for the accumulation of diversity at *HLA-B* should be sufficient to allow linked neutral diversity to be lost at moderate recombination rates (Figure 3). We have shown that the effects of selection can be seen near *HLA-B* and possibly extend as far as *TNF*, some 200 kb away from *HLA-B*, where the diversity is double that found at other neutral loci. The recombination frequency that gives a good fit to this diversity is 0.15% per megabase per generation, one-sixth of the human genomic average (1% per megabase). The high degree of linkage disequilibrium seen in neutral polymorphisms appears to favor the view that *HLA-B* evolves in a manner reflecting "frozen haplotypes" rather than the recombination hotspot models (KLEIN *et al.* 1991). This mode of evolution may be common among *Mhc* loci. Preliminary data on class II polymorphisms indicate that at the *HLA-DQB1* locus, as at the *HLA-B* locus, adjacent neutral regions show a higher than expected level of neutral polymorphism (HORTON *et al.* 1998). According to this view, the *Mhc* is evolving in a conserved fashion, without high levels of mutation or recombination, but simply by the accumulation of polymorphism over long time periods under balancing selection.

FIGURE 4.—Incompatibility matrix in the region encompassing the *HLA-B* locus. Locations of sequenced fragments are indicated on the top. On the side scale, the same number shows compatible regions. The asterisks in the matrix indicate that the pair of sites are phylogenetically incompatible.

LITERATURE CITED

- ABRAHAM, L. J., C. LEELAYUWAT, G. GRIMSLEY, M. A. DEGLI-ESPOSTI, A. MANN *et al.*, 1992 Sequence differences between HLA-B and TNF distinguish different MHC ancestral haplotypes. *Tissue Antigens* **39**: 117–121.
- ANDERSSON, L., and S. MIKKO, 1995 Generation of MHC class II diversity by intra- and intergenic recombination. *Immunol. Rev.* **143**: 5–12.
- ANDERSSON, L., K. GUSTAFSSON, A.-K. JONSSON and L. RASK, 1991 Concerted evolution in a segment of the first domain exon of polymorphic MHC class II b loci. *Immunogenetics* **33**: 235–242.
- ANDO, H., N. MIZUKI, M. OTA, M. YAMAZAKI, S. OHNO *et al.*, 1997 Allelic variants of the human MHC class I chain-related B gene (MICB). *Immunogenetics* **46**: 499–508.
- BAHRAM, S., M. BRESNAHAN, D. E. GERAGHTY and T. SPIES, 1994 A second lineage of mammalian major histocompatibility complex class I genes. *Proc. Natl. Acad. Sci. USA* **91**: 6259–6263.
- BAUER, S., V. GROH, J. WU, A. STEINLE, J. H. PHILLIPS *et al.*, 1999 Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285**: 727–729.
- BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- BODMER, J. G., S. G. E. MARSH, E. D. ALBERT, W. F. BODMER, R. E. BONTROP *et al.*, 1999 Nomenclature for factors of the HLA system, 1998. *Eur. J. Immunogenet.* **26**: 81–116.
- CATTLEY, S. K., N. LONGMAN, R. L. DAWKINS, S. GAUDIERI, J. K. KULSKI *et al.*, 1999 Phylogenetic analysis of primate MIC (PERB11) sequences suggests that the representation of the gene family differs in different primates: comparison of MIC (PERB11) and C4. *Eur. J. Immunogenet.* **26**: 233–238.
- CEREB, N., Y. KONG, S. LEE, P. MAYE and S. YANG, 1996 Nucleotide sequences of MHC class I introns 1, 2, and 3 in humans and intron 2 in nonhuman primates. *Tissue Antigens* **47**: 498–511.
- CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- DAWKINS, R. L., M. P. DEGLI-ESPOSTI, J. ABRAHAM, W. ZHANG and F. T. CHRISTIANSEN, 1991 Conservation versus polymorphism of the MHC in relation to transplantation, immune responses and autoimmune disease, pp. 391–402 in *Molecular Evolution of the Major Histocompatibility Complex*, edited by J. KLEIN and D. KLEIN. Springer-Verlag, Heidelberg, Germany.
- FELSENSTEIN, J., 1974 The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- FODIL, N., L. LALOUX, V. WANNER, P. PELLET, G. HAUPTMANN *et al.*, 1996 Allelic repertoire of the human MHC class I MICA gene. *Immunogenetics* **44**: 351–357.
- GAUDIERI, S., C. LEELAYUWAT, K. T. TAY, D. C. TOWNEND and R. L. DAWKINS, 1997 The major histocompatibility complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. *J. Mol. Evol.* **45**: 17–23.
- GERAGHTY, D. E., B. H. KOLLER, J. A. HANSEN and H. T. ORR, 1992 The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. *J. Immunol.* **149**: 1934–1946.
- GUILLAUMEUX, T., M. JANER, G. K. WONG, T. SPIES and D. E. GERAGHTY, 1998 The complete genomic sequence of 424,015 bp at the centromeric end of the HLA class I region: gene content and polymorphism. *Proc. Natl. Acad. Sci. USA* **95**: 9494–9499.
- GUSTAFSSON, K., and L. ANDERSSON, 1994 Structure and polymorphism of horse MHC class II *DRB* genes: convergent evolution in the antigen binding site. *Immunogenetics* **39**: 355–358.
- GUSTAFSSON, K., U. BRUNSBURG, S. SIGURDARDOTTIR and L. ANDERSSON, 1991 A phylogenetic investigation of MHC class II *DRB* genes reveals convergent evolution in the antigen binding site, pp. 119–130 in *Molecular Evolution of the Major Histocompatibility Complex*, edited by J. KLEIN and D. KLEIN. Springer-Verlag, Heidelberg, Germany.
- GYLLENSTEN, U., M. SUNDVALL and H. A. ERLICH, 1991 Allelic diversity is generated by intraexon sequence exchange at the *DRB1* locus of primates. *Proc. Natl. Acad. Sci. USA* **88**: 3686–3690.
- HORTON, R., D. NIBLETT, S. MILNE, S. PALMER, B. TUBBY *et al.*, 1998 Large-scale sequence comparisons reveal unusually high levels of variation in the HLA-DQB1 locus in the class II region of the human MHC. *J. Mol. Biol.* **282**: 71–97.
- HUDSON, R. R., and N. L. KAPLAN, 1988 The coalescent process in models with selection and recombination. *Genetics* **120**: 831–840.
- IRIS, F., L. BOUGUELERET, S. PRIEUR, D. CATERINA, G. PRIMAS *et al.*, 1993 Dense Alu clustering and a potential new member of the NFKappaB family within a 90 kilobase HLA class III segment. *Nat. Genet.* **3**: 137–145.
- KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The “hitchhiking effect” revisited. *Genetics* **123**: 887–899.
- KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**: 111–120.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- KLEIN, J., and C. O'HUIGIN, 1995 Class II B *Mhc* motifs in an evolutionary perspective. *Immunol. Rev.* **143**: 89–111.
- KLEIN, J., C. O'HUIGIN, M. KASAHARA, V. VINCEK, D. KLEIN *et al.*, 1991 Frozen haplotypes in *Mhc* evolution, pp. 261–286 in *Molecular Evolution of the Major Histocompatibility Complex*, edited by J. KLEIN and D. KLEIN. Springer-Verlag, Heidelberg, Germany.
- KLEIN, J., Y. SATTA, C. O'HUIGIN and N. TAKAHATA, 1993 The molecular descent of the major histocompatibility complex. *Annu. Rev. Immunol.* **11**: 269–295.
- KRIENER, K., C. O'HUIGIN, H. TICHY and J. KLEIN, 2000 Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics* **51**: 169–178.
- KUHNER, M. K., and M. J. PETERSON, 1992 Genetic exchange in the evolution of the human MHC class II loci. *Tissue Antigens* **39**: 209–215.
- KUMAR, S., K. TAMURA and M. NEI, 1993 MEGA: Molecular Evolutionary Genetic Analysis, Version 1.01. The Pennsylvania State University, University Park, PA.
- LEELAYUWAT, C., L. J. ABRAHAM, H. TABARIAS, F. T. CHRISTIANSEN and R. L. DAWKINS, 1992 Genomic organization of a polymorphic duplicated region centromeric of HLA-B. *Immunogenetics* **36**: 208–212.
- LEELAYUWAT, C., L. J. ABRAHAM, M. PINELLI, D. C. TOWNEND, A. F. WILKS *et al.*, 1996 The primate MHC contains sequences related to the fibroblast growth factor receptor gene family. *Tissue Antigens* **48**: 59–64.
- LI, W., and L. SADLER, 1991 Low nucleotide diversity in man. *Genetics* **129**: 513–523.
- MAYNARD SMITH, J., and J. HAIGH, 1974 The hitchhiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- MCADAM, S. N., J. E. BOYSON, X. LIU, T. L. GARBER, A. L. HUGHES *et al.*, 1994 A uniquely high level of recombination at the HLA-B locus. *Proc. Natl. Acad. Sci. USA* **91**: 5893–5897.
- MCDÉVITT, H. O., 1995 Evolution of MHC Class II allelic diversity. *Immunol. Rev.* **143**: 113–122.
- MIZUKI, N., H. ANDO, M. KIMURA, S. OHNO, S. MIYATA *et al.*, 1997 Nucleotide sequence analysis of the HLA class I region spanning the 237-kb segment around the HLA-B and -C genes. *Genomics* **42**: 55–66.
- NORDBORG, M., B. CHARLESWORTH and D. CHARLESWORTH, 1996 The effect of recombination on background selection. *Genet. Res.* **67**: 159–174.
- O'HUIGIN, C., 1995 Quantifying the degree of convergence in primate *Mhc-DRB* genes. *Immunol. Rev.* **143**: 123–140.
- PELLET, P., M. RENAUD, N. FODIL, L. LALOUX, H. INOKO *et al.*, 1997 Allelic repertoire of the human MICB gene. *Immunogenetics* **46**: 434–436.
- PRASAD, V. K., and S. Y. YANG, 1996 Allele assignment for HLA-A, -B, and -C genes to the Tenth International Histocompatibility Workshop cell lines. *Tissue Antigens* **47**: 538–546.
- ROBINSON, D. F., and L. R. FOULDS, 1981 Comparison of phylogenetic trees. *Math Biosci.* **53**: 131–147.
- ROZAS, J., and R. ROZAS, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Comput. Appl. Biosci.* **13**: 307–311.
- SATTA, Y., and N. TAKAHATA, 2000 Polymorphism in the HLA class I region, pp. 178–185 in *The Major Histocompatibility Complex: Evolution, Structure, and Function*, edited by M. KASAHARA. Springer, Tokyo.
- SATTA, Y., C. O'HUIGIN, N. TAKAHATA and J. KLEIN, 1993 The synon-

- ymous substitution rate of the major histocompatibility complex loci in primates. *Proc. Natl. Acad. Sci. USA* **90**: 7480–7484.
- SATTA, Y., Y. J. LI and N. TAKAHATA, 1998 The neutral theory and natural selection in the *HLA* region. *Front. Biosci.* **3**: 459–467.
- SHE, J.-X., S. A. BOEHME, T. W. WANG, F. BONHOMME and E. K. WAKELAND, 1991 Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. *Proc. Natl. Acad. Natl. Sci. USA* **88**: 453–457.
- SHIINA, T., G. TAMIYA, A. OKA, T. YAMAGATA, N. YAMAGATA *et al.*, 1998 Nucleotide sequencing analysis of the 146-kilobase segment around the *IkBL* and *MICA* genes at the centromeric end of the *HLA* class I region. *Genomics* **47**: 372–382.
- TAKAHATA, N., and Y. SATTA, 1997 Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. USA* **94**: 4811–4815.
- TAKAHATA, N., and Y. SATTA, 1998a Footprints of intragenic recombination at *HLA* loci. *Immunogenetics* **47**: 430–441.
- TAKAHATA, N., and Y. SATTA, 1998b Selection, convergence, and intragenic recombination in *HLA* diversity. *Genetica* **102/103**: 157–169.
- VISSER, C. J. T., M. G. J. TILANUS, V. SCHAEFFER, Z. TATARI, R. TAMOUZA *et al.*, 1998 Sequencing-based typing reveals six novel *MHC* class I chain-related gene B (*MICB*) alleles. *Tissue Antigens* **51**: 649–652.
- VOGEL, F., and M. MOTULSKY, 1997 *Human Genetics, Problems and Approaches*, Ed. 3. Springer, New York.
- YAO, Z., A. VOLGGER, W. HELMBERG, E. KELLER, L. A. FAN *et al.*, 1999 Definition of new alleles of *MICA* using sequencing based typing. *Eur. J. Immunogenet.* **26**: 225–232.

Communicating editor: N. TAKAHATA