Locus Specificity of Polymorphic Alleles and Evolution by a Birth-and-Death Process in Mammalian MHC Genes

Xun Gu and Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University

We have conducted an extensive phylogenetic analysis of polymorphic alleles from human and mouse major histocompatibility complex (MHC) class I and class II genes. The phylogenetic tree obtained for 212 complete human class I allele sequences (*HLA-A, -B,* and *-C*) has shown that all alleles from the same locus form a single cluster, which is highly supported by bootstrap values, except for one *HLA-B* allele (*HLA-B**7301). Mouse MHC class I loci did not show locus-specific clusters of polymorphic alleles. This was considered to be because of either interlocus genetic exchange or the confusing designation of loci in different haplotypes at the present time. The locus specificity of polymorphic alleles was also observed in human and mouse MHC class II loci. It was therefore concluded that interlocus recombination or gene conversion is not very important for generating MHC diversity, with a possible exception of mouse class I loci. According to the phylogenetic trees of complete coding sequences, we classified human MHC class I (*HLA-A, -B,* and *-C*) and class II (*DRB1*) alleles into three to five major allelic lineages (groups), which were monophyletic with high bootstrap values. Most of these allelic groups remained unchanged even in phylogenetic trees based on individual exons, though this does not exclude the possibility of intralocus recombination involving short DNA segments. These results, together with the previous observation that MHC loci are subject to frequent duplication and deletion, as well as to balancing selection, indicate that MHC evolution in mammals is in agreement with the birth-and-death model of evolution, rather than with the model of concerted evolution.

Introduction

The major histocompatibility complex (MHC) plays an important role in the regulation of immune response in vertebrates. MHC molecules can be divided into two groups: class I and class II molecules. The class I MHC molecule consists of an α chain and a β_2-microglobulin (β_2m). The class I α chain has three extracellular domains (α_1, α_2, and α_3), a transmembrane portion, and a cytoplasmic tail. The α_3 domain associates noncovalently with β_2m. The class II MHC molecule consists of noncovalently associated α and β chains, which are encoded by the class II α-chain (*A*) and β-chain (*B*) genes, respectively. Each chain is composed of two extracellular domains (designated as α_1 and α_2 in the α chain and as β1 and β2 in the β chain), a transmembrane portion, and a cytoplasmic tail (Klein 1986).

A remarkable feature of MHC genes is the extremely high degree of genetic polymorphism within loci, and some allelic lineages apparently have coexisted in the population for a long evolutionary time encompassing several speciation events (Klein and Figueroa 1986; Lawlor et al. 1988; Mayer et al. 1988; McConnell et al. 1988; Klein, Takahata and Ayala 1993). The primary factor for maintaining this high degree of polymorphism seems to be overdominant selection operating at the antigen recognition sites of MHC molecules (Hughes and Nei 1988). However, to explain the diversity and evolution of MHC and immunoglobulin (Ig) genes, a number of authors invoked concerted evolution, in which polymorphism is assumed to be generated by interlocus recombination or by gene conversion (e.g., Ohta 1983; Weiss et al. 1983; Lawlor et al. 1990). However, this view has been questioned on the ground that the member genes of Ig and MHC multigene families are not necessarily more closely related to one another than to the genes from different species (Gojobori and Nei 1984; Hughes and Nei 1989*a*; Nei and Hughes 1991; Ota and Nei 1994). By contrast, Nei and Hughes (1992) and Ota and Nei (1994) proposed the birth-and-death model of evolution to explain the long-term evolution of MHC and Ig genes. In this model, new genes are created by repeated gene duplications, and some duplicate genes are maintained in the genome for a long time, but others are deleted or have become nonfunctional through deleterious mutations. A similar mechanism called the accordion model was also proposed by Klein et al. (1993). Evidence supporting the birth-and-death model has accumulated (Nei and Hughes 1992; Ota and Nei 1994; Nei, Gu, and Sitinikova 1997), but the controversy over the two competing hypotheses is still continuing (e.g., Pease et al. 1993; Hogstrand and Bohme 1994; Brunsberg et al. 1996; Yun, Melvold, and Pease 1997).

Nei, Gu, and Sitinikova (1997) attempted to resolve this controversy by examining the pattern of phylogenetic trees of polymorphic alleles from different loci of the MHC. If concerted evolution is the major mechanism for generating MHC polymorphism, one would expect that alleles from the same locus would not form a monophyletic cluster because some alleles at a locus should have been derived from other loci by interlocus recombination or gene conversion. By contrast, if birth-and-death evolution and overdominant selection are important mechanisms, as claimed by Nei and Hughes (1992), the alleles from different loci are expected to

Key words: MHC polymorphism, multigene families, birth-and-death evolution, interlocus recombination, gene conversion, phylogenetic analysis.

Address for correspondence and reprints: Xun Gu, Iowa Computational Molecular Biology Laboratory and Department of Zoology and Genetics, Science II building, Iowa State University, Iowa 50011. E-mail: xgu@iastate.edu.

Mol. Biol. Evol. 16(2):147–156. 1999

form different clusters. Nei, Gu, and Sitinikova (1997) showed that this is indeed the case, but their analysis was based on relatively small data sets. In this paper, we present results from a more comprehensive study using almost all data. Our purpose is not to enumerate every possible case of gene conversion or recombination but to evaluate the relative importance of gene conversion and point mutation in generating genetic diversity within and between populations. Note that in natural populations, many new mutant alleles are eliminated by purifying selection and genetic drift (Nei 1987), and only a minority of mutant alleles are incorporated into the population.

Materials and Methods

In humans, MHC (HLA) class I and class II genes are located on chromosome 6 and form two separate clusters. The class I MHC comprises three highly expressed and highly polymorphic loci, *HLA-A, -B,* and *-C,* which are called classical class I genes and are usually designated as Ia. In addition, there are many nonclassical class I genes with limited expression and low polymorphism (designated as Ib) and many pseudogenes in the MHC class I gene cluster (e.g., *E, F,* and *G* genes; Klein and O'Huigin 1994). The definition of nonclassical genes is not always consistent in the literature, and pseudogenes are often included in the Ib group. The function of Ib genes has not been well understood, though some of them have existed in the genome for a long time (Nei, Gu, and Sitinikova 1997). In this paper, we are primarily interested in the polymorphic pattern of three Ia loci (*HLA-A, -B,* and *-C*). The nucleotide sequences of the *HLA-A, -B,* and *-C* loci were obtained from the database established by P. Parham (http://www.swmed.edu), which contains a total of 67 alleles from locus *HLA-A,* 149 alleles from locus *HLA-B,* and 39 alleles from locus *HLA-C.* Excluding partial sequences, we used 56 *HLA-A* alleles (20 subtypes), 122 *HLA-B* alleles (30 subtypes), and 34 *HLA-C* alleles (14 subtypes) in this study.

The human class II gene cluster contains six major gene regions: *DP, DN, DM, DO, DQ,* and *DR* (Trowsdale 1995). Each of the *DP, DM, DQ,* and *DR* regions consists of at least one *A*-gene locus and at least one *B*-gene locus, which encode the α and β subunits of the class II MHC molecules, respectively. Here, we are concerned with loci *DQA1, DQB1, DRA, DRB1, DPA1,* and *DPB1,* which are universally expressed, highly polymorphic, and present in all haplotypes. Nucleotide sequences for these genes were obtained from the database in the network Histo (http://histo.cryst.bbk.ac.uk and http://ebi.ac.uk). The numbers of alleles so far available are 15, 25, 2, 135, 8, and 65 for loci *DQA1, DQB1, DRA, DRB1, DPA1,* and *DPB1,* respectively (see also Parham and Ohta 1996). We used 64 alleles whose complete sequences were available.

In mice, the MHC (H-2) gene cluster is located on chromosome 17. There are three Ia loci, *K, L,* and *D,* but most haplotypes do not have locus *L.* The mouse Ia genes are not orthologous with human Ia genes (Klein and Figueroa 1986; Hughes and Nei 1989*a*). The mouse MHC class II gene cluster includes α-gene loci (*A*α and *E*α) and β-gene loci (e.g., *A*β1, *A*β2, and *A*β3 and *E*β1 and *E*β2). The mouse loci *A*α, *A*β1, *E*α, *E*β, and *A*β2 are orthologous to the human loci *DQA, DQB, DRA, DRB,* and *DOB,* respectively (Klein 1986, Hughes and Nei 1990). Most of the mouse *DP* region has been lost, and only one pseudogene (*A*β3, also called *Pb*) remains (Trowsdale 1995). The *A*β3 gene was excluded from the present study because only one partial sequence was available. Nucleotide sequences for mouse MHC genes were obtained from the database in Histo (http://histo.cryst.bbk.ac.uk) or from GenBank.

The genetic diversity of MHC loci has been extensively studied in many different organisms (e.g., Watkins, Hodi, and Letvin [1988] studied New World Monkeys; Zoorob et al. [1993] studied chicken; Sato et al. [1993] studied *Xenopus*; Edwards, Wakeland and Potts [1995] studied songbirds; and Hashimoto, Nakanishi, and Kurosawa [1992] studied sharks). However, large-sequence data sets that can be used for a comprehensive phylogenetic analysis are available only in humans and mice. Note that the complete allelic sequences are not only important for conducting a reliable phylogenetic analysis but also useful for studying interlocus or interallelic recombination. In some organisms, cDNA sequences are available, but they are not appropriate for our study because it is unclear whether they come from the same locus or from different loci.

Results

Locus Specificity of Human Classical Class I Alleles

Figure 1 shows the phylogenetic tree of human MHC class Ia genes and alleles (*HLA-A, -B,* and *-C*). This tree was constructed by the neighbor-joining method (Saitou and Nei 1987) from complete coding sequences (1,113 bp). The tree was rooted by using mouse Ia genes *K* and *L* as outgroups. Strikingly, all alleles from the same locus form a single cluster in the phylogenetic tree, and the genetic distances between loci are much larger than the average distance within loci. The bootstrap values for the monophyletic groups of *HLA-A* alleles and of *HLA-C* alleles are both 100%. All *HLA-B* alleles form a single cluster, but the bootstrap value is very low (<50%). This low bootstrap value is caused by one outlier allele, *HLA-B**7301, and if we eliminate this allele, the bootstrap value for *HLA-B* alleles becomes 93%. These results indicate that interlocus recombination (or gene conversion) has not played an important role in the generation of genetic diversity at these loci. A similar conclusion has been reached by Parham et al. (1988) in their study of sequence homology.

Allele *HLA-B**7301 Is a Product of Interlocus Recombination

The phylogenetic position of *HLA-B**7301 in figure 1 is uncertain and could be due to interlocus recombination. (The first two digits of the allele number at loci *HLA-A, HLA-B,* and *HLA-C* designates the allelic sub-

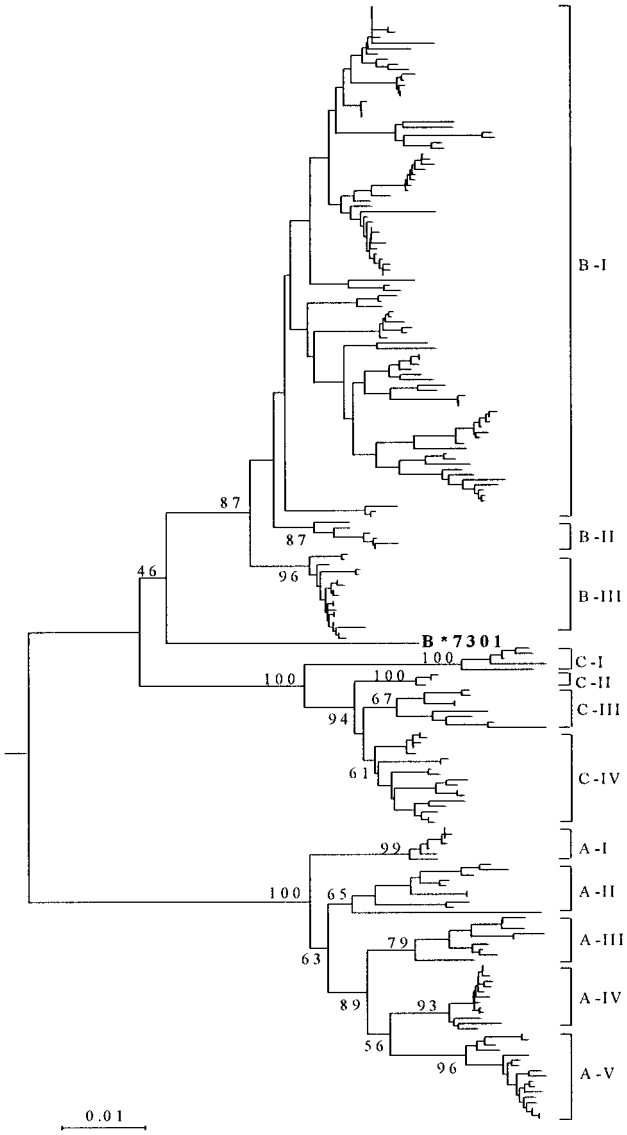

FIG. 1.—Phylogenetic tree of human MHC class Ia genes and alleles (*HLA-A, -B* and *-C*). It was constructed by the neighbor-joining method (Saitou and Nei 1987), with Jukes-Cantor (1969) distances for complete coding sequences (1,113 bp). This tree was rooted by mouse class Ia genes as outgroups. The numbers for interior branches refer to the bootstrap values with 500 replications; values below 50% are not given. At locus *HLA-A*, five (monophyletic) groups can be identified. Group *A-I* consists of subtypes *A*23 and *A*24; group *A-II* consists of *A*01, *A*03, *A*11, *A*30, *A*36, and *A*80; group *A-III* consists of *A*29, *A*31, *A*32, *A*33, and *A*74; group *A-IV* consists of *A*25, *A*26, *A*34, *A*43, and *A*66; and group *A-V* consists of *A*02, *A*68, and *A*69. At locus *HLA-B*, three (monophyletic) groups can be identified. Group *B-I* consists of subtypes *B*07, *B*08, *B*13, *B*15, *B*18, *B*27, *B*35, *B*37, *B*40, *B*41, *B*42, *B*44, *B*46, *B*47, *B*48, *B*51, *B*52, *B*53, *B*57, *B*58, *B*78, and *B*81; group *B-II* consists of *B*54, *B*55, *B*56, and *B*59; and group *B-III* consists of *B*14, *B*38, *B*39, and *B*67. At locus *HLA-C*, four (monophyletic) groups can be identified. Group *C-I* consists of subtype *C*07; group *C-II* consists of *C*03; group *C-III* consists of *C*01, *C*04, *C*14, and *C*18; and group *C-IV* consists of *C*02, *C*05, *C*06, *C*08, *C*12, *C*13, *C*15, and *C*16.

type of each allele, so that *HLA-B**7301 refers to allele 01 that belongs to subtype *HLA-B*73). This allele is clustered with other *HLA-B* alleles with a bootstrap value of 46%, and the average nucleotide difference between *HLA-B**7301 and other *HLA-B* alleles (~6%) is about twice as large as the average nucleotide difference for other *HLA-B* alleles (~3%). Parham et al. (1994) speculated that *HLA-B**7301 may represent an ancient lineage, but it contains several nucleotide substitutions previously observed only in some *HLA-A* and *-C* alleles (Vilches et al. 1994).

We therefore compared the DNA sequence of allele *HLA-B**7301 with the sequences of other alleles. This comparison showed that while this allele is similar to other *HLA-B* alleles in exons 2 and 3, it is more similar to *HLA-C* alleles in exons 4–8 than to other *HLA-B* alleles. The chimeric structure of *HLA-B**7301 can be shown by exon-based phylogenetic analysis. Figure 2 shows the phylogenetic trees constructed from each of exons 2, 3, and 4, which encode class I MHC domains $\alpha 1$, $\alpha 2$, and $\alpha 3$, respectively, and from the region including exons 5 to 8, which encode the transmembrane portion and the cytoplasmic tail. It is interesting to note that exons 2 and 3 of *HLA-B*7301* clearly belong to the *HLA-B* allele cluster (panels *A* and *B*), whereas exon 4 and exons 5–8 are clustered with those of *HLA-C*-alleles with a 97% bootstrap value (panels *C* and *D*). Thus, *HLA-B**7301 seems to be a hybrid allele with *HLA-B*–like exons 2–3 and *HLA-C*–like exons 4–8, which was generated by interlocus recombination. This interlocus recombination may have occurred shortly after the separation of human Ia loci *HLA-B* and *-C* when the sequence divergence between them was relatively small; exon 4 of allele *HLA-B**7301 differs considerably from that of *HLA-C*-alleles (data not shown, but see fig. 2*c* and *d*). Interestingly, a similar intron-mediated interlocus recombination has been proposed by Hughes and Nei (1989*b*) for the formation of the present *HLA-A* locus from the ancient *HLA-E* and *HLA-A* loci that apparently occurred about 20 million years ago.

In short, among 212 complete sequences obtained from human MHC loci *HLA-A, -B,* and *-C*, there was only one allele (*HLA-B**7301) that was apparently created by interlocus recombination. Therefore, the contribution of interlocus recombination to genetic diversity of the present-day human MHC class I genes seems to be generally small. Of course, the phylogenetic analysis used here would not detect a small number of gene conversions involving a few nucleotide changes. Therefore, the present study does not completely rule out the possible occurrence of gene conversion. However, if gene conversion between different loci occurs so frequently as to enhance the extent of genetic variability at a locus, the locus specificity of polymorphic alleles as shown here would not be observed.

Major Lineages of Human Class Ia Alleles

On the basis of the phylogenetic analysis shown in figure 1, we can classify polymorphic alleles from the same locus into several monophyletic groups, i.e., major allelic lineages. *HLA-A* alleles can be classified into five groups (*A-I* to *A-V*), which are supported by bootstrap values of 65% to 99%. Similarly, *HLA-C* alleles can be classified into four groups (*C-I* to *C-IV*), with high bootstrap values. Most of these groups include many allelic subtypes (see the legend for fig. 1), and different allelic

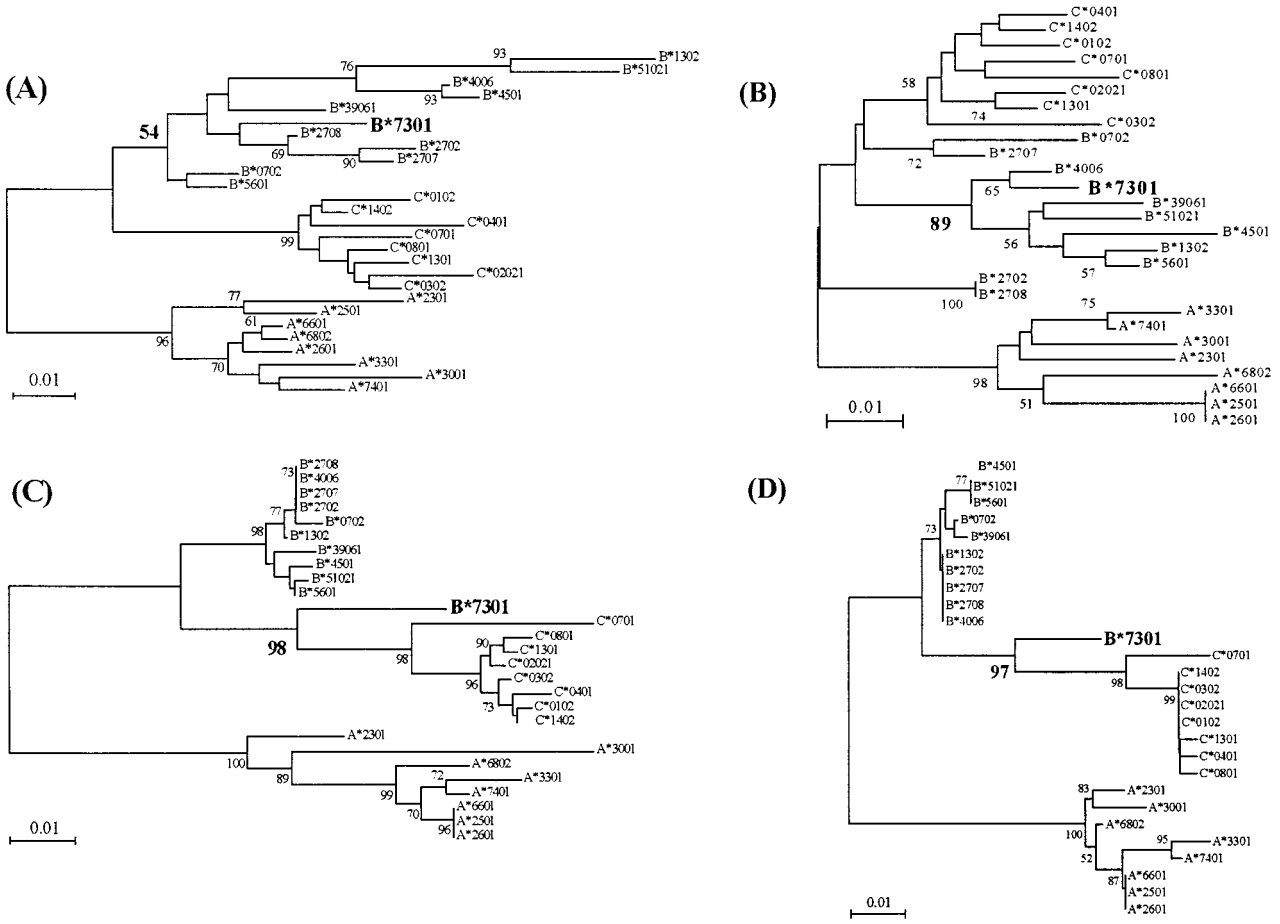

FIG. 2.—Phylogenetic positions of *HLA-B**7301 exons. Phylogenetic trees were constructed for several representative alleles from loci *HLA-A, -B,* and *-C* based on exon 2 (270 bp; panel *A*), exon 3 (276 bp; panel *B*), exon 4 (276 bp; panel *C*), and exons 5–8 (207 bp; panel *D*). In panel *B*, the exon 3 sequences of *B* alleles do not form a monophyletic cluster, but the bootstrap values are very low.

subtypes may be intermingled within each group. The fact that there are highly significant monophyletic clusters of polymorphic alleles at the *HLA-A* and *HLA-C* loci supports that at these loci, the frequency of occurrence of interallelic recombinations, if any, is relatively low. This observation is consistent with the conclusion obtained by McAdam et al. (1994) and by Yeager and Hughes (1996). Furthermore, our analysis has shown that with a few exceptions, the major allelic clusters at locus *HLA-A* or *HLA-C* largely remained unchanged in phylogenetic trees based on different exons (data not shown).

At the *HLA-B* locus, however, the phylogenetic relationships among different alleles are not very clear (fig. 1). We can only classify them into three groups, i.e., *B-I*, comprising most of the *B* allele subtypes; *B-II*, comprising subtypes *B54*, *B55*, *B56*, and *B59* (87% bootstrap value); and *B-III*, comprising subtypes *B14*, *B38*, *B39*, and *B67* (96% bootstrap value). The *B-I* group consists of several allelic clusters, but they are not statistically significant even when the complete coding sequences are used. This low resolution of *HLA-B* allelic lineages is probably caused by a relatively high rate of interallelic recombination as documented at this locus (e.g., Watkins et al. 1992; Santos et al. 1996; Marcos et al. 1997).

Polymorphic Pattern of Human Class II Loci

We constructed phylogenetic trees of alleles from three polymorphic class II *A* loci: *DQA1*, *DRA*, and *DPA1* (Klein and Figueroa 1986), using complete coding sequences. As shown in figure 3, all alleles from the same locus exclusively form a single cluster in phylogenetic analysis, with a bootstrap value of 100%. The same pattern was observed in human class II *B* genes (*DRB1*, *DPB1*, and *DQB1*; fig. 4). Therefore, there is no clear-cut indication of interlocus recombination in human class II *A* and *B* genes. Our results are more reliable than those based on only exon 2 or 3 sequences (e.g., Gyllensten and Erlich 1989; Gyllensten, Sundvall, and Erlich 1991; Brunsberg et al. 1996). Brunsberg et al. (1996) reported that alleles from different pig MHC II *B* loci were intermingled and regarded this as evidence for interlocus recombination. Since their sequence data were from cDNA clones, Brunsberg et al. (1996) conducted a phylogenetic analysis for locus identification. We have reanalyzed their data and found that the bootstrap values of the tree are quite low (data not shown).

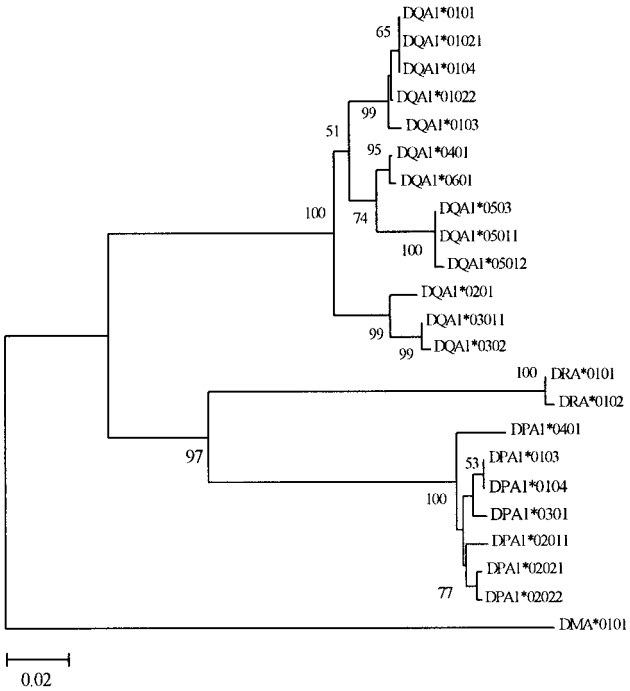

FIG. 3.—Phylogenetic tree of alleles from human MHC class II *A* loci, using complete coding sequences. *DMA* is known to be related to other genes (Nei, Gu, and Sitinikova 1997).

In general, our results support the earlier conclusions of Klein et al. (1989) and of Nei, Gu, and Sitinikova (1997) that interlocus recombination has not played an important role for generating a high degree of polymorphism at human class II loci.

According to the phylogenetic tree in figure 4, *DRB1* alleles can be classified into five groups (*DRB1-I* to *DRB1-V*), supported by bootstrap values of 82% to 100% (see the legend for fig. 4 for details). These allelic groups do not correspond to the five human haplotypes known in the DRB region (Svensson et al. 1996). Interestingly, when phylogenetic trees are constructed separately for exons 2, 3, or 4, these five major groups always remain monophyletic (data not shown). These results suggest that at least the intron-mediated interallelic recombination does not occur frequently at the human *DRB1* locus. However, the allelic classification for human class II alleles should be based on complete coding sequences since that method is statistically more reliable than that based on single exons (e.g., Klein and O'Huigin 1995). Gyllensten, Sundvall, and Erlich (1991) suggested that the part of the sequences of exon 2 encoding β-sheet and the part of those encoding α-helix have different evolutionary histories and interpreted this difference as being due to interallelic recombination or gene conversion. One cannot exclude this possibility, but we should be cautious

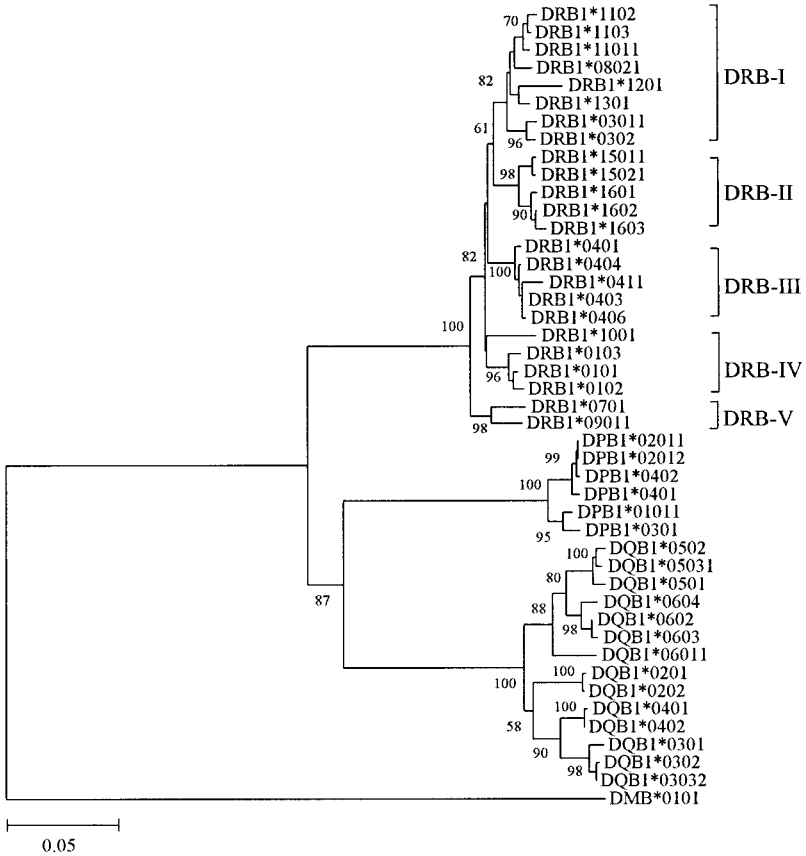

FIG. 4.—Phylogenetic tree of alleles from human MHC class II *B* loci, using complete coding sequences. *DMB* is known to be related to other genes. *DRB1* alleles can be classified into five allelic groups. *DRB-I* includes subtypes *DRB1*-03, -08, -11, -12, and -13. *DRB-II* includes subtypes *DRB*-15 and -16. *DRB-III* includes subtype *DRB*-04. *DRB-IV* includes subtype *DRB*-01. *DRB-V* includes subtypes *DRB1*-07 and -09.

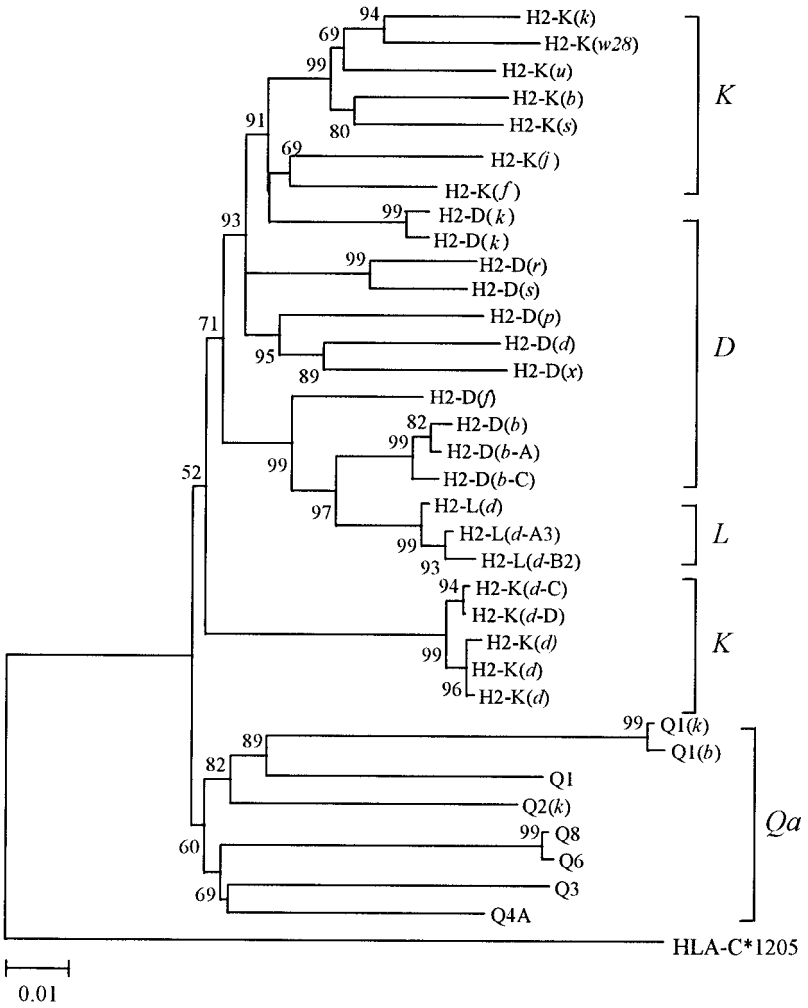

FIG. 5.—Phylogenetic tree of alleles from mouse MHC class I loci, using complete coding sequences. Human *HLA-C**1205 was used as an outgroup. This tree is different from that of Hughes (1991), partly because different allelic sequences are used and partly because Hughes (1991) used nonsynonymous substitutions for constructing the tree.

about their conclusion because they did not conduct any statistical test of their phylogenetic tree. Since the numbers of nucleotides in the α-helix and β-sheet regions are only 54 and 106, respectively, the phylogenetic difference found by Gyllensten, Sundvall, and Erlich (1991) could be caused by stochastic errors. In addition, phylogenetic reconstruction based on only the peptide-binding region could be misleading because positive selection and/or convergent evolution apparently occurs in this region.

Polymorphic Pattern of Mouse Class I Loci

In mice, the genomic organization of class Ia loci varies with inbred strain or haplotype because of gene deletion and duplication, and it is often difficult to identify the same genetic loci among different haplotypes (Stroynowski 1990). For example, the *L* locus is missing in many haplotypes, whereas the *D* locus is duplicated in some haplotypes. The number of Ib *Qa* loci also varies extensively with haplotype. Yet, the polymorphic alleles at the *K, D,* and *L* loci have been studied at the DNA sequence level. The phylogenetic tree for alleles for which the complete sequences are available is presented in figure 5. Unlike the previous trees, this tree does not show a locus specificity of polymorphic alleles. Thus, the alleles at the *K* locus form two separate clusters, whereas those of the *D* locus are not monophyletic. Only the alleles that belong to locus *L* show a clear monophyly. Furthermore, the genes from Ib *Qa* loci are clearly separate from the alleles at the Ia loci.

A simple interpretation of these results would be that the locus specificity of polymorphic alleles has been lost by interlocus recombination between loci. In fact, there are a number of reports that suggest the occurrence of interlocus recombination or gene conversion at mouse class I loci (e.g., Weiss et al. 1983; Pease et al. 1983, 1993; Yun, Melvold, and Pease 1997). However, it is also possible that some of the above results are due to misidentification of alleles at different loci (Hughes 1991). To clarify the evolutionary relationship of alleles from different class Ia loci in mice, it seems to be necessary to establish the genomic maps of genes in all haplotypes.

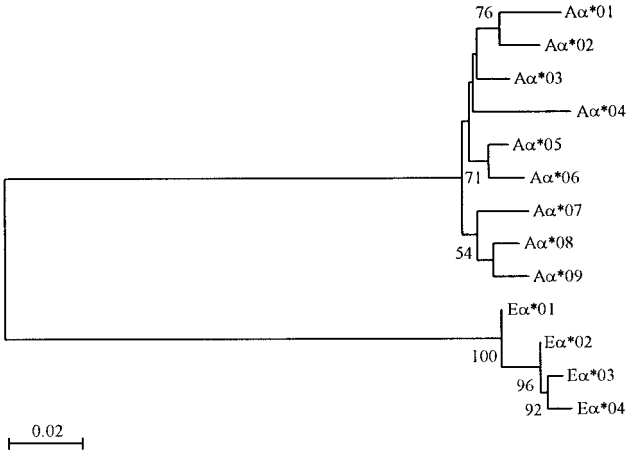

FIG. 6.—Phylogenetic tree of alleles from mouse MHC class II-*A* loci, using complete coding sequences. The traditional designations for these alleles are quite complicated. Here, we use simplified designations for general readers. A detailed list of these alleles is available upon request.

Polymorphic Pattern of Mouse Class II Loci

In mice, there are nine and four complete sequences available for the class II *A*α and *E*α loci, respectively. The phylogenetic tree for these sequences, presented in figure 6, clearly shows that *A*α and *E*α alleles form separate clusters, as in the case of the alleles of human class II loci, and there is no indication of interlocus recombination.

As indicated before, the mouse *A*β1, *A*β2 and *A*β3 loci are orthologous to the human *DQB1*, *DOB*, and *DPB* loci, whereas the mouse *E*β1 locus is orthologous to the human *DRB1* locus and the mouse *E*β2 is an unexpressed duplicated gene of *E*β1. We collected 32 and 12 complete sequences from loci *A*β1 and *E*β1, respectively, and collected several *A*β2 and *E*β2 sequences and conducted a phylogenetic tree (fig. 7). In this tree, all the alleles from the same locus form a single cluster with a 100 percent bootstrap value, and there is no overlap of alleles among different loci. Therefore, the polymorphic pattern of mouse MHC class II genes indicates

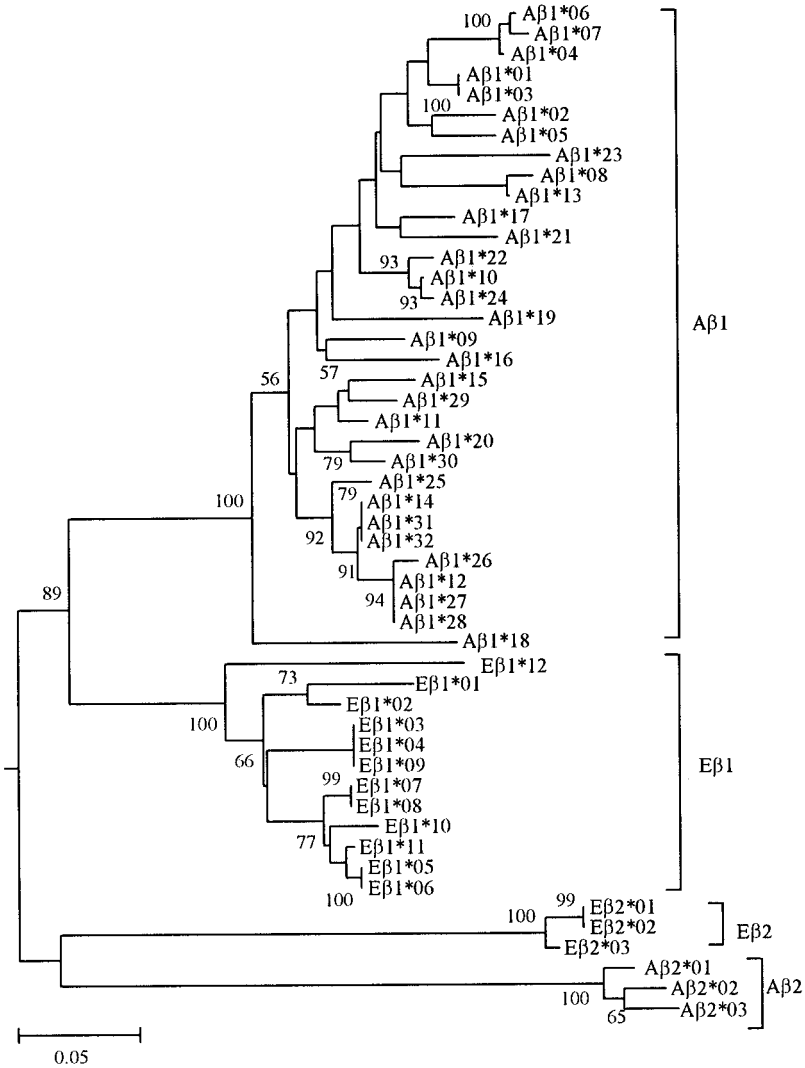

FIG. 7.—Phylogenetic tree of alleles in mouse MHC class II *B* genes, using complete coding sequences. See the legend for figure 6 for the designations of alleles. The root of this tree was determined by using chicken MHC class II *B* sequences (see Nei, Gu, and Sitinikova 1997, fig. 6).

no extensive interlocus recombination, as in the case of human genes. Incidentally, it appears that the alleles at the *E*β2 and *A*β2 loci have evolved faster than the alleles at the *E*β1 and *A*β1 loci.

The clear-cut locus specificity of polymorphic alleles at the mouse class II β loci is contradictory with the experimental result that an allele at the *E*β locus was converted by an allele at the *A*β locus (Hogstrand and Bohme 1994). Possible reasons for this discrepancy will be discussed later.

Discussion

We have seen that in human and mouse MHC loci, polymorphic alleles from the same locus generally form a monophyletic cluster, except in the mouse class I loci, where the relationship of different alleles from different loci is not well established. This indicates that the effect of interlocus recombination or gene conversion on MHC diversity is reasonably small, though it does not completely rule out the possible occurrence of these events. As mentioned earlier, this observation is somewhat contradictory with some experimental evidence that suggests the occurrence of interlocus gene conversion in mice (Hogstrand and Bohme 1994; Yun, Melvold, and Pease 1997). Hogstrand and Bohme (1994) estimated that the frequency of gene conversion between loci *E*β and *A*β is 2×10^{-6} per sperm per generation. If this estimate is representative, one would expect that the alleles from loci *E*β1 and *A*β1 are intermingled when phylogenetic analysis is conducted. However, the phylogenetic tree given in figure 7 shows that the alleles from the two loci are clearly separate. This suggests either that the current estimate of the frequency of gene conversion is too high or that many gene conversions observed in the mouse experiment are not fixed in natural populations of mice because of selective disadvantage compared with wild-type alleles. As mentioned earlier, most MHC alleles are subject to overdominance or to some other type of balancing selection. It is then possible that many mutant alleles generated by gene conversion are not selectively advantageous and are eliminated in natural populations. If this is the case, the contribution of interlocus recombination to MHC diversity would not be as large as one might expect from experimental data.

Another problem that occurs with the gene conversion hypothesis is the long persistence of polymorphic alleles. Many MHC alleles are known to persist in the population for tens of millions of years (Klein and Figueroa 1986; Nei and Hughes 1991). If gene conversion occurs very often, the selective advantage of an allele (in heterozygous condition) would change from time to time, and the allele would not persist in the population for a long time (Takahata and Nei 1990).

Some authors have argued that some gene conversion may involve a small number of nucleotide changes, often as small as one or two (e.g., Watkins et al. 1992). Particularly when two consecutive nucleotide changes (e.g., AG → GA) are observed, gene conversion is usually invoked (Yun, Melvold, and Pease 1997). In these cases, however, the distinction between gene conversion and point mutation becomes moot, because a single mutational event may generate two or three nucleotide changes (Ripley 1991; Wolfe and Sharp 1993).

So far, we have considered interlocus recombination or gene conversion. In recent years, many authors (e.g., Gyllensten, Sundvall, and Erlich 1991; Belich et al. 1992; Watkins et al. 1992; McAdam et al. 1994; Zangenberg et al. 1995; Marcos et al. 1997) have reported cases in which intralocus gene conversion apparently occurred. These reports have been concerned primarily with highly polymorphic loci, such as the *HLA-B* locus. The phylogenetic analysis in this study also suggested that at the *HLA-B* locus, interallelic recombination apparently occurred with a relatively high frequency. At other loci, however, allelic clusters with high bootstrap values are observed, suggesting that interallelic recombination or gene conversion is not as frequent as that at the *HLA-B* locus.

Theoretically, unequal crossover or gene conversion is a mechanism that homogenizes allelic variation both within and between loci. If there is no mutation and no balancing selection, it will eventually wipe out all genetic variation in finite populations. Genetic variation at a locus is increased by gene conversion only when genetic variation from other divergent loci is introduced. Therefore, in the absence of mutation and selection, genetic variability at a locus is expected to decrease, on the average, though new allelic sequences may occasionally appear through recombination or gene conversion. In other words, it is important to realize that the primary factors that enhance genetic variability at an MHC locus are mutation and balancing selection. However, the mutation rate is not particularly higher than that for many other loci (Klein and Figueroa 1986; Nei and Hughes 1991). Therefore, the major factor for the high degree of MHC polymorphism seems to be overdominant or balancing selection.

The present study has shown that interlocus recombination is relatively rare and that the genetic variation between MHC loci is enhanced primarily by mutation and selection. As shown by Klein and Figueroa (1986), Hughes and Nei (1989*a*, 1990), Nei and Hughes (1992), and Nei, Gu, and Sitinikova (1997), MHC genes are also subject to frequent duplication and deletion, and some duplicate loci may become nonfunctional. Therefore, the long-term evolution of MHC genes appears to be dictated by the birth-and-death model of evolution rather than by concerted evolution.

Acknowledgments

We thank Kei Takahashi for his help in data analysis. We are also grateful to Sudhir Kumar, Tatyana Sitnikova, Jianzhi Zhang, and Ingrid Jakobsen for their comments on an earlier version of this manuscript. This work was supported by an NIH grant and an NSF grant to M.N.

LITERATURE CITED

BELICH, M. P., J. A. MADRIGAL, W. H. HILDEBRAND, J. ZEMMOUR, R. C. WILLIAMS, R. LUZ, M. L. PETZL-ERLER, and

P. PARHAM. 1992. Unusual *HLA-B* alleles in two tribes of Brazilian Indians. Nature **357**:326–329.

BRUNSBERG, U., I. EDFORS-LIJA, L. ANDERSSON, and K. GUSTAFSSON. 1996. Structure and organization of pig MHC class II *DRB* genes: evidence for genetic exchange between loci. Immunogenetics **44**:1–8.

EDWARDS, S. V., E. K. WAKELAND, and W. K. POTTS. 1995. Contrasting histories of avian and mammalian MHC genes revealed by class IIB sequences from songbirds. Proc. Natl. Acad. Sci. USA. **92**:12200–12204.

GOJOBORI, T., and M. NEI. 1984. Concerted evolution of the immunoglobulin V_H gene family. Mol. Biol. Evol. **1**:1947–1957.

GYLLENSTEN, U. B., and H. ERLICH. 1989. Ancient roots for polymorphism at the *HLA-DQ*α locus in primates. Proc. Natl. Acad. Sci. USA **86**:9986–9990.

GYLLENSTEN, U. B., M. SUNDVALL, and H. ERLICH. 1991. Allelic diversity is generated by intraexon sequence exchange at the DRB1 locus of primates. Proc. Natl. Acad. Sci. USA **88**:3686–3690.

HASHIMOTO, K., T. NAKANISHI, and Y. KUROSAWA. 1992. Identification of a shark sequence resembling the major histocompatibility complex class I α3 domain. Proc. Natl. Acad. Sci. USA. **89**:2209–2212.

HOGSTRAND K., and J. BOHME. 1994. A determination of the frequency of gene conversion in unmanipulated mouse sperm. Proc. Natl. Acad. Sci. USA. **91**:9921–9925.

HUGHES, A. L. 1991. Independent gene duplications, not concerted evolution, explain relationships among class I MHC genes of murine rodents. Immunogenetics **33**:367–373.

HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167–170.

———. 1989*a*. Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol. Biol. Evol. **6**:559–579.

———. 1989*b*. Ancient interlocus exon exchange in the history of the *HLA-A* locus. Genetics **122**:681–686.

———. 1990. Evolutionary relationships of class II major histocompatibility complex genes in mammals. Mol. Biol. Evol. **7**:491–514.

KLEIN, J. 1986. Natural history of the major histocompatibility complex. Wiley and Sons, New York.

KLEIN, J., and F. FIGUEROA. 1986. Evolution of the major histocompatibility complex. Crit. Rev. Immunol. **6**:295–386.

KLEIN, J., M. KASAHARA, J. GUTKNECKT, and F. FIGUEROA. 1989. Origin and function of MHC polymorphism. Hum. Immunol. **26**:107–121.

KLEIN, J., and C. O'HUIGIN. 1994. The conundrum of nonclassical major histocompatibility complex genes. Proc. Natl. Acad. Sci. USA **91**:6251–6252.

———. 1995. Class II *B* MHC motifs in an evolutionary perspective. Immunol. Rev. **143**:88–111.

KLEIN, J., H. ONO, D. KLEIN, and C. O'HUIGIN. 1993. The accordion model of MHC evolution. Prog. Immunol. **8**:137–143.

KLEIN, J., N. TAKAHATA, and F. J. AYALA. 1993. MHC polymorphism and human origins. Sci. Am. **Dec.**:78–83.

LAWLOR, D. A., F. E. WARD, P. D. ENNIS, A. P. JACKSON, and P. PARHAM. 1988. *HLA-A* and *B* polymorphisms predate the divergence of human and chimpanzee. Nature **335**:268–271.

LAWLOR, D. A., J. ZEMMOUR, P. D. ENNIS, and P. PARHAM. 1990. Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. Annu. Rev. Immunol. **8**:23–63.

MARCOS, C. Y., M. A. FERNANDEZ-VINA, A. M. LAZARO, C. J. NULF, E. H. RAIMONDI, and P. STASTNY. 1997. Novel *HLA-B*35 subtypes: putative gene conversion events with donor sequences from alleles common in Native Americans (*HLA-B**4002 or *B**4801). Hum. Immunol. **53**:148–155.

MAYER, W. E., M. JONKER, D. KLEIN, P. IVANYI, G. VAN SEVENTER, and J. KLEIN. 1988. Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. EMBO J. **7**:2765–2774.

MCADAM, S., J. E. BOYSON, X. LIU, T. L. GARBER, A. L. HUGHES, R. E. BONTROP, and D. WATKINS. 1994. A uniquely high level of recombination at the *HLA-B* locus. Proc. Natl. Acad. Sci. USA. **91**:5893–5897.

MCCONNELL, T. L., W. S. TALBOT, R. A. MCINDOE, and E. K. WAKELAND. 1988. The origin of MHC class II gene polymorphism within the genus *Mus.* Nature **332**:651–654.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., X. GU, and T. SITINIKOVA. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. USA. **94**: 7799–7806.

NEI, M., and A. L. HUGHES. 1991. Polymorphism and evolution of the major histocompatibility complex loci in mammals. Pp. 227–247 *in* R. SELANDER, A. CLARK, and T. WHITTAM, eds. Evolution at the molecular level. Sinauer Associates, Sunderland, Mass.

———. 1992. Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. Pp. 27–38 *in* K. TSUJI, M. AIZAWA, and T. SASAZUKI, eds. Proceeding of the 11th histocompatibility workshop and conference, Vol. 2. Oxford University Press, Oxford, England.

OHTA, T. 1983. On the evolution of multigene families. Theor. Pop. Biol. **23**:216–240.

OTA, T., and M. NEI. 1994. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. Mol. Biol. Evol. **11**:469–482.

PARHAM, P., K. L. ARNETT, E. J. ADAMS, L. D. BARBER, J. D. DOMENA, D. STEWART, W. H. HILDEBRAND, and A. M. LITTLE. 1994. The *HLA-B*73 antigen has a most unusual structure that defines a second lineage of *HLA-B* alleles. Tissue Antigens **43**:302–313.

PARHAM, P., C. E. LOMEN, D. A. LAWLOR, J. P. WAYS, N. HOLMS, H. L. COPPIN, R. D. SALTER, A. M. WAN, and P. D. ENNIS. 1988. Nature of polymorphism in *HLA-A, -B,* and *-C* molecules. Proc. Natl. Acad. Sci. USA. **85**:4005–4009.

PARHAM, P., and T. OHTA. 1996. Population biology of antigen presentation by MHC class I molecules. Science **272**:67–74.

PEASE, L. R., R. M. HORTON, J. K. PULLEN, and T. J. YUN. 1993. Unusual mutation clusters provide insight into class I gene conversion mechanisms. Mol. Cell. Biol. **13**:4374–4381.

PEASE, L. R., D. H. SCHULZE, G. M. PFAFFENBACH, and S. G. NATHENSON. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. Proc. Natl. Acad. Sci. USA. **80**:242–246.

RIPLEY, L. 1991. Concerted mutagenesis: its potential impact on interpretation of evolutionary relationships. Pp. 63–94 *in* J. KLEIN and D. KLEIN, eds. Molecular evolution of the major histocompatibility complex., Springer-Verlag, Berlin.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:949–967.

SANTOS, S., J. L. VICARIO, J. L. MERINO, J. L., and A. BALAS. 1996. Characterization of a new *HLA-B* allele (*B**3702) generated by an intronic recombination event. Immunogenetics **43**:171–172.

SATO, K., M. F. FLAJNIK, L. DU PASQUIER, M. KATAGIRI, and M. KASAHARA. 1993. Evolution of the MHC: isolation of class II β-chain cDNA clones from the amphibian *Xenopus laevis*. J. Immunol. **150**:2831–2843.

STROYNOWSKI, I. 1990. Molecules related to class-I major histocompatibility complex antigens. Annu. Rev. Immunol. **8**: 501–530.

SVENSSON, A.-C., N. SETTERBLAD, U. PIHLGREN, L. RASK, and G. ANDERSON. 1996. Evolutionary relationship between human major histocompatibility complex HLA-DR haplotypes. Immunogenetics **43**:304–314.

TAKAHATA, N., and M. NEI. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics **124**:967–978.

TROWSDALE, J. 1995. ''Both man & bird & beast'': comparative organization of MHC genes. Immunogenetics **41**:1–17.

VILCHES, C., R. DE PABLO, M. J. HERRERO, M. E. MORENO, and M. KREISLER. 1994. *HLA-B*73: an atypical *HLA-B* molecule carrying a Bw6-epitope motif variant and a *B* pocket identical to *HLA-B*27. Immunogenetics **40**:166.

WATKINS, D. I., F. S. HODI, and N. L. LETVIN. 1988. A primate species with limited major histocompatibility complex class I polymorphism. Proc. Natl. Acad. Sci. USA. **85**:7714–7718.

WATKINS, D. I., S. N. MCADAMS, X. LIU, et al. (13 co-authors.) 1992. New recombinant *HLA-B* alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. Nature **357**:329–333.

WEISS, E. H., A. L. MELLOR, L. GOLDEN, K. FAHRNER, E. SIMPSON, J. HURST, and R. A. FLAVELL. 1983. The structure of a mutant *H*-2 gene suggests that the generation of polymorphism in *H*-2 genes may occur by gene conversion-like events. Nature **301**:671–674.

WOLFE, K. H., and P. M. SHARP. 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J. Mol. Evol. **37**:441–456.

YEAGER, M., and A. L. HUGHES. 1996. Interallelic recombination has not played a major role in the history of the *HLA-C* locus. Immunogenetics **44**:128–133.

YUN, T. J., R. W. MELVOLD, and L. R. PEASE. 1997. A complex major histocompatibility complex D locus variant generated by an unusual recombination mechanism in mice. Proc. Natl. Acad. Sci. USA. **94**:1384–1389.

ZANGENBERG, G., M. HUANG, N. ARNHEIM, and H. ERLICH. 1995. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat. Genet. **10**:407–414.

ZOOROB, R., A. BERNOT, D. M. RENOIR, F. CHOUKRI, and C. AUFFRAY. 1993. Chicken major histocompatibility complex II B genes: analysis of interallelic and inter-locus sequence variation. Eur. J. Immunol. **23**:1139–1145.

SAITOU NARUYA, reviewing editor

Accepted October 2, 1998