

NATURAL SELECTION AND THE EVOLUTIONARY HISTORY OF MAJOR HISTOCOMPATIBILITY COMPLEX LOCI

Austin L. Hughes and Meredith Yeager

Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park PA 16802 USA

Received 5/11/98 Accepted 5/22/98

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Balancing selection at MHC loci
 - 3.1. Explaining MHC polymorphism
 - 3.2 Trans-species polymorphism
4. Duplication of loci
5. The role of recombination
 - 5.1. Intralocus recombination in the class I MHC
 - 5.2 intralocus recombination the class II MHC
6. Acknowledgments
7. References

1. ABSTRACT

The major histocompatibility complex (MHC) is a multi-gene family unique to the vertebrates, whose products function to present peptides to T cells. Certain MHC loci are highly polymorphic, and this polymorphism is maintained by a form of balancing selection, probably overdominant selection. This selection has several consequences for MHC biology that make these genes different from neutrally evolving genes: an enhanced rate of nonsynonymous nucleotide substitution in codons encoding the peptide-binding region; long-lasting (“trans-species”) polymorphism; and homogenization of introns relative to exons as a result of recombination and subsequent genetic drift. The MHC also reveals evidence of processes shared with other multi-gene families, including gene duplication and deletion and a low level of inter-locus recombination.

2. INTRODUCTION

The major histocompatibility complex (MHC) of vertebrates has become an important paradigm for molecular evolutionary studies. First, certain polymorphic MHC loci represent probably the best documented examples of positive Darwinian selection at the molecular level. In particular, these loci are examples of balancing selection; that is, selection that acts to maintain polymorphism. Second, the genes of the MHC are members of a multi-gene family. Attempts to unravel the evolutionary history of these loci have revealed examples of evolutionary processes that are common to eukaryotic multi-gene families in general. These processes include gene duplication and the functional divergence of duplicated genes; the silencing of duplicate genes by mutation and deletion of certain duplicate genes, leading to

the turnover of loci over evolutionary time; and inter-locus recombination.

The purpose of the present paper is to review some of the major findings regarding the evolution of MHC loci. We will emphasize in particular the evidence regarding the role of natural selection and the MHC as a paradigmatic case of a multi-gene family. It is by now widely accepted that most polymorphisms in natural populations are selectively neutral (1,2). Thus, the very fact that polymorphisms at certain MHC loci are selectively maintained makes these loci unusual. Most traits of MHC genes at the DNA sequence level that have puzzled investigators can be seen as consequences of the balancing selection acting at these loci (3). If the MHC is an atypical multi-gene family – and in certain respects it is – this is a consequence of the fact that it contains loci subject to an unusual type of natural selection.

3. BALANCING SELECTION AT MHC LOCI

3.1. Explaining MHC Polymorphism

There are four independent lines of evidence that polymorphic MHC loci are subject to balancing selection: (i) The distribution of allelic frequencies does not fit the neutral expectation. (ii) The rate of nonsynonymous nucleotide substitution significantly exceeds the rate of synonymous substitution in the codons encoding the peptide-binding region of the molecule. (iii) Polymorphisms at certain MHC loci have been maintained for long periods of time, sometimes predating speciation events. (iv) Introns have been homogenized relative to exons over evolutionary time, as expected when balancing selection acts to maintain diversity in the latter but not the

Natural selection and the evolutionary history of MHC loci

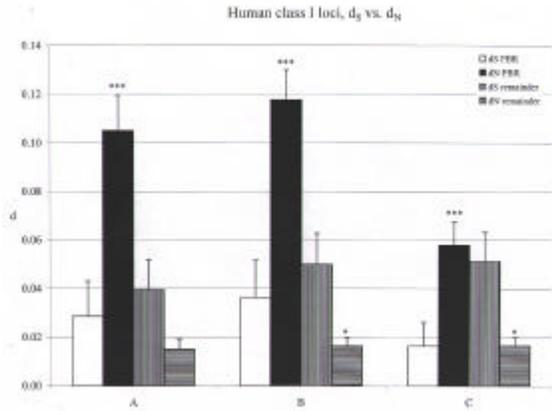


Figure 1. Mean numbers of nucleotide substitutions per synonymous (d_S) and per nonsynonymous (d_N) site (50,51) at the human class I MHC loci *HLA-A* (A), *HLA-B* (B), and *HLA-C* (C). Tests of the hypothesis that $d_S = d_N$: * $P < 0.05$; *** $P < 0.001$.

former. Here we briefly discuss each of these lines of evidence.

Under selective neutrality, polymorphism is a transient phenomenon in evolutionary terms (4). For neutral alleles that become fixed, the average time to fixation is $4N_e$ generations, where N_e is the effective population size (5). At most selectively neutral loci, we expect to see one allele with high frequency at one or a few others of lower frequency. This occurs because most alleles other than the most frequent allele represent new mutations which are still at a low frequency or older alleles that are on the way to extinction. In the case of the polymorphic MHC loci, however, there are a large number of alleles of intermediate frequency (6). Such an allelic distribution is inconsistent with selective neutrality but easily explained under balancing selection (6).

The polymorphism of MHC class I loci was initially discovered through these gene's role in transplant rejection, and initially the function of the molecules they encode was unknown. Therefore, in proposing hypotheses to account for MHC polymorphism, biologists also tried to account for the molecules' function as well. Because the self-incompatibility genes of plants were the only other system known with such high polymorphism, Thomas (7) came up with the idea that the MHC might be a kind of self-incompatibility system for animals. This led to the ideas that MHC molecules might function in maternal-fetal interaction or in mate choice. Even though there has never been good experimental evidence for either of these ideas (8,9), they continue to attract adherents. Ironically, right around the same time as Thomas's exercise in mistaken analogy, Zinkernagel and Doherty (10) provided solid evidence for the MHC's real function; namely, in presenting peptides to T cells (11).

Soon afterward, Doherty and Zinkernagel (12) presented the first hypothesis to account for MHC polymorphism in terms of the molecules' immune function,

which they based on their observation that different MHC allelic products differ with respect to the antigens (i.e., peptides) they can bind and present to T cells. To re-express their argument in terms of our current understanding of MHC function, they proposed that since different allelic products bind different sets of peptides, a heterozygote at all or most MHC loci will have an advantage in a population exposed to more than one pathogen. Because a heterozygote can bind a wider array of peptides, it will be able to deal with a broader array of pathogens. These pathogens might be different species or antigenically distinct strains of the same species (13). Thus, Doherty and Zinkernagel proposed that MHC polymorphism is maintained by heterozygote advantage (overdominant selection).

By now, there is abundant evidence from direct sequencing of peptides bound by both class I and class II molecules that MHC allelic products do indeed differ strikingly with respect to their peptide-binding specificities (14). As regards the hypothesis of overdominant selection, it will be very difficult to obtain evidence regarding it from population studies. In outbred species such as human and mouse, the vast majority of individuals are heterozygous at MHC loci. Thus, a conventional population approach to testing for heterozygote advantage, in which fitnesses of homozygotes and heterozygotes are compared in a natural population, is unlikely to be practical.

An alternative approach to testing for natural selection at MHC loci by makes use of the evolutionary information in DNA sequence comparisons (9,15-17). In most genes, the number of synonymous (or silent) nucleotide substitutions per synonymous site (d_S) exceeds the number of nonsynonymous (or replacement) substitutions per nonsynonymous site (d_N). This occurs because most nonsynonymous mutations are deleterious to protein structure and thus are quickly eliminated by so-called purifying selection. Synonymous substitutions, because they do not change the amino acid, are likely to be selectively neutral or nearly so. On the other hand, if natural selection favors amino acid replacements in a certain protein region, d_N can exceed d_S . Overdominant selection is expected to produce such a pattern of nucleotide substitution because this type of selection accelerates the rate of amino acid replacement (18); and certain other types of balancing selection may have a similar effect.

In the case of both class I and class II MHC, estimation of rates of nucleotide substitution revealed that d_N is significantly greater than d_S in the codons that encode the peptide-binding region (PBR) of the molecule, whereas in the remainder of the molecule d_S exceeds d_N as is true of most genes (9,15-17). Figures 1 and 2 illustrate this pattern of nucleotide substitution in the case of human class I and class II MHC genes. Such analyses provide strong evidence that MHC polymorphism is selectively maintained. They also demonstrate that the selection maintaining this polymorphism is focused on the PBR. Thus they support Doherty and Zinkernagel's (12) hypothesis that the main force driving the selection is the

Natural selection and the evolutionary history of MHC loci

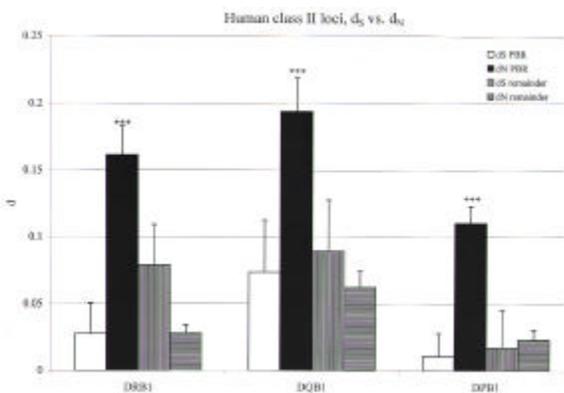


Figure 2. Mean numbers of nucleotide substitutions per synonymous (d_S) and per nonsynonymous (d_N) site (50,51) at the human class II MHC loci HLA-DRB1, HLA-DQB1, and HLA-DPB1. Tests of the hypothesis that $d_S = d_N$: *** $P < 0.001$.

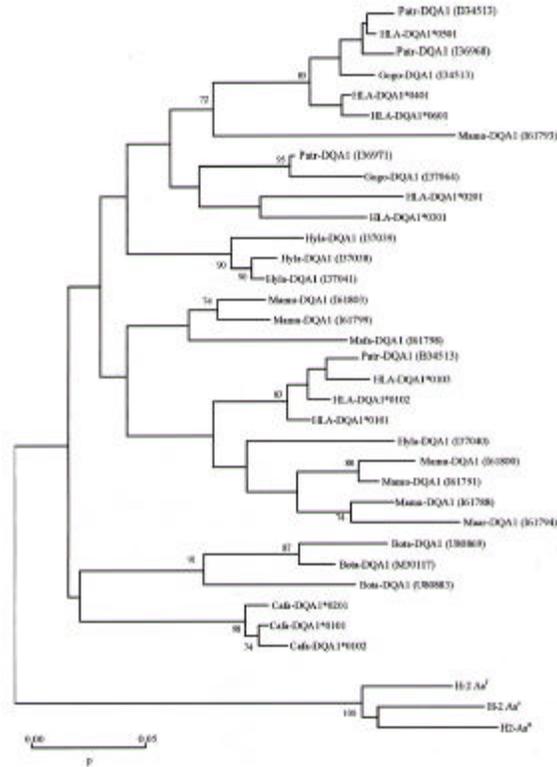


Figure 3. Phylogenetic tree (52) of DQA1 alleles based on the proportion of amino acid difference (p) in the $\alpha 1$ domain. Species used are human (HLA-), chimpanzee (Patr-), Gorilla (Gogo-), gibbon (Hyla-), rhesus monkey (Mamu-), crab-eating macaque (Mafa-), red-faced stump-tailed macaque (Maar-), bovine (Bota-), dog (Cafa-), and mouse (H-2). PIR database accession numbers are in parentheses. Numbers of branches represent percentages of 1000 bootstrap replicates (derived by pseudosampling sites with replacement) that supported a given branch; only values $>50\%$ are shown.

advantage conferred by being able to bind a variety of peptides and thus to resist a variety of pathogens.

3.2. Trans-species Polymorphism

One of the characteristics of MHC loci is the existence of so-called trans-species polymorphism (19-21). In other words, MHC polymorphisms have been maintained for long periods of time, often pre-dating speciation events. Figure 3 shows an example of this phenomenon in the case of the *DQA1* locus of primates, which encodes a class II MHC chain. In this tree the human alleles *HLA-DQA1*0101*, *-DQA1*0102*, and *-DQA1*0103* cluster with certain chimpanzee alleles; and these alleles cluster with alleles from gibbon and rhesus monkey (figure 3). Certain other human alleles cluster with other chimpanzee alleles; for example, *HLA-DQA1*0501* (figure 3). The phylogenetic tree thus indicates that there are allelic lineages at the human *DQA1* locus that have persisted since before human and chimpanzee diverged (5-7 million years ago) or even since before human and old world monkeys diverged (about 25 million years ago). Because selectively neutral polymorphisms are not expected to be maintained for long periods of time, the long persistence of MHC polymorphisms is evidence that they are selectively maintained (22).

Computer simulations by Takahata and Nei (22) showed that polymorphisms could be maintained for long periods of time both by overdominant selection and by one model of frequency-dependent selection which those authors called "minority advantage." In this model, it was assumed that a genotype has an advantage when it becomes rare in a population. Mathematically, this model is very similar to that of overdominant selection, but biologically it is rather different. Consider a parasite species that is well controlled by members of its host species having a particular MHC allele (A1) whose product can bind and present a peptide from the parasite. Under selective pressure from this parasite, the A1 allele will increase in frequency. Suppose a mutation occurs so that the same peptide is no longer bound by the A1 allelic product, giving the parasite an advantage. Then suppose a mutation occurs, giving rise to a new MHC allele (A2), which can bind and present another peptide from the same parasite. The A2 allele will now have an advantage and increase in frequency. When the A2 allele becomes common and the A1 allele scarce, the minority advantage model requires that the A1 allele again have a selective advantage. But it is hard to see that this will happen in the case of the MHC. If the parasite mutates again so that the peptide bound by the A2 allelic product is no longer bound, what likelihood is there that it will mutate in such a way that the parasite now contains a peptide bound by the A1 allelic product. Thus, the minority advantage model does not seem applicable to the MHC. For this reason, it seems likely that selection at MHC loci is overdominant, as originally proposed by Doherty and Zinkernagel (12).

3.3. Introns of Class I Genes

Recently sequences of introns from the human class I MHC loci *HLA-A*, *-B*, and *-C* have become available (23). Comparison of these introns with the

Natural selection and the evolutionary history of MHC loci

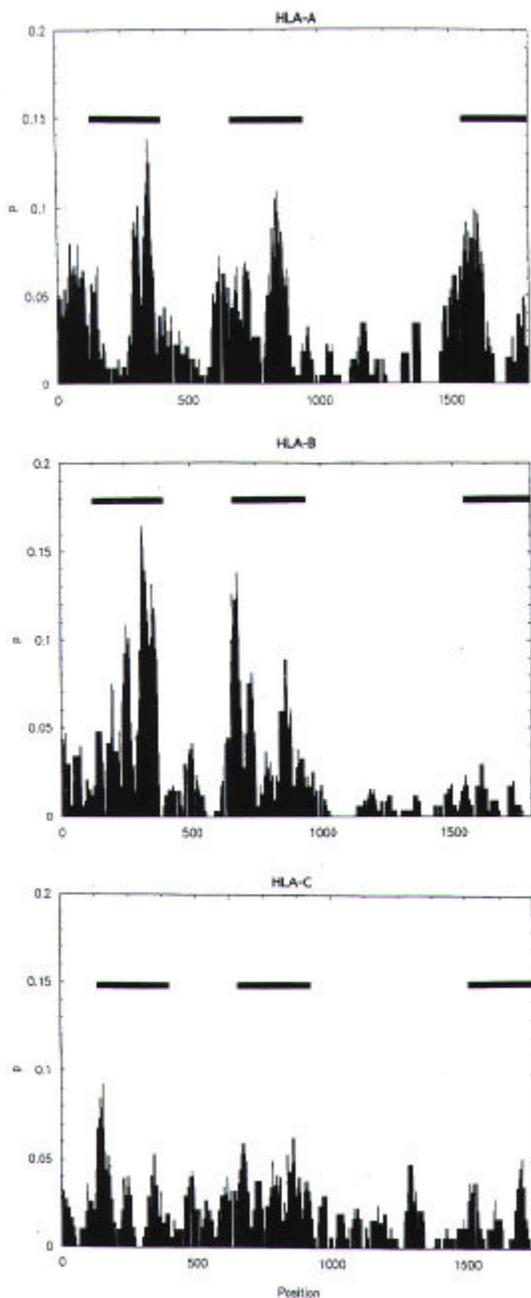


Figure 4. Proportion of nucleotide difference (p) in a sliding window of 30 base pairs in all pairwise comparisons among HLA-A, -B, and -C alleles. Horizontal bars indicate the positions of exons 2, 3, and 4. From ref. 23.

adjoining exons reveals an additional line of evidence that polymorphisms at these loci are maintained by balancing selection. Figure 4 shows the proportion of nucleotide difference in a sliding window of 30 aligned base pairs along the sequence of alleles at the human class I loci *HLA-A*, *-B*, and *-C* in the region from intron 1 to exon 4. Exons 2 and 3 encode the $\alpha 1$ and $\alpha 2$ domains, respectively, of the

protein. The $\alpha 1$ and $\alpha 2$ domains include the codons of the PBR, and most polymorphism is found in these exons; exon 4 encodes the conserved $\alpha 3$ domain. The exons, particularly exons 2 and 3, show substantially more sequence divergence than do the introns, particularly the long third intron.

This pattern of nucleotide divergence in exons and introns closely fits what is expected under balancing selection (24-26). As first shown by Hughes and Nei (15) the balancing selection at these loci operates on nonsynonymous sites in the PBR codons. Being closely linked to these sites, other sites in exons 2 and 3 are expected to “hitch-hike” along with these selected sites. Thus the polymorphism in these exons will be relatively ancient. Introns, whose linkage to the PBR codons is less tight, will have less ancient polymorphism than exons because introns have been homogenized relative to the exons by recombination and subsequent genetic drift (23). The fact that the long third intron is the most homogenized among alleles (figure 3) provides support for this model because linkage between sites in this intron is expected to be less than in the shorter introns 1 and 2 (23).

4. Duplication of Loci

It is by now well established that gene duplication plays a major role in the adaptive evolution of organisms (27,28), and patterns of gene duplication have been reconstructed for a number of multi-gene families of eukaryotic organisms. In the case of the MHC, an ancestral gene duplication early in vertebrate history gave rise to the class I and class II MHC loci. This event must have also included rearrangement of exons, since class I and class II molecules have different structures (29). In mammals, phylogenetic analyses have revealed different patterns of gene duplication in class I and class II (30,31). In the class I MHC of mammals, duplication and deletion of genes has occurred frequently so that orthologous relationships do not exist between class I loci of mammals of different orders (30). The process of duplication and deletion of loci occurs more slowly in the class II MHC. Orthologous relationships do occur between class II loci of mammals belonging to different eutherian (placental) orders (31). However, no orthologous relationships are known between class II MHC loci in different classes of vertebrates.

The class I MHC of mammals includes a number of different loci differing in expression and function. The class Ia loci or classical class I loci are typically highly polymorphic; they are expressed in all nucleated somatic cells and function to present peptides to cytotoxic T cells (11). The class Ia loci are characterized by an enhanced rate of nonsynonymous nucleotide substitution in the PBR codons, indicating that polymorphism at these loci is selectively maintained. In addition, there are certain loci expressed at lower levels, which also show relatively low levels of polymorphism; these are known as class Ib or nonclassical class I loci (32). The function of class Ib loci is not clearly understood, although at least some of them have been shown to bind peptides (33). The class Ib loci do not have an enhanced rate of nonsynonymous substitution in the PBR codons; thus what little

Natural selection and the evolutionary history of MHC loci

polymorphism is seen at these loci appears to be neutral polymorphism (30).

From an evolutionary point of view, the most remarkable feature of class Ib loci is that they do not show orthologous relationships between mammals of different orders. Thus these loci have arisen independently in different mammalian lineages (30). Class Ib loci have evidently evolved as a result of duplication of class Ia loci followed by a change in expression pattern due to mutation in the promoter region (30). The class Ia and class Ib genes thus do not represent mutually exclusive groups over evolutionary time. Dramatic evidence for this is the fact that the class Ia genes of New World primates are homologous to the human class Ib locus HLA-G (34,35). This locus was evidently present in the common ancestor of New World monkeys and Old World monkeys, apes, and hominids, but in the former it has been duplicated to form separate class Ia loci. In spite of the independent origin of class Ib loci in different mammals, there is evidence that they can evolve similar functions convergently. For example the class Ib HLA-E molecule of humans and other primates has convergently evolved features of the PBR that are similar to those of the mouse class Ib molecule H2-Qa-1a (36), and there is evidence that these molecules may bind similar peptides (37).

5. THE ROLE OF RECOMBINATION

5.1 Intralocus Recombination in the Class I MHC

Alignment of DNA sequences of MHC genes has suggested the possibility that recombination has played a role in the evolution of this gene family. Both recombination among alleles at a locus (intralocus or interallelic recombination) and recombination between loci (interlocus recombination) have been proposed as mechanisms in the diversification of these genes. However, these hypotheses have been controversial. First, it has been noted that many of the cases that have been attributed to recombination might be just as easily explained by convergent or parallel evolution (38). Natural selection can greatly increase the probability of convergent or parallel evolution over the neutral case. Because of the positive selection acting on PBR codons of polymorphic MHC loci, convergent or parallel evolution would not be surprising in this case, and indeed there is evidence of convergent evolution of similar sequence motifs in the class II PBR of mammals of different orders (39).

In addition, most proposed cases of recombination have not been subjected to any statistical test but have only involved visual inspection of sequences. It is desirable to have some way of discriminating statistically between the amount of similarity between two sequences that might be observed by chance and the degree of similarity that indicates a past recombination event. Unfortunately, although a number of methods have been proposed to test for recombination in molecular sequence data, there may be problems with applying them to the MHC because the balancing selection at these loci often causes violations of the methods' underlying assumptions.

We can distinguish two types of inter-allelic recombination: (i) large-scale recombination involves the exchange of a long stretch of sequence between two alleles,

such as may occur by an ordinary crossing-over mechanism; and (ii) small-scale recombination involves the exchange of a short sequence motif or "mini-cassette" and may involve a "gene conversion" mechanism. A number of fairly reliable methods can be used to test for large-scale recombination, including construction of phylogenetic trees for different gene regions and estimation of numbers of nucleotide substitutions per site in different gene regions (40). This approach cannot be used in the case of small-scale recombination. Available methods, such as that of Stephens (41), identify clusters of sites shared between sequences. If these sites are nonsynonymous, such a cluster might result from convergent evolution at the amino acid level, rather than recombination. Thus, this method cannot always discriminate between the hypotheses of recombination and gene conversion. Nonetheless, in some cases, recombination "mini-cassettes" involving synonymous sites have been identified, giving strong support to the hypothesis of interallelic recombination (40).

Although it may be difficult to decide in some specific cases whether or not interallelic recombination has been involved, studies employing a variety of methods have been able to detect major trends. One type of trend that has been observed is a difference between loci. In the case of the class I MHC of humans, it is clear that interallelic recombination has played a far more prominent role in the history of the HLA-B locus than of either HLA-A or HLA-C (40,42,43). It is so far uncertain whether this difference has occurred because the actual rate of recombination is higher at the HLA-B locus than at the other two loci or whether more recombinants at the HLA-B locus have been selectively favored and thus have increased in the population. Some evidence favors the latter hypothesis. First, putative recombination events that have occurred at human class I loci disproportionately involve the PBR codons, as is consistent with their being selectively favored. Second, although HLA-C is closely related to HLA-B, it has a low level of recombination like HLA-A. This might not be expected if the high observed rate of recombination at HLA-B is simply a consequence of some aspect of its sequence, since HLA-B and HLA-C are similar in sequence (43).

Immunologists may have a difficult time understanding the fact that interallelic recombination leads to homogenization within loci in class I introns (23), but can enhance diversity in exons. However, this is exactly what population genetics theory predicts in the case of balancing selection; it certainly would not be true in the absence of such selection. In introns, recombination breaks up linkage of intron sites with nonsynonymous PBR sites that are under balancing selection, thus allowing genetic drift to reduce sequence diversity in introns. In exons, polymorphism has been selectively favored, evidently because of the wide immune surveillance of heterozygotes, and over time this selection has given rise to numerous quite distinct alleles. A recombinant between two very different alleles may create a new form of the PBR, which may confer an immediate selective advantage; and such a favored recombinant will increase in frequency. Thus, recombination alone does not enhance diversity; only

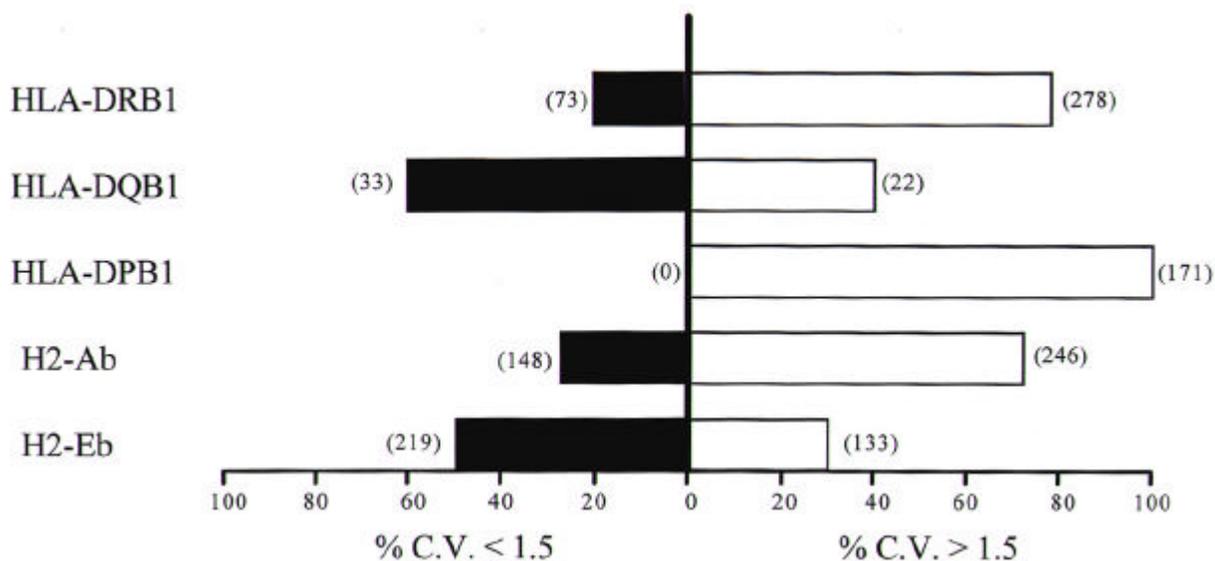


Figure 5. Coefficient of variation (C.V.) in proportion of nucleotide difference in a sliding window of 10 in exon 2 (excluding PBR codons) in pairwise comparisons among human (HLA-) and mouse (H2-) class II b chain alleles. For each locus, the data are represented as the percentage of comparisons with a C.V. greater than 1.5 and the proportion of comparisons with a C.V. < 1.5. Loci with greater percentages of comparisons with C.V. > 1.5 show greater incidence of interallelic recombination.

recombination combined with balancing selection as seen in the case of the MHC PBR.

5.2 Intralocus Recombination the the Class II MHC

So far, the levels of recombination at MHC class II loci have not been studied as extensively as those at class I loci. We developed a simple method to address this problem. For a given pair of alleles, we computed the proportion of nucleotide difference (p) within a sliding window of 10 nucleotide positions in exon 2 of the gene, which includes the PBR codons. The coefficient of variation (C.V.) of p was used as an index of the non-homogeneity of the difference between the two sequences. This in turn provided an indication of the rate of past recombination. For example, two sequences very similar in one region but very different in another would have a high C.V. of p ; and such a pattern might have resulted from recombination. However, non-homogeneity of p could conceivably be caused by factors other than recombination. In the case of the MHC, the main such factor is natural selection enhancing the rate of non-synonymous substitution in the PBR codons. Therefore, we removed the PBR codons prior to this analysis.

The results (figure 5) show marked differences among the class II b chain loci of human and mouse with respect to the C.V. of p and thus the putative level of past recombination. In humans, these results suggest that there has been a much higher level of recombination at HLA-DRB1 and HLA-DPB1 than at HLA-DQB1 (figure 5). In the mouse, they suggest that there has been a much greater rate of recombination at H2-Ab than at H2-Eb (figure 5). These conclusions are consistent with those of previous studies using other methods. For example, phylogenetic

analysis of different portions of exon 2 suggested that recombination has occurred at high levels in the human HLA-DRB1 locus (44) and the mouse H2-Ab locus (45).

5.3 Interlocus Recombination

There is much less evidence of interlocus recombination in the MHC than of intralocus recombination. However, as in most multi-gene families, sequence analyses have suggested that such events have occurred in the past. For example the human class Ia HLA-A locus arose from a recombination between (1) a 5' region (exon 1 through exon 3) of a gene related to the other human class Ia loci HLA-B and -C; and (2) a 3' region (from intron 3 to the end of the gene) from a gene closely related to the human class I pseudogene HLA-70 (46).

A persistent theme in the history of MHC biology has been the search for amazing or unusual molecular mechanisms to explain the high polymorphism of these loci. For example, it was proposed that MHC loci have a high mutation rate (47). Now it is known that in fact the mutation rate at these loci is lower than the mammalian average (9). Similarly, it was argued that some unusual mechanism of interlocus recombination must exist to create diversity at these loci (48), but sequence analyses have failed to support this hypothesis (49). Rather, the MHC seems to represent a rather ordinary multi-gene family in most respects. The only really dramatic difference is that the MHC includes some loci that are subject to balancing (most likely overdominant) selection of a type that has persisted since the origin of the MHC itself early in vertebrate history and most likely will persist as long as vertebrates are exposed to parasites of any sort. All unique

Natural selection and the evolutionary history of MHC loci

or unusual features of the MHC genes are best understood as consequences of that selection.

6. ACKNOWLEDGMENTS

This research was supported by grants R01-GM34940 and K04-GM00614 from the National Institutes of Health.

7. REFERENCES

1. M. Kimura: The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge (1983)
2. M. Nei: Molecular Evolutionary Genetics. Columbia University Press, New York (1987)
3. A.L. Hughes, M Yeager & M. Carrington: Peptide-binding function and the paradox of HLA disease associations. *Immunol Cell Biol* 74, 444-448 (1996)
4. M. Kimura: Evolutionary rate at the molecular level. *Nature* 217, 624-626 (1968)
5. M. Kimura & T. Ohta: The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61, 763-771 (1969)
6. P.W. Hedrick & G. Thomson: Evidence for balancing selection at HLA. *Genetics* 104, 449-456 (1983)
7. L Thomas: Biological signals for self-identification. In: Progress in Immunology II. Eds: Brent L, Holborrow J, North-Holland, Amsterdam (1974)
8. M Nei & AL Hughes: Polymorphism and evolution of the major histocompatibility complex loci in mammals. In: Evolution at the Molecular Level. Eds: Selander RK, Whittam TS, Clarke AG, Sinauer, Sunderland MA (1990)
9. A.L. Hughes & M.K. Hughes: Natural selection on the peptide-binding regions of the major histocompatibility complex molecules. *Immunogenetics* 42, 233-243 (1995)
10. R.M. Zinkernagel & P.C. Doherty: Immunological surveillance against altered self components by sensitized T lymphocytes in lymphocytic choriomeningitis. *Nature* 251, 547-548 (1974)
11. J. Klein: The Natural History of the Major Histocompatibility Complex. Wiley, New York (1986)
12. P.C. Doherty & R. Zinkernagel: Enhanced immune surveillance in mice heterozygous at the H-2 gene complex. *Nature* 256, 50-52 (1975)
13. A.L. Hughes & M. Nei: Models of host-parasite interaction and MHC polymorphism. *Genetics* 132, 863-864 (1992)
14. H.-G. Rammensee, T. Friede & S. Stevanovic. MHC ligands and peptide motifs: first listing. *Immunogenetics* 41, 178-228 (1995)
15. A.L. Hughes & M. Nei: Pattern of nucleotide substitution at MHC class I loci reveals overdominant selection. *Nature* 335, 167-170 (1988)
16. A.L. Hughes & M. Nei: Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 86, 958-962 (1989)
17. A.L. Hughes, M.K. Hughes, C.Y. Howell & M. Nei: Natural selection at the class II major histocompatibility complex loci of mammals. *Phil Trans R Soc Lond B* 345, 359-367 (1994)
18. T. Maruyama & M. Nei: Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* 98, 441-459 (1981)
19. W.E. Mayer, D. Jonker, D. Klein, P. Ivanyi, G. van Seventer & J. Klein: Nucleotide sequence of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J* 7, 2765-2774 (1988)
20. D.A. Lawlor, F.E. Ward, P.D. Ennis, A.P. Jackson & P. Parham: HLA-A,-B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 335, 268-271 (1988)
21. U.B. Gyllensten & H.A. Erlich. Ancient roots for polymorphism at the HLA-Dq locus in primates. *Proc Natl Acad Sci USA* 86, 9986-9990 (1989)
22. N. Takahata & M. Nei: Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124, 967-978 (1990)
23. N. Cereb, A.L. Hughes & S.Y. Yang: Locus-specific conservation of the HLA Class I introns by intra-locus homogenization. *Immunogenetics* 47, 30-36 (1997)
24. M. Nei & W.-H. Li: Non-random association between electromorphs and inversion chromosomes in finite populations. *Genet Res* 35, 65-83 (1980)
25. C. Strobeck: Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* 103, 545-555 (1983)
26. M. Kreitman & R.R. Hudson: Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* 127, 565-582 (1991)
27. S Ohno: Evolution by Gene Duplication. Springer-Verlag, New York (1970)

Natural selection and the evolutionary history of MHC loci

28. A.L. Hughes: The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B* 256, 119-124 (1994)
29. A.L. Hughes & M. Nei: Evolutionary relationships of the classes of major histocompatibility complex genes. *Immunogenetics* 37, 337-346 (1993)
30. A.L. Hughes & M. Nei: 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)
31. A.L. Hughes & M. Nei: Evolutionary relationships of class II MHC genes in mammals. *Mol Biol Evol* 7, 491-514 (1990)
32. J. Klein & F. Figueroa: Evolution of the major histocompatibility complex. *CRC Crit Rev Immunol* 6, 295-386 (1986)
33. E.G. Pamer, C.-R. Wang, L. Flaherty, K. Fischer-Lindahl & M.J. Bevan: H-2M3 presents a *Listeria monocytogenes* peptide to cytotoxic T lymphocytes. *Cell* 70: 215-223 (1992)
34. D.I. Watkins, Z.W. Chen, A.L. Hughes, M.G. Evans, T.F. Tedder & N.L. Letvin: Evolution of the MHC class I genes of a New World primate from ancestral homologues of human non-classical genes. *Nature* 346, 60-63 (1990)
35. L.F. Cadavid, C. Shufflebotham, F.J. Ruiz, M. Yeager, A.L. Hughes & D.I. Watkins: Evolutionary instability of the major histocompatibility complex loci in New World primates. *Proc Natl Acad Sci USA* 94, 14536-14541 (1997)
36. M. Yeager, S. Kumar & A.L. Hughes: Sequence convergence in the peptide-binding region of primate and rodent MHC class Ib molecules. *Mol Biol Evol* 14, 1035-1041 (1997)
37. A. DeCloux, A.S. Woods, R.J. Cotter, M.J. Soloski & J. Forman: Dominance of a single peptide bound to the class Ib molecules, Qa1b. *J Immunol* 158, 2183-2191 (1997)
38. J. Klein & C. O'hUigin: Class II B *Mhc* motifs in an evolutionary perspective. *Immunol Rev* 143, 235-243 (1991)
39. L. Andersson, S. Sigurdardottir, C. Borsch & K. Gustafsson: Evolution of MHC polymorphism: extensive sharing of sequence motifs between human and bovine DRB alleles. A.L. Hughes, M.K. Hughes & D.I. Watkins: Contrasting roles of interallelic recombination at the HLA-A and HLA-B loci. *Genetics* 133, 669-680 (1993)
40. J.C. Stephens: Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. *Mol Biol Evol* 2, 539-556 (1985)
41. S.T. McAdam, J.E. Boyson, X.Liu, T.L. Garber, A.L. Hughes, R.E. Bontrop & D.I. Watkins: A uniquely high level of recombination at the *HLA-B* locus. *Proc Natl Acad Sci USA* 91, 5893-5897 (1994)
42. M. Yeager & A.L. Hughes: Interallelic recombination has not played a major role in the history of the *HLA-C* locus. *Immunogenetics* 44, 128-133 (1996)
44. U.B. Gyllensten, M. Sundvall & H.A. Erlich: Allelic diversity is generated by intraexon sequence exchange at the DRB1 locus of primates. *Proc Natl Acad Sci USA* 88, 3686-3690 (1991)
45. J.X. She, S.A. Boehme, T.W. Wang, F. Bonhomme & E.K. Wakeland: Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. *Proc Natl Acad Sci USA* 88, 453-457 (1991)
46. A.L. Hughes: Origin and evolution of HLA class I pseudogenes. *Mol Biol Evol* 12, 247-258 (1995)
47. D.W. Bailey & H.I. Kohn: Inherited histocompatibility changes in progeny of irradiated and unirradiated inbred mice. *Genet Res* 6, 330-340 (1965)
48. T. Ohta: Some models of gene conversion for treating the evolution of multigene families. *Genetics* 106, 517-528 (1984)
49. M. Nei, X. Gu & T. Sitnikova: Evolution by the birth-and-death process in multi-gene families of the vertebrate immune system. *Proc Natl Acad Sci USA* 94, 7799-7809 (1997)
50. M. Nei & T. Gojobori: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3, 418-426 (1986)
51. M. Nei & L. Jin: Variances of the average numbers of nucleotide substitutions within and between populations. *Mol Biol Evol* 6, 209-300 (1989)
52. N. Saitou & M. Nei: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425 (1987)

Key Words: HLA-DR, Major histocompatibility complex, MHC, T cells, polymorphism

Send correspondence to: Dr Austin L. Hughes, Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park PA 16802 USA, Tel:(814)-865-5013, Fax:(814)-865-9131, E-mail: austin@hugans3.bio.psu.edu