PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Natural Selection at the Class II Major Histocompatibility Complex Loci of Mammals [and Discussion]

Austin L. Hughes, Marianne K. Hughes, Carina Y. Howell, Masatoshi Nei, J. C. Howard and P. Higgs

Phil. Trans. R. Soc. Lond. B 1994 **346**, 359-367
doi: 10.1098/rstb.1994.0153

Natural selection at the class II major histocompatibility complex loci of mammals

AUSTIN L. HUGHES, MARIANNE K. HUGHES, CARINA Y. HOWELL AND MASATOSHI NEI

Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park, PA 16802, U.S.A.

SUMMARY

The role of natural selection at major histocompatibility complex (MHC) loci was studied by analysis of molecular sequence data from mammalian class II MHC loci. As found previously for the class I MHC molecule and a hypothetical model of the class II molecule, the rate of non-synonymous nucleotide substitution exceeded that of synonymous substitution in the codons encoding the antigen recognition site of polymorphic class II molecules. This pattern is evidence that the polymorphism at these loci is maintained by a form of balancing selection, such as overdominant selection. By contrast, in the case of monomorphic class II loci, no such enhancement of the rate of non-synonymous substitution was observed. Phylogenetic analysis indicates that, in contrast to monomorphic ('non-classical') class I MHC loci, some monomorphic class II loci of mammals are quite ancient. The *DMA* and *DMB* loci, for example, diverged before all other known mammalian class II loci, possibly before the divergence of tetrapods from bony fishes. Analysis of the patterns of sharing of polymorphic residues at class II MHC loci by mammals of different species revealed that extensive convergent evolution has occurred at these loci; but no support was found for the hypothesis that MHC polymorphisms have been maintained since before the divergence of orders of eutherian mammals.

1. INTRODUCTION

The major histocompatibility complex (MHC) is a multi-gene family whose products are cell-surface glycoproteins that function to present intracellularly processed peptides to T cells. The MHC genes are remarkable for several reasons. Their most striking characteristic is the high level of polymorphism found at certain MHC loci, which are among the most polymorphic known in organisms. Furthermore, certain of these loci are characterized by 'trans-species' polymorphism (Lawlor *et al.* 1988; Mayer *et al.* 1988); that is, polymorphic allelic lineages may be shared by related species, such as human and chimpanzee, having apparently been maintained in both species since before speciation. By contrast, certain MHC loci are monomorphic or nearly so, and the function of most of the monomorphic MHC loci is unknown.

In the following, after a brief introduction to the structure and function of MHC molecules, we present evidence from statistical analysis of DNA sequence data that addresses a number of questions regarding MHC evolution. While reviewing previous studies of class I MHC molecules, we present the results of new analyses of class II MHC sequence data. We address the following questions: (1) the mechanism of maintenance of MHC polymorphism; (2) the evolution of monomorphic class II MHC loci; and (3) the contributions of common ancestry and convergent evolution to sharing of polymorphic amino acid motifs among different mammalian species. Because extensive sequence data are available from orthologous class II loci from several orders of mammals, data from the class II MHC are particularly appropriate for addressing this last question.

2. STRUCTURE AND FUNCTION OF MHC MOLECULES

Class I MHC molecules, which are expressed on almost all kinds of nucleated cells, present peptides to cytotoxic T cells (Bjorkman & Parham 1990; Klein 1986). Class II molecules have a much more restricted expression, being expressed primarily on antigen-presenting cells of the immune system. The complex of self class II MHC molecule and foreign peptide on an antigen-presenting cell is recognized by a helper T cell; the latter releases lymphokines that stimulate B cells (which produce antibodies) and macrophages to respond to the infection.

Both class I and class II molecules are heterodimers with four extracellular domains, but they achieve a similar molecular structure in different ways. The class I α chain includes three extracellular domains

Phil. Trans. R. Soc. Lond. B (1994) **345**, 359–367
Printed in Great Britain

($\alpha_1, \alpha_2, \alpha_3$). The α_3 domain associates non-covalently with β_2-microglobulin. β_2-microglobulin is encoded by a gene outside the MHC region; but the gene has a distant evolutionary relation to MHC genes. Portions of the α_1 and α_2 domains form a groove at the top of the molecule called the antigen recognition site (ARS), in which peptides are bound and presented to T cells (Bjorkman *et al.* 1987*a*,*b*). The ARS groove is formed by two α helices bounding a β pleated sheet.

The class II heterodimer consists of an α chain with two extracellular domains (α_1 and α_2) and a β chain with two extracellular domains (β_1 and β_2), both encoded by genes within the MHC region. A hypothetical structure for the class II molecule was initially proposed on the basis of analogy with the class I structure (Brown *et al.* 1988). Recently, the structure of the HLA–DR1 class II heterodimer has been worked out (Brown *et al.* 1993), and it is similar to the previous hypothetical model. The class II molecule has an ARS similar to that of the class I molecule; but in this case one α helix and about half of the β pleated sheet are contributed by the α_1 domain of the α chain, while the β_1 domain of the β chain contributes the other α helix and the remainder of the β pleated sheet.

The class II MHC genes of eutherian mammals are arranged in a number of separate chromosomal regions, each of which typically contains at least one α chain gene and one or more β chain genes. The genes encoding α chains are generally designated *A* (e.g. *DRA*), and those encoding β chains are designated *B* (e.g. *DRB1*). In humans, there are three regions including polymorphic class II loci, designated *DR*, *DQ* and *DP*. Each of these regions contains at least one polymorphic β chain locus. In the *DR* region, there are four functional β chain loci (*DRB1*, *DRB3*, *DRB4* and *DRB5*). Between one and three of these β chain loci are present in an individual haplotype. The *DQ* and *DP* regions each contain a single functional β chain locus (*DQB1* and *DPB1*, respectively). Only in the *DQ* region is there a substantial degree of polymorphism in the α chain gene (*DQA1*). However, sequences of a number of alleles are also available for *DPA1*.

In a number of other eutherian mammals, homologues of the human class II loci have been discovered, and these are usually named following their human homologues. However, because the class II loci of the mouse were discovered independently of those of humans, a different system of nomenclature has been used for genes in the mouse MHC (called the *H-2* complex). The symbols *a* and *b* are used respectively for class II α and β chain genes. The regions corresponding to human *DR* and *DQ* are known as *E* and *A*, respectively; the mouse lacks functional *DP* homologues. Thus the mouse locus homologous to *HLA–DRA* is called *H-2Ea*.

3. MAINTENANCE OF MHC POLYMORPHISM

When the function of the class I MHC in immune recognition was worked out, its role in antigen presentation suggested to Doherty & Zinkernagel (1975) a mechanism whereby overdominant selection (heterozygote advantage) might operate at MHC loci. They had evidence that products of different alleles at MHC loci differ in their ability to bind and present specific foreign peptides. Given such differences, in a population exposed to a variety of pathogens, an individual heterozygous at all or most MHC loci would presumably be able to present more kinds of foreign peptides than a homozygote. Thus, a heterozygote would be resistant to a wider array of pathogens than would a homozygote.

Hughes & Nei (1988, 1989*a*) approached the question of selection at MHC loci by comparing rates of synonymous and non-synonymous (amino acid altering) nucleotide substitution. They used the method of Nei & Gojobori (1986) to estimate the number of synonymous nucleotide substitutions per synonymous site (d_S) and the number of non-synonymous nucleotide substitutions per non-synonymous site (d_N) in different regions of MHC genes. Hughes & Nei (1988) reasoned that, if MHC polymorphisms are maintained by overdominant selection relating to peptide binding and thus to pathogen resistance, such selection would act to favour amino acid differences in the ARS. Therefore d_N should exceed d_S in the codons encoding the ARS. This pattern is the reverse of that seen in most functional genes, in which d_S exceeds d_N. This occurs because most amino acid changes disrupt protein structure and are harmful to the organism, and thus most non-synonymous mutations will be eliminated by conservative or 'purifying' natural selection.

As predicted, Hughes & Nei (1988) observed values of d_N significantly higher than d_S in the class I ARS. By contrast, they found that d_S generally exceeded d_N in the rest of the class I gene. Thus, the class I gene is subject to overdominant selection or some similar form of balancing selection, favouring diversity in the ARS, whereas the remainder of the molecule is subject to purifying selection. Similar results were found for class II genes (Hughes & Nei 1989*b*). In the latter, although a crystallographic structure was not available, a hypothetical structure based on analogy with the class I molecule had been proposed (Brown *et al.* 1988).

Here we present the results of new analyses of class II MHC sequences from human and mouse, based on the HLA–DR1 structure (Brown *et al.* 1993). Tables 1 and 2 show estimates for mean d_S and d_N in pairwise comparisons among alleles from polymorphic human and mouse class II MHC. These values were estimated separately for the ARS codons; for the remainder of D1 (α_1 or β_1), excluding the ARS codons; and for D2 (α_2 or β_2). At polymorphic β chain loci in both species, d_N is enhanced in the ARS (table 1). The same pattern is seen at the *Aa* locus of the mouse, but not at *DPA1* or *DQA1* of humans (table 1). For *DPA1* there is no significant difference between d_S and d_N in the ARS (table 1). Overall the level of sequence divergence among *DPA1* alleles is low compared with that found at other human polymorphic class II loci. For example, the d_S values for the non-ARS portion of Domain 1 and for Domain 2 are lower for

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *Mean numbers of synonymous (d_S) and non-synonymous (d_N) nucleotide substitutions per 100 sites among class II MHC alleles of human and mouse*

		antigen recognition site		remainder, D1		D2	
locus	(*n*)	d_S	d_N	d_S	d_N	d_S	d_N
β chain loci							
mouse							
Ab	(9)	0.6 ± 0.8	$21.7 \pm 4.0^{***}$	4.3 ± 1.9	6.0 ± 1.3	9.2 ± 2.5	$1.2 \pm 0.5^{**}$
Eb	(6)	5.2 ± 3.8	$22.9 \pm 4.4^{**}$	1.3 ± 1.0	3.1 ± 0.9	1.0 ± 0.8	0.5 ± 0.3
human							
DPB1	(6)	1.7 ± 2.4	$9.8 \pm 3.0^{*}$	1.8 ± 1.4	1.9 ± 0.7	4.3 ± 1.9	0.7 ± 0.4
DQB1	(9)	7.3 ± 4.2	$19.3 \pm 3.6^{*}$	11.1 ± 2.9	5.7 ± 1.2	5.4 ± 2.0	1.7 ± 0.6
DRB1	(23)	6.1 ± 3.9	$24.7 \pm 3.5^{***}$	9.5 ± 2.5	$3.4 \pm 0.8^{*}$	6.1 ± 1.7	$2.3 \pm 0.6^{*}$
DRB3	(4)	4.5 ± 4.5	10.3 ± 3.3	1.3 ± 1.3	1.2 ± 0.7	4.3 ± 1.9	0.9 ± 0.5
DRB5	(4)	4.3 ± 4.4	10.8 ± 3.7	2.7 ± 1.9	0.8 ± 0.6	2.7 ± 1.6	0.6 ± 0.4
all *DRB*	(32)	5.6 ± 3.9	$26.9 \pm 3.3^{***}$	9.7 ± 2.2	$3.8 \pm 0.7^{*}$	7.4 ± 1.7	$2.7 \pm 0.6^{**}$
α chain loci							
mouse							
Aa	(6)	3.6 ± 3.3	$15.8 \pm 3.7^{*}$	2.7 ± 1.6	4.1 ± 1.0	7.6 ± 2.2	$0.7 \pm 0.4^{***}$
human							
DPA1	(4)	0.0 ± 0.0	2.1 ± 1.5	5.1 ± 2.6	0.7 ± 0.5	2.4 ± 1.4	0.7 ± 0.4
DQA	(6)	23.2 ± 12.0	7.2 ± 2.7	9.7 ± 3.0	8.4 ± 1.7	4.1 ± 1.7	1.4 ± 0.5

Tests of the hypothesis that $d_S = d_N$: $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. D1 is β_1 or α_1 domain; D2 is β_2 or α_2 domain. Comparison of all *DRB* includes one *DRB4* sequence. *n*, Number of sequences.

comparisons among *DPA1* alleles than for most comparisons among alleles at other human class II loci analysed (table 1). Thus, the *DPA1* alleles seem to have diverged from each other more recently than have alleles at the other loci. These results suggest that *DPA1* polymorphism is probably selectively neutral.

The case of *DQA1* is somewhat different. At this locus, d_S and d_N do not differ significantly in the ARS (table 1). However, as measured by d_S values among alleles, *DQA1* alleles are quite divergent from each other (table 1), suggesting that polymorphism at this locus has been maintained for a long time. The existence of trans-species polymorphism at this locus in humans and other Old World primates was supported by phylogenetic analyses (Gyllensten & Erlich 1989; Nei & Rzhetsky 1991). At this locus in humans and great apes, alleles belong to four lineages, within which there are minor variants; and these lineages seem to have been maintained at least since before the divergence of human and gorilla (Nei & Rzhetsky 1991). Thus at the *DQA1* locus of higher primates a somewhat different type of balancing selection may be occurring than at other polymorphic MHC loci. This selection may act to maintain the four allelic lineages but may not favour major new allelic forms of the ARS. By contrast, the homologous *Aa* locus of the mouse shows a pattern of nucleotide substitution similar to that found at polymorphic β

Table 2. *Numbers of synonymous (d_S) and non-synonymous (d_N) nucleotide substitutions per 100 sites in comparisons between orthologous class II MHC loci of human and mouse*

		antigen recognition site		remainder, D1		D2	
comparisons	(*n*)	d_S	d_N	d_S	d_N	d_S	d_N
β chain loci							
monomorphic loci							
DMB vs. *Mb*	(2)	56.8 ± 25.3	31.0 ± 9.1^{b}	76.9 ± 20.6	$17.2 \pm 3.6^{**}$	61.8 ± 13.3	$11.9 \pm 2.5^{***}$
DOB vs. *Ob*	(2)	45.2 ± 20.4	7.8 ± 3.9	58.1 ± 15.7	$13.7 \pm 3.1^{**}$	84.6 ± 18.7	$22.7 \pm 3.6^{**}$
polymorphic loci							
DQB vs. *Ab*	(18)	54.1 ± 21.7	32.5 ± 4.9^{c}	30.4 ± 7.9	17.6 ± 8.9	53.6 ± 11.0	$10.5 \pm 2.2^{***}$
DRB vs. *Eb*	(38)	47.8 ± 20.8	51.2 ± 8.5^{c}	42.4 ± 10.3	$11.8 \pm 2.4^{**}$	54.5 ± 11.5	$9.3 \pm 2.0^{***}$
α chain loci							
monomorphic loci							
DMA vs. *Ma*	(2)	47.5 ± 26.0	21.4 ± 8.0^{a}	79.1 ± 20.0	$15.9 \pm 3.2^{**}$	71.8 ± 14.6	$9.7 \pm 2.1^{***}$
DNA vs. *Na*	(2)	98.4 ± 53.3	19.8 ± 7.1^{a}	59.6 ± 16.2	$11.5 \pm 3.0^{**}$	51.4 ± 11.5	$8.2 \pm 2.0^{***}$
DRA vs. *Ea*	(2)	37.5 ± 23.8	2.0 ± 2.0	120.5 ± 38.2	$16.5 \pm 3.6^{**}$	60.3 ± 13.3	$11.9 \pm 2.5^{***}$
polymorphic locus							
DQA vs. *Aa*	(12)	71.0 ± 33.4	29.3 ± 7.2^{c}	74.1 ± 18.0	$25.6 \pm 4.1^{**}$	33.0 ± 7.6	$15.9 \pm 2.9^{*}$

Tests of the hypothesis that $d_S = d_N$: $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. Tests of the hypothesis that d_N equals that for *DRA* versus *Ea* comparison: $^{a}P < 0.05$; $^{b}P < 0.01$; $^{c}P < 0.001$. *n*, Number of sequences.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

chain loci. Here d_N is significantly higher than d_S in the ARS (table 1), which suggests that selection is acting to favour amino acid diversity in this region, as with the polymorphic β chain loci.

4. EVOLUTION OF MONOMORPHIC LOCI

Class I MHC molecules are divided into two subgroups: (1) the class Ia or class I 'classical' loci, which are typically polymorphic and are expressed on almost all cells; and (2) the class of Ib or 'non-classical' class I loci, which are monomorphic or have very low polymorphism and have a much restricted tissue distribution (Klein & Figueroa 1986; Howard 1987). The function of the class Ib loci remains mysterious, although it is now known that at least some of them can have an antigen-presenting function (Pamer *et al.* 1992). Phylogenetic analysis shows that class Ib genes of mammals of one order are more closely related to class Ia genes of the same order than they are to class Ib genes of mammals of other orders (Hughes & Nei 1989*b*). This indicates that class Ib loci have evolved independently in different orders of mammals.

The class II MHC also includes certain monomorphic or nearly monomorphic loci. Certain α chain loci (such as human *DRA*) are monomorphic but form heterodimers with highly polymorphic β chains. In addition, there are also some class II loci that seem more closely analogous with class I 'non-classical' loci in that they are monomorphic and lack a known function. In the mouse the *Na* and *Ob* genes encode a nearly monomorphic heterodimer (Karlsson & Peterson 1992). Homologues of these genes (*DNA* and *DOB*) exist in humans, but in this case it has been suggested that the heterodimer is not expressed because of defects in the *DNA* mRNA (Trowsdale & Kelly 1985). The *DMA* and *DMB* genes (*Ma* and *Mb* in mouse) constitute another pair of genes encoding an almost monomorphic heterodimer (Kelly *et al.* 1991; Cho *et al.* 1991).

To understand phylogenetic relations among polymorphic and monomorphic class II MHC loci in mammals, we constructed a phylogenetic tree based on the conserved Domain 2 amino acid sequences of α and β chain genes (figure 1). Since the gene duplication that gave rise to separate α and β chains occurred early in the history of the MHC, the α chain tree serves to root the β chain tree, and vice versa. As with previous phylogenetic analyses (Hughes & Nei 1990), the tree shows that the different class II regions diverged prior to the divergence of the eutherian orders; thus, for example, *DRB* genes from Primates, Rodentia, Carnivora and Artiodactyla all cluster together, and the cluster is supported by a statistically significant branch (figure 1). Indeed some of these regions may have diverged before the divergence of marsupial (red-necked wallaby) and eutherian mammals. One of two marsupial β chain genes clusters with the *DRB* genes of placental mammals, while the other clusters with *DQB* (figure 1).

The tree shows a contrast between monomorphic class II loci and class Ib loci; unlike the latter, some monomorphic class II loci seem to have diverged early

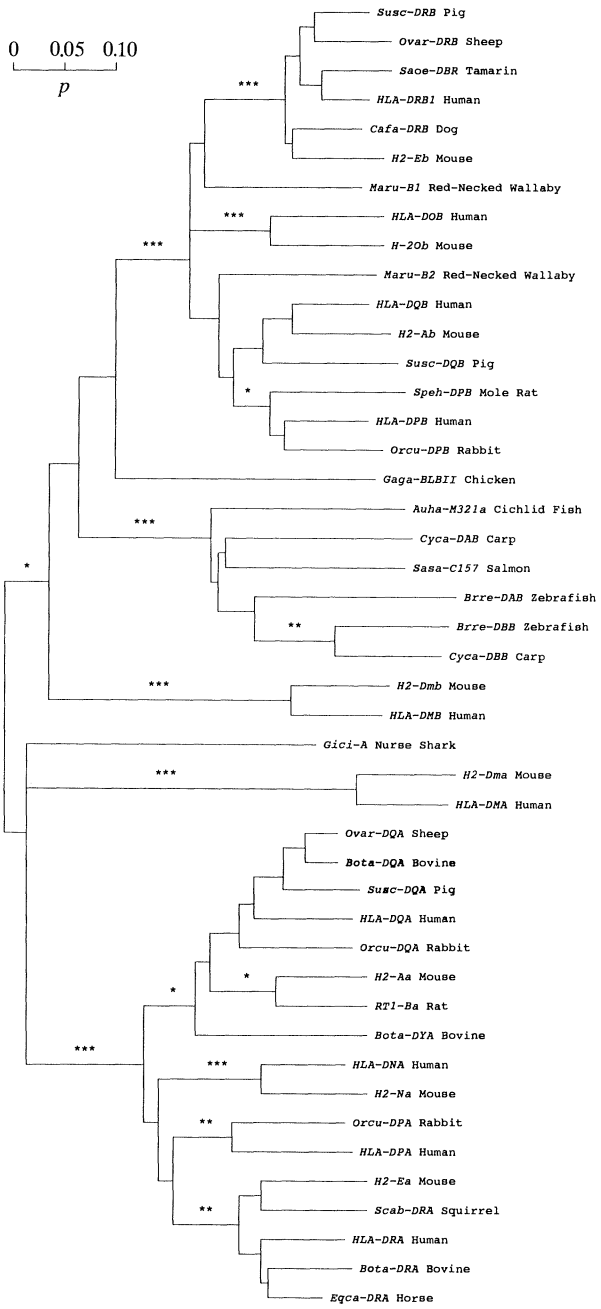

Figure 1. Phylogenetic tree of class II MHC α_2 and β_2 domains, constructed by the minimum evolution method (Rzhetsky & Nei 1992) based on the proportion of amino acid difference (p). Tests of the significance of internal branches: $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

in mammalian evolution. Both human *DOB* and *DNA* cluster with their mouse homologues (figure 1). More remarkably, *DMA* clusters outside all other mammalian class II α chain genes, and *DMB* clusters outside all mammalian class II β chain genes. In each case, this pattern is supported by a significant internal branch (figure 1). For class II β chain genes, chicken and bony fish sequences are available, and the *DMB* genes cluster outside both of these, although here the branches are not statistically significant (figure 1). The tree thus suggests that the *DMA*–*DMB* heterodimer may be a very ancient feature of the vertebrate class II MHC, having arisen by a tandem duplication of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

α and β chain genes possibly before the divergence of tetrapods from bony fishes.

The *DYA* locus is a non-polymorphic class II α chain locus so far known only from Artiodactyla (Van der Poel *et al.* 1990). This locus clusters with *DQA*, a pattern that is supported by a significant internal branch (figure 1). Thus, this locus appears to have evolved by duplication of a *DQA* gene before radiation of the eutherian orders.

To understand the type of natural selection acting on monomorphic class II loci, we computed d_S and d_N between human and mouse for orthologous pairs of class II genes (table 2). For the polymorphic loci, d_S and d_N in the ARS are about equal in the human–mouse comparison, as observed previously (Hughes & Nei 1988, 1989*a*). This slowdown in the non-synonymous rate relative to the synonymous rate in between-species comparisons is evidently due to a saturation effect at non-synonymous sites (Nei & Hughes 1992). For some monomorphic loci such as *DMB*, d_N in the ARS between human and mouse is nearly as high as that between polymorphic loci. *DOB* and *DRA*, however, show much lower d_N values in the ARS than do the other loci examined. Because the number of synonymous sites in the ARS is small, d_S values have high standard errors; thus, the difference between d_S and d_N in the ARS is significant in none of the human–mouse comparisons; but the fact that d_N in the ARS of *DRA* and *DOB* is significantly lower than that in the other loci (table 2) suggests that this region has been subject to purifying selection in these genes in contrast to the polymorphic loci. Outside the ARS, in comparisons of monomorphic class II genes, d_S is significantly greater than d_N in all comparisons. This indicates that these loci are subject to purifying selection and thus probably encode functional products in human, mouse, or both.

5. POLYMORPHISMS SHARED BETWEEN SPECIES

The existence of long-lasting polymorphisms at MHC loci, leading to sharing of allelic lineages by related species, is now well established (Nei & Rzhetsky 1991). Theoretical studies suggest that there is an upper limit to the amount of time that a polymorphism can be maintained that depends on the effective population size and selection coefficient (Takahata & Nei 1990). Andersson *et al.* (1991) noted that certain polymorphic amino acid residues are shared by DRB molecules of human and bovine. They hypothesized that such sharing of polymorphic residues by mammals of different orders was due to independent evolution, possibly as a result of selection for similar amino acids in the ARS in different species ('convergent evolution').

Lundberg & McDevitt (1992) compared class II β sequences from human and mouse and concluded that shared polymorphism between human and mouse 'represents direct descent of ancestral sequences rather than convergent evolution' (Lundberg & McDevitt 1992, p. 6545). This is a surprising conclusion since primates and rodents probably diverged 80–100 Ma ago (Li *et al.* 1990). Lundberg & McDevitt found that human and mouse used the same codon at variable β chain positions significantly more frequently than expected on the overall pattern of codon usage found for each species in the database. However, Lundberg & McDevitt did not consider the fact that codon usage in the β_1 domain of human and mouse class II MHC genes is extraordinarily biased. For human and mouse genes in the database (Wada *et al.* 1992), mean third position G + C is 62.0%, whereas that for human and mouse *DQB* (*Ab*) genes is 88.8% and that for human and mouse *DRB* (*Eb*) is 80.8%. When expectations are adjusted for the G + C content bias, there is no significant tendency to share codons between human and mouse or human and bovine (data not shown).

To obtain additional evidence regarding trans-species sharing of polymorphic residues at class II loci, we compared among mammalian species three types of shared amino acid polymorphism.

(1) Shared polymorphism at an *individual residue position* was defined to occur when at least two residues at a position were found to be identical in the two species. For example, at position 9 in the DQ β chain the residues F and Y occur in both human and chimpanzee alleles.

(2) Shared polymorphism at a *two amino acid motif* occurred when, within a sliding window of six aligned residue positions, any two positions were polymorphic in both species and at least two identical amino acid motifs were found in both species. For example, at positions 84 and 85 of the DQ β chain, the sequence motifs EV and QL occur in both human and chimpanzee.

(3) Shared polymorphism at a *three amino acid motif* occurred when, within a sliding window of six aligned residue positions, any two positions were polymorphic in both species and at least two identical amino acid motifs were found in both species. For example, at positions 53, 55 and 57 of the DQ β chain, the sequence motifs QRV and LPA occur in both human and chimpanzee.

The reason for examining these three types of polymorphism was that we expected to find a much lower level of sharing of two and three amino acid motifs when sharing of polymorphism was due to independent evolution rather than to common ancestry. Lundberg & McDevitt note that interallelic recombination may obscure allelic lineages over time, but recombination is unlikely to break up a high proportion of polymorphic motifs within a six amino acid window. Indeed, most evidence on recombination at class I and class II MHC loci suggests that recombination can serve to shuffle polymorphic motifs but not to break them up (Gyllensten *et al.* 1991; She *et al.* 1991; Hughes *et al.* 1993; McAdam *et al.* 1994). For example, putative recombinant alleles at the *DRB1* locus of humans and other Old World primates tend to involve recombination between the β pleated sheet and the α helix of the β chain ARS (Gyllenstein *et al.* 1991). Therefore, it is unlikely that recombination would break up many of the two and three amino acid motifs that we considered.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 3. *Sharing of amino acid polymorphisms in β_1 domains of class II MHC β chains by mammalian species*

species (n)	estimated divergence time/Ma	individual positions		2 amino acid motifs		3 amino acid motifs	
		shared	not shared	shared	not shared	shared	not shared
DR β chains							
human (32)							
vs chimpanzee (19)	5–7	32	0	80	0	71	0
vs tamarin (10)	30–40	22	9[c]	14	40[c]	3	38[c]
vs squirrel monkey (11)	30–40	18	16[c]	3	60[c]	0	43[c]
vs dog (9)	60–80	16	4[c]	3	16[c]	0	9[c]
vs bovine (10)	60–80	20	9[c]	6	44[c]	0	40[c]
vs mouse (6)	80–100	13	7[c]	3	18[c]	0	8[c]
DQ β chains							
human (9)							
vs chimpanzee (5)	5–7	12	4	16	3	17	0
vs rhesus monkey (19)	15–25	23	3	22	19	17	17[c]
vs crab-eating macaque (5)	15–25	16	6	16	15[c]	14	6[b]
vs mouse (52)	80–100	13	6	4	20[c]	0	18[c]
vs rat (6)	80–100	8	13	2	16[c]	0	6[c]
chimpanzee (5)							
vs rhesus monkey (19)	15–25	18	0[a]	23	0	19	0
vs crab-eating macaque (5)	15–25	13	4	16	4	15	3
rhesus monkey (19)							
vs crab-eating macaque (5)	1–3	30	0[a]	56	0[b]	49	0
mouse (52)							
vs rat (6)	15–25	12	25[b]	13	23[c]	2	24[c]

Tests of the hypothesis that the proportion shared is the same as that for the human–chimpanzee comparison at the same locus: [a]$P < 0.05$; [b]$P < 0.01$; [c]$P < 0.001$ (Fisher's exact test). n, Number of sequences.

The results (table 3) reveal that there is a remarkably high proportion of shared polymorphic residues between mammalian species. For example, human and chimpanzee shared polymorphisms at every one of the 32 sites in D1 of DR β chains that were polymorphic in both species and at 12 of 16 (75.0%) of such sites in DQ β chains. Even in comparisons between different orders, the proportion of shared residues remains quite high. For example, human and bovine share polymorphisms at 20 of 29 (69.0%) of DR β sites polymorphic in both species, and human and mouse share polymorphisms at 13 of 20 (65.0%) of DR β sites. At the DQB locus, there is not a significant difference between the human–chimpanzee comparison and the human–mouse comparison with respect to the proportion of sites at which there are shared polymorphisms (table 3).

The pattern seen for two and three amino acid motifs is strikingly different. Here, although a high proportion of polymorphism was shared when closely related species such as human and chimpanzee were compared, the proportion of shared polymorphisms was much lower when species of different orders were compared. Indeed, no example of shared polymorphism involving a three amino acid motif between mammals of different orders was found (table 3). Only a very small proportion of three amino acid motif polymorphisms were shared between the human and two New World monkey species, the cottontop tamarin (3 of 41 or 7.3%) and the squirrel monkey (0 of 43) (table 3).

These results are most consistent with the hypothesis that sharing of amino acid polymorphisms between mammals of different orders involves independent evolution rather than common ancestry. By contrast, in closely related species such as human and chimpanzee, where there is a high proportion of sharing of polymorphisms both for single residues and for two- and three amino acid motifs, common ancestry is the likely explanation for most shared polymorphism, although here also some independent evolution of the same residues probably has occurred as well.

6. CONCLUSIONS

Clearly, MHC polymorphism is maintained by some form of balancing selection that favours diversity in the ARS. It has been proposed that this selection relates to the advantage such diversity confers in enhancing the individual's ability to bind a wide array of foreign peptides and thus to resist a wide array of pathogens (Doherty & Zinkernagel 1975; Hughes & Nei 1988, 1989*a*). The discovery of a human class I MHC allele and a class II haplotype associated with resistance to *Plasmodium falciparum* malaria in West Africa (Hill *et al.* 1991) provides further support for the hypothesis that selection on MHC loci is driven by pathogens. The fact that natural selection is focused on the antigen-binding portion of the MHC molecule is not consistent with some older hypotheses for MHC polymorphism such as that of mate choice (Thomas 1974) or maternal–fetal incompatibility (Clarke & Kirby 1966).

Overdominant selection can maintain polymorphisms for millions of years, but maintenance of MHC polymorphisms for over 80 million years, as alleged by

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Lundberg & McDevitt, seems unlikely on theoretical grounds (Takahata & Nei 1990). Consistent with this view, our analyses suggest that sharing of polymorphic residues by mammals of different orders is much more likely to be due to independent evolution than to maintenance of an ancestral polymorphism. As regards this issue, the comparisons between DR β chains of human and New World monkeys are particularly instructive. In this case, relatively few two- and three-amino-acid motifs are shared. These species (separated by 30–40 Ma) may be close to the upper limit for sharing of polymorphism through common ancestry. Interestingly, theoretical studies (Takahata & Nei 1990) suggest a maximum of 30–40 Ma for maintenance of MHC polymorphisms in mammals.

This research was supported by grants from the National Institutes of Health to A.L.H. and to M.N. and by a grant from the National Science Foundation to M.N. C.Y.H. is a Howard Hughes scholar.

REFERENCES

Andersson, L., Sigurdardottir, S., Borsch, C. & Gustafsson, K. 1991 Evolution of MHC polymorphism: extensive sharing of polymorphic sequence motifs between human and bovine *DRB* alleles. *Immunogenetics* **33**, 188–193.

Bjorkman, P.J. & Parham, P. 1990 Structure, function, and diversity of class I major histocompatibility complex molecules. *A. Rev. Biochem.* **59**, 253–288.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. 1987*a* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature, Lond.* **329**, 506–512.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. 1987*b* The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature, Lond.* **329**, 512–518.

Brown, J.H., Jardetzky, T., Saper, M.A., Samraoui, B., Bjorkman, P.J. & Wiley, D.C. 1988 A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. *Nature, Lond.* **332**, 845–850.

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. & Wiley, D.C. 1993 Three dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature, Lond.* **364**, 33–39.

Cho, S., Attaya, M. & Monaco, J.J. 1991 New class II-like genes in the mouse MHC. *Nature, Lond.* **353**, 573–576.

Clarke, B. & Kirby, D.R.S. 1966 Maintenance of histocompatibility complex polymorphisms. *Nature, Lond.* **211**, 999–1000.

Doherty, P.C. & Zinkernagel, R. 1975 Enhanced immunologic surveillance in mice heterozygous at the H-2 gene complex. *Nature, Lond.* **256**, 50–52.

Gyllensten, U.B. & Erlich, H.A. 1989 Ancient roots for polymorphism at the HLA–DQα locus in primates. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9986–9990.

Gyllensten, U.B., Sundvall, M. & Erlich, H.A. 1991 Allelic diversity is generated by intraexon sequence exchange at the *DRB1* locus of primates. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3686–3690.

Hill, A.V.S., Allsopp, C.E.M., Kwiatkowski, D. *et al.* 1991 Common West African HLA antigens are associated with protection from severe malaria. *Nature, Lond.* **352**, 595–600.

Howard, J.C. 1987 MHC organization of the rat: evolutionary considerations. In *Evolution and vertebrate immunity* (ed. G. Kelsoe & D. H. Schulze), pp. 397–411. Austin: University of Texas Press.

Hughes, A.L. & Nei, M. 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature, Lond.* **335**, 167–170.

Hughes, A.L. & Nei, M. 1989*a* Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. natn. Acad. Sci. U.S.A.* **86**, 958–962.

Hughes, A.L. & Nei, M. 1989*b* Evolution of the major histocompatibility complex: independent origin of non-classical class I genes in different groups of mammals. *Molec. Biol. Evol.* **6**, 559–579.

Hughes, A.L. & Nei, M. 1990 Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. *Molec. Biol. Evol.* **7**, 491–514.

Hughes, A.L., Hughes, M.K. & Watkins, D.I. 1993 Contrasting roles of interallelic recombination at the *HLA-A* and *HLA-B* loci. *Genetics* **133**, 669–680.

Karlsson, L. & Peterson, P.A. 1992 The α chain of H-2O has an unexpected location in the major histocompatibility complex. *J. exp. Med.* **176**, 477–483.

Kelly, A.P., Monaco, J.J., Cho, S. & Trowsdale, J. 1991 A new human HLA class II-related locus *DM*. *Nature, Lond.* **353**, 571–576.

Klein, J. 1986 *Natural history of the major histocompatibility complex*. New York: Wiley.

Klein, J. & Figueroa, F. 1986 Evolution of the major histocompatibility complex. *CRC crit. Rev. Immunol.* **6**, 295–386.

Lawlor, D.A., Ward, F.E., Ennis, P.D., Jackson, A.P. & Parham, P. 1988 HLA-A, B polymorphisms predate the divergence of humans and chimpanzees. *Nature, Lond.* **335**, 268–271.

Li, W.-H., Gouy, M., Sharp, P.M., O'hUigin, C. & Yang, Y.W. 1990 Molecular phylogeny of Rodentia, Primates, Artiodactyla, and Carnivora and molecular clocks. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6703–6707.

Lundberg, A.S. & McDevitt, H.O. 1992 Evolution of major histocompatibility complex class II allelic diversity: direct descent in mice and humans. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6545–6549.

McAdam, S.N., Boyson, J.E., Liu, X., Garber, T.L., Hughes, A.L., Bontrop, R.E. & Watkins, D.I. 1994 A uniquely high level of recombination at the *HLA-B* locus. *Proc. natn. Acad. Sci. U.S.A.* **91**, 5893–5897.

Mayer, W.E., Jonker, D., Klein, D., Ivanyi, P., van Seventer, G. & Klein, J. 1988 Nucleotide sequence of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J.* **7**, 2765–2774.

Nei, M. & Gojobori, T. 1986 Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Molec. Biol. Evol.* **3**, 418–426.

Nei, M. & Hughes, A.L. 1992 Balanced polymorphism and evolution by the birth-and-death process in the MHC loci. In *Proceedings of the 11th histocompatibility workshop and conference* (ed. K. Tsuji, M. Aizawa & T. Sasazuki), pp. 27–38. Oxford: Oxford University Press.

Nei, M. & Rzhetsky, A. 1991 Reconstruction of phylogenetic trees and evolution of major histocompatibility complex genes. In *Molecular evolution of the major histocompatibility complex* (ed. J. Klein & D. Klein), pp. 13–27. Berlin: Springer-Verlag.

Pamer, E.G., Wang, C.-R., Flaherty, L., Fischer Lindahl, K. & Beran, M.J. 1992 H-2M3 presents a *Listeria*

monocytogenes peptide to cytotoxic T lymphocytes. *Cell* **70**, 215–223.

Rzhetsky, A. & Nei, M. 1992 A simple method for estimating and testing minimum evolution trees. *Molec. Biol. Evol.* **9**, 945–967.

She, J.X., Boehme, S.A., Wang, T.W., Bonhomme, F. & Wakeland, E.K. 1991 Amplification of major histocompatibility complex class II gene diversity by intra-exonic recombination. *Proc. natn. Acad. Sci. U.S.A.* **88**, 453–457.

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

Thomas, L. 1974 Biological signals for self-identification. In *Progress in immunology II* (ed. L. Brent & J. Holborrow), pp. 239–247. Amsterdam: Elsevier.

Trowsdale, J. & Kelly, A. 1985 The human HLA class II α chain gene DZα is distinct from genes in the DP, DQ, and DZ subregions. *EMBO J.* **9**, 2231–2237.

Van der Poel, J.J., Groenen, M.A., Dijkhof, R.J., Ruyter, D. & Giphart, M.J. 1990 The nucleotide sequence of the bovine MHC class II alpha genes: *DRA*, *DQA*, and *DYA*. *Immunogenetics* **31**, 29–36.

Wada, K., Wada, Y., Ishibashi, F., Gojobori, T. & Ikemura, T. 1992 Codon usage tabulated from GenBank genetic sequence data. *Nucl. Acids Res.* **20** suppl., 2111–2118.

Discussion

J. C. Howard (*Institute for Genetics, University of Cologne, Germany*). In his paper Professor Hughes expressed what he called 'the gene conversion theory' as a form of mutational model for the presence of multiple alleles in the MHC, a model in which allelic frequencies owe their existence to mutational pressure from a conversion-based hypermutator. In this model selection plays no part. This is both old-fashioned and implausible, and not 'the gene conversion theory' which anybody else espouses in this field. I suspect that he has set this crude mutational model up as a straw man, and I for one should be as happy as him to tear it down again.

The gene conversion theory which seems to me to be of interest relates not to the question of allelic frequency (which I take to be regulated by natural and neutral selection in the usual way) but to the nature of the substrates on which selection acts. With selection favouring the rare type, an unusual amount of interest falls on the nature of the process by which rare types are introduced into the population. In such a selective environment, the mutational mechanism which generates the higher overall frequency of successful new alleles will be favoured. If one may distinguish between 'strong' and 'weak' forms of gene conversion theory, the strong asserts that the gene conversion mechanism in the MHC is an adaptive response to the need to generate effective new alleles; the weak form merely eliminates the adaptive component from the argument and asserts that the point mutational mutator exists side by side in the MHC with a gene conversional mutator, perhaps because of accidental features of genomic structure. In the weak form, natural selection may use the products of the gene conversional mutator more effectively, that is, the 'take-up' rate of mutations derived by this mechanism may be higher, but this does not imply that the conversional mutator itself is sustained by natural selection.

If plant gene-for-gene resistance systems also generate new resistance factors by a recombinational mechanism rather than by point substitution (as seems increasingly likely to be the case, see Ian Crute's paper) I shall consider the strong form of the conversion theory to be highly plausible.

A. L. Hughes. I have two reasons for doubting the 'strong' version of the gene conversion theory, one empirical and one theoretical. The empirical reason is that we have no evidence that the rate of interlocus recombination ('gene conversion') is particularly high in the MHC. Certainly comparison of DNA sequences has brought to light a number of cases suggestive of past interlocus recombination. The same is true, however, of virtually every multi-gene family for which we have data, and the MHC is not exceptional in this regard. Indeed, if anything, the rate of interlocus recombination may be somewhat lower in the case of the MHC than in the case of other multi-gene families. It is usually assumed that 'gene conversion' involves the formation of a heteroduplex. Because positive selection has acted to diversify MHC genes, the ability to form such a heteroduplex may be reduced in the MHC in comparison to other multi-gene families lacking such diversifying selection.

The theoretical reason is based on arguments by evolutionary biologists that an 'adaptively high' mutation rate will not evolve by natural selection. Most mutations in coding regions are selectively deleterious, and the same is no doubt true of 'gene conversions' as well. Thus, even if a high rate of mutation or 'gene conversion' might be in some way advantageous to the species, it will be disadvantageous to the individual. But natural selection acts primarily at the level of the individual, and thus such a 'group-level' adaptation is not likely to evolve.

In a historical context, it is interesting to note that, in suggesting that there would be a high mutation rate in immune system genes, Haldane relied on a 'group selection' argument. This was at a time when the mechanism by which immunoglobulin diversity is generated was unknown. Using Haldane's argument, one might have predicted that immunoglobulin diversity is generated by a high rate of germline mutation; indeed, many immunologists made exactly this prediction. However, following the logic of natural selection acting at the individual level, one would have predicted rather that somatic mutation or somatic rearrangement of gene segments must be the mechanism, a prediction that has of course turned out to be correct.

P. Higgs (*Department of Physics, University of Sheffield, U.K.*). How typical is MHC as an example of polymorphic loci in general? In this case the polymorphism has been maintained over a long period. There are presumably many other loci at which new alleles are appearing by mutation all the time and old ones are disappearing due to random drift, thus maintaining a fairly large amount of polymorphism. Is it known what fraction of loci have a long established polymorphism such as MHC? Loci involved in disease resistance may be rather few, but they potentially have a large effect on the fitness of the individuals. If one is interested in population genetics theory with fluctuating selective pressures, as in host–parasite coevolution, one needs to know what are the most important selective pressures acting. Is selection mostly due to a small number of loci with large effects, or due to the sum of a very large number of loci which may each be nearly neutral?

A. L. Hughes. Among vertebrate genes, the MHC is likely to be unique both in its extraordinarily high level of polymorphism and in its long-lasting polymorphism. Outside the vertebrates, we know of a few loci that parallel these

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

features of the MHC. Some of the loci encoding surface proteins of malaria parasites and the self-incompatibility loci of plants are examples for which a certain amount of sequence data are now available. I am sure that other such examples remain to be discovered, but I expect them to account for only a small proportion of all loci. The type of 'trans-species' polymorphism seen at MHC loci will occur only under balancing selection such as overdominant selection or some type of frequency-dependent selection. However, molecular data suggest that the vast majority of polymorphisms are selectively neutral.

It is somewhat tangential to the question of MHC polymorphism, but my own feeling is that important adaptive breakthroughs generally involve a small number of loci with large effects.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES