Evolutionary Relationships of Class II Major-Histocompatibility-Complex Genes in Mammals[1]

Austin L. Hughes[2] *and Masatoshi Nei*[2]

Center for Demographic and Population Genetics, The University of Texas Health Science Center at Houston

The major histocompatibility complex (MHC) class II molecule consists of noncovalently associated α and β chains. In mammals studied so far, the class II MHC can be divided into a number of regions, each containing one or more α-chain genes (*A* genes) and β-chain genes (*B* genes), and it has been known for some time that orthologous relationships exist between genes in corresponding regions from different mammalian species. A phylogenetic analysis of DNA sequences of class II *A* and *B* genes confirmed these relationships; but no such orthologous relationship was observed between the *B* genes of mammals and those of birds. Thus, the class II regions have diverged since the separation of birds and mammals (~300 Mya) but before the radiation of the placental mammalian orders (60–80 Mya). Comparison of the phylogenetic trees for *A* and *B* genes revealed an unexpected characteristic of *DP*-region genes: *DPB* genes are most closely related to *DQB* genes, whereas *DPA* chain genes are most closely related to *DRA*-chain genes. Thus, the *DP* region seems to have originated through a recombinational event which brought together a *DQB* gene and a *DRA* gene (perhaps ~120 Mya). The 5′ untranslated region of all class II genes includes sequences which are believed to be important in regulating class II gene expression but which are not conserved in known pseudogenes. These sequences are conserved to an extraordinary degree in the human *DQB1* gene and its mouse homologue *Aβ1*, suggesting that regulation of expression of this locus may play a key role in expression of the entire class II MHC.

Introduction

The major histocompatibility complex (MHC) of vertebrates is a multigene family with a number of unique properties. The MHC contains many loci belonging to two very distantly related groups, the class I and class II MHC. Within each of these groups, loci differ markedly in degree of expression, extent of polymorphism, and (perhaps) function. So far the evolutionary relationships among different MHC loci are not well understood. Having previously considered the evolutionary relationships among class I MHC genes (Hughes and Nei 1989*a*), here we examine evolutionary relationships among class II MHC genes.

The class II MHC molecule is a heterodimer, consisting of noncovalently associated α and β chains, which are encoded by separate genes, generally designated *A* and *B* genes, respectively. The class II MHC of mammals consists of a number of

1. Key words: gene phylogeny, multigene families, pseudogenes, regulatory sequences, MHC genes.

2. Current address: Department of Biology and Institute of Molecular Evolutionary Genetics, 208 Mueller Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802.

Address for correspondence and reprints: Dr. Austin L. Hughes, Department of Biology, 208 Mueller Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802.

Mol. Biol. Evol. 7(6):491–514. 1990.

0737-4038/90/0706-0001$02.00

gene regions, each typically containing one or more *A* genes and one or more *B* genes. Distantly related mammalian species do not show orthologous relationships among their class I MHC loci, suggesting that there is a relatively rapid turnover of class I loci over evolutionary time, as class I loci duplicate and are eliminated by unequal crossing-over (Hughes and Nei 1989*a*). On the other hand, it is well known that orthologous relationships exist between the class II MHC gene regions in mammals of different orders (Klein and Figueroa 1986). It is not known, however, whether these orthologous relationships extend to other vertebrates. Furthermore, the evolutionary relationships among the class II regions of mammals are not well known, though a number of hypotheses have been proposed (Figueroa and Klein 1986; Klein and Figueroa 1986; Blanck and Strominger 1988). Finally, like the class I MHC, the class II MHC includes a number of loci which are poorly expressed or unexpressed, and the evolutionary origin of these genes is not well understood. In the present paper, we use DNA sequence analysis to address these unsolved problems regarding the class II MHC.

DNA Sequences Analyzed

The class II MHC molecule is a cell-surface glycoprotein expressed on antigen-presenting cells of the immune system; it functions to present intracellularly processed foreign peptides to helper T cells, which then stimulate an appropriate immune response (Kappes and Strominger 1988). The α and β chains of the class II heterodimer each consist of two extracellular domains [domain 1 (D1) and domain 2 (D2)], a transmembrane portion, and a cytoplasmic tail. A typical class II gene consists of five or six exons. Exon 1 encodes the leader peptide and the first two to four residues of D1; exon 2 encodes the bulk of D1 (~90 codons); exon 3 encodes D2 (94 codons); and exons 4–6 encode the transmembrane portion and cytoplasmic tail (~40–50 codons). The foreign-peptide-binding function of the class II molecule has been shown experimentally to reside in D1 of both α and β chains (Folsom et al. 1985; Germain et al. 1985), and most polymorphic residues are in D1. By analogy with the class I MHC molecule, Brown et al. (1988) proposed a model for the antigen recognition site (ARS) of the class II molecule; it consists of 19 or 20 residues in D1 of the α chain and of 15 or 16 residues in D1 of the β chain. At polymorphic class II *B* (β-chain) and *A* (α-chain) loci, the rate of nonsynonymous nucleotide substitution exceeds that of synonymous substitution in the codons encoding the ARS; this unusual pattern suggests that positive (probably overdominant) selection is acting on the ARS of polymorphic class II genes (Hughes and Nei 1989*b*).

Figure 1 shows schematic maps of the class II MHC in the human and the mouse. In the human, the class II MHC is divided into five regions: *DP, DN* (also called *DZ*), *DO, DQ,* and *DR*. In most cases, each region contains one or more *B* genes and one or more *A* genes, and expressed class II heterodimers usually contain a β and α chain from the same region. The only exceptions are that the *DN* region contains only a single *A* gene and that the *DO* region contains only a single *B* gene. These two regions are sometimes considered a single region (*DN/DO*). However, protein products for *DN* and *DO* have not been found in any species (Kappes and Strominger 1988); thus, it is uncertain whether products of these regions form heterodimers together or whether they ever did so in the evolutionary past. In the present paper, we will sometimes refer to them as the *DN/DO* region for convenience, without implying that any functional relationship exists between the two regions. Recently adopted nomenclature for human class II loci employs the convention that α-chain loci are designated *A* and that β-

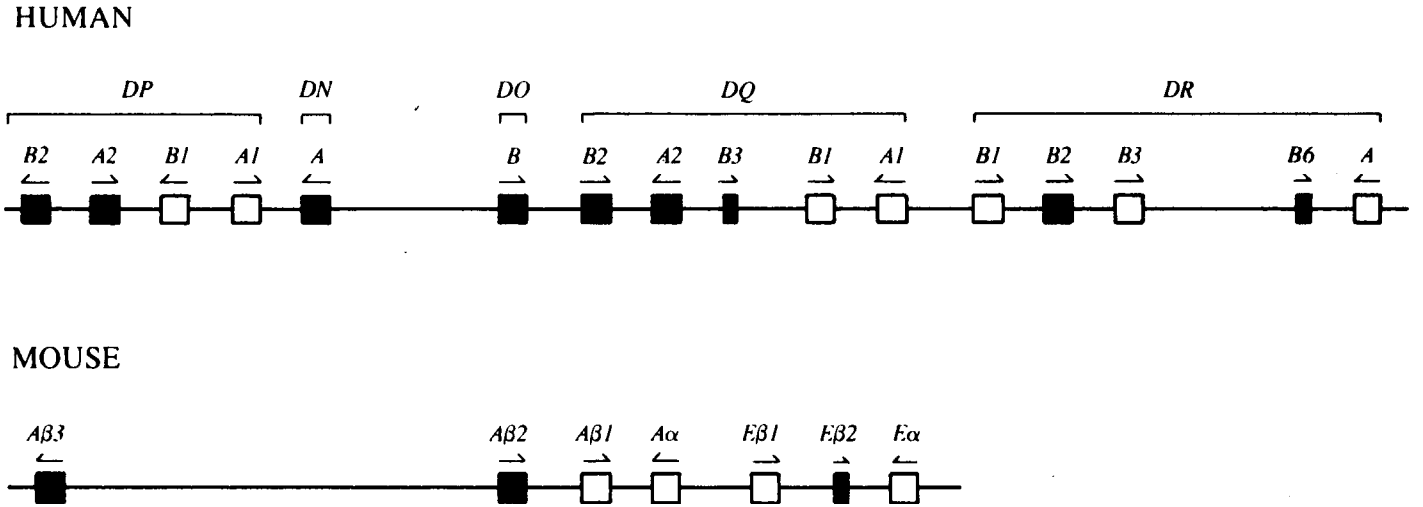

FIG. 1.—Gene arrangement of MHC class II genes in the human and mouse. Open boxes represent functional genes, whereas black boxes represent nonexpressed genes. The location of the centromere is to the left. The number of *B* genes in the human *DR* region varies with the haplotype. The figure illustrates a haplotype (such as *DR3*) having three full-length *B* genes (*B1–B3*) and one partial *B* gene (*DRB6*).

chain loci are designated *B* (Bodmer et al. 1990). We use this nomenclature here for human loci and for other mammalian loci except those of the mouse and the rat. In the case of the latter two species, a different system of nomenclature is generally used, and we retain the traditional nomenclature for these species.

In the *DP* region, both *DPB2* and *DPA2* genes of humans are known to be pseudogenes, since they both have frameshift mutations. *DNA* and *DOB* genes are known to be transcribed, though in one cell line they are transcribed at a level 5–10 times lower than that of the expressed *DQ*- and *DR*-region genes (Jonsson and Rask 1989). However, neither *DNA* nor *DOB* seems to be translated; in the case of *DNA*, defects in the signals for processing of the mRNA have been identified and these may prevent translation (Trowsdale and Kelly 1985). In the *DQ* region, *DQB2* and *DQA2* (also known as *DXB* and *DXA*, respectively) do not have any traits of pseudogenes, but neither mRNA nor protein products have been found for either of these genes (Jonsson et al. 1987). *DQB3* (also known as *DVB*) is a truncated pseudogene (Ando et al. 1989). In the *DR* region, the number of *B* genes varies with haplotype. Human *DR* loci are designated *DRB1–DRB5*. *DRB2* is a pseudogene; the other named loci

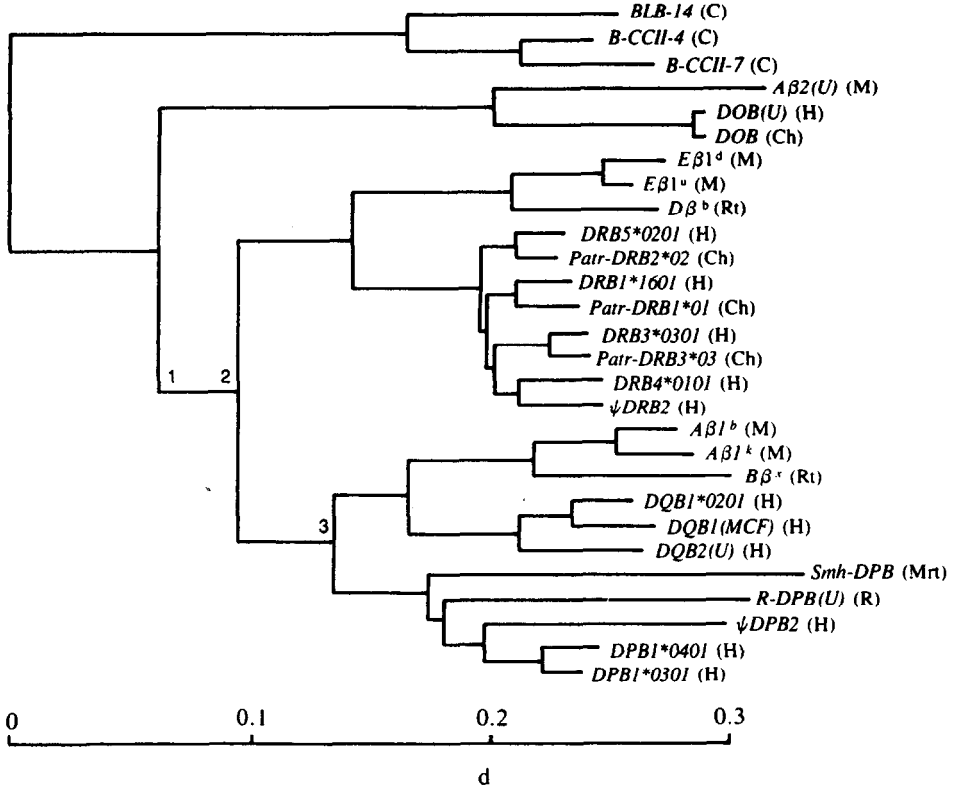

FIG. 2.—Phylogenetic tree constructed by the neighbor-joining method for mammalian and avian MHC class II *B* genes. The tree is based on 528 nucleotides from exons 2 and 3. The length of the internal branch separating the *DOB/Aβ2* cluster from other mammalian genes, d_{12}, is 0.030 ± 0.008; this branch length is significantly different from zero at the 0.1% level. The length of the internal branch separating the *DQB/DPB* cluster from other mammalian genes, d_{23}, is 0.041 ± 0.010; this branch length is also significantly different from zero at the 0.1% level. C = chicken; Ch = chimpanzee; H = human; M = mouse; Mrt = mole rat; R = rabbit; Rt = rat; d = branch length.

are functional, but only two are present in most haplotypes (e.g., the *DR3* haplotype illustrated in fig. 2). In addition to the five named *DR* loci, there is an isolated exon 2 of a *DR*β gene located directly centromeric to the *DRA* gene (Meunier et al. 1986). In the present paper, we refer to this truncated gene as ψ*DRB6*.

The class II MHC of the mouse is much simpler than that of the human. Traditionally, the class II loci of the mouse have been given names very different from those used for human class II genes, and those of the rat have been given still different names. Nonetheless, orthologous relationships exist between class II regions in all mammals known so far. In the mouse, the *DP* region is represented only by a β-chain gene (called *Aβ3*), for which only a partial sequence is available (Widera and Flavell 1985). *Aβ2* is orthologous to *DOB;* like *DOB* it is apparently transcribed but not translated (Larhammar et al. 1985). The *Aβ1* and *Aα* genes are expressed and orthologous to the *DQ*-region genes, while *Eβ1* and *Eα* are orthologous to *DR*-region genes; no sequence data are available for the unexpressed *Eβ2* gene.

Table 1 lists mammalian DNA sequences used in phylogenetic analyses in the present paper. Sequences are grouped to show the orthologous relationships between class II regions in different mammalian species. (See table 1 for conventions in designating nonexpressed class II genes.)

The class II MHC has not been mapped at the molecular level in any mammalian species besides the human and the mouse. Genes from other mammalian species are usually given the names of the orthologous human region (table 1). The rat is an exception. In the rat, the *B*β and *B*α loci are orthologous, respectively, to *Aβ1* and *Aα* of the mouse (and to *DQ*-region genes of humans); and *D*β and *D*α are orthologous, respectively, to *Eβ1* and *Eα* in the mouse (and to *DR*-region genes in humans). In the case of the rabbit, sequences from both *DP* and *DQ* regions are available. However, it is still uncertain whether rabbit *DP* genes are expressed; *DP* protein products have been difficult to detect in this species. Although rabbit *DP*-region genes lack obvious deleterious mutations, the *R-DPB(U)* gene has a sequence in the vicinity of the initiation codon which is known to reduce efficiency of translation (Sittisombut et al. 1988). A single *DPB* gene from the mole-rat *Spalax ehrenbergi* is available; it is not known whether this gene is expressed (Schöpfer et al. 1987). In addition to the mammalian genes listed in table 1, the following three class II *B* genes from the chicken were used in phylogenetic analyses: *B-LB14* (Bourlet et al. 1988) and *CCII-4* and *CCII-7* (Xu et al. 1989).

A number of hypotheses have been proposed regarding the relationships among the different class II regions. Figueroa and Klein (1986) proposed that the *DP* region was the first of the class II regions to diverge. They based this hypothesis on observations about the orientation of *B* and *A* genes in the *DP* region and in the other class II regions. In the *DP* region, both pairs of genes are in a tail-to-tail configuration, whereas in other class II regions *B* and *A* genes are in a head-to-head configuration (fig. 1). Figueroa and Klein (1986) argued that the tail-to-tail configuration is the primitive condition for the class II MHC and that thus the *DP* region must have been the first to diverge. Klein and Figueroa (1986) presented a different scenario, according to which the *DN/DO* region diverged first, then the *DR* region diverged, and finally *DP* and *DQ* diverged from each other. On the other hand, Blanck and Strominger (1988) proposed that class II evolution can be explained by the repeated duplication of a set of three genes arranged *ABB;* like Figueroa and Klein (1986) they also proposed that the *DP* region was the first to diverge.

Table 1
Mammalian Class II MHC DNA Sequences Used in Phylogenetic Tree Construction

Source	Class II Region[a]			
	DP	*DN/DO*	*DQ*	*DR*
β-Chain genes:				
Human[b]	*DPB1*0301* (1)	*DOB(U)* (3)	*DQB1*0201* (4)	*DRB1*1601* (8)
	*DPB1*0401* (2)		*DQB1(MCF)* (5)	ψ*DRB2* (9)
	ψ*DPB2* (2)		*DQB2(U)* (6)	*DRB5*0201* (8)
			{ψ*DQB3*} (7)	*DRB3*0301* (10)
				*DRB4*0101* (10)
				{ψ*DRB6*} (11)
Chimpanzee		*DOB* (12)		*Patr-DRB1*01* (13)
				*Patr-DRB2*02* (13)
				*Patr-DRB3*03* (13)
Rabbit	*R-DPB(U)* (14)			
Mouse	{ψ*Aβ3*} (15)	*Aβ2(U)* (16)	*Aβ1*^{b} (17)	*Eβ1*^{d} (18)
			Aβ1^k (17)	*Eβ1*^{u} (19)
Rat			*Bβ*^{x} (20)	*Dβ*^{b} (21)
Mole-rat	*Smh-DPB* (22)			
α-Chain genes:				
Human[b]	*DPA1* (23)	*DNA(U)* (24)	*DQA1*0301* (25)	*DRA* (28)
	ψ*DPA2* (19)		*DQA1*0102* (26)	
			DQA2(U) (27)	
Chimpanzee				*DRA* (13)
Rabbit	*R-DPA(U)* (14)		*R-DQA* (29)	*DRA* (30)
Cat	ψ*DPA2* (31)			
Mouse			*Aα*^{d} (32)	*Eα* (33)
			Aα^k (32)	
Rat			*Bα* (34)	*Dα* (35)

[a] ψ Indicates a pseudogene; (*U*) indicates an apparently unexpressed gene that lacks pseudogene criteria; { } indicate partial sequences. Numbers denote references as follows: (1) Kappes et al. 1984; (2) Gustafsson et al. 1987; (3) Servenius et al. 1987; (4) Boss and Strominger 1984; (5) So et al. 1987; (6) Jonsson et al. 1987; (7) Ando et al. 1989; (8) Lee et al. 1987; (9) Andersson et al. 1987; (10) Bell et al. 1987; (11) Meunier et al. 1986; (12) Kasahara et al. 1989; (13) Fan et al. 1989; (14) Sittisombut et al. 1988; (15) Widera and Flavell 1985; (16) Larhammar et al. 1985; (17) Choi et al. 1983; (18) Saito et al. 1983; (19) Ayane et al. 1986; (20) Eccles and McMaster 1985; (21) Robertson and McMaster (1985); (22) Schöpfer et al. 1987; (23) Servenius et al. 1984; (24) Trowsdale and Kelly 1985; (25) Auffray et al. 1982; (26) Auffray et al. 1984; (27) Auffray et al. 1987; (28) Schamboeck et al. 1983; (29) LeGuern et al. 1987; (30) Laverriè.e et al. 1989; (31) Verhoeven et al. 1988; (32) Benoist et al. 1983; (33) Hyldig-Nielsen et al. 1983; (34) Barran and McMaster 1987; (35) Holowachuk et al. 1987.

[b] When they are available, internationally accepted names (Bodmer et al. 1990) are used for alleles at human loci; when they are not, alleles are designated by the source cell line or clone.

Results

Phylogenetic Trees

Figure 2 shows a phylogenetic tree for MHC class II *B* genes that is constructed by the neighbor-joining method (Saitou and Nei 1987). The tree was based on the number of nucleotide substitutions per site (d), corrected for multiple hits by Jukes and Cantor's (1969) method. In the present study only exons 2 and 3 were used, because other exons could not always be aligned when the two genes were distantly related and because the sequences of other exons were not always available. The phylogenetic tree confirms that there are orthologous relationships between class II *B* loci of mammals belonging to different orders. However, no orthologous relationships are

detected between avian and mammalian *B* loci. The *DOB/Aβ2* locus is shown to have been the first of the mammalian β-chain loci to diverge (fig. 2). The *DPB* and *DQB* clusters are more closely related to each other than either is to *DRB* (fig. 2).

In the case of the class II *A* genes (fig. 3), no avian sequences are available for analysis; so it is not possible to determine whether orthologous relationships exist among mammalian and avian *A* genes. Since there is no outgroup, an unrooted tree is presented. Like the class II *B*-gene tree, the *A*-gene tree confirms the existence of orthologous relationships among *A* genes of mammals belonging to different orders (fig. 3). However, there is one major difference in topology between *A*- and *B*-chain trees. Among the *A* genes the *DR* and *DP* genes form a clade (fig. 3), whereas among the *B* genes *DQ* and *DP* form a clade (fig. 2).

Thus it appears that the *DP* region as it exists today in mammals has arisen as a result of a recombinational event between the *DQ* and *DR* regions. The *B* gene of the ancestral *DP* region appears to have been derived by duplication of a *DQB* gene, whereas the *A* gene was derived by duplication of a *DRA* gene. Presumably the ancestral *DPB* and *DPA* genes have since duplicated, giving rise to the two pairs of *DP* genes now found in humans (fig. 1). The tail-to-tail orientation of both pairs of *DPB* and *DPA* genes may be a result of this unusual origin. It is possible to imagine a number of scenarios by which such a hybrid class II region might have arisen. Two possibilities are illustrated in figure 4; either of these would account for both the hybrid nature of the *DP* region and the tail-to-tail orientation of genes there. If the *A*- and *B*-chain genes were translocated to the *DP* region by separate events, both genes must have been inverted independently, either at the time of translocation (as in fig. 4a) or at

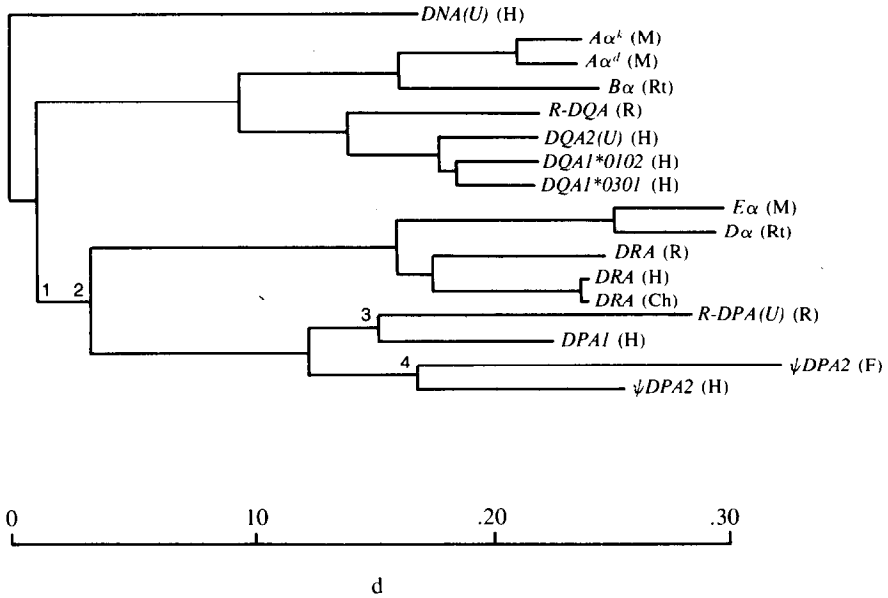

FIG. 3.—Phylogenetic tree constructed by the neighbor-joining method for mammalian MHC class II *A* genes. The tree is based on 519 nucleotides from exons 2 and 3. The length of the internal branch separating the *DRA/DPA* cluster from the *DQα* cluster, d_{12}, is 0.022 ± 0.006; this branch length is significantly different from zero at the 0.1% level. The length of the internal branch separating the *DPA1* cluster from the *DPA2* cluster, d_{34}, is 0.072 ± 0.012; this branch length is also significantly different from zero at the 0.1% level. Ch = chimpanzee; F = cat; H = human; M = mouse; R = rabbit; Rt = rat.

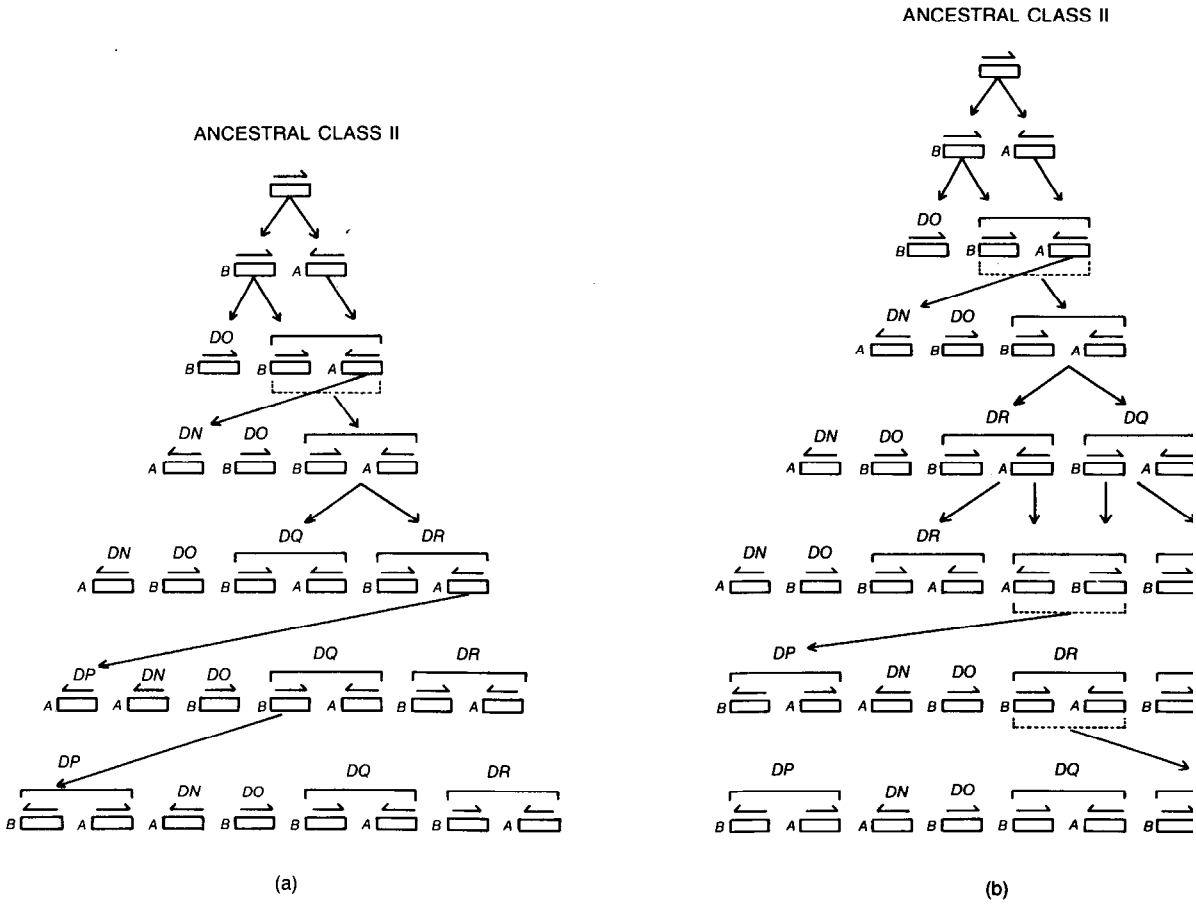

FIG. 4.—Hypothetical series of events leading to the differentiation of the mammalian MHC class II regions, illustrating two way the *DP*-region genes might have arisen from a *DQB* gene and a *DRA* gene, either (a) by two independent events of duplication plus or (b) by a single duplication event.

some later time (as proposed by Klein and Figueroa 1986). An alternative possibility is that the *A*- and *B*-chain genes were inverted and translocated simultaneously (fig. 4b).

The most appealing explanation for the fact that the human *DP* region contains two pairs of *B*- and *A*-chain genes is that an original pair were duplicated in tandem. However, this hypothesis may not be true. In the β-chain tree, the human *DPB1* and ψ*DPB2* genes are more closely related to each other than either is to the rabbit *R-DPB(U)* and the mole-rat *Smh-DPB* (fig. 2). This suggests that the human *DPB1* and ψ*DPB2* may have separated since the divergence of the mammalian orders. For exon 3 of the mouse ψ*Aβ3*, a sequence is available which is homologous to human *DPB* genes (Widera and Flavell 1985). A phylogenetic tree of exon 3 sequences from ψ*Aβ3* and other *DPB* genes [fig. 5(a)] indicates that ψ*Aβ3* diverged from the human *DPB* genes prior to the *DPB1*/ψ*DPB2* split. By contrast, the tree of *DP*-region *A* genes presents a somewhat different picture. In this case there is a significant internal branch separating the *DPA1* genes of the human and the rabbit from the *DPA2* genes of the human and the cat (fig. 3), suggesting that the duplication of the *DPA* genes has predated the radiation of mammalian orders. Thus, the results suggest that *DPA* and *DPB* were duplicated by separate events, widely separated in time, as previously noted by Gustafsson et al. (1987). This conclusion would be stronger if more data were available; for example, the hypothesis that ψ*DPB2* has evolved in the primates will be strengthened if it is shown that no orthologous locus exists in other orders of mammals. Note also that, if the duplication of *DPA*-chain genes predated the mammalian radiation, it does not necessarily follow that ψ*DPA2* has been a pseudogene since prior to the mammalian radiation, though it is now a pseudogene in both the human and the cat. However, the fact that the same deletions occur in the cat and the human ψ*DPA2* supports the hypothesis that this gene became nonfunctional before the separation of the cat and the human (Verhoeven et al. 1988).

Other class II pseudogenes and poorly expressed genes include some of ancient origin and some of very recent origin. *DNA(U)* and *DOB(U)* both seem to have diverged prior to the mammalian radiation. *DOB*/*Aβ2* is shared not only by the mouse and the human but by the bovine as well (Andersson and Rask 1988), and *DNA* is found in the rabbit (Kaluga et al. 1987). *DQB2(U)*, *DQA2(U)*, and ψ*DRB2* have a very recent origin, presumably within the primates. The evidence for a recent origin of these genes is that they cluster closer to other human genes than they do to any nonhuman genes (figs. 2 and 3). For example, *DQB2(U)* is much closer to *DQB1* than either is to orthologous rodent genes (fig. 2). Similarly, *DQA2(U)* clusters with *DQA1* (fig. 3) and ψ*DRB2* clusters with *DRB1, DRB3,* and *DRB5*.

ψ*DRB6* is another example of a gene that has apparently arisen by duplication sometime since the radiation of mammals. A phylogenetic tree of exon 2 sequences from *DR*-region genes [fig. 5(b)] groups ψ*DRB6* with other primate *DRB* genes. The branch for ψ*DRB6* is very long, reflecting the numerous substitutions that have accumulated in this nonfunctional gene fragment [fig. 5(b)]. Andersson et al. (1987) note that *DRB* genes are duplicated in mammalian species of several different orders, including not only the human and the mouse (fig. 1) but the bovine and the rabbit. For this reason, they proposed that the duplication of *DRB* genes occurred prior to the mammalian radiation; they explain the high similarity of sequences at the different human *DRB* loci as the result of homogenization by gene conversion. Our phylogenetic tree [fig. 5(b)] argues strongly against this hypothesis. ψ*DRB6* is extremely divergent

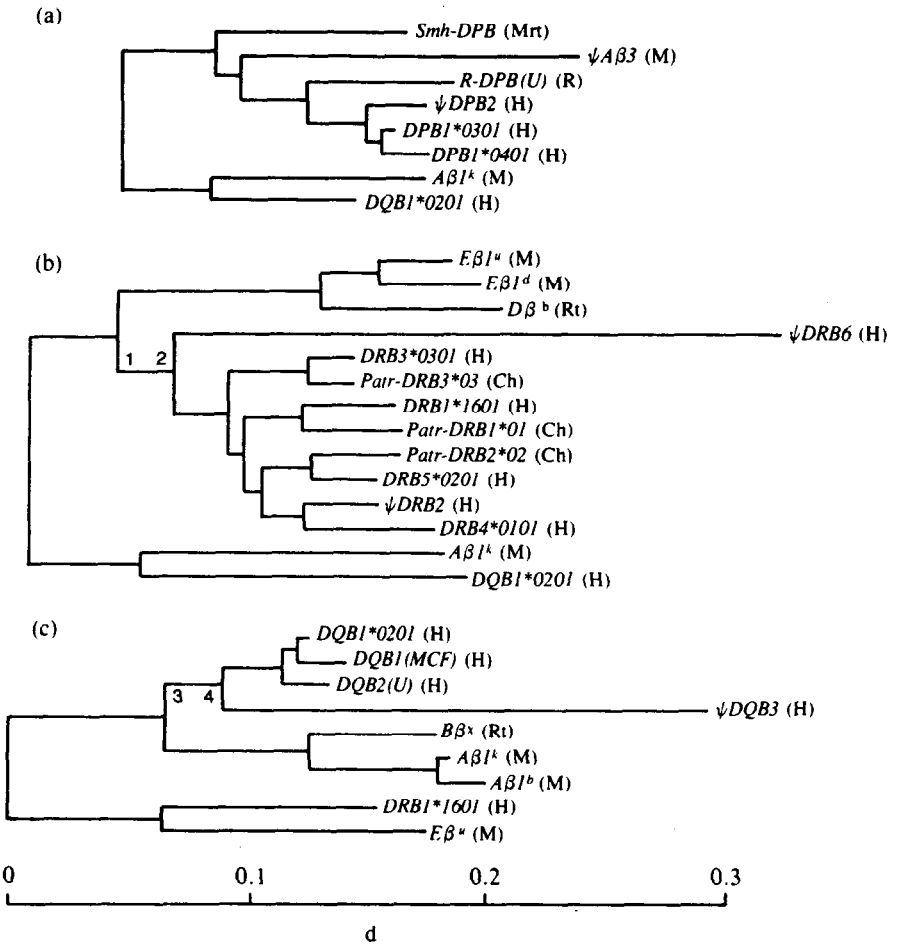

FIG. 5.—Phylogenetic trees constructed by the neighbor-joining method to examine the relationships of some class II *B* pseudogenes for which only partial sequences are available. (a), Phylogenetic tree based on 261 nucleotides from exon 3, showing the relationship of the mouse ψ*Aβ3* to other *DP*-region *B* genes (with *DQ*-region *B* genes as outgroup). (b), Phylogenetic tree based on 249 nucleotides from exon 2, showing the relationships of the human truncated pseudogene ψ*DRB6* (with *DQ*-region *B* genes as outgroup). The length of the internal branch separating the primate *DRB* genes (including ψ*DRB6*) from nonprimate *DRB* genes, d_{12}, is 0.022 ± 0.009; this branch length is significantly different from zero at the 5% level. (c), Phylogenetic tree based on 267 nucleotides from exon 3, showing relationships of the human truncated pseudogene ψ*DQB3*. The length of the internal branch separating the human *DQB* genes from those of other mammals, d_{34}, is 0.022 ± 0.009; this branch length is significantly different from zero at the 5% level. Species abbreviations are as in fig. 2.

from the other *DRB* genes, indicating that it has not been subject to a homogenizing process; and yet it clusters with other primate *DRB* genes.

In the *DQ* region, ψ*DQB3* is a pseudogene lacking exons 1 and 2. Just as ψ*DRB6* is very divergent from other human *DRB* genes, ψ*DQB3* is very divergent from other human *DQB* genes. However, like ψ*DRB6*, ψ*DQB3* seems to have arisen as the result of a gene duplication after the primate and rodent lineages diverged [fig. 5(c)]. Thus, in both the *DR* and *DQ* regions, the results argue against Blanck and Strominger's (1988) hypothesis that the primitive arrangement of genes was *ABB*. Rather, the

evidence supports the hypothesis that in each class II region there was originally a single pair of *B* and *A* genes and that duplications of loci within regions have taken place after the mammalian radiation.

Divergence Times of Class II *B* Genes

The class II *B* genes provide a unique opportunity to examine the relationship between time since divergence and degree of nucleotide difference in MHC genes. Since orthologous pairs of genes exist for different mammalian species and since we have at least rough estimates of the time since these species diverged, it is possible to obtain a number of independent estimates of the rate of change of class II MHC genes at the DNA level. These, in turn, may be used to estimate divergence times of the different class II regions.

The following groups of *B* genes are available for this purpose (see fig. 2): (1) three pairs of orthologous human and chimpanzee alleles at *DR* loci (estimated divergence time 7 Mya); (2) four pairs of orthologous loci in different mammalian orders—namely, *DPB1* in the human and *R-DPB(U)* in the rabbit; *DOB(U)* in the human and *Aβ2(U)* in the mouse; *DQB1* in the human and *Aβ1* in the mouse; and *DRB1, DRB3,* and *DRB5* in the human and *Eβ1* in the mouse (estimated time of mammalian radiation 80 Mya); and (3) *B* genes of mammals and birds (estimated divergence time 250 Mya). For divergence time estimates see Nei (1987, pp. 9–11) and Carroll (1988, p. 362). Pseudogenes were not used in these analyses, but the apparently unexpressed genes *DOB(U)*, *Aβ2(U)*, and *R-DPB(U)* were used. The comparison between orthologous rat and mouse genes was not used because preliminary analyses showed that the rate of evolution between these was higher than in other genes; this is consistent with data suggesting that genes of muroid rodents evolve at unusually high rates (Catzeflis et al. 1987).

For each of these comparisons, we estimated d, nucleotide substitutions per synonymous site (d_S), and nucleotide substitutions per nonsynonymous site (d_N) in exons 2 and 3. We used Nei and Gojobori's (1986) method I to estimate d_S and d_N. We then computed regression coefficients of d and d_N on time (t) separately for exons 2 and 3 and for the highly conserved exon 3 only. We also computed regression coefficients of d_S on time for exons 2 and 3; we did not compute the regression of d_S on time for exon 3 only, because the number of synonymous sites in this exon is quite small. In each case we fitted the line d (or d_S or d_N) $= a + bt$. Often in molecular evolutionary studies, a line through the origin ($a = 0$) is fitted. In the case of the MHC, however, most loci are polymorphic at the time of species divergence; thus it is more appropriate to fit a line with a y-intercept than to assume that the line passes through the origin.

For the regression of d_N in exons 2 and 3 on time, R^2 was only 81.9% (regression not shown). On the other hand, in all other cases the regression coefficients were highly significant, with R^2 values being 95%–99% (fig. 6). Polymorphic class II loci are known to be subject to positive Darwinian selection acting on the antigen recognition site (Hughes and Nei 1989*b*). Monomorphic, nonfunctional class II genes such as *DOB(U)* and *Aβ2(U)* are apparently free from such selection at the present time (Hughes and Nei 1989*b*), but they may not have been so in the past. The effect of selection may explain the fact that there is not a very good linear relationship between d_N in exons 2 and 3 (including the ARS) and time, whereas there is a very good linear relationship between d_N in the conserved exon 3 and time. On the basis of these regressions, we estimated that the rate of synonymous nucleotide substitution per

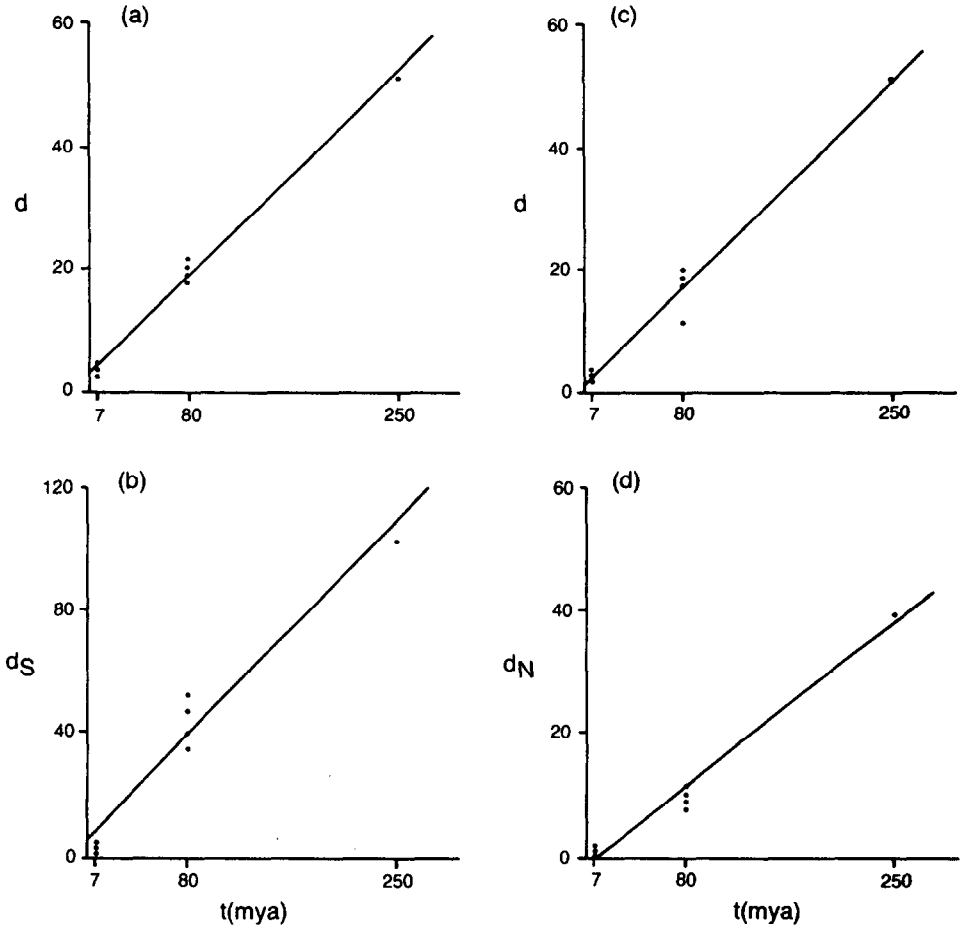

FIG. 6.—Regressions of (a) d in exons 2–3, (b) d_S in exons 2 and 3, (c) d in exon 3, and (d) d_N in exon 3 on estimated time since divergence (t). d and d_S are expressed as percentages. Regression equations are as follows: (a), $d = 3.485 + 0.197t$ ($R^2 = 99.5\%$); (b), $d_S = 6.082 + 0.408t$ ($R^2 = 95.4\%$); (c), $d = 1.039 + 0.198t$ ($R^2 = 97.5\%$); (d) $d_N = -1.305 + 0.154t$ ($R^2 = 97.6\%$).

lineage ($\lambda = b/2$) in exons 2 and 3 is 2×10^{-9}/site/year, while the overall rate is 1×10^{-9}/site/year.

Given that there is a linear relationship between the number of nucleotide substitutions and time, the same data can be used to provide estimates of divergence times of class II gene regions (table 2). By this method, the divergences of the different mammalian class II regions are estimated to have taken place ~100–200 Mya. Note that d_S generally gives somewhat lower estimates of divergence times than does d or d_N; but estimates based on d_S are much less reliable, because the number of sites is small. This method generally produced estimates greater than those obtained by Klein and Figueroa (1986) using a different method. Since the conserved second domain of class II B and A genes can easily be aligned, we used d and d_N in this region to estimate the time of divergence of the B and A genes (table 2). We obtained an estimate of a more recent divergence between B and A genes by using d (446 Mya) than we obtained by using d_N (521 Mya). Both estimates are consistent with the

Table 2
***d* and d_S in Comparisons between Mammalian Class II MHC *B* Genes from Different Regions and between *B* and *A* Genes (mean ± SE), with Divergence-Time Estimates Based on Them**

PARAMETER[a]	COMPARISON			
	DPB/DQB	*DRB/DPB-DQB*	*DOB/DPB-DQB-DRB*	*B/A*[b]
$d \pm$ SE:				
Exons 2 and 3 ($N = 528$):				
d	29.8 ± 2.1	32.7 ± 2.0	43.7 ± 2.6	. . .
d_S	51.4 ± 6.1	59.5 ± 6.2	96.2 ± 10.4	. . .
Exon 3 ($N = 282$):				
d	26.5 ± 2.8	32.3 ± 2.9	36.9 ± 3.1	91.4 ± 6.8
d_N	18.2 ± 2.6	21.5 ± 2.7	26.1 ± 2.9	80.7 ± 7.6
Divergence time (Mya) ± SE, based on:				
Exons 2 and 3:				
d	133 ± 7	148 ± 7	204 ± 9	. . .
d_S	104 ± 20	128 ± 20	214 ± 23	. . .
Exon 3:				
d	127 ± 15	156 ± 15	178 ± 16	446 ± 24
d_N	125 ± 15	146 ± 15	175 ± 16	521 ± 32
Klein and Figueroa 1986 (exons 1–6)	80	107	161–190	. . .

[a] N = number of nucleotides compared. For estimation of d_S, see Nei and Gojobori (1986); for estimation of SE, see Nei and Jin (1989). Time estimates are based on regression coefficients of time on d, d_S, or d_N.

[b] *A* and *B* genes can be aligned only in exon 3.

observation that amphibians, which diverged 370 Mya, have both class II β and class II α chains (Kaufman et al. 1985). If the larger estimate is accurate, it would suggest that separate class II β and α chains are present in all vertebrates.

Regulatory Sequences

Three highly conserved, putative-regulatory sequences have been identified in the 5′ untranslated region of all class II *B* and *A* genes so far studied. These sequences (fig. 7) are designated H (for heptamer), A, and B (Saito et al. 1983; Kelly and Trowsdale 1985; Thanos et al. 1988). It has been shown experimentally that these sequences are involved in binding nuclear factors that are believed to regulate transcription of class II genes (Celada et al. 1988). Because evolution of these sequences may shed light on the evolution of gene expression with the class II MHC, we compared the *d* in these 5′ conserved sequences (5′CS) with that in exons 2 and 3, for all class II *B* and *A* genes for which data were available. In three cases, allelic sequences were available for polymorphic class II loci. In these cases, the 5′CS are highly conserved in comparison with exon 2 and are at least as highly conserved as exon 3 (table 3). Alleles at polymorphic class II loci show numerous, largely nonsynonymous nucleotide substitutions in exon 2, which includes the codons encoding the ARS; these codons are subject to positive selection enhancing the rate of nonsynonymous substitution (Hughes and Nei 1989*b*). Exon 3, on the other hand, encodes the conserved second domain.

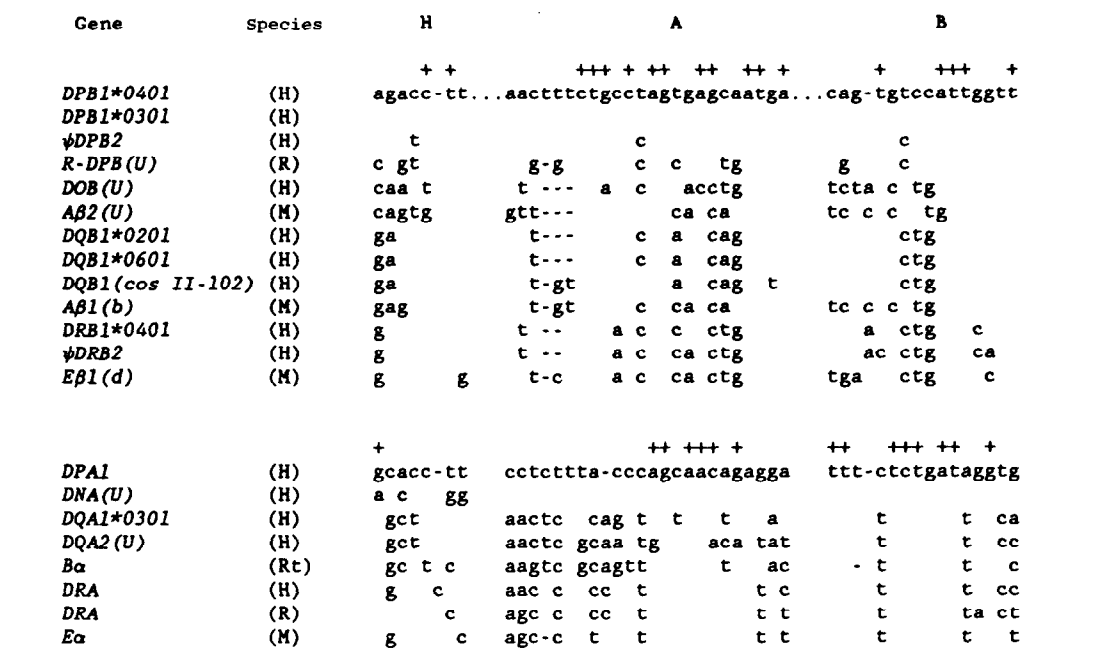

FIG. 7.—Sequence of three 5′ conserved sequences (H, A, and B) believed to be involved in the regulation of expression of MHC class II *B* and *A* genes in the mouse and the human. Sequences were aligned by eye, following Kelly and Trowsdale (1985) with modifications. For *B* genes, only differences from the top sequence, *DPB1*0401*, are shown. A plus sign (+) indicates a nucleotide position conserved in all known functional mammalian *B* genes. For *A* genes, only differences from *DPA1* are shown; here + indicates a nucleotide position conserved in all functional mammalian *A*-chain genes. Species abbreviations are as in fig. 2. A minus sign (−) indicates a gap in alignment.

Sequences homologous to the 5′CS are found even in unexpressed genes and in known pseudogenes (fig. 7). In the comparison between expressed loci and closely related pseudogenes or nonexpressed loci, the 5′CS are not conserved (table 3). *DRB1* and ψ*DRB2* are about as divergent from each other in exon 2 as are alleles at polymorphic class II loci (table 3; also see Hughes and Nei 1989*b*), but they are more divergent in the 5′CS than are alleles at class II loci. Likewise, *DQA1* and *DQA2(U)* are about as divergent in exon 2 as are alleles at the *DQA1* locus, but they are much more divergent in the 5′CS. Unfortunately, a full sequence is not available for the 5′CS of *DQB2(U)*, but the available portion (34 nucleotides from A and B) shows several substitutions relative to *DQB1*. However, *DQB2(U)* is not as different from *DQB1* in the 5′CS ($d = 6.1 \pm 4.4$) as *DQA2(U)* is from *DQA1* in the same 34 nucleotide positions ($d = 42.3 \pm 14.1$). *DPB1(4)* and ψ*DPB2* are somewhat more divergent from each other than are alleles at a typical class II locus, but they are also divergent in the 5′CS.

When we compared available 5′CS sequences from class II genes known to be expressed at normal levels, we found (*a*) five nucleotide positions that are conserved in all *A* and *B* genes, (*b*) an additional 13 positions conserved in all *B* genes, and (*c*) a nonoverlapping set of 10 positions conserved in all *A* genes. Most unexpressed or poorly expressed class II genes have one or more nucleotide substitutions at these positions (table 4). Without experimental evidence it is not possible to say whether these substitutions are responsible for reduced levels of transcription of genes such as

Table 3
***d* ± SE in Comparisons of 5′ CS and in Exons 2 and 3 of Class II MHC Genes**

Comparison[a]	No. of Comparisons	$d \pm$ SE for[b] 5′CS ($N = 41$)	Exon 2 ($N = 243$ and 255)	Exon 3 ($N = 279$ and 282)
Alleles at polymorphic loci:				
DPB1 (H)	1	0.0 ± 0.0	6.1 ± 1.6***	2.6 ± 1.0***
DQB1 (H)	3	3.2 ± 2.4	11.3 ± 1.8**	1.6 ± 0.6
Eβ1 (M)	1	0.0 ± 0.0	9.0 ± 1.9***	0.7 ± 0.5
Expressed vs. nonexpressed loci:				
DPB1 vs. ψ*DPB2* (H)	2	7.7 ± 4.5	26.8 ± 3.6***	5.8 ± 1.4
DRB1 vs. ψ*DRB2* (H)	1	7.7 ± 4.5	8.5 ± 1.9	7.5 ± 1.7
DQA1 vs. *DQA2* (*U*) (H)	1	33.2 ± 11.0	16.0 ± 2.8	1.4 ± 0.7**
Orthologous loci in different species:				
DPB1 (H) vs. *R-DPB*(*U*) (R)	2	29.9 ± 10.0	31.1 ± 4.1	12.0 ± 2.2
DOB(*U*) (H) vs. *Aβ2*(*U*) (M)	1	29.9 ± 10.0	19.6 ± 3.1	19.5 ± 2.9
DQB1 (H) vs. *Aβ1* (M)	3	4.2 ± 2.4	22.4 ± 3.0***	18.9 ± 2.8***
DRB1 (H) vs. *Eβ1* (M)	2	26.0 ± 8.9	23.3 ± 3.2	17.5 ± 2.7
DQA1 (H) vs. *Bα* (Rt)	1	29.5 ± 10.0	38.4 ± 4.9	19.8 ± 2.9
DRA (H) vs. *Eα* (M)	1	13.3 ± 6.3	25.8 ± 3.7	20.3 ± 3.0
Paired β- and α-chain genes:				
DPB1 vs. *DPA1* (H)	2	103.4 ± 30.8	. . .	82.5 ± 8.9
DOB(*U*) vs. *DNA*(*U*) (H)	1	94.3 ± 27.4	. . .	99.0 ± 11.1
DQB1 vs. *DQA1* (H)	3	69.1 ± 19.2	. . .	99.9 ± 11.2
DRB1 vs. *DRA* (H)	1	45.6 ± 13.4	. . .	87.9 ± 9.6*
Eβ1 vs. *Eα* (M)	2	26.0 ± 8.9	. . .	85.7 ± 9.4***

[a] The sequence references used in table 1 are given, as well as *DQB1*(*cos II-102*) (Jonsson et al. 1987), *DQB1*0601* (Tsukamoto et al. 1987), *DRB1*0401* (Andersson et al. 1987), and *Eβ1*[b] (Widera and Flavell 1984). Species abbreviations are as follows: H = human; M = mouse; R = rabbit; Rt = rat.

[b] N = number of nucleotides compared. In exon 2, $N = 243$ for β chains and 255 for α chains; in exon 3, $N = 279$ for β chains and 282 for α chains. d in exon 2 or exon 3 is significantly different from d in 5′CS at 5% (*), 1% (**), or 0.1% (***) levels.

DNA(*U*) and *DOB*(*U*) or for the absence of transcription of genes such as *DQA2*(*U*). However, they are consistent with the hypothesis that class II genes may suffer reduced expression as a result of mutations in the promoter region. Thus, the 5′CS in class II pseudogenes and in nonexpressed genes resemble 5′ regulatory sequences of mouse class I nonclassical MHC genes (Hughes and Nei 1989*b*). In both of these cases, regulatory sequences are less conserved in poorly expressed genes than in fully expressed genes.

In comparisons of orthologous class II loci in mammals of different orders, the 5′CS usually evolve at roughly the same rate as exons 2 and 3 (table 3). There is one exception, however. In the comparison between *DQB1* and *Aβ1*, the 5′CS are extraordinarily conserved. This high degree of conservation suggests that the 5′CS in the *DQB*/*Aβ1* region may play some especially important role in regulation of class II expression.

Typically, paired *A* and *B* genes are about as distant in the 5′CS as in exon 3 (table 3). However, in the *DR* region of humans the *A* and *B* genes are about half as distant in the 5′CS as in exon 3; and in the mouse the *Eβ1* and *Eα* are less than one-third as distant in the 5′CS as they are in exon 3 (table 3). In the case of the mouse,

Table 4
Number of Nucleotide Substitutions in 5′CS of Nonexpressed Class II MHC *B* and *A* Genes at Positions Which Are 100% Conserved in Functional Class II Genes

Gene (species[b])	Positions Conserved in[a]		
	All *B* and *A* ($N = 5$)	All *B* ($N = 13$)	All *A* ($N = 10$)
ψ*DPB2* (H)	0	0	. . .
R-DPB(U) (R)	0	0	. . .
DOB(U) (H)	1	3	. . .
Aβ2(U) (M)	1	1	. . .
DQB2(U)[c] (H)	0	0	. . .
ψ*DRB2* (H)	0	1	. . .
DNA(U) (H)	0	. . .	1
DQA2(U) (H)	1	. . .	1

[a] For conserved positions, see fig. 7.
[b] Abbreviations are as in table 3.
[c] Only a partial sequence is available, excluding two of the 13 positions conserved in all expressed *B* genes.

this unusual resemblance appears to be the result either of a recombinational event unique to the mouse lineage or of convergent evolution between the 5′CS of *Eβ1* and *Eα*. There is some evidence against the hypothesis of recombination. Outside the 5′CS there is no significant similarity between the 5′ untranslated regions of *Eβ1* and *Eα*. It seems very unlikely that a recombinational event could affect the three separate 5′CS without affecting the surrounding region. On the basis of present knowledge, it is difficult to imagine any sort of selective pressure that might favor convergent evolution between the 5′CS of *Eβ1* and *Eα*; and, since the number of nucleotides involved is small, it is possible that this convergence is the result of chance.

Amino Acid Replacements in Nonexpressed Class II Genes

Comparison of available sequences for functional class II *A* and *B* genes revealed that certain amino acid residues in D1 (encoded mainly by exon 2) and D2 (encoded by exon 3) are highly conserved. (For sequences used in determining conserved residues, see references cited in the present paper and in Hughes and Nei 1989*b*; also see Golubic et al. 1987; Figueroa et al. 1988). In the case of β chains, it is possible to determine which residues are conserved in both avian and mammalian sequences, whereas, in the case of α chains, only data for mammals are available. In D2, certain residues are conserved in both mammalian β chains and mammalian α chains. To examine the loss of functionality in nonexpressed class II genes, we looked for replacements at conserved positions in nonexpressed *B* and *A* genes (table 5). All available nonexpressed gene sequences showed such replacements, except for *DQA2(U)* (table 5). In this respect, nonexpressed class II MHC genes resemble nonclassical class I MHC genes, which also show amino acid replacements at conserved positions (Hughes and Nei 1989*b*).

The number of amino acid replacements at conserved positions seems to be roughly proportional to the time since a class II gene became nonexpressed. For example, the ancestor of the human and feline ψ*DPA2* genes is believed to have become

Table 5
Number of Amino Acid Replacements in Nonfunctional Class II MHC β and α Chains at Positions Which Are 100% Conserved in Functional Class II Chains

	Exon 2 Positions Conserved in[b]			Exon 3 Positions Conserved in[c]			
Gene[a]	All β Chains (N = 14)	All Mammalian β Chains (N = 7)	All Mammalian α Chains (N = 17)	All Mammalian β and α Chains (N = 13)	All β Chains (N = 21)	All Mammalian β Chains (N = 5)	All Mammalian α Chains (N = 22)
R-DPβ(U) (R)	3	3	. . .	1	1	1	. . .
ψ*DPB2* (H)	2	2	. . .	1	2	1	. . .
DOB(U) (H)	1	1	. . .	1	2	1	. . .
Aβ2(U) (M)	2	2	. . .	0	1	1	. . .
DQB2(U) (H)	1	1	. . .	0	1	1	. . .
ψ*DRB2* (H)	0	0	. . .	1	1	0	. . .
R-DPA(U) (R)	. . .	. . .	2	1	. . .	. . .	3
ψ*DPA2* (H)	. . .	. . .	5	6	. . .	. . .	3
ψ*DPA2* (F)	. . .	. . .	4	1	. . .	. . .	4
DNA(U) (H)	. . .	. . .	4	0	. . .	. . .	3
DQA2(U) (H)	. . .	. . .	0	0	. . .	. . .	0

[a] F = cat; all other species abbreviations are as in table 3.

[b] Conserved in all β chains: 15, 20, 21, 42, 43, 49, 54, 62, 69, 79, 82, 83, 93, and 94. Conserved in all mammalian β chains: 35, 36, 39, 64, 72, 76, and 80. Conserved in all mammalian α chains: 8, 24, 30–34, 36, 38, 39, 47, 58, 61, 62, 68, 74, and 83.

[c] Conserved in all mammalian α and β chains (numbered according to β-chain numbering; α-chain numbering for exon 3 = β-chain numbering + 5): 97, 113, 115, 124, 134, 152, 155, 161, 165, 171, 173, 175, and 183. Conserved in all mammalian β chains: 117, 119, 122, 123, 127, 129, 132, 137, 138, 144, 145, 150, 151, 154, 156, 158, 159, 162, 168, 177, and 186. Conserved in all mammalian α chains: 95, 100, 101, 103, 107, 110, 111, 114, 119, 122, 125, 126, 136, 141, 154, 160, 166, 172, 174, 180, and 181.

a pseudogene prior to the mammalian radiation; and, of all nonexpressed genes analyzed, these two sequences have accumulated by far the most replacements at conserved positions (table 5). Genes such as ψ*DPB2*, *DQB2*(*U*), *DQA2*(*U*), and ψ*DRB2*, all of which seem to have arisen by gene duplication within the primates, have accumulated considerably fewer replacements at conserved positions. If we assume that the number of amino acid replacements at conserved positions is proportional to time since loss of expression, we can use this relationship to provide a rough estimate of the time since *DOB*(*U*) and *DNA*(*U*) have lost expression. Since these genes have accumulated an intermediate number of replacements, it seems likely that they have been expressed more recently than ψ*DPA2* but not as recently as *DQB2*(*U*) and *DQA2*(*U*).

Discussion

Duplication and Deletion of Class II Loci

The class I and class II MHC genes of mammals differ with respect to both the proportion of loci which are fully expressed and the extent to which orthologous relationships exist among loci of different species. In the class I MHC, only a very small proportion of loci (the classical class I loci) are fully expressed and function in antigen presentation. The other loci are expressed at lower levels and have no known function (nonclassical class I loci) or are pseudogenes. For example, in the mouse only ~10% of class I loci are classical, and in humans only ~20% of class I loci are classical (Klein and Figueroa 1986; Orr et al. 1987). In the class II MHC, on the other hand, in both the human and the mouse approximately half the loci are functional (fig. 1).

In the class I MHC, no orthologous relationships exist between loci in distantly related mammalian species (Hughes and Nei 1989*a*). Yet the human and the chimpanzee share three orthologous classical class I loci and at least one orthologous nonclassical class I locus (Lawlor et al. 1988). In the class II MHC, on the other hand, orthologous relationships among class II regions are found in all orders of mammals so far examined (figs. 2 and 3).

In the case of the class I MHC, Hughes and Nei (1989*a*) argued that both the high proportion of poorly expressed loci and the lack of orthologous relationships between loci in mammals of different orders are consequences of (*a*) relatively frequent duplication of class I loci by unequal crossing-over, (*b*) loss of full expression of certain loci because of mutations in the promoter region, and (*c*) eventual elimination of certain duplicate loci by unequal crossing-over. The data presented in the present paper suggest that the same processes occur in the class II MHC—but at a much slower rate. The class II MHC differs from the class I MHC in that there are orthologous relationships among the class II regions in mammals of different orders; but within each class II region, loci have sometimes duplicated independently in different mammalian groups, so that there are not always orthologous relationships among individual loci. For example, certain nonfunctional class II loci in humans [such as *DQB2*(*U*), *DQA2*(*U*), and ψ*DRB2*] have evolved since the radiation of the mammalian orders.

Furthermore, over very long periods of evolutionary time, even the class II regions are not conserved. The mammalian class II regions have arisen since the separation of birds and mammals, and avian class II β-chain genes have themselves duplicated independently a number of times (fig. 2). Within the mammals, class II regions may also be lost, as the mouse has lost the *DP* region except for one β-chain pseudogene

and as the mole-rat has apparently lost both *DN/DO* and *DR* regions (Schöpfer et al. 1987).

Evolution of Nonfunctional Class II Loci

As in the case of class I loci, class II loci may sometimes lose functionality as a result of mutation. There are several different levels at which mutations can conceivably affect expression of class II genes. (1) It is possible that mutations in the promoter region can affect transcription. While there are no experimentally demonstrated cases of this at present, *DQA2(U)* seems a likely candidate. This gene is not transcribed, yet it has no other known defects except for a highly aberrant promoter region (tables 3–5). (2) In certain class II loci, transcription occurs but translation is prevented because of defects in the mRNA. The rabbit *R-DPB(U)* (Sittisombut et al. 1988) and the human *DNA(U)* (Trowsdale and Kelly 1985) seem to be examples of this, and the human *DOB(U)* and the mouse *Aβ2(U)* genes may have similar defects. (3) The protein product may be truncated or nonfunctional because of a frameshift or premature stop codon. Certain alleles at the mouse *Eβ1* and *Eα* loci include examples of both of the last two kinds of crippling mutation (Tacchini-Cottier and Jones 1988; Vu et al. 1989). In the class I MHC, most nonclassical genes and pseudogenes are of relatively recent origin (Hughes and Nei 1989*a*); in the class II MHC also, most nonexpressed genes are of a recent origin (figs. 2, 3, and 5). However, one unique aspect of the class II MHC is that certain nonexpressed loci [such as *ψDPA2, DNA(U)*, and *DOB(U)*] are evidently quite ancient, having diverged prior to the mammalian radiation.

Once a duplicated *A* or *B* gene accumulates mutations which prevent it from forming a heterodimer, the regulatory elements necessary for its transcription and the splicing signals necessary for its translation would no longer be conserved, and it would no longer be expressed. Alternatively, if a duplicated gene accumulates mutations in the promoter region or elsewhere which prevent expression, its coding sequence will no longer be conserved. The fact that nonexpressed class II genes show mutations in conserved positions both in the putative-regulatory sequences and in the coding regions (tables 4 and 5) is evidence that these processes have occurred repeatedly in the history of nonexpressed class II genes.

The Origin of the *DP* Region

In most cases, the events that gave rise to duplicate genes within class II regions are easily reconstructed. In the human *DR* region, for example, the *B* genes have duplicated several times, whereas in the *DQ* region both a duplication of a *B* gene (giving rise to *ψDQB3*) and a tandem duplication of both *B* and *A* gene [giving rise to *DQB2(U)* and *DQA2(U)*] have taken place. However, it is much more difficult to reconstruct the history of the human *DP* region. Gustafsson et al. (1987) proposed that the *DPA1/ψDPA2* split antedated the *DPB1/ψDPB2* split. Our results are consistent with this hypothesis [figs. 2 and 5(a)], but more sequences are needed before a definite conclusion can be reached. The *DP* region of humans is unique not only for the unusual pattern of relationship among the four genes which constitute it but also for the tail-to-tail orientation of the functional pair. Furthermore, our phylogenetic analysis has revealed still another unique feature of the *DP* region—namely, that it arose through an event which combined a *DQB* gene and a *DRA* gene. Any explanation of *DP* evolution should take into account all of these features.

Several authors have presented ingenious theories regarding *DP* evolution. For

example, Gustafsson et al. (1987) argued that the original *DP* molecule was encoded by the *DPB1* and ψ*DPA2* genes, which are oriented head-to-head (fig. 1). One problem with this hypothesis is that it is difficult to explain how *DPA1* (supposed to have arisen by duplication of *DPA2*) is located telomeric to *DPB1* and how ψ*DPB2* (supposed to have arisen by duplication of *DPB1*) is located centromeric to *DPA2*. The hypothesis that the original pair of *DP* genes had a tail-to-tail arrangement and then duplicated in tandem would seem simpler. However, this hypothesis does not explain the evidence suggesting that the two *DPA* genes diverged considerably prior to the *DPB1*/ψ*DPB2* split. One possible resolution of this difficulty would be to hypothesize that, sometime after a tandem duplication of an original tail-to-tail pair, the duplicate *DPB* gene was replaced by a second, more recently duplicated gene.

Klein and Figueroa (1986) explained *DP* evolution as follows: The original *DP* pair were oriented head-to-head but in an inverted relationship, with the *A* gene located on the centromeric side and the *B* gene on the telomeric side. In the *DQ* and *DR* regions, by contrast, the *B* gene is on the centromeric side, and the *A* chain gene is telomeric (fig. 1). These authors suggest that this original *DP* pair duplicated in tandem, giving rise to two head-to-head pairs. Finally, the centromerically located *A* gene and the telomerically located *B* gene were deleted, leaving a single pair oriented tail-to-tail. This pair constitutes the present *DPB1* and *DPA1*, and the other two *DP* genes are presumed to have arisen by later duplications of these same two genes. One problem with this hypothesis is that the sequence of an initial duplication followed by two deletions and further duplication seems rather complicated.

Both Gustafsson et al. (1987) and Klein and Figueroa (1986) agree in proposing that the original *DP*-region *A* and *B* genes were oriented head-to-head. If this is true, each must have been duplicated without inversion. An alternative possibility is that inversion accompanied duplication, as illustrated in the two scenarios shown in figure 4. At present, it is not possible to decide among these hypotheses. Data on the genomic organization of the class II MHC in different mammalian species may be helpful. For example, since *DPB* seems to have originated around the time of the placental/marsupial split, examination of the class II MHC of marsupials may yield important clues regarding the origin of the *DP* region.

Acknowledgments

This study was supported by research grants from the National Institutes of Health and the National Science Foundation.

LITERATURE CITED

ANDERSSON, G., D. LARHAMMAR, E. WIDMARK, B. SERVENIUS, P. A. PETERSON, and L. RASK. 1987. Class II genes of the human major histocompatibility complex: organization and evolutionary relationship of the DRβ genes. J. Biol. Chem. **262**:8748–8758.

ANDERSSON, L., and L. RASK. 1988. Characterization of the MHC class II region in cattle: the number of *DQ* genes varies between haplotypes. Immunogenetics **27**:110–120.

ANDO, A., J. KAWAI, M. MAEDA, K. TSUJI, J. TROWSDALE, and H. INOKO. 1989. Mapping and nucleotide sequence of a new *HLA* class II light chain gene. Immunogenetics **30**:243–249.

AUFFRAY, C., A. J. KORMAN, M. ROUX-DOSSETO, R. BONO, and J. L. STROMINGER. 1982. cDNA clone for the heavy chain of the human B cell alloantigen DC1: strong sequence homology to the HLA-DR heavy chain. Proc. Natl. Acad. Sci. USA 79:6337–6341.

AUFFRAY, C., J. W. LILLIE, D. ARNOT, D. GROSSBERGER, D. KAPPES, and J. L. STROMINGER.

1984. Isotypic and allotypic variation of the class II human histocompatibility antigen α chain genes. Nature **308**:327–333.

AUFFRAY, C., J. W. LILLIE, A. J. KORMAN, J. M. BOSS, N. FRÉCHIN, F. GUILLEMOT, J. COOPER, R. C. MULLIGAN, and J. L. STROMINGER. 1987. Structure and expression of *HLA-DQα* and *-DXα* genes: interallelic alternate splicing of the *HLA-DQα* gene and functional splicing of the *HLA-DXα* gene using a retroviral vector. Immunogenetics **26**:63–73.

AYANE, M., L. MENGLE-GAW, H. O. MCDEVITT, C. BENOIST, and D. MATHIS. 1986. $E\alpha^u$ and $E\beta^u$ chain association: where lies the anomaly? J. Immunol. **137**:948–951.

BARRAN, P. A., and W. R. MCMASTER. 1987. DNA sequence analysis of the rat *RT1.Bα* gene. Immunogenetics **26**:56–62.

BELL, J. I., D. DENNEY JR., L. FOSTER, T. BELT, J. A. TODD, and H. O. MCDEVITT. 1987. Allelic variation in the DR subregion of the human major histocompatibility complex. Proc. Natl. Acad. Sci. USA **84**:6234–6238.

BENOIST, C. O., D. J. MATHIS, M. R. KANTER, V. E. WILLIAMS II, and H. O. MCDEVITT. 1983. Regions of allelic hypervariability in the murine Aα immune response gene. Cell **34**: 169–177.

BLANCK, G., and J. L. STROMINGER. 1988. Molecular organization of the DQ subregion (DO-DX-DV-DQ) of the human MHC and its evolutionary implications. J. Immunol. **141**:1734–1737.

BODMER, J. G., S. G. G. MARSH, and E. ALBERT. 1990. Nomenclature for factors of the HLA system, 1989. Immunol. Today **11**:3–10.

BOSS, J. M., and J. L. STROMINGER. 1984. Cloning and sequence analysis of the human major histocompatibility complex gene *DC-3β*. Proc. Natl. Acad. Sci. USA **81**:5199–5203.

BOURLET, Y., G. BÉHAR, F. GUILLEMOT, N. FRÉCHIN, A. BILLAUT, A.-M. CHAUSSÉ, R. ZOOROB, and C. AUFFRAY. 1988. Isolation of chicken major histocompatibility complex class II (B-L) β chain sequences: comparison with mammalian β chains and expression in lymphoid organs. EMBO J. **7**:1031–1039.

BROWN, J. H., T. JARDETZKY, M. A. SAPER, B. SAMRAOUI, P. J. BJORKMAN, and D. C. WILEY. 1988. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. Nature **322**:846–850.

CARROLL, R. L. 1988. Vertebrate paleontology and evolution. Freeman, New York.

CATZEFLIS, F. M., F. H. SHELDON, J. E. AHLQUIST, and C. G. SIBLEY. 1987. DNA-DNA hybridization evidence of the rapid rate of muroid rodent DNA evolution. Mol. Biol. Evol. **4**:242–253.

CELADA, A., M. SHIGA, M. IMAGAWA, J. KOP, and R. A. MAKI. 1988. Identification of a nuclear factor that binds to a conserved sequence of the I-Aβ gene. J. Immunol. **140**:3995–4002.

CHOI, E., K. MCINTYRE, R. GERMAIN, and J. G. SEIDMAN. 1983. Murine I-Aβ chain polymorphism: nucleotide sequences of three allelic I-Aβ sequences. Science **221**:283–286.

ECCLES, S. J., and W. R. MCMASTER. 1985. DNA sequence of a rat RT1 class II Aβ gene. Immunogenetics **22**:653–663.

FAN, W., M. KASAHARA, J. GUTTKNECHT, D. KLEIN, W. E. MAYER, M. JONKER, and J. KLEIN. 1989. Shared class II MHC polymorphisms between humans and chimpanzees. Hum. Immunol. **26**:107–121.

FIGUEROA, F., E. GÜNTHER, and J. KLEIN. 1988. MHC polymorphism predating speciation. Nature **335**:265–267.

FIGUEROA, F., and J. KLEIN. 1986. The evolution of MHC class II genes. Immunol. Today **7**: 78–81.

FOLSOM, V., D. GAY, and S. TONEGAWA. 1985. The β1 domain of the mouse Eβ chain is important for restricted antigen presentation to helper T-cell hybridomas. Proc. Natl. Acad. Sci. USA **82**:1678–1682.

GERMAIN, R. N., J. D. ASHNELL, R. I. LECHLER, P. N. MARGULIES, K. M. NICKERSON, G. SUZUKI, and J. Y. L. TON. 1985. "Exon-shuffling" maps control of antibody- and T-cell-

recognition sites to the NH2-terminal domain of the class II major histocompatibility polypeptide Aβ. Proc. Natl. Acad. Sci. USA **82**:2940–2944.

GOLUBIC, M., O. BUDIMER, R. SCHREPFER, M. KASAHARA, W. E. MAYER, F. FIGUEROA, and J. KLEIN. 1987. Nucleotide sequence analysis of class II genes borne by mouse *t* chromosomes. Genet. Res. **50**:137–146.

GUSTAFSSON, K., E. WIDMARK, A.-K. JONSSON, B. SERVENIUS, D. H. SACHS, D. LARHAMMAR, L. RASK, and P. A. PETERSON. 1987. Class II genes of the human major histocompatibility complex: evolution of the DP region as deduced from nucleotide sequences of the four genes. J. Biol. Chem. **262**:8778–8786.

HOLOWACHUK, E. W., M.-K. GREER, and D. R. MARTIN. 1987. The complete sequence of the MHC class II chain RT1.Dα^u of the diabetic BB rat: mRNA levels of RT1.Dα in lymphocytes. Nucleic Acids Res. **15**:10551–10567.

HUGHES, A. L., and M. NEI. 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*. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**:958–962.

HYLDIG-NIELSEN, J. J., L. SCHENNING, U. HAMMERLING, E. WIDMARK, E. HELDIN, P. LIND, B. SERVENIUS, T. LUND, R. FLAVELL, J. S. LEE, J. TROWSDALE, P. H. SCHREIER, F. ZABLITZKY, D. LARHAMMAR, P. A. PETERSON, and L. RASK. 1983. The complete nucleotide sequence of the I-Eα^d immune response gene. Nucleic Acids Res. **11**:5055–5071.

JONSSON, A. K., J. J. HYLDIG-NIELSEN, B. SERVENIUS, D. LARHAMMAR, G. ANDERSSON, F. JÖRGENSEN, P. A. PETERSON, and L. RASK. 1987. Class II genes of the human major histocompatibility complex: comparisons of the DQ and DX α and β genes. J. Biol. Chem. **262**:8767–8777.

JONSSON, A. K., and L. RASK. 1989. Human class II *DNA* and *DOB* genes display low sequence variability. Immunogenetics **29**:411–413.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KAPPES, D., and J. L. STROMINGER. 1988. Human class II major histocompatibility complex genes and proteins. Annu. Rev. Biochem. **57**:991–1028.

KAPPES, D. J., D. ARNOT, K. OKADA, and J. L. STROMINGER. 1984. Structure and polymorphism of the HLA class II SB light chain genes. EMBO J. **3**:2985–2993.

KASAHARA, M., D. KLEIN, and J. KLEIN. 1989. Nucleotide sequence of a chimpanzee *DOB* cDNA clone. Immunogenetics **30**:66–68.

KAUFMAN, J. F., M. F. FLAJNIK, L. DU PASQUIER, and P. RIEGERT. 1985. *Xenopus* MHC class II molecules. I. Identification and structural characterization. J. Immunol. **134**:3248–3257.

KELLY, A., and J. TROWSDALE. 1985. Complete nucleotide sequence of a functional HLA-DPβ and the region between the *DPβ1* and *DPα1* genes: comparison of the 5′ ends of HLA class II genes. Nucleic Acids Res. **13**:1607–1621.

KLEIN, J., and F. FIGUEROA. 1986. Evolution of the major histocompatibility complex. CRC Crit. Rev. Immunol. **6**:295–389.

KULAGA, H., J. A. SOGN, J. D. WEISSMAN, P. N. MARCHE, C. LEGUERN, E. O. LONG, and T. J. KINDT. 1987. Expression patterns of MHC class II genes in rabbit tissues indicate close homology to human counterparts. J. Immunol. **139**:587–592.

LARHAMMAR, D., U. HAMMERLING, L. RASK, and P. A. PETERSON. 1985. Sequence of genes and cDNA encoding murine major histocompatibility complex class II gene Aβ2. J. Biol. Chem. **260**:14111–14119.

LAVERRIÈRE, A., H. KULAGA, T. J. KINDT, C. LEGUERN, and P. N. MARCHE. 1989. A rabbit class II MHC gene with strong similarities to *HLA-DRA*. Immunogenetics **30**:137–140.

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 humans and chimpanzees. Nature **335**:268–271.

LEE, B. S. M., N. A. RUST, A. J. MCMICHAEL, and H. O. MCDEVITT. 1987. HLA-DR2 subtypes form an additional supertypic family of DRβ alleles. Proc. Natl. Acad. Sci. USA **84**:4591–4595.

LEGUERN, C., J. D. WEISSMAN, P. N. MARCHE, E. JOUVIN-MARCHE, A. LAVERRIÈRE, M. R. BAGNATO, and T. J. KINDT. 1987. Sequence determination of a transcribed rabbit class II gene with homology to HLA-DQα. Immunogenetics **25**:104–109.

MEUNIER, H. F., S. CARSON, W. F. BODMER, and J. TROWSDALE. 1986. An isolated β1 exon next to the *DRα* gene in the *HLA-D* region. Immunogenetics **23**:172–180.

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

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

NEI, M., and L. JIN. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. Mol. Biol. Evol. **6**:290–300.

ORR, H. T., B. H. KOLLER, D. GERAGHTY, and R. DEMARS. 1987. Organization of the HLA class I gene family: evolutionary considerations. Pp. 349–361 *in* G. KELSO and D. H. SCHULZE, eds. Evolution and vertebrate immunity: the antigen-receptor and MHC gene families. University of Texas Press, Austin.

ROBERTSON, K. A., and W. R. MCMASTER. 1985. Complete structure of rat RT1 Eβ chain: extensive conservation of MHC class II β chains. J. Immunol. **135**:4095–4099.

SAITO, H., R. A. MAKI, L. K. CLAYTON, and S. TONEGAWA. 1983. Complete primary structure of the Eβ chain and gene of the mouse major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80**:5520–5524.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SCHAMBOECK, A., A. J. KORMAN, A. KAMB, and J. L. STROMINGER. 1983. Organization of the transcriptional unit of a human class II histocompatibility antigen: HLA-DR neavy chain. Nucleic Acids Res. **11**:8663–8675.

SCHÖPFER, R., F. FIGUEROA, D. NIŽETIĆ, E. NEVO, and J. KLEIN. 1987. Evolutionary diversification of class II *P* loci in the MHC of the mole-rat *Spalax ehrenbergi*. Mol. Biol. Evol. **4**:287–299.

SERVENIUS, B., K. GUSTAFSSON, E. WIDMARK, E. EMMOTH, G. ANDERSSON, D. LARHAMMAR, L. RASK, and P. A. PETERSON. 1984. Molecular map of the human HLA-SB (HLA-DP) region and sequence of an SBα (DPα) pseudogene. EMBO J. **3**:3209–3214.

SERVENIUS, B., L. RASK, and P. A. PETERSON. 1987. Class II genes of the human major histocompatibility complex: the DOβ gene is a divergent member of the class II β gene family. J. Biol. Chem. **262**:8759–8766.

SITTISOMBUT, N., J. MORDACQ, and K. L. KNIGHT. 1988. Rabbit MHC. II. sequence analysis of the R-DP α- and β-genes. J. Immunol. **140**:3237–3243.

SO, A. K. L., J. LINDSAY, J. BODMER, and J. TROWSDALE. 1987. Molecular variation of human major histocompatibility complex DQw3β-chains. J. Immunol. **139**:3506–3511.

TACCHINI-COTTIER, F. M., and P. J. JONES. 1988. Defective Eβ expression in three mouse *H-2* haplotypes resulting from aberrant RNA splicing. J. Immunol. **141**:3647–3653.

THANOS, D., G. MAVROTHALASSITIS, and J. PAMATHEAKIS. 1988. Multiple regulatory regions on the 5′ side of the mouse Eα gene. Proc. Natl. Acad. Sci. USA **85**:3075–3079.

TROWSDALE, J., and A. KELLY. 1985. The human HLA class II α chain gene DZα is distinct from genes in the DP, DQ, and DR subregions. EMBO J. **9**:2231–2237.

TSUKAMOTO, K., M. YASUNAMI, A. KIMURA, H. INOKO, A. ANDO, T. HIROSE, S. INAYAMA, and T. SASAZUKI. 1987. *DQW1B* gene from *HLA-DR2-Dw12* consists of six exons and expresses multiple DQw1B polypeptides through alternative splicing. Immunogenetics **25**: 343–346.

VERHOEVEN, A. J. M., M. H. KOOS, and H. L. PLOEGH. 1988. A feline class II α gene with striking similarity to the *HLA-DPA* pseudogene. Immunogenetics **28**:406–411.

VU, T. H., A. B. BEGOVICH, F. M. TACCHINI-COTTIER, and P. J. JONES. 1989. Molecular defects in the non-expressed *H-2* Eα genes of the *f* and *q* haplotypes. J. Immunol. **142**:2936–2942.

WIDERA, G., and R. A. FLAVELL. 1984. The nucleotide sequence of the murine I-Eβ^d immune response gene: evidence for gene conversion events in class II genes of the major histocompatibility complex. EMBO J. **3**:1221–1226.

———. 1985. The I region of the C57BL/10 mouse: characterization and physical linkage to *H-2K* of an SBβ-like class II pseudogene, ψAβ3. Proc. Natl. Acad. Sci. USA **82**:5500–5504.

XU, Y., J. PITCOVSKI, L. PETERSON, C. AUFFRAY, Y. BOURLET, B. M. GERNDT, A. W. NORDSKOG, S. J. LAMONT, and C. M. WARNER. 1989. Isolation and characterization of three class II MHC genomic clones from the chicken. J. Immunol. **142**:2122–2132.

JAN KLEIN, reviewing editor

Received March 10, 1990; revision received May 18, 1990

Accepted June 7, 1990