Immunogenetics (1996) 44: 36–48

ORIGINAL PAPER

Robert Graser · Colm O'hUigin · Vladimir Vincek
Axel Meyer · Jan Klein

Trans-species polymorphism of class II *Mhc* loci in danio fishes

Received: 6 December 1995 / Revised: 6 February 1996

Abstract A characteristic feature of the major histocompatibility complex (*Mhc*) polymorphism in mammals is the existence of allelic lineages shared by related species. This trans-species polymorphism has thus far been documented only in primates, rodents, and artiodactyls. In this communication we provide evidence that it also exists in cyprinid (bony) fishes at the class II *A* and *B* loci coding for the α and β polypeptide chains of the class II α:β heterodimers. The study has focused on three species of the family Cyprinidae, subfamily Rasborinae: the zebrafish (*Danio rerio*), the giant danio (*D. malabaricus*), and the pearl danio (*D. albolineatus*). The polymerase chain reaction was used to amplify and then sequence intron 1 and exon 2 of the class II *B* loci and exon 2 of the class II *A* loci in these species. Phylogenetic analysis of the sequences revealed the existence of allelic lineages whose divergence predates the divergence of the three species at both the *A* and *B* loci. The lineages at the *B* locus in particular are separated by large genetic distances. The polymorphism is concentrated in the peptide-binding region sites and is apparently maintained by balancing selection. Sharing of this unique *Mhc* feature by both bony fishes and mammals suggests that the main function of the *Mhc* (presentation of peptides to T lymphocytes) has not changed during the last 400 million years of its evolution.

Introduction

The feature that sets the major histocompatibility complex (*Mhc*) loci apart from all other known loci of the animal kingdom is their polymorphism, specifically the large genetic distances between some of the alleles (Klein and Figueroa 1986). The acquisition of over 30 nucleotide differences between alleles could theoretically be explained either by very fast substitution (mutation) rate or long persistence time of allelic lineages (J. Klein et al. 1993). The available evidence fails to support the former but strongly supports the latter explanation. The estimated average nonsynonymous substitution rate for the peptide-binding region (PBR) of the *Mhc* loci is 5.9×10^{-9} substitutions per nonsynonymous site per year (Satta et al. 1991). The average synonymous substitution rate of the *Mhc* class II loci is 1.18×10^{-9} substitutions per synonymous site per year. Both estimates are within the range of those reported for non-*Mhc* loci (J. Klein et al. 1993). The alternative explanation requires allelic lineages to be much older than 2–4 million years (my), which is believed to be the average life span of a species (Stanley 1975): it requires the polymorphism to evolve trans-specifically (Klein 1980, 1987).

The concept of trans-species evolution of the *Mhc* polymorphism was first formulated to explain the occurrence of serologically indistinguishable allelic products (allomorphs) in geographically isolated populations and subspecies of the house mouse (Klein 1980). The first biochemical evidence for the concept was provided by Arden and Klein (1982) who demonstrated that serologically identical H2 molecules from *Mus musculus domesticus*, *M. m. molossinus*, and *M. m. castaneus*, three subspecies believed to have diverged more than 1 my ago, were also indistinguishable by tryptic peptide mapping. Clear evidence for the existence of *Mhc* class II allelic lineages which predated the divergence of mice and rats dated to more than 10 my ago was obtained by Figueroa and Klein (1987) and by Figueroa and co-workers (1988) from a combination of serological and molecular data. Molecular evidence for the allelic lineages predating the divergence of different species of *Mus* was also obtained by Wakeland and co-workers (1987) and McConnell and co-workers (1988). Shortly afterwards, nucleotide sequence analysis demonstrated striking similarities between alleles in differ-

C. O'hUigin · J. Klein (✉)
Max-Planck-Institut für Biologie, Abteilung Immungenetik,
Corrensstr. 42, D-72076 Tübingen, Germany

R. Graser · V. Vincek · J. Klein
Department of Microbiology and Immunology,
University of Miami School of Medicine, Miami, FL 33106, USA

A. Meyer
Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, NY 11794-6246, USA

ent species of primates at both class I (Mayer et al. 1988; Lawlor et al. 1988, and others) and class II (Fan et al. 1989; Gyllensten and Erlich 1989; Gyllensten et al. 1991; Kenter et al. 1992; Mayer et al. 1992; Kupfermann et al. 1992; Slierendregt et al. 1992, and others) loci. These studies have led to the conclusion that in primates, allelic lineages at some of the polymorphic loci are probably more than 20 my old (Klein et al. 1990b; Klein and Takahata 1990).

In mammals, trans-species persistence of allelic lineages at the *Mhc* loci is therefore well documented. But for other vertebrate classes the data on trans-species polymorphism are either fragmentary or not available. Only in the cichlid fishes of the Great Lakes in East Africa (D. Klein et al. 1993; Ono et al. 1993a; E. Malaga and J. Klein, unpublished data) and in the Darwin's finches on the Galápagos Islands (V. Vincek, Y. Satta, C. O'hUigin, P. T. Boag, P. R. Grant, B. R. Grant, and J. Klein, unpublished data) is there clear evidence for the interspecific sharing of *Mhc* alleles. But these species are of recent origin and they therefore do not provide any information about long-term persistence of allelic lineages. In an attempt to obtain such information, we decided to compare the polymorphism of the zebrafish (*Danio rerio*) with that of two related fish species, pearl danio (*D. albolineatus*) and giant danio (*D. malabaricus*).

The three species belong to the order Cypriniformes (carp-like fishes), family Cyprinidae and subfamily Rasborinae, which is distributed over most of India, Pakistan, Bangladesh, Sri Lanka, Thailand, the Malay Peninsula, Sumatra, and the Yunnan Province of China (Nelson 1984). The taxonomy of the family has been revised several times. Early on, *D. rerio* was assigned to a separate genus and referred to as *Brachydanio rerio* (Hamilton-Buchanan 1822). Later, the genus *Brachydanio* was fused with the genus *Danio* and the species *D. malabaricus* rendered synonymous with *D. aequipinnatus* (Barman 1991). These fusions are supported by molecular data obtained in a study of mitochondrial (mt) DNA (Meyer et al. 1993, 1995) and of the *Mhc* (this communication). The designation *Danio rerio* for the zebrafish is now used almost uniformly by developmental biologists. Hence to avoid further confusion, here we introduce the designation *Dare* for the *Mhc* of the zebrafish to replace the former designation *Brre* (Ono et al. 1992, 1993a, b; Sültmann et al. 1993, 1994a, b). We retain, however, the name *D. malabaricus* for the giant danio because most recent studies (Meyer et al. 1995) suggest that it might be a distinct species from *D. aequipinnatus* after all.

In our earlier studies (Ono et al. 1992; Sültmann et al. 1993, 1994a, b), we demonstrated the existence of four class II *A* and six class II *B* loci in the zebrafish. Of the four class II *A* loci, *Dare-DAA* is linked to *Dare-DAB*, *Dare-DCA* is linked to *Dare-DCB*, and *Dare-DEA* is linked to *Dare-DEB* (Sültmann et al. 1993, 1994a, b; J. Bingulac-Popovic, F. Figueroa, J. Postlethwait, and J. Klein, unpublished data); the location of the *Dare-DBA* locus is not known. Three of the class II *B* loci (*Dare-DBB*, *-DCB*, and *-DEB*) seem to be present only in some stocks or subspecies of zebrafish (they were originally detected in a genomic library provided by a commercial supplier and the origin of the stock used could no longer be established); they are absent from the stocks maintained in our aquarium. Of the three remaining loci, *DAB* and *DDB* are closely linked, whereas *DFB* is located on a chromosome different from that of these two loci (J. Bingulac-Popovic, F. Figueroa, J. Postlethwait, and J. Klein, unpublished data). Only the *DAB* locus has thus far been found to be highly polymorphic and active (represented in a cDNA library). Four allelic lineages have been identified at this locus, which will be referred to here as *Dare-DAB1*01* through *-DAB1*04*. In the present study, we have focused on the *DAA*, *DBA*, and *DAB* loci.

Materials and methods

Source and isolation of genomic DNA

Specimens of *D. albolineatus* and *D. malabaricus* were obtained from three sources: local dealers in Miami and in Tübingen (Aquarium Rio, Egelsbach, Germany), as well as from the wild. Genomic DNA was isolated as described previously (Ono et al. 1993b with modifications according to Laird et al. 1991).

Polymerase chain reaction (PCR)

The class II *B* intron 1/exon 2 region was amplified using primers based on previously reported *D. rerio* sequences; the primers were provided by F. Figueroa of the Max-Planck-Institut für Biologie, Tübingen. The forward (sense) primer Tu385 (5′-TGCTGTCGA-(A/G)CATTTACTGGAAC-3′) corresponded to the 3′ region of exon 1; the reverse (anti-sense) primer Tu360 (5′-TCGTTTAT-CACG(G/T)ACAGCTGA-3′) corresponded to the 3′ region of exon 2. PCR was performed in 50 µl of reaction mixture containing 2 mM $MgCl_2$, 200 µM dNTP, and 2.5 units *Taq* polymerase (Perkin Elmer, Norwalk, CT). The Tu385/Tu360 primer pair was used initially at an annealing temperature of 55 °C or at 53 °C in some amplifications. Thermocycler conditions were as follows: an initial denaturation for 2 min at 94 °C, then at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 2 min, 40 cycles. The final extension at 72 °C was for 10 min. Exon 2 of class II *A* genes was amplified using the primer pair A20290 and A20291 based on known *D. rerio* cDNA and genomic sequences (Sültmann et al. 1993). A20291 is a forward (sense) primer (5′-G(C/T)T(A/C)A(A/T)G(C/T)TG- (A/T)(A/G)CA(C/T)(A/G)(A/G)GGAT-3′) corresponding to the 5′ end of exon 2; A20290 is a reverse (anti-sense) primer (5′-TT(T/G)-CTCT(A/T)(A/C)(A/T)GG(A/T)GA(G/T)TTGTA(G/T)GCCTT-3′).
PCR was performed using the Opti-prime kit (Stratagene, La Jolla, CA) to optimize buffer conditions. The reactions were as above with a 10 × buffer consisting of 3.5 mM $MgCl_2$, 10 mM Tris-HCl, and 75 mM KCl, pH 8.8. The cycling parameters were 94 °C for 3 min, 50 °C for 2 min, and 72 °C for 1.5 min in the first cycle; then 30 cycles for 1 min at 50 °C and at 72 °C for 1.5 min, followed by a final extension at 72 °C for 10 min.

Subcloning and sequencing

The PCR products were resolved by electrophoresis on 1% agarose gels and the fragments of interest excised. The agarose blocks were minced and centrifuged through silanized glass wool columns at 6500 rpm for 20–25 min. The isolated DNA was subcloned into a *Sma* I-digested pUC19 vector using the SureClone ligation kit (Pharmacia, Piscataway, NJ). After blunt-end ligation, the reactions were transformed into XL1-Blue MRF′ bacteria by standard methods and plated on LB agar containing ampicillin (50 µg/mL), X-GAL (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), and IPTG (isopropyl-β-D-thiogalactopyranoside). Positive (white) colonies were

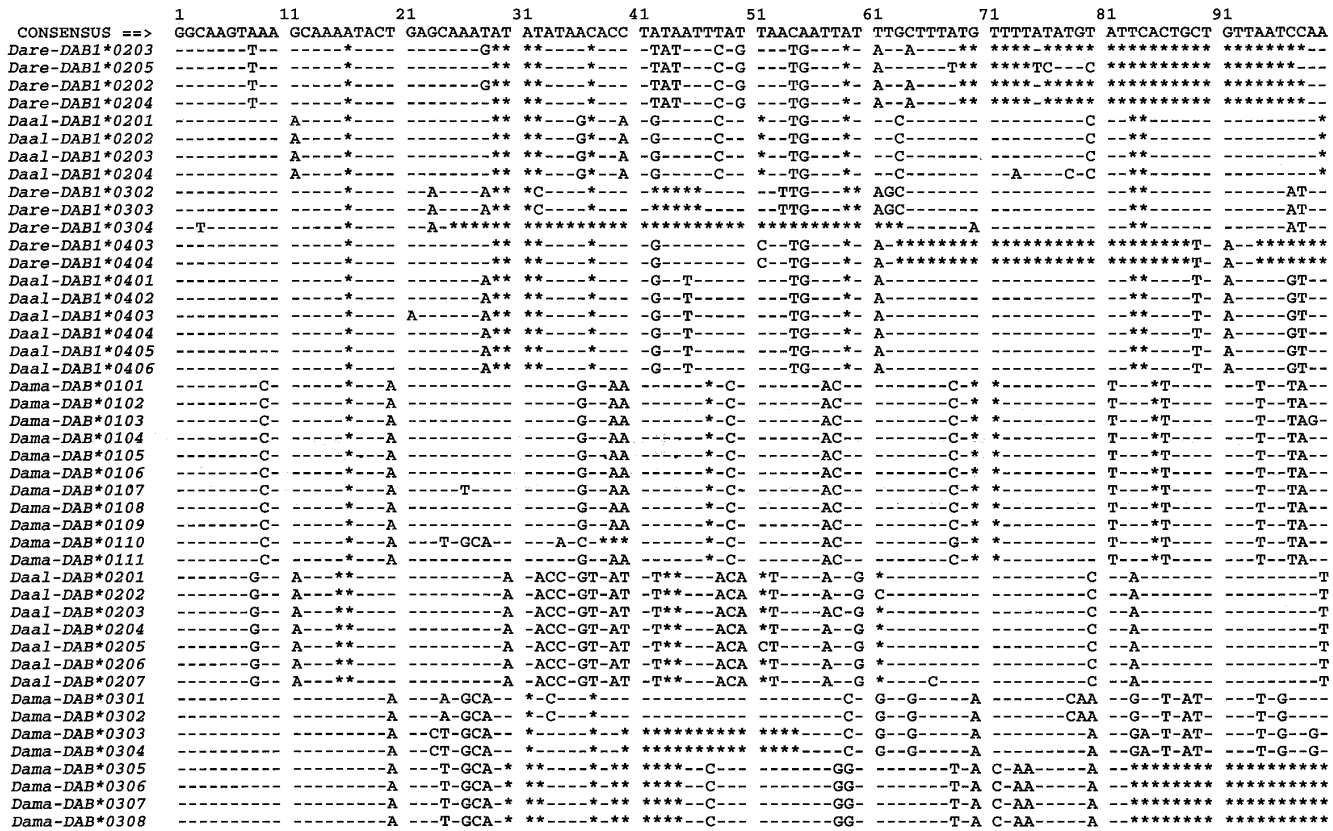


Fig. 1 (For continuation and legend see p. 39)

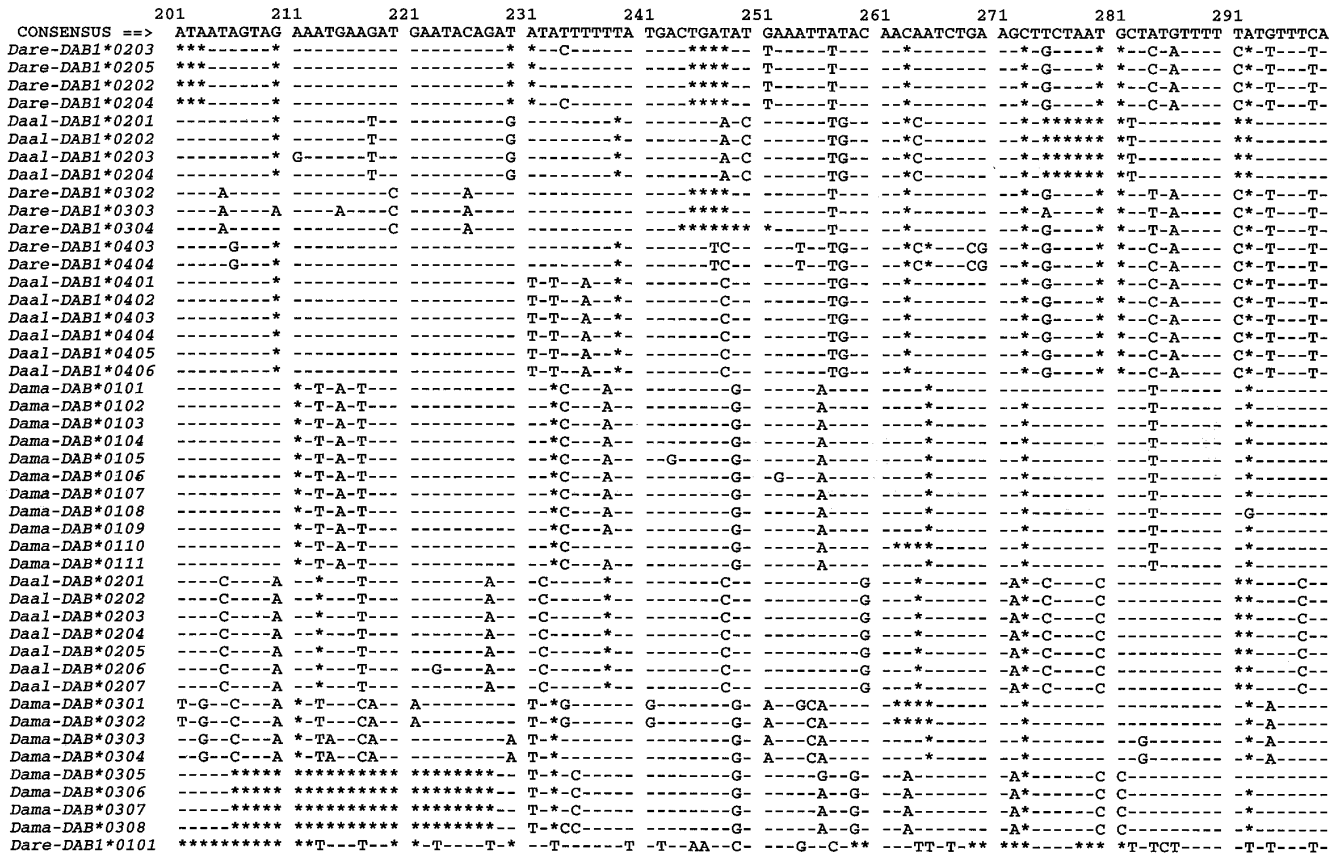


Fig. 1 Nucleotide sequences of intron 1 from *Mhc* class II *B* genes of *D. rerio* (*Dare*), *D. malabaricus* (*Dama*), and *D. albolineatus* (*Daal*). The *Dare* sequences are from Ono and co-workers (1992, 1993b). *Dashes* (–) indicate identity with the simple-majority consensus at the top; *dots* (.) unavailability of sequence information; and *asterisks* (*) indels introduced to achieve optimal sequence alignment

picked and grown overnight in LB⁻ ampicillin broth and minipreps were performed according to standard protocol (Maniatis et al. 1991). The insert size was then checked by double digestion with *Eco* RI and *Hin* dIII. The inserts were sequenced using the dideoxy nucleotide chain termination method (Sanger et al. 1977) and the Sequenase v2.0 kit (Amersham Buchler, Braunschweig, Germany). Sequencing reactions were electrophoresed on 5% polyacrylamide gels made from 50% Long Ranger gel concentrate (AT Biochem, Malvern, PA) for 4 h and 2 h. Gels were then dried and exposed to X-ray film.

Construction of dendrograms

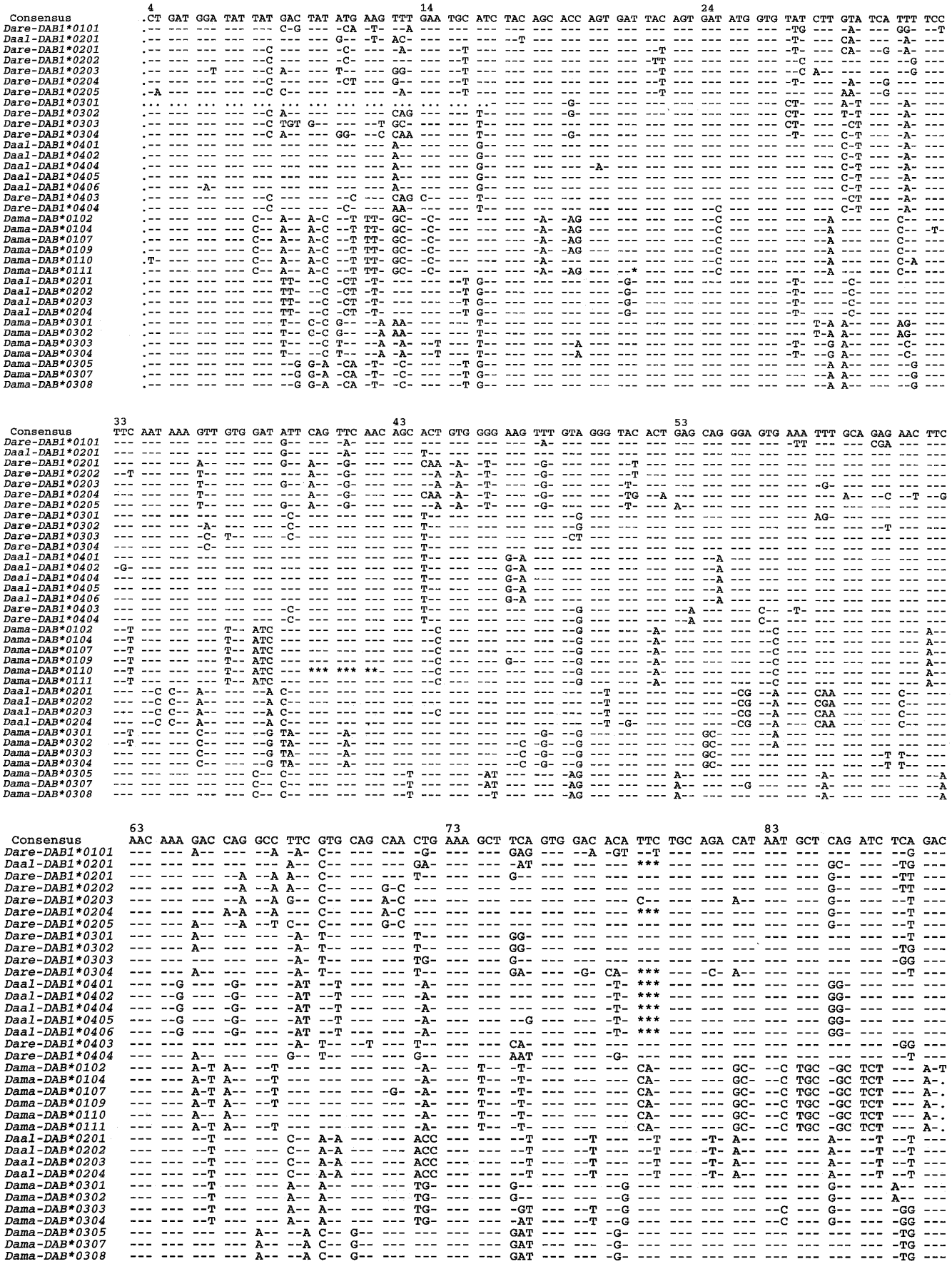
The nucleotide sequences and inferred protein sequences were aligned using the GCG computer program (Genetic Computer Group, Madison, WI) and the CLUSTAL V program (Higgins et al. 1992). Nucleic acid distances were measured using the two-parameter method (Kimura 1980) for noncoding regions. For coding regions, nonsynonymous substitutional differences were corrected using the Jukes and Cantor method of the MEGA package (Kumar et al. 1993). The evolutionary relationships were then evaluated by the neighbor-joining algorithm (Saitou and Nei 1987). Five-hundred bootstrap replications were performed to determine the reliability of the branching order.

Results and Discussion

Class II B polymorphism

Genomic DNA was isolated from 28 individuals of *D. albolineatus* and *D. malabaricus* (14 from each species), amplified by PCR using the primer pair Tu385/Tu360, and the amplification products were sequenced. Thirty-six unique sequences were obtained and their identity confirmed by sequencing multiple clones in both directions. The primers were such that they directed the amplification of intron 1 and exon 2 of the β chain-encoding (*B*) class II genes. The length of the amplified products (sequences) ranged from 466 to 520 base pairs (bp), the length variation being effected by the presence of insertions/deletions (indels) in intron 1 – different ones in the various sequence groups. The sequences are designated *Daal* for *Danio albolineatus* and *Dama for Danio malabaricus*, in accordance with the proposed nomenclature (Klein et al. 1990a). The *Daal* and *Dama* nucleotide sequences, together with the *Dare* (*Danio rerio*) sequences previously reported (Ono et al. 1992; Sültmann et al. 1994a, b) are given in Figures 1 (intron 1) and 2 (exon 2); the translated amino acid sequences of the exon 2-encoded β1 domain are given in Figure 3. A phylogenetic tree of intron 1 sequences is shown in Figure 4. A tree of exon sequences from the various *Danio* species, together with homologous sequences of other bony fish, is presented in Figure 5.

The exon codons that specify the residues of the PBR are subject to balancing selection (Hughes and Nei 1988, 1989). These codons are major contributors to the observed


Consensus
Dare-DAB1*0101
Daal-DAB1*0201
Dare-DAB1*0201
Dare-DAB1*0202
Dare-DAB1*0203
Dare-DAB1*0204
Dare-DAB1*0205
Dare-DAB1*0301
Dare-DAB1*0302
Dare-DAB1*0303
Dare-DAB1*0304
Daal-DAB1*0401
Daal-DAB1*0402
Daal-DAB1*0404
Daal-DAB1*0405
Daal-DAB1*0406
Dare-DAB1*0403
Dare-DAB1*0404
Dama-DAB*0102
Dama-DAB*0104
Dama-DAB*0107
Dama-DAB*0109
Dama-DAB*0110
Dama-DAB*0111
Daal-DAB*0201
Daal-DAB*0202
Daal-DAB*0203
Daal-DAB*0204
Dama-DAB*0301
Dama-DAB*0302
Dama-DAB*0303
Dama-DAB*0304
Dama-DAB*0305
Dama-DAB*0307
Dama-DAB*0308

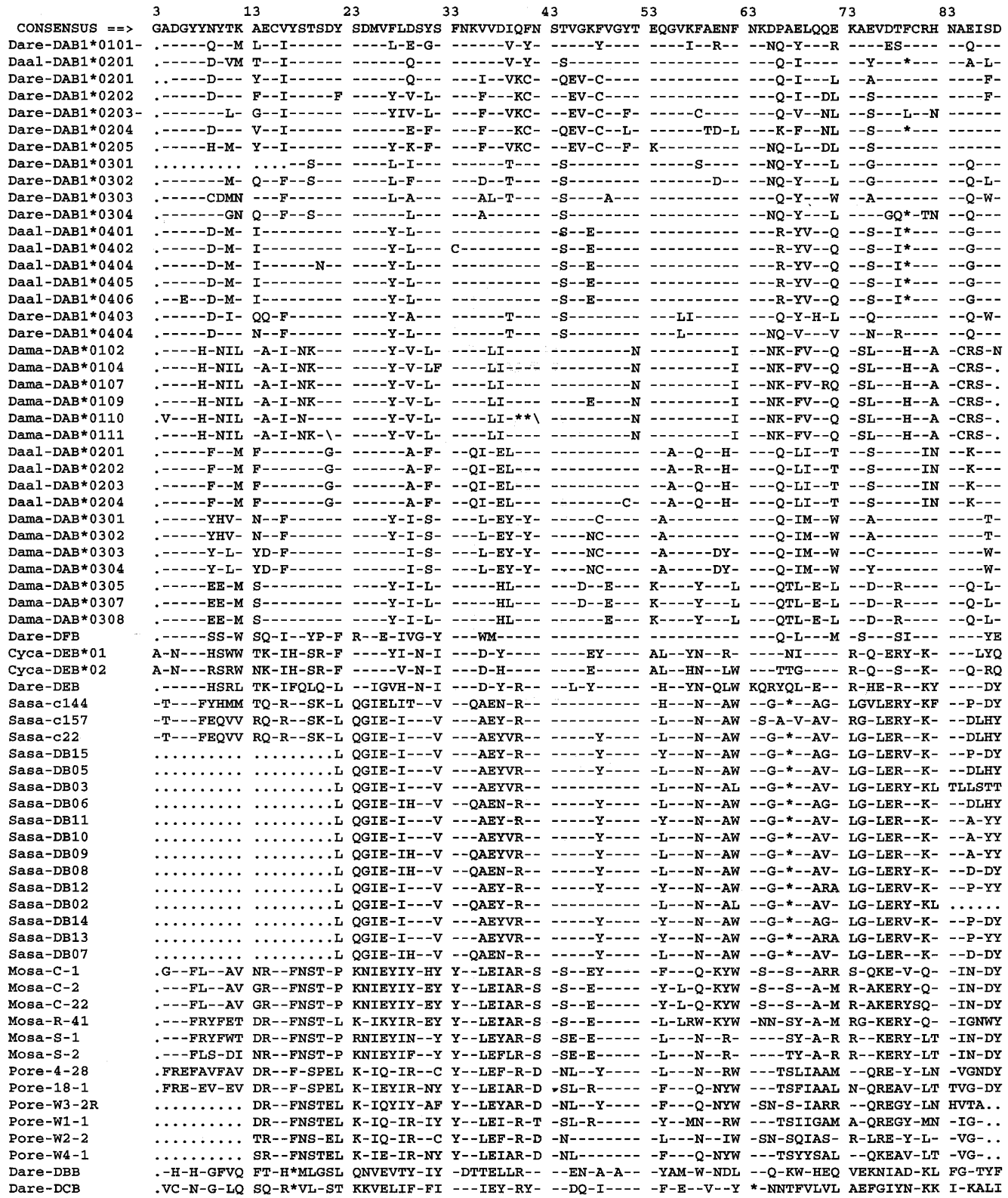


Fig. 2 Nucleotide sequences of exon 2 from *Mhc* class II *B* genes of *D. rerio* (*Dare*), *D. malabaricus* (*Dama*), and *D. albolineatus* (*Daal*). The numbering follows Ono and co-workers (1992). For symbol explanations, see Figure 1

Fig. 3 Amino acid sequences translated from nucleotide sequences in Figure 2. A *backslash* (\) indicates a frameshift mutation. Symbols are explained in Figure 1. The amino acid residues are given in the international *single-letter code. Cyca* is the carp (Ono et al. 1993c), *Mosa* the striped bass (Walker and McConnell 1994), *Sasa* the Atlantic salmon (Grimholt et al. 1994), and *Pore* the guppy (Sato et al. 1996)

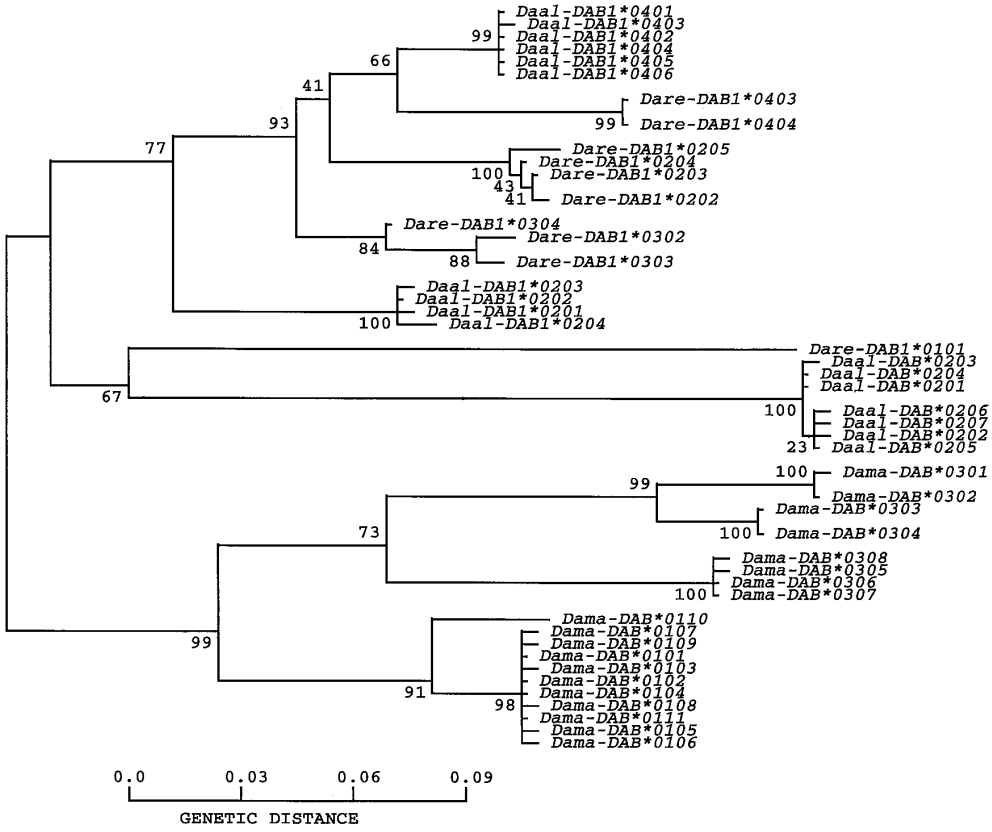


Fig. 4 Phylogenetic tree based on genetic distances between intron 1 nucleotide sequences in Figure 2. Genetic distances were determined from the number of substitutions between sequences (Kimura 1980) and the dendrogram was constructed using the neighbor-joining method of the MEGA package (Kumar et al. 1993). Numbers on branches are bootstrap values

convergences between *Mhc* alleles in primates (O'hUigin 1995; Klein and O'hUigin 1995) and their influence tends to reduce apparent evolutionary distances between distant alleles which happen to share PBR motifs, an influence that is most obvious when short exons are used to measure distances. Although the PBR region is not yet defined in fish, the most polymorphic codons are implicated in its construction. We constructed trees in which all codons are included (not shown) and others in which the ten most polymorphic codons were removed. The trees resulting from the latter approach were more conservative with respect to the extent of transspecies polymorphisms and these were therefore used in this manuscript. The most polymorphic codons assessed by the method of Wu and Kabat (1970) for the α chains are those at positions 12, 14, 25, 32, 53, 55, 58, 61, 62, and 68. For the β chains, codons 9, 11, 13, 29, 31, 39, 68, 72, 75, and 85 are excluded.

In an earlier study (Ono et al. 1992), the *Dare-DAB* sequences could be divided into clusters, originally designated *DAB1* through *DAB4*. At that stage it could not be decided whether the four clusters constituted allelic lineages at a single locus or alleles at separate loci. More recent results from linkage and segregation analysis in haploids (J. Bingulac-Popovic, F. Figueroa, and J. Klein, unpublished data) indicate that the sequences represent alleles at a single locus which we designate *Dare-DAB1*, the lineages being *DAB1*01* through *DAB1*04*.

In Figures 4 and 5, the cluster of alleles in the *Dare-DAB1* lineages is interspersed with sequences from *Daal* (and to a lesser degree with those of *Dama*). We conclude that these are allelic products of the *Daal-DAB1* locus (and possibly *Dama-DAB1*). A cluster of six *Daal-DAB1* alleles shows a higher degree of similarity to *Dare-DAB1*04* than to other lineages. We designated these alleles *Daal-DAB1*0401* through *Daal-DAB1*0406*. A second group of *Daal* alleles clusters close to the *Dare-DAB1*02* lineage and these are named *Daal-DAB1*0201* through *Daal-DAB*0204*. The *Dare-DAB1*0101* allele is the most distinctive of the *Dare-DAB1* alleles, clustering at the base of the *DAB1* tree. Although several *Daal* (Fig. 4) and *Dama* (Fig. 5) alleles cluster with the *Dare-DAB*0101* sequence, the clades are not robust and we prefer leaving the locus designation indeterminate in these cases. Among the remaining alleles we distinguish three groups corresponding to the *Daal-DAB*02*, *Dama-DAB*03*, and *Dama-DAB*01* lineages/loci. The closer similarity of *Dare-* and *Daal-DAB1* alleles to each other rather than to *Dama* alleles supports the sister group relationship of *Dare* and *Daal* proposed by Meyer and co-workers (1993, 1995).

No *Dama* or *Daal* genes corresponding to the *Dare-DBB*, *-DCB*, *-DDB*, *-DEB*, and *-DFB* genes were found, possibly because the primers were originally chosen to amplify preferentially the *DAB* genes. While no exon 2 sequences were obtained for *DDB*, as seen in Figure 5, the other four genes are quite distinct from the *DAB* genes. On the other hand, *DAB* is apparently the main, if not the only, functional class II *B* family of *Danio rerio*. The facts that lead us to this conclusion are first, that *DAB* genes are

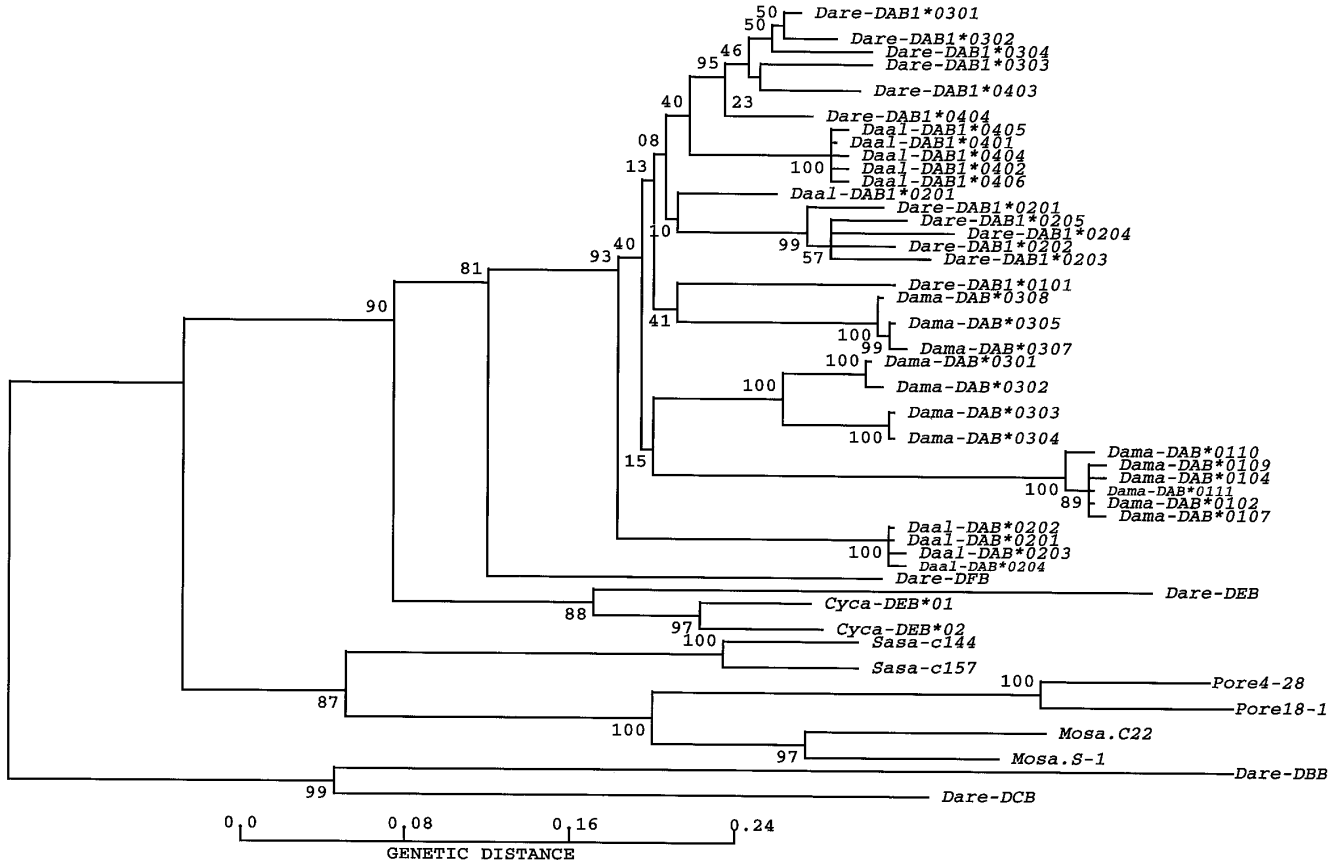


Fig. 5 Phylogenetic tree based on nonsynonymous genetic distances between sequences in Figure 3, as well as other bony fish exon 2 sequences. Genetic distances were determined from the number of nonsynonymous substitutions between sequences following exclusion of the ten most polymorphic residues. The dendrogram was constructed using the neighbor-joining method of the MEGA package (Kumar et al. 1993). Numbers on branches are bootstrap values. Species designations are given in Figures 2 and 3. *Mosa, Morone saxatilis* (striped bass; Hardee et al. 1995); *Sasa, salmo salar* (Atlantic Salmon; Grimholt et al. 1994)

found in all individuals tested, while some of the other loci may be absent in some stocks or species (Sültmann et al. 1994a, b). Second, *DAB1* is thus far the only *Dare* class II *B* locus for which transcripts (cDNA) could be found (Ono et al. 1992, Sültmann et al. 1994a, b). Third, the *Dare-DAB1* gene does not contain any apparent defect which could interfere with its functionality (Ono et al. 1992). And fourth, the *DAB* locus is polymorphic in all three species tested, while the other loci are either mono- or oligomorphic (Ono et al. 1992; Sültmann et al. 1994a, b; and the present study).

The intermingling of *DAB1* allelic lineages from different species in Figures 4 and 5 constitutes evidence of trans-species evolution of class II *B* polymorphism in cyprinid fishes. Had the polymorphism not evolved trans-specifically, one would expect the *Dare*, *Daal*, and *Dama* sequences to form three separate clades. An alternative explanation – that the relatedness of allelic lineages in the different species is the result of convergent evolution (similar mutations being selected for independently in the three species) – is not supported by the data. This type of relationship could not have occurred at intron 1 sequences of the *DAB1* locus, yet the sequences (Fig. 1) and the phylogenetic tree based on them (Fig. 4) reveal both similar clustering and similar affinities among the clusters at the exon 2 (β1 domain) sequence. The same is also true for the exon 2 synonymous sites (data not shown). Nonsynonymous PBR sites, the primary target of convergent evolution in the *Mhc*, are largely omitted in exon 2 trees by the removal of the ten most polymorphic sites. We conclude therefore that the trans-species mode of evolution of *Mhc* polymorphism is not limited to mammals, but occurs also in bony fishes.

Table 1 The number of synonymous (K_s) and nonsynonymous (K_a) substitutions per site for PBR and non-PBR codons of all pairwise comparisons of 18 *Dare* and *Daal DAB1* alleles*

	PBR	Non-PBR
No. of codons	16	69
K_a	0.397	0.078
K_s	0.556	0.169

Average g = 1.64

* The location of the PBR sites in fish *Mhc* genes is uncertain. The Brown (1988) model of the PBR in mammals is preferred to the Brown (1993) model because it more fully predicts the most polymorphic codons in fish. γ is the ratio of PBR nonsynonymous substitution frequency (per site) to the overall synonymous substitution frequency. γ values greater than 1.0 are indicative of positive selection

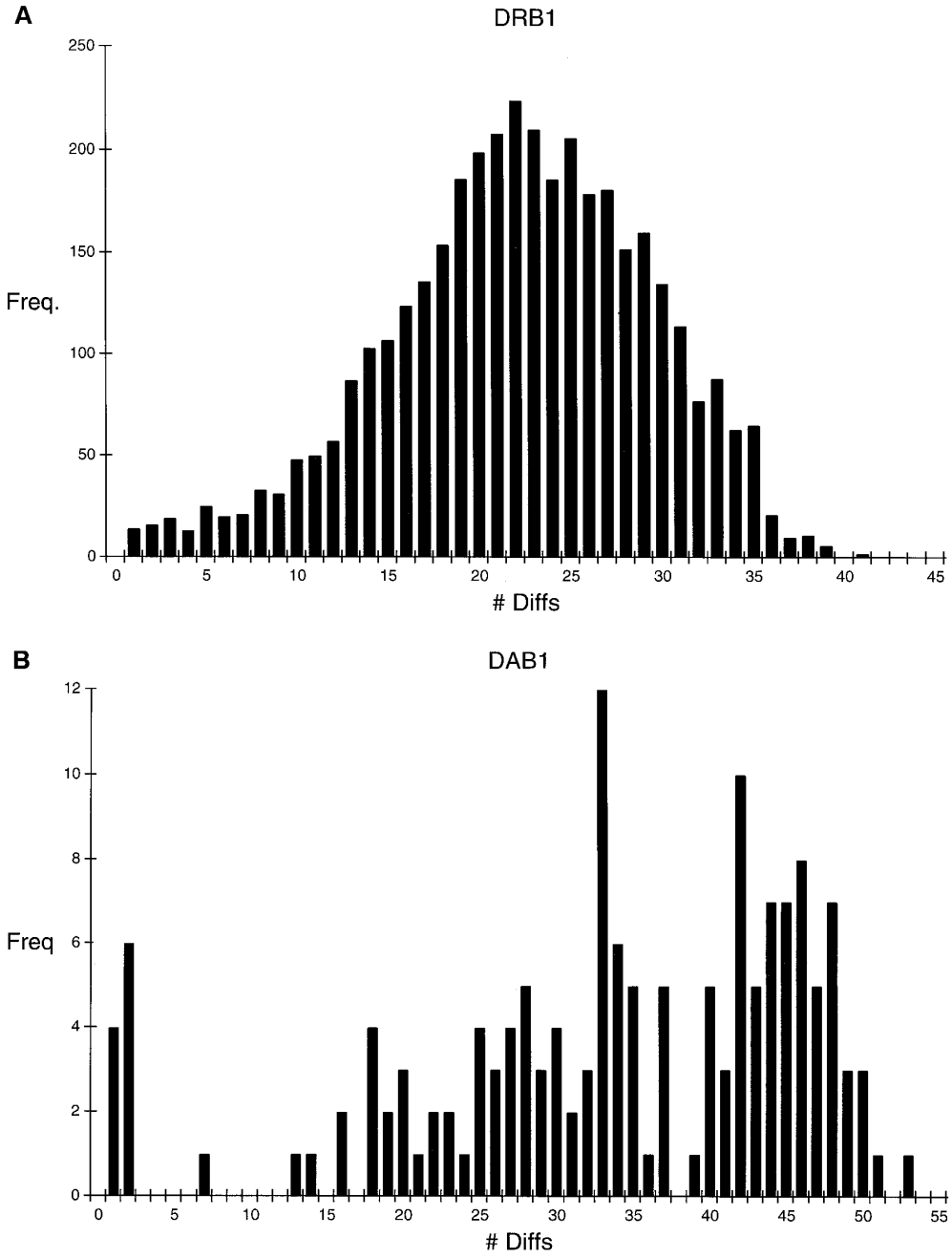


Fig. 6 A, B Pairwise comparison of differences between **A** the *Dare* (*Daal*)-*DAB1* exon 2 sequences in Figure 2 or **B** 87 primate *DRB1* exon 2 sequences (consisting of Old World monkey, ape, and human sequences). The number of combinations differing at a given number of nucleotides are plotted on the *y* and *x* axes, respectively

In mammals, *Mhc* polymorphism is maintained over long periods of time by positive (balancing) selection at the nonsynonymous sites specifying the PBR of the Mhc molecule. Evidence for the effect of balancing selection is provided by the observation that at the PBR sites (but not in the remainder of the gene), the rate of nonsynonymous substitutions (K_a) significantly exceeds that of synonymous substitutions (K_s; Hughes and Nei 1988, 1989). Table 1 indicates that at the *DAB1* locus, too, the K_a/K_s ratio is greater than 1, as would be expected if the locus were evolving under balancing selection.

The allelic lineages at the primate *DRB1* and certain other class II loci are characterized by the presence of sequence motifs at the sites specifying the PBR (Klein and O'hUigin 1995; O'hUigin 1995). Sequence motifs are also present in the *Dare/Daal/Dama* lineages (Fig. 3) at positions presumably constituting the PBR. None of the fish motifs, however, resembles any of the primate class II motifs, suggesting that the fish and primate PBRs deal with different sets of peptides. (In primates, similar motifs apparently arise independently by convergent evolution in distantly related genes; see O'hUigin 1995 and Klein and O'hUigin 1995).

In comparison with primate class II *B* allelic lineages, the *Dare/Daal/Dama* lineages appear to encompass large sequence differences. In pairwise comparisons of all the

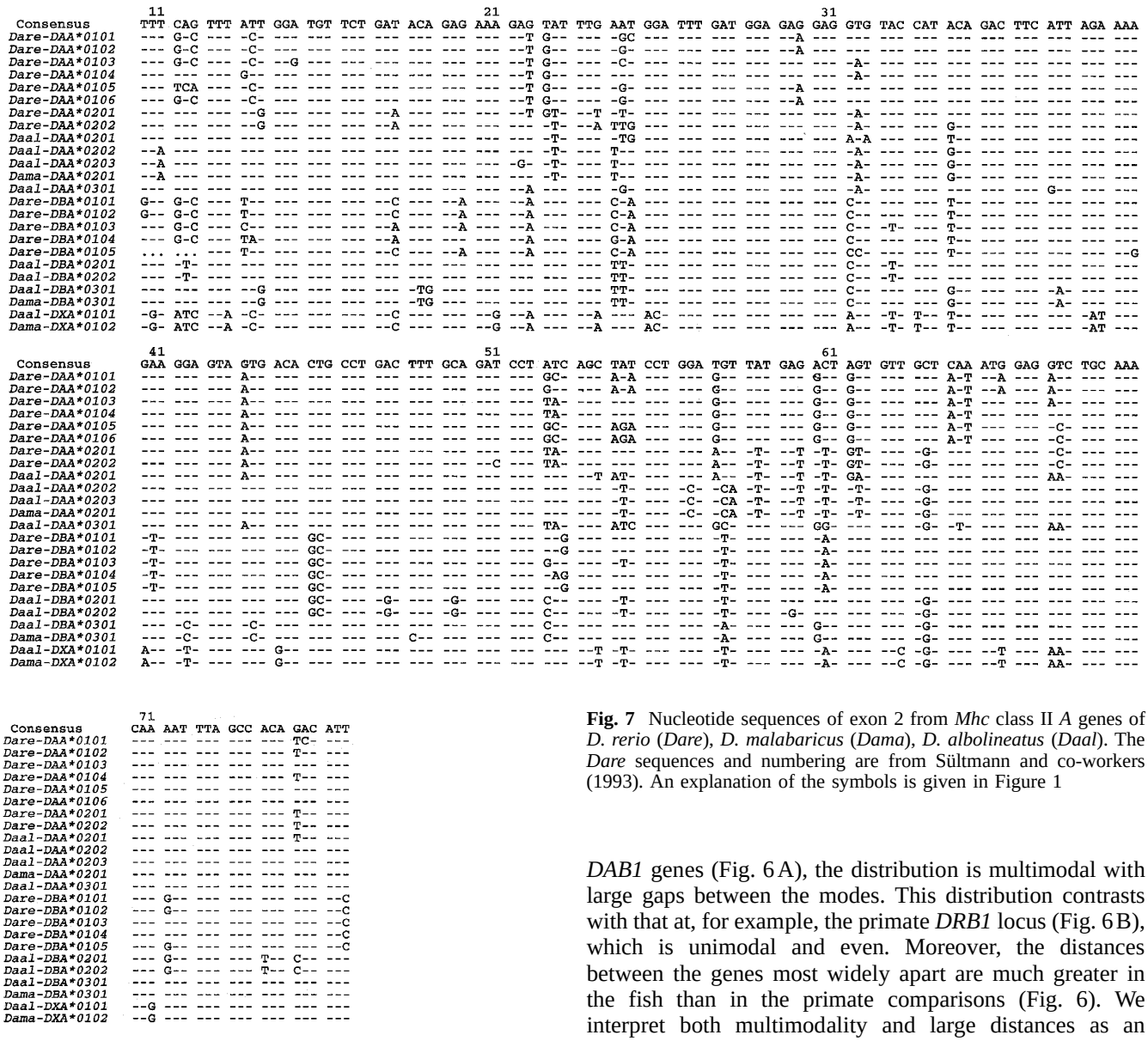
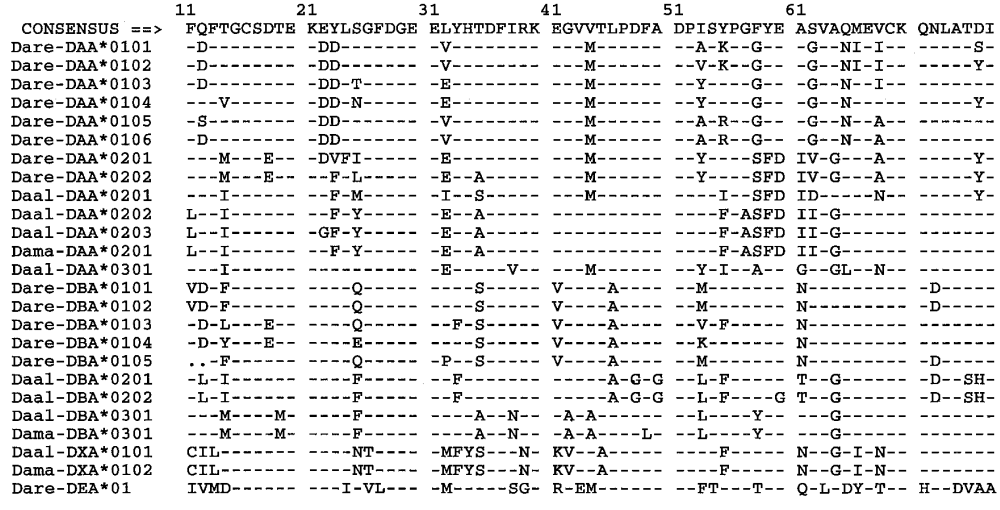


Fig. 7 Nucleotide sequences of exon 2 from *Mhc* class II *A* genes of *D. rerio* (*Dare*), *D. malabaricus* (*Dama*), *D. albolineatus* (*Daal*). The *Dare* sequences and numbering are from Sültmann and co-workers (1993). An explanation of the symbols is given in Figure 1

DAB1 genes (Fig. 6 A), the distribution is multimodal with large gaps between the modes. This distribution contrasts with that at, for example, the primate *DRB1* locus (Fig. 6 B), which is unimodal and even. Moreover, the distances between the genes most widely apart are much greater in the fish than in the primate comparisons (Fig. 6). We interpret both multimodality and large distances as an

Fig. 8 Amino acid sequences translated from the nucleotide sequences in Figure 7. An explanation of the symbols is given in Figures 1 and 3



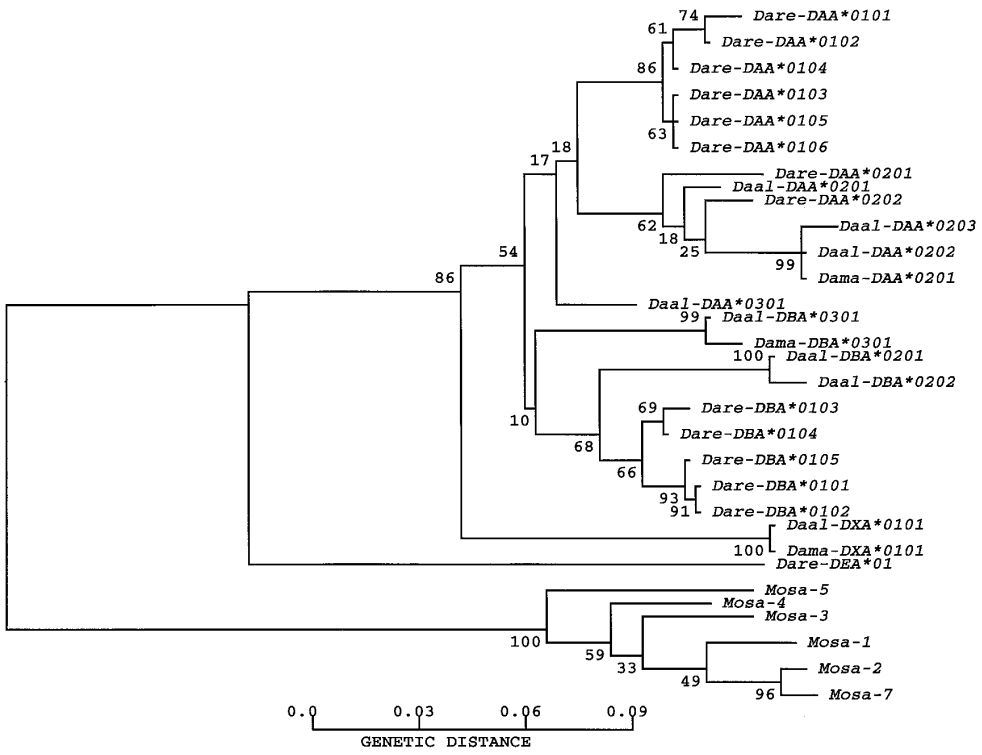


Fig. 9 Phylogenetic tree based on nonsynonymous genetic distances between sequences in Figure 8, as well as those reported by Hardee and co-workers (1995) for the striped bass (*Mosa*). The ten most polymorphic residues were omitted in tree construction using the MEGA package (Kumar 1993). An explanation of the symbols is given in Figure 5. Numbers on branches are bootstrap values

indication that the fish genes studied here are more diverged from each other than the corresponding ape and Old World monkey genes. A large divergence time will result in species-specific lineages and multimodality in pairwise comparisons. This does not automatically imply that the three fish species themselves are very old; on the contrary, the shallow topology of each clade suggests a relatively recent origin. The three species may have diverged from a common ancestor a long time ago (but after the divergence of the main allelic lineages) and there may have subsequently been several divergences in each species lineage, but the latest occurred relatively recently. The common ancestor of the three species contained alleles ancestral to the main extant lineages and during the long separation time the lineages diverged from one another considerably. But each species had a much shorter time for its lineages to diversify because of its relatively recent origin.

There is no good estimate of the evolutionary rate of the fish *Mhc*. The rate estimated for primate class II *DRB* genes at synonymous sites is 1.2×10^{-9} per site per year (Satta et al. 1993) but this is probably too slow to apply to fish, since primates have a slow molecular clock for nuclear genes and the *DRB* genes are among the slowest evolving sequences at synonymous sites. We consider the average mammalian synonymous rate (4.7×10^{-9} per site per year; Li et al. 1985) to be more appropriate. For exon sequences, a minimum value of 0.12 synonymous substitutions per site is observed between *Daal-* and *Dare-DAB1* alleles. This suggests that the most recently shared ancestor of these two species lived some 13 million years ago. The minimum synonymous distances between *Dama* and *Dare* or *Daal* are greater by a factor of 2 than the *Daal-Dare* distances. This again supports the sister grouping of *Dare* and *Daal* as suggested by Meyer and co-workers (1993, 1995).

Class II A polymorphism

Using the primer pair A20290 and A20291, we were able to amplify a 202 bp exon 2 fragment and obtain 13 unique sequences from 12 specimens of *D. malabaricus* and another 12 specimens of *D. albolineatus*. The nucleotide and the translated amino acid sequences of the amplification products are given in Figures 7 and 8, respectively, together with previously published and unpublished *D. rerio* sequences. A phylogenetic tree of these and relevant other sequences appears in Figure 9. The assignment of these sequences to loci remains ambiguous. The majority of the sequences appear to belong to the *DAA* locus; no sequences belonging to the *Dare-DCA* and *Dare-DEA* loci have been found. Two sequences lie between the *DEA* and the *DAA/DBA* clades and are therefore designated *Daal-DXA*0101* and *Dama-DXA*0101* ("X" for "locus unknown"). Four sequences appear to be related to the *DBA* locus, but it could not be decided with certainty whether they are an allelic lineage at this locus or alleles at another, closely-related locus. Another possibility is that all the sequences, in fact, belong to a single locus. Linkage data for the putative *DBA* locus are not available and the locus assignments were made on the basis of genetic distance and differences in the 3′ untranslated regions (Sültmann et al. 1993, 1994 a, b). If we take the most conservative stand and consider only the sequences that belong clearly to either the *DAA* locus group or to the *DBA* group, we find unambig-

uous evidence of trans-species polymorphism. The *DAA* locus sequences fall into three lineages, here designated *DAA*01* through **03*. Trans-species polymorphism is indicated by the intermingling of the lineages from the three species (i.e., the lack of segregation into three clusters according to the species) and the presence of sequences from different species in the same allelic lineage. Thus the *DAA*02* lineage contains sequences from *D. malabaricus*, *D. albolineatus*, and *D. rerio*. The *DBA*03* lineage contains sequences from both *D. malabaricus* and *D. albolineatus*. The interspecies genetic distances of the alleles in each of these two lineages are very small. In general, the branch lengths of the class II *A* sequences are shorter and the segregation into lineages less distinct than those of the class II *B* sequences (compare Figures 5 and 9). This observation may be an indication of a more conservative evolution of the class II *A* loci in comparison with the class II *B* loci. Such a tendency has also been noted for mammalian class II loci, its extreme example being provided by the *DR* gene family in which the *DRB* loci are highly polymorphic, while the *DRA* locus is virtually monomorphic (reviewed in Klein 1986). Either the *DAA*, the *DBA* or both (assuming they represent one locus) appear to be the main functional class II *A* loci of the *Danio* fish group (transcripts have thus far been found only for these genes; Sültmann et al. 1993, 1994a, b). It may, therefore, be presumed that one or both of their products associate with those of the *DAB* locus to form functional $\alpha:\beta$ heterodimers. Both the *DAA* and *DBA* loci, like *DAB*, are under positive selection as indicated by the relative prevalence of nonsynonymous over synonymous substitutions at the PBR sites, and the polymorphism is concentrated in the putative PBR (data not shown).

In conclusion, we find strong evidence for trans-species evolution of class II gene polymorphism at both the *A* and *B* loci in cyprinid fishes. The polymorphism is restricted to apparently functional loci; it is focused on the PBR sites and is apparently maintained by balancing selection. These observations suggest that the *Mhc* carries out the same function in bony fishes as it does in mammals and that this function relies on the complexing of peptides with the Mhc molecules.

Acknowledgments We thank Ms. Lynne Yakes for editorial assistance. This work was supported by grant no. AI 23667 from the National Institutes of Health, Bethesda, MD, and grant no. MCB-9418046 from the National Science Foundation, Washington, D.C.

References

Arden, B. and Klein, J. Biochemical comparison of major histocompatibility complex molecules from different subspecies of *Mus musculus*: evidence for trans-specific evolution of alleles. *Proc Natl Acad Sci USA 79:* 2342–2346, 1982

Barman, R. P. A taxonomic revision of the Indo-Burmese species of *Danio* Hamilton Buchanan (Pisces: Cyprinidae). *Rec Zool Surv India 137:* 1–91, 1991

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

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

Fan, W., Kasahara, M., Gutknecht, J., Klein, D., Mayer, W. E., Jonker, M., and Klein, J. Shared class II *Mhc* polymorphisms between human and chimpanzees. *Hum Immunol 26:* 107–121, 1989

Figueroa, F. and Klein, J. Origin of H-2 polymorphism. *In* C. S. David (ed.): *H-2 Antigens, Genes, Molecules, Function. Series A: Life Sciences 144,* pp. 61–76, Plenum Publishing Corporation, 1987

Figueroa, F., Günther, E., and Klein, J. MHC polymorphism pre-dating speciation. *Nature 335:* 265–267, 1988

Grimholt, U., Olsaker, I., De Vries Linstroem, C., and Lie, Ø. A study of variability in the MHC class II β1 and the MHC class I α2 domain exons of Atlantic salmon (*Salmo salar*). *Anim Genet 25:* 147–153, 1994

Gyllensten, U., Sundvall, M., Ezcurra, I., and Erlich, H. A. Genetic diversity at class II DRB loci of primate Mhc. *J Immunol 146:* 4368–4376, 1991

Gyllensten, U. B. and Erlich, H. A. Ancient roots for polymorphism at the HLA-DQα locus in primates. *Proc Natl Acad Sci USA 86:* 9986–9990, 1989

Hamilton-Buchanan, F. *An Account of the Fishes Found in the River Ganges and its Branches,* Archibald Constable, Edinburgh 1822

Hardee, J. J., Godwin, U., Benedetto, R., and McConnell, T. J. Major histocompatibility complex class II *A* gene polymorphism in the striped bass. *Immunogenetics 41:* 229–238, 1995

Higgins, D. G., Bleasby, A. J., and Fuchs, R. CLUSTAL V: improved software for multiple sequences alignment. *Cabios 8:* 189–191, 1992

Hughes, A. L. and Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature 355:* 167–170, 1988

Hughes, A. L. and Nei, M. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA 86:* 958–962, 1989

Kenter, M., Otting, N., Anholts, J., Jonker, M., Schipper, R., and Bontrop, R. E. *Mhc-DRB* diversity of the chimpanzee. *Immunogenetics 37:* 1–11, 1992

Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol 16:* 111–120, 1980

Klein, D., Ono, H., O'hUigin, C., Vincek, V., Goldschmidt, T., and Klein, J. Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature 364:* 330–334, 1993

Klein, J. Generation of diversity at MHC loci: implications for T-cell receptor repertoires. *In* M. Fougereau and J. Dausset (eds.): *Immunology 80,* pp. 239–253, Academic Press, London, 1980

Klein, J. *Natural History of the Major Histocompatibility Complex,* John Wiley, New York, 1986

Klein, J. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum Immunol 19:* 155–162, 1987

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

Klein, J. and O'hUigin, C. Class II B *Mhc* motifs in an evolutionary perspective. *Immunol Rev 143:* 89–111, 1995

Klein, J. and Takahata, N. The major histocompatibility complex and the quest for origins. *Immunol Rev 113:* 5–25, 1990

Klein, J., Bontrop, R. E., Dawkins, R. L., Erlich, H. A., Gyllensten, E. B., Heise, E. R., Jones, P. P., Parham, P., Wakeland, E. K., and Watkins, D. I. Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics 31:* 217–219, 1990a

Klein, J., Kasahara, M., Gutknecht, J., and Figueroa, F. Origin and function of *Mhc* polymorphism. *Chem Immunol 49:* 35–50, 1990b

Klein, J., Satta, Y., O'hUigin, C., and Takahata, N. The molecular descent of the major histocompatibility complex. *Annu Rev Immunol 11:* 269–295, 1993

Kumar, S., Tamura, K., and Nei, M. *MEGA: Molecular Evolutionary Genetic Analysis Version 1.0,* The Pennsylvania State University, University Park, PA, 1993

Kupfermann, H., Mayer, W. E., O'hUigin, C., Klein, D., and Klein, J. Shared polymorphism between gorilla and human major histocompatibility complex *DRB* loci. *Hum Immunol 34:* 267–278, 1992

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R., and Berns, A. Simplified mammalian DNA isolation procedure. *Nucleic Acids Res 19:* 4293, 1991

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

Li, W.-H., Wu, C.-I., and Luo, C.-C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol 2:* 150–174, 1985

Maniatis, T., Fritsch, E. F., and Sambrook, J. *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1991

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

Mayer, W. E., O'hUigin, C., Zaleska-Rutczynska, Z., and Klein, J. Trans-species origin of *Mhc-DRB* polymorphism in the chimpanzee. *Immunogenetics 37:* 12–23, 1992

McConnell, T. J., Talbot, W. S., McIndoe, R. A., and Wakeland, E. K. The origin of *Mhc* class II gene polymorphism within the genus *Mus. Nature 332:* 651–654, 1988

Meyer, A., Biermann, C. H., and Ortí, G. The phylogenetic position of the zebrafish (*Danio rerio*), a model system in developmental biology: an invitation to the comparative method. *Proc R Soc Lond (Biol) 252:* 231–236, 1993

Meyer, A., Ritchie, P. A., and Witte, K.-E. Predicting developmental processes from evolutionary patterns: a molecular phylogeny of the zebrafish (*Danio rerio*) and its relatives. *Phil Trans R Soc Lond (Biol) 349:* 103–111, 1995

Nelson, J. S. *Fishes of the World,* John Wiley, New York, 1984

O'hUigin, C. Quantifying the degree of convergence in primate Mhc-DRB genes. *Immunol Rev 143:* 123–140, 1995

Ono, H., Klein, D., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H., and Klein, J. Major histocompatibility complex class II genes of zebrafish. *Proc Natl Acad Sci USA 89:* 11886–11890, 1992

Ono, H., O'hUigin, C., Tichy, H., and Klein, J. Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. *Mol Biol Evol 10:* 1060–1072, 1993a

Ono, H., O'hUigin, C., Vincek, V., and Klein, J. Exon-intron organization of fish major histocompatibility complex class II *B* genes. *Immunogenetics 38:* 223–234, 1993b

Ono, H., O'hUigin, C., Vincek, V., Stet, R. J. M., Figueroa, F., and Klein, J. New β chain-encoding *Mhc* class II genes in the carp. *Immunogenetics 38:* 146–149, 1993c

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

Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74:* 5463–5467, 1977

Sato, A., Figueroa, F., O'hUigin, C., Reznick, D. N., and Klein, J. Major histocompatibility complex genes of the guppy, *Poecilia reticulata:* identification and cloning. *Immunogenetics 43:* 38–49, 1996

Satta, Y., Takahata, N., Schönbach, C., Gutknecht, J., and Klein, J. Calibrating evolutionary rates at major histocompatibility complex loci. *In* J. Klein and D. Klein (eds.): *Molecular Evolution of the Major Histocompatibility Complex,* pp. 51–62, Springer, Heidelberg, 1991

Satta, Y., O'hUigin, C., Takahata, N., and Klein, J. The synonymous substitution rate of the major histocompatibility complex loci in primates. *Proc Natl Acad Sci USA 90:* 7480–7484, 1993

Slierendregt, B. L., van Noort, J., Bakas, R. M., Otting, N., Jonker, M., and Bontrop, R. E. Evolutionary stability of trans-species major histocompatibility complex class II *DRB* lineages in man and rhesus monkey. *Hum Immunol 35:* 29–39, 1992

Stanley, S. M. A theory of evolution above the species level. *Proc Natl Acad Sci USA 72:* 646–650, 1975

Sültmann, H., Mayer, W. E., Figueroa, F., O'hUigin, C., and Klein, J. Zebrafish *Mhc* class II α chain-encoding genes: polymorphism, expression, and function. *Immunogenetics 38:* 408–420, 1993

Sültmann, H., Mayer, W. E., Figueroa, F., O'hUigin, C., and Klein, J. Organization of *Mhc* class II *B* genes in the zebrafish (*Brachydanio rerio*). *Genomics 23:* 1–14, 1994a

Sültmann, H. Organisation and Charakterisierung von *Mhc* Klasse II *A* und *B* Genen des Zebrabärblings (*Brachydanio rerio*). Universität Tübingen, 1994b

Wakeland, E. K., Tarruzzee, R. W., Lu, C.-C., Potts, W., McIndoe, R. A., Talbot, W. S., and McConnell, T. J. *In* C. S. David (ed.): *H-2 Antigens, Genes, Molecules, Function,* pp. 139–153, Plenum Press, New York, 1987

Walker, R. B. and McConnell, T. J. Variability in an *MhcMosa* class II β chain-encoding gene in the striped bass (*Morone saxatilis*). *Dev Comp Immunol 18:* 325–342, 1994

Wu, T. T. and Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med 132:* 211–250, 1970