Molecular Genetic Diversity and Evolution at the MHC DQB Locus in Four Species of Pinnipeds

A. Rus Hoelzel,* J. Claiborne Stephens,† and Stephen J. O’Brien†

*Department of Biological Sciences, University of Durham, Durham, England; and †Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland

Variation was investigated at exon 2 (including part of the putative peptide-binding region) of the class II major histocompatibility complex (MHC) DQB locus for two congeneric phocid seal species and two congeneric otariid seal species. Polymorphism in one phocid species, the southern elephant seal (Mirounga leonina), was comparable to that seen in human populations, while the other phocid, the northern elephant seal (Mirounga angustirostris), has been through a severe population bottleneck and exhibited much less variation at this locus. A phylogenetic comparison of the four species was consistent with the trans-specific pattern of evolution described for other taxa at this locus, and relative nonsynonymous and synonymous substitution rates suggest the maintenance of polymorphisms by natural selection. A comparison of sequence patterns also suggested that some variation could have been generated through recombinational events, primarily within genera. These results suggest a pattern of evolution of the immune response in pinnipeds similar to that in terrestrial mammal species.

Introduction

Genes in the human class II major histocompatibility complex (MHC) at the DR, DQ, and DP loci encode alpha and beta glycopeptide chains. These form cell surface heterodimers on antigen-presenting cells which bind processed antigen peptide fragments (see reviews in Germain and Margulies 1993; Harding 1996). In many species, these loci are highly polymorphic, especially in the second exons of the beta genes, and the variation is thought to be maintained by natural selection through heterozygote advantage (Hughes and Nei 1989; Thurst et al. 1997) or frequency-dependent selection (Takahata and Nei 1990; Hill et al. 1992). Natural selection for diversity is typically thought to promote pathogen recognition. Recognition of the complex between antigen-presenting cells and invading peptides by the T-cell receptor of CD4+ lymphocytes (in humans) leads to an immune response through T-cell activation (Guillet et al. 1987; Sette et al. 1987). MHC diversity at these and other immune system loci may reflect the pathogen environment of a given species (e.g., Potts and Slev 1995), and various associations between specific infectious diseases and MHC genotypes have been identified, including Marek’s disease (Longenecker et al. 1977), malaria (Hill et al. 1991), and HIV-1 (Kaslow et al. 1996). Evidence for the maintenance of alleles over long periods of evolutionary time (see Edwards and Hedrick 1998) and a high proportion of nonsynonymous change in the peptide-binding region (Hughes and Nei 1989; Yuhki and O’Brien 1990; Miller, Withler, and Beacham 1997) support the contention that polymorphism is maintained by natural selection.

It has been suggested that the pathogen environment of marine mammals may provide a diminished selective pressure for maintaining MHC polymorphism (Trowsdale, Groves, and Arnason 1989; Slade 1992; Murray, Malik, and White 1995), due to a relatively low prevalence of infectious disease in the marine environment. Slade (1992) analyzed restriction fragment length polymorphisms (RFLPs) for four restriction enzymes and 24 southern elephant seals from two populations. He used heterologous probes for MHC class I (HLA-B7) and class II genes (DQA, DQB, and DRB) and compared the average percentage of difference between banding patterns with those seen in various other species. In general, there was apparently less variation for all probes in marine mammals than in terrestrial mammals. The highest level of variation for the southern elephant seal (SES) was for the DQB probe, but this was still much lower than comparable values seen for other species. As Slade (1992) points out, however, the comparison of levels of variation between taxa can be misleading, especially for relatively low resolution genetic analyses. Furthermore, the function of the immune response is too poorly understood for most non-human taxa to permit the identification of which loci are critical, and the critical sequences could be short enough to miss detection by the RFLP method.

In this study, we investigate sequence variation at exon 2 of the DQB gene, including part of the putative peptide-binding region, for four species of pinnipeds. We chose this locus because it is known to be highly polymorphic in primates (e.g., Bugawan and Erlich 1991) and a variety of other terrestrial species. We compare the level of variation in the SES at this locus to that seen in human populations and investigate the relationship between alleles in different pinniped species to evaluate whether pinnipeds have a uniformly reduced level of diversity at the MHC relative to terrestrial mammals. The study species include two congeneric species in the superfamily Phocidae and two congeneric species in the superfamily Otariidae (classification as in Riedman 1990).
Materials and Methods

PCR Amplification and Single-strand Conformational Polymorphism Analysis

Samples were collected from live animals by biopsy sampling as described in Hoelzel et al. (1993), and the DNA was extracted by standard protocol. Northern elephant seal (NES) samples were collected at Año Nuevo, California; SES samples were collected at Punta Delgada, Argentina; New Zealand fur seals (NFS; Arctocephalus fosteri) were collected at Kangaroo Island, South Australia; and Antarctic fur seals (AFS; Arctocephalus gazella) were collected at Bird Island in the South Atlantic. Primers for the exon 2 region of the DQB gene were designed from a comparison of cow, human, and dog sequences from the Gen-EMBL database: sense—5’-TCGTGATCCAGTTAAGG; antisense—5’-ACGTCTTCTGCTGTCTCCA. A 142-bp segment was amplified by PCR from each individual, and 33P alpha dATP was incorporated into the product under the following assay conditions: 1.5 mM MgCl2; 10 mM Tris; 50 mM KCl; 100 μM dTTP, dCTP, dGTP; 5 μM dATP; 1 μCi alpha 33P dATP; and 50 pM of each primer. Labeled PCR product was denatured at 95°C for 5 min, chilled on ice for 1 min, and loaded onto a nondenaturing gel (4.5% acrylamide [37.5/1 acrylamide/bis], 10% glycerol) run at 200V onto a nondenaturing gel (4.5% acrylamide [37.5/1 acrylamide/bis], 10% glycerol) run at 200V.

DNA Sequencing and Phylogenetic Analysis

PCR product was prepared for cloning using the double gene-clean-cloning kit (Bio 101). This entailed polishing the ends using T4 DNA kinase and polymerase I. Insert DNA was then cloned into blue script sk+ vector and transformed (after Chung and Miller 1988) into the XL1 blue strain of Escherichia coli using blue/white selection. Five to seven clones were sequenced in both directions for each representative of a putative genotype identified through SSCP analysis. A given putative genotype was sequenced from two to four individuals with identical SSCP phenotypes. Sequencing was by the dye-primer method for the ABI automated sequencing system.

Mitochondrial DNA from the control region was amplified using published primers (Hoelzel and Green 1992) and sequenced on the automated ABI system using the same primers. Four unique NFS sequences and three AFS sequences were determined from six individuals of each species (accession numbers AF111046-52) and compared with published data for NES and SES (Hoelzel et al. 1993).

Unique alleles were compared for phylogenetic relationships using maximum-parsimony (Swofford and Berlocher 1987) and neighbor-joining (Saitou and Nei 1987) analyses with the PAUP and PHYLIP computer programs. The transition/transversion ratio for the DQB locus was approximately 1:1 (determined empirically). For maximum parsimony, a consensus tree was derived by bootstrap analysis with 1,000 replications. Branches with greater than 50% support were retained. For neighbor-joining analysis, genetic distances were derived using the Jukes-Cantor correction, and a bootstrap analysis with 1,000 replications was used.

Recombination and Substitution Rate Analyses

The sequences were inspected for potential recombination events, a process facilitated by the application of published algorithms (Stephens 1985; Hudson and Kaplan 1985) and their updated versions (see Stephens 1985 for statistical methods). Results from the statistical tests were used to guide the subsequent inspection of sequences and their interpretation. In addition, all variable nucleotide sites were mapped onto the phylogenetic trees suggested by PAUP and Neighbor-Joining to see if recombination was a sensible alternative to purely substitutional processes in the generation of these alleles. The inferred mosaic nature of several alleles is best seen by juxtaposing the alternative trees derived from these recombination analyses.

The number of synonymous substitutions per synonymous site (dS) and the number of nonsynonymous substitutions per nonsynonymous site (dN) (Nei and Gojobori 1986) and the standard error of those measures (Nei and Jin 1989) were derived using the MEGA program (Kumar, Tamura, and Nei 1993) based on an alignment of unique alleles.

Results

Variation

For the purpose of this study, we concentrated on the phocid species for an assessment of levels of variation. Table 1 summarizes the diversity of genotypes in all four species. All four species were polymorphic, and the absolute sequence difference between alleles was up to 8.5% within species and up to 8.5% between species. Among 109 NES individuals, there were eight alleles, while there were only two alleles among 69 NES indi-
Sequence Alignment and Phylogenetic Analyses

An alignment comparing the DNA sequences of all 16 DQB alleles and three outgroups is presented in figure 1, and an alignment of translated sequences is presented in figure 2. The DNA sequence alignment provided by Pileup (Devereux, Haeberli, and Smithies 1984) did not include the indels at base pairs 118 and 125; however, the inclusion of these two indels conserved 3 bp of sequence. The amino acid alignment illustrates the conservation of motifs between the pinniped and human DQB1 sequences that are not shared with the human DQB2 sequence. This and the fact that multiple clones never revealed more than two sequences at a locus suggest that we were successful at amplifying just one locus from the pinniped species and that the amplified locus is the homolog to the human DQB1 locus.

Two maximum-parsimony consensus trees are shown in figure 3. Figure 3 A shows a phylogeny based on 142 bp of mtDNA sequence from the 5'9 end of the control region (base pairs 210–351 in Hoelzel, Hancock, and Dover 1993). This tree shows a clear reciprocal monophyletic distinction between genera and between species, and strong bootstrap support between the Phocidea and Otarioidea superfamilies. Figure 3B shows a phylogeny derived from the 142-bp DQB sequence. For DQB, there were 34 equally parsimonious trees. Bootstrap analysis supported five of the nodes, shown in figure 3B. There is no clear distinction between alleles from congenic species, which share terminal clades for both genera. Alleles from the two seal superfamilies are intermingled in a single main clade. The neighbor-joining tree for the DQB alleles is shown in figure 4. This tree suggests some distinction between alleles from different seal superfamilies; however, the nodes separating alleles from the two lineages are very shallow and poorly supported by the bootstrap analysis, and one allele from the fur seal species (NFS4) groups with the elephant seals.

Recombination

Figure 5 illustrates the mapping of nucleotide sequence variants (based on variable sites; see fig. 1) to a phylogenetic tree suggested by the neighbor-joining algorithm, but in which almost all instances of homoplasy among seals are accounted for by recombination. Figure 5A reflects the first 29 variable sites and conforms largely with the neighbor-joining tree of figure 4 and the logical taxonomic groupings. As noted in figure 4, one fur seal sequence (NFS4) clusters with the elephant seal sequences. Four instances of homoplasy (circled in fig. 5A) are explained by three putative recombination events. The first of these involves AFS1 (fig. 6): the 5'9 end of AFS1 is identical to that of NFS1 but is otherwise identical to AFS3 and AFS4 through nucleotide 100 (variable site 29). Similarly, NFS4 is essentially an elephant seal sequence: it is nearly identical to SES7, differing only by a unique change at site 21, an inferred parallel transition at site 29, and retention of the G at site 1 seen in all fur seals. Hence, we invoke recombination in which a typical fur seal sequence recombined with a sequence like SES7 to produce NFS4. Our third proposed recombination event, in SES2/NES1, invokes
replacement of the 5' end by a typical southern elephant seal sequence, changing variable site 10 from C to G.

Each of these events invokes a recombination donor that currently exists in the population. However, consideration of additional events is also warranted. Figure 5B shows a cladogram for the 3' end of the sequence (variable sites 30–53). Several events are suggested by the contrast between figure 5A and B. First, note that in figure 5B, AFS3 and AFS4 are clustered with several northern and southern elephant seal sequences, separate from all other fur seal sequences. There are several combinations of recombination events that could lead to this shift in affinities, but it seems clear that at least two events would need to occur. One of these would seem to involve the common ancestor of AFS3 and AFS4 and would lead to the acquisition of variants more typical of elephant seals, as well as some unique variants (figs. 5B and 6). In particular, note that the 3' clustering of variants unique to AFS3 and AFS4 is highly significant ($P < 0.006$) by Stephens' (1985) algorithm. At least one more recombination event is needed to explain SES3 and SES4's position with SES2 and the two NES sequences, distinct from the main body of SES sequences, which now includes NFS1 in addition to NFS4.
Diversity and Evolution of Pinniped DQB

**Fig. 5.** Variable sites are labeled (1–53) starting from the 5' end (see fig. 1). A, Phylogeny based on the first 29 variable sites. Three instances of recombination are inferred involving the 5' ends of AFS1, NFS4, and the common sequence of SES2/NES1 and are shown encircled. Three instances of inferred parallel change between an outgroup and a seal variant are designated with a “P” next to the variant. A fourth instance of parallel change, between NFS4 and three AFS sequences, is also denoted in this manner. B, Phylogeny based on variable sites 30–53. Three parallelisms between seals and the outgroups are inferred, as well as a reversal within seals. No recombination events within this region are depicted, although several are inferred to explain the difference between parts A and B (see text).

**Synonymous and Nonsynonymous Change**

To identify a putative peptide-binding domain, the DQB allele AFS3 was aligned with human DRB1–1105 (Apple et al. 1993). The alignment showed 80.9% homology and corresponded to amino acids 14–60 (based on comparison with the sequence presented in Otting et al. 1992). The crystalline structure of the human DR alpha and beta chains has been established, and residues in the antigen-binding groove have been identified (Brown et al. 1993). Assuming that the structure of DQB is similar, the same residues were considered in isolation of the rest of the pinniped DQB exon 2 sequence (see fig. 2).

The number of synonymous substitutions per synonymous site ($d_S$) and the number of nonsynonymous substitutions per nonsynonymous site ($d_N$), with the standard error of those measures for the entire sequence, are $d_N = 0.0652 \pm 0.0137$ and $d_S = 0.0375 \pm 0.017$. The $d_S/d_S$ ratio for the whole sequence (1.74) is similar to that seen for the PBR region at MHC class I loci in the domestic cat (1.81; Yuhki and O'Brien 1990). Our sample size is small for the statistical comparison of within the PBR versus outside the PBR. However, of the 15 amino acid residue sites among the pinniped alleles showing DNA polymorphism (residues 1, 4, 12, 13, 15, 17, 24, 29, 34, 37, 38, 40, 42, 43, and 44; see fig. 2) most (13 out of 15) are dominated by nonsynonymous substitutions, and most synonymous change is found at just one residue (number 43 in fig. 2). Site 43 is one of the eight putative PBR sites, and all of the synonymous

**Fig. 6.** Mosaic structure of AFS1. AFS1 is piecemeal identical to NFS1, AFS4, and AFS2. Note that in the central region (sites 14–29), AFS1 is identical to AFS4 but differs from AFS2 at nine nucleotide sites (not shown). Diagnostic variable sites are given above AFS4.
changes within the PBR sites sequenced in this study are at this residue.

Discussion

We investigated the radiation of alleles at the pinniped DQB1 locus and their levels of polymorphism within species. The results show relatively high levels of variability and a pattern of evolution that is consistent with observations from a variety of other mammalian species, including humans. For those species, natural selection, recombination, and trans-specific polymorphism at this locus have been invoked. Previous studies have suggested relatively low levels of MHC variation in several marine mammal species and large terrestrial mammal species (Trowsdale, Groves, and Arnason 1989; Slade 1992; Murray, Malik, and White 1995); however, for one of the species investigated for DQB variation in this study (SES), we found substantial levels of variation.

We suggest that variation in SES is comparable to that in humans for the following reasons. In a study of allelic diversification at the DQB1 locus of humans (at a 120-bp sequence within exon 2 that overlaps with the region sequenced in our study), Gyllensten, Lashkari, and Erlich (1990) found 12 alleles among 30 individuals of differing ethnic origins. The same study compared eight primate species and found percentages of genetic differences between alleles ranging from 0.8% to 19.2% for this 120-bp sequence. For the same 120-bp sequence, we found eight alleles among 109 individual SESs from a single breeding population, and genetic differentiation among the alleles of four species ranging from 0.8% to 11.7%. The greatest sequence divergence over this range within a species (AFS) was 10.8%, compared with 14.2% maximum difference between human alleles at this locus (Gyllensten, Lashkari, and Erlich 1990). While the level of variation among pinniped alleles is consistently lower, it appears higher at this locus than that seen in previous studies of MHC loci in marine mammals (e.g., Trowsdale, Groves, and Arnason 1989; Slade 1992; Murray, Malik, and White 1995), and nearly as high as that seen for human populations. We interpret this to indicate a high level of variation at this locus among the four pinniped species investigated, particularly within the SES.

The phylogenetic relationship among primate DQB alleles (Gyllensten, Lashkari, and Erlich 1990) shows a pattern similar to that seen for the pinniped species. The main difference is that, while the primate alleles fell into distinct lineages (DQB1–DQB4), there is no similar structure seen among the pinniped alleles. Instead, the relationship appears to be an extended polymony, with some structure (especially in the neighbor-joining tree) reflecting the two main groups within the pinnipeds: the otariids and the phocids. Otariids and phocids fall into distinct lineages in phylogenies using mtDNA (Arnason et al. 1995; this study) and nuclear DNA markers such as aldolase (Slade, Moritz, and Heideman 1994). The distinction between species in these two superfamilies is also well established based on morphological criteria (see King 1983). The nature of the radiation of otariid and phocid pinnipeds has been controversial (see discussion in Riedman 1990), but is likely to be at least as early as the radiation among, for example, Old World primate families (see Slade, Moritz, and Heideman 1994; Stewart and Disotell 1998). However, pinniped phylogenies derived using the DQB loci do not reflect these taxonomic differences, in that different species and different genera are mixed together within the same allelic lineages. This pattern of evolution has been referred to as “transpecific” (Klein 1987). It is thought to indicate the operation of natural selection through either overdominance or frequency dependence (see Hughes and Nei 1989; Takahata and Nei 1990).

The distribution of mutations across the DQB sequence is discontinuous, which can lead to a shallow tree when relatively few sites are informative. In this case, there are 53 variable sites, of which 21 are polymorphic among the pinniped species and 17 are variant in two or more sequences. A longer nuclear DNA sequence that includes a similar number of clustered variable sites shows a phylogenetic pattern with 97% bootstrap support separating the phocid and otariid lineages (using 400 bp from the Aldolase-A gene; Slade, Moritz, and Heideman 1994). Our mtDNA comparison included a sequence showing a similar level and pattern of diversity and the same overall length. MtDNA is haploid and therefore has a smaller effective population size; however, the comparison with the DQB sequence shows a pattern that is more distinct than would be expected due to this difference alone. The DQB phylogeny suggests that selection has maintained alleles for longer than the average speciation event interval, although we cannot conclude unequivocally that this is the case.

Hughes and Nei (1988, 1989) reported on the relative rates of nonsynonymous and synonymous substitutions in primate species at MHC loci. They proposed that since overdominant selection is known to increase both the extent of polymorphisms and the rate of codon substitution (Maruyama and Nei 1981), a greater proportion of nonsynonymous substitutions within (compared to outside) the codons of the peptide-binding region would indicate that the polymorphism was caused mainly by overdominant selection. Although our study did not focus on the region in which PBR sites are found in the highest concentration in primates, most polymorphic sites (13 out of 15) showed predominantly nonsynonymous change. Among these, four are putative PBR sites. It is also possible that PBR sites other than those determined by reference to the human data are equally or more important, as the exact structure has not been determined for these or for closely related taxa.

The recombinational inferences suggested in the results explain the statistical associations (Stephens 1985), and could explain most of the homoplasy seen among seal sequences (figs. 3–5). It is difficult to argue rigorously that recombination is heavily favored over recurrent mutation to explain the homoplasy among seal sequences, and several authors have indeed inferred sequence convergence in the PBR (O’Uigin 1995; Edwards and Hedrick 1998). Clearly, many instances of homoplasy in the seal data set are due to one or two
variants being shared between otherwise “unrelated” sequences (fig. 5), so processes of parallel or reverse mutation cannot be excluded. One could argue that much of the homoplasy seen is due to hypervariability at several sites. Indeed, six of the variable sites have three or even four nucleotides present among seal sequences, and an additional eight sites have three or more distinct nucleotides when outgroups are considered. Even so, there is only a modest amount of homoplasy between the outgroups and seals (fig. 5), which suggests that a portion of the homoplasy seen among seal sequences is due to other factors, such as recombination. Our strategy has been to determine whether recombination is a plausible mechanism given the sequences currently seen among these species. Most of the recombination events we invoke are plausible in that potential recombination donors currently exist in the population.

Whether recombination or recurrent nucleotide substitution is a better explanation for the pattern of allelic diversity seen for this locus depends on a number of factors, especially the relative rates of such processes. Currently, we can only guess that the recombination rate for this region is on the order of $10^{-6}$, extrapolating from the genomewide human estimate (see Lander and Schork 1994). This would seem to be two or three orders of magnitude higher than genomewide substitution rates, although bona fide hot spots for mutation may also exist. Regardless of which process contributes most to this pattern of variation, both are time-dependent processes, and it is key to note that considerable time must have elapsed to accumulate this level of variation. The fact that some of this variation appears to be trans-specific further suggests that the invoked recombination events transpired in the relevant common ancestors: that of both fur seals, that of both elephant seals, and even in that of otariids and phocids. The central point is that if we assume that the recombination rate for this region is small, the numerous instances of homoplasy among seals suggest that many of the variants have coexisted for a substantial period of time, allowing recombination to shuffle them into the combinations we see today. Such maintenance of alternative alleles over long periods of time suggests the action of natural selection, as in terrestrial mammals (see Satta 1997).

One of the two species for which a large number of individuals were included in this study (NES) is known to have been through a severe population bottleneck approximately 110 years ago. Earlier studies have investigated molecular genetic variation in the NES and found greatly reduced levels of variation at both mitochondrial (Hoelzel et al. 1993) and nuclear markers (Bonnell and Selander 1974; Hoelzel et al. 1993). A simulation model approach incorporating both demographic and genetic data indicated that the population consisted of only 10–30 seals at its nadir (Hoelzel et al. 1993). In the current study, we find only two DQB alleles in the northern species, compared with eight alleles in the closely related SES. The relatively high level of variation in the SES and the relationship between alleles among species suggest that the low level of variation in the NES is more a reflection of the demographic history of the species than a consequence of living in a marine environment. Low levels of variation in the MHC in endangered species have been implicated in affecting their response to infectious disease (O’Brien and Eversmann 1988), although no such effect has yet been proposed for the NES.

In summary, these data indicate a pattern of genetic diversity at the class II MHC DQB locus that is comparable to that seen for terrestrial mammal species. There is no indication that this locus in particular exhibits less variation for these species than expected based on this kind of a comparative analysis.

Acknowledgments

We thank Claudio Campagna for providing tissue samples of SESs and for assistance in obtaining samples from NESs, Tom Arnbom for samples of SESs, Burney Le Boeuf for samples of NESs, Peter Shaughnessy for samples of NFSs, and Ian Boyd and the British Antarctic Survey for samples of AFSSs.

LITERATURE CITED


Axel Meyer, reviewing editor

Accepted January 8, 1999