Immunogenetics (2006) 58: 283–296
DOI 10.1007/s00251-006-0080-y

ORIGINAL PAPER

C. S. Baker · M. D. Vant · M. L. Dalebout ·
G. M. Lento · S. J. O'Brien · N. Yuhki

# Diversity and duplication of *DQB* and *DRB*-like genes of the MHC in baleen whales (suborder: Mysticeti)

Received: 5 September 2005 / Accepted: 21 December 2005 / Published online: 28 March 2006
© Springer-Verlag 2006

**Abstract** The molecular diversity and phylogenetic relationships of two class II genes of the baleen whale major histocompatibility complex were investigated and compared to toothed whales and out-groups. Amplification of the *DQB* exon 2 provided sequences showing high within-species and between-species nucleotide diversity and uninterrupted reading frames consistent with functional class II loci found in related mammals (e.g., ruminants). Cloning of amplified products indicated gene duplication in the humpback whale and triplication in the southern right whale, with average nucleotide diversity of 5.9 and 6.3%, respectively, for alleles of each species. Significantly higher nonsynonymous divergence at sites coding for peptide binding (32% for humpback and 40% for southern right) suggested that these loci were subject to positive (overdominant) selection. A population survey of humpback whales detected 23 alleles, differing by up to 21% of their inferred amino acid sequences. Amplification of the *DRB* exon 2 resulted in two groups of sequences. One was most similar to the *DRB*3 of the cow and present in all whales screened to date, including toothed whales. The second was most similar to the *DRB*2 of the cow and was found only in the bowhead and right whales. Both loci showed low diversity among species and apparent loss of function or altered function including interruption of reading frames. Finally, comparison of inferred protein sequence of the *DRB*3-like locus suggested convergence with the *DQB*, perhaps resulting from intergenic conversion or recombination.

**Keywords** Overdominant selection · Diversity · Cetaceans · Dolphins

**Electronic Supplementary Material** Supplementary material is available for this article at http://dx.doi.org/10.1007/s00251-006-0080-y

C. S. Baker (✉) · M. D. Vant · M. L. Dalebout · G. M. Lento
School of Biological Sciences,
University of Auckland,
Private Bag 92019,
Auckland, New Zealand
e-mail: cs.baker@auckland.ac.nz
Tel.: +64-9-3737599
Fax: +64-9-3737417

S. J. O'Brien · N. Yuhki
Laboratory of Genomic Diversity,
National Cancer Institute,
Frederick, MD, USA

## Introduction

The class II genes are attractive candidates for comparative evolutionary studies of the mammalian major histocompatibility complex (MHC) because of their diversity and homologous relationships (Klein 1986). In placental mammals, five MHC class II genes coding for peptide binding proteins have been identified: the *DQ*, *DR*, *DN*, *DO*, and *DP*, although the role of the *DO* is less certain (Beck and Trowsdale 1999). Individual class II alpha and beta-chain genes are thought to be orthologous across placental mammals, having diverged (perhaps through duplication) early in mammalian evolution (Takahashi et al. 2000). In contrast, class I genes seem to have evolved independently in different mammalian orders and homologues can be identified only in closely related species (Yeager and Hughes 1999). Class II glycopeptide chains are highly polymorphic in some species. In particular, exon 2 of the beta chains, which includes the Antigen Recognition Sites or Peptide Binding Region (PBR), shows high levels of genetic variation in some species (Parham and Ohta 1996). The proposed mechanism behind maintenance of this variation is overdominant selection or heterozygote advantage (Hughes and Nei 1988). This hypothesis is supported by two findings: (1) a high proportion of nonsynonymous (amino acid altering) relative to synonymous changes in the peptide binding site (Hughes and Nei 1988) and (2) the persistence of alleles in the population for much longer than expected under neutral coalescence (Parham and Ohta 1996). This persistence can extend across the speciation process, resulting in transspecific sharing of alleles as observed in class I genes of humans and chimpanzees (Lawlor et al. 1988).

However, not all class II alpha and beta-chain genes are found in all mammals, and not all show high diversity (Parham 1999). Of the four or five found in most mammals, the *DQB* and *DRB* are the most widely studied and often the most diverse (e.g., Takahashi et al. 2000). Diversity of the *DQB* or *DRB* has been characterized in humans and other primates, mice and domesticated mammals, including cattle, pigs, sheep, goat, red deer, and cats (Yuhki and O'Brien 1997; Mikko et al. 1999; Parham 1999). Surveys of wild populations have revealed a few notable exceptions to the general pattern of high MHC diversity in mammals. For example, the cheetah and the Asiatic lion have considerably reduced variation in MHC compared to other populations of carnivores (O'Brien et al. 1985; Yuhki and O'Brien 1990). It has been suggested that MHC diversity in these species was lost as the result of one or more population "bottlenecks" (Yuhki and O'Brien 1990). Reduced MHC diversity in the moose and the roe deer has been attributed to relatively ancient (100,000 years ago) and recent (200–500 years ago) bottlenecks, respectively (Mikko et al. 1999).

More surprising are reports of reduced MHC diversity in marine mammals, including baleen whales (suborder Mysticeti; Trowsdale et al. 1989), and pinnipeds (subfamily Phocidae; Slade and McCallum 1992) revealed by Southern blot analyses. These reports suggest that cetaceans, which evolved from an ungulate lineage, and pinnipeds, which evolved from a carnivore lineage, have both lost MHC diversity as a result of adapting to the marine environment. One explanation for this loss is a decrease in the exposure to microparasitic diversity in the marine environment, or an "escape" from microparasites adapted to terrestrial mammals (Slade and McCallum 1992). Alternative (nonexclusive) explanations include demographic factors, such as population bottlenecks due to recent human exploitation (Baker and Clapham 2004), a reduced mutation rate due to large body size (Martin and Palumbi 1993), or protection by the marine environment from mutagens (e.g., gamma rays; Schlotterer et al. 1991).

Evidence of a reduced pathogen load in the marine environment (McCallum et al. 2004) or of a reduced MHC diversity for cetaceans remains debatable. Only three surveys of population diversity of cetacean class II genes (*DQ* or *DR*) have been reported for toothed whales, the beluga (*Delphinapterus leucas*), the narwhal (*Monodon monoceros*) (family Monodontidae; Murray et al. 1995; Murray and White 1998), and the baiji (*Lipotes vexillifer*) or Chinese river dolphin (Yang et al. 2005). A wider survey of sequence diversity in *DQB* of whales and dolphins by Hayashi et al. (2003) included only one individual for most species. Class I sequences from cetaceans have been published only for the gray whales (*Eschrichtius robustus*; Flores-Ramirez et al. 2000) and the bottlenose dolphin (*Tursiops truncatus*; Shirai et al. 1998). All studies have found evidence of positive Darwinian selection, as demonstrated by high levels of nonsynonymous substitutions at residues of the PBR, but, for the class II loci of the beluga and narwhal, population level diversity was low compared to most other terrestrial mammals. For these two sister taxa, Murray and White (1998) concluded that the overdominant selection against pathogens was a force in evolution of the class II genes but that this was limited by a slow overall rate of mutation or a small effective population size during relatively recent speciation. In contrast, the baiji displayed high diversity at the *DQB*, perhaps attributed to the immunological challenges of life in a riverine, rather than marine, habitat (Yang et al. 2005).

In this study, we present results from a comparative investigation of molecular diversity of class II genes among baleen whales. We attempted to account for a potential organismal specific bias in evolutionary histories of cetaceans by choosing species (e.g., the humpback and southern right whale) from two baleen whale families and by including toothed whales, ruminants, primates, and carnivores as out-groups, where appropriate. Our results both confirm and contrast with some previous studies of the cetacean MHC, contributing to a greater phylogenetic representation of class II gene diversity among mammals.

## Materials and methods

### Tissue samples and DNA extraction

Tissue samples were collected from beach-cast or stranded whales and dolphins and from free-swimming individuals using a small stainless steel biopsy dart deployed from a crossbow (Lambertsen 1987; Baker et al. 1993). A list of all species, including both their common and scientific names, is shown in Table 1. For the bowhead whale, *Balaena mysticetus*, tissue was obtained from the aboriginal/subsistence hunt through the North Slope Borough, Alaska, USA. For the northern bottlenose whale, *Hyoperoodon ampullatus*, DNA was obtained from the tooth of an animal taken in the historic Norwegian commercial hunt for this species (for details, see Dalebout et al. 2001). Two out-group species were used routinely in amplification and sequencing: the New Zealand sea lion, *Phocarctos hookeri*, and the American bison, *Bison bison*. Total cellular DNA was isolated from tissue by digestion with proteinase K, followed by standard phenol/chloroform extraction, as described previously (Baker et al. 1994). Records of provenance for all tissue and DNA samples are held in an archival database system maintained by the first author (CSB).

For noncetacean out-groups (e.g., ruminants, carnivores, and primates), published sequences from presumably orthologous loci were derived from GenBank or EMBL, often after performing BLAST searches (e.g., Altschul et al. 1990) using experimental cetacean sequences. GenBank and EMBL sequences are generally referred to by the common name, the accepted MHC nomenclature, and the accession number, except where alleles have been numbered (e.g., for humans, the cow in some cases, the beluga, and the narwhal).

### Amplification

Fragments of the *DQB* and *DRB* genes were amplified with polymerase chain reaction (PCR) under standard condi-

**Table 1** Species source and number of sequences (length in base pairs, excluding primers) collected from cetacean MHC Class II genes (*DQB* and *DRB* exon 2) in this study

| Species | *DQB* (172 bp) Number of sequences, *n* | *DRB* (217–219 bp) Number of sequences, *n* |
|---|---|---|
| Suborder: Mysticeti (baleen whales) | | |
| Family Balaenidae | | |
| Euau, *Eubalaena australis*, southern right whale | 5 | 2 |
| Bamy, *Balaena mysticetus*, bowhead whale | 3 | 3 |
| Family Balaenopteridae | | |
| Bamu, *Balaenoptera musculus*, blue whale | – | 3 |
| Baph, *Balaenoptera physalis*, fin whale | 2 | 3 |
| Esro, *Eschrichtius robustus*, gray whale | 2 | 1 |
| Meno, *Megaptera novaeangliae*, humpback whale | 23 | 7 |
| Suborder: Odontoceti (toothed whales) | | |
| Family Ziphiidae (beaked whales) | | |
| Hyam, *Hyperoodon ampullatus*, northern bottlenose | 1 | 1 |
| Meca, *Mesoplodon carlhubbsi*, Hubbs' beaked | – | 1 |
| Meeu, *Mesoplodon europaeus*, Gervais' beaked | 2 | – |
| Megr, *Mesoplodon grayi*, Gray's beaked | 1 | – |
| Mest, *Mesoplodon stejnegeri*, Stejneger's beaked | 1 | – |
| Family Physeteridae, | | |
| Phma, *Physeter macrocephalus*, sperm whale | | 1 |
| Family Delphinidae | | |
| Cehe, *Cephalorhynchus hectori*, Hector's dolphin | 2 | 1 |
| Out-groups | | |
| Bibi, *Bison bison*, American bison | – | 1 |
| Phho, *Phocarctos hookeri*, New Zealand sea lion | – | 1 |
| Total | 42 | 25 |
| GenBank no. | DQ354623-DQ354664 | DQ354665-DQ354689 |

Cetacean nomenclature follows Rice (1998) in most cases except where noted in text. Four-letter codes for each species follow those in all other figures and tables

tions: 1× PCR buffer II, 2.5 mM $MgCl_2$, 0.5 U Ampli*Taq* or *Taq*Gold (Applied Biosystems), 0.24 mM dNTPs (Amersham Pharmacia Biotech), 0.2 μM primers, and approximately 25–100 ng of DNA. Most amplifications were conducted with a PE Applied Biosystems (ABI) model 9700 thermocycler. Thermocycle profiles differed somewhat for each gene as described below.

*DQB* A 172-bp fragment (excluding primers) from exon 2 of the *DQB* gene was amplified with the universal primers *DQB*2 (5′-CATGTGCTACTTCACCAACGG-3′) and *DQB*1 (5′-CTGGTAGTTGTGTCTGCACAC-3′) as reported by Murray et al. (1995). Amplification of this locus used *Taq*Gold and the following "touch-down" thermocycling profile: initial denaturation/activation at 94°C for 10 min; 10 cycles of 94°C denaturation for 30 s, 64°C annealing for 20 s, and 72°C extension for 30 s, with the annealing temperature decreasing by 1°C on each cycle; 25 cycles of 94°C for 40 s, 54°C for 30 s, and 72°C for 30 s. A final 10-min soak at 72°C was used to allow for complete extension of the PCR fragments.

*DRB* Initial efforts to amplify exon 2 of the *DRB* gene using published primers designed from the intron/exon borders of the cow (Ammer et al. 1992; Groenen et al. 1990) and used by Murray and White (1998) for the beluga and narwhal were not successful. These included (with slight modifications and corrections), forward primer DRB219 (5′-CCACACAGCACGTTTCYTG-3′) and reverse primers DRB61a (5′-CCGCTGCACTGTGA AGCT-3′) and DRB3b (5′-CTCGCCGCTGCACAGTGA ARC-3′). We note these primers overlap with LA31 (for DRB219) and LA32 (for DRB61a and DRB3b) described by Sigurdardottir et al. (1991) and used by Mikko et al. (1999) on a number of ruminant species. Although a PCR product of the correct size was generated (approximately 280 bp, including primers), this fragment did not provide a readable sequence. A new forward primer, DRBex2f2b, (5′-CRGTTTAAGAGCGAGTGTC-3′) was designed based on the aligned sequence of the beluga and the cow and nested just within DRB219. The thermocycling profile had an initial denaturation period of 94°C for 3 min followed by 35 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 60 s. Under these conditions, the combination of primers DRBex2f2b and DRB61a provided strong amplification of a fragment of 217–219 bp (excluding primers) and clean sequence from both amplified and cloned fragments.

## Sequencing and cloning

Purified PCR products were either sequenced directly or, for some loci, cloned into P-GEM-T Easy vectors (Promega) following manufacturer's recommendations and transformed into JM109 competent cells (Invitrogen). Cells (100 μl) were plated on LB Agar plates containing ampicillin (100 μg/ml), isopropylthiogalactoside (0.5 mM), and X-Gal (100 μg/ml). After incubation at 37°C overnight, white colonies were picked and the insert

was amplified using the M13 forward (5′ GTAAAACG ACGGCCAG 3′) and M13 reverse (5′ CAGGAAACAGC TATGAC 3′) primers (as provided by Promega). PCR reaction conditions were as described above. The cycling profile consisted of 94°C for 3 min, followed by 35 cycles of 30 s at 94°C, 50°C, and 72°C each. A final extension period of 72°C for 10 min was included at the end of the reaction. Amplified fragments of the correct size were sequenced on an ABI 377 automated sequencer or an ABI 3100 automated sequencer using BigDye Terminator Chemistry (ABI). Sequences were aligned and edited using Sequencher (v4.1, Gene Codes Corporation). Sequences differing by a single nucleotide from common allele sequences were assumed to be due to Ampli*Taq* error (e.g., Tindall 1988; Barnes 1992) unless the variable site was shared with other alleles or found in more than one clone.

## Automated single-strand conformation polymorphism

Automated single-strand conformation polymorphism (SSCP) analysis followed the methods of (Lento et al. 2003). Briefly, forward and reverse primers were labeled with the fluorescent dyes 6′-FAM and TET, respectively, and were amplified following standard conditions. Samples were prepared for SSCP following the guidelines of Atha et al. (1998) and the ABI 310 manual: a 2 µl aliquot of a 1:5 dilution of the *DQB* amplicon in ddH2O was mixed with 10.5 µl deionized formamide (Gibco), 0.5 µl of freshly made 0.3 N NaOH, and 1.0 µl GeneScan-500 TAMRA dye standard (ABI Prism). The samples were denatured at 96°C for 2–4 min and then snap-cooled on wet ice slurry for 2–4 min before electrophoresis on an ABI 310 automated capillary sequencer running 3% GeneScan (ABI Prism) polymer with 10% glycerol in 1× TBE at a constant temperature of 29°C. SSCP peak profiles were analyzed using ABI 310 data collection and Genotyper software (ABI, v2.1). Analyses presented here refer only to allelic separation of amplified products as represented by one or more peaks, rather than allele-specific genotyping.

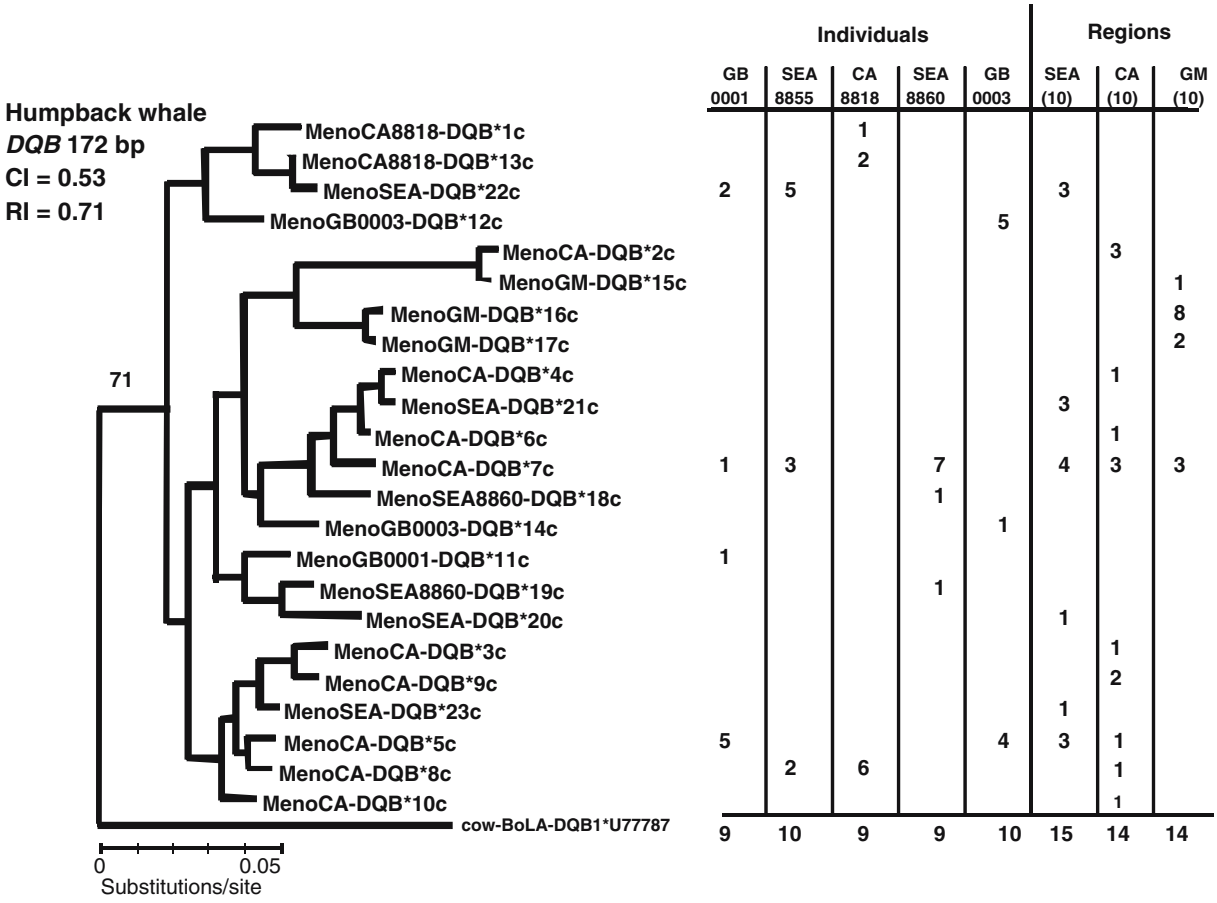


**Fig. 1** Neighbor-joining reconstruction of cloned humpback whale *DQB* exon 2 sequences with frequencies of clones recovered from five individual whales and three pooled samples representing southeastern Alaska (SEA, including Glacier Bay, GB), California (CA), and the Gulf of Maine (GM). *Number* above internal node uniting humpback whale sequences indicates bootstrap value >50% from 1,000 NJ resamplings of these data. *CI* Consistency index, *RI* retention index. Distances are adjusted using best-fit model (F81+I=0.2929+G=0.4101) indicated by MODELTEST vs 3.6

Statistical and phylogenetic analysis

Sequences were aligned initially with the ruminant *DRB*3 sequence from Mikko et al. (1999) and the beluga *DRB* and *DQB* sequences from Murray et al. (1995). We followed Mikko et al. (1999) in the identification of codons of the PBR of ruminants, noting that these authors do not consider positions 26 or 58 indicated by Brown et al. (1993), and included by Murray et al. (1995). Inclusion of these additional positions did not affect the significance of statistical tests presented here. Nucleotide diversity was calculated using the program MEGA2 (Kumar et al. 2001). Synonymous and nonsynonymous divergence ($d_S$ and $d_N$) and standard errors were calculated using the Jukes and Cantor correction of Nei and Gojobori (1986). The ratio of nonsynonymous to synonymous divergence ($d_N/d_S$) was tested for departure from the neutral expectation of unity using the *z* statistic available in MEGA2. Other differences in proportions of substitution or divergence were tested with a *t* test (Nei 1987). Amino acid sequences were inferred using the translation function in the program, MacClade (v4.0, Maddison and Maddison 2000).

Phylogenetic relationships among *DRB* and *DQB* sequences were reconstructed using maximum parsimony or neighbor-joining (NJ; Saitou and Nei 1987) methods, as available in the program PAUP* (v4.0b0, Swofford 2000). For NJ reconstructions, genetic distances were adjusted using the best-fit model of substitution indicated by the hierarchical likelihood ratio test implemented in the program MODELTEST vs3.6 (Posada and Crandall 1998), and branch support

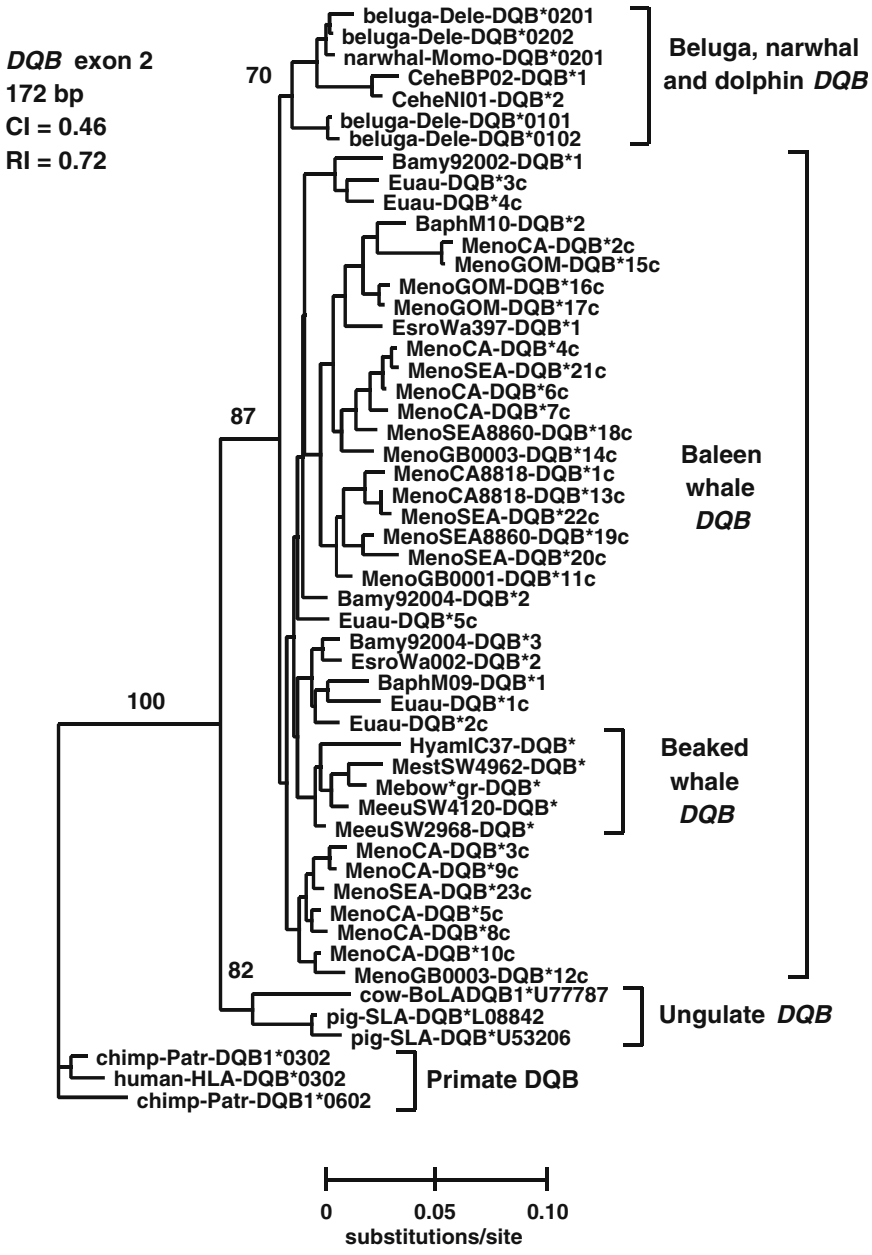


**Fig. 2** Neighbor-joining reconstruction of the cetacean *DQB* exon 2 sequences and outgroups. *Numbers* adjacent to some internal nodes indicate bootstrap values >50% from 1,000 resamplings of these data. *CI* Consistency index, *RI* retention index. Distances are adjusted using best-fit model (F81+I=0.2929+G=0.4101) indicated by MODELTEST vs 3.6

was assessed with 1,000 bootstrap replications. Topologies and branch support of phylogenies with best-fit models were also examined for consistency with Jukes and Cantor and Kimura 2-parameter models. The large number of sequences prevented use of maximum likelihood or exhaustive searches under maximum parsimony, but NJ trees were evaluated under parsimony criteria, including calculation of consistency and retention indices (CI and RI in Figs. 1, 2, 3, and 4) with the program MacClade.

## Allele nomenclature

We followed the standard four-letter abbreviations for species origins recommended for MHC nomenclature (Bodmer et al. 1997). For reasons described in the results, we did not attempt to assign codes to different groups of alleles at any of the three loci. Instead, we retained the individual-specific code used in the tissue sample and DNA archival database. Following the four-letter species prefix, this code usually specified the geographic location of the collection, the year of collection, and an accession number. Where alleles were derived from a pooled population sample, we assigned species code followed by the geographic location of the population (e.g., SEA for southeastern Alaska) and simple accession number following the alphabetic ordering of the sample of clone. All sequences derived from cloning were indicated by the letter "c" at the end of the code.

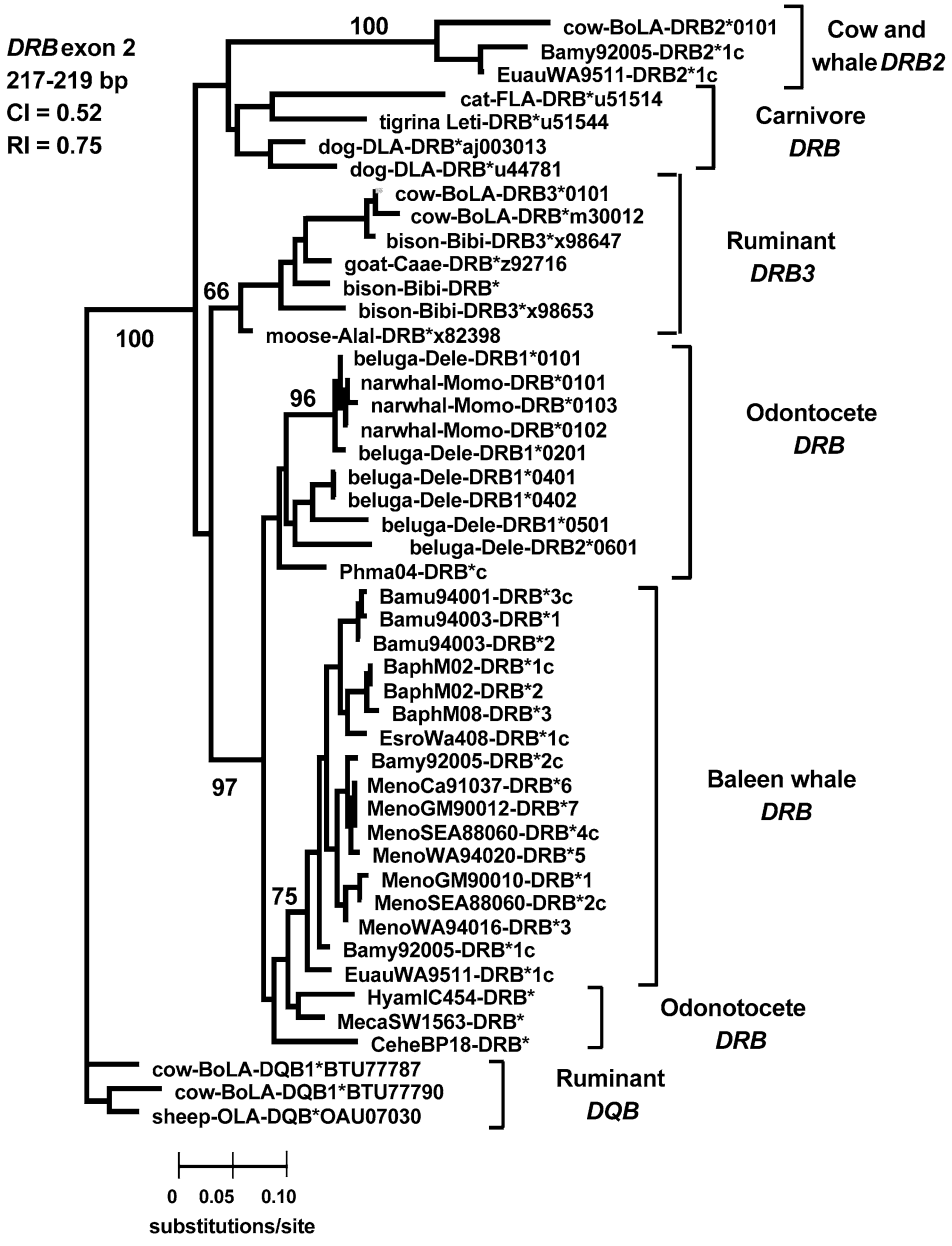


**Fig. 3** Neighbor-joining reconstruction of the cetacean *DRB*-like exon 2 sequences and out-groups, including the cow BoLA-*DRB*2 and ruminant *DBQ*. *Numbers* adjacent to some internal nodes indicate bootstrap values >50% from 1,000 NJ resamplings of these data. *CI* Consistency index, *RI* retention index. Distances are adjusted using best-fit model (F81+G=0.6896) indicated by MODELTEST vs 3.6

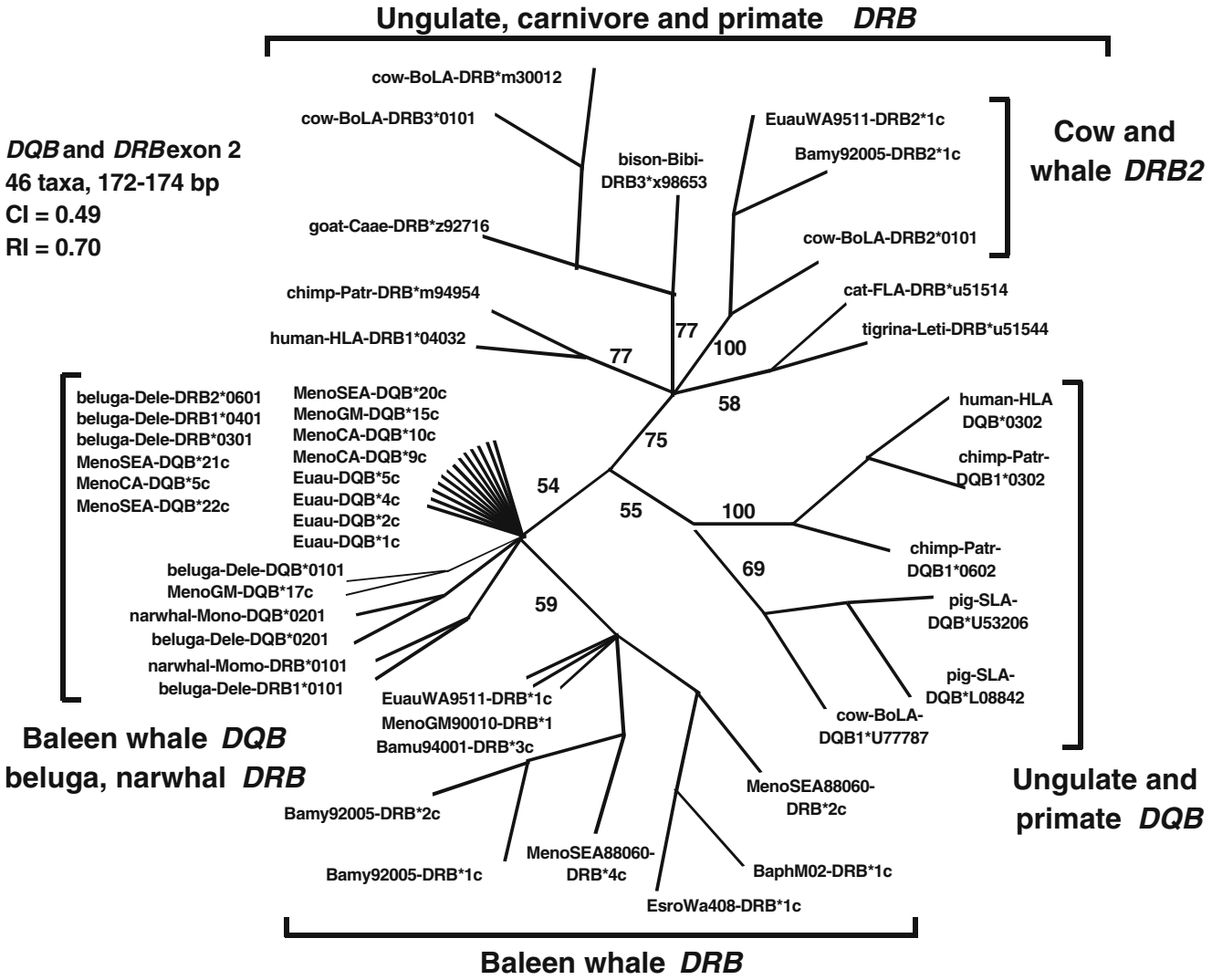


**Fig. 4** Unrooted neighbor-joining reconstruction of the cetacean *DQB* and *DRB*-like exon 2 sequences and out-groups. *Numbers* adjacent to internal nodes indicate bootstrap values >50% for relevant internal branches from 1,000 resamplings of these data. *CI* consistency index, *RI* retention index. Distances are adjusted using best-fit model (F81+G=0.6896) indicated by MODELTEST vs 3.6

## Results

### *DQB* characterization

Variation in exon 2 of a *DQB* gene in baleen whales, beaked whales, and dolphins was surveyed by direct sequencing of genomic PCR products, cloning of PCR products, and fluorescent SSCP (Table 2). The universal primers *DQB*1 and *DQB*2 provided strong amplifications of fragments of the expected size (172 bp, excluding primers) and reliable direct sequencing for all species screened. A BLAST search of GenBank showed greatest similarity of these sequences to the published *DQB* exon 2 sequence of the beluga and narwhal, and various ungulates, including the cow (for exception, see below). Aligned sequences showed no insertion/deletion events (indels) and were in reading frame with other mammalian *DQB* sequences extending from coding position 21 to 77 (Table 2). Examination of sequencing electropherograms revealed numerous sites with secondary peaks, suggesting that many individuals were heterozygotes. Fluorescent SSCP with standard conditions produced complex, variable profiles that did not show a simple correspondence to alleles inferred from direct sequencing.

To evaluate potential amplification of duplicate loci, PCR products from five humpback whales and one right whale were cloned and sequenced. For the five humpbacks, sequencing of 9–10 clones from each individual resulted in a total of 11 unique sequences (Fig. 1). A duplication of the *DQB* gene in the humpback whale was indicated by the number of alleles resolved for each individual; three unique sequences were obtained from each of the four whales, and four unique sequences were obtained from the fifth whale. For the right whale, five unique sequences were found among the eight clones, suggesting three copies of the *DQB* gene. All cloned sequences were in reading frame with published mammalian *DQB* alleles and showed no unusual stop or start codons.

**Table 2** Inferred reading frame of *DQB* exon 2 sequences from baleen whales aligned to published beluga *DQB*, cow *DQB1*, and human *DQB* (underlined), with comparison to selected whale and cow *DRB3* alleles (italics)

```
[                   21       30        40        50        60        70       ]
[                          * *      **                  *   *     *  **  *   ]
MenoCA8818-DQB*1c     TERVRLVVRHIYNREEFARFDSDVGEYRAVTELGRPSAKYWNSQKDLLEQTRAELDT
MenoCA-DQB*2c        .......E.D.......L..................I.EN..........R..AV..
MenoCA-DQB*3c        .......T.Y....................S.....D.E.......I..E...AV..
MenoCA-DQB*4c        .....A.E.Y......Y.............S.....D.E..............AV..
MenoCA-DQB*5c        .......T.Y......Y.............S.....D.E.......I..........
MenoCA-DQB*6c        .....A.E.Y....................S.....D.E..............AV..
MenoCA-DQB*7c        .....A.E.Y......Y.............S.....D.E.......I..E...AV..
MenoCA-DQB*8c        .......T.Y....................S.....D.E..................
MenoCA-DQB*9c        .......T.Y......Y.............S.....D.E.......I..E...AV..
MenoCA-DQB*10c       .......T.Y..........................D.E..................
MenoGB0001-DQB*11c   ................Y.............S.....D.E.......I..........
MenoGB0003-DQB*12c   .......T.Y.............................................
MenoCA8818-DQB*13c   ................Y...................E...........R...V..
MenoGB0003-DQB*14c   .......E.D......Y.............S.....D.E.......I..........
MenoGM-DQB*15c       .......E.D.......L..................I.EN..........R...V..
MenoGM-DQB*16c       ......LE.Y....................S.....D.E...........R..AV..
MenoGM-DQB*17c       ......LE.Y....................S.....D.E...........R...V..
MenoSEA8860-DQB*18c  .....A.E.Y......Y.............S.....D.E.......I..........
MenoSEA8860-DQB*19c  ................Y.............S.....D.E.......I..E...AV..
MenoSEA-DQB*20c      ................Y...................E.......I..E...AV..
MenoSEA-DQB*21c      .....A.E.Y....................S.....D.E..............AV..
MenoSEA-DQB*22c      ................Y...............................R...V..
MenoSEA-DQB*23c      .......T.Y....................S.....D.E.......I..E.......
Bamy92002-DQB*1      .....A.T.Y......Y.............S.......E.......I...E..AV..
Bamy92004-DQB*2      .......SSY......Y.............S.....D.E.......I..........
Bamy92004-DQB*3      .....Y.SSY......Y.............S.....D.E.......I..........
BaphM09-DQB*1        .....Y.T.Y................F...S.....D.............R...V..
BaphM10-DQB*2        .......E.Y....................S.....V.EK..........R...V..
EsroWa397-DQB*1      .......E.Y......Y...................D............KR......
EsroWa002-DQB*2      .....Y.S.Y......Y...................D.........I..........
Euau-DQB*1c          .....Y.T........YV...........................H...R......
Euau-DQB*2c          .....Y.T........YV............S.....D.E.......I..........
Euau-DQB*3c          .......T.Y......LV............S.....D.E.......I..RE..AV..
Euau-DQB*4c          .......SSY......Y.............S.....D.E.......I..RE..AV..
Euau-DQB*5c          .......T.Y......LV............S.....D.E.......I..........
beluga-Dele-DQB*0101 .......S.Y......LVH................D.E.......I..R.......
cow-BoLA-DQB1*U77787 .....Y.T.Y...Q..YV.............P....P.EH......F.......V.S
human-HLA-DQB*0302   .......T.Y......Y........V.....P..P.A.E......EV..R.......
BaphM02-DRB*1c       ................Y.............S.....D............R...V.R
cow-BoLA-DRB3*0101   .....FLD.YYT.G..TV.....W..F........QD.E.......F..EK...V.R
```

The cetacean sequences starts at position 21 relative to the human and beluga *DQB* (Murray et al. 1995). Positions thought to be involved in Peptide Binding (*) follow Brown et al. (1993) as shown by Mikko et al. (1999). Species abbreviations follow Table 1. Alleles from humpback whale (Meno) originated from three populations: southeastern Alaska (SEA, including Glacier Bay, GB), California (CA), and Gulf of Maine (GM)

To survey *DQB* diversity across humpback whale populations, PCR products from a sample of ten individuals were pooled and cloned for each of three regions (southeast Alaska and California in the North Pacific and the Gulf of Maine in the North Atlantic). Sequencing of 43 clones ($n$=14–15 from each of the three pooled regional samples) resulted in 16 unique alleles, four of which matched the 11 cloned alleles obtained from cloning of the five individual whales. Phylogenetic reconstruction of the 23 alleles recovered from the humpback whales (Fig. 1) showed no strong bootstrap support for groups that might represent individual loci. Maximum and average nucleotide diversity among these alleles (11.3%; average=5.9%, SE=1.2%) was nearly twice that observed among beluga or narwhal *DQB* alleles (6.1%; average=3.7, SE=1.0%). Other than one pair of alleles that differed at only a single synonymous site, all alleles differed from each other at one or more codons. Maximum divergence across the inferred protein sequence was 21% (i.e., 79% amino acid identity).

Over the entire sequence, nonsynonymous substitutions exceeded those of synonymous substitutions ($d_N/d_S$=1.48), although this was not significantly different from unity ($p$=0.30). However, among the ten codons of the PBR, the greater divergence of nonsynonymous substitutions was highly significant ($d_N/d_S$=7.45; $p$<0.001). For the five right whale *DQB* alleles, the divergence of nonsynonymous substitutions was significantly greater than synonymous over the entire exon ($d_N/d_S$=8.3; $p$<0.0005). This effect was more pronounced at the codons of the PBR, where nonsynonymous divergence was 40% and there were no synonymous substitutions (Table 4; $p$<0.0001).

To investigate phylogenetic relationships of the cetacean *DQB*, a further 14 alleles were inferred from direct sequencing of PCR products from other baleen whales, beaked whales, and dolphins. Adding these to the 28

cloned alleles obtained from humpback and right whales gave a total of 42 *DQB* alleles. Phylogenetic reconstructions of the cetacean *DQB* sequences (including five published beluga and narwhal alleles) and out-groups showed strong bootstrap support (87%) for a cetacean clade distinct from the ungulates and primates, but little or no support for clades uniting members of the same family or species (Fig. 2). Results were similar with inclusion of additional cetacean *DQB* sequences from Hayashi et al. (2003) and Yang et al. (2005). Given the extensive transspecific sharing of similar alleles, we did not attempt to classify and number *DQB* alleles according to species-specific phylogenetic groupings, as is the convention (Bodmer et al. 1997).

## *DRB* characterization

Nucleotide variation of the putative second exon of a *DRB*-like gene of baleen whales was surveyed by direct sequencing of genomic PCR products, cloning of PCR products, and fluorescent SSCP. The survey included genomic PCR sequences for eight individuals from each of six species of baleen whales, two species of beaked whales, a Hector's dolphin, sperm whale, and an out-group, the American bison (Table 1). The combination of primers, DRBex2f2b and DRB3b, provided strong amplifications and reliable direct sequencing of 217–219 bp for most balaenopterid whales and all odontocete species surveyed (see Supplemental Information). A BLAST search of GenBank showed greatest similarity of the balaenopterid sequences to the published *DRB* exon 2 sequence of the beluga and narwhal (85–91% identity), and various ungulates, including the cow *DRB*3 gene (78–80% identity to *DRB*3*0101; Groenen et al. 1990).

Alignment of genomic PCR sequences of several balaenopterid whales revealed interruptions due to several indels. For the humpback and fin whales, indels were polymorphic, and heterozygous individuals were recognized by sudden appearance of double peaks at specific positions in electropherograms. For the gray whale, a 1-bp indel was fixed in the eight individuals surveyed. SSCP profiles (data not shown) of the amplified *DRB*-like exon from balaenopterid whales were consistent with a Mendelian pattern of inheritance for a single locus, with a small number of alleles for some species (see below). For balaenid whales, however, complex SSCP profiles suggested amplification of more than one heterozygous locus. No indels were observed in the odontocete species surveyed here or the published study of the beluga and narwhal (Murray and White 1998).

To investigate the potential for multiple *DRB* loci and to confirm sequences of the indel alleles, amplified products from one individual of each of the six baleen whales were cloned. For the humpback, fin, right, and bowhead whales, individuals selected for cloning were first identified as likely heterozygotes for indel alleles based on SSCP profiles and genomic PCR sequences. Eight positive

**Table 3** Inferred amino acid translation of the *DRB* exon 2 nucleotide sequences from baleen whales based on "standard" alignment to published beluga *DRB* and cow *DRB3* and *DRB2* (underlined), with comparison to selected whale and cow *DQB* alleles (italics)

```
[                    9         +20        30        ++40        50        60 ^      70        80       ]
[                    * * *              * *      **                   *   *     *  **  *   *   *   *
MenoGM90010-DRB*1#            FS NETERVRLVVRHIYNREEYA  RFDSDVGEYRAVSELGRPDAEYWNSQKDLLEQTRAAVDTYCRHNYGAYE
MenoSEA88060-DRB*2c           .. ....................  .......................^........R..E............
MenoWA94016-DRB*3             .. ....................  .......................^........R..E...V........
MenoSEA88060-DRB*4c           ..+......Q.............  ............T.....S....^........R......V........
MenoWA94020-DRB*5             ..+......Q.............  ...................S.............R......V........
MenoCa91037-DRB*6             ..+......Q.............  ............T.....S.............R......V........
MenoGM90012-DRB*7             ..+......Q.............  ............T.....S.............R......V........
Bamu94001-DRB*3c#             .. ..................F.  ............................I...R..E...V.....WVG.
Bamu94003-DRB*1#              .. ..................F.  ............................I...R..E...V.....WVG.
Bamu94003-DRB*2#              .. ..................F.  ............................I...R..E...V......VG.
BaphM02-DRB*1c#               .. ....................  ....................K...........R..E..R.......VG.
BaphM02-DRB*2                 .. ....................  ....................K..^........R..E..R.......VG.
BaphM08-DRB*3#                .. ..................F.  ....................K...........R..E..R.......VV.
EsroWa408-DRB*1c              .. ....................  ....................R..^....I...R..E...V......VG.
Bamy92005-DRB*1c              .. ......Q.............++............T.....S.K.......H...R..EL..V......VV.
Bamy92005-DRB*2c              .. ......Q.............++............T.....S.K.......H...R..EL..V........
EuauWA9511-DRB*1c             .. .V....G.............++............T.....I.KH......H...R......V......VV.
beluga-Dele-DRB1*0101 YQFKGECR.. .G......T.....G..FM  .Y......C...T....RT..SL.....F..RR..E...V......VV.
cow-BoLA-DRB3*0101    EYS.S..H.F .G.....FLD.YYT.G..TV  .....W..F...T....Q..........F..EK..E..RV......GM.
Bamy92005-DRB2*1c             .. .GM.Q..FLA.YT..AHRRM  ........FT..T........S..Q...FM..M..K...V..SS.WGIG
EuauWA9511-DRB2*1c#           .. .GM.Q..FLA.YT..AQ.D.  H.......FT..T........S..Q...FM..M..K...V..SS.WGIG
cow-BoLA-DRB2*0101    H......... .GL..M.FFA.Y...TQ.DV  H.......FT.LT....L......Q...FM..M..K...L..P..QGIG
MenoSEA-DQB22*c                    ..................  ............T.....S.K...........R..E...
beluga-Dele-DQB*0101               .......S.Y......LV  H...........T...............I..R...EL..
cow-BoLA-DQB1*U77787  ....CLCY.T .G.....Y.T.Y...Q...V  ............TP....P..H......F......E..SV.....QLEL
```

Translations inferred from nucleotide sequences in reading frame with the cow are marked with a number sign (#). Pseudotranslations of indel sequences are shown for comparative purposes. The standard reading frame was maintained by ignoring codons with single (+) and double (++) nucleotide insertions or deletions (ˆ) relative to the cow. The baleen whale sequences starts at position 17 relative to the human *DRB* exon 2 and the cow, BoLA-DRB*0101. Positions thought to be involved in Peptide Binding (*) follow Brown et al. (1993), as shown by Mikko et al. (1999). Species abbreviations follow Table 1

colonies were isolated and sequenced for each individual. Among the balaenopterid whales, results were consistent with the amplification of only a single locus, similar to the cow *DRB*3. No more than two alleles were detected in the eight clones derived from any one individual. Clonal sequences confirmed the indel alleles suggested by SSCP profiles and were consistent with individuals identified as homozygotes or heterozygotes.

The combined cloning and genomic PCR sequencing from the survey of eight individuals for each species detected a total of three alleles for the blue whale, a single allele for the gray whale, three alleles for the fin whale, and seven alleles for the humpback (see Supplemental Information). The three blue whale alleles were in reading frame with the cow *DRB*3, extending from inferred amino acid position 17 to 87 (Table 3) and differed from each other at two synonymous and one nonsynonymous positions (1.0–1.4% divergence). The single gray whale allele had a deletion at position 165 relative to the cow *DRB*3. Two of the three fin whale alleles were in reading frame with the cow *DRB*3, differing from each other by a single substitution. The third fin whale allele shared with the gray whale a deletion at nucleotide position 165. One humpback allele (Meno-*DRB**1) was in reading frame with the cow *DRB*3 sequence. The other six were defined by an insertion at position 33, a deletion at position 165 or both, as well as nucleotide substitutions (0.0–3.8%, excluding indels). Interspecific differences among the 14 alleles found among these three balaenopterid species ranged from 3.3–6.4% (excluding indels). Attempts to infer translation of amino acids for the indel alleles confirmed the interruption of the reading frame relative to other mammals (Table 3).

For the two individuals representing balaenid species (right and bowhead whales), clonal sequences indicated two *DRB*-like loci (Table 3). For the right whale EuauWA9511, two divergent sequences were found in nine clones: DRB*1c was most similar to the cow BoLA-*DRB*3*0101 (and the *DRB*3-like sequences of the balaenopterid whales) except for a GC insertion at position 95–96; DRB2*1c showed greatest similarity to a cow *DRB*2 gene and was in reading frame with this locus (e.g., GenBank BTU77068; Muggli-Cockett and Stone 1989). For bowhead Bamy92005, two similar and one divergent sequences were found in nine clones: DRB*1c and DRB*2c were most similar to the right whale *DRB3*-like sequence, sharing the GC insertion at position 95–96; DRB2*1c was most similar to the cow *DRB*2 and right whale *DRB2*-like sequence, but had an AC insertion at position #95–96. The bowhead and right *DRB2*-like sequences differed from each other by 4.3% (excluding indels), from the cow *DRB*2 by 12–16%, and from the other whale *DRB3*-like alleles by 29–34%. The right whale *DRB2*-like sequence (EuauWA9511-DRB*1c) was in reading frame, but like the cow *DRB2*, the inferred amino acid sequence included intervening "start" codons.

In total, six of the 14 alleles characterized from the balaenopterids (three blue, two fin, and one humpback whale) were in reading frame with the cow *DRB3* and other mammalian *DRB* sequences. The 71 amino acid residues inferred from the nucleotide sequences differed at 27 positions (32 changes) in comparison to the cow and 22 positions (24 changes) in comparison to the beluga (Table 3). A large number of these substitutions were found among codons of the PBR: 14 substitutions relative to the cow and 13 relative to the beluga. Among the six alleles, there were 11 variable amino acid positions of which seven were found at positions within the PBR. Unlike the *DQB* exon 2, however, the ratio of nonsynonymous to synonymous substitutions among the six alleles of the balaenopteriid whales was not significant for any comparison (Table 4). Instead, divergence at both nonsynonymous and synonymous positions was highly localized in the PBR ($d_N$=23% and $d_S$=21%) compared to the non-PBR ($d_N$=2.1% and $d_S$=1.6%). This tenfold difference in divergence at PBR and non-PBR was highly significant ($t$=2.74, $p$<0.01).

Phylogenetic reconstructions of the baleen and toothed whale *DRB3*-like sequences showed strong bootstrap support for a cetacean-specific clade (98%), distinct from the ungulates and carnivores (Fig. 3) and exclusive of ungulate *DQB* sequences used here as an out-group (although see "DRB and DQB phylogeny"). Among the cetaceans, there was moderate bootstrap support (75%) for a baleen whale clade, but little support for an odontocete clade (i.e., bootstrap scores less than 50%). Among the baleen whales, species-specific clades were well-supported for alleles of the blue and fin whale, but not for the humpback, right, or bowhead whales. The bowhead and right whale *DRB*2-like sequences grouped with each other and the cow *DRB*2 (100% bootstrap), exclusive of all other *DRB* sequences and the ungulate *DQB* out-group sequences.

## *DRB* and *DQB* phylogeny

BLAST searches of GenBank using some (but not all) baleen whale *DRB*-like sequences showed greater similarity to shorter-length sequences of other cetacean *DQB* sequences (e.g., beluga and narwhal) than to the putative *DRB* orthologues of other species. To investigate the potential for convergence or gene conversion between these loci, we aligned all cetacean *DQB* and *DRB*-like sequences with the out-group sequences used in the analyses described above (Tables 2 and 3). Codon position 21 of the cetacean *DQB* aligned unambiguously with codon 21 of the cow *DRB*3, and required no gaps to be inserted except for those already included in the cetacean *DRB* multiple sequence alignment. The aligned sequences were trimmed to the *DQB* consensus length of 174 bp (172 bp plus 2 bp insertion for right and bowhead *DRB* sequences). To facilitate analysis, the full data set of cetacean *DRB* and *DQB* sequences was reduced to 31 sequences representing the range of diversity among cetacean class II genes. The final data set, including out-groups, consisted of 46 sequences: nine baleen whale *DRB*3-like alleles, two baleen whale *DRB*2-like alleles,

**Table 4** Nucleotide divergence (and standard errors; Kumar et al. 2001) among baleen whales and in relationship to out-groups for all sites ($d$) and for nonsynonymous ($d_N$) and synonymous ($d_S$) differences at the Peptide Binding Region (PBR) and nonbinding sites (*nc* Not calculated due to uncertain reading frame)

| Species | Number of alleles (length, in bp) | $d$ SE | Overall | | PBR | | non-PBR | |
|---|---|---|---|---|---|---|---|---|
| | | | $d_N$ SE | $d_S$ SE | $d_N$ SE | $d_S$ SE | $d_N$ SE | $d_S$ SE |
| *DQB* exon 2 | | | | | | | | |
| Humpback whale | 23 (172) | 0.059 | 0.065 | 0.046 | 0.304* | 0.043 | 0.022 | 0.047 |
| | | 0.012 | 0.019 | 0.020 | 0.099 | 0.052 | 0.010 | 0.022 |
| Right whale | 5 (172) | 0.064 | 0.083* | 0.010 | 0.396* | 0.000 | 0.031 | 0.012 |
| | | 0.014 | 0.023 | 0.011 | 0.133 | 0.000 | 0.017 | 0.012 |
| Humpback whale to right whale | 23, 5 | 0.074 | 0.087 | 0.043 | 0.461 | 0.031 | 0.027 | 0.043 |
| | | 0.015 | 0.024 | 0.020 | 0.150 | 0.036 | 0.014 | 0.024 |
| Baleen whales (average) to cow | 28, 2 | 0.148 | 0.138 | 0.153 | 0.404 | 0.203 | 0.091 | 0.144 |
| | | 0.027 | 0.039 | 0.057 | 0.161 | 0.218 | 0.038 | 0.059 |
| *DRB*3 exon2 | | | | | | | | |
| Balaenopteriids (three species) | 6 (213) | 0.051 | 0.055 | 0.044 | 0.230 | 0.210 | 0.021 | 0.016 |
| | | 0.012 | 0.018 | 0.024 | 0.117 | 0.165 | 0.001 | 0.016 |
| Humpback whale to right whale | 1, 1 | 0.083 | nc | nc | nc | nc | nc | nc |
| | | 0.014 | | | | | | |
| Balaenopteriids (average) to cow *DRB*3 | 3, 1 | 0.242 | 0.227 | 0.291 | 0.596 | 0.382 | 0.162 | 0.279 |
| | | 0.037 | 0.051 | 0.098 | 0.280 | 0.318 | 0.043 | 0.107 |
| | | 0.037 | 0.051 | 0.098 | 0.280 | 0.318 | 0.043 | 0.107 |
| Right whale to cow *DRB2* | 1, 1 | 0.124 | nc | nc | nc | nc | nc | nc |
| | | 0.024 | | | | | | |
| | | 0.024 | | | | | | |

*$p<0.05$ denotes values of $d_N$ that were significantly greater than $d_S$, indicating overdominant selection
For the Balaenopteriid *DRB*3 exon, the divergence of synonymous and nonsynonymous substitutions combined is significantly greater for the PBR than for the non-PBR (PBR, $d$=0.222; SE=0.0735; non-PBR, $d$=0.019; SE=0.0084; $p<0.01$)

eight humpback and four right whale *DQB* alleles, eight of the published alleles from the beluga and narwhal, and 15 alleles representing ungulate, primate and carnivore out-groups, including the cow *DRB*2.

The phylogenetic analysis confirmed the similarity of cetacean class II sequences (Fig. 4) across this 174-bp fragment (except the *DRB*2 sequences, see below). In contrast to the phylogenetic analysis of the longer *DRB* sequences shown in Fig 3, the shorter cetacean *DQB* and *DRB*3 sequences grouped together with weak bootstrap support (54%) and appeared most closely related to the primate and ungulate *DQB*. These clades grouped together to the exclusion of primate, ungulate, and carnivore *DRB* sequences, with moderately strong bootstrap support (75%). Among the cetaceans, there was weak bootstrap support (59%) for a group comprising the whale *DRB*3-like sequences, but this was nested within the cetacean *DQB* alleles and did not include the beluga and narwhal *DRB* sequences. Only the putative orthology of the bowhead and southern right whale sequences (Bamy92005-DRB2*1c and EuauWA9511-DRB2*1c) with the cow *DRB*2 was supported by the phylogenetic reconstruction of the reduced length sequences (see Figs 3 and 4).

As suggested by the phylogenetic analysis of nucleotide sequences, the greatest similarity was between the cetacean Class II loci, rather than to either the *DQB* or *DRB* of other mammals. This was also apparent in the inferred amino acid sequences. The two most similar baleen whale *DQB* and *DRB* sequences (fin whale BaphM02-*DRB**1c and humpback whale MenoSEA-*DQB**22c) differed at 17–18 positions (30–31%) from the cow *DBR*3 and 12–13 positions (21–23%) from the cow *DQB*2, but at only three positions (5.3%) from each other. At the nucleotide level, these whales differed by 21% from the cow *DBR*3 and 14–17% from the cow *DQB*2, but only 5.5% from each other. This pattern was even more striking at the ten codons of the PBR, where the whales differed at seven to eight residues from the cow *DBR*3 and at six from the cow *DQB*2, but at only one from each other. Of the nine codons of the PBR shared between the cetacean *DQB* and *DRB*, seven differed from one or both of these genes in the cow.

## Discussion

### High diversity and duplication of the whale *DQB*

All *DQB* sequences recovered from baleen and toothed whales were in reading frame with, and showed greatest similarity to the cow *DQB1*, as noted previously by Murray et al. (1995) and Hayashi et al. (2003). However, unlike the previous surveys by Murray et al. (1995) and Hayashi et al. (2003), we found evidence of gene duplication as well as high allelic and nucleotide diversity for at least some baleen

whale *DQB*. Cloning experiments provided strong evidence for two *DQB* loci in the humpback whale and preliminary evidence for three in the right whale. These duplications could be analogous or homologous with the cow, which is also known to have two or three transcribed *DQB* loci (Ellis 1999). However, we found no significant groupings of alleles that would indicate divergence of the duplicate genes, and so we were unable to attribute alleles to individual loci. Multiple copies of the *DQB* are also reported in the baiji or Chinese river dolphin, although some inflation of allelic diversity due to cloning artifacts could not be discounted in this study (Yang et al. 2005). The finding of duplicate *DQB* genes in the baleen whales (suborder Mysticeti) and the Chinese river dolphin, an early divergence of the toothed whales (suborder Odonotoceti; Cassens et al. 2000), is consistent with retention of an ancestral condition shared with the ruminants and loss in the more derived cetaceans such as the beluga, the narwhal (family Monodontidae) and the true dolphins (family Delphinidae).

Allelic and nucleotide diversity of the humpback and southern right whale *DQB* approached or exceeded that observed in the usually more diverse *DRB* gene of other mammals. With additional population sampling (e.g., from the Southern Hemisphere), it seems likely that the total number of humpback *DQB* alleles will exceed the 40 reported in cattle (Ellis 1999) and approach the more than 60 reported for humans (EMBL-EBI 2005). Average nucleotide diversity among *DQB* alleles of the humpback (5.9%) and southern right whales (6.3%) was greater than that of the beluga (3.7%, calculated here) and comparable to the *DRB*3 exon 2 of wild ruminants (ranging from 1.8 to 10.9% for several species) and the *DRB* of humans (7.7%; Mikko et al. 1999).

The distribution of nucleotide and inferred amino acid substitutions was consistent with the hypothesis of over-dominant (or frequency dependent; Slade and McCallum 1992) selection acting on the PBR (Hughes and Nei 1988). As a likely result of constraints on non-PBR and over-dominant selection on the PBR, the phylogeny showed a pattern of transspecific sharing of similar alleles as reported for some primate class I and II genes (Gyllensten et al. 1991; Lawlor et al. 1988), rather than the species-specific pattern reported for some ruminants (Mikko et al. 1999) and observed here for the *DRB*-like alleles of whales (see below). It remains controversial whether such patterns of apparent transspecific sharing of alleles are due to common lineages (i.e., retention of ancestral alleles) or convergence of independent lineages (Yeager and Hughes 1999). If it is the former, sharing of similar alleles by common ancestry among balaenid and balaenopterid whales would require their preservation in each of these lineages for at least 10–15 million years (Barnes et al. 1985).

## Low diversity and duplication of whale *DRB*-like genes

We found two divergent *DRB*-like genes among the baleen whales. One gene was also present in our survey of toothed whales and is almost certainly an ortholog of the *DRB* locus described previously in the beluga and narwhal (Murray et al. 1995; Murray and White 1998). This locus showed greatest sequence similarity with the cow *DRB*3 and is presumably an ortholog of this ancestral ruminant gene. In contrast to the cow and some other ruminants, where this gene is highly polymorphic (with 68 alleles currently described for the cow and 49 for the red deer (Mikko et al. 1999)), the *DRB*3-like gene of the baleen whales was low in allelic and nucleotide diversity. A second *DRB*-like gene, found only in balaenid whales, showed greatest sequence similarity with the cow *DRB*2 locus, and is presumably an ortholog of this gene or pseudogene. Murray and White (1998) also reported a second *DRB* locus in the beluga (*DRB2**0601 in our Fig. 3), but this sequence is closely related to the other *DRB*3-like genes, not to the *DRB*2 of the cow.

For both *DRB*-like loci, there was mixed evidence of functionality. Sequences in the reading frame with the cow *DRB*3 exon 2 were found in blue, fin, and humpback whales. For the latter two species, SSCP showed these sequences were found in an allelic relationship with other sequences showing interrupted or altered reading frames (i.e., indel alleles). Similarly, the *DRB*2-like sequences of the right and bowhead whales included at least one sequence in reading frame with the cow, but others with interrupted reading frames. No stop or intervening start codons were observed and, ignoring the indels, a number of positions in the inferred protein sequences were conserved among the baleen whales but differed in comparison to toothed whales and the cow. Changes at two sites within the PBR (positions 20 and 38) uniquely group the baleen whales together relative to the toothed whales and ruminant out-groups (Table 3). Finally, distribution of nucleotide divergence was highly localized, with a tenfold higher divergence at the PBR, compared to the non-PBR for the uninterrupted baleen whale alleles.

Overall, these characteristics suggest directional selection of the *DRB* within the Balaenopteridae, accompanied by a reduction in diversity, rather than a complete loss of function. The high rate of both nonsynonymous and synonymous substitutions at the PBR seems best explained by gene conversion or recombination (Parham and Ohta 1996), perhaps with the more diverse *DQB* loci (see below). The indels observed in the alleles of both the *DRB*3- and *DRB*2-like genes could be a consequence of these mechanisms.

It is possible that we have amplified only pseudogenes of the *DRB* and that other fully functional and diverse *DRB* loci remain to be found in baleen whales. The cow, for example, is thought to have nine *DRB* genes, pseudogenes, or gene fragments (Ellis 1999). Only one of these, the *DRB*3, is highly polymorphic and expressed, although the *DRB*2 is transcribed at low levels in lymphocyte tissue. Further investigation by alternate methods (e.g., RT-PCR, screening of genomic libraries) is needed to fully characterize the diversity and function of the whale DRB, although obtaining tissue known to express Class II genes remains an obstacle.

*DQB* and *DRB* convergence

The *DQB* and *DRB* genes are thought to have arisen early in the placental mammals and evolved independently such that sequences of each gene can be recognized as orthologous across lineages (e.g., Ellis 1999; Groenen et al. 1990). Thus, sequences from either gene of different mammals should group together with their orthologs in phylogenetic reconstruction, exclusive of sequences from paralogous loci. Although this pattern was observed for the whale *DRB*3-like sequences when the full length of the fragment was compared with *DRB* and *DQB* sequences of out-groups (Fig 3), it was not the case when sequence length was truncated for comparison to the other cetacean *DQB*. For this shorter fragment, alignment of cetacean *DQB* and *DRB*3 showed a surprising similarity between some sequences (Table 2 and 3) and phylogenetic reconstruction showed weak support for grouping of sequences from both cetacean genes (Fig. 4). A similar pattern of *DQB*/*DRB* convergence is reported in the canids, although in this case, the similarity was not sufficient to disrupt the expected monophyly of alleles from each locus (Seddon and Ellegren 2002). Possible explanations for this pattern include experimental or analytical artifacts, such as nonspecific amplification or insufficient phylogenetic signal, as well as the evolutionary hypothesis of convergent evolution (Yeager and Hughes 1999). Nonspecific amplification of both genes was considered unlikely given the poor match of the *DQB*2 primers to the cetacean *DRB*3 sequences and the contrasting phylogenetic pattern observed for the full-length whale *DRB*3-like sequences. Similarly, the grouping of the shorter whale *DRB*3 sequences together, but within the larger diversity of the cetacean *DQB* sequences, argues against simple laboratory error or contamination as a contributing factor. Although phylogenetic signal was limited by the shorter length of the cetacean *DQB* and the large number of taxa included in the analysis, this limitation was not sufficient to obscure the expected orthologous relationship of the *DQB* and *DRB* sequences of the out-groups (e.g., carnivores, ungulates, and primates; bootstrap support 75%).

Several mechanisms are potentially responsible for convergence between species at orthologous class II genes: point mutations and strong selection, large-scale intergenic recombination, or small-scale intergenic conversions (Andersson and Mikko 1995; Parham and Ohta 1996; Yeager and Hughes 1999). Of these, small-scale conversion of the *DRB* by *DQB* alleles seems most consistent with the available evidence. First, for the fin whale (BaphM02-*DRB**1c) and humpback whale (MenoSEA-*DQB**22c), nucleotide sequences were identical at nine of the ten codons of the PBR, while the small number of other nucleotide differences were scattered across the exon. The identity of both synonymous and nonsynonymous nucleotide positions would be unexpected if the convergence was due to point mutation and selection. Second, the tenfold higher divergence at the PBR, compared to the non-PBR (for both $d_N$ and $d_S$) of the baleen whales *DRB*3-like alleles, suggests small-scale conversion at these codons. Finally, the nesting of the whale *DRB*3 sequences within the *DQB* sequences in the phylogenetic reconstruction suggests directionality consistent with gene conversion.

**Acknowledgements** This study was supported by a grant to CSB, GML, and SJO from the Marsden Fund of New Zealand. We thank the following: A. Dizon and K. Robertson, Southwest Fisheries Science Center, for access to beaked whale samples; S. K. Hooker of Dalhousie University and I. Christensen for access to northern bottlenose whale samples; J. Bannister and C. Burton, Museum of Western Australia, for access to right whale samples; T. Albert, North Slope Borough, Alaska, for access to bowhead whale samples; the New Zealand Department of Conservation for the collection of beach-cast whales and dolphins; J. Calambokidis, J. Straley, and M. Weinrich for assistance in collection of humpback whale samples; J. Martenson, T. Beck, and V. David for assistance at NCI-LGD; and J. McEntee, Z. Velickovic, and D. Steel for assistance at University of Auckland. Three anonymous reviewers provided valuable comments on an earlier draft of the manuscript. All samples were collected under an animal experimentation protocol approved by the University of Auckland Animal Ethic Committee and were transferred with appropriate national permits.

## References

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment tool. J Mol Biol 215:403–410

Ammer H, Schwaiger F, Kammerbauer C, Gomolka M, Arriens A, Lazary S, Epplen J (1992) Exonic polymorphism vs intronic simple repeat hypervariability in MHC-DRB genes. Immunogenetics 35:332–340

Andersson L, Mikko S (1995) Generation of MHC class II diversity by intra- and intergenic recombination. Immunol Rev 143:5–12

Atha DH, Wenz H-M, Morehead H, Tiang J, O'Connell CD (1998) Detection of p53 point mutations by single strand conformation polymorphism: analysis by capillary electrophoresis. Electrophoresis 19:172–179

Baker CS, Clapham P (2004) Modelling the past and future of whales and whaling. Trends Ecol Evol 19:365–371

Baker CS, Perry A, Bannister JL, Weinrich MT, Abernethy RB, Calambokidis J, Lien J, Lambertsen RH, Urban-Ramirez J, Vasquez O, Clapham PJ, Alling A, O'Brien SJ, Palumbi SR (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. Proc Natl Acad Sci U S A 90:8239–8243

Baker CS, Slade RW, Bannister JL, Abernethy RB, Weinrich MT, Lien J, Urban-Ramirez J, Corkeron P, Calambokidis J, Vasquez O, Palumbi SR (1994) Hierarchical structure of mitochondrial DNA gene flow among humpback whales Megaptera novaeangliae, world-wide. Mol Ecol 3:313–327

Barnes WM (1992) The fidelity of Taq polymerase catalysing PCR is improved by an N-terminal deletion. Gene 112:29–35

Barnes LG, Domming DP, Ray CF (1985) Status of studies on fossil marine mammals. Mar Mamm Sci 1:15–53

Beck S, Trowsdale J (1999) Sequence organisation of the class II region of the human MHC. Immunol Rev 167:201–210

Bodmer JG, Marsh SGE, Albert ED, Bodmer WF, Bontrop RE, Charron D, Dupont B, Erlich HA, Fauchet R, Mach B, Mayr WR, Parham P, Sasazuki T, Schreuder GMT, Strominger JL, Svejgaard A, Terasaki PI (1997) Nomenclature for factors of the HLA system 1996. Hum Immunol 53:98–128

Brown J, Jardetzky T, Gorga J, Stern L, Urban R, Strominger J, Wiley D (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR. Nature 364:33–39

Cassens I, Vicario S, Waddell VG, Balchowsky H, Van Belle D, Ding W, Fan C, Lal Mohan RS, Simones-Lopes PC, Bastida R, Meyer A, Stanhope MJ, Milinkovitch MC (2000) Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages. Proc Natl Acad Sci U S A 97:11343–11347

Dalebout M, Hooker S, Christensen I (2001) Genetic diversity and population structure among northern bottlenose whales. Can J Zool 79:478–484

Ellis G (1999) Cattle MHC: evolution in action? Immunol Rev 167:159–168

EMBL-EBI (2005) IMGT/HLA Sequence Database. http://www.ebi.ac.uk/imgt/hla/stats.html

Flores-Ramirez S, Urban-Ramirez J, Miller RD (2000) Major histocompatibility complex class I loci from the gray whale (*Eschrichtius robustus*). J Hered 91:279–282

Groenen MAM, van der Poel JJ, Dijkhof RJM, Giphart MJ (1990) The nucleotide sequence of bovine MHC class II DQB and DRB gene. Immunogenetics 31:37–44

Gyllensten U, Sundvall M, Ezcurra I, Erlich H (1991) Genetic diversity at class II DRB loci of the primate MHC. J Immunol 146:4368–4376

Hayashi K, Nishida S, Yoshida H, Goto M, Pastene LA, Koike H (2003) Sequence variation of the DQB allele in the cetacean MHC. Mamm Study 28:89–96

Hughes AL, Nei M (1988) Patterns of nucleotide substitution at major histocompatibility complex class I loci reveal overdominant selection. Nature 335:167–170

Klein J (1986) The natural history of the major histocompatibility complex. Wiley, New York

Kumar S, Tamura K, Jacobsen IB, Nei M (2001) MEGA, molecular evolutionary genetics analysis. Arizona State University, Tempe, AZ, USA

Lambertsen RH (1987) A biopsy system for large whales and its use for cytogenetics. J Mammal 68:443–445

Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335:268–271

Lento GM, Baker CS, David V, Yuhki N, Gales NJ, O'Brien SJ (2003) Automated single-strand conformation polymorphism reveals low diversity of a major histocompatibility complex class II gene in the threatened New Zealand sea lion. Mol Ecol Notes 3:346–349

Maddison DR, Maddison WP (2000) MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, Massachusetts

Martin A, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci U S A 90:4087–4091

McCallum HI, Kuris A, Harvell CD, Lafferty KD, Smith GW, Porter J (2004) Does terrestrial epidemiology apply to marine systems? Trends Ecol Evol 19:585–591

Mikko S, Road K, Schmutz S, Andersson L (1999) Monomorphism and polymorphism at the Mhc DRB loci in domestic and wild ruminants. Immunol Rev 167:169–178

Muggli-Cockett NE, Stone RT (1989) Partial nucleotide sequence of a bovine major histocompatibility class II DR beta-like gene. Anim Genet 20:361–369

Murray BW, White BN (1998) Sequence variation at the major histocompatibility complex DRB loci in beluga (*Delphinapterus leucas*) and narwhal (*Monodon monoceros*). Immunogenetics 48:242–252

Murray BW, Malik S, White BN (1995) Sequence variation at the major histocompatibility complex locus DQB in beluga whales (*Delphinaterus leucas*). Mol Biol Evol 12:582–593

Nei M (1987) Molecular evolutionary genetics. Columbia Univ. Press, New York

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

O'Brien SJ, Roelke ME, Marker L, Newman A, Winkler CA, Meltzer D, Cody L, Everman JF, Bush M, Wildt DE (1985) Genetic basis for species vulnerability in the cheetah. Science 227:1428–1434

Parham P (1999) Virtual reality in the MHC. Immunol Rev 167: 5–15

Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class 1 molecules. Science 272:67–74

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Rice DW (1998) Marine mammals of the world: systematics and distribution. Society for Marine Mammalogy, Lawrence, Kansas

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

Schlotterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. Nature 354: 63–65

Seddon JM, Ellegren H (2002) MHC class II genes in European wolves: a comparison with dogs. Immunogenetics 54:490–500

Shirai K, Sakai T, Oike T (1998) Molecular cloning of bottle-nosed dolphin (*Tursiops truncatus*) MHC class I cDNA. J Vet Med Sci 60:1093–1096

Sigurdardottir S, Borsch C, Gustafsson K, Andersson L (1991) Cloning and sequence analysis of 14 DRB alleles of the bovine major histocompatibility complex by using the polymerase chain reaction. Anim Genet 22:199–209

Slade RW, McCallum HI (1992) Overdominant vs. frequency-dependent selection at MHC loci. Genetics 132:861–862

Swofford DL (2000) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA

Takahashi K, Rooney A, Nei M (2000) Origins and divergence times of mammalian class II MHC gene clusters. J Hered 91:198–204

Tindall (1988) Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry 27:6008–6013

Trowsdale J, Groves V, Arnason A (1989) Limited MHC polymorphism in whales. Immunogenetics 29:19–24

Yang G, Yan J, Zhou K, Wei F (2005) Sequence variation and gene duplication at the MHC DQB Loci of baiji (Lipotes vexillifer), a Chinese river dolphin. J Hered 96:310–317

Yeager M, Hughes AL (1999) Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev 167:45–58

Yuhki N, O'Brien SJ (1990) DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. Proc Natl Acad Sci U S A 87:836–840

Yuhki N, O'Brien SJ (1997) Nature and origin of polymorphism in feline MHC class II DRA and DRB genes. J Immunol 158:2822–2833