J Mol Evol (2009) 69:260–275
DOI 10.1007/s00239-009-9272-z

Sequence Polymorphism and Evolution of Three Cetacean MHC Genes

Shi Xia Xu · Wen Hua Ren · Shu Zhen Li · Fu Wen Wei · Kai Ya Zhou · Guang Yang

Received: 23 September 2008 / Accepted: 2 August 2009 / Published online: 20 August 2009

Abstract Sequence variability at three major histocompatibility complex (MHC) genes (*DQB*, *DRA*, and MHC-I) of cetaceans was investigated in order to get an overall understanding of cetacean MHC evolution. Little sequence variation was detected at the *DRA* locus, while extensive and considerable variability were found at the MHC-I and *DQB* loci. Phylogenetic reconstruction and sequence comparison revealed extensive sharing of identical MHC alleles among different species at the three MHC loci examined. Comparisons of phylogenetic trees for these MHC loci with the trees reconstructed only based on non-PBR sites revealed that allelic similarity/identity possibly reflected common ancestry and were not due to adaptive convergence. At the same time, trans-species evolution was also evidenced that the allelic diversity of the three MHC loci clearly pre-dated species divergence events according to the relaxed molecular clock. It may be the forces of balancing selection acting to maintain the high sequence variability and identical alleles in trans-specific manner at the MHC-I and *DQB* loci.

S. X. Xu · W. H. Ren · S. Z. Li · K. Y. Zhou · G. Yang (✉)
Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, 1 Wenyuan Road, Nanjing 210046, China
e-mail: gyang@njnu.edu.cn

S. X. Xu · F. W. Wei
Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, 1 Datun Road, Beijing 100101, China

S. X. Xu
Graduate School of Chinese Academy of Sciences, Beijing 100049, China

Keywords Cetacean · MHC · Sequence polymorphism · Trans-species evolution · Balancing selection

Introduction

The major histocompatibility complex (MHC) consists of a group of closely linked genes that constitute the most important genetic component of the mammalian immune system (Klein 1986). The MHC encodes cell-surface glycoproteins that bind antigens derived from pathogens or parasite and present them to T-lymphocytes which trigger the appropriate immune response (Klein and Horejsi 1997; Dengjel et al. 2005). The MHC family is divided into three classes, of which classes I and II genes exhibit the highest polymorphism (Trowsdale 1996). Class I genes encode molecules that recognize intracellular antigens and class II genes encode molecules that recognize extracellular antigens, such as those of invading parasites and other pathogens. A prominent feature of MHC genes is the extremely high degree of polymorphism, both in terms of number of alleles and nucleotide differences (Klein 1986). Great polymorphism at MHC genes is expected to confer immunity against a wide range of pathogens, leading to potentially strong positive selection for allelic and genotypic variability (Penn and Potts 1999). Therefore, investigations of patterns and levels of variability at MHC loci can yield critical insights into the effects of selection on the origin and maintenance of MHC genetic diversity.

MHC genes have received much more attention than before from evolutionary geneticists, as they are considered as a prime example of the effects of balancing selection (Bernatchez and Landry 2003; Garrigan and Hedrick 2003). There are several lines of evidence that polymorphic MHC loci are subject to balancing selection. An excess of

nonsynonymous over synonymous substitutions in the peptide-binding region (PBR) was considered as a primary indication of balancing selection (Bernatchez and Landry 2003). MHC alleles can be maintained for long evolutionary periods, with an origin predating speciation events, resulting in pronounced similarities between alleles of different but closely related species (Klein et al. 1998). This phenomenon, known as trans-species polymorphism (or evolution), is regarded as another evidence for balancing selection (Klein et al. 1998). Up to now, trans-species sharing of identical MHC sequences has been documented in primates (Suárez et al. 2006; Huchard et al. 2006), ungulates (Radwan et al. 2007), and rodents (Cutrera and Lacey 2007). It was noted that case of total identity amongst MHC alleles are generally being restricted to species from the same genus and more rarely from the same family (Suárez et al. 2006; Huchard et al. 2006; Cutrera and Lacey 2007). In addition, some sequence motifs or even identical MHC alleles were also found in some distantly related species generally above genus level, which was explained as convergent evolution (Gustafsson and Andersson 1994; Kriener et al. 2000; Suárez et al. 2006; Xu et al. 2008). The similarities that are shared in the case of convergent evolution are not the result of evolution from a common ancestor but the common adaptive solutions to similar environmental and ecological pressures (Kriener et al. 2000).

MHC variation has been examined in a few species of cetaceans but revealed inconsistent results. Earlier studies demonstrated low level of MHC genetic diversity in fin whales (*Balaenoptera physalus*) and sei whales (*Balaenoptera borealis*) (Trowsdale et al. 1989) which was regarded as a consequence of comparatively weak pathogenic pressure in marine habitats. Sequence analysis of beluga whales (*Delphinapterus leucas*) MHC-II loci (including *DQB* and *DRB*), however, revealed low but measurable polymorphism (Murray et al. 1995; Murray and White 1998). Additionally, recent sequencing analysis of cetacean populations revealed that considerable sequence variation in some species of the baleen whales and toothed whales (Baker et al. 2006; Xu et al. 2007, 2008; Yang et al. 2005a). All studies also found evidence of positive Darwinian selection, as demonstrated by high levels of nonsynonymous substitutions at residues of the PBR.

However, our understanding of cetacean MHC genes is far from systematic. For example, class I sequences have been reported only for gray whales (*Eschrichtius robustus*, Flores-Ramirez et al. 2000), bottlenose dolphin (*Tursiops truncatus*, Shirai et al. 1998), finless porpoises and baiji (Xu et al. 2007, 2008), whereas *DRA* sequence variation has been available only for finless porpoises and baiji (Xu et al. 2007, 2008). Although some recent surveys on *DQB* and *DRB* loci by Hayashi et al. (2003, 2006) and Baker et al. (2006) involved in 22 cetacean species of six families but only having one sample for most species. In particular, several pairs of identical alleles were identified in the baiji and finless porpoise, and possible explanations for this pattern including trans-specific polymorphism and adaptive convergence were put forward in our previous study (Xu et al. 2008). In the present study, sequence of exon 2 at three MHC loci (i.e. MHC-I, *DRA* and *DQB*) were determined for some more representative species. These sequences were combined with those published data in order to get an overall understanding of MHC variation and evolution in cetaceans. Specially, we investigated the relationship between alleles in different cetacean species to evaluate whether certain allele sequences were shared by different cetaceans inhabiting similar or different environments. Further, by investigating trans-specific polymorphism and adaptive convergence in more cetacean species, we sought to provide evidence of which mechanism on these identical alleles that would complement previous analyses of variation at positive selection on MHC genes in cetaceans (Xu et al. 2008; Baker et al. 2006). Moreover, by comparison with homologous sequences of other terrestrial mammals, we expected to evaluate whether the effect of the natural selection on the level of adaptive genetic variation in the cetacean species is consistent with other terrestrial mammals.

Materials and Methods

Samples and DNA Isolation

Nucleotide sequences of three MHC loci (*DRA*, *DQB*, and MHC-I) of 10 cetacean species (nine toothed whales and one baleen whale, see Table 1) were specially sequenced in the present study. Through combining with relevant sequences of other species downloaded from GenBank (Table 1), *DRA* genes from 12 species, *DQB* genes from 28 species, and MHC-I genes from 11 species were available for subsequent analyses.

Genomic DNA was extracted from myologic and skeletal samples by using the DNeasy Tissue Kit (Qiagen) and Gene Clean for Ancient DNA Kit (Q. Biogene), respectively, following the manufacturer's protocol.

Amplification, Cloning and Sequencing

The primers used to amplify the *DRA*, *DQB* and MHC-I genes were described in Xu et al. (2007), Murray et al. (1995), and Flores-Ramirez et al. (2000). Polymerase chain reaction (PCR) was carried out in a total volume of 50 μl comprising 2.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.4), 50 mM KCl, 0.2 mM each dNTP, 0.4 μM each primer, 1.0

Table 1 Samples of cetacean species analyzed in this study

Suborder	Family	Scientific name (abbreviation)[a]	Common name	Sampling locality	Sample size	Number of sequences			Reference
						DRA	*DQB*	MHC-I	
Odontoceti (toothed whales)	Lipotidae	*Lipotes vexillifer* (Live)	Baiji	Yangtze River, China	15	2	8[b]	6	Xu et al. (2008) Yang et al. (2005a)
	Platanistidae	*Platanista gangetica* (Plga)	Indian River dolphin	Bengal	2	1	–[c]	–	This study
	Pontoporiidae	*Pontoporia blainvillei* (Pobl)	La Plata dolphin	Argentina	2	1	2	2	This study
	Phocoenidae	*Neophocaena phocaenoides* (Neph)	Finless porpoise	Yangtze River, Yellow Sea, South China Sea, China, and Japanese Coast	245	5	17	34	Xu et al. (2007) Hayashi et al. (2003)
		Phocoena phocoena (Phph)	Harbor porpoise	Presented by SWFSC, USA	5	2	4	2	This study
		Phocoenoides dalli (Phda)	Dall's porpoise	Presented by SWFSC, USA	5	2	3	2	This study
		Phocoena sinus (Phsi)	Gulf of California endemic porpoise	na	25	–	1	–	Munguia-Vega et al. (2007)
	Delphinidae	*Stenella coeruleoalba* (Stco)	Striped dolphin	South China Sea, China	15	6	9	8	This study
		S. attenuata (Stat)	Pantropical spotted dolphin	South China Sea, China	4	2	1	2	This study
		Tursiops aduncus (Tuad)	Indian Ocean bottlenose dolphin	South China Sea, China	4	2	5	2	This study
		T. truncatus (Tutr)	Bottlenose dolphin	na ($n = 1$) Taiwan coast ($n = 42$)	43	–	7	–	Hayashi et al. (2003) Yang et al. (2008)
		Globicephala macrorhynchus (Glma)	Short-finned pilot whale	na	2	–	3	–	Hayashi et al. (2003)
		Grampus griseus (Grgr)	Risso's dolphin	South China Sea, China	1	2	1	2	This study
		Delphinus capensis (Deca)	Long-beaked common dolphin	East China Sea, China	1	2	1	3	This study
		Cephalorhynchus hectori (Cehe)	Hector's dolphin	na	na	–	2	–	Baker et al. (2006)
	Monodontidae	*Monodon monoceros* (Mono)	Narwhal	High Arctic	4	–	1	–	Murray et al. (1995)
		Delphinapterus leucas (Dele)	Beluga	St Lawrence Estuary, Hudson Bay, Eastern Beaufort Sea, High Arctic	9	–	5	–	Murray et al. (1995)

Table 1 continued

Suborder	Family	Scientific name (abbreviation)[a]	Common name	Sampling locality	Sample size	Number of sequences			Reference
						DRA	*DQB*	MHC-I	
	Ziphiidae	*Hyperoodon ampullatus* (Hyam)	Northern bottlenose whale	Davis Strait off the Northern Labrador	1	–	1	–	Baker et al. (2006)
		Mesoplodon stejnegeri (Mest)	Stejneger's beaked whale	na	na	–	1	–	Baker et al. (2006)
		M. grayi (Megr)	Gray's beaked whale	na	na	–	1	–	Baker et al. (2006)
		M. europaeus (Meeu)	Gervais' beaked whale	na	na	–	2	–	Baker et al. (2006)
Mysticeti (baleen whales)	Balaenopteridae	*Balaenoptera omurai* (Baom)	Omura's baleen whale	East China Sea, China	1	1	–	–	This study
		B. physalus (Baph)	Fin whale	na	na	–	3	–	Baker et al. (2006) Hayashi et al. (2003)
		B. bonaerensis (Babo)	Antarctic minke whale	na	na	–	6	–	Hayashi et al. (2003)
		B. acutorostrata (Baac)	Minke whale	na	9	–	1	–	Hayashi et al. (2003)
		Megaptera novaeangliae (Meno)	Humpback whale	na	na	–	7	–	Baker et al. (2006)
	Eschrichtiidae	*Eschrichtius robustus* (Esro)	Gray whale	Guerrero Negro Lagoon BCS, Mexico	9	–	2	14	Baker et al. (2006) Flores-Ramirez et al. (2000, 2004)
	Balaenidae	*Balaena mysticetus* (Bamy)	Bowhead whale	North Slope Borough, Alaska, USA	na	–	3	–	Baker et al. (2006)
		Eubalaena australis (Euau)	Southern right whale	na	na	–	4	–	Baker et al. (2006)

SWFSC The Southwestern Fisheries Science Centre of USA

[a] Abbreviation of each species (in the parentheses) for alleles

[b] Only those alleles identified from at least two different clones or two individuals were included

[c] PCR amplification unsuccessful for this species

unit Ex-Taq DNA polymerase (Takara), and 10–100 ng DNA template. The amplification profile consisted of 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 55°C and 30 s at 72°C, with a final extension of 8 min at 72°C. PCR products were purified using Wizard PCR Preps DNA Purification Kit (Promega) according to the manufacturer's protocol.

Purified PCR products were cloned into a pMD-18T vector (Takara) following manufacturer's recommendations and transformed into *Escherichia coli* DH-5α competent cells. Positive transformants containing insert of approximate length were identified by PCR screening and agarose electrophoresis. Five to ten clones were selected from each individual to sequence in the forward and/or reverse directions using the BigDye terminator cycle sequencing ready reaction kit (ABI) on an ABI 3730 automated genetic analyzer.

Data Analysis

MHC Diversity Analyses

All the cetacean MHC sequences were aligned by CLUSTAL X 1.83 (Thompson et al. 1997) and translated into the corresponding amino acid sequences using the program MEGA 4 (Tamura et al. 2007). The average pairwise nucleotide distances (Kimura 2-parameter model, K2P), Poisson-corrected amino acid distances were computed in MEGA 4 (Tamura et al. 2007). Standard errors of the estimates were obtained through 1000 bootstrap replicates.

Detecting Selection

Evidence for positive selection was assessed using the CODEML subroutine contained in the PAML 4 program suite (Yang 2007). This procedure, which is considered to be more sensitive than other methods for detecting molecular-level evidence of selection (Anisimova et al. 2003), uses maximum likelihood estimation to examine heterogeneity in ω ($\omega = d_N/d_S$, Goldman and Yang 1994) among codons within a sequence. Positive selection is indicated by $\omega = d_N/d_S > 1$. Estimates of ω were generated following the protocol of Yang et al. (2000). Six different models that allow for different intensity of selection among sites (and deduced from the data) were tested in this study: M0 (one ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (β model) and M8 (β and ω) (Yang et al. 2000, details on models M1a and M2a in Yang et al. 2005c). We carried out three likelihood-ratio tests comparing nested models (M0 vs. M3, M1a vs. M2a, and M7 vs. M8), in which the alternative models (M2a, M3, and M8) suggest the presence of sites with $\omega > 1$. All three tests may therefore be considered tests for positive selection (Yang et al. 2000). Posterior probabilities for site classes have been calculated by Bayes empirical Bayes (BEB) method in models M2a and M8 (Yang et al. 2005c). If the posterior means of ω for some site classes are >1 (calculated as the average of ω over all site classes weighted by the posterior probabilities), those sites are likely to be under positive selection (Yang et al. 2005c). In addition, amino acid positions involved in the peptide binding were identified by comparing with the peptide binding groove structure of the human class I and class II molecules (Bjorkman et al. 1987; Brown et al. 1993).

Phylogenetic Reconstruction

To analyze the relationship of MHC alleles of different species and to test for the trans-species evolution in MHC allelic lineages, phylogenetic trees of three MHC loci were separately constructed according to their nucleotide sequences. Phylogenetic analyses were performed using Bayesian analyses in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). Modeltest 3.7 (Posada and Crandall 1998) was used to determine the best-fit model (based on Akike Information Criterion, AIC) for each of the three MHC loci datasets. AIC analyses revealed HKY+G to be the most appropriate evolutionary model for the *DRA*, GTR+G for the *DQB*, and TVM+G for the MHC-I. These evolutionary models were respectively applied in the phylogenetic reconstruction of three MHC loci. Four Metropolis-coupled Markov chain Monte Carlo (three of them 'heated', temperature = 0.20) were run for 1×10^6 generations and sampled every 1000 generations. The first 10% of trees were discarded as burn-in, resulting in 900 sampled trees. To calculate the posterior probability of each bipartition, the majority-rule consensus tree was computed from these 900 sampled trees. Sequences of bovine (*Bos indicus*) MHC-I, pig (*Sus scrofa*) and cattle (*Bos taurus*) *DQB*, cattle and sheep (*Ovis aries*) *DRA* were used as outgroups.

Results

Amplification and Sequence Variation

DRA

The samples from the 10 cetacean species sequenced in this study all had successful amplification of the *DRA* gene, and 21 *DRA* alleles identified in this study have been deposited in GenBank with accession numbers EF375585–EF375605. In combination with relevant sequences of the baiji and finless porpoise previously published, a total of 28 *DRA* sequences were analyzed. The number of unique sequences in each species is shown in Table 1. Each unique sequence

was identified from at least two replicates of PCR products or independent clones or individuals. No indels (insertions/deletions) or stop codons were detected, suggesting that all sequences might come from functional molecules in the genome. There were no more than two sequences for each individual at the *DRA* locus, suggesting that one single locus was amplified using the primers. Sequence comparison revealed that the absolute sequence difference between alleles was up to 3.17% within species and 7.94% between species at the nucleotide level (Fig. 1a). Maximum

(a) *DRA*

```
                1111111222 2222222333 3333333444 4444444555 5555555666 6666666777 777
                3456789012 3456789012 3456789012 3456789012 3456789012 3456789012 345
                       * ** *   **          *      * *** *  *  * ** *
Plga-DRA*01 SLSPDQSNEF MFDFDGDEIF HVDMEKRETV WRLKEFGNFA SFQAQGALAN MAVGKANLDI LIK
Phph-DRA*01 .......... .......... .......... ...E...... .......... .......... ...
Pobl-DRA*01 .......S.. .......... ......K... .......... .......... .......... ...
Phph-DRA*02 .......... .......... .......... .......... .......... .......... ...
Phda-DRA*01 .......... .......... .......... .......... .......... .......... ...
Phda-DRA*02 .......... .......... .......... ...E...... .......... .......... ...
Stco-DRA*01 ....Y..... .......... ......K... .......... .......... ....R..... ...
Stco-DRA*02 ....Y..... .......... ......K... .......... .......... ....R..... ...
Stco-DRA*03 ....Y..... .......... ......K... .......... .......... ....R..... M..
Stco-DRA*04 ....Y..... .......... ......K... .......... .......... ....R..... M..
Stco-DRA*05 ....Y..... .......... ......K... ........L. .......... ....R..... ...
Stco-DRA*06 .......... .......... .......... .......... .......... .......... ...
Stat-DRA*01 ....Y..... .......... ......K... .......... .......... ....R..... ...
Stat-DRA*02 .......... .......... .......... .......... .......... .......... ...
Tuad-DRA*01 .......... .......... .......... .......... .......... .......... ...
Tuad-DRA*02 ....Y..... .......... ......K... .......... .......... ....R..... ...
Grgr-DRA*01 ....Y..... .......... ......K... .......... .......... ....R..... ...
Grgr-DRA*02 ....Y..... .......... ......K... .......... .......... ....R..... ...
Deca-DRA*01 ....Y..... .......... ......K... .......... .......... ....R..... M..
Deca-DRA*02 .......... .......... .......... .......... .......... .......... ...
Baom-DRA*01 ..V....G.. .......... ...L.RK... ...E...... .......... I......... M..
```

(b) *DQB*

```
                2222222223 3333333334 4444444445 5555555556 6666666667 7777777
                1234567890 1234567890 1234567890 1234567890 1234567890 1234567
                     ** *   **        *        *  **  * **** *
Pobl-DQB*01 TERVRFVSRY IYNREEFVRY DSDVGEYRAV TELGRPDAEY WNSQKDIMEQ TRAELDT
Pobl-DQB*02 .......... ......L... .......... .....RT... .......L.R ....V..
Phph-DQB*01 .......E.H .......L.F .......... .....RI..N ..G...LL.R K..V...
Phph-DQB*02 .....L.E.L ......Y..F .......... .....RT... ..G...LL.. K..A...
Phph-DQB*03 .....L.E.. ......Y..F ......H... .....RT... ..G...LL.. K..A...
Phph-DQB*04 .....L.E.H .......L.F .......... .....RI..N ..G....L.R K......
Phda-DQB*01 .......E.. ......Y..F .......... ........K. ..G....L.. K......
Phda-DQB*02 .....L.E.. ......Y..F .......... .....RT..N ..G...LL.. K..A...
Phda-DQB*03 .......E.. .........F .......... .....RI..N ..G....L.R K..V...
Stco-DQB*01 .......D.. .........F ......H... .....RI..D .......V.R K......
Stco-DQB*02 .......... .......... .......... .......... .......... .......
Stco-DQB*03 .....L...S .........F ......F... .......... .......L.R K..AV..
Stco-DQB*04 .....V.D.S ......L..F ......F... .......... ......FL.R K......
Stco-DQB*05 .....H.... ......Y..F ......F..M .......... F......L.. E..DV.R
Stco-DQB*06 .....H.... ......Y..F ......F... .......... F......L.E E..DV.R
Stco-DQB*07 .....H.... .........F ......F... .......... ......FL.. E..DV.R
Stco-DQB*08 .......D.. .........F ......F... .......... .......L.. EW.DG..
Stco-DQB*09 .......D.. .........F ......F... .......... .......L.R EW.DG..
Stat-DQB*01 .......D.. .........F ......F... .......... .......L.R EW.DG..
Tuad-DQB*01 .....V.D.S ......L..F ......F... .......... ......FL.R K......
Tuad-DQB*02 .....H.... .........F ......F... ........K. ....E..L.R K..DV..
Tuad-DQB*03 .....H.... ......Y..F ......F... .....RT... ....E..L.R K..DV..
Tuad-DQB*04 .......... .......... .......... .......... .......... .......
Tuad-DQB*05 .....L.E.. .........F ......F... ........T. ..G.E..L.E E..A...
Grgr-DQB*01 .......D.. .........F ......F... .......... .......L.R EW.DG..
Deca-DQB*01 .....L...S .........F ......F... .......... .......L.R K..AV..
```

(c) MHC-I

```
                3333333444 4444444555 5555555666 6666666777 7777777888
                4567890123 4567890123 4567890123 4567890123 4567890123 456789012
                           *          ** **  ** * * ** * **
Pobl-I*01 LRFDGDAPNP RMELWAPWVE QKGPEYWDRN PRIYKHHAQT FRVNLNTLC
Pobl-I*02 V......... .C.PR..... .E....EAE T.LF.GV..M Y..E..K.R
Phph-I*01 V...S..... .K.PR...M. .V.......E TQ.S.DA..V Y.....K.R
Phph-I*02 V...S..... ...PR..... .E.......E T.NF.EN..F Y.ED....R
Phda-I*01 V...S..... .K.PR..... .E......LE TQ.S.AA..V Y.....K.R
Phda-I*02 V...S..... .G.PR..... .E.......E TQ.F.DA..V Y.E.....R
Stco-I*01 .......... ...P...... .......... .........
Stco-I*02 M......... .C.PR..... .E....EEE T.LF.GA..M Y..D....R
Stco-I*03 M......... .C.PR..... .E.......E T.NF.ER..M Y.AG.DI.R
Stco-I*04 V......... ...PR..... .E....EAE T.LF.GA..M YL....K.R
Stco-I*05 V...S..... .K.PR..... .E.......E T.VS.EN..S Y.RS.DN.R
Stco-I*06 V...S..... .K.PR..... .E.......E T..S.EN..S ..AG.DN.R
Stco-I*07 S......... ...P...... .......... ......Y... .........
Stco-I*08 .......... ...P...... .......... ......Y... .........
Stat-I*01 V...S..... .K.PR..... .E.......E T..S.EN..S ..AG.DN.R
Stat-I*02 V...S..... .G.PR...M. .E.......E T.NF.EN..F Y.ED....R
Tuad-I*01 VW..S..... .K.PR..... .E....KEE T..S.ER..I Y.......R
Tuad-I*02 VW..S..... .K.PR...M. .E....EEE T..S.DA..I Y.....N.R
Grgr-I*01 VW..S..... .K.PR..... .E....KEE T..S.ER..I Y.......R
Grgr-I*02 V...S..... .K.PR..... .E.......E T.VS.EN..S Y.RS.DN.R
Deca-I*01 V...S..... .E.PR..... .E....EEQ TQVC.DT..I ..G...N.R
Deca-I*02 V...S..... .E.PR..... .E....EEQ T..S.GD..I Y.RS.DN.R
Deca-I*03 .......... ...P...... .......... .......... .........
```

Fig. 1 Putative amino acid sequences translated from *DRA* (**a**), *DQB* (**b**) and MHC-I (**c**) exon 2 alleles of the 10 cetacean species sequenced in this study. *Dots* (…) indicate identity to the reference sequence. Putative peptide-binding region (PBR) (Brown et al. 1993; Bjorkman et al. 1987) are marked with *asterisks* (*)

divergence across the inferred protein sequence was 15.87% between species, i.e. *Baom-DRA*01* vs. *Stco-DRA*05*. It was noted that the minimum nucleotide and amino acid distance was zero between alleles from different species, indicating that some alleles were shared between species. Further analyses showed that five *DRA* sequences were shared by different species. Of the five sharing sequences, one was found in nine species (i.e. *Stat-DRA*02* vs. *Live-DRA*01* vs. *Phph-DRA*02* vs. *Deca-DRA*02* vs. *Phda-DRA*01* vs. *Tuad-DRA*01* vs. *Plga-DRA*01* vs. *Neph-DRA*01* vs. *Stco-DRA*06*), another one found in three species (*Stco-DRA*01* vs. *Stat-DRA*01* vs. *Grgr-DRA*01*), and the rest three found in only two species (*Phda-DRA*02* vs. *Phph-DRA*01*; *Grgr-DRA*02* vs. *Stco-DRA*02*; *Deca-DRA*01* vs. *Stco-DRA*04*). In addition to these identical alleles, some alleles from different species differed at only one single nucleotide, e.g. *Live-DRA*01* vs. *Phph-DRA*01*, *Stco-DRA *03* vs. *Deca-DRA*01*.

DQB

Amplification of the *DQB* locus was successful in 8 of the 10 species examined in this study, but failed in the Indian River dolphin and Omura's baleen whale. In addition to 26 *DQB* alleles (GenBank accession nos. EU698952–EU698977) determined from the above eight species (Fig. 1b), other 75 *DQB* alleles downloaded from the GenBank were also included. Consequently, a total of 101 *DQB* alleles were obtained from a total of 28 cetacean species (Table 1). There were no more than two sequences for each individual from all species except for the baiji and finless porpoise at the *DQB*. All the *DQB* sequences identified in this study might come from functional molecules in the genome as indicated by no indels or stop codons in all the sequences. The absolute sequence difference between alleles was up to 12.28% within species and 14.62% between species at the nucleotide level (Fig. 1b). In addition, the maximum amino acid sequence divergence was up to 29.82% between the bottlenose dolphin (*Tutr-DQB*02*) and the harbor porpoise (*Phph-DQB*01*). Similar to the *DRA* locus, seven *DQB* allele sequences were shared by different species as evidence by the zero nucleotide and amino acid distance between alleles. One sequence corresponding to the striped dolphin (*Stco-DQB*02*) was shared with the La Plata dolphin (*Pobl-DQB*01*) and Indian Ocean bottlenose dolphin (*Tuad-DQB*04*), and another allele sequence of the striped dolphin (*Stco-DQB*09*) was shared with the pantropical spotted dolphin (*Stat-DQB*01*) and Risso's dolphin (*Grgr-DQB*01*). Interspecific identities were also found in other five cases, i.e. *Phda-DQB*01* vs. *Neph-a*, *Tuad-DQB*01* vs. *Stco-DQB*04*, *Stco-DQB*03* vs. *Deca-DQB*01*, *Neph-DQB*4* vs. *Live-DQB*14*, *Neph-DQB*03* vs. *Live-DQB*16*. Besides the above identical alleles, some alleles showed unexpectedly high similarity between different species. For example, one pair of alleles separately from the finless porpoise and baiji (i.e. *Neph-DQB*14* vs. *Live-DQB*28*) differed only at one single base, and *Momo-DQB*0201* and *Dele-DQB*0201* was detected only with two nucleotide differences.

MHC-I

Similar to the *DQB*, amplification for the MHC-I gene were unsuccessful in the Indian River dolphin and Omura's baleen whale, and 23 MHC-I unique sequences (GenBank accession nos. EU698978–EU699000) were identified in remaining eight species examined in this study. In combination with those published sequences of baiji, finless porpoises, and gray whales, 76 MHC-I sequences from a total of 11 cetacean species were obtained. Three or more MHC-I loci were possibly active because 3–5 distinct sequences were detected from most species and individuals. None of the sequences contained indels or stop codons. Extensive variability was detected at this locus, with the absolute sequence difference up to 20.41% within species and 23.13% between species at the nucleotide level (Fig. 1c). These values indicated the divergence both within and between loci as we were not able to distinguish alleles of particular loci at present. Taking into account the fact that MHC polymorphism might be not only by true locus allelism, we considered all sequences as if they would be alleles of one locus in further analyses. The number of pairwise nucleotide differences between pairs of MHC-I sequence in 11 cetacean species ranged from 0 to 34, and the number of amino acid differences ranged from 0 to 20. Maximum divergence across the inferred protein sequence up to 40.82% was found between the striped dolphin (*Stco-I*03*) and the gray whale (*Esro64.6*). As was found at the *DRA* and *DQB* loci, extensive sharing of alleles between species was observed for the MHC-I gene. A total of eight MHC-I sequences were shared by different species, i.e. *Pobl-I*01* vs. *Neph-I*14* vs. *Live-I*04*, *Stco-I*05* vs. *Grgr-I*02*, *Stco-I*06* vs. *Stat-I*01*, *Stco-I*01* vs. *Deca-I*03*, *Grgr-I*01* vs. *Tuad-I*01*, *Neph-I*19* vs. *Live-I*01*, *Neph-I*33* vs. *Live-I*02*, *Neph-I*04* vs. *Stat-I*02*. In addition to these identical sequences, some alleles have high similarity between species. For instance, one allele of the long-beaked common dolphin (*Deca-I*03*) only has one base difference with one allele (*Stco-I*08*) from the striped dolphin.

Patterns of Selection

Examination of the amino acid chains encoded by the MHC-I and *DQB* alleles provided potential evidence of positive selection on the two loci. Although the mean

pairwise number of differences between nucleotide sequences was greater than that between amino acid chains, the corrected percent divergence between variants was greater for amino acids (Table 2). The greater corrected divergence among amino acid sequences suggests that nonsynonymous substitutions are more frequent than synonymous changes, as expected for MHC alleles that are currently or have evolved under positive selection.

Additionally, the positive selection was also evidenced by the maximum-likelihood-based random-sites models analysis using PAML 4. Likelihood ratio tests revealed that for the *DQB* and MHC-I genes, all models of codon evolution that included selection (i.e., M2a, M3, and M8) fit the data significantly better than models (i.e. M1a, M0, and M7) that did not incorporate selection (Table 3). At the *DQB* locus, seven codons were each identified by model M2a model M8 as subject to significant positive selection (Table 4). For MHC-I exon 2, six sites were significantly positively selected under the model M2a and seven sites under the M8 (Table 4). These codons are thought to be involved in or adjacent to PBR in the associated MHC class I and II molecule (Brown et al. 1993; Bjorkman et al. 1987). Thus, for MHC-I and *DQB* loci, both nucleotide substitution rates and the locations of non-neutral codons are consistent with the effects of positive selection.

While for the *DRA* locus, all the six models had nearly equivalent likelihood scores (Table 4). In addition, no site was identified as subject to significant positive selection by model M2a and M8 (Table 4), which provided no evidence for positive selection acting on this locus.

Phylogenetic Reconstruction

The phylogenetic relationships were reconstructed in terms of the nucleotide sequences determined in this study and those downloaded from GenBank. The phylogenetic trees showed that all cetacean MHC sequences in each of three loci formed a monophyletic group relative to outgroups with high bootstrap supports (Fig. 2). For the *DRA* locus, the cetacean alleles were further separated into two clades respectively corresponding to toothed whales (suborder Odontoceti) and baleen whales (suborder Mysticeti) (Fig. 2a). However, monophyletic groups for two cetacean suborders were not found at the *DQB* and MHC-I loci (Fig. 2b, c). Further, it was noted that the alleles were not separated according to species but were intermixed with one another for all the three loci examined. Some alleles were more closely related with those from other species rather than with intraspecific alleles. In the MHC-I tree, alleles from closely related species, such as species of the family Phocoenidae or Delphinidae, fell into one cluster (Fig. 2c). Two alleles of the Risso's dolphin (*Grgr-I*01* and *02*) respectively clustered with the allele of the Indian Ocean bottlenose dolphin (*Tuad-I*01*) and striped dolphin (*Stco-I*05*) with 100 and 97% bootstrap support. Similarly, two alleles of the pantropical spotted dolphin (*Stat-I*01*, *02*) separately grouped with that of the striped dolphin (*Stco-I*06*) and finless porpoise (*Neph-I*04*) with high bootstrap values. In addition, some alleles from distantly related species also grouped together in the phylogenetic trees. For example, some alleles of the baiji included in the family Lipotidae were always clustered with those of the finless porpoise in the family Phocoenidae. Also, some alleles separately from the La Plata dolphin (*Pobl-I*01*) and baiji (*Live-I*04*) grouped together, then clustered with an allele of the finless porpoise (*Neph-I*14*). Similar results were found in the phylogenetic reconstruction for the *DQB* (Fig. 2b) and *DRA* (Fig. 2a) loci. For instance, for the *DRA* tree, alleles from nine species (i.e. *Live-DRA*01*, *Neph-DRA*01*, *Stco-DRA*06*, *Tuad-DRA*01*, *Phda-DRA*01*, *Deca-DRA*02*, *Stat-DRA*02*, *Plga-DRA*01* and *Phph-DRA*02*) grouped together although were not supported with significant bootstrap values (Fig. 2a).

Table 2 Mean nucleotide and amino acid divergence among alleles at three MHC loci in cetaceans, with standard errors given in parentheses

	Locus		
	DRA	*DQB*	MHC-I
Nucleotide sequences			
Mean Kimura-2P distance (%)	2.00 (0.54)	8.33 (1.36)	12.33 (1.73)
Mean no. of differences	3.69 (0.97)	12.61 (1.83)	16.44 (2.08)
Amino acid sequences			
Mean Poisson-corrected distance (%)	4.42 (1.54)	16.74 (3.61)	26.84 (5.14)
Mean no. of differences	2.68 (0.88)	7.63 (1.44)	11.33 (1.82)

Table 3 Summary of test statistics for the likelihood-ratio test of codon evolution at exon 2 of the *DRA*, *DQB*, and MHC-I genes in cetacean

Locus	Models compared	df	Test statistic	Significance (*P*)
DRA	M2a vs. M1a	2	2.706	0.258
	M3 vs. M0	4	4.164	0.384
	M8 vs. M7	2	2.718	0.257
DQB	M2a vs. M1a	2	87.618	<0.001
	M3 vs. M0	4	480.584	<0.001
	M8 vs. M7	2	82.908	<0.001
MHC-I	M2a vs. M1a	2	87.864	<0.001
	M3 vs. M0	4	294.766	<0.001
	M8 vs. M7	2	87.550	<0.001

df degrees of freedom

Test statistics was computed as 2 ($L_b - L_a$), where L_a and L_b are log-likelihood values for each of the nested models being compared

Table 4 Results of maximum likelihood analyses of models of codon evolution for the MHC *DRA*, *DQB* and MHC-I loci in 30 species of cetaceas

Model code	*P*	Log-likelihood	Parameter estimates	Positively selected sites
DRA locus				
M0 (one ratio)	1	−411.278	$\omega = 0.935$, $K = 5.960$	None
M1a (nearly neutral)	1	−410.549	$p_0 = 0.327$, $p_1 = 0.673$, $K = 5.522$, $\omega_0 = 0$, $\omega_1 = 1$	Not allowed
M2a (positive selection)	3	−409.196	$p_0 = 0.907$, $p_1 = 0$, $p_2 = 0.538$, $K = 6.422$, $\omega_0 = 0.538$, $\omega_1 = 1$, $\omega_2 = 6.306$	None
M3 (discrete)	5	−409.196	$p_0 = 0.102$, $p_1 = 0.805$, $p_2 = 0.093$, $K = 6.422$, $\omega_0 = 0.538$, $\omega_1 = 0.538$, $\omega_2 = 6.306$	Not analyzed
M7 (β)	2	−410.556	$p = 0.012$, $q = 0.005$, $K = 5.574$	Not allowed
M8 (β and ω)	4	−409.197	$p_0 = 0.907$, $p_1 = 0.093$, $p_2 = 99.000$, $q = 84.745$, $\omega = 6.308$, $K = 6.422$	None
DQB locus				
M0 (one ratio)	1	−2011.662	$\omega = 0.825$, $K = 2.156$	None
M1a (nearly neutral)	1	−1820.012	$p_0 = 0.693$, $p_1 = 0.307$, $K = 2.640$, $\omega_0 = 0.047$, $\omega_1 = 1$	Not allowed
M2a (positive selection)	3	−1776.203	$p_0 = 0.619$, $p_1 = 0.242$, $p_2 = 0.139$, $K = 2.566$, $\omega_0 = 0.056$, $\omega_1 = 1$, $\omega_2 = 4.226$	**26F**, **28N**, *29R*, **30N**, **37Y**, **57T**, **71K**
M3 (discrete)	5	−1771.370	$p_0 = 0.638$, $p_1 = 0.226$, $p_2 = 0.135$, $K = 2.479$, $\omega_0 = 0.091$, $\omega_1 = 1.745$, $\omega_2 = 6.086$	Not analyzed
M7 (β)	2	−1817.740	$p = 0.147$, $q = 0.263$, $K = 2.625$	Not allowed
M8 (β and ω)	4	−1776.286	$p_0 = 0.860$, $p_1 = 0.140$, $p_2 = 0.167$, $q = 0.377$, $\omega = 3.875$, $K = 2.590$	**26F**, **28N**, **29R**, **30N**, **37Y**, **57T**, **71K**
MHC-I				
M0 (one ratio)	1	−1715.475	$\omega = 1.077$, $K = 2.965$	None
M1a (nearly neutral)	1	−1614.660	$p_0 = 0.628$, $p_1 = 0.372$, $K = 2.831$, $\omega_0 = 0.084$, $\omega_1 = 1$	Not allowed
M2a (positive selection)	3	−1570.728	$p_0 = 0.514$, $p_1 = 0.301$, $p_2 = 0.185$, $K = 3.083$, $\omega_0 = 0.115$, $\omega_1 = 1$, $\omega_2 = 4.911$	**45K**, **67S**, **70A**, **73I**, **77N**, **80I**
M3 (discrete)	5	−1568.092	$p_0 = 0.604$, $p_1 = 0.269$, $p_2 = 0.127$, $K = 3.062$, $\omega_0 = 0.202$, $\omega_1 = 1.971$, $\omega_2 = 7.523$	Not analyzed
M7 (β)	2	−1616.168	$p = 0.244$, $q = 0.359$, $K = 2.793$	Not allowed
M8 (β and ω)	4	−1572.393	$p_0 = 0.813$, $p_1 = 0.187$, $p_2 = 0.308$, $q = 0.427$, $\omega = 4.618$, $K = 3.096$	**45K**, *52V*, **67S**, **70A**, **73I**, **77N**, **80I**

Note: Analyses were completed using CodeML (included in the PAML 3.14 program suite). Positively selected sites were identified in models M2a and M8 by the Bayes empirical Bayes procedure (Yang et al. 2005c). Sites inferred to be under selection at the 99% level are listed in bold; those at the 95% level are italicized. For models M7 and M8, p and q are the shape parameters of the β function. *P* number of parameters in the ω distribution, *K* estimated transition/transversion rate ratio, ω selection parameter, and p_n proportion of sites that fall into the ω_n site class

Further analyses found that the alleles from different species but having close relationship in the phylogenetic reconstructions are those alleles which shared sequence or had unexpected high nucleotide similarity between species as revealed above. Most identical alleles were found in species within the same family (e.g. Delphinidae or Phocoenidae). For instance, three pairs of MHC-I alleles (i.e. *Stco-I*05* vs. *Grgr-I*02*; *Stco-I*06* vs. *Stat-I*01*; *Deca-I*03* vs. *Stco-I*01*) of the family Delphinidae, one pair of *DQB* alleles (*Phda-DQB*01* vs. *Neph-a*) of the family Phocoenidae, three pairs of *DQB* alleles (*Stco-DQB*04* vs. *Tuad-DQB*01*; *Stco-DQB*09* vs. *Stat-DQB*01* vs. *Grgr-DQB*01*; *Stco-DQB*03* vs. *Deca-DQB*01*) of the family Delphinidae, one pair of *DRA* alleles of the family Phocoenidae (*Phda-DRA*02* vs. *Phph-DRA*01*), and three pairs of *DRA* alleles of the family Delphinidae (*Deca-DRA*01* vs. *Stco-DRA*04*; *Grgr-DRA*02* vs. *Stco-DRA*02*; *Stco-DRA*01* vs. *Stat-DRA*01* vs. *Grgr-DRA*01*) were respectively identical. More interestingly, case of identical alleles also came from some very distantly related species (e.g. in different families and/or superfamilies). For example, one allele of the finless porpoise (family Phocoenidae) (*Neph-I*04*) was identical to *Stat-I*02* identified in the pantropical spotted dolphin (family

Delphinidae). In addition, species from the families Phocoenidae, Delphinidae, Pontoporiidae, and Lipotidae shared one sequence at the MHC-I and *DRA* loci, i.e. MHC-I alleles: *Neph-I*14* vs. *Pobl-I*01* vs. *Live-I*04*, and *DRA* alleles: *Neph-DRA*01* vs. *Phph-DRA*02* vs. *Phda-DRA*01* vs. *Stat-DRA*01* vs. *Tuad-DRA*01* vs. *Deca-DRA*02* vs. *Stco-DRA*06* vs. *Plga-DRA*01* vs. *Live-DRA*01*. However, identical alleles were not found between toothed whales and baleen whales in this study.

Discussion

Sequence Variability at Three Cetacean MHC Loci

Three MHC loci were successfully amplified for 40 samples from 10 cetacean species examined with an exception for *DQB* and MHC-I genes of the Indian River dolphin and Omura's baleen whale. No more than two sequences were identified in each individual at the *DRA*, suggesting that only a single locus was amplified for this gene. For the *DQB*, all species except for the baiji and finless porpoise had no more than two sequences. Considering the sample sizes for most species examined in this study were less than 5, duplication of this gene might be revealed if more samples are examined. In contrast, 3–5 MHC-I unique sequences were examined in at least 5 of 10 species, which strongly suggested at least three copies of MHC-I gene amplified using the present primer sets. This is consistent with the data in some artiodactyls with three MHC-I loci (al-Murrani et al. 1994; Velten et al. 1999; Renard et al. 2001). However, we found no significant groupings of alleles that would indicate divergence of the duplicate genes, and thus we were unable to further attribute alleles to individual loci.

The three loci examined in this study differed in variability at both the nucleotide and amino acid levels (Table 2). As found in many other mammal species, little sequence variation was found at the *DRA* locus. The sequence difference between alleles varied from zero to at most 15 nucleotide substitutions ($\bar{x} = 3.69$, SE = 0.97) between species. Compared with the *DRA*, the variability of the *DQB* is much higher ($\bar{x} = 12.61$, SE = 1.83), with the absolute sequence difference between alleles up to 12.28% within species and 14.62% between species at the nucleotide level. The level of genetic variation at the *DQB* locus in cetacean was greater than those of some other marine mammal species. For example, in the four species of pinnipeds, including southern elephant seal *Mirounga leonina*, northern elephant seal *Mirounga angustirostris*, Antarctic fur seal *Arctocephalus gazella*, and New Zealand fur seal *Arctocephalus fosteri*, the absolute sequence difference was up to 8.5% within and between species (Hoelzel et al. 1999). Further comparisons of the extent of *DQB* exon 2 nucleotide divergences resulted in comparable values between cetacean species and some terrestrial mammal species/populations. Four species were chosen for comparison based on availability of data, including human (*Homo sapiens*), dog (*Canis familiaris*), pig (*S. scrofa*) and cow (*B. taurus*) (see Table 5). We found the sequence difference between alleles varied from zero to at most 35 nucleotide substitutions among the four species. The mean nucleotide variability was separately $\bar{x} = 13.63$ (SE = 2.25), $\bar{x} = 13.32$ (SE = 2.04), $\bar{x} = 13.72$ (SE = 1.95), and $\bar{x} = 11.08$ (SE = 1.70) in the dog, pig, cow, and human *DQB* sequences. This result was contradicted by the earlier study which demonstrated limited MHC polymorphism in whales (Trowsdale et al. 1989), possibly due to the small panel studied of the earlier study. By contrast with *DRA* and *DQB*, MHC-I was most variable in the cetacean species. Nucleotide sequence variation among pairwise comparisons of all cetacean MHC-I sequences ranged from 0 to 23.13% ($\bar{x} = 16.44$, SE = 2.08), leading to an average sequence difference of 11.33 amino acids (about 23.12% divergence). These values were in the range of observed polymorphism of MHC class I polymorphic genes in artiodactyls (al-Murrani et al. 1994; Velten et al. 1999; Renard et al. 2001).

Positive Selection on *DQB* and MHC-I Sequences

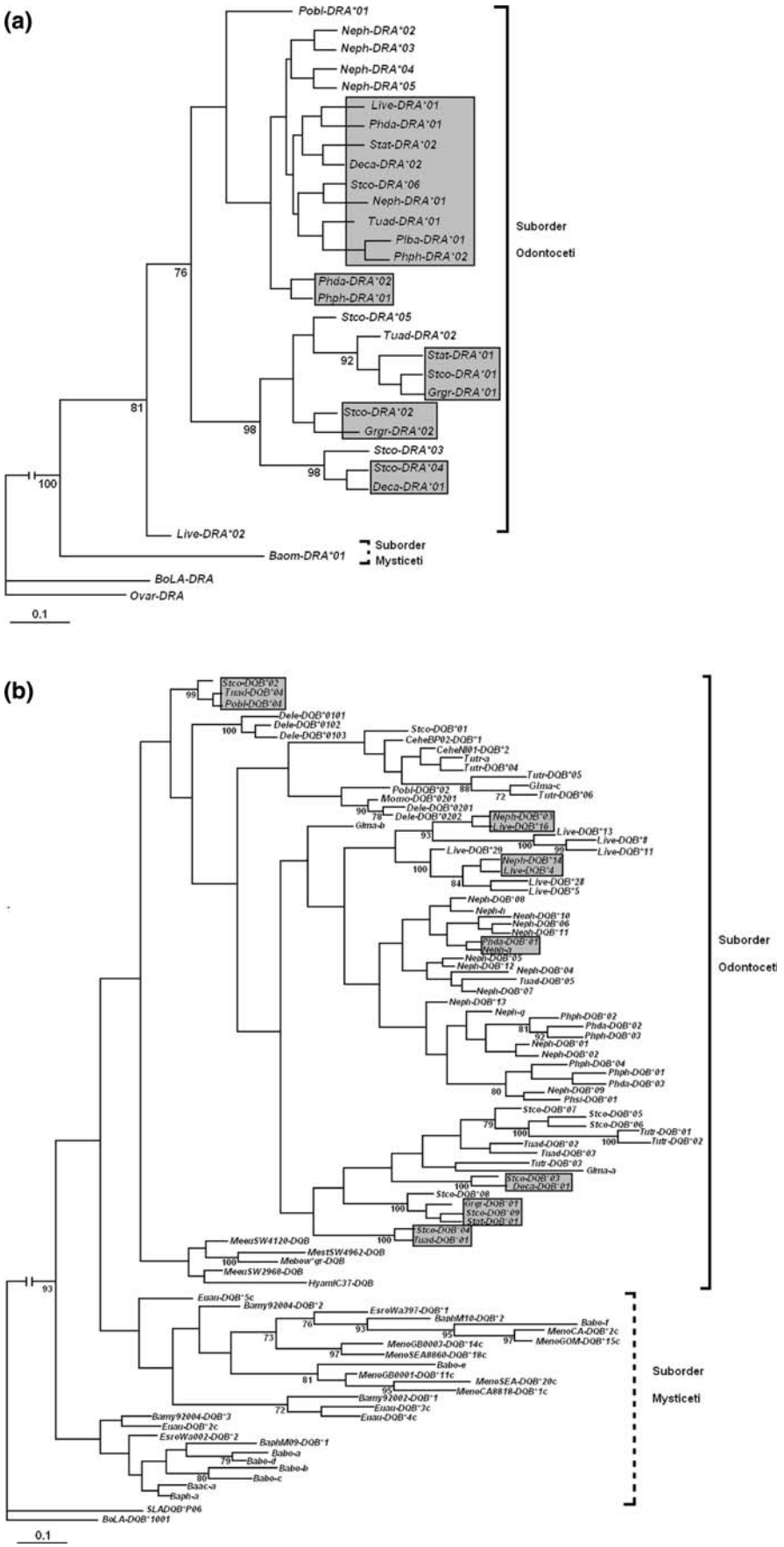
We detected clear signs of balancing selection acting on the assayed fragment of the *DQB* and MHC-I loci. Maximum likelihood models allowing selection fitted the data significantly better than models that considered only neutral or conserved sites at the *DQB* and MHC-I loci. The selected sites identified by Bayesian analysis coincided largely with sites that aligned with PBR in human sequences and which represent most of the variability among the sequences. For the *DQB* locus, seven sites were identified with a strong signal of positive selection separately by model M2a and M8 (Table 4) from which four correspond to potential PBRs (28N, 30N, 37Y, 71K) according to the comparison with the human *DQB* sequences (Brown et al. 1993). For the MHC-I locus, six sites were significantly positively selected under the model M2a and all of these sites corresponded to amino acid residues lying in the PBR. Model M8 included all the sites identified by M2a; M8 added one more sites (52V) which corresponded to a conserved residue with the human MHC-I sequences (Bjorkman et al. 1987). These facts suggested that sequence diversity in the *DQB* and MHC-I genes may be, to a large extent, generated by positive selection on PBRs. Furthermore, the balancing selection was also found from the pattern of trans-species evolution at the *DQB* and MHC-I loci (see below discussion on the

genetic mechanisms for the identical alleles among different species for details).

However, in the case of the *DRA* locus, a non-significant test for positive selection was examined using PAML software, which provided no evidence for positive selection acting on this locus. In addition, the synonymous substitutions is greater than nonsynonymous substitutions in the PBR (data not shown), suggesting that purifying selection acting on the *DRA* gene. Purifying selection refers to selection against nonsynonymous substitutions at the DNA level (Kimura 1977). In such a case, the evolutionary distance based on synonymous substitutions is expected to be greater than that based on nonsynonymous substitutions. This is known to be true of other mammals such as human (*H. sapiens*), mouse (*Mus musculus*), and California sea lion (*Zalophus californianus*) (Pimtanothai et al. 2001; Janitz et al. 1998; Bowen et al. 2004). It is suggested that different MHC loci were under different degrees of selection pressure, which led to distinct polymorphisms among loci from highly conservative (*DRA*) to considerably variable (*DQB* and MHC-I).

To test whether sites inferred under positive selection in cetacean species largely correspond to the sites found under selection in other mammals, or whether these sites are mostly distinct, we mainly analyzed the *DQB* and MHC-I homology sequences of four terrestrial mammal species, i.e. human, cow, dog, and pig. Each dataset was analyzed using the same method at the *DQB* and MHC-loci, respectively. In every dataset, similar tendency to the cetacean species, models M2a and M8, which allow for sites under positive selection, fitted the data significantly better than models M1a and M7, which do not allow for such sites (data not shown). Positive selection sites identified under model M2a and M8 from data of different species are summarized in Table 5. At the *DQB* locus, of the seven sites inferred to be under positive selection in the cetacean dataset, six of them are also detected in at least one other species. Similar pattern was also detected at the MHC-I locus. Sites inferred to be under positive selection in the cetacean dataset were also similar to those detected in other terrestrial mammal species at the MHC-I locus. These shared sites detected to be under positive selection among species suggested that the similarity in the selective pressure on amino acid residues across species in both the *DQB* and MHC-I sequences. This result was in agreement with recent summaries of diversifying selection on the highly polymorphic *DRB* gene in five mammal species by Furlong and Yang (2008). Thus, we could conclude that MHC class I and II genes perform more or less the same function of antigen binding in different species despite differences in their habitats/pathogens. However, the number of sites under positive selection was different in theses species. For example, at the *DQB* locus, five and six sites were detected at a high posterior probability (>0.99) under model M8 in the porcine and canine dataset, respectively, while only three and two sites were separately found in the bovine and human datasets. On the other hand, sites inferred to be under positive selection were remarkably similar to that estimate, given the small size of their dataset. Hence, the number of sites detected in each species is variable, possibly reflecting the size and information content of the datasets as suggested by Furlong and Yang (2008). In the current analysis, only the part of exon 2 sequences (<200 bp) was included. Therefore, more MHC loci and longer MHC sequence are needed to analyze in the future.

Fig. 2 Phylogenetic trees reconstructed using Bayesian inference method in MrBayes 3.12, based on *DRA*, *DQB* and MHC-I nucleotide sequences. Posterior probabilities greater than 70% are given for each clade. The identical alleles were marked in the frames. The Odontoceti and Mysticeti clades are separately indicated by the *straight line* and *broken line*. **a** *DRA*. In addition to 21 *DRA* sequences identified in this study, the following sequences were also included: *N. phocaenoides* (*Neph-DRA*01-05*: DQ843609-DQ843613); *L. vexillifer* (*Live-DRA*01-02*: DQ851844-DQ851846); *B. taurus* (*BoLA-DRA*: M30120); *O. aries* (*Ovar-DRA*: M73983). **b** *DQB*. In addition to 26 *DQB* sequences identified in this study, the following sequences were also included: *N. phocaenoides* (*Neph-DQB*01-14*: DQ843614-DQ843623 and EF056477-EF056480; *Neph-a*: AB164212; *Neph-g*: AB164218; *Neph-h*: AB164219); *T. truncatus* (*Tutr-a*: AB164221; *Tutr-DQB*01-06*: EF690293- EF690298); *L. vexillifer* (*Live-DQB*4*: AY177153; *Live-DQB*5*: AY177283; *Live-DQB*8*: AY177286; *Live-DQB*11*: AY177289; *Live-DQB*13*: AY177291; *Live-DQB*16*: AY333383; *Live-DQB*28*: AY333395; *Live-DQB*29*: AY333396); *B. bonaerensis* (*Babo-a-f*: AB164202-AB164207); *B. acutorostrata* (*Baac-a*: AB164201); *B. mysticetus* (*Bamy92002-DQB*1-3*: DQ354623-DQ354625); *B. physalus* (*Baph-a*: AB164199; *BaphM09-DQB*1*: Q354626; *BaphM10-DQB*2*: DQ354627); *C. hectori* (*CeheBP02-DQB*1*: DQ354628; *CeheNI01-DQB*2*: DQ354629); *E. australis* (*Euau-DQB*2c-5c*: DQ354631-DQ354634); *E. robustus* (*EsroWa397-DQB*1*: Q354635; *EsroWa002-DQB*2*: DQ354636); *H. ampullatus* (*HyamIC37-DQB*: DQ354637); *M. stejnegeri* (*MestSW4962-DQB*: DQ354638); M. grayi (*Mebow*gr-DQB*: DQ354639); *M. europaeus* (*MeeuSW2968-DQB*: DQ354640; *MeeuSW4120-DQB*: DQ354641); *M. novaeangliae* (*MenoCA8818-DQB*1c*: DQ354642; *MenoCA-DQB*2c*: DQ354643; *MenoGB0001-DQB*11c*: DQ354652; *MenoCA8818-DQB*13c*: DQ354654; *MenoGB0003-DQB*14c*: DQ354655; *MenoGOM-DQB*15c*: DQ354656; *MenoSEA8860-DQB*18c*: DQ354659; *MenoSEA-DQB*20c*: DQ354661); *D. leucas* (*Dele-DQB*0101-0103*: U16986-U16988; *Dele-DQB*0201-0202*: U16989-U16990); *M. monoceros* (*Momo-DQB*0201*: U16991); *P. sinus* (*Phsi-DQB*01*: AY170897); *G. macrorhynchus* (*Glma-a-c*: AB164226-AB164228); *S. scrofa* (*SLADQB*P06*: AF272715); *B. taurus* (*BoLA-DQB*1001*: U62318). **c** MHC-I. In addition to 23 MHC-I sequences identified in this study, some sequences of other cetaceans were from GenBank: *N. phocaenoides* (*Neph-I*01-34*: DQ843624-DQ843657); *L. vexillifer* (*Live-I*01-06*: DQ851847-DQ851852); *E. robustus* (*Esro3.1.4*: AF149216; *Esro3.2.2*: AF149217; *Esro3.2.3*: AF149218; *Esro3.5.2*: AF149219; *Esro3.5.3*: AF149220; *Esro6.1.1*: AF149221; *Esro3.1.1*: AF149222; *Esro6.4.1*: AF149223; *Esro4.2*: AF188615; *Esro5.1*: AF188616; *Esro64.6*: AF188617; *Esro7.2*: AF188618; *Esro64.2*: AF361229); *B. indicus* (*BOVMHAWA*: M69206; *BOVMHBKNA*: M69204)


(a)
Pobl-DRA*01
Neph-DRA*02
Neph-DRA*03
Neph-DRA*04
Neph-DRA*05
Live-DRA*01
Phda-DRA*01
Stat-DRA*02
Deca-DRA*02
Stco-DRA*06
Neph-DRA*01
Tuad-DRA*01
Plba-DRA*01
Phph-DRA*02
Phda-DRA*02
Phph-DRA*01
Stco-DRA*05
Tuad-DRA*02
Stat-DRA*01
Stco-DRA*01
Grgr-DRA*01
Stco-DRA*02
Grgr-DRA*02
Stco-DRA*03
Stco-DRA*04
Deca-DRA*01
Live-DRA*02
Baom-DRA*01
BoLA-DRA
Ovar-DRA
Suborder
Odontoceti
Suborder
Mysticeti
76
81
98
92
98
100
0.1
(b)
Stco-DQB*02
Tuad-DQB*04
Pobl-DQB*04
Dele-DQB*0101
Dele-DQB*0102
Dele-DQB*0103
Stco-DQB*01
CeheBP02-DQB*1
CeheN01-DQB*2
Tutr-a
Tutr-DQB*04
Tutr-DQB*05
Glma-c
Tutr-DQB*06
Pobl-DQB*02
Momo-DQB*0201
Dele-DQB*0201
Dele-DQB*0202
Glma-b
Neph-DQB*03
Live-DQB*16
Live-DQB*13
Live-DQB*8
Live-DQB*11
Live-DQB*20
Neph-DQB*14
Live-DQB*4
Live-DQB*28
Live-DQB*5
Neph-DQB*08
Neph-h
Neph-DQB*10
Neph-DQB*06
Neph-DQB*11
Phda-DQB*01
Neph-a
Neph-DQB*05
Neph-DQB*12
Neph-DQB*04
Tuad-DQB*05
Neph-DQB*07
Neph-DQB*13
Neph-g
Phph-DQB*02
Phda-DQB*02
Phph-DQB*03
Neph-DQB*01
Neph-DQB*02
Phph-DQB*04
Phph-DQB*01
Phda-DQB*03
Neph-DQB*09
Phsi-DQB*01
Stco-DQB*07
Stco-DQB*05
Stco-DQB*06
Tutr-DQB*01
Tutr-DQB*02
Tuad-DQB*02
Tuad-DQB*03
Tutr-DQB*03
Glma-a
Stco-DQB*03
Deca-DQB*01
Stco-DQB*08
Grgr-DQB*01
Stco-DQB*09
Stat-DQB*01
Stco-DQB*04
Tuad-DQB*01
MeeuSW4120-DQB
MestSW4962-DQB
Mebow*gr-DQB
MeeuSW2968-DQB
HyamIC37-DQB
Euau-DQB*5c
Bamy92004-DQB*2
EsroWa397-DQB*1
BaphM10-DQB*2
Babo-f
MenoCA-DQB*2c
MenoGOM-DQB*15c
MenoGB0003-DQB*14c
MenoSEA8860-DQB*18c
Babo-e
MenoGB0001-DQB*11c
MenoSEA-DQB*20c
MenoCA8818-DQB*1c
Bamy92002-DQB*1
Euau-DQB*3c
Euau-DQB*4c
Bamy92004-DQB*3
Euau-DQB*2c
EsroWa002-DQB*2
BaphM09-DQB*1
Babo-a
Babo-d
Babo-b
Babo-c
Baac-a
Baph-a
SLADQB*P06
BoLA-DQB*1001
Suborder
Odontoceti
Suborder
Mysticeti
0.1

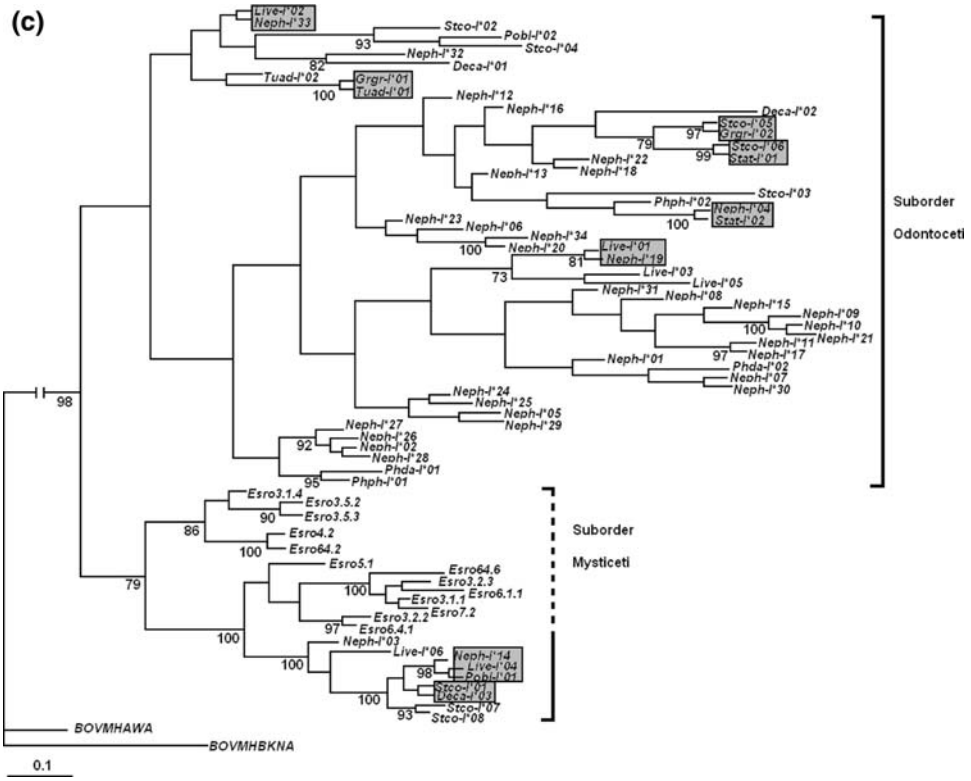


Fig. 2 continued

Genetic Mechanisms for Generation and Maintenance of Identical Alleles Between Species

Extensive sharing of both MHC allele lineages and identical MHC alleles were detected among the 28 species of cetacean examined (Fig. 2). A total of 20 pairs of identical alleles, i.e. five at *DRA*, seven at *DQB*, and eight at the MHC-I, were detected in multiple species (Fig. 2). Several hypotheses were proposed to explain these identical alleles among different species: (1) they can be PCR artifacts or sample contamination; (2) retention of ancestral alleles (i.e. trans-species evolution); (3) convergent evolution; or (4) the result of introgressive hybridization (introgression) between species.

Table 5 Summary statistics of positively selected sites identified in models M2a and M8 in the *DQB* and MHC-I exon 2 sequences of different species

Locus	Species	Positively selected sites	
		M2a (positive selection)	M8 (β & ω)
DQB	Cetacea ($N = 101$)	**26F**, **28N**, *29R*, **30N**, **37Y**, **57T**, **71K**	**26F**, **28N**, **29R**, **30N**, **37Y**, **57T**, **71K**
	Human ($N = 94$)	**26L**, **57D**	**26L**, **57D**
	Cow ($N = 45$)	**26Y**, **57D**, *67F*	**26Y**, **57D**, 67F
	Dog ($N = 36$)	*29K*, *30Y*, **37Y**, **57S**, **67E**	**29K**, **30Y**, **37Y**, *47Y*, **57S**, **67E**, **75L**
	Pig ($N = 43$)	**26G**, *28A*, *37H*, *57T*, *71K*	**26G**, **28A**, *30W*, **37H**, **57T**, **71K**, *75V*
MHC-I	Cetacea ($N = 76$)	**45K**, **67S**, **70A**, **73I**, **77N**, **80I**	**45K**, *52V*, **67S**, **70A**, **73I**, **77N**, **80I**
	Human ($N = 87$)	*66K*, **73T**	**66K**, **73T**
	Cow ($N = 56$)	*45E*, **67Y**, **73T**, **77Y**	**45E**, *62R*, *66I*, **67Y**, *69D*, *70T*, *71A*, **73T**, *76V*, **77Y**, *81L*
	Dog ($N = 48$)	**45M**, **62R**, **66T**, **67A**, **73R**, *77D*	**4M5**, **62R**, **66T**, **67A**, **73R**, *74Y*, **77D**
	Pig ($N = 44$)	**62R**, **66K**, **67Q**, **73T**, **74Y**, **75R**, **77G**, **79K**	**62R**, *63E*, **66K**, **67Q**, *70T*, **73T**, **74Y**, **75R**, **77G**, **79K**

Note: Positively selected sites were identified in models M2a and M8 by the Bayes empirical Bayes procedure (Yang et al. 2005c) using CodeML (included in the PAML 3.14 program suite). Sites inferred to be under selection at the 99% level are listed in bold; those at the 95% level are italicized. The residues are numbered according to the human reference sequence. *N* in the parentheses is the total number of *DQB* or MHC-I sequences in each species, and these sequences were downloaded from http://www.ebi.ac.uk/imgt/hla and http://www.ebi.ac.uk/ipd/mhc

The identical alleles were unlikely to be resulted from PCR artifacts or methodological errors. Firstly, we attempted to reamplify, clone and sequence those sharing sequences, and the same results were obtained. Secondly, these identical alleles were found not in one or two samples, but in a number of clones or samples. Two species in our data set, i.e. baiji and finless porpoise, were surveyed intensively as part of population-level analyses of MHC variation in these animals. For the *DRA* locus, *Live-DRA*01* is identical to *Neph-DRA*01*, the former of which was identified from 30 clones of 10 baiji samples, whereas the latter of which was detected in 94 clones of 35 finless porpoises. This sequence was also found in seven other cetacean species. Two pairs of identical *DQB* alleles were shared by 16 baiji individuals and four finless porpoises, respectively. Finally, one pair of identical *DQB* alleles, i.e. *Phda-DQB*01* & *Neph-a*, the former was identified in the Dall's porpoise examined in this study, and the latter was identified in the finless porpoise examined by Hayashi et al. (2003). Actually, mitochondrial control region of some species (e.g. baiji, finless porpoise, bottlenose dolphin, striped dolphin, etc.) were examined for the same DNA extraction, and no haplotype was shared between them (Yang et al. 2002, 2003, 2005b). For these reasons, we should exclude the possibility that the identity or similarity among cetacean species may be due to 'PCR artifacts' or sample contamination. Also, the great number of shared alleles in all loci examined and the common existence of shared alleles identified in so many species strongly excluded the possibility that these shared alleles evolved by chance.

To investigate whether the shared MHC alleles between different cetacean species were retention of ancestral alleles or adaptive convergence, phylogenetic trees were also reconstructed based on non-PBR sites of the three MHC loci (data not shown). A tree constructed using non-PBR sites should provide a more reasonable approximation of the true relationship among the investigated species. Compared with the trees constructed using the alleles of all-sites (PBR + non-PBR), the non-PBR sites trees were in the similar topology. In other words, interspecific allele sharing was also grouped together in these non-PBR trees (data not shown). The concordance between the all-sites and non-PBR trees provided evidence that allelic similarities/identity possibly reflected common ancestry and were not due to convergence. Future analysis of flanking sequence or intron regions may help to establish the extent to which trans-species evolution has shaped those identical alleles.

In addition, we also estimated the relative age of the three MHC loci using the relaxed molecular clock obtained from the primates. The relative age of polymorphism at a locus is given by the number of synonymous substitutions at non-PBR sites, since these sites are thought not to be subject to diversifying selection. Applying the estimated synonymous substitution rate of 1.2×10^{-9} million years (Klein et al. 1993) to the class II loci in this study, the polymorphism at the *DRA* locus was more recent (mean age of alleles 31.66 ± 15 million years) than at the more-variable beta loci (*DQB* 65 ± 28.4 million years). While for the MHC-I genes, the polymorphism arose approximately 56.94 ± 24.82 million years according to the synonymous substitution rate of 1.37×10^{-9} million years to the class I loci estimated by Satta et al. (1993). According to the relaxed molecular clock based on mtDNA data, estimates for dates of divergence between species of toothed whales varied from 33.52 million years to 35.43 million years (Xiong et al. 2009). Although it might be inappropriate to directly compare these evolutionary time scales because they were generated from different models, the very old age of allelic diversity of three MHC loci examined strongly implied that the generation of cetacean MHC diversities through trans-species evolution predated the divergence of toothed whales. This was in line with the hypothesis that the upper limit for taxa to share alleles was estimated to be 30–40 million years (Takahata and Nei 1990), and Klein et al.'s (1993) opinion that the positive selection pressure was responsible for the trans-species persistence of allelic lineages at functional MHC loci for more than 40 million years.

Finally, the existence of introgressive hybridization (or introgression) between different species is another possibility that cannot be entirely ruled out. Cetaceans may have the potential to produce viable hybrid offspring more easily than other mammals, since it is unusual for this animal group to display prominent karyological uniformity (Amaral et al. 2007). In captivity, some cases of hybridization strongly support this hypothesis. For example, a bottlenose dolphin and a rough-toothed dolphin (*Steno bredanensis*) produced hybrid offspring (http://www.messybeast.com/genetics/hybrid-marine-mammals.htm), another case from a captive-born intergeneric hybrid of bottlenose dolphin and common dolphin (Zornetzer and Duffield 2003). Although it is more difficult to directly identify hybrids in the wild, some cases have been suspected to support field introgression with assistance from molecular markers, such as mtDNA sequences. For instance, Dall's porpoises and harbour porpoises have hybridized in the wild (Baird et al. 1998). Another reported hybrid in the field was found between blue whale (*B. musculus*) and fin whale (Bérubé and Aguilar 1998). Thus, the identical alleles shared in related species also might be the result of introgressive hybridization. It is needed to further investigate with more field evidences and other molecular markers to find out whether these species have some hidden hybridized offspring.

Acknowledgments This research was financially supported by the key project of National Natural Science Foundation of China (NSFC) grant no. 30830016, NSFC project grant no. 30670294, the Program for New Century Excellent Talents in University (grant no. NCET-07-0445), the Ministry of Education of China, the Specialized Research Fund for the Doctoral Program of Higher Education (grant no. SRFDP 20060319002), the Ministry of Education of China, and the major project of the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (grant no. 07KJA18016). We thank Dr. Anli Gao, Mr. Xin-Rong Xu, Mr. Hua Chen, and Dr. Qing Chang for their assistance and contribution to collecting samples for many years.

References

al-Murrani SW, Glass EJ, Hopkins J (1994) BoLA class I charge heterogeneity reflects the expression of more than two loci. Anim Genet 25:165–172

Amaral R, Sequeira M, Martínez-Cedeira J, Coelho MM (2007) New insights on population genetic structure of *Delphinus delphis* from the northeast Atlantic and phylogenetic relationships within the genus inferred from two mitochondrial markers. Mar Biol 151:1967–1976

Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236

Baird RW, Willis PM, Guenther TJ, Wilson PJ, White BN (1998) An intergeneric hybrid in the family Phocoenidae. Can J Zool Rev Can Zool 76:198–204

Baker CS, Vant MD, Dalebout ML, Lento GM, O'Brien SJ, Yuhki N (2006) Diversity and duplication of *DQB* and *DRB*-like genes of the MHC in baleen whales (suborder: Mysticeti). Immunogenetics 58:283–296

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: What have we learned about natural selection in 15 years? J Evol Biol 16:363–377

Bérubé M, Aguilar A (1998) A new hybrid between a blue whale, *Balaenoptera musculus*, and a fin whale, *B. physalus*: frequency and implications of hybridization. Mar Mamm Sci 14:82–98

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) Structure of the human class I histocompatibility antigen, HLA-A2. Nature 329:506–512

Bowen L, Aldridge BM, Gulland F, Van Bonn W, DeLong R, Melin S, Lowenstine LJ, Sott JF, Johnson ML (2004) Class II multiformity generated by variable MHC-DRB region configurations in the California sea lion (*Zalophus californianus*). Immunogenetics 56:12–27

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) 3-Dimensional structure of the human class-II histocompatibility antigen HLA-DR1. Nature 364:33–39

Cutrera AP, Lacey EA (2007) Trans-species polymorphism and evidence of selection on class II MHC loci in tuco-tucos (Rodentia: Ctenomyidae). Immunogenetics 59:937–948

Dengjel J, Schoor O, Fischer R, Reich M, Kraus M, Muller M, Kreymborg K, Altenberend F, Brandenburg J, Kalbacher H, Brock R, Driessen C, Rammensee HG, Stevanovic S (2005) Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc Natl Acad Sci USA 102:7922–7927

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

Flores-Ramirez S, Miller RD, Urban-Ramirez J (2004) Major histocompatibility. Complex I polymorphism in a cetacean: the gray whale (*Eschrichtius robustus*). Mar Mamm Sci 20:262–273

Furlong RF, Yang ZH (2008) Diversifying and purifying selection in the peptide binding region of DRB in mammals. J Mol Evol 66:384–394

Garrigan D, Hedrick PW (2003) Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evolution 57:1707–1722

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Gustafsson K, Andersson L (1994) Structure and polymorphism of horse MHC class II *DRB* genes: convergent evolution in the antigen binding site. Immunogenetics 39:355–358

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

Hayashi K, Yoshida H, Nishida S, Goto M, Pastene LA, Kanda N, Baba Y, Koike H (2006) Genetic variation of the MHC DQB locus in the finless porpoise (*Neophocaena phocaenoides*). Zool Sci 23:147–153

Hoelzel AR, Stephens JC, O'Brien SJ (1999) Molecular genetic diversity and evolution at the MHC DQB locus in four species of pinnipeds. Mol Biol Evol 16:611–618

Huchard E, Cowlishaw G, Raymond M, Weill M, Knapp LA (2006) Molecular study of *Mhc-DRB* in wild chacma baboons reveals high variability and evidence for trans-species inheritance. Immunogenetics 58:805–816

Janitz M, Reiners-Schramm L, Lauster R (1998) Expression of the H2-Ea gene is modulated by a polymorphic transcriptional enhancer. Immunogenetics 48:266–272

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276

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

Klein J, Horejsi V (1997) Immunology, 2nd edn. Blackwell Science, Oxford

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

Klein J, Sato A, Nagl S, O'Huigin C (1998) Molecular trans-species polymorphism. Annu Rev Ecol Syst 29:1–21

Kriener K, O'hUigin C, Tichy H, Klein J (2000) Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. Immunogenetics 51:169–178

Munguia-Vega A, Esquer-Garrigos Y, Rojas-Bracho L, Vazquez-Juarez R, Castro-Prieto A, Flores-Ramirez S (2007) Genetic drift vs. natural selection in a long-term small isolated population: major histocompatibility complex class II variation in the Gulf of California endemic porpoise (*Phocoena sinus*). Mol Ecol 16:4051–4065

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 (*Delphinapterus leucas*). Mol Biol Evol 12:582–593

Penn DJ, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. Am Nat 153(2):145–164

Pimtanothai N, Hurley CK, Leke R, Klitz W, Johnson AH (2001) HLA-DR and -DQ polymorphism in Cameroon. Tissue Antigens 58:1–8

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

Radwan J, Kawalko A, Wojcik JM, Babik W (2007) MHC-DRB3 variation in a free-living population of the European bison, *Bison bonasus*. Mol Ecol 16:531–540

Renard C, Vaiman M, Chiannilkulchai N, Cattolico L, Robert C, Charwn P (2001) Sequence of the pig major histocompatibility region containing the classical class I genes. Immunogenetics 53:490–500

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

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

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

Suárez CF, Patarroyo ME, Trujillo E, Estupiñán M, Baquero JE, Parra C, Rodriguez R (2006) Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. Immunogenetics 58:542–558

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

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Trowsdale M (1996) Molecular genetics of HLA class I and class II regions. In: Browning MB, McMichael A (eds) HLA and MHC: genes, molecules and function. BIOS Scientific Publishers, Oxford, pp 23–36, 329–351

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

Velten FW, Renard C, Rogel-Gaillard C, Vaiman M, Chardon P (1999) Spatial arrangement of pig MHC class I sequences. Immunogenetics 49:919–930

Xiong Y, Brandley MC, Xu SX, Zhou KY, Yang G (2009) Seven new dolphin mitochondrial genomes and a time-calibrated phylogeny of whales. BMC Evol Biol 9(1):20

Xu SX, Sun P, Zhou KY, Yang G (2007) Sequence variability at three MHC loci of finless porpoises (*Neophocaena phocaenoides*). Immunogenetics 59:581–592

Xu SX, Chen BY, Zhou KY, Yang G (2008) High sequence similarity at three MHC loci between baiji and finless porpoise: trans-species or convergent evolution? Mol Phylogenet Evol 47:36–44

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, Nielsen R, Goldman N, Pedersen AMK (2000) Codon substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Yang G, Ren WH, Niu MH, Zhou KY (2002) Variability of the complete mitochondrial control region of striped dolphins (*Stenella coeruleoalba*). Acta Zool Sin 48:131–134

Yang G, Liu S, Ren WH, Zhou KY, Wei FW (2003) Mitochondrial control region variability of baiji and the Yangtze finless porpoises, two sympatric small cetaceans in the Yangtze River. Acta Theriol 48:469–483

Yang G, Yan J, Zhou KY, Wei FW (2005a) Sequence variation and gene duplication at MHC DQB loci of Baiji (*Lipotes vexillifer*), a Chinese river dolphin. J Hered 96:310–317

Yang G, Ji GQ, Ren WH, Zhou KY, Wei FW (2005b) Pattern of genetic variation of bottlenose dolphins in Chinese waters. Raffles B Zool 53:189–196

Yang Z, Wong WSW, Nielsen R (2005c) Bayes Empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Yang WC, Hu JM, Chou LS (2008) Sequence variation of MHC class II DQB gene in bottlenose dolphin (*Tursiops truncatus*) from Taiwanese waters. Taiwania 53:42–50

Zornetzer HR, Duffield DA (2003) Captive-born bottlenose dolphin × common dolphin (*Tursiops truncatus* × *Delphinus capensis*) intergeneric hybrids. Can J Zool 81:1755–1762