Allelic polymorphism, gene duplication and balancing selection of the MHC class II *DAB* gene of *Cynoglossus semilaevis* (Cynoglossidae)

T.J. Xu*, Y.N. Sun* and R.X. Wang

Key Laboratory for Marine Living Resources and Molecular Engineering,
College of Marine Science, Zhejiang Ocean University, Zhoushan, P.R. China

*These authors contributed equally to this study.
Corresponding author: R.X. Wang
E-mail: wangrixin1123@126.com

ABSTRACT. Major histocompatibility complex (MHC) genes play an important role in the immune response of vertebrates. Allelic polymorphism and evolutionary mechanism of MHC genes have been investigated in many mammals, but much less is known in teleosts. We examined the polymorphism, gene duplication and balancing selection of the MHC class II *DAB* gene of the half-smooth tongue sole (*Cynoglossus semilaevis*); 23 alleles were found in this species. Gene duplication manifested as three to six distinct sequences at each domain in the same individuals. Non-synonymous substitutions occurred at a significantly higher frequency than synonymous substitutions in the PBR domain, suggesting balancing selection for maintaining polymorphisms at the MHC II *DAB* locus. Many positive selection sites were found to act very intensely on antigen-binding sites of MHC class II *DAB* gene.

Key words: Major histocompatibility complex; *Cynoglossus semilaevis*; Allelic polymorphism; Balancing selection

INTRODUCTION

The major histocompatibility complex (MHC) is a genomic region that encodes cellular proteins that are involved in the immune response in vertebrates. The main functions of MHC encode cell-surface glycoproteins responsible for binding foreign peptides for the presentation proteins to T-cell receptors (TCR) in the adaptive immune system, and thereby triggering a specific immune response against the pathogen from which the peptides are derived. The MHC class I molecule, consisting of one alpha chain and β2-microglobuin, presents foreign peptide products by the degradation of intracellular pathogens and presenting them to cytotoxic CD8[+] T cells. The MHC class II molecule, consisting of one alpha chain and one beta chain, presents foreign peptides derived from extracellular pathogens to helper CD4[+] T cells. Surprisingly, unlike mammals and other vertebrates, the class I and class II genes were found to reside on different linkage groups in teleosts (Graser et al., 1996). Thus, some scientists believe that the major histocompatibility (MH) genes are more appropriate designation in teleosts (Stet et al., 2002).

The hallmark of MHC genes is the extremely high degree of polymorphism, which results from the presence of multiple loci and numerous alleles at each given locus within populations (Xu SX et al., 2009; Archie et al., 2010; Goüy and Leirs, 2010). The high polymorphism found in the MHC class II genes is generally confined to exon 2, which encodes the peptide-binding region (PBR) in the α1- or β1-domain. Three major types of mechanisms can explain the evolution and maintenance of allelic polymorphism in MHC genes within vertebrate populations (Hedrick, 2002; Aguilar and Garza, 2007). There are several lines of evidence that polymorphic MHC loci are subject to balancing selection (see review in Hughes and Yeager, 1998a; Ekblom et al., 2010; Eimes et al., 2010; Karaiskou et al., 2010), and the pressure exerted by pathogens and parasites is considered to be a major component of balancing selection, influencing MHC polymorphism (Hedrick, 1998, 2002; Stear et al., 2005). None of the other selection pressures were proposed for the maintenance of MHC polymorphism (see review in Apanius et al., 1997). Thus, operating evolutionary mechanisms are of great interest (Hedrick 1998). High rates of non-synonymous (d_N) to synonymous (d_S) substitutions are regarded as good indicators for an ongoing adaptive evolutionary process (Parham and Ohta, 1996; Axtner and Sommer, 2007). Moreover, it becomes more and more evident that not only allelic diversity within one single locus is important, but also gene duplication seems to be an important evolutionary tool to increase allelic diversity within a species (Axtner and Sommer, 2007). Gene conversion plays an important role in the generation of new alleles and the evolution of MHC genes (Hughes and Yeager, 1998b). MHC loci have sometimes duplicated independently in different mammalian groups, and half of the loci are functional (Hughes and Nei, 1990). Gene duplication plays a major role in the adaptive evolution of organisms (Hughes, 1994). BS, the "birth and death" evolution model, produces in mammals new orthologous genes by duplication, which diverge by accumulating mutations over time. Some of these alleles persist in the genome, and some get deleted or become pseudogenes (Nei and Rooney, 2005).

Due to its important role in the vertebrate immune system, most studies of the MHC are primarily interested in the extent of polymorphism within loci. Understanding the evolution of MHC genes in lower vertebrates may provide insights into the origin of the MHC. The half-smooth tongue sole (*Cynoglossus semilaevis*) is an important economic marine fish in China, and its evolution has attracted the attention of some scientists, where three genes of

the MHC gene family have been reported (Xu TJ et al., 2009; Xu et al., 2010). To explore the ongoing evolutionary mechanisms that determine adaptive MHC polymorphism in the half-smooth tongue sole, we decided to analyze the polymorphism of exon 2 of the MHC II class *DAB* gene, which encodes the PBR in the α1-domain and is the most polymorphic region of the MHC class II *DAB* gene. Many methods can be applied to the characterization of MHC polymorphism; we chose sequence analysis to study the polymorphism of the MHC II *DAB* gene, because it is a precise technique for investigating nucleotide variations. Our specific goals were to investigate the polymorphism of the MHC II *DAB* gene and to analyze the evolution pattern and gene duplication.

MATERIAL AND METHODS

Animals and DNA extraction

Specimens of half-smooth tongue sole were collected in the Zhoushan fishing ground of Zhejiang province along the East Sea coast in China. Fin tissue samples were taken and stored in 95% ethanol. Genomic DNA was extracted from fin samples with the phenol-chloroform method as descried elsewhere (Xu et al., 2008). The quality and concentration of DNA were assessed by agarose gel electrophoresis and measured with a GeneQuant Pro (Pharmacia Biotech Ltd.) RNA/DNA spectrophotometer. Finally, DNA was normalized to a concentration of 100 ng/μL for subsequent polymerase chain reaction (PCR) amplifications.

PCR procedures

Genetic investigations focused on exon 2 of MHC class II *DAB* gene, because it contains most of the PBR loci and, therefore, shows the most likely chance for detecting balancing selection acting on the MHC class II gene (Axtner and Sommer, 2007). We analyzed 29 individuals using PCR with specific primers. To isolate the complete exon 2 of the MHC class II *DAB* gene, specific primers (Cyse-*DAB*-F: 5'-TTCCTCCTCACGGTCCACA-3' and Cyse-*DAB*-R: 5'-TGATTTAGCCAAACCAGTC-3') were designed according to the complete genomic sequence of the MHC class II *DAB* gene as described by Xu TJ et al. (2009). All PCRs were carried out in a final volume of 50 μL, each containing 5.0 μL 10X *Taq* polymerase buffer, 0.2 mM dNTP$_S$, 0.2 μM of the forward and reverse primers, 2 U *Taq* plus polymerase and 1 μL DNA template. Cycling conditions were 94°C for 4 min followed by 30 cycles of 94°C for 40 s, 52°C for 40 s and 72°C for 2 min to reduce artifact formation, ending with 1 cycle of 72°C for 10 min; PCR was performed on a PTC-200. The PCR products were electrophoresed on a 1% agarose gel to check integrity and visualized by the Molecular Imager Gel Doc XR system (Biorad).

Cloning and sequencing

PCR products were purified using the QIAEX II Gel Extraction Kit (Qiagen). The purified fragments were ligated into PMD18-T vectors (Takara) and cloned to TOP10 cells according to the standard protocol. Positive clones were screened via PCR with M13+/- primers. Amplicons from 5-6 positive clones per individual were sequenced using the ABI

3730 automated sequencer with M13+/- primer. The number of clones was sufficient to determine the genotype of each individual and to avoid reading errors or recombinant sequences generated during PCR amplifications.

Sequence analysis

The sequences were edited and aligned using Clustal W multiple alignment. We have previously reported 15 alleles of the MHC class II *DAB* gene (Xu TJ et al., 2009), and here, the new alleles were identified and designated according to accepted nomenclature rules (Davies et al., 1997). Polymorphic values were determined using the DnaSP v5.0 analysis software and the DAMBE software, and nucleotide diversity was estimated with a sliding window of 100 bp using DnaSP v5.0. To test for recombination events between alleles, we used the DnaSP software to determine the recombination loci. The MEGA software was used to calculate the rate of d_S and d_N, according to the method of Nei and Gojobori (1986), and the phylogenetic trees were constructed using nucleotide sequences via the neighbor-joining method.

Detecting positive selection

Evidence for positive selection was assessed using the likelihood ratio model in the CODEML program, which is contained in the PAML 3.14 program suite. We used maximum likelihood models to examine heterogeneity in ω (d_N/d_S; Goldman and Yang, 1994) among site partitions. For these analyses, we employed the random-site models, assuming several heterogeneous sites with different ω parameter without *a priori* knowledge of the class (neutral, purifying, or positive selection) to which a given codon belongs. The models implemented in this study are M0, M1a, M2a, M3, M7, and M8 (Yang et al., 2005). The likelihood ratio tests were used to compare model M0 with M3, M1a with M2a, and M7 with M8, respectively. When the alternative models M2a, M3 and M8 suggest the presence of codons with $\omega > 1$, this can be considered evidence of positive selection. Posterior probabilities for site classes have been calculated by the Bayes empirical Bayes (BEB) method in M2a and M8 (Yang et al., 2005). The posterior means of ω for some site classes are >1 (calculated is the average of ω over all site classes weighted by the posterior probabilities), and those sites are likely to be under positive selection (Yang et al., 2005).

RESULTS

Polymorphism of MHC class II *DAB*

Of the 29 individuals, 20 Cyse-*DAB* alleles were defined, and eight of these alleles were novel and designated Cyse-*DAB**0103, Cyse-*DAB**0104, Cyse-*DAB**0105, Cyse-*DAB**1001, Cyse-*DAB**1101, Cyse-*DAB**1201, Cyse-*DAB**1202, and Cyse-*DAB**1301 (GenBank accession No. GU186865-GU186872), according to accepted nomenclature rules (Figure 1). Fifteen alleles (GenBank accession No. FJ372722, GU186873-GU186886), which we have previously reported, were also used in this study. A fragment of 270 bp of the complete exon 2 of the MHC class II *DAB* gene, corresponding to the α1-domain of the protein was sequenced and aligned with class II *DAB* alleles described in previous studies (Xu TJ et al., 2009). The

number of alleles per individual varied from 2 to 6, as expected from the duplication gene locus, and more than two alleles were detected for most of the individuals, suggesting that at least three loci of the MHC class II *DAB* gene exist in the half-smooth tongue sole.

```
              1          11         21         31         41         51         61         71         81
Cyse-DAB*0101 GATGGATTTC TAAATTTTAG GACAGTCACC TGTCTGTTCA ACTCTACAAA AGCCGATGAT ATCGCGTACA TTGACTCCTA TTACCTCAAC
Cyse-DAB*0102 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0103 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0104 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0105 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0201 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0202 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0301 .......... ..T...C..T ......G... .......... .......... ......C... .......... ....G..... ....T.....
Cyse-DAB*0401 .......... ..T...C..T ......G... .......... .......... ......C... .......T.. ....G..... ....T.....
Cyse-DAB*0402 .......... ..T...C..T ......G... .......... .......... ......C... .......T.. ....G..... ....T.....
Cyse-DAB*0501 .......... ..C...C..T ......G... .......... .......... ......C... .......... ....G..... ....T.....
Cyse-DAB*0502 .......... ..C...C..T ......G... .......... .......... ......C... .......... ....G..... ....T.....
Cyse-DAB*0601 ..G......G AGG.A.A.GA AGTG.A..GA ...CG....T .T..C..TG. .C.G..C... ...CA..... ..T.....C. C..TTA....
Cyse-DAB*0701 ..G......G AGG.A.A..C AGTG.A..GA ...CG....T .T..C..TG. .C.G..C... ...CA..... ..T.....C. C..TTA....
Cyse-DAB*0801 ..G......G AGG.A.A.GC AGTG.A..GA ...CG....T .T..C..T.. .C.G..C... ...CA..... ..T.....C. C..TTA....
Cyse-DAB*0901 ..G......G AGG.A.A.GC AGTG.A..GA ...CG....T .T..C..T.. .C.G..C... ...CA..... ..T.....C. C..TTA....
Cyse-DAB*0902 ..G......G AGG.A.A.GC AGTG.A..GA ...CG....T .T..C..T.. .C.G..C... ...CA..... ..T.....C. C..TTA....
Cyse-DAB*0903 ..G......G AGG.A.A.GC AGTG.A..GA ...CG....T .A..C..T.. .C.G..C... ...CA..... ..T.....C. C..TTA....
Cyse-DAB*1001 .......... ..C....... ......G... .......... .......... ......C... .......... ....G..... ..........
Cyse-DAB*1101 .......... ..C....... ......G... .......... .......... .......... .......... .......... ..........
Cyse-DAB*1201 .......... ...G..C.GT ....AC.... .......... .......... ......C... .......... ..C.G..... ..........
Cyse-DAB*1202 .......... ...G..C.GT ....AC.... .......... .......... ......C... .......... ..C.G..... ..........
Cyse-DAB*1301 .......... ..T...C.GT ......G... .......... .......... .......... .......... ....G...C. ..........
Recombination           a          a                                      c                    c        e
Recombination                    b  b                                                        d  d        e

              91         101        111        121        131        141        151        161        171
Cyse-DAB*0101 AAACTGAAGA TCATATCGTT CGACAGCAGA GTGGGGAAAT TTGTGGGATA CACTGAGTTT GGAGTGTCCA ACGCTAAGAG ATGGAATGAA
Cyse-DAB*0102 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0103 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0104 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0105 .......... .......... .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0201 ........T. .......... .......... .......... A......... ......A... ........A. .......... ..........
Cyse-DAB*0202 ....A...T. .......... .......... .......... A......... ......A... ........A. .......... ..........
Cyse-DAB*0301 ........T. .......... .......... .......... .......... ......A... .......... .......... ..........
Cyse-DAB*0401 ........T. .G........ .......... .......... A......... .......... .......... .......... ..........
Cyse-DAB*0402 ........T. .G........ .......... .......... A......... .......... .......... .......... ..........
Cyse-DAB*0501 ........T. .......... .......... .......... A......... .......... .......... .......... ..........
Cyse-DAB*0502 ........T. .......... .......... .......... A......... .......... .......... .......... ..........
Cyse-DAB*0601 ..GAG.G.AT ATG.CAG... .......... .......... .......... ......A... .......... .......... ..........
Cyse-DAB*0701 ..GAG.G.AT AT..CAG... .......... ......AG.. A......... ......A... ....A..... .......... ..........
Cyse-DAB*0801 ..GAG.G.AT AT..CAG... .......... ......AG.. A......... ......A... .......... .......... ..T.....
Cyse-DAB*0901 ..GAG.G.AT ATG.CAG... .......... ......AG.. A......... ......A... .......... .......... ..........
Cyse-DAB*0902 ..GAG.G.AT ATG.CAG... .......... ......AG.. A......... ......A... .......... .......... ..........
Cyse-DAB*0903 ..GAG.G.AT ATG.CAG... .......... ......AG.. A......... ......A... .......... .......... ..T.....
Cyse-DAB*1001 .......... .......... .......... .......... .......... .......... .......... .......... ..T.....
Cyse-DAB*1101 .......... .......... .......... .......... .......... ........A. .......... .......... ..........
Cyse-DAB*1201 ........T. .......... .......... .......... .......... ........A. .......... .......... ..........
Cyse-DAB*1202 ........T. .......... .......... .......... .......... ........A. .......... .......... ..........
Cyse-DAB*1301 ........T. .......... .......... .......... .......... ........A. .......... .......... ..........
Recombination             f                          f        g                g                   i
Recombination                                                           h                       h

              181        191        201        211        221        231        241        251        261
Cyse-DAB*0101 GGTCCAGAAG TAACCCGAGC GAGGAATGAG AAGGAGAGGT ACTGTGTACA CAACGTCGGA CTGGACTACA GGACTGGTTT GGCTAAATCA
Cyse-DAB*0102 .......... .......... .......... .......... .......... .......... .......... ...C...... ..........
Cyse-DAB*0103 .......... .......... ......G... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0104 .......... .........A .......... .......... .......... .......... .......... .......... ..........
Cyse-DAB*0105 .......... .......AG ....C..... .......... ......GA.. .......... .......... .......... ..........
Cyse-DAB*0201 .......... ...T..A..T .......... .......... ......GA.. .C....... A.TT..A... ACG...C.C. ..A.......
Cyse-DAB*0202 .......... ...T..A..T .......... .......... ......GA.. .C....... A.TT..A... ACG...C.C. ..A.......
Cyse-DAB*0301 .......... ...T..A.AG .......... .......... ......GA.. .......... AAT...... ACCA..C.G. ..A.......
Cyse-DAB*0401 .......... .......... .......... .......... ......GA.. .......... A.T...... ACG...C.C. ..A.......
Cyse-DAB*0402 .......... .......... .......... .......... ......GA.. .......... A.T...... ACG...C.C. ..A.......
Cyse-DAB*0501 .......... ...T..A.AG ....C.AC.. .......... ......GA.. .......... AAT...... ACCA..C.G. ..A.......
Cyse-DAB*0502 .......... ...T..A.AG ....C.AC.. .......... ......GG.. .......... AAT...... ACCA..C.G. ..A.......
Cyse-DAB*0601 .......... ...T..A.AG ....C..... .........T .......... .......... A.T...... ACG...C.C. ..A.......
Cyse-DAB*0701 .......... ...T..A.AG ....C..C.. .......... ......GA.. .C....... A.TT..A... ACG...C.C. ..A.......
Cyse-DAB*0801 .......... ...T..A.AG .......... .......... ......GA.. .C....... G.TT..A... ACG...C.C. ..A.......
Cyse-DAB*0901 .......... ...T..A.AG ....C..C.. .......... .........T .......... A.T...... ACG...C.C. ..A.......
Cyse-DAB*0902 .......... ...T..A..G ....C..C.. .......... .........T .......... A.T...... ACG...C.C. ..A.......
Cyse-DAB*0903 .......... ...T..A.AG ....C..C.. .......... .........T .......... A.T...... ACG...C.C. ..A.......
Cyse-DAB*1001 .......... .......AG ....C..... .......... ......GA.. .......... .......... .......... ..........
Cyse-DAB*1101 .......... ...T..A.A. .......... .......... ......GA.. .......... .......... .......... ..........
Cyse-DAB*1201 .......... ...T..A.AG T......... .......... .......G.. .......... .......... .......... ..........
Cyse-DAB*1202 .......... ......A.AG T......... .......... .......G.. .......... .......... .......... ..........
Cyse-DAB*1301 .......... ...T..A.AG .......... .......... ......GA.. .......... .......... .......... ..........
Recombination   i  j j              l              l
Recombination      k              k            m            m
```

Figure 1. Allele sequences of the MHC class II *DAB* gene of the half-smooth tongue sole. Dots indicate identity with the top sequence. Regions that have been shown to be involved in recombination events as revealed by DnaSP are noted with identical letters at the beginning and end of the region.

The full alignment of the 270 bp exon 2 of MHC class II *DAB* gene did not show any gaps, and none of the sequences contained insertions/deletions or stop codons, suggesting that all sequences may come from functional molecules in the genome. The nucleotide sequence corresponded to a putative peptide of 90 amino acids. There were 81 variable sites in exon 2, of which 81 parsimony informative sites were observed. Most sequences consisted of highly conserved motifs with a number of variable residues concentrated in particular sites (Figures 1, 2). The putative amino acid residues involved in antigen binding in the half-smooth tongue sole were based on the corresponding antigen-binding sites identified in humans (Brown et al., 1993), and indicated that the polymorphic sites in the fish alleles differed only slightly from those in human (Figure 2). Analysis of all the 23 alleles showed that 30.0% of the nucleotides were variable and that 45.7% of the amino acids were variable. Within the polymorphic α1-domain of the 23 alleles, the pairwise nucleotide distance ranged from 0.004 to 0.258 with a mean of 0.145, and the pairwise distance of amino acid sequences ranged from 0.011 to 0.511 with a mean of 0.258. In addition, the neighbor-joining phylogenetic tree revealed that the alleles of the half-smooth tongue sole constituted a distinct clade, separated from the flounder (Paol-*DAB**0101, EU380332) and turbot (Scma, DQ001730) (Figure 3). Phylogenetic analysis of the sequences revealed two distinct allele clusters (clusters I and II), and subsequently, polymorphism analyses were also performed on each cluster separately. The polymorphic value was determined using the DnaSP v5.0 software. To further characterize patterns of polymorphism, analyses were carried out on whole sequences and each cluster separately. Table 1 summarizes the measures of polymorphism in both all sequences and each cluster. The average number of nucleotide diversity Pi (π) and Theta-W value (per site) of all sequences of exon 2 were 0.1259 and 0.0813, respectively, and the average number of nucleotide differences (k) was 34.00. The exon 2 of MHC II *DAB* revealed high nucleotide diversity. The spatial distribution of nucleotide diversity is shown in Figure 4. There were two peaks upstream located between 13 and 103 bp, and there was lower nucleotide diversity in the middle of exon 2.

Figure 2. Putative amino acid sequences of alleles for MHC class II *DAB* exon 2 of the half-smooth tongue sole. Dots indicate identity with the top sequence. Asterisks indicate the codons involved in the antigen-binding region. Arrows indicate the limits of the β-sheet region and the α-helix region.
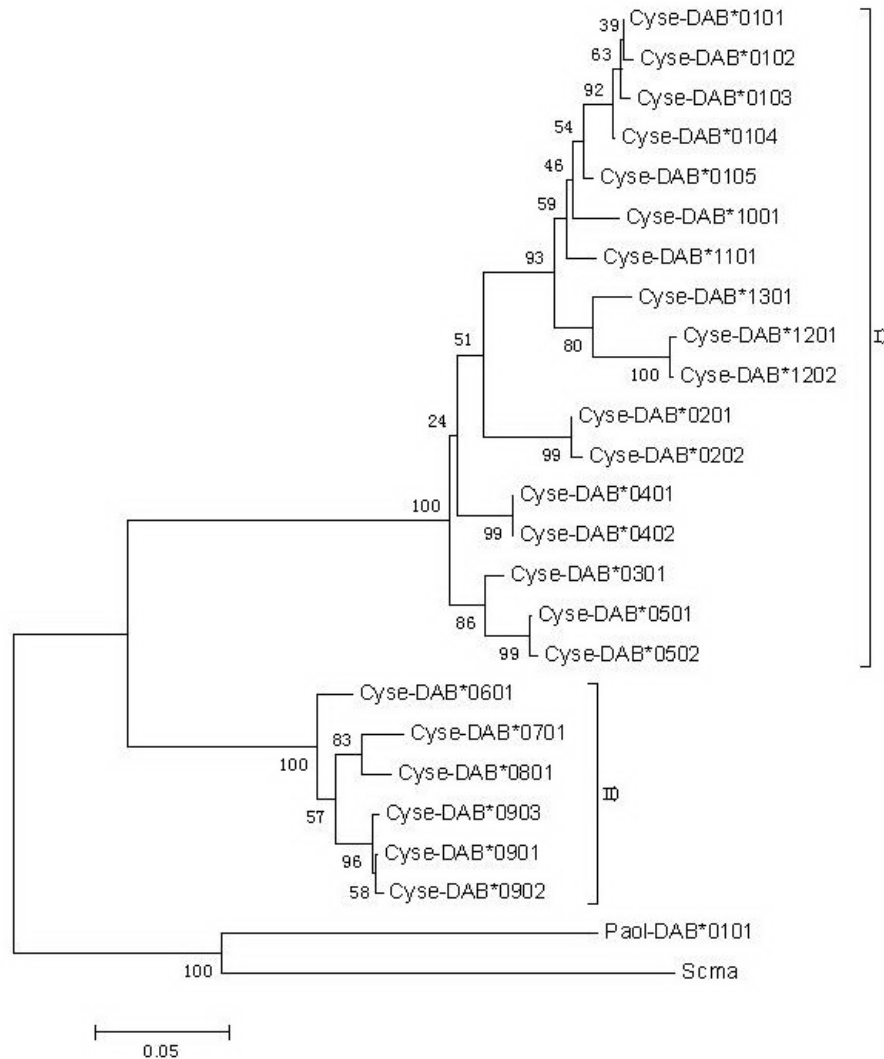
Figure 3. Phylogenetic tree of exon 2 of the MHC class II *DAB* gene in the half-smooth tongue sole, flounder, and turbot. Bootstrap values are as shown in branch.

Table 1. Polymorphism statistics in the exon 2 of half-smooth tongue sole MHC class II *DAB* gene.

Locus	No.	S	k	Pi	Theta-W (per site)	Theta-W (per sequence)	Diversity	
							nt	AA
All alleles	23	81	34.00	0.1259	0.0813	21.946	0.145 (0.004-0.320)	0.258 (0.011-0.511)
Cluster I	17	45	17.68	0.0655	0.0493	13.311	0.069 (0.004-0.129)	0.159 (0.011-0.266)
Cluster II	6	19	7.933	0.0294	0.0308	9.821	0.030 (0.004-0.054)	0.082 (0.011-0.143)

S = number of variable nucleotide sites; k = the average number of nucleotide difference; Pi = the average number of nucleotide diversity.
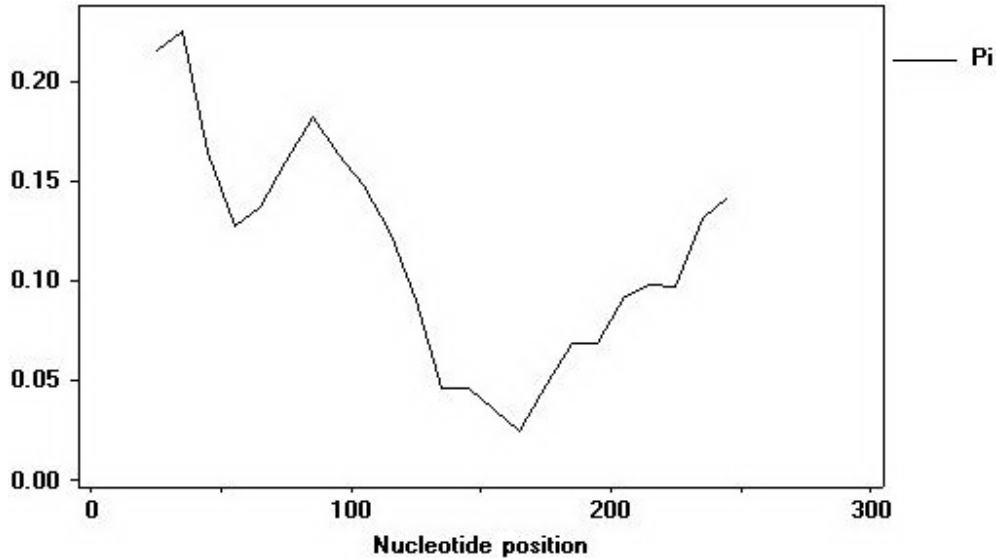
Figure 4. The nucleotide diversity within exon 2 sequences of MHC class II *DAB* genes at the alleles denoted by Pi (π). Sliding window length: 100; step size: 10.

Patterns of positive selection

The rates of d_S and d_N, nucleotide substitutions and their ratio ($\omega = d_N/d_S$) for the PBR and non-PBR of the MHC class II *DAB* sequence are presented in Table 2. This high level of non-synonymous substitutions leads in both clusters I and II to an ω significantly higher than the one in analysis over all sites. In all sequences, the rate of d_N was 2.083 times higher than the d_S in the non-PBR. Within clusters I and II, there were differences between the ω at the PBR and the non-PBR, both also showing strong positive selection pressure. For cluster II, d_S at the PBR was 0, thus, impossible to test the ω value.

Table 2. Synonymous (d_S) and nonsynonymous (d_N) substitution ratio in the putative peptide-binding region (PBR) and non-peptide-binding region (non-PBR) with standard errors (SE) obtained through 1000 bootstrap replicates in parentheses, and results of the Z-test of positive selection.

Locus	All domain				PBR				Non-PBR			
	d_N (SE)	d_S (SE)	ω	Z (P)	d_N (SE)	d_S (SE)	ω	Z (P)	d_N (SE)	d_S (SE)	ω	Z (P)
Exon 2												
All sequence	0.025	0.012	2.083	0.214	0.070	0.003	23.333	0.004	0.011	0.015	0.733	0.737
	(0.007)	(0.008)			(0.068)	(0.003)			(0.005)	(0.010)		
Cluster I	0.083	0.026	3.192	0.005	0.249	0.021	11.857	$1*10^{-5}$	0.033	0.027	1.222	0.766
	(0.017)	(0.013)			(0.055)	(0.016)			(0.011)	(0.016)		
Cluster II	0.035	0.014	2.500	0.107	0.104	0.000	∞	$2*10^{-4}$	0.014	0.019	0.737	0.719
	(0.009)	(0.010)			(0.027)	(0.000)			(0.006)	(0.015)		

Maximum likelihood-based random-site models that revealed positive selection fitted the data significantly better than models that did not incorporate selection: models M3 and M8 better than models M0 and M7 (Table 3). The random-site models demonstrated extreme vari-

ability in selective pressure among sites and the presence of a number of sites under positive selection. These positive selection sites were identified using the BEB procedure in the PAML software. In total, 10 codons were significantly positively selected under the model M2a, 32 codons under the model M3, and 15 codons under the model M8 (Table 4). These positively selected sites corresponded almost exactly to amino acid residues lying in the PBR (Figure 5).

Table 3. Summary of statistical test for the likelihood-ratio tests of exon 2 of the *DAB* gene in half-smooth tongue sole.

Model compared	d.f.	Statistical test	Significance
M1a *vs* M2a	2	40.482	P < 0.001
M3 *vs* M0	4	95.360	P < 0.001
M8 *vs* M7	2	40.814	P < 0.001

d.f. = degrees of freedom.

Table 4. Results of maximum likelihood models of exon 2 of the *DAB* gene.

Model	P	Log-likelihood	Parameter estimates	Positively selected sites
M0	1	-1123.419	$K = 1.403$, $\omega = 1.498$	None
M1a (nearly neutral)	1	-1098.226	$K = 1.108$, $p_0 = 0.463$, $p_1 = 0.536$, $\omega_0 = 0$, $\omega_1 = 1$	Not allowed
M2a (positive selection)	3	-1077.985	$K = 1.719$, $p_0 = 0.413$, $p_1 = 0.294$, $p_2 = 0.294$, $\omega_0 = 0$, $\omega_1 = 1$, $\omega_2 = 5.494$	**5, 7, 9,** *25, 50,* **67, 69,** *70, 71, 81*
M3 (discrete)	5	-1075.739	$K = 1.682$, $p_0 = 0.535$, $p_1 = 0.427$, $p_2 = 0.039$, $\omega_0 = 0.065$, $\omega_1 = 3.569$, $\omega_2 = 14.236$	**4, 5, 6, 7, 8, 9,** *10,* **17, 22,** **25, 27, 29,** **32,** *34, 36, 43,* **44, 50, 53, 58, 65, 66, 67, 69, 70, 77,** *78,* **81,** *82, 83,* **84, 85**
M7 (beta)	2	-1098.400	$K = 1.105$, $P = 0.0058$, $q = 0.0050$	Not allowed
M8 (beta and omega)	4	-1077.993	$K = 1.720$, $P_0 = 0.7012$, $p_1 = 0.2988$, $p = 0.050$, $q = 0.0075$, $\omega = 5.4365$	**5,** *6,* **7, 9,** *25,* **29,** *32, 44,* **50, 67, 69, 70, 77,** **81,** *85*

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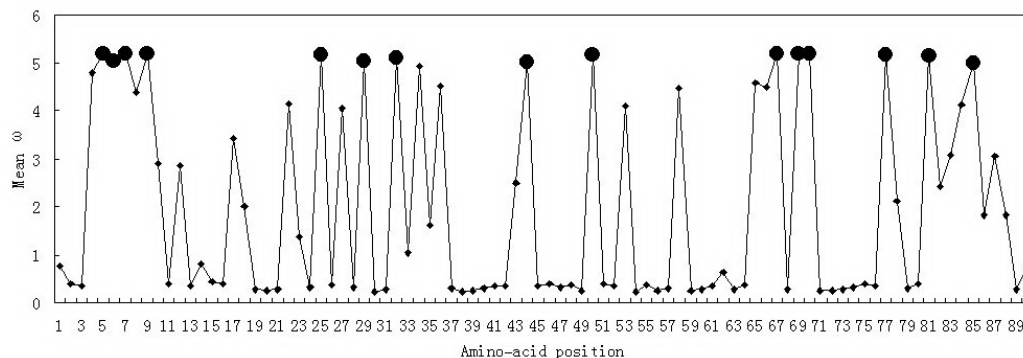

Figure 5. Posterior means of ω, calculated as the average of ω over the 11 site classes, weighted by the posterior probabilities under the random-site model M8 (β and ω). Sites inferred to be under positive selection are indicated by filled circles.

DISCUSSION

Because MHC genes are significant elements of adaptive immunity in vertebrates, they have attracted the attention of many scientists. High polymorphism is one of the most characteristic features of the MHC genes; allelic MHC molecules bind and present specificity peptides, and the response of an organism toward certain pathogens can be influenced by MHC alleles present in different haplotypes. Allelic polymorphisms of MHC genes have been investigated in many species (Graser et al., 1996; Axtner and Sommer, 2007). For example, 76 alleles in 60 flounder individuals (Xu et al., 2008), 26 alleles in 36 *Clethrionomys glareolus* individuals (Axtner and Sommer, 2007), 27 alleles in 146 *Apodemus flavicollis* individuals (Meyer-Lucht and Sommer, 2005), and 34 alleles in 40 *Gerbillurus paeba* individuals (Harf and Sommer, 2005) were found. However, a very low allelic diversity in some species was also demonstrated, for example, only 5 alleles in 100 individuals sampled from four populations of the water vole (Oliver and Piertney, 2006). To evaluate allelic variation and detect patterns of selection of MHC class II *DAB* gene in the half-smooth tongue sole, we report here many alleles of the MHC class II *DAB* gene. To our knowledge, this is the first study of the allelic polymorphism and balancing selection of exon 2 of the MHC class II *DAB* gene in the half-smooth tongue sole.

In the present study, of high interest is the fact that there is high polymorphism of the MHC class II *DAB* gene. Twenty alleles in 29 individuals were detected in this study; a total of 23 alleles were discovered in 33 individuals of half-smooth tongue sole, including findings of our previous study (Xu TJ et al., 2009). Most alleles differed at more than five amino acids. Concerning the number of MHC class II *DAB* genes, the present study revealed the presence of up to six different sequences in a single individual, suggesting that at least three loci of MHC II class *DAB* existed in half-smooth tongue sole. Some strong evidence of MHC gene duplication has been found in many species (Flores-Ramirez et al., 2000; Baker et al., 2006). In Japanese flounder, at least three duplication loci of the MHC class IIB gene were also found (Xu et al., 2008). Grimholt et al. (2003) showed that the presence of multiple loci is one of the reasons for the extremely high degree of MHC polymorphism. There are two major theories explaining the evolution of the MHC gene. One hypothesis is the birth and death model of evolution (Nei et al., 1997), which suggests that the MHC genes have been produced by duplications. Some genes diverge functionally through the accumulation of the mutations over a long period; some alleles become fixed, and some pseudogenes are deleted from the genome due to deleterious mutations. The second hypothesis is gene conversion between alleles of one or different loci leading to new alleles using segmental exchange (Parham and Ohta, 1996). In this study, we found events for recombination loci in the half-smooth tongue sole, suggesting that the observed diversity in sequences is probably generated by recombination/gene conversion.

Gene duplication of MHC genes is a common feature; duplicated MHC loci usually are governed by diversifying selection over shorter evolutionary time, which may favor their specialization (Hughes, 1999). The d_N/d_S test, a commonly used measure of selection, demonstrated that there is strong evidence of the PBR within *DAB* exon 2 in the half-smooth tongue sole. d_N/d_S is greater than one indicating balancing selection (Hughes and Nei, 1988, 1989), and thus, the MHC class *DAB* locus in the half-smooth tongue sole seems to be subject to a strong balancing selection. In this study, where the d_N/d_S ratio is much higher at PBR sites than at non-PBR sites, the differences are very significant (Table 2). However, different results for the MHC class II genes in some species are unexpected (Axtner and Sommer, 2007). The unexpected

results could be caused by the unknown detailed structure of MHC proteins in the majority of species studied and incorrectly identified PBR sites (Bryja et al., 2006). In this study, the putative amino acid residues involved in antigen binding were based on the corresponding antigen-binding sites identified in human HLA protein (Brown et al., 1993). The structure of the PBR sites is variable between different alleles at one locus, and thus, it is very important to identify the PBR sites under various types of selection using maximum likelihood models. The results of this study reveal that the observed diversity in the MHC class II *DAB* gene may be, to a large extent, generated by positive selection on PBR sites. The effect of positive, balancing selection may increase in regions involved in recombination. The positive selection sites identified by Bayesian analysis coincided largely with sites that compared with PBR sites in humans. Fifteen sites were identified with a strong signal of positive selection by model M8 (Table 2; Figure 5) of which nine corresponded to potential PBR sites (5N, 7R, 25D, 29L, 44F, 67A, 77H, 81L, 85T). Model M2a included all the 10 positive selection sites completely identified by M8, but M3 added 17 more sites than model M8, nine of them potentially the PBR sites (27Y, 34I, 53S, 58W, 65T, 70E, 78N, 82D, 84R). The fact that sites were identified as PBR sites using maximum likelihood models suggests that they are implicated in peptide binding in this species.

ACKNOWLEDGMENTS

REFERENCES

Aguilar A and Garza JC (2007). Patterns of historical balancing selection on the salmonid major histocompatibility complex class II beta gene. *J. Mol. Evol.* 65: 34-43.

Apanius V, Penn D, Slev PR, Ruff LR, et al. (1997). The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* 17: 179-224.

Archie EA, Henry T, Maldonado JE, Moss CJ, et al. (2010). Major histocompatibility complex variation and evolution at a single, expressed DQA locus in two genera of elephants. *Immunogenetics* 62: 85-100.

Axtner J and Sommer S (2007). Gene duplication, allelic diversity, selection processes and adaptive value of MHC class II DRB genes of the bank vole, *Clethrionomys glareolus*. *Immunogenetics* 59: 417-426.

Baker CS, Vant MD, Dalebout ML, Lento GM, et al. (2006). Diversity and duplication of DQB and DRB-like genes of the MHC in baleen whales (suborder: Mysticeti). *Immunogenetics* 58: 283-296.

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, et al. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 364: 33-39.

Bryja J, Galan M, Charbonnel N and Cosson JF (2006). Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). *Immunogenetics* 58: 191-202.

Davies CJ, Andersson L, Ellis SA and Hensen EJ (1997). Nomenclature for factors of the BoLA system, report of the ISAG BoLA Nomenclature Committee. *Anim. Genet.* 28: 159-168.

Eimes JA, Bollmer JL, Dunn PO, Whittingham LA, et al. (2010). *Mhc* class II diversity and balancing selection in greater prairie-chickens. *Genetica* 138: 265-271.

Ekblom R, Saether SA, Fiske P, Kalas JA, et al. (2010). Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. *Genetica* 138: 453-461.

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

Goldman N and Yang Z (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol.*

Biol. Evol. 11: 725-736.

Goüy de BJ and Leirs H (2010). Polymorphism and signatures of selection in the multimammate rat DQB gene. *Immunogenetics* 62: 59-64.

Graser R, O'huigin C, Vincek V, Meyer A, et al. (1996). Trans-species polymorphism of class II Mhc loci in danio fishes. *Immunogenetics* 44: 36-48.

Grimholt U, Larsen S, Nordmo R, Midtlyng P, et al. (2003). MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics* 55: 210-219.

Harf R and Sommer S (2005). Association between major histocompatibility complex class II DRB alleles and parasite load in the hairy-footed gerbil, *Gerbillurus paeba*, in the southern Kalahari. *Mol. Ecol.* 14: 85-91.

Hedrick PW (1998). Balancing selection and MHC. *Genetica* 104: 207-214.

Hedrick PW (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution* 56: 1902-1908.

Hughes AL (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* 256: 119-124.

Hughes AL (1999). Adaptive Evolution of Genes and Genomes. Oxford University Press, New York.

Hughes AL and Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170.

Hughes AL and Nei M (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. U. S. A.* 86: 958-962.

Hughes AL and Nei M (1990). Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. *Mol. Biol. Evol.* 7: 491-514.

Hughes AL and Yeager M (1998a). Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* 32: 415-435.

Hughes AL and Yeager M (1998b). Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* 32: 415-435.

Karaiskou N, Moran P, Georgitsakis G and Abatzopoulos TJ (2010). High allelic variation of MHC class II alpha antigen and the role of selection in wild and cultured *Sparus aurata* populations. *Hydrobiologia* 638: 11-20.

Meyer-Lucht Y and Sommer S (2005). MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Mol. Ecol.* 14: 2233-2243.

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

Nei M and Rooney AP (2005). Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39: 121-152.

Nei M, Gu X and Sitnikova T (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. U. S. A.* 94: 7799-7806.

Oliver MK and Piertney SB (2006). Isolation and characterization of a MHC class II DRB locus in the European water vole (*Arvicola terrestris*). *Immunogenetics* 58: 390-395.

Parham P and Ohta T (1996). Population biology of antigen presentation by MHC class I molecules. *Science* 272: 67-74.

Stear MJ, Innocent GT and Buitkamp J (2005). The evolution and maintenance of polymorphism in the major histocompatibility complex. *Vet. Immunol. Immunopathol.* 108: 53-57.

Stet RJM, Vries B, Mudde K, Hermsem T, et al. (2002). Unique haplotypes of co-segregating major histocompatibility complex class II A and class II B alleles in Atlantic salmon *Salmo salar* give rise to diverse class II genotypes. *Immunogenetics* 54: 320-331.

Xu SX, Ren WH, Li SZ, Wei FW, et al. (2009). Sequence polymorphism and evolution of three cetacean MHC genes. *J. Mol. Evol.* 69: 260-275.

Xu TJ, Chen SL, Ji XS and Tian YS (2008). MHC polymorphism and disease resistance to *Vibrio anguillarum* in 12 selective Japanese flounder (*Paralichthys olivaceus*) families. *Fish Shellfish Immunol.* 25: 213-221.

Xu TJ, Chen SL, Ji XS and Sha ZX (2009). Molecular cloning, genomic structure, polymorphism and expression analysis of major histocompatibility complex class IIA and IIB genes of half-smooth tongue sole (*Cynoglossus semilaevis*). *Fish Shellfish Immunol.* 27: 192-201.

Xu TJ, Sha ZX and Chen SL (2010). Unexpected variations of beta2-microglobulin gene in the half-smooth tongue sole. *Fish Shellfish Immunol.* 28: 212-215.

Yang Z, Wong WS and Nielsen R (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22: 1107-1118.