

RESEARCH ARTICLE

Open Access



# Polymorphism of MHC class IIB in an acheilognathid species, *Rhodeus sinensis* shaped by historical selection and recombination

Hyung-Bae Jeon<sup>1,2</sup>, Hari Won<sup>1</sup> and Ho Young Suk<sup>1\*</sup> 

## Abstract

**Background:** *Rhodeus sinensis* is a bitterling species occurring throughout the numerous freshwater systems on the East Asia. Here, we analyzed the diversity of the MHC class IIB (*DAB*) genes from this species, which may offer meaningful insights into evolutionary processes in this species as well as other bitterlings.

**Results:** Using cDNA and gDNA samples from 50 individuals, we discovered classical 140 allelic sequences that could be allocated into either *DAB1* (*Rhsi-DAB1*) or *DAB3* (*Rhsi-DAB3*). *DAB* sequences completely lacking the intron, but identical or similar to *Rhsi-DAB1*, were also discovered from our gDNA samples, and this intron loss likely originated from the retrotransposition events of processed mRNA. The  $\beta 1$  domain was the most polymorphic in both *Rhsi-DAB1* and *-DAB3*. Putative peptide binding residues (PBRs) in *Rhsi-DAB1*, but not in *Rhsi-DAB3*, exhibited a significant  $d_N/d_S$ , presumably indicating that different selection pressures have acted on those two *DABs*. Recombination between different alleles seemed to have contributed to the increase of diversity in *Rhsi-DABs*. Upon phylogenetic analysis, *Rhsi-DAB1* and *-DAB3* formed independent clusters. Several alleles from other species of Cypriniformes were embedded in the clade of *Rhsi-DAB1*, whereas *Rhsi-DAB3* clustered with alleles from the wider range of taxa (Cyprinodontiformes), indicating that these two *Rhsi-DABs* have taken different historical paths.

**Conclusions:** A great deal of MHC class IIB allelic diversity was found in *R. sinensis*, and gene duplication, selection and recombination may have contributed to this diversity. Based on our data, it is presumed that such historical processes have commonly or differently acted on the polymorphism of *Rhsi-DAB1* and *-DAB3*.

**Keywords:** *Rhodeus sinensis*, Major histocompatibility complex, Trans-species polymorphism, Balancing selection, Bitterlings, Intronless

## Background

The major histocompatibility complex (MHC) is a set of genes encoding cell membrane glycoproteins responsible for the initiation of adaptive immune response by displaying antigenic peptide to T lymphocytes in vertebrates [1]. MHC genes are classified into class I and II; class II genes are only expressed on professional antigen-presenting cells (e.g., dendritic cells, macrophages, B lymphocytes), whereas class I genes are expressed on all nucleated cells [1, 2]. A

peptide fragment loaded onto an MHC class II molecule is derived from antigens endocytosed, digested within lysosomes and presented to the specific receptors on the surface of CD4<sup>+</sup> helper T cells [1, 2]. MHC class II molecules are heterodimers consisting of two noncovalently associated homogeneous  $\alpha$  ( $\alpha 1$  and  $\alpha 2$  domains) and  $\beta$  chains ( $\beta 1$  and  $\beta 2$  domains; [1, 2]). A peptide fragment loaded onto an MHC class I molecule is derived from cytosolic proteins of infected cells, and is presented to specific receptors on the surface of CD8<sup>+</sup> cytotoxic T cells [1, 2]. MHC class I molecules are also heterodimers composed of three  $\alpha$  domains and  $\beta_2$ -microglobulin [1, 2].

\* Correspondence: [hsuk@ynu.ac.kr](mailto:hsuk@ynu.ac.kr)

<sup>1</sup>Department of Life Sciences, Yeungnam University, 280 Daehak-ro, Gyeongsan, Gyeongsangbuk-do 38541, South Korea

Full list of author information is available at the end of the article



Classical MHC class II genes are known to be among the most variable in vertebrate genomes [3–5]. The greatest polymorphism can be found in the  $\beta$ 1 domain, where peptide-binding residues (PBRs) are located [3, 4, 6]. In contrast, the level of polymorphism in the  $\alpha$ 1 domain is relatively low, with a few exceptions [7, 8], even though this domain also plays a role in binding with the antigenic peptide. Three major evolutionary forces are known to contribute to the enormous levels of polymorphism observed in the  $\beta$ 1 domain [9]: (i) negative frequency-dependent selection [10–12], (ii) heterozygote advantage [13, 14] and (iii) preferences for MHC dissimilar mates [15–17]. Another well-known evolutionary signature of MHC genes is trans-species polymorphism (TSP), which refers to genetic variants whose origin predates speciation, resulting in the occurrence of shared or similar alleles between different, but related, taxa [18–22]. The existence of TSP also means that there must be common alleles to secure survival or high adaptability even in different species under a specific environment [23]. To learn if TSP exist in a family or order and to infer the related evolutionary factors, however, information about a very certain phylogenetic structure among species in the taxa should be available.

Since the first analysis was attempted in carp [24], MHC genes have been characterized in a wide variety of teleost species [3, 25–28]. Teleost MHC class II can be divided into three major groups, namely, *DA*, *DB* and *DE*, based on their sequence features and phylogenetic clustering patterns [28]. Classical MHC class II genes are only found in *DA* (*DAA*: MHC II $\alpha$  chain; *DAB*: MHC II $\beta$  chain), whereas *DB* and *DE* generally comprise non-classical MHC genes [28, 29]. Therefore, MHC class II genes found in *DA* show tremendous polymorphisms among individuals, and conservative residues that are thought to form hydrogen bonds with antigen peptides [28]. These features do not perfectly appear in MHC class II genes belonging to *DB* and *DE* [28].

The primary aim of this study was to identify the signature of evolutionary forces that have acted on the MHC class IIB (*DAB*) sequences of Korean *Rhodeus sinensis*, one of the most widely distributed bitterling species (Acheilognathidae). Analyzing MHC sequences can offer meaningful insights into evolutionary processes in this or other bitterling species. First, because bitterlings spawn on freshwater mussels, which serve as intermediate host of many infectious organisms in freshwater ecosystems, response to pathogens and immunity may have played an important role in evolutionary processes in this species [30–32]. Second, bitterlings are a representative fish group in which the diversity and evolutionary patterns of *DAB* genes have not been properly characterized. In fact, only a partial investigation has been conducted in *R. ocellatus* [33] and *Pseudorhodeus tanago* [34]. Third, the population sizes of several bitterling species have been declining or at

the brink of extinction, due to the introduction of exotic species, climate change and devastation of many natural habitats [35–37], which serves as a good opportunity to investigate how size changes in the population and the resulting genetic drift affects MHC allelic diversity. Finally, *R. sinensis* is found in a variety of rivers with very different environmental features [38–40], providing an excellent opportunity to study the differences in selection pressures acting on MHC genes.

This study consisted of three stages. First, the nearly-complete sequences of *DAB* genes were identified from 50 individuals of *R. sinensis* collected from five different drainages, and the structural and functional characteristics were examined. Second, individual and inter-locus variabilities were examined to discover a signature of evolutionary processes acting on *DAB* diversities in this species. Finally, phylogenetic analyses were conducted to infer the evolutionary history of *DAB* genes in this species upon comparison with other vertebrates.

## Results

### Structure and diversity of MHC class IIB

Among primers used (Table 1), only a single pair, SP-F1 and TM-R1, successfully amplified all 50 individual cDNA samples. This primer pair was designed to anneal the sequences of signal peptide (exon 1) and transmembrane region (exon 6). The intronic sequences linking the six exons were identified from gDNA isolated from 20 Nakdong River samples. The amplification with the SP-F1 and TM-R1 yielded 293 sequences, and a total of 140 novel *DAB* alleles of *Rhodeus sinensis* were detected. Based on the blasting search and comparison with *DAB* sequences of other cypriniform species, the alleles were allocated into either *DAB1* (*Rhsi-DAB1*;  $N = 104$ ) or *DAB3* (*Rhsi-DAB3*;  $N = 36$ ; Additional file 8: Figure S1 and Additional file 9: Figure S2). All allelic sequences identified in this study were deposited to NCBI GenBank with the accession numbers of MG989278 to MG989423.

The sequence of exon 2 was confirmed to be  $\beta$ 1 domain region with PBRs through cDNA and gDNA sequencing (Additional file 8: Figure S1 and Additional file 9: Figure S2). Several conserved residues were found and may be related to the function of  $\beta$ 1 domain region based on the comparison with mammalian classical *DRB* structure [42]. For example, H81 and N82 are predicted to form hydrogen bonds with antigen peptides (Additional file 8: Figure S1 and Additional file 9: Figure S2). N38, S39 and T40 seem to be responsible for *N*-linked glycosylation, and two conserved cysteine residues, 11C and 76C were thought to form disulfide bridges (Additional file 8: Figure S1 and Additional file 9: Figure S2). In addition, N30 and N59 are residues that are found without exception in all jawed vertebrates, and G46 and Y47 are known to be ray-finned fish specific ([28]; Additional file 8: Figure

**Table 1** PCR primers used for the amplification of DAB sequences in *Rhodeus sinensis*. Data comprise primer name, direction (D), sequence, priming position (Position) and reference (Ref)

Name	D	Sequence (5' - 3')	Position	Ref
FishIIBEx2-1F	F	CTGATGCTGTCTGCTTTCACCTGGA	SP	1
FIBE-F1	F	CTGATGCTGTCTGCTTTCAC	SP	P
ROSI-R1	R	CGGGATCCCAGTCTACGATG	TM	P
ROSI-R2	R	GATCCCAGTCTACGATGATG	TM	P
ROSI-R3	R	GAGTGATTCTAGGTGAGAGTGG	exon3 3' end	P
ROSI-R4	R	CAGGTGCGAGTGAATCTGGT	exon3 middle	P
ROSI-R5	R	CTGAGCTTGACCTTTGGTGC	exon3 5' end	P
Rooc-Fw	F	CCCATAGTTGACATGATGTCATCTG	intron1	2
Rooc-Rev	R	CATGTGTGACAGGAGGATCAG	intron2	2
DAB1-Rev374	R	CAAGAGTTTCCCGTGTGACAG	intron2	2
DAB1-Fw18	F	AGTTGACATGATGCCATCTGA	intron1	2
DAB-Ex1-Fw	F	CATCCATACTGATGCTGTCTGC	SP	2
Rhod-GSP2	R	CCAGTCTCCGTTAGGCATCTCC	exon3	2
SP-F1	F	CTGATGCTGTCKRCWTTYACYGGA	SP	P
TM-R1	R	TGGTACCAGGATCCTCCC	TM	P
TM-R2	R	TGGKACCAGGATYCTCCC	TM	P
MHC2-TMr	R	TCAGTTTGGTACCAGGATCCTCC	TM	P
Exon-3R	R	TGAGCTTGACCTCTGGTGCCAC	exon3 5' end	P
THMHCI-3R	R	GCCAGCGTGATCCACCATAC	TM	P

(Ref) 1: Ottová et al., 2007 [41]; 2: Agbali et al., 2010 [33]; P: present study

S1 and Additional file 9: Figure S2). *Rhsi-DAB1\*04:02* and *Rhsi-DAB3\*06:01–03* showed a single codon insertion and deletion, respectively, in exon 2 (Additional file 8: Figure S1 and Additional file 9: Figure S2). From exon 3 to 6, overall amino acid sequences were highly conserved with no length variation (Additional file 10: Figure S3, Additional file 11: Figure S4, Additional file 12: Figure S5 and Additional file 13: Figure S6). Exon 3 contained a conserved amino acid motif (from 49 to 65) that seemed to be responsible for binding to CD4 molecule (Additional file 10: Figure S3 and Additional file 11: Figure S4). Two conserved cystein residues (C23 and C29) were predicted to form disulfide bridges in this domain (Additional file 10: Figure S3 and Additional file 11: Figure S4).

In our gDNA analysis, MHC class IIB sequences lacking introns (single exon gene, SEG) were discovered from all 20 individuals collected in the Nakdong River. These sequences were similar or completely identical (*Rhsi-DAB1\*01:09*, *-DAB1\*03:03*, *-DAB1\*03:04*, *-DAB1\*03:09*, *-DAB1\*03:10*, *-DAB1\*03:16*, *-DAB1\*05:03*, *-DAB1\*05:15* and *-DAB1\*07:01*) to the *Rhsi-DAB1* alleles obtained from the cDNA samples. No SEGs similar or identical to *Rhsi-DAB3* were found.

#### Signature of recombination

The result of RDP analysis showed that five *Rhsi-DAB* alleles likely occurred by recombination events in  $\beta 1$

domain region (Table 2). All seven algorithms used in this analysis supported that *Rhsi-DAB1\*04:01* and *-DAB3\*07:01* were formed from the recombination between two different alleles (Table 2). *Rhsi-DAB1\*03:07*, *-DAB1\*03:17* and *-DAB1\*10:02* could be considered as the recombinant, but not supported by all algorithms. The signature of recombination in *Rhsi-DAB1\*04:01* and *-DAB3\*07:01* could also be viewed in the network tree analysis, because they did not form clusters with the same allelic groups but instead were located in the middle of the recombination origins (Fig. 1).

#### Signature of positive selection

The non-synonymous substitution level ( $d_N/d_S$ ) of  $\beta 1$  domain was significantly high in both *Rhsi-DAB1* ( $Z = 2.967$ ,  $P = 0.037$ ) and *-DAB3* ( $Z = 1.290$ ,  $P = 0.045$ ), whereas no significance was found in other exonal regions (*Rhsi-DAB1*:  $Z = -2.443$ ,  $P = 1.000$ ; *Rhsi-DAB3*:  $Z = -1.911$ ,  $P = 1.000$ ; Fig. 2 (a)). Two positive selection models, *M2a* and *M8*, were likely to fit the data significantly better than nearly neutral (*M1a*) and  $\beta$  distribution (*M7*), respectively, in both *Rhsi-DAB1* (Additional file 2: Table S2 and Additional file 3: Table S3) and *Rhsi-DAB3* (Additional file 4: Table S4 and Additional file 5: Table S5). The average  $d_N/d_S$  level of PBRs was specifically higher than that of other codons (non-PBRs) in  $\beta 1$  domain region of *Rhsi-DAB1* (Fig. 2 (b)), though no such a pattern was observed in *Rhsi-DAB3*

**Table 2** Five putative recombinants detected using seven different testing algorithms (R: RDP, C: CHIMAERA, G: GENECONV, S: SISCAN, B: BOOTSCAN, T: 3SEQ, M: MAXCHI) implemented in RDP4. Significance was indicated by '+'. 'Unknown' means that a definite parent allele was not estimated, but alleles with the highest probability were indicated in parentheses

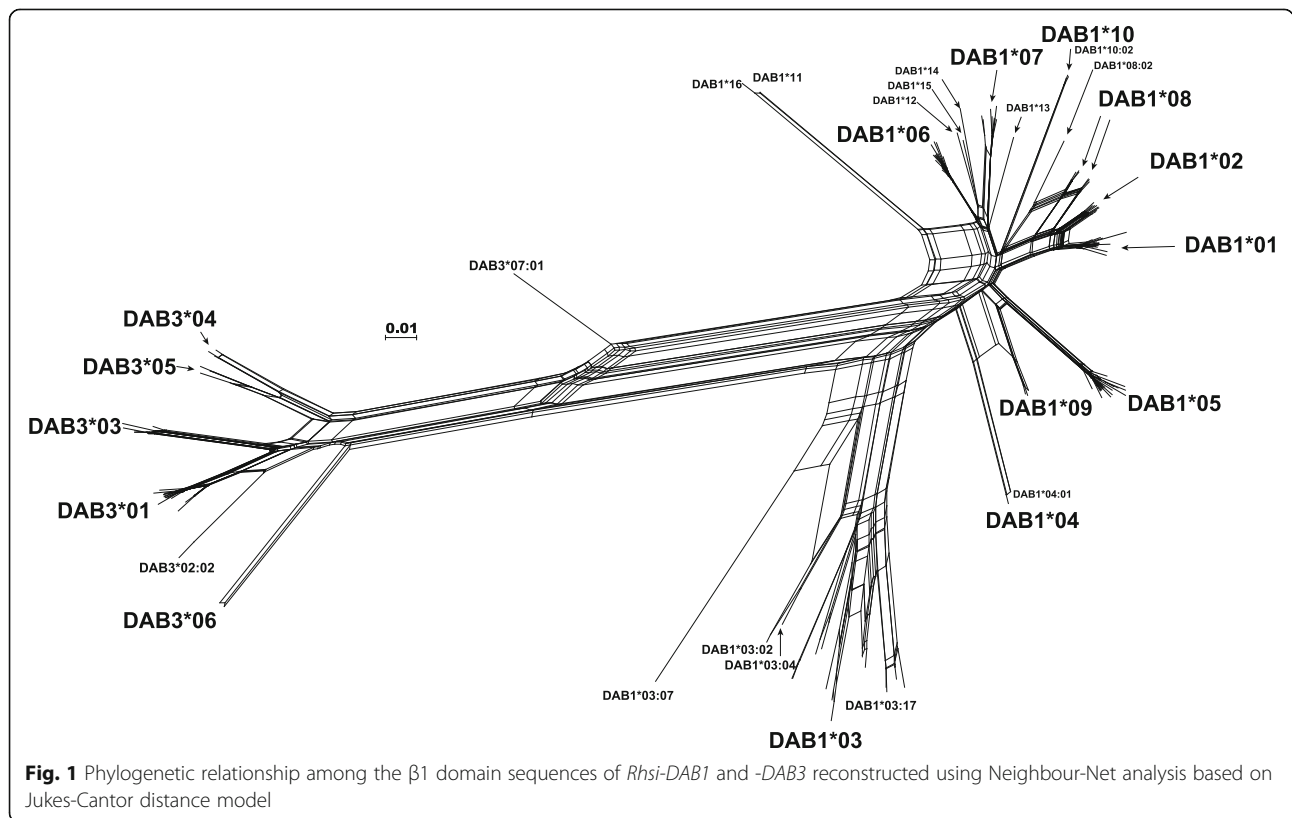
Recombinant	Breakpoints		Parent alleles		Algorithm						
	Begin	End			R	G	B	M	C	S	T
DAB3*07:01	1	144	DAB1*09:01	DAB3*01:07	+	+	+	+	+	+	+
DAB1*04:01	232	276	DAB1*01:13	Unknown (DAB1*03:06)	+	+	+	+	+	+	+
DAB1*03:07	225	274	DAB1*03:06	Unknown (DAB1*05:04)	+	+	-	+	+	-	+
DAB1*03:17	165	213	DAB1*03:02	DAB1*03:13	+	-	-	-	-	-	-
DAB1*10:02	82	172	DAB1*09:01	DAB1*03:05	+	-	-	+	+	-	+

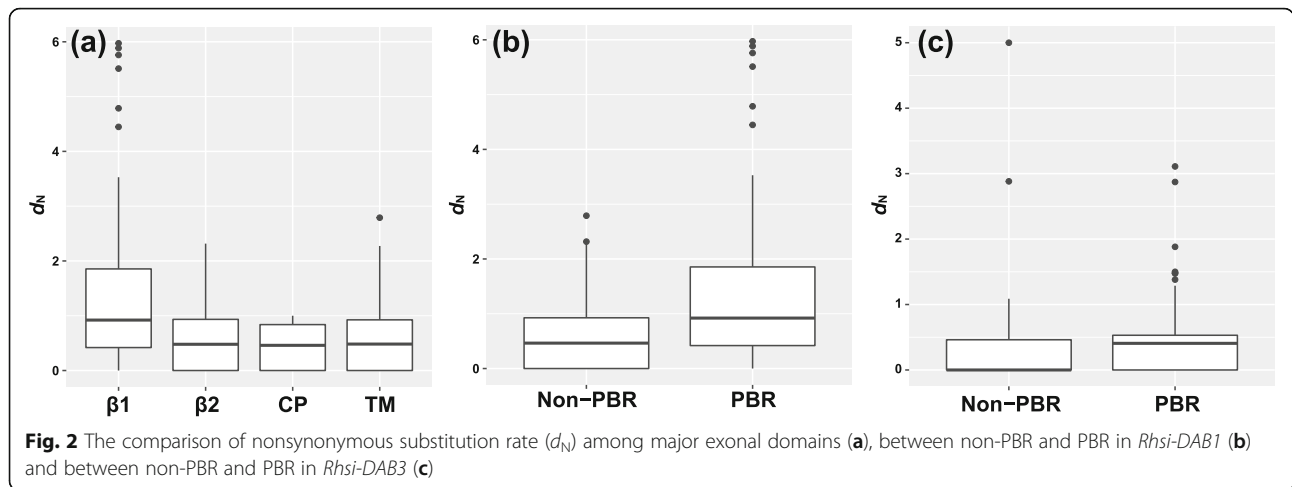
(Table 3; Fig. 2(c)). The BEB analysis in CODEML revealed that twenty-three codons in *Rhsi-DAB1* β1 (Additional file 2: Table S2 and Additional file 3: Table S3) and fifteen codons in *Rhsi-DAB3* β1 (Additional file 4: Table S4 and Additional file 5: Table S5) showed the signature of positive selection (Table 4; Additional file 6: Table S6). Thirteen putative PBRs were included among those codons in *Rhsi-DAB1* (Additional file 14: Figure S7 and Additional file 15: Figure S8). In *Rhsi-DAB3*, however, only five putative PBRs were included in these fifteen codons (Additional file 16: Figure S9 and Additional file 17: Figure S10). Four different codon-based maximum likelihood tests yielded slightly different results. For example,

SLAC and FEL indicated three and ten codons showing the signature of positive selection, respectively, in *Rhsi-DAB1* β1 (Table 4), while no codon was indicated in *Rhsi-DAB3* β1 (Additional file 7: Table S7). MEME detected the signature of positive selection from 23 and three codons in *Rhsi-DAB1* β1 and in *Rhsi-DAB3* β1, respectively (Table 4). FUBAR detected the signature of positive selection from 17 and 10 codons in *Rhsi-DAB1* β1 and in *Rhsi-DAB3* β1, respectively (Table 4).

**Phylogenetic analysis**

Upon NJ tree analysis based on the β1 domain of *DABs*, a total of 16 and seven allelic groups were identified in





*Rhsi-DAB1* and in *Rhsi-DAB3*, respectively. The allelic IDs were determined based on this NJ tree clustering (Additional file 18: Figure S11). In BI tree, however, some of the allelic groups identified in NJ tree failed to form a clear monophyletic cluster (Fig. 3). Like NJ tree, *Rhsi-DAB1* and *-DAB3* alleles formed completely independent clusters in the BI tree (Fig. 3). Within the *Rhsi-DAB1* allelic groups, *Rhsi-DAB1\*05* was found to form a sister to other groups (Fig. 3). Several alleles from other Cypriniformes species, such as *Hymo-DAB*, *Hyam DAB1*, *Ctid-DAB* and *Cyca-DAB*, were embedded in the clade of *Rhsi-DAB1* allelic groups (Fig. 3). *Rhsi-DAB3* allelic groups clustered with alleles from the wider range of taxa (Cyprinodontiformes) including *Dare-DAB1*, *Dare-DAB2*, *Dare-DAB4*, *Cyca-DAB3*, *Cyca-DAB4*, *Ximu-DXB*, *Xipy-DXB*, *Tata-DAB3* and *Hyam DAB3* (Fig. 3). *Rhsi-DAB* alleles did not cluster with the alleles from other teleost orders, such as Salmoniformes, Siluriformes, Perciformes, Pleuronectiformes and Syngnathiformes as well as non-teleost vertebrates (Fig. 3).

**Discussion**

Our analysis of MHC class IIB variation in *Rhodeus sinensis* revealed very large numbers of *DAB* alleles, which was only from 50 individuals. Characterization of

*DAB* was attempted in two other bitterling species, *R. ocellatus* [33] and *Pseudorhodeus tanago* [34], as mentioned in Introduction, and only 17 [33] and 16 [34] sequence variants were identified, respectively. However, direct comparison of diversity with these species are not likely to be possible, since *Pseudorhodeus tanago* was listed as critically endangered species [34] and the MHC class IIB of *R. ocellatus* was analyzed simply because of the need to gather information on MHC allele diversity in behavioral experiments [33]. Our study aimed to analyze nearly complete *DAB* sequences, and all 140 alleles characterized included a signal peptide, two extra-cellular domains ( $\beta 1$  and  $\beta 2$ ), a connecting peptide, a trans-membrane region and a cytoplasmic domain. Upon comparison with other teleost or vertebrate species at the codon level, major residues predicted to be responsible for the processes in adaptive immune response were completely identified, indicating that all of the *Rhsi-DAB* sequences could be regarded as functional and classical alleles.

One of the surprising observations from this study was that SEGs (MHC class IIB lacking the introns) were discovered from gDNA samples. Intron loss has been reported across many vertebral genes (e.g., *hsp70* in zebrafish [43], the histone gene family in vertebrates [44], claudin and olfactory receptor genes in teleosts [45]). Intron loss in MHC class IIB was reported in two teleost species, *Gasterosteus aculeatus* and *Tetraodon nigroviridis* [45]. It has been proposed that intron loss could be mediated by random or homologous recombination of ancestral genome with the DNA fragment reverse-transcribed from processed mRNA [46–48]. Considering that partial intron loss was not observed in our study, these SEGs likely originated from the insertion of processed cDNA into a nonhomologous genome location (retrotransposition event; see [45]). Nine variants of SEGs were found to be identical or similar to *Rhsi-DAB1*, while no SEG similar to *Rhsi-DAB3* was found, indicating that this event has probably occurred since the

**Table 3** The signature of positive selection in  $\beta 1$  domain regions of two *Rhsi-DABs*. Data comprise the number of residues (Codon), ratio of nonsynonymous vs synonymous substitution ( $d_N/d_S$ ) and the significance of Z test (*P*)

Region	DAB1 $\beta 1$			DAB3 $\beta 1$		
	Codon	$d_N/d_S$	<i>P</i>	Codon	$d_N/d_S$	<i>P</i>
PBR	19	2.929	<b>0.001</b>	19	1.940	0.123
Non-PBR	72	1.091	0.395	72	0.798	1.000
Total	91	1.533	<b>0.037</b>	91	1.034	<b>0.045</b>

Statistical significance (< 0.05) was highlighted with bold



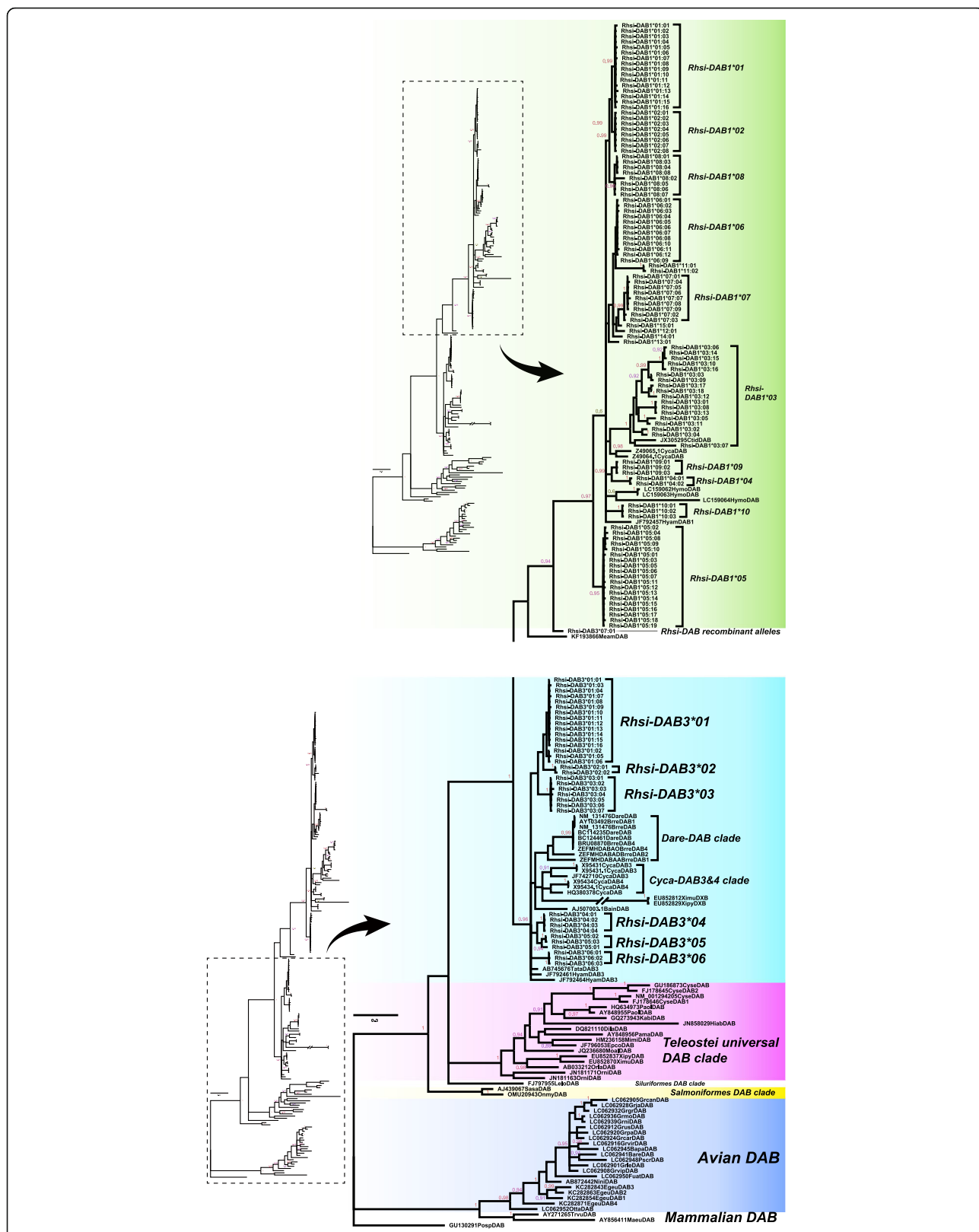
**Table 4** Identification of the codons showing the signature of positive select based on six different models implemented RDP and CODEML. The codons predicted to be PBR were highlighted by red

	5	7	8	9	12	17	22	25	27	28	29	30	33	34	35	36	37	38	41	44	50	52	53	
<b>Rhsi-DAB1</b>																								
SLAC		+																						
FEL	+	+											+		+									+
MEME	+	+			+		+		+		+				+	+	+	+						+
FUBAR	+	+		+				+		+			+		+									+
M2a	+	+		+				+	+	+				+	+									+
M8	+	+		+				+	+	+				+	+								+	+
<b>Rhsi-DAB3</b>																								
SLAC																								
FEL																								
MEME				+		+		+																
FUBAR				+				+																+
M2a								+																+
M8								+																+
	54	57	58	60	61	63	64	65	67	68	71	72	74	75	77	78	82	83	84	85	86	88	90	91
<b>Rhsi-DAB1</b>																								
SLAC													+						+					
FEL								+		+			+	+					+					
MEME		+						+		+	+		+	+	+				+	+	+		+	+
FUBAR		+		+				+		+			+	+	+				+					
M2a			+	+				+		+				+	+				+	+				
M8			+	+				+		+		+	+		+				+	+		+		+
<b>Rhsi-DAB3</b>																								
SLAC																								
FEL																								
MEME																								
FUBAR	+		+					+		+			+						+					
M2a			+	+	+	+	+	+		+	+		+						+		+		+	
M8			+	+	+	+	+	+		+	+		+						+		+		+	

divergence between *DAB1* and *DAB3* were completed. Given that intron loss is more common in highly expressed genes, such as housekeeping genes [46–48], the reason why intron loss was observed only in *Rhsi-DAB1* can be deduced from the difference in expression level. However, it is premature to evaluate the difference in expression level without comparing mRNA levels. It remains to assure that intron loss is prevalent in all *R. sinensis* populations, because all of the gDNA samples used were obtained from the Nakdong River. It is thus necessary to investigate the existence of SEG from other drainages in future study.

The phylogenetic separation between *DAB1* and *DAB3* was evident in our BI tree reconstructed using the  $\beta 1$  domain as shown in a study of the Cypriniformes species [49]. Multiple loci may exist within *Rhsi-DAB1*, given that more than three alleles were frequently found in a

single individual sample. Assumption based only on the BI tree revealed that *Rhsi-DAB1\*05* is the most prominent locus candidate, and the remaining allelic groups were likely divided into five; (1) *Rhsi-DAB1\*01*, *-\*02* and *-\*08*; (2) *Rhsi-DAB1\*06*, *-\*07*, *-\*11*, *-\*12*, *-\*13*, *-\*14* and *-\*15*; (3) *Rhsi-DAB1\*03*; (4) *Rhsi-DAB1\*04* and *-\*09*; (5) *Rhsi-DAB1\*10*. No individuals with all six locus-candidates were observed, and the number of *Rhsi-DAB1* loci possessed by each individual appears to be different, given that the individuals showed considerably different composition of locus-candidates. Although its diversity was not as high as *Rhsi-DAB1*, there were roughly three locus-candidates in *Rhsi-DAB3*, if speculating based solely on the BI tree. In contrast to the enormous diversity observed in our results, no evidence of gene duplication was reported in bitterling species to dates [17, 33]. Although



**Fig. 3** Phylogenetic placement of *Rhsi-DAB* β1 domain sequences among the MHC class II B alleles of major telosts and vertebrates. A total of 44 vertebrate species were used as outgroup. Bayesian inference tree was reconstructed under GTR + I + G model

gene duplication of *DABs* may be thought of as a species-specific evolutionary event that occurred in *R. sinensis*, it may be wise to suspend judgment until tests using various primers are conducted. It has often been found that slightly different primer sets, which were designed for almost the same region, exhibited amplification bias yielding different sequences or different number of alleles.

In both *Rhsi-DAB1* and *-DAB3*, the  $\beta 1$  domain region was the most polymorphic, and the signature of positive selection was clearly detected. The amino acid sequences of the remaining regions were highly conservative with no clear sign of positive selection. In the  $\beta 1$  domain region, the polymorphism level of *Rhsi-DAB3* was less than that of *Rhsi-DAB1*, which was also observed in other studies of Cypriniformes species [50, 51]. In addition, *Rhsi-DAB1*, but not *Rhsi-DAB3*, showed a large and significant  $d_N/d_S$ , when considering only PBRs. Presumably, different types of selection pressures have acted on *Rhsi-DAB1* and *-DAB3* (see also [52, 53]). Our results suggest that while *Rhsi-DAB1* is responsible for the binding to a wide range of pathogenic peptides, *Rhsi-DAB3* may be associated with detection of some specialized antigenic peptides. If so, highly variable *Rhsi-DAB1* might have been specialized to interact with many different types of pathogens, which has probably led to the specific pathogen-mediated selections (see also [54, 55]). This explanation is highly plausible, considering that a slight difference in amino acid sequence among MHC class IIB alleles may be associated with adaptation to different pathogens (e.g., [56–59]). If the pathogenic fauna and allelic variation at the population level are studied, it will be possible to examine whether the allelic diversity of *Rhsi-DAB1* is the result of adaptation to different types of pathogens.

Recombination between different alleles seemed to have at least partially contributed to the increase or maintenance of diversity in *Rhsi-DABs*. Several other studies also reported examples of MHC class IIB alleles originating from recombination [3, 60–62]. In this study, some of the recombination events occurred between *Rhsi-DAB1* and *-DAB3*. Since *Rhsi-DAB1* and *-DAB3* were expected to differ in terms of their diversity and functional range, as mentioned above, further studies should be conducted to determine if the alleles originating from interlocus recombinations have resulted in adaptive benefits in their habitats.

The existence of the phylogenetic lineages containing alleles of *Rhsi-DAB1*, *Hyam-DAB1*, *Hymo-DAB* and *Ctid-DAB* probably suggests that the occurrence of *Rhsi-DAB1* alleles should predate the ages of diversification into the present species in Cypriniformes. Specifically, the sister relationship between *Rhsi-DAB1\*03* and *Ctid-DAB* showed that the *Rhsi-DAB1\*03* allelic group likely have existed before the divergence of Leuciscidae and Acheilognathidae that could be estimated to be around 66 million years ago (MYA) when speculating based on the divergence time

between the representative genera of those two families, *Ctenopharyngodon* and *Rhodeus* [63, 64]. *Rhsi-DAB3* formed a cluster with Cyprinodontiformes, which probably means that the sequence structure of *Rhsi-DAB3* has been maintained as more conserved rather than that of *Rhsi-DAB1*. Previous studies estimated that Cyprinodontiformes diverged from other teleosts at around 229.9 MYA [63, 64]. It thus can be predicted that the strength of selection pressure acting on *DAB3* might definitely be weaker than that acting on *DAB1*, considering that the habitat range of Cyprinodontiformes is much wider than Cypriniformes.

Our analysis revealed the evolutionary processes that have at least partially contributed to the formation of *DAB* diversity in *R. sinensis*. What needs to be revealed or confirmed in the future can be summarized as follows. First, what adaptive importance does the alleles identified here have? Second, did the alleles were born before or after the birth of this species? These questions can also be resolved at least in part through the investigation of allelic frequency among *R. sinensis* populations inhabiting different environments or direct comparison among species of the same genus (i.e., *Rhodeus*) and with other major genus species (i.e., *Tanakia* or *Acheilognathus*) in Acheilognathidae. Finally, are the alleles found here products of the same gene or other (duplicated or somewhat distant) genes? Genomic analysis, which has become highly advanced and inexpensive in recent years, will provide a close answer to this question.

## Conclusion

Using cDNA and gDNA samples from 50 individuals of *R. sinensis*, a great deal of MHC class IIB allelic diversity was found, and gene duplication, selection and recombination may have contributed to this diversity. A total of 140 allelic sequences could be allocated into two different loci, *Rhsi-DAB1* and *-DAB3*. Numerous variants of MHC IIB lacking the introns were found from our gDNA samples and these sequences appeared to be historically derived from the retrotransposition events of processed mRNA. Upon robust phylogenetic analysis, *Rhsi-DAB1* and *-DAB3* formed completely independent clusters. Based on our data, it is presumed that such historical processes have commonly or differently acted on the polymorphism of *Rhsi-DAB1* and *-DAB3*.

## Methods

### Sampling

Ten individuals of *Rhodeus sinensis* were collected each from five different rivers (Han, Hyeongsan, Mangyeong, Nakdong and Tamjin) on the Korean Peninsula for the extraction RNA samples. Twenty individuals were additionally collected from the Nakdong River and used for the analysis of genomic sequence structure.



### Extraction of RNA and DNA

RNA was isolated from the brain tissue of each individual using TRIzol (Invitrogen, Carlsbad, CA, USA) reagent according to the manufacturer's protocol. Before tissue removal for the RNA extraction, each individual fish was euthanized with MS-222 (250 mg/L; Sigma-Aldrich, St. Louis, MO, USA). Brain tissues were used since they were the samples that should also be analyzed to investigate the expression of non-immune genes. Complementary DNA (cDNA) was synthesized using RNA (800–2000 ng) extracted, oligo-dT and GoScript™ Reverse Transcriptase (Promega, Madison, WI, USA). Genomic DNA (gDNA) was extracted from muscle using a DNeasy Blood and Tissue kit in accordance with the manufacturer's protocol (Qiagen, Dusseldorf, Germany). All individuals used for DNA and RNA extraction are housed in the Department of Life Science at Yeungnam University as a 95% ethanol sample. Our sampling and experimental procedure were approved by the Yeungnam University Institutional Animal Care and Use Committee (Protocol # 2015013). The concentration of genetic samples was measured using Maestro-Nano (Maestrogen, Hsinchu City, Taiwan).

### Primers design, PCR, cloning and sequencing

The universal primers designed for cyprinids in previous studies [33, 41] were used to obtain the draft *DAB* sequences of *R. sinensis* (Table 1). Specific primers were also designed based on the draft sequences obtained (Table 1). Each 50  $\mu$ L mixture for PCR amplification contained 50–100 ng of DNA (cDNA or gDNA), 1  $\times$  PCR buffer, 3 mM  $MgCl_2$ , 0.25 mM of each forward and reverse primer, 0.2 mM dNTP and 0.25 unit of *Taq* DNA polymerase (Solgent, Daejeon, South Korea). PCR amplification was performed using GenePro (Bioer, Hangzhou, China) under the following program setup: 94 °C for 10 min, 35 cycles of 40s at 94 °C, 45 s at 52–64 °C (depending on the primers) and 50s at 72 °C, and 72 °C for 10 min. The amplified products were ligated into the pGEM-T Easy vector (Promega) and transformed into *E. coli* DH5 $\alpha$ . Ten to sixteen white colonies were selected from each individual for the amplification by SP6 and T7 primer set ( $T_a = 56$  °C). Being successfully identified on a 2% agarose gel, the PCR products were purified and sent for the commercial sequencing to Macrogen Inc. (Seoul, South Korea).

The nucleotide sequences obtained from cDNA or gDNA samples were regarded as a valid allelic sequence only when identified in at least two separated clones and two different individuals to avoid the possibility of artifacts. To examine the possibility of cross-over contamination, a negative control tube containing purified water instead of a DNA sample was placed in every amplification sample set. All data used in the analyses included only those with no amplification reaction at the negative control. Analyses of cDNA and gDNA samples were performed in a completely separated state in

time and space. The sequences identified as valid were aligned using CLUSTALX [65] implemented in GENEIOUS v.9.1.8 [66]. The *DAB* alleles were named following the nomenclature (locus\*allelic group;protein sequence) [67].

### Tests of recombination

RDP v.4.5 (Recombination Detection Program; [68, 69]) was used to identify the signature of gene recombination in  $\beta 1$  domain region between different allelic sequences based on seven different algorithms including RDP [68], Chimaera [70], Geneconv [71], SiScan [72], Bootscan [69], 3Seq [73] and Maxchi [74]. For the visualization of the recombination events, phylogenetic network of *DAB* alleles was reconstructed using Neighbour-Net analysis based on Jukes-Cantor distance model implemented in SplitsTree4 [75].

### Tests of positive selection

The putative PBRs in  $\beta 1$  domain region were identified based on the comparison with the sequences characterized in the previous studies [42, 76, 77]. The ratio of non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions ( $\omega$ ) was estimated by Nei-Gojobori method [78] with 2000 bootstrap replicates and modified under Juke Cantor corrections, which was used to determine the strength of historical selection pressure acting on *DAB* sequences.

The signature of positive selection was detected for each codon in the *DAB* allelic sequences using HyPhy package [79] implemented in DataMonkey web server (<http://www.datamonkey.org/>), where four different codon-based maximum likelihood tests, namely, SLAC (Singel Likelihood Ancestry Counting), FEL (Fixed Effects Likelihood), MEME (Mixed Effects Model of Evolution) and FUBAR (Fast Unconstrained Bayesian AppRoximation) were used. For this purpose, phylogenetic relationship was reconstructed under the default setting in DataMonkey.

Each codon was also tested for the signature of positive selection using CODEML implemented in PAMLX package [80, 81]. For this test, Bayesian posterior probability (BPP) was calculated based on the Bayes empirical Bayes (BEB) method, and the codon was regarded as showing the significant signature of positive selection when the BPP was greater than 95%. Maximum likelihood (ML) phylogenetic tree was reconstructed using rapid bootstrap analyses (1000 replicates) implemented in RAxML-GUI v.1.5 [82], which was used for the input resource in CODEML. The likelihood ratio tests (LRT) were carried out to compare codon-based models, for example, between *M1a* (nearly neutral) and *M2a* (positive selection), between *M7* (beta distribution) and *M8* (beta distribution and positive selection), and between *M0* (one-ratio) and *M3* (discrete).

### Phylogenetic structure

Neighbor-joining (NJ) tree analysis [83] was performed using the MEGA v.7 [84] under 1000 bootstrapping to check whether *DAB* alleles were divided into evolutionary lineages. To detect TSP across vertebrate, Bayesian inference (BI) phylogenetic analysis was performed using MrBayes v.3.2.3 [85] under the following options: four heated chains, 40,000,000 generations and sampling tree at each 1000 generation. GTR + I + G model was selected as the best fit for the BI analysis using jModelTest v.2.0 [86] under Akaike Information Criterion (AIC; [87]). The first 25% generations on each run was discarded as burn-in. FigTree v.1.4.2 [88] was used to visualize the consensus phylogenetic trees and posterior probabilities on the nodes. A total of 44 vertebrate species were used as outgroup taxa (Additional file 1: Table S1).

### Additional files

**Additional file 1: Table S1.** The information of 44 vertebrate species used as outgroup in the BI tree phylogenetic analyses. (XLSX 12 kb)

**Additional file 2: Table S2.** Summary of models that were used to discover the signature of selection across codons of *Rhisi-DAB1*  $\beta 1$  domain. Data were from six models and comprise the logarithm of the likelihood (lnL), codons showing the signature of positive selection and number of parameters estimated in the model. (XLSX 10 kb)

**Additional file 3: Table S3.** The likelihood ratio tests (LRT) performed to verify the signature of positive selection in each nested codon-based model comparison for *Rhisi-DAB1*  $\beta 1$  domain. (XLSX 9 kb)

**Additional file 4: Table S4.** Summary of models that were used to discover the signature of selection across codons of *Rhisi-DAB1*  $\beta 3$  domain. Data were from six models and comprise the logarithm of the likelihood (lnL), codons showing the signature of positive selection and number of parameters estimated in the model. (XLSX 10 kb)

**Additional file 5: Table S5.** The likelihood ratio tests (LRT) performed to verify the signature of positive selection in each nested codon-based model comparison for *Rhisi-DAB3*  $\beta 1$  domain. (XLSX 9 kb)

**Additional file 6: Table S6.** The likelihood ratio tests (LRT) performed to verify the signature of positive selection in each nested codon-based model comparison for *Rhisi-DAB1*  $\beta 1$  domain. (XLSX 11 kb)

**Additional file 7: Table S7.** The likelihood ratio tests (LRT) performed to verify the signature of positive selection in each nested codon-based model comparison for *Rhisi-DAB3*  $\beta 1$  domain. (XLSX 11 kb)

**Additional file 8: Figure S1.** Amino acid sequences of *Rhisi-DAB1*  $\beta 1$  domain aligned using CLUSTALX. Dashes were added to represent gaps made in alignment procedure. Various color shading represent the functional regions; black: peptide-backbone interacting residues predicted based on mammalian studies, red: positions 81H and 82 N, yellow: cysteines predicted to form a disulfide bridge, cyan: conserved N-glycosylation motifs, green: highly conserved residues among jawed vertebrates, pink: peptide binding region predicted based on human DRB study, and orange: ray-finned specific residues. (DOCX 1440 kb)

**Additional file 9: Figure S2.** Amino acid sequences of *Rhisi-DAB3*  $\beta 1$  domain aligned using CLUSTALX. Dashes were added to represent gaps made in alignment procedure. Various color shading represent the functional regions; black: peptide-backbone interacting residues predicted based on mammalian studies, red: positions 81H and 82 N, yellow: cysteines predicted to form a disulfide bridge, cyan: conserved N-glycosylation motifs, green: highly conserved residues among jawed vertebrates, pink: peptide binding region predicted based on human DRB study, and orange: ray-finned specific residues. (DOCX 513 kb)

**Additional file 10: Figure S3.** Amino acid sequences of *Rhisi-DAB1*  $\beta 2$  domain aligned using CLUSTALX. Specific functional regions with conserved sequences were highlighted by color shadings; pink: residues binding to CD4, and yellow: cysteines predicted to form a disulfide bridge. (DOCX 1270 kb)

**Additional file 11: Figure S4.** Amino acid sequences of *Rhisi-DAB3*  $\beta 2$  domain aligned using CLUSTALX. Specific functional regions with conserved sequences were highlighted by color shadings; pink: residues binding to CD4, and yellow: cysteines predicted to form a disulfide bridge. (DOCX 446 kb)

**Additional file 12: Figure S5.** Amino acid sequences of *Rhisi-DAB1* CP, TM and CY domain regions aligned using CLUSTALX. (DOCX 846 kb)

**Additional file 13: Figure S6.** Amino acid sequences of *Rhisi-DAB3* CP, TM and CY domain regions aligned using CLUSTALX. (DOCX 467 kb)

**Additional file 14: Figure S7.** The positive selection signatures in the codons of *Rhisi-DAB1*  $\beta 1$  region were represented based on the mean  $\omega$  weighted by the posterior probabilities under the models, *M2a* (a) and *M8* (b). (DOCX 78 kb)

**Additional file 15: Figure S8.** The signature of positive selection was tested with the posterior probability for each codon in *Rhisi-DAB1*  $\beta 1$  region under the models, *M2a* (a) and *M8* (b). (DOCX 130 kb)

**Additional file 16: Figure S9.** The positive selection signatures in the codons of *Rhisi-DAB3*  $\beta 1$  region were represented based on the mean  $\omega$  weighted by the posterior probabilities under the models, *M2a* (a) and *M8* (b). (DOCX 75 kb)

**Additional file 17: Figure S10.** The signature of positive selection was tested with the posterior probability for each codon in *Rhisi-DAB3*  $\beta 1$  region under the models, *M2a* (a) and *M8* (b). (DOCX 131 kb)

**Additional file 18: Figure S11.** Phylogenetic relationship among *Rhisi-DAB*  $\beta 1$  alleles reconstructed based on NJ tree analysis. (DOCX 696 kb)

**Additional file 19:** Supplementary data. All sequences (in fasta format) reported in this study. (DOCX 25 kb)

### Abbreviations

BI: Bayesian inference; cDNA: Complementary DNA; gDNA: Genomic DNA; LRT: Likelihood ratio test; MHC: Major histocompatibility complex; ML: Maximum likelihood; MYA: Million years ago; NJ: Neighbour-joining; PBR: Peptide binding residue; SEG: Single exon gene; TSP: Trans-species polymorphism

### Acknowledgements

We thank S J Cho and A Hanado for assisting with sample collection.

### Author's contributions

HBJ and HYS designed this study. HBJ and HW carried out all the experiments, sampling and data analysis. HYS supervised the whole processes in this study. HBJ and HYS wrote the manuscript. All authors read and approved the final manuscript.

### Funding

This work was funded by the grant from the National Research Foundation of Korea (2015R1D1A2A01058987 and 2018R1D1A1A02086230). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Availability of data and materials

All sequences reported in this study can be retrieved from NCBI GenBank (accession numbers: MG989278 to MG989423) or 'Additional file 19: Supplementary data' in this article.

### Ethics approval and consent to participate

The fish samples used in this study were collected in accordance with the rules described in the Inland Water Fisheries Act and Wildlife Protection and Management Act of Republic of Korea. Our sampling and experimental procedure were approved by the Yeungnam University Institutional Animal Care and Use Committee (Protocol # 2015013).

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no conflict of interests.

**Author details**

<sup>1</sup>Department of Life Sciences, Yeungnam University, 280 Daehak-ro, Gyeongsan, Gyeongsangbuk-do 38541, South Korea. <sup>2</sup>Department of Biology, Concordia University, 7141 Sherbrooke W, Montreal, Quebec H4B 1R6, Canada.

Received: 2 April 2019 Accepted: 30 August 2019

Published online: 13 September 2019

**References**

- Klein J, Figueroa F. The evolution of class I MHC genes. *Immunol Today*. 1986;7:41–4.
- Marsh SGE, Parham P, Barber LD. The HLA factsbook. London: Academic Press; 1999.
- Reusch TBH, Langefors Å. Inter- and intralocus recombination drive MHC class II B gene diversification in a teleost, the three-spined stickleback *Gasterosteus aculeatus*. *J Mol Evol*. 2005;61:531–41.
- Pitcher TE, Neff BD. MHC class II B alleles contribute to both additive and nonadditive genetic effects on survival in Chinook salmon. *Mol Ecol*. 2006;15:2357–65.
- Alcaide M, Edwards SV, Negro JJ. Characterization, polymorphism, and evolution of MHC class II B genes in birds of prey. *J Mol Evol*. 2007;65:541–54.
- Hosomichi K, Shiina T, Suzuki S, Tanaka M, Shimizu S, Iwamoto S, et al. The major histocompatibility complex (Mhc) class II B region has greater genomic structural flexibility and diversity in the quail than the chicken. *BMC Genomics*. 2006;7:322.
- Musolf K, Meyer-Lucht Y, Sommer S. Evolution of MHC-DRB class II polymorphism in the genus *Apodemus* and a comparison of DRB sequences within the family Muridae (Mammalia: Rodentia). *Immunogenetics*. 2004;56:420–6.
- Wegner KM, Kalbe M, Rauch G, Kurtz J, Schaschl H, Reusch TBH. Genetic variation in MHC class II expression and interactions with MHC sequence polymorphism in three-spined sticklebacks. *Mol Ecol*. 2006;15:1153–64.
- Hedrick PW. Balancing selection and MHC. *Genetica*. 1998;104:207–14.
- Hedrick PW. Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity*. 1999;82:126–33.
- Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? *J Evol Biol*. 2003;16:363–77.
- Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool*. 2005;2:16.
- Langefors Å, Von Schantz T, Widegren B. Allelic variation of Mhc class II in Atlantic salmon; a population genetic analysis. *Heredity*. 1998;80:568–75.
- Wegner KM, Reusch TBH, Kalbe M. Multiple parasites are driving major histocompatibility complex polymorphism in the wild. *J Evol Biol*. 2003;16:224–32.
- Penn DJ. The scent of genetic compatibility: sexual selection and the major histocompatibility complex. *Ethology*. 2002;108:1–21.
- Milinski M. The major histocompatibility complex, sexual selection, and mate choice. *Annu Rev Ecol Syst*. 2006;37:159–86.
- Reichard M, Spence R, Bryjová A, Bryja J, Smith C. Female rose bitterling prefer MHC-dissimilar males: experimental evidence. *PLoS One*. 2012;7:e40780.
- Klein J, Sato A, Nagl S, O'huigin C. Molecular trans-species polymorphism. *Annu Rev Ecol Syst*. 1998;29:1–21.
- Li L, Zhou X, Chen X. Characterization and evolution of MHC class II B genes in ardeid birds. *J Mol Evol*. 2011;72:474–83.
- Lenz TL, Eizaguirre C, Rotter B, Kalbe M, Milinski M. Exploring local immunological adaptation of two stickleback ecotypes by experimental infection and transcriptome-wide digital gene expression analysis. *Mol Ecol*. 2013;22:774–86.
- Luo M, Pan H. MHC II DRB variation and trans-species polymorphism in the golden snub-nosed monkey (*Rhinopithecus roxellana*). *Chin Sci Bull*. 2013;58:2119–27.
- Eibes JA, Townsend AK, Sepil I, Nishiumi I, Satta Y. Patterns of evolution of MHC class II genes of crows (*Corvus*) suggest trans-species polymorphism. *PeerJ*. 2015;3:e853.
- Těšický M, Vinkler M. Trans-species polymorphism in immune genes: general pattern or MHC-restricted phenomenon? *J Immunol Res*. 2015;2015:1–10.
- Hashimoto K, Nakanishi T, Kurosawa Y. Isolation of carp genes encoding major histocompatibility complex antigens. *Proc Natl Acad Sci*. 1990;87:6863–7.
- Dixon B, van Erp SHM, Rodrigues PNS, Egberts E, Stet RM. Fish major histocompatibility complex genes: an expansion. *Dev Comp Immunol*. 1995;19:109–33.
- Miller KM, Withler RE. Sequence analysis of a polymorphic Mhc class II gene in Pacific salmon. *Immunogenetics*. 1996;43:337–51.
- Clark MS, Shaw L, Kelly A, Snell P, Elgar G. Characterization of the MHC class I region of the Japanese pufferfish (*Fugu rubripes*). *Immunogenetics*. 2001;52:174–85.
- Dijkstra JM, Grimholt U, Leong J, Koop BF, Hashimoto K. Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. *BMC Evol Biol*. 2013;13:260.
- Wilson AB. MHC and adaptive immunity in teleost fishes. *Immunogenetics*. 2017;69:521–8.
- Yamano H, Yamauchi T, Hosoya K. A new host record of *Ichthyoxenus amurensis* (Crustacea: Isopoda: Cymothoidae) from the Amur bitterling *Rhodeus sericeus* (Cypriniformes: Cyprinidae). *Limnology*. 2010;12:103–6.
- Dávidová M, Blažek R, Trichkova T, Koutrakis E, Gaygusuz Ö, Ercan E, et al. The role of the European bitterling (*Rhodeus amarus*, Cyprinidae) in parasite accumulation and transmission in riverine ecosystems. *Aquat Ecol*. 2011;45:377–87.
- Huber V, Geist J. Glochidial development of the freshwater swan mussel (*Anodonta cygnea*, Linnaeus 1758) on native and invasive fish species. *Biol Conserv*. 2017;209:230–8.
- Agbali M, Reichard M, Bryjová A, Bryja J, Smith C. Mate choice for nonadditive genetic benefits correlate with MHC dissimilarity in the rose bitterling (*Rhodeus ocellatus*). *Evolution*. 2010;64:1683–96.
- Kubota H, Watanabe K. Loss of genetic diversity at an MHC locus in the endangered Tokyo bitterling *Tanakia tanago* (Teleostei: Cyprinidae). *Zool Sci*. 2013;30:1092–101.
- Kawamura K. Low genetic variation and inbreeding depression in small isolated populations of the Japanese rosy bitterling, *Rhodeus ocellatus kurumeus*. *Zool Sci*. 2005;22:517–24.
- Kim WJ, Shin EH, Kong HJ, Kim HS, Kim BS, Nam BH, et al. Characterization of novel microsatellite markers derived from Korean rose bitterling (*Rhodeus uyekii*) genomic library. *Genet Mol Res*. 2014;13:8147–52.
- Kwon Y-S, Li F, Chung N, Bae M-J, Hwang S-J, Byoen M-S, et al. Response of fish communities to various environmental variables across multiple spatial scales. *Int J Environ Res Public Health*. 2012;9:3629–53.
- Akai Y, Arai R. *Rhodeus sinensis*, a senior synonym of *R. lighti* and *R. uyekii* (acheilognathinae, cyprinidae). *Ichthyol Res*. 1998;45:105–10.
- Chang C-H, Li F, Shao K-T, Lin Y-S, Morosawa T, Kim S, et al. Phylogenetic relationships of Acheilognathidae (Cypriniformes: Cyprinoidea) as revealed from evidence of both nuclear and mitochondrial gene sequence variation: evidence for necessary taxonomic revision in the family and the identification of cryptic spec. *Mol Phylogenet Evol*. 2014;81:182–94.
- Kim IS, Choi Y, Lee CL, Lee YJ, Kim BJ, Kim JH. Illustrated book of Korean fishes. Seoul: Kyo-Hak Publ Co; 2005.
- Ottová E, Šímková A, Morand S. The role of major histocompatibility complex diversity in vigour of fish males (*Abramis brama* L.) and parasite selection. *Biol J Linn Soc*. 2007;90:525–38.
- Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*. 1993;364:33–9.
- Sambrook JG, Figueroa F, Beck S. A genome-wide survey of major histocompatibility complex (MHC) genes and their paralogues in zebrafish. *BMC Genomics*. 2005;6:152.
- Holmes WF, Braastad CD, Mitra P, Hampe C, Doenecke D, Albig W, et al. Coordinate control and selective expression of the full complement of replication-dependent histone H4 genes in normal and cancer cells. *J Biol Chem*. 2005;280:37400–7.
- Tine M, Kuhl H, Beck A, Bargelloni L, Reinhardt R. Comparative analysis of intronless genes in teleost fish genomes: insights into their evolution and molecular function. *Mar Genomics*. 2011;4:109–19.
- Coulombe-Huntington J, Majewski J. Characterization of intron loss events in mammals. *Genome Res*. 2007;17:23–32.
- Zhu T, Niu D-K. Frequency of intron loss correlates with processed pseudogene abundance: a novel strategy to test the reverse transcriptase model of intron loss. *BMC Biol*. 2013;11:23.
- Catania F. From intronization to intron loss: how the interplay between mRNA-associated processes can shape the architecture and the expression of eukaryotic genes. *Int J Biochem Cell Biol*. 2017;91:136–44.

49. Šimková A, Ottová E, Morand S. MHC variability, life-traits and parasite diversity of European cyprinid fish. *Evol Ecol*. 2006;20:465–77.
50. Rakus Kł, Wiegertjes GF, Jurecka P, Walker PD, Pilarczyk A, Imnarow I. Major histocompatibility (MH) class II B gene polymorphism influences disease resistance of common carp (*Cyprinus carpio* L.). *Aquaculture*. 2009;288:44–50.
51. Osborne MJ, Turner TF. Isolation and characterization of major histocompatibility class II $\beta$  genes in an endangered north American cyprinid fish, the Rio Grande silvery minnow (*Hypognathus amarus*). *Fish Shellfish Immunol*. 2011;30:1275–82.
52. Ottová E, Šimková A, Martin J-F, De Belloccq JG, Gelnar M, Allienne J-F, et al. Evolution and trans-species polymorphism of MHC class II $\beta$  genes in cyprinid fish. *Fish Shellfish Immunol*. 2005;18:199–222.
53. Seifertová M, Šimková A. Structure, diversity and evolutionary patterns of expressed MHC class IIB genes in chub (*Squalius cephalus*), a cyprinid fish species from Europe. *Immunogenetics*. 2011;63:167–81.
54. Eizaguirre C, Lenz TL. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *J Fish Biol*. 2010;77:2023–47.
55. Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proc R Soc B Biol Sci*. 2010;277:979–88.
56. Eizaguirre C, Yeates SE, Lenz TL, Kalbe M, Milinski M. MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol Ecol*. 2009;18:3316–29.
57. Evans ML, Neff BD. Major histocompatibility complex heterozygote advantage and widespread bacterial infections in populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Mol Ecol*. 2009;18:4716–29.
58. Fraser BA, Neff BD. Parasite mediated homogenizing selection at the MHC in guppies. *Genetica*. 2010;138:273–8.
59. Pavey SA, Sevellec M, Adam W, Normandeau E, Lamaze FC, Gagnaire P, et al. Nonparallelism in MHCII $\beta$  diversity accompanies nonparallelism in pathogen infection of lake whitefish (*Coregonus clupeaformis*) species pairs as revealed by next-generation sequencing. *Mol Ecol*. 2013;22:3833–49.
60. Ohta T. Effect of gene conversion on polymorphic patterns at major histocompatibility complex loci. *Immunol Rev*. 1999;167:319–25.
61. Aguilar A, Garza JC. Patterns of historical balancing selection on the salmonid major histocompatibility complex class II  $\beta$  gene. *J Mol Evol*. 2007;65:34–43.
62. Blais J, Rico C, Van Oosterhout C, Cable J, Turner GF, Bernatchez L. MHC adaptive divergence between closely related and sympatric African cichlids. *PLoS One*. 2007;2:e734.
63. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics*. 2006;22:2971–2.
64. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol*. 2015;32:835–45.
65. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. 2003;00:2.3.1–2.3.22.
66. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
67. Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, et al. Nomenclature for the major histocompatibility complexes of different species: a proposal. In: *The HLA System in Clinical Transplantation*. Berlin, Heidelberg, Germany: Springer; 1993. p. 407–11.
68. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics*. 2000;16:562–3.
69. Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retrovir*. 2005;21:98–102.
70. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci*. 2001;98:13757–62.
71. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology*. 1999;265:218–25.
72. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*. 2000;16:573–82.
73. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*. 2007;176:1035–47.
74. Smith JM. Analyzing the mosaic structure of genes. *J Mol Evol*. 1992;34:126–9.
75. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23:254–67.
76. Dixon B, Nagelkerke LAJ, Sibbing FA, Egberts E, Stet RJM. Evolution of MHC class II  $\beta$  chain-encoding genes in the Lake Tana barbel species flock (*Barbus intermedius* complex). *Immunogenetics*. 1996;44:419–31.
77. Landry C, Bernatchez L. Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). *Mol Ecol*. 2001;10:2525–39.
78. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 1986;3:418–26.
79. Pond SLK, Muse SV. HyPhy: hypothesis testing using phylogenies. In: *Statistical methods in molecular evolution*. New York: Springer; 2005. p. 125–81.
80. Xu B, Yang Z. PAMLX: a graphical user interface for PAML. *Mol Biol Evol*. 2013;30:2723–4.
81. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91.
82. Silvestro D, Michalak I. raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol*. 2012;12:335–7.
83. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4:406–25.
84. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
85. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.
86. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012;9:772.
87. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19:716–23.
88. Rambaut A. FigTree v1.4. <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed Oct 2017.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

