Developmental and Comparative Immunology 76 (2017) 361–369

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

A novel C-type lectin gene is a strong candidate gene for Benedenia disease resistance in Japanese yellowtail, *Seriola quinqueradiata*

Masatoshi Nakamoto [a], Yusuke Takeuchi [a], Kazuki Akita [a], Ryo Kumagai [a], Junpei Suzuki [a], Takashi Koyama [a], Tsutomu Noda [b], Kazunori Yoshida [b], Akiyuki Ozaki [c], Kazuo Araki [c], Takashi Sakamoto [a, *]

[a] Department of Aquatic Marine Biosciences, Tokyo University of Marine Science and Technology, Tokyo 108-8477, Japan
[b] Goto Laboratory of the Seikai National Fisheries Research Institute, Japan Fisheries Research and Education Agency, Nagasaki 853-0508, Japan
[c] National Research Institute of Aquaculture, Japan Fisheries Research and Education Agency, Mie 516-0193, Japan

ARTICLE INFO

Article history:
Received 31 March 2017
Received in revised form
8 July 2017
Accepted 8 July 2017
Available online 10 July 2017

Keywords:
Japanese yellowtail
Seriola quinqueradiata
Benedenia disease
Parasitic disease resistance
C-type lectin
Marker-assisted selection

ABSTRACT

Little is known about mechanisms of resistance to parasitic diseases in marine finfish. Benedenia disease is caused by infection by the monogenean parasite *Benedenia seriolae*. Previous quantitative trait locus (QTL) analyses have identified a major QTL associated with resistance to Benedenia disease in linkage group Squ2 of the Japanese yellowtail/amberjack *Seriola quinqueradiata*. To uncover the bioregulatory mechanism of Benedenia disease resistance, complete Illumina sequencing of BAC clones carrying genomic DNA for the QTL region in linkage group Squ2 was performed to reveal a novel C-type lectin in this region. Expression of the mRNA of this C-type lectin was detected in skin tissue parasitized by *B. seriolae*. Scanning for single nucleotide polymorphisms (SNPs) uncovered a SNP in the C-type lectin/C-type lectin-like domain that was significantly associated with *B. seriolae* infection levels. These results strongly suggest that the novel C-type lectin gene controls resistance to Benedenia disease in Japanese yellowtails.

1. Introduction

Marine fish of the genus *Seriola*, commonly known as yellowtails or amberjacks, are globally distributed and aquaculturally important in countries such as Japan, China, South Korea, the United States, New Zealand and Australia (Sicuro and Luzzana, 2016). In yellowtail aquaculture production, Benedenia disease is a serious parasitic disease caused by infection by the monogenean parasite *Benedenia seriolae*. Eggs of *B. seriolae* drift freely in water and attach to fish cages in aquaculture farms, where their free-swimming larvae parasitize the skin surface of yellowtails. Yellowtails infected by *B. seriolae* show reduced growth and also scrape their bodies against the cages to remove parasites, leading to secondary infection due to viral or bacterial disease.

Previous quantitative trait locus (QTL) analyses of Benedenia disease resistance using wild F_1 strains of the Japanese yellowtail/amberjack *Seriola quinqueradiata* identified two major QTLs: *BDR-1* and *BDR-2* (Ozaki et al., 2013). The highest detected LOD (logarithm of odds) score was for *BDR-1*, which was located within a 5.5 cM interval (at the 95% confidence level) of the proximal region of linkage group Squ2. *BDR-1* was found to explain 20.1%–21.4% of the phenotypic variance. The location of *BDR-2* was narrowed to an 11-cM interval (95% confidence level) at the edge of linkage group Squ20. *BDR-2* explained 12.8%–14.1% of the phenotypic variance. Although yellowtails use the same genetically regulated immune mechanisms to reject and/or protect against *B. seriolae* infection, the bioregulatory mechanisms underlying Benedenia disease resistance are unclear. *In vitro* experiments on the parasitism of *B. seriolae* have indicated that extracts of yellowtail skin epithelia induce the attachment and deciliation of *B. seriolae*, while some sugar-related compounds such as lectin suppress the attachment capacity (Yoshinaga et al., 2002). The results of these experiments suggest that sugar-related compounds are also involved in host specificity. In addition, parasite–host fish interactions via skin mucus chemicals have been reported in various fishes (Buchmann, 1999; Buchmann and Lindenstrøm, 2002; Alvarez-Pellitero, 2008). These results suggest that the chemical environment of yellowtail skin is critical for protecting these fish against parasites. Specific

* Corresponding author. Konan 4-5-7, Minato-ku, Tokyo 108-8477, Japan.
E-mail address: takashis@kaiyodai.ac.jp (T. Sakamoto).

http://dx.doi.org/10.1016/j.dci.2017.07.010
0145-305X/

chemical components may thus be present in the skin mucus of yellowtail individuals exhibiting Benedenia disease resistance.

Many parasitic diseases exist in vertebrates; these include the especially well-known parasitic diseases of humans caused by malaria and hookworm, where these protozoa parasitize the skin and ultimately the intestines through the bloodstream (Hotez et al., 2004; Sinnis and Zavala, 2012). Another example, in an aquaculture species, is whirling disease caused by infection by the myxosporean parasite *Myxobolus cerebralis*. A QTL for resistance against this disease has been identified in rainbow trout, *Oncorhynchus mykiss* (Baerwald et al., 2011). Relatively few researchers have focused on parasitic diseases compared with viral and bacterial infections. Consequently, the bioregulatory mechanisms underlying resistance against parasitic diseases are still unclear.

Resistance to Benedenia disease in yellowtails is a good model for studying parasitic diseases in vertebrates. Because bioregulatory mechanisms against parasitic diseases may be evolutionally conserved among vertebrates, information about bioregulation against Benedenia disease may aid the elucidation of the mechanism(s) of bioregulation against other parasitic diseases.

The aim of the present study was to reveal bioregulatory mechanisms against parasitic diseases caused by monogenean parasites. To achieve this goal, in this study we performed a detailed analysis of the QTL region of linkage group Squ2 in Japanese yellowtail and identified a strong candidate gene for Benedenia disease resistance. Complete sequencing of the QTL region of linkage group Squ2 uncovered a novel C-type lectin. The expression of this C-type lectin mRNA was detected in skin tissue parasitized by *B. seriolae*. A search for single nucleotide polymorphisms (SNPs) in the open reading frame of this C-type lectin revealed that one SNP located in the C-type lectin/C-type lectin-like domain was associated with infection levels of *B. seriolae*. These results strongly suggest that the novel C-type lectin gene is responsible for Benedenia disease resistance in yellowtails. This is the first report of a gene controlling resistance to parasitic diseases in teleost fish and provides the first genetic evidence that the protein conformation of C-type lectin in skin is important for resistance and protection against *B. seriolae*.

2. Materials and methods

2.1. Yellowtail source and measurements

Japanese yellowtail/amberjack (*S. quinqueradiata*) individuals from a wild population were provided by the Goto Laboratory of the Seikai National Fisheries Research Institute (Nagasaki, Japan). Individual yellowtails were labeled with PIT (Passive Integrated Transponder) tags. To count infection numbers of *B. seriolae* acquired under natural aquaculture conditions at the fish farm, fish were individually dipped in a freshwater tank. Parasites released from the fish bodies were collected and counted. These parasite measurements were performed on December 27 in 2007 and February 5, March 18, June 4 and July 9 in 2008. The sum of the five measurements was used as the number of parasites. Parasite numbers recorded for 40 yellowtail individuals are given in Supplemental Table 1. After the final measurement, the caudal fin of each yellowtail was fixed in 100% ethanol. All animal experiments were carried out with the approval of the Institutional Animal Care and Use Committee of the Tokyo University of Marine Science and Technology.

2.2. Screening for BAC clones

A BAC (Bacterial Artificial Chromosome) library was constructed from genomic DNA of sperm cells taken from one male yellowtail (Fuji et al., 2014). BAC library three-dimensional pools were generated according to a PCR-based protocol (Bruno et al., 1995; Bouzidi et al., 2006). The PCR amplifications were carried out using Takara Ex *Taq* (Takara Bio, Shiga, Japan) under the following cycling conditions: initial denaturation for 2 min at 95 °C, followed by 30 cycles of 30 s at 95 °C, 30 s at 56 °C and 30 s at 72 °C, with a final extension of 3 min at 72 °C. To determine BAC end sequences, BAC DNA extracted using a PureLink HiPure Plasmid Miniprep kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions was sequenced from both SP6 and T7 ends on an Applied Biosystems 3730xl DNA Analyzer (Life Technologies, Carlsbad, CA, USA). Primer sets for BAC end markers are shown in Supplemental Table 2. Novel simple sequence repeat (SSR) markers in linkage group Squ2 were developed from the BAC clones using the microsatellite hybrid capture technique according to a previous protocol (Prochazkal, 1996). Microsatellite genotyping was performed as previously described using the same mapping family as in that study (Ozaki et al., 2013). QTL analysis based on simple interval mapping was conducted using R/qtl software (Broman et al., 2003).

2.3. Sequencing and gene detection in BDR-1

Sequencing libraries with insert sizes of approximately 350 bp were individually constructed and subjected to 101-bp paired-end sequencing on a HiSeq2000 system (Illumina, San Diego, CA, USA) by InfoBio (Tokyo, Japan). The generated sequencing data were submitted to the DDBJ Sequence Read Archive (DRA) under accession number DRA005631. After filtering low-quality reads from the data, we performed *de novo* assembly on each clone separately using two programs: Velvet version 1.2.03 and Newbler version 2.9 (Roche Diagnostics, Mannheim, Germany), which were based respectively on *de Bruijn* graph (Zerbino and Birney, 2008) and overlap consensus algorithms. The sequence read assembly in Velvet was performed using *k*-mer sizes of 93 for #066_e22, 95 for #013_p20 and #101_o14, and 97 for #102_l17, #090_d18, #013_a02 and #070_j15, while the *de novo* assembler Newbler was used with default parameters. For finishing, gaps within contigs and ambiguous nucleotides were sequenced by Sanger sequencing using the primers shown in Supplemental Table 3. Names and GenBank accession numbers of the generated BAC clone sequences are as follows: #102_l17 (AP017998), #013_p20 (AP017999), #066_e22 (AP018000), #090_d18 (AP018001), #013_a02 (AP018002), #070_j15 (AP018003) and #101_o14 (AP018004). For *ab initio* gene detection, repetitive elements were identified from the genomic sequence and masked by RepeatMasker (Smit et al., 2013). To predict genes in the QTL region of linkage group Squ2, we used GENSCAN (Burge and Karlin, 1997; Burge, 1998; GENSCAN Web server at MIT http://genes.mit.edu/GENSCAN.html), Augustus version 3.1.0 (Stanke et al., 2008) and GrimmerHMM version 3.0.3 (Majoros et al., 2004) with default parameters. Augustus was trained using stickleback genomes (Ensembl Release 81; Yates et al., 2016). GrimmerHMM was used with a zebrafish training model. Amino acid sequences of predicted genes were searched and annotated against the NCBI protein database using Blastp with an *E*-value cutoff of 1×10^{-5}.

2.4. cDNA cloning

Yellowtail skins were fixed in RNAlater solution (Thermo Fisher Scientific). Total RNA was then extracted from the fixed skin using an RNeasy Mini kit (Qiagen, Hilden, Germany) and used for first-strand cDNA synthesis with an Ominiscript RT kit (Qiagen) and oligo-dT primers. The cDNA fragments of the yellowtail C-type lectin were amplified by RT-PCR using Takara Ex *Taq* (Takara Bio)

along with forward primers Squ_clec-1F, Squ_clec-2F, Squ_clec-3F and Squ_clec-4F and reverse primers Squ_clec-5R, Squ_clec-6R, Squ_clec-7R and Squ_clec-8R that were designed based on the predicted coding sequence (Supplemental Table 4). PCR conditions were 2 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 58 °C and 1 min at 72 °C, with a final extension of 5 min at 72 °C. The amplified fragments were cloned into a pGEM-T easy vector (Promega, Fitchburg, WI, USA) for sequencing on both strands on an Applied Biosystems 3730xl DNA Analyzer (Life Technologies).

To obtain the full-length cDNA of the yellowtail C-type lectin, we carried out 5′- and 3′ rapid amplification of cDNA ends (RACE). SMART cDNAs were synthesized using a SMART RACE cDNA Amplification kit (Clontech, Mountain View, CA, USA) according to the manufacturer's instructions with gene-specific primers RACE-5 5′-CTGGATCGGAGGTTACTACTTTCAG-3′ and RACE-nested-5 5′-TTCTGATTATGTCTGTTGTCCTCTG-3′ for the 5′-RACE PCR and RACE-3 5′-ACTGCCTGACCTGACACACGAATGG-3′ and RACE-nested-3 5′-ACCGTGAGCCATCTTCCCATCTCCA-3′ for the 3′-RACE PCR. The gene-specific primers were designed according to the partial cDNA sequence. For 5′-RACE, a first round of PCR amplifications was performed using the following conditions: 1 min at 94 °C, followed by 30 s at 94 °C, 30 s at 55 °C and 2 min at 72 °C. Nested PCR was then performed using 1.25 μl of the product of the first PCR round according to the following cycling protocol: 1 min at 94 °C, followed by 20 cycles of 30 s at 94 °C, 30 s at 55 °C and 2 min at 72 °C. For 3′-RACE, first-round PCR conditions consisted of 1 min at 94 °C, followed by 5 cycles of 30 s at 94 °C, 30 s at 70 °C and 2 min at 72 °C and then 15 cycles of 30 s at 94 °C, 30 s at 68 °C and 2 min at 72 °C, with a subsequent nested PCR consisting of 1 min at 94 °C followed by 30 cycles of 30 s at 94 °C, 30 s at 68 °C and 2 min at 72 °C.

2.5. Molecular phylogenetic analysis

The deduced amino acid sequence of the candidate gene was aligned with C-type lectin amino acid sequences of other teleost fish using Muscle of MEGA 7 (Kumar et al., 2016). Phylogenetic trees of teleost C-type lectins were constructed in MEGA 7 using maximum-likelihood and neighbor-joining methods. Bootstrap resampling was repeated 1000 times for each method. Names and accession numbers of analyzed sequences are shown in Supplemental Table 5. Amino acid sequences of other C-type lectins annotated in genome sequences of zebrafish, medaka, pufferfish and stickleback (*Gasterosteus aculeatus*) were obtained from Ensembl Release 86 (Yates et al., 2016); their Ensembl Protein IDs are listed in Supplemental Table 5.

2.6. RT-PCR for expression analysis

Tissues (fin, gill, spleen, heart, stomach, gut, skin, muscle, liver, kidney, head kidney and brain) of three uninfected yellowtail juveniles were fixed in RNAlater solution (Qiagen). Total RNA was extracted using an RNeasy Mini kit (Qiagen). First-strand cDNA was synthesized using an Ominiscript RT kit (Qiagen) with oligo-dT primers and Takara Ex *Taq* (Takara Bio). PCR conditions were 1 min at 96 °C, followed by 40 cycles of 10 s at 98 °C, 30 s at 54 °C and 1 min at 68 °C. The C-type lectin primers Squ_clec-9F and Squ_clec-10R amplified a 990-bp DNA fragment (Supplemental Table 3). As an internal control, β-actin mRNA was amplified using the primer set Beta-Actin_1F 5′-AATCGCCGCACTGGTTGTTG-3′ and Beta-Actin_1R 5′-ATACCGAGGAAGGAAGGCTG-3′ under the same PCR conditions mentioned above except with a cycle number of 30.

2.7. Analysis of DNA sequence polymorphism

Genomic DNA of yellowtails from the wild population was extracted from caudal fins using a DNeasy tissue and blood kit (Qiagen) according to the manufacturer's instructions. For SNP genotyping, the genomic region of the candidate gene for Benedenia disease resistance was amplified using primer set Squ_clec-11F and Squ_clec-12R and directly sequenced using sequencing primers Squ_clec-9F, Squ_clec-10R, Squ_clec-13F, Squ_clec-14F, Squ_clec-15F, Squ_clec-16F, Squ_clec-17F, Squ_clec-18R, Squ_clec-19R, Squ_clec-20R and Squ_clec-21R (Supplemental Table 3). To determine repeat numbers of repetitive elements, the genomic region of the C-type lectin was amplified using primer set Squ_clec-9F and Squ_clec-10R and cloned into a pGEM-T easy vector. The genomic clone was sequenced on both strands using primers Squ_clec-15F and Squ_clec-19R.

3. Results and discussion

3.1. Sequencing of BAC clones in BDR-1

To isolate the BAC clones covering the QTL region for Benedenia disease resistance in linkage group Squ2, BAC walking was performed starting from SSR markers Sequ0125TUF and Sequ1066TUF located at opposite ends of the QTL 95% confidence interval (Ozaki et al., 2013). Four clones (#013_k04, #102_l17, #028_c14 and #019_k17) were screened by Sequ0125TUF and five (#101_o14, #077_f22, #078_k13, #075_e05 and #114_l03) were screened by Sequ1066TUF. Neighboring clones were then screened according to the BAC end sequence of these nine clones. Finally, the QTL region in Squ2 was covered by a total of 18 BAC clones (Fig. 1). A minimal tiling path for the QTL region was constructed using clones #102_l17, #013_p20, #066_e22, #090_d18, #013_a02, #070_j15 and #101_o14. To confirm the accuracy of BAC walking, novel SSR markers were developed using the BAC clones comprising the *BDR-1* minimal tiling path; 18 of these markers showed polymorphism (Supplemental Table 6). All novel SSR markers were mapped to the vicinity of the QTL region of linkage group Squ2 by simple interval mapping using a mapping panel from a previous QTL analysis (Ozaki et al., 2013) (Fig. 1). The accuracy of our BAC screening was thus verified by linkage analysis. The peak LOD value of Sequ1094TUF (LOD = 4.71) was substantially higher than the genome-wide LOD significance threshold value of 3.27 determined by permutation testing ($Pg < 0.01$, where Pg is the genome-wide LOD *P*-value) (data not shown). This QTL explained 21% of the phenotypic variance. The 95% confidence interval of the QTL was localized to 4.4 cM between Sequ0125TUF and Sequ1085TUF.

3.2. Sequencing of BAC clones to construct a minimal tiling path

BAC clones #102_l17, #013_p20, #066_e22, #090_d18, #013_a02, #070_j15 and #101_o14 were sequenced separately on a HiSeq2000 system (Illumina) with 101-bp paired-end reads, which yielded an average of 28,278,959 ± 1,386,761 raw reads. Raw read counts of each BAC clone are shown in Supplemental Table 7, and the assembly results are summarized in Table 1. We next arranged contigs according to the BAC end markers (Fig. 1). For finishing, contig ends and ambiguous nucleotide regions were sequenced towards the gap by Sanger sequencing and manually aligned into the assembly. We then aligned overlapping BAC regions and obtained a 519,869-bp complete sequence of the QTL region for Benedenia disease resistance in linkage group Squ2.

3.3. Identification of a gene candidate for the Benedenia disease resistance trait

Coding sequence predictions were performed using GENSCAN, Augustus and GrimmerHMM, which predicted 26, 22 and 84

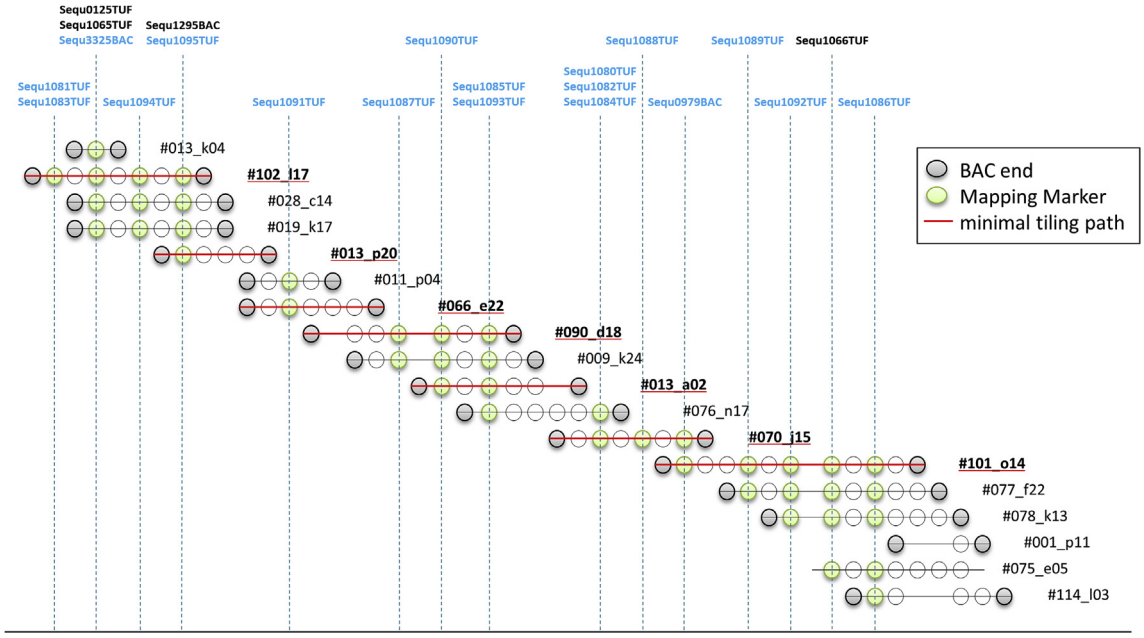

Fig. 1. Summary of the order of BAC (Bacterial Artificial Chromosome) clones in the QTL region for Benedenia disease resistance in linkage group Squ2. Red lines indicate BAC clones subjected to Illumina sequencing and used to construct a minimal tiling path. Gray and green circles indicate the position of BAC end markers and mapping markers, respectively. Mapping markers developed in this study are shown in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of assembly results.

Clone name	Assembler	Number of contigs	Mean length	Longest contig length
#102_l17	Velvet	5	23,952	66,574
	Newbler	12	9100	29,263
#013_p20	Velvet	10	11,068	28,503
	Newbler	20	5204	16,592
#066_e22	Velvet	8	15,328	38,997
	Newbler	29	3854	14,933
#090_d18	Velvet	22	4842	55,905
	Newbler	23	3742	11,702
#013_a02	Velvet	25	4283	44,972
	Newbler	18	4794	30,354
#070_j15	Velvet	81	2503	87,189
	Newbler	24	4105	13,089
#101_o14	Velvet	12	7901	28,880
	Newbler	19	4606	13,260

proteins, respectively. Because proteins predicted by GrimmerHMM were over-represented, we used proteins predicted by GENSCAN and Augustus for subsequent analyses. Of the 26 coding sequences predicted by GENSCAN, 14 showed similarity to known proteins in the NCBI database according to Blast searching. Among the 22 coding sequences predicted by Augustus, 15 showed similarity to known proteins (Table 2). No other predicted coding sequences showed similarity to known proteins. Aside from overlapping and transposon-related genes, 13 genes encoding the following proteins were suggested as candidates in this region: glycine receptor beta Z subunit, deoxycytidylate deaminase, E3 ubiquitin-protein ligase DTX1/4, F-box protein 8, Hand2, nattectin, N-lysine methyltransferase SETD8, pentatricopeptide repeat-containing protein, Pol polyprotein, POL5, PPUP8783 and teneurin-3. We focused on the predicted gene that showed similarity to nattectin, a C-type lectin containing a carbohydrate recognition domain and binding carbohydrate structures in a Ca^{2+}-dependent manner (Lopes-Ferreira et al., 2011), because lectins secreted in skin are critical for parasite–host-fish recognition and rejection (Buchmann, 1999; Buchmann and Lindenstrøm, 2002; Alvarez-Pellitero, 2008). These results suggested that differences in lectin molecular structure and/or expression patterns control resistance to Benedenia disease. We therefore cloned the candidate gene similar to a C-type lectin and analyzed its sequence polymorphism and expression patterns.

3.4. Candidate gene cloning

Cloning of the candidate gene exhibiting similarity to a C-type lectin by 5′ and 3′ RACE yielded a 1092-bp full-length cDNA composed of eight exons and encoding 274 amino acids (GenBank accession number LC223121; Fig. 2A and B). The candidate gene was located at positions 70,831 to 73,301 in the QTL region for Benedenia disease resistance in linkage group Squ2. Sequ1094TUF, the SSR marker located at the peak of the QTL, was positioned at 81,771. The physical distance between this peak QTL marker and the candidate gene was approximately 8.5 kb. The deduced peptide of this gene contained a C-type lectin (CTL)/C-type lectin-like (CTLD) domain between amino acid residues 159 to 269 with an *E*-value of 2.70×10^{-26} based on the NCBI conserved domain database (Marchler-Bauer et al., 2015); this indicates that the candidate gene belongs to the C-type lectin family. A TAVED repetitive element was located at amino acid positions 46 to 120. According to a conserved domain search, this element showed a weak similarity to the bacterial histone H1-like nucleoprotein HC2 domain, which suggests that polymorphism of this repetitive element affects binding to other molecules. A Blastp search revealed that the candidate gene shared approximately 40%–60% identity with C-type lectins of teleost fish. A multiple alignment was performed between the deduced amino acid sequence of the candidate gene and C-type lectin amino acid sequences of the following teleost fish: *Thalassophryne nattereri*, *Larimichthys crocea* (large yellow croaker), *Danio rerio* (zebrafish), *Oreochromis niloticus* (Nile tilapia), *Takifugu rubripes* (pufferfish) and *Salmo salar* (Atlantic

Table 2
Candidate genes for trait of Benedenia disease resistance.

Gene name	e-value
Predicted by Genscan	
Glycine receptor betaZ subunit	1.0e-175
Transposase	6.0e-91
Pentatricopeptide repeat-containing protein	9.0e-05
Nattectin	4.0e-28
E3 ubiquitin-protein ligase DTX4	0
Deltex1	4.0e-34
Transposase	2.0e-97
Reverse transcriptase	1.0e-119
Transposase	3.0e-25
ReO_6 protein	9.0e-12
F-box protein 8	2.0e-77
Deoxycytidylate deaminase	4.0e-83
Transposon Ty3-I Gag-Pol polyprotein	6.0e-13
Teneurin-3	0
Predicted by Augustus	
TC1-like transposase	2.0e-15
Glycine receptor betaZ subunit	0
Putative reverse transcriptase	2.0e-12
Transposase	2.0e-18
Pol polyprotein	1.0e-29
Nattectin	3.0e-47
deltex-4	0
Hand2	4.0e-96
ELE RTJK	9.0e-37
N-lysine methyltransferase SETD8	4.0e-19
F-box only protein 8	0
Deoxycytidylate deaminase	9.0e-87
POL5	4.0e-33
Teneurin-3	0
PPUP8783	1.0e-5

salmon). Names and accession numbers of analyzed sequences are given in Supplemental Table 5. This alignment revealed that the amino acid sequences of the N-terminal and C-type lectin domains were evolutionally conserved, whereas the repetitive element was not (Fig. 3). In the C-type lectin superfamily, four cysteine residues that form two disulfide bridges in the C-type lectin domain are evolutionally conserved and critical for protein folding (Zelensky and Gready, 2005). These cysteine residues were conserved in the C-type lectin of Japanese yellowtail (Fig. 3, white arrowheads). Molecular phylogenetic analyses using maximum-likelihood and neighbor-joining methods loosely grouped the candidate gene with nattectin, ladderlectin, rheacalcin-1 and struthiocalcin-1 (Supplemental Fig. 1A and B, dashed circle), but some branches in this cluster had low bootstrap support using either method. The reliability of the phylogenetic tree was therefore low, suggesting that amino acid sequences of C-type lectins belonging to this cluster are highly diverse. Taking all of these results into consideration, we concluded that the candidate gene was a novel C-type lectin and named it Benedenia disease resistance-related C-type lectin, *brcl*.

3.5. Expression of mRNA of the candidate gene in Japanese yellowtail

The expression of *brcl* mRNA was detected in fin, gill, spleen, heart, stomach, gut, skin, kidney and head kidney tissue by RT-PCR. The observed expression of *brcl* in skin, a tissue infected by *B. seriolae* (Fig. 4), supports our hypothesis that *brcl* in skin mucus is the gene controlling Benedenia disease resistance. Strong expression was detected in gill, spleen, gut and kidney tissue. Spleens and kidneys are organs in which many immune-related cells accumulate, while gills and gut tissue are in contact with the outside environment and secrete mucus. The above expression patterns suggest that *brcl* plays a role in the immune system. A comparison of *brc1* skin expression in fish with different resistance levels would further confirm our hypothesis. Antigen-presenting cells sense the microenvironment through several types of receptors that recognize pathogen-associated molecular patterns. In particular, C-type lectin receptors, which are expressed by distinct subsets of dendritic cells and macrophages, recognize and internalize specific carbohydrate antigens in a Ca^{2+}-dependent manner (Vázquez-Mendoza et al., 2013). The targeting of these receptors is becoming an effective strategy for parasite recognition. The C-type lectin nattectin may directly activate macrophages in *Thalassophryne nattereri* (Ishizuka et al., 2012), while nattectin-like protein exhibits bacterial agglutination activity in a Ca^{2+}-dependent manner in large yellow croaker (Lv et al., 2016). *Brcl* may thus be involved in protection against some pathogens via these immune pathways in Japanese yellowtail.

3.6. Sequence polymorphism of brcl in a wild yellowtail population

We compared repeat numbers of the repetitive element of brcl among yellowtail individuals from a wild population. Two to 11 copies of the TAVED repeat were observed in the wild population of 40 fish (Fig. 5A). The repeat number of this element was highly polymorphic. Nevertheless, an association between the repeat number of the TAVED element and the total number of *B. seriolae* per fish was not detected (correlation coefficient $r^2 = 0.037$, Fig. 5B).

We next identified SNPs within the *brcl* cDNA sequence (Table 3). Ten of 18 SNPs were located in the untranslated region. C308T in exon 2, C1643T in exon 6, and T2121G and A2175G in exon 7 were synonymous SNPs, causing no changes in amino acid sequences (Fig. 2A, white arrowheads). C279A in exon 2 involved an amino-acid change from leucine to methionine. In exon 6, A1637C involved a non-synonymous amino acid change from glutamic acid to aspartic acid while C1680A was associated with a leucine to isoleucine alteration. C2174T in exon 7 caused an amino-acid change between proline and leucine. A1637C, C1680A and C2174T were located within a C-type lectin/C-type lectin-like domain (Fig. 2A, black arrowheads). The leucine and proline residues affected by C1680A and C2174T seemed to be weakly conserved among teleost C-type lectins, whereas glutamic acid associated with A1637C in yellowtail *brcl* was not shared with C-type lectins of other teleost fish (Fig. 3, black arrowheads).

We also investigated the association between the genotypes of these non-synonymous SNPs and the number of *B. seriolae* per fish. The number of *B. seriolae* in yellowtail individuals possessing one or two leucine alleles (L/P or L/L) at C2174T (mean = 149.8 ± 17.8, $n = 29$) was significantly less ($P = 0.041$, one-sided Mann-Whitney *U* test; Fig. 5C) than in individuals homozygous (P/P) for the proline allele (mean = 200.8 ± 21.7, $n = 11$). No significant association between SNP genotype and levels of *B. seriolae* was detected for any of the other non-synonymous SNPs (data not shown). These results suggest that the leucine-containing *brcl* allele contributes to Benedenia disease resistance. We genotyped SNP C2174T in the 40 yellowtails from the wild population and determined that 29 fish harbored the leucine allele (L/L and L/P) and 11 fish possessed only the proline allele (P/P). The SNP C2174T genotypes of the 40 wild yellowtail individuals are given in Supplemental Table 8. The frequency of the leucine allele of *brcl* was high in the wild yellowtail population. This polymorphism should therefore be a useful DNA marker in marker-assisted selection for Benedenia disease resistance from wild populations. In general, the cyclic side chain of the proline residue confers conformational rigidity and is important for protein secondary structure (Rauscher et al., 2006). A change in amino acids, i.e., from proline containing the cyclic side chain to leucine containing an isobutyl side chain, would be expected to

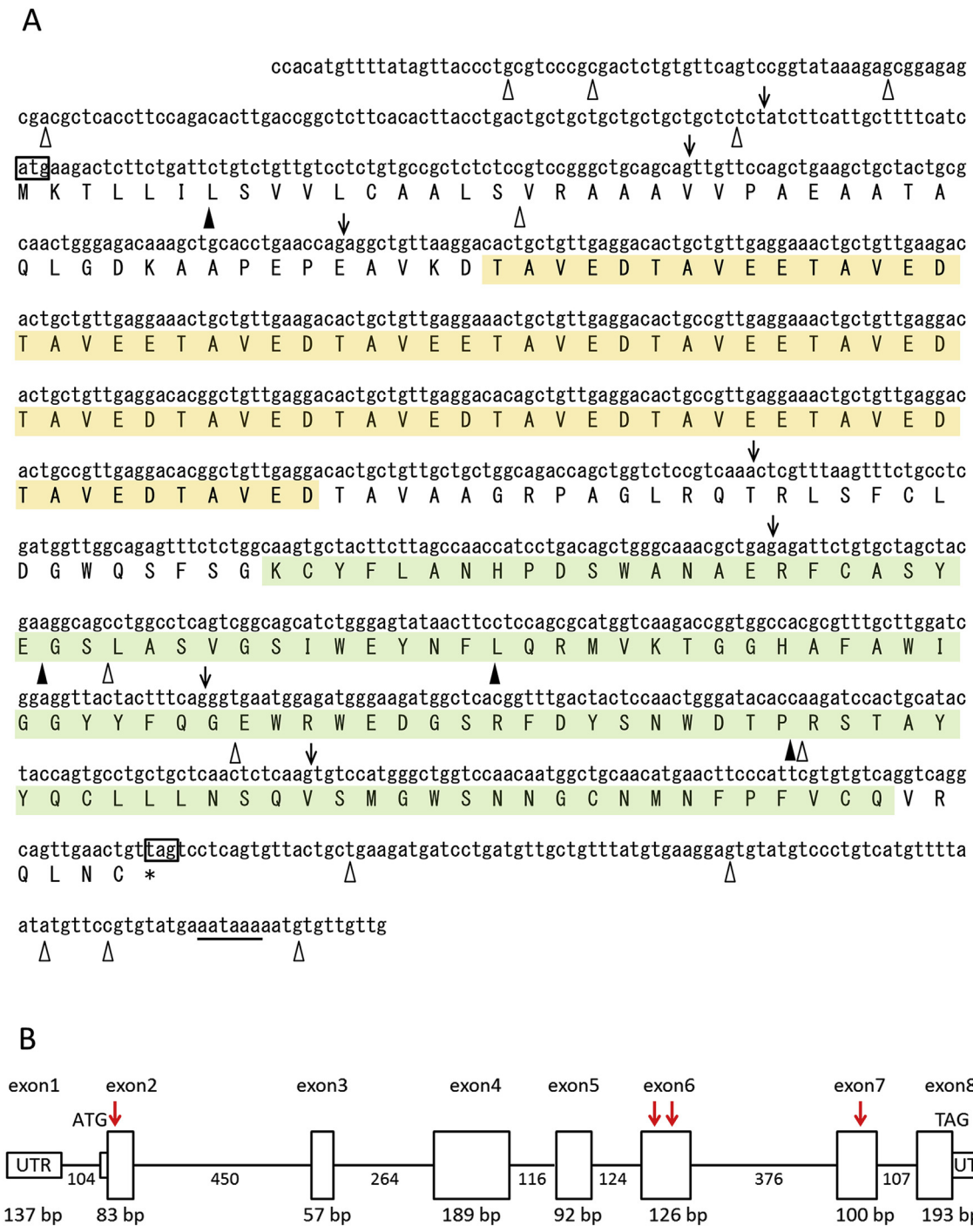

Fig. 2. The candidate gene for Benedenia disease resistance. (A) Sequence of the cDNA of the candidate gene for Benedenia disease resistance. Boxes indicate start and stop codons. The green-shaded region indicates the C-type lectin (CTL)/C-type lectin-like (CTLD) domain, and orange-shaded regions correspond to repetitive elements. The polyadenylation signal is underlined. White and black arrowheads indicate positions of synonymous and non-synonymous SNPs, respectively. Arrows mark exon–intron boundaries. (B) Schematic of the gene structure of the candidate gene. Numbers refer to sizes of exons and introns. Red arrows indicate the positions of SNPs causing amino acid changes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

affect the conformation of the brcl protein. C-type lectins act as pathogen-recognizing receptors and induce various immune reactions (Alvarez-Pellitero, 2008). Our results therefore suggest that the leucine-containing allele of *brcl* has a high affinity to sugar-related molecules on the surface of *B. seriolae* in yellowtails retaining the QTL for Benedenia disease resistance. Further study is needed to functionally analyze the mutated brcl protein. Some lectins suppress the attachment capacity of *B. seriolae* (Yoshinaga et al., 2002). Brcl in skin mucus of yellowtails may thus directly inhibit *B. seriolae* infection. Furthermore, the interaction between brcl and *B. seriolae* activates various immune pathways that may cause the immune system to reject *B. seriolae* on skin. Biochemical characterization and functional analysis of yellowtail *brcl* will need to be performed in future studies.

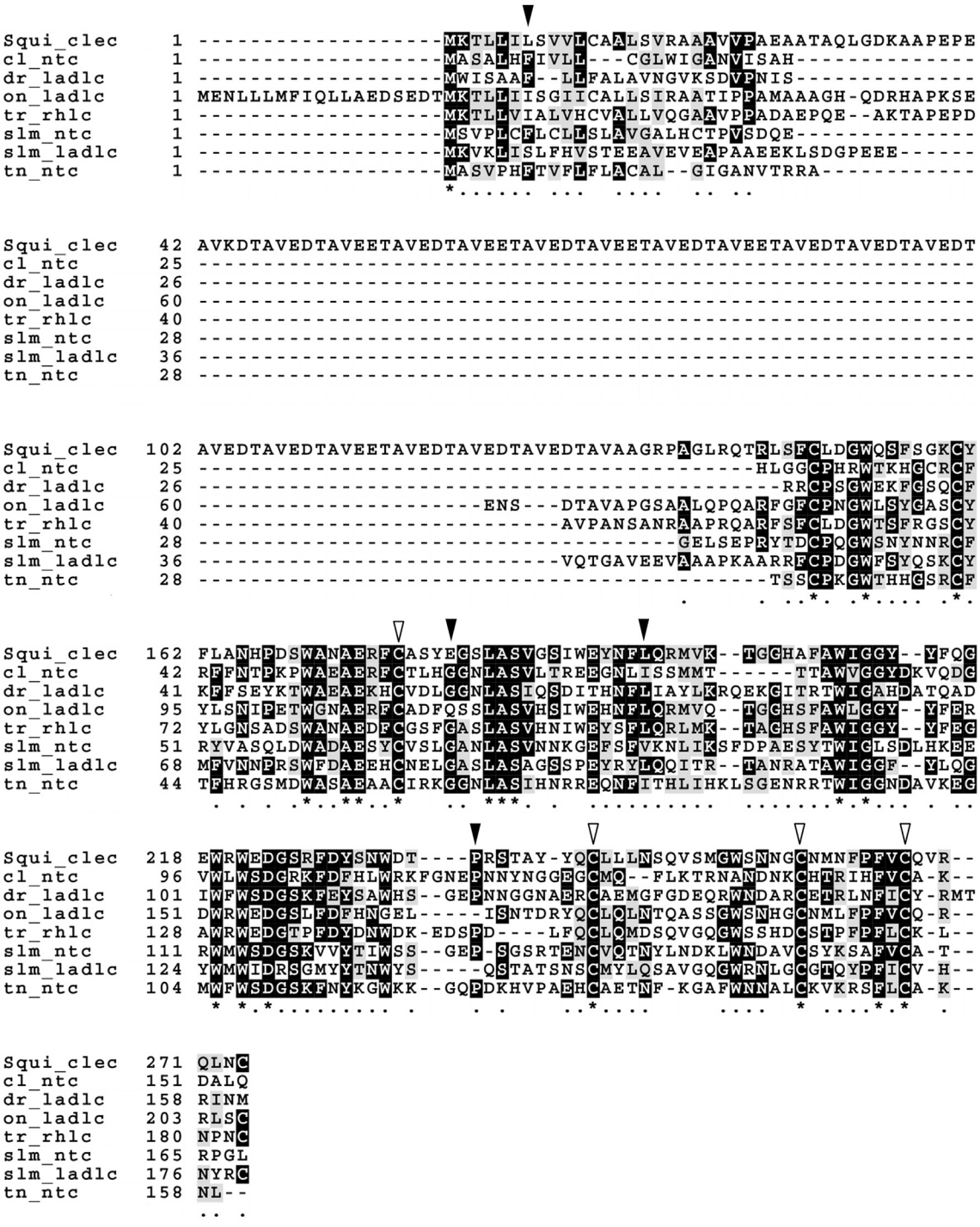

Fig. 3. Multiple alignment of the deduced amino acid sequence of the candidate gene for Benedenia disease resistance (*Squi_clec*) with C-type lectins of other teleost fish. Sequences were aligned using Muscle of Mega 7 and were shaded using BOXSHADE software (https://sourceforge.net/projects/boxshade/). Black arrowheads indicate the positions of SNPs causing amino acid changes. White arrowheads indicate evolutionally conserved cysteine residues critical for protein folding of C-type lectins.

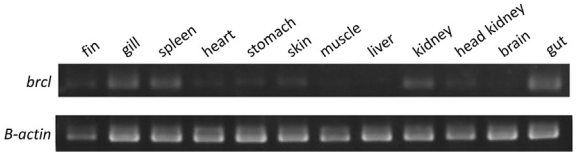

Fig. 4. Expression of yellowtail *brcl* in various tissues according to RT-PCR. The β-actin gene was used as an internal control.

In conclusion, we screened and sequenced BAC clones carrying genomic DNA of the QTL region for Benedenia disease resistance in linkage group Squ2. C-type lectin *brcl* was identified in this QTL region. One SNP located in the C-type lectin domain of *brcl* was associated with *B. seriolae* infection levels. Our findings suggest that *brcl* is a gene controlling Benedenia disease resistance in Japanese yellowtail. These results represent a first step toward identification of molecular mechanisms for resistance against monogenean

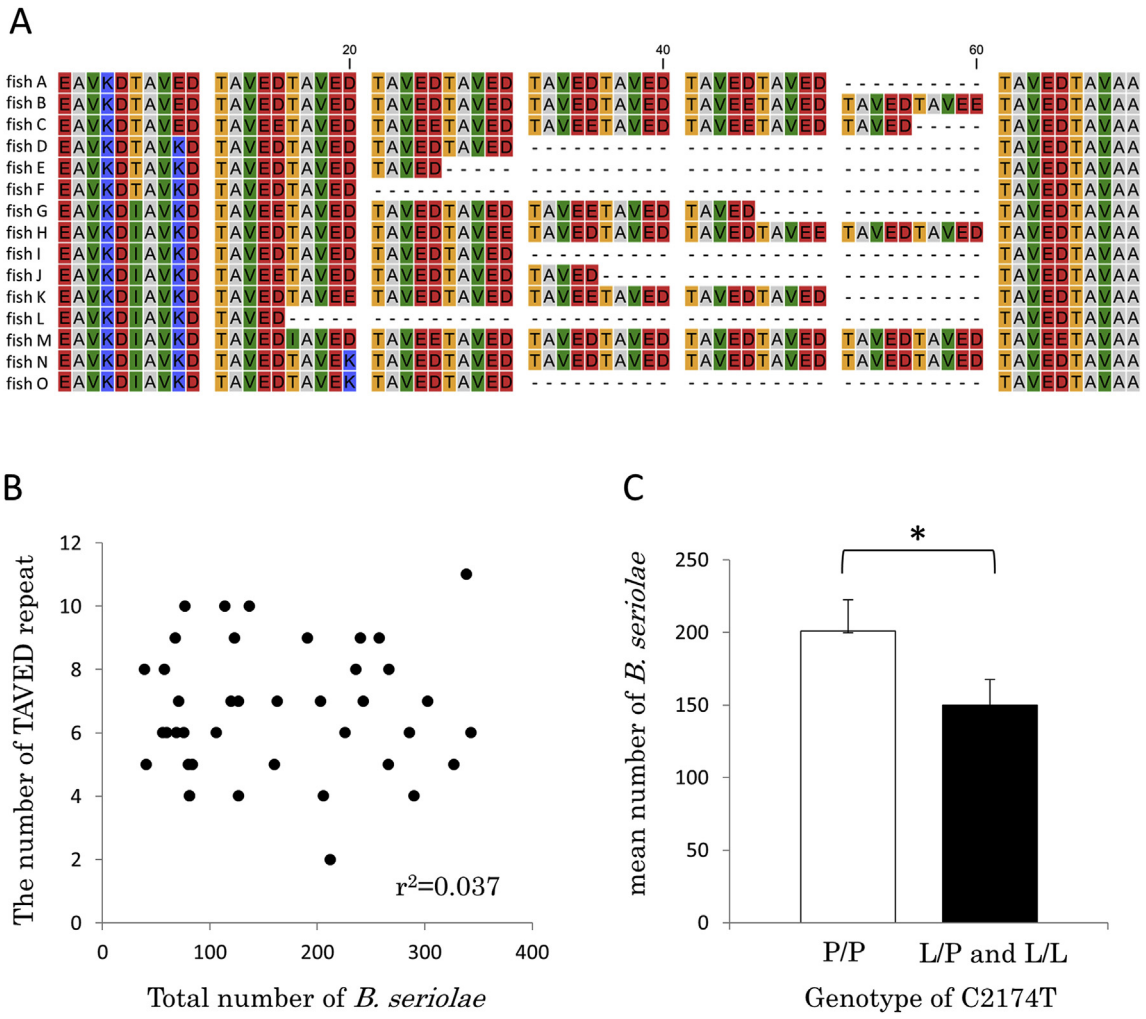

Fig. 5. Sequence polymorphism of *brcl* in a wild yellowtail population. (A) Repeat numbers of the TAVED repetitive element in *brcl* in a wild Japanese yellowtail population. (B) Relationship between TAVED element repeat number (y-axis) and the total number of *B. seriolae* per fish (x-axis). These two factors were not correlated. (C) Association between the SNP C2174T genotype and total numbers of *B. seriolae*. Fewer *B. seriolae* were present in individuals carrying the leucine (L)-containing allele ($P = 0.041$, one-sided Mann-Whitney *U* test).

Table 3
Summary of variations in yellowtail*brcl* gene.

	Location	Position	Base change	Amino acid change
G25A	5′ UTR	25	G to A	N/A
C31T	5′ UTR	31	C toT	N/A
G59T	5′ UTR	59	G to T	N/A
A69T	5′ UTR	69	A to T	N/A
T135G	5′ UTR	135	T to G	N/A
C279A	exon2	279	C to A	L to M
C308T	exon2	308	C to T	no change
A1637C	exon6	1637	A to C	E to D
C1643T	exon6	1643	C to T	no change
C1680A	exon6	1680	C to A	L to I
T2121G	exon7	2121	T to G	no change
C2174T	exon7	2174	C to T	P to L
A2175G	exon7	2175	A to G	no change
T2419G	3′ UTR	2419	T to G	N/A
G2455A	3′ UTR	2455	G to A	N/A
A2438G	3′ UTR	2438	A to G	N/A
C2489T	3′ UTR	2489	C to T	N/A
G2507C	3′ UTR	2507	G to C	N/A

parasite infection in marine aquaculture fish. Analysis of other QTL regions for Benedenia disease resistance, such as linkage groups Squ8 and Squ20, will be needed to further reveal the molecular mechanisms of Benedenia disease resistance. Nevertheless, we have provided the first direct evidence of a gene conferring resistance to parasitic diseases in teleost fish. The polymorphism in the candidate gene may be useful as a DNA marker in marker-assisted selection for Benedenia disease resistance.

In conventional marker-assisted selection in aquaculture, indirect selection for a target trait is based on a DNA marker linked to the trait. The effectiveness of a DNA marker thus depends on the family used for the QTL mapping and is not applicable to other families. If the gene and variants responsible for the target trait can be identified and used as a DNA marker, however, the target trait can be directly selected for using the DNA sequence of the responsible gene. Such gene selection markers are not family-specific and can be applied to any individuals, including members of wild populations as well as closely related species. Our findings will facilitate gene selection for Benedenia disease resistance in Japanese yellowtail and other members of the genus *Seriola*.

Acknowledgments

This study was supported by a grant from the Ministry of

Agriculture, Forestry and Fisheries of Japan (Research Project for the Development of Low-Cost and Stable Production Technology for Promoting the Export of Cultured Yellowtails). We thank Barbara Goodson, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.dci.2017.07.010.

References

Alvarez-Pellitero, P., 2008. Fish immunity and parasite infections: from innate immunity to immunoprophylactic prospects. Vet. Immunol. Immunopathol. 126, 171–198.

Baerwald, M.R., Petersen, J.L., Hedrick, R.P., Schisler, G.J., May, B., 2011. A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). Heredity 106, 920–926.

Bouzidi, M.F., Franchel, J., Tao, Q., Stormo, K., Mraz, A., Nicolas, P., Mouzeyar, S., 2006. A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions. Theor. Appl. Gene 113, 81–89.

Broman, K.W., Wu, H., Sen, Ś., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19, 889–890.

Bruno, W.J., Knill, E., Balding, D.J., Bruce, D.C., Doggett, N.A., Sawhill, W.W., Stallings, R.L., Whittaker, C.C., Torney, D.C., 1995. Efficient pooling designs for library screening. Genomics 26, 21–30.

Buchmann, K., 1999. Immune mechanisms in fish skin against monogeneans - a model. Folia Parasitol. 46, 1–9.

Buchmann, K., Lindenstrøm, T., 2002. Interactions between monogenean parasites and their fish hosts. Int. J. Parasitol. 32, 309–319.

Burge, C., 1998. Modeling dependencies in pre-mRNA splicing signals. In: Salzberg, S., Searls, D., Kasif, S. (Eds.), Computational Methods in Molecular Biology. Elsevier Science, Amsterdam, pp. 127–163.

Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94.

Fuji, K., Koyama, T., Kai, W., Kubota, S., Yoshida, K., Ozaki, A., Aoki, J., Kawabata, Y., Araki, K., Tsuzaki, T., Okamoto, N., Sakamoto, T., 2014. Construction of a high-coverage bacterial artificial chromosome library and comprehensive genetic linkage map of yellowtail *Seriola quinqueradiata*. BMC Res. Notes 7, 200.

Hotez, P.J., Brooker, S., Bethony, J.M., Bottazzi, M.E., Loukas, A., Xiao, S., 2004. Hookworm infection. N. Engl. J. Med. 351, 799–807.

Ishizuka, E.K., Ferreira, M.J., Grund, L.Z., Coutinho, E.M.M., Komegae, E.N., Cassado, A.A., Bortoluci, K.R., Lopes-Ferreira, M., Lima, C., 2012. Role of interplay between IL-4 and IFN-γ in the in regulating M1 macrophage polarization induced by Nattectin. Int. Immunopharmacol. 14, 513–522.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evo. 33, 1870–1874.

Lopes-Ferreira, M., Magalhães, G.S., Fernandez, J.H., Junqueira-de-Azevedo, I.L.M., Ho, P.L., Lima, C., Valente, R.H., Moura-da-Silva, A.M., 2011. Structural and biological characterization of Nattectin, a new C-type lectin from the venomous fish *Thalassophryne nattereri*. Biochimie 93, 971–980.

Lv, C., Zhang, D., Wang, Z., 2016. A novel C-type lectin, Nattectin-like protein, with a wide range of bacterial agglutination activity in large yellow croaker *Larimichthys crocea*. Fish Shellfish Immunol. 50, 231–241.

Majoros, W.H., Pertea, M., Salzberg, S.L., 2004. TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. Bioinformatics 20, 2878–2879.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43 (Database issue), D222–D226.

Ozaki, A., Yoshida, K., Fuji, K., Kubota, S., Kai, W., Aoki, J., Kawabata, Y., Suzuki, J., Akita, K., Koyama, T., Nakagawa, M., Hotta, T., Tsuzaki, T., Okamoto, N., Araki, K., Sakamoto, T., 2013. Quantitative trait loci (QTL) associated with resistance to a monogenean parasite (*Benedenia seriolae*) in yellowtail (*Seriola quinqueradiata*) through genome wide analysis. PLoS One 8, e64987.

Prochazkal, M., 1996. Microsatellite hybrid capture technique for simultaneous isolation of various STR markers. Genome Res. 6, 646–649.

Rauscher, S., Baud, S., Miao, M., Keeley, F.W., Pomès, R., 2006. Proline and Glycine control protein self-organization into elastomeric or amyloid fibrils. Structure 14, 1667–1676.

Sicuro, B., Luzzana, U., 2016. The state of Seriola spp. Other than yellowtail (*S. quinqueradiata*) farming in the world. Rev. Fish. Sci. Aquac. 24, 314–325.

Sinnis, P., Zavala, F., 2012. The skin: where malaria infection and the host immune response begin. Semin. Immunopathol. 34, 787–792.

Smit, A.F.A., Hubley, R., Green, P., 2013. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Stanke, M., Diekhans, M., Baertsch, R., Haussler, D., 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24, 637–644.

Vázquez-Mendoza, A., Carrero, J.C., Rodriguez-Sosa, M., 2013. Parasitic infections: a role for C-Type lectins receptors. Biomed. Res. Int. 2013, 456352.

Yates, A., Ridwan Amode, W.A.M., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Garcín Girón, C., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., HRiat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S.J., Cunningham, F., Aken, B.L., Zerbino, D.R., Flicek, P., 2016. Ensembl (2016). Nucleic Acids Res. 44 (Database issue), D710–D716.

Yoshinaga, T., Nagakura, T., Ogawa, K., Fukuda, Y., Wakabayashi, H., 2002. Attachment-inducing capacities of fish skin epithelial extracts on oncomiracidia of *Benedenia seriolae* (Monogenea: Capsalidae). Int. J. Parasitol. 32, 381–384.

Zelensky, A.N., Gready, J.E., 2005. The C-type lectin-like domain superfamily. FEBS J. 272, 6179–6621.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.