

RESEARCH ARTICLE

Identification of Immune-Related Genes and Development of SSR/SNP Markers from the Spleen Transcriptome of *Schizothorax prenanti*

Hui Luo^{1,2,3,4}, Shijun Xiao³, Hua Ye², Zhengshi Zhang², Changhuan Lv³, Shuming Zheng^{2*}, Zhiyong Wang^{3*}, Xiaoqing Wang^{1,4*}

1 College of Animal Science & Technology, Hunan Agricultural University, Changsha, Hunan, China, **2** Fisheries Breeding and Healthy Cultivation Research Centre, Southwest University, Chongqing, China, **3** Key Laboratory of Healthy Mariculture for the East China Sea, Ministry of Agriculture, P.R. China, Fisheries College, Jimei University, Xiamen, Fujian, China, **4** Collaborative Innovation Center for Efficient and Health Production of Fisheries in Hunan Province, Changde, Hunan, China

☞ These authors contributed equally to this work.

* zhsm199@163.com (SZ); zywang@jmu.edu.cn (ZW); wangxiao@hunau.net (XW)



OPEN ACCESS

Citation: Luo H, Xiao S, Ye H, Zhang Z, Lv C, Zheng S, et al. (2016) Identification of Immune-Related Genes and Development of SSR/SNP Markers from the Spleen Transcriptome of *Schizothorax prenanti*. PLoS ONE 11(3): e0152572. doi:10.1371/journal.pone.0152572

Editor: Shoba Ranganathan, Macquarie University, AUSTRALIA

Received: October 19, 2015

Accepted: March 16, 2016

Published: March 28, 2016

Copyright: © 2016 Luo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: To make the raw sequence available to other researchers, short reads were deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRR2241952, SRR2241953, and SRR2241954, respectively, and the assembled sequences have been submitted as a supplementary file ([S1 Text](#)). The gene lists of each category that have matches in NCBI Nr/Nt and Swiss-prot have been submitted as supplementary files ([S2 Text](#)).

Funding: This research was supported by the National Natural Science Foundation of China

Abstract

Schizothorax prenanti (*S. prenanti*) is mainly distributed in the upstream regions of the Yangtze River and its tributaries in China. This species is indigenous and commercially important. However, in recent years, wild populations and aquacultures have faced the serious challenges of germplasm variation loss and an increased susceptibility to a range of pathogens. Currently, the genetics and immune mechanisms of *S. prenanti* are unknown, partly due to a lack of genome and transcriptome information. Here, we sought to identify genes related to immune functions and to identify molecular markers to study the function of these genes and for trait mapping. To this end, the transcriptome from spleen tissues of *S. prenanti* was analyzed and sequenced. Using paired-end reads from the Illumina HiSeq2500 platform, 48,517 transcripts were isolated from the spleen transcriptome. These transcripts could be clustered into 37,785 unigenes with an N50 length of 2,539 bp. The majority of the unigenes (35,653, 94.4%) were successfully annotated using non-redundant nucleotide sequence analysis (nt), and the non-redundant protein (nr), Swiss-Prot, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. KEGG pathway assignment identified more than 500 immune-related genes. Furthermore, 7,545 putative simple sequence repeats (SSRs), 857,535 single nucleotide polymorphisms (SNPs), and 53,481 insertion/deletion (InDels) were detected from the transcriptome. This is the first reported high-throughput transcriptome analysis of *S. prenanti*, and it provides valuable genetic resources for the investigation of immune mechanisms, conservation of germplasm, and molecular marker-assisted breeding of *S. prenanti*.

(31402302, U1205122), Scientific Research Fund of Hunan Provincial Education Department (15A087), Fundamental Research Funds for the Central Universities (XDK2015C034, XDK2014C152), and the project Supported by Scientific research initiation project aided by special fund, Southwest university Rongchang campus (20700208). The authors thank Meishan Aquaculture Company for providing *S. prenanti*. Thanks also to Lingbing Zeng, in Yangtze River Fisheries Research Institute Chinese Academy of Fishery Sciences, and Bin Chen, in College of Animal Science & Technology, Hunan Agricultural University, for providing constructive suggestions on experiments.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Fish of the subfamily Schizothoracinae (Teleostei: Cyprinidae) are mainly distributed in rivers on the Qinghai-Tibetan Plateau and its peripheral areas [1] in China. These fish are well adapted to the harsh conditions of the Plateau [2] and have evolved many specific traits to adapt to an environment that exposes them to low temperatures, high levels of radiation, and hypoxia [1]. As a consequence, Schizothoracinae fish are regarded as excellent models to study high altitude adaptations in animals [3]. *Schizothorax prenanti* is a member of the subfamily Schizothoracinae and is mainly distributed in the upstream regions of the Yangtze River and its tributaries in China. The species lives in a cold-water environment with a gravel riverbed [4]. Given the extreme environmental changes in its habitat areas, *S. prenanti* offers an excellent model to investigate the effects of historical and contemporary environmental changes [5]. *S. prenanti* is also important commercially in west China because of its high flesh quality and good flavor. In recent years, it has become one of the most important cold-water aquaculture species in China. However, overly dense stocking levels and rapid expansion of aquaculture has led to several problems affecting the sustainable development of the industry, such as the frequent outbreak of infectious diseases [6,7]. Concurrently, the wild resources and populations of *S. prenanti* have rapidly declined because of water pollution and the construction of hydropower stations [8–10]. In order to reduce economic losses caused by infective agents and to protect germplasm resources, it will be necessary to identify genes that have a role in economically important traits, including the immune system. The development of molecular markers for use in selective breeding programs will also be of importance.

Transcriptome sequences can be used to identify genes and to develop genetic molecular markers [11]. The recent advances in next generation sequencing (NGS) technologies have enabled the transcriptomes of non-model species to be analyzed in a high-throughput manner. The ability to sequence all transcripts in one experiment and to assess gene expression levels has led to the application of high-throughput methods to species of importance for aquaculture [12–15]. To date, RNA sequencing (RNA-Seq) has been employed in a wide range of aquatic animals [11,16] to examine immune responses [17–19], growth and development [20–23], evolution [3,24,25], and toxicology [26,27]. Taxonomy, diversity, geographic distribution, and disease prevention have been examined in *S. prenanti* [28]. In recent years, microsatellite markers have been developed from a small number of expressed sequence tags (ESTs), and several genes involved in growth, metabolism, and immunity have been characterized [5,28–33]. However, a detailed transcriptome analysis has still not been undertaken in this important fish species. The identification of genes involved in the immune system and the development of molecular markers will significantly advance investigations into the *S. prenanti* immune defense mechanisms and also contribute to selective breeding programs for this species.

In this study, we used the Illumina Hiseq2500 platform to analyze the spleen transcriptome of *S. prenanti*. Genes involved in immune pathways were identified; the majority of these are reported for the first time. In addition, we identified and analyzed SSR markers and small variants (SNPs/InDels) in the transcriptome. This is the first analysis of a transcriptome from *S. prenanti* and will enable large-scale molecular marker development.

Materials and Methods

Fish and sample collection

Eighteen one-year-old *S. prenanti* individuals (average body weight 105 g, average body length 18.7 cm) were collected from an aquaculture farm in Meishan, Sichuan Province, China. The fish were reared in a recirculating freshwater system (18 individuals/tank; tank dimensions 100 × 48 × 60 cm) at the Fisheries Breeding and Healthy Cultivation Research Centre of

Southwest University. The fish were maintained at $19 \pm 1^\circ\text{C}$ in aerated water for two weeks before the experiment. They were fed a commercial diet (Sichuan Giastar Group; particle diameter 2 mm) twice a day. The experimental protocols used here were approved by the institutional animal care and use committee of Southwest University. In order to reduce stress, fish were anesthetized using tricaine methanesulfonate (MS222) before dissection. Spleen tissues were randomly mixed into three samples (6 individuals for each sample) and stored in 1 mL Sample Protector for RNA (TaKaRa, Dalian, China) at 4°C overnight. The samples were then transferred to a -80°C ultra-low freezer until preparation of RNA.

RNA extraction, cDNA library construction, and Illumina sequencing

Total RNA was extracted using TRIzol reagent (Invitrogen, USA) and incubated for 1 h at 37°C with 10 units of DNase I (TaKaRa, Dalian, China) to eliminate genomic DNA. RNA quality and quantity were analyzed using a BioAnalyzer 2100 (Agilent Technology, Santa Clara, CA) and NanoDrop 2000 spectrophotometer (Infinigen Biotechnology Inc., City of Industry, CA), respectively. In order to evaluate the reliability of the libraries, we constructed three cDNA libraries using spleen tissue from six fish per RNA library. Poly(A)⁺ RNA was purified with oligo(dT) magnetic beads and fragmented into short sequences. First-strand cDNA was synthesized using random hexamer primers and Superscript III (Invitrogen, Carlsbad, CA, USA); this was followed by second-strand cDNA synthesis, end repair, and adaptor ligation. Finally, libraries with insert lengths of ~ 280 bp were created by PCR amplification and purification. Each library was sequenced on an Illumina HiSeq2500 in 125PE mode (Illumina Inc., San Diego, CA, USA). Short reads were deposited in the NCBI Sequence Read Archive (SRA) under Accession numbers SRR2241952, SRR2241953, and SRR2241954.

Transcriptome de novo assembly and annotation

In order to ensure reliable assembly results, the quality of raw reads was checked by FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and filtered by the high-throughput quality control (HTQC) toolkit [34]. The following quality-filtering criteria were applied: 1) the 5-bp window was omitted if the average quality was lower than 20; 2) reads were removed if the percentage of unknown bases was higher than 10%; 3) read pairs were filtered if any one read end was shorter than 50 bp. The resulting cleaned reads were used in the following bioinformatics pipelines. The Trinity package was used for transcript assembly using default parameters [35]. The assembled transcripts were then processed through the Evigene package (<http://arthropods.eugenes.org/EvidentialGene/>) to eliminate sequence redundancy with default parameters [36]. The resulting transcripts that showed significant similarities ($>90\%$) were then clustered and the longest transcripts for each group were selected as representative unigenes, which were then used for functional annotation. Sequence-length statistics of the assembled transcriptome were performed using our own Python scripts. The assembled sequences have been submitted as [S1 Text](#). For gene annotation, BLAST package (with an E-value threshold of 1×10^{-5}) of all unigenes were performed in the National Center for Biotechnology Information (NCBI) non-redundant nucleotide sequence (nt) and non-redundant protein (nr) databases [37]. Transcripts were further annotated using the Swiss-Prot, Gene Ontology (GO), EC (Enzyme Code), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases with Blast2GO [38]. The gene annotations in NCBI Nr/Nt and Swiss-prot were available in [S2 Text](#).

SSR, SNP, and InDel discovery

To develop SSR markers from the transcriptome of *S. prenanti*, sequences with repeat unit lengths from di- to hexa-nucleotides were detected using the MicroSATellite application

(MISA, <http://pgrc.ipk-gatersleben.de/misa/>). The parameters were set to identify di-, tri-, tetra-, penta-, and hexa-nucleotide motifs with a minimum of six, five, five, five, and five repeats, respectively. To eliminate non-genomic artificial SSR loci caused by the sequence complexity of the transcripts, mono-nucleotide SSR markers were excluded in this work. For the development of putative SNPs and InDels, BWA 0.7.6a [39], SAMTools 1.19 [40] and GATK 2.8–1 [41] pipeline methods were applied. Only SNP and InDel markers with a depth greater than 5 and a quality score higher than 100 were selected as reliable loci for subsequent analysis.

Immune gene PCR validation

To validate the reliability of the assembly of the transcriptome, 15 annotated unigenes related to immunity were selected and validated using PCR experiments. Primer set was designed based on RNA-Seq unigenes sequences by Primer Premier 5.0. Total RNA was extracted from spleen tissues of 18 fish (3 biological replicate sample pools (n = 6 fish for each pools)). First strand cDNA was synthesized from 4 µg total RNA and used as a template for PCR with gene-specific primers. The PCR analysis was performed on the Applied Biosystems 2720 Thermal Cycler using Ex Takara DNA Polymerase according to the manufacture's protocol. PCR amplifications were performed in 10 µL reactions, containing genomic DNA 1.0 µL, 10× PCR buffer 1.0 µL, 15 mM MgCl₂ 1.0 µL, 0.2 µM dNTPs, 0.2 µM of each primer, 5 U Taq enzymes 0.1 µL and ddH₂O 7.3 µL. The PCR reaction procedure was that 94°C for 5 min, 30 cycles at 94°C for 0.5 min, annealing temperature (from 55°C to 62°C) for 0.5 min, 72°C for 0.5 min and a final extension step at 72°C for 10 min.

Results and Discussion

Transcriptome sequencing and assembly

The three cDNA libraries yielded a total of 99.6 million raw reads with a read length of 125 bp, resulting in a total of 12.5 Gb. Sequencing quality was assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to determine the Phred quality score of each base in the raw reads. More than 93% of the bases had a Phred quality greater than 20, with 88% having a score greater than 30 (S1 Fig). After read quality evaluation and length trimming (see [Method and Material](#) for detail), 97.3 Mb cleaned pair-ended reads were obtained and used for the bioinformatics analysis.

The Trinity package [35] was used for *de novo* transcriptome assembly from the cleaned reads. An additional step with Evigene [36] was used to remove redundancy in the assembled contigs. As a result, we obtained 48,517 transcripts ranging from 201 to 27,365 bp with an average length of 1,491 bp (Table 1, Fig 1). The transcripts clustered into 37,785 unigenes with an average length of 1,323 bp (ranging from 201 bp to 27,365 bp) (Table 1). The N50 lengths of transcripts and unigenes were 2,700 and 2,539 bp, respectively; these estimates are comparable with those reported previously for other fish species [11,42]. Approximately 50% of transcripts ranged from 201 to 500 bp (Fig 1). An estimated 21,704 (44.7%) transcripts exceeded 1,000 bp and 18,199 (26.3%) exceeded 2,000 bp; these proportions are consistent with a previous study [43].

Table 1. Assembled transcripts and unigenes obtained from transcriptome analysis.

Terms	Transcripts	Unigenes
Total number	48,517	37,785
Shortest length (bp)	201	201
Longest length (bp)	27,365	27,365
N50 length (bp)	2,700	2,539

doi:10.1371/journal.pone.0152572.t001

To validate the reliability of the assembly of the transcriptome, 15 annotated unigenes related to immunity were selected to be analyzed using PCR experiments. Primers were designed based on corresponding unigenes sequences, their putative gene names, primer sequences and expected PCR product sizes are shown in [Table 2](#). As we expected, PCR experiments exhibited amplicons of expected sizes for all 15 unigenes ([Fig 2](#)). The results not only testified the reliability of the assembly of the transcriptome and transcript annotation, but also indicated that it could be useful for further research.

Transcript annotation

Functional annotation of the transcriptome was carried out by searching the transcripts against nucleotide and protein databases. As a result, 34,813 (92.1%), 25,222 (66.7%), and 20,765

transcript length distribution

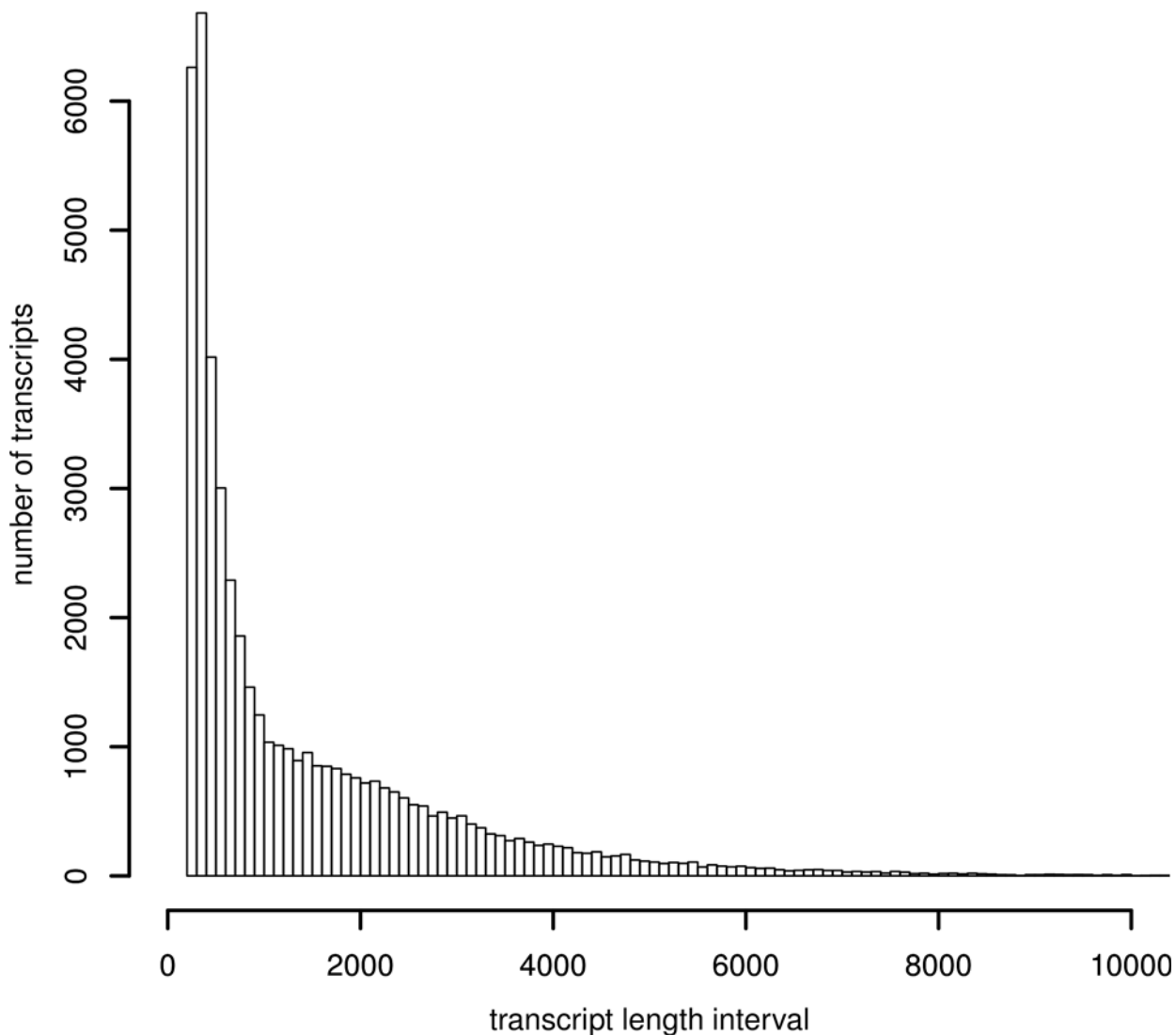


Fig 1. Length distribution of assembled transcripts. Assembled transcript numbers (Y-axis) were plot against length interval (X-axis).

doi:10.1371/journal.pone.0152572.g001

Table 2. Putative gene name and primer sequences and the expected size for PCR of the 15 unigenes.

Number	Putative gene name	Forward primer	Reverse primer	Product size (bp)
unigene8059	Interleukin 12A	TGCTGGAGATGGACCACTTG	ACACCGTTTGCCTCCTGAT	81
unigene11048	Complement C1q subcomponent subunit A	TTTTCTTTGGGTGCTGCTTC	GACTCCATTACGCCATCTTT	82
unigene25066	Toll-like receptor 5	TGATATTAAGTGGGTAGAAAGAGCG	CAAAGCAGCAGCGGAGTGT	83
unigene24999	Major histocompatibility complex, class I	CCCATCCTAAACCAGTACACCA	GCCTTTTCGGACCTCCTCAT	99
unigene10116	C-C motif chemokine 21	ATGGTTTACTGGTGTCTATGCTTCT	CCAAGTGACAATAACTGGGTGAG	101
unigene24705	Interleukin 6 receptor	CCACCTGTAACCATAAGAAAAGAAC	TTGCTCAAAATCTGTCCCAT	120
unigene12568	Myeloid differentiation primary response protein MyD88	AGACTGAACGCAACCTGAAGC	GAAATGACCACCACCATCCTC	121
unigene19017	Interleukin 1 receptor type II	TACAGGAGATTGCGAGTGAACA	AGATGGGTGAGACGGGAGGA	126
unigene16797	Heat shock 70kDa protein 1/8	TGGAGGGAAGCCGAAAG	TCATTGAAATAGGCAGGAACTG	166
unigene_22539	Toll-like receptor 4	AACAACATGCCCATGACCTTT	GCTTATCCGCCATCCCTATCT	175
unigene17508	Tumor necrosis factor, alpha-induced protein 3	ACAAATCAAACCTCGTCCTCG	CACAGCCCGTCATACTCCAC	182
unigene11117	Toll-interacting protein	ATCCAGAGCACTGTCCCACC	TGCCTTCTTTGTCATCACCCCT	184
unigene24689	Integrin alpha 5	GGAGTTGGGTCGGGTCTATG	TGAGCACCAGTCCTTGTGTT	200
unigene17711	TNF receptor-associated factor 6	CTCAGTTGCCGCCATTCT	TCGGGTGTCTCCTCCAAA	232
unigene23056	Toll-like receptor 3	GCAGTTCCAACGGACTTACCC	GCCCATTAGCTTCAGCCTATT	283

doi:10.1371/journal.pone.0152572.t002

(54.9%) unigenes showed significant similarities (E value 1×10^{-5}) to the NCBI nucleotide (nt), protein (nr) and Swiss-Prot databases, respectively (Table 3). We found that the majority (19,272, ~76%) of homology hits in the nr search were to zebrafish (*Danio rerio*), a result that is consistent with the fact that both *S. prenanthi* and zebrafish belong to the Cyprinidae family and therefore have a close evolutionary relationship. In addition, 4.8% (1,219) of hits in the nr database were from *Astyanax mexicanus*, 2.4% (617) from *Oncorhynchus mykiss*, 1.8% (463) from *Oreochromis niloticus*, and 1.2% (315) from *Cyprinus carpio* (Fig 3).

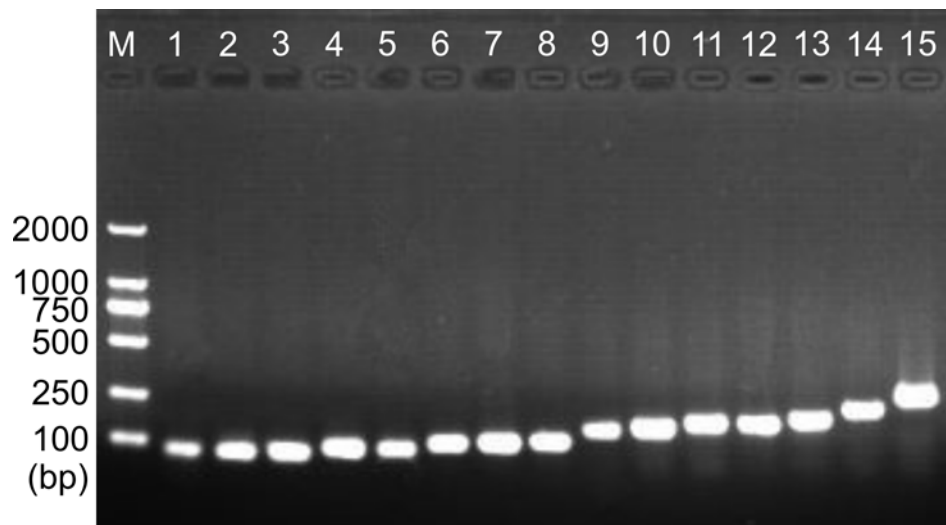


Fig 2. PCR amplification and agarose gel (1%) electrophoresis of 15 unigenes. The corresponding run lanes of unigene 8059, unigene 11048, unigene 25066, unigene 24999, unigene 10116, unigene 24705, unigene 12568, unigene 19017, unigene 16797, unigene 22539, unigene 17508, unigene 11117, unigene 24689, unigene 17711 and unigene 23056 are from 1 to 15, respectively.

doi:10.1371/journal.pone.0152572.g002

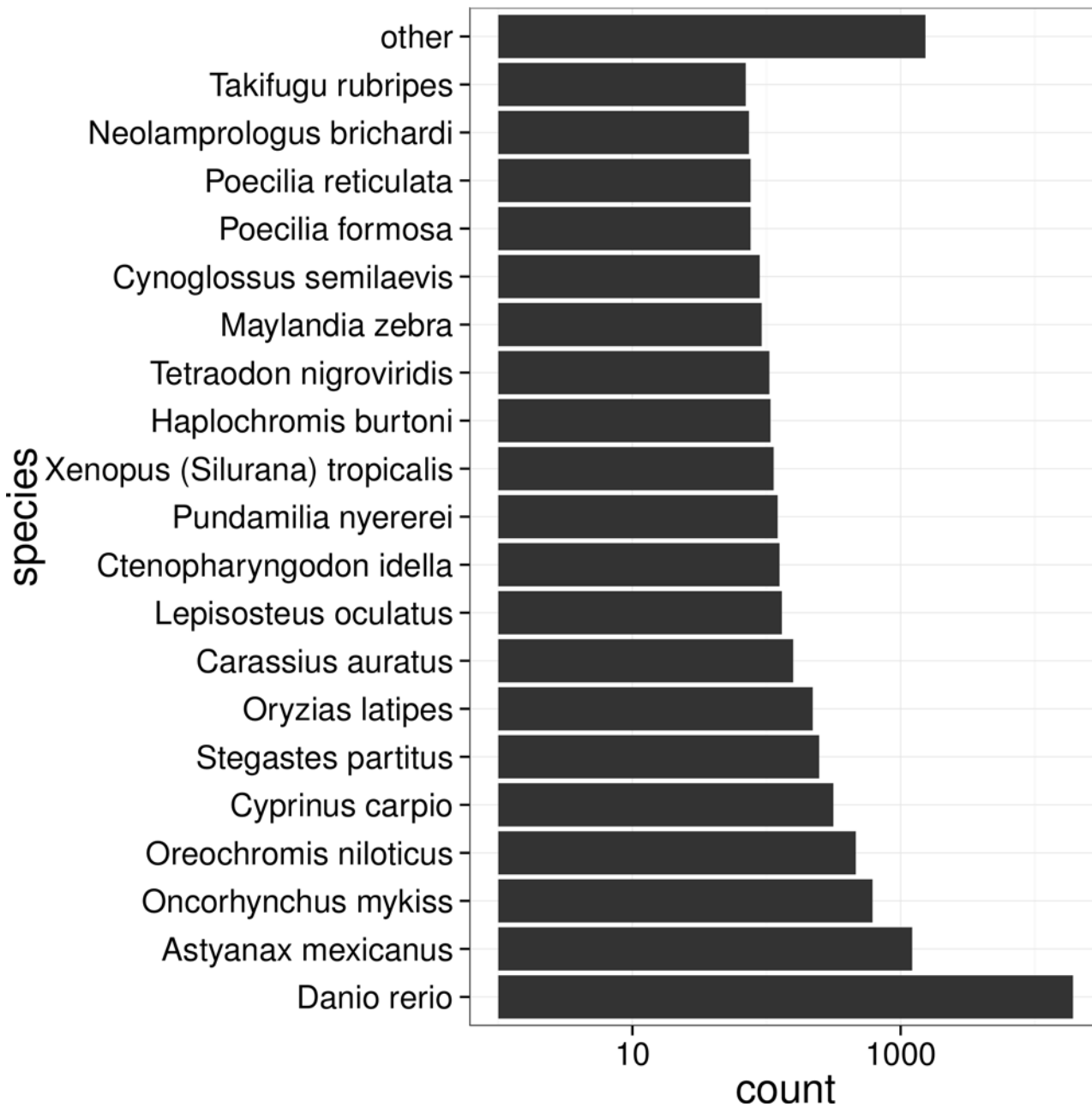


Fig 3. Species distribution for NCBI nr databases annotation. Note that only the best hits for unigenes were used in the analysis.

doi:10.1371/journal.pone.0152572.g003

Table 3. Unigenes annotation by various databases.

Database	Hit number	Percentage (%)
Nr	25,222	66.7
Nt	34,813	92.1
Swiss-prot	20,765	54.9
GO	21,760	57.6
KEGG	5,445	14.4
EC	2,648	7.1
Total	35,653	94.4

doi:10.1371/journal.pone.0152572.t003

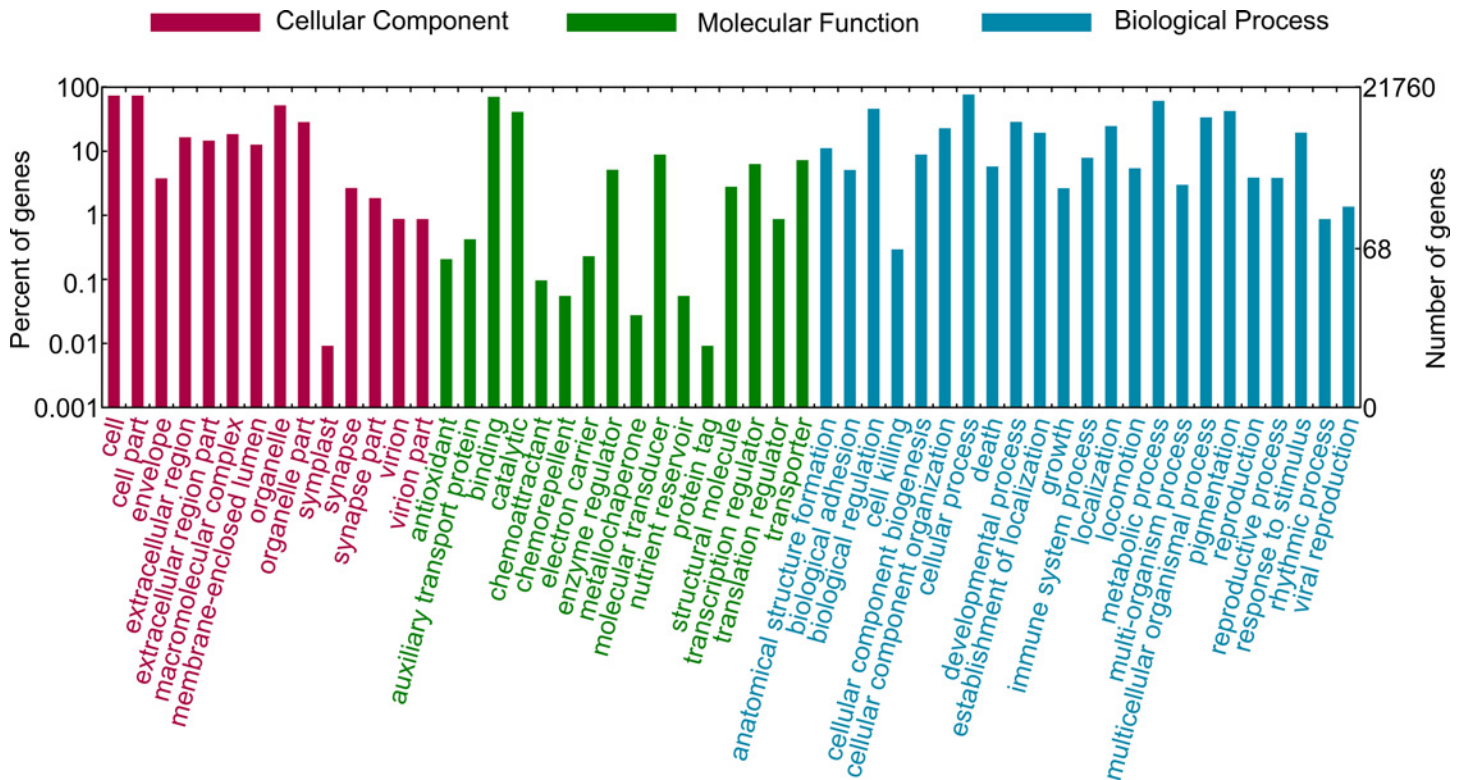


Fig 4. GO annotation of prenat's schizothoracin transcriptome. Unigenes were annotated by Gene Ontology (GO) terms which belong to three main categories: biological process, cellular component, or molecular function.

doi:10.1371/journal.pone.0152572.g004

The potential functions of the unigenes were determined using Gene Ontology (GO) databases. GO classification generally assigns the functions of genes and their products in organisms. The spleen transcriptome unigenes were annotated to three major GO categories—16,982 unigenes (44.9%) were assigned to Cell Component (CC), 19,229 (50.9%) to Molecular Function (MF), and 18,948 unigenes (50.1%) to Biological Process (BP) (Table 3; Fig 4). The most enriched components in CC terms were cell (16,101 unigenes, GO: 0005623), cell part (16,100 unigenes, GO: 0044464), and organelle (11,324 unigenes, GO: 0043226). For MF terms, a large number of unigenes were assigned to binding (15,426 unigenes, GO: 0005488), catalytic activity (8,963 unigenes, GO: 0003824), and molecular transducer (1,930 unigenes, GO: 0060089). In the BP category, most of the unigenes were related to the terms cellular process (16,734 unigenes, GO: 0009987), metabolic process (13,315 unigenes, GO: 0008152), and biological regulation (10,027 unigenes, GO: 0065007) (Fig 4). These results indicated that the annotated unigenes were assigned to various terms of the biological process category as has been reported in previous transcriptome analyses of the larger yellow croaker (*Larimichthys crocea*), the naked carp (*Gymnocypris przewalskii*), and the blunt snout bream (*Megalobrama amblycephala*) [11,42,44].

Next, the unigenes were mapped to reference canonical pathways in the KEGG database; 5,445 (14.4%) unigenes were assigned to KEGG Orthology (KO) terms and grouped into 365 pathways. The annotated pathways were clustered into six major categories: Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, and Human Diseases. The detailed pathways and distributions in each major pathway category are shown in Fig 5A. KEGG pathway-based analysis facilitates the

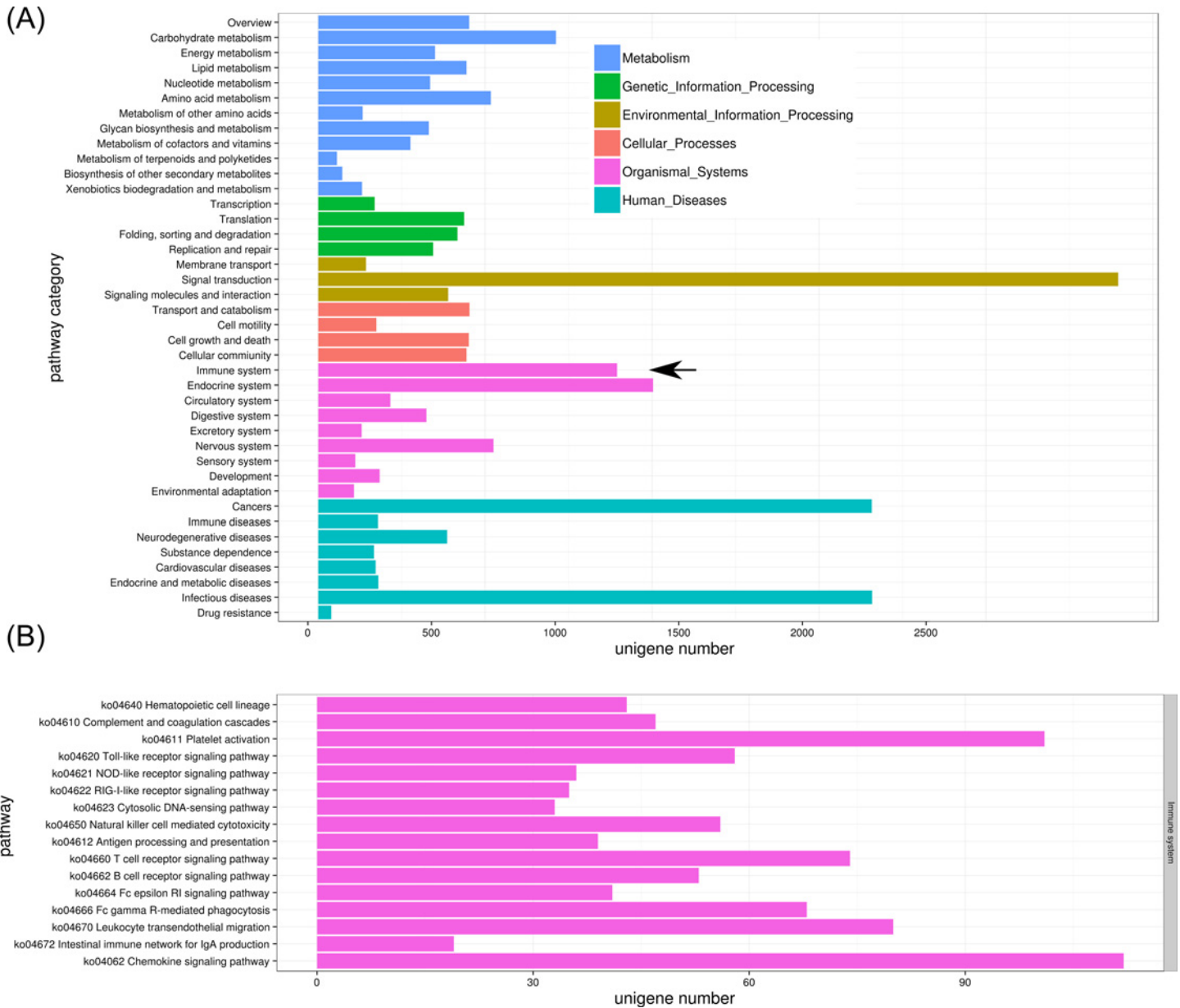


Fig 5. Identified KEGG pathways of assembled unigenes. Unigene numbers distribution in six major categories: Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, Human Diseases. Immune system was indicated by black arrow(A); the detailed unigenes distribution in 16 immune-related pathways (B).

doi:10.1371/journal.pone.0152572.g005

systematic study of intricate metabolic pathways and the biological behavior of functional molecules. The largest pathway in the present annotation was ‘Pathways in cancer’ (ko05200), which contained 269 unigenes. Other major pathways were ‘PI3K-Akt_signaling_pathway’ (218 unigenes, ko04151), ‘Purine metabolism’ (200 unigenes, ko00230), ‘HTLV-I infection’ (193 unigenes, ko05166), and ‘Biosynthesis of amino acids’ (183 unigenes, ko01230). These results are similar to those reported for the Eastern Oyster (*Crassostrea virginica*) [18]. In addition, Enzyme Commission (ECs) were assigned to 2,648 (7.1%) unigenes according to KEGG mapping results.

Identification of immune-related genes

KEGG pathway assignments were used to identify functional unigenes involved in immune processes and their interactions. In total, 511 immune-related unigenes were identified in 16 KEGG immune pathways (Fig 5B). Many of these unigenes are reported for the first time in *S. prenanti*. The number of immune genes identified here is similar to that reported in the Miiuy Croaker (*Miichthys miiuy*) [45] and rainbow trout (*Oncorhynchus mykiss*) [46]. Two functional subcategories of immune response, 'Chemokine signaling pathway' (112 unigenes, ko04062) and 'Platelet activation' (101 unigenes, ko04611), had most unigenes. Other important immune pathways with large numbers of unigenes included 'Toll-like receptor signaling pathway' (58 unigenes, ko04620), 'Complement and coagulation cascades' (47 unigenes, ko04610), 'Leukocyte transendothelial migration' (80 unigenes, ko04670), 'T-cell receptor signaling pathway' (74 unigenes, ko04660), 'B-cell receptor signaling pathways' (53 unigenes, ko04662), and 'Fc gamma R-mediated phagocytosis' (68 unigenes, ko04666). The detailed pathways, KO terms, putative functions, and expression of these immune-related genes are summarized in S1 Table. To systematically identify genes involved in these pathways, we identified genes in several representative networks, including Toll-like receptor signaling pathway, Complement and coagulation cascades, and Chemokine signaling pathway.

Toll-like receptor signaling pathway

Toll-like receptors (TLRs) are a type of pattern recognition receptor (PRR) and are key components of the innate immune system. They were the first PRRs to be characterized [47]. TLRs detect the presence of pathogens through recognition of pathogen associated molecular patterns (PAMPs) and trigger innate immune responses [46,48]. To date, 17 TLRs have been identified in teleost species [49]. Bacterial PAMPs are mainly detected by TLR1, TLR2, TLR4, TLR5, TLR6, TLR7, and TLR9 [50] and nucleic acids are recognized by TLR7/8, TLR3, and TLR9 [51]. Here, we identified five unigenes showing high similarity to TLR3, TLR4, TLR5, TLR7, and TLR8. Notably, TLR4 has been reported only in zebrafish and the Chinese rare minnow (*Gobiocypris rarus*) and not in other teleosts [49,52]. Our results suggest that *S. prenanti* is likely to be a teleost species that possesses TLR4, similar to zebrafish and Chinese rare minnow. TLR4 plays an important role in functional links between TLRs and complement in mammalian immune systems. In a tissue-damage model of the mouse intestine, it was observed that TLR4 stimulation could modulate local C3 and factor B synthesis [53]. TLR4-mediated signaling is also involved in T-helper 17 cell differentiation [54]; however TLR4 has rarely been reported in fish and the cross-talk between complement and TLRs has not yet been described in teleosts. We found that *S. prenanti* had a TLR4 homologous to that in zebrafish; therefore, investigation of the teleost-specific relationship between complement and TLRs will be an interesting topic in fish immunological studies. The gene identification from the *S. prenanti* transcriptome may not represent the complete set of TLRs in this species and further studies are needed to identify more TLR genes for a better understanding of the molecular immune mechanism. All unigenes involved in the Toll-like receptor signaling pathway are listed in S2 Fig and S1 Table.

Complement and coagulation cascades

In general, the complement system is considered the first line of defense against microbial invaders [55]. The complement system includes over 20 soluble and cell-surface proteins that respond to the presence of foreign antigens by activating a regulated cascade of reactions [56]. There are three different ways to initiate the complement system on the surface of invading pathogens—the classical, alternative, and mannan-binding lectin pathways. Although activation of each pathway depends on different factors, they all produce the same anti-infection effects [57]. In this study, three unigenes were found to have high similarity to C3, C4, and C5.

Complement components C3, C4, and C5 belong to the alpha-2 macroglobulin superfamily of thioester-containing proteins [58]. C3 is the key element of the complement system and when activated can split into C3a and C3b in the three different pathways [59]. C4 serves as a link in the initiation of the lectin and classical pathways [56]. C5 plays a pivotal role in the formation of the membrane attack complex (MAC), which results in cell lysis [59,60]. In addition to the above complement members, we also found many other key genes in complement and coagulation, such as C6, C7, C8, C9, complement C1q subcomponent subunit A (C1QA), complement C1q subcomponent subunit C (C1QG), Factor B, and Factor D. However, excessive complement activation can cause serious damage to the host tissue, resulting in anaphylaxis and cell damage. Therefore, complement activation needs to be controlled in multiple reaction steps by several different regulatory elements. For instance, C1-inhibitor (C1INH), factor H, membrane-cofactor protein (MCP), C4-binding protein (C4bp), S-protein, and CD59 are key regulatory elements [55]. In this study, we identified several regulatory factors: C1-inhibitor (C1INH), S-protein, and CD59. Information on unigenes involved in Complement and coagulation cascades is included in [S3 Fig](#) and [S1 Table](#).

Chemokine signaling pathway

Chemokines are chemotactic cytokine family components that are secreted by tissues at an early stage of infection. These secreted chemokines are small heparin-binding molecules that recruit neutrophils, monocytes, and other effector cells from vessels towards the focus of infection [57]. The two most important and studied families of chemokines are CC chemokines (characterized by two adjacent cysteine residues next to the N-terminus) and CXC chemokines (characterized by two cysteine residues separated by one amino acid next to the N-terminus). These chemokine families and their receptors have been found in many bony fish species [57]. Here, we found seven, five, one, and four unigenes showing high similarity to CC, CC chemokine receptors (CCR), CXC, and CXC chemokine receptors (CXCR), respectively. The number of predicted CC chemokines from *S. prenanti* was significantly less than that (26) in catfish [61], suggesting that further transcriptome analysis under infectious conditions will be needed to identify further genes in the Chemokine signaling pathway. Several CXC chemokines, including CXC12 and CXC14, have been identified in fish [62,63]; however, only one *S. prenanti* unigene showed high similarity to CXC11 and none were homologous to CXC12 or CXC14. In addition, we identified three CXCRs (CXCR3, CXCR4, and CXCR5) in *S. prenanti*. Unigenes involved in the chemokine signaling pathway are listed in [S4 Fig](#) and [S1 Table](#).

Although some representative immune genes, such as myeloid differentiation factor 88 (MYD88) gene [33], were identified here, most of the immune-related genes found in this study have not previously been reported in *S. prenanti*. These immune-related genes offer a valuable resource for further gene function investigations to explore the detailed immune mechanism in this species. We also noted that some important immune genes identified in other fish were absent in our transcriptome; the absence of these genes may indicate that they are species-specific; alternatively, since all the sampled fish in this study were healthy, then genes with a zero or low expression level would not be included in the RNA-Seq and gene assembly. It will be necessary to repeat this analysis under different physiological conditions to identify further genes and to examine their expression patterns.

SSR, SNP, and InDel discovery

SSRs are useful molecular markers for genetic and breeding studies. The development of genetic markers is the first step in the application of genomic resources to improve a broodstock [64]. At the present time, only a few SSR markers are available for *S. prenanti* [5,30].

Table 4. Repeat numbers and unit length distribution of putative SSR markers in the transcriptome.

Repeat numbers	Motif length					Total	Percent %
	Di	Tri	Tetra	Penta	Hexa		
5	-	1,368	227	6	5	1,606	20.08
6	1,999	507	22	1	1	2,530	31.63
7	1,090	314	-	-	-	1,404	17.55
8	867	15	-	-	1	883	11.04
9	891	1	-	-	-	892	11.15
10	535	2	-	-	1	538	6.73
11	137	-	-	-	-	137	1.71
≥12	3	4	1	-	-	8	0.10
Total	5,522	2,211	250	7	8	7,998	100
Percent %	69.04	27.64	3.13	0.09	0.10	100	

doi:10.1371/journal.pone.0152572.t004

Here, we detected potential SSR markers using the MISA package (<http://pgrc.ipk-gatersleben.de/misa/>). As it was difficult to distinguish true mononucleotide repeats from polyadenylation sites and false-positive mononucleotide repeats generated by sequencing errors, we did not include mononucleotide repeats in the following analysis.

A total of 7,545 SSRs of 2–6 bp unit length were identified (Table 4); this number of SSRs corresponds to a frequency of about one SSR per 9.6 kb of expressed sequences. We identified 5,168 (68.5%) di-nucleotide repeats, 2,131 (28.2%) tri-nucleotide repeats, and 246 (3.2%) tetra-/penta-/hexa-nucleotide repeats. The AC/GT sequence was the most common among the di-nucleotide repeats motifs, followed by AG/CT and AT/AT (Fig 6). Ten types of tri-nucleotide repeats were found; ATC/ATG was the most abundant, followed by AGG/CCT and AAT/ATT (Fig 6 and S2 Table). The complete set of SSR units and repeat number distributions is listed in S2 Table. The pattern of SSRs found here is consistent with that reported in previous studies in blunt snout bream (*Megalobrama amblycephala*) [64], but differs from those in yellow catfish (*Pelteobagrus fulvidraco*) [65] and half-smooth tongue sole (*Cynoglossus semilaevis*) [66], indicating that SSR repeat unit distributions are likely to be species-specific in teleosts. As is shown in Table 4, the majority of SSRs (69.27%) have a repeat number lower than 8.

Small variants, including single nucleotide polymorphisms (SNPs) and insertion and deletions (InDels), are very useful markers for mapping important traits and for whole-genome association studies because of their wide distributions and abundant polymorphisms [15]. In the present study, we identified 857,535 SNPs (533,503 transitions and 324,032 transversions) and 53,481 InDels by mapping sequencing reads to 37,785 assembled unigenes. As is shown in Fig 7, the numbers of the two transition types A/G and C/T were similar, and the numbers of the transversion types A/T, A/C, and G/T were likewise similar; however, the transversion C/G and InDels represented the smallest types. This variation might be due to differences in base structure and hydrogen bond interactions between the base pairs [67]. The transition/transversion (Ts/Tv) ratio was about 1.65, which is comparable to the ratios reported in other fish species [11,68–70]. These SNP and InDel loci provide an abundant marker resource for investigating population genetic structures, wild population conservation, mapping important economic traits, and for performing association studies in *S. prenanti*.

Conclusion

In this work, we sequenced mRNA fragments in spleen tissues and assembled the transcriptome into 48,517 transcripts and 37,785 unigenes. To our knowledge, this is the first

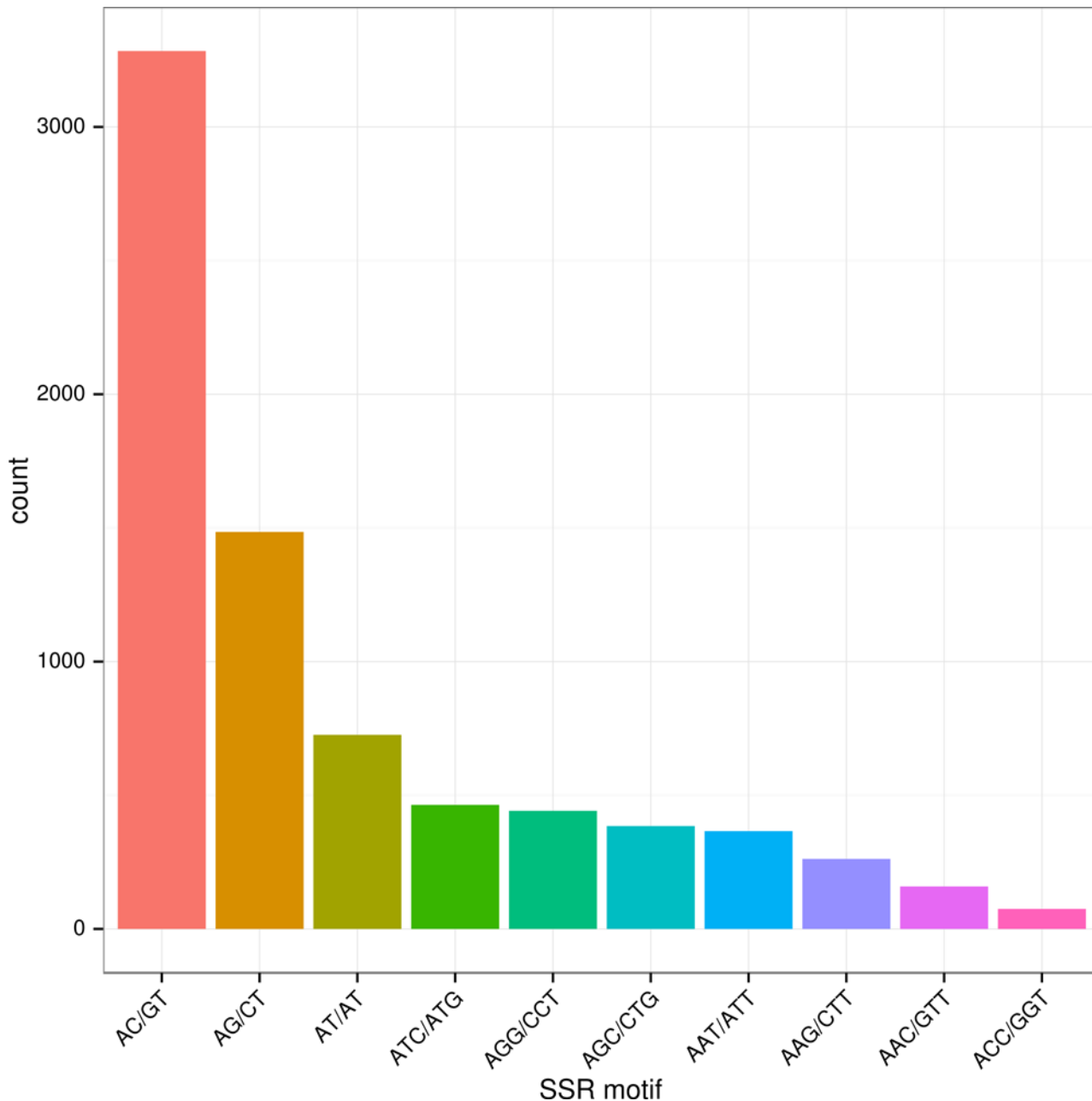


Fig 6. Frequency distribution of the top ten most abundant SSRs based on motif sequence types. Each histogram represented one detected SSR type in transcriptome of *S. prenanti*. Sequence complementary was considered during SSR type classification.

doi:10.1371/journal.pone.0152572.g006

transcriptome sequencing and *de novo* analysis of *S. prenanti* using the Illumina sequencing platform. By searching against known nucleotide and protein databases, 35,653 unigenes were successfully annotated. The 2,132 unigenes that failed to generate homologous hits may be non-coding RNAs, new genes, or species-specific sequences. Among the identified genes, more than 500 putative immune-related genes were identified in 16 signaling pathways. Most of the immune-related genes were reported for the first time and could provide important resources to understand the immune systems in *S. prenanti*. Additionally, 7,545 SSRs, 857,535 SNPs, and 53,481 InDels were identified from the transcriptome data. The transcriptome and molecular markers not only offer

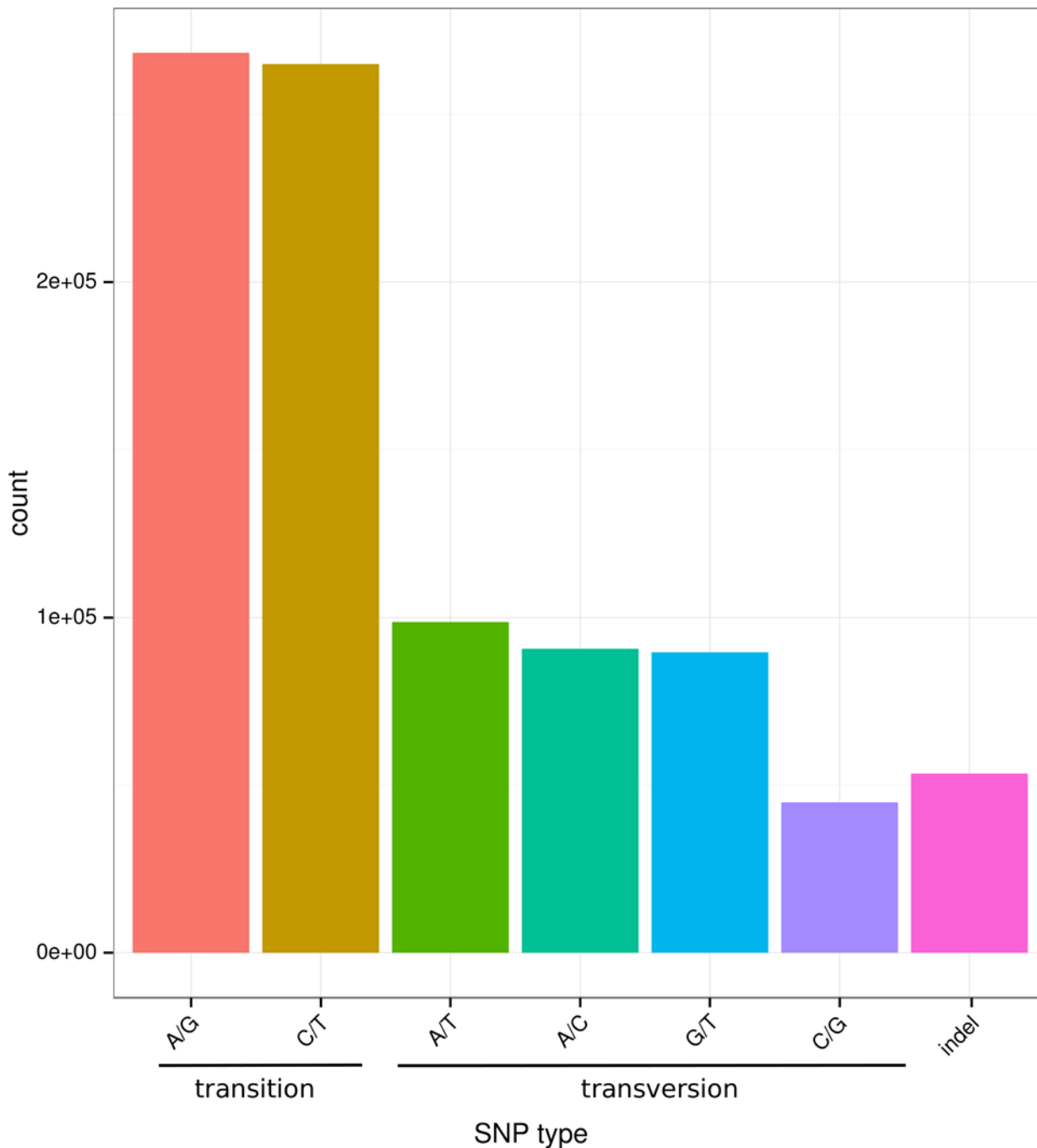


Fig 7. Frequency distribution of SNPs/InDels based on mutation types. The detected SNPs in transcriptome of *S. prenanti* were classified into transition and transversion, and the transition/transversion (Ts/Tv) ratio was estimated as ~1.65.

doi:10.1371/journal.pone.0152572.g007

precise sequence information for the functional gene analysis, but also provide valuable marker resources for conservation and molecular assisted selection of *S. prenanti*.

Supporting Information

S1 Fig. Base quality distribution for the raw RNA-sequencing data. The six separate plots represent paired-end sequencing runs for the three RNA libraries. (DOCX)

S2 Fig. Gene annotation in Toll-like receptor signaling pathway using the KEGG database. Identified genes are highlighted by the green background.

(DOCX)

S3 Fig. Gene annotation in complement and coagulation cascades using the KEGG database. Identified genes are highlighted by the green background.

(DOCX)

S4 Fig. Gene annotation in the Chemokine signaling pathway using the KEGG database. Identified genes are highlighted by the green background.

(DOCX)

S1 Table. Detailed annotation information for immune-related genes. (Included in a separate Excel file)

(XLS)

S2 Table. SSR unit size distribution with different repeat numbers in the *S. prenanti* transcriptome. (Included in a separate Excel file)

(XLS)

S1 Text. The list of 37,785 unigenes that were identified in the transcriptome of the *S. prenanti*.

(FASTA)

S2 Text. The gene lists of each category that have matches in NCBI Nr/Nt and Swiss-prot.

(XLSX)

Acknowledgments

We thank Meishan Aquaculture Company for providing *S. prenanti*. We also thank Lingbing Zeng, in Yangtze River Fisheries Research Institute Chinese Academy of Fishery Sciences, and Bin Chen, in College of Animal Science & Technology, Hunan Agricultural University, for providing constructive suggestions on experiments.

Author Contributions

Conceived and designed the experiments: HL HY SZ XW ZW. Performed the experiments: HY SZ CL ZZ. Analyzed the data: SX HL. Contributed reagents/materials/analysis tools: HY SZ CL SX HL ZZ. Wrote the paper: HL SX. Designed the software used in analysis: SX.

References

1. Wu Y, Wu C. The Fishes of the Qinghai-Xizang Plateau. Chengdu China: Sichuan Publishing House of Science and Technology; 1992.
2. He D, Chen Y, Chen Y, Chen Z. Molecular phylogeny of the specialized schizothoracine fishes (Teleostei: Cyprinidae), with their implications for the uplift of the Qinghai-Tibetan Plateau. Chinese Sci Bull. 2004; 49: 39–48. doi: [10.1007/BF02901741](https://doi.org/10.1007/BF02901741)
3. Yang L, Wang Y, Zhang Z, He S. Comprehensive Transcriptome Analysis Reveals Accelerated Genic Evolution in a Tibet Fish, *Gymnodiptychus pachycheilus*. Genome Biol Evol. 2015; 7: 251–261. doi: [10.1093/gbe/evu279](https://doi.org/10.1093/gbe/evu279)
4. Ding R. The Fishes of Sichuan. Chengdu China: Sichuan Publishing House of Science and Technology. 1994.
5. Wu J, Hou F, Wang Y, Wu B, Wei Q, Song Z. Isolation and characterization of twenty tetranucleotide microsatellite DNA makers in a schizothoracin fish, *Schizothorax prenanti* (Tchang). Conserv Genet Resour. 2013; 5: 891–894. doi: [10.1007/s12686-013-9926-4](https://doi.org/10.1007/s12686-013-9926-4)

6. Du Z, Huang X, Wang K, Deng Y, Li H. Isolation and Identification of Etiology of Skin Ulcer in *Schizothorax prenanti*. J Sichuan Agric Univ. 2011; 29: 274–279.
7. Geng Y, Wang KY, Huang XL, Chen DF, Li CW, Ren SY, et al. *Streptococcus agalactiae*, an Emerging Pathogen for Cultured Ya-Fish, *Schizothorax prenanti*, in China. Transbound Emerg Dis. 2012; 59: 369–375. doi: [10.1111/j.1865-1682.2011.01280.x](https://doi.org/10.1111/j.1865-1682.2011.01280.x) PMID: [22146014](https://pubmed.ncbi.nlm.nih.gov/22146014/)
8. Song J, Song Z, Yue B, Zheng W. Assessing genetic diversity of wild populations of Prenant's schizothoracin, *Schizothorax prenanti*, using AFLP markers. Environ Biol Fishes. 2006; 77: 79–86. doi: [10.1007/s10641-006-9056-x](https://doi.org/10.1007/s10641-006-9056-x)
9. Liang J, Liu Y, Zhang X, Zhang X, Yue B, Song Z. An observation of the loss of genetic variability in prenat's schizothoracin, *Schizothorax prenanti*, inhabiting a plateau lake. Biochem Syst Ecol. 2011; 39: 361–370. doi: [10.1016/j.bse.2011.05.005](https://doi.org/10.1016/j.bse.2011.05.005)
10. Zhang X, Gao X, Wang J, Cao W. Extinction risk and conservation priority analyses for 64 endemic fishes in the Upper Yangtze River, China. Environ Biol Fishes. 2015; 98: 261–272. doi: [10.1007/s10641-014-0257-4](https://doi.org/10.1007/s10641-014-0257-4)
11. Xiao S, Han Z, Wang P, Han F, Liu Y, Li J, et al. Functional Marker Detection and Analysis on a Comprehensive Transcriptome of Large Yellow Croaker by Next Generation Sequencing. PLoS One. 2015; 10: e0124432. doi: [10.1371/journal.pone.0124432](https://doi.org/10.1371/journal.pone.0124432) PMID: [25909910](https://pubmed.ncbi.nlm.nih.gov/25909910/)
12. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: A Method for Comprehensive Transcriptome Analysis. Current Protocols in Molecular Biology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010. pp. 1–13.
13. Malone JH, Oliver B. Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol. 2011; 9: 34. doi: [10.1186/1741-7007-9-34](https://doi.org/10.1186/1741-7007-9-34) PMID: [21627854](https://pubmed.ncbi.nlm.nih.gov/21627854/)
14. Xia JH, Wan ZY, Ng ZL, Wang L, Fu GH, Lin G, et al. Genome-wide discovery and in silico mapping of gene-associated SNPs in Nile tilapia. Aquaculture. 2014; 432: 67–73. doi: [10.1016/j.aquaculture.2014.04.028](https://doi.org/10.1016/j.aquaculture.2014.04.028)
15. Cui J, Wang H, Liu S, Zhu L, Qiu X, Jiang Z, et al. SNP Discovery from Transcriptome of the Swimbladder of *Takifugu rubripes*. PLoS One. 2014; 9: e92502. doi: [10.1371/journal.pone.0092502](https://doi.org/10.1371/journal.pone.0092502) PMID: [24651578](https://pubmed.ncbi.nlm.nih.gov/24651578/)
16. Luo H, Ye H, Xiao S, Zheng S, Wang X, Wang Z. Application of transcriptomics technology to aquatic animal research. J Fish China. 2015; 39: 598–607. doi: [10.11964/jfc.20141109541](https://doi.org/10.11964/jfc.20141109541)
17. Wang R, Sun L, Bao L, Zhang J, Jiang Y, Yao J, et al. Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. BMC Genomics. 2013; 14: 929. doi: [10.1186/1471-2164-14-929](https://doi.org/10.1186/1471-2164-14-929) PMID: [24373586](https://pubmed.ncbi.nlm.nih.gov/24373586/)
18. Zhang L, Li L, Zhu Y, Zhang G, Guo X. Transcriptome Analysis Reveals a Rich Gene Set Related to Innate Immunity in the Eastern Oyster (*Crassostrea virginica*). Mar Biotechnol. 2014; 16: 17–33. doi: [10.1007/s10126-013-9526-z](https://doi.org/10.1007/s10126-013-9526-z) PMID: [23907648](https://pubmed.ncbi.nlm.nih.gov/23907648/)
19. Rebl A, Korytář T, Köbis JM, Verleih M, Krasnov A, Jaros J, et al. Transcriptome Profiling Reveals Insight into Distinct Immune Responses to *Aeromonas salmonicida* in Gill of Two Rainbow Trout Strains. Mar Biotechnol. 2014; 16: 333–348. doi: [10.1007/s10126-013-9552-x](https://doi.org/10.1007/s10126-013-9552-x) PMID: [24122123](https://pubmed.ncbi.nlm.nih.gov/24122123/)
20. Lanes CFC, Bizuayehu TT, de Oliveira Fernandes JM, Kiron V, Babiak I. Transcriptome of Atlantic Cod (*Gadus morhua* L.) Early Embryos from Farmed and Wild Broodstocks. Mar Biotechnol. 2013; 15: 677–694. doi: [10.1007/s10126-013-9527-y](https://doi.org/10.1007/s10126-013-9527-y) PMID: [23887676](https://pubmed.ncbi.nlm.nih.gov/23887676/)
21. Chapman RW, Reading BJ, Sullivan C V. Ovary Transcriptome Profiling via Artificial Intelligence Reveals a Transcriptomic Fingerprint Predicting Egg Quality in Striped Bass, *Morone saxatilis*. PLoS One. 2014; 9: e96818. doi: [10.1371/journal.pone.0096818](https://doi.org/10.1371/journal.pone.0096818) PMID: [24820964](https://pubmed.ncbi.nlm.nih.gov/24820964/)
22. Sellars MJ, Trewin C, McWilliam SM, Graves RSE, Hertzler PL. Transcriptome Profiles of *Penaeus (Marsupenaeus japonicus)* Animal and Vegetal Half-Embryos: Identification of Sex Determination, Germ Line, Mesoderm, and Other Developmental Genes. Mar Biotechnol. 2015; 17: 252–265. doi: [10.1007/s10126-015-9613-4](https://doi.org/10.1007/s10126-015-9613-4) PMID: [25634056](https://pubmed.ncbi.nlm.nih.gov/25634056/)
23. Ulloa PE, Rincón G, Islas-Trejo A, Araneda C, Iturra P, Neira R, et al. RNA Sequencing to Study Gene Expression and SNP Variations Associated with Growth in Zebrafish Fed a Plant Protein-Based Diet. Mar Biotechnol. 2015; 17: 353–363. doi: [10.1007/s10126-015-9624-1](https://doi.org/10.1007/s10126-015-9624-1) PMID: [25702041](https://pubmed.ncbi.nlm.nih.gov/25702041/)
24. Rebl A, Verleih M, Köbis JM, Kühn C, Wimmers K, Köllner B, et al. Transcriptome Profiling of Gill Tissue in Regionally Bred and Globally Farmed Rainbow Trout Strains Reveals Different Strategies for Coping with Thermal Stress. Mar Biotechnol. 2013; 15: 445–460. doi: [10.1007/s10126-013-9501-8](https://doi.org/10.1007/s10126-013-9501-8) PMID: [23547003](https://pubmed.ncbi.nlm.nih.gov/23547003/)
25. Zhang R, Ludwig A, Zhang C, Tong C, Li G, Tang Y, et al. Local adaptation of *Gymnocypris przewalskii* (Cyprinidae) on the Tibetan Plateau. Sci Rep. 2015; 5: 9780. doi: [10.1038/srep09780](https://doi.org/10.1038/srep09780) PMID: [25944748](https://pubmed.ncbi.nlm.nih.gov/25944748/)

26. Huang Q, Dong S, Fang C, Wu X, Ye T, Lin Y. Deep sequencing-based transcriptome profiling analysis of *Oryzias melastigma* exposed to PFOS. *Aquat Toxicol.* 2012; 120–121: 54–58. doi: [10.1016/j.aquatox.2012.04.013](https://doi.org/10.1016/j.aquatox.2012.04.013) PMID: [22613580](https://pubmed.ncbi.nlm.nih.gov/22613580/)
27. Guo H, Ye C-X, Wang A-L, Xian J-A, Liao S-A, Miao Y-T, et al. Transcriptome analysis of the Pacific white shrimp *Litopenaeus vannamei* exposed to nitrite by RNA-seq. *Fish Shellfish Immunol.* 2013; 35: 2008–2016. doi: [10.1016/j.fsi.2013.09.019](https://doi.org/10.1016/j.fsi.2013.09.019) PMID: [24055647](https://pubmed.ncbi.nlm.nih.gov/24055647/)
28. Dai Y, Xiao H. Review of Studies on the Germplasm Resources of the Schizothoracinae Fishes. *Chinese Agric Sci Bull.* 2011; 27: 38–46.
29. Li C, Chen X, Zhang Y, Ye H, Liu T. Molecular and expression characterization of growth hormone/prolactin family genes in the Prenant's schizothoracin. *Mol Biol Rep.* 2011; 38: 4595–4602. doi: [10.1007/s11033-010-0592-2](https://doi.org/10.1007/s11033-010-0592-2) PMID: [21468655](https://pubmed.ncbi.nlm.nih.gov/21468655/)
30. Wu J, Hou F, Liang J, Zhang X, Song Z, Wei Q. Development of ten microsatellite DNA markers in a hexaploid fish, *Schizothorax prenanti* (Tchang). *Conserv Genet Resour.* 2013; 5: 545–547. doi: [10.1007/s12686-012-9848-6](https://doi.org/10.1007/s12686-012-9848-6)
31. Wang T, Zhou C, Yuan D, Lin F, Chen H, Wu H, et al. *Schizothorax prenanti* corticotropin-releasing hormone (CRH): molecular cloning, tissue expression, and the function of feeding regulation. *Fish Physiol Biochem.* 2014; 40: 1407–1415. doi: [10.1007/s10695-014-9935-6](https://doi.org/10.1007/s10695-014-9935-6) PMID: [24696302](https://pubmed.ncbi.nlm.nih.gov/24696302/)
32. Yuan D, Zhou C, Wang T, Lin F, Chen H, Wu H, et al. Molecular characterization and tissue expression of peptide YY in *Schizothorax prenanti*: Effects of periprandial changes and fasting on expression in the hypothalamus. *Regul Pept.* 2014; 190–191: 32–38. doi: [10.1016/j.regpep.2014.03.004](https://doi.org/10.1016/j.regpep.2014.03.004) PMID: [24681121](https://pubmed.ncbi.nlm.nih.gov/24681121/)
33. Ye H, Zhu C, Zheng Z, Wu Q, Zheng S. The cDNA sequence and characterization of myeloid differentiation factor 88 (MyD88) gene in *Schizothorax prenanti*. *Freshw Fish.* 2014; 44: 14–19.
34. Yang X, Liu D, Liu F, Wu J, Zou J, Xiao X, et al. HTQC: a fast quality control toolkit for Illumina sequencing data. *BMC Bioinformatics.* 2013; 14: 33. doi: [10.1186/1471-2105-14-33](https://doi.org/10.1186/1471-2105-14-33) PMID: [23363224](https://pubmed.ncbi.nlm.nih.gov/23363224/)
35. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson D a, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011; 29: 644–652. doi: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
36. Nakasugi K, Crowhurst R, Bally J, Waterhouse P. Combining transcriptome assemblies from multiple de novo assemblers in the allo-tetraploid plant *nicotiana benthamiana*. *PLoS One.* 2014; 9: e91776. doi: [10.1371/journal.pone.0091776](https://doi.org/10.1371/journal.pone.0091776) PMID: [24614631](https://pubmed.ncbi.nlm.nih.gov/24614631/)
37. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids Res.* 1997; 25: 3389–3402. Available: <http://nar.oxfordjournals.org/content/25/17/3389.short> PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
38. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005; 21: 3674–3676. doi: [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610) PMID: [16081474](https://pubmed.ncbi.nlm.nih.gov/16081474/)
39. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25: 1754–1760. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25: 2078–2079. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
41. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20: 1297–1303. doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110) PMID: [20644199](https://pubmed.ncbi.nlm.nih.gov/20644199/)
42. Tran NT, Gao Z-X, Zhao H-H, Yi S-K, Chen B-X, Zhao Y-H, et al. Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. *Fish Shellfish Immunol.* 2015; 45: 72–82. doi: [10.1016/j.fsi.2015.01.034](https://doi.org/10.1016/j.fsi.2015.01.034) PMID: [25681750](https://pubmed.ncbi.nlm.nih.gov/25681750/)
43. Li C, Ling Q, Ge C, Ye Z, Han X. Transcriptome characterization and SSR discovery in large-scale loach *Paramisgurnus dabryanus* (Cobitidae, Cypriniformes). *Gene.* 2015; 557: 201–208. doi: [10.1016/j.gene.2014.12.034](https://doi.org/10.1016/j.gene.2014.12.034) PMID: [25528212](https://pubmed.ncbi.nlm.nih.gov/25528212/)
44. Tong C, Zhang C, Zhang R, Zhao K. Transcriptome profiling analysis of naked carp (*Gymnocypris przewalskii*) provides insights into the immune-related genes in highland fish. *Fish Shellfish Immunol.* 2015; 46: 366–377. doi: [10.1016/j.fsi.2015.06.025](https://doi.org/10.1016/j.fsi.2015.06.025) PMID: [26117731](https://pubmed.ncbi.nlm.nih.gov/26117731/)
45. Che R, Sun Y, Sun D, Xu T. Characterization of the miiuy croaker (*Miichthys miiuy*) transcriptome and development of immune-relevant genes and molecular markers. *PLoS One.* 2014; 9: e94046. doi: [10.1371/journal.pone.0094046](https://doi.org/10.1371/journal.pone.0094046) PMID: [24714210](https://pubmed.ncbi.nlm.nih.gov/24714210/)

46. Ali A, Rexroad CE, Thorgaard GH, Yao J, Salem M. Characterization of the rainbow trout spleen transcriptome and identification of immune-related genes. *Front Genet.* 2014; 5: 1–17. doi: [10.3389/fgene.2014.00348](https://doi.org/10.3389/fgene.2014.00348) PMID: [25352861](https://pubmed.ncbi.nlm.nih.gov/25352861/)
47. Akira S, Uematsu S, Takeuchi O. Pathogen Recognition and Innate Immunity. *Cell.* 2006; 124: 783–801. doi: [10.1016/j.cell.2006.02.015](https://doi.org/10.1016/j.cell.2006.02.015) PMID: [16497588](https://pubmed.ncbi.nlm.nih.gov/16497588/)
48. O'Neill L a J, Golenbock D, Bowie AG. The history of Toll-like receptors—redefining innate immunity. *Nat Rev Immunol.* 2013; 13: 453–460. doi: [10.1038/nri3446](https://doi.org/10.1038/nri3446) PMID: [23681101](https://pubmed.ncbi.nlm.nih.gov/23681101/)
49. Rebl A, Goldammer T, Seyfert HM. Toll-like receptor signaling in bony fish. *Vet Immunol Immunopathol.* 2010; 134: 139–150. doi: [10.1016/j.vetimm.2009.09.021](https://doi.org/10.1016/j.vetimm.2009.09.021) PMID: [19850357](https://pubmed.ncbi.nlm.nih.gov/19850357/)
50. Kumar H, Kawai T, Akira S. Pathogen Recognition by the Innate Immune System. *Int Rev Immunol.* 2011; 30: 16–34. doi: [10.3109/08830185.2010.529976](https://doi.org/10.3109/08830185.2010.529976) PMID: [21235323](https://pubmed.ncbi.nlm.nih.gov/21235323/)
51. Gerlier D, Lyles DS. Interplay between Innate Immunity and Negative-Strand RNA Viruses: towards a Rational Model. *Microbiol Mol Biol Rev.* 2011; 75: 468–490. doi: [10.1128/MMBR.00007-11](https://doi.org/10.1128/MMBR.00007-11) PMID: [21885681](https://pubmed.ncbi.nlm.nih.gov/21885681/)
52. Takano, T., Kondo, H., Hirono, I., Endo, M., Saito-Taki, T. and Aoki T. Toll-like receptors in teleosts [Internet]. Diseases in Asian Aquaculture VII. Fish Health Section, Asian Fisheries Society, Selangor, Malaysia. 2011. Available: http://www.fhs-afs.net/daa_vii_files/17.pdf
53. Pope MR, Hoffman SM, Tomlinson S, Fleming SD. Complement regulates TLR4-mediated inflammatory responses during intestinal ischemia reperfusion. *Mol Immunol.* 2010; 48: 356–364. doi: [10.1016/j.molimm.2010.07.004](https://doi.org/10.1016/j.molimm.2010.07.004) PMID: [20800895](https://pubmed.ncbi.nlm.nih.gov/20800895/)
54. Fang C, Zhang X, Miwa T, Song W. Complement promotes the development of inflammatory T-helper 17 cells through synergistic interaction with Toll-like receptor signaling and interleukin-6 production. *Blood.* 2009; 114: 1005–1015. doi: [10.1182/blood-2009-01-198283](https://doi.org/10.1182/blood-2009-01-198283) PMID: [19491392](https://pubmed.ncbi.nlm.nih.gov/19491392/)
55. Nakao M, Tsujikura M, Ichiki S, Vo TK, Somamoto T. The complement system in teleost fish: Progress of post-homolog-hunting researches. *Dev Comp Immunol.* 2011; 35: 1296–1308. doi: [10.1016/j.dci.2011.03.003](https://doi.org/10.1016/j.dci.2011.03.003) PMID: [21414344](https://pubmed.ncbi.nlm.nih.gov/21414344/)
56. Roozendaal R, Carroll MC. Emerging Patterns in Complement-Mediated Pathogen Recognition. *Cell.* 2006; 125: 29–32. doi: [10.1016/j.cell.2006.03.018](https://doi.org/10.1016/j.cell.2006.03.018) PMID: [16615887](https://pubmed.ncbi.nlm.nih.gov/16615887/)
57. Zhu L, Nie L, Zhu G, Xiang L, Shao J. Advances in research of fish immune-relevant genes: A comparative overview of innate and adaptive immunity in teleosts. *Dev Comp Immunol.* 2013; 39: 39–62. doi: [10.1016/j.dci.2012.04.001](https://doi.org/10.1016/j.dci.2012.04.001) PMID: [22504163](https://pubmed.ncbi.nlm.nih.gov/22504163/)
58. Dodds MW, Alex Law SK. The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and alpha2-macroglobulin. *Immunol Rev.* 1998; 166: 15–26. doi: [10.1111/j.1600-065X.1998.tb01249.x](https://doi.org/10.1111/j.1600-065X.1998.tb01249.x) PMID: [9914899](https://pubmed.ncbi.nlm.nih.gov/9914899/)
59. Boshra H, Li J, Sunyer JO. Recent advances on the complement system of teleost fish. *Fish Shellfish Immunol.* 2006; 20: 239–262. doi: [10.1016/j.fsi.2005.04.004](https://doi.org/10.1016/j.fsi.2005.04.004) PMID: [15950490](https://pubmed.ncbi.nlm.nih.gov/15950490/)
60. Muller-Eberhard H. The killer molecule of complement. *J Invest Dermatol.* 1985; 85: 47–52.
61. Bao B, Peatman E, Peng X, Baoprasertkul P, Wang G, Liu Z. Characterization of 23 CC chemokine genes and analysis of their expression in channel catfish (*Ictalurus punctatus*). *Dev Comp Immunol.* 2006; 30: 783–796. doi: [10.1016/j.dci.2005.10.007](https://doi.org/10.1016/j.dci.2005.10.007) PMID: [16510183](https://pubmed.ncbi.nlm.nih.gov/16510183/)
62. Liu Y, Chen S- L, Meng L, Zhang Y- X. Cloning, characterization and expression analysis of a novel CXC chemokine from turbot (*Scophthalmus maximus*). *Fish Shellfish Immunol.* 2007; 23: 711–720. doi: [10.1016/j.fsi.2007.01.008](https://doi.org/10.1016/j.fsi.2007.01.008) PMID: [17604647](https://pubmed.ncbi.nlm.nih.gov/17604647/)
63. Wan X, Chen X. Molecular cloning and expression analysis of a CXC chemokine gene from large yellow croaker *Pseudosciaena crocea*. *Vet Immunol Immunopathol.* 2009; 127: 156–161. doi: [10.1016/j.vetimm.2008.09.009](https://doi.org/10.1016/j.vetimm.2008.09.009) PMID: [18963007](https://pubmed.ncbi.nlm.nih.gov/18963007/)
64. Gao Z, Luo W, Liu H, Zeng C, Liu X, Yi S, et al. Transcriptome Analysis and SSR/SNP Markers Information of the Blunt Snout Bream (*Megalobrama amblycephala*). *Fuentes J, editor. PLoS One.* 2012; 7: e42637. doi: [10.1371/journal.pone.0042637](https://doi.org/10.1371/journal.pone.0042637) PMID: [22880060](https://pubmed.ncbi.nlm.nih.gov/22880060/)
65. Chen X, Mei J, Wu J, Jing J, Ma W, Zhang J, et al. A Comprehensive Transcriptome Provides Candidate Genes for Sex Determination/Differentiation and SSR/SNP Markers in Yellow Catfish. *Mar Biotechnol.* 2015; 17: 190–198. doi: [10.1007/s10126-014-9607-7](https://doi.org/10.1007/s10126-014-9607-7) PMID: [25403497](https://pubmed.ncbi.nlm.nih.gov/25403497/)
66. Zhang X, Wang S, Chen S, Chen Y, Liu Y, Shao C, et al. Transcriptome analysis revealed changes of multiple genes involved in immunity in *Cynoglossus semilaevis* during *Vibrio anguillarum* infection. *Fish Shellfish Immunol.* 2015; 43: 209–218. doi: [10.1016/j.fsi.2014.11.018](https://doi.org/10.1016/j.fsi.2014.11.018) PMID: [25543033](https://pubmed.ncbi.nlm.nih.gov/25543033/)
67. Ma K, Qiu G, Feng J, Li J. Transcriptome Analysis of the Oriental River Prawn, *Macrobrachium nipponense* Using 454 Pyrosequencing for Discovery of Genes and Markers. *Liu Z, editor. PLoS One.* 2012; 7: e39727. doi: [10.1371/journal.pone.0039727](https://doi.org/10.1371/journal.pone.0039727) PMID: [22745820](https://pubmed.ncbi.nlm.nih.gov/22745820/)

68. Hale MC, McCormick CR, Jackson JR, DeWoody JA. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. *BMC Genomics*. 2009; 10: 203. doi: [10.1186/1471-2164-10-203](https://doi.org/10.1186/1471-2164-10-203) PMID: [19402907](https://pubmed.ncbi.nlm.nih.gov/19402907/)
69. Renaut S, Nolte AW, Bernatchez L. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus spp. Salmonidae*). *Mol Ecol*. 2010; 19: 115–131. doi: [10.1111/j.1365-294X.2009.04477.x](https://doi.org/10.1111/j.1365-294X.2009.04477.x) PMID: [20331775](https://pubmed.ncbi.nlm.nih.gov/20331775/)
70. Vera M, Alvarez-Dios JA, Fernandez C, Bouza C, Vilas R, Martinez P. Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *Int J Mol Sci*. 2013; 14: 5694–5711. doi: [10.3390/ijms14035694](https://doi.org/10.3390/ijms14035694) PMID: [23481633](https://pubmed.ncbi.nlm.nih.gov/23481633/)