Fish and Shellfish Immunology 84 (2019) 825–833

Contents lists available at ScienceDirect

Fish and Shellfish Immunology

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

Full length article

Liver transcriptome analysis of the *Sparus macrocephlus* in response to *Vibrio parahaemolyticus* infection

Xing-Wei Xiang[a,b], Jin-Xing Xiao[b], Yu-Fang Zhou[b], Bin Zheng[a,b,*], Zheng-Shun Wen[a,**]

[a] College of Food Science and Pharmacy, Zhejiang Ocean University, Haida Road 1, New Town, Zhoushan, Zhejiang Province, 316000, China
[b] Zhejiang Marine Development Research Institute, Tiyu Road 10, New Town, Zhoushan, Zhejiang Province, 316000, China

ARTICLE INFO

Keywords:
Sparus macrocephlus
Vibrio parahaemolyticus
Infection
Transcriptome
Immune mechanism

ABSTRACT

The black seabream (*Sparus macrocephlus*) is an economically pivotal aquaculture species cultured in China and Southeast Asian countries. To understand the molecular immune mechanisms underlying the response to *Vibrio parahaemolyticus*, a comparative gene transcription analysis were performed with utilized fresh livers of *V. parahaemolyticus*-immunized *Sparus macrocephlus* with a control group through RNA-Seq technology. A total of 256663 contigs were obtained after excluded the low-quality sequences and assembly. The average length of contigs collected from this research is 1066.93 bp. Furthermore, blast analysis indicates 30747 contigs were annotated based on homology with matches in the NT, NR, gene, and string databases. A gene ontology analysis was employed to classify 21598 genes according to three major functional categories: molecular function, cellular component, and biological process. A total of 14470 genes were discovered in 303 KEGG pathways. RSEM and EdgeR were introduced to estimate 3841 genes significantly different expressed (False Discovery Rate < 0.001) which includes 4072 up-regulated genes and 3771 down-regulated genes. A significant enrichment analysis of these differentially expressed genes and isogenes were conducted to reveal the major immune-related pathways which refer to the toll-like receptor, complement, coagulation cascades, and chemokine signaling pathways. In addition, 92175 potential simple sequence repeats (SSRs) and 121912 candidate single nucleotide polymorphisms (SNPs) were detected and identified sequencely in the *Sparus macrocephlus* liver transcriptome. This research characterized a gene expression pattern for normal and the *V. parahaemolyticus* -immunized *Sparus macrocephlus* for the first time and not only sheds new light on the molecular mechanisms underlying the host-*V. parahaemolyticus* interaction but contribute to facilitate future studies on *Sparus macrocephlus* gene expression and functional genomics.

1. Introduction

The black seabream (*Sparus macrocephlus*) cultured in the coastal areas of China, Japan, Korean and some other countries of Southeast Asia, is a pivotal commercial fish species [1]. With rapid development of aquaculture, bacteria, parasites and viruses had caused severe outbreaks of infectious diseases increasingly [2–4]. The market had suffered significant economic loss caused by infectious diseases. However, information of the immune mechanisms of the *Sparus macrocephlus* response to *V. parahaemolyticus* is limited.

As an effective and efficient method for genome studies and functional gene identification, transcriptome prolifing analysis have experienced swift and violent developments especially in the high-throughput deep sequencing technologies aspect, which provide an overwhelming increase in transcriptome data [5,6]. RNA-Seq is a relatively new technology for transcriptomic researches across the entire genome of aquaculture species [7]. In recent years, several studies had reported the transcriptome profile of fish following exposure to pathogenic microorganisms, including the *Danio rerio* [8], *Cynoglossus semilaevis* [9], *Epinephelus coioides* [10], *Megalobrama amblycephala* [11], *Concholepas* [12], *Ctenopharyngodon idella* [13], *Takifugu rubripes* [14] and *Pseudosciaena crocea* [15]. Eventhough, there is less information available on the gene expression profile for the transcriptome of *Sparus macrocephlus* in response to *V. parahaemolyticus*, there is still many investigation of gene expression pattern after *V. parahaemolyticus* in other models, such as, *Danio rerio* [16], *Branchiostoma belcheri* [17], *Meretrix petechialis* [18], *Litopenaeus vannamei* [19], *Macrobrachium rosenbergii* [20], *Scylla paramamosain* [21], *Sinonovacula constricta* [22],

* Corresponding author. Zhejiang Ocean University (Zhejiang Marine Development Research Institute), Zhoushan, 316000, China.
** Corresponding author.
E-mail addresses: yfzhou@yeah.net (Y.-F. Zhou), xxw11086@126.com (B. Zheng), zswenmr@163.com (Z.-S. Wen).

https://doi.org/10.1016/j.fsi.2018.09.057
Received 17 July 2018; Received in revised form 16 September 2018; Accepted 20 September 2018
Available online 21 September 2018

Exopalaemon carinicauda [23]. Until now, molecular studies on the immune response to *V. parahaemolyticus* in *Sparus macrocephlus* remain rare.

To systematically and comprehensively comprehend the molecular immune mechanisms underlying the response to *V. parahaemolyticus*, Illumina/Solexa sequencing technology were employed for analyzing the transcriptome profiling of *Sparus macrocephlus* infected with *V. parahaemolyticus* firstly. Gene expression at the transcriptome level, immune-related genes and pathways were measured and identified. SSRs and SNPs were also detected respectively in this research. These data can contribute to further investigation into the molecular mechanism underlying the host-*V. parahaemolyticus* interaction.

2. Materials and methods

2.1. Sparus macrocephlus experiment

Sparus macrocephlus (weight: 25.5 ± 5.1 g, length: 10.2 ± 1.3 cm) were purchased from Zhoushan (Zhejiang, China) and acclimatized for two weeks prior to the experiment. The pathogenic microorganism was previously isolated from *Sparus macrocephlus* suffering from a deadly *V. parahaemolyticus* infection and identified by morphology and molecular biology methods. The numbers of larvae was calculated by an inverted microscope. 36 healthy individuals were immersed in the suspension containing the pathogens with the final concentration of 10^7 CFU/mL for 4 h, subsequently transferred into a new cement pool containing treated seawater, and daily observations were conducted at the same time. Three days after infection, the samples were transferred to another new cement pool to avoid reinfection with *V. parahaemolyticus*. The fresh liver tissue was collected separately from five unchallenged fishes and five fishes with 3-days infection. The fresh tissues were placed in RNAfixer (TaKaRa, Japan) immediately after collection and then stored at −20 °C for preservation before use.

2.2. Library construction and sequencing

Total RNA of five *Sparus macrocephlus* per group were extracted by RNAiso Plus (TaKaRa, Japan) in accordance with the manufacturer's instructions and then treated with RNase-free DNase I (TaKaRa, Japan) to remove the DNA contaminants. Subsequently, the quality and quantity of purified RNA were determined by a Nanodrop spectrophotometer (LabTech, USA), and only RNA samples with an A260/A280 from 1.8 to 2.0 and an A260/A230 from 2.0 to 2.5 were employed for subsequent analyses. The RNA pool was obtained from each group following the verification of the RNA integrity with utilizing an Agilent 2100 Bioanalyzer (RIN of all samples ≥ 8) (Agilent Technologies, USA). Subsequently, poly (A) mRNA was isolated from the total RNA with poly (dT) oligo-attached magnetic beads, and the cDNA libraries were prepared using the TruSeq RNA Sample Preparation Kit (Illumina) following the TruSeq protocol. The cDNA libraries were sequenced by the Illumina HiSeq2000 sequencing platform 110 with 2 × 101 bp paired-end (PE) reads.

2.3. Transcriptome data analysis

The raw reads from the Illumina Solexa sequencing experiment were first preprocessed by eliminating adaptor sequences and low quality reads. The clean reads were assembled and stitched using the Trinity program (http://trinityrnaseq.sourceforge.net/) into contigs [24]. Subsequently, Trinity (http://trinityrnaseq.sourceforge.net/analysis/extract_proteins_from_trinity_transcripts.html) was used to analyze the open reading frame (ORF) [24]. The assembled transcripts were divided into two categories (contigs with ORFs or without ORFs) based on these ORF predictions. The isogenes were annotated using local BLAST programs with the NCBI non-redundant (nr) protein database as well as nt, gene, and string databases. The Blast2GO program was used to obtain GO annotation based on BLASTx hits with the NCBI Nr database [25]. The number of isogenes associated with each GO term (FDR ≤ 0.05) was calculated by the categories molecular function, cellular component, and biological process [26]. The Blastx/blastp 2.2.24 + tools were used to analyze the KEGG pathways, and KO (KEGG Orthology) annotations were obtained based on the BLASTx hits with the KEGG database (http://www.genome.jp/kegg/genes.html) [27].

Table 1
Summary statistics for sequencing results from two transcriptomes.

Terms	LIVER	LIVER-VP
Number of raw sequences (n)	50841136	47686808
Raw bases (bp)	7626170400	7153021200
Number of clean sequences (n)	50628716	47285670
Base of clean sequencing reads (bp)	7329235140	6799353679
Good ratio	99.58	99.16

2.4. Differentially expressed genes and enrichment analysis

The clean sequencing reads from each of the two libraries (Liver and Liver-Vp) were mapped back to the transcriptome assembly using the software Bowtie2 with the default parameters [28]. Differentially expressed sequences between the two libraries were sifted using RSEM (http://deweylab.biostat.wisc.edu/rsem/) [29], and EdgeR (http://www.bioconductor.org/packages/elease/bioc/tml/edgeR.html) [30] was used to analyze differential expression based on a threshold false FDR < 0.001. DESeq is an R package to analyze count data from transcriptome sequencing data and test for differential expression (http://bioconductor.org/packages/release/bioc/html/DESeq.html), and the FDR < 0.001 and Fold Change > 2 as the threshold to judge the significance of gene expression differences. Heatmap and cluster analysis were performed using R package. Subsequently, a significant enrichment analysis of the differentially expressed genes and isogenes was performed to analyze the immune-related genes through hypergeometric distribution testing using the software Goatools (https://github.com/tanghaibao/goatools) and KOBAS (http://kobas.cbi.pku.edu.cn/home.do).

2.5. Experimental validation using RT-qPCR

To investigate the correction of data from the transcriptome, six differentially expressed genes (Toll-like receptor 1, Jun proto-oncogene, C—C motif chemokine 25, C-X-C motif chemokine 12, coagulation factor VIII, tyrosine-protein kinase ZAP-70) were selected randomly for validation using RT-qPCR amplification. Quantitative real-time PCR was performed using an ABI Quantstudio 6 Flex system with SYBR® Premix Ex TaqTM (TaKaRa, Japan) in accordance with the manufacturers' instructions. Primer sequences were carefully designed based on each identified gene sequence from the transcriptome library using the Primer Premier 6 software (Premier Biosoft, USA) (Table S3). PCR amplification experiments were performed in triplicate under the following conditions: 95 for 30 s, followed by 40 cycles of 95 for 5 s, 55 for 30 s, and 72 for 30 s. The results were normalized using GAPDH for each sample and the $2^{-\triangle\triangle CT}$ method. The RT-qPCR data were analyzed using one-way analysis of variance (ANOVA) and the PASW Statistics 18 software.

2.6. Detecting SSRs and SNPs

Using Msatcommander (http://code.google.com/p/msatcommander/) with the default settings, the assembled *Sparus macrocephlus* transcriptomes were screened for microsatellites [31]. The parameters were designed to identify perfect mono-hexa-nucleotide motifs with a minimum of ten repetitions for mononucleotides, six repetitions for

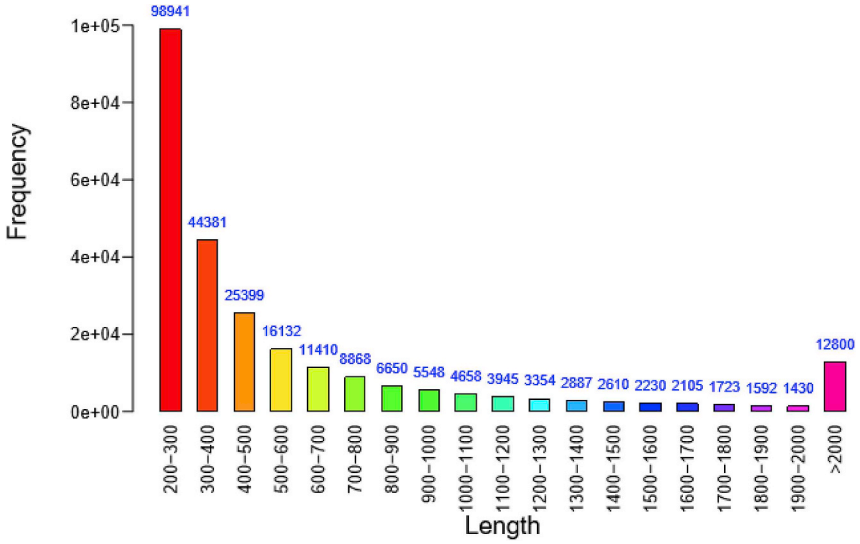

Fig. 1. Sequence length distributions for all contigs from the LIVER and LIVER-VP experiments.

Table 2
Summary statistics for assembly results from two transcriptomes.

Terms	Number
Total genes	214513
Total isogenes	256663
Total residues	163035909
Average length	635.21
Largest isogene	37437
Smallest isogene	201

dinucleotides, and five repetitions for tri-hexa nucleotides. Potential SNPs were filtered using the program Samtools (http://samtools.sourceforge.net/) and VarScan v.2.2.7 (http://varscan.sourceforge.net/) with the default parameters [32,33].

3. Results and discussion

3.1. RNA-seq and transcriptome sequence assembly

An immediate application for the transcriptome sequence data included expression profiling of experimental and control groups. Sequencing the normal liver library and infected library transcriptomes produced 50841136 and 47686808 raw reads, respectively, with the total (average) lengths 76.26 Mb and 71.53 Mb, respectively (Table 1). The raw reads were assembled into transcripts using the SOAP *de novo* assembly software package (http://soap.genomics.org.cn/soapdenovo.html) with the default settings, which reduced the total raw sequencing reads to 50628716 high-quality sequences for LIVER and 47285670 sequences for LIVER-Vp (Table 1). An analysis of the nucleotide content showed that the overall guanine-cytosine (GC) content of the transcriptome was 49%. These sequences include reads fully or partially assembled into contigs and reads that we thought were derived from repeat regions. The clean reads were assembled and stitched using

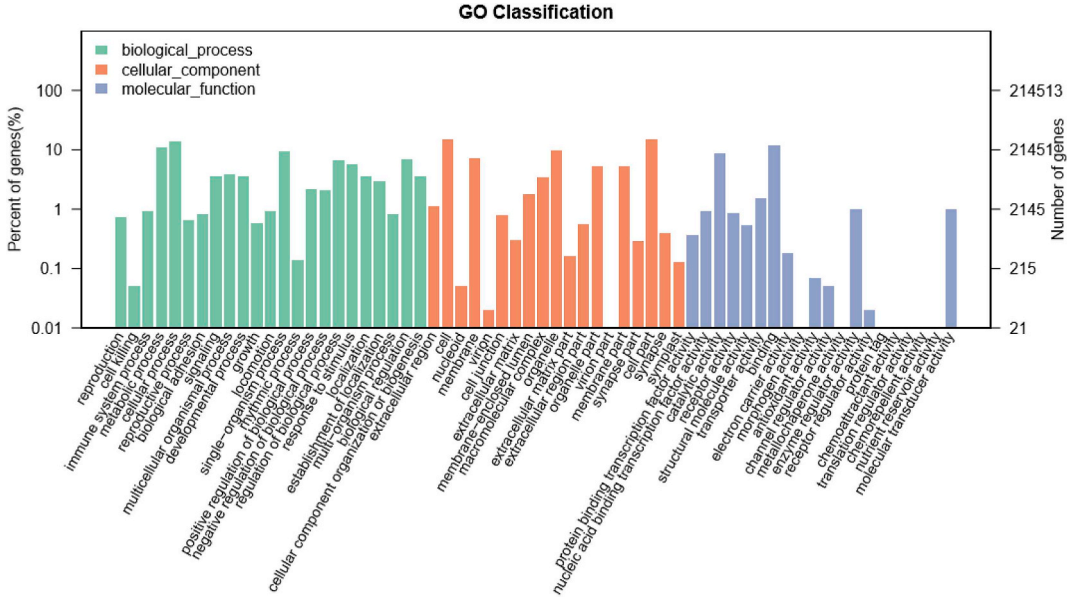

Fig. 2. Gene ontologies. Transcript counts for the transcriptome gene ontology classifications into biological process (green), cellular component (blue), molecular function (pink) categories. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
The top 20 statistically significant KEGG classifications.

Pathway ID	Pathway definition	Group	Number of seqs
ko03010	Ribosome	Genetic Information Processing	354
ko05200	Pathways in cancer	Human Diseases	335
ko04151	PI3K-Akt signaling pathway	Environmental Information Processing	332
ko01200	Carbon metabolism	Metabolism	313
ko00230	Purine metabolism	Metabolism	307
ko01230	Biosynthesis of amino acids	Metabolism	286
ko05166	HTLV-I infection	Human Diseases	283
ko04141	Protein processing in endoplasmic reticulum	Genetic Information Processing	273
ko05016	Huntington's disease	Human Diseases	264
ko03013	RNA transport	Genetic Information Processing	254
ko00190	Oxidative phosphorylation	Metabolism	254
ko05169	Epstein-Barr virus infection	Human Diseases	248
ko04144	Endocytosis	Cellular Processes	237
ko05203	Viral carcinogenesis	Human Diseases	235
ko04010	MAPK signaling pathway	Environmental Information Processing	230

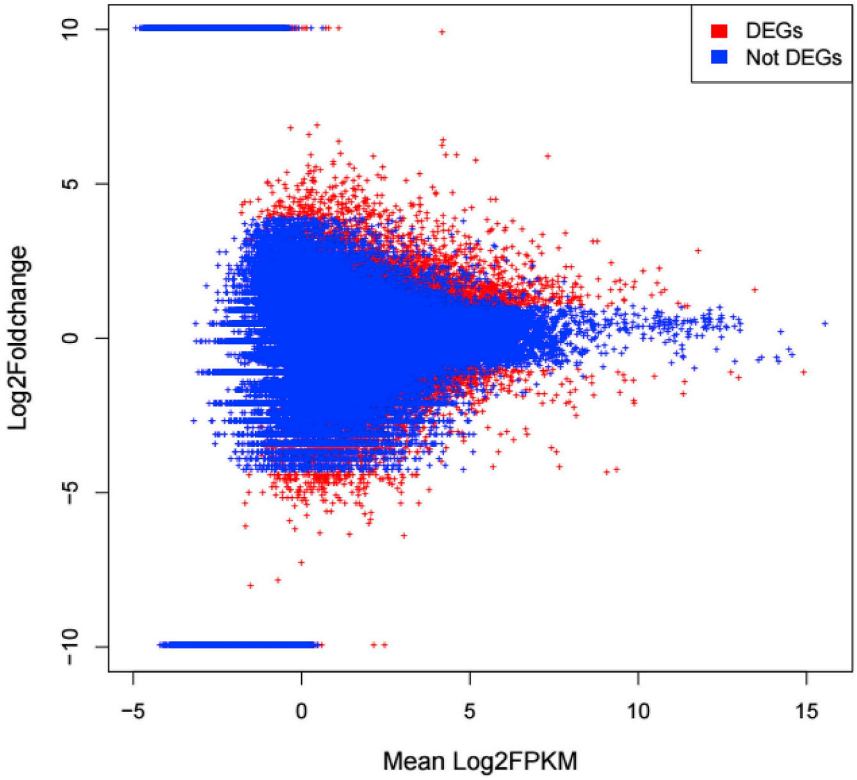

Fig. 3. An MA plot of differentially expressed genes identified in the normal liver library (LIVER) and infected library (LIVER-VP).

Trinity software (http://trinityrnaseq.sourceforge.net/), and the data were assembled into 256663 contigs with an average length of 635.21 bp, 55.84% (143322) of which included lengths of 200–400 bp, 16.18% (41531) were 401–600 bp, 7.91% (20278) were 601–800 bp, 4.75% (12198) were 801–1000 bp, 3.35% (8603) were 1001–1200 bp, and 11.97% (30731) were > 1200 bp (Fig. 1, Table 2). The assembly results indicate that the length distribution pattern and mean length of the contigs was similar to previous transcriptome studies using Illumina sequencing [9,15]. The sequence mass information obtained not only filled the gene expression profile gap for *V. parahaemolyticus*-immunized *Sparus macrocephlus* but facilitated mining of the systematic genetic information resources.

3.2. Transcriptome data functional annotation and classification

ORF predictions were generated using Trinity (http://trinityrnaseq.sourceforge.net) following the assembly, and 214847 contigs were used for the ORF predictions. Subsequently, to annotate the sequences, protein sequences were used for Blastp alignments (E-value < 10^{-5}) with the NR, gene, and string, and 25694 contigs featured a corresponding annotation. Further, the remanent nucleotide sequences were used for a Blastx alignment with NT, NR, gene, and string, and 5053 contigs featured a corresponding annotation (Table S1).

Gene ontology (GO) terms for the transcriptome were analyzed using Blast2GO, which provides information on the “Biological Process”, “Cellular Component”, and “Molecular Function” for each sequence (Fig. 2); 41886 genes were annotated. In the “biological processes” categories, which feature 23 subtypes, most corresponding genes were involved in cellular processes, metabolic processes, biological regulation and biological process regulation. In addition, 19 subtypes were annotated with“cellular components”, and most corresponding genes were involved in the cell, organelles and cell parts. In the “molecular function” category, which featured 20 subtypes, most corresponding genes were involved in binding and catalytic activity.

Table 4
Representative immune-related genes differentially expressed after *V. parahaemolyticus* infection.

Gene name	Description	Change	Fold	P-value	FDR
Toll-like receptor signaling pathway					
Jun	Jun proto-oncogene	Down	0.42	6.20E-05	0.000798
TLR-5	Toll-like receptor 5	Down	0.062	1.73E-100	3.17E-98
TLR-1	Toll-like receptor 1	Up	2.59	2.67E-12	1.31E-10
TNF- α	tumor necrosis factor α	Up	25.2/0	2.13E-07	4.80E-06
IRF7	Interferon regulatory factor 7	Up	6.28	0	0
Pik3c	Phosphatidylinositol-4,5-bisphosphate 3-kinase	Down	0.19	1.22E-05	0.000196
Pik3r	Phosphoinositide-3-kinase, regulatory subunit	Up	2.48	3.33E-12	1.55E-10
LBP	lipopolysaccharide-binding protein	Down	0.30	1.96E-304	3.94E-302
complement and coagulation cascades					
MASP2	mannose-binding protein-associated serine protease 2	Up	2.24	3.26E-12	1.53E-10
C1qA	complement C1q subcomponent subunit A	Up	5.93	7.75E-12	3.23E-10
PLAU	urokinase plasminogen activator	Up	2.87	7.99E-15	6.37E-13
CFD	component factor D	Up	4.22	3.23E-12	1.52E-10
C4	complement C4-like	Up	4.48	6.62E-14	4.40E-12
F8	coagulation factor VIII	Down	0.304	2.32E-30	2.87E-28
TFPI	tissue factor pathway inhibitor	Up	2.05	5.14E-12	2.22E-10
Serping1	serpin peptidase inhibitor, clade G (C1 inhibitor)	Down	0.072	3.57E-41	5.01E-39
C1qB	complement C1q subcomponent subunit B	Up	2.70	1.05E-11	4.23E-10
Antigen processing and presentation					
LGMN	legumain	Up	5.23	9.02E-12	3.68E-10
β2M	beta-2-microglobulin	Up	2.67	6.48E-11	2.33E-09
CD8α	CD8α antigen, alpha polypeptide	Up	2.38	9.83E-12	3.97E-10
CTSB	cathepsin B	Down	0/15.35	1.58E-08	4.22E-07
TAPBP	TAP binding protein	Up	3.38	5.44E-13	3.07E-11
NFYα	nuclear transcription factor Y, alpha	Up	2.30	6.69E-12	2.82E-10
CD8β	CD 8βantigen, beta polypeptide	Up	10.71	9.48E-09	2.62E-07
MHC2	major histocompatibility complex, class II	Up	2.52	1.55E-12	7.92E-11
Chemokine signaling pathway					
WAS	Wiskott-Aldrich syndrome protein	Down	0.23	2.65E-05	0.000387
PXN	paxillin	Down	0.45	8.28E-06	0.000138
GNG10	guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-10	Up	2.07	2.45E-06	4.56E-05
ITK	IL2-inducible T-cell kinase	Up	2.15	6.77E-13	3.71E-11
PREXI	phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchanger 1 protein	Up	2.18	5.49E-12	2.35E-10
CCL25	C−C motif chemokine 25	Up	3.82	4.31E-12	1.91E-10
CXCL12	C-X-C motif chemokine 12	Up	6.69	3.63E-12	1.66E-10
CXCR3	C-X-C chemokine receptor type 3	Up	2.35	2.40E-05	0.000353
Pik3c	Phosphatidylinositol-4,5-bisphosphate 3-kinase	Down	0.19	1.22E-05	0.000196
Pik3r	Phosphoinositide-3-kinase, regulatory subunit	Up	2.48	3.33E-12	1.55E-10
Natural killer cell mediated cytotoxicity					
CD3Z	CD3Z antigen, zeta polypeptide	Up	9.42	7.94E-13	4.27E-11
HCST	Hematopoietic cell signal transducer	Up	5.79	5.24E-13	2.98E-11
TNFSF6, FASL	tumor necrosis factor ligand superfamily member 6	Up	3.15	4.55E-14	3.22E-12
GZMB	granzyme B	Up	2.36	3.00E-10	1.00E-08
ZAP70	tyrosine-protein kinase ZAP-70	Up	2.09	5.88E-13	3.29E-11
Pik3c	Phosphatidylinositol-4,5-bisphosphate 3-kinase	Down	0.19	1.22E-05	0.000196
Pik3r	Phosphoinositide-3-kinase, regulatory subunit	Up	2.48	3.33E-12	1.55E-10
MAPK signaling pathway					
Jun	Jun proto-oncogene	Down	0.42	6.20E-05	0.000798
STMN1	stathmin	Up	2.75	2.97E-12	1.43E-10
TNFSF6, FASL	tumor necrosis factor ligand superfamily member 6	Up	3.15	4.55E-14	3.22E-12
PDGFRβ	platelet-derived growth factor receptor beta	Up	2.47	2.36E-09	7.02E-08
CACNα1H	voltage-dependent calcium channel T type alpha-1H	Down	0.33	2.85E-05	0.000410053
MKNK	MNK MAP kinase interacting serine/threonine kinase	Down	0.43	1.25E-21	1.33E-19
CACNα1G	voltage-dependent calcium channel T type alpha-1G	Down	0/16.11	5.37E-06	9.26E-05
MAP4K4	HGK mitogen-activated protein kinase kinase kinase kinase 4	Up	2.15	1.42E-09	4.35E-08

The genes from the merged groups were categorized using the KEGG (Kyoto Encyclopedia of Genes and Genomes) database to obtain more information to predict the unigene functions [34]; 25798 genes ware classified into 358 KEGG pathways (Table S2.1). The top 15 statistically significant KEGG classifications are shown in Table 3. “Mitogen-activated protein kinase (MAPK) signaling pathway” and “PI3K-Akt signaling pathway” play a key role in signal transduction and immunoregulation.

3.3. Differentially expressed genes and clustering

In this study, false discovery rate (FDR) < 0.001, fold change ≥ 2 and p-value ≤ 0.05 were used to define significantly up-regulated or down-regulation genes following immunization. Based on this standard, 131846 contigs were differentially expressed, and 7843 genes were significantly differentially expressed at greater than two-fold, including 4072 up-regulated genes and 3771 down-regulated genes (Table S2.2). The substantial sequence information compensates for the incomplete information on the immune-related genes for *V. parahaemolyticus* -immunized *Sparus macrocephlus*. The magnitude distribution for the differentially expressed genes is illustrated using an MA plot analysis (Fig. 3) [35]. To visualize the gene expression profiles, we generated a heatmap using the Hcluster algorithm, where the color changes from red to green with a decrease in expression.

3.4. Identification of differentially expressed immune-related genes

The up- and down-regulated genes were mainly annotated with the

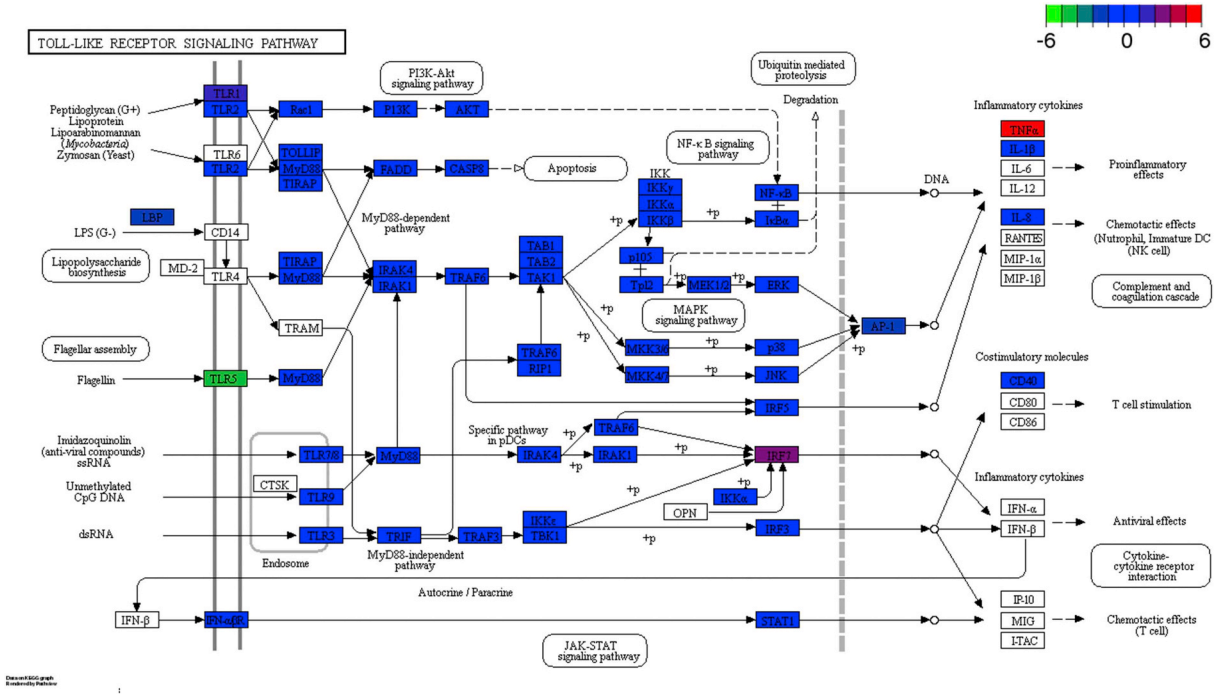

Fig. 4. Significantly differentially expressed genes identified by KEGG as involved in Toll-like receptor signaling pathway. Red boxes indicate significantly increased expression; Green boxes indicate significantly decreased expression; Blue boxes indicate unchanged expression. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

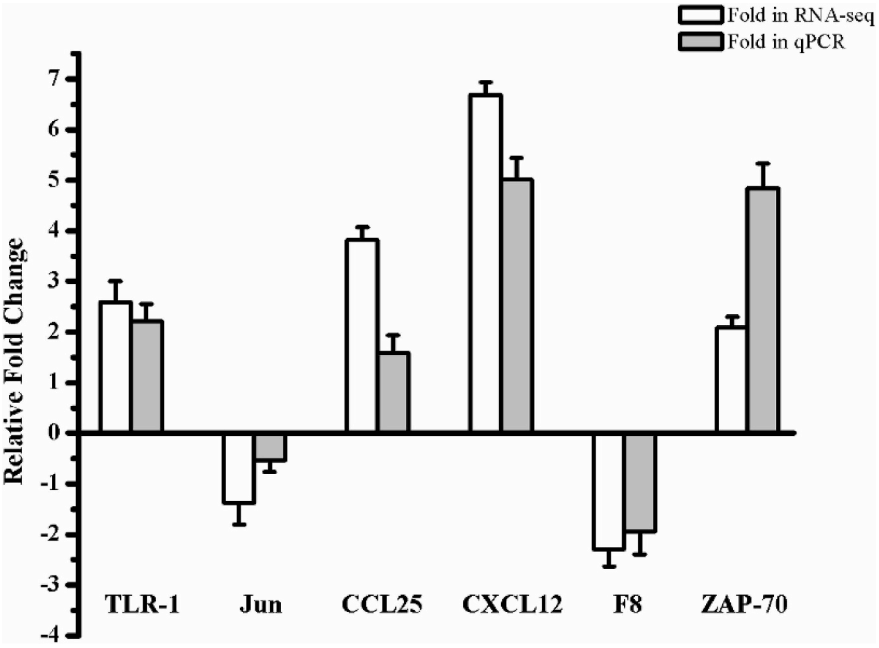

Fig. 5. Comparison of relative fold change of RNA-Seq and QPCR results between the normal and VP-immunized liver. The expression levels of selected genes were each normalized to that of the GAPDH gene.

GO terms for biological processes (cell killing, immune system process, response to stimulus), cellular processes, binding and metabolism (Table S2.3). For the KEGG analyses, 2708 differentially expressed isoforms were annotated with 323 immune and signaling transduction related pathways (Table S2.4). For example, for the 'immune system', we identified 19 pathways (containing 204 genes), including pathways associated with complement and coagulation cascades (10 genes, ko04610), antigen processing and presentation (14 genes, ko04612), toll-like receptor signaling (8 genes, ko04620), T cell receptor signaling (16 genes, ko04660), B cell receptor signaling (7 genes, ko04662), natural killer cell mediated cytotoxicity (13 genes, ko04650), RIG-I-like receptor signaling (6 genes, ko04622), NOD-like receptor signaling (4 genes, ko04621), chemokine signaling (11 genes, ko04062), peroxisome (14 genes, ko04146), phagosome (18 genes, ko04145), apoptosis (11 genes, ko04210), MAPK signaling (13 genes, ko04010), PPAR signaling (18 genes, ko03320), Jak-STAT signaling (12 genes, ko04630), ErbB signaling (5 genes, ko04012), Fc gamma R-mediated phagocytosis (6 genes, ko04666), leukocyte transendothelial migration (11 genes,

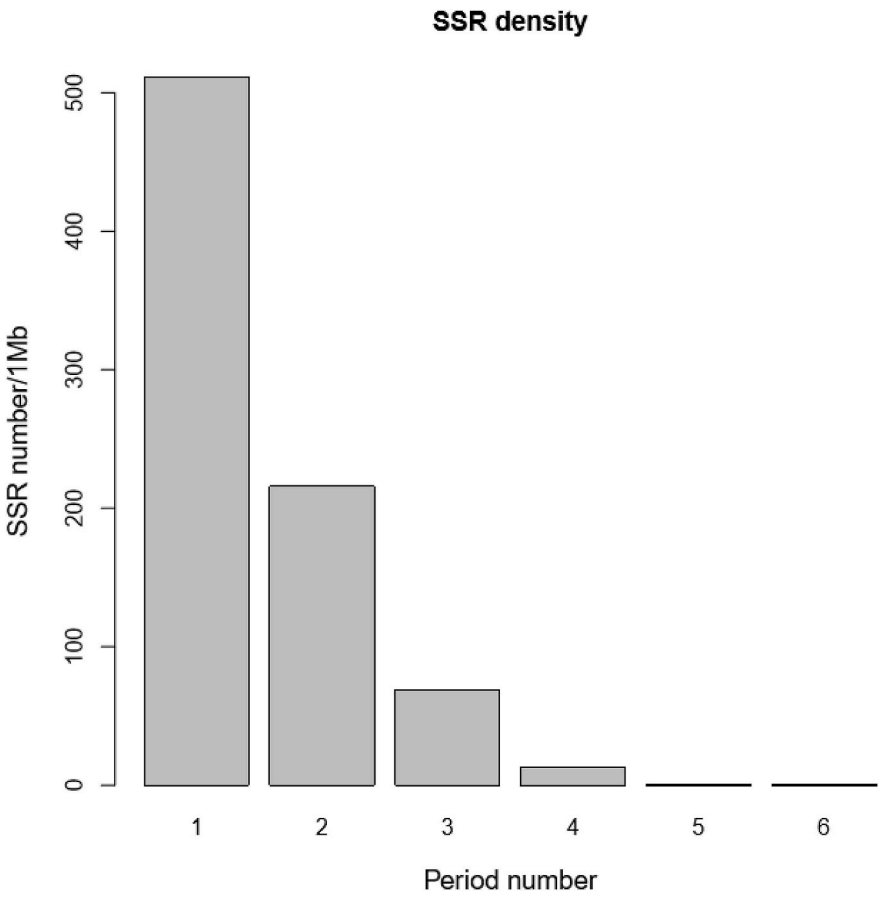

Fig. 6. A summary of the SSRs identified from the *Sparus macrocephlus* transcriptome.

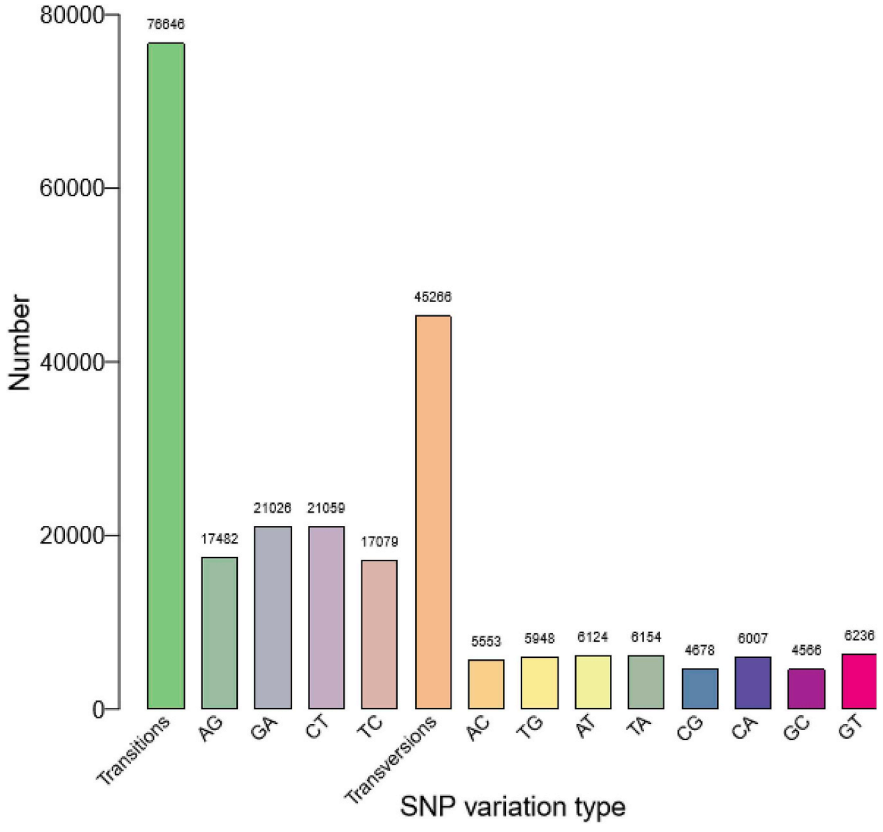

Fig. 7. Distribution of putative SNP in the *Sparus macrocephlus* transcriptome.

ko04670), and ECM-receptor interaction (7 genes, ko04512). Certain significantly differentially expressed genes are shown in Table 4. These pathways and genes that are related to the immune system, signaling transduction and disease processes are similar to previous studies in *Danio rerio* [8], *Epinephelus coioides* [10], *Megalobrama amblycephala* [11] and *Pseudosciaena crocea* [15]. Compared with previously analyzed single genes, the immune-related pathways and genes aid in systematically comprehending the molecular immune mechanisms underlying the *V. parahaemolyticus* response.

3.5. Toll-like receptor signaling pathway

To fully understand the molecular mechanisms underlying the response to *V. parahaemolyticus*, we first analyzed the liver transcriptome profile of the *Sparus macrocephlus* after *V. parahaemolyticus* infection and obtained gene information. Subsequently, a gene expression profile analysis was performed and identified many immune-related genes. The innate immune system is the first line of defense against pathogenic microorganism infections in fish. TLRs, which compose one group of well-known PRRs, can monitor antigen molecules from pathogenic microorganisms and activate the innate immune response. Eight representative immune-related genes were significantly differentially expressed in the toll-like receptor signaling pathway (Table 4, Fig. 4). Among them, *TLR1* gene was upregulated, which has been reported can recognizes the flagellin protein component of bacterial flagella [36]. We speculate that *TLR1* gene might be involved in identifying the antigen composition to enhance host resistance to *V. parahaemolyticus*. *TNF-α* gene was significantly up-regulated following *V. parahaemolyticus* infection, but the gene was not identified in the control group. We speculate that the *TNF-α* gene plays a key role in resistance to *V. parahaemolyticus*. Upon stimulation by *V. parahaemolyticus*, *IRF7* gene was significantly up-regulated, which promoted pro inflammatory cytokine and inflammatory cytokine expression, such as for IL-1β, IL-6, IFN-α and IFN-β. During early infection, the inflammatory response provides a major contribution to anti-pathogenic microorganism mechanisms [37]. Further, expression for the *PIK3C* and *LBP* genes was significantly down-regulated after stimulation by *V. parahaemolyticus*, which may be the negative regulation mechanism for an over-inflammatory response. We obtained similar and different results with other bacterial pathogen, which suggests that certain specific genes in the toll-like receptor signaling pathway play a pivotal role in the response to *V. parahaemolyticus*.

3.6. Validation of differentially expressed genes using real-time PCR

In this study, six of the differentially expressed genes in the infected *Sparus macrocephlus* transcriptome that are involved in the immune system were randomly selected for real-time PCR validation. As shown in Fig. 5, the real-time PCR results significantly correlated with the RNA-Seq results.

3.7. SSR and SNP discovery

Among various molecular markers, SSRs feature many putative functions and are widely used in certain aspects of parentage, genetic diversity, linkage mapping, and marker-assisted breeding [38,39]. To develop new molecular markers for *Sparus macrocephlus*, the 214513 transcripts were used to mine potential microsatellites that were defined as mononucleotide to hexanucleotide SSRs with a minimum of four repetitions for each motifs. Using the Msatcommander, 92175 potential SSRs were identified, and 33922 potential SSRs contained repeats of more than two nucleotides (Fig. 6, Table S2.5). Among the 33922 SSRs, the dinucleotides repeats numbered 24591 (72.49%) followed by trinucleotide (7847, 23.13%), tetranucleotide (1439, 4.24%), pentanucleotide (31, 0.091%) and hexanucleotide repeat motifs (14, 0.041%). Of the dinucleotide SSRs, AC/GT was most common and accounted for 73.61% and CG (0.03% for the present study). The most common trinucleotide repeat was AGG/CCT and AAG/CTT, which accounted for 29.63% and 18.58%, respectively. The most common tetranucleotide repeat was AAAG/CTTT which accounted for 20.15%.

SNPs compose the most abundant type of DNA sequence polymorphism and have been increasingly used as molecular markers in quantitative trait loci (QTL) mapping, linkage map construction, and association studies [40,41]. Compared with the assembled transcript sequences using Samtools and VarScan v.2.2.7, 121912 candidate SNPs were identified (Table S2.6). Of these SNP candidates, 76646 SNPs were putative transitions (Ts), and 45266 SNPs were putative transversions (Tv) with a mean Ts: Tv ratio of 1.66. The detail information is shown in Fig. 7. The SNPs were then categorized 4 types, including class 1 (C/A, A/C, T/G and G/T) at 23744 (19.48%), class 2 (C/T, G/A, T/C and A/G) at 76646 (62.87%), class 3 (C/G and G/C) at 9244 (7.58%), and class 4 (A/T and T/A) at 12278 (10.07%).

4. Conclusions

In this study, we first performed a comparative gene transcription analysis for *V. parahaemolyticus*-immunized *Sparus macrocephlus* to explore the molecular immune mechanisms and develop molecular markers. We generated 256663 contigs though *de novo* assembly using multiple programs and steps. Gene expression in *V. parahaemolyticus*-infected *Sparus macrocephlus* revealed significant changes in multiple immune-related pathways, such as the toll-like receptor signaling pathway, chemokine signaling pathways, and complement and coagulation cascades. In addition, 33922 SSRs and 121912 SNP were identified in the *Sparus macrocephlus* liver transcriptome, which is helpful for subsequent marker development, genetic linkage and QTL analysis. In conclusion, this study provides deep insight into the immune defense mechanisms of *Sparus macrocephlus* against *V. parahaemolyticus*, which may aid in developing an effective solution for fighting this serious fish disease.

Acknowledgments

This work was supported by grants from International Science & Technology Cooperation Program of China (No. 2015DFA30980).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fsi.2018.09.057.

References

[1] F. Zhou, J.X. Xiao, Y. Hua, B.O. Ngandzali, Q.J. Shao, Dietary L-methionine requirement of juvenile black sea bream (*Sparus macrocephalus*) at a constant dietary cystine level, Aquacult. Nutr. 17 (5) (2011) 469–481.
[2] W.S. Kim, K.H. Kong, J.O. Kim, S.J. Jung, J.H. Kim, M.J. Oh, Amoebic gill disease outbreak in marine fish cultured in Korea, Official Publication of the American Association of Veterinary Laboratory Diagnosticians Inc, J. Vet. Diagn. Invest. 29 (3) (2017) 1040638717690783.
[3] W.S. Kim, M.J. Oh, Hirame rhabdovirus (HIRRV) as the cause of a natural disease outbreak in cultured black seabream (*Acanthopagrus schlegeli*) in Korea, Arch. Virol. 160 (12) (2015) 3063–3066.
[4] L. Li, S.L. Lin, L. Deng, Z.G. Liu, Potential use of chitosan nanoparticles for oral delivery of DNA vaccine in black seabream *Acanthopagrus schlegelii* Bleeker to protect from *Vibrio parahaemolyticus*, J. Fish. Dis. 36 (12) (2013) 987–995.
[5] G. Li, Y. Zhao, Z. Liu, C. Gao, F. Yan, B. Liu, J. Feng, De novo assembly and characterization of the spleen transcriptome ofcommon carp (*Cyprinus carpio*) using Illumina paired-end sequencing, Fish Shellfish Immunol. 44 (2) (2015) 420–429.
[6] X. Han, X. Wu, W.Y. Chung, T. Li, A. Nekrutenko, N.S. Altman, G. Chen, H. Ma, Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing, Proc. Natl. Acad. Sci. U.S.A. 106 (31) (2009) 12741–12746.
[7] S. Liu, Z. Yu, Z. Zhou, W. Geoff, F. Sun, J. Lu, J. Zhang, Y. Jiang, Z. Hao, X. Wang, Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote, BMC Genom. 13 (1) (2012) 595.
[8] A.J. Lü, X.C. Hu, Y. Wang, A.H. Zhu, L.L. Shen, J. Tian, Z.Z. Feng, Z.J. Feng, Skin immune response in the zebrafish, *Danio rerio* (Hamilton), to Aeromonas hydrophila

infection: a transcriptional profiling approach, J. Fish. Dis. 38 (2) (2015) 137–150.
[9] X. Zhang, S. Wang, S. Chen, Y. Chen, Y. Liu, C. Shao, Q. Wang, Y. Lu, G. Gong, S. Ding, Transcriptome analysis revealed changes of multiple genes involved in immunity in *Cynoglossus semilaevis* during *Vibrio anguillarum* infection, Fish Shellfish Immunol. 43 (1) (2015) 209–218.
[10] Y. Huang, X. Huang, Y. Yan, J. Cai, Z. Ouyang, H. Cui, P. Wang, Q. Qin, Transcriptome analysis of orange-spotted grouper (*Epinephelus coioides*) spleen in response to Singapore grouper iridovirus, BMC Genom. 12 (2011) 556.
[11] N.T. Tran, Z.X. Gao, H.H. Zhao, S.K. Yi, B.X. Chen, Y.H. Zhao, L. Lin, X.Q. Liu, W.M. Wang, Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*, Fish Shellfish Immunol. 45 (1) (2015) 72–82.
[12] L. Cárdenas, R. Sánchez, D. Gomez, G. Fuenzalida, C. Gallardo-Escárate, A. Tanguy, Transcriptome analysis in *Concholepas concholepas* (Gastropoda, Muricidae): mining and characterization of new genomic and molecular markers, Marine Genomics 4 (3) (2011) 197–205.
[13] J. Chen, C. Li, R. Huang, F. Du, L. Liao, Z. Zhu, Y. Wang, Transcriptome analysis of head kidney in grass carp and discovery of immune-related genes, BMC Vet. Res. 8 (2012) 108.
[14] J. Cui, H. Wang, S. Liu, X. Qiu, Z. Jiang, X. Wang, Transcriptome analysis of the gill of *Takifugu rubripes* using Illumina sequencing for discovery of SNPs, Comp. Biochem. Physiol. Genom. Proteonomics 10 (1) (2014) 44–51.
[15] P. Wang, J. Wang, Y.Q. Su, Y. Mao, J.S. Zhang, C.W. Wu, Q.Z. Ke, K.H. Han, W.Q. Zheng, N.D. Xu, Transcriptome analysis of the *Larimichthys crocea* liver in response to *Cryptocaryon irritans*, Fish Shellfish Immunol. 48 (2016) 1–11.
[16] Q. Zhang, C. Ji, J. Ren, Q. Zhang, X. Dong, Y. Zu, L. Jia, W. Li, Differential transcriptome analysis of zebrafish (*Danio rerio*) larvae challenged by *Vibrio parahaemolyticus*, J. Fish. Dis. 41 (7) (2018) 1049–1062.
[17] Q.L. Zhang, Q.H. Zhu, M.Z. Liang, F. Wang, J. Guo, X.Y. Deng, J.Y. Chen, Y.J. Wang, L.B. Lin, Comparative transcriptomic analysis provides insights into antibacterial mechanisms of *Branchiostoma belcheri* under *Vibrio parahaemolyticus* infection, Fish Shellfish Immunol. 76 (2018) 196–205.
[18] F. Jiang, X. Yue, H. Wang, B. Liu, Transcriptome profiles of the clam *Meretrix petechialis* hepatopancreas in response to *Vibrio* infection, Fish Shellfish Immunol. 62 (2017) 175–183.
[19] Z. Qin, V.S. Babu, Q. Wan, M. Zhou, R. Liang, A. Muhammad, L. Zhao, J. Li, J. Lan, L. Lin, Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei*) challenged by *Vibrio parahaemolyticus* reveals unique immune-related genes, Fish Shellfish Immunol. 77 (2018) 164–174.
[20] R. Rao, Y.B. Zhu, T. Alinejad, S. Tiruvayipati, K.L. Thong, J. Wang, S. Bhassu, RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection, Gut Pathog. 7 (2015) 6.
[21] C. Xie, Y. Chen, W. Sun, J. Ding, L. Zhou, S. Wang, S. Wang, Y. Zhang, D. Zhu, X. Wen, Transcriptome and expression profiling analysis of the hemocytes reveals a large number of immune-related genes in Mud Crab *Scylla paramamosain* during *Vibrio parahaemolyticus* infection, PLoS One 9 (12) (2014) e114500.
[22] X. Zhao, X. Duan, Z. Wang, W. Zhang, Y. Li, C. Jin, J. Xiong, C. Li, Comparative transcriptome analysis of *Sinonovacula constricta* in gills and hepatopancreas in response to *Vibrio parahaemolyticus* infection, Fish Shellfish Immunol. 67 (2017) 523–535.
[23] Q. Ge, J. Li, J. Wang, J. Li, H. Ge, Q. Zhai, Transcriptome analysis of the hepatopancreas in *Exopalaemon carinicauda* infected with an AHPND-causing strain of *Vibrio parahaemolyticus*, Fish Shellfish Immunol. 67 (2017) 620–633.
[24] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (7) (2011) 644–652.
[25] A. Conesa, S. Götz, J.M. Garcíagómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (18) (2005) 3674–3676.
[26] S. Götz, J.M. Garcíagómez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, M. Robles, M. Talón, J. Dopazo, A. Conesa, High-throughput functional annotation and data mining with the Blast2GO suite, Nucleic Acids Res. 36 (10) (2008) 3420–3435.
[27] C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C.Y. Li, L. Wei, KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases, Nucleic Acids Res. 39 (Web Server issue) (2011) 316–322.
[28] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359.
[29] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinf. 12 (2011) 323.
[30] M. Robinson, D. McCarthy, G. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (1) (2010) 139–140.
[31] B. Faircloth, msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design, Molecular Ecology Resources 8 (1) (2008) 92–94.
[32] H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, Bioinformatics 27 (21) (2011) 2987–2993.
[33] D. Koboldt, K. Chen, T. Wylie, D. Larson, M. McLellan, E. Mardis, G. Weinstock, R. Wilson, L. Ding, VarScan: variant detection in massively parallel sequencing of individual and pooled samples, Bioinformatics 25 (17) (2009) 2283–2285.
[34] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30.
[35] L. Wang, Z. Feng, X. Wang, X. Wang, X. Zhang, DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, Bioinformatics 26 (1) (2010) 136–138.
[36] D. Pietretti, G.F. Wiegertjes, Ligand specificities of Toll-like receptors in fish: indications from infection studies, Dev. Comp. Immunol. 43 (2) (2014) 205–222.
[37] A. Iwasaki, R. Medzhitov, Toll-like receptor control of the adaptive immune responses, Nat. Immunol. 5 (10) (2004) 987–995.
[38] D. Dawson, A. Ball, L. Spurgin, D. Martín-Gálvez, I. Stewart, G. Horsburgh, J. Potter, M. Molina-Morales, A. Bicknell, S. Preston, R. Ekblom, J. Slate, T. Burke, High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species, BMC Genom. 14 (2013) 176.
[39] D.A. Chistiakov, B. Hellemans, F.A.M. Volckaert, Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics, Aquaculture 255 (1) (2006) 1–29.
[40] H. Yu, W. Xie, J. Wang, Y. Xing, C. Xu, X. Li, J. Xiao, Q. Zhang, Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers, PLoS One 6 (3) (2011) e17595.
[41] S. Hubert, B. Higgins, T. Borza, S. Bowman, Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*), BMC Genom. 11 (2010) 191.