Fish and Shellfish Immunology 88 (2019) 237–243

Contents lists available at ScienceDirect

Fish and Shellfish Immunology

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

Full length article

Transcriptome analysis and discovery of genes involved in immune pathways in *Solen strictus* (Gould, 1861) under *Vibrio anguillarum*

Qiang Wang[a], Jing Li[b], Hongmei Guo[a,*]

[a] Department of Animal Nutrition and Aquaculture, Shandong Vocational Animal Science and Veterinary College, 88 Shengli East Street, Weifang 261061, China
[b] College of Economics and Management in Shanghai Ocean University, Shanghai 201306, China

ARTICLE INFO

Keywords:
Solen strictus
Vibrio anguillarum
Transcriptome
Immune response

ABSTRACT

The *Solen strictus* (Gould, 1861) has been recognized as an important marine economic bivalve in Eastern and Southeast Asia. To gain a better understanding of the *S. strictus* immune system and its related genes in response to bacterial infections, we performed a comparative gene transcription analysis from *S. strictus* with *Vibrio anguillarum* through RNA-Seq technology, meanwhile the differentially expressed genes (DEGs) were investigated. After assembly, a total of 195,774 transcripts with an average length of 996 bp were obtained. Total 153,038 unigenes were annotated in the nr, Swiss-Prot, KEGG, COG, KOG, GO and Pfam databases, and 56,597 unigenes (36.98%) were annotated in at least one database. After bacterial challenge, there were 1588 significant differentially expressed genes (DEGs) between the challenged and control groups, including 999 up-regulated and 589 down-regulated genes. All the DEGs were classified into three gene ontology categories, and allocated to 225 KEGG pathways. Immune-related genes were detected from immune system pathways among the top 20 enriched pathways, such as Toll-like receptor signaling, RIG-I-like receptor signaling and NOD-like receptor signaling pathway. In addition, 56,079 potential simple sequence repeats (SSRs) and 1,031,521 candidate single nucleotide polymorphisms (SNPs) were detected and identified in the *S. strictus* transcriptome. Results of the present study will provide valuable theoretical resources for future genetic and genomic research on *S. strictus*. The research results will be helpful for improving the efficiency and quality of artificial breeding, establishing genetic linkage map, and enhancing health management for this species.

1. Introduction

As invertebrates, mollusks rely exclusively on the innate immune system to execute cellular and humoral immune reactions, which plays crucial roles in immune response, overall homeostasis and fundamental defense mechanism [1,2]. In recent years, transcriptome analysis has been a powerful tool for understanding the underlying pathways and mechanisms of host to diseases, pathogens and environmental challenges [3]. And the high-throughput transcriptome sequencing has facilitated the functional genomic study of organisms and provided insights into the extremely environment effects on aquatic animals, especially for the non-model organisms and those lacking reference genomes [4]. Infectious bacteria diseases such as *Listonella anguillarum*, *Edwardsiella tarda*, *Vibrio harveyi* and *Vibrio aestuarianu* have seriously hindered the development of aquaculture. In recent years, the Gram-negative bacterium *V. anguillarum* is the main reason of high mortality of mollusks leading to the enormous economic losses. Recently, many bivalve species have characterized the mechanism of host-pathogen interactions and immune responses using transcriptome sequencing, such as *Ruditapes philippinarum*, *Concholepas concholepas* and *Sinonovacula constricta* [5–7]. However, knowledge of the immune system of *Solen strictus* and different signaling pathways implicated in its immune response remains incomplete.

The razor shell *Solen strictus* Gould, 1861 belongs to Mollusca, Bivalvia, Heterodonta, Veneroida and Solenidae, which is distributed in temperate zone and tropical zone. The *S. strictus* distributes widely along the coasts of the Bohai Sea and the Yellow Sea in China, which is a commercial and potential mariculture species due to its rapid growth, delicious taste, and good market price. Thus, understanding *S. strictus*'s biology and genome resources is valuable. Furthermore, understanding the molecular mechanisms of *S. strictus* adapts to variable environments is necessary for its breeding and conservation.

In this study, we characterized the transcriptional profiles of gills in razor shell *Solen strictus* with *Vibrio anguillarum*. RNA-Seq was performed using Illumina Hiseq 4000 platform, the transcripts were assembled to unigenes, and KOG, GO, KEGG enrichment analyses

* Corresponding author.
E-mail address: sdmyghm@163.com (H. Guo).

https://doi.org/10.1016/j.fsi.2019.01.024
Received 2 November 2018; Received in revised form 10 January 2019; Accepted 20 January 2019
Available online 18 February 2019

revealed unigenes that are involved in different biological metabolism and signaling pathways. In addition, differentially expressed genes (DEGs) were identified, and an enrichment analysis was conducted based on comparisons in two *S. strictus* libraries. This study provides valuable information for future genetics and genomics studies on *S. strictus* and the molecular mechanism underlying the host-*V. anguillarum* interaction and may contribute to the genetic improvement of the growth and disease resistance of *S. strictus* in the future.

2. Materials and methods

2.1. Experimental S. strictus and treatment

Live *S. strictus* with an average weight of 12 g and a length of 10 cm, were obtained from Shandong Weifang Longwei Industrial CO.LTD in Weifang, Shandong Province, China. *S. strictus* were cleaned to remove any fouling and were acclimated in aerated 100 L plastic tanks, containing water at 20 °C with salinity of 30 ppt. All the clams were fed with Spirulina powder daily for two days before bacterial challenge, and water was exchanged once per day to discharge waste products. *S. strictus* was randomly distributed into 2 tanks with 30 individuals per tank. For bacterial challenge, *S. strictus* were challenged with 50 μl V. *anguillarum* (5×10^6 cfu ml^{-1} in PBS) for 24 h. For blank control, *S. strictus* received an injection of 50 μl PBS. At 24 h post challenge, the gill was collected from 3 individuals from each tank, and then all these samples were stored at −80 °C for the subsequent experiment.

2.2. RNA extraction and sequencing

Total RNA was extracted using TRIzol Reagent (Aidlab, China) treated with RNase-Free DNase (Promega, USA). The RNA concentration and structural integrity were evaluated using a NanoDrop ND1000 spectrophotometer (Thermo Scientific, USA). For RNA library preparation, the RNA of each amphioxus individual from both control and treatment groups was diluted to the same concentration with RNase-free water, and these were pooled in equal amounts to obtain the control and treatment groups (each group contains three individuals).

The NEBNext Ultra RNA Library Prep Kit (NEB, USA) was used to generate a cDNA library. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations. First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activity. Following adenylation of the 3′ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. To select cDNA fragments of 150–200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then, 3 μL USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37 °C for 15 min, followed by 5 min at 95 °C prior to PCR. PCR was then performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) primer. Finally, PCR products were purified (AMPure, XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. Illumina HiSeq 4000 sequencing was carried out by Novogene, Tianjin, China.

2.3. De novo assembly of transcriptom, annotation, and function enrichment

The assembly of two sets of gill was performed by using RNA-Seq *de novo* programs Trinity *de novo* software with default parameters [8]. Clean reads were obtained after removing the low quality reads (Quality score > 20) and short reads (Read length > 10bp). The clean reads were clustered to form contigs, and then these contigs were processed to obtain unigenes, which were finally connected to obtain transcript. Expression abundance was measured using the reads per kilobase per million mapped reads (RPKM) method [9].

After *de novo* assembly of transcriptome, further annotation of transcripts was performed on the basis of sequence similarity with previously described genes. In brief, Unigenes were used for Blast and annotation against seven databases, including NCBI non-redundant protein database (NR, evalue = 1e-5), NCBI non-redundant nucleotide sequences (NT, evalue = 1e-5), Pfam (http://pfam.sanger.ac.uk/, e-value = 0.01), The Eukaryotic Orthologous Groups (KOG/COG: http://www.ncbi.nlm.nih.gov/COG/, e-value = 1e-3), Swiss-prot (http://www.ebi.ac.uk/uniprot/, e-value = 1e-5), KEGG (http://www.genome.jp/kegg, evalue = 1e-10) [10] and GO (http://www.geneontology.org/, evalue = 1e-6) [11]. Unigenes were defined according to top hits against known sequence that were retrieved from blast search. The Blast2GO program was used to obtain gene ontology (GO) annotation of unigenes based on NR annotation, and then WEGO software was used to perform GO functional classification [12]. The KEGG pathway and KEGG Orthology (KO) of unigenes were analyzed based on blastx hits against the KEGG database.

2.4. Analysis of differentially expressed genes

The expression levels of genes in each library were calculated using RSEM (RNA-seq by Expectation Maximization) based on the RPKM value [10,13]. Readcount for each gene was obtained from the mapping results. Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeaR program package through one scaling normalized factor [14]. Differential expression analysis of two libraries was performed using DEGseq R package [15]. *P*-value was adjusted using q-value. Q-value < 0.005 & |log2(foldchange)| > 1 was set as the threshold for significantly differential expression. DEGs were then carried out into GO functional enrichment analysis and KEGG pathway analysis. GO enrichment analysis of DEGs was implemented by GOseq R package based Wallenius non-central hyper-genometric distribution, which can adjust for gene length bias in DEGs [16]. KOBAS software was used to test the statistical enrichment of DEGs in KEGG pathway [17].

2.5. Identification of SSR and SNP

Simple sequence repeats (SSRs) of the transcriptome were identified using MISA software (http://pgrc.ipk-gatersleben.de/misa/misa.html) and primer for each SSR was designed by Primer (http://primer3.sourceforge.net/releases.php). The parameters were adjusted to identify mono-, di-, tri-, tetra-, penta- and hexanucleotide motif with a unit size of 10, 6, 5, 5 and 5 repeats, respectively. Picard-tools v1.41 and samtools v0.1.18 were used to sort, remove duplicated reads and merge the bam alignment results of each sample. GATK3 software was used to perform single nucleotide polymorphisms (SNPs) calling [18].

2.6. Validation of RNA-seq data and survey of gene expression by qRT-PCR

To validate our RNA-Seq sequencing data, 15 putative DEGs were randomly selected for qPCR analysis, using the same RNA samples as for the RNA-Seq profiling. Total RNA purification and evaluation were performed based on the above mentioned methods, and this was performed three times independently to generate three biological replicates. Primers used in the qRT-PCR were designed automatically using primer primer 5 (Table S1). For preparation of cDNA libraries, first-strand cDNA was synthesized from total RNA using the SYBR® Prime Script™ RT-PCR Kit II (Takara, Japan). Then, we used RNase-free water to dilute the preliminary cDNA solution (final concentration, 100 ng/μl). qRTPCR was performed on the ABI PRISM 7300 Fast Real-Time PCR System (Applied Biosystems, USA) with a SYBR® Premix Ex Taq II Kit (Takara, Japan). The *β*-actin gene was used as a reference gene. The qRT-PCR reaction of each gene in each sample was repeated

three times, and experiments were performed in three biological replicates. The $2^{-\triangle\triangle CT}$ method was used to normalize the expression results of each gene, and final results are shown as means ± standard deviations (SD).

3. Results and discussion

3.1. De novo assembly of the Solen strictus gill transcriptome

In the present study, we used RNA-seq to elucidate the immune mechanisms of *Solen strictus* after *V. anguillarum* infection. In aquatic organisms, gills, as the primary organ for respiration and immuneregulation [19–21], present a large surface area in direct contact with the surrounding environment, thus acting as a fast response to bacterial stress. The gills after *V. anguillarum* infection were selected to study the *S. strictus* transcriptome. Results of the transcriptomic analysis provided more information for further researches on the immune defense mechanisms in *S. strictus*.

After filtering the raw data, a total of 161.8 and 175.6 million clean reads were obtained for the treatment and control groups, respectively (Table S1, SRA: PRJNA500072). The results showed that more than 97% and 93% of each sample had the quality scores of Q20 and Q30 (nucleotides with quality value larger than 20 and 30), respectively. Besides, the content of GC was about equal (39%) per library. All the clean reads were spliced into 195,774 transcripts (total length of 195,034,869) with mean length of 996, N50 length of 1931 and N90 length of 354 (Table S2). These results indicated that the sequences data were high-quality and the transcriptome analysis results were reliable. The length of assembled transcripts revealed that the most of the transcripts (100,985) were less 500 bp, 37,331 of transcripts ranged from 500 bp to 1000 bp, 30,392 of transcripts ranged from 1000 bp to 2000 bp and 27,093 of transcripts were over 2000 bp (Fig. 1A).

3.2. Functional annotation and classification

In order to obtain comprehensive gene function information, we used seven different databases to annotate the unigenes, including NR, NT, Pfam, KOG/COG, Swiss-prot, KEGG, and GO. Total of 153,038 unigenes were annotated (total length of 184,158,114) with mean length of 1203, N50 length of 2065 and N90 length of 482 were further assembled (Fig. 1B). Meanwhile, Fig. 1B also revealed that the most of the unigenes (58,348) ranged from 200 bp to 500 bp, 37,207 unigenes ranged from 500 bp to 1000 bp, 30,390 of unigenes ranged from 1000 bp to 2000 bp, and 27,093 of unigenes (as the same as transcripts) were over 2000 bp. Out of 153,038 unigenes 56,597 (36.98%) were annotated in at least one database, and 3465 genes (2.26%) were annotated in all seven databases. The annotation success rate of unigenes was 43,249 in NR (28.26%), 6,562 in NT (4.28%), 16,143 in KO (10.54%), 32,488 in SwissProt (21.22%), 44,036 in Pfam (28.77%), 44,183 in GO (28.87%), and 17,516 in KOG (11.44%) (Table 1).

Table 1
The annotated gene number in different database.

	Number of Unigenes	Percentage (%)
NR	43249	28.26
NT	6562	4.28
KO	16143	10.54
SwissProt	32488	21.22
PFAM	44036	28.77
GO	44183	28.87
KOG	17516	11.44
All Databases	3465	2.26
At least one Database	56597	36.98
Total Unigenes	153038	100

Among these unigenes, about 35% of unigenes possessed an Evalue more than 1e-30 (Fig. 2A) and 59% of unigenes shared more than 60% similarities (Fig. 2B) with the annotated results in the NCBI database. The result of species classification indicated that *S. strictus* has the highest number of hits to the *Crassostrea gigas* (53%), followed by *Lottia gigantean* (11.8%), *Aplysia californica* (8.1%), *Branchiostoma floridae* (2.0%), *Capitella teleta* (1.8%) and others (23.4%) (Fig. 2C).

3.3. Functional classification, enrichment, and pathway analysis

To further investigate the biological importance of the unigenes, Blast2GO was performed to determine the functional categories. 44,183 unigenes were classified into three main categories (molecular functional, cellular component, and biological process) and 56 subcategories based on the NR annotation from NCBI. Within biological process category, most unigenes were assigned into cellular process (25,468, 57.64%), binding (24,231, 54.84%), single-organism process (20,962, 47.44%) and metabolic process (20,617, 46.66%) subcategories (Fig. S1). From GO classification, we can know that 7814 (17.69%) and 606 (1.37%) unigenes were assigned to response to stimulus and immune system process, respectively.

In the present study, to predict the assembled unigenes possible functions and classify them, they were further aligned to the KOG database and 17,516 unigenes were classified into 26 functional

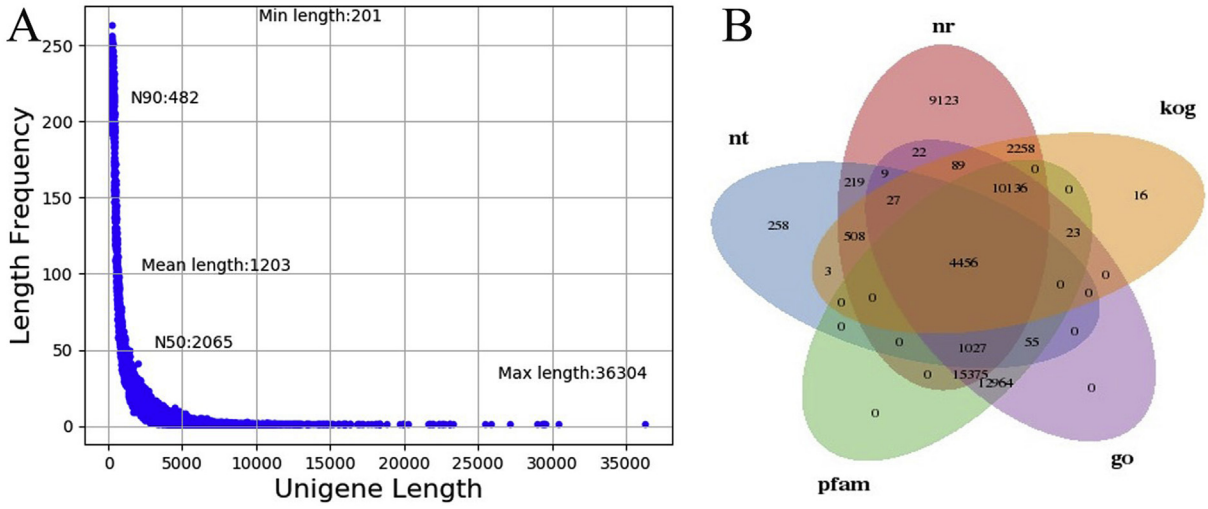

Fig. 1. Functional annotations of UniGenes. (A) Length distribution of Unigenes. (B) Venn Diagram: summary of annotation results.

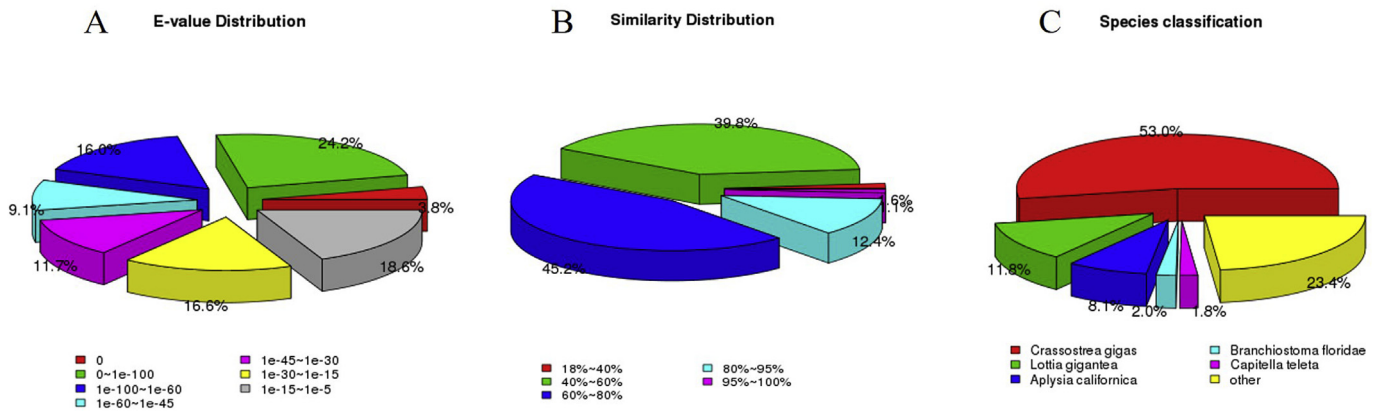

Fig. 2. Comparison of *S. strictus* transcriptomic sequences with the known sequences in database.

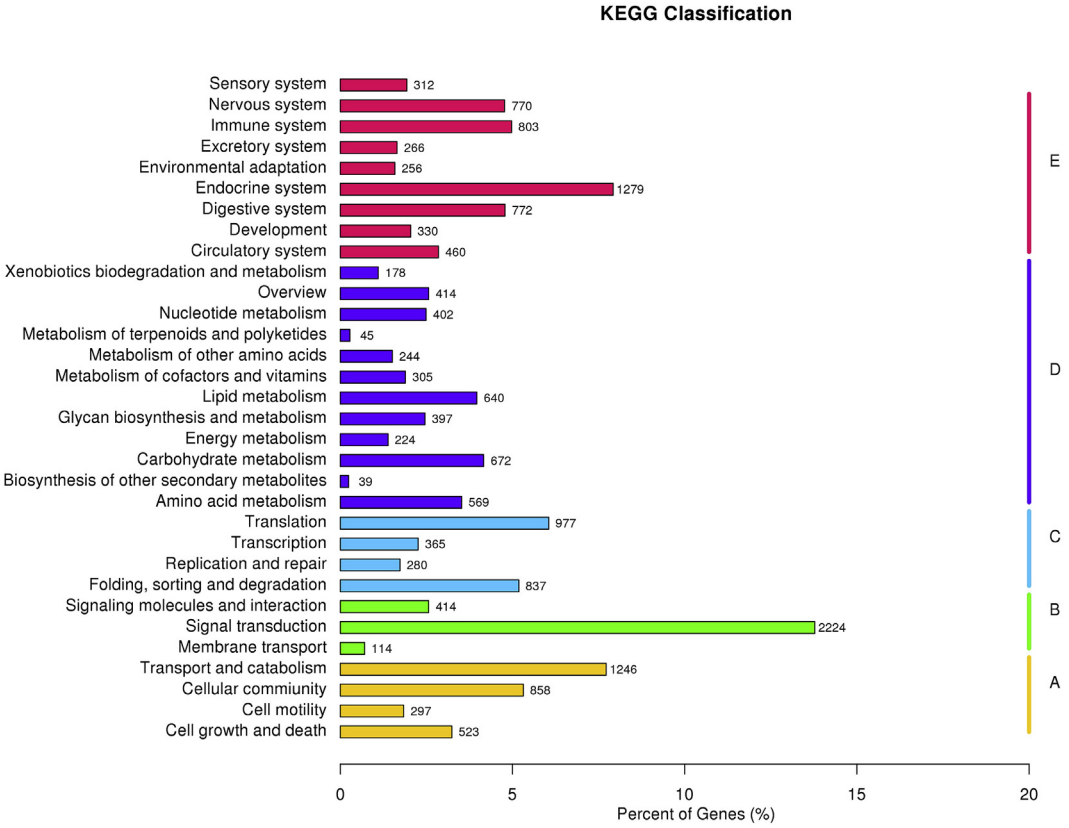

Fig. 3. KEGG annotation of assembled unigenes.

categories (Fig. S2, Table S3). Among these, the categories of signal transduction mechanisms (T, 3205, 18.30%) contained the greatest number of unigenes, followed by general function prediction only (R, 2653, 15.15%) and posttranslational modification, protein turnover and chaperones (O, 1855, 10.59%). The three categories unnamed protein, cell motility and nuclear structure contained the lowest number of unigenes, accounting 4, 56, and 70 unigenes, respectively. A number of unigenes (V, 232, 1.32%) were assigned to cluster of defense mechanisms, implying these unigenes might be involved in immune defense of *S. strictus*.

KEGG pathway analysis was also performed for all unigenes as a bioinformatic base for the systematic analysis of gene functions. In the present study, a total of 16,143 unigenes were grouped into 230 known pathways and the largest group is Rap1 signaling pathway, containing 425 unigenes, followed by Ribosome (423), Endocytosis (402), Focal adhesion (390), PI3K-Akt signaling pathway (366) and Ras signaling pathway (359) (Table S4). The 230 pathways were divided into five categories: cellular processes, environmental information processing, genetic information processing, metabolism and organismal systems (Fig. 3). Within organismal systems, the two most abundant subcategories were endocrine system (1279, 7.92%) and immune system (803, 4.97%). In the metabolism, carbohydrate metabolism (672, 4.16%), lipid metabolism (640, 3.96%) and amino acid metabolism (569, 3.52%) were the most common categories. The majority of Environmental Information Processing was Signal transduction (2224, 13.78%).

3.4. Identification and analysis of differentially expressed genes (DEGs)

Total of 1588 significant DEGs (|log2 fold change|) > 1, adjusted q-value < 0.005 were identified between the *V. anguillarum* challenged and PBS control group with DEG-seq analysis, 999 unigenes were up-

Table 2
Summary of 25 significant immune-related signaling pathways.

Pathway	DEGs number
Antigen processing and presentation	6
NOD-like receptor signaling pathway	10
Toll-like receptor signaling pathway	11
Chemokine signaling pathway	8
Natural killer cell mediated cytotoxicity	3
Hematopoietic cell lineage	1
B cell receptor signaling pathway	6
T cell receptor signaling pathway	6
RIG-I-like receptor signaling pathway	12
Fc epsilon RI signaling pathway	3
Cytosolic DNA-sensing pathway	7
Leukocyte transendothelial migration	4
Fc gamma R-mediated phagocytosis	4
MAPK signaling pathway - fly	9
MAPK signaling pathway	11
Platelet activation	2
PI3K-Akt signaling pathway	8
Ras signaling pathway	8
TNF signaling pathway	12
Jak-STAT signaling pathway	2
Apoptosis	17
PPAR signaling pathway	3
ErbB signaling pathway	2
Leukocyte transendothelial migration	4
ECM-receptor interaction	1

regulated and 589 uningenes were down-regulated. A Venn diagram of gene expression displaying the number of shared and exclusively expressed genes between the two groups was also mapped in the present study There were 112,443 and 106,826 expressed genes detected from the *V. anguillarum* challenged group and PBS control group, respectively. Among them, 87,879 were shared, 24,564 were exclusive to the *V. anguillarum* groups, and 18,947 were present only in the PBS control group (Fig. S1). A heat map was created to visualize the quantitative differences in the expression levels of the DEGs between the *V. anguillarum* challenged group and PBS group (Fig. S4).

3.5. Analysis of DEGs of immune-related genes

In total, 26 immune system-related KEGG pathways were detected among 225 pathways mapped to the KEGG database, and a number of DEGs in the 25 pathways were showed in Table 2. The top 20 enriched KEGG pathways ($q \leq 0.05$) were showed in Fig. 4. Among them, the toll like receptor signaling pathway, TNF signaling pathway, RIG-I-like receptor signaling pathway, Apoptosis and NOD-like receptor signaling pathway were related to the immune system. Among them, we identified 23 immune related DEGs and classified them into four main groups, including complement and coagulation cascades, PRRs, cytokines and regulators, and adaptors and signal transducers (Table 3). Among these immune related DEGs, 20 were up-regulated to varying degrees after *V. anguillarum*. It was suggested that these pathways played crucial roles in the pathogenic mechanism of *S. strictus* during infection of *V. anguillarum*. These pathways and genes that are related to the immune system, signaling transduction and disease processes are similar to previous studies in *Crassostrea gigas* [22], *Meretrix petechialis* [23] and *Sinonovacula constricta* [24]. Compared with analyzing single gene, the

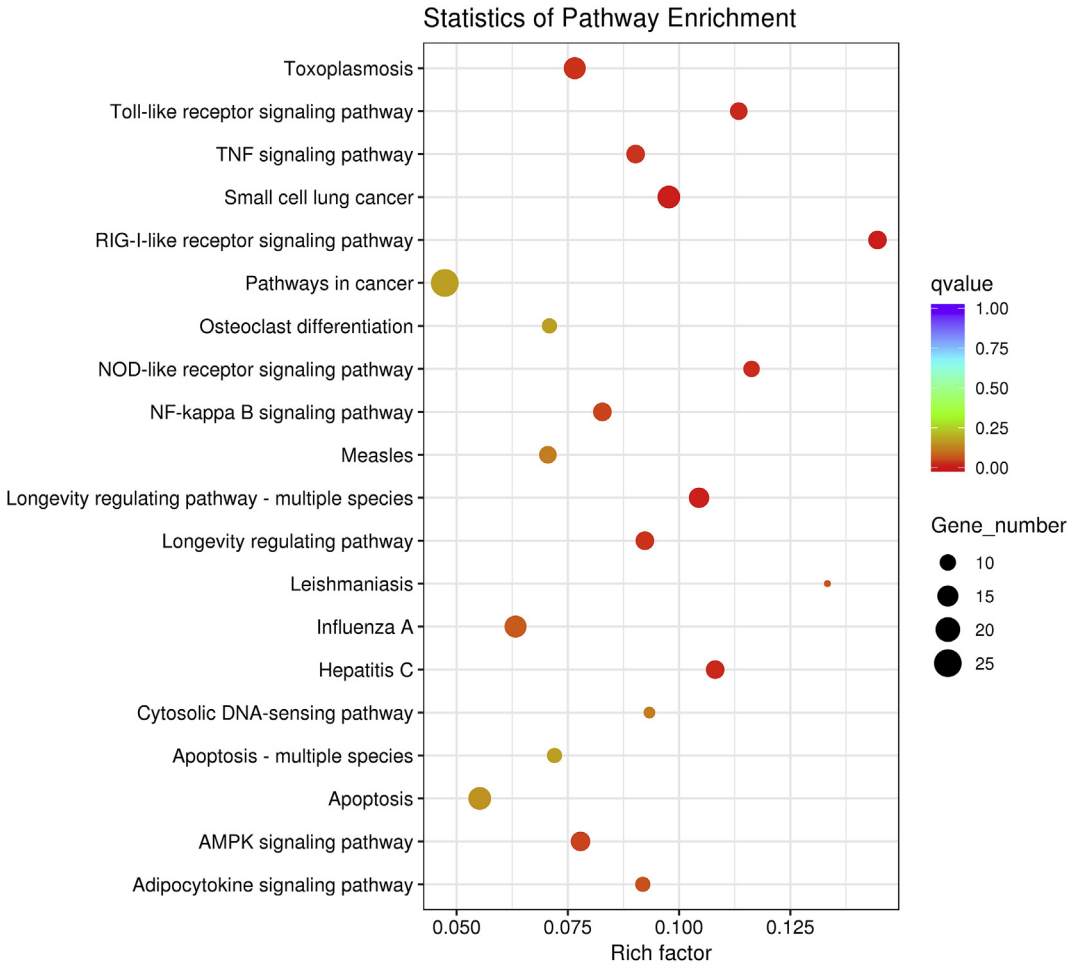

Fig. 4. The top 20 pathways with the largest numbers of differentially expressed genes based on KEGG classifications.

Table 3
DEGs typically related to the immune responses.

Category and Gene name	Log2Ratio (Treatment/Control)	Diff	FDR
Complement and coagulation cascades			
coagulation factor II	−1.4209	Down	0.000491
complement component 1	−1.4421	Down	0.000164
Pattern recognition receptors			
toll-like receptor 4	1.2123	Up	2.45E-07
Scavenger receptor cysteine-rich domain	3.3157	Up	8.42E-14
Scavenger receptor cysteine-rich domain	2.0068	Up	8.35E-08
Scavenger receptor cysteine-rich domain	7.4207	Up	2.36E-05
Scavenger receptor class F member 1	3.808	Up	8.52E-05
Cytokines and regulators			
Interleukin 5	2.8184	Up	7.54E-06
Interleukin-4 receptor	1.5047	Up	2.95E-05
Interleukin-1	1.5594	Up	0.00027
Interferon-induced 6–16 family	1.2109	Up	1.80E-08
stimulator of interferon protein 1	1.453	Up	2.75E-08
Interferon-induced transmembrane protein	2.0656	Up	7.43E-08
Interferon-regulatory factor 3	2.997	Up	3.44E-07
Interferon-induced protein 44-like protein	1.2239	Up	3.36E-05
Interferon-inducible GTPase	2.4138	Up	5.27E-05
Tumor necrosis factor	1.6098	Up	0.000251
Adaptors and signal transducers			
TNF receptor-associated factor 2	1.1063	Up	0.000369
TNF receptor-associated factor 3	3.0617	Up	1.65E-15
TNF receptor-associated factor 6	1.2559	Up	1.32E-05
tumor necrosis factor receptor superfamily member 16-like	2.8367	Up	4.03E-05
Tumor necrosis factor receptor superfamily member 6B	1.8163	Up	0.000147
Tumor necrosis factor ligand superfamily member 13	−1.117	Down	0.000263

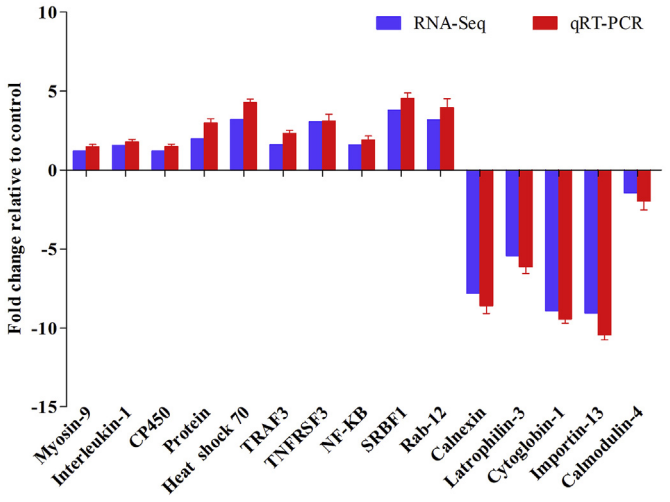

Fig. 5. Comparison of the fold change expression of 15 putative DEGs as determined by RNA-seq and qPCR. The fold change in gene expression was normalized to β-actin gene and relative to the control group samples. Blue bars represent gene expression data from RNA-Seq analysis. Red bars represent gene expression data obtained by qRT-PCR. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

immune-related pathways and genes aid in systematically comprehending the molecular immune mechanisms underlying the *V. anguillarum* response.

Table 4
Statistics of SNP type in transcriptome library.

Type	Count	Frequency per kb
Transition		
C/T	276146	1.5
A/G	275107	1.49
Transversion		
A/T	231835	1.26
A/C	104926	0.57
T/G	105156	0.57
C/G	38351	0.21
Total	1031521	5.6
SNP position in codon		
First	115901	
Second	34454	
Third	120390	

Mollusks have established complete innate immunity that consisted of humoral and cellular components to defense against outside invaders [25]. As the first reactor to initiate the immune response, pattern recognition receptors (PRRs) are the sensory receptors of organisms to resist invading pathogens [26]. Some PRRs in mollusks have been identified, including Toll-like receptors (TLRs), C-type lectins, galectins, lipopolysaccharide-B-1,3-glucan binding protein (LGBP), and peptidoglycan recognition protein (PGRP) [27–29], etc. As well known that innate immune system is the first line of defense against pathogenic microorganism infections in mollusk [30,31], TLRs, as the key components of the innate immune system in mollusk, can recognize the presence of pathogens and activate the immune responses. In this study, a total of 11 DEGs transcripts are involved in the TLR signaling pathway, such as TLR1 gene, TNF receptor-associated factor 3, NFκB and IkB, suggesting that certain specific genes in the toll-like receptor signaling pathway of *S. strictus* play a pivotal role in the response to *V. anguillarum*.

3.6. Validation of 15 DEGs by qPCR

Following RNA-seq transcriptomic profiling of pooled data, expression of fifteen putative DEGs was re-assessed by qPCR using individual biological replicates (n = 3 per treatment) and normalized to *β*-actin gene. The results confirm the differential expression of all 15 genes which showed a similar trend in expression pattern to RNA-Seq pooled library results in all cases, further supporting the reliability of the genes identified by RNA-seq to be differentially expressed (Fig. 5).

3.7. Identification of SSRs and SNPs

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 [32,33]. In this study, a total of 56,079 SSRs from 153,038 unigenes were identified, and 10,620 sequences containing more than one potential SSR were identified using MISA software (Table S5, Fig. S5). Among the 56,079 SSRs, the dinucleotides repeats numbered 10,147 (18.09%), trinucleotide numbered (3663, 6.53%), tetranucleotide numbered (734, 1.31%), pentanucleotide numbered (17, 0.003%) and hexanucleotide repeat motifs (3). Of the dinucleotide SSRs, AC/GT was most common and accounted for 40.81% and CG (0.5% for the present study). The most common trinucleotide repeat was AAC/GTT and ATC/ATG, which accounted for 30.8% and 25.59%, respectively. The most common tetranucleotide repeat was AAAT/ATTT which accounted for 28.07%.

As the most abundant type of DNA sequence polymorphism, SNPs have been commonly used as molecular markers in quantitative trait loci (QTL) mapping, linkage map construction, and association studies [34,35]. A total of 1,031,521 candidate SNP were identified from the *S. strictus* transcriptome libraries. Among these SNPs, 551,253 transition

and 480,268 transversion types were identified in the transcriptome libraries, and SNPs were typically located in the first (115,901) and third (120,390) codon positions (Table 4). The potential SNPs and SSRs identified in this transcriptome dataset can provide important information for further genetic conservation and diversity markers in *S. strictus*.

4. Conclusion

In the present study, we constructed and sequenced the transcriptomic profile from the gill of *Solen strictus* in response to *V. anguillarum* injection. Illumina paired-end RNA sequencing technology was used to sequence. In total, 337,533,440 clean reads were obtained by sequencing with 50.63G clean bases, which then assembled into 195,774 transcripts and 150,038 unigenes. The unigenes were annotated by seven databases, NR, NT, Pfam, KOG, Swiss-prot, KEGG, and GO, and we obtained comprehensive gene function information. Through GO and KEGG enrichment, large numbers of DEGs and pathways involved in immunity were identified for the first time. We identified 1588 significantly DEGs including 999 significantly up-regulated and 589 down-regulated DEGs. In addition, a total of 56,079 SSRs were detected and a large amount of candidate SNP were identified in the transcriptome of *S. strictus*, which is helpful for subsequent marker development, genetic linkage and quantitative trait loci (QTL) analysis. Overall, this study provides valuable information on the antiviral mechanism in *S. strictus*.

Acknowledgments

This work was supported by the Weifang science and technology development project (2018GX054). We would like to thank Shandong Weifang Longwei Industrial CO.LTD for providing *S. strictus*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fsi.2019.01.024.

References

[1] L. Song, L. Wang, H. Zhang, M. Wang, The immune system and its modulation mechanism in scallop, Fish Shellfish Immunol. 46 (1) (2015) 65–78.
[2] L. Wang, X. Song, L. Song, The oyster immunity, Dev. Comp. Immunol. 80 (2018) 99–118.
[3] K.O. Mutz, A. Heilkenbrinker, M. Lonne, J.-G. Walter, F. Stahl, Transcriptome analysis using next-generation sequencing, Curr. Opin. Biotechnol. 24 (1) (2013) 22–30.
[4] N.M. Dheilly, C. Adema, D.A. Raftos, B. Gourbal, C. Grunau, L.D. Pasquier, No more non-model species: the promise of next generation sequencing for comparative immunology, Dev. Comp. Immunol. 45 (2014) 56–66.
[5] Y. Ren, J. Xue, H. Yang, B. Pan, W. Bu, Transcriptome analysis of *Ruditapes philippinarum*, hepatopancreas provides insights into immune signaling pathways under *Vibrio anguillarum* infection, Fish Shellfish Immunol. 64 (2017) 14–23.
[6] C. Détrée, E. Lópezlandavery, C. Gallardoescárate, F.L. Lafargade, Transcriptome mining of immune-related genes in the muricid snail *Concholepas concholepas*, Fish Shellfish Immunol. 71 (2017) 69–75.
[7] 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.
[8] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data, Nat. Biotechnol. 29 (7) (2011) 644.
[9] C. Trapnell, B.A. Williams, G.A. Pertea, Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (5) (2010) 511–515.
[10] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, K.L. Madden, Blast +: architecture and applications, BMC Bioinf. 10 (1) (2009) 421.
[11] A. Conesa, S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 2 (2005) 3674–3676.
[12] Y. Jia, F. Lin, H. Zheng, Z. Yong, C. Jie, Z. Zhang, J. Wang, S. Li, R. Li, L. Bolund, J. Wang, WEGO: a web tool for plotting go annotations, Nucleic Acids Res. 34 (Web Server issue) (2006) 293–W297.
[13] B. Li, C. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinf. 12 (2011) 323.
[14] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[15] L. Wang, Z. Feng, X. Wang, X. Zhang, DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, Bioinformatics 26 (2010) 136–138.
[16] M.D. Young, M.J. Wakefield, G.K. Smyth, A. Oshlack, Gene ontology analysis for RNA-seq: accounting for selection bias, Genome Biol. 11 (2) (2010) R14.
[17] X. Mao, T. Cai, J.G. Olyarchuk, L. Wei, Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary, Bioinformatics 21 (2005) 3787–3793.
[18] G.A. Auwera, M.O. Carneiro, C. Hartl, R. Poplin, G.D. Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K.V. Garimella, D. Altshuler, S. Gabriel, M.A. DePristo, From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline, Curr. Protoc. Bioinf. 43 (1) (2013) 11-10.
[19] M. Havas, B.O. Rosseland, Response of zooplankton, benthos, and fish to acidification: an overview, Water Air Soil Pollut. 85 (1) (1995) 51–62.
[20] L. Austbø, I.B. Aas, M. König, S.C. Weli, M. Syed, K. Falk, E.O. Koppang, Transcriptional response of immune genes in gills and the interbranchial lymphoid tissue of Atlantic salmon challenged with infectious salmon anaemia virus, Dev. Comp. Immunol. 45 (1) (2014) 107–114.
[21] W. Jiang, F. Lin, J. Fang, Y. Gao, M. Du, J. Fang, W. Li, Z. Jiang, Transcriptome analysis of the Yesso scallop, *Patinopecten yessoensis* gills in response to water temperature fluctuations, Fish Shellfish Immunol. 80 (2018) 133–140.
[22] Z. Liu, L. Wang, Z. Zhou, Y. Liu, M. Dong, W. Wang, X. Song, M. Wang, Q. Gao, L. Song, Transcriptomic analysis of oyster *Crassostrea gigas* larvae illustrates the response patterns regulated by catecholaminergic system upon acute heat and bacterial stress, Dev. Comp. Immunol. 73 (2017) 52–60.
[23] F. Jiang, Y. Xin, H. Wang, B. Liu, Transcriptome profiles of the clam *Meretrix petechialis*, hepatopancreas in response to Vibrio, infection, Fish Shellfish Immunol. 62 (2017) 175–183.
[24] 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.
[25] C.J. Bayne, Phagocytosis and non-self recognition in invertebrates: phagocytosis appears to be an ancient line of defense, Bioscience 40 (10) (1990) 723–731.
[26] O. Takeuchi, S. Akira, Pattern recognition receptors and inflammation, Cell 140 (6) (2010) 805–820.
[27] L.J. De, R. Zenagui, R.D. Rosa, D. Piquemal, E. Bachère, Whole transcriptome profiling of successful immune response to vibrio infections in the oyster Crassostrea gigas by digital gene expression analysis, PLoS One 6 (8) (2011) e23142.
[28] J. Yang, L. Qiu, L. Wang, X. Wei, H. Zhang, Y. Zhang, L. Liu, L. Song, CfLGBP, a pattern recognition receptor in *Chlamys farreri* involved in the immune response against various bacteria, Fish Shellfish Immunol. 29 (5) (2010) 825–831.
[29] Y. Zhang, Z. Yu, The first evidence of positive selection in peptidoglycan recognition protein (PGRP) genes of *Crassostrea gigas*, Fish Shellfish Immunol. 34 (5) (2013) 1352–1355.
[30] L. Song, L. Wang, H. Zhang, M. Wang, The immune system and its modulation mechanism in scallop, Fish Shellfish Immunol. 46 (1) (2015) 65.
[31] L. Wang, X. Song, L. Song, The oyster immunity, Dev. Comp. Immunol. 80 (2018) 99–118.
[32] 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 Genomics 14 (2013) 176.
[33] 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.
[34] 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.
[35] S. Hubert, B. Higgins, T. Borza, S. Bowman, Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*), BMC Genomics 11 (2010) 191.