

RESEARCH ARTICLE

# 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

Chuping Xie<sup>1,2</sup>, Yaping Chen<sup>3</sup>, Wanwei Sun<sup>1,2</sup>, Jun Ding<sup>1,2</sup>, Lizhen Zhou<sup>1,2</sup>, Shasha Wang<sup>1,2</sup>, Shuqi Wang<sup>1,2</sup>, Yueling Zhang<sup>1,2</sup>, Dashi Zhu<sup>1,2</sup>, Xiaobo Wen<sup>1</sup>, Songnian Hu<sup>3</sup>, Shengkang Li<sup>1,2\*</sup>

1. Guangdong provincial key laboratory of marine biology, Shantou University, Shantou, China, 2. Marine Biology Institute, Shantou University, Shantou, China, 3. The CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

\*[lisk@stu.edu.cn](mailto:lisk@stu.edu.cn)



CrossMark  
click for updates

## OPEN ACCESS

**Citation:** Xie C, Chen Y, Sun W, Ding J, Zhou L, et al. (2014) 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): e114500. doi:10.1371/journal.pone.0114500

**Editor:** Kristin Michel, Kansas State University, United States of America

**Received:** June 4, 2014

**Accepted:** November 7, 2014

**Published:** December 8, 2014

**Copyright:** © 2014 Xie 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:** The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

**Funding:** The work was supported by grants from National Natural Science Foundation of China (41376176, 41176103), Guangdong provincial project of Science and Technology (2012B020308007). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

## Abstract

**Background:** Mud crab *Scylla paramamosain* is an economically important marine species in China. However, frequent outbreaks of infectious diseases caused by marine bacteria, such as *Vibrio parahaemolyticus*, result in great economic losses. **Methodology/Principal Findings:** Comparative *de novo* transcriptome analysis of *S. paramamosain* infected with *V. parahaemolyticus* was carried out to investigate the molecular mechanisms underlying the immune response to pathogenic bacteria by using the Illumina paired-end sequencing platform. A total of 52,934,042 clean reads from the hemocytes of *V. parahaemolyticus*-infected mud crabs and controls were obtained and assembled into 186,193 contigs. 59,120 unigenes were identified from 81,709 consensus sequences of mud crabs and 48,934 unigenes were matched proteins in the Nr or Swissprot databases. Among these, 10,566 unigenes belong to 3 categories of Gene Ontology, 25,349 to 30 categories of KEGG, and 15,191 to 25 categories of COG database, covering almost all functional categories. By using the Solexa/Illumina's DGE platform, 1213 differentially expressed genes ( $P < 0.05$ ), including 538 significantly up-regulated and 675 down-regulated, were detected in *V. parahaemolyticus*-infected crabs as compared to that in the controls. Transcript levels of randomly-chosen genes were further measured by quantitative real-time PCR to confirm the expression profiles. Many differentially expressed genes are involved in various immune processes,

including stimulation of the Toll pathway, Immune Deficiency (IMD) pathway, Ras-regulated endocytosis, and proPO-activating system.

**Conclusions/Significance:** Analysis of the expression profile of crabs under infection provides invaluable new data for biological research in *S. paramamosain*, such as the identification of novel genes in the hemocytes during *V. parahaemolyticus* infection. These results will facilitate our comprehensive understanding of the mechanisms involved in the immune response to bacterial infection and will be helpful for diseases prevention in crab aquaculture.

---

## Introduction

*Scylla paramamosain* (Crustacea: Decapods: Brachyura), an economically important mariculture commonly known as mud crab, is widely distributed in China and several parts of the world. However, recent rapid development of mud crab farming industry has led to increasingly severe outbreaks of infectious disease [1]. For example, the “milky disease” of mud crab, caused by pathogenic infection of bacteria *V. parahaemolyticus* in the last few years in Southern China, had resulted in more than 60% of mortality and thus brought in great economic losses [2–3].

Although much has been reported on the investigations of the innate immunity [4] and the characterization of immune-related genes in mud crab [5–9], little is known about the global molecular mechanisms underlying the immune response to such a pathogenic bacterial infection in this species. Due to a lack of genetic and genomic information, comprehensive gene expression analyses in the immune system have not been performed in mud crab. Cellular function of genes can be evaluated by the transcriptome analysis, which is a powerful method for assessing the relative importance of genes in cell, tissue, or organism under various conditions [10]. During the last few years, several methods, including ESTs sequencing [11–13], cDNA-AFLP [14–15], suppression subtractive hybridization (SSH) [16–17], and microarrays [18–19], have been applied to the study of transcriptome in maricultures. However, cDNA-AFLP and SSH have incomplete coverage [20], while EST sequencing techniques have limitations in the depth of transcriptome that can be sampled [21]. Microarrays are limited by background and cross hybridization problems. Further, microarrays measure the relative abundance of transcripts, and only predefined sequences can be detected [22].

When genome sequence is unavailable, transcriptome sequencing can be applied to decode the genomes of non-model organisms, providing valuable information to the understanding of gene function, cell responses and evolution [23–25]. High-throughput sequencing of RNA (RNA-Seq) has stimulated an unprecedented increase in transcriptome data [26]. The next-generation sequencing platforms, such as the Solexa/Illumina Genome Analyzer, ABI/SOLiD

Gene Sequencer, and Roche/454 Pyrosequencer, can sequence in parallel massive amounts of DNA molecules derived directly from mRNA and produce millions or even billions of high-quality short reads [27]. The latest paired-end sequencing of RNA-Seq techniques have further improved the efficiency and expanded short read lengths, permitting a better understanding of the transcriptome data [28]. Recently, by using this technology, significant progress has been made in understanding the transcripts of various maricultures [29–32]. These results contribute greatly to the in-depth study of candidate genes in specific biological processes and provide rich sources for identification of novel genes. However, despite the economical importance of *S. paramamosain*, immunological investigation of the transcriptome has not been reported.

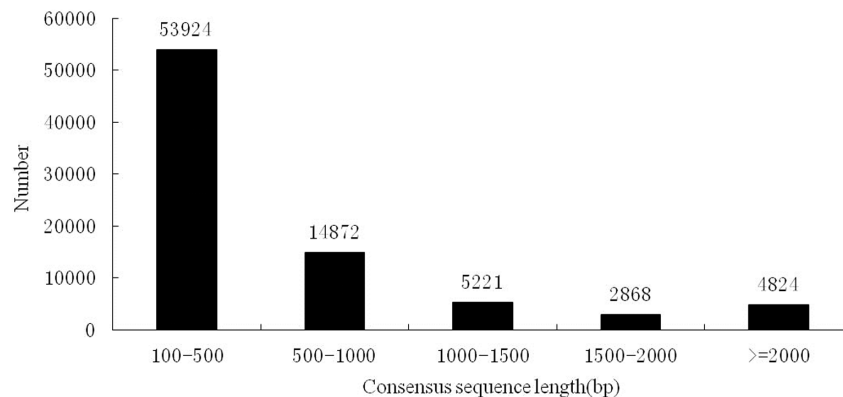
In this study, a transcriptome library of the hemocytes isolated from *V. parahaemolyticus*-infected crabs was constructed for massive parallel mRNA sequencing. Using the *de novo* transcriptome assembly software, we ultimately obtained a transcriptome database containing 81,709 unigenes. Quantitative gene expression analysis was performed using the Solexa/Illumina's DGE technology, which identified 538 significantly up-regulated and 675 down-regulated genes from pathogen-infected crabs. A large number of the differentially expressed genes were immune-related and were further highlighted in signalling cascades.

## Results

### Transcriptome profile of mud crab *S. paramamosain*

Total RNA was extracted from the hemocytes of crabs at several time points, including 0, 12, 24, and 48 h after *V. parahaemolyticus* infection. Equal quantities of RNA were mixed together to construct a cDNA library for Illumina sequencing, which yielded a total of 62,004,374 reads. The reads containing more than five consecutive bases with a quality <13 were removed. The overall Illumina clean reads and bases are 52,934,042 and 4.76G, respectively. Files containing these data were deposited in the Sequence Read Archive of the National Center for Biotechnology Information (NCBI) with the accession number of SRP040563. The high-quality clean reads were assembled into 186,193 contigs by using the SOAP *de novo* software, which were further assembled into 81,709 consensus sequences with a mean length of 656 bp (range: 205~23,885 bp). The length statistics of all consensus sequences is presented in [Fig. 1](#).

59,120 unigenes were identified from 81,709 consensus sequences of mud crabs. After searching the Nr and Swissprot databases, 48,934 unigenes (31,482 in Nr and 27,498 in Swissprot) were identified. Among these, 10,566 unigenes belong to 3 categories of Gene Ontology, 25,349 to 30 categories of KEGG, and 15,191 to 25 categories of COG database, covering almost all functional categories ([Table 1](#)). The remaining 10,186 unigenes failed to match proteins in the Nr database or Swissprot databases and therefore represented potentially novel genes.



**Figure 1. Length statistic of all consensus sequences obtained from the mud crab *S.paramamosain* transcriptome library.**

doi:10.1371/journal.pone.0114500.g001

### GO, COG and KEGG Classification

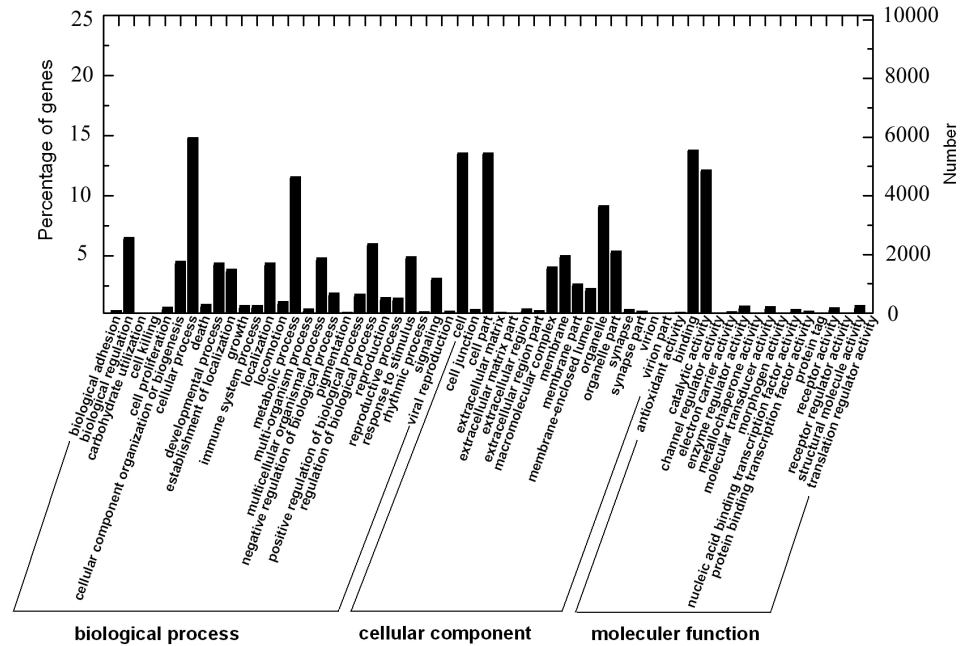
Gene ontology (GO) analysis was performed using the web-based Database for Annotation, Visualization, and Integrated Discovery (DAVID) [32–33]. Among the 59,120 assembled unigenes, 10,566 of them were successfully annotated by GO assignments to one or more of the three categories: biological process, cellular component and molecular function (Fig. 2), which fell further into 27, 17 and 17 subcategories with the largest ones being the “primary metabolic process” (59.1%), “cell” (53.9%) and “binding” (55.0%), respectively. In summary, these terms account for a large fraction of the overall assignments in the *S. paramamosain* transcriptome data.

Phylogenetic classifications of the predicted CDSs of unigenes were carried out by searching against the COG database (Fig. 3). A quantity of 15,191 unigenes were classified and subdivided into 25 COG categories (23 of them had known functions, while the other two were “general function prediction only” and “function unknown”), among which “General function prediction only”

**Table 1. Annotation of non-redundant unigenes.**

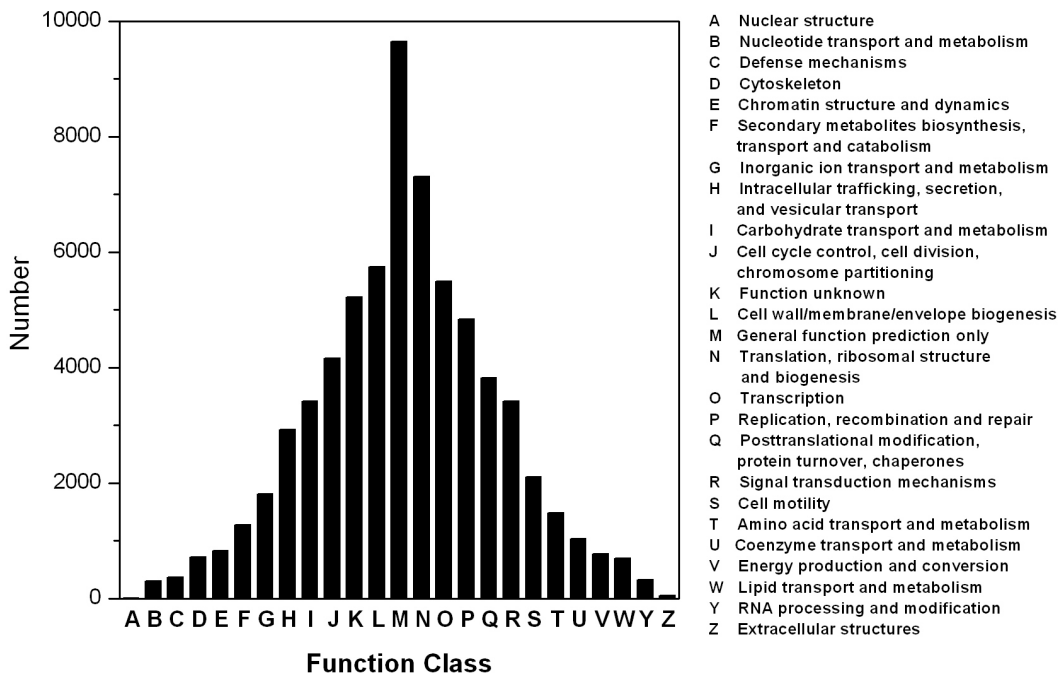
Database	Number of annotated consensus sequences	Percentage of annotated consensus sequences
<b>Nr</b>	31,482	38.53%
<b>Swissprot</b>	27,498	33.65%
<b>Nt</b>	37,234	45.57%
<b>KEGG</b>	25,349	31.02%
<b>GO</b>	10,139	12.41%
<b>COG</b>	15,191	18.59%
<b>ALL</b>	59,120	

doi:10.1371/journal.pone.0114500.t001



**Figure 2. GO annotation of the unigenes in the *V. parahaemolyticus* infected crabs transcriptome. Most unigenes can be divided into three major categories, including biological process, cellular component, and molecular function.**

doi:10.1371/journal.pone.0114500.g002



**Figure 3. Histogram presentation of clusters of orthologous groups (COG) classification in *S. paramamosain*. All putative proteins were aligned to the COG database and could be classified into at least 25 molecular families.**

doi:10.1371/journal.pone.0114500.g003

represented the largest group (9650 matched unigenes, 13.52%), followed by “Translation, ribosomal structure and biogenesis” (2,279, 14.08%) and “Cell wall/membrane/envelope biogenesis” (1,483, 9.16%).

A total of 25,349 unigenes in mud crab transcriptome were mapped to 30 statistically remarkable categories in KEGG ( $P < 0.05$ ) (S1 Table). Metabolic pathways (2701, 10.66%) and regulation of actin cytoskeleton (1610, 6.35%) were two of the most statistically significant categories. Notably, three immune related categories, mitogen-activated protein kinase (MAPK) signalling pathway (668, 2.64%), ubiquitin mediated proteolysis (594, 2.34%), and chemokine signalling pathway (761, 3.00%), were also identified as statistically significant.

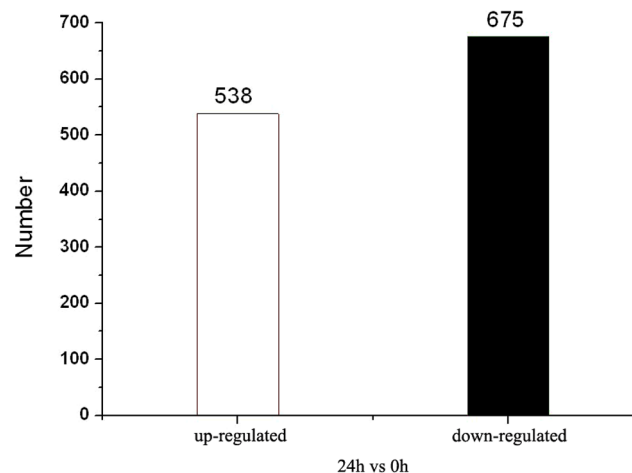
### Global changes in gene expression upon *V.parahaemolyticus* infection

To characterize the immune response of mud crab to bacterial infection, two DGE libraries were constructed using mRNA from the haemocytes of crabs infected with *V.parahaemolyticus* and the controls, respectively. After removal of low-quality tags, adaptor tags, and one copy number tag, a total of 7,562,126 and 7,468,808 clean tags were obtained from the libraries of bacterium infected and control crabs, respectively (S2 Table), which then were mapped to the transcriptome database described above. A total of 1213 differentially expressed genes ( $P < 0.05$ ) were found, including 538 up-regulated and 675 down-regulated genes, in the haemocytes of crabs infected with *V.parahaemolyticus* (Fig. 4). Files containing these differentially expressed genes upon *V.parahaemolyticus* infection were deposited in the Gene Expression Omnibus (GEO) of NCBI with the accession number of GSE45400.

To achieve a functional annotation of the infection-responsive genes, GO classifications were assigned to the differentially expressed genes by using DAVID, indicating that the up- and down-regulated genes were involved in immunity, transcription, translation regulations, and biological regulation (S3 Table). Among the 1213 differentially expressed genes, 961 (79.23%) genes were well annotated, whereas the remaining 252 (20.77%) genes had low sequence homology to known sequences in public databases, suggesting that they might be putative novel genes in *S.paramamosain* involved in immune response to *V.parahaemolyticus* infection.

### Candidate genes involved in mud crab immune response

Of these 1213 differentially expressed genes, 96 of them were known to be involved in immunity and immune related signal transduction in mud crabs after *V. parahaemolyticus* infection (S4 Table). Most of these immune-related genes were enriched in GO terms “Protein metabolic” and “immune system related”, which contained up to 40 and 20 genes, respectively. The rest three categories were “cytokine and chemokine” (18 genes), “Transporter and regulator activity” (10 genes), and “Cell proliferation and apoptosis” (8 genes).



**Figure 4. The differentially expressed genes upon *V. parahaemolyticus* infection.** By using Solexa/Illumina's DGE platform, 1213 differentially expressed genes were detected in comparative analysis of the expression profiles between *V. parahaemolyticus*-infected crabs and control crabs, including 538 up-regulated genes and 675 down-regulated genes ( $P < 0.05$ ).

doi:10.1371/journal.pone.0114500.g004

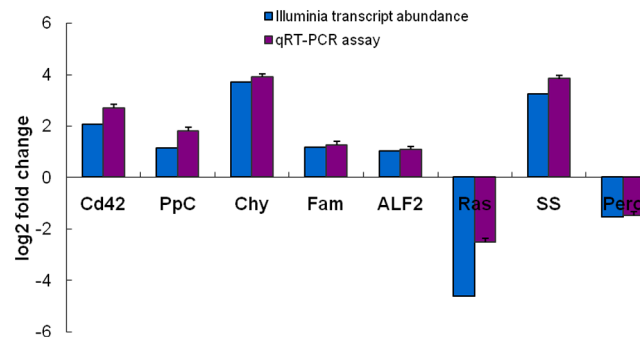
Genes involved in the Toll pathway and the Immune deficiency (IMD) pathway, such as Toll-like receptor (TLR), interferon (IFN) and tumor necrosis factor receptor 6 (TNFR6), were found significantly induced in mud crab upon *V. parahaemolyticus* infection. Antimicrobial proteins, such as Crustin, ALF2, and ALF6, were also significantly up-regulated, indicating they may play important roles in crab's defence against *V. parahaemolyticus* infection.

Genes annotated as Ras superfamily, including Ras, Rab and myosin, were also activated by *V. parahaemolyticus* infection. Many genes in the transcription regulation group and genes encoding cell proliferation and apoptosis, including Tob1 and Syntaxin, as well as genes involved in antioxidant activity, such as AST, CAT, CathepsinA and CathepsinD, were found up-regulated in crabs infected with *V. parahaemolyticus*.

Of these differentially expressed genes, a large number of immune related genes belong to the proPO-activating system, in which the most significantly differentially expressed genes were serine proteinases, clip domain serine proteinase, serine protease-like protein, chymotrypsin, and their inhibitors, such as serpin and pacifastin.

### Verification of transcriptome data by real-time RT-PCR

Relative quantitative real-time RT-PCR (qRT-PCR) analysis was performed on randomly chosen genes (*Chy*, *Fam*, *ALF2*, *Ras*, *Cd42*, *PpC*, *SS*, *Pero*) to confirm the expression profiles of mud crabs in response to *V. parahaemolyticus* infection (Fig. 5). The results showed that qRT-PCR analyse of these genes was in good agreement with the Illumina RNA-seq analyse, indicating the transcriptome data



**Figure 5. Comparison of the expression profiles of randomly selected genes as determined by Illuminia sequencing and qRT-PCR analysis.** Target gene abbreviations are as follows: Cd42, cell division cycle protein 42; PpC, phospholipid phospholipase C; Chy, chymotrypsin; Fam, farnesoic acid O-methyltransferase; ALF2, antilipoplysaccharide factor 2; Ras, Ras GTPase-activating protein; SS, son of sevenless; Pero, peroxinectin

doi:10.1371/journal.pone.0114500.g005

can reflect an actual gene expression profile in *V. parahaemolyticus* infected mud crabs.

## Discussion

*S. paramamosain* has received increasing attention over the years because of its wonderful flavor, and whose aquaculture production has reached 110,000 tons in China in 2011 [34]. *V. parahaemolyticus* infection is a main bacterial cause of crab mortality in cultured *S. paramamosain* in southern China during the last few years [2]. Although there are many reports on isolating and characterising immune-related genes in mud crabs, molecular studies on the immune response to pathogens in mud crab are still rare. To increase our knowledge of host responses to bacterial infection, we used Illuminia RNA-Seq to investigate the transcriptome profile of crab after *V. parahaemolyticus* infection, showing that a large proportion of unigenes from *S. paramamosain* were new transcripts compared to known genes in public databases by bioinformatics analysis. Comparative analysis of transcriptome changes between *V. parahaemolyticus*-infected and non-infected crabs revealed a large number of differentially expressed genes. Our sequencing data analyses indicate that *V. parahaemolyticus* infection has a significant impact on the transcriptional profile of *S. paramamosain* hemocytes. Previously, transcriptome data obtained from the 454 pyrosequencing platform have identified growth performance-related genes in mud crabs [34]. In another study, using the 454 GS FLX sequencing system, the authors identified 4,021 gonad-differentially, 10,522 ovary-specifically, and 19,013 testis-specifically expressed genes after comparing libraries of *S. paramamosain* [35]. Comparing our Illuminia RNA-Seq data with those from the 454 pyrosequencing platform, we found that the two sequencing platforms produced equivalent numbers of unigenes, however, a larger number of immune-related genes were identified in



our study, while more growth-related and gonad-specific genes could be retrieved by the 454 pyrosequencing platform.

Among the differentially expressed genes found in this study, several, such as C-type lectin and catalase, have been previously reported to be involved in the host response against white spot syndrome virus (WSSV) and taura syndrome virus (TSV) [36–37]. In invertebrates, C-type lectins are involved in the non-self immune recognition and pathogen phagocytosis through opsonization [38]. It is indicated that catalase activity or increase in gene expression parallels immune-stimulant challenge or pathogen infections in maricultures [39–40]. In this study, we found several unigenes homologous to C-type lectins and 2 unigenes annotated as catalase, whose expression exhibited significant changes after *V. parahaemolyticus* infection. Similarly, heat shock protein,  $\alpha$ 2-macroglobulin, peroxinectin and alcohol dehydrogenase, have been reported to be strongly up-regulated in invertebrates when challenged with stressful conditions [7, 41–45]. In the present study, up-regulation of these genes has also been observed in crabs challenged with *V. parahaemolyticus*. Meanwhile, other significantly changed genes in mud crab during *V. parahaemolyticus* infection (such as calmodulin and integrin) have not been previously reported of roles in immune systems in aquatic animals. The immunological functions of these genes need further investigations.

The well-documented signaling pathways involved in the innate immune response of invertebrates against pathogens include the Toll pathway and immune deficiency (IMD) pathway, which can induce the expression of antimicrobial peptide (AMP) genes and regulate the host humoral response [30, 45–47]. The differentially expressed genes identified in crab's transcriptome data included Toll-like receptor (TLR), interferon (IFN) and tumor necrosis factor receptor 6 (TNFR6). Similar observations have been made in previous reports [30]. Further, many downstream effectors of Toll-IMD pathways, including *ALF2*, *ALF6*, crustins and other antimicrobial related peptides (S4 Table), were identified from the transcriptome data of crabs upon *V. parahaemolyticus* infection.

The Ras superfamily, a group of small GTPases exhibiting high-affinity binding for GDP and GTP, can be divided into several families based on their structures and functions [48]. By regulating actin and myosin recruitment, these GTPases are essential for phagocytosis [49–50]. Some subtypes of GTPases, such as Ras, Rab and myosin, were significantly induced (S4 Table), indicating that Ras-regulated phagocytosis was participated in the process of *V. parahaemolyticus* infection in mud crabs. Cell apoptosis plays significant roles in the control of immune response, the deletion of immune cells recognizing self-antigens, and cytotoxic killing [51]. Of these differentially expressed genes in mud crabs, genes involved in cell apoptosis/proliferation, including *Tab1*, *Cathepsin D* and *Syntaxin*, were activated by *V. parahaemolyticus* infection in mud crabs.

The proPO-activating system is composed of genes such as serine proteinases and their inhibitors (serpins), prophenoloxidase-activating enzyme (PPA), proPO and its active form, phenoloxidase (PO) [52]. In the present study, a number of differentially expressed genes found during *V. parahaemolyticus* infection in mud crabs were annotated to be tentative members of the proPO-activating system (S4

Table). These genes were mainly kinds of serine proteinases, including clip domain serine proteinase, serine protease-like protein, chymotrypsin, and their inhibitors, such as serpin and pacifastin. Hemocyanin is another well-known immune-related protein previously reported to possess multiple bioactive functions [53]. Hemocyanins have the defense-related functions that are mediated through phenoloxidase activity. Peroxinectin is present as a biologically inactive protein in the hemocytes in invertebrates and becomes converted into its active form, a cell adhesion molecule, when the proPO system is activated [52]. Hemocyanins, cryptocyanins and peroxinectin were found significantly induced upon *V. parahaemolyticus* infection in mud crabs. These data provided evidences that many genes in the proPO-activating cascade were stimulated by the pathogenic bacterium, indicating that the expression and regulation of different types of proPOs constituted a response to different kinds of pathogens in mud crabs. Last but not least, most of differentially expressed genes identified in the transcriptome of mud crabs in our study were annotated as unknown or hypothetical proteins, which would need more investigations in the future. Further identification of these genes and their functions might provide new insights into the immune response to *V. parahaemolyticus* infection.

In conclusion, the present study focused on the difference of the transcriptome of *V. parahaemolyticus*-infected and non-infected crabs, aiming at discovery of underlying mechanisms involved in host defense against pathogenic bacterium infection. Comparative transcriptome analysis of crabs of *V. parahaemolyticus*-infected and control groups showed significant differences in gene expression. Many immune-related genes as well as Toll and IMD pathway, Ras-regulated endocytosis, and proPO-activating genes, were affected by the infection and probably participated in antimicrobial process. In addition, this study provided a detailed data for identification of novel genes in the hemocytes in mud crab *S. paramamosain*, especially under the situation that the whole genome sequence of crab is not currently available.

## Methods

### Ethics statements

Mud crabs used in this study were taken under permission from a local crab farm (Niutianyang, Shantou, Guangdong, China) (23.3543°N. latitude and 116.6458°E. longitude). Crabs were processed according to “the Regulations for the Administration of Affairs Concerning Experimental Animals” established by the Guangdong Provincial Department of Science and Technology on the Use and Care of Animals.

### *S. paramamosain* sampling and the challenge experiment

Samples collection and the pathogen challenge had been described previously [54]. Briefly, forty healthy mud crabs (100 ± 10 g in weight) were collected from

Niutianyang area in Shantou, China. Crabs were acclimated for 3 days in 1 m<sup>3</sup> tanks (ten crabs per tank) in conditions (salinity: 8‰; temperature: 28°C) similar to those of culture ponds from which crabs were obtained. Sea water was changed twice a day. Crabs were then fed with shellfishes for one week before sampling. *V. parahaemolyticus* isolated from diseased crabs (from Shantou crab farms) were cultured in 2216E medium at 28°C. 30 crabs were inoculated individually in the ventral hind legs with 100 µL of *V. parahaemolyticus* ( $2.4 \times 10^7$  cells/mL), while the other 10 crabs were injected with 100 µL of saline (0.9%) as the control group. 10 crabs were harvested for haemocyte sampling at 12 h, 24 h, 48 h after the challenge. Haemocytes were collected from blood samples after mixing with anti-coagulant solution and centrifugation at 800 g at 4°C for 20 min. Haemocytes were immediately frozen in liquid nitrogen and then stored at -80°C prior to RNA extraction.

### RNA isolation

Total RNA was extracted from 50 to 100 mg of tissue with TRIZOL Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The RNA samples were incubated for 30 min at 37°C with 10 units of DNase I (Takara, Dalian, China) to remove residual genomic DNA. The quality and quantity of the purified RNA were determined by measuring the absorbance at 260 nm/280 nm (A260/A280) using a ND-1000 spectrophotometer (LabTech, Holliston, MA) and integrity was ensured through analysis on a 1.5% (w/v) agarose gel.

### Library preparation and sequencing

cDNA libraries were constructed using RNA Sample Preparation Kit (Illumina, San Diego, CA) following manufacturer's instructions. Briefly, mRNA was purified from total RNA using polyATtract mRNA isolation systems (Promega, Madison, WI). Fragmentation was carried out using divalent cations under elevated temperature in Illumina proprietary fragmentation buffer. First strand cDNA was synthesized using random oligonucleotides and SuperScript II (Promega, Madison, WI). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H (Promega, Madison, WI). Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. In order to select cDNA fragments of preferentially 200 bp in length, the library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly). Products were quantified using the Agilent high sensitivity DNA assay on the Agilent Bioanalyzer 2100 system. Finally, the library preparations were sequenced on an Illumina HiSeq 2000 platform which generates 100 bp paired-end reads.

### Assembly of transcripts and annotation

Raw sequencing reads were quality trimmed, and adaptor sequences were removed before the assembly. After removal of low quality reads, processed reads

were assembled using SOAP *de novo* software [55] with default parameters. The overall assembly was performed using the combined sequence data for both treated crabs and the control groups. The contigs and singletons were generally referred to as unigenes. As a result, 81,709 unigenes were generated. Subsequently, the unigenes were subjected to BLASTX similarity search against NCBI non-redundant protein database and the swissprot database using BLASTALL program with an E-value threshold of  $10^{-5}$ . All annotated unigenes were used to determine the COG term, GO term and KEGG pathway with a cut-off E-value of  $10^{-5}$  using BLASTX [29]. The assembled contigs were used as a reference for annotating the DGE tags.

### Identification of differentially expressed genes

Gene expression was measured by counting tags from normal and pathogen-infected crabs and normalized to the total high-quality reads. High-throughput sequencing was performed using the Solexa/Illumina Genome Analyzer. The raw data was filtered to remove low quality sequences, including ambiguous nucleotides and adaptor sequences. The number of reads mapping to each unigene in the two populations analyzed (infected and non-infected) was counted and used as an approximate estimation of gene expression level in the corresponding individuals. For unigenes with multiple transcripts, the number of reads were summed and assumed to be the corresponding final read count. To link the expressed signatures to known genes from *crab*, the unigene dataset ([http://trace.ncbi.nlm.nih.gov/Traces/sra\\_sub/sub.cgi?login=pda](http://trace.ncbi.nlm.nih.gov/Traces/sra_sub/sub.cgi?login=pda)) from the transcriptome of *S. paramamosain* was used as a reference. The transcription level of each gene was deduced by determining the total number of reads mapped to each gene using Picard tools (<http://picard.sourceforge.net/>). To provide a relative assessment of transcript-abundance, the numbers of raw reads that were mapped to individual contigs were normalized for sequence length (i.e., Fragments Per Kilobase of transcript per Million mapped reads, FPKM). To investigate differences in gene expression profiles, FDR (false discovery rate)  $<0.001$  was used as the threshold of *P*-value in multiple test to judge the significance of gene expression difference [56]. Genes were considered differentially expressed in a given library when the *P*-value  $<0.001$  and a greater than two-fold change (absolute value of  $\log_2$  ratio  $>1$ ) in expression across libraries was observed.

### Quantitative real-time PCR

Quantitative real-time PCR was performed using the ABI Prism 7300 Detection System (Applied Biosystems, Foster City, CA) with SYBR Green I as the fluorescent dye according to the manufacturer's protocol (Takara, Dalian, China). First-strand cDNA was synthesized from 2  $\mu\text{g}$  of total RNA as described above and used as a template for real-time PCR with specific primers (S5 Table). Real-time PCR was performed in a total volume of 20  $\mu\text{l}$ , and cycling conditions were 95°C for 5 min, followed by 40 cycles of 94°C for 5 s, 60°C for 31 s. All reactions

were performed in biological triplicates, and the results were expressed relative to the expression levels of  $\beta$ -actin in each sample by using the  $2^{-\Delta\Delta CT}$  method [57]. Each sample was first normalized for the amount of template added by comparison with the abundance of  $\beta$ -actin mRNA [5, 58].

### Statistical analysis

One way ANOVA analysis was used to analyze the means differences among samples. Pairwise t-test was then applied to compare gene expression levels between two crab populations. All statistical analyses (avo functions and pairwise t-test functions in R) were conducted in the R computation environment (<http://www.r-project.org>) [59]. *P*-values smaller than 0.05 were considered statistically significant.

### Supporting Information

#### S1 Table.

[doi:10.1371/journal.pone.0114500.s001](https://doi.org/10.1371/journal.pone.0114500.s001) (XLS)

#### S2 Table.

[doi:10.1371/journal.pone.0114500.s002](https://doi.org/10.1371/journal.pone.0114500.s002) (DOC)

#### S3 Table.

[doi:10.1371/journal.pone.0114500.s003](https://doi.org/10.1371/journal.pone.0114500.s003) (XLS)

#### S4 Table.

[doi:10.1371/journal.pone.0114500.s004](https://doi.org/10.1371/journal.pone.0114500.s004) (DOC)

#### S5 Table.

[doi:10.1371/journal.pone.0114500.s005](https://doi.org/10.1371/journal.pone.0114500.s005) (XLS)

### Acknowledgments

We thank Prof. Chiju Wei of Shantou University for critical discussion and proof-reading of the manuscript.

### Author Contributions

Conceived and designed the experiments: SKL CPX WWS. Performed the experiments: CPX. Analyzed the data: YPC JD LZZ SSW SQW YLZ SNH SKL DSZ. Contributed reagents/materials/analysis tools: YPC WWS. Wrote the paper: SKL XBW CPX.

### References

1. Ou J, Meng Q, Li Y, Xiu Y, Du J, et al. (2012) Identification and comparative analysis of the *Eriocheir sinensis* microRNA transcriptome response to *Spiroplasma eriocheiris* infection using a deep sequencing approach. *Fish Shellfish Immunol* 32: 345–352.

2. Li Y, Xia X, Wu Q, Liu W, Lin Y (2008) Infection with *Hematodinium* sp. in mud crabs *Scylla serrata* cultured in low salinity water in southern China. *Dis Aquat Organ* 82: 145–150.
3. Li S, Sun L, Wu H, Hu Z, Liu W, et al. (2012) The intestinal microbial diversity in mud crab (*Scylla paramamosain*) as determined by PCR-DGGE and clone library analysis. *J Appl Microbiol* 113: 1341–1351.
4. Mangkalan S, Sanguanrat P, Utairangsri T, Sritunyalucksana K, Krittanai C (2014) Characterization of the circulating hemocytes in mud crab (*Scylla olivacea*) revealed phenoloxidase activity. *Dev Comp Immunol* 44: 116–123.
5. Li S, Zhang Z, Li C, Liu W, Zhang Y, et al. (2012) Molecular cloning and expression profiles of nitric oxide synthase (NOS) in mud crab *Scylla paramamosain*. *Fish Shellfish Immunol* 32: 503–512.
6. Zhu L, Lan JF, Huang YQ, Zhang C, Zhou JF, et al. (2014) SpALF4: a newly identified anti-lipopolysaccharide factor from the mud crab *Scylla paramamosain* with broad spectrum antimicrobial activity. *Fish Shellfish Immunol* 36(1): 172–180.
7. Du ZQ, Ren Q, Huang AM, Fang WH, Zhou JF, et al. (2013) A novel peroxinectin involved in antiviral and antibacterial immunity of mud crab, *Scylla paramamosain*. *Mol Biol Rep* 40: 6873–6881.
8. Zhang QX, Liu HP, Chen RY, Shen KL, Wang KJ (2013) Identification of a serine proteinase homolog (Sp-SPH) involved in immune defense in the mud crab *Scylla paramamosain*. *PLoS One* 8: e63787.
9. Li XC, Zhu L, Li LG, Ren Q, Huang YQ, et al. (2013) A novel myeloid differentiation factor 88 homolog, SpMyD88, exhibiting SpToll-binding activity in the mud crab *Scylla paramamosain*. *Dev Comp Immunol* 39: 313–322.
10. Mu Y, Ding F, Cui P, Ao J, Hu S, et al. (2013) Transcriptome and expression profiling analysis revealed changes of multiple signaling pathways involved in immunity in the large yellow croaker during *Aeromonas hydrophila* infection. *BMC Genomics* 11: 506.
11. Cao D, Kocabas A, Ju Z, Karsi A, Li P, et al. (2001) Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles of the head kidney. *Anim Genet* 32: 169–188.
12. Martin SA, Caplice NC, Davey GC, Powell R (2002) EST-based identification of genes expressed in the liver of adult Atlantic salmon (*Salmo salar*). *Biochem Biophys Res Commun* 293: 578–585.
13. Zou Z, Zhang Z, Wang Y, Han K, Fu M, et al. (2011) EST analysis on the gonad development related organs and microarray screen for differentially expressed genes in mature ovary and testis of *Scylla paramamosain*. *Comp Biochem Physiol Part D* 6: 150–157.
14. Taris N, Lang RP, Reno PW, Camara MD (2009) Transcriptome response of the Pacific oyster (*Crassostrea gigas*) to infection with *Vibrio tubiashii* using cDNA-AFLP differential display. *Anim Genet* 40: 663–677.
15. Taboada X, Robledo D, Del Palacio L, Rodeiro A, Felip A, et al. (2012) Comparative expression analysis in mature gonads, liver and brain of turbot by cDNA-AFLPs. *Gene* 492: 250–261.
16. Gao W, Tan B, Mai K, Chi S, Liu H, et al. (2012) Profiling of differentially expressed genes in hepatopancreas of white shrimp (*Litopenaeus vannamei*) exposed to long-term low salinity stress. *Aquaculture* 364–365: 186–191.
17. He N, Liu H, Xu X (2004) Identification of genes involved in the response of haemocytes of *Penaeus japonicus* by suppression subtractive hybridization (SSH) following microbial challenge. *Fish Shellfish Immunol* 17: 121–128.
18. Bayne CJ, Gerwick L, Wheeler PA, Thorgaard GH (2006) Transcriptome profiles of livers and kidneys from three rainbow trout (*Oncorhynchus mykiss*) clonal lines distinguish stocks from three allopatric populations. *Comparative Biochemistry and Physiology* 1: 396–403.
19. Boswell MG, Wells MC, Kirk LM, Ju Z, Zhang Z, et al. (2009) Comparison of gene expression responses to hypoxia in viviparous (*Xiphophorus*) and oviparous (*Oryzias*) fishes using a medaka microarray. *Comp Biochem Physiol C Toxicol Pharmacol* 149: 258–265.
20. Vuylsteke M, Daele HVD, Vercauteren A, Zabeau M, Kuiper M (2006) Genetic dissection of transcriptional regulation by cDNA-AFLP. *Plant J* 45: 439–446.
21. Hanriot L, Keime C, Gay N, Faure C, Dossat C, et al. (2008) A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. *BMC Genomics* 9: 418

22. **t Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, et al.** (2008) Deep sequencing based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* 36: e141.
23. **Ramayo-Caldas Y, Mach N, Esteve-Codina A, Corominas J, Castello A, et al.** (2012) Liver transcriptome profile in pigs with extreme phenotypes of intramuscular fatty acid composition. *BMC Genomics* 13: 547.
24. **Sanchez CC, Weber GM, Gao G, Cleveland BM, Yao J, et al.** (2011) Generation of a reference transcriptome for evaluating rainbow trout responses to various stressors. *BMC Genomics* 12: 626.
25. **Wang XW, Luan JB, Li JM, Su YL, Xia J, et al.** (2011) Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. *BMC Genomics* 12: 458.
26. **Han X, Wu X, Chung WY, Li T, Nekrutenko A, et al.** (2009) Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proc Natl Acad Sci USA* 106: 12741–12746.
27. **Morozova O, Hirst M, Marra MA** (2009) Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* 10: 135–151.
28. **Fullwood MJ, Wei C-L, Liu ET, Ruan Y** (2009) Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome research* 19: 521–532.
29. **Zeng D, Chen X, Xie D, Zhao Y, Yang C, et al.** (2013) Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura Syndrome Virus (TSV) experimental infection. *PLoS ONE* 8: e57515.
30. **Li S, Zhang X, Sun Z, Li F, Xiang J** (2013) Transcriptome analysis on Chinese shrimp (*Fenneropenaeus chinensis*) during WSSV acute infection. *PLoS ONE* 8: e58627.
31. **He L, Jiang H, Cao D, Liu L, Hu S, et al.** (2013) Comparative transcriptome analysis of the accessory sex gland and testis from the Chinese mitten crab (*Eriocheir sinensis*). *PLoS ONE* 8: e53915.
32. **Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al.** (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotech* 29: 644–652.
33. **Pertea G, Huang X, Liang F, Antonescu V, Sultana R, et al.** (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19: 651–652.
34. **Ma H, Ma C, Li S, Jiang W, Li X, et al.** (2014) Transcriptome Analysis of the Mud Crab (*Scylla paramamosain*) by 454 Deep Sequencing: Assembly, Annotation, and Marker Discovery. *PLoS ONE* 9(7): e102668.
35. **Gao J, Wang X, Zou Z, Jia X, Wang Y, Zhang Z** (2014) Transcriptome analysis of the differences in gene expression between testis and ovary in green mud crab (*Scylla paramamosain*). *BMC Genomics* 15: 585.
36. **Gross PS, Bartlett TC, Browdy CL, Chapman RW, Warr GW** (2001) Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, *Litopenaeus vannamei*, and the Atlantic White Shrimp, *L. setiferus*. *Dev Comp Immunol* 25: 565–577.
37. **Dhar AK, Dettori A, Roux MM, Klimpel KR, Read B** (2003) Identification of differentially expressed genes in shrimp (*Penaeus stylirostris*) infected with White spot syndrome virus by cDNA microarrays. *Arch Virol* 148: 2381–2396.
38. **Ma TH, Tiu SH, He JG, Chan SM** (2007) Molecular cloning of a C-type lectin (LvLT) from the shrimp *Litopenaeus vannamei*: early gene down-regulation after WSSV infection. *Fish Shellfish Immunol* 23: 430–437.
39. **Liu HP, Chen FY, Gopalakrishnan S, Qiao K, Bo J, Wang KJ** (2010) Antioxidant enzymes from the crab *Scylla paramamosain*: gene cloning and gene/protein expression profiles against LPS challenge. *Fish Shellfish Immunol* 28, 862–871.
40. **Liu J, Pan LQ, Zhang L, Miao JJ, Wang J** (2009) Immune responses, ROS generation and the haemocyte damage of scallop *Chlamys farreri* exposed to Aroclor 1254. *Fish Shellfish Immunol* 26, 422–428.
41. **Gaston JSH** (2002) Heat shock proteins and innate immunity. *Clin Exp Immunol* 127: 1–3.
42. **Armstrong PB, Quigley JP** (1999)  $\alpha$ 2-macroglobulin: an evolutionarily conserved arm of the innate immune system. *Dev Comp Immunol* 23: 375–390.

43. **Laudanna C, Bolomini-Vittori M** (2009) Integrin activation in the immune system. *Wiley Interdiscip Rev Syst Biol Med.* 1: 116–127.
44. **Schlenke TA, McKean KA** (2005) A role for alcohol dehydrogenase in the *Drosophila* immune response? *Insect Mol Biol.* 14: 175–178.
45. **Tanji T, Ip YT** (2005) Regulators of the Toll and Imd pathways in the *Drosophila* innate immune response. *Trends Immunol* 26(4): 193–198.
46. **Naitza S, Ligoxygakis P** (2004) Antimicrobial defences in *Drosophila*: the story so far. *Mol Immunol* 40(120): 887–896.
47. **Leclerc V, Reichhart JM** (2004) The immune response of *Drosophila melanogaster*. *Immunol Rev* 198: 59–71.
48. **Wennerberg K, Rossman KL, Der CJ** (2005) The Ras superfamily at a glance. *J Cell Sci* 118(5): 843–846.
49. **Hoppe AD, Swanson JA** (2004) Cdc42, Rac1, and Rac2 display distinct patterns of activation during phagocytosis. *Mol Biol Cell* 15(8): 3509–3519.
50. **Kim JG, Moon MY, Kim HJ, Li Y, Song DK, et al.** (2012) Ras-related GTPases Rap1 and RhoA collectively induce the phagocytosis of serum-opsonized zymosan particles in macrophages. *J Biol Chem* 287(7): 5145–5155.
51. **Ekert PG, Vaux DL** (1997) Apoptosis and the immune system. *Br Med Bull* 53: 591–603.
52. **Cerenius L, Soderhall K** (2004) The prophenoloxidase-activating system in invertebrates. *Immunol Rev* 198: 116–126.
53. **Yan F, Zhang Y, Jiang R, Zhong M, Hu Z, et al.** (2011) Identification and agglutination properties of hemocyanin from the mud crab (*Scylla serrata*). *Fish Shellfish Immunol.* 30: 354–360.
54. **Li S, Zhu S, Li C, Zhang Z, Zhou L, et al.** (2013) Characterization of microRNAs in mud crab *Scylla paramamosain* under *Vibrio parahaemolyticus* infection. *PLoS ONE* 8(8): e73392.
55. **Luo R, Liu B, Xie Y, Li Z, Huang W, et al.** (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1: 18.
56. **Storey J** (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society* 64: 19.
57. **Livak KJ, Schmittgen TD** (2001) Analysis of relative gene expression data using real time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25: 402–408.
58. **Cao J, Lin Z, Zhang Y, Zhang X, Li S, et al.** (2013) Cloning and characterization of the SpLRR cDNA from green mud crab, *Scylla paramamosain*. *Fish Shellfish Immunol* 34: 129–135.
59. **Zhang Q, Liu L, Zhu F, Ning Z, Hincke MT, et al.** (2014) Integrating *De Novo* transcriptome assembly and cloning to obtain chicken ovocleidin-17 full length cDNA. *PLoS ONE* 9(3): e93452.