Transcriptome Analysis and Discovery of Genes Involved in Immune Pathways from Hepatopancreas of Microbial Challenged Mitten Crab *Eriocheir sinensis*

Xihong Li¹,², Zhaoxia Cui¹,³, Yuan Liu¹, Chengwen Song¹,², Guohui Shi¹,²

¹ Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, ² Graduate University of the Chinese Academy of Sciences, Beijing, China, ³ National & Local Joint Engineering Laboratory for Ecological Mariculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

Abstract

**Background:** The Chinese mitten crab *Eriocheir sinensis* is an important economic crustacean and has been seriously attacked by various diseases, which requires more and more information for immune relevant genes on genome background. Recently, high-throughput RNA sequencing (RNA-seq) technology provides a powerful and efficient method for transcript analysis and immune gene discovery.

**Methods/Principal Findings:** A cDNA library from hepatopancreas of *E. sinensis* challenged by a mixture of three pathogen strains (Gram-positive bacteria *Micrococcus luteus*, Gram-negative bacteria *Vibrio alginolyticus* and fungi *Pichia pastoris*; 10⁶ cfu·mL⁻¹) was constructed and randomly sequenced using Illumina technique. Totally 39.76 million clean reads were assembled to 70,300 unigenes. After ruling out short-length and low-quality sequences, 52,074 non-redundant unigenes were compared to public databases for homology searching and 17,617 of them showed high similarity to sequences in NCBI non-redundant protein (Nr) database. For function classification and pathway assignment, 18,734 (36.00%) unigenes were categorized to three Gene Ontology (GO) categories, 12,243 (23.51%) were classified to 25 Clusters of Orthologous Groups (COG), and 8,983 (17.25%) were assigned to six Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Potentially, 24, 14, 47 and 132 unigenes were characterized to be involved in Toll, IMD, JAK-STAT and MAPK pathways, respectively.

**Conclusions/Significance:** This is the first systematical transcriptome analysis of components relating to innate immune pathways in *E. sinensis*. Functional genes and putative pathways identified here will contribute to better understand immune system and prevent various diseases in crab.

Citation: Li X, Cui Z, Liu Y, Song C, Shi G (2013) Transcriptome Analysis and Discovery of Genes Involved in Immune Pathways from Hepatopancreas of Microbial Challenged Mitten Crab *Eriocheir sinensis*. PLoS ONE 8(7): e68233. doi:10.1371/journal.pone.0068233

Editor: Gen Hua Yue, Temasek Life Sciences Laboratory, Singapore

Received January 29, 2013; Accepted June 1, 2013; Published July 17, 2013

Copyright: © 2013 Li et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported by the National Natural Science Foundation of China (41276165) and the Chinese National ‘863’ Project (No. 2012AA10A409). 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.

* E-mail: zhxcui@yahoo.com.cn

Introduction

Chinese mitten crab *Eriocheir sinensis* (Crustacea: Decapoda: Grapsidae, Eriocheir) (Henri Milne Edwards, 1854) is one of the important economic aquaculture species in China. However, with rapid development of large-scale culture, frequent outbreaks of diseases caused by viruses, bacteria and rickettsia-like organisms have led to catastrophic economic losses in cultured *E. sinensis* stocks [1–3]. Characterizing immune molecules and understanding defense mechanism are useful to health management and disease control in crab aquaculture.

Like other invertebrates, *E. sinensis* lacks adaptive immune system and mainly depends on innate immunity. Innate immune system provides a first line for host to defense against invading pathogens. It is composed of cellular responses like phagocytosis and encapsulation, and humoral responses that produce immune-related factors. Immune relevant genes, such as crustin [4], cathepsin L [5], prophenoloxidase (proPO) [6], C-type lectin [7] and anti-lipopolysaccharide factor (ALF) [8], have been separately cloned and characterized from *E. sinensis*. However, knowledge about immune system of *E. sinensis* is still fragmentary and different signaling pathways implicated in immune response also remain incomplete.

To date, genome sequence of any crab species is still unavailable, which limits resources of molecular information. In recent years, high-throughput RNA-sequencing (RNA-Seq), including Solexa/Illumina, Roche/454 and ABI/SOLiD, has offered high-effective technology for analysis of gene expression, discovery of novel transcripts, identification of differentially expressed genes and others [9]. The powerful technology provides a new opportunity for studies of genome reference-free species and non-model organisms. With development of this technology, RNA-Seq has been widely applied in various invertebrates, such as *Eriocheir sinensis* [10,11], *Litopenaeus vannamei* [12], *Bactrocera dorsalis* [13], *Pinctada martensi* [14] and *Crassostrea gigas* [15].
In crustacean, apart from functioning as a digestive gland, hepatopancreas is also an important immune organ that functions as a primary site to synthesize and excrete immune molecules, such as beta-1,3-glucan binding protein (LGBP) [16], antibacterial peptide (AMP) [17], lectin or lectin related proteins and others [18]. Expressed sequence tag (EST) analysis and gene discovery of *L. vannamei* and *L. setiferus* also demonstrated that hepatopancreas played a crucial role in innate immunity and hepatopancreas cDNA library appeared to be more diverse than hemocytes library [18]. Hence, large-scale identification of immune genes from hepatopancreas is of great value and necessity to study immune mechanism in crustacean. Previously, Jiang et al [19] constructed a nonnormalized hepatopancreas cDNA library of *E. sinensis* and characterized immune-associated genes by EST approach. It aided to understand biological function of hepatopancreas and served a basis for in-depth investigation of Chinese mitten crab.

In the present study, by Illumina sequencing and bioinformatics analysis, we analyzed hepatopancreas transcriptome of *E. sinensis* that was infected with a mixture of three pathogen strains (Gram-positive bacteria *Micrococcus luteus*, Gram-negative bacteria *Vibrio alginolyticus* and fungi *Pichia pastoris*; 10⁸ cfu·mL⁻¹). Main objective of this study was to annotate functional genes from this transcriptome analysis and identify potential immune molecules of different signaling pathways, such as Toll, immune deficiency (IMD), janus kinase (JAK)-signal transducers and activators of transcription (STAT) and mitogen-activated protein kinase (MAPK) pathways.

### Materials and Methods

**Ethic Statement**

This study was strictly performed in accordance with the Guide for Care and Use of Laboratory Animals by Chinese Association for Laboratory Animal Sciences (No. 2011-2).

**Mitten crab and microbial challenge**

Healthy mature female mitten crabs were obtained from a commercial farm in Panjin, China and acclimated in oxygenated seawater at 15±1°C for a week before processing. During whole period of the experiment, all crabs were fed with clam meat and the water was changed every day. For immune challenge experiment, we prepared a mixture of three pathogen strains (Gram-positive bacteria *Micrococcus luteus*, Gram-negative bacteria *Vibrio alginolyticus* and fungi *Pichia pastoris*), which were suspended in 0.1 mol/L PBS (pH 7.0) with the final pathogens concentration of 10⁸ cfu·mL⁻¹. The crabs were injected at arthrodial membrane of the last walking leg with 100 µL, the mixture of pathogens and returned into seawater tanks for 8 h. Hepatopancreas of treated crabs were collected and kept in liquid nitrogen for RNA extraction.

**Table 1. Summary of sequences analysis.**

| Description                  | Number   |
|------------------------------|----------|
| Before trimming              |          |
| Raw reads                    | 40.78×10⁶|
| After trimming               |          |
| Clean reads                  | 39.76×10⁶|
| Clean bases (Mb)             | 4.02×10³ |
| Average length of clean reads (bp) | 101.10   |
| GC content (%)               | 46.01    |
| Q20 percentage (%)           | 97.95    |
| After assembly               |          |
| Unigenes                     | 70,300   |
| Min length (bp)              | 201      |
| Max length (bp)              | 16,874   |
| Average length (bp)          | 967      |
| N50 (bp)                     | 1,834    |
| N90 (bp)                     | 354      |

**Figure 1. Length distribution of assembled unigenes.**

**Figure 2. E-value and score distribution of unigenes matched to Nr database.**

(A) E-value distribution of annotated unigenes. (B) Score distribution of annotated unigenes.

That is, 10⁸ cfu·mL⁻¹, the final pathogens concentration of 10⁸ cfu·mL⁻¹. The crabs were injected at arthrodial membrane of the last walking leg with 100 µL, the mixture of pathogens and returned into seawater tanks for 8 h. Hepatopancreas of treated crabs were collected and kept in liquid nitrogen for RNA extraction.

![Figure 1. Length distribution of assembled unigenes.](https://doi.org/10.1371/journal.pone.0068233.g001)

![Figure 2. E-value and score distribution of unigenes matched to Nr database.](https://doi.org/10.1371/journal.pone.0068233.g002)
RNA isolation and cDNA library construction

Total RNA was isolated with Trizol Reagent (Invitrogen), after which the concentration, quality and integrity were determined with a NanoDrop spectrophotometer and an Agilent 2100 Bioanalyzer. Poly-(A)-containing mRNA was purified using oligo(dT) magnetic beads and Oligotex mRNA Kits (Qiagen). The mRNA was fragmented and used as template to synthesize first-stranded cDNA with reverse transcriptase and random hexamer-primers. Second-stranded cDNA was synthesized using RNase H and DNA polymerase I. These double-stranded cDNA fragments underwent process of end repair, addition of a single ‘A’ base and ligation of adapters. Adaptor modified fragments were selected by gel purification and amplified through PCR to create the final cDNA library.

Illumina sequencing, assembly, and annotation

Transcriptome sequencing was carried out on an Illumina HiSeq 2000 platform that generated about 100 bp paired-end (PE) raw reads [Novogene Bioinformatics Technology Co.Ltd]. Raw sequences were deposited to NCBI Short Read Archive (SRA) database (http://www.ncbi.nlm.nih.gov/Traces/sra/). After removing adaptor sequences, ambiguous ‘N’ nucleotides (with the ratio of ‘N’ to be more than 10%) and low quality sequences (with quality score to be less than 5), the remaining clean reads were assembled using Trinity software as described for de novo transcriptome assembly without reference genome [20].

For homology annotation, non-redundant sequences were subjected to public databases including NCBI (http://www.ncbi.nlm.nih.gov/) non-redundant protein (Nr) and non-redundant nucleotide (Nt), Swiss-Prot (http://www.ebi.ac.uk/uniprot/), Gene Ontology (GO) (http://www.geneontology.org/), Clusters of Orthologous Groups (COG) (http://www.ncbi.nlm.nih.gov/COG/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/). If results of different databases were conflicted, a priority order of alignments from Nr, Nt, KEGG, Swiss-Prot, GO and COG databases was followed. Comparing to Nr, Nt and Swiss-Prot databases was carried out using BlastX algorithm with an E-value cut-off of 10^-10. GO terms at 2nd level was used to perform GO annotation. COG and KEGG classification were done using BlastX with an E-value cut-off of 10^-5.

Immune gene identification

Immune genes belonging to different signaling pathways were manually identified according to annotated sequences in above databases. Protein coding sequences (CDSs) were also predicted by Trinity software and multiple sequence alignment was carried out using ClustalX.

Gene expression validation

Genes identified in this transcriptome sequencing analysis were validated and quantified by real-time PCR (RT-PCR). Primers

Table 2. BLAST analysis of non-redundant unigenes against public databases.

| Database  | Number of annotated unigenes | Percentage of annotated unigenes |
|-----------|------------------------------|----------------------------------|
| Nr        | 17,617                       | 33.83%                           |
| Nt        | 5,033                        | 9.67%                            |
| Swiss-Prot| 14,862                       | 28.54%                           |
| KEGG      | 8,983                        | 17.25%                           |
| GO        | 18,734                       | 36.00%                           |
| COG       | 12,243                       | 23.51%                           |

For homology annotation, non-redundant sequences were subjected to public databases including NCBI (http://www.ncbi.nlm.nih.gov/) non-redundant protein (Nr) and non-redundant nucleotide (Nt), Swiss-Prot (http://www.ebi.ac.uk/uniprot/), Gene Ontology (GO) (http://www.geneontology.org/), Clusters of Orthologous Groups (COG) (http://www.ncbi.nlm.nih.gov/COG/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/). If results of different databases were conflicted, a priority order of alignments from Nr, Nt, KEGG, Swiss-Prot, GO and COG databases was followed. Comparing to Nr, Nt and Swiss-Prot databases was carried out using BlastX algorithm with an E-value cut-off of 10^-10. GO terms at 2nd level was used to perform GO annotation. COG and KEGG classification were done using BlastX with an E-value cut-off of 10^-5.

Figure 3. GO categorization of non-redundant unigenes. Each annotated sequence was assigned at least one GO term. doi:10.1371/journal.pone.0068233.g003
(Table S1) were designed according to Illumina sequencing data with Primer Premier 5. Prepared total RNA used in RT-PCR analysis was isolated from the same treated crab hepatopancreas as that in Illumina sequencing. Reversed cDNA was also synthesized using the same method as described in Illumina sequencing preparation.

RT-PCR was performed in an ABI 7300 Real-time Detection System (Applied Biosystems). β-actin of *E. sinensis* was used as an internal control to normalize the expression level and all experiments were performed in triplicate. The reaction was carried out in a total volume of 10 μL, containing 5 μL of 2× SYBR Premix Ex Taq™ II (Takara), 0.2 μL of 50× ROX Reference Dye, 2 μL of diluted cDNA mix, 0.2 μL of each primer (10 mM) and 2.4 μL of Milli-Q water. Thermal profile for SYBR Green RT-PCR was 95°C for 5 min, followed by 40 cycles of 95°C for 5 s and 60°C for 31 s. To confirm that only one PCR product was amplified and detected, dissociation curve analysis of amplification products was performed at the end of each PCR reaction. After the PCR program, data were analyzed with ABI 7300 SDS software (Applied Biosystems). The comparative CT method (2−ΔΔCT method) was used to analyze the expression level of different genes.

**Results**

**Transcriptome sequencing and assembly**

Illumina sequencing data from microbial challenged *E. sinensis* hepatopancreas were deposited to NCBI SRA database under accession number of SRA068878. Approximately 40.78 million Illumina PE raw reads were generated (Table 1). After removing adaptor sequences, ambiguous nucleotides and low-quality sequences, 39.76 million clean reads with an average length of 101.10 bp remained. Assembly of clean reads resulted in 70,300 unigenes that ranged from 201 bp to 16874 bp with a N50 length of 1834 bp (Table 1). Length statistics of assembled unigenes were displayed (Figure 1).

**Blast analysis**

After eliminating repeated and short-length sequences, 52,074 non-redundant unigenes were subjected to public databases for similarity searching. 17,617 (33.83%) and 5,033 (9.67%) non-redundant unigenes (Table 2) showed identity with sequences in NCBI Nr and Nt databases, respectively. E-value and score distribution of best hits in Nr database revealed that 55.45% (9,481) of matched sequences showed high homology with an E-value <1E-50 and 55.79% (9,539) with a score ≥500 (Figure 2).
Our results also showed that 28.54% (14,862) of non-redundant unigenes demonstrated similarity to known genes in Swiss-Prot database (Table 2).

**Functional annotation and pathway assignment**

According to Gene Ontology (GO), an international standardized gene functional classification system, 18,734 non-redundant unigenes were classified into three major functional categories (biological process, cellular component and molecular function) and 46 subcategories (Figure 3). In the category of biological process, dominant subcategories were ‘cellular process’ (10,321, 23.84%) and ‘metabolic process’ (9,268, 21.41%). Of sequences categorized as cellular component, ‘cell’ (9,864, 29.69%) and ‘cell part’ (9,864, 29.69%) were most represented, followed by ‘organelle’ (4,315, 13.08%) and ‘macromolecular complex’ (2,776, 8.36%). Among molecular function terms, they showed a significant proportion of clusters assigned to ‘binding’ (10,562, 42.78%) and ‘catalytic activity’ (8,086, 32.75%). However, within each of the three categories, few genes were assigned to subcategories of ‘growth’, ‘cell junction’ and ‘receptor regulator activity’.

| KEGG category         | KEGG subcategory                           | No. of unigenes |
|-----------------------|--------------------------------------------|----------------|
| Metabolism            | Amino acid metabolism                      | 368            |
|                       | Biosynthesis of other secondary metabolites | 59             |
|                       | Carbohydrate metabolism                    | 513            |
|                       | Energy metabolism                          | 302            |
|                       | Glycan biosynthesis and metabolism         | 247            |
|                       | Lipid metabolism                           | 300            |
|                       | Metabolism of cofactors and vitamins        | 149            |
|                       | Metabolism of other amino acids            | 120            |
|                       | Metabolism of terpenoids and polyketides   | 48             |
|                       | Nucleotide metabolism                      | 381            |
|                       | Xenobiotics biodegradation and metabolism  | 95             |
| Genetic information   | Folding                                    | 629            |
|                       | Replication and repair                      | 523            |
|                       | Transcription                              | 294            |
|                       | Translation                                | 707            |
| Environmental         | Membrane transport                         | 31             |
| information processing| Signal transduction                        | 776            |
|                       | Signaling molecules and interaction         | 96             |
| Cellular processes    | Cell communication                         | 282            |
|                       | Cell growth and death                      | 597            |
|                       | Cell motility                              | 90             |
|                       | Transport and catabolism                   | 466            |
| Organismal systems    | Circulatory system                         | 68             |
|                       | Development                                | 116            |
|                       | Digestive system                           | 264            |
|                       | Endocrine system                           | 339            |
|                       | Environmental adaptation                   | 48             |
|                       | Excretory system                           | 126            |
|                       | Immune system                              | 418            |
|                       | Nervous system                             | 478            |
|                       | Sensory system                             | 53             |
| Human diseases        | Cancers                                    | 782            |
|                       | Cardiovascular diseases                    | 111            |
|                       | Endocrine and metabolic diseases           | 22             |
|                       | Immune diseases                            | 82             |
|                       | Infectious diseases                        | 1,366          |
|                       | Neurodegenerative diseases                 | 443            |
|                       | Substance dependence                       | 149            |
Figure 5. Predicted domain and multiple sequence alignment of MyD88. (A) Putative conserved domain of *E. sinensis* MyD88. (B) Multiple sequence alignment of *E. sinensis MyD88* with homologs from other thirteen organisms. Species and GenBank accession numbers of other *MyD88* sequences were as follow: *Litopenaeus vannamei* (AFP49300.1), *Camponotus floridanus* (EFN62977.1), *Pediculus humanus corporis* (XP_002431564.1), *Solenopsis invicta* (EFZ13196.1), *Acromyrmex echinatior* (EGL65212.1), *Harpegnathos saltator* (EFN82696.1), *Bombus terrestris* (XP_003394201.1), *Aedes aegypti* (XP_001658635.1), *Megachile rotundata* (XP_003705811.1), *Bombus impatiens* (XP_003489479.1), *Daphnia pulex* (EFX88460.1), *Anopheles gambiae* (XP_314167.4), *Culex quinquefasciatus* (XP_001868621.1).

doi:10.1371/journal.pone.0068233.g005

To classify orthologous gene products, 12,243 (23.51%) non-redundant unigenes (Table 2) were subdivided into 25 COG classifications. Among them, the cluster of ‘general function prediction only’ (2,086, 14.92%) represented the largest group, followed by ‘signal transduction mechanisms’ (1,913, 13.68%), ‘post-translational modification, protein turnover, chaperon’ (1,167, 8.35%) and ‘transcription’ (984, 7.04%), whereas ‘cell mobility’ (23, 0.16%) was the smallest group (Figure 4).

Using KEGG, 8983 unigenes (Table 2) were assigned to six classifications. Among them, the cluster of ‘general function prediction only’ represented the largest group, followed by ‘signal transduction mechanisms’, ‘cell growth and death’, ‘nervous system’, and ‘amino acid metabolisms’. Dominant subcategories of other terms were ‘carbohydrate metabolisms’, ‘nucleotide metabolisms’ and ‘organism system, human diseases, genetic information processing’. Twenty-six non-redundant unigenes were identified in metabolism and main metabolism associated factor 6 (TRAF6) appeared to be absent in this study. In putative Toll pathway, microbial components triggered activating of Spaetzle and Toll, which initiated signaling pathway by recruiting MyD88 and other molecules (Tube, Pelle and TRAF6). Then, it induced nuclear translocation of Dorsal/Dif. The key adaptor protein coordinating Toll pathway, MyD88, contained a death domain (DD) and a Toll/interleukin-1 receptor (TIR)

Immune gene and pathway analysis
High-throughput sequencing effort revealed that a large number of molecules were highly enriched in immune processes and signaling pathways. Among them, we focused on key genes involved in Toll, IMD, JAK-STAT and MAPK signaling pathways. Main components of these immune pathways were described as follow.

**Toll pathway.** Twenty-six non-redundant unigenes were identified with identity to main molecules of Toll pathway, including Spaetzle, Toll, myeloid differentiation factor 88 (MyD88), Pelle, Cactus, Dorsal/ Dorsal-related immunity factor (Dif) (Table 4, Figure S1). Tube and tumor necrosis factor receptor-associated factor 6 (TRAF6) appeared to be absent in this study. In putative Toll pathway, microbial components triggered activating of Spaetzle and Toll, which initiated signaling pathway by recruiting MyD88 and other molecules (Tube, Pelle and TRAF6).

Using KEGG, 8983 unigenes (Table 2) were assigned to six classifications. Among them, the cluster of ‘general function prediction only’ represented the largest group, followed by ‘signal transduction mechanisms’, ‘cell growth and death’, ‘nervous system’, and ‘amino acid metabolisms’. Dominant subcategories of other terms were ‘carbohydrate metabolisms’, ‘nucleotide metabolisms’ and ‘organism system, human diseases, genetic information processing’. Twenty-six non-redundant unigenes were identified in metabolism and main metabolism associated factor 6 (TRAF6) appeared to be absent in this study. In putative Toll pathway, microbial components triggered activating of Spaetzle and Toll, which initiated signaling pathway by recruiting MyD88 and other molecules (Tube, Pelle and TRAF6). Then, it induced nuclear translocation of Dorsal/Dif. The key adaptor protein coordinating Toll pathway, MyD88, contained a death domain (DD) and a Toll/interleukin-1 receptor (TIR)

Immune gene and pathway analysis
High-throughput sequencing effort revealed that a large number of molecules were highly enriched in immune processes and signaling pathways. Among them, we focused on key genes involved in Toll, IMD, JAK-STAT and MAPK signaling pathways. Main components of these immune pathways were described as follow.

**Toll pathway.** Twenty-six non-redundant unigenes were identified with identity to main molecules of Toll pathway, including Spaetzle, Toll, myeloid differentiation factor 88 (MyD88), Pelle, Cactus, Dorsal/ Dorsal-related immunity factor (Dif) (Table 4, Figure S1). Tube and tumor necrosis factor receptor-associated factor 6 (TRAF6) appeared to be absent in this study. In putative Toll pathway, microbial components triggered activating of Spaetzle and Toll, which initiated signaling pathway by recruiting MyD88 and other molecules (Tube, Pelle and TRAF6). Then, it induced nuclear translocation of Dorsal/Dif. The key adaptor protein coordinating Toll pathway, MyD88, contained a death domain (DD) and a Toll/interleukin-1 receptor (TIR)

**Table 4. Putative immune genes involved in Toll pathway.**

| Signaling molecular | Unigene   | ID             | E-value  | Description                                      |
|---------------------|-----------|----------------|----------|-------------------------------------------------|
| Spaetzle            | comp20879_c1 | AEL23015.1     | 2.97E-32 | protein spaetzle [Cherax quadricarinatus]        |
|                     | comp20655_c0 | ACD36030.1     | 6.58E-61 | spatzle protein [Fenneropenaeus chinensis]       |
|                     | comp30857_c1 | ACD36030.1     | 6.58E-61 | spatzle protein [Fenneropenaeus chinensis]       |
| Toll                | comp28394_c0 | XP_971999.1    | 0        | PREDICTED: similar to toll [Trichobothrium castaneum] |
|                     | comp31809_c0 | XP_971999.1    | 0        | PREDICTED: similar to toll [Trichobothrium castaneum] |
|                     | comp31880_c0 | XP_971999.1    | 0        | PREDICTED: similar to toll [Trichobothrium castaneum] |
|                     | comp401835_c0 | XP_971999.1   | 0        | PREDICTED: similar to toll [Trichobothrium castaneum] |
|                     | comp6372_c0 | XP_971999.1    | 0        | PREDICTED: similar to toll [Trichobothrium castaneum] |
|                     | comp30733_c0 | XP_003209739.1 | 5.18E-43 | PREDICTED: toll-like receptor 13-like [Meleagris gallopavo] |
|                     | comp15258_c0 | BAF99007.1     | 0        | toll receptor [Marsupenaeus japonicas]          |
|                     | comp18754_c0 | BAF99007.1     | 1.33E-42 | toll receptor [Marsupenaeus japonicas]          |
|                     | comp2035_c0 | BAF99007.1     | 0        | toll receptor [Marsupenaeus japonicas]          |
|                     | comp27278_c0 | BAF99007.1     | 0        | toll receptor [Marsupenaeus japonicas]          |
|                     | comp27278_c1 | BAF99007.1     | 1.33E-42 | toll receptor [Marsupenaeus japonicas]          |
|                     | comp24095_c0 | ABK88278.1     | 3.33E-17 | toll-like receptor [Carcinoscorpius rotundicauda] |
| MyD88              | comp23475_c0 | EFN62977.1     | 1.33E-52 | Myeloid differentiation primary response protein MyD88 [Camponotus floridanus] |
| Pelle              | comp30621_c0 | JN180645.1     | 4.34E-13 | Litopenaeus vannamei pelle mRNA, complete cds [Litopenaeus vannamei] |
|                     | comp29278_c0 | XP_002431275.1 | 2.65E-11 | conserved hypothetical protein [Pediculus humanus corporis] |
| Cactus             | comp400958_c0 | XP_001927565.2 | 3.02E-47 | PREDICTED: tonsoku-like protein-like isoform 1 [Sus scrofa] |
| Dorsal/Dif         | comp115581_c0 | ACZ98167.1     | 0        | dorsal [Litopenaeus vannamei]                     |
|                     | comp1171_c0 | ACZ98167.1     | 0        | dorsal [Litopenaeus vannamei]                     |
|                     | comp15051_c0 | ACZ98167.1     | 0        | dorsal [Litopenaeus vannamei]                     |
|                     | comp28273_c0 | ADM14334.1     | 7.09E-178 | short gastrulation protein [Parhyale hawaiensis] |
|                     | comp415679_c0 | ADM14334.1     | 7.09E-178 | short gastrulation protein [Parhyale hawaiensis] |

doi:10.1371/journal.pone.0068233.t004
domain (Figure 5A). Multiple sequence alignment of MyD88 from E. sinensis and other thirteen species revealed that they were more similar at N-terminus and less conserved at C-terminus (Figure 5B). Sequence of E. sinensis MyD88 showed highest identity (65%) to homolog from L. vannamei.

**IMD pathway.** Fourteen non-redundant unigenes showed similarity to signaling molecules of IMD pathway, such as IMD, transforming growth factor beta-activated kinase (dTAK1), inhibitor of nuclear factor kappa-B kinase (IKK), Dredd and Relish, while Fas-associated death domain protein (dFADD) was not detected (Table 5, Figure S2). In putative IMD pathway, bacterial components activated the adaptor protein IMD, causing signaling cascade and finally leading to activation of Relish. Relish then regulated expression of antimicrobial peptide (AMP) and other immune-related genes.

**JAK-STAT pathway.** Various molecules involved in JAK-STAT signaling pathway were characterized in our analysis (Table 6, Figure S3). Through JAK-STAT pathway, different aspects of hematopoiesis and immune response were mediated by cytokines, including interleukin (IL), interferon (IFN), growth hormone (GH) and thyroid peroxidase (TPO) [21]. In putative JAK-STAT pathway, after interaction of cytokine and cytokine receptor (CytokineR), STAT was activated by JAK, dimerizing, and signal regulated kinase (ERK) family, and regulated different processes by protease cascade. In this putative pathway, each signal regulated kinase (ERK) family, and regulated different processes by protease cascade. In this putative pathway, each

**MAPK pathway.** Putative MAPK signaling pathway containing 122 non-redundant unigenes was also analyzed (Table 7, Figure S4). MAPKs were composed of three different major families – c-Jun N-terminal kinase (JNK) family, p38/stress-activated protein kinase (p38/SAPK) family and extracellular-signal regulated kinase (ERK) family, and regulated different processes by protease cascade. In this putative pathway, each cascade was triggered by extracellular signals and resulted in activation of MAPK kinase kinase (MAPKKK/MEKK), followed by activation of MAPK kinase (MAPKK/MEK/MKK) and MAPK/ERK, finally leading to function of diverse substrates and NF-kB proteins.

**Validation of Illumina sequencing results by RT-PCR**

Quantitative RT-PCR was used to confirm the expression profiles of genes that were identified in Illumina sequencing analysis. As was shown in Table S1, analyzed members of the study contained some pathway-associated components, including IMD, SOCS, Spaetzle (Spa), filamin (FLNA), p21-activated kinase 1 (Pak1), cytosolic phospholipase A2 (cPLA2), dual-specificity MAP kinase phosphatase (MKP), heat shock 70kDa protein (HSP), Ras-related C3 botulinum toxin substrate 1 (Rac1), hepatocyte growth factor (HGF), vascular endothelial growth factor receptor (VEGFR). Many other important immune-related genes, such as arginase (ARG), chitinase (Chi), Integrin (Int), lysozyme (LZM), peroxinectin (Pero), thymosin (Thy), anti-lipopolysaccharide factor (ALF), mannose receptor (ManR), scavenger-receptor (ScaR), mannose-binding protein (MBL), masquerade-like protein (MasL), glutathione peroxidase (GPX), glutathione S-transferase (GST), thioredoxin reductase (TrxR), trypsin-like serine protease (TrySP), chymotrypsin-like serine protease (ChySP), were also identified and analyzed (Table S1).

Results of RT-PCR revealed different expression abundances of the analyzed genes (Figure 6). Among them, TrySP, ChySP and ALF showed highest expression level, followed by Thy and TrxR, while LZW displayed the lowest level. It was consistent with the
### Table 6. Putative immune genes involved in JAK-STAT pathway.

| Signaling molecular | Unigene | ID           | E-value | Description                                                                                     |
|---------------------|---------|--------------|---------|-------------------------------------------------------------------------------------------------|
| TPO                 | comp107905_c0 | XP_002431634.1 | 0       | Thyroid peroxidase precursor, putative [Pediculus humanus corporis]                              |
|                     | comp13197_c0 | XP_002431634.1 | 0       | Thyroid peroxidase precursor, putative [Pediculus humanus corporis]                              |
|                     | comp13197_c1 | XP_002431634.1 | 0       | Thyroid peroxidase precursor, putative [Pediculus humanus corporis]                              |
|                     | comp170056_c0 | XP_002431634.1 | 0       | Thyroid peroxidase precursor, putative [Pediculus humanus corporis]                              |
|                     | comp267202_c0 | XP_002431634.1 | 0       | Thyroid peroxidase precursor, putative [Pediculus humanus corporis]                              |
| CytokinR            | comp14984_c0 | EFN76806.1    | 1.02E-75| Cytokine receptor [Harpegnathos saltator]                                                        |
|                     | comp23666_c0 | EFN76806.1    | 1.02E-75| Cytokine receptor [Harpegnathos saltator]                                                        |
|                     | comp23666_c1 | EFN76806.1    | 1.02E-75| Cytokine receptor [Harpegnathos saltator]                                                        |
|                     | comp97600_c0 | EEZ97840.1    | 5.51E-22| hypothetical protein TcasGA2_TCD00299 [Tribolium castaneum]                                       |
|                     | comp155435_c0 | ADV57398.1    | 3.34E-69| leptin receptor protein [Eriocheir sinensis]                                                     |
|                     | comp30992_c0 | ADV57398.1    | 3.34E-69| leptin receptor protein [Eriocheir sinensis]                                                     |
| JAK                 | comp11170_c0 | XP_002425471.1 | 3.44E-49| tyrosine-protein kinase jak2, putative [Pediculus humanus corporis]                             |
|                     | comp13215_c0 | XP_002425471.1 | 3.44E-49| tyrosine-protein kinase jak2, putative [Pediculus humanus corporis]                             |
|                     | comp29254_c0 | XP_002425471.1 | 5.67E-122| tyrosine-protein kinase jak2, putative [Pediculus humanus corporis]                             |
|                     | comp29596_c1 | XP_002425471.1 | 5.67E-122| tyrosine-protein kinase jak2, putative [Pediculus humanus corporis]                             |
| STAT                | comp25948_c0 | ACA79939.1    | 0       | STAT long form [Peneaus monodon]                                                                |
| STAM                | comp28940_c0 | XP_003398833.1 | 1.29E-104| PREDICTED: signal transducing adapter molecule 1-like [Bombus terrestris]                     |
| CBL                 | comp10185_c0 | XP_002426825.1 | 0       | E3 ubiquitin-protein ligase CBL, putative [Pediculus humanus corporis]                          |
|                     | comp10185_c1 | XP_002426825.1 | 0       | E3 ubiquitin-protein ligase CBL, putative [Pediculus humanus corporis]                          |
| PIAS                | comp22323_c0 | EFN82639.1    | 5.21E-134| E3 SUMO-protein ligase PIAS2 [Harpegnathos saltator]                                            |
| CBP                 | comp25273_c0 | XP_002423797.1 | 1.26E-88| CREB-binding protein, putative [Pediculus humanus corporis]                                    |
| SOCS                | comp30717_c0 | EFN64132.1    | 6.83E-111| CREB-binding protein [Camponotus floridanus]                                                    |
| Pim                 | comp126988_c0 | NP_001090165.1 | 1.67E-89| pim-3 oncogene [Xenopus laevis]                                                                |
|                     | comp188941_c0 | NP_001090165.1 | 1.67E-89| pim-3 oncogene [Xenopus laevis]                                                                |
| SHP2                | comp24266_c1 | XP_002430772.1 | 2.00E-18| tyrosine-protein phosphatase corkscrew, putative [Pediculus humanus corporis]                  |
| GRB2                | comp30475_c0 | XP_002426825.1 | 0       | E3 ubiquitin-protein ligase CBL, putative [Pediculus humanus corporis]                          |
|                     | comp30475_c1 | XP_002426825.1 | 0       | E3 ubiquitin-protein ligase CBL, putative [Pediculus humanus corporis]                          |
| SOS                 | comp27781_c0 | XP_002428152.1 | 0       | ras GTP exchange factor, son of sevenless, putative [Pediculus humanus corporis]               |
|                     | comp30785_c2 | XP_002428152.1 | 0       | ras GTP exchange factor, son of sevenless, putative [Pediculus humanus corporis]               |
| PI3K                | comp31023_c0 | ADE44091.1    | 0       | phosphoinositide 3-kinase isofrom b [Panulirus argus]                                            |
|                     | comp24583_c0 | XP_00160345.1 | 0       | PREDICTED: similar to MGC80357 protein [Nasonia vitripennis]                                    |
| AKT                 | comp29921_c0 | ADM87425.3    | 0       | Akt [Gecarcinus lateralis]                                                                     |
| CycD                | comp12912_c0 | NP_001089817.3 | 1.95E-55| cyclin D2 [Xenopus laevis]                                                                     |
|                     | comp138262_c0 | NP_001089817.3 | 1.95E-55| cyclin D2 [Xenopus laevis]                                                                     |
|                     | comp28643_c0 | XP_002431634.1 | 5.44E-50| PREDICTED: similar to cyclin d [Tribolium castaneum]                                            |
| Myc                 | comp28588_c0 | EFX79343.1    | 9.02E-16| Myc, dMyc-like protein [Daphnia pulex]                                                           |
|                     | comp30540_c3 | EEZ97840.1    | 2.98E-15| hypothetical protein TcasGA2_TCD000123 [Tribolium castaneum]                                      |
| BclXL               | comp27005_c0 | EG69168.1    | 2.27E-20| Bcl-2-like protein 1 [Acromyrmex echinatior]                                                     |
| Spred               | comp25472_c2 | XP_002414876.1 | 3.24E-41| sprouty protein evh1 domain-containing protein, putative [Hades scapularis]                    |
|                     | comp28599_c0 | XP_002414876.1 | 3.24E-41| sprouty protein evh1 domain-containing protein, putative [Hades scapularis]                    |
|                     | comp28599_c4 | NP_001164144.1 | 1.05E-46| sprouty-related protein with EVH-1 domain [Tribolium castaneum]                                |
| Sprouty             | comp22483_c0 | EFX18471.1    | 4.90E-45| hypothetical protein SINV_11790 [Solenopsis invicta]                                             |

doi:10.1371/journal.pone.0068233.t006
| Signaling molecular | Unigene   | ID                  | E-value     | Description                                                                 |
|---------------------|-----------|---------------------|-------------|-----------------------------------------------------------------------------|
| CACN                | comp371_c0| XP_003251102.1      | 0           | PREDICTED: voltage-dependent calcium channel subunit alpha-2/delta-3-like [Apis mellifera] |
|                     | comp21877_c0| XP_003251102.1      | 0           | PREDICTED: voltage-dependent calcium channel subunit alpha-2/delta-3-like [Apis mellifera] |
|                     | comp232994_c0| XP_003251102.1     | 0           | PREDICTED: voltage-dependent calcium channel subunit alpha-2/delta-3-like [Apis mellifera] |
|                     | comp398239_c0| XP_003251102.1      | 0           | PREDICTED: voltage-dependent calcium channel subunit alpha-2/delta-3-like [Apis mellifera] |
|                     | comp10818_c0| XP_002168351.1      | 3.42E-15    | PREDICTED: similar to calcium channel, voltage-dependent, alpha2/delta subunit 1, partial [Hydra magnipapillata] |
|                     | comp241026_c0| XP_002742061.1      | 1.68E-62    | PREDICTED: calcium channel, voltage-dependent, alpha2/delta subunit 3-like [Saccoglossus kowalevskii] |
|                     | comp409663_c0| XP_001807530.1      | 8.88E-45    | PREDICTED: similar to voltage-gated calcium channel alpha 1 subunit [Triolobium castaneum] |
| EGFR                | comp15094_c1| XP_003395927.1      | 0           | PREDICTED: epidermal growth factor receptor-like [Bombus terrestris] |
|                     | comp19914_c0| XP_003395927.1      | 0           | PREDICTED: epidermal growth factor receptor-like [Bombus terrestris] |
|                     | comp1108074_c0| XP_003395927.1      | 0           | PREDICTED: epidermal growth factor receptor-like [Bombus terrestris] |
|                     | comp190699_c0| XP_003395927.1      | 0           | PREDICTED: epidermal growth factor receptor-like [Bombus terrestris] |
| FGFR                | comp3098_c0| XP_003401483.1      | 0           | PREDICTED: fibroblast growth factor receptor homolog 1-like [Bombus terrestris] |
|                     | comp15223_c0| XP_003401483.1      | 0           | PREDICTED: fibroblast growth factor receptor homolog 1-like [Bombus terrestris] |
|                     | comp127712_c0| XP_003401483.1      | 0           | PREDICTED: fibroblast growth factor receptor homolog 1-like [Bombus terrestris] |
|                     | comp143525_c0| XP_003401483.1      | 0           | PREDICTED: fibroblast growth factor receptor homolog 1-like [Bombus terrestris] |
|                     | comp226442_c0| XP_003401483.1      | 0           | PREDICTED: fibroblast growth factor receptor homolog 1-like [Bombus terrestris] |
|                     | comp290904_c0| XP_003401483.1      | 0           | PREDICTED: fibroblast growth factor receptor homolog 1-like [Bombus terrestris] |
|                     | comp20492_c0| CAH03726.1          | 1.41E-49    | TPA: FGF receptor-like protein 1a [Takifugu rubripes] |
|                     | comp25479_c0| CAH03726.1          | 1.41E-49    | TPA: FGF receptor-like protein 1a [Takifugu rubripes] |
|                     | comp127203_c0| CAH03726.1          | 1.41E-49    | TPA: FGF receptor-like protein 1a [Takifugu rubripes] |
|                     | comp22034_c0| NP_001012263.2      | 1.30E-22    | fibroblast growth factor receptor-like 1b [Danio rerio] |
| PDGFR               | comp4550_c0| XP_002422693.1      | 4.28E-19    | alpha platelet-derived growth factor receptor precursor, putative [Pediculus humanus corporis] |
| GRB2                | comp30475_c0| XP_009998.1         | 2.03E-105   | PREDICTED: similar to AGAP011768-PA [Triolobium castaneum] |
| SOS                 | comp27781_c0| XP_002428152.1      | 0           | ras GTP exchange factor, son of sevenless, putative [Pediculus humanus corporis] |
|                     | comp30785_c2| XP_002428152.1      | 0           | ras GTP exchange factor, son of sevenless, putative [Pediculus humanus corporis] |
| Ras                 | comp23549_c0| XP_0077376.2        | 1.94E-37    | PREDICTED: similar to MRA52, putative [Triolobium castaneum] |
|                     | comp24659_c0| XP_0077154.1        | 3.98E-20    | PREDICTED: similar to MRA52, putative [Triolobium castaneum] |
|                     | comp27508_c0| AAK14389.1          | 1.34E-89    | Ras [Marsupenaeus japonicus] |
|                     | comp162483_c0| XP_0039895.2        | 1.14E-78    | ras-related protein M-Ras-like [Apis mellifera] |
|                     | comp29718_c0| XM_003506401.1      | 3.01E-14    | PREDICTED: Cricetulus griseus ras-related protein R-Ras-like [Cricetulus griseus] |
| G12                 | comp16378_c0| EFX86199.1          | 8.35E-18    | guanine nucleotide binding protein, gamma subunit [Daphnia pulex] |
|                     | comp18258_c0| EFX86199.1          | 8.35E-18    | guanine nucleotide binding protein, gamma subunit [Daphnia pulex] |
|                     | comp26578_c0| EG64184.1           | 3.12E-133   | Guanine nucleotide-binding protein subunit alpha-like protein [Acromyrmex echinatior] |
| Gap1m               | comp24182_c3| XP_0001945701.2     | 0           | PREDICTED: probable Ras GTPase-activating protein-like [Acrythosiphon pisonum] |
|                     | comp193916_c0| XP_0001945701.2     | 2.60E-111   | PREDICTED: probable Ras GTPase-activating protein-like [Acrythosiphon pisonum] |
|                     | comp30738_c0| XP_0001942745.1     | 0           | PREDICTED: ras GTPase-activating protein 1-like [Acrythosiphon pisonum] |
| NF1                 | comp16823_c0| XP_003402236.1      | 3.29E-64    | PREDICTED: neurofibromin-like [Bombus terrestris] |
|                     | comp29735_c1| XP_000160269.1      | 0           | PREDICTED: similar to neurofibromin [Nasonia vitripennis] |
| CNrasGEF            | comp21203_c0| XP_0001952587.1     | 0           | PREDICTED: ras guanine nucleotide exchange factor 2-like [Acrythosiphon pisonum] |
|                     | comp2527_c0| XP_000737277.1      | 3.29E-28    | PREDICTED: Ras guanine nucleotide exchange factor 2-like [Saccoglossus kowalevskii] |
| PKA                 | comp11911_c1| XP_000243550.1      | 0           | cAMP-dependent protein kinase catalytic subunit, putative [Pediculus humanus corporis] |
| Signaling molecular | Unigene ID | E-value | Description |
|---------------------|------------|---------|-------------|
| cPLA2               | XP_002127884.1 | 9.58E-53 | PREDICTED: similar to Cytosolic phospholipase A2 (cPLA2) (Phospholipase A2 group IVA) [Ciona intestinalis] |
| MNK1/2              | ACY6641.1 | 2.83E-167 | Map kinase-interacting serine/threonine kinase 1 [Crassostrea gigas] |
| RSK2                | NP_001036922.1 | 1.90E-158 | MAP kinase-activated protein kinase 2 [Saccoglossus kowalevskii] |
| Elk-1               | XP_002429096.1 | 8.37E-67 | Protein C-ets-1-B, putative [Haplographa lactea] |
| Rap1                | ACJ66625.1 | 7.77E-90 | Ras protein [Fenneropenaeus chinensis] |
| IKK                 | AAX56336.1 | 3.37E-90 | ikk-like protein [Pinctada fucata] |
| NF-kB               | ADM14334.1 | 1.03E-168 | Relish [Eriocheir sinensis] |
| PKC                 | NP_001036922.1 | 0 | Mitogen-activated protein kinase 2 [Nasonia vitripennis] |
| ERK                 | NP_001036922.1 | 1.90E-158 | MAP kinase-ERK kinase [Saccoglossus kowalevskii] |
| Tau                 | XP_001955318.1 | 1.60E-56 | GF18699 [Drosophila melanogaster] |
| STMN                | EGI59233.1 | 8.12E-49 | Stathmin-4 [Calanus finmarchicus] |
| SRF                 | ADM45311.1 | 3.38E-27 | Serum Response Factor [Artemia franciscana] |
| MKP                 | XP_002430571.1 | 4.24E-115 | Dual specificity protein phosphatase, putative [Pediculus humanus corporis] |
| PPP3C               | XM_001369081.1 | 0 | PREDICTED: similar to calcineurin A [Nasonia vitripennis] |
| FASL                | AEN6525.1 | 3.86E-80 | TNFSF [Litopenaeus vannamei] |
| FAS                 | AEK66527.1 | 8.05E-34 | TNFSF [Litopenaeus vannamei] |
| TGFBR               | XP_002127884.1 | 9.25E-53 | PREDICTED: transforming growth factor-beta receptor type I, putative [Haplographa lactea] |
| CASP                | ADM45311.1 | 3.38E-27 | Caspase [Eriocheir sinensis] |
| DAXX                | PE_002735791.1 | 9.16E-29 | PREDICTED: death-domain associated protein-like [Saccoglossus kowalevskii] |
| ECSIT               | AEK40021.1 | 4.42E-114 | Evolutionarily conserved signaling intermediate in Toll pathways [Marsupenaeus japonicus] |
| PPP2CB              | NP_001008030.1 | 1.21E-134 | Protein phosphatase, Mg2+/Mn2+ dependent, 1B [Xenopus (Silurana) tropicalis] |
| cdc42/Rac           | XP_002428346.1 | 3.00E-98 | RAC GTPase, putative [Pediculus humanus corporis] |
| Dystrobrevin        | XP_002428346.1 | 3.00E-98 | RAC GTPase, putative [Pediculus humanus corporis] |
Table 7. Cont.

| Signaling molecular | Unigene | ID             | E-value   | Description                                      |
|---------------------|---------|----------------|-----------|--------------------------------------------------|
| HGK                 | comp23927_c1 | XP_003403321.1 | 1.79E-16 | PREDICTED: mitogen-activated protein kinase kinase kinase 4-like isoform 1 (Bombus terrestris) |
| PAK1/2              | comp14636_c1 | EG164863.1     | 6.98E-114| Serine/threonine-protein kinase PAK 1 (Acromyrmex echinatior) |
|                    | comp25614_c0 | EG164863.1     | 6.98E-114| Serine/threonine-protein kinase PAK 1 (Acromyrmex echinatior) |
|                    | comp26148_c0 | XP_002251334.1 | 0         | PREDICTED: serine/threonine-protein kinase PAK 1 isoform 2 (Apis mellifera) |
|                    | comp26487_c0 | XP_002246989.1 | 1.33E-139| CDC42 GTPase-activating protein, putative (Pediculus humanus corporis) |
| MST1/2              | comp28042_c0 | EG57844.1      | 1.02E-157| Serine/threonine-protein kinase 3 (Acromyrmex echinatior) |
|                    | comp98224_c0 | EG57844.1      | 1.02E-157| Serine/threonine-protein kinase 3 (Acromyrmex echinatior) |
| MEKK1               | comp27862_c0 | XP_002408296.1 | 1.02E-21 | mitogen activated protein kinase kinase 1, MAPKK1, MEKK1, putative (xodes scapularis) |
|                    | comp29482_c2 | XP_424734.2    | 1.27E-44 | PREDICTED: similar to MEK kinase 1 (Gallus gallus) |
|                    | comp64508_c0 | XP_424743.2    | 1.27E-44 | PREDICTED: similar to MEK kinase 1 (Gallus gallus) |
| LTK                 | comp262893_c0 | XP_003396640.1 | 1.42E-164| PREDICTED: mitogen-activated protein kinase kinase kinase 13-like isoform 2 (Bombus terrestris) |
| TAK1                | comp10992_c0 | EFN79813.1     | 1.13E-92 | Mitogen-activated protein kinase kinase 7 (Harpegnathos saltator) |
| MEKK4               | comp81382_c0 | EDL02704.1     | 0         | mCG16678 [Mus musculus] |
| TAO                 | comp28070_c0 | XP_002246013.1 | 0         | predicted protein (Pediculus humanus corporis) |
| FLNA                | comp27356_c0 | EFX70014.1     | 2.10E-12 | hypothetical protein DAPPUDRAFT_328543 (Daphnia pulex) |
|                    | comp31197_c0 | XP_002243351.1 | 1.41E-35 | Filamin-C, putative (Pediculus humanus corporis) |
| JIP3                | comp6235_c0 | XM_003354659.1 | 5.17E-37 | PREDICTED: Sus scrofa mitogen-activated protein kinase 8 interacting protein 3, transcript variant 2 (MAPK8IP3) (Sus scrofa) |
|                    | comp30482_c0 | XP_003395970.1 | 0         | PREDICTED: LOW QUALITY PROTEIN: JNK-interacting protein 3-like (Bombus terrestris) |
| HSP72               | comp8227_c0 | XP_002246982.1 | 4.32E-80 | molecular chaperone (Enteroctozaon bieneusi H348) |
|                    | comp133984_c0 | AA57912.1      | 4.91E-65 | 70 kDa heat shock cognate protein 1 (Vigna radiata) |
|                    | comp197937_c0 | ACB70177.1     | 2.10E-49 | 70 kDa heat shock protein (Capparis spinosa) |
|                    | comp406329_c0 | XP_002532297.1 | 1.19E-173| heat shock protein, putative (Ricinus communis) |
|                    | comp25846_c0 | ACF98297.1     | 0         | heat shock protein 70 (Eriocheir sinensis) |
| ARRB                | comp28931_c4 | XM_001867259.1 | 8.49E-26 | Culex quinquefasciatus beta-arrestin 1 |
| Crk                 | comp28991_c0 | XP_002427596.1 | 5.89E-106| Adapter molecule Crk, putative (Pediculus humanus corporis) |
| MKK4                | comp25769_c0 | EFN81517.1     | 4.92E-152| Dual specificity mitogen-activated protein kinase kinase 4 (Harpegnathos saltator) |
| JNK                 | comp20212_c0 | BAIB7826.1     | 0         | c-jun N-terminal kinase (Marsupenaeus japonicus) |
| JUN                 | comp28673_c0 | EG6820.1       | 1.90E-49 | Transcription factor AP-1 (Acromyrmex echinatior) |
| AKT                 | comp31162_c1 | EG6820.1       | 1.90E-49 | Transcription factor AP-1 (Acromyrmex echinatior) |
|                    | comp29921_c0 | ADM87425.3     | 0         | Akt (Gecarcinus lateralis) |
| PP5                 | comp29577_c0 | XP_971407.1    | 0         | PREDICTED: similar to protein phosphatase-5 (Tribolium castaneum) |
| ATF2                | comp30056_c0 | XP_001515843.1 | 1.13E-16 | PREDICTED: similar to activating transcription factor 2 (Ornithorhynchus anatinus) |
| p38                 | comp24454_c0 | ADT91683.1     | 8.07E-169| p38 mitogen-activated protein kinase (Apis cerana cerana) |
| p53                 | comp24379_c1 | AGCQS385.1     | 2.54E-13 | p53 and DNA damage-regulated protein 1 (Anoploparma fimbria) |
|                    | comp204793_c0 | XP_96601.2     | 2.59E-102| PREDICTED: similar to apoptosis stimulating of P53 (Tribolium castaneum) |
| MAX                 | comp25261_c3 | XP_003401810.1 | 7.59E-38 | PREDICTED: protein max-like isoform 1 (Bombus terrestris) |
| MEF2C               | comp18140_c0 | XP_971771.1    | 1.56E-82 | PREDICTED: similar to myocyte-specific enhancer factor 2d (Tribolium castaneum) |
|                    | comp18140_c1 | XP_971771.1    | 1.56E-82 | PREDICTED: similar to myocyte-specific enhancer factor 2d (Tribolium castaneum) |
| MSK1/2              | comp21333_c0 | XP_002431024.1 | 3.20E-21 | Ribosomal protein 56 kinase alpha-5, putative (Pediculus humanus corporis) |
| NLK                 | comp17966_c0 | XP_000248311.1 | 0         | GJ13897 [Drosophila virilis] |
| MAPKAPK             | comp15249_c0 | ABC25082.1     | 7.31E-87 | MAP kinase activated protein-kinase-2 (Glossina morsitans morsitans) |
|                    | comp30931_c0 | ABC25082.1     | 7.31E-87 | MAP kinase activated protein-kinase-2 (Glossina morsitans morsitans) |
|                    | comp144964_c0 | ABC25082.1     | 7.31E-87 | MAP kinase activated protein-kinase-2 (Glossina morsitans morsitans) |

doi:10.1371/journal.pone.0068233.t007
results of Illumina sequencing data (Table S1), which not only validated the expression profile of different identified immune genes, but also verified the reliability and accuracy of our transcriptome analysis.

**Discussion**

In this context, considerable efforts have been made to research hepatopancreas transcriptome of microbial challenged *E. sinensis* by high-throughput sequencing technology (Solexa/Illumina). Comparing with EST analysis of hepatopancreas from *E. sinensis* with traditional method [19,22], our study produces more sequencing reads and assembled unigenes. It largely enriches transcriptome data of *E. sinensis* and indicates enormous advantage of high-throughput technology. Although a comparative transcriptome analysis of haemocytes from *E. sinensis* under normal condition and in response to *Spiroplasma eriocheiris* infection indicates certain microRNAs may be essential in interaction between host and pathogen [11], only miRNAs are identified and analyzed for the expression pattern. In our study, various immune genes and pathways are annotated from hepatopancreas of *E. sinensis* after immune challenge. The analysis increases molecular information and genomic resources of *E. sinensis* in response to microorganism stimulation.

Toll pathway was initially identified in genetic screen of genes involved in early embryonic development of *Drosophila* [23] and gradually studied of importance in innate immunity. In economic crustaceans, many genes related to Toll pathway, such as Spatzle [24], Toll [24], MyD88 [25], Pelle [26] and TRAF6 [27], have been reported from shrimp, while only SpToll of *Scylla paramamosain* [28] has been cloned and characterized from crab. In the present study, we are first to find various key members of Toll pathway in *E. sinensis*. This suggests the existence of putative Toll pathway in crab and indicates its crucial function in antimicrobial response. Moreover, different from mammalian Toll-like receptors (TLRs) directly functioning as a pattern recognition receptor (PRR) to recognize pathogen-associated molecular patterns (PAMPs) [29], DmToll of *Drosophila melanogaster* uses the cytokine-like molecule Spatzle as a ligand [24,30,31]. In *E. sinensis*, identification of Spatzle in our study also suggests that the Chinese mitten crab Toll may be activated by functioning with Spatzle. In addition, in spite of different MyD88 variants in human, mice, chicken and other vertebrates, only a MyD88 variant gene is found in an invertebrate species *L. vannamei* [23]. Here, we find one MyD88 sequence of *E. sinensis* that shows highest similarity to homolog from *L. vannamei*. This will provide a foundation for further study of MyD88 in crab.

Gram-negative bacteria-yielded diaminopimelic acid (DAP)-type peptidoglycan can be recognized by peptidoglycan recognition protein (PGRP)-LE and PGRP-LC receptor complex, which then activate IMD and cause activation of signaling cascade to trigger Relish [32]. Experiments of *Drosophila* also reveal that infection by Gram-negative bacteria activates IMD pathway, but not Toll pathway [30,31]. In this context, among different molecules relevant to IMD pathway, not only caspase and Relish previously reported [33,34] are identified, but also IMD, dTAK1 and IKK are first found in microbial challenged *E. sinensis*. Similarly, LvIMD of *L. vannamei* and FcRelish of *Fenneropenaeus chinensis* are identified after immune challenge and characterization of them implies that they can induce expression of some antimicrobial peptides (AMPs), which are integral components of innate immune system and exhibit great activities to defense against pathogens [35,36]. Taken these reports together, investigation of principal component molecules will promote researching on innate immune mechanism and immune pathway of *E. sinensis*.

A large number of molecules involved in JAK-STAT signaling pathway such as four JAKs, seven STATs and more than 30 cytokines are widely found in mammals [21]. However, only SOCS and leptin receptor protein (LEPR) have been cloned from *E. sinensis* [37,38]. In the present study, along with SOCS and LEPR, many other genes including CytokineR, JAK, STAT, downstream genes and regulatory molecules (CIS, SHP1, SOCS,
PIAS and STAM) are first fully and systemically identified in crab. Considering different aspects of cell development and host response activated by JAK-STAT pathway [39], there is no surprise that lots of regulators are found to control this pathway. Expression and regulation of components in JAK-STAT pathway are also reported in transcriptome analyses of microbial infected Pseudociona crassa [40] and Laodelphax crispellus [41]. These reports together increase knowledge of JAK-STAT pathway on microbial stimulation and provide valuable information for further study of immune response against pathogen infection. Additionally, researchers have compared the one single STAT gene from invertebrates with seven STATs from vertebrates by phylogenetic analysis [42]. This comparison supports the hypothesis that STAT genes duplicate before splitting in invertebrates and vertebrates and shows difference between them. In mammals and other vertebrates, JAK-STAT pathway plays a crucial function in lots of biological processes of both innate and adaptive immunity, such as apoptosis, proliferation, differentiation, hematopoiesis, oncogene-sis and immune defense [39,43]. However, in crustaceans, only the antibacterial or antiviral activities of several JAK/STAT genes are known so far [37,44,45]. It is still unknown whether the pathway has other functions and needs us to do more efforts for its complete function research.

MAPK pathway widely exists in all eukaryotes from yeast to human. Through a conserved three-kinase cascade that finally phosphorylates intracellular substrates and transcription factors, it transduces extracellular cues to cytoplasm and nucleus to control physiological processes [46]. Currently, similar with JAK-STAT pathway, most knowledge of MAPK pathway is also focused on vertebrate system. In invertebrates, this pathway is multifunctional and plays a key role in anti-stress, reproduction, cell development, differentiation and inflammation [46]. In shrimp, anti-lipopoly-saccharide factor treatment can regulate Trichomonas vaginalis-induced proinflammatory cytokines through MAPK pathway [47]. Interestingly, in the crab Chasmagnathus, MAPK pathway participates in neural plasticity, which can only be found in rodents and mollusks before, and is necessary for long-term memory consolidation of this crab model [48]. Thus, MAPK pathway may have many different functions in various species of vertebrates and crustaceans. However, for the reason that knowledge about MAPK pathway in aquatic invertebrates is largely unclear, it still needs deep research to fully clarify the role of this pathway. Our detection of numerous genes involved in MAPK pathway, such as ERK, JNK, p38, MEK, MEKK, Elk, Crk and CREB, will offer valuable reference in crab and other important crustaceans.

In conclusion, numerous genes from hepatopancreas of microbial challenged E. sinensis are characterized to be associated with Toll, IMD, JAK-STAT and MAPK pathways. Accuracy of Illumina sequencing data and expression profile of the identified genes are also further confirmed by RT-PCR. This research will be not only helpful to fully research host-pathogen interaction and comprehensively understand immune system of crab, but also beneficial to prevent diseases appeared in crab culture.

Supporting Information

Figure S1 Putative Toll pathway. Putative Toll pathway of E. sinensis was constructed based on knowledge in Drosophila and shrimps. Proteins appearing in hepatopancreas of microbial challenged E. sinensis were represented in grey circle and absent proteins in grey square. However, most interactions have to be confirmed experimentally. (TIF)

Figure S2 Putative IMD pathway. Putative IMD pathway of E. sinensis was constructed based on knowledge in Drosophila and shrimps. Proteins appearing in hepatopancreas of microbial challenged E. sinensis were represented in grey circle and absent proteins in grey square. However, most interactions have to be confirmed experimentally. (TIF)

Figure S3 Putative JAK-STAT pathway. Putative JAK-STAT pathway of E. sinensis was constructed based on KEGG reference pathway. Proteins appearing in hepatopancreas of microbial challenged E. sinensis were represented in circle and absent proteins in square. However, most interactions have to be confirmed experimentally. (TIF)

Figure S4 Putative MAPK pathway. Putative MAPK pathway of E. sinensis was constructed based on KEGG reference pathway. Proteins appearing in hepatopancreas of microbial challenged E. sinensis were represented in circle and absent proteins in square. However, most interactions have to be confirmed experimentally. (TIF)

Table S1 Genes and specific primers used for real-time PCR.

Author Contributions

Conceived and designed the experiments: ZC. Performed the experiments: XL ZC YL CS GS. Analyzed the data: XL CS. Wrote the paper: XL ZC.

References

1. Wang W, Gu Z (2002) Rickettsia-like organism associated with tremor disease and mortality of the Chinese mitten crab Eriocheir sinensis. Diseases of Aquatic Organisms 49: 149–153.
2. Wang W, Zhu N, Gu Z, Du K, Xu Z (2002) Study on the transmission of tremor disease (TD) in the Chinese mitten crab, Eriocheir sinensis (Crustacea: Decapoda). Journal of Invertebrate Pathology 81: 202–204.
3. Bonami JR, Zhang S (2011) Viral diseases in commercially exploited crabs: a review. J Invertebr Pathol 106: 6–17.
4. Fu Z, Zeng Y, Li Y, Guo XN, et al. (2012) Expression and regulation of components in JAK-STAT pathway of E. sinensis. Journal of Invertebrate Pathology 106: 142–147.
5. Li WW, Jin XK, He L, Jiang H, Gong YN, et al. (2010) Molecular cloning, characterization, expression and activity analysis of cathepsin L in Chinese mitten crab, Eriocheir sinensis. Fish & Shellfish Immunology 29: 1010–1018.
6. Gai Y, Zhao J, Song L, Li C, Zheng P, et al. (2006) A prophenoloxidase from the Chinese mitten crab Eriocheir sinensis: Gene cloning, expression and activity analysis. Fish & Shellfish Immunology 24: 156–167.
Infected by Colletotrichum gloeosporioides. Journal of Phytopathology 160: 449–460.

13. Hsu JC, Chien TY, Hu CC, Chen MJ, Wu WJ, et al. (2012) Discovery of Genes Related to Insecticide Resistance in Bactrocera dorsalis by Functional Genomic Analysis of a De Novo Assembled Transcriptome. PLoS One 7: e40950.

14. Zhao X, Wang Q, Jiao Y, Huang R, Deng Y, et al. (2012) Identification of Genes Potentially Related to Biomimarization and Immunity by Transcriptome Analysis of Pearl Sac in Pearl Oyster Pinctada maxima. Mar Biotechnol (NY) 14: 739–749.

15. Gavry M, Roberts SB (2012) Characterizing short read sequencing for gene discovery and RNA-Seq analysis in Crassostrea gigas. Comp Biochem Physiol Part D Genomics Proteomics 7: 94-99.

16. Roux MM, Pain A, Klimpel KR, Dhar AK (2002) The Lipopolysaccharide and β1,3-Glucan Binding Protein Gene Is Upregulated in White Spot Virus-Infected Shrimp (Penaeus stylirostris). Journal of Virology 76: 7140–7149.

17. Ried C, Wahl C, Morthile T, Wellnhofer G, Landgraf C, et al. (1996) High Affinity Endotoxin-binding and Neutralizing Peptides Based on the Crystal Structure of Reombinant Limulus Anti-liopolysaccharide Factor. Journal of Biological Chemistry 271: 28120–28127.

18. Gross P, Bartlett T, Bwolf C, Chapman R, Warr G (2003) Immune gene discovery by expressed sequence tag analysis of hemocytes and hematopoeises in the Pacific White Shrimp, Litopenaeus vannamei, and the Atlantic White Shrimp, L. setiferus. Developmental and Comparative Immunology 25: 365–377.

19. Jiang H, Cai YM, Chen LQ, Zhang XW, Hu SN, et al. (2009) Functional annotation and analysis of expressed sequence tags from the hematopoeises of mitten crabs (Eriocheir sinensis). Mar Biotechnol (NY) 11: 317–326.

20. 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. Nat Biotechnol 29: 644–652.

21. Kisselova T, Bhattacharya S, Braunstein J, Schindler C (2002) Signaling through the JAK-STAT pathway, recent advances and future challenges. Gene 285: 1–24.

22. Jiang H, Yin Y, Zhang X, Hu S, Wang Q (2009) Clasings relationships between nutrition and reproduction: A comparative transcriptome analysis of hepato-pancreas and testis from Eriocheir sinensis. Comp Biochem Physiol Part D Genomics Proteomics 4: 227–234.

23. Morisato D, Anderson KV (1995) Signaling pathways that establish the dorsalventral pattern of the Drosophila embryo. Annual review of genetics 29: 371–399.

24. Wang PH, Jiang JP, Gu ZH, Wan DH, Weng SP, et al. (2012) Molecular cloning, characterization and expression analysis of two novel Tolls (LiToll1 and LiToll3) and three putative Spatzle-like Toll ligands (LvSpz1-3) from Litopenaeus vannamei. Dev Comp Immunol 36: 359–371.

25. Zhang S, Li C, Yan H, Qiu W, Chen Y, et al. (2012) Identification and Function of Myeloid Differentiation Factor 88 (MyD88) in Litopenaeus vannamei. PLoS One 7: e40738.

26. Wang PH, Gu ZH, Wan DH, Zhang MY, Weng SP, et al. (2011) The shrimp NF-kappaB pathway is activated by white spot syndrome virus (WSSV) 449 to facilitate the expression of WSSV069 (erl), WSSV303 and WSSV371. PLoS One 6: e24773.

27. Wang PH, Wan DH, Gu ZH, Deng XX, Weng SP, et al. (2011) Litopenaeus vannamei tumor necrosis factor receptor-associated factor 6 (TRAF6) responds to Vibrio alginolyticus and white spot syndrome virus (WSSV) infection and activates antimicrobial peptide genes. Dev Comp Immunol 35: 105–114.

28. Liu Z, Qiao J, Zhang Y, Guo L, Huang H, et al. (2012) Cloning and characterization of the SpToll gene from green mud crab, Scylla paramamosain. Dev Comp Immunol 37: 164–175.

29. Kawai T, Akira S (2006) TLR signaling. Cell Death Differ 13: 816-825.