Transcriptome Analysis of juvenile Pacific White Shrimp (*Litopenaeus vannamei*) with symptom of stunted growth

Huaiyi Fang^{2#}, Jingjing Song^{2#}, Bin Gong^{1*}, Tingcai Pang², Chunyan Peng¹

1. Guangxi Key Laboratory of Marine Disaster in the Beibu Gulf, Beibu Gulf University, Qinzhou, China, 535000

2. Guangxi Key Laboratory of Beibu Gulf Marine Biodiversity Conservation, Beibu Gulf University, Qinzhou, China, 535000

Huaiyi Fang and Jingjing Song contribute equally to this article

*Correspondent author.

E-mail addresses: 342965691@aliyun.com (Bin Gong)

Abstract

In search for viruses associated with *Litopenaeus vannamei* with symptom of stunted growth, we have undertaken a comparative transcriptome analysis of total RNA sequences using Illumina based high-throughput sequencing (HTS). We obtain 89000242 and 95126160 high-quality clean reads from cDNA libraries of *L.vannamei* with stunted growth symptom (sick group) and health shrimp (health group control), respectively. Totally, 108221 unigenes with an average length of 716 bp were obtained from RNA-seq data. The unigenes were annotated against NR, NT, KO, KOG, SwissProt, PFAM and GO databases. 3975 (3.67%) showed significant matches in all the above databases and 39812 (36.78%) unigenes were annotated in at least one database. Comparative transcriptomic analysis revealed that 807 significantly differentially expressed unigenes were identified, including 282 down-regulated genes and 525 up-regulated genes. Thirteen up-regulated virus-related genes were only discovered in the sick shrimp groups, but not in health ones. Five of them were closely related to virus family Picornaviridae. From phylogenetic tree, we could find five viral genes were closely related to genus Ampvirus, Falcovirus, Aquamavirus (Seal picornavirus). Some up-regulated genes in the Sick groups mainly included genes involved in virus infecting processes concerning antigen processing and presenting, apoptosis, lysosome, phagosome and inflammation. While, many down-regulated genes in the Sick groups mainly involved in the process of starch and sucrose metabolism, and protein digestion and absorption. Our research provided some useful information about pathogenic factor and mechanism on the stunted growth symptom of *L.vannamei*, and was significant for the control and prevention of this disease.

Key words: *Litopenaeus vannamei*, transcriptome, Illumina, symptom of stunted growth

Introduction

Litopenaeus vannamei are economically important and widely farmed aquaculture species. It currently provides about 52% of the total penaeid shrimp production in the world [1-2]. However, the existence of viral or bacterial disease has posed a serious threat to the shrimp aquaculture industry and caused enormous economic losses [3-4]. The stunted growth symptom of *L.vannamei* was one of shrimp diseases appeared frequently in many places in China during the latest several years [5]. Although this significantly impact the shrimp aquaculture and gained considerable attention, the detail for this was not worked out until now.

Previous research indicated that the stunting of *Penaeus* shrimp was closely related to the presence of some known viruses. Statistical analysis showed a correlation between stunting of *Penaeus monodon* and severity of hepatopancreatic parvovirus (HPV) infection [6]. Rai et al. (2009) detected four kinds of viruses in 72 *P. monodon* samples with slow growth symptom by PCR and suggested the presence of infectious hypodermal and hematopoietic necrosis virus (IHHNV) could be one of the causes of slow growth in cultured *P. monodon* in India [7]. Monodon baculovirus (MBV) was also found a correlation with slow growth of shrimp since the mean length of MBV infected shrimp was significantly shorter than uninfected shrimp from the same pond [8]. In 2001-2002, an unusual retarded growth of *P.monodon*, called monodon slow growth syndrome (MSGs), was found in Thailand. This disease resulted in decreased production of shrimp and approximately US\$ 300 million economic losses in 2002 in Thailand [9]. Laem-Singh virus (LSNV) was suggested associating with MSGs [10]. A novel integrase-containing element may interact with LSNV to cause MSGs [11]. RNAi technology using dsRNA of LSNV was proven to be effective in reducing LSNV and restoring normal growth in LSNV-infected shrimp [12].

The specific pathogen free (SPF) shrimp ensure that the *L.vannamei* is free of economically important viruses. This is an important program to ensure the success of shrimp aquaculture. However, new emerging viruses make it very urgent to develop novel and effective method to discover them. In our present work, some *L.vannamei* shrimp samples with stunted growth symptom were collected from Qinzhou and Beihai city in China. The HPV, IHHNV, MBV and LSNV could not be detected in all the samples by PCR test. In order to find out what pathogen agents caused the stunting of shrimp and its pathogenesis, high-throughput sequencing (HTS) technology based RNA-seq and comparative transcriptome analysis were exploited to discover whether new or unusual viruses infected sick shrimp and the molecular mechanism of this disease.

Our research may provide some useful information about pathogenic factor and mechanism on the stunted growth symptom of *L.vannamei*, and will be significant for the establishment of more testing standards and maintaining SPF shrimp stocks.

Materials and Methods

Shrimp sample

Total of 20 *Litopenaeus vannamei* with symptom of stunted growth (grew for 100 days and the lengths of body were only about 2-3 cm) were collected from shrimps breeding farms (Qinzhou and Beihai city in China), transferred to laboratory immediately and maintained in the aerated recirculating seawater system (temperature 25-26°C, salinity of 32-ppt) for six days prior to the experiment. In parallel, the negative control group was juvenile shrimps collected from a specific pathogen-free (SPF) line with PCR confirmation (without infection of HPV, IHHNV, MBV and LSNV). Then the two shrimp groups: two sick groups with stunted growth (Sick1 and Sick2), and two negative health control groups (each group containing 15 shrimps) (Health1 and Health2) were used in the following experiment. Five individuals from each group were frozen immediately in liquid nitrogen, and then they were stored at -80°C until RNA extraction.

RNA isolation, cDNA library construction and Illumina sequencing

The above frozen shrimp samples (Sick and Health) were delivered to Novogene Company (Beijing, China) by dry ice for total RNA extraction, cDNA library construction and Illumina sequencing. The total RNA was extracted with TRIzol reagent (Invitrogen, Life Technologies, Grand Island, NY, USA). The RNA purity, concentration, degradation, contamination and integrity was assessed by conventional approach [13-14]. 1.5 µg total RNA was used to generate sequencing libraries by NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). Poly-T oligo-attached magnetic beads were used to isolate mRNA from total RNA. The purified mRNA was cleaved into fragments in Reaction Buffer using divalent cations under elevated temperature. First-strand cDNA was synthesized with random hexamer primer and RNase H, and the second-strand cDNA was produced using RNase H and DNA Polymerase I. Then the double-strand cDNAs were subjected to end repair and addition of A and adapter. The cDNA fragments 150-200 bp in length were selected and purified with AMPure XP system (Beckman Coulter, Beverly, USA). PCR amplification were performed for library construction. After necessary purification and qualification, the cDNA libraries were sequenced on an Illumina HiSeq platform and paired-end reads were generated.

De novo transcriptome assembly and gene function annotation

All the generated raw data (raw reads) were filtered to get high quality clean reads by removing adapter and low quality reads. Meanwhile, Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. Transcriptome assembly was accomplished using Trinity [15] with min_kmer_cov set to 2 by default and all other parameters set default. Gene function of the unigenes was annotated based on databases: Nr (NCBI non-redundant protein sequences); Nt (NCBI non-redundant nucleotide sequences); Pfam (Protein family, <http://pfam.xfam.org/search>) [16]; KOG/COG (<http://www.ncbi.nlm.nih.gov/COG/>) [17]; Swiss-Prot (<http://www.ebi.ac.uk/uniprot/>) and GO (<http://www.geneontology.org/>). The protein function annotation about Unigenes was achieved by protein with the highest sequence similarity in the databases.

Differential expression analysis

RSEM was used to estimate the gene expression levels in each sample [18]. The DESeq R package (1.10.1) was used to analyse the different expression pattern between group sick and health. DESeq software package determine differential expression in digital gene expression data by statistical methods. In order to control the false discovery rate (FDR), the P values adjusted and P-value <0.05 was assigned as a criteria for differential expression.

Enrichment analysis of GO and KEGG

Enrichment analysis of differentially expressed genes (DEGs) was conducted by using Gene Ontology (GO) and KEGG. GO seq R packages based Wallenius non-central hyper-geometric distribution [19] was conducted to adjust for gene length bias in DEGs. The database KEGG was used to understand the gene functions and their activity in cell and organism (<http://www.genome.jp/kegg/>) [20]. Then the KOBAS software was used to test statistical enrichment of DEGs in KEGG pathways [21].

Gene expression validation

In order to validate the correction of Illumina sequencing, eight differentially expressed genes were chosen for quantitative analysis by real-time PCR (qRT-PCR) (Table 1). These genes included heat shock 70kDa protein (HSP70), caspase 7 (CASP7), dynein heavy chain 1(DHC-1), low-density lipoprotein receptor (LDLR), trypsin (PRSS), trehalose 6-phosphate synthase (otsA), alpha-amylase (AMY) and collagen (COL1A), where 18s rRNA gene was used as the reference gene. The primers of PCR were designed using Primer Premier 5 software and sequences were indicated in Table 1. Five RNA samples in each group (Sick and Health) were used for gene

expression validation. 1 mg of total RNA was used as template for the synthesis of first strand cDNA using reverse transcriptase M-MuLV (NEB). The qRT-PCR reaction mixture (20 µL) contain 2 µL of primer (10 µM), 10 µL 2×SYBR Green PCR buffer, 1 µL of cDNA template. The PCR cycling parameters as follows: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 20 s, 60°C for 1 min and a final 72°C for 1 min. All the reactions were done in three triplicates.

Table 1. Primers utilized in the gene expression validation experiment.

Gene id	Similarity	qRT-PCR
		Primers sequences
c73441_g2	heat shock 70kDa protein (HSP70)	CGCAGTAGCTTTCCAAACCG GCCATTTTGAGCGGACAAGG
c72347_g1	caspase 7 (CASP7)	CGAAACCCCACTGGACGTAA ACGAGAGAGCAGCTGTTGTC
c74367_g1	dynein heavy chain 1(DHC-1)	GGACGAAGACAGGAACACGT AGTCGAATGGCCGATGATCC
c81817_g2	low-density lipoprotein receptor (LDLR)	CCTTAATGGCGACGGGTCTT GGTGGCGAGGATGAAGTTGA
c68281_g1	trypsin (PRSS)	AAACCGGCCCCACTTATCAG TGAGCGGGAATATCGATGGC
c76080_g1	trehalose 6-phosphate synthase (otsA)	AACCGAAGCTGCATGTCTGA CGCAGGCGATGACATTTTCAG
c66739_g1	alpha-amylase (AMY)	GGTCTGCTTGTGGACGAAGA TTCTCCCTCGCCCATGACTA
c74810_g1	collagen (COL1A)	GGGAAGATGGGAGAGGGTCT AACATCTACACCATGCGGCA

Results

Illumina sequencing and *De novo* assembly of shrimp transcriptome

For the group Sick, 89000242 clean reads were generated from 92277132 raw reads. The mean values of Q20 percentage, Q30 percentage, GC percentage are 96.84%, 92.14%, 49.62%, respectively, and the error rate is 0.02%. For the health samples, 95126160 clean reads were generated from 98615280 raw reads. The Q20 percentage, Q30 percentage, GC percentage are 96.8%, 91.99%, 49.59%, respectively. The clean bases were performed *de novo* assembly using Trinity software. A total of 130544 transcripts with 120846130 nucleotides were assembled. The length of the transcripts ranged from 201 bp to 38462 bp with an average length of 926 bp, and the N50 and N90 were 2080 and 301, respectively. After removing the low-quality and short-length

155 sequences, these transcripts were further assembled and clustered into 108221 unigenes with an
156 average length of 716 bp and N50 of 1361.

157

158 Gene functional annotation

159 The unigenes were submitted to alignments against the public databases for gene functional
160 annotation, and the unigenes annotated in NR, NT, KO, KOG, SwissProt, PFAM and GO were
161 26005 (24.02%), 12468 (11.52%), 12194 (11.26%), 16806 (15.52%), 23050 (21.29%), 29084
162 (26.87%) and 29911 (27.63%), respectively. 3975 (3.67%) showed significant matches in all the
163 above databases and 39812 (36.78%) unigenes were annotated in at least one database.

164 According to GO analysis, a total of 29912 unigenes were classified into three major
165 functional categories (Fig 1A). The majority kinds of unigenes (1235) involved in the biological
166 process of shrimp cells, followed by 369 kinds of unigenes were responsible for synthesis cellular
167 component. 173 kinds of unigenes had certain molecular function such as catalytic activity,
168 molecular function regulator, transporter activity and binding. In order to determine the biological
169 pathways in both Sick and Health shrimp, all the unigenes were mapped to the referential
170 canonical pathways in the KEGG database. A total of 15205 unigenes were assigned to 229
171 KEGG pathways. Among them, the top 10 KEGG pathways are: signal transduction (1526),
172 Translation (1438), Transport and catabolism (1194), Endocrine system (1051), protein Folding,
173 sorting and degradation (939), Carbohydrate metabolism (780), Immune system (703), Amino
174 acid metabolism (615), Cellular commiunity (589) and Digestive system (529) (Fig 1B). KOG
175 classification of the unigenes is important for functional annotation and evolutionary studies. A
176 total of 16806 unigenes were categorized into 26 KOG classifications, among which the (R)
177 General Function Prediction Only (2892) was dominant, followed by (O) Posttranslational
178 modification (2389) and (T) Signal transduction mechanisms (2111) (Fig 1C).

179 A.

180 B.

181 C.

182 **Figure 1. Gene functional annotation by Go, KEGG and KOG analysis.** (A) GO
183 categorization of non-redundant unigenes, and each annotated unigene was
184 assigned at least one GO term. (B) KEGG classification of non-redundant unigenes. (C)
185 KOG annotation of putative proteins.

186

187 The Differential expressed genes between Sick and Health shrimp

188 In our work, an absolute value of $\log_2\text{Ratio} \geq 1$ was utilized as the filtering threshold to determine
189 down-regulated or up-regulated genes between Sick and Health shrimp groups. 807 significantly
190 differentially expressed unigenes were identified between Sick and Health groups of shrimp,
191 including 282 down-regulated genes and 525 up-regulated genes (Figure 2). The different
192 expression pattern of Sick and Health were further discovered by a hierarchical cluster analysis.
193 By using this analytical method, a clustered pattern of the DEGs between the two groups was

194 distinctively illustrated (Figure 3), and it was found that the genes expressed in Sick group were
195 distinguishable from that of the Health group.

196 **Figure 2. Differentially expressed unigenes were identified between Sick and**
197 **Health groups of shrimp.**

Cluster analysis of differentially expressed genes

Figure 3. The different expression pattern between Sick and Health groups discovered by a hierarchical cluster analysis.

It was noticed that 35 of 807 DEGs were virus related genes. Among them, 27 viral related genes were up-regulated, while seven were down-regulated (Table 2). It was worth specially mentioning that 13 up-regulated virus-related genes were only discovered in the Sick shrimp groups but not in health ones. Moreover, all the down-regulated virus-related genes appeared in the health shrimp also could be detected in the sick groups. In the 13 up-regulated virus-related genes that only appeared in the sick shrimp groups, it was found that 5 of them were related to genes of virus family Picornavirus (incertae sedis (phylum)/incertae sedis (class)/Picornavirales (order)/Picornaviridae (family)). These viral gene fragments encoded proteins of Picornavirus including RNA-directed RNA polymerase, virus capsid protein and RNA helicase (Table 3). By constructing Neighbour-Joining tree of these five viral gene fragments, it was found that three of them exhibited the same pattern as figure 4A, and the other two displayed trees as figure 4B and 4C, respectively (Figure 4). From all the above phylogenetic tree, we could find a similar feature that the five viral genes were closely related to genus Ampvirus, Falcovirus, Aquamavirus (Seal picornavirus) in the family Picornaviridae. However, the similarity of sequence between the five viruses and that existing in the database was low (about 20%-40%).

Table 2. The annotation details about down and up regulated virus-related genes.

Gene ID	Sick_read count	Health_read count	Swissprot Description	PFAM description	Cellular Component description
Up-regulated Genes					
c75311_g1	352.52	0	Replicase polyprotein OS=Cricket paralysis virus	RNA helicase	RNA-directed RNA polymerase complex
c78682_g2	351.11	0	--	Equine arterivirus Nsp2-type cysteine proteinase	Extracellular region//membrane
c76407_g1	63.83	0	--	Adenovirus IVa2 protein	--
c73240_g1	46.94	0	--		Viral capsid
c56446_g1	73.69	0		Viral RNA	RNA-directed RNA

				helicase//Viral methyltransferase	polymerase complex
c65243_g1	23.47	0	--	Baculovirus polyhedron envelope protein	Viral capsid//viral Envelope
c68785_g1	21.592	0		Picornavirus capsid protein	Viral capsid
c69554_g1	60.55	0		Picornavirus capsid protein	Viral capsid
c20203_g1	29.57	0	--	Herpes virus major outer envelope glycoprotein (BLLF1)	Viral envelope
c79968_g1	122.51	0	--	African swine fever virus multigene family 360 protein	--
c57532_g1	86.83	0	--	Viral polyprotein N-terminal	Viral envelope
c79999_g1	3179.27	0	Replicase polyprotein OS=Cricket paralysis virus	Picornavirus capsid protein	RNA-directed RNA polymerase complex//viral capsid
c79905_g1	180.25	0	--	Rotavirus non-structural protein NSP3	Membrane
c63553_g1	33.79	1.06	--	Herpesvirus latent membrane protein 1 (LMP1)	Integral component of membrane
c74575_g1	51.164	1.52		Picornavirus 2B protein	RNA-directed RNA Polymerase complex//virion
c70839_g1	52.57	4.15	--	Viral A-type inclusion protein repeat	Viral envelope//viral capsid
c75896_g1	215.92	5.53	--	Baculovirus polyhedron envelope protein, PEP, C terminus	Viral capsid
c71244_g1	89.65	5.99		Cytomegalovirus glycoprotein L	Viral envelope
c78209_g2	114.53	31.88	Retrovirus-relat ed Pol polyprotein from	Integrase core domain//Integrase core domain	--

			transposon		
			--	Baculovirus polyhedron envelope protein, PEP, C terminus//Herpesvirus UL6 like	Viral capsid
c66170_g1	361.91	35.93			
c78881_g1	147.39	35.98	--	Chitin binding domain	Viral capsid
c74177_g1	188.23	58.17	--	Picornavirus coat protein (VP4)	Viral capsid
			Retrovirus-relat ed Pol polyprotein from transposon	Integrase core domain//SCAN domain	Transcription factor complex
c81816_g1	783.90	283.31			
c74330_g1	1266.45	360.8		Herpes virus major outer envelope glycoprotein (BLLF1)	Viral envelope
c69319_g1	1985.58	458.79	--	Chitin binding domain	Viral capsid
Down-regulated Genes					
c70393_g1	13.61	76.35	--	Arterivirus GL envelope glycoprotein	Viral envelope
c73935_g2	33.79	283.39	--	Influenza C virus M1 protein	Viral capsid
c77375_g3	108.43	4173.53	--	Carlavirus endopeptidase	RNA-directed RNA polymerase complex
c73420_g1	240.33	690.86	--	Occlusion-derived virus envelope protein ODV-E18	Viral envelope
c51579_g1	306.52	839.47	--	Alphaherpesvirus tegument protein US9	Viral tegument
c79156_g2	477.84	3651.03	--	Pneumovirus matrix protein	Viral envelope
c66739_g1	1626.48	9945.80	Alpha-amylase OS=Pecten maximus	Occlusion-derived virus envelope protein ODV-E18	Viral envelope

217

218

219

Table 3. PFAM analysis of five virus-related genes fragments appeared only in the <i>L.vannamei</i> groups with stunted growth symptom.		
Fragment number	Length of gene	PFAM prediction
F1	6868	RNA helicase/RNA-directed RNA polymerase
F2	4707	Viral Methytransferase/RNA helicase/RNA-directed RNA polymerase
F3	1518	RNA helicase/RNA-directed RNA polymerase/Picornavirus Capsid protein
F4	4392	Picornavirus Capsid protein/Dicistro VP4/CRPV Capsid
F5	9609	RNA helicase/RNA-directed RNA polymerase

220

221

A

B

C

222 **Figure 4. Phylogenetic tree constructed by Neighbour-Joining method using five**
223 **viral gene fragment sequences and genome sequence of viruses in the family**
224 **Picornaviridae.**

225

226 The DEGs between Sick and Health groups were enriched by KEGG pathway, and a total of

10 significantly changed KEGG pathways (q values < 0.05) were detected (Table 4). Some main up-regulated genes in the Sick groups were concerning biological processes of antigen processing and presenting, apoptosis, lysosome, phagosome, rheumatoid arthritis, systemic lupus erythematosus, and legionellosis. 20 genes enriched in antigen processing and presenting (ko04612) were significantly up regulated, and they encoded HSP 70, HSP 90, legumain and cathepsin B, respectively. 19 genes enriched in lysosome (ko04142) were significantly up regulated, and they encoded clathrin light chain A, lysosomal alpha-mannosidase, legumain and carboxypeptidase C, respectively. 17 genes enriched in phagosome (ko04145) were significantly up regulated, and they encoded actin beta/gamma 1, dynein heavy chain 1, tubulin alpha, tubulin beta and cathepsin L, respectively. 9 genes enriched in rheumatoid arthritis (ko05323) were significantly up regulated, and they all encoded cathepsin L. While, many down-regulated genes in the Sick groups mainly involved in the process of starch and sucrose metabolism, Influenza A infection and protein digestion and absorption. 7 genes enriched in protein digestion and absorption metabolism pathway (ko04974) were significantly down regulated, and they encoded trypsin, neprilysin and collagen, respectively. 6 genes enriched in starch and sucrose metabolism (ko00500) were significantly down regulated, and they encoded trehalose 6-phosphate synthase, glucose-6-phosphate isomerase, trehalose 6-phosphate phosphatase, 1,4-alpha-glucan branching enzyme and alpha-amylase, respectively. 8 genes enriched in Influenza A infecting process (ko05164) were significantly down regulated, and they encoded trypsin, HSP 70 and influenza virus NS1A-binding protein, respectively.

Table 4. List of 10 significantly changed KEGG pathways (q values < 0.05) in *L.vannamei* with stunted growth symptom.

Pathway ID	pathway_term	rich_factor	Q value	gene_number
ko04612	Antigen processing and presentation	0.088	2.31E-06	25
ko04210	Apoptosis	0.074	1.60E-05	23
ko04142	Lysosome	0.050	0.004	27
ko04145	Phagosome	0.045	0.020	15
ko05323	Rheumatoid arthritis	0.069	0.023	7
ko05322	Systemic lupus erythematosus	0.083	0.024	10
ko05134	Legionellosis	0.0556	0.024	13
ko04974	Protein digestion and absorption	0.06	0.002	7
ko05164	Influenza A	0.04	0.003	8

ko00500	Starch and sucrose metabolism	0.06	0.003	6
---------	-------------------------------	------	-------	---

248

249 Verification of transcriptomic data by qRT-PCR

250 Eight DEGs were selected for validating the transcriptomic data by real-time PCR analysis. It was
 251 found that the quatitive test of up/down-regulated genes indicated by the qRT-PCR and the
 252 Illumina sequences demonstrated the similar trends. The qRT-PCR analysis confirmed that
 253 transcriptomic data of different expression pattern between Sick and Health shrimp indicated by
 254 Illumina sequences were reliable.

255 **Figure 5. Validation results of RNA-seq profiles by real time PCR.** Target gene
 256 abbreviations are as follows: HSP 70 - heat shock 70kDa protein, GASP 7 - caspase 7,
 257 DHC-1 - dynein heavy chain 1, LDLR - low-density lipoprotein receptor, PRSS - trypsin,
 258 otsA - trehalose 6-phosphate synthase, AMY - alpha-amylase, COL1A - collagen.

259

260 Discussion

261 Illumina based HTS developed rapidly during the last few years and has been applied abroad in
 262 the fields of biological and medical research. It has been employed and proven to be a useful
 263 method to discover viruses that are not detected by existing protocols. HTS has been utilized to
 264 identify novel plant viruses [22-23]; for instance, HTS was used to find novel virus of the etiology
 265 of stem-pitting disease of nectarine [24], viruses associated with decline symptoms of Syrah
 266 grapevines [25], apassion fruit woodiness virus isolate from Australia [26]. Such mining of RNA
 267 sequence data has also been utilized to identify viruses associated with animal diseases [27-28]. In
 268 this study, we demonstrated the potential of the HTS technology platform based comparative
 269 transcriptome analysis to identify pathogenic viruses and its pathogenic mechanism involved in

stunted growth symptom of *L.vannamei*. We divided the shrimp into two groups: shrimp with stunted growth symptom (Sick group) and shrimp from the normal health pond as negative control (Health group). The total RNA was extracted from intestinal contents and tissue of whole shrimp. By utilized Illumina sequencing, 13 virus gene fragments were only detected in Sick group, but not in Health group. Moreover, the DEGs enriched in stunting shrimp were genes involved in antigen processing and presenting, apoptosis, lysosome, phagosome and inflammation, and many previous research indicated that some of these genes were associated with the process of viral infection. Chen et al. (2013) studied the impact of WSSV infection on host gene expression in the hepatopancreas of *L.vannamei* through the use of RNA-Seq, they discovered some genes involved in the processes of animal defense against pathogens such as apoptosis, antigen processing and presentation and MAPK signaling pathway [29]. It should be mentioned that some genes such as cathepsin B, heat shock protein 70, cathepsin L and legumain were also significantly up regulated in the *L.vannamei* with stunted growth symptom. Li et al. (2013) employed next-generation sequencing (NGS) and bioinformatic techniques to observe the transcriptome differences of the shrimp *Fenneropenaeus chinensis* between latent infection stage and acute infection stage of WSSV [30]. They found many genes involved in acute infection of WSSV, including protein digestion and absorption, phagosome, lysosome, and starch and sucrose metabolism. Coincidentally, these genes were also found in our present work and participated in pathological effect. Similar results were illustrated in other previous research as well [31-34]. Therefore, our obtained data will shed a light on the reason for the occurrence the stunted growth symptom of *L.vannamei* in some place of China, and strongly suggested it may originate from viral infection.

In present work, five Picornavirus (family Picornaviridae) viruses related gene fragments were detected in the stunting *L.vannamei*. Picornaviruses are small non-enveloped viruses, 25-30 nm icosahedral, single messenger-active (+)-RNA, and possess a single long ORF [35]. They could infect and replicate in the cytoplasm of cell, and cause a variety of disease in human and animals [36-37]. Until now, only one member of Picornavirus (family Picornaviridae), Taura syndrome virus (TSV), had been reported in penaeid shrimp [38]. In marine crustacea, only three Picornavirus representatives have been identified, all in crabs [39]. The single strain RNA of TSV genome is 10205 nucleotides in length, containing two large open reading frames (ORFs). The ORF1 represented motifs of a helicase, a protease and an RNA-dependent RNA polymerase; another ORF2 revealed a baculovirus IAP repeat (BIR) domain of inhibitor of apoptosis proteins [40]. In our work, the PFAM analysis indicated that five viruses related genes fragments also contain some conservative domain of Picornavirus genes (e.g. RNA helicase, RNA-directed RNA polymerase and Picornavirus Capsid protein). However, the similarity of sequence between the

five viruses and that existing in the database was low (about 20%-40%). Therefore, one interesting result in our work was that some novel Picornavirus may be discovered from *L.vannamei* with stunted growth symptom in China. Their correlation with pathogenesis of stunting of *L.vannamei* require our more in-depth research in the future.

Our another significant finding was that we illustrated some hint about the mechanism of stunting of *L.vannamei* at the molecular level. The fact that viruses infection could result in stunted growth of cell commonly occurred in animal and plant [41-43]. Tang et al. (2016) report that ZIKV virus directly infects human cortical neural progenitor cells with high efficiency, resulting in stunted growth of this cell population and transcriptional dysregulation [44]. Their excellent work solved an urgent global health concern and established a link between ZIKV infection and microcephaly. Qi et al. (2003) discovered a U257A change in the CCR of family Pospiviroidae converted the intermediate strain PSTVdInt to a lethal strain that caused severe growth stunting and premature death of infected plants [45]. Many factor, such as parasite *Enterocytozoon hepatopenaei* (EHP) [46], nutrient elements, and viruses infection could cause slow growth of shrimp. Among them, viruses infection was the most important one and attract widespread attention in the last several years. Although many viral infection (e.g. IHHNV, LSNV, MBV, HPV) were discovered leading to the slow or stunting growth of *Penaeus* shrimp in previous research, the mechanism has been not illustrated until now [6-8]. In our study, the result that some viruses were only detected in *L.vannamei* with stunted growth symptom made us hypothesize that this disease may have correlation with viral infection. This hypothesis was further supported by gene expression profiles, since many genes of shrimp, which participate in the infecting process of viruses, were up-regulated in sick shrimp with stunting symptom. Moreover, we have found the reason why *L.vannamei* infected by viruses could not grow up to the normal size. This may originated from the reason that the shrimp infected by viruses could not metabolize protein and sugar (starch and sucrose) very well. The correlation between viral infection and metabolism of protein and sugar has been well studied in plant [47-50]. However, there were not any information about this in animal. Anyway, the depressed metabolism of protein and sugar could undoubtedly inhibit the growth of cell in plant or animal [51-53].

Although our research provided some useful information about pathogenic factor and mechanism on the stunted growth symptom of *L.vannamei*, more work should be carried out in the future. HTS based comparative transcriptome analysis was a helpful method for discovering novel viruses infecting shrimp, but what can not be neglected, this method was more effective in detecting RNA viruses than DNA viruses. We are not sure about whether DNA viruses are significant in the occurrence of this shrimp disease. Thus, metagenomic assay will be utilized for

further work on the identification and proven of other pathogenic factor [54-55]. Moreover, the cDNA probe and RT-PCR methods will be useful tools for detecting Picornavirus in more sick shrimp samples and its correlation with the emerging of stunted growth symptom. Anyway, our work was important for the establishment of more testing standards and maintain SPF shrimp stocks.

Acknowledge

Supported by the National Natural Science Foundation of China (31560727); Guangxi Natural Science Foundation (2014GXNSFBA118135, 2018JJA130187).

Reference

1. Sukgyoung Kim, Pang Z , Hyung Jhel Seo (2014) Effect of bioflocs on growth and immune activity of Pacific white shrimp, *Litopenaeus vannamei* Postlarvae. Aquaculture Research 45: 362-371.
2. Ai HS, Huang YC, Li SD, Weng SP, Yu XQ, He JG (2008) Characterization of a prophenoloxidase from hemocytes of the shrimp *Litopenaeus vannamei* that is down regulated by white spot syndrome virus. Fish and Shellfish Immunology 25: 28-29.
3. Naylor RL, Goldburg RJ, Primavera JH, Kautsky N, Beveridge MC (2000) Effect of aquaculture on world fish supplies. Nature 405: 1017-1024.
4. Thitamadee S, Prachumwat A, Srisala J, Jaroenlak P, Salachan PV, Sritunyalucksana K, Itsathitphaisarn O (2016) Review of current disease threats for cultivated penaeid shrimp in Asia. Aquaculture 452: 69-87.
5. Shi H, Xie J, Wang G, & Xu W (2017) Pathogenicity of *Penaeus vannamei* slow growth syndrome in intensively cultured penaeid shrimp in zhoushan. Journal of Fishery Sciences of China, 24: 387-394.
6. Flegel TW, Thamavit V, Pasharawipas T, Alday-Sanz V (1999) Statistical correlation between severity of hepatopancreatic parvovirus infection and stunting of farmed black tiger shrimp (*Penaeus monodon*). Aquaculture 174: 197-206.
7. Rai P, Pradeep B, Karunasagar I, & Karunasagar I (2009) Detection of viruses in *Penaeus monodon* from India showing signs of slow growth syndrome. Aquaculture 289: 231-235.
8. Flegel TW, Nielsen L, Thamavit V, Kongtim S, & Pasharawipas T (2004) Presence of multiple viruses in non-diseased, cultivated shrimp at harvest. Aquaculture 240: 55-68.
9. Chayaburakul K, Nash G, Pratanpipat P, Sriurairatana S, & Withyachumnarnkul B (2004) Multiple pathogens found in growth-retarded black tiger shrimp *Penaeus monodon* cultivated

- in Thailand. Diseases of aquatic organisms 60: 89-96.
10. Pratoomthai B, Sakaew W, Sriurairatana S, Wongprasert K, Withyachumnarnkul B (2008) Retinopathy in stunted black tiger shrimp *Penaeus monodon* and possible association with Laem-Singh virus (LSNV). Aquaculture, 284: 53-58.
11. Panphut W, Senapin S, Sriurairatana S, Withyachumnarnkul B, Flegel TW (2011) A novel integrase-containing element may interact with Laem-Singh virus (LSNV) to cause slow growth in giant tiger shrimp. BMC veterinary research 7: 18.
11. Saksmerprome V, Thammasorn T, Jitrakorn S. (2013) Using double-stranded RNA for the control of Laem-Singh Virus (LSNV) in Thai *P.monodon*. Journal of Biotechnology 164: 449-453.
13. Qin Z, Babu VS, Wan Q (2018) Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei*) challenged by, *Vibrio parahaemolyticus*, reveals unique immune-related genes. Fish & Shellfish Immunology 2018: S1050464818301475.
14. Chen K, Li E, Li T, Xu C, Wang X, Lin H, Chen L (2015) Transcriptome and molecular pathway analysis of the hepatopancreas in the Pacific white shrimp *Litopenaeus vannamei* under chronic low-salinity stress. PLoS One 10: e0131503.
15. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Chen Z.(2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644-652.
16. Finn RD, John T, Jaina M, Coghill PC, Stephen John S, Hans-Rudolf H (2008) The pfam protein families database. Nucleic Acids Research 36: 281-288.
17. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biology 5: R7.
18. Li B, Dewey C (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323.
19. Young MD, Wakefield MJ, Smyth GK, & Alicia O (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology 11: R14.
20. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M (2007) KEGG for linking genomes to life and the environment. Nucleic Acids Research 36(Database): D480-D484.
21. Mao X, Cai T, Olyarchuk JG, Wei L (2005) Automated genome annotation and pathway identification using the kegg orthology (ko) as a controlled vocabulary. Bioinformatics, 21: 3787-3793.

- 406 22. Prabha K, Baranwal VK, Jain RK (2013) Applications of next generation high throughput
407 sequencing technologies in characterization, discovery and molecular interaction of plant
408 viruses. Indian Journal of Virology 24: 157-165.
- 409 23. Massart S, Olmos A, Jijakli H, Candresse T (2014) Current impact and future directions
410 of high throughput sequencing in plant virus diagnostics. Virus Research 188: 90-96.
- 411 24. Villamor DEV, Mekuria TA, Pillai SS, Eastwell KC (2016) High-throughput sequencing
412 identifies novel viruses in nectarine: insights to the etiology of stem-pitting
413 disease. Phytopathology 106: 519-527.
- 414 25. Rwahnih MA, Daubert S, Golino D, Rowhani A (2009) Deep sequencing analysis of
415 RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection
416 that includes a novel virus. Virology 387: 395-401.
- 417 26. Wylie SJ, Jones MGK (2011) The complete genome sequence of a passion fruit woodiness
418 virus isolate from Australia determined using deep sequencing, and its relationship to other
419 potyviruses. Archives of Virology 156: 479-482.
- 420 27. Sparks ME, Gundersen-Rindal DE (2011) The *lymantria dispar* iplb-ld652y cell line
421 transcriptome comprises diverse virus-associated transcripts. Viruses 3: 2339-2350.
- 422 28. Dresang LR, Teuton JR, Feng H, Jacobs JM, Camp DG, & Purvine SO (2011) Coupled
423 transcriptome and proteome analysis of human lymphotropic tumor viruses: insights on the
424 detection and discovery of viral genes. BMC Genomics 12: 625.
- 425 29. Chen X, Zeng D, Chen X, Xie D, Zhao Y, Yang C (2013) Transcriptome analysis of
426 *litopenaeus vannamei* in response to white spot syndrome virus infection. PLoS ONE 8:
427 e73218.
- 428 30. Li S, Zhang X, Sun Z, Li F, Xiang J (2013) Transcriptome analysis on chinese shrimp
429 *Fenneropenaeus chinensis* during WSSV acute infection. PLoS ONE 8: e58627.
- 430 31. Zeng D, Chen X, Xie D, Zhao Y, Yang C, Li Y, & Wang H (2013) Transcriptome analysis
431 of
432 Pacific white shrimp (*Litopenaeus vannamei*) hepatopancreas in response to Taura syndrome
433 Virus (TSV) experimental infection. PloS one 8 : e57515.
- 434 32. Qin Z, Babu VS, Wan Q, Zhou M, Liang R, Muhammad A, Lin L (2018) Transcriptome
435 analysis of Pacific white shrimp (*Litopenaeus vannamei*) challenged by *Vibrio*
436 *parahaemolyticus* reveals unique immune-related genes. Fish & shellfish immunology 77 :
437 164-174.
- 438 33. Sookruksawong S, Sun F, Liu Z, & Tassanakajon A (2013) RNA-Seq analysis reveals genes
439 associated with resistance to Taura syndrome virus (TSV) in the Pacific white shrimp

- 440 *Litopenaeus vannamei*. Developmental & Comparative Immunology 41 : 523-533.
- 441 34. Xue S, Liu Y, Zhang Y, Sun Y, Geng X, Sun J (2013) Sequencing and de novo analysis of
- 442 the hemocytes transcriptome in *Litopenaeus vannamei* response to white spot syndrome virus
- 443 infection. PLoS One 8 : e76718.
- 444 35. Witwer C, Rauscher S, Hofacker IL, & Stadler PF (2001) Conserved RNA secondary
- 445 structures in Picornaviridae genomes. Nucleic acids research 29: 5079-5089.
- 446 36. Kapoor A, Victoria J, Simmonds P, Slikas E, & Delwart E (2008) A highly prevalent and
- 447 genetically diversified picornaviridae genus in south asian children. Proceedings of the
- 448 National Academy of Sciences 105: 20482-20487.
- 449 37. Kim MC, Kwon YK, Joh SJ, Lindberg AM, Kwon JH, & Kim JH (2006) Molecular analysis
- 450 of duck hepatitis virus type 1 reveals a novel lineage close to the genus parechovirus in the
- 451 family picornaviridae. Journal of General Virology 87: 3307-3316.
- 452 38. Bonami JR, Hasson KW, Mari J, Poulos BT, & Lightner DV (1997) Taura syndrome of
- 453 marine penaeid shrimp: characterization of the viral agent. Journal of General Virology 78:
- 454 313-319.
- 455 39. Bonami JR, & Zhang S (2011) Viral diseases in commercially exploited crabs: a
- 456 review. Journal of Invertebrate Pathology 106: 0-17.
- 457 40. Mari J, Poulos BT, Lightner DV, & Bonami JR (2002) Shrimp Taura syndrome virus:
- 458 genomic characterization and similarity with members of the genus Cricket paralysis-like
- 459 viruses. Journal of general virology 83: 915-926.
- 460 41. Whenham RJ, & Fraser RSS (1990) Plant growth regulators, viruses and plant growth. In
- 461 Recognition and response in plant-virus interactions (pp. 287-310). Springer, Berlin,
- 462 Heidelberg.
- 463 42. Burton RA, Gibeaut DM, Bacic A, Findlay K, Roberts K, Hamilton A, Fincher GB (2000)
- 464 Virus-induced silencing of a plant cellulose synthase gene. The Plant Cell 12: 691-705.
- 465 43. Fraser RSS, Whenham RJ (1982) Plant growth regulators and virus infection: a critical
- 466 review. Plant Growth Regulation 1: 37-59.
- 467 44. Tang H, Hammack C, Ogden SC, Wen Z, Qian X, Li Y, Christian KM (2016) Zika virus
- 468 infects human cortical neural progenitors and attenuates their growth. Cell stem cell 18:
- 469 587-590.
- 470 45. Qi Y, Ding B (2003) Inhibition of cell growth and shoot development by a specific
- 471 nucleotide sequence in a noncoding viroid RNA. The Plant Cell 15: 1360-1374.
- 472 46. Kesavan K, Mani R, Toshiaki I, Sudhakaran R (2017) Quick report on prevalence of
- 473 shrimp microsporidian parasite *Enterocytozoon hepatopenaei* in India. Aquaculture

Research 48: 3980-3984.

47. Técsi LI, Maule AJ, Smith AM, Leegood RC (1994) Metabolic alterations in cotyledons of Cucurbita pepo infected by cucumber mosaic virus. Journal of Experimental Botany 45: 1541-1551.
48. Almon E, Horowitz M, Wang HL, Lucas WJ, Zamski E, Wolf S (1997) Phloem-specific expression of the tobacco mosaic virus movement protein alters carbon metabolism and partitioning in transgenic potato plants. Plant Physiology 115: 1599-1607.
49. Yan SL, Lehrer AT, Hajirezaei MR, Springer A, & Komor E (2008) Modulation of carbohydrate metabolism and chloroplast structure in sugarcane leaves which were infected by Sugarcane Yellow Leaf Virus (SCYLV). Physiological and Molecular Plant Pathology 73: 78-87.
50. Arias MC, Lenardon S, Taleisnik E (2003) Carbon metabolism alterations in sunflower plants infected with the Sunflower chlorotic mottle virus. Journal of Phytopathology 151: 267-273.
51. Torrissen KR, & Shearer KD (1992) Protein digestion, growth and food conversion in Atlantic salmon and Arctic charr with different trypsin-like isozyme patterns. Journal of Fish Biology 41: 409-415.
52. Tang GQ, Sturm A (1999) Antisense repression of sucrose synthase in carrot (Daucus carota L.) affects growth rather than sucrose partitioning. Plant molecular biology 41: 465-479.
53. Coleman HD, Ellis DD, Gilbert M, Mansfield SD (2006) Up-regulation of sucrose synthase and UDP-glucose pyrophosphorylase impacts plant growth and metabolism. Plant Biotechnology Journal 4: 87-101.
54. Foulongne V, Sauvage V, Hebert C, Dereure O, Cheval J, Gouilh MA (2012) Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. PLoS ONE 7: e38499.
55. Hall RJ, Mily LM, Jing W, Xiaoyun R, Moore NE, Brooks CR (2013) Metagenomic detection of viruses in aerosol samples from workers in animal slaughterhouses. PLoS ONE 8: e72226.