TRANSCRIPTOME ANALYSIS OF IMMUNE RESPONSES IN *Macrobrachium rosenbergii* AGAINST WHITE SPOT SYNDROME VIRUS AND *Vibrio parahaemolyticus*

RAMARAO A/L SERIRAMALU

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

TRANSCRIPTOME ANALYSIS OF IMMUNE RESPONSES IN *Macrobrachium rosenbergii* AGAINST WHITE SPOT SYNDROME VIRUS AND *Vibrio parahaemolyticus*

RAMARAO A/L SERIRAMALU

THESIS SUBMITTED IN FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR

OF PHILOSOPHY

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

*(Please delete this part) Depending on the medium of thesis/dissertation, pick either the English or Bahasa Malaysia version and delete the other.

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: (I.C/Passport No:)

Registration/Matric No:

Name of Degree:

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Field of Study:

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

The Malaysian giant freshwater prawn, *Macrobrachium rosenbergii*, an economically important crustacean worldwide are being affected by pathogens such as the white spot syndrome virus (WSSV) and *Vibrio* species. The information regarding the regulation of innate immune system in this species is lacking, however it is critical in providing solutions to control and minimize the loss of production due to diseases. An investigation regarding immune response of *M. rosenbergii* towards *V. parahaemolyticus* and WSSV infection were undertaken by performing a transcriptome profiling of *M. rosenbergii* hepatopancreas infected with these two pathogens by using the 'Next Generation' sequencing method (Illumina HiSeq^{TM}2000). The reads were assembled de *novo* and were annotated against databases using the BLASTX search (E-value $<10^{-5}$). A total of 22,455 unigenes (34.86% of all unigenes) were successfully annotated. The unigene differential expression analysis by FPKM (Fragments Per kb per Million fragments) method revealed 14,569 unigenes that were differentially expressed in *Vibro*-infected shrimp and 14,416 in WSSV-infected ones when compared to the controls. Several differentially expressed unigenes were shown to be involved in various animal immune functions. Interestingly, we also found 79 immune related unigenes that were up-regulated in *M. rosenbergii* when independently infected with each of the pathogen. Among those unigenes is the STAT gene, a key component of the Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway. We further characterised the full-length cDNA of STAT (designated as MrSTAT) obtained from the transcriptome database by using various bioinformatics tools. Time course analysis of the transcription level of MrSTAT in hepatopancreas after WSSV and *V.parahaemolyticus* challenge showed significant ($p < 0.05$) increase expression, highest are at 6 hours for WSSV- infected and 24 hour for *V.parahaemolyticus*- infected ones. All these results may imply the potential role of JAK/STAT pathway in the immune response against bacteria and virus in *M. rosenbergii*.

ABSTRAK

Macrobrachium rosenbergii merupakan spesies ternakan aquakultur yang penting di serata dunia. Namun demikian, jumlah ternakan *M. rosenbergii* mengalami kemerosotan disebabkan jangkitan white spot syndrome virus (WSSV) and spesies *Vibrio*. Informasi berkenaan regulasi sistem imun semula jadi bagi spesies ini kurang diketahui. Oleh itu, adalah penting untuk mengkaji sistem imun spesies tersebut untuk mencari penyelesaiannya. Bagi kajian ini, kami memprofil transcriptome daripada hepatopancreas *M. rosenbergii* dijangkiti *V. parahaemolyticus* dan WSSV menggunakan kaedah penjujukan 'Generasi Baru' (Illumina HiSeq^{TM}2000). Data jujukan diatur secara *de novo* dan dianotasi terhadap pelbagai pangkalan data dengan menggunakan carian BLASTX (E-nilai $<10^{-5}$). Sebanyak 22.455 unigenes (34.86% daripada semua unigenes) berjaya dianotasi. Analisis expresi unigene melalui kaedah FPKM melaporkan sebanyak 14.569 unigenes menunjukkan perbezaan expresi bagi *M. rosenbergii* dijangkiti dengan *V. parahaemolyticus* manakala hanya 14.416 unigene bagi *M. rosenbergii* dijangkiti WSSV apabila dibandingkan dengan *M. rosenbergii* kawalan. Beberapa unigenes yang berbeza expresi terlibat dalam pelbagai fungsi imun haiwan. Menariknya, kami juga mendapati sebanyak 79 unigenes imun mengalami peningkatan expresi apabila *M.rosenbergii* dijangkiti dengan kedua-dua patogen secara berasingan. Antara unigenes tersebut, gen STAT, merupakan komponen utama bagi sistem kinase Janus (JAK) / transduser isyarat dan pengaktif transkripsi (STAT). Kami menggunakan analisis bioinformatik untuk mendapatkan maklumat lengkap berkenaan cDNA STAT (MrSTAT). Tahap expresi MrSTAT dalam hepatopancreas *M.rosenbergii* selepas jangkitan WSSV dan *V.parahaemolyticus* menunjukkan peningkatan signifikan tertinggi ($p<0.05$) pada 6 jam untuk jangkitan WSSV dan 24 jam untuk jangkitan *V. parahaemolyticus*. Semua keputusan ini boleh membayangkan peranan berpotensi sistem JAK / STAT dalam tindak balas imun terhadap bakteria dan virus bagi *M. rosenbergii*.

ACKNOWLEDGEMENTS

A moment of prayers is extended to the Almighty for I am very thankful to HIM for giving me the opportunity to complete this research within the time given. I would like to express my sincere gratitude and deep appreciation to these people for their contribution to make my theses as successful.

I am grateful beyond words to Associate. Prof. Dr Subha Bhassu for accepting me to conduct a research on this title. Her guidance, support as well as positive comments kept me on track and focused.

Many thanks to my co-supervisor, Assoc. Prof. Dr Sharifah Binti Syed Hassan, for her kind encouragement, guidance and support for my work.

Heartiest thanks to Ministry of Higher Education (MOHE), under the MyPhd scheme for sponsoring my studies throughout the whole Phd's candidature.

Not forgetting my dear lab mates. Our friendship will be cherished forever.

Not forgetting also in the list, with deepest appreciation are my parents and siblings for their love and encouragement while undertaking my research.

TABLE OF CONTENTS

Chapter Six: STAT GENE CHARACTERIZATION AND EXPRESSION IN *MACROBRACHIUM ROSENBERGII* AGAINST *VIBRIO PARAHAEMOLYTICUS* AND WHITE SPOT SYNDROME VIRUS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

M. rosenbergii	*Macrobrachium rosenbergii*
P. monodon	*Penaeus monodon*
L. vannamei	*Litopenaeus vannamei*
F. chinensis	*Fenneropenaeus chinensis*
M. japonicas	*Marsupenaeus japonicas*
D. pulex	*Daphnia pulex*
T. castaneum	*Tribolium castaneum*
C. elegans	*Caenorhabditis elegans*
P. h. corporis	*Pediculus humanus corporis*
H. sapiens	*Homo sapiens*
V. parahaemolyticus	*Vibrio parahaemolyticus*
V. harveyi	*Vibrio harveyi*
WSSV	White spot syndrome virus
IHHNV	Infectious Hypodermal And Hematopoietic Necrosis Virus
TSV	Taura Syndrome Virus
KEGG	Kyoto Encyclopaedia of Genes and Genomes
FPKM	Fragments Per kb per Million fragments
GO	Gene Ontology
COG	Cluster of Orthologous Groups of Proteins
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
DNA	Deoxyribonucleic Acid
cDNA	Complementary Deoxyribonucleic acid
mRNA	Messenger RNA

PCR	Polymerase Chain Reaction
RT-PCR	Reverse Transcriptase PCR
qRT-PCR	Quantitative real time PCR
dNTP	Deoxynucleotide
Bp	Base pair
Aa	Amino acid
BLASTx	Basic Local Alignment Tool
ORF	Open reading frame
UTR	Untranslated region
NCBI	National Centre for Biotechnology Information
SRA	Short Read Archive
DEPC	Diethylpyrocarbonate
Ng	Nano gram
µM	Micro molar
L	Litre
Ml	Mililitre
NaCl	Sodium chloride
OIE	World Organization for Animal Health
EF1-A	Elongation factor 1-alpha
FDR	False Discovery Rate
DEG	Differentially expressed genes
NR	Non redundant
EST	Expressed sequence tag
SSH	Suppression subtractive hybridization
DH	Differential hybridization
ANOVA	Analysis of Variance
NGS	Next generation sequencing

ROS	Reactive oxygen species
MAPK	Mitogen-activated protein kinase
IMD	Immune Deficiency
TGICL	TIGR Gene Indices clustering tools

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

The giant river prawn, *Macrobrachium rosenbergii* (*M. rosenbergii*) is a species of freshwater shrimp. Although *M. rosenbergii* is considered a freshwater species, the early stage of development of the shrimp (larva stage) takes place in brackish water. Once the individual shrimp has grown beyond the planktonic stage and become a juvenile, it will live entirely in freshwater. This species has emerged as an economically important and a commercial species in Asia due to high demand of its delicious taste (New, 2008). Thus freshwater prawn farming has gain pace dramatically in the last decade around the world and also in Malaysia.

The major problem in *M. rosenbergii* aquaculture industry are the occurrence of two main acute epizootics diseases known as "White spot disease" and "Vibriosis" which contributes to massive mortality among prawns, leading to economic losses in commercial aquaculture (Walker & Mohan, 2009). The "White spot disease" is caused by a virus known as White Spot Syndrome Virus (WSSV) whereas the "Vibriosis" is due to infection by *Vibrio* species.

WSSV, which was first reported in Fujian, China, in 1992, is classified under the genus *Whispovirus* of the virus family *Nimaviridae* (Pradeep *et al.*, 2012). This deadly virus has an enveloped, non-occluded, rod-shaped capsid that contains a circular, double-stranded DNA of about 300 kb (van Hulten *et al.*, 2001). The virus has a diverse range of potential hosts such as marine shrimps, crayfish, crab, spiny lobsters, insects and freshwater prawns (Chang *et al.*, 1998). The clinical signs of WSSV are; reduction in food consumption, lethargy, loose cuticle and often reddish discolouration, and the presence of white spots of 0.5 to 2.0 mm in diameter on the inside surface of the carapace,

appendages and cuticle over the abdominal segments (Chang *et al*., 1998; Rajendran *et al.*, 1999).

On the other hand *Vibrio* is a Gram-negative halophilic bacterium found abundantly in marine and estuarine environments (Ramesh *et al.*, 1990; Thompson *et al.*, 2004). Among the different species, *Vibrio parahaemolyticus* has emerged as an important pathogen for *M. rosenbergii* (Khuntia *et al.*, 2008). Several other marine shrimps such as *Penaeus monodon*, *Penaeus japonicas* and *Litopenaeus vannamei* were also found to be susceptible to *Vibrio* infection. *M. rosenbergii* suffering from Vibriosis may appear black in colour on the carapace, with red discolouration of the exoskeleton and loss of appendages within six days, leading to an 80% mortality rate (Nash *et al.,* 1992).

M. rosenbergii possesses an innate immune system that functions as the first line defence mechanism against pathogen infection (Iwanaga & Lee, 2005). Unlike other vertebrates which contain adaptive immune system to fight pathogenic infection, there are no effective chemicals or drugs to boost this shrimp's resistance against diseases. The common strategies used in combating shrimp diseases include quarantine and environmental management (Bachère, 2000). These approaches are non-specific in combating infectious diseases and cannot improve the shrimp's ability to cope with future infection even with the same pathogen.

Another alternative approach available to fight pathogenic infection in shrimp is by studying into the host pathogen interaction. To understand the pathogenesis of any disease, knowledge of the interactions between pathogen and host is vital. Host-pathogen interactions may provide immune responses against the invader or could also result in changes in the expression levels of host genes that favour virus replication and bacteria

growth (Waddell *et al.*, 2007; Wise *et al.*, 2007). The identification of these molecules (genes) might provide a promising strategy for protection and treatment of diseases.

A cutting edge technology has emerged recently, that has revolutionised modern genomics research, known as 'Next Generation Sequencing' (NGS), was introduced in 2004. This platform provides a faster and cheaper way to sequence large amounts of data with greater depth and breadth. At present, the three dominant commercial platforms available are the Roche 454 Genome Sequencer, the Illumina Genome Analyzer, and the Life Technologies SOLiD System (Zhou *et al*., 2010). This sequencing technology has a wide range of applications, such as genome sequencing, metagenomics, epigenetics, discovery of non-coding RNAs, protein-binding sites and transcriptomics analysis (MacLean *et al.*, 2009). Transcriptome analysis using the NGS approach has been successfully utilised to identify immune-related genes in marine shrimps against the Taura Syndrome Virus (TSV) and WSSV (Li *et al.*, 2013; Zeng *et al.*, 2013).

At present, there is limited discovery related to immune genes in *M. rosenbergii* which primarily involved in WSSV and *Vibrio* infection. Therefore in this study, the NGS sequencing technology (transcriptomics analysis) was applied to obtain a detail overview of the immune-related genes involved in WSSV and *V. parahaemolyticus* infection which leads to a better understanding regarding host-pathogen interaction between *M. rosenbergii* with these two main pathogens. We further compared the differentially expressed immune related genes caused by WSSV or *Vibrio* infection with with a particular focus on the role of immune molecules of different signalling pathways Lastly, we reported the bioinformatics analysis of STAT gene, a key component of Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway and its expression pattern in hepatopancreas following challenge with WSSV and *V. parahaemolyticus*

using qRT-PCR to understand its role in antiviral and antibacterial immunity of *M. rosenbergii*.

1.1 Objectives of study

The study is done to fulfil the following objectives:

1. To perform a transcriptomics analysis on *M. rosenbergii* hepatopancreas experimentally challenged with *V. parahaemolyticus* and White Spot Syndrome Virus (WSSV). (Chapter 3 and 4)
2. To identify candidate immune-related genes involved in *V. parahaemolyticus* and WSSV infection which can provide a better insight into the host-pathogen interaction between *M. rosenbergii* and these two pathogens. (Chapter 3 and 4)
3. To compare differentially expressed immune related genes caused by *Vibrio* or WSSV infection with a particular focus on the role of immune molecules of different signalling pathways. (Chapter 5)
4. To characterize the STAT gene and investigate its expression in *M. rosenbergii* following challenge with WSSV and *V. parahaemolyticus* at high-intensity sampling points by using qRT-PCR. (Chapter 6)

CHAPTER 2: LITERATURE REVIEW

2.1 *Macrobrachium rosenbergii*

There are about 150 species of *Macrobrachium* being reported in the world, of which 49 are available commercially. In Asia and Pacific region alone, twenty-seven commercial *Macrobrachium* species were found. *Macrobrachium* species mostly lives in freshwater but a few species were found in brackishwater in the mouths of rivers (New, 1990). *Macrobrachium rosenbergii* also known as the Malaysian giant freshwater prawn is spread abundantly in the tropical and subtropical waters of the Indo- Pacific region such as Malaysia, Thailand, the Philippines, India, Sri Lanka, Bangladesh, Myanmar, Indonesia and Vietnam (New, 2002; Wisely, 1984).

M. rosenbergii like other *Macrobrachium* species lives in freshwater, in ponds, rivers, lakes, ditches, canals, depressions, low-lying floodplains and river mouths. This species spend their early life in brackishwater before moving upstream, entering lakes and even paddy fields. This pattern of migration is observed in *M. rosenbergii* and also in other species of *Macrobrachium* (Chowdhury *et al.*, 1993).

Among all *Macrobrachium* species, *M. rosenbergii* is the most popular species used in research and has been introduced to many new countries for commercial culture. Fujimura and Okamoto (1970) were successful in producing post-larvae (FL) of *M. rosenbergii* in large numbers in Hawaii which later led to large-scale farming of this prawn in many parts of the world which includes Hawaii, Honduras. Mauritius, Taiwan, Thailand, Philippines and Malaysia (Fujimura *et al.*, 1970; New, 2002). It has been reported that the contribution of shrimp industry has touched approximately 3.4 million

tons per year worldwide, and Asia has been listed as a major contributor (FAO, 2008). The worldwide annual production of the giant freshwater prawn has begun to surpass 200,000 tons from the year 2002 (FAO, 2008).

However, emergence of several new viral pathogens have successively contributed to mass mortalities and threatening the sustainability of aquaculture industry in Asia and the Americas since 1981 (Walker & Mohan, 2009). With expansion of the shrimp industry, the number of viral pathogens of shrimp has expended as well, reaching a new high of 22 viruses (Walker & Mohan, 2009). Among the viruses discovered in shrimp, white spot disease caused by white spot syndrome virus (WSSV) has had the most devastating impact on shrimp culture and continues to harm a sustainable shrimp farming (Flegel, 2012).

2.1.1 Taxonomic Position of *Macrobrachium rosenbergii*

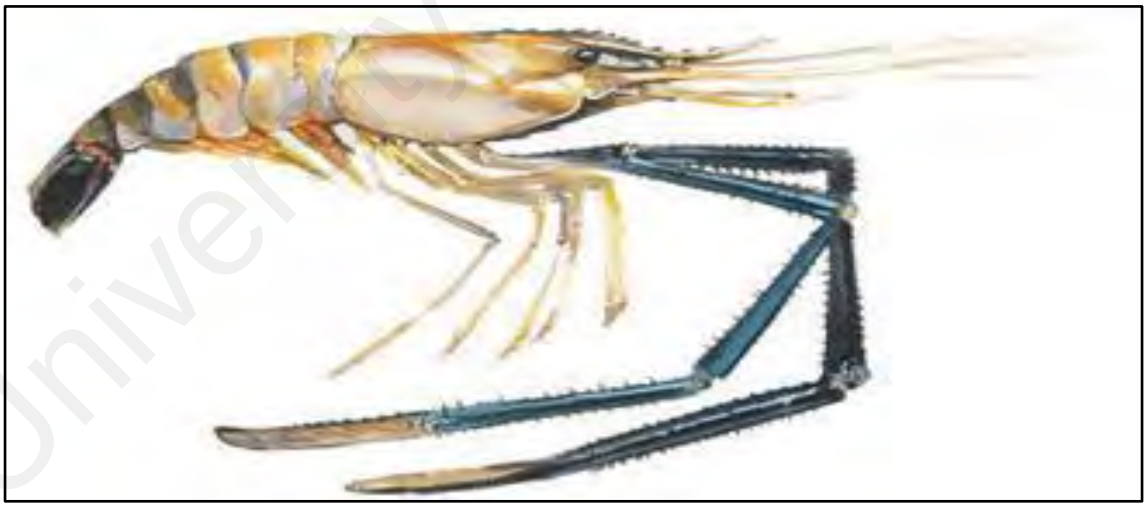


Figure 2.1: Image of *Macrobrachium rosenbergii*

*****Graphic structure of *M. rosenbergii* is adapted from Jayachandran, K. (2001)

Kingdom:	Animalia
Phylum:	Arthropoda
Subphylum:	Crustacea
Class:	Malacostraca
Order:	Decapoda
Infraorder:	Caridea
Family:	Palaemonidae
Species:	Macrobrachium rosenbergii

2.1.2 Life Cycle

The life cycle of *M. rosenbergii* consists of four stages: egg, larva, juvenile and adult (Figure 2.2). They undergo moulting, however the number of moults and the durations of intermoults are not fixed and depends on environmental factors such as temperature and the availability of food. The mating process of *M. rosenbergi* takes place all year round in the natural environment. A successful mating takes place between soft shelled females and hard shelled males. The number of eggs produced depending on the size of the female. Generally, a fully grown female weighs between 50-100 g lays around 50,000-100,000 eggs (New, 2002, 2008).

The larvae generally hatches from the egg during the night. At this stage, it requires brackish water to survive. In the wild, they generally consume zooplankton, small insects and larvae of other aquatic invertebrates. Larvae found in hatchery require a minimum of 26 days to metamorphose into post-larvae (PL). Postlarvae adapt to a wide

range of salinity but freshwater is their preference habitat. After metamorphosis, the PL travels against the current and move towards lakes and rivers. At this stage, they discard their planktonic habit and turn into omnivorous, feeding on aquatic insects and their larvae, fruit, small Mollusca and crustacean, fish flesh, slaughterhouse waste and animal remains. They basically move by crawling and generally swim rapidly with their dorsal side uppermost (New, 2002, 2008).

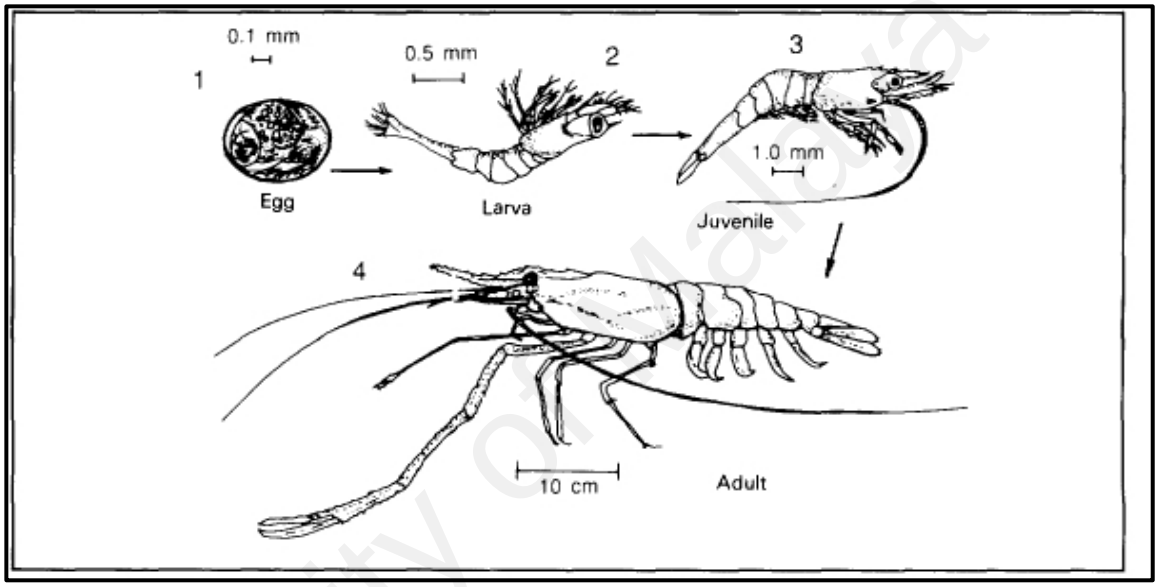


Figure 2.2: Life cycle of *M. rosenbergii*

* Graphic structure of *M. rosenbergii* is adapted from Jayachandran, K. (2001)

2.1.3 *Macrobrachium rosenbergii* Distinguishing Characters

The largest recorded size for males and females *M. rosenbergii* were 33 cm and 29 cm, respectively. Generally, the matured male prawns are bigger and heavier than the females. *M. rosenbergii*'s cephalothorax has five distinct segments in the head and eight at the thoracic region. The abdomen consists of six distinct segmented movable terga with each has a pair of pleopods (swimming leg) (Figure 2.3) (Jayachandran, 2001).

The cephalothorax part consists of a pair of stalked eyes, 2 pairs of antennae, 3 pairs of jaws, 3 pairs of maxillipids and 5 pairs of walking legs. *M. rosenbergii* can be distinguished from penaeid shrimp by having the second and largest abdominal tergum laterally overlaps the first and the third. The dorsal surface of *M. rosenbergii* abdominal carapace is smooth and rounded. The rostrum is slender, curved upwards which extends beyond the antennal scale and has 11-14 dorsal teeth and 8-10 on the ventral side. The mature male prawns' head and second pair of chelipeds is proportionately larger than in the female. When compared to other *Macrobrachium* species, the left and right chellipeds of *M. rosenbergii* are equal in size. They grow extremely long and reach well beyond the tip of the rostrum in adult males. The genital pores of the males are situated at the base of fifth walking leg whereas in females is located at the base of the third walking legs. The tip of the telson extends beyond the posterior telson spines (Jayachandran, 2001).

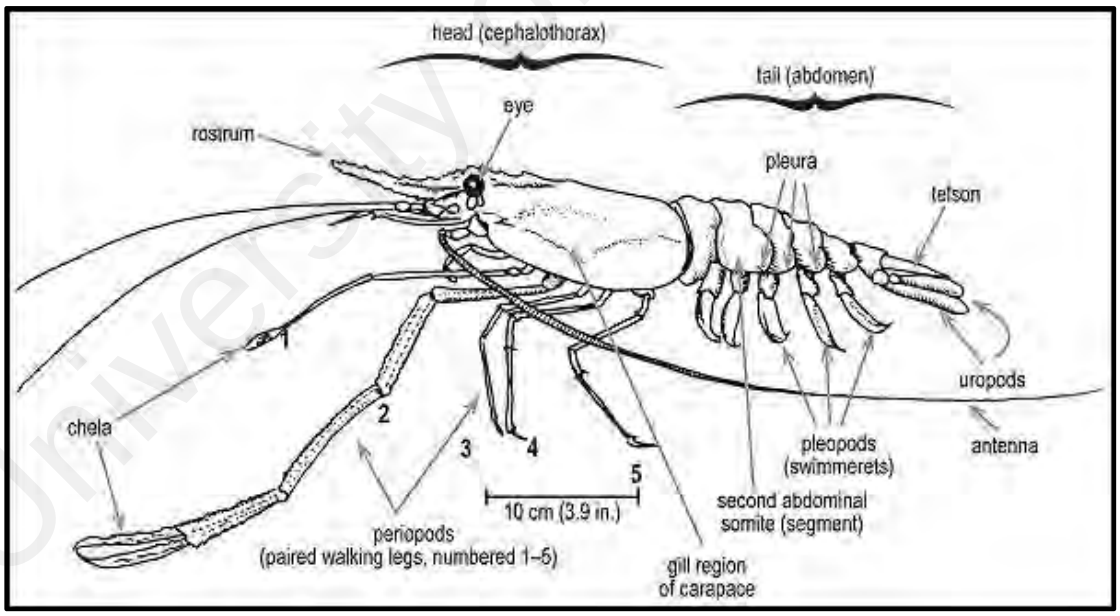


Figure 2.3: Features of *M. rosenbergii*

* Graphic structure of *M. rosenbergii* is adapted from Jayachandran, K. (2001)

2.1.4 Morphotypes

M. rosenbergii males exists in three different morphotypes. The beginning stage is called "small male" (SM), where the males have the smallest size and possesses nearly translucent claws. When conditions permits, the small males will metamorphose into "orange claws" (OC), having a large orange claws on their second chelipeds with a length of 0.8 to 1.4 to their body size. The final form is referred as 'blue claw' males (BC) which is characterised by the presence of blue claws and the second chelipeds may become twice as long as their body (Cohen *et al.*, 1981; Kuris *et al.*, 1987).

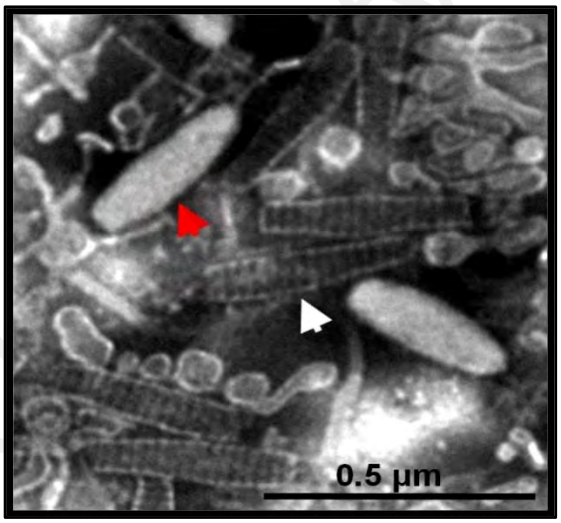
A hierarchy system is being strictly followed by male *M. rosenbergii*. The BC males dominate the OCs whereas the OCs does similarly to SMs. The growth of SMs and the metamorphosis of OCs to BCs are being delayed by the presence of BC males. The different stages of male are sexually active, will mate with females who have undergone their pre-mating moult. The BC males protect the female until their shell has hardened which is not observed similarly in OCs and SMs (Barki *et al.*, 1992).

2.2 White Spot Syndrome Virus

This deadly virus was first described in China and Taiwan in 1992, spread rapidly throughout East, South-East and South Asia creating a panzootic by 1994. A second panzootic wave were created when this virus was reported in whole North America during the mid-90s (Escobedo-Bonilla *et al.*, 2008). At present, WSSV is present in all shrimp producing countries except for Australia and some African countries. The disease prevalence is highly variable and seasonal, increases during cold and/or rainy season in captive and wild populations (Withyachumnarnkul *et al.*, 2003).

2.2.1 Morphology and Ultrastructure

By using the electron microscopy, WSSV virions are found to be enveloped with rod shaped capsid. The nucleocapsids have a bacilliform to ovoidal particles of about 275 nm in length and 120 nm in width (Figure 2.4). These virions have a tail-like appendage at one end with no knowledge regarding the function and composition (Durand *et al.*, 1997). The nucleocapsid, located inside the envelope contains the viral genome and are made up of the WSSV encoded proteins VP664, VP26, VP24 and VP15 (Figure 2.5) (Leu *et al.*, 2005; Goldbach*, et al.*, 2000; van Hulten *et al.*, 2002; Westenberg*, et al.*, 2000). Another further 40 WSSV encoded minor proteins were discovered by protein sequencing of individual bands after applying purified WSSV virions on a SDS-PAGE gel (Xie *et al.*, 2006). WSSV isolated from various geographic areas were found to be very similar in terms of morphology and proteome composition (Nadala Jr & Loh, 1998; Wang *et al.*, 2000).



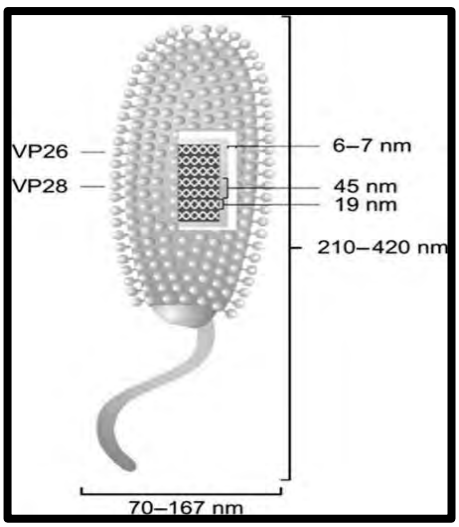


Figure 2.4 and 2.5: Morphology of WSSV

* Graphic structure of WSSV is adapted from Duran *et al*., 1997.

2.2.2 WSSV Genome and Classification

The WSSV genome was identified to be a circular and supercoiled with double-stranded DNA of approximately 300 kilobasepairs (kbp) (van Hulten *et al*., 2001). In addition, WSSV isolated from various region varies in terms of genome size (Thailand 293 kbp, China 305 kbp, Taiwan 307 kbp) (van Hulten *et al*., 2001; Yang *et al.*, 2001). Sequence analysis performed on WSSV genome found it to be unique with there is no known similarities with any known viruses (Lo *et al.*, 1997). Hence, it has been grouped to a new family (*Nimaviridae*) and genus (*Whispovirus*) (van Hulten & Vlak, 2001).

2.2.3 Clinical Signs and Transmission

The onset of WSSV infection leads to a high cumulative rate up to 100 % within 3–10 days (Lightner, 1996). The clinical signs of infected animals includes lethargic behaviour, reduction in food consumption, reduced preening activities, anorexia, loose cuticle and reddish to pink body discoloration (Chou *et al.*, 1995). A distinct characteristic of WSSV infection is the presence of white spots on the exoskeleton especially on the carapace and last abdominal segment. These white spots are the outcome of calcified deposits that range in size from a few mm to 1 cm or more in diameter (Figure 2.6) (Chou *et al*., 1995). There are two modes of WSSV transmission in natural environment which through vertical and horizontal route. Horizontal transmissions includes ingestion of dead infected shrimp and contacting with water containing infected animals or free virus particles (Chou *et al.*, 1998; Poh Shing *et al.*, 1996). Vertical transmission involves transferring of viral particles from the mother to offspring at the time of spawning and then ingested by larvae at first feeding. However whether the WSSV virions are present inside the shrimp eggs remains unresolved (Lo *et al*., 1997; Peng *et al.*, 1998).

Figure 2.6: *M. rosenbergii* infected with WSSV

* Graphic structure of *M. rosenbergii* infected with WSSV is adapted from Escobedo-Bonilla *et al.,* (2008).

2.2.4 Host Range

WSSV can be detected in a wide host range which includes all cultured, wild marine shrimps, crabs, lobsters, crayfishes, squilla, copepods and freshwater cultures species of *M. rosenbergii* (Chang *et al.,* 1998).

2.2.5 Histopathology and Cytopathology

WSSV targets mostly tissues originating from both ectoderm and mesoderm in *M. rosenbergii* such as cuticular epithelium, gills, lymphoid organs, antennal gland, hematopoietic tissues, connective tissue, ovary and the ventral nerve cord (Wonteerasupaya *et al.,* 1995). The typical histological signs of infection is the appearance of enlarged nuclei in cells of susceptible tissues. In the late stages of infection, the epithelia of the stomach, gills and integument may become severely damaged which will lead to multiple organ failure and finally mortality (PohShing *et al.,* 1996; Wang *et al.*, 1999).

The replication of WSSV is initiated in the nucleus which later causes the appearance of homogeneous hypertrophied nuclei and marginated chromatin (Lightner, 1996; Wang *et al.*, 2000). The virus morphogenesis begins with the formation of viral envelopes in the nucleoplasm. The nucleocapsids are formed when the extended, empty, long tubules breaks up into fragments of 12–14 rings to form empty nucleocapsid shells. Later, the empty capsids are enclosed by the envelope leaving at one end an open extremity. The viral DNA together with the nucleoproteins enters the empty capsid through its open end. Once the open end are sealed and a tail-like extension of the envelope forms, the mature form of the virus are obtained (Durand *et al.,* 1997; Wang *et al.,*2000). The transmission of WSSV virion from one cell to another could possibly happen by budding or by rupture of the nuclear envelope or cell membrane of infected cell (Pradeep *et al.,* 2012).

2.3 Vibrio species

Besides viral infection, global shrimp cultivation is also seriously hampered by bacterial related diseases. Bacteria are known as opportunistic pathogen when natural defence mechanisms are suppressed, causing harmful diseases such as Vibriosis and Necrotizing Hepatopancreastitis. Bacteria from the *Vibrio* species, are widely distributed in culture facilitates throughout the world. *Vibrio*-related infections frequently occur in hatcheries, but epizootics also commonly occur in pond reared shrimp species. Vibriosis is caused by gram-negative bacteria in the family *Vibrionaceae* (Saulnier *et al.*, 2000). Outbreaks may occur when environmental factors trigger the rapid multiplication of bacteria which have been tolerated at low levels within shrimp blood, or by bacterial penetration of host barriers (Sizemore & Davis, 1985)

2.3.1 Taxonomic position of *Vibrio parahaemolyticus*

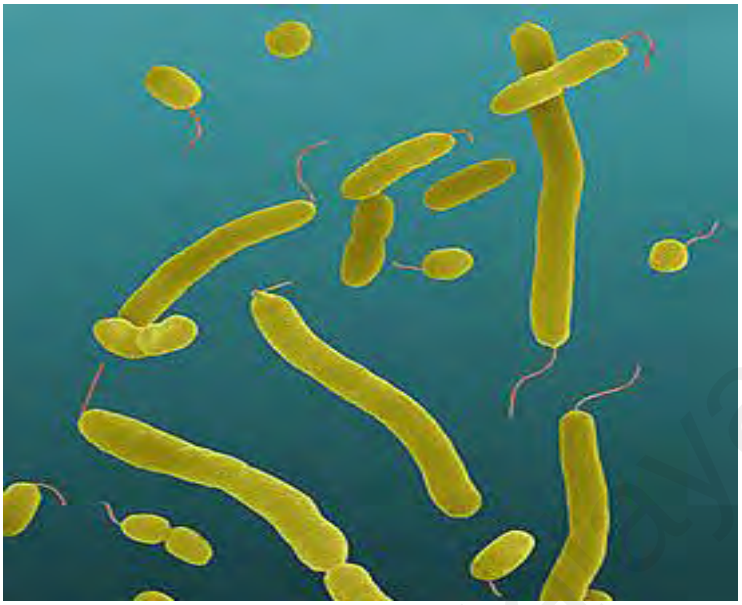


Figure 2.7: Image of *V. parahaemolyticus*

* Graphic structure of *V. parahaemolyticus* is adapted from Ramesh *et al.* (1990)

Kingdom:	Bacteria
Phylum:	Proteobacteria
Class:	Gamma Proteobacteria
Order:	Vibrionales
Family:	Vibrionaceae
Genus:	Vibrio
Species:	Vibrio parahaemolyticus

Vibrio parahaemolyticus is a curved, rod-shaped, gram-negative halophilic bacteria found abundantly in marine and estuarine environments (Figure 2.7). *V.*

parahaemolyticus like other *Vibrio* spp. are facultative anaerobes which tested positive for oxidase and do not form spores (Sizemore *et al.*, 1975). They are motile and have polar flagella with sheaths, similarly with other *Vibrio* spp. They grow optimally between 35°C to 39°C and flourish under salinity concentrations between 0.8% and 3% in marine environments (DePaola *et al.*, 2000; Joseph *et al.*, 1982).

V. parahaemolyticus strains produces two types of toxin which are thermostable direct hemolysin (encoded by tdh genes) and thermostable direct hemolysin-related (encoded by trh genes). The beta-hemolysis of human erythrocytes, also known as the Kanagawa phenomenon are caused by the encoding of thermostable direct hemolysin virulence factors (Honda & Iida, 1993).

2.3.2 Clinical Signs

High mortalities due to Vibriosis occur when shrimps are stressed by these factors: poor water quality, crowding, high water temperature, low DO and low water exchange (Adams, 1991; Lewis, 1973). Adult shrimps suffering from Vibriosis may feature these characteristics: hypoxic, reddening of the body with red to brown gills, reduce feeding and may be observed swimming lethargically at the edges and surface of ponds (Figure 2.8) (Anderson *et al.*, 1988; Nash *et al*., 1992).

Figure 2.8: Shrimp suffering from Vibriosis

* Graphic structure of shrimp suffering from Vibriosis is adapted from Rao, A. V. (2008).

2.3.3 Histopathology of Infected Shrimp

Systemic Vibriosis typically results in the formation of septic haemocytic nodules in the lymphoid organ, heart and connective tissues of the gills, hepatopancreas, antennal gland, nerve cord, telson and muscle (Anderson *et al.*, 1988; Jiravanichpaisal *et al.*, 1994; Liuxy *et al.*, 1996).

2.4 Defence Mechanism against Pathogens

Invertebrate's immune system like crustaceans is deprived of highly adaptive immune system of vertebrate which consists of lymphocytes, immunoglobulins and immunological memory. They rely primarily on innate defence mechanism consisting of various strategies to combat invading pathogen (Lee & Söderhäll, 2002). This approach is less specific but is highly efficient in eliminating the invading pathogens.

The innate defence mechanism involves both cellular and humoral components (Figure 2.9) which works synergistically in detecting and removing harmful pathogens

(Pikul Jiravanichpaisal *et al.*, 2006). The cellular component encompass activities such as encapsulation, nodule formation, and phagocytosis (Johansson & Soderhall, 1989). Whereas the humoral part comprises secretion, activation and release anticoagulant molecules, agglutinins, phenoloxidase enzyme, antimicrobial peptides, antiviral proteins, protease inhibitors, histones, lysosomal enzymes, lipopolysaccharide, β- 1, 3-glucan binding protein and recognition proteins (Holmblad & Söderhäll, 1999; Lemaitre & Hoffmann, 2007).

Besides those above mention processes, additional two more processes; coagulation and melanisation activates upon invasion or injury and plays a crucial role in protecting the host. Melanization is a process involves formation phenolic intermediates and melanin taking place at the site of infected resulting immobilizing and elimination of pathogens from the host system (Söderhäll & Cerenius, 1998).

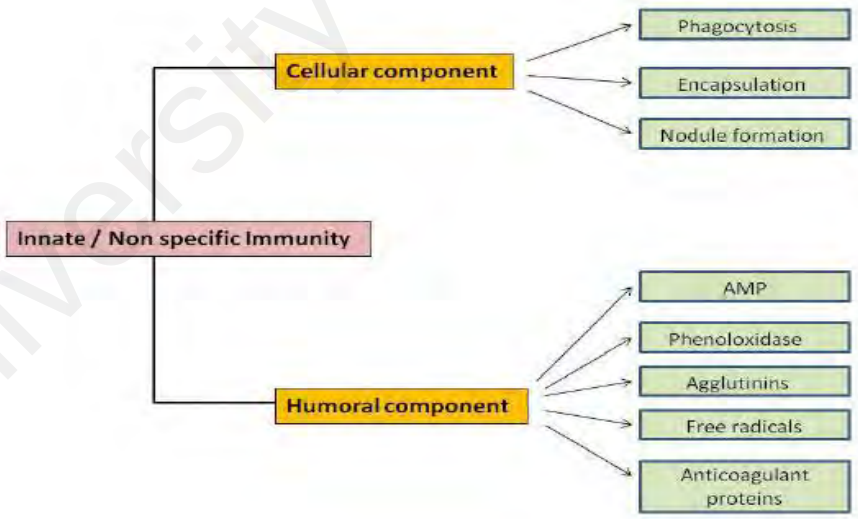


Figure 2.9: Cellular and humoral components of innate immune system in crustaceans.

* Graphic is adapted from Martinez, 1997.

2.5 Host-pathogen Interaction

The host-pathogen interaction has shaped the invertebrate immune system over the past million years. Based on the evolutionary Red Queen hypothesis, pathogens are persistently evolving new strategies to evade the immune response of hosts which caused the host to upgrade their immune system in self-defence to these pathogens (Van Valen, 1974). Therefore, outcome of an imbalance of host–pathogen interactions may either be a successful elimination of the infectious agents or causing harm to the host. This is particularly true for farmed shrimp that have suffered frequent and severe mortality events worldwide (Flegel, 2012).

At present, there is no any effective cure for both WSSV and *V. parahaemolyticus* infections in crustaceans, the best management practice for shrimp aquaculture industry worldwide is to prevent the occurrence of viral and bacterial diseases to some extent. Shrimps like *M. rosenbergii* are well-known to lack an adaptive immune system and they rely entirely on innate mechanism to resist against the invasion of pathogen (Aguirre-Guzman *et al.*, 2009; Iwanaga & Lee, 2005). Shrimp defence mechanisms against viral and bacterial are not well understood. Therefore, studies had been undertaken to identify, characterize and analyse the expression of the immune related genes which may lead to uncover the mechanism involved in host–pathogen interactions and the innate immune system of shrimp.

Insight into the differential gene expression pattern between the host against viral and bacterial pathogens forms the basis in understanding the response of shrimp toward these pathogens, which may aid in developing control strategies to prevent different infectious diseases. Expressed sequence tag (EST) analysis has been used to uncover the immune-related genes in haemocytes and the hepatopancreas of *L. vannamei* and *L.*

setiferus (Gross *et al.*, 2001). This similar approach was utilised in search of immune genes in haemocytes of shrimps *M. japonicus*, *P. monodon* and *F. chinensis* (Leu *et al.*, 2011; Rojtinnakorn *et al.*, 2002; Supungul *et al.*, 2004). In *M. rosenbergii*, there are 11 expressed sequence tags (ESTs) available in public EST database (accessed on June 2, 2010) which is inferior to the EST sequencing of the oriental river prawn, *M. nipponense* containing 3,256 ESTs (Mohd-Shamsudin *et al*., 2013). The shrimp ESTs provided the first description of the host differential expression at the transcriptional level under different pathogen infection.

Besides EST analysis, differential display PCR (DD-PCR) and suppression subtractive hybridization (SSH) techniques were also used in gene expression analysis. The DD-PCR method was used to investigate changes in gene expression patterns in hemocytes of *P. monodon* and hepatopancreas of *P. stylirostris* infected with *V. harveyi* and WSSV(Astrofsky *et al.*, 2002; Somboonwiwat *et al.*, 2006). Meanwhile, the SSH method was applied to reveal the differentially expressed immune genes in *F. indicus* against WSSV and in *P. japonicas* against microbial infection (He *et al.*, 2004; James *et al.*, 2010). Nevertheless, these techniques were able to characterise only a small number of immune genes which is still insufficient as larger numbers of genes need to be identified and functionally characterized to better understand shrimp immunity.

The introduction of cDNA microarrays, a high throughput gene expression platforms enables the expression of thousands of genes to be assayed simultaneously which allows the description of biological phenomena in a global manner. By using the microarray method, the differential gene expression profiles were obtained for *F. chinensis* challenged with viral and bacterial pathogens (Wang *et al.*, 2006; Wang *et al*., 2008). Besides *F. chinensis*, the similar technology was utilised to discover differential immune related gene expression in *P. monodon* and *P. stylirostris* infected with

WSSV(Dhar *et al.*, 2003; Wongpanya *et al.*, 2007). There is no available publication regarding use of microarray in *M. rosenbergii* at present. Even though this technology was found to be effective in studying gene expression, it faces problems related to background and cross-hybridization. Besides it measures only the relative abundance of transcripts which lacks accuracy (Kothapalli *et al.*, 2002; Murphy, 2002). Furthermore,it is costly and requires only predefined sequences to be detected by using this approach (Simon *et al.*, 2003). Therefore its use within organisms lacking complete genome information such as shrimp is still limited.

A far more superior technology known as the 'next generation sequencing technologies' (NGS), has provided a powerful tool for transcriptome analysis with great advantages over previous conventional methods (Mutz *et al.*, 2013). NGS technology enables detail analysis of the complexity of whole transcriptome (complete repertoire of expressed RNA transcripts) and allows discovering of millions of expressed genes with or without genomic information. When compared to previous conventional methods, this latest technology offers advantages like minimum bias, a greater dynamic range, a lower frequency of false positive and higher reproducibility (Mutz *et al*., 2013). By studying the transcriptome, it offers an efficient mean of characterizing the molecular basis of host–pathogen relationship which may unlock the mystery surrounding this field of study (Waddell *et al*., 2007; Wise *et al*., 2007). Recently, transcriptome analysis using the NGS approach has been successfully utilised to identify immune-related genes in marine shrimps against the Taura Syndrome Virus (TSV) and WSSV(Chen *et al*., 2013; Zeng *et al*., 2013).

2.6 Next Generation Sequencing Technology

In 1977, Fred Sanger and Alan R. Coulson introduced a rapid method to sequence the DNA which would go on to transform biology as a whole by providing a tool for deciphering complete genes and later entire genomes. The chain-termination method invented is commonly referred to as Sanger or dideoxy sequencing (Sanger *et al.*, 1977). This method involved reduced handling of toxic chemicals and radioisotopes which made it the only DNA sequencing method used for the next 30 years.

The capillary sequencer which is based on Sanger's chain-termination chemistry such as ABI 3730 (Applied Biosystems) was still applied during sequencing of PCR products and other mini-scale sequencing projects (Mardis, 2008). However, although with major improvement in chemistries and the robust performance of those instruments, the application of relatively expensive Sanger sequencing to large sequencing projects has remained beyond the feasibility of the typical grant-funded investigator (Morozova & Marra, 2008). Thus, most sequencing operations today are performed on next-generation instruments.

At present, there are five commercially available new-generation sequencing technologies which are Roche/454, Illumina, Applied Biosystems SOLiD, Ion Torrent and most recently released HelicosHeliScope (Table 2.1) (Mardis, 2013; Shokralla *et al.*, 2012). These platforms generates an abundance of short reads at a much higher throughput in compared to the state-of-the-art Sanger sequencer (Shokralla *et al*., 2012). The focus will be on Illumina genome analyser as this platform was used for this study.

High-end sequencing Platform	Sequencing Chemistry	Read lengths	Run Time	Template Preparation
Roche 454	Pyrosequencing	400bp 400mb/run	10 hours	Emulsion PCR
Illumina/Solexa –HiSeq 2000	Reversible terminator chemistry	2x100bp	11.5 days	Solid-phase
ABI/Life Technology-SOLiD 5550XL	Sequencing by ligation	2x60bp 15gb/day	8 days	Emulsion PCR
HelicosBiotechnologies	Reversible terminator chemistry	25-55bp 28gb/run	>1gb/hour	Single molecule
Roche 454-GS Junior	Pyrosequencing	400bp 50mb/run	10 hours	Emulsion PCR
Illumina/Solexa- MiSeq	Reversible terminator chemistry	2x150bp 1.0-1.4 gb	26 hours	Solid-phase
ABI/Life Techonology-Ion Torrent	H+ Ion sensitive transistor	320mb/run	8 hours	Emulsion PCR
PacBio	Real-time sequencing	3-15 mb 3gb/run	20 minutes	Single molecule

Table 2.1: Types of NGS platforms

2.6.1 Illumina Genome Analyzer

The concept of 'sequencing by synthesis' (SBS) introduced by Illumina Genome Analyzer (GA) in 2006 produces sequence reads between 32–40 bp from tens of millions of surface amplified DNA fragments simultaneously. By 2010, Illumina launched the HiSeq^{TM}2000, which adopts similar strategy with GA but produces longer read length reaching 100 bp and faster data generation rate.

The Illumina sequencing process involves using a microfluidic cluster station to add a mixture of single-stranded, adaptor oligo-ligated DNA fragments to the surface of a glass flow cell. The flow cell contains eight separate lanes with the interior surfaces having covalently attached oligos which complements to the specific adapters attached onto the library fragments. The library fragments were hybridized to the oligos of the flow cell using an active heating and cooling step, subsequently followed by an incubation

step with reactants and an isothermal polymerase that amplifies the fragments in a discrete area or 'cluster' on the flow cell surfaces. Once completed, the flow cell is inserted into a fluidics cassette within the sequencer. Each cluster is provided with polymerase enzyme and four differentially labelled fluorescent nucleotides with their 3'-OH chemically inactivated to ensure that only a single base is incorporated per cycle. During base incorporation cycle, an imaging step is included to detect the incorporated nucleotide at each cluster and followed by a chemical step that removes the fluorescent group and deblocks the 3' end for the next base incorporation cycle (Figure 2.10). Once the sequencing run ends (~4 days), the sequence of each cluster is computed and filtered to eliminate low-quality reads as specified by the user. A typical sequencing run produces approximately 40–50 million such reads (Zhou *et al*., 2010).

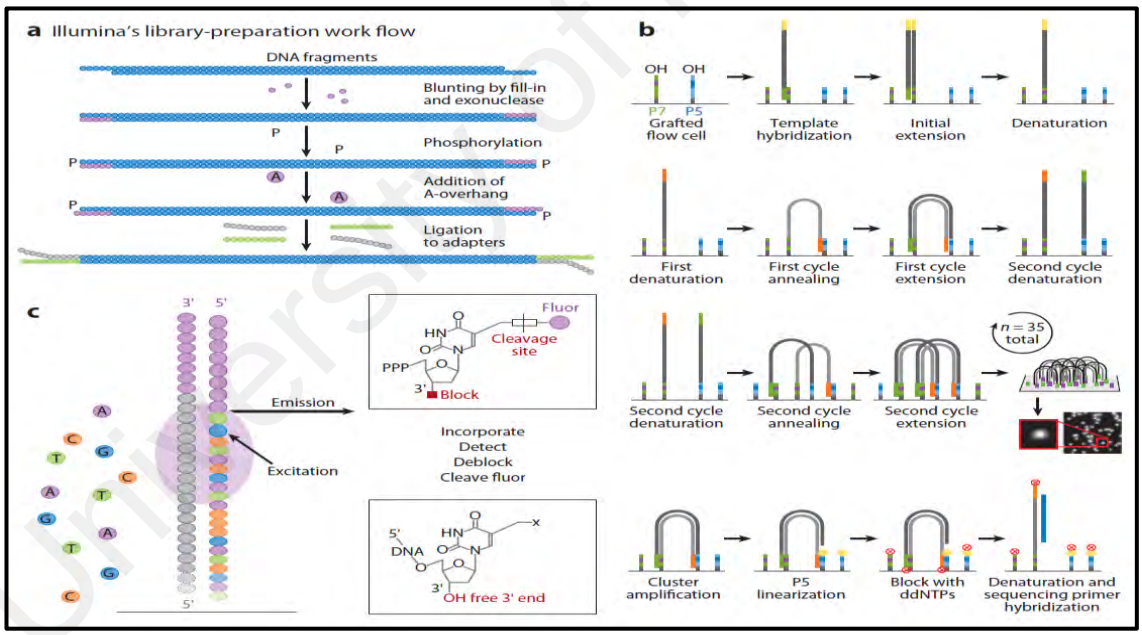


Figure 2.10: (a) Illumina library-construction process. (b) Illumina cluster generation by bridge amplification. (c) Sequencing by synthesis with reversible dye terminators.

* Graphic is adapted from Gullapalli, R. R. (2012).

2.6.2 Bioinformatics Tools for Data Analysis

At present, there is a diverse software tools available specifically to analyse next-generation sequencing data. If a reference genome is present for the target transcriptome analysis, the transcriptome assembly can be built using a splice-aware aligner such as TopHat, SpliceMap, MapSplice, GSNAP, Cufflinks and Scripture. A *de novo* assembly can be undertaken when a reference genome is not available or is incomplete. The rna-seq reads can be *de novo* assembled by using a handful of *de novo* transcriptome assemblers such as the Rnnotator, Multiple-k, Trans-ABySS, Oases and Trinity (Miller *et al.*, 2010; Shendure & Ji, 2008).

2.6.3 General Applications of NGS Platform

NGS technologies provides novel, rapid ways for genome-wide characterisation and sequencing of mRNAs, small RNAs, transcription factor regions, chromatin structure and DNA methylation patterns which revolutionised in the field of molecular biology and genetics (Table 2.2) (MacLean *et al*., 2009; Mardis, 2008).

Category	Examples of application
Complete genome sequencing	Comprehensive polymorphism and mutation discovery in individual human genome
Reduced representation sequencing	Large-scale polymorphism discovery
Targeted genomic resequencing	Targeted polymorphism and mutation discovery
Paired end sequencing	Discovery of inherited and acquired structural variation
Metagenomic sequencing	Discovery of infectious and commensal flora
Transcriptome sequencing	Quantification of gene expression and alternative splicing; transcript annotation; discovery of transcribed SNPs or somatic mutations
Small RNA sequencing	microRNA profiling
Sequencing of bisulfate-treated DNA	Determining pattern of cytosine methylation in genomic DNA

Chromatin immunoprecipitation sequencing (ChIP-Seq)	Genome-wide mapping of protein-DNA interactions
Nuclease fragmentation and sequencing	Nucleosome positioning
Molecular barcoding	Multiplex sequencing of samples from multiple individual

Table 2.2: Application of NGS

2.7 NGS Application within Shrimp Aquaculture

In the field of shrimp aquaculture, application of NGS in shrimps was focused on transcriptome characterization, functional annotation, gene expression profiles analysis and gene-associated markers identification. The use of this method has led to identification of candidate genes or Quantitative Trait Loci (QTLs) which is relevant for aquaculture industry, such as reproduction, sex determination, growth, immunity and tolerance against environmental stress (Santos *et al*., 2014).

2.7.1 Gene Discovery and Differential Expression Analysis

The use of NGS in transcriptome and differential expression analysis for crustaceans have increased lately, search in the Sequence Read Archive Database of the National Center for Biotechnology Information (SRA-NCBI) indicates only 28 deposits of data generated by NGS, mainly for penaeid shrimps such as *L. vannamei*, *L. stylirostris* and *P. monodon* (Santos *et al.*, 2014). Whereas, only three deposited NGS data are reported for freshwater prawn in SRA-NCBI database.

Literature search showed that research in this field has been primarily working on uncovering genes related to immunity, especially for White Spot Syndrome Virus (WSSV) and the Taura Syndrome Virus (TSV) diseases (Chen *et al.*, 2013; Li *et al.*, 2013; Sookruksawong *et al.*, 2013; Zeng *et al*., 2013). These diseases had contributed massive economic losses for the shrimp industry throughout the past few decades. Although

crustaceans do not possess advance immune system, many candidate immune-related genes have been obtained from hemolymph and hepatopancreas tissues such as hemocyanin, lectins, antimicrobial peptide, pattern recognition proteins, superoxide dismutase, caspase and heat shock proteins. By differential gene expression analysis, more up-regulated genes were observed during WSSV infection whereas the opposite was observed during TSV infection. During exposure to nitrite, numerous candidate genes associated with immune response, detoxification, apoptosis pathway were identified in *L. vannamei* (Guo *et al*., 2013).

Besides discovering immune related genes in shrimp, the transcriptome analysis by NGS was undertaken to identify novel genes related to growth and muscle development in *M. rosenbergii*. Novel genes such as cyclophilin, intracellular fatty acid-binding proteins (FABPs),O-methyltransferase (OMT), profilin, actin, tropomyosin and troponin were identified and facilitates genomics approaches to improve the breeding process for aquaculture (Jung *et al*., 2011). Many putative sex determination genes were found in *M. nipponense* by using this technology which improves the understanding of sex-determination mechanisms in shrimp (Jin *et al.*, 2013; Ma *et al.*, 2012).

2.8 Role of Signalling Pathways in Regulating Shrimp Innate Immune Responses

Shrimp possesses an innate immune response which provides defence against pathogen infections (Hoffmann *et al.*, 1999). The entry of pathogens triggers various humoral and cellular activities via signal transduction pathways which are also conserved for both insects and mammals (Borregaard *et al.*, 2000). The initiation of signal transduction began with the binding of extracellular signalling molecular such as pathogen-associated molecular patterns (PAMPs) or a viral protein antigen to cell-surface receptors which later creates a chain reaction inside the cell.

In turn, this receptor alters intracellular molecular creating a response. A second messenger transmits the signal into the cell, triggers a physiological response. The combination of messenger with receptor causes a change in the conformation of the receptor, known as receptor activation. The signal can be amplified and one signalling molecular can cause many responses (effects). The responses includes encapsulation, phagocytosis, coagulation, and melanization performed by different lymph gland cells, and humoral immune reaction includes recognition to microbes, signal transduction including Toll pathway immune deficiency (IMD), Wnt pathway, Janus kinase (JAK)-signal transducers and activators of transcription (STAT), mitogen-activated protein kinase (MAPK) pathway and ubiquitin proteasome pathway (Lemaitre & Hoffmann, 2007). The Toll pathway is mainly involved in defence against fungi, Gram-positive bacteria, and viruses (Lemaitre *et al.*, 1996; Zambon *et al.*, 2005), while the IMD pathway plays key roles in controlling Gram negative bacterial infections and virus infection (Avadhanula *et al.*, 2009; Lemaitre *et al.*, 1995), whereas mitogen-activated protein kinase (MAPK) signalling plays a key role in animal defence during virus infection and also contributed to virus replication in animal cells (Andrade *et al*., 2004). Lastly the JAK/STAT pathway functions in antiviral and antibacterial defence in shrimps and insects (Dostert *et al.*, 2005; Tzou *et al.*, 2002).

2.8.1 JAK/STAT pathway

The Janus kinase (JAK)/signal transducer and activator of transcription (JAK/STAT) pathway consists of three main cellular components: the receptor Domeless, the Janus Kinase (JAK) Hopscotch and the STAT transcription factor (Agaisse & Perrimon, 2004). This pathway is active in regulating the immune system and other biological processes (Agaisse & Perrimon, 2004; Arbouzova & Zeidler, 2006; Dearolf, 1999). The activation of this pathway involves binding of interferon, interleukin, growth factors, or other

chemical messengers to the receptor, which triggers the kinase function of JAK. JAK undergoes self-phosphorylation which then allows the STAT protein to bind to the phosphorylated region and itself undergo phosphorylation. The phosphorylated STAT protein proceeds to dimerize with another phosphorylated STAT protein and translocate into the cell nucleus. In the nucleus, it binds to DNA and promotes transcription of genes responsive to STAT (Dearolf, 1999; Leonard & O'Shea, 1998).

Lately various studies pointing towards crucial role of the JAK/STAT pathway in the innate immune response against both bacteria and virus. The discovery of Ag-STAT, from the mosquito *Anopheles gambiae*, provide the very first evidence that the JAK/STAT pathway is involved in the immune response of insects (Barillas-Mury *et al.*, 1999). Furthermore, the other two mosquito STATs, AaSTAT, and CtSTAT, also responded to LPS stimulation, providing further proof that the JAK/STAT pathway involved in antibacterial immunity (Lin *et al.*, 2004). Besides mosquito, the JAK/STAT pathway in *Drosophila* was responsive to septic injury, which further suggest role in antibacterial responses (Agaisse *et al.*, 2003). The JAK/STAT cascades were similarly activated in Drosophila when challenged with *Drosophila* C virus (Dostert *et al*., 2005).

In crustaceans, Chen *et al*., (2008) discovered a STAT homolog in *P. monodon* which contain similar domains reported in insects such as the DNA binding domain, SH2 domain and C-terminal transactivation domain. Another discovery of STAT homolog in *F. chinensis* further indicates presence of this pathway in shrimp.The transcription of STAT in shrimp was modulated after WSSV and *V. anguillarum* infection, which indicated that JAK/STAT pathway could play crucial role against both virus and bacterial infection in shrimp (Sun *et al.*, 2011). The role of this pathway in *M. rosenbergii* is not known and would be beneficial to elucidate its role in this species.

CHAPTER 3

RNA-SEQ ANALYSIS OF *MACROBRACHIUM ROSENBERGII HEPATOPANCREAS* IN RESPONSE TO *VIBRIO PARAHAEMOLYTICUS* INFECTION

3.1 Introduction

The Malaysian giant freshwater prawn, *Macrobrachium rosenbergii*, is an economically important crustacean worldwide. However, production of this prawn is facing a serious threat from Vibriosis disease caused by *Vibrio* species such as *Vibrio parahaemolyticus*. Unfortunately, the mechanisms involved in the immune response of this species to bacterial infection are not fully understood. We therefore used a high-throughput deep sequencing technology to investigate the transcriptome and comparative expression profiles of the hepatopancreas from this freshwater prawn infected with *V. parahaemolyticus* to gain an increased understanding of the molecular mechanisms underlying the species' immune response to this pathogenic bacteria.

A total of 59,122,940 raw reads were obtained from the control group, and 58,385,094 reads from the *Vibrio*-infected group. Via *de novo* assembly by Trinity assembler, 59,050 control unigenes and 73,946 *Vibrio*-infected group unigenes were obtained. By clustering unigenes from both libraries, a total of 64,411 standard unigenes were produced. The standard unigenes were annotated against the NCBI non-redundant, Swiss-Prot, Kyoto Encyclopaedia of Genes and Genome pathway (KEGG) and Orthologous Groups of Proteins (COG) databases, with19,799 (30.73%), 16,832 (26.13%), 14,706 (22.83%%) and 7,856 (12.19%) hits respectively, giving a final total of 22,455 significant hits (34.86% of all unigenes). A Gene Ontology (GO) analysis search using the Blast2GO program resulted in 6,007 unigenes (9.32%) being categorized

into 55 functional groups. A differential gene expression analysis produced a total of 14,569 unigenes aberrantly expressed, with 11,446 unigenes significantly up-regulated and 3,103 unigenes significantly down-regulated. The differentially expressed immune genes fall under various processes of the animal immune system.

This study provided an insight into the antibacterial mechanism in *M. rosenbergii* and the role of differentially expressed immune genes in response to *V. parahaemolyticus* infection. Furthermore, this study has generated an abundant list of transcript from *M. rosenbergii* which will provide a fundamental basis for future genomics research in this field.

3.2 Literature Review

The Malaysian giant freshwater prawn, *Macrobrachium rosenbergii* (locally known as 'udang galah'), belongs to the genus *Macrobrachium,* which is the largest genus of the family *Palaemonidae* (De Grave *et al.,* 2008). They are found in most inland freshwater areas, including lakes, rivers, swamps, estuarine areas, ponds, canals as well as in irrigation ducts (New, 2002). *M. rosenbergii* spends its adult life in freshwater, but requires brackishwater during the initial stages of its life cycle (Wowor *et al.,* 2009). High demand from the aquaculture industry has led to large-scale farming of this prawn in many countries; the major producers being Bangladesh, Brazil, China, Ecuador, India, Thailand, Taiwan Province of China, and Malaysia (New, 2010).

The global production of this prawn had increased to over 200 000 tonnes/year by 2002, and income in Asia alone is now worth US$1 billion per annum (New, 2005; Schwantes *et al.,* 2009). In Malaysia, the production of cultured *M. rosenbergii* reached 281 metric tonnes by 1998 (New, 2010). Generally, *M. rosenbergii* is assumed to be less resistant towards diseases than penaeid shrimp (Nash *et al.,* 1987). However, with the rise

of large-scale high density prawn aquaculture techniques, production of this species worldwide is facing a serious threat from fatal diseases caused by nodaviruses and bacteria, particularly from the *Vibrio* species (Bonami & Sri Widada, 2011; Tonguthai, 1995). The emergence of these pathogens has had a detrimental impact on the *M. rosenbergii* farming industry, causing considerable economic losses.

Vibrio is a Gram-negative halophilic bacterium found abundantly in marine and estuarine environments (Ramesh *et al.,* 1990; Thompson *et al.,* 2004). Among the different species, *Vibrio parahaemolyticus* has emerged as an important pathogen for *M. rosenbergii* (Khuntia *et al.,* 2008). Several other marine shrimps such as *P. monodon, P. japonicas* and *L. vannamei* have also been found to be susceptible to *Vibrio* infection (Lightner, 1996). Severe *V. parahaemolyticus* infection in prawns leads to a disease known as 'Vibriosis'(Ruangpan & Kitao, 1991; Xu *et al.,* 1994). *M. rosenbergii* suffering from vibriosis may appear black in colour on the carapace, with red discolouration of the exoskeleton and loss of appendages within six days, leading to an 80% mortality rate (Khuntia *et al.,* 2008).

Acquiring and establishing knowledge regarding host pathogen interactions is necessary to unlock the pathogenesis of a particular disease. Host pathogen interactions can result in acute and adaptive immune responses against an invader; however, this has been lacking in *M. rosenbergii* (Kimbrell & Beutler, 2001). The species defends itself against pathogen invasion using an innate immune system involving the cellular and humoral mechanisms (Jiravanichpaisal *et al.,* 2006; Young Lee & Söderhäll, 2002). Recently, some progress has been made in analysing the molecular mechanisms of shrimp-pathogen interactions, and several immune genes from shrimp have been discovered such as lectins, antimicrobial peptides, prophenoloxidase and manganese superoxide dismutase, using methods such as suppression subtractive hybridization

(SSH) and expressed sequence tags (EST) (Leu *et al.,* 2011; Pan *et al.,* 2005; Zhao *et al.,* 2007). However, these two methods have been found to be laborious and costly, which limits their use for the production of large-scale transcripts (Morozova *et al.,* 2009).

A cutting edge technology has emerged recently, known as Next Generation Sequencing technology (NGS). Currently, there are three established platforms which uses NGS technology: the Illumina Genome Analyzer, the Roche/454 Genome Sequencer FLX Instrument, and the ABI SOLiD System (Ansorge, 2009; Metzker, 2010). These platforms have proven versatile and cost-effective tools for advanced research in various genomic areas, such as genome sequencing and re-sequencing, DNA methylation analysis, miRNA expression profiling, and also in non-model organisms as the *de novo* transcriptome sequencing (Varshney *et al.,* 2009). By using the NGS platform, transcriptome analysis can be performed faster and more easily, because it does not require any bacterial cloning of cDNAs (Martin & Wang, 2011). NGS sequencing has the further advantage of generating greater depth of short reads with minimum error rates (Reis-Filho, 2009). Moreover, it is more reliable and efficient than previous methods in measuring transcriptome composition, revealing RNA expression patterns, and discovering new genes on a larger scale (Mutz *et al.,* 2013). The superiority of this technology also lies in its sensitivity, which allows the detection of low-abundance transcripts (Asmann *et al.,* 2008).

Previous studies have been performed on whole transcriptome sequencing of the hepatopancreas, gill and muscle tissues of *M. rosenbergii* using the Illumina Genome AnalyzerIIx platform (Illumina). They successfully produced a comprehensive transcript data for this freshwater prawn, leading to the discovery of new genes (Mohd-Shamsudin *et al.,* 2013). This present study utilised a similar approach to analyse transcriptome data obtained from the hepatopancreas of *M. rosenbergii* experimentally infected with *V.*

parahaemolyticus. The aim was to discover, and determine the role of, immune genes in *M. rosenbergii* involved in *V. parahaemolyticus* infection, which in turn could provide insights into the host-pathogen interactions between these two organisms.

3.3 Material and Methods

3.3.1 *M. rosenbergii* and *V. parahaemolyticus* PCV08-7 Challenge

M. rosenbergii juvenile prawns (5-8g body weight) (equal number of males and females) purchased from a local hatchery (Kuala Kangsar, Perak, Malaysia) were acclimatized at 28 ± 1 ºC in aerated and filtered freshwater for one week prior to challenge with *V. parahaemolyticus*. During the challenge experiment, the prawns (n=10) were intramuscularly injected with 100 µl $1X10^5$ cfu cultured *V. parahaemolyticus* (Tiruvayipati *et al.,* 2013) whereas another batch of prawns (n=10) were injected with 100µl 2% NaCl (1:10, w/v) solution which serves as negative control group. The hepatopancreas tissues of the prawns were dissected at 12 hours post-infection. The tissues were rapidly frozen in liquid nitrogen and stored at -80^{o}C until total RNA extraction. The 12 hour time point was chosen based on our previous work regarding immune related genes from *M. rosenbergii* in response to pathogen such as viruses showing significant gene expression at this time point (Arockiaraj *et al.,* 2012a, 2012b; Arockiaraj *et al.,* 2011a, 2011b).

3.3.2 Total RNA Extraction and Next-Generation Sequencing

Total RNA (~20 mg) was isolated from both the *V. parahaemolyticus*-challenged and negative control group hepatopancreases. The RNA extraction process was performed by using the Macherey-Nagel NucleoSpin RNA II extraction kit in accordance with the manufacturer's protocols and stored at -80^{o}C prior to RNA sequencing. The purity and

integrity of the RNA was assessed by using the Bioanalyzer 2100 (Agilent technologies, USA). In each group, the total RNA samples were pooled from 10 prawns after which cDNA was synthesized followed by sequencing. The sequencing run was conducted on an Illumina HiSeq^{TM} 2000 platform at the Beijing Genome Institute, Shenzhen, China. The sequencing data constituted 90bp paired end read data, with ~117 million raw reads.

3.3.3 Assembly and Functional Annotation

The raw reads were primarily quality filtered to remove adaptor sequences followed by removal of ambiguous 'N' nucleotides (with a ratio of 'N' more than 10%) and sequences with a phred quality score of less than 20 before proceeding to *de novo* assembly by using the Trinity software (Haas *et al.,* 2013). The Trinity programme assembles the reads into contigs and these contigs were assembled to unigenes. Finally, the TIGR Gene Indices clustering tools (TGICL) (Pertea *et al.,* 2003) with default parameters was applied to cluster the unigenes from both groups which produces non-redundant unigenes. Bacterial sequence contamination was investigated using the web-based version of DeconSeq (Schmieder & Edwards, 2011), with a query coverage and sequence identity threshold of 90%.

The non-redundant unigene sequences were aligned to databases which included NCBI non-redundant (Nr), Swissprot (Boeckmann *et al.,* 2003), Cluster of Orthologous Groups (COG) (Tatusov *et al.,* 2000) and Kyoto Encyclopaedia of Genes and Genome (KEGG) (Kanehisa & Goto, 2000) using BLASTX (Mount, 2007) with an E-value cut-off of 10^{-5}. Gene Ontology (GO) was conducted utilizing default parameters using the BLAST2GO software (Ashburner *et al.,* 2000; Conesa *et al.,* 2005). It was from the above mentioned databases that the gene direction of the unigenes which were annotated and the coding sequence were determined from the BLAST results. The prediction for the

coding sequence and the gene direction was performed by ESTscan (Iseli *et al.,* 1999) for those sequences with no defined annotation by using BLAST predicted coding sequence data as the training set.

3.3.4 Identification of Differentially Expressed Unigenes

The transcript expression levels of the unigenes were measured using the FPKM method (Fragments Per kb per Million fragments). The calculation of Unigene expression uses FPKM method (Fragments Per kb per Million fragments), the formula is shown below:

$$FPKM = \frac{10^6 C}{NL/10^3}$$

Set FPKM to be the expression of Unigene A, and C to be number of fragments that uniquely aligned to Unigene A, N to be total number of fragments that uniquely aligned to all Unigenes, and L to be the base number in the CDS of Unigene A. The FPKM method is able to eliminate the influence of different gene length and sequencing level on the calculation of gene expression. Therefore the calculated gene expression can be directly used for comparing the difference of gene expression between samples. A FDR (False Discovery Rate) of <0.001 was used as the threshold of the p-value in multiple tests to judge the degree of differences in gene expression (Reiner *et al.,* 2003). FDR (False Discovery Rate) control is a statistical method used in multiple hypothesis testing to correct for p-value. In practical terms, the FDR is the expected False Discovery Rate; for example, if 1000 observations were experimentally predicted to be different, and a maximum FDR for these observation was 0.1, then 100 out of these observations

would be expected to be false discovered. Genes were considered differentially expressed in a given library when the p-value was less than 0.001 and a greater than two-fold change in expression across libraries was observed.

3.3.5 Quantitative RT-PCR Analysis

We selected seven differentially expressed *M. rosenbergii* unigenes (arginine kinase 1, anti-lipopolysaccharide factor, inhibitor of apoptosis protein, caspase, heat shock protein 21, lectin 1, and NF-kappa B inhibitor alpha) for quantitative RT-PCR analysis (qRT-PCR) to evaluate our Illumina sequencing result. The primer design for the seven unigenes was performed by using Primer3 (Untergasser, A *et al*., 2012) and listed in Table 3.1. Using 1 µg of RNA, first strand cDNA synthesis was carried out (similar to the sample used for transcriptome sequencing) by using the ImProm-II™ Reverse Transcription System (Promega). The qRT-PCR reaction (20µl) consisted of a 10 µl TaqMan Universal RT-PCR Master Mix (Applied Biosystems, Foster City, CA, USA), a 1 µl of primers/probe set containing 900 nM of forward reverse primers, a 300 nM probe and 2 µl of template cDNA. The qRT-PCR program was set with an incubation step at 50°C for 2 min, 40 cycles at 95°C for 10 min, 95°C for 15 sec, and 60°C for 1 min, carried out by using Step One Plus Real-Time PCR System® (Applied Biosystems). Similar qRT-PCR cycle profile was applied for the internal control gene, Elongation factor 1-alpha (primer sequences are listed in Table 3.1). The expression level of the seven immune genes were analysed by using the comparative CT method ($2^{-\Delta\Delta CT}$ method) (Schmittgen & Livak,2008).

Table 3.1. List of genes, primers and probes used in real-time TaqMan PCR assay.

Gene ID	Gene	Gene Abbreviations	NCBI NO	Forward Primer Sequence	Reverse Primer Sequence	Probe Sequence
Unigene22686_All	Arginine Kinase	AK1	NA	GTCTGGTGATCGCAACCTTCA	GTAGATACCGCGGCCTTCAG	TCGCGCTCCATACCTG
Unigene4120_All	Anti-lipopolysaccharide factor	ALF	NA	TGAAGCTCAGGGTTGGGAAGT	ATACCATTTGGTCGTCCACCC	GTTGCCGAAAAGATC
CL6019.Contig2_All	Inhibitor of apoptosis protein	IAPs	NA	CGCACACTATCAAGTCCACAAAATT	GCAGCATTATTGTTAGCTTCAGTAGCT	CAAGGCTGGCACCACC
CL6385.Contig2_All	Caspase	Casp	NA	GAGCAGATCCAGCGATTCTTCA	AACACACACAGCTAAACAAGATACGA	CTGACACGTAAATTTT
Unigene3736_All	Heat shock protein 21	HSP21	NA	ATCAGTCCCTCAACCAGACC	ATTGCTCTGCTGTGTAACCG	TCCCAGCAGACACAGCAAACACA
Unigene10978_All	Lectin 1	LT1	NA	ACTGGACCGACGCTAAGGA	CCCTATCAGATTGGCGTCATCAAAT	ACGGGCTGCATACTCA
Unigene13073_All	NF-kappa B inhibitor alpha	NFκBI-α	NA	GCAAGACATCTGGTCGTTTGTG	GAGTGCAAGGGCGTGTTG	CCTGCGGTCTCTCG
	Internal control					
	Elongation factor 1-alpha	EF1-A	EL609261.1	ACTGCGCTGTGTTGATTGTAGCT	ACAACAGTACGTGTTCACGGGTC	TACTGGTGAGTTTGAAGCT

-NA = not available

3.4 Results

3.4.1 Illumina Sequencing and Assembly

The task of profiling all the immune-related genes involved in *V. parahaemolyticus* infection began with sequencing the two cDNA libraries prepared from pooled mRNAs obtained from the hepatopancreases of the control and infected groups using the Illumina HiSeq^{TM} 2000 platform. A total of 59,122,940 raw reads were obtained from the control group, and 58,385,094 reads from the *Vibrio*-infected group. The raw reads were further filtered to remove adaptor sequences, ambiguous reads and low quality reads, thereby generating 90-bp of 54,708,014 and 54,295,342 clean reads for the control and infected groups respectively (Q_{20}~98% and percentage of unknown nucleotide is 0%). All sequencing reads were deposited into the Short Read Archive (SRA) of the National Centre for Biotechnology Information (NCBI), and can be accessed under the accession number SRR1424572 for control and SRR1424574 for *Vibrio*-infected ones.

All the clean reads were subjected to *de novo* assembly using the Trinity program which uses three independent software modules – Inchworm, Chrysalis, and Butterfly – applied sequentially to process the huge sequencing data of RNA-seq reads. The assembly of the reads produced 95,645 contigs (with an N_{50} of 467 bp and mean length of 313 bp) for the control group and 123,141 contigs (with an N_{50} of 482 bp and mean length of 318 bp) for *Vibrio*-infected group. These contigs were further assembled into unigenes, producing 59,050 control unigenes (with an $N_{50}$of 685 bp and mean length of 479 bp) and 73,946 infected group unigenes (with an N_{50} of 829 bp and mean length of 532 bp). The length distribution of control and *Vibrio*-infected contigs and unigenes are shown in Appendix. By clustering unigenes from both libraries, a total of 64,411 standard unigenes

were produced, with a mean size of 698 bp and an N_{50} of 1137 bp. An overview of the sequencing and assembly is shown in Table 3.2.

Table 3.2: Summary of the control and *V. parahaemolyticus* infected transcriptome sequencing

	Control	*V. parahaemolyticus* Infected
Total number of reads	54,708,014	54,295,342
Total base pairs (bp)	4,923,721,260	4,886,580,780
Q_{20} value	97.73%	97.77%
Total number of contigs	95,645	123,141
Mean length of contigs (bp)	313	318
Total number of unigenes	59,050	73,946
Mean length of unigenes (bp)	479	532
NCBI Nr annotated	19,799	
Swiss-Prot annotated	16,832	
KEGG annotated	14,706	
COG annotated	7,856	
GO annotated	6,007	

The standard unigenes were annotated by searching the sequences using BLASTX against the NCBI non-redundant, Swiss-Prot, Kyoto Encyclopaedia of Genes and Genome pathway (KEGG) and Orthologous Groups of Proteins (COG) databases, which produced 19,799 (30.73%), 16,832 (26.13%), 14,706 (22.83%%) and 7,856 (12.19%) hits respectively, giving a final total of 22,455 significant hits (34.86% of all unigenes). The size distribution profile for the coding sequences (CDS) and identified proteins are shown in Appendix B. A Gene Ontology (GO) analysis search using the Blast2GO program

resulted in 6,007 unigenes (9.32%) being categorized into 55 functional groups. The unigenes without hits using the BLASTX analysis were subjected to an ESTScan, producing 4,977 unigenes (7.82%) predicted to contain coding sequences. The size distribution of the ESTs and proteins are shown in Appendix B.

The species distribution of the unigenes using the BLASTX results is shown in Figure 3.1. The *M. rosenbergii* unigenes were matched against *D.pulex* sequences (10.3%), *T. castaneum* (6.1%) and *P. h. corporis* (4.2%). The unigenes showed a match with those of *D.pulex* and *T. castaneum* probably because of their closer phylogenetic relationship and the availability of vast genomic information. The remaining unigenes (66.5%) which matched were similar to other species due to limited genome information in crustaceans.

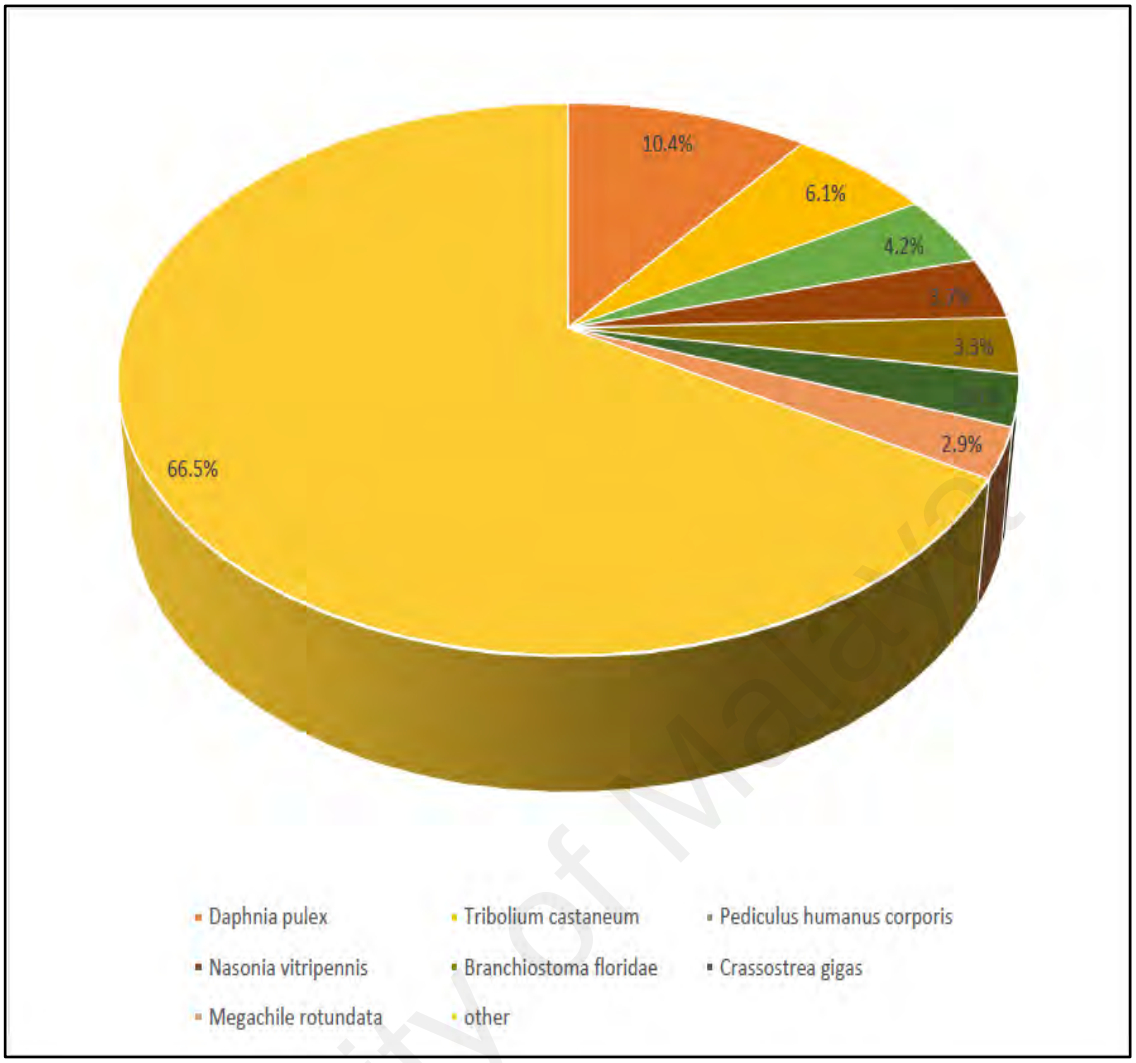


Figure 3.1. Species distribution of the BLASTX matches of the transcriptome unigenes. This figure shows the species distribution of unigene BLASTX matches against the nr protein database (cutoff value E<10^{-5}) and the proportions for each species.

3.4.2 Functional Annotation

The standard unigenes were analysed using the COG database to classify them and predict their functions. A total of 7,856 unigenes were assigned to COG classifications and functionally classified into 25 protein families which mainly were

involved in cellular structure, biochemistry metabolism, molecular processing, and signal transduction (Figure 3.2). The cluster predicted for general function (3,431, 43.67%) emerged as the largest group, followed by translation, ribosomal structure, biogenesis cluster (1,760, 22.40%) and the replication, recombination, repair clusters (1,413, 17.98%). The clusters with the lowest number of unigenes were nuclear structure and extracellular structures (<1% in each cluster).

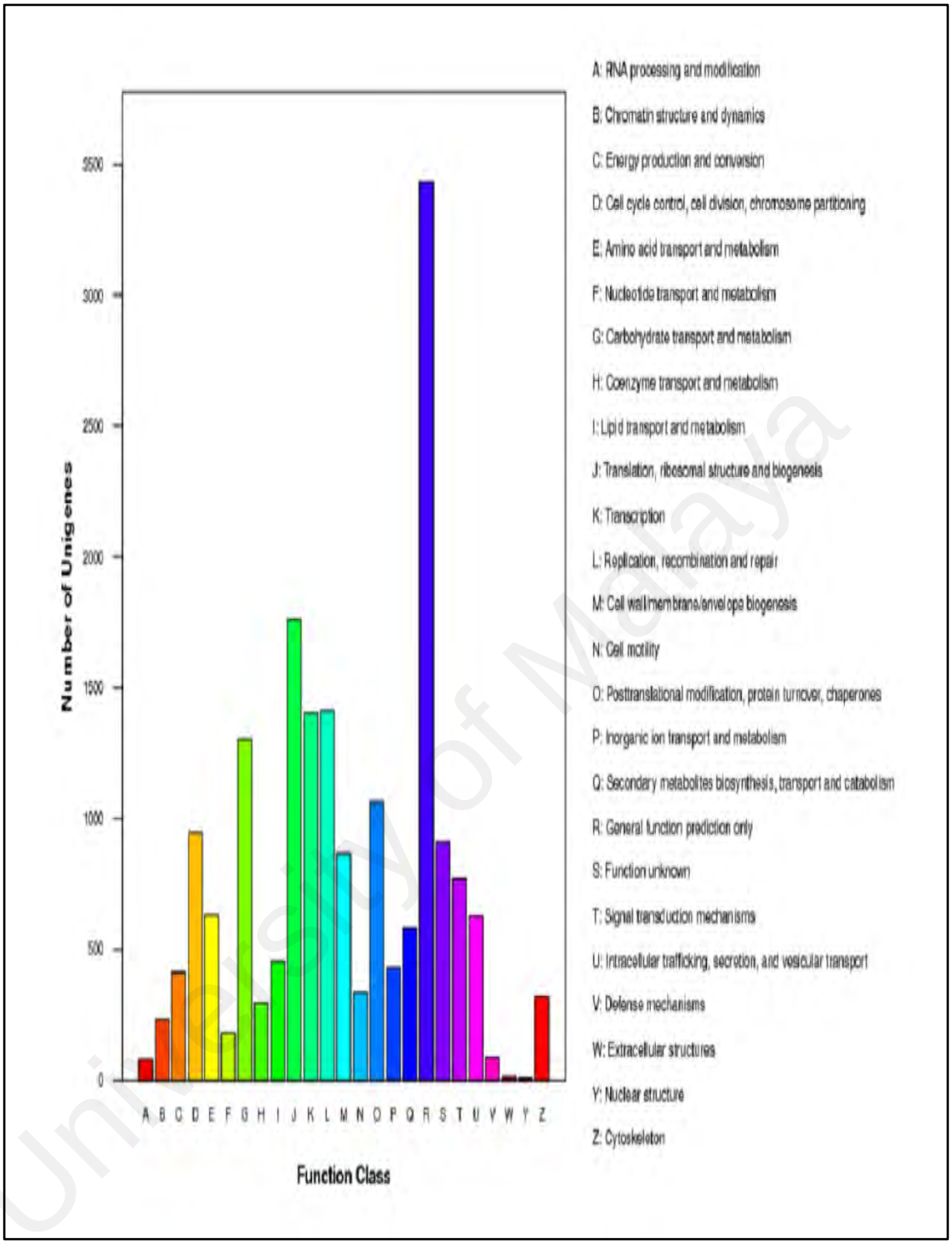


Figure 3.2. Histogram presentation of Cluster of Orthologus Groups (COG) classification of 7,856 known protein annotated unigenes. Each bar represents the number of unigenes classified into each of the 25 COG functional categories.

.

The standard unigenes with Nr annotations were subjected to Gene Ontology (GO) analysis, which provides a dynamic controlled vocabulary and hierarchical relationship for the representation of information on molecular functions, cellular components and biological processes, allowing a coherent annotation of gene products. The GO annotations produced 6,007 unigenes (biological process: 2,177 unigenes, cellular component: 1,563 unigenes and molecular function: 2,267 unigenes) which were assigned to 55 GO ontology sub-categories (Figure 3.3). In the molecular function category, most of the genes fell into the 'binding' and 'catalytic activity' groups, whereas in the biological process category, the majority of the genes fell into the categories of 'cellular processes', 'metabolic processes' and 'single-organism processes'. Finally, in the cellular component category, a high percentage of the genes fell into the 'cell', 'cell part' and 'organelle' sub-categories.

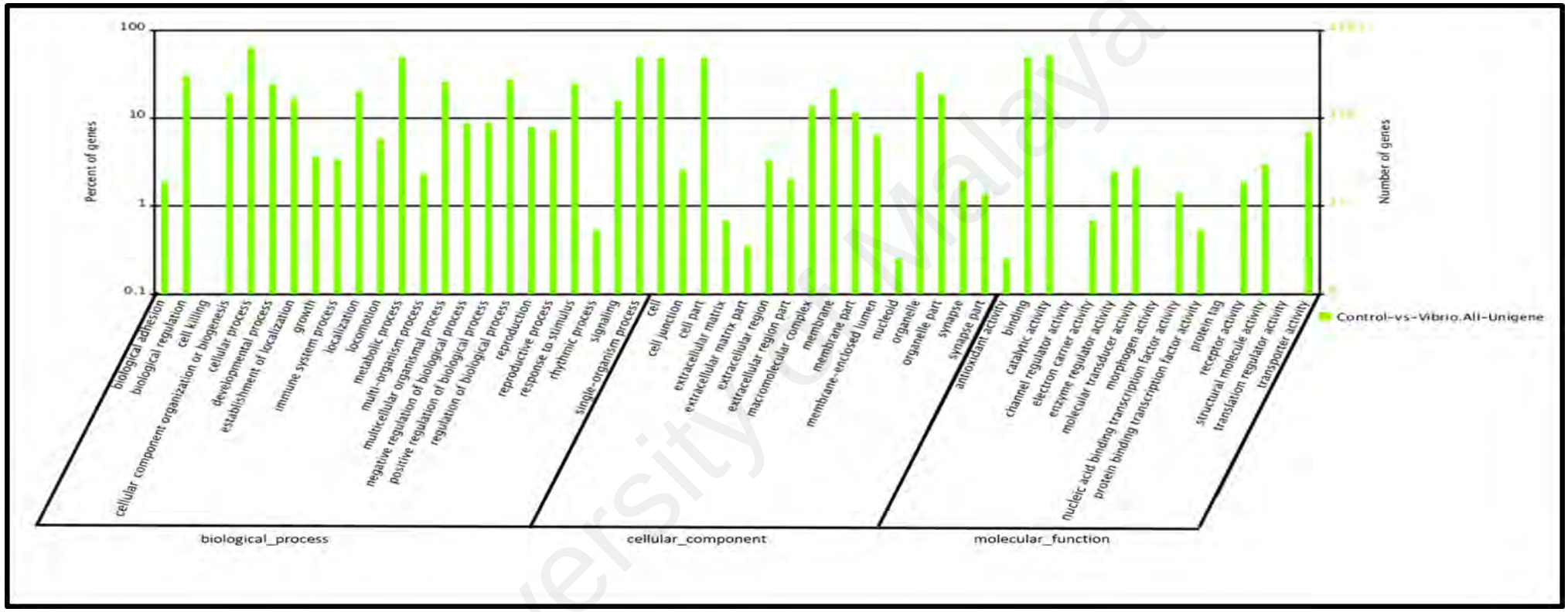


Figure 3.3. Gene ontology (GO) classification of the 6,007 protein annotated unigenes. Unigenes sequences were systematically classified into GO sub-categories under the biological process, cellular component and molecular function gene ontology catalogue system. Each bar represents the relative abundance of unigenes classified under each sub-category.

The Kyoto Encyclopaedia of Genes and Genomes (KEGG) Pathway is a collection of manually drawn pathway maps representing knowledge on molecular interactions and reaction networks. Pathway-based analysis is helpful in identifying biological functions and gene interactions in the pathway. Using the KEGG database, 14,706 unigenes were grouped into 253 pathways. The majority of the unigenes fell into the categories of "metabolic pathways" (2192 members, 14.91%), "regulation of actin cytoskeleton" (557 members, 3.79%), "spliceosome" (509 members, 3.46%), "RNA transport" (498 members, 3.38%) and "focal adhesion" (475 members, 3.22% each). The least represented pathways, with less than 10 unigenes categorized in each pathway, were "biotin metabolism", "phenylalanine, tyrosine and tryptophan biosynthesis", "vitamin B6 metabolism", "lipoic acid metabolism" and "thiamine metabolism". The top twenty of these KEGG biological pathway classifications are shown in Figure 3.4.

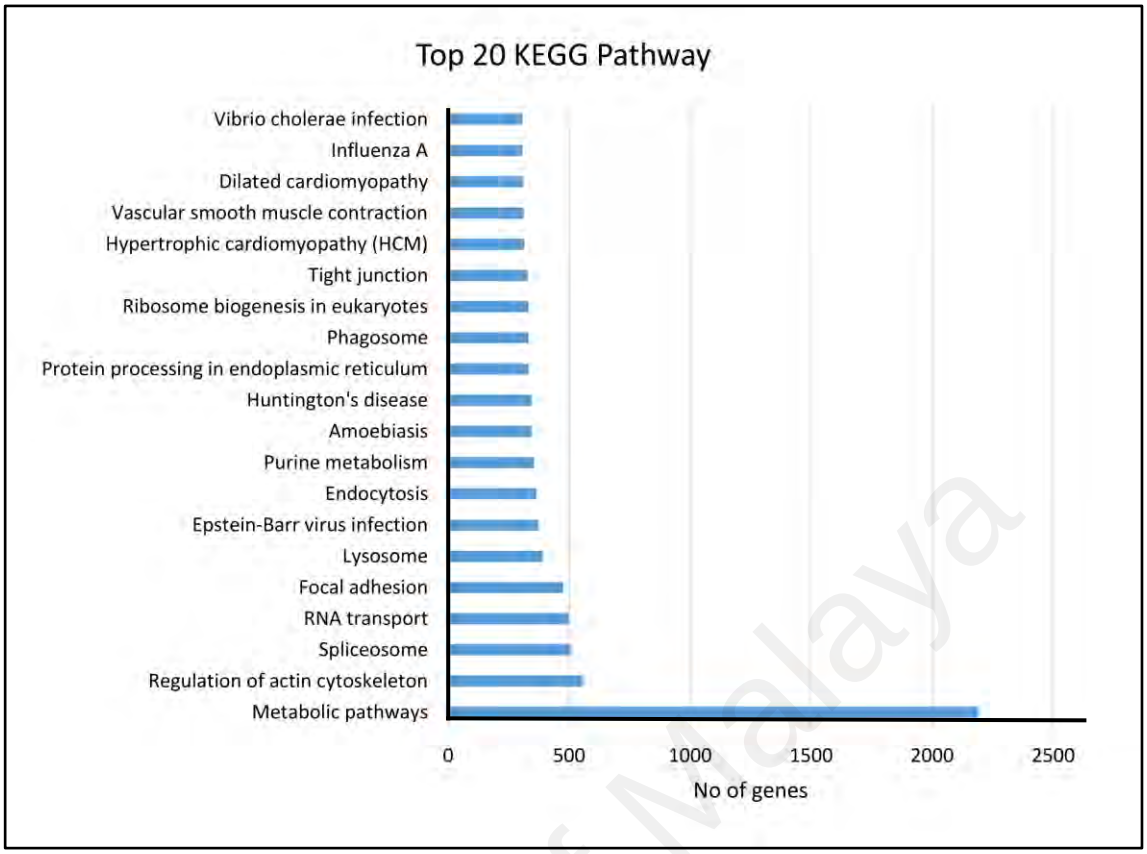


Figure 3.4. Top 20 KEGG biological pathway classification histograms for annotated unigenes.

3.4.3 Identification of Aberrantly Expressed Genes

We identified differentially expressed genes between these two groups by comparing the relative transcript abundance in each unigene by using the FPKM method (Fragments Per kb per Million fragments). A total of 14,569 unigenes were found to be aberrantly expressed; 11,446 of these were significantly up-regulated, whereas 3,103 unigenes were significantly down-regulated (Figure 3.5). The differentially expressed genes were annotated against the NR, Swiss-Prot, GO, COG and KEGG databases by BLASTX with a cut-off E-value of 10^{-5}. The 9,469 (65%) unigenes containing low sequence homology to known sequences in public databases could represent non-coding

RNA, misassembled unigenes or unknown genes of *M. rosenbergii* which responded to *V. parahaemolyticus*-infection.

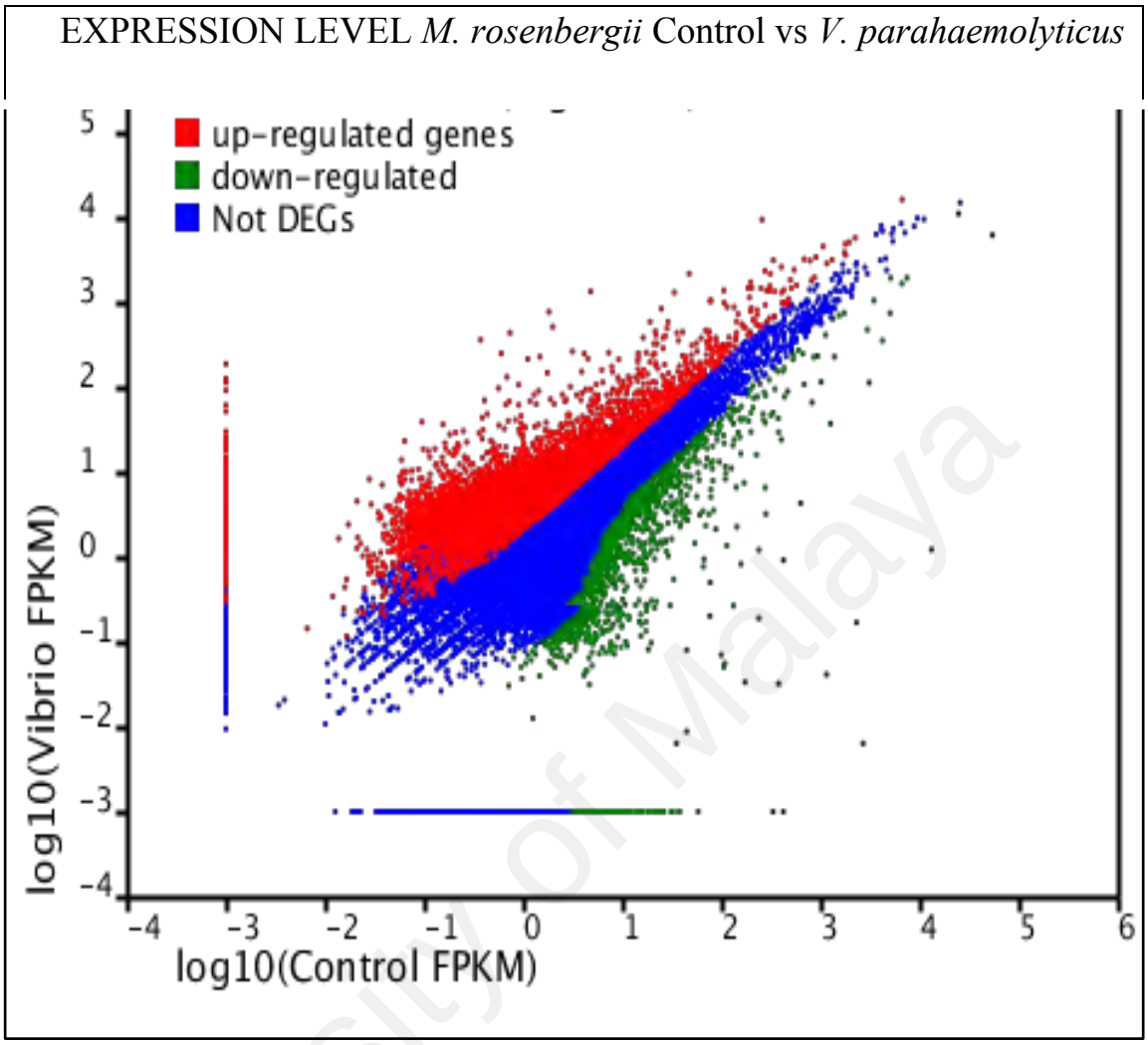


Figure 3.5. Digital gene expression between control group and *V. parahaemolyticus* infected group. Each point represents a unigene. The x- and y-axis are the log10 of the normalized expression level (FPKM) of unigene between the two groups. Red and green points indicate significant change at the absolute value of log2 (FPKM ratio in two groups) ≥1 and fdr =0.001. Red points indicate up-regulated unigenes and green points indicate down-regulated unigenes in the two groups which its expression level is represented by the y-axis. Blue points indicate insignificant differentially expressed unigenes.

For validation of the Illumina sequencing result, seven unigenes were chosen randomly for quantitative real time-PCR (qRT-PCR) analysis. The qRT-PCR results showed similar trends for all the genes to the sequencing data (Figure 3.6). For example, based on the Illumina sequencing analysis, arginine kinase 1, anti-lipopolysaccharide factor, inhibitor of apoptosis protein, caspase, heat shock protein 21, lectin 1 and NF-kappa B inhibitor alpha were up-regulated 4.67, 4.13, 1.02, 3.08, 4.61, 4.4 and 3.11 log2-fold respectively; and the same elements showed 2.5, 4.4, 1.5, 2.1, 4.8, 3.6 and 2.7 log2-fold change respectively in the qRT-PCR analysis. While the results from these two analyses did not match perfectly, perhaps due to sequencing biases, the qRT-PCR analysis broadly confirmed the direction of change obtained from the Illumina sequencing analysis.

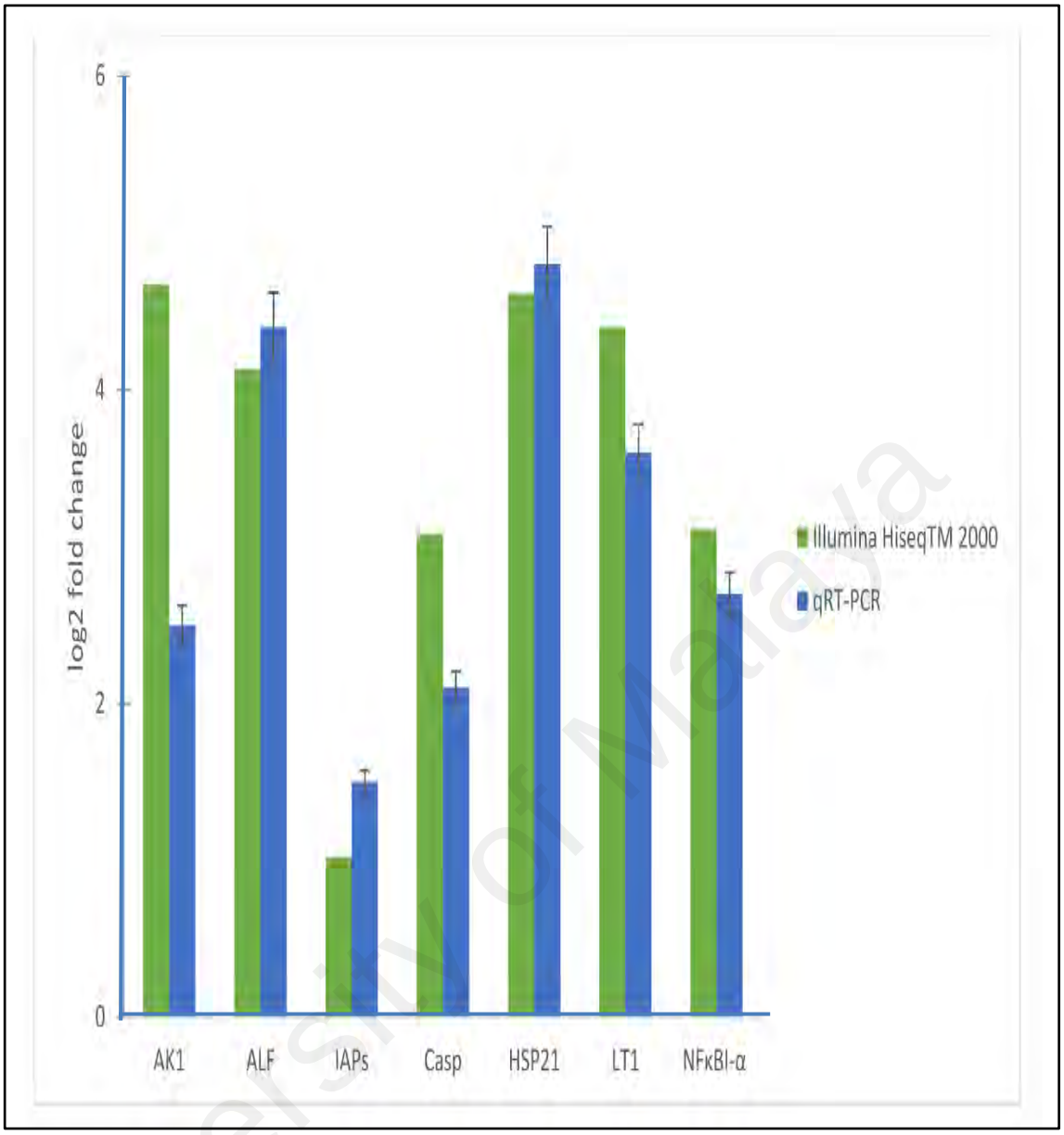


Figure 3.6. Comparison of the expression profiles of selected genes as determined by Illumina Hiseq^{TM} 2000 sequencing (green) and qRT-PCR (blue). Target gene abbreviations are as follows: AK1 - arginine kinase, ALF- anti-lipopolysaccharide factor, IAPs- apoptosis inhibitor, Casp- caspase, HSP21- heat shock protein 21, LT1- lectin 1, NFκBI-α - NF-kappa B inhibitor alpha. Error bars indicated standard deviations of averages from three replicates.

3.4.4 Potential Immune-related Genes Involved in *M. rosenbergii* Immune Response.

Many of the aberrantly expressed genes found in the *V. parahaemolyticus*-infected groups compared to the control group are known to belong to various processes clustered under the animal immune system (Table 3.3). These immune genes are grouped into 11 functions, including antimicrobial proteins (6 unigenes), proteases and proteinases (19 unigenes), signal transduction (Toll pathway (7 unigenes), Wnt signalling pathway (6 unigenes), MAPK signalling pathway (13 unigenes), JAK/STAT pathway (5 unigenes), IMD pathway (2 unigenes) and other signal transduction genes (11 unigenes)), blood clotting system (4 unigenes), cell death (7 unigenes) , cytoskeletal (7 unigenes), heat shock proteins (4 unigenes), oxidative stress (6 unigenes), pathogen recognition immune receptors (16 unigenes), prophenoloxidase system (3 unigenes) and other immune genes (14 unigenes). The role of these groups is described in more detail in the next (Discussion) section.

Table 3.3. Candidate genes involved in *M. rosenbergii* immune response against *V. parahaemolyticus*.

Category or gene id	Homologues function	Species	FC*
Antimicrobial			
Unigene4120_All	Anti-lipopolysaccharide factor	*Macrobrachium rosenbergii*	4.13
Unigene23546_All	Anti-lipopolysaccharide factor 2	*Macrobrachium rosenbergii*	2.84
CL1276.Contig1_All	Anti-lipopolysaccharide factor 3	*Macrobrachium rosenbergii*	2.15
Unigene37309_All	Crustin	*Macrobrachium rosenbergii*	5.13
Unigene26338_All	Lysozyme	*Portunus trituberculatus*	1.92
Unigene13073_All	NF-kappa B inhibitor alpha	*Macrobrachium rosenbergii*	3.11
Blood Clotting system			
Unigene13048_All	Clottable protein	*Marsupenaeus japonicus*	1.42
Unigene34308_All	Transglutaminase	*Macrobrachium rosenbergii*	11.34
Unigene36567_All	Proclotting enzyme	*Harpegnathos saltator*	11.22
Unigene41253_All	Coagulation factor XII	*Hageman factor*	10.61
PRPs			
Unigene10978_All	Lectin 1	*Macrobrachium rosenbergii*	4.4
CL4516.Contig1_All	Lectin 2	*Macrobrachium rosenbergii*	1.05
Unigene7825_All	Lectin 3	*Macrobrachium rosenbergii*	2.54
CL3039.Contig4_All	Lectin 4	*Macrobrachium rosenbergii*	4.55
Unigene9635_All	C-type lectin	*Penaeus monodon*	1.14
CL1600.Contig1_All	C-type lectin-2	*Litopenaeus vannamei*	1.69
CL5230.Contig3_All	C-type lectin 5	*Fenneropenaeus chinensis*	2
Unigene1391_All	Hemolectin	*Papilio xuthus*	5.81
Unigene25701_All	M-type lectin	*Marsupenaeus japonicus*	1.69
Unigene23215_All	Perlucin-like protein	*Mytilus galloprovincialis*	-1.92

Category or gene id	Homologues function	Species	FC*
CL1075.Contig4_All	Tachylectin	*Macrobrachium rosenbergii*	-1.71
Unigene19793_All	Lectin B isoform 2	*Marsupenaeus japonicus*	3.55
Unigene28944_All	Ficolin	*Branchiostoma floridae*	1.31
Unigene37041_All	Mannose-binding protein	*Procambarus clarkii*	11.69
CL1124.Contig1_All	Glucan pattern-recognition lipoprotein	*Fenneropenaeus chinensis*	1.12
Unigene23671_All	lipopolysaccharide and beta-1,3-glucan binding protein	*Macrobrachium rosenbergii*	1.71
Proteinases and Proteinases inhibitors			
Unigene1509_All	Cathepsin B	*Pandalus borealis*	1.87
Unigene22859_All	Cathepsin C	*Fenneropenaeus chinensis*	2.33
CL1587.Contig2_All	Cathepsin D	*Penaeus monodon*	-1.56
Unigene26111_All	Cathepsin L	*Pandalus borealis*	1.41
CL5824.Contig1_All	Serpin serine protease inhibitor	*Fenneropenaeus chinensis*	12.24
CL6038.Contig1_All	26S protease regulatory	*Nasonia vitripennis*	1.44
Unigene12869_All	Alpha-2-macroglobulin	*Macrobrachium rosenbergii*	5.09
CL2365.Contig1_All	Caspase	*Marsupenaeus japonicus*	1.04
Unigene13293_All	Astacin	*Strongylocentrotus purpuratus*	1.64
Unigene13633_All	Serine protease	*Fenneropenaeus chinensis*	3.08
Unigene17259_All	Caspase 8	*Ictalurus punctatus*	2.41
CL2565.Contig1_All	Serine proteinase inhibitor 6	*Penaeus monodon*	4.51
CL1118.Contig1_All	Serpin B	*Marsupenaeus japonicus*	7.02
Unigene26757_All	Hemocytekazal-type proteinase inhibitor	*Penaeus monodon*	3.87
CL1127.Contig2_All	Kazal-type serine proteinase inhibitor 4	*Procambarus clarkii*	-6.01
CL5487.Contig2_All	Aminopeptidase N	*Camponotus floridanus*	1.4

Category or gene id	Homologues function	Species	FC*
Unigene5621_All	Masquerade-like serine proteinase-like protein 2	*Penaeus monodon*	5.39
Unigene21512_All	Serine proteinase inhibitor	*Macrobrachium rosenbergii*	6.91
Unigene28987_All	CUB-serine protease	*Panulirus argus*	2.92
Heat shock Proteins			
Unigene3736_All	Heat shock protein 21	*Macrobrachium rosenbergii*	4.61
Unigene23034_All	Heat shock protein 40	*Frankliniella occidentalis*	1.8
Unigene14757_All	Heat shock protein 70	*Portunustri tuberculatus*	2.26
CL4309.Contig3_All	Heat shock protein 90	*Scylla paramamosain*	1.93
Unigene16858_All	Small heat shock protein	*Fenneropenaeus chinensis*	2.41
Oxidative stress			
CL353.Contig2_All	Glutathione S transferase	*Procambarus clarkii*	1.01
Unigene19405_All	Cu/Zn superoxide dismutase	*Bombyx mori*	1.61
CL5590.Contig3_All	Glutamine synthetase	*Panulirus argus*	-3.03
CL559.Contig1_All	Farnesoic acid O-methyltransferase	*Nilaparvata lugens*	1.81
CL2477.Contig1	Catalase	*Litopenaeus vannamei*	1.6
CL2787.Contig2	Thioredoxin reductase	*Branchiostoma floridae*	1.09
Cytoskeletal			
Unigene10471	Chitinase	*Trichomonas vaginalis*	6.16
CL6026.Contig5_All	Actin	*Marsupenaeus japonicus*	2.92
Unigene13021_All	Calponin	*Chironomus riparius*	2.32
CL5710.Contig1_All	Profilin	*Fenneropenaeus chinensis*	1.45
CL461.Contig8_All	Tubulin	*Homarus americanus*	1.39
Unigene19173_All	Beta-integrin	*Fenneropenaeus chinensis*	2.78
CL2046.Contig1_All	Integrin	*Litopenaeus vannamei*	3.03

Category or gene id	Homologues function	Species	FC*
Signal transduction			
Toll pathway			
Unigene25507_All	Toll-like receptor 3	*Daphnia pulex*	2.59
CL2421.Contig1_All	Toll-like receptor 8	*Ixodes scapularis*	1.48
CL2438.Contig1_All	Caspase	*Eriocheir sinensis*	3.08
Unigene25799_All	Myd88 protein	*Artemia sinica*	3.25
CL1915.Contig5_All	Toll interacting protein	*Marsupenaeus japonicus*	1.73
CL5402.Contig1_All	Toll receptor 2	*Marsupenaeus japonicus*	-2.46
Unigene14481_All	Dorsal	*Fenneropenaeus chinensis*	3.19
Wnt signalling pathway			
CL2430.Contig2_All	beta-catenin	*Parhyale hawaiensis*	1.62
Unigene36938_All	Lipoprotein receptor	*Callinectessapidus*	10.94
Unigene15960_All	Prickle	*Saccoglossuskowalevskii*	1.71
Unigene15970_All	Low-density lipoprotein receptor-related protein 6	*Daphnia pulex*	2.05
Unigene12737_All	Glypican 4	*Daphnia pulex*	1.14
Unigene25938_All	Secreted frizzled-related protein 5	*Daphnia pulex*	1.82
MAPK signalling pathway			
CL4432.Contig2_All	Max protein	*Daphnia pulex*	-3.11
Unigene9636_All	Serine/threonine-protein kinase TAO1	*Acromyrmex echinatior*	1.31
CL5614.Contig1_All	JNK-interacting protein 3	*Apis mellifera*	4.89
Unigene26817_All	Mitogen-activated protein kinase kinasekinase 1	*Oryctolaguscuniculus*	2.79
Unigene102_All	Map kinase-interacting serine/threonine	*Scylla paramamosain*	2.48
Unigene110_All	Mitogen-activated protein kinase 8 interacting protein 3	*Bombusterrestris*	2.94
CL726.Contig3_All	Mitogen-activated protein kinase kinasekinasekinase 4	*Bombus terrestris*	1.65

Category or gene id	Homologues function	Species	FC*
CL3174.Contig1_All	Mitogen-activated protein kinase kinasekinase 7	*Danaus plexippus*	1.41
Unigene29569_All	Mitogen-activated protein kinase kinasekinase 13	*Bombus terrestris*	4.03
Unigene14757_All	Heat shock protein 70	*Portunus trituberculatus*	2.26
Unigene3873_All	Protein phosphatase 5	*Bombusterrestris*	1.21
Unigene102_All	Mapkinase-interacting serine/threonine	*Scylla paramamosain*	2.48
Unigene11596_All	Raf homolog serine/threonine-protein kinase phl-like	*Nasoniavitripennis*	2.41
JAK/STAT pathway			
Unigene25562_All	STAT long form	*Penaeus monodon*	2.84
Unigene25762_All	Signal transducing adaptor molecule	*Megachile rotundata*	1.98
Unigene16474_All	Tumor susceptibility gene 101 protein	*Harpegnathos saltator*	2.46
CL1133.Contig6_All	Domeless	*Tribolium castaneum*	2.94
Unigene28876_All	Thyroid peroxidase	*Pediculus humanus corporis*	3.66
IMD pathway			
CL5784.Contig1_All	Relish	*Litopenaeus vannamei*	2.53
Unigene16147_All	IKKbeta	*Litopenaeus vannamei*	5.39
Other signal transduction genes			
CL1409.Contig1_All	Innexin 3	*Cancer borealis*	10.29
Unigene30035_All	cAMP-dependent protein kinase type II regulatory subunit	*Camponotus floridanus*	1.05
CL3406.Contig2_All	Casein kinase II subunit alpha	*Danaus plexippus*	1.75
CL4247.Contig1_All	Rab-protein 14	*Tribolium castaneum*	1.85
Unigene37520_All	Afadin	*Pediculus humanus corporis*	2.89
Unigene34256_All	TBC1 domain family member 10B	*Crassostrea gigas*	11.52
Unigene33128_All	REM2- and Rab-like small GTPase 1-like	*Cavia porcellus*	3.21
Unigene22971_All	Cullin-2	*Harpegnathos saltator*	1.25

Category or gene id	Homologues function	Species	FC*
Unigene6699_All	Ubiquitin-protein ligase E3C	*Camponotus floridanus*	1.37
Unigene10240_All	Cullin-associated NEDD8-dissociated protein 1	*Crassostrea gigas*	1.95
CL1584.Contig1_All	Beta 1,4-endoglucanase	*Cherax quadricarinatus*	2.37
ProPO system			
Unigene12734_All	Prophenoloxidase	*Macrobrachium rosenbergii*	3.15
Unigene7353_All	Prophenoloxidase activating factor	*Fenneropenaeus chinensis*	3.79
Unigene18337_All	Prophenoloxide activating enzyme III	*Macrobrachium rosenbergii*	4.99
Cell death			
Unigene2555_All	Beclin 1	*Megachile rotundata*	11.89
CL1843.Contig2_All	ALG-2 interacting protein x *Penaeusmonodon*	*Penaeus monodon*	3.2
Unigene14045_All	Program cell death 5-like	*Penaeus monodon*	1.61
CL2477.Contig1_All	Catalase	*Litopenaeus vannamei*	1.6
Unigene3163_All	inhibitor of apoptosis protein	*Litopenaeus vannamei*	1.02
Unigene9776_All	Caspase	*Fenneropenaeus merguiensis*	3.08
Unigene36052_All	DNA fragmentation factor subunit beta	*Danio rerio*	3.26
Other immune genes			
CL4920.Contig1_All	Calmodulin	*Procambarus clarkii*	1.92
CL1939.Contig4_All	Ferritin	*Litopenaeus vannamei*	-2.87
CL4071.Contig2_All	Peritrophin	*Macrobrachium nipponense*	-2.18
Unigene10835_All	Selenoprotein W	*Cricetulus griseus*	-1.98
Unigene26739_All	Metallothionein I	*Macrobrachium rosenbergii*	-2.75
Unigene29279_All	Selenoprotein L	*Ciona intestinalis*	1.06
CL6324.Contig2_All	Arginine kinase 1	*Macrobrachium rosenbergii*	4.67
CL5294.Contig2_All	Hemocyanin	*Macrobrachium nipponense*	1.21

Category or gene id	Homologues function	Species	FC*
Unigene26326_All	Tetraspanin-like protein CD9	*Fenneropenaeus chinensis*	1.02
CL4316.Contig1_All	Crustacyanin-like lipocalin	*Macrobrachium rosenbergii*	-4.12
CL6297.Contig2_All	E cadherin	*Tenebrio molitor*	3.19
CL2339.Contig2_All	Adenosine deaminase	*Caligus rogercresseyi*	-1.35
CL3196.Contig2_All	Calnexin	*Penaeus monodon*	1.57
CL2701.Contig1_All	Ubiquitin-conjugating enzyme	*Macrobrachium nipponense*	1.33

*Fold changes (Log2 ratio) in gene expression.PRPs-pattern recognition proteins, ProPO- prophenoloxidase.

3.5 Discussion

Apart from viral diseases, *Vibrio* infections causing Vibriosis is another factor hindering the shrimp aquaculture industry worldwide (Tonguthai, 1995). This fatal disease has contributed to mass mortality and severe economic losses in India, Thailand, Philippines, Japan, Ecuador, Peru, Colombia and Central America (Lightner, 1996). Knowledge about the interaction between *M. rosenbergii* and *Vibrio* species is in its infancy, and in-depth study is urgently needed to address this issue. Discovery of the molecular mechanisms surrounding the innate immune system against *Vibrio* infection in freshwater prawns should be beneficial to both scientific research and the aquaculture industry. Identifying and quantifying immune-related gene expression on a large scale is a promising method to investigate the host response against pathogens and provide a platform for further studies in this area.

Microarray and EST analyses have long been used to study the molecular mechanisms underlying the innate immune system and to identify genes aberrantly expressed during infection (Aoki *et al.,* 2011; Gross *et al.,* 2001; Supungul *et al.,* 2002). However, the most recent NGS platforms such as the Illumina HiSeq^{TM} 2000 appear much better at quantifying transcripts expressed at low levels than microarrays or EST analysis (Asmann *et al.,* 2008). This is because this revolutionary technique verifies direct transcript profiling without compromise or bias, unlike previous methods (Reis-Filho, 2009). Furthermore, this technology has been successfully used in transcriptome profiling studies on non-model organisms where there is no complete genome database (Feldmeyer *et al.,* 2011; Garg *et al.,* 2011; Xu *et al.,* 2013). The introduction of NGS technology has led to various studies on host-viral interactions in shrimps to identify potential immune-related genes (Chen *et al.,* 2013; Xue *et al.,* 2013; Zeng *et al.,* 2013)– but not so far on

the interaction between the freshwater prawn and *Vibrio* species. To our knowledge, this study could be the very first to use the Illumina HiSeq^{TM} 2000 platform to explore the immune-related gene response in *M. rosenbergii* against *V. parahaemolyticus*.

Taking advantage of the Illumina HiSeq^{TM} 2000 platform's capability to sequence with a high throughput data providing more candidate genes, the total RNA extracted at the 12th hour time point from a pool of control and infected hepatopancreases was sequenced and assembled using the Trinity assembler. The overall analysis yielded 14,549 differentially expressed unigenes, with 11,446 unigenes significantly up-regulated and 3,103 unigenes significantly down-regulated. The sequencing data analyses obtained clearly showed a significant impact of *V. parahaemolyticus* infection on the *M. rosenbergii* transcriptome.

M. rosenbergii possesses an innate immune system, consisting of cellular and humoral components which work individually or cooperatively to protect the species from invading pathogens such as *V. parahaemolyticus* (Jiravanichpaisal *et al.,* 2006; Young Lee & Söderhäll, 2002). This immune response is activated when the animal detects the invading pathogen through pattern recognition proteins (PRPs)(Akira *et al.,* 2006). Two important PRPs molecules identified in shrimp are lectins and the beta-1,3-glucan binding protein (LGBP) (Kumar *et al.,* 2011; Dong*, et al.,* 2007; Ma *et al.,* 2007; Sritunyalucksana *et al.,* 2002). Besides pathogen recognition, lectins are also involved in phagocytosis through opsonisation in crustaceans (Marques & Barracco, 2000). Our data showed that challenge by *V. parahaemolyticus* greatly affects the expression of PRPs, as observed in previous studies using different *Vibrio* strains (Adams, 1991; Cheng *et al.,* 2005; Liu *et al.,* 2009; Soonthornchai *et al.,* 2010; Sun *et al.,* 2008). The activation of LGBP molecules triggers the melanisation process, a prophenoloxidase-activating system

(proPO-AS) which is an enzymatic cascade involving several enzymes, including the key enzyme phenoloxidase (PO) (Amparyup *et al.,* 2012; Soltanian *et al.,* 2009; Vargas-Albores & Yepiz-Plascencia, 2000). The active PO converts phenols into quinones. These build a non-specific crosslink between neighbouring molecules to form melanin, which provides defence against invading microorganisms (Perazzolo & Barracco, 1997). Increased activity of prophenoloxidase against *Vibrio* species has also been noted in *F. indicus* (Sarathi *et al.,* 2007), *L. vannamei* (Liu *et al.,* 2004) and *P. monodon* (Nayak *et al.,* 2010).

Antimicrobial peptides (AMPs) play a pivotal role in killing or clearing infected pathogens, especially *Vibrio* species (Bachère, 2003). Notable shrimp AMPs, such as penaeidins, lysozymes, crustins, anti-lipopolysaccharide factors (ALFs) and stylicins have been identified and characterized previously in shrimps (De la Vega *et al.,* 2008; Destoumieux *et al.,* 1997; Rolland *et al.,* 2010; Tassanakajon *et al.,* 2011; Zhang *et al.,* 2007). However, only four types of AMPs – lysozyme, crustin, NF-kappa B inhibitor alpha and ALF (3 isoforms) – were detected in our transcriptome data and found to be highly expressed. The up-regulation of these AMPs correlated with previous studies showing their antimicrobial properties against *Vibrio* species and other bacteria (Arockiaraj *et al.,* 2012; Arockiaraj *et al.,* 2013; Burge *et al.,* 2007; Somboonwiwat *et al.,* 2005). Blood clotting is vital in crustaceans to prevent excess blood loss from a wound and prevent micro-organisms from invading the wound (Maningas *et al.,* 2008). We found four molecules of the blood clotting system – transglutaminase, clottable protein, proclotting enzyme and coagulation factor XII – to be highly induced in our transcriptome data after challenge by *V. parahaemolyticus*. A similar expression of these molecules after bacterial challenge has been reported in previous studies (Wang*, et al.,* 2007; Sarathi *et al.,* 2007; Yeh *et al.,* 2009).

Stress conditions such as bacterial infections lead to an accumulation of reactive oxygen species (ROS) in a cell (Bandyopadhyay *et al.,* 1999). Increased levels of ROS causes oxidative damage to important cellular macromolecules (lipids, proteins, carbohydrates and nucleotides) which are components of the membranes, cellular enzymes and DNA (Yu, 1994). In order to restrict the production of ROS, antioxidant genes are activated to produce antioxidant enzymes which eliminate ROS. Several antioxidant enzymes have been isolated and characterised in the penaeid shrimp in previous studies (Cheng *et al.,* 2006; Zhang *et al.,* 2008; Zhou *et al.,* 2009). In this study, we found six antioxidant unigenes to be over-expressed after *V. parahaemolyticus* challenge – the exception being glutamine synthetase. High expression levels of these genes have similarly been observed in other shrimps and scallops after *Vibrio* challenge (Li *et al.,* 2008; Li *et al.,* 2006; Ren *et al.,* 2009; Tian *et al.,* 2011). The up-regulation of actin and tubulin genes play a crucial role in a wide range of cellular functions such as nodule formation, phagocytosis, encapsulation, as well as cell shape change, cell motility and adhesion, all of which may aid in clearing the pathogen (He *et al.,* 2004).

Heat shock proteins (HSPs) are highly generated when induced by stress. They are known to play a major role in protein folding, the protection of proteins from denaturation or aggregation, and aiding protein transport through membrane channels (Horváth *et al.,* 2008; Vabulas *et al.,* 2010). In addition to molecular chaperones, HSPs have been reported to play important roles in innate immune responses, and have been well studied in crustaceans (Cui *et al.,* 2010; Jiang *et al.,* 2009; Lo *et al.,* 2004). In this study, we noted higher levels of expression of all heat shock proteins in *M. rosenbergii* when challenged by *V. parahaemolyticus*. The increased expression of these HSPs is in line with previous reports, which tends to confirm the important role of these proteins in protecting this species from the stress induced by *Vibrio* challenge (He *et al.,* 2004;

Rungrassamee *et al.,* 2010; Zhou *et al.,* 2010).The general higher expression of proteinases and their inhibitors was to be expected in our data, as these are known to modulate elements of the innate immune system such as haemolymph coagulation, antimicrobial peptide synthesis, cell adhesion, and melanisation (Robalino *et al.,* 2007).

Bacteria like *Vibrio* are known to induce cell apoptosis through a variety of mechanisms such as pore-forming proteins, secretion of protein synthesis inhibitors, molecules activating the endogenous death machinery in an infected cell, lipopolysaccharides, and other superantigens (Weinrauch & Zychlinsky, 1999). Increased levels of apoptosis contribute to the degradation of DNA and RNA, which may contribute to shrimp mortality (Chen & Zychlinsky, 1994). In our transcriptome data, an up-regulation of genes involved in apoptosis was observed, similar to the trends reported in previous studies (Chang *et al.,* 2008; Fall *et al.,* 2010; Nayak *et al.,* 2010). Apoptosis may also serve as host defence against bacterium by allowing other healthy cells to phagocytise apoptotic bodies containing bacteria from target cells, effectively clearing the pathogen (Finlay & McFadden, 2006).

The signalling pathways involved in the *M. rosenbergii* innate immune response against *V. parahaemolyticus* were observed to be highly induced in our transcriptome data. The Toll protein initially identified in *Drosophila* had been reported to play a key role in the anti-fungal and anti-Gram-positive bacterial responses of flies in the Toll pathway (Belvin & Anderson, 1996). Other Toll components have also been found to be activated in penaeid shrimps when challenged with *Vibrio* species (Fall *et al.,* 2010; Li *et al.,* 2010; Yang *et al.,* 2008). In *C. elegans*, the mitogen-activated protein kinase (MAPK) pathways are transcriptionally up-regulated by the pore-forming toxin released by *B. thuringiensis*, which provide a cellular defence against this toxin (Huffman *et al.,* 2004).

This could explain the higher expression of this signalling pathway in our data, as *V. parahaemolyticus* is known to release a thermostable direct haemolysin, which is a pore-forming toxin (Honda *et al.,* 1992). The Janus kinase (JAK) and signal transducer and activator of transcription (STAT) pathways have been reported to be activated when *F. chinensisis* challenged with *V. anguillarum,* which suggest that these pathways are important for immune responses against bacteria (Sun *et al.,* 2011). In addition, Rab-related proteins have been reported to regulate the hemocytic phagocytosis of bacteria in *M. japonicas* (Zong *et al.,* 2008).

Several genes in the other immune gene group were found to be aberrantly expressed in our transcriptome data. Calmodulin, which plays an important role in calcium-dependent signal transduction pathways, was over-expressed in our data – as was the case in *L. vannamei* when challenged with *V. parahaemolyticus* (Ji *et al.,* 2011). Ferritin, an iron storage protein crucial for the metabolism of iron and maintaining iron homeostasis in a cell, was found to be down-regulated. The reduced expression of this gene could possibly lead to prawn mortality, as increased expression of this gene has been found to protect *P.monodon* from *V. harveyi* (Maiti *et al.,* 2010). Arginine kinase (AK), a phosphagen kinase in the invertebrate energy metabolism, has previously been reported to play an immune role against viral infection (Arockiaraj *et al.,* 2011a). However, we observed a higher expression of this gene in *M. rosenbergii* after challenge by *V. parahaemolyticus,* which could suggest that AK plays a similar role in bacterial infection. Haemocyanin, an important immune gene in crustaceans, is involved in prophenoloxidase activity (Yan *et al.,* 2008). It has antiviral properties against WSSV(Zhang *et al.,* 2004), and its increased expression in our transcriptome data tends to bear out its importance as a defence molecule against challenge by *V. parahaemolyticus*. Finally, metallothioneins, a metal-binding protein, was found to be highly expressed in our transcriptome data. The

increased expression of this gene was to be expected, as it is known as a scavenger of reactive oxygen intermediates and generally shows higher expression levels during immune responses in invertebrates against pathogens (Gross *et al.,* 2001).

3.6 Conclusion

We utilised the Illumina HiSeq^{TM} 2000 platform and Trinity assembler package to perform a *de novo* transcriptome profiling of the hepatopancreases isolated from *M. rosenbergii* challenged with *V. parahaemolyticus.* The differential expression analysis between *V. parahaemolyticus*-infected and control groups revealed significant differences in the gene expression with 11,446 unigenes found to be significantly up-regulated and 3,103 unigenes observed to be significantly down-regulated. This study provided a valuable insight into antibacterial mechanisms of freshwater prawn against *V. parahaemolyticus* with majority of the differentially expressed unigenes were grouped into 11 animal immune system categories. Furthermore, this study has generated an abundant list of transcript from *M. rosenbergii* which will provide a fundamental basis for future genomics research in this field

CHAPTER 4

A TRANSCRIPTOME STUDY ON *MACROBRACHIUM ROSENBERGII* HEPATOPANCREAS EXPRERIMENTALLY CHALLENGED WITH WHITE SPOT SYNDROME VIRUS (WSSV)

4.1 Introduction

The world production of shrimp such as the Malaysian giant freshwater prawn, *Macrobrachium rosenbergii* is seriously affected by the white spot syndrome virus (WSSV). There is an urgent need to understand the host pathogen interaction between *M. rosenbergii* and WSSV which will be able to provide a solution in controlling the spread of this infectious disease and lastly save the aquaculture industry. Now, using Next Generation Sequencing (NGS), we will be able to capture the response of the *M. rosenbergii* to the pathogen and have a better understanding of the host defence mechanism.

Two cDNA libraries, one of WSSV-challenged *M. rosenbergii* and a normal control one, were sequenced using the Illumina HiSeq^{TM} 2000 platform. After *de novo* assembly and clustering of the unigenes from both libraries, 63,584 standard unigenes were generated with a mean size of 698 bp and an N_{50} of 1137 bp. We successfully annotated 35.31% of all unigenes by using BLASTX program (E-value $<10^{-5}$) against NCBI non-redundant (Nr), Swiss-Prot, Kyoto Encyclopedia of Genes and Genome pathway (KEGG) and Orthologous Groups of proteins (COG) databases. Gene Ontology (GO) assessment was conducted using BLAST2GO software. Differentially expressed genes (DEGs) by using the FPKM method showed 8,443 host genes were significantly

up-regulated whereas 5,973 genes were significantly down-regulated. The differentially expressed immune related genes were grouped into 15 animal immune functions. The present study showed that WSSV infection has a significant impact on the transcriptome profile of *M. rosenbergii*'s hepatopancreas, and further enhanced the knowledge of this host-virus interaction. Furthermore, the high number of transcripts generated in this study will provide a platform for future genomic research on freshwater prawns.

4.2 Literature Review

Viral diseases remain a thorn in the side of the shrimp aquaculture industry. Among the viruses reported for shrimps, the white spot syndrome virus (WSSV) stands out as the most devastating, causing high mortality and severe economic losses in the shrimp farming industry throughout the world (Escobedo-Bonilla *et al*., 2008; Flegel, 1997). WSSV, which was first reported in Fujian, China, in 1992, is classified under the genus *Whispovirus* of the virus family *Nimaviridae* (Marks *et al.*, 2004). This deadly virus has an enveloped, non-occluded, rod-shape that contains a circular, double-stranded DNA of about 300 kb (Chen *et al*., 2002; van Hulten *et al*., 2001).The virus has a diverse range of potential hosts such as marine shrimps, crayfish, crab, spiny lobsters, insects and freshwater prawns (Imada & Leonard, 2000; Lo *et al.*, 1996; Wang *et al.*, 1998).

The giant freshwater prawn, *Macrobrachium rosenbergii,* is an economically important crustacean, being cultured on a large-scale in different parts of the world. *M. rosenbergii* is native to Malaysia and other Asian countries, including Vietnam, Cambodia, Thailand, Myanmar, Bangladesh, India, Sri Lanka, and the Philippines (Hameed, 2009). Being popular for its delicious flesh and high nutritive value, demand for this species has escalated, generating revenue worth US$1 billion per annum in Asia alone (Schwantes *et al.*, 2009). *M. rosenbergii* is generally considered less prone to

disease in culture when compared to the penaeid shrimp (Bonami & Sri Widada, 2011). However, with exception of the adult stage, the other developmental stages of *M. rosenbergii* were found to be more susceptible to WSSV (Kiran *et al.*, 2002).

M. rosenbergii, like other crustaceans, possesses an innate immune system which provides defence against pathogenic agents (Gross *et al.*, 2001). A greater understanding of the immune system in this species is necessary to develop methods to control and minimize the loss of production due to infectious diseases. The understanding of host pathogen can be achieved by studying the transcriptome or the complete repertoire of expressed RNA transcripts. Transcriptome profiling offers a strategy to assess the relative importance of gene products in any chosen cell, tissue, organism, or given condition (Mu *et al*., 2010). Considerable progress has been made to understand the molecular mechanism of the shrimp-virus interaction using conventional transcriptome techniques such as suppression subtractive hybridization (SSH), differential hybridization (HD), microarrays and expressed sequence tag analysis (EST) (James *et al.*, 2010; Leu *et al.*, 2007; Leu *et al.*, 2011; Nanhai *et al.*, 2005; Pongsomboon *et al.*, 2011; Wang *et al.*, 2006; Zhao *et al.*, 2007). However, microarrays and subtractive hybridization methods are hindered by background and cross-hybridization problems, and measure only the relative abundance of transcripts (Hoen *et al*., 2008; Moody, 2001). Furthermore, only predefined sequences can be detected using this approach (Watson *et al.*, 1998). The EST sequencing technique is moreover laborious; and it has limitations in the depth of the transcriptome that can be sampled, which could possibly prevent the detection of transcripts with low abundance (Hanriot *et al*., 2008).

A far superior technology that has revolutionised modern genomics research, known as 'Next Generation Sequencing' (NGS), was introduced in 2004. This platform

provides a faster and cheaper way to sequence large amounts of data with greater depth and breadth. At present, the three dominant commercial platforms available are the Illumina Genome Analyzer, the Roche/454 Genome Sequencer FLX Instrument, and the ABI SOLiD System (Metzker, 2010). This sequencing technology has a wide range of applications, such as genome sequencing, metagenomics, epigenetics and the discovery of non-coding RNAs, protein-binding sites and transcriptomics analysis (Morozova & Marra, 2008). Transcriptome analysis using the NGS approach has been successfully utilised to identify immune-related genes in marine shrimps against the Taura Syndrome Virus (TSV) and WSSV (Chen *et al.*, 2013; Li *et al.*, 2013; Xue *et al.*, 2013; Zeng *et al.*, 2013).

At present, there is limited evidence on the immune-related genes of *M. rosenbergii* which are primarily involved in WSSV infection. In this study, we performed a transcriptome analysis from the hepatopancreas of *M. rosenbergii* experimentally infected with WSSV, using a high-throughput sequencing method (Illumina HiSeq^{TM}2000). The aim was to identify candidate immune-related genes involved in WSSV infection which can provide a better insight into the virus-shrimp interaction between *M. rosenbergii* and WSSV. In addition, the high-throughput sequencing method revealed a large number of new transcripts, which could point the way for future genomics research into the freshwater prawn.

4.3 Materials and Methods

4.3.1 *M. rosenbergii* and WSSV challenge

M. rosenbergii juvenile prawns (5-8g body weight) (equal number of males and females) were purchased from a hatchery at Kuala Kangsar, Perak, Malaysia. The prawns were

acclimatized for 1 week prior to being challenged by WSSV in flat-bottomed glass tanks (300 L) with aerated and filtered freshwater at 28 ± 1 ºC in the laboratory. The prawns were randomly sampled and tested by PCR to ensure they were free from WSSV. In the challenge experiment, two prawn groups were used, each of 10 prawns: one WSSV challenge group and one negative control group. The challenged group were intramuscularly injected with 100 µl filtered supernatant obtained from WSSV infected *Penaeus monodon* ($1X10^5$copies/ml). Similarly, the negative control group were injected with 2% NaCl (1:10, w/v) solution. At 12 hours after infection, the hepatopancreas tissues of the prawns were dissected and immediately frozen in liquid nitrogen before storage at -80ºC until RNA extraction. Our previous studies working on immune related genes from *M. rosenbergii* in response to infectious hypodermal and hematopoietic necrosis virus (IHHNV) showed significant gene expression at 12 hour time point, hence this time point was chosen for this study (Arockiaraj *et al.*, 2012b, 2012c; Arockiaraj*, et al.*, 2011a; Arockiaraj *et al.*, 2011b)

4.3.2 Total RNA Extraction

Total RNA (~20 mg) was extracted from the WSSV-challenged and negative control group hepatopancreases using a Macherey-Nagel NucleoSpin RNA II extraction kit in accordance with the manufacturers' protocols. This was then dissolved in diethylpyrocarbonate (DEPC) treated water, to give a final concentration of >750ng/ml. The integrity and size distribution of the dissolved RNA samples was checked with Agilent 2100 Bioanalyzer (Agilent technologies, USA) before storage at -80ºC. The total RNA samples were equally pooled from 10 prawns in each negative control and WSSV-challenged groups prior to cDNA synthesis and sequencing.

4.3.3 Illumina Sequencing

The cDNA library preparation and sequencing were conducted using Illumina sequencing technology at the Beijing Genome Institute, Shenzhen, China. A Sera-mag Magnetic Oligo (dT) Beads (Illumina) was used to isolate poly A+ mRNA from the total RNA. Small fragments of extracted RNA (~150-200 bp) were generated from the extracted RNA to avoid priming bias. These small fragments of RNA were transcribed into double stranded cDNA using reverse transcriptase, RNase H (Invitrogen) and DNA polymerase I (New England BioLabs) primed with random hexamers. Two sequencing cDNA libraries of pooled WSSV-infected and pooled negative control hepatopancreas were prepared according to the manufacturer's protocols (Illumina), before being subjected to 90 bp paired-end sequencing using an Illumina HiSeq™ 2000 platform.

4.3.4 *De novo* Assembly and Assessment

After removing adaptor sequences, ambiguous 'N' nucleotides (with a ratio of 'N' more than 10%) and low quality sequences (with a quality score less than 20), the remaining clean reads were assembled using Trinity software (Haas *et al*., 2013). The Trinity software combines three independent software modules: Inchworm, Chrysalis, and Butterfly, applied sequentially to process large volumes of RNA-seq reads. Trinity partitions the sequence data into many individual de Bruijn graphs, each representing the transcriptional complexity at a given gene or locus, and then processes each graph independently to extract full-length splicing isoforms and to tease apart transcripts derived from paralogous genes. The unigenes from both groups were then clustered using the TIGR Gene Indices clustering tools (TGICL) (Pertea *et al*., 2003), with default parameters. Bacterial sequence contamination was investigated using the web-based

version of DeconSeq (Schmieder & Edwards, 2011), with a query coverage and sequence identity threshold of 90%.

4.3.5 Protein Coding, Gene Identification and Classification

The clustered unigene sequences were annotated using homology search (BLASTX) (Mount, 2007) with an E-value cut-off of 10^{-5} against an NCBI non-redundant (Nr) database, Swissprot (Boeckmann *et al.*, 2003), Cluster of Orthologous Groups database (COG) (Tatusov *et al.*, 2000) and Kyoto Encyclopedia of Genes and Genome (KEGG) database (Kanehisa & Goto, 2000). Gene Ontology (GO) (Ashburner *et al*., 2000) assignment was conducted using BLAST2GO software (Conesa *et al*., 2005) with default parameters. The coding sequence and the direction of the annotated unigenes were determined based on the BLAST results from the four above mentioned databases. For those sequences without annotation, the coding sequence and direction were predicted using an ESTscan (Iseli *et al.*, 1999),with BLAST predicted coding sequence data as the training set.

4.3.6 Identification of Differentially Expressed Unigenes

The transcript expression levels of the unigenes were measured using the FPKM method (Fragments Per kb per Million fragments). The calculation of Unigene expression uses FPKM method (Fragments Per kb per Million fragments), the formula is shown below:

$$FPKM = \frac{10^6 C}{NL/10^3}$$

Set FPKM to be the expression of Unigene A, and C to be number of fragments that uniquely aligned to Unigene A, N to be total number of fragments that uniquely aligned to all Unigenes, and L to be the base number in the CDS of Unigene A. The FPKM method is able to eliminate the influence of different gene length and sequencing level on the calculation of gene expression. Therefore the calculated gene expression can be directly used for comparing the difference of gene expression between samples. A FDR (False Discovery Rate) of <0.001 was used as the threshold of the p-value in multiple tests to judge the degree of differences in gene expression (Reiner *et al.,* 2003). FDR (False Discovery Rate) control is a statistical method used in multiple hypothesis testing to correct for p-value. In practical terms, the FDR is the expected False Discovery Rate; for example, if 1000 observations were experimentally predicted to be different, and a maximum FDR for these observation was 0.1, then 100 out of these observations would be expected to be false discovered. Genes were considered differentially expressed in a given library when the p-value was less than 0.001 and a greater than two-fold change in expression across libraries was observed.

4.3.7 Phylogenetic Analyses

Three immune genes, alpha-2-macroglobulin (Unigene12869_All), catalase (CL5831.Contig2_All) and heat shock protein 70 (Unigene14757_All) were randomly selected for phylogenetic analysis by using Maximum likelihood tree on MEGA 6 software (Tamura *et al.*, 2013). The node support was assessed using a bootstrap procedure based on 10,000 replicates.

4.3.8 Quantitative RT-PCR Analysis

Validation of our Illumina sequencing results involved the selection of ten differentially expressed *M. rosenbergii* unigenes (arginine kinase 1, anti-lipopolysaccharide factor, inhibitor of apoptosis protein, caspase, chaperonin, heat shock protein 21, lectin 1, manganese superoxide dismutase, NF-kappa B inhibitor alpha and peroxiredoxin) for quantitative RT-PCR analysis (qRT-PCR). Primers for qRT-PCR were designed using Primer3 software (Untergasser, A *et al*., 2012) (Table 4.1). The first strand of cDNA was synthesized from 1µg of RNA (similar sample used for transcriptome sequencing) using the ImProm-II™ Reverse Transcription System (Promega). The qRT-PCR reaction (20 µl) consisted of 10 µl TaqMan Universal RT-PCR Master Mix (Applied Biosystems, Foster City, CA, USA), 1 µl of primers/probe set containing 900 nM of forward reverse primers , 300 nM probe and 2 µl of template cDNA. The RT-PCR program consisted of incubation at 50°C for 2 min, and 40 cycles at 95°C for 10 min, 95°C for 15 sec, and 60°C for 1 min with the Step One Plus Real-Time PCR System® (Applied Biosystems). Concurrently, elongation factor 1-alpha primers were applied to normalize the amount of template cDNA added to each reaction (Dhar *et al.*, 2009). Each sample was analysed in triplicate and the comparative CT method ($2^{-\Delta\Delta CT}$ method) was used to obtain the expression level of the ten immune genes (Schmittgen & Livak, 2008).

Table 4.1. List of genes, primers and probes used in real-time TaqMan PCR assay.

Gene ID	Gene	Gene Abbreviations	NCBI NO	Forward Primer Sequence	Reverse Primer Sequence	Probe Sequence
Unigene22686_All	Arginine Kinase	AK1	NA	GTCTGGTGATCGCAACCTTCA	GTAGATACCGCGGCCTTCAG	TCGCGCTCCATACCTG
Unigene4120_All	Anti-lipopolysaccharide factor	ALF	NA	TGAAGCTCAGGGTTGGGAAGT	ATACCATTTGGTCGTCCACCC	GTTGCCGAAAAGATC
CL6019.Contig2_All	Inhibitor of apoptosis protein	IAPs	NA	CGCACACTATCAAGTCCACAAAATT	GCAGCATTATTGTTAGCTTCAGTAGCT	CAAGGCTGGCACCACC
CL6385.Contig2_All	Caspase	Casp	NA	GAGCAGATCCAGCGATTCTTCA	AACACACACAGCTAAACAAGATACGA	CTGACACGTAAATTTT
Unigene13644_All	Chaperonin	Chap	NA	ATTGGTGATGTCGCCACTCA	CACGTCCAGCACAAAACATGTC	CCTGTCTGCAAAGTAC
Unigene3736_All	Heat shock protein 21	HSP21	NA	ATCAGTCCCTCAACCAGACC	ATTGCTCTGCTGTGTAACCG	TCCCAGCAGACACAGCAAACACA
Unigene10978_All	Lectin 1	LT1	NA	ACTGGACCGACGCTAAGGA	CCCTATCAGATTGGCGTCATCAAAT	ACGGGCTGCATACTCA
Unigene19405_All	Manganese superoxide dismutase	MnSOD	NA	GGCGTGAAAGGGTCAGGTT	CGGCAACTTTGTCGTTCTTAGGA	TAGCCAAGCCAGCCCC
Unigene13073_All	NF-kappa B inhibitor alpha	NFκBI-α	NA	GCAAGACATCTGGTCGTTTGTG	GAGTGCAAGGGCGTGTTG	CCTGCGGTCTCTCG
Unigene20577_All	Peroxiredoxin	PRDX	NA	GGTGGACTGAGATCAAAGCGATATT	ACTGTCTTTCTAAATCACATTCAACGGT	CACCTTGCCATCTTCA
	Internal control					
	Elongation factor 1-alpha	EF1-A	EL609261.1	ACTGCGCTGTGTTGATTGTAGCT	ACAACAGTACGTGTTCACGGGTC	TACTGGTGAGTTTGAAGCT

NA= not available

4.4 Results

4.4.1 Illumina Sequencing and *De novo* Assembly

The mission to identify immune-related genes which are crucial for *M. rosenbergii* defence against WSSV infection involved preparing two cDNA libraries using pooled mRNAs obtained from the hepatopancreases of control and WSSV-infected groups. The cDNA libraries underwent sequencing using the Illumina HiSeq^{TM} 2000 platform, and the sequencing data were then subjected to *de novo* assembly using the Trinity program.

Once the adaptor sequence and the low-quality reads were filtered out, a total of 54,708,014 high quality clean reads of 90-bp in length were generated, consisting of 4,923,721,260 nucleotides for the control group and 38,671,944 reads with 3,480,474,960 nucleotides for the WSSV-infected group. The sequencing reads were found to be a high quality reads with the Average Q_{20} value of 97.76 % and their percentage of unknown nucleotide (N percentage) of 0 %. All sequencing reads were deposited into the Short Read Archive (SRA) of the National Centre for Biotechnology Information (NCBI), and can be accessed under the accession number SRR1424572 for control and SRR1424575 for WSSV-infected ones. The *de novo* assembly of the reads produced 95,645 contigs, with an N_{50} of 467bp (i.e. 50% of the assembled bases were grouped into contigs of 467bp or longer) for the control group and 97,151 contigs with an N_{50} of 426bp for the WSSV-infected group respectively. An overview of the sequencing and assembly is shown in Table 4.2.

Table 4.2. Summary of the control and WSSV-infected transcriptome sequencing.

	Control	WSSV Infected
Total number of reads	54,708,014	38,671,944
Total base pairs (bp)	4,923,721,260	3,480,474,960
Q_{20} value	97.73%	97.79%
Total number of contigs	95,645	97,151
Mean length of contigs (bp)	313	307
Total number of unigenes	59,050	65,625
Mean length of unigenes (bp)	479	427
NCBI Nr annotated	19,799	
Swiss-Prot annotated	16,382	
KEGG annotated	14,706	
COG annotated	7,856	
GO annotated	6,935	

The size distribution of the control and WSSV-infected contigs are shown in Appendix C. The control group had a total of 95,645 contigs, with a mean length of 313 bp and a majority of the contigs (55,489) falling between 100 to 200 bp, and 12,661 contigs with a length of more than 500 bp. The WSSV-infected group generated 97,151 contigs, with a mean length of 307 bp, 56.28% of the contigs (54,673) falling between 100 to 200 bp and 11,986 contigs greater than 500 bp. The contigs were further assembled into unigenes, which produced a total of 59,050 unigenes and an N_{50} of 685 bp in the control group, and 65,625 unigenes with an N_{50} of 525 bp in the WSSV-infected group.

The length distribution of the control and WSSV-infected unigenes is shown in Appendix C. The control group contained a total of 59,050 unigenes, with a mean length of 479 bp and with the bulk of the unigenes (73.77%) less than 500 bp in length and only

a small percentage (2.40%) longer than 2 kbp in length. The 65,625 unigenes from the WSSV-infected group had a mean length of 427 bp, with the majority (76.91%) less than 500 bp in length and only a tiny proportion (1.07%) longer than 2 kbp. The clustering of the unigenes from both libraries generated 63,584 standard unigenes with a mean size of 698 bp and an N_{50} of 1137 bp.

4.4.2 Similarity Searches

To achieve protein identification and gene annotation, a search was made on standard unigenes in the NCBI non-redundant (Nr) (19,799 unigenes, 31.14%), Swiss-Prot (16,832 unigenes, 26.47%), Kyoto Encyclopedia of Genes and Genome pathway (KEGG) (14,706 unigenes, 23.12%) and Orthologous Groups of proteins (COG) databases (7,856 unigenes, 12.36%) using the BLASTX program (E-value $<10^{-5}$). This BLASTX search yielded a total of 22,455 significant hits (35.31% of all unigenes). The size distribution profile for the coding sequences (CDS) and identified proteins are shown in Appendix. A GO annotation search using the Blast2GO resulted in 6,935unigenes (10.9%) being categorized into 54 functional groups. The unigenes with no hits in the BLASTX analysis were subjected to an ESTscan, producing 4,977 unigenes (7.82%) predicted to contain coding sequences. The size distribution of the ESTs and proteins are shown in Appendix.

The species distribution of the unigenes using the BLASTX results is shown in Figure 4.1. The *M. rosenbergii* unigenes produced 10.35 % matches against *D.pulex* sequences, and 6.06% matches with *T. castaneum* and *P.h. corporis* (4.23%). The unigenes were closely matched with *D.pulex* and *T. castaneum* probably due to their closer phylogenetic relationship and the vast amount of genomic information available. The remaining 66.51% of matched unigenes were similar to other species because of the limited genome information in crustaceans.

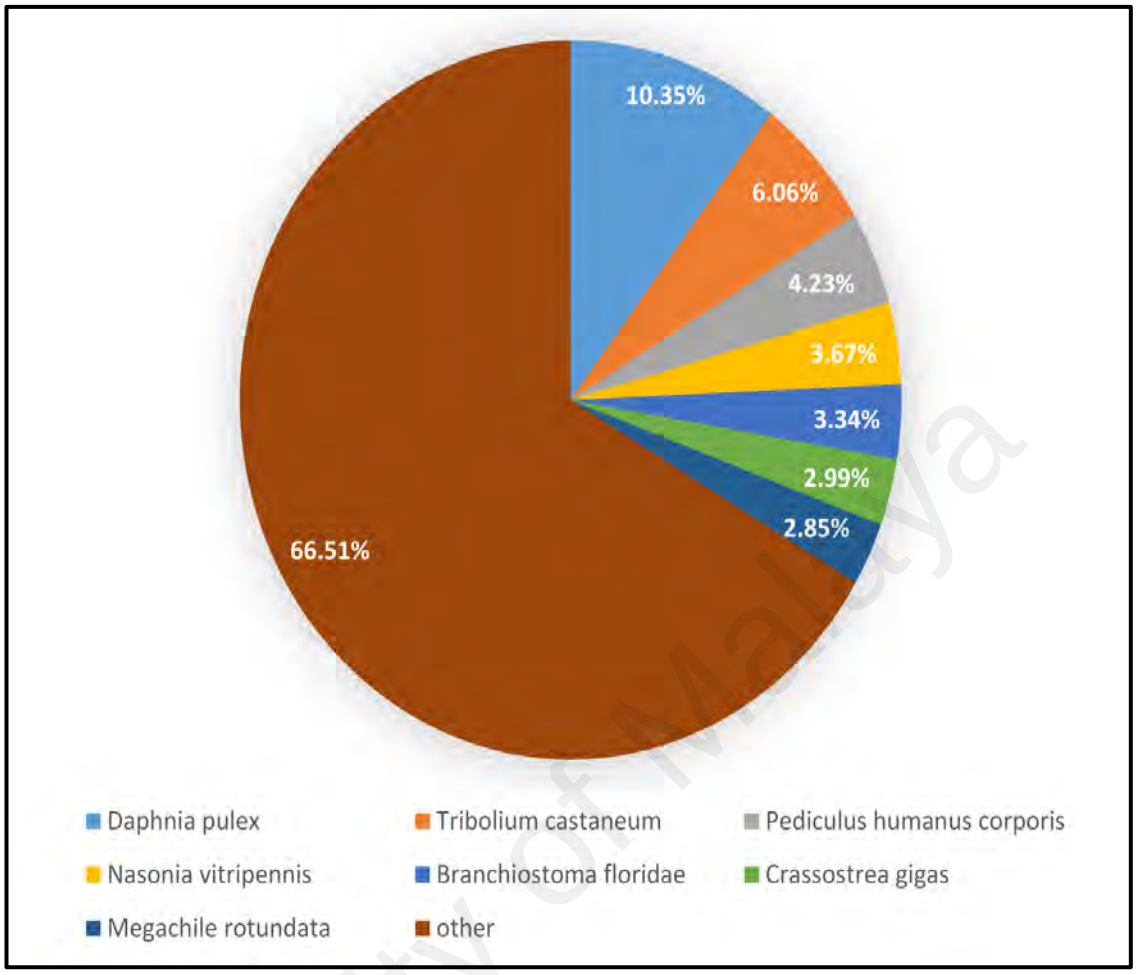


Figure 4.1. Species distribution of the BLASTX matches of the transcriptome unigenes. This figure shows the species distribution of unigene BLASTX matches against the Nr protein database (cut-off value E<10^{-5}) and the proportions for each species.

4.4.3 Functional Annotation

The COG is a database containing classified orthologous gene products. Every protein in COG is assumed to originate from an ancestor protein; thus the database is developed based on coding proteins with complete genomes as well as the system evolution relationships of bacteria, algae and eukaryotic creatures. The standard unigenes were aligned to the COG database to predict and classify their possible roles. A total of 7,856 unigenes had COG classifications, distributed among the 25 COG categories. The highest proportions of unigenes fell into the general function prediction only cluster (3,431, 43.67%), followed by the translation, ribosomal structure and biogenesis cluster (1,760, 22.40%) and then the replication, recombination and repair cluster (1,413, 17.98%). The smallest proportions of unigenes were grouped in the nuclear structure and extracellular structure clusters respectively (<1% in each cluster) (Figure 4.2).

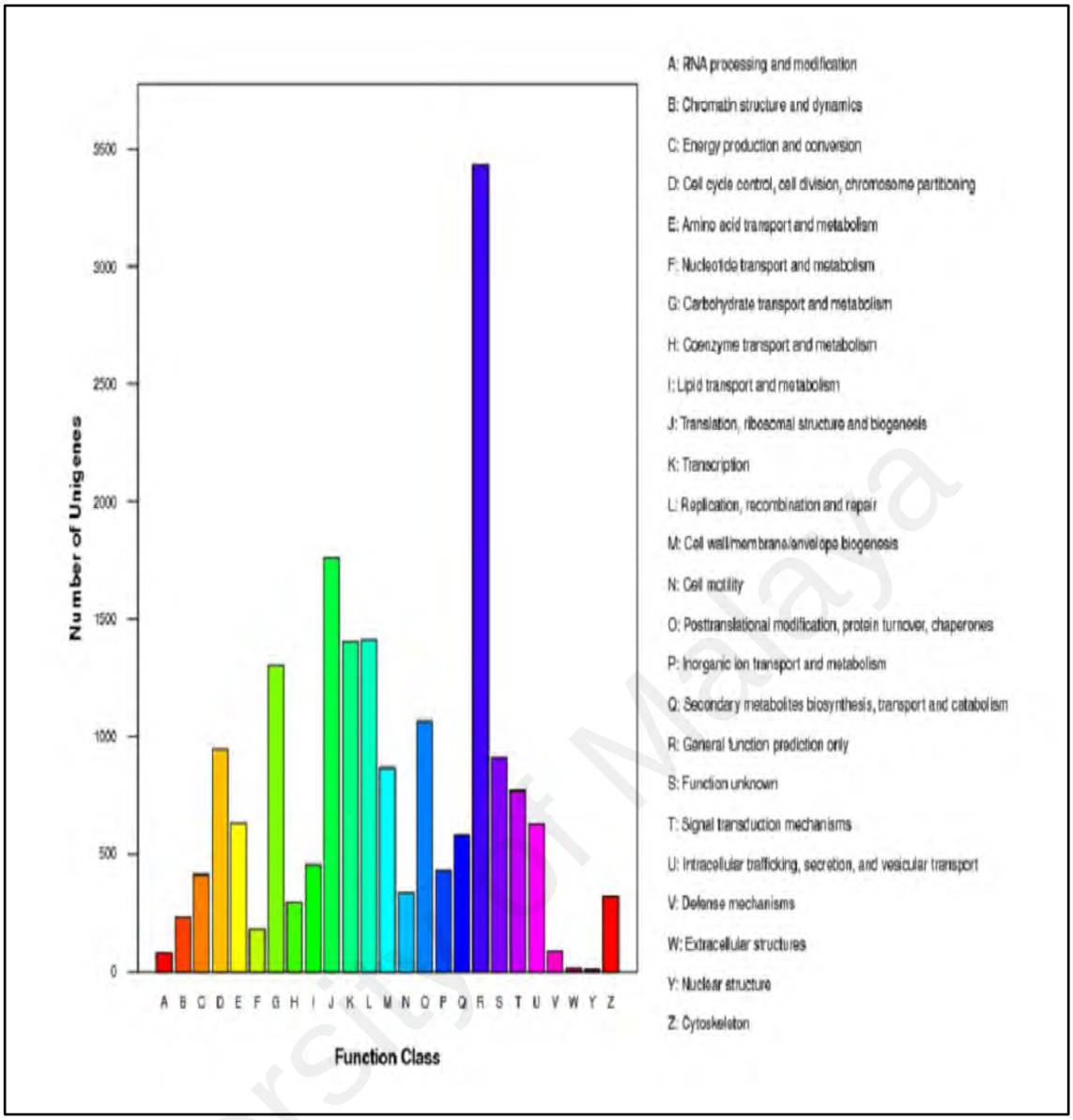


Figure 4.2. Histogram presentation of Cluster of Orthologus Groups (COG) classification of 7,856 known protein annotated unigenes. Each bar represents the number of unigenes classified into each of the 25 COG functional categories.

Gene Ontology (GO) is a unified gene functional classification system which offers a systematic and updated vocabulary in a strictly defined concept to describe the properties of genes and their products in any organism. GO has three ontologies: molecular function, cellular component and biological process. The basic unit of GO is 'GO-term', which is a type of ontology. The unigenes with Nr annotations were subjected

to GO annotations using the BLAST2GO program for functional unigene classification and to obtain the macro-level gene functional distribution for this species. The GO annotations produced 6,935 unigenes (Biological process: 2,477 unigenes; Cellular component: 1,896 unigenes; and Molecular function: 2,562 unigenes), which were assigned to 54 GO ontology (Figure 4.3). Most of the molecular function genes were involved in 'binding' and 'catalytic activity', whereas in the biological process categories the majority of the genes were grouped into 'cellular processes, 'metabolic processes' and 'single-organism processes'. A high percentage of genes fell under 'cell', 'cell part' and 'organelle' within the cellular component category.

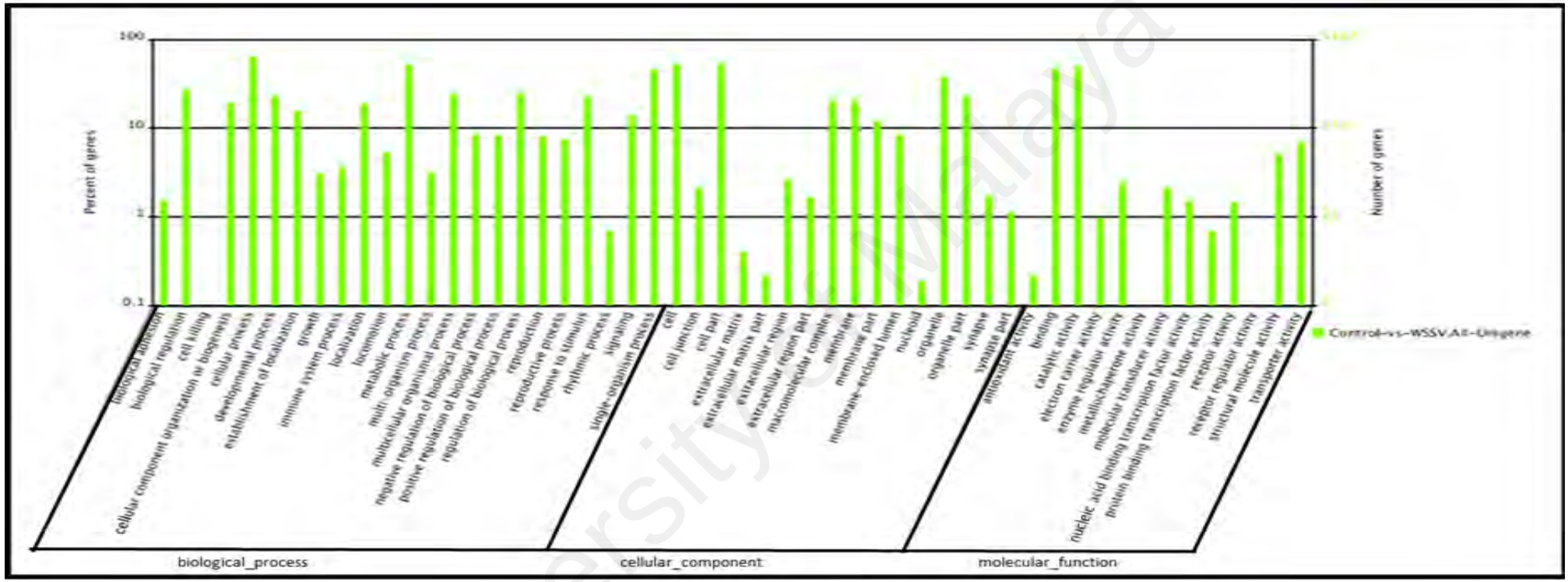


Figure 4.3. Gene ontology (GO) classification of the 6,935 protein annotated unigenes. Unigenes sequences were systematically classified into GO sub-categories under the biological process, cellular component and molecular function gene ontology catalogue system. Each bar represents the relative abundance of unigenes classified under each sub-category.

The understanding of biological functions and gene interaction can be achieved by performing a pathway-based analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) offers detailed information on networks related to intracellular molecular interactions and their organism-specific variations. The biological pathways involved in WSSV infection were identified when we mapped the annotated sequences to the reference canonical pathways contained in the KEGG database. A total of 14,706 unigenes were assigned into 253 pathways. The highly represented categories were ‘‘metabolic pathways’’ (2,192 members, 14.91%), “regulation of actin cytoskeleton” (557 members, 3.79%), “spliceosome” (509 members, 3.46%), “ubiquitin mediated proteolysis (508 members, 3.45%) and “RNA transport” (498 members, 3.39%). The least represented pathways with less than 10 unigenes found in each pathway were “biotin metabolism”, “vitamin B6 metabolism”, “cyanoamino acid metabolism”, “phenylalanine, tyrosine and tryptophan biosynthesis”, “D-Glutamine and D-glutamate metabolism” and “type I diabetes mellitus”. The top twenty of these KEGG biological pathway classifications are shown in Figure 4.4.

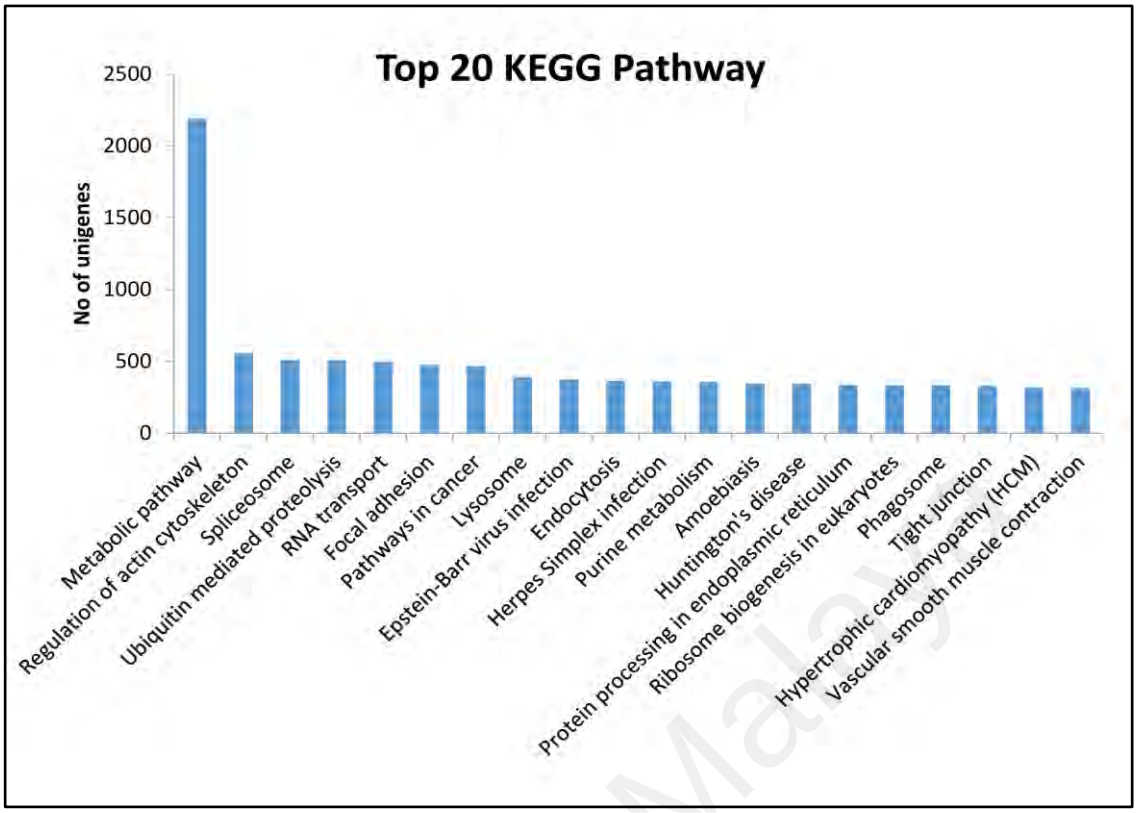


Figure 4.4. Top 20 KEGG biological pathway classification histograms for annotated unigenes. The x-axis shows pathways from KEGG classification and the y-axis shows the number of the matched unigenes.

4.4.4 Identification of Aberrantly Expressed Genes

In order to identify the genes expressed differentially between these two groups, a comparison of the relative transcript abundance in each unigene was carried out using the FPKM method (Fragments Per kb per Million fragments). A total of 14,416 unigenes were found to be aberrantly expressed; with 8,443 unigenes significantly up-regulated and 5,973 unigenes significantly down-regulated (Figure 4.5). The differentially expressed genes were annotated against the NR, Swiss-Prot, GO, COG and KEGG databases by BLASTX, with a cut-off E-value of 10^{-5}. The 9,469 (65%) unigenes containing low sequence homology to known sequences in public databases could represent non-coding

RNA, misassembled unigenes, or unknown genes of *M. rosenbergii* which responded to WSSV-infection.

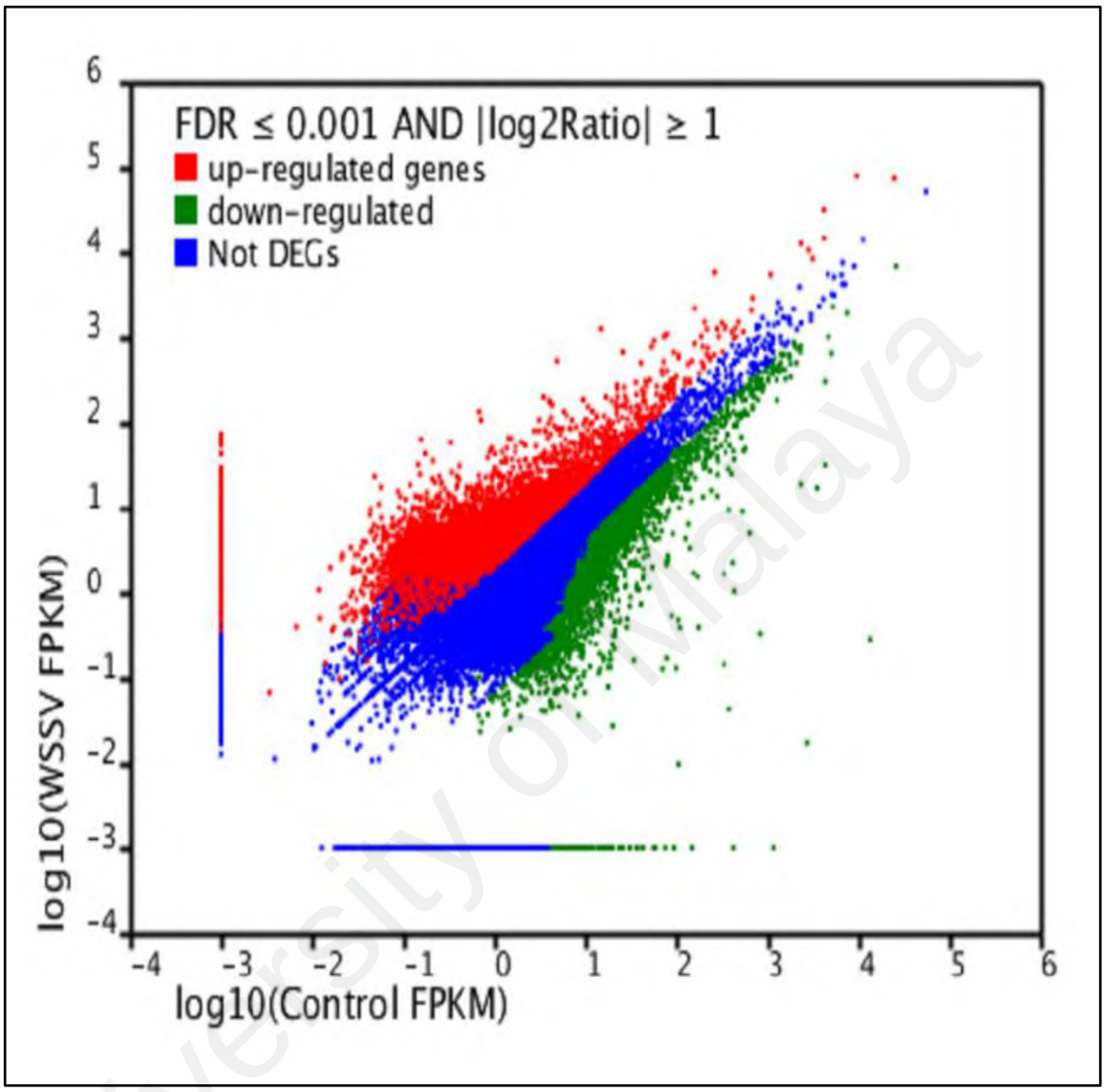


Figure 4.5. Digital gene expression between control group and WSSV infected group. Each point represents a unigene. The x- and y-axis are the log10 of the normalized expression level (FPKM) of unigene between the two groups. Red and green points indicate significant change at the absolute value of log2 (FPKM ratio in two groups) ≥1 and fdr =0.001. Red points indicate up-regulated unigenes and green points indicate down-regulated unigenes in the two groups which its expression level is represented by the y-axis. Blue points indicate insignificant differentially expressed unigenes.

4.4.5 Phylogenetic Analyses

The reason behind the phylogenetic analyses using the contigs or unigenes was to test the accuracy of assembly (Sadamoto *et al*., 2012). Maximum likelihood tree for three unigenes, alpha-2-macroglobulin (Unigene12869_All), catalase (CL5831.Contig2_All) and heat shock protein 70 (Unigene14757_All) were found to be closely related with other crustacean and was evolutionarily distant from that of other phylum represented in the tree (Figure 4.6).

A

B

Figure 4.6. The phylogenetic trees for alpha-2-macroglobulin (Unigene12869_All) (A), catalase (CL5831.Contig2_All) (B) and heat shock protein 70 (Unigene14757_All) (C). The phylogenetic trees include the sequences with species names and their accession numbers.

4.4.6 qRT-PCR Analysis

Ten unigenes were chosen randomly for quantitative real time-PCR (qRT-PCR) analysis, to validate the sequencing results. The resulting qRT-PCR results showed similar trends for all the genes to those of the sequencing data (Figure 4.7). For example, based on the Illumina sequencing analysis, arginine kinase 1, anti-lipopolysaccharide factor, inhibitor of apoptosis protein, caspase, heat shock protein 21, lectin 1, manganese superoxide dismutase, NF-kappa B inhibitor alpha and peroxiredoxin were up-regulated 5.33, 1.66, 2.26, 2.41, 4.61, 2.83, 1.61, 1.12, and 1.35 log2-fold changes respectively and 6.51, 2.5, 1.53, 1.8, 3.35, 2.05, 1.38, 2.68 and 1.8 log2-fold changes respectively in the qRT-PCR analysis. The chaperonin unigene meanwhile was down-regulated by -2.01 log2-fold changes in our transcriptome data, and was similarly found to be down-regulated by -3.33 log2-fold changes in the qRT-PCR analysis. Although the results from both analyses did not match perfectly, perhaps due to sequencing biases (Hoen *et al.,* 2008; Grabherr *et al.*, 2011), the qRT-PCR analysis substantially confirmed the direction of changes obtained from the Illumina sequencing analysis.

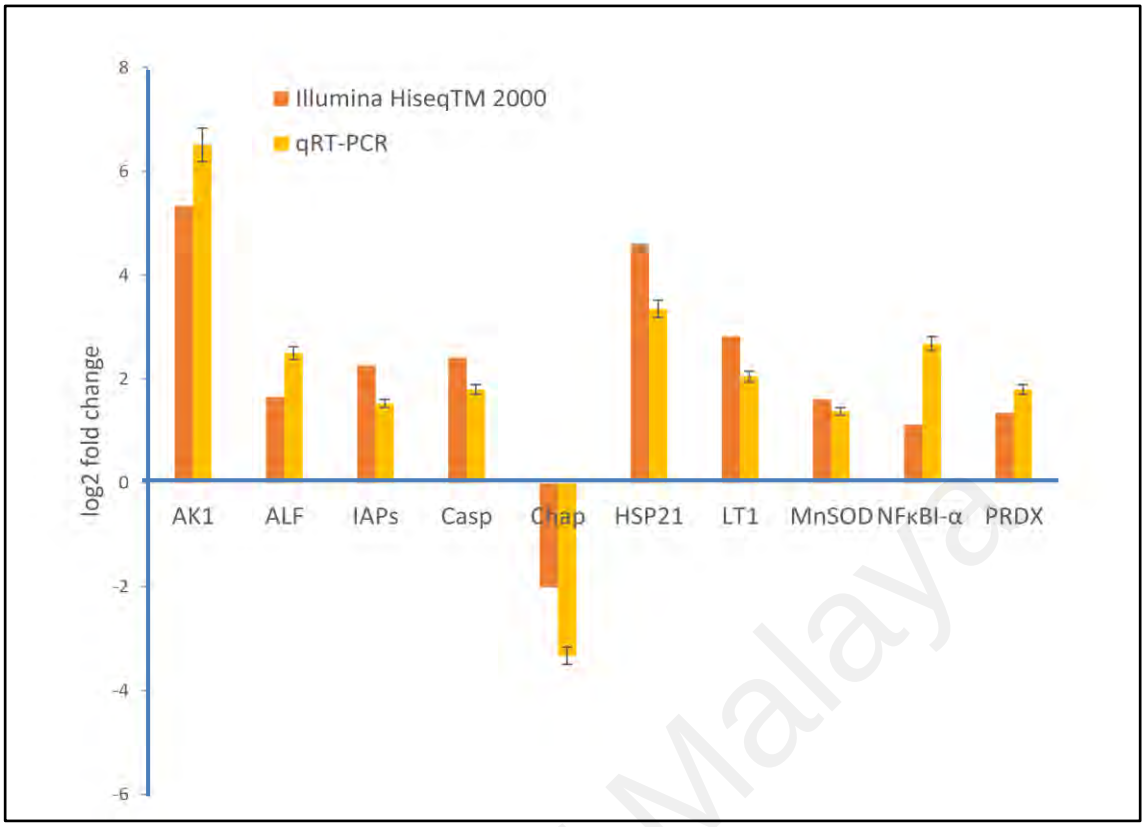


Figure 4.7. Comparison of the expression profiles of selected genes as determined by Illumina Hiseq™ 2000 sequencing (orange) and qRT-PCR (yellow). Target gene abbreviations are as follows: AK1 – arginine kinase, ALF- anti-lipopolysaccharide factor, IAPs- apoptosis inhibitor, Casp- caspase, Chap- chaperonin, HSP21- heat shock protein 21, LT1- lectin 1, MnSOD- manganese superoxide dismutase, NfκBI-α – NF-kappa B inhibitor alpha, PRDX- peroxiredoxin. Error bars indicated standard deviations of averages from three replicates.

4.4.7 Potential Immune-related Genes Involved in *M. rosenbergii* Immune Response

Numerous genes found to be aberrantly expressed in the WSSV-infected groups compared to the control groups were known to be part of various processes clustered under the animal immune system (Table 4.3). These immune genes are grouped into 15 functions, which include antiviral proteins (2 unigenes), antimicrobial proteins (4

unigenes), proteases (10 unigenes), protease inhibitors (3 unigenes), blood clotting system (2 unigenes), transcriptional control (4 unigenes), cell death (11 unigenes), cell adhesion (4 unigenes), heat shock proteins (6 unigenes), RNA silencing (3 unigenes), oxidative stress (7 unigenes), pathogen recognition immune receptors (9 unigenes), prophenoloxidase system (2 unigenes), signalling pathway; (Toll pathway-5 unigenes, Wnt signalling pathway-3 unigenes, MAPK signalling pathway-7 unigenes, JAK/STAT pathway-2 unigenes, Ubiquitin Proteasome Pathway-10 unigenes) and other immune genes (17 unigenes). Overall a total of 74 immune related unigenes were found to be significantly up-regulated whereas 37 were significantly down-regulated. The role of these groups is elaborated in detail in the next (Discussion) section.

Table 4.3. Candidate genes involved in *M. rosenbergii* immune response against WSSV.

Category or gene id	Homologues function	Species	FC*
Antiviral			
Unigene19164_All	Zinc finger CCCH-type antiviral protein 1	*Anolis carolinensis*	1.20
Unigene10846_All	Antiviral protein	*Litopenaeus vannamei*	-1.19
Antimicrobial			
Unigene26338_All	Lysozyme	*Portunus trituberculatus*	2.24
Unigene13073_All	NF-kappa B inhibitor alpha	*Macrobrachium rosenbergii*	1.12
Unigene4120_All	Anti-lipopolysaccharide factor	*Macrobrachium rosenbergii*	1.66
Unigene25224_All	Crustin type I	*Macrobrachium rosenbergii*	-14.90
Blood Clotting system			
Unigene13048_All	Clottable protein	*Marsupenaeus japonicus*	3.67
Unigene6536_All	Transglutaminase	*Fenneropenaeus chinensis*	3.05
Proteases			
CL2946.Contig3_All	Clip domain serine proteinase 1	*Penaeus monodon*	11.30
CL5686.Contig1_All	Cathepsin L	*Penaeus monodon*	-3.59
CL1587.Contig2_All	Cathepsin D	*Penaeus monodon*	-2.83
CL1474.Contig1_All	cathepsin C	*Fenneropenaeus chinensis*	1.08
CL6038.Contig1_All	26S protease regulatory	*Nasonia vitripennis*	-1.44
Unigene7896_All	ATP-dependent Clp protease proteolytic	*Caenorhabditis briggsae*	-2.46
Unigene12869_All	Alpha-2-macroglobulin	*Macrobrachium rosenbergii*	2.90
Unigene13293_All	Astacin	*Strongylocentrotus purpuratus*	-1.03
Unigene29347_All	Serine carboxypeptidase 1	*Saccoglossus kowalevskii*	-4.21
Unigene13633_All	Serine protease	*Fenneropenaeus chinensis*	2.37
Protease inhibitor			
CL1127.Contig2_All	Kazal-type serine proteinase inhibitor 4	*Procambarus clarkii*	-4.57
Unigene22656_All	Pacifastin-related serine protease inhibitor	*Portunus trituberculatus*	6.79

Category or gene id	Homologues function	Species	FC*
CL816.Contig12_All	Serpin B	*Amblyomma americanum*	-1.68
Signal transduction			
Toll pathway			
Unigene25507_All	Toll-like receptor 3	*Daphnia pulex*	3.86
CL2421.Contig1_All	Toll-like receptor 8	*Ixodes scapularis*	2.11
Unigene11753_All	Toll protein	*Litopenaeus vannamei*	11.00
CL2438.Contig1_All	Caspase	*Eriocheir sinensis*	1.54
Unigene25799_All	Myd88 protein	*Artemia sinica*	1.35
CL1915.Contig5_All	Toll interacting protein	*Marsupenaeus japonicus*	-1.21
CL5402.Contig1_All	Toll receptor 2	*Marsupenaeus japonicus*	-1.37
Wnt signaling pathway			
CL2430.Contig2_All	beta-catenin	*Parhyale hawaiensis*	13.0
Unigene23651_All	Ring box protein	*Bombyx mori*	-2.21
Unigene16346_All	Supernumerary limbs	*Tribolium castaneum*	1.27
Unigene25990_All	Casein kinase II subunit beta	*Acromyrmex echinatior*	-1.58
Unigene15960_All	Prickle	*Saccoglossuskowalevskii*	1.65
Unigene15970_All	Low-density lipoprotein receptor-related protein 6	*Daphnia pulex*	1.35
Unigene25938_All	Secreted frizzled-related protein 5	*Daphnia pulex*	2.37
MAPK signaling pathway			
CL1473.Contig2_All	Heat shock protein 70 cognate	*Fenneropenaeus chinensis*	4.11
CL4432.Contig2_All	Max protein	*Daphnia pulex*	-5.67
Unigene29777_All	Serine/threonine-protein kinase PAK 1	*Harpegnathos saltator*	3.43
Unigene9636_All	Serine/threonine-protein kinase TAO1	*Acromyrmex echinatior*	1.44
CL3877.Contig1_All	MAPK kinase 1 interacting protein 1	*Penaeus monodon*	-1.62
CL5614.Contig1_All	JNK-interacting protein 3	*Apis mellifera*	4.99
Unigene26817_All	Mitogen-activated protein kinase kinasekinase 1	*Oryctolaguscuniculus*	2.07
Unigene102_All	Map kinase-interacting serine/threonine	*Scylla paramamosain*	1.46

Category or gene id	Homologues function	Species	FC*
CL726.Contig3_All	Mitogen-activated protein kinase kinasekinasekinase 4	*Bombus terrestris*	1.24
CL3174.Contig1_All	Mitogen-activated protein kinase kinasekinase 7	*Danaus plexippus*	1.04
Unigene29569_All	Mitogen-activated protein kinase kinasekinase 13	*Bombus terrestris*	2.89
Unigene14757_All	Heat shock protein 70	*Portunus trituberculatus*	2.43
Unigene3873_All	Protein phosphatase 5	*Bombusterrestris*	11.61
Unigene11596_All	Raf homolog serine/threonine-protein kinase phl-like	*Nasoniavitripennis*	2.01
JAK/STAT pathway			
Unigene25562_All	STAT long form	*Penaeus monodon*	2.78
CL1133.Contig5_All	Domeless	*Tribolium castaneum*	1.38
Unigene25762_All	Signal transducing adaptor molecule	*Megachilerotundata*	1.36
Unigene16474_All	Tumor susceptibility gene 101 protein	*Harpegnathos saltator*	1.64
Ubiquitin Proteasome Pathway			
Unigene548_All	Ubiquitin	*Procambarus clarkii*	-1.61
Unigene22971_All	Cullin-2	*Harpegnathos saltator*	1.29
Unigene6699_All	Ubiquitin-protein ligase E3C	*Camponotus floridanus*	1.03
CL5401.Contig1_All	E3 ubiquitin-protein ligase HUWE1	*Acromyrmex echinatior*	10.58
Unigene6165_All	Rho-related BTB domain-containing protein 2	*Acromyrmex echinatior*	4.44
Unigene29739_All	Baculoviral IAP repeat-containing protein 6	*Crassostrea gigas*	2.80
Unigene10240_All	Cullin-associated NEDD8-dissociated protein 1	*Crassostrea gigas*	1.66
CL1584.Contig1_All	Beta 1,4-endoglucanase	*Cherax quadricarinatus*	1.98
Unigene3861_All	U88	*Brugia malayi*	1.39
Unigene20413_All	RING finger protein	*Eriocheir sinensis*	-1.75
IMD Pathway			
CL5784.Contig1_All	Relish	*Litopenaeus vannamei*	1.79
Unigene16147_All	IKKBeta	*Litopenaeus vannamei*	3.70
Transcriptional control			
CL965.Contig3_All	RNA-binding region-containing protein	*Danaus plexippus*	3.03

Category or gene id	Homologues function	Species	FC*
Unigene7122_All	DNA-directed RNA polymerase III subunit RPC2-like	*Cavia porcellus*	-2.02
CL5643.Contig1_All	Microphthalmia-associated transcription factor-like	*Megachile rotundata*	3.43
CL4672.Contig1_All	Replication protein A	*Callorhinchus milii*	1.65
Cell death			
Unigene28705_All	Ribosomal protein S3	*Plecoglossus altivelis*	-1.03
CL1843.Contig2_All	ALG-2 interacting protein x Penaeus monodon	*Penaeus monodon*	2.88
CL6385.Contig2_All	Caspase	*Harpegnathos saltator*	2.41
Unigene2555_All	Beclin 1	*Megachile rotundata*	11.10
Unigene14045_All	Program cell death 5-like	*Penaeus monodon*	1.36
Unigene325_All	Prohibitin	*Penaeus monodon*	-2.17
Unigene10321_All	Oncoprotein nm23	*Litopenaeus vannamei*	1.18
Unigene19853_All	Deoxyribonuclease I	*Marsupenaeus japonicus*	-2.63
CL6019.Contig2_All	Inhibitor of apoptosis protein	*Litopenaeus vannamei*	2.26
Unigene16154_All	Zinc finger protein ush-like	*Nasonia vitripennis*	5.08
Unigene9836_All	Autophagy protein 5	*Callinectes sapidus*	-1.16
Cell adhesion			
Unigene13411_All	ADP-ribosylation factor 1	*Marsupenaeus japonicus*	-1.52
Unigene16833_All	Adhesion regulating molecule 1	*Scylla paramamosain*	-1.63
CL2046.Contig1_All	Integrin	*Litopenaeus vannamei*	1.85
Unigene23158_All	Tetraspanins-like protein-8	*Fenneropenaeus chinensis*	-1.72
Oxidative stress			
Unigene7354_All	Thioredoxin 1	*Bombus impatiens*	-2.06
Unigene20577_All	Peroxiredoxin	*Eriocheir sinensis*	1.35
CL353.Contig2_All	Glutathione S transferase D1	*Procambarus clarkii*	-2.94
CL1891.Contig1_All	O-methyltransferase	*Litopenaeus vannamei*	-1.71
Unigene26204_All	Glutathione peroxidase	*Metapenaeus ensis*	1.85
CL5831.Contig2_All	Catalase	*Macrobrachium rosenbergii*	3.56

Category or gene id	Homologues function	Species	FC*
Unigene19405_All	Manganese superoxide dismutase	*Macrobrachium rosenbergii*	1.61
Heat Shock proteins			
Unigene20010_All	Chaperonin 10	*Scylla paramamosain*	-2.87
CL4288.Contig1_All	Heat shock protein 90	*Macrobrachium nipponense*	1.43
Unigene13644_All	Chaperonin	*Macrobrachium rosenbergii*	-2.01
Unigene14757_All	Heat shock protein 70	*Portunus trituberculatus*	2.43
Unigene16617_All	Heat shock protein 60	*Litopenaeus vannamei*	1.20
Unigene3736_All	Heat shock protein 21	*Macrobrachium rosenbergii*	4.61
RNA silencing			
CL1315.Contig3_All	Dicer 2	*Litopenaeus vannamei*	1.48
Unigene26246_All	Argonaute 1 isoform B	*Marsupenaeus japonicus*	4.03
CL3019.Contig2_All	Argonaute 2	*Marsupenaeus japonicus*	1.56
PRPs			
Unigene1391_All	Hemolectin	*Papilio xuthus*	5.85
CL1124.Contig1_All	Glucan pattern-recognition lipoprotein	*Fenneropenaeus chinensis*	1.53
Unigene17200_All	Lectin B isoform 1	*Marsupenaeus japonicus*	-1.65
CL4516.Contig1_All	Lectin 2	*Macrobrachium rosenbergii*	-1.16
CL51.Contig1_All	Lectin D	*Marsupenaeus japonicus*	-1.82
CL5230.Contig3_All	C-type lectin 5	*Fenneropenaeus chinensis*	2.00
Unigene10978_All	Lectin 1	*Macrobrachium rosenbergii*	2.83
Unigene23671_All	Lipopolysaccharide and beta-1,3-glucan binding protein	*Macrobrachium rosenbergii*	1.20
CL3039.Contig4_All	Lectin 4	*Macrobrachium rosenbergii*	2.41
ProPO system			
Unigene12734_All	Prophenoloxidase	*Macrobrachium rosenbergii*	4.54
Unigene7353_All	Prophenoloxidase activating factor	*Fenneropenaeus chinensis*	1.83
Other immune genes			
Unigene26739_All	Metallothionein I	*Macrobrachium rosenbergii*	1.12

Category or gene id	Homologues function	Species	FC*
CL4071.Contig2_All	Peritrophin	*Macrobrachium nipponense*	2.56
Unigene7335_All	Profilin	*Penaeus monodon*	1.15
Unigene1349_All	Copper-specific metallothionein CuMT-II	*Callinectes sapidus*	-3.19
Unigene10835_All	Selenoprotein W	*Cricetulus griseus*	-2.76
CL4159.Contig2_All	Selenoprotein M	*Penaeus monodon*	-1.16
CL6324.Contig2_All	Arginine kinase 1	*Macrobrachium rosenbergii*	5.33
CL5195.Contig1_All	Hemocyanin	*Macrobrachium nipponense*	3.04
Unigene12758_All	Tetraspanins-like protein-8	*Fenneropenaeus chinensis*	-1.53
CL1939.Contig4_All	Ferritin	*Litopenaeus vannamei*	1.84
CL6297.Contig1_All	E cadherin	*Tenebrio molitor*	1.45
CL2214.Contig1_All	Chitinase	*Pandalopsis japonica*	1.97
CL6026.Contig5_All	Actin	Marsupenaeus japonicus	3.69
Unigene13021_All	Calponin	Chironomus riparius	1.41
CL2339.Contig2_All	Adenosine deaminase	Caligus rogercresseyi	-1.23
Unigene19713_All	Pacifastin heavy chain	*Macrobrachium rosenbergii*	*1.27*
CL4920.Contig1_All	Calmodulin	*Procambarus clarkii*	1.26

*Fold changes (Log2 ratio) in gene expression. PRPs- pattern recognition proteins, ProPO- prophenoloxidase.

4.5 Discussion

The white spot syndrome disease resulting from WSSV infection is the most serious viral disease in cultured commercial shrimps worldwide (Escobedo-Bonilla, *et al*., 2008; Flegel, 1997). The disease has caused high mortality and consequent economic losses to the shrimp culture industry in Thailand (Wongteerasupaya *et al*., 1995), Taiwan (Wang *et al.,*1999), India (Rajan *et al.*, 2000; Sunarto, 2001) and Malaysia (Wang *et al.*, 1999). Knowledge about the interaction between *M. rosenbergii* and the virus is limited, despite the numerous published studies on the characterisation and detection of WSSV available (Huang *et al.*, 2001; Otta *et al.*, 1999; Pradeep *et al.*, 2008; Sithigorngul *et al.*, 2006; van Hulten *et al.*, 2001). A better understanding of viral-host interaction is urgently needed to help develop a pro-active approach to fighting and suppressing this disease.

The application of NGS platforms such as Illumina HiSeq^{TM} 2000 has allowed the fast and cost-effective generation of massive amounts of sequence data, which has revolutionised the field of transcriptomics. This platform, together with the *de novo* assembler, has been successfully utilised to discover novel genes as well as to reveal the expression levels of genes in non-model organisms which lack a complete genome database (Kawahara-Miki *et al.*, 2011; Li *et al.*, 2012; Mohd-Shamsudin *et al.*, 2013; Shen *et al.*, 2011; Jian Xu *et al.*, 2013). The introduction of NGS technology has led to studies on host-viral interaction in shrimps in order to identify candidate immune-related genes (Chen *et al.,* 2013; Li *et al.,* 2013; Xue, *et al.,* 2013; Zeng *et al.,* 2013). We had successfully used this sequencing platform to identify molecular mechanisms surrounding the innate immune system against *V. parahaemolyticus* in *M. rosenbergii* (Rao *et al*., 2015). In the present study, we utilised a similar approach to elucidate the transcriptional responses of *M. rosenbergii* towards WSSV infection.

The total RNA for sequencing was sampled from pooled hepatopancreases of both groups, taken at 12th hour time point. The overall analysis yielded 14,416 unigenes differentially expressed, with 8,443 unigenes significantly up-regulated and 5,973 unigenes significantly down-regulated. The differences in terms of number of aberrantly expressed genes in this study and our previous study against *V. parahaemolyticus* cleared showed that different approach was taken by *M. rosenbergii* against viral and bacterial pathogens. In addition, the number of differentially expressed unigenes obtained was higher than from *L. vannamei* (315 unigenes in hemocytes and 1,496 unigenes in hepatopancreas) (Chen *et al.,* 2013; Xue *et al.,* 2013). This could be due to a number of reasons: the type of tissue samples used, different time-points of tissue sampling, the use of different NGS platforms, different *de novo* assembly approaches, and lastly different species of shrimp. However, the number of up-regulated genes obtained was much larger than the down-regulated genes, as was the case with studies done on *L. vannamei* challenged by WSSV.

Crustaceans like *M. rosenbergii* rely on their innate immune system to defend themselves against various pathogens and environmental stress. This defence system is initiated by detecting the invading pathogen through pattern recognition proteins (PRPs) (Akira *et al.*, 2006). The hepatopancreas of *M. rosenbergii*, similar to *L. vannamei* and *L. setiferus,* seems to be the primary production site for two PRPs molecules, namely the beta-1,3- glucan binding protein (LGBP) and C-type lectins (Gross *et al.*, 2001). The LGBP protein functions to recognize and eliminate pathogens, whereas C-type lectins facilitate recognition and pathogen phagocytosis through opsonisation (Cerenius & Söderhäll, 2004; Vargas-Albores & Yepiz-Plascencia, 2000). In this study, PRPs expression was affected by the WSSV challenge, similar to what has been reported in previous studies (Ma *et al.*, 2007; Roux *et al.*, 2002; Song *et al.*, 2010). The initial immune

response towards pathogen invasion also activates the melanization process. Melanization is achieved by the prophenoloxidase-activating (proPO-AS) system, which is an enzymatic cascade involving several enzymes including the key enzyme phenoloxidase (PO) (Ratcliffe *et al.*, 1984; Söderhäll & Cerenius, 1998). The up-regulation of proPO-AS components in our results showed that this system was activated by WSSV infection.

Besides the melanization process, anti-microbial peptides (AMPs) also function as the first line of defence to fight against invading pathogens (Hancock & Diamond, 2000). Several families of shrimp AMPs, such as penaeidins, lysozymes, crustins, anti-lipopolysaccharide factors (ALFs) and stylicins, have been identified and characterized previously (Bartlett *et al.*, 2002; Destoumieux *et al.*, 2000; Rolland *et al.*, 2010; Tassanakajon *et al.*, 2011). In our data, we found only three members of the AMPs, lysozymes, ALFs and crustins. The lysozymes and ALFs were significantly up-regulated, whereas the crustins were significantly down-regulated. An increased expression of lysozyme and ALF was also reported in *F. chinensis*, *L. stylirostris* and *L. vannamei,* suggesting that these two genes are involved in the innate response of shrimps to WSSV (De la Vega *et al.*, 2008; Mai & Wang, 2010; Wang *et al.,* 2006). A reduced expression of crustin type 1 was also noted in *P. monodon* challenged by WSSV (Philip *et al.*, 2011; Singh*, et al.*, 2011). Since crustin type 1 was reported to be down-regulated in both P. *monodon* and *M. rosenbergii* during WSSV infection, therefore this isoform of crustin may not be required for antiviral defence of shrimp against WSSV.

Blood coagulation is a conserved defence mechanism among invertebrates, which possess open circulatory systems and hence rely more extensively on this system than vertebrates (Theopold *et al.*, 2004). Transglutaminase (TGase) and clotting protein (CP) are molecules responsible for blood clotting in shrimps (Maningas *et al.*, 2008). The

clotting process involves the polymerization of the plasma clotting protein catalyzed by a Ca^{+2} dependent transglutaminase (TGase) released from the haemocytes in response to tissue damage or the invasion of microbial pathogens. The released TGase produces the Ɛ - (γ-glutamyl) - lysine crosslinks between glutamine and lysine of the clottable protein (CP) (Kopacek *et al.*, 1993). Higher expression levels for both components of blood coagulation were observed in this study. Both the TGase gene and clotting protein were observed to be differentially expressed in response to viral and bacterial pathogens, which suggests that these genes have an important role in shrimp immunity (Wang *et al.*, 2010; Liu *et al.*, 2007; Song *et al.*, 2003). Several families of heat shock proteins (Hsps) – Hsp90s (83–99 kDa), Hsp70s (68–80 kDa), Hsp60s, and the small Hsps (25–28 kDa), function as chaperones, correcting the folding of proteins and repairing and forming multiprotein assemblies (Pockley, 2003; Stewart & Young, 2004). The differential expression of heat shock proteins in our data is similar to that reported by other researchers in WSSV-infected shrimps (Wang, et al., 2006; Yan *et al.*, 2010). All these results point to a potential role for heat shock proteins as anti-virus responses to WSSV in shrimps.

In our transcriptome data, three unigenes identified as important components of the RNA interference (RNAi) pathway, Argonaute 1, Argonaute 2 and Dicer 2 were highly expressed. Previous research reported a significant inhibition of virus replication via the dsRNA (Ongvarrasopone *et al.*, 2008; Robalino *et al.*, 2005; Sarathi *et al.*, 2008; Yodmuang *et al.*, 2006) and siRNA (Westenberg *et al.*, 2005; Wu *et al.*, 2007; Xu *et al.*, 2007) of specific genes, suggesting that an RNAi mechanism against virus infection possibly exists in shrimps. Subsequently, the genes encoding Argonaute (Unajak *et al.*, 2006), Dicer (Su *et al*., 2008) and other members of the RNAi pathway were revealed in shrimps. The increased expression of these genes in our data might imply that they have

an antiviral role against WSSV. The zinc finger CCCH-type antiviral protein 1-like was found to be significantly up-regulated in our data. This protein has been reported to inhibit the replication of several viruses which could possibly act similarly to WSSV infection (Bick *et al.*, 2003; Gao *et al.*, 2002; Müller *et al.*, 2007).

Apoptosis also plays a major role in regulating the innate cellular response, by minimising virus replication and suppressing the spread of progeny virus in the host (Roulston *et al.*, 1999). Caspase, an effector molecule that mediates the apoptotic process (Wang *et al.*, 2008), was found to be up-regulated in our data. We note that the up-regulation of this gene has been similarly observed when *M. rosenbergii* are challenged by the infectious hypodermal and haematopoietic necrosis virus (IHHNV) (Arockiaraj *et al.*, 2012a) and in other shrimps in response to WSSV as well (Phongdara *et al.*, 2006; Wongprasert *et al.*, 2007). The inhibitors of apoptosis proteins (IAPs) regulate the apoptosis mechanism by blocking the activity of caspase (Deveraux & Reed, 1999). Furthermore, IAPs are also reported to participate in signal transduction regulation during immune responses in insects and mammals (Deveraux & Reed, 1999; Earnshaw *et al.*, 1999; Silke & Meier, 2013). IAPs were highly expressed in our transcriptome data, in *M. rosenbergii* challenged with IHHNV (Arockiaraj *et al.,* 2011a) and in *L. vannamei* challenged by WSSV (Wang *et al*., 2013), which means they could perhaps play a protective role against viral infection. We observed the expression of oncoprotein nm23, beclin, prohibitin and deoxyribonuclease I in our data similar to previous reports, which may reaffirm the role of these genes in controlling the extent of apoptosis in response to WSSV (Chen *et al.,* 2013). The role of other genes in this group remains vague and needs further study. The differentially expressed genes in cell adhesion, transcriptional control, proteases and proteases inhibitors groups are generally assumed to modulate phagocytic

events, the recruitment of immune cells to sites of infection, cellular remodelling, and extracellular immune cascades such as the melanization response (Robalino *et al*., 2007).

Antioxidant enzymes play a crucial role in shrimp defences against pathogens by removing potentially harmful excess reactive oxygen species generated during the immune response (De la Fuente & Victor, 2000). In this study we reported seven unigenes in this group to be differentially expressed after WSSV infection. Aberrant expression levels of these genes have been noted in other shrimps as well, but in a differential manner according to the types of pathogen (Castex *et al.*, 2010; Mathew *et al.*, 2007; Mohankumar & Ramasamy, 2006; Zhang *et al.*, 2007). We also observed that the five signal transduction pathways, Toll, Janus kinase-signal transducer and activator of transcription (JAK/STAT), IMD, mitogen-activated protein kinase (MAPK) and Wnt signalling pathway were greatly affected by WSSV infection. These pathways are known to be activated to play important roles in viral recognition and replication (Andrade *et al*., 2004; Chen *et al*., 2008; Gan *et al*., 2001; Li & Xiang, 2013; Wang *et al*., 2012). However, the underlying molecular mechanisms of these pathways in response to WSSV infection remain unclear. Therefore further studies are necessary to uncover the roles of those pathways in *M. rosenbergii* innate immune system.

In addition to the five above mentioned pathways, we also observed the activation of ubiquitin proteasome pathway (UPP) in *M. rosenbergii* with E3 ubiquitin-protein ligase HUWE1 (+10.58 fold) showing the highest fold changes. The UPP pathway regulates degradation of short-lived proteins in various cellular processes such as cell cycle control, transcriptional regulation, signal transduction, antigen presentation and induction of the inflammatory response, degradation from the ER, membrane trafficking, receptor endocytosis and down-regulation, apoptosis, and development in vertebrate

(Loureiro & Ploegh, 2006). Furthermore, this pathway is also involved in viral infection and facilitates activities required for various aspects of virus life cycle, from entry through replication and enhanced cell survival to viral release (Chen & Gerlier, 2006). Lately, various reports have emerged showing the ability of WSSV to hijack the host UPP pathway to modulate cellular intrinsic antiviral activities and innate immunity (Jeena *et al.*, 2012; Keezhedath *et al.*, 2013; Vidya *et al.*, 2013). The higher expression of E3 ubiquitin-protein ligase HUWE1 could possibly be due to ability of virus to redirect host ubiquitin E3 ligases to target new substrate proteins for its own gain (Banks *et al.*, 2003). However further study is required to fully understand the interaction between WSSV and this protein.

Several of the differentially expressed genes in other immune gene group have been reported to be involved in immune responses in shrimps against WSSV. Hemocyanins, the oxygen-transporting proteins in arthropods and molluscs, play a defence-related role mediated through phenoloxidase activity (Markl & Decker, 1992; Zhang *et al.*, 2004). Several previous studies have found hemocyanins in shrimp and crayfish to be highly expressed during WSSV infection (Lei *et al.*, 2008; Pan *et al.*, 2005; Wang *et al.*, 2013). Ferritin, a major iron storage protein involved in iron metabolism, was up-regulated in the hepatopancreases of the shrimps challenged by the white spot syndrome virus (WSSV) (Pan *et al.,* 2005; Zhang *et al.*, 2006). Profilins are a group of ubiquitous proteins commonly found in animals, plants and viruses. They activate the sequestering of actin monomers and prevent actin polymerization. Besides their actin binding activity, profilins are also involved in cellular processes such as membrane trafficking, small-GTPase signalling and nuclear activities, in addition to neurological diseases and tumor formation (Rawe *et al.*, 2006; Witke 2004). The up-regulation of this gene has also been observed in *L. vannamei* and *F. chinensis* challenged with WSSV

(Clavero-Salas *et al*., 2007; Kong *et al*., 2009). Calmodulin, an important intracellular Ca^{2+} sensor and a Ca-binding protein involved in calcium-dependent signal transduction pathways, has been induced in *L. vannamei* challenged with WSSV (Ji *et al.*, 2011). Arginine kinase (AK) is a phosphotransferase that plays a crucial role in the coupling of energy production during energy metabolism in invertebrates (Arockiaraj *et al.,* 2011b; Yao *et al.*, 2009). Recently, shrimp AK has been reported to play an important role in WSSV infections by interacting with the VP14 protein of WSSV in infected *L. vannamei* (Ma *et al.*, 2014). All of these genes were up-regulated in our transcriptome data, indicating that they may play important roles in *M. rosenbergii* defence against WSSV.

4.6 Conclusion

In this study, we investigated the transcriptome profile of *M. rosenbergii* hepatopancreas when infected with WSSV using the Illumina HiSeq^{TM}2000 technology. We observed that WSSV infection has effected the transcriptome profile of *M. rosenbergii*'s hepatopancreas, with 8,443 host genes were significantly up-regulated whereas 5,973 genes were significantly down-regulated. The differentially expressed immune related genes were grouped into 15 animal immune functions. However to further enhance our understanding regarding molecular interactions between *M.rosenbergii* and WSSV, further studies on the functionality of these genes will provide valuable information to find effective strategies to prevent viral disease. Besides, with whole genome sequence of this species is still unavailable, large number of transcripts obtained from this study could provide a strong basis for future genomic research on *M. rosenbergii*.

CHAPTER 5

COMPARATIVE TRANSCRIPTOMICS ANALYSIS OF *MACROBRACHIUM ROSENBERGII* HEPATOPANCREAS AGAINST *VIBRIO PARAHAEMOLYTICUS* AND WHITE SPOT SYNDROME VIRUS

5.1 Introduction

The global production of *M. rosenbergii* is seriously affected by bacterial and viral infections. Two major culprits have been identified as *Vibrio sp*. and the white spot syndrome virus (WSSV). Detailed information on the immune system of *M. rosenbergii* is still lacking or vague, especially regarding the role of different signalling pathways implicated in immune responses.

In previous work, we performed a transcriptomics analysis, using the Illumina HiSeq^{TM} 2000 sequencing platform, as well as a bioinformatics analysis on the responses of the *M. rosenbergii* hepatopancreas to *V. parahaemolyticus* and WSSV infection. We were able to generate a significant number of immune genes in response to these two pathogens from the *M. rosenbergii* hepatopancreas. In the present study, we compared the expression of 156 immune-related unigenes from *M. rosenbergii* when challenged with *V. parahaemolyticus* and WSSV. The expression profiles of these unigenes were different against each of these two pathogens. We also observed that 79 unigenes were up-regulated and 7 unigenes down-regulated against both pathogens. Several of these unigenes were key components of five signalling pathways: the Toll pathway immune deficiency (IMD), the Wnt pathway, the Janus kinase (JAK)-signal transducers and

activators of transcription (STAT), the Mitogen-activated protein kinase (MAPK) pathway, and finally the Ubiquitin proteasome pathway.

5.2 Literature Review

Macrobrachium rosenbergii is one the most popular cultured freshwater prawns in Malaysia, and contributes significantly to the local aquaculture industry. As in other parts of the world, the production of *M. rosenbergii* in Malaysia has been seriously affected by pathogens, mostly because of bacterial and viral infections. Two main culprits, *Vibrio sp.* (Lavilla-Pitogo *et al.,* 1990; Muñoz *et al.,* 2000) and the white spot syndrome virus (WSSV) (Hameed *et al.*, 2003; Yang *et al.,* 2001), have been determined as the main pathogens causing serious mortality and severe economic losses in shrimp production.

It is well-known that shrimps have no adaptive immune systems. Instead, their defence systems are believed to rely entirely on innate mechanisms to resist the invasion of pathogens (Aguirre-Guzman *et al.,* 2009). However, our understanding of the immune system of *M. rosenbergii* is still fragmentary, and various signalling pathways implicated in immune responses also remain incomplete. Proper health management of these crustaceans requires an improved understanding of their molecular responses to pathogens.

In crustaceans, the hepatopancreas plays an important role as an immune organ as well functioning as a digestive gland. It is a primary site for synthesizing and excreting immune molecules, such as antibacterial peptides (AMP), beta-1,3-glucan binding proteins (LGBP), lectin or lectin-related proteins, and others (X. Li *et al*., 2013). This was further confirmed by the gene discovery in *L. vannamei* and *L. setiferus*, using the Expressed sequence tag (EST) analysis, which showed that the hepatopancreas plays a

crucial role in innate immunity and that its cDNA library appears to be more diverse than its hemocytes library (Gross *et al*., 2001). Hence, the large-scale identification of immune genes from the hepatopancreas of crustaceans will be of great value, indeed a necessity, in order to further our knowledge of the immune mechanisms of these species.

It was this belief that led us to perform a transcriptomics analysis, using the Illumina HiSeq^{TM} 2000 sequencing platform, as well as a bioinformatics analysis on the responses of the *M. rosenbergii* hepatopancreas to *V. parahaemolyticus* and WSSV infection. We were able to identify a significant number of immune-related unigenes in the *M. rosenbergii* hepatopancreas in response to these two pathogens.

The aim of the present study is to compare differentially expressed immune-related unigenes caused by WSSV or *Vibrio* infection, with a particular focus on the role of immune molecules in different signalling pathways such as the Toll, immune deficiency (IMD) and Wnt pathways, the Janus kinase (JAK)-signal transducers and activators of transcription (STAT), and the Ubiquitin proteasome (UPP) and Mitogen-activated protein kinase (MAPK) pathways.

5.3 Materials and Methods

The transcript expression levels of the immune related unigenes were measured by using the FPKM method (fragments per kb per million fragments) were retrieved from the transcriptome databases. An FDR (False Discovery Rate) of <0.001 was used as the threshold of the p-value in the multiple test to judge the significance of the difference in gene expression (Reiner, Yekutieli, & Benjamini, 2003). The genes were considered to be differentially expressed within a given library when the p-value was less than 0.001, and a greater than two-fold change in expression across libraries was observed. The

FPKM values for the immune related unigenes from Table 3.3 and Table 4.3 (control and WSSV- and *V. parahaemolyticus*-infected samples) were converted into a heatmap using the R package (Gentleman *et al.*, 2004).

5.4 Results and Discussion

A total of 156 immune-related unigenes were obtained from the transcriptome data in response to WSSV and *V. parahaemolyticus* infections. The expression of all these unigenes is shown in a heatmap (Figure 5.1). In response to *V. parahaemolyticus* infection alone, 40 unigenes were up-regulated and 5 unigenes were found to be down-regulated. This compares with 19 unigenes found to be up-regulated and 33 unigenes down-regulated in response to WSSV infection. Interestingly, we also found 79 unigenes to be up-regulated and 7 unigenes down-regulated in response to both pathogens. Unigenes which were differentially expressed in response to *V. parahaemolyticus* and WSSV infections, or expressed only in response to one of these pathogens, are shown in a Venn diagram (Figure 5.2). These 156 unigenes were clustered under various processes of the animal immune system. Several were found to play an important role in six important signalling pathways: the Toll pathway (8 unigenes), Immune deficiency (IMD) (2 unigenes), the Wnt pathway (9 unigenes), Janus kinase (JAK)-signal transducers and activators of transcription (STAT) (4 unigenes), the Mitogen-activated protein kinase (MAPK) pathway (16 unigenes), and the Ubiquitin proteasome pathway (10 unigenes) (Table 5.1).

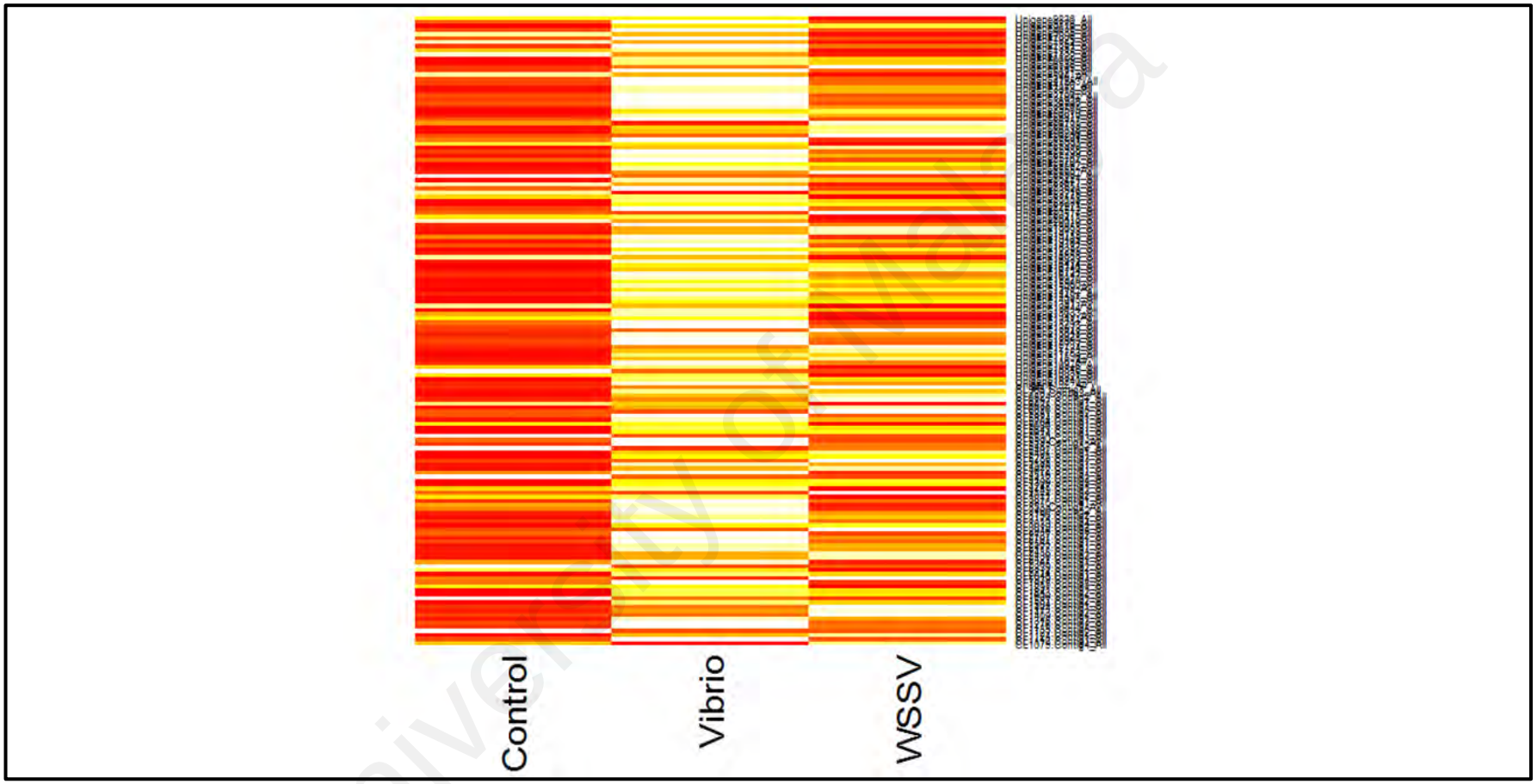


Figure 5.1: Heatmap representing 156 immune-related unigene expression profiles of *M. rosenbergii* in response to WSSV and *V. parahaemolyticus*.

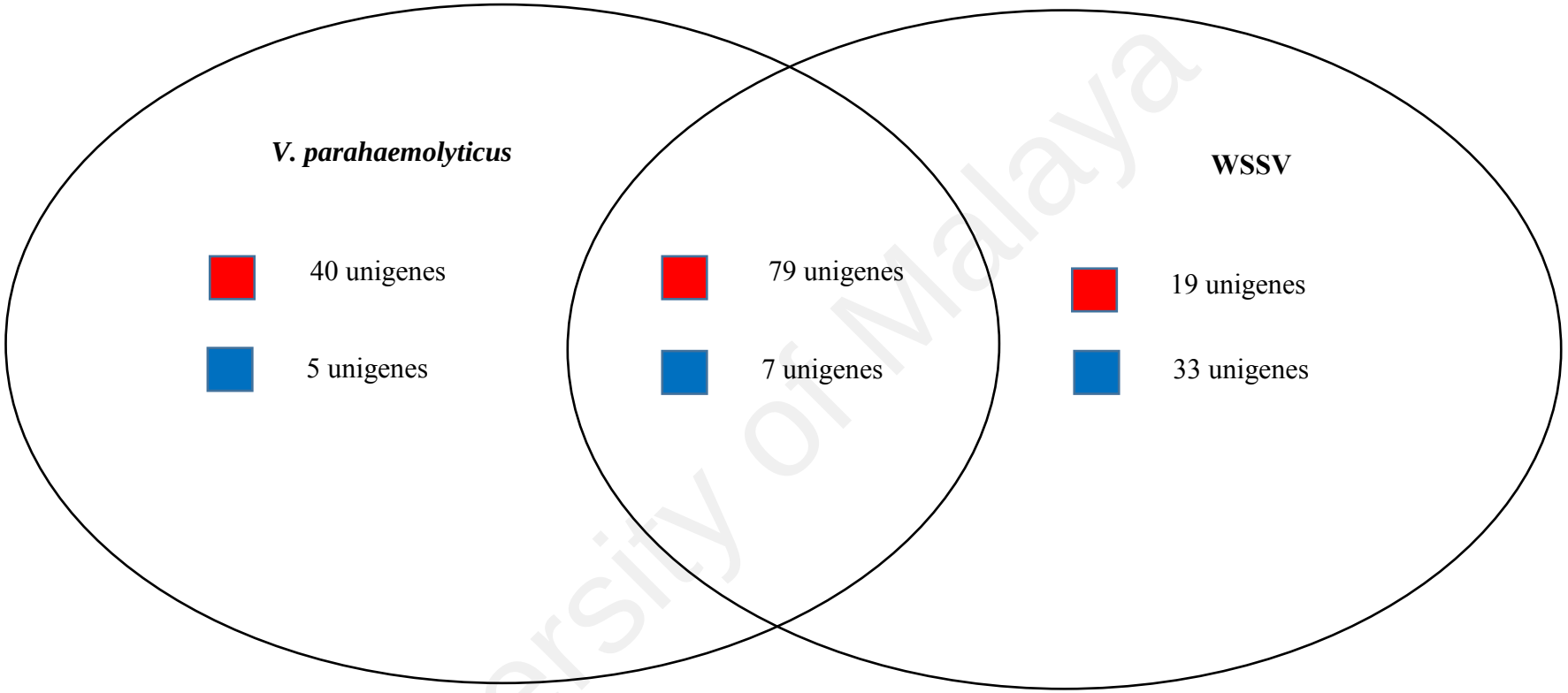


Figure 5.2: Venn diagram showing genes up- or down-regulated by *V.parahaemolyticus* and WSSV infection.

Up- regulated

Down- regulated

Table 5.1: Candidate signalling pathway genes involved in *M. rosenbergii* immune response against *V. parahaemolyticus* and WSSV.

Category or gene id	Homologue function	Species	WSSV FC*	*V. parahaemolyticus* FC*
Signal transduction				
Toll pathway				
Unigene25507_All	Toll-like receptor 3	*Daphnia pulex*	3.86	2.59
CL2421.Contig1_All	Toll-like receptor 8	*Ixodes scapularis*	2.11	1.48
Unigene11753_All	Toll protein	*Litopenaeus vannamei*	11.00	-
CL2438.Contig1_All	Caspase	*Eriocheir sinensis*	1.54	3.08
Unigene25799_All	Myd88 protein	*Artemia sinica*	1.35	3.25
CL1915.Contig5_All	Toll interacting protein	*Marsupenaeus japonicus*	-1.21	1.73
CL5402.Contig1_All	Toll receptor 2	*Marsupenaeus japonicus*	-1.37	-2.46
Unigene14481_All	Dorsal	*Fenneropenaeus chinensis*	-	3.19
Wnt signaling pathway				
CL2430.Contig2_All	beta-catenin	*Parhyale hawaiensis*	13.0	1.62
Unigene23651_All	Ring box protein	*Bombyx mori*	-2.21	-
Unigene16346_All	Supernumerary limbs	*Tribolium castaneum*	1.27	-
Unigene25990_All	Casein kinase II subunit beta	*Acromyrmex echinatior*	-1.58	-
Unigene36938_All	Lipoprotein receptor	*Callinectessapidus*	-	10.94
Unigene15960_All	Prickle	*Saccoglossuskowalevskii*	1.65	1.71
Unigene15970_All	Low-density lipoprotein receptor-related protein 6	*Daphnia pulex*	1.35	2.05
Unigene12737_All	Glypican 4	*Daphnia pulex*	-	1.14
Unigene25938_All	Secreted frizzled-related protein 5	*Daphnia pulex*	2.37	1.82
MAPK signaling pathway				

Category or gene id	Homologue function	Species	WSSV FC*	*V. parahaemolyticus* FC*
CL1473.Contig2_All	Heat shock protein 70 cognate	*Fenneropenaeus chinensis*	4.11	-
CL4432.Contig2_All	Max protein	*Daphnia pulex*	-5.67	-3.11
Unigene29777_All	Serine/threonine-protein kinase PAK 1	*Harpegnathos saltator*	3.43	-
Unigene9636_All	Serine/threonine-protein kinase TAO1	*Acromyrmex echinatior*	1.44	1.31
CL3877.Contig1_All	MAPK kinase 1 interacting protein 1	*Penaeus monodon*	-1.62	-
CL5614.Contig1_All	JNK-interacting protein 3	*Apis mellifera*	4.99	4.89
Unigene26817_All	Mitogen-activated protein kinase kinasekinase 1	*Oryctolaguscuniculus*	2.07	2.79
Unigene102_All	Map kinase-interacting serine/threonine	*Scylla paramamosain*	1.46	2.48
Unigene110_All	Mitogen-activated protein kinase 8 interacting protein 3	*Bombusterrestris*	-	2.94
CL726.Contig3_All	Mitogen-activated protein kinase kinasekinasekinase 4	*Bombus terrestris*	1.24	1.65
CL3174.Contig1_All	Mitogen-activated protein kinase kinasekinase 7	*Danaus plexippus*	1.04	1.41
Unigene29569_All	Mitogen-activated protein kinase kinasekinase 13	*Bombus terrestris*	2.89	4.03
Unigene14757_All	Heat shock protein 70	*Portunus trituberculatus*	2.43	2.26
Unigene3873_All	Protein phosphatase 5	*Bombusterrestris*	11.61	1.21
Unigene102_All	Mapkinase-interacting serine/threonine	*Scylla paramamosain*	-	2.48
Unigene11596_All	Raf homolog serine/threonine-protein kinase phl-like	*Nasoniavitripennis*	2.01	2.41
JAK/STAT pathway				
Unigene25562_All	STAT long form	*Penaeus monodon*	2.78	2.84
CL1133.Contig5_All	Domeless	*Tribolium castaneum*	1.38	2.94
Unigene25762_All	Signal transducing adaptor molecule	*Megachilerotundata*	1.36	1.98
Unigene16474_All	Tumor susceptibility gene 101 protein	*Harpegnathos saltator*	1.64	2.46

Category or gene id	Homologue function	Species	WSSV FC*	*V. parahaemolyticus* FC*
Ubiquitin Proteasome Pathway				
Unigene548_All	Ubiquitin	*Procambarus clarkii*	-1.61	-
Unigene22971_All	Cullin-2	*Harpegnathos saltator*	1.29	1.25
Unigene6699_All	Ubiquitin-protein ligase E3C	*Camponotus floridanus*	1.03	1.37
CL5401.Contig1_All	E3 Ubiquitin-protein ligase HUWE1	*Acromyrmex echinatior*	10.58	-
Unigene6165_All	Rho-related BTB domain-containing protein 2	*Acromyrmex echinatior*	4.44	-
Unigene29739_All	Baculoviral IAP repeat-containing protein 6	*Crassostrea gigas*	2.80	-
Unigene10240_All	Cullin-associated NEDD8-dissociated protein 1	*Crassostrea gigas*	1.66	1.95
CL1584.Contig1_All	Beta 1,4-endoglucanase	*Cherax quadricarinatus*	1.98	2.37
Unigene3861_All	U88	*Brugia malayi*	1.39	-
Unigene20413_All	RING finger protein	*Eriocheir sinensis*	-1.75	-
IMD pathway				
CL5784.Contig1_All	Relish	*Litopenaeus vannamei*	1.79	2.53
Unigene16147_All	IKKbeta	*Litopenaeus vannamei*	3.70	5.39

*FC= Fold changes in gene expression

The Toll pathway has been reported to play a significant role during Gram-positive bacterial and fungal invasion, by regulating a large set of genes (including antimicrobial peptide genes, many small peptides with unknown functions, as well as components of the melanization and clotting cascades). In addition, this pathway has also been shown to be involved in immune responses to viruses in both Drosophila (Sabin, Hanna, & Cherry, 2010; Zambon *et al*., 2005) and *Aedes aegypti* (Ramirez & Dimopoulos, 2010; Sim & Dimopoulos, 2010).

Based on our results, this pathway is clearly activated in *M. rosenbergii* during *V. parahaemolyticus* and WSSV infection. The major components of this pathway, the Toll-like receptors (TLRs), play a crucial role in the innate immune system. In *M. japonicas*, MjToll gene expression was significantly up-regulated in lymphoid organs stimulated with peptidoglycan at 12 h (Mekata *et al*., 2008). In the *L. vannamei* gills, only LvToll1 was up-regulated during *Vibrio* challenge, whereas LvToll1, LvToll2 and LvToll3 were all up-regulated in similar tissues after WSSV challenge (Wang *et al.*, 2012). During *Vibrio* or WSSV stimulation in the *F. chinensis* lymphoid organ, the expression level of FcToll was modulated (Yang *et al.*, 2008).

Based on our transcriptome data, we found that the dorsal, a key component of the Toll pathway, was up-regulated only during *V. parahaemolyticus* infection – which may suggest that it plays an important role against bacterial infection only. However, in other studies, the FcDorsal in *F. chinensis* was shown to have increased expression levels in the hemocytes and lymphoid organs (Oka) when stimulated by either bacteria or WSSV (F. Li *et al.,* 2010). The myeloid differentiation factor 88 (MyD88) is known to act as an important adapter protein which links members of the toll-like receptors (TLR) to the downstream components which activate related signalling pathways (Wen *et al*., 2013).

MyD88 was found to be highly expressed in our data, as similarly reported in *F. chinensis* MyD88 when injected with *V. anguillarum* (Gram-negative bacteria, G−), *M. lysodeikticu* (Gram-positive bacteria, G+) and WSSV (Wen *et al.,* 2013).

Previous research has found the expression of antimicrobial peptide genes, especially in response to Gram-negative bacteria in Drosophila, shrimp and crayfish, to be regulated by the IMD pathway (De Gregorio *et al*., 2002; Lemaitre & Hoffmann, 2007). We observed two unigenes in our database belonging to this pathway, Relish and Ikkbeta, to be up-regulated in response to both pathogens. A similar expression of Relish has been reported in *F. chinensis* and *L. vannamei*, while the highest expression of this unigene has been reported in *F. chinensis* hemocytes and lymphoid organs in response to *V. anguillarum*, *Micrococcus lysodeikticus* (*M. lysodeikticus*) stimulation (F. Li *et al.,* 2009) and WSSV (F. Li *et al.,* 2010).

The Ubiquitin-mediated proteasome pathway (UPP) regulates the degradation of short-lived proteins in various cellular processes, including cell development and pathogen defence (Loureiro & Ploegh, 2006). Viruses such as WSSV have been found to produce proteins to hijack cellular components of UPP, modifying their substrate specificity and getting rid of nonessential cellular proteins, in particular inhibitors of the cell cycle (Jeena *et al*., 2012; Keezhedath *et al.,* 2013; Vidya, Gireesh-Babu, & Prasad, 2013). They are also known to retard specific types of protein degradation by inhibiting several UPP components or by utilising the system to manage the level of expression of their native proteins. In shrimps, this pathway has also been found to be triggered in hemocytes of *L. vannamei* in response to WSSV infection (Xue *et al.*, 2013). Our data shows that this pathway was clearly activated only during WSSV infection. This lends

support to previous research pointing to the ability of WSSV to trigger the host UPP pathway to modulate cellular intrinsic antiviral activities and innate immunity.

Mitogen-activated protein kinase (MAPK) signalling is another noteworthy pathway. We found that this pathway was among those activated in our transcriptome data in response to the two pathogens. This important pathway plays a key role in animal defence against pathogens, especially during virus infection; it also contributes to virus replication in animal cells (Andrade *et al.*, 2004). Studies have found it to be activated in *L. vannamei* also when challenged with WSSV and TSV viruses (Xue *et al.,* 2013; Zeng *et al.*, 2013). With regard to to its antibacterial properties, activation of this pathway has also been observed in *E. sinensis* in the course of three different microbial challenges (X. Li *et al.,* 2013).

Along with the MAPK pathway, our data showed that the Wnt pathway too was activated against WSSV and *V. parahaemolyticus*. In multicellular organisms, the Wnt pathway has been shown to be involved in development processes such as mitogenic stimulation, cell fate specification and differentiation (Huelsken & Behrens, 2002). This pathway has moreover been found to be evolutionarily conserved across a wide range of species, ranging from the freshwater polyp Hydra to worms, flies and vertebrates (Huelsken & Birchmeier, 2001). In arthropods, information on the role of the Wnt pathway in inflammation and immunity was obtained from a study on Drosophila, which found this pathway to regulate the phagocytosis of WSSV in Drosophila (Zhu & Zhang, 2013). Another study showed that this signalling pathway was also activated in *L. vannamei* during WSSV infection (Xue *et al.,* 2013). In addition, a key component of this pathway, β-Catenin, was found to be significantly up-regulated in *M. japonicas* at various

points in time after infection with *S.aureus*, *V. anguillarum* and WSSV (Xie *et al*., 2015). Our data similarly found this gene to be up-regulated against the two pathogens tested.

Finally, the JAK/STAT pathway is involved in various biological processes of both innate and adaptive immunity in mammals and other vertebrates, such as apoptosis, proliferation, differentiation, hematopoiesis, oncogenesis and immune defence (Rawlings, Rosler, & Harrison, 2004). Our transcriptome data also found this pathway to be up-regulated in response to infection by the two pathogens. This finding is consistent with other reports suggesting that the JAK/STAT pathway plays a pivotal role in antiviral and antibacterial defence in insects and crustaceans (Barillas-Mury *et al*., 1999; Souza-Neto, Sim, & Dimopoulos, 2009; Sun *et al*., 2011).

5.5 Conclusion

In this study, we compared the expression of 156 immune-related unigenes from *M. rosenbergii* when challenged with *V. parahaemolyticus* and WSSV. Separate infection by each of these two pathogens produced different numbers of up-regulated and down-regulated unigenes. At the same time, we found 79 unigenes that were up-regulated and 7 unigenes that were down-regulated in response to both pathogens. Several of these unigenes were key components of six signalling pathways: the Toll pathway immune deficiency (IMD), the Wnt pathway, Janus kinase (JAK)-signal transducers and activators of transcription (STAT), the Mitogen-activated protein kinase (MAPK) pathway, the IMD pathway and the Ubiquitin proteasome pathway.

While our study produced a range of useful additional information on the immune responses in *M. rosenbergii*, and specifically on the above pathways, our findings do not allow us to draw firm conclusions at this stage as to whether and how the above various

pathways might be triggered or activated, either specifically or generally, by viral and bacterial challenge. Further work is therefore needed to unravel the role of these pathways in regulating the host immune system.

Chapter 6

STAT GENE CHARACTERIZATION AND EXPRESSION IN *MACROBRACHIUM ROSENBERGII* AGAINST *VIBRIO PARAHAEMOLYTICUS* AND WHITE SPOT SYNDROME VIRUS

6.1 Introduction

Crustaceans such as the *Macrobrachium rosenbergii* are known to possess an innate immune system, which consists of humoral and cellular activities that function as a defence against pathogen invasion. The regulation of the immune response involves several signalling pathways, one of which is the Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway. Various reports have suggested that the JAK/STAT pathway could play important role in both antibacterial and antiviral immunities.

In this study we report the full-length cDNA of STAT obtained from the transcriptome database of the freshwater prawn *M. rosenbergii*, designated as MrSTAT. The MrSTAT full length cDNA is 2907 base pairs (bp), with an open reading frame (ORF) of 2436 bp encoding for an 811 amino acid long protein. It contains 102 bp 5' untranslated regions (UTRs) and 240 bp 3' UTRs, including a stop codon. The MrSTAT coding region contains four domains: a STAT_INT domain, STAT alpha domain, STAT_bind domain and SH2 domain. 24 high probability common motifs were also found in the amino acid sequence of MrSTAT. A phylogenetic analysis reveals that the MrSTAT is clustered, with both STAT5s from vertebrates and STATs from invertebrates. The three-dimensional structure of the MrSTAT, modelled by the I-TASSER program, was compared with other experimentally identified structures deposited in the Protein

Data Bank to identify similarities. The results showed that the MrSTAT exhibits maximum structural identity with unphosphorylated STAT1 (PDB id 1yvl) from *Homo sapiens*. The minimum free energy of the predicted RNA structure of MrSTAT is -753.40 kcal/mol.

A tissue distribution analysis by quantitative real-time PCR showed a wide distribution of all the detected tissues for MrSTAT. A time course analysis of the transcription level of MrSTAT in the hepatopancreas after challenge by WSSV and *V. parahaemolyticus* showed a significant ($p < 0.05$) increase in expression, with the highest levels of expression occurring at 6 hours for the WSSV infected specimens and 24 hours for the *V. parahaemolyticus* infected ones. All these results suggest a potential role for the JAK/STAT pathway in immune response against bacteria and virus for *M. rosenbergii*.

6.2 Literature Review

Shrimp aquaculture represents a very large industry, which has been greatly affected by a growing number of diseases caused by virus and bacteria. Among the 22 viruses reported so far, the white spot syndrome virus (WSSV) has had a particularly devastating impact on shrimp culture, and continues to hinder sustainable shrimp farming (Bachère, 2000, van Hulten *et al*., 2001). In addition, another disease known as Vibriosis, caused by *Vibrio* sp. bacteria such as *Vibrio parahaemolyticus*, has also caused similar damaging effects on the industry (Rajendran *et al.*, 1999). A better understanding of the immune system of shrimps should be very helpful to tackling these severe disease problems in shrimp aquaculture.

In both invertebrates and vertebrates, the innate immune response forms the first line of defence against various pathogen infections (Iwanaga and Lee, 2005, Janeway Jr and Medzhitov, 2002). Once activated via the signal transduction pathway, this response leads to various humoral and cellular activities to clear the pathogen and ensure the survival of the host. The signal transduction involves the binding of extracellular signalling molecular patterns, such as pathogen-associated molecular patterns (PAMPs), or of a viral protein antigen to cell-surface receptors. These trigger various events inside the cell as an immune reaction, including encapsulation, phagocytosis, coagulation, melanisation and the secretion of potent antimicrobial peptides (AMPs). During this process, several signalling cascades such as the Toll pathway, Imd pathway and JAK/STAT pathway are known to play an important role (Li and Xiang, 2013).

The Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway has been previously reported to exist in many multicellular organisms, such as mammals, fishes, insects, crustaceans and nematodes (Rawlings *et al*., 2004, Guo *et al*., 2009, Agaisse and Perrimon, 2004, Chen *et al*., 2008, Plowman *et al*., 1999). This pathway plays an important role in various biological processes, such as cell proliferation and differentiation, developmental regulation, growth control, stem cell maintenance, and immune responses (Rawlings *et al*., 2004). The JAK/STAT pathway consists of three main cellular components: the receptor Domeless, the Janus kinase (JAK) Hopscotch and the STAT transcription factor (Chen *et al*., 2002). In mammals, the JAK/STAT pathway has been well studied, and seven STAT family members have been discovered, namely STAT1, STAT2, STAT3, STAT4, STAT5 (STAT5A and STAT5B) and STAT6 (Chang *et al.*, 1998). All these STAT family members in mammals have similar characteristics such as structural functional motifs, an N-terminal domain, a coiled coil domain, a DNA binding domain, a linker domain, an Src homology 2 (SH2) domain and a transcription

activation domain (Chang *et al*., 1998). The binding of extracellular cytokines induces activation of the intracellular Janus kinase. The activated Janus kinase phosphorylates a specific tyrosine residue in the STAT protein which later initiates the dimerization of STAT monomers via their SH2 domain. After dimerization, the dimer moves to the cell nucleus to bind to specific DNA sites and initiates the transcription mechanism.

The JAK/STAT pathway also plays a pivotal role in antiviral and antibacterial defence in insect (Barillas - Mury *et al.,* 1999, Kingsolver *et al.,* 2013). In *A. gambiae*, the STAT is highly induced during bacterial challenge, further enhancing its importance as an antibacterial defence (Barillas - Mury *et al.,* 1999). During bacterial challenge, the JAK/STAT pathway in *D. melanogaster* is reported to activate two gene families, TEP and TOT, which play a significant role in innate immunity (Thompson *et al.*, 2004). Finally, the transcriptional profile of Drosophila C Virus-infected *D. melanogaster* has also revealed the involvement of this pathway in antiviral mechanisms (Khuntia *et al.*, 2008).

Chen *et al.,* (2008) reported the first STAT homolog in *P. monodon*, with typical functional domains including the DNA binding domain, SH2 domain and C-terminal transactivation domain, which were found conserved with those similar domains reported in insect. Sun *et al*., (2011) also isolated the full-length cDNA of the STAT homolog from specimens of *F. chinensis* which had responded during WSSV and *V. anguillarum* infections, thus indicating its importance in shrimp responses to viral and bacterial infections.

Unlike the penaeid shrimp, however, knowledge regarding the STAT protein is still lacking in the palaemonid shrimp such as the giant freshwater prawn *Macrobrachium rosenbergii*. The results from our transcriptome data, obtained after challenge by WSSV

and *V. parahaemolyticus*, showed the STAT gene to be significantly up-regulated at the 12th hour post infection. In order to gain further insights into the characterization of the STAT gene and its role in *M. rosenbergii*, we report below the bioinformatics analysis and tissue-specific mRNA expression of the STAT gene. We also investigate its expression pattern in the hepatopancreas following challenge with WSSV and *V. parahaemolyticus* at high-intensity sampling points (0, 3, 6, 12, 24, 48 h post-injection), using qRT-PCR.

6.3 Materials and Methods

6.3.1 *M. rosenbergii*

Throughout the experiment, the *M. rosenbergii* juvenile prawns (5-8g body weight) used were purchased from a hatchery at Kuala Kangsar, Perak, Malaysia. The prawns were acclimatized in flat-bottomed glass tanks (300 L) with aerated and filtered freshwater at 28 ± 1 ºC in the laboratory for a week. A total of 20 prawns were maintained in each tank during the experiment. The prawns were randomly screened by the PCR method to ensure they were free from WSSV and *V. parahaemolyticus*.

6.3.2 Identification of *M. rosenbergii* STAT

A full length MrSTAT gene was retrieved from the *M. rosenbergii* transcriptome unigene database, generated by Illumina's Hiseq^{TM} 2000 sequencing platform. Briefly, mRNA isolated from *M. rosenbergii* infected with WSSV and *V. parahaemolyticus* was sequenced using the Illumina Hiseq^{TM} 2000 platform in order to obtain short read sequences. These short sequences were *de novo* assembled using the Trinity assembler to obtain a set of unigenes.

The sequences representing the STAT gene were then identified through a BLAST homology search against the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast).

6.3.3 MrSTAT Sequence Characterization

The full-length MrSTAT sequence was compared with other sequences available on the NCBI database, and the similarities were analysed. The open reading frame (ORF) and amino acid sequence of MrSTAT were obtained using DNAssist 2.2 (Lin *et al.*, 2004). Characteristic domains or motifs were identified using the PROSITE profile database (Hulo *et al*., 2006). Identity, similarity and gap percentages were calculated using the FASTA program (Dostert *et al.*, 2005). The N-terminal transmembrane sequence was determined by the DAS transmembrane prediction program (Cserzö *et al*., 1997). A signal peptide analysis was conducted using the SignalP (Peterson *et al*., 2011). The cellular localization of the depicted MrSTAT protein was predicted by the MultiLoc2 predictor online tool. The "MultiLoc2-HighRes (Animal), 9 localizations" prediction method was selected so as to predict the cellular localizations at a higher resolution (Blum *et al.*, 2009). Pair-wise and multiple sequence alignment were constructed in the Clustal Omega Program and edited using BioEdit (ver. 7.1.3.0) (Thompson *et al*., 2012). The phylogenetic tree was conducted using the Maximum Likelihood method on MEGA 6 (Tamura *et al.*, 2013). The secondary structure of the MrSTAT was elucidated utilising Polyview (Porollo, Adamczak & Meller, (2004). The 3D structure of the MrSTAT was determined using the I-TASSER program (Zhang, 2008). The RNA structure of the MrSTAT was predicted, along with the minimum free energy (MFE), using the RNA Fold Server program (Lorenz *et al*., 2011).

6.3.4 Disease Challenge

Three groups of *M. rosenbergii* were prepared for WSSV and *V. parahaemolyticus* induced MrSTAT gene expression analysis. The three sets of prawns in question were intramuscularly injected with 100 µl filtered supernatant obtained from WSSV infected *Penaeus monodon* ($1X10^5$ copies/ml), 100 µl $1X10^5$ cfu cultured *V. parahaemolyticus* PCV08-7, and 100µl 2% NaCl (1:10, w/v) solution, respectively. The last of these groups served as a negative control group. Samples were collected before (i.e. at 0 h) and at intervals of 3, 6, 12, 24 and 48 h after injection. The tissues were immediately snap-frozen in liquid nitrogen and stored at -80°C until total RNA was isolated.

6.3.5 RNA Extraction and cDNA Conversion

The total RNA isolation of the *M. rosenbergii* tissues was performed using TRIZOL reagent, following the manufacturer's protocol (Life Technologies, Carlsbad, CA). DNA contamination was eliminated by treating the total RNA with RNase free DNAse set (5 Prime GmbH, Hamburg, Germany). The total RNA concentration was measured spectrophotometrically (NanoVue Plus Spectrophotometer, GE Healthcare UK Ltd, England). First-strand cDNA was then synthesized from the total RNA by M-MLV reverse transcriptase (Promega, USA) followed by oligo-dTVN primer, in accordance with the manufacturer's protocol.

6.3.6 qRT-PCR Analysis of MrSTAT

The relative expression of MrSTAT in the eye stalk, gills, pleopod, stomach, hepatopancreas, muscle, intestine and haemocyte were measured by quantitative real time polymerase chain reaction (qRT-PCR). The primer design for the MrSTAT was performed by using Primer3 software (Untergasser, A *et al*., 2012) and listed in Table

6.1. The qRT-PCR reaction (20 µl) consisted of 10 µl TaqMan Universal RT-PCR Master Mix (Applied Biosystems, Foster City, CA, USA), 1 µl of primers/probe set containing 900 nM of forward reverse primers, a 300 nM probe and 2 µl of template cDNA. The RT-PCR program consisted of incubation at 50°C for 2 min, and 40 cycles at 95°C for 10 min, 95°C for 15 sec, and 60°C for 1 min, with the Step One Plus Real-Time PCR System® (Applied Biosystems). The same qRT-PCR cycle profile was used for the internal control gene, elongation factor 1-alpha (ELF-1) primers (Dhar *et al.,* 2009).

Table 6.1. List of genes, primers and probes used in real-time TaqMan PCR assay

Gene	Gene Abbreviations	NCBI NO	Forward Primer Sequence	Reverse Primer Sequence	Probe Sequence
***M. rosenbergii* STAT**	MrSTAT	NA	ACCAACCCTCAATTCCCATA	TTCAACTTTCCACCAACCAA	CCCGTTTCACTGCCATAGTGAGGC
Elongation factor 1-alpha	EF1-A	EL609261.1	ACTGCGCTGTGTTGATTGTAGCT	ACAACAGTACGTGTTCACGGGTC	TACTGGTGAGTTTGAAGCT

NA= not available

After the PCR program, the data were analysed with ABI 7500 SDS software (Applied Biosystems). To maintain consistency, the baseline was set automatically by the software. The comparative CT method ($2^{-\Delta\Delta CT}$ method) was used to analyse the expression level of the MrSTAT (Schmittgen and Livak, 2008). All samples were analysed in three duplications, with the overall results expressed in terms of one sample through the mean ± standard deviation (Schmittgen and Livak, 2008). One way ANOVA with post-hoc Tukey HSD test was used for statistical analysis. Values were considered to be statistically significant at $P < 0.05$.

6.4 Results

6.4.1 cDNA and Amino Acid Sequences Analysis of MrSTAT

A full length MrSTAT sequence was retrieved from the *M. rosenbergii* transcriptome unigenes obtained by Illumina's HiSeq^{TM}2000 sequencing technology, and was submitted to the EMBL GenBank database under the accession number KT380661. The nucleotide and deduced amino acid structure of the MrSTAT is given in Figure 6.1. The MrSTAT full length cDNA was 2907 base pairs (bp), with an open reading frame (ORF) of 2436 bp encoding for an 811 amino acid long protein with a theoretical molecular mass of 93 kDa and an isoelectric point of 6.04. The signal peptide prediction program revealed no putative signal peptide.

The MrSTAT cDNA contained 102 bp 5' untranslated regions (UTRs) and 240 bp 3' UTRs including a stop codon. The MrSTAT coding region contained 4 domains: a STAT_INT domain (amino-terminal domain) (17aa-145aa), a STAT alpha domain (coiled-coil domain) (157aa-352aa), a STAT_bind domain (354aa-604aa) and an SH2 domain (608aa-701aa) (Figure 6.2). 24 high probability common motifs were found in the amino acid sequence of the MrSTAT (Table 6.1). The cellular location prediction analysis indicated (with an estimated probability of 0.86) that the MrSTAT protein was most likely to be present in the nuclear region.

1 taacacttca gccattttca cggcttcata aacctcaagg gatttcccag agaaagacgt

61 gttattgtca caaccagatc gtgatatctc gcttctttac ca

103	atg	agt	cct	aac	aaa	gta	ggc	att	gaa	gcc	aag	gtc	aag	acc	aca	
	M	S	P	N	K	V	G	I	E	A	K	V	K	T	T	15
148	atg	tcg	ttg	tgg	aat	aga	gca	caa	cag	tta	ccc	caa	gac	gcg	ctt	
	M	S	L	W	N	R	A	Q	Q	L	P	Q	D	A	L	30
193	agg	cag	gtt	caa	aat	gta	tac	aat	gaa	cag	ttc	ccc	atc	gaa	gtg	
	R	Q	V	Q	N	V	Y	N	E	Q	F	P	I	E	V	45
238	cga	cat	tac	ttg	gct	gga	tgg	att	gaa	gaa	aaa	ata	cat	cag	tgg	
	R	H	Y	L	A	G	W	I	E	E	K	I	H	Q	W	60
283	aat	gaa	att	gat	cca	gac	aac	cca	gca	cac	agc	cag	tat	gct	cac	
	N	E	I	D	P	D	N	P	A	H	S	Q	Y	A	H	75
328	act	atc	gta	tca	cag	ctt	atc	caa	gag	atg	gaa	aat	aaa	tca	ttg	
	T	I	V	S	Q	L	I	Q	E	M	E	N	K	S	L	90
373	agc	tat	gtt	aac	aat	gaa	gac	cta	ttt	tta	gtg	cgg	atg	cgt	tta	
	S	Y	V	N	N	E	D	L	F	L	V	R	M	R	L	105
418	aat	gaa	gct	gct	aat	tta	ttt	aag	act	cgg	tac	ctg	aat	act	aat	
	N	E	A	A	N	L	F	K	T	R	Y	L	N	T	N	120
463	cca	ttg	gcc	ctt	gtg	tca	atc	att	cgt	aat	tgt	cta	aat	aca	gaa	
	P	L	A	L	V	S	I	I	R	N	C	L	N	T	E	135
508	ctt	aac	ctc	gtc	caa	caa	cac	gaa	agc	atg	ctg	ggt	ggc	gta	ggg	
	L	N	L	V	Q	Q	H	E	S	M	L	G	G	V	G	150

553	cca	ggt	gta	aac	atg	atc	gtt	gaa	ccc	tgt	aca	gaa	att	gtt	cag	
	P	G	V	N	M	I	V	E	P	C	T	E	I	V	Q	165
598	gaa	tta	gag	gtg	ctg	cac	aga	cgt	acc	cgg	gaa	act	gca	gat	gag	
	E	L	E	V	L	H	R	R	T	R	E	T	A	D	E	180
643	ttg	cga	cag	cta	gag	cag	gaa	cag	gaa	tca	ttt	gcc	ctt	cag	tac	
	L	R	Q	L	E	Q	E	Q	E	S	F	A	L	Q	Y	195
688	cat	gac	tgc	gca	aag	atc	aat	gct	cat	tta	tcc	cac	ata	caa	tca	
	H	D	C	A	K	I	N	A	H	L	S	H	I	Q	S	210
733	caa	gag	aga	aca	cca	cag	aat	aga	gat	gtg	gaa	atg	aat	tta	cgc	
	Q	E	R	T	P	Q	N	R	D	V	E	M	N	L	R	225
778	aag	cga	aaa	gaa	gtt	ggg	gag	caa	cag	tta	gca	caa	aag	gtt	tct	
	K	R	K	E	V	G	E	Q	Q	L	A	Q	K	V	S	240
823	ggg	tta	ttg	caa	cgg	cgg	atg	gcg	cta	gca	gaa	aaa	cat	aaa	gga	
	G	L	L	Q	R	R	M	A	L	A	E	K	H	K	G	255
868	acc	ata	gac	cga	ctc	aac	agt	tta	caa	cag	cgt	att	tta	gac	gaa	
	T	I	D	R	L	N	S	L	Q	Q	R	I	L	D	E	270
913	gag	ttg	atc	aac	tgg	aag	aga	gaa	caa	caa	atg	gct	ggg	aat	ggc	
	E	L	I	N	W	K	R	E	Q	Q	M	A	G	N	G	285
958	cga	cct	ttc	aac	caa	aat	aag	ctt	gat	caa	ata	caa	gag	tgg	tgt	
	R	P	F	N	Q	N	K	L	D	Q	I	Q	E	W	C	300
1003	gag	gct	ctg	gca	gaa	ata	att	tgg	cta	aac	cga	cac	caa	ata	aaa	
	E	A	L	A	E	I	I	W	L	N	R	H	Q	I	K	315
1048	gag	tgt	gaa	cga	cat	cag	acc	aag	att	cca	ata	gcc	cct	cca	gga	
	E	C	E	R	H	Q	T	K	I	P	I	A	P	P	G	330
1093	ggt	gtt	gat	atg	tta	cca	acc	ctc	aat	tcc	cat	atc	act	cgt	ctt	
	G	V	D	M	L	P	T	L	N	S	H	I	T	R	L	345
1138	ctc	tct	tca	ctt	gtt	acc	agc	aca	ttc	ata	ata	gag	aaa	cag	cct	
	L	S	S	L	V	T	S	T	F	I	I	E	K	Q	P	360

1183	cca	caa	gtt	atg	aag	acc	aac	acc	cgt	ttc	act	gct	aca	gtg	agg	
	P	Q	V	M	K	T	N	T	R	F	T	A	T	V	R	375
1228	cta	ttg	gtt	ggt	gga	aag	ttg	aat	gta	aat	atg	act	cca	cct	caa	
	L	L	V	G	G	K	L	N	V	N	M	T	P	P	Q	390
1273	gta	cgg	gtg	tcc	atc	atc	agt	gag	gcg	cag	gca	aat	gct	ctc	cta	
	V	R	V	S	I	I	S	E	A	Q	A	N	A	L	L	405
1318	aag	aat	gat	cag	atg	aac	aaa	gga	gaa	cag	tcg	ggt	gaa	att	cta	
	K	N	D	Q	M	N	K	G	E	Q	S	G	E	I	L	420
1363	aat	aat	aca	ggc	acc	atg	gaa	tac	cac	cag	ggt	acg	agg	cag	ctt	
	N	N	T	G	T	M	E	Y	H	Q	G	T	R	Q	L	435
1408	tcc	gtc	agt	ttc	cgc	aat	atg	cag	tta	aga	aaa	att	aaa	cga	gct	
	S	V	S	F	R	N	M	Q	L	R	K	I	K	R	A	450
1453	gaa	aag	aaa	gga	aca	gag	tct	gtg	atg	gat	gaa	aag	ttt	tcg	ctt	
	E	K	K	G	T	E	S	V	M	D	E	K	F	S	L	465
1498	ctt	ttc	caa	tca	caa	ttc	agt	gta	gga	gga	gga	gaa	tta	gta	ttc	
	L	F	Q	S	Q	F	S	V	G	G	G	E	L	V	F	480
1543	cag	gtg	tgg	acg	cta	tcc	cta	cct	gtg	gtg	gtc	att	gtc	cac	ggt	
	Q	v	W	T	L	S	L	P	V	V	V	I	V	H	G	495
1588	aat	cag	gag	cca	cat	gct	tgg	gca	act	gtc	tct	tgg	gat	aat	gca	
	N	Q	E	P	H	A	W	A	T	V	S	W	D	N	A	510
1633	ttt	gct	gaa	caa	ggt	cgc	ata	cct	ttc	aca	gtc	cca	gaa	aag	gta	
	F	A	E	Q	G	R	I	P	F	T	V	P	E	K	V	525
1678	cct	tgg	cca	cag	ata	gct	gaa	atg	tta	gac	aca	aaa	ttc	aag	gct	
	P	W	P	Q	I	A	E	M	L	D	T	K	F	K	A	540
1723	gca	act	ggc	agg	ggt	ctt	aca	gaa	gat	aac	ctg	aag	ttc	tta	gca	
	A	T	G	R	G	L	T	E	D	N	L	K	F	L	A	555
1768	ggc	aaa	gcc	ttt	cgt	ctt	gat	agc	tct	caa	gtt	caa	gac	ttc	act	
	G	K	A	F	R	L	D	S	S	Q	V	Q	D	F	T	570

1813	aac	atg	ttg	ctg	tca	tgg	tca	cag	ttc	tgt	aag	gag	cca	ctt	tct	
	N	M	L	L	S	W	S	Q	F	C	K	E	P	L	S	585
1858	gag	cgt	aat	ttc	act	ttt	tgg	gaa	tgg	ttc	ttt	gct	gtt	atg	aag	
	E	R	N	F	T	F	W	E	W	F	F	A	V	M	K	600
1903	gta	acc	aga	gaa	cac	ctt	cgt	caa	cct	tgg	aat	gat	ggt	tct	atc	
	V	T	R	E	H	L	R	Q	P	W	N	D	G	S	I	615
1948	atg	ggc	ttt	gtt	gga	cgt	cgc	ccg	gca	gag	gag	atg	ctg	aaa	aat	
	M	G	F	V	G	R	R	P	A	E	E	M	L	K	N	630
1993	tcg	aaa	agt	gga	acg	ttc	ctc	tta	aga	ttt	tcc	gac	tca	gaa	ttg	
	S	K	S	G	T	F	L	L	R	F	S	D	S	E	L	645
2038	gga	ggg	gtt	acc	att	gca	tgg	atg	tat	gaa	gat	act	aca	aaa	ggt	
	G	G	V	T	I	A	W	M	Y	E	D	T	T	K	G	660
2083	gac	cag	cgg	gat	gtc	ttc	atg	ttg	cag	cct	ttc	aca	agc	aag	gct	
	D	Q	R	D	V	F	M	L	Q	P	F	T	S	K	A	675
2128	ttt	gca	atc	cgt	cca	cta	gct	gat	gtt	att	gct	gac	ttg	aag	tat	
	F	A	I	R	P	L	A	D	V	I	A	D	L	K	Y	690
2173	ttg	ctc	tat	tta	tat	cca	aat	gtg	cct	aaa	gag	cag	gca	ttt	gga	
	L	L	Y	L	Y	P	N	V	P	K	E	Q	A	F	G	705
2218	aag	tac	tac	act	cca	atg	gga	gga	gag	cag	ccc	aca	aat	aat	gga	
	K	Y	Y	T	P	M	G	G	E	Q	P	T	N	N	G	720
2263	tat	gtg	aaa	cca	cat	ctt	atc	act	cat	gta	cct	gga	tgg	tca	gtg	
	Y	V	K	P	H	L	I	T	H	V	P	G	W	S	V	735
2308	gct	gga	ggg	tcg	atg	gat	tct	tat	ccc	aac	aca	ccg	caa	cct	atg	
	A	G	G	S	M	D	S	Y	P	N	T	P	Q	P	M	750
2353	tac	cca	atg	cat	gac	agc	aat	atg	ggt	gat	cct	ccg	tcc	gtc	agt	
	Y	P	M	H	D	S	N	M	G	D	P	P	S	V	S	765
2398	tcc	aat	ccc	tcc	gac	agt	gtc	tca	aca	atg	ccc	cca	tac	aac	aac	
	S	N	P	S	D	S	V	S	T	M	P	P	Y	N	D	780

2443	acc	gat	tat	cct	gac	atc	ttg	gaa	aat	cta	cca	gat	act	gac	ttc	
	T	D	Y	P	D	I	L	E	N	L	P	D	T	D	F	795
2488	act	gat	atc	aac	ctt	gac	ttt	ctt	caa	acc	aac	ttc	atg	aag	ccc	
	T	D	I	N	L	D	F	L	Q	T	N	F	M	K	P	810
2533	cag	taa														
	Q	*														811

2539 tcagatgaag attttttta attgtcatac tgtatataag cattgtttat tggatagcct

2600 tgcttttaat atcttaggat ggtctttaat tgaattttgc atttaaaaag gcaaagaaaa

2661 agaagagtgt attttcagag gatgatgtgt aaccatagtg tcttaggggt atgctttaat

2722 gttaggtgaa tatgaaatcg ttgttacaag aattctgcaa acgtgtatgt aggaattacc

Figure 6.1. Nucleotide and deduced amino acid sequence of *M. rosenbergii* STAT gene. The nucleotide sequence is numbered from 5' end, and the single letter amino acid code is shown below the corresponding codon. The start codon (ATG) and stop codon (TGA) are highlighted in red shade. The termination code is marked with an asterisk.

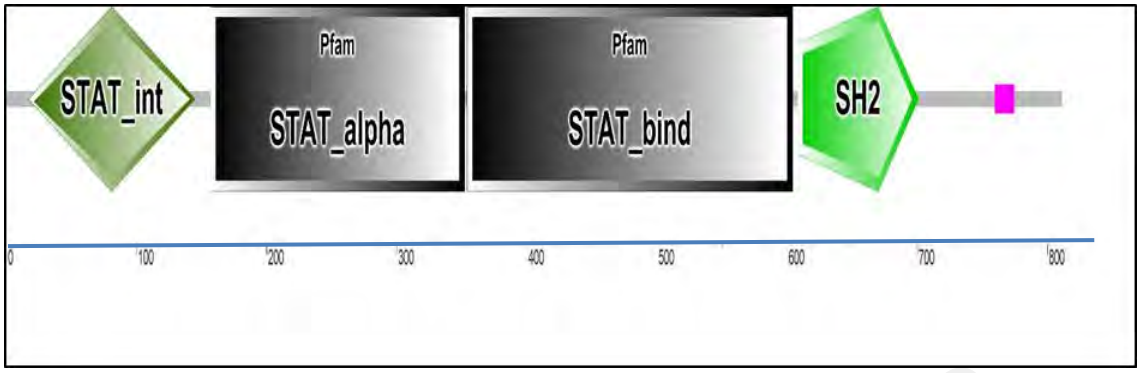


Figure 6.2. Position of MrSTAT domains.

Table 6.2: List of common motifs present in MrSTAT predicted by PROSITE scan analysis

Motif (numbers)	Position of the amino acid
Src homology 2 (SH2) domain profile(1)	610 - 711
N-glycosylation site(4)	87 - 90, 421-424, 588-591, 779-782
N-myristoylation site (5)	148-153, 379-384, 613-618, 647-652, 738-743
Casein kinase II phosphorylation site(6)	177-180, 457-460, 520-523, 590-593, 641-644, 658-661
Protein kinase C phosphorylation site(6)	373 - 375, 438-440, 542-544, 585-587, 657-659, 672-674
cAMP- and cGMP-dependent protein kinase phosphorylation site (1)	452-455
Amidation site(1)	619-622

6.4.2 Homologous Analysis of MrSTAT

A homology analysis using a BLAST search showed the MrSTAT protein to have a high similarity with the STATs from other crustaceans such as *S. paramamosain* (88%), tiger shrimp *P. monodon* (87%), white shrimp *L. vannamei* (87%), kuruma shrimp *M. japonicus* (86%) and *F. chinensis* (89%). No homology with MrSTAT from *M. rosenbergii* was detected. MrSTAT was aligned with seven homologous sequences using ClustalW multiple analyses (Figure 6.3). Interestingly, the results showed MrSTAT to have the longest amino acid sequence (811aa) compared with rest, with the highest degree of similarity and conservation of amino acids occurring mainly in the STAT bind domain.

```
MrSTAT          MSPNKVGIEAKVKTTMSLWNRAQQLPQDALRQVQNVYNEQFPIEVRHYLAGWIEEKIHQW 60
SpSTAT          ---------------MSLWNRAQQLPPDALREVQNVYGEQFPIEVRHYLAGWIEDKMHQW 45
PmSTAT          ---------------MSLWNRAQQLPADDLRRVQGIYGEQFPIEVRHYLAGWIEDKMQQW 45
LvSTAT          ---------------MSLWNRAQQLPPXDLRRVQGIYGEQFPIEVRHYLAVWIEDKMQQW 45
MjSTAT          ---------------MSLWNRAQQLPPDDLRRVQGIYGEQFPIEVRHYLAGWIEDKMQQW 45
FcSTAT          ---------------MSLWNRAQQLPPDDLRRVQGIYGEQFPIEVRHYLAGWIEDKMQQW 45
BmSTAT          ---------------MAVWIQAQQLQGDALHQMQALYGQHFPIEVRHYLSQWIES--QAW 43
HsSTAT          ---------------MAGWIQAQQLQGDALRQMQVLYGQHFPIEVRHYLAQWIES--QPW 43
                                *: * :****    *:.:* :*.::*********: ***.  : *

MrSTAT          NEIDPDNPAHSQYAHTIVSQLIQEMENKSLSYVNNEDLFLVRMRLNEAANLFKTRYLNTN 120
SpSTAT          NEIDPENVSHSQYAHSLVSQLIQEIENKALNYGSNEDLFLVRIRLDEAANMFKTRYLNSN 105
PmSTAT          NEIDPDNPSHSQYAQSLVSQLIQEIENKALSYANNEDLFLVRMRLDEAATSFRTRYLNSN 105
LvSTAT          NEIDPDNPSHTQYAQSLVSQLIQEIENKALSYANNEDLFLVRMRLDEAATSFRTRYLNSN 105
MjSTAT          NEIDPDNPSHSQYAQSLVSQLIQEIENKALSYANNEDLFLVRMRLDEAATSFKTRYLNSN 105
FcSTAT          NEIDPDNPSHSQYAQSLVSQLIQEIENKALSYANNEDLFLVRMRLDEAATSFRTRYLNSN 105
BmSTAT          DSIDLDNPQENIKATQLLEGLVQELQKKAEHQVG-EDGFLLKIKLGHYATQLQNTYDR-C 101
```

```
HsSTAT          DAIDLDNPQDRAQATQLLEGLVQELQKKAEHQVG-EDGFLLKIKLGHYATQLQKTYDR-C 101
                : ** :*  .   *  ::. *:**:::*:    . ** **::::*.. *. ::. * .

MrSTAT          PLALVSIIRNCLNTELNLVQQHESMLGG-VGPGVNMIVEPCTEIVQELEVLHRRTRETAD 179
SpSTAT          PLVLVGIIQNCLKIELQLVQRHENMLG--APHTTNMVIEPCAEIVNELDILHRRTRETAD 163
PmSTAT          PLGLVGIIRQCLNTEHNLVQQNENMLGGGVSHATNMVIEPCAEIEQELRILHERTRETAN 165
LvSTAT          PLGLVGIIRQCLNTEHNLVQQNENMLGGGVSHATNMVIEPCAEIEQELRILHERTRETAN 165
MjSTAT          PLGLVGIIRQCLNTEHNLVQQNENMLGGSVPHATNMVVEPCAEIEQELRILHERTRETAN 165
FcSTAT          PLGLVGIIRQCLNTEHNLVQQNESMLGGGVSHATNMVIEPCAEIEQELRILHERTRETAN 165
BmSTAT          PMELVRCIRHILYNEQRLVREANNGTSP-AGSLADAMSQKHLQINQTFEELRLVTQDTEN 160
HsSTAT          PLELVRCIRHILYNEQRLVREANNCSSP-AGILVDAMSQKHLQINQTFEELRLVTQDTEN 160
                *: **  *:: *  * .**:. :.  .  .   .: : :  :* : :  *:  *::* :

MrSTAT          ELRQLEQEQESFALQYHDCAKINAHLSHIQSQERTPQNRDVEMNLRKRKEVGEQQLAQKV 239
SpSTAT          ILRQLEQEQESFALQYHDCTKINAHLSHIQSQEKTPQNREVEANLRRRKQLGEQQLTEKV 223
PmSTAT          ELRHLEQEQESFALQYHDCAKINAHLSHIQSQERTQQNREMEQSLRRRKELGEQQLAQKV 225
LvSTAT          ELRHLEQEQESFALQYHDCAKINAHLSHIQSQERTQQNREMEQSLRRRKELGEQQLAQKV 225
MjSTAT          ELRHLEQEQESFALQYHDCAKINAHLSHIQSQERTQQNREMEQSLRRRKELGEQQLAQKV 225
FcSTAT          ELRHLEQEQESFALQYHDCAKINAHLSHIQSQERTQQNREMEQSLRRRKELGEQQLAQKV 225
BmSTAT          ELKKLQQTQEYFIIQYQESLRIQAQFAQL-AQLNPQERLSRETALQQKQVSLEAWLQREA 219
HsSTAT          ELKKLQQTQEYFIIQYQESLRIQAQFAQL-AQLSPQERLSRETALQQKQVSLEAWLQREA 219
                 *:*:* ** * :**::. :*:*::::: :*  . :. . *  *::::   *  * .:.

MrSTAT          SGLLQRRMALAEKHKGTIDRLNSLQQRILDEELINWKREQQMAGNGRPFNQNKLDQIQEW 299
SpSTAT          SGLLQRRMDLAEKHKGTIDRLNILQQRILDEELINWKREQQMAGNGKPFNGNKLDTIQEW 283
PmSTAT          SGLLQLRMALADKHKGTIDRLNSLQQRILDEELINWKRDQQMHGNGKPFNPNKLDQIQEW 285
LvSTAT          SGLLQLRMALADKHKGTIDRLNSLQQRILDEELINWKRDQQMHGNGKPFNPNKLDQIQEW 285
MjSTAT          SGLLQLRMALADKHKGTIDRLNSLQQRILDEELINWKRDQQMHGNGKPFNQNNWDQIQEW 285
FcSTAT          SGLLQLRMALADKHKGTIDRLNSLQQRILDEELINWKRDQQMHGNGKPFNPNKLDQIQEW 285
BmSTAT          QTLQQYRVELAEKHQKTLQLLRKQQTIILDDELIQWKRRQQLAGNGGPPEGS-LDVLQSW 278
HsSTAT          QTLQQYRVELAEKHQKTLQLLRKQQTIILDDELIQWKRRQQLAGNGGPPEGS-LDVLQSW 278
```

```
                . * * *: **:**: *:: *.  *  ***:***:*** **: *** * : .  * :*.*

MrSTAT          CEALAEIIWLNRHQIKECERHQTKIPIAPPGGVDMLPTLNSHITRLLSSLVTSTFIIEKQ 359
SpSTAT          CEALAEIIWLNRHQIKECERHQTKIPITPHGGVDMLPTLNSHITRLLSSLVTSTFIIEKQ 343
PmSTAT          CEALAEIIWLNRHQIKGCERHQTKIPITPPGGVDMLPTLNSHITRLLSSLVTSTFIIEKQ 345
LvSTAT          CEALAEIIWLNRHQIKECERHQTKIPIAPPGGVDMLPTLNSHITRLLSSLVTSTFIIEKQ 345
MjSTAT          CEALAEIIWLNRHQIKECERHQTKIPIAPPGGVDMLPTLNSHITRLLSSLVTSTFIIEKQ 345
FcSTAT          CEALAEIIWLNRHQIKECERHQTKIPIAPPGGVDMLPTLNSHITRLLSSLVTSTFIIEKQ 345
BmSTAT          CEKLAEIIWQNRQQIRRAEHLCQQLPIPGP-VEEMLAEVNATITDIISALVTSTFIIEKQ 337
HsSTAT          CEKLAEIIWQNRQQIRRAEHLCQQLPIPGP-VEEMLAEVNATITDIISALVTSTFIIEKQ 337
                ** ****** **:**: .*:   ::**.     :**. :*: ** ::*:***********

MrSTAT          PPQVMKTNTRFTATVRLLVGGKLNVNMTPPQVRVSIISEAQANALLKNDQMNKGEQSGEI 419
SpSTAT          PPQVMKTNTRFSATVRLLVGGKLNVNMTPPQVRVSIISEAQANALLKNDQMSKGEMSGEI 403
PmSTAT          PPQVMKTNTRFTATVRLLVGGKLNVNMTPPQVRVSIISEAQANALLKNDQMNKGEQSGEI 405
LvSTAT          PPQVMKTNTRFTATVRLLVGGKLNVNMTPPQVRVSIISEAQANALLKNDQMNKGEQSGEI 405
MjSTAT          PPQVMKTNTRFTATVRLLVGGKLNVNMTPPQVRVSIISEAQANALLKNDQMNKGEQSGEI 405
FcSTAT          PPQVMKTNTRFTATVRLLVGGKLNVNMTPPQVRVSIISEAQANALLKNDQMNKGEQSGEI 405
BmSTAT          PPQVLKTQTKFAATVRLLVGGKLNVHMNPPQVKATIISEQQAKSLLKNENTRN-DYSGEI 396
HsSTAT          PPQVLKTQTKFAATVRLLVGGKLNVHMNPPQVKATIISEQQAKSLLKNENTRN-ECSGEI 396
                ****.**:*:*:************:*.****:.:**** **::****::  : : ****

MrSTAT          LNNTGTMEYHQGTRQLSVSFRNMQLRKIKRAEKKGTESVMDEKFSLLFQSQFSVGGGELV 479
SpSTAT          LNNTGTMEYHQSSRQLSVSFRNMQLRKIKRAEKKGTESVMDEKFSLLFQSQFSVGGGELV 463
PmSTAT          LNNTGTMEYNQTSRQLSVSFRNMQLRKIKRAEKKGTESVMDEKFSLLFQSQFSVGGGELV 465
LvSTAT          LNNTGTMEYNQTSRQLSVSFRNMQLRKIKRAEKKGTESVMDEKFSLLFQSQFSVGGGELV 465
MjSTAT          LNNTGTMEYNQGTRQLSVSFRNMQLRKIKRAEKKGTESVMDEKFSLLFQSQFSVGGGELA 465
FcSTAT          LNNTGTMEYNQTSRQLSVSFRNMQLRKIKRAEKKGTESVMDEKFSLLFQSQFSVGGGELV 465
BmSTAT          LNNCCVMEYHQATGTLSAHFRNMSLKRIKRSDRRGAESVTEEKFTILFESQFSVGGNELV 456
HsSTAT          LNNCCVMEYHQATGTLSAHFRNMSLKRIKRADRRGAESVTEEKFTVLFESQFSVGSNELV 456
                ***  .***:* :  **. ****.*::***::::*:*** :***::**:******..**.
```

```
MrSTAT          -FQVWTLSLPVVVIVHGNQEPHAWATVSWDNAFAEQGRIPFTVPEKVPWPQIAEMLDTKF 538
SpSTAT          -FQVWTLSLPVVVIVHGNQEPHAWATVSWDNAFAEQGRIPFTVPDKVPWPQVADMLDAKF 522
PmSTAT          -FQVWTLSLPVVVIVHGNQEPHAWATVSWDNAFAEQGRIPFTVPEKVPWPQIADMLDTKF 524
LvSTAT          -FQVWTLSLPVVVIVHGNQEPHAWATVSWDNAFAEQGRIPFTVPEKVPWPQIADMLDTKF 524
MjSTAT          PLQVWTLSLPVVVIVHGNQEPHAWATVSWDNAFAEQGRIPFTVPEKVPWPQIADMLDTKF 525
FcSTAT          -FQVWTLSLPVVVIVHGNQEPHAWATVSWDNAFAEQGRIPFTVPEKVPWPQIADMLDTKF 524
BmSTAT          -FQVKTLSLPVVVIVHGSQDNNATATVLWDNAFAEPGRVPFAVPDKVLWPQLCEALNMKF 515
HsSTAT          -FQVKTLSLPVVVIVHGSQDHNATATVLWDNAFAEPGRVPFAVPDKVLWPQLCEALNMKF 515
                 :** ************.*: :* *** ******* **:**:**:** ***:.: *: **

MrSTAT          KAATG--RGLTEDNLKFLAGKAFRLD-SSQVQDFTNMLLSWSQFCKEPLSERNFTFWEWF 595
SpSTAT          KSATG--RGLTEDNLRFLAGKAFRLDRSPQVQDFTNMMLSWSQFCKEPLSERNFTFWEWF 580
PmSTAT          KAATG--RGLTEDNLKFLAGKAFRNP---QVQDFTNMMLSWSQFCKEPLSERNFTFWEWF 579
LvSTAT          KAATG--RGLTEDNLKFLAGKAFRNP---QVQDFTNMMLSWSQFCKEPLSERNFTFWEWF 579
MjSTAT          KAATG--RGLTEDNLKFLAGKAFRNP---QVQDFTNMMLSWSQFCKEPLSERNFAFWEWF 580
FcSTAT          KAATG--RGLTEDNLKFLAGKAFRNP---QVQDFTNMMLSWSQFCKEPLSERNFTFWEWF 579
BmSTAT          KAEVQSNRGLTKENLVFLAQKLFNSS-SSHLEDYNGMSVSWSQFNRENLPGRNYTFWQWF 574
HsSTAT          KAEVQSNRGLTKENLVFLAQKLFNNS-SSHLEDYSGLSVSWSQFNRENLPGWNYTFWQWF 574
                *: .  ****::** *** * *.     :::*:..: :***** :* *.  *::**:**

MrSTAT          FAVMKVTREHLRQPWNDGSIMGFVGRRPAEEMLKNSKSGTFLLRFSDSELGGVTIAWMYE 655
SpSTAT          FAVMKVTKEHLRQPWNDGSIMGFVGRRPAEEMLKNSKSGTFLLRFSDSELGGVTIAWMYE 640
PmSTAT          FAVMKVTREHLRQQWNDGSIMGFVGRRQAEEMLKNSKSGTFLLRFSDSELGGVTIAWMYE 639
LvSTAT          FAVMKVTREHLRQQWNDGSIMGFVGRRQAEEMLKNSKSGTFLLRFSDSELGGVTIAWMYE 639
MjSTAT          FAVMKVTRGHLRPQWNDGSIMGFVGRRQAEEMLKNSKSGTFLLRFSDSELGGVTIAWMYE 640
FcSTAT          FAVMKVTREHLRQQWNDGSIMGFVGRRQAEEMLKNSKSGTFLLRFSDSELGGVTIAWMYE 639
BmSTAT          DGVMEVLKKHLKPHWNDGAILGFVNKQQAHDLLINKPDGTFFW---DSEIGGITIAWKFD 631
HsSTAT          DGVMEVLKKHHKPHWNDG---GFVNKQQAHDLLINKPDGTFLLRFSDSEIGGITIAWKFD 631
                 .**:* : * :  ****   ***.:: *.::* *. .***:    ***:**:**** ::
```

```
MrSTAT          DTTK-GDQRDVFMLQPFTSKAFAIRPLADVIADLKYLLYLYPNVPKEQAFGKYYTPMGG- 713
SpSTAT          DTSKAGDQRDVFMLQPFTSKSFAIRPLADVIADLKYLLYLYPNIPKEQVFGKYYTPIG-- 698
PmSTAT          DTTK-ACQRDVFMLQPFTSKAFAIRPLADVIADLNYLLYLYPNVPKDQAFGKYYTPLGE- 697
LvSTAT          DTTK-ACQRDVFMLQPFTSKAFAIRPLADVIADLNYLLYLYPNVPKDQAFGKYYTPLGE- 697
MjSTAT          DTTK-ACQRDVFMLQPFTSKAFAIRPLADVIADLNYLLYLYPNVPKDQAFGKYYTPQGE- 698
FcSTAT          DTTK-ACQRDVFMLQPFTSKAFAIRPLADVIADLNYLLYLYPNVPKDQAFGKYYTPLGE- 697
BmSTAT          SQER-----MFWNLMPFTTRDFSIRSLADRLGDLSYLIYVFPDRPKDEVYSKYYTPVPCE 686
HsSTAT          SPER-----NLWNLKPFTTRDFSIRSLADRLGDLSYLIYVFPDRPKDEVFSKYYTPV--- 683
                .  :      .: * ***:: *:**.*** :.**.**:*::*: **::.:.*****

MrSTAT          -EQPTNN-GYVKPHLITHVPGWSVAGGSMDSYPNTPQPMYPMHDSNMGDPPSVSSNPSDS 771
SpSTAT          -EHPTNN-GYVKPHLITHVPGWPGRG-STESYPNTPQPMYPMHDTSLGDPPSVSSNPSDC 755
PmSTAT          -QQPTTNNGYVKPQLKTHVPGWSGDP--MDSYPNTPQTMY----GVMGGPPSVSSNPSDC 750
LvSTAT          -QQPTTNNGYVKPQLKTHVPGWSGDP--MDSYPNTPQTMY----GGMGGPPSVSSNPSDC 750
MjSTAT          -QQPTTNNGYVKPQLKTHVPGWSGDP--MDSYPNTPQTIY----GGMGGPPSVSSNPSDC 751
FcSTAT          -QQPTTNNGYVKPQLKTHVPGWSGDP--MDSYPNTPQTMY----GGMGGPPSVSSNPSDC 750
BmSTAT          PATAKAVDGYVKPQIKQVVPEFVSAS--ADSAGGSATYMDQAPSPAVCPQPHYNMYPQNP 744
HsSTAT          --LAKAVDGYVKPQIKQVVPEFVNAS--ADAGGSSATYMDQAPSPAVCPQAPYNMYPQNP 739
                   ..   *****::   ** :       ::  .:.  :       :   .  .  *.:

MrSTAT          VSTMPPYNDTDYPDILENL-PDTDFTDINLDFLQTNFMKPQ---------- 811
SpSTAT          VSTMPPYNDTDYPDILENL-PDPDFTNFSLDFLHTNFMKQQ---------- 795
PmSTAT          VSTVPPYNENEYDQFLENL-SEADFPDMNFDFLQTNFMKPQ---------- 790
LvSTAT          VSTVPTYNECEYDQFLENL-SEADFPDMNFDFLQTNFMKPQ---------- 790
MjSTAT          VSTVPTYNECEYDPWRT-C-SEADFPDMNFDFSSDQLHEASVKFWKYVHRHL 801
FcSTAT          VSTDQK--------------PTLDSPLFDAANVLSDFS------------- 774
BmSTAT          DPVLDNDGDFDLDDTIDVARRVEELLGRPMDS--QWIPHPQS---------- 784
HsSTAT          DHVLDQDGEFDLDETMDVARHVEELLRRPMDSLDSRLSPPAGLFTSARGSLS 791
                  .                   :           :
```

Figure 6.3. Multiple sequence alignments of *M. rosenbergii* STAT with seven other homologous STAT amino acid sequences. STAT of *S. paramamosain* (AHH2935.1), *P. monodon* (ACA79939.1), *L. vannamei* (ADQ43368.1), *M. japonicas* (BAI 49681.1), *F. chinensis* (ACH 70130.1), *B. mori* (ACR61178.1) and *H. sapiens* (CAD19637.1) are shown. Asterisk marks indicate identical amino acids and numbers to the right indicate the amino acid position of catalase in the corresponding species. Conserved substitutions are indicated by (:) and semi-conserved substitutions are indicated by (.). Deletions are indicated by dashes. GenBank accession numbers for the amino acid sequences of STAT given in the parentheses.

6.4.3 Phylogenetic Tree

A phylogenetic tree was built to investigate the evolutionary relationship between MrSTAT and other homologous forms such as crustaceans, insects, fish and mammals (Figure 6.4). Bootstrap sampling was reiterated 10,000 times. The maximum likelihood tree showed that each group formed a separate clad. The MrSTAT clustered together with the other members of the crustacean STAT protein, and formed a sister group with the insect STAT protein. The fish and mammals groups were not closely related to the MrSTAT, and thus formed the out-group.

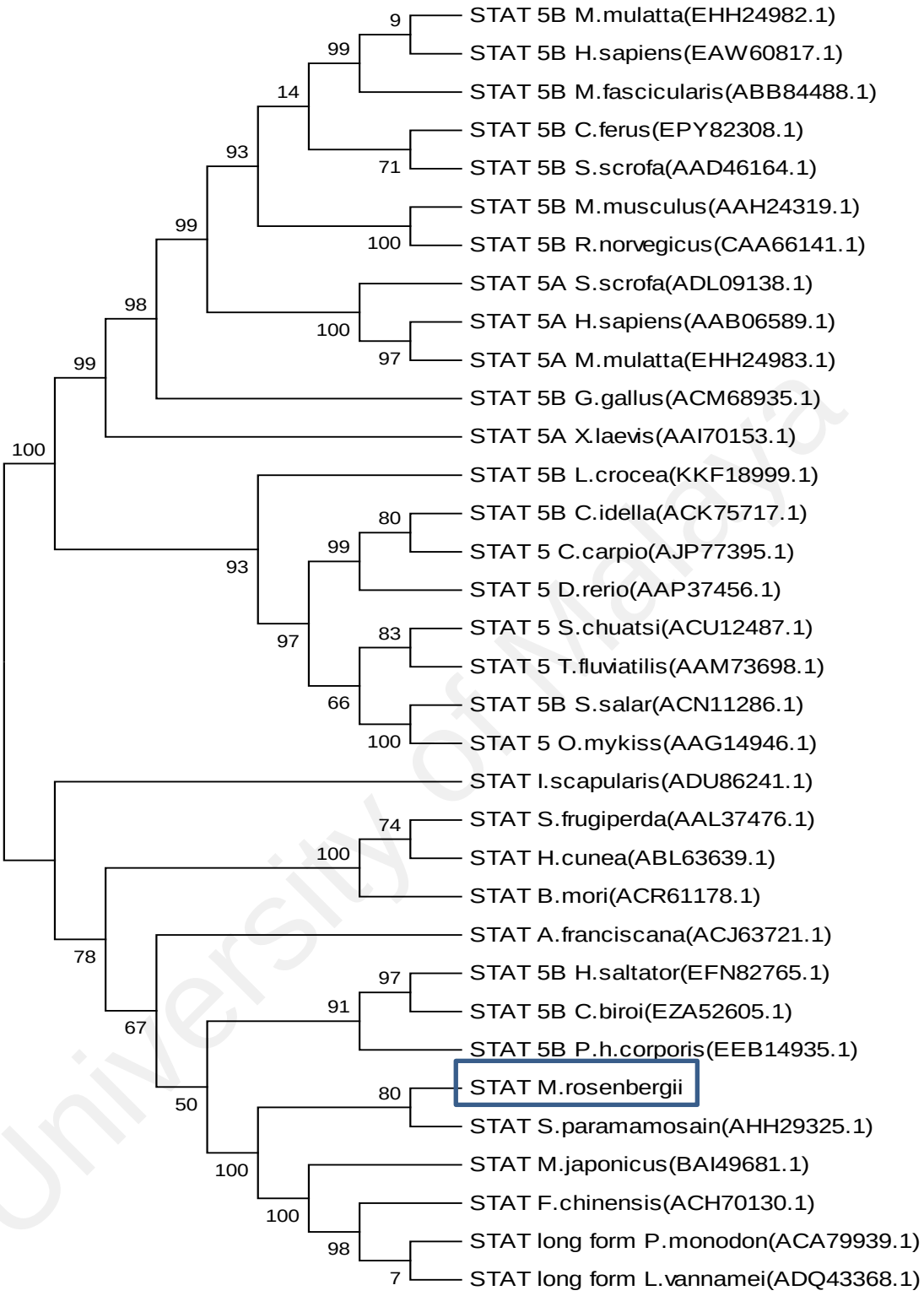


Figure 6.4. Phylogenetic tree of MrSTAT with other homologous was re-constructed by the Maximum Likelihood Method. The percentage values of each internal branch show the bootstrap analysis of 10,000 replications.

6.4.4 Structural Analysis of MrSTAT

A secondary structural analysis of MrSTAT showed that the protein contains 45.13% α-helical region (366 amino acid residues), 46.73% coils (379 amino acid residues) and 8.14% β-sheet region (66 amino acid residues) (Figure 6.5). The I-TASSER program predicted five different models for the MrSTAT protein. The best 3D structure for MrSTAT was obtained from the I-TASSER online server (Figure 6.6), based on an analysis of the C-score (-1.06) and Ramachandran plot program (data not shown). Furthermore, an analysis of the RCSB database showed MrSTAT to have the greatest structural identity with unphosphorylated STAT1 (PDB id 1yvl) from *H. sapiens*, experimentally obtained using the X-Ray diffraction method. The RNA structure of MrSTAT was predicted using MFE. The results are presented in Figure 6.7. The MFE of the predicted RNA structure of MrSTAT was -753.40 kcal/mol. The prediction also shows that the RNA is mostly paired, with very few nucleotides left unpaired.

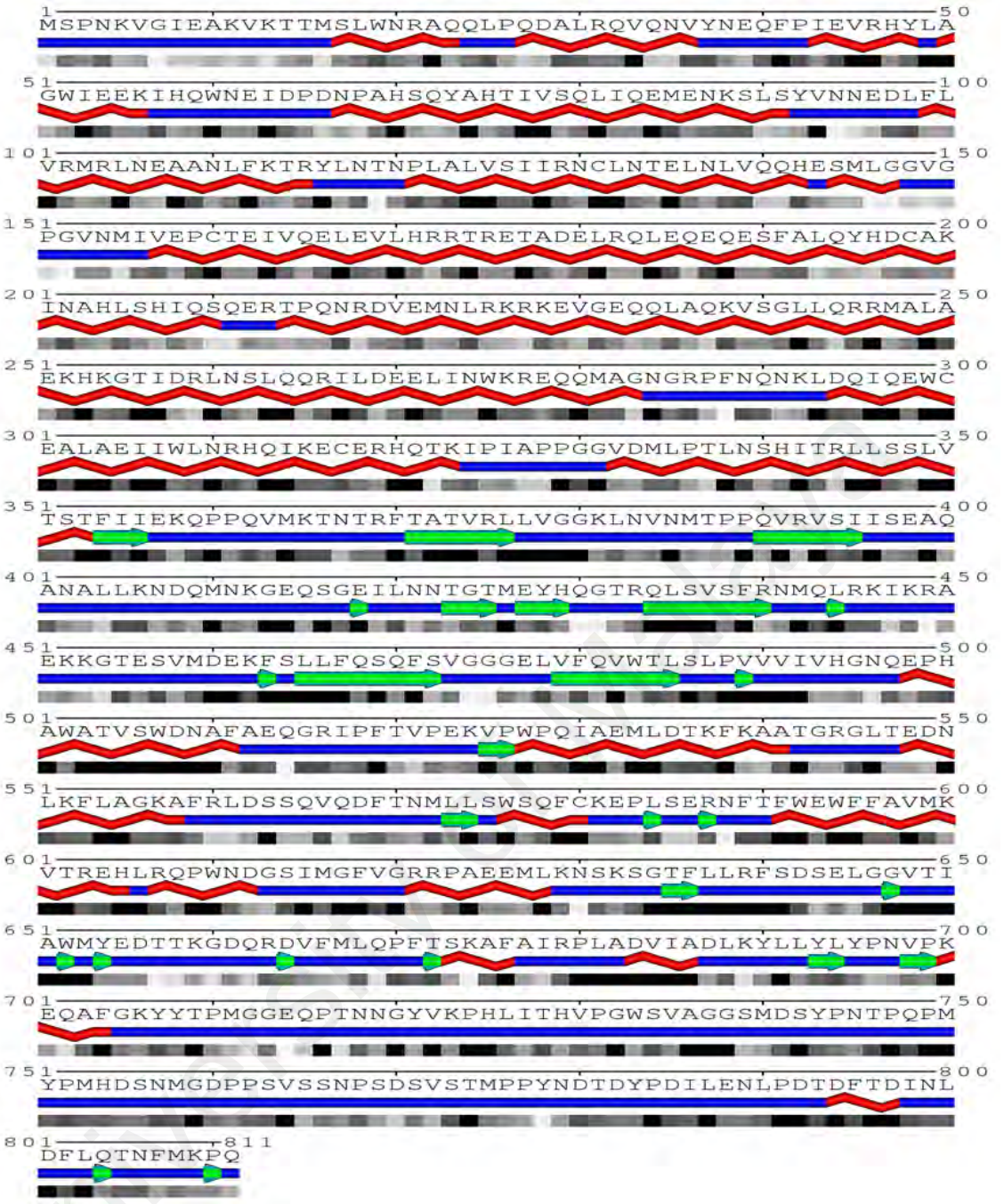


Figure 6.5. Two dimensional structure of MrSTAT predicted and analysed on POLYVIEW protein structure visualization online server. The number of amino acids starts at the left and terminates at the right margin, respectively. The a-helices, antiparallel b-sheets and random coils are indicated by red, green and blue colour, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

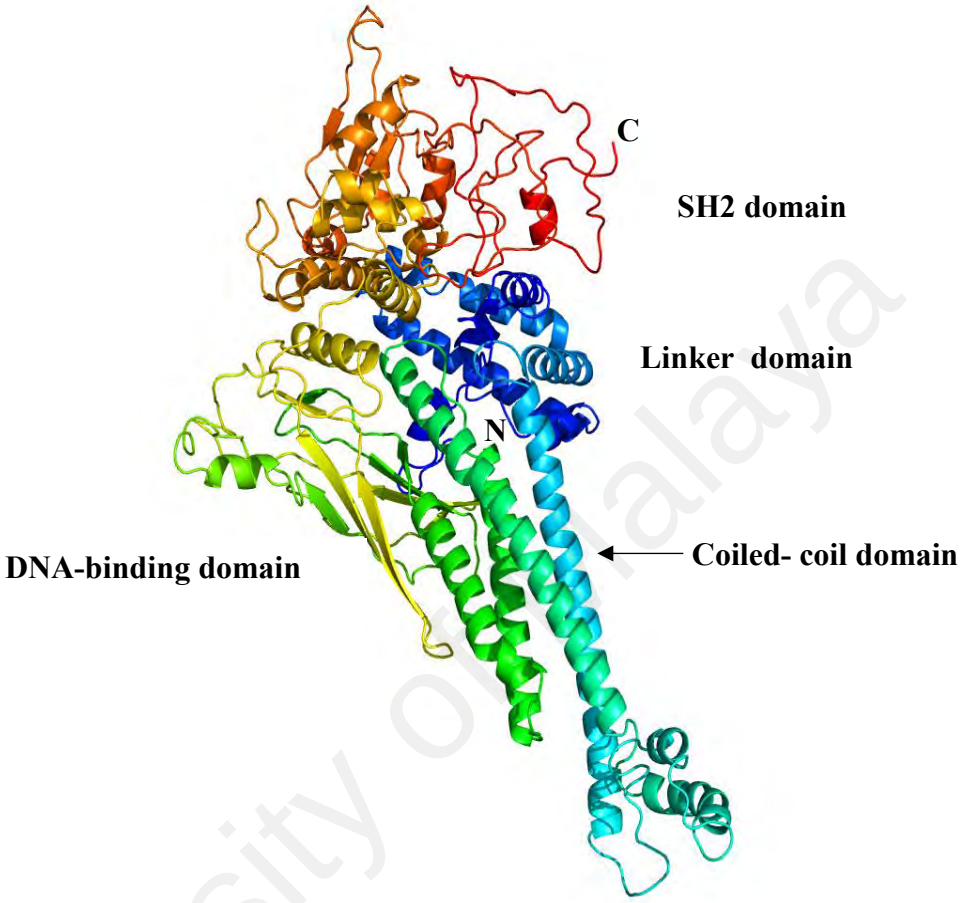


Figure 6.6. A complete three dimensional structure of MrSTAT predicted by I-Tasser prediction algorithm program based on the similarities with other homologous STAT available in protein database and analysed on PyMol viewer.

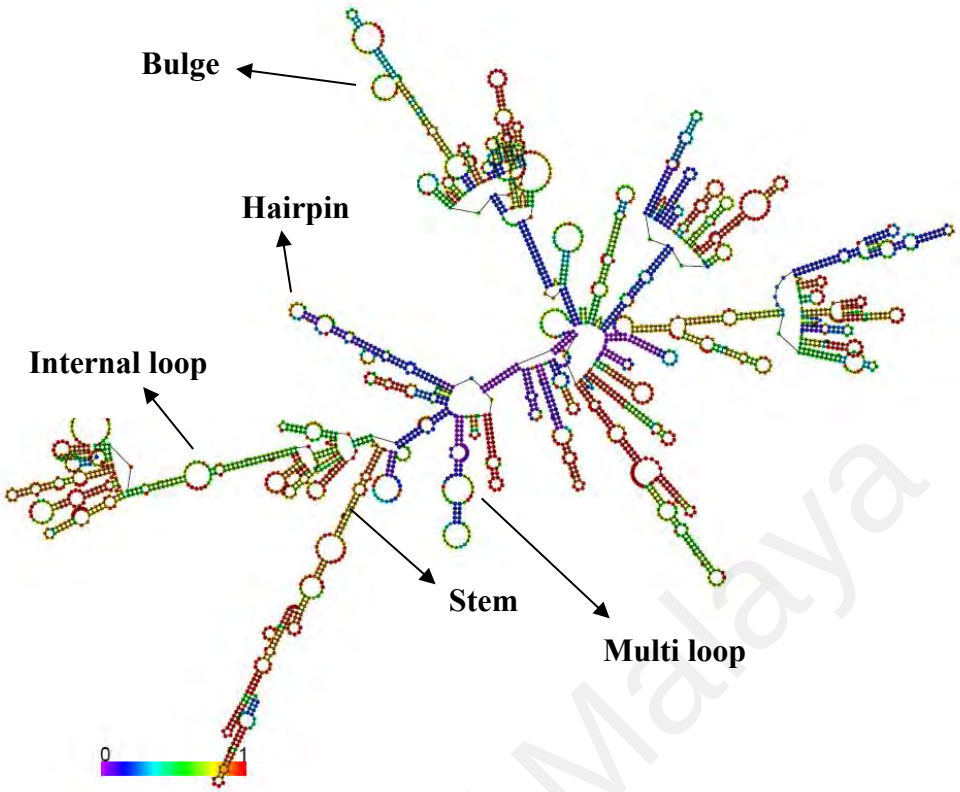


Figure 6.7. The predicted structure of MrSTAT RNA folds with MFE using RNA fold server. The scale bar (colour map) from violet to red denotes the probability of nucleotides being unpaired.

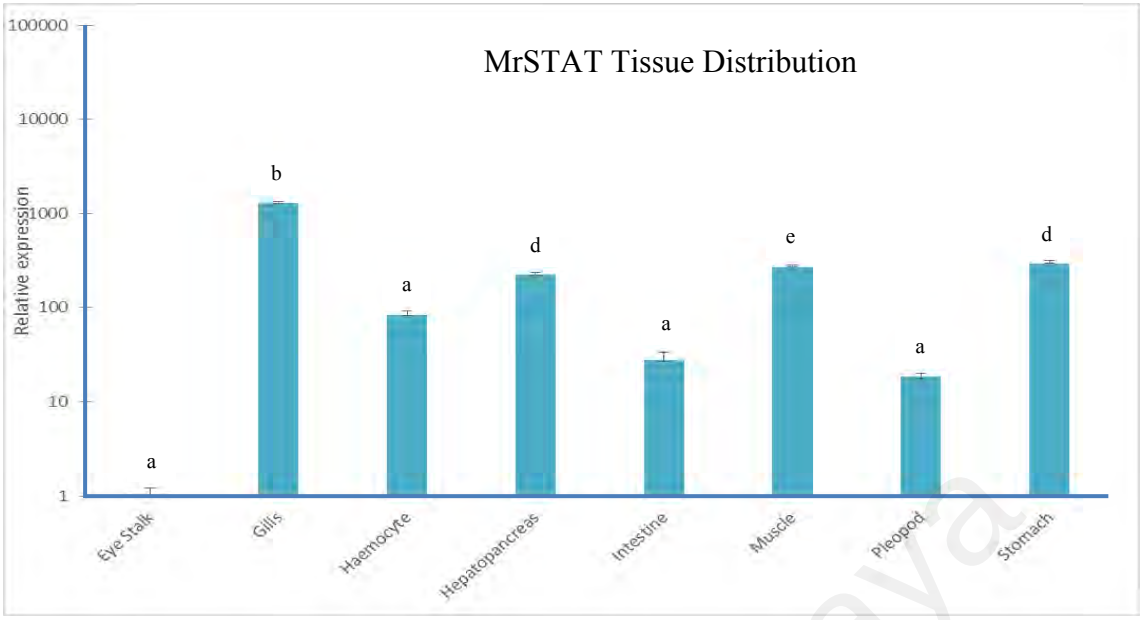
6.4.5 Tissue-Specific mRNA Expression and Transcriptional Changes of MrSTAT upon WSSV and *V. parahaemolyticus* Infection

A tissue distribution analysis of MrSTAT was carried out by real-time PCR using cDNA (as a template) synthesized from the extracted total RNA of the eye stalk, gills, pleopod, stomach, hepatopancreas, muscle, intestine and haemocyte, together with gene-specific primers and other reaction mixtures. The elongation factor 1-alpha (ELF-1) was used as the housekeeping gene. The tissue-specific mRNA transcription was measured by comparing the expression quantity found in the tissues taken for analysis with that from

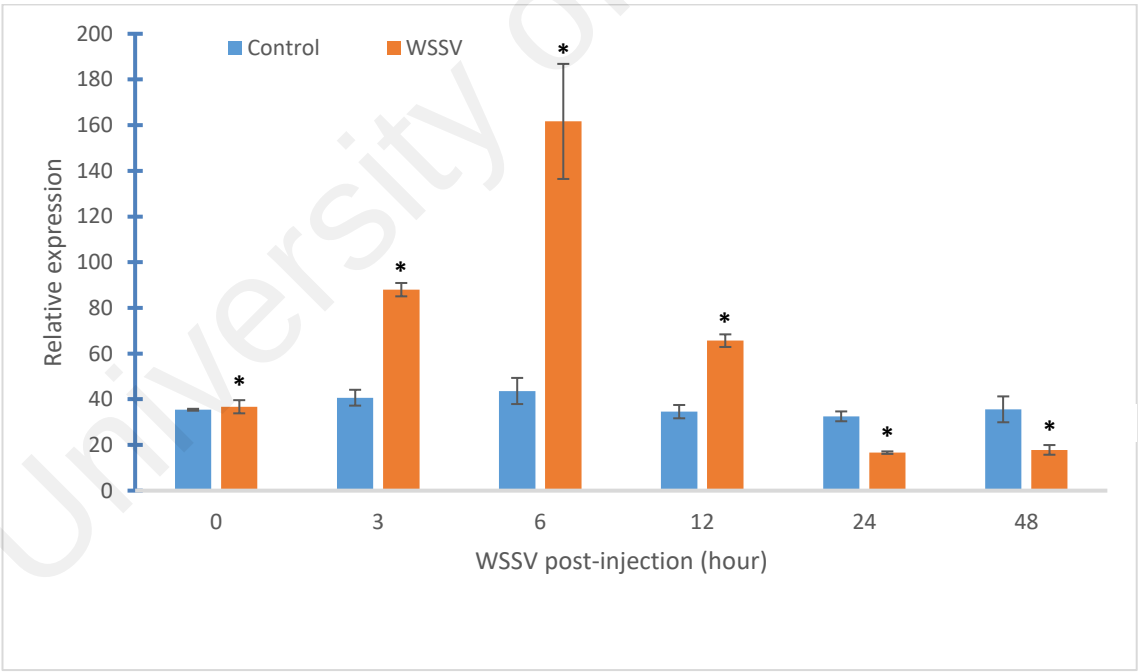
the eye stalk (Figure 6.8A). The gene expression results showed that the MrSTAT transcripts were expressed in all the tissues taken for examination. The highest expression was observed in the gills, followed by the stomach, muscle, hepatopancreas, haemocytes, intestine, pleopod and eye stalk, and these results were significant at the level of $P < 0.05$.

We then studied MrSTAT mRNA expression in the hepatopancreas of *M. rosenbergii* after WSSV and bacterial *V. parahaemolyticus* infection. We chose the hepatopancreas tissue because our previous work on transcriptome profiling in the hepatopancreas after infection with WSSV and *V. parahaemolyticus* had showed significant up-regulation of MrSTAT at the 12-hour time point. The transcriptional expression of MrSTAT after WSSV stimulation in the hepatopancreas is shown in Figure 6.8B. After the injection with WSSV, the MrSTAT expression levels showed a significant ($p < 0.05$) increase from 0 to 6 hours post-injection (p.i), with a peak value of 3.7 times greater than the control group at 6h. The levels of MrSTAT expression then gradually decreased at 12h p.i. In contrast, these levels did not show any significant expression in the control groups. Meanwhile, in the *V. parahaemolyticus* challenged prawn, the transcription levels rose steadily from 3h p.i., reaching their highest level at 24h p.i and decreasing thereafter at 48h p.i. (Figure 6.8C). Again, the respective controls showed no significant fluctuation in MrSTAT activity throughout the 48 hour test period.

A



B



C

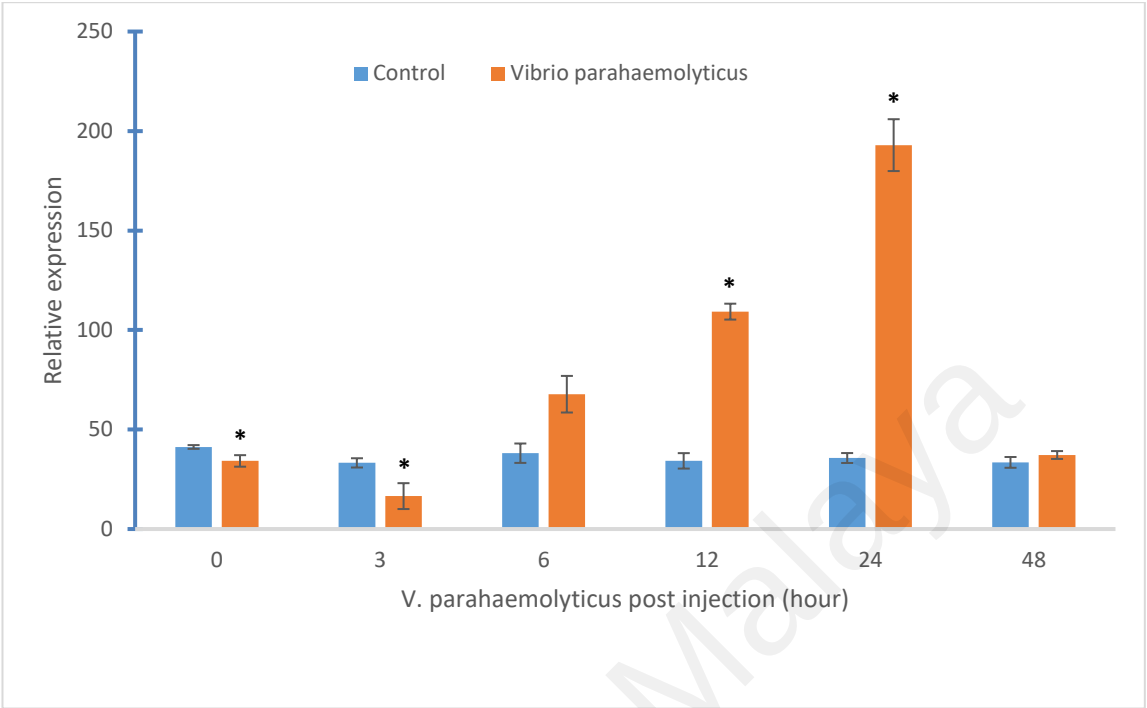


Figure 6.8. Relative gene expression analysis of MrSTAT by qRT-PCR. A: Tissue-specific mRNA expression of MrSTAT. Data is expressed as a ratio to MrSTAT mRNA expression in eyestalk. The different alphabets are statistically significant at P< 0.05 level by one-way ANOVA with post-hoc Tukey HSD Test. B-C: the time course of MrSTAT MRNA expression in hepatopancreas at 0, 3, 6, 12, 24, and 48 h post-injection with WSSV and *V. parahaemolyticus* along with their controls respectively. The significant difference (P <0.05) of MrSTAT expression between the challenged and the control group were indicated with asterisks.

6.5 Discussion

Total world fishery production has declined significantly in recent years, while at the same time human consumption of aquatic organisms has increased sharply (Garcia and Rosenberg, 2010, Morato *et al*., 2006). The reduction in fish catches has been partly compensated by the fast growth of the aquaculture industry, of which the commercial culture of shrimp is an important part in tropical and subtropical countries (Bostock *et al*., 2010, Primavera, 1997). However, the spread of viral and bacterial diseases has significantly limited the growth of this promising industry (Aguirre Guzman and Ascencio Valle, 2000, Fegan, 1992). Unfortunately, no effective method of controlling these pathogens has been found so far, hence the importance of gaining a holistic understanding of shrimp immune responses, which could help in the development of a strategy to control these diseases in crustacean aquaculture.

Previous research has suggested that the JAK/STAT pathway may be involved in regulating innate immunity. In insects such as mosquitoes and Drosophila, the JAK/STAT pathway is involved in antiviral activity, although no interferon has been detected (Agaisse and Perrimon, 2004, Souza-Neto *et al*., 2009). In the *Culex* mosquito, this pathway induces the expression of Vago, which is regarded as an antiviral factor, to restrict the West Nile virus infection (Paradkar *et al.*, 2012). In addition to its antiviral activity, this pathway has also been found to be activated in the mosquito *Anopheles gambiae* in response to bacterial challenge. The discovery of a STAT homolog in *P. monodon* and *F. chinensis* suggests that a putative JAK/STAT pathway might exist in shrimps (Sun *et al*., 2011, Chen *et al*., 2008).

As the JAK/STAT pathway in *M. rosenbergii* is not reported yet, we reported a full length cDNA of a novel MrSTAT obtained from a constructed trancriptome library

of *M. rosenbergii*, sequenced by Illumina's HiSeq^{TM}2000 sequencing technology. MrSTAT has an ORF of 2436 bp that coded for 811 aa (Figure 6.1), which showed a high degree of similarity to known STAT genes in shrimps and other predicted proteins (Figure 6.2 and Table 6.2). MrSTAT showed an 88% homology with the mud crab *S. paramamosain*, between 50.0% and 63.0% identity with insect STAT, and approximately 48.0% identity with fish STAT, and human HsSTAT5a and HsSTAT5b, respectively. Interestingly, MrSTAT has the longest amino acid sequence of all the shrimp species examined. The four domains found in MrSTAT are highly conserved, and are also similar to the domains found in the STAT from other species. The existence of these seemingly functional residues in MrSTAT is further evidence that shrimp STAT may function as a regulation mechanism as it does in vertebrates.

All known STATs are reported to be classified into two groups: the ancient STAT family, comprising all the insect STATs (AgSTAT, SfSTAT, DmSTAT) and vertebrate STAT5a, STAT5b and STAT6; and a second, newer STAT family derived from the ancient STAT through duplication events (Copeland *et al*., 1995). Our phylogenetic tree derived for MrSTAT (Figure 6.4) suggests that this protein is most closely related to the ancient STAT family, and specifically SpSTAT. We therefore concluded that shrimp STAT belongs to the ancient STAT family.

The predicted 3D model of MrSTAT was modelled based on human unphosphorylated STAT1 (PDB id 1yvl) – the best template available, in the absence of data on structural models of STAT from crustaceans. Like human unphosphorylated STAT1 structure, MrSTAT contains four long helices (α1–4), which we predicted to be the coiled coil domain. The DNA binding domain in MrSTAT also contains several β-sheets that are folded in a similar way to those found in the DNA binding domains of the transcription factors nuclear factor κB (NF-κB) or p53. The SH2 domain of MrSTAT also

consists of an anti-parallel β sheet flanked by two α-helices. The high degree of similarity between the structures of MrSTAT and of the human unphosphorylated STAT1 is further evidence that these structures have been evolutionarily conserved in both vertebrates and invertebrates.

A stable RNA is a prerequisite for regulating the levels of cellular mRNA transcripts. Based on the RNA prediction program, the MFE value of the predicted RNA structure for MrSTAT is -753.40 kcal/mol. The MFE value of mRNA depends on AU/GC base pairing, with paired base pairs receiving a negative value and unpaired base pairs a positive value. The prediction also indicated that the vast majority of the nucleotides are paired in the structure, with very few nucleotides left unpaired, thus making MrSTAT mRNA highly stable.

A tissue distribution analysis of MrSTAT showed that this gene was expressed in all the tissues examined, with the highest levels of expression observed in the gills and stomach. In *F. chinensis* and *M. japonicas*, this gene was also found to be expressed in all tissues, with the stomach showing the highest level of expression in *F. chinensis* and the muscle in *M. japonicas* (Okugawa *et al*., 2013, Sun *et al*., 2011). STAT has also been found to be expressed in various tissues of *S. chuatsi*, except in muscle, which suggests that STATs are widely distributed in tissues in both invertebrates and vertebrates (Guo *et al*., 2009). However, the precise reasons for the varied levels of expression in different types of tissue in different species are not known.

A time-course expression analysis by qPCR clearly showed that MrSTAT gene expression was up-regulated in the hepatopancreas post-induction by WSSV and *V. parahaemolyticus*. These results clearly support our previous transcriptome data, which also showed a higher expression of MrSTAT in response to WSSV and *V.*

parahaemolyticus. An up-regulation of STAT was moreover similarly reported in *F. chinensis* against WSSV and *V. anguillarum* (Sun *et al*., 2011*)*. In addition, an enhanced expression of STAT was observed following stimulation with peptidoglycan and polycytidylic acid in *M. japonicas* (Okugawa *et al*., 2013).

These results all point to an important role for STAT in shrimp immunity against viral and bacterial infection. However, further investigation is necessary to understand fully the role of STAT in shrimp defence mechanisms against various pathogens.

6.6 Conclusion

A full-length cDNA of MrSTAT was obtained from *M. rosenbergii* and further characterised using various bioinformatics tools. A phylogenetic analysis showed MrSTAT to have a close relationship with an ancient class of the STAT family. A quantitative real-time PCR analysis of MrSTAT showed enhanced expression of the gene across a wide range of tissues and at various time points during WSSV and *V. parahaemolyticus* challenge, which indicates that MrSTAT plays a role in the antiviral and antibacterial immunity of *M. rosenbergii*.

Chapter 7

CONCLUSION AND RECOMMENDATION

Total world fisheries production has decreased, while human consumption of aquatic organisms has increased (FAO, 2006). This reduction in capture fisheries has been partly compensated for by the fast growth of the aquaculture industry, within which the commercial culture of penaeid shrimp and freshwater prawn is an important sector in tropical and subtropical countries (FAO, 2006). Of all the commercial crustacean species, the giant freshwater prawn, *Macrobrachium rosenbergii*, has emerged one of the most popular, which is cultivated on a large-scale in different parts of the world. However, the production of this species has been in decline with the rise in both viral and bacterial diseases affecting it.

The two major pathogens affecting the *M. rosenbergii* aquaculture industry are the White spot syndrome virus (WSSV) and *V. parahaemolyticus*. When infected with WSSV, *M. rosenbergii* develops white spot syndrome which leads to a high cumulative mortality rate of up to 100 % within 3–10 days. *V. parahaemolyticus*, meanwhile, causes Vibriosis, which is similarly a fatal disease in *M. rosenbergii*. The prevention and control of these diseases have therefore become top priorities to secure the long-term production and economic sustainability of this industry. To achieve this, understanding the shrimp's innate immune system and its molecular mechanism is crucial.

To this end, we subjected *M. rosenbergii* hepatopancreases challenged by WSSV and *V. parahaemolyticus* to a transcriptomic analysis using the Illumina HiSeq^{TM} 2000 platform, a *de novo* assembly using Trinity software, and a bioinformatics approach.

A comparison of the transcriptome analysis of the WSS-infected and control specimens revealed that 8,443 host genes were significantly up-regulated, whereas 5,973 genes were significantly down-regulated. These differentially expressed genes fell into 15 categories of animal immune functions, including antiviral proteins, antimicrobial proteins, proteases, protease inhibitors, signal transduction, blood clotting system, transcriptional control, cell death, cell adhesion, heat shock proteins, RNA silencing, oxidative stress, pathogen recognition immune receptors, prophenoloxidase system and other immune genes.

An analysis of the differential gene expression between *V. parahaemolyticus*-infected and control specimens showed a total of 14,569 unigenes aberrantly expressed, with 11,446 unigenes significantly up-regulated and 3,103 unigenes significantly down-regulated. These immune genes fell into 11 categories of functions, including antimicrobial proteins, proteases and proteinases, signal transduction, blood clotting system, cell death, cytoskeletal, heat shock proteins, oxidative stress, pathogen recognition immune receptors, prophenoloxidase system and other immune genes.

We then undertook a comparative analysis to compare the expression of 156 immune-related unigenes obtained from the *M. rosenbergii* transcriptome data when challenged with *V. parahaemolyticus* and WSSV respectively. This produced rather different results between the two types of infection: there were 40 unigenes up-regulated and 5 unigenes down-regulated in response to *V. parahaemolyticus* infection alone, whereas during WSSV infection 19 unigenes were found to be up-regulated and 33 unigenes down-regulated. Interestingly, we also found 79 unigenes which were up-regulated and 7 unigenes which were down-regulated in response to both pathogens. Several of these unigenes were key components of five signalling pathways: the Toll pathway immune deficiency (IMD), Wnt pathway, IMD pathway, Ubiquitin proteasome

pathway, and Mitogen-activated protein kinase (MAPK) pathway, as well as Janus kinase (JAK)-signal transducers and activators of transcription (STAT).

The results from our transcriptome data obtained in response to WSSV and *V. parahaemolyticus* showed the STAT gene to be significantly up-regulated at the 12th hour post infection. This led us to isolate a full-length cDNA of MrSTAT from the above transcriptomic database, which we then further characterised using various bioinformatics tools. A phylogenetic analysis showed a close relationship between MrSTAT and an ancient class of the STAT family, with SpSTAT being the closest relative. A 3D model structure prediction of MrSTAT showed a high similarity to the structure of human unphosphorylated STAT1. The predicted RNA structure for MrSTAT is highly stable, as most of the nucleotides are paired in the structure with very few left unpaired. A tissue distribution analysis of MrSTAT by quantitative real-time PCR showed expression in all the examined tissues. The enhanced expression profile in the hepatopancreas observed at various time points during WSSV and *V. parahaemolyticus* challenge points to the involvement of MrSTAT in the antiviral and antibacterial immunity of *M. rosenbergii*.

From this current study, we were able to successfully annotate only a small number of unigenes (approximately 35% of the unigenes) by using the BLASTX search. Further work might therefore focus on characterising and annotating the function of those unannotated unigenes, to obtain a better understanding of the response in *M. rosenbergii* to the two pathogens in question. In addition, sequencing the whole genome of *M. rosenbergii* would be useful to increase the number of annotated unigenes for this species.

The data generated from this study provides a solid foundation for further future investigation into the role of the target immune genes involved in the viral and bacterial defence mechanisms of *M. rosenbergii*. Further work on the *in vivo* effects of

differentially expressed unigenes in *M. rosenbergii*, using the SiRNA and gene silencing approaches, could also help to bring out further the importance of these immune genes in the defence mechanisms of this species. *In vivo* biological activity-related experiments on individual target genes in *M. rosenbergii* could also aid in the characterisation and better understanding of many of the results obtained from our study.

The current study used hepatopancreasas as the sole targeted tissue for transcriptomics sequencing. In future, it would interesting to include other tissues such as haemocytes, gills and muscle for sequencing so as to gain a more comprehensive understanding of *M. rosenbergii* immune responses to the two pathogens. With regard to the role of the JAK/STAT pathway in the *M. rosenbergii* defence mechanism, it would be worthwhile studying other components of this pathway, such as its receptors, the Januse Kinase protein and the pathway inhibitors, in order to improve our understanding of its role in immunity.

REFERENCES

Adams, A. (1991). Response of penaeid shrimp to exposure to Vibrio species. *Fish & Shellfish Immunology,* 1(1), 59-70.

Agaisse, H., & Perrimon, N. (2004). The roles of JAK/STAT signalling in Drosophila immune responses. *Immunological Reviews,* 198(1), 72-82.

Aguirre Guzman, G., & Ascencio Valle, F. (2000). Infectious disease in shrimp species with aquaculture potential. *Recent Research Developments in Microbiology*, 333-348.

Aguirre-Guzman, G., Sanchez-Martinez, J. G., Campa-Cordova, A. I., Luna-González, A., & Ascencio, F. (2009). Penaeid shrimp immune system. *The Thai Journal of Veterinary Medicine,* 39(3), 205-215.

Anderson, I., Shamsudin, M., Shariff, M., & Nash, G. (1988). Bacterial septicemia in juvenile tiger shrimp, Penaeus monodon, cultured in Malaysian brackishwater ponds. *Asian Fisheries Science Journal,* 2, 93-108.

Andrade, A., Silva, P., Pereira, A., De Sousa, L., Ferreira, P., Gazzinelli, R.T, Kroon, E.G., Catherine, R.,& Bonjardim, C. (2004). The vaccinia virus-stimulated mitogen-activated protein kinase (MAPK) pathway is required for virus multiplication. *Biochemical Journal,* 381, 437-446.

Antony, S. P., Philip, R., Joseph, V., & Singh, I. B. (2011). Anti-lipopolysaccharide factor and crustin-III, the anti-white spot virus peptides in Penaeus monodon: Control of viral infection by up-regulation. *Aquaculture,* 319(1), 11-17.

Arockiaraj, J., Easwvaran, S., Vanaraja, P., Singh, A., Othman, R. Y., & Bhassu, S. (2012a). Immunological role of thiol-dependent peroxiredoxin gene in Macrobrachium rosenbergii. *Fish & Shellfish Immunology,* 33(1), 121-129.

Arockiaraj, J., Easwvaran, S., Vanaraja, P., Singh, A., Othman, R. Y., & Bhassu, S. (2012b). Molecular cloning, characterization and gene expression of an antioxidant enzyme catalase (Mr Cat) from Macrobrachium rosenbergii. *Fish & Shellfish Immunology,* 32(5), 670-682.

Arockiaraj, J., Easwvaran, S., Vanaraja, P., Singh, A., Othman, R. Y., & Bhassu, S. (2012c). Effect of infectious hypodermal and haematopoietic necrosis virus (IHHNV) infection on caspase 3c expression and activity in freshwater prawn Macrobrachium rosenbergii. *Fish & Shellfish Immunology*, 32(1), 161-169.

Arockiaraj, J., Vanaraja, P., Easwvaran, S., Singh, A., Alinejaid, T., Othman, R. Y., & Bhassu, S. (2011a). Gene profiling and characterization of arginine kinase-1 (MrAK-1) from freshwater giant prawn (Macrobrachium rosenbergii). *Fish & Shellfish Immunology,* 31(1), 81-89.

Arockiaraj, J., Vanaraja, P., Easwvaran, S., Singh, A., Othman, R. Y., & Bhassu, S.

(2011b). Bioinformatic characterization and gene expression pattern of apoptosis inhibitor from Macrobrachium rosenbergii challenged with infectious hypodermal and hematopoietic necrosis virus. *Fish & Shellfish Immunology,* 31(6), 1259-1267.

Astrofsky, K., Roux, M., Klimpel, K., Fox, J., & Dhar, A. (2002). Isolation of differentially expressed genes from white spot virus (WSV) infected Pacific blue shrimp (Penaeus stylirostris). *Archives of Virology,* 147(9), 1799-1812.

Banks, L., Pim, D., & Thomas, M. (2003). Viruses and the 26S proteasome: hacking into destruction. *Trends in Biochemical Sciences,* 28(8), 452-459.

Barillas-Mury, C., Han, Y. S., Seeley, D., & Kafatos, F. C. (1999). Anopheles gambiae Ag-STAT, a new insect member of the STAT family, is activated in response to bacterial infection. *The EMBO Journal,* 18(4), 959-967.

Barki, A., Karplus, I., & Goren, M. (1991). Morphotype related dominance hierarchies in males of Macrobrachium rosenbergii (Crustacea, Palaemonidae). *Behaviour*, 117(3), 145-160.

Barki, A., Karplus, I., & Goren, M. (1992). Effects of size and morphotype on dominance hierarchies and resource competition in the freshwater prawn Macrobrachium rosenbergii. *Animal Behaviour,* 44, 547-555.

Blum, T., Briesemeister, S., & Kohlbacher, O. (2009). MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics, 10*(1), 274.

Bostock, J., McAndrew, B., Richards, R., Jauncey, K., Telfer, T., Lorenzen, K., Little, D., Ross, L., Handisyde, N., Gatward, I., & Corner, R. (2010). Aquaculture: global status and trends. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 365(1554), 2897-2912.

Chen, L.L., Wang, H.C., Huang, C.J., Peng, S.E., Chen, Y.G., Lin, S.J., Chen, W.Y., Dai, C.F., Yu, H.T., & Wang, C.H. (2002). Transcriptional analysis of the DNA polymerase gene of shrimp white spot syndrome virus. *Virology*, 301(1), 136-147.

Chen, M., & Gerlier, D. (2006). Viral hijacking of cellular Ubiquitination pathways as an anti-innate immunity strategy. *Viral Immunology,* 19(3), 349-362.

Chen, W. Y., Ho, K. C., Leu, J. H., Liu, K. F., Wang, H. C., Kou, G. H., & Lo, C.F. (2008). WSSV infection activates STAT in shrimp. *Developmental & Comparative Immunology,* 32(10), 1142-1150.

Chen, X., Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., Ma, N., Li, M., & Yang, Q. (2013). Transcriptome analysis of Litopenaeus vannamei in response to white spot syndrome virus infection. *PloS One*, 8(8).

Chou, H., Huang, C., Lo, C., & Kou, G. (1998). Studies on transmission of white spot syndrome associated baculovirus (WSBV) in Penaeus monodon and Penaeus japonicus via waterborne contact and oral ingestion. *Aquaculture,* 164(1), 263-

276.

Chou, H.Y., Huang, C.Y., Wang, C.H., Chiang, H.C., & Lo, C.F. (1995). Pathogenicity of a baculovirus infection causing white spot syndrome in cultured penaeid shrimp in Taiwan. *Diseases of Aquatic Organisms,* 23(3), 165-173.

Chowdhury, R., Angell, C., & Bhattacharjee, H. (1993). A manual for operating a small-scale recirculation freshwater prawn hatchery*: Bay of Bengal Programme Report,* 13, 1-37.

Cohen, D., Raanan, Z., & Brody, T. (1981). Population profile development and morphotypic differentiation in the giant freshwater prawn Macrobrachium rosenbergii (de Man). *Journal of the World Mariculture Society,* 12(2), 231-243.

Copeland, N. G., Gilbert, D. J., Schindler, C., Zhong, Z., Wen, Z., Darnell, J. E., Mui, A.L.F., Miyajima, A., Quelle, F.W., Ihle, J.N., & Jenkins, N.A. (1995). Distribution of the mammalian Stat gene family in mouse chromosomes. *Genomics,* 29(1), 225-228.

Cserzö, M., Wallin, E., Simon, I., von Heijne, G., & Elofsson, A. (1997). Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engineering*, 10(6), 673-676.

De Gregorio, E., Spellman, P. T., Tzou, P., Rubin, G. M., & Lemaitre, B. (2002). The Toll and Imd pathways are the major regulators of the immune response in Drosophila. *The EMBO Journal,* 21(11), 2568-2579.

DePaola, A., Kaysner, C. A., Bowers, J., & Cook, D. W. (2000). Environmental investigations of Vibrio parahaemolyticus in oysters after outbreaks in Washington, Texas, and New York (1997 and 1998). *Applied and Environmental Microbiology,* 66(11), 4649-4654.

Dhar, A. K., Bowers, R. M., Licon, K. S., Veazey, G., & Read, B. (2009). Validation of reference genes for quantitative measurement of immune gene expression in shrimp. *Molecular Immunology,* 46(8), 1688-1695.

Dhar, A., Dettori, A., Roux, M., Klimpel, K., & Read, B. (2003). Identification of differentially expressed genes in shrimp (Penaeus stylirostris) infected with White spot syndrome virus by cDNA microarrays. *Archives of Virology,* 148(12), 2381-2396.

Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C., Hoffmann, J. A., & Imler, J.L. (2005). The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila. *Nature Immunology*, 6(9), 946-953.

Durand, S., Lightner, D., Redman, R., & Bonami, J. R. (1997). Ultrastructure and morphogenesis of white spot syndrome baculovirus (WSSV). *Diseases of Aquatic Organisms,* 29(3), 205-211.

Escobedo-Bonilla, C. M., Alday-Sanz, V., Wille, M., Sorgeloos, P., Pensaert, M., &

Nauwynck, H. (2008). A review on the morphology, molecular characterization, morphogenesis and pathogenesis of white spot syndrome virus. *Journal of Fish Diseases,* 31(1), 1-18.

Fegan, D. F. (1992). Recent developments and issues in the penaeid shrimp hatchery industry. *Proceeding of the Special Session on Shrimp Farming. The World Aquaculture Society, Baton Rouge, Belgium*, 55-70.

Flegel, T. W. (2012). Historic emergence, impact and current status of shrimp pathogens in Asia. *Journal of Invertebrate Pathology,* 110(2), 166-173.

Flegel, T., & Alday-Sanz, V. (1998). The crisis in Asian shrimp aquaculture: current status and future needs. *Journal of Applied Ichthyology,* 14(3-4), 269-273.

Fujimura, T., Okamoto, H., & Council, I. P. F. (1970). *Notes on progress made in developing a mass culturing technique for Macrobrachium rosenbergii in Hawaii*: Indo-Pacific Fisheries Council, FAO Regional Office for Asia and the Far East.

Garcia, S. M., & Rosenberg, A. A. (2010). Food security and marine capture fisheries: characteristics, trends, drivers and future perspectives. *Philosophical Transactions of the Royal Society of London B: Biological Sciences,* 365(1554), 2869-2880.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., & Hornik, K. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology,* 5(10), R80.

Gross, P., Bartlett, T., Browdy, C., Chapman, R., & Warr, G. (2001). Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, Litopenaeus vannamei, and the Atlantic White Shrimp, Litopenaeus setiferus. *Developmental & Comparative Immunology,* 25(7), 565-577.

Guo, C.J., Zhang, Y.F., Yang, L.S., Yang, X.B., Wu, Y.Y., Liu, D., Chen, W.J., Weng, S.P., Yu, X.Q., & He, J.G. (2009). The JAK and STAT family members of the mandarin fish Siniperca chuatsi: molecular cloning, tissues distribution and immunobiological activity. *Fish & Shellfish Immunology*, 27(2), 349-359.

Guo, H., Ye, C.X., Wang, A.L., Xian, J.A., Liao, S.A., Miao, Y.T., & Zhang, S.P. (2013). Transcriptome analysis of the Pacific white shrimp Litopenaeus vannamei exposed to nitrite by RNA-seq. *Fish & Shellfish Immunology*, 35(6), 2008-2016.

He, J. G., Chen, Y. G., Deng, M., Yao, P., Zhou, H. M., Weng, S. P. (2003). *Natural host range and pathogenicity of white spot syndrome virus infection in shrimp and crab species in China.* Paper presented at the American Fisheries Society Symposium.

He, N., Liu, H., & Xu, X. (2004). Identification of genes involved in the response of haemocytes of Penaeus japonicus by suppression subtractive hybridization (SSH) following microbial challenge. *Fish & Shellfish Immunology,* 17(2), 121-128.

Honda, T., & Iida, T. (1993). The pathogenicity of Vibrio parahaemolyticus and the role of the thermostable direct haemolysin and related haemolysins. *Reviews in Medical Microbiology,* 4(2), 106-113.

Huang, C., Zhang, X., Lin, Q., & Xu, X. (2002). Characterization of a novel envelope protein (VP281) of shrimp white spot syndrome virus by mass spectrometry. *Journal of General Virology,* 83(10), 2385-2392.

Huang, C., Zhang, X., Lin, Q., Xu, X., Hu, Z., & Hew, C. L. (2002). Proteomic analysis of shrimp white spot syndrome viral proteins and characterization of a novel envelope protein VP466. *Molecular & Cellular Proteomics,* 1(3), 223-231.

Huang, R., Xie, Y., Zhang, J., & Shi, Z. (2005). A novel envelope protein involved in white spot syndrome virus infection. *Journal of General Virology,* 86(5), 1357-1361.

Huelsken, J., & Behrens, J. (2002). The Wnt signalling pathway. *Journal of Cell Science,* 115(21), 3977-3978.

Huelsken, J., & Birchmeier, W. (2001). New aspects of Wnt signaling pathways in higher vertebrates. *Current Opinion in Genetics & Development,* 11(5), 547-553.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., & Sigrist, C. J. (2006). The PROSITE database. *Nucleic Acids Research*, 34(suppl 1), D227-D230.

Imada, K., & Leonard, W. J. (2000). The Jak-STAT pathway. *Molecular Immunology,* 37(1), 1-11.

Iwanaga, S., & Lee, B. L. (2005). Recent advances in the innate immunity of invertebrate animals. *BMB Reports,* 38(2), 128-150.

James, R., Thampuran, N., Lalitha, K., Rajan, L. A., & Joseph, T. C. (2010). Differential gene expression profile of the hepatopancreas of white spot syndrome virus infected Fenneropenaeus indicus by suppression subtractive hybridization. *Fish & Shellfish Immunology,* 29(5), 884-889.

Janeway Jr, C. A., & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology,* 20(1), 197-216.

Jayachandran, K. (2001). *Palaemonid prawns: Biodiversity, Taxonomy, Biology and Management*: Science Publishers Inc., Enfield (NH), USA.

Jeena, K., Prasad, K. P., Pathan, M. K., & Babu, P. G. (2012). Expression profiling of WSSV ORF 199 and shrimp Ubiquitin conjugating enzyme in WSSV infected Penaeus monodon. *Asian-Australasian Journal of Animal Sciences,* 25(8), 1184.

Jin, S., Fu, H., Zhou, Q., Sun, S., Jiang, S., Xiong, Y., Gong, Y., Qiao, H., & Zhang, W. (2013). Transcriptome analysis of androgenic gland for discovery of novel genes from the oriental river prawn, Macrobrachium nipponense, using Illumina Hiseq 2000. *PloS One*, 8(10).

Jiravanichpaisal, P., Miyazaki, T., & Limsuwan, C. (1994). Histopathology, biochemistry, and pathogenicity of Vibrio harveyi infecting black tiger prawn Penaeus monodon. *Journal of Aquatic Animal Health,* 6(1), 27-35.

Joseph, S. W., Colwell, R. R., & Kaper, J. B. (1982). Vibrio parahaemolyticus and related halophilic vibrios. *Critical Reviews in Microbiology,* 10(1), 77-124.

Jung, H., Lyons, R. E., Dinh, H., Hurwood, D. A., McWilliam, S., & Mather, P. B. (2011). Transcriptomics of a giant freshwater prawn (Macrobrachium rosenbergii): de novo assembly, annotation and marker discovery. *PloS One,* 6(12).

Keezhedath, J., Kurcheti, P. P., Pathan, M. K., Babu, G. P., Tripathi, G., Sudhagar, A., & Rao, S. P. (2013). Expression profile of Penaeus monodon ubiquitin conjugating enzyme (PmUbc) at protein level in white spot syndrome virus challenged shrimp. *Indian Journal of Virology,* 24(1), 48-53.

Kisseleva, T., Bhattacharya, S., Braunstein, J., & Schindler, C. (2002). Signalling through the JAK/STAT pathway, recent advances and future challenges. *Gene,* 285(1), 1-24.

Kothapalli, R., Yoder, S. J., Mane, S., & Loughran, T. P. (2002). Microarray results: how accurate are they? *BMC Bioinformatics,* 3(1), 22.

Kuris, A. M., Ra'anan, Z., Sagi, A., & Cohen, D. (1987). Morphotypic differentiation of male Malaysian giant prawns, Macrobrachium rosenbergii. *Journal of Crustacean Biology*, 7(2), 219-237.

Lemaitre, B., & Hoffmann, J. (2007). The host defense of Drosophila melanogaster. *Annual Review of Immunology,* 25, 697-743.

Leu, J.H., Chen, S.H., Wang, Y.B., Chen, Y.C., Su, S.Y., Lin, C.Y., Ho, J.M., & Lo, C.F. (2011). A review of the major penaeid shrimp EST studies and the construction of a shrimp transcriptome database based on the ESTs from four penaeid shrimp. *Marine Biotechnology*, 13(4), 608-621.

Leu, J.H., Tsai, J.M., Wang, H.C., Wang, A. H.J., Wang, C.H., Kou, G.H., & Lo, C.F. (2005). The unique stacked rings in the nucleocapsid of the white spot syndrome virus virion are formed by the major structural protein VP664, the largest viral structural protein ever found. *Journal of Virology*, 79(1), 140-149.

Lewis, D. (1973). *Response of Brown Shrimp to Infection with Vibrio sp. 1*. Paper presented at the Proceedings of the annual workshop-World Mariculture Society, 333-338.

Li, F., & Xiang, J. (2013). Signalling pathways regulating innate immune responses in shrimp. *Fish & Shellfish Immunology,* 34(4), 973-980.

Li, F., Wang, D., Li, S., Yan, H., Zhang, J., Wang, B., Zhang, J., & Xiang, J. (2010). A Dorsal homolog (FcDorsal) in the Chinese shrimp Fenneropenaeus chinensis is responsive to both bacteria and WSSV challenge. *Developmental & Comparative Immunology,* 34(8), 874-883.

Li, F., Yan, H., Wang, D., Priya, T. J., Li, S., Wang, B., Zhang, J., & Xiang, J. (2009). Identification of a novel relish homolog in Chinese shrimp Fenneropenaeus chinensis and its function in regulating the transcription of antimicrobial peptides. *Developmental & Comparative Immunology,* 33(10), 1093-1101.

Li, S., Zhang, X., Sun, Z., Li, F., & Xiang, J. (2013). Transcriptome analysis on Chinese shrimp Fenneropenaeus chinensis during WSSV acute infection. *PloS One,* 8(3).

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).

Lightner, D. (1996). A Handbook of Shrimp Pathology and Diagnostic Procedures for Diseases of Cultured Penaeid Shrimp. *World Aquaculture Society, Baton Rouge, LA, USA*.

Lightner, D. V. (1985). *A review of the diseases of cultured penaeid shrimps and prawns with emphasis on recent discoveries and developments.* Paper presented at the Proceedings of the First International Conference on the Culture of Penaeid Prawns/Shrimps, 4-7 December 1984, Iloilo City, Philippines.

Lightner, D., & Redman, R. (1998). Shrimp diseases and current diagnostic methods. *Aquaculture,* 164(1), 201-220.

Liuxy, P. C., Lee, K. K., & Chen, S. N. (1996). Pathogenicity of different isolates of Vibrio harveyi in tiger prawn, Penaeus monodon. *Letters in Applied Microbiology,* 22(6), 413-416.

Lo, C., Ho, C., Chen, C., & Liu, K. (1997). Detection and tissue tropism of white spot syndrome baculovirus (WSBV) in captured brooders of Penaeus monodon with a special emphasis on reproductive organs. *Diseases of Aquatic Organisms,* 30(1), 53-72.

Lorenz, R., Bernhart, S. H., Zu Siederdissen, C. H., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1), 26.

Loureiro, J., & Ploegh, H. L. (2006). Antigen presentation and the ubiquitin-proteasome system in host–pathogen Interactions. *Advances in Immunology,* 92, 225-305.

Ma, K., Qiu, G., Feng, J., & Li, J. (2012). Transcriptome analysis of the oriental river prawn, Macrobrachium nipponense using 454 pyrosequencing for discovery of genes and markers. *PloS One,* 7(6).

MacLean, D., Jones, J. D., & Studholme, D. J. (2009). Application of' next-generation sequencing technologies to microbial genetics. *Nature Reviews Microbiology, 7*(4), 287-296.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics,* 24(3), 133-141.

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry,* 6, 287-303.

Mekata, T., Kono, T., Yoshida, T., Sakai, M., & Itami, T. (2008). Identification of cDNA encoding Toll receptor, MjToll gene from kuruma shrimp, Marsupenaeus japonicus. *Fish & Shellfish Immunology,* 24(1), 122-133.

Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics,* 95(6), 315-327.

Mohd-Shamsudin, M. I., Kang, Y., Lili, Z., Tan, T. T., Kwong, Q. B., Liu, H., Zhang, G., Othman, R. Y., & Bhassu, S. (2013). In-depth tanscriptomic analysis on giant freshwater prawns. *PloS one*, 8(5).

Morato, T., Watson, R., Pitcher, T. J., & Pauly, D. (2006). Fishing down the deep. *Fish and Fisheries,* 7(1), 24-34.

Morozova, O., & Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics,* 92(5), 255-264.

Murphy, D. (2002). Gene expression studies using microarrays: principles, problems, and prospects. *Advances in Physiology Education,* 26(4), 256-270.

Mutz, K. O., Heilkenbrinker, A., Lönne, M., Walter, J. G., & Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology,* 24(1), 22-30.

Nadala Jr, E., & Loh, P. C. (1998). A comparative study of three different isolates of white spot virus. *Diseases of Aquatic Organisms,* 33(3), 231-234.

Nash, G., Nithimathachoke, C., Tungmandi, C., Arkarjamorn, A., Prathanpipat, P., & Ruamthaveesub, P. (1992). Vibriosis and its control in pond-reared Penaeus monodon in Thailand. *Diseases in Asian Aquaculture,* 1, 143-155.

New, M. B. (1990). Freshwater prawn culture: a review. *Aquaculture,* 88(2), 99-143.

New, M. B. (2002). Farming freshwater prawns: a manual for the culture of the giant river prawn (Macrobrachium rosenbergii). *Food & Agriculture Org*, 148, 1-11.

New, M. B. (2009). History and global status of freshwater prawn farming. *Freshwater Prawns: Biology and Farming*, 194, 16-40.

Okugawa, S., Mekata, T., Inada, M., Kihara, K., Shiki, A., Kannabiran, Kono, T., Sakai, M., Yoshida, T., Itami, T., & Sudhakaran, R. (2013). The SOCS and STAT from JAK/STAT signaling pathway of kuruma shrimp Marsupenaeus japonicus: molecular cloning, characterization and expression analysis. *Molecular and Cellular Probes,* 27(1), 6-14.

Paradkar, P. N., Trinidad, L., Voysey, R., Duchemin, J.-B., & Walker, P. J. (2012). Secreted Vago restricts West Nile virus infection in Culex mosquito cells by activating the Jak-STAT pathway. *Proceedings of the National Academy of*

Sciences, 109(46), 18915-18920.

Patterton, H.-G., & Graves, S. (2000). DNAssist: the integrated editing and analysis of molecular biology sequences in windows. *Bioinformatics,* 16(7), 652-653.

Pearson, W. R. (1994). Using the FASTA program to search protein and DNA sequence databases. *In Computer Analysis of Sequence Data: Part 1*, 307-331.

Peng, S., Lo, C., Ho, C., Chang, C., & Kou, G. (1998). Detection of white spot baculovirus (WSBV) in giant freshwater prawn, Macrobrachium rosenbergii, using polymerase chain reaction. *Aquaculture, 164*(1), 253-262.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8(10), 785-786.

Plowman, G. D., Sudarsanam, S., Bingham, J., Whyte, D., & Hunter, T. (1999). The protein kinases of Caenorhabditis elegans: a model for signal transduction in multicellular organisms. *Proceedings of the National Academy of Sciences,* 96(24), 13603-13610.

Poh S. C., Chu F., L., Yu C., W., & Guang H., K. (1996). Identification of white spot syndrome associated baculovirus (WSBV) target organs in the shrimp Penaeus monodon by in situ hybridization. *Diseases of Aquatic Organisms,* 27(2), 131-139.

Porollo, A. A., Adamczak, R., & Meller, J. (2004). POLYVIEW: a flexible visualization tool for structural and functional annotations of proteins. *Bioinformatics*, 20(15), 2460-2462.

Pradeep, B., Rai, P., Mohan, S. A., Shekhar, M. S., & Karunasagar, I. (2012). Biology, Host Range, Pathogenesis and Diagnosis of White spot syndrome virus. *Indian Journal of Virology,* 23(2), 161-174.

Primavera, J. H. (1997). Socio-economic impacts of shrimp culture. *Aquaculture Research,* 28(10), 815-827.

Ramirez, J. L., & Dimopoulos, G. (2010). The Toll immune signaling pathway control conserved anti-dengue defenses across diverse Aedes aegypti strains and against multiple dengue virus serotypes. *Developmental & Comparative Immunology,* 34(6), 625-629.

Rao, A.V. 2015. Vibriosis in Shrimp Aquaculture. Available at http://www.neospark.com (Accessed on 20th February 2016).

Rao, R., Zhu, Y. B., Alinejad, T., Tiruvayipati, S., Thong, K. L., Wang, J., & Bhassu, S. (2015). RNA-seq analysis of Macrobrachium rosenbergii hepatopancreas in response to Vibrio parahaemolyticus infection. *Gut Pathogens,* 7(1), 6.

Rawlings, J. S., Rosler, K. M., & Harrison, D. A. (2004). The JAK/STAT signaling pathway. *Journal of Cell Science,* 117(8), 1281-1283.

Reiner, A., Yekutieli, D., & Benjamini, Y. (2003). Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics,* 19(3), 368-375.

Rojtinnakorn, J., Hirono, I., Itami, T., Takahashi, Y., & Aoki, T. (2002). Gene expression in haemocytes of kuruma prawn, Penaeus japonicus, in response to infection with WSSV by EST approach. *Fish & Shellfish Immunology,* 13(1), 69-83.

Rosenberry, B. (2000). World shrimp farming 2000. (San Diego: Shrimp News International; 2000).

Sabin, L. R., Hanna, S. L., & Cherry, S. (2010). Innate antiviral immunity in Drosophila. *Current Opinion in Immunology,* 22(1), 4-9.

Sadamoto, H., Takahashi, H., Okada, T., Kenmoku, H., Toyota, M., & Asakawa, Y. (2012). De novo sequencing and transcriptome analysis of the central nervous system of mollusc Lymnaea stagnalis by deep RNA sequencing. *PloS One,* 7(8).

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,* 74(12), 5463-5467.

Santos, C. A., Blanck, D. V., & de Freitas, P. D. (2014). RNA-seq as a powerful tool for penaeid shrimp genetic progress. *Livestock Genomics,* 5, 298.

Saulnier, D., Haffner, P., Goarant, C., Levy, P., & Ansquer, D. (2000). Experimental infection models for shrimp vibriosis studies: a review. *Aquaculture,* 191(1), 133-144.

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols,* 3(6), 1101-1108.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology,* 26(10), 1135-1145.

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology,* 21(8), 1794-1805.

Sim, S., & Dimopoulos, G. (2010). Dengue virus inhibits immune responses in Aedes aegypti cells. *PloS One,* 5(5).

Simon, R., Radmacher, M. D., Dobbin, K., & McShane, L. M. (2003). Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *Journal of the National Cancer Institute,* 95(1), 14-18.

Sizemore, R. K., & Davis, J. W. (1985). Source of Vibrio spp. found in the hemolymph of the blue crab, Callinectes sapidus. *Journal of Invertebrate Pathology,* 46(1), 109-110.

Sizemore, R., Colwell, R., Tubiash, H., & Lovelace, T. (1975). Bacterial flora of the

hemolymph of the blue crab, Callinectes sapidus: numerical taxonomy. *Applied Microbiology,* 29(3), 393-399.

Somboonwiwat, K., Supungul, P., Rimphanitchayakit, V., Aoki, T., Hirono, I., & Tassanakajon, A. (2006). Differentially expressed genes in hemocytes of Vibrio harveyi-challenged shrimp Penaeus monodon. *Journal of Biochemistry and Molecular Biology,* 39(1), 26-36.

Sookruksawong, S., Sun, F., Liu, Z., & Tassanakajon, A. (2013). RNA-seq analysis reveals genes associated with resistance to Taura syndrome virus (TSV) in the Pacific white shrimp Litopenaeus vannamei. *Developmental & Comparative Immunology,* 41(4), 523-533.

Souza-Neto, J. A., Sim, S., & Dimopoulos, G. (2009). An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. *Proceedings of the National Academy of Sciences,* 106(42), 17841-17846.

Sun, C., Shao, H. L., Zhang, X. W., Zhao, X. F., & Wang, J. X. (2011). Molecular cloning and expression analysis of signal transducer and activator of transcription (STAT) from the Chinese white shrimp Fenneropenaeus chinensis. *Molecular Biology Reports,* 38(8), 5313-5319.

Supungul, P., Klinbunga, S., Pichyangkura, R., Hirono, I., Aoki, T., & Tassanakajon, A. (2004). Antimicrobial peptides discovered in the black tiger shrimp Penaeus monodon using the EST approach. *Diseases of Aquatic Organisms,* 61(1-2), 123-135.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution,* 30(12), 2725-2729.

Thompson, J. D., Gibson, T., & Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*, 2.3. 1-2.3. 22.

Tiruvayipati, S., Bhassu, S., Kumar, N., Baddam, R., Shaik, S., Gurindapalli, A. K., Thong, K. L., & Ahmed, N. (2013). Genome anatomy of the gastrointestinal pathogen, Vibrio parahaemolyticus of crustacean origin. *Gut Pathogens*, 5(1), 37.

Tsai, J. M., Wang, H. C., Leu, J. H., Hsiao, H. H., Wang, A. H. J., Kou, G. H., & Lo, C. F. (2004). Genomic and proteomic analysis of thirty-nine structural proteins of shrimp white spot syndrome virus. *Journal of Virology,* 78(20), 11360-11370.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115-e115.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 418-426.

van Hulten, M. C., & Vlak, J. M. (2001). Identification and phylogeny of a protein kinase gene of white spot syndrome virus. *Virus Genes,* 22(2), 201-207.

van Hulten, M. C., Goldbach, R. W., & Vlak, J. M. (2000). Three functionally diverged major structural proteins of white spot syndrome virus evolved by gene duplication. *Journal of General Virology,* 81(10), 2525-2529.

van Hulten, M. C., Reijns, M., Vermeesch, A. M., Zandbergen, F., & Vlak, J. M. (2002). Identification of VP19 and VP15 of white spot syndrome virus (WSSV) and glycosylation status of the WSSV major structural proteins. *Journal of General Virology,* 83(1), 257-265.

van Hulten, M. C., Westenberg, M., Goodall, S. D., & Vlak, J. M. (2000). Identification of two major virion protein genes of white spot syndrome virus of shrimp. *Virology,* 266(2), 227-236.

van Hulten, M. C., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R. K., & Vlak, J. M. (2001). The white spot syndrome virus DNA genome sequence. *Virology*, 286(1), 7-22.

Van Valen, L. (1974). Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution,* 3(2), 89-101.

Vidya, R., Gireesh-Babu, P., & Prasad, K. P. (2013). White spot syndrome virus manipulates ubiquitin gene expression in Penaeus monodon. *Indian Journal of Virology,* 24(1), 82-84.

Waddell, S. J., Butcher, P. D., & Stoker, N. G. (2007). RNA profiling in host–pathogen interactions. *Current Opinion in Microbiology,* 10(3), 297-302.

Walker, P. J., & Mohan, C. (2009). Viral disease emergence in shrimp aquaculture: origins, impact and the effectiveness of health management strategies. *Reviews in Aquaculture,* 1(2), 125-154.

Wang, B., Li, F., Dong, B., Zhang, X., Zhang, C., & Xiang, J. (2006). Discovery of the genes in response to white spot syndrome virus (WSSV) infection in Fenneropenaeus chinensis through cDNA microarray. *Marine Biotechnology, 8*(5), 491-500.

Wang, B., Li, F., Luan, W., Xie, Y., Zhang, C., Luo, Z., Gui, L., Yan, H., & Xiang, J. (2008). Comparison of gene expression profiles of Fenneropenaeus chinensis challenged with WSSV and Vibrio. *Marine Biotechnology*, 10(6), 664-675.

Wang, C. H., Yang, H. N., Tang, C. Y., Lu, C. H., Kou, G. H., & Lo, C. F. (2000). Ultrastructure of white spot syndrome virus development in primary lymphoid organ cell cultures. *Diseases of Aquatic Organisms,* 41(2), 91-104.

Wang, P. H., Liang, J. P., Gu, Z. H., Wan, D. H., Weng, S. P., Yu, X. Q., & He, J. G. (2012). Molecular cloning, characterization and expression analysis of two novel Tolls (LvToll2 and LvToll3) and three putative Spätzle-like Toll ligands (LvSpz1–3) from Litopenaeus vannamei. *Developmental & Comparative Immunology,* 36(2), 359-371.

Wang, Q., Poulos, B., & Lightner, D. (2000). Protein analysis of geographic isolates of

shrimp white spot syndrome virus. *Archives of Virology,* 145(2), 263-274.

Wang, Y., Hassan, M., Shariff, M., Zamri, S., & Chen, X. (1999). Histopathology and cytopathology of white spot syndrome virus (WSSV) in cultured Penaeus monodon from peninsular Malaysia with emphasis on pathogenesis and the mechanism of white spot formation. *Diseases of Aquatic Organisms,* 39(1), 1-11.

Wen, R., Li, F., Sun, Z., Li, S., & Xiang, J. (2013). Shrimp MyD88 responsive to bacteria and white spot syndrome virus. *Fish & Shellfish Immunology,* 34(2), 574-581.

Wise, R. P., Moscou, M. J., Bogdanove, A. J., & Whitham, S. A. (2007). Transcript profiling in host-pathogen interactions. *Annual Review of Phytopathology,* 45, 329-369.

Wisely, B. (1984). Freshwater prawn farming. A manual for the culture of Macrobrachium rosenbergii: MB New and S. Singholka. FAO Fish. Tech. Pap. 225. FAO, Rome, 1982. 116 pp.

Withyachumnarnkul, B., Boonsaeng, V., Chomsoong, R., Flegel, T. W., Muangsin, S., & Nash, G. L. (2003). Seasonal variation in white spot syndrome virus-positive samples in broodstock and post-larvae of Penaeus monodon in Thailand. *Diseases of Aquatic Organisms,* 53(2), 167-171.

Wongpanya, R., Aoki, T., Hirono, I., Yasuike, M., & Tassanakajon, A. (2007). Analysis of gene expression in haemocytes of shrimp Penaeus monodon challenged with white spot syndrome virus by cDNA microarray. *Science Asia,* 33, 165-174.

Wonteerasupaya, C., Vickers, J. E., Sriurairatana, S., Nash, G. L., Akarajamorn, A., Boonsaeng, V., Panyim, S., Tassanakajon, A., Withyachumnarnkul, B., & Flegel, T. (1995). A non-occluded, systemic baculovirus that occurs in cells of ectodermal and mesodermal origin and causes high mortality in the black tiger prawn Penaeus monodon. *Diseases of Aquatic Organisms*, 21(1), 69-77.

Xie, Y. K., Ding, D., Wang, H. M., & Kang, C. J. (2015). A homologue gene of β-catenin participates in the development of shrimps and immune response to bacteria and viruses. *Fish & Shellfish Immunology,* 47(1), 147-156.

Xue, S., Liu, Y., Zhang, Y., Sun, Y., Geng, X., & Sun, J. (2013). Sequencing and de novo analysis of the hemocytes transcriptome in Litopenaeus vannamei response to white spot syndrome virus infection. *PloS One,* 8(10).

Yang, C., Zhang, J., Li, F., Ma, H., Zhang, Q., Priya, T. J., Zhang, X., & Xiang, J. (2008). A Toll receptor from Chinese shrimp Fenneropenaeus chinensis is responsive to Vibrio anguillarum infection. *Fish & Shellfish Immunology,* 24(5), 564-574.

Yang, F., He, J., Lin, X., Li, Q., Pan, D., Zhang, X., & Xu, X. (2001). Complete genome sequence of the shrimp white spot bacilliform virus. *Journal of Virology,* 75(23), 11811-11820.

Zambon, R. A., Nandakumar, M., Vakharia, V. N., & Wu, L. P. (2005). The Toll pathway is important for an antiviral response in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America,* 102(20), 7257-7262.

Zeng, D., Chen, X., Xie, D., Zhao, Y., Yang, C., Li, Y., Ma, N., Peng, M., Yang, Q., & Liao, Z. (2013). Transcriptome analysis of pacific white shrimp (Litopenaeus vannamei) hepatopancreas in response to Taura Syndrome Virus (TSV) experimental infection. *PloS One*, 8(2).

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics,* 9(1), 40.

Zhou, X., Ren, L., Meng, Q., Li, Y., Yu, Y., & Yu, J. (2010). The next-generation sequencing technology and application. *Protein & Cell,* 1(6), 520-536.

Zhu, F., & Zhang, X. (2013). The Wnt signaling pathway is involved in the regulation of phagocytosis of virus in Drosophila. *Scientific Reports,* 3.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Rama Rao, Ya Bing Zhu, Tahereh Alinejad, Suma Tiruvayipati, Kwai Lin Thong, Jun Wang and Subha Bhassu. RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection. *Gut Pathogens, 7*(1), 6/2015.

Rama Rao, Subha Bhassu, Robin Zhu Ya Bing, Tahereh Alinejad, Sharifah Syed Hassan, Jun Wang (2016). A transcriptome study on *Macrobrachium rosenbergii* hepatopancreas experimentally challenged with white spot syndrome virus (WSSV). *Journal of Invertebrate Pathology*, 136, 10-22.

Paper / Poster presentations:

Poster presentation entitled:

"Transcriptome Analysis of *Macrobrachium Rosenbergii* Hepatopancreas experimentally challenged with White Spot Syndrome Virus (WSSV)" at 38th Annual Conference of the Malaysian Society For Biochemistry and Molecular Biology, Putrajaya, August, 2013.

"RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection" at 19th Biological Sciences Graduate Congress (BSGC) National University of Singapore, Singapore, December, 2014.

Paper presentation entitled:

"Stat gene characterization and expression in *Macrobrachium rosenbergii* against WSSV and *Vibrio parahaemolyticus*" at 22nd Annual Scientific Meeting, RMIC, University Malaya, Kuala Lumpur, September, 2015.

POSTER 56

TRANSCRIPTOME ANALYSIS OF *Macrobrachium rosenbergii* HEPATOPANCREAS EXPERIMENTALLY CHALLENGED WITH WHITE SPOT SYNDROME VIRUS (WSSV)

RAMARAO SERIRAMALU, TAHERAH ALINEJAD, AND SUBHA BHASSU

Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University Malaya, 50603 Kuala Lumpur

White spot syndrome (WSSV) is a deadly viral disease effecting *Macrobrachium rosenbergii* (*M.rosenbergii*) shrimp farming industry worldwide. *M.rosenbergii* possesses an innate immune system characterized by lack of immunoglobulin and memory, which provides defence against pathogenic agents. Knowledge regarding the immune system regulation in this species is necessary to develop methods to control and minimize the loss of production due to this viral disease. In this study, we performed a transcriptome analysis of the hepatopancreas from *M.rosenbergii* experimentally infected with WSSV, using the 'Next Generation' high-throughput sequencing method (Illumina HiSeq2000). A total of 38,671,944 and 54,708,014 high-quality reads obtained from WSSV-infected and control *M.rosenbergii* cDNA libraries. The overall de novo assembly of cDNA sequence data generated 63,584 unigenes, with an average length of 648 bp. Based on BLASTX search (E-value $<10^{-5}$) against NR, Swissprot, GO, COG and KEGG databases, 22,455 unigenes (35.31% of all unigenes) were annotated with gene descriptions, gene ontology terms, or metabolic pathways. The unigene differential expression analysis revealed that 14,415 unigenes differentially expressed in the infected shrimp compared to the controls, including 8,443 up- and 5,972 down- regulated unigenes. Several differentially expressed genes are involved in various animal immune functions, such as antiviral, antimicrobial, proteases, protease inhibitors, signal transduction, transcriptional control, cell death and cell adhesion. Results can contribute to the in-depth study of candidate genes in shrimp immunity, and improves our current understanding of this host-virus interaction.

19th Biological Sciences Graduate Congress (BSGC)

12-14th December 2014

National University of Singapore, Singapore

Part II: Abstract

Abstract Title:

RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection

Abstract:

RNA-seq analysis of *Macrobrachium rosenbergii* hepatopancreas in response to *Vibrio parahaemolyticus* infection

Rama Rao and Subha Bhassu

Animal Genomics and Evolutionary Biology, Genetics and Molecular Biology and Institute of Biological Sciences, Faculty of Science and CEBAR, University Malaya, 50603 Kuala Lumpur.

The Malaysian giant freshwater prawn, *Macrobrachium rosenbergii*, is an economically important crustacean worldwide, being affected by Vibriosis disease caused by Vibrio strains such as *Vibrio parahaemolyticus*. *M. rosenbergii* possesses an innate immune system which provides defence against pathogenic agents. Knowledge regarding the immune system regulation in this species is necessary to

develop methods to control and minimize the loss of production due to this bacterial disease. In this study, we performed a transcriptomics analysis of *M. rosenbergii* hepatopancreas infected with *V. parahaemolyticus* using the 'Next Generation' sequencing method (Illumina HiSeqTM2000). A total of 58,385,094 and 54,708,014 high-quality reads obtained from *Vibrio*-infected and control *M. rosenbergii* cDNA libraries. The overall de novo assembly and clustering of both reads generated 64,411 unigenes, with an average length of 698 bp. Based on BLASTX search (E-value <10-5) against NR, Swissprot, GO, COG and KEGG databases, 22,455 unigenes (34.86% of all unigenes) were annotated with gene descriptions, gene ontology terms, and metabolic pathways. The unigene differential expression analysis revealed 14,569 unigenes were differentially expressed in the infected shrimp compared to the controls. Several differentially expressed genes are involved in various animal immune functions. The large number of transcripts obtained in this study would provide valuable resources for further genomic research into freshwater prawn.

Oral Presentation

STAT gene Characterization and Expression in *Macrobrachium rosenbergii* Against WSSV and *Vibrio parahaemolyticus*

RAMARAO SERIRAMALU AND SUBHA BHASSU

Animal Genomics and Evolutionary Biology, Genetics and Molecular Biology and Institute of Biological Sciences, Faculty of Science and CEBAR, University Malaya, 50603 Kuala Lumpur.

M.rosenbergii possesses an innate immune system which is being regulated by signaling pathways, one of which is the Janus kinase (JAK)/signal transducer and activator of transcription (STAT) pathway. Here we reported the full-length cDNA of STAT (designated as MrSTAT) obtained from the transcriptome database of freshwater prawn *M. rosenbergii*. MrSTAT full length cDNA was 2907 base pairs (bp) with an open reading frame (ORF) of 2436 bp encoding for an 811 amino acids long protein. MrSTAT cDNA contains 102 bp 5' untranslated region (UTR) and 240 bp 3' UTR including a stop codon. MrSTAT coding region contains 4 domains STAT_INT domain, STAT alpha domain, STAT_bind domain and SH2 domain. 24 high probability common motifs were also found in the amino acid sequence of MrSTAT. Phylogenetic analysis reveals that the MrSTAT is clustered with STAT5s from vertebrates and STATs from invertebrates. The three-dimensional structure of MrSTAT modelled by the I-Tasser program was compared with structures deposited in the Protein Data Bank to find out the structural similarity of MrSTAT with experimentally identified structures. The results showed that the MrSTAT exhibits maximum structural identity with unphosphorylated STAT1 (PDB id 1yvl) from Homo sapiens. The Minimum free energy of the predicted RNA structure of MrSTAT is -753.40 kcal/mol. Tissue distribution analysis by quantitative real-time PCR showed wide distribution in all detected tissues for MrSTAT. Time course analysis of the transcription level of MrSTAT in hepatopancreas after WSSV and *V.parahaemolyticus* challenge showed significant ($p < 0.05$) increase expression with highest are at 6 hours for WSSV infected and 24 hour for *V.parahaemolyticus* infected ones. All these results may imply the potential role of JAK/STAT pathway in the immune response against bacteria and virus in *M. rosenbergii*.

APPENDIX

APPENDIX A: PURITY AND INTEGRITY ASSESSMENT OF THE RNA SAMPLES BY USING THE BIOANALYZER 2100 (AGILENT TECHNOLOGIES, USA)

华大基因 BGI

File Code/Version Number: BGI-NGS-R10-036/A0

BGI Sample Testing Report

1. Project Information

Report No.: TSZe13022605

Project Name				Project No.	HKC11023		
Customer Name	DR SUBHA BHASSU			Customer Unit	UNIVERSITY MALAYA		
Lab Sample Collector	Cherry			Lab Sample Receiving Date	20130225		
Lab Sample Tester	Cherry			Lab Sample Testing Date	20130326		
Reported by	Kevin	Inspected by	Ziyin HUANG	Approved by	Ziyin HUANG	Report Date	20130226

2. Sample Test Method

①Method of concentration determination: □ QubitFluorometer, ■ Agilent 2100, □ NanoDrop, □MicroplateReader;
②Method of 28S/18S & 23S/16S test: ■ Agilent 2100;
③Method of RIN test: ■ Agilent 2100;

3. Sample Test Result

No.	Sample Name	Sample Number	Tube No.	Concentration (ng/μL)	Volume (μL)	Total Mass (μg)	RIN	28S/18S	Library Type	Test Result	Remark
1	MR HP CONTROL TRANSCRIPTOME	0130117841	1	1955	53	103.62	6.4	0.0	RNA-Seq (Transcriptome)	Level A	
									Small RNA	Level A	

华大基因 BGI

File Code/Version Number: BGI-NGS-R10-036/A0

BGI Sample Testing Report

1. Project Information

Report No.: TSZe13050207

Project Name	Flagship collaborate project of bananas genome de novo sequence for University Malaya			Project No.	HKC11023		
Customer Name	DR SUBHA BHASSU			Customer Unit	UNIVERSITY MALAYA		
Lab Sample Collector	Sliy			Lab Sample Receiving Date	20130502		
Lab Sample Tester	Qixuan CAI			Lab Sample Testing Date	20130502		
Reported by	Sliy	Inspected by	Qixuan CAI	Approved by	Qixuan CAI	Report Date	20130502

2. Sample Test Method

①Method of concentration determination: □ QubitFluorometer, ■ Agilent 2100, □ NanoDrop, □MicroplateReader;
②Method of 28S/18S & 23S/16S test: ■ Agilent 2100;
③Method of RIN test: ■ Agilent 2100;

3. Sample Test Result

No.	Sample Name	Sample Number	Tube No.	Concentration (ng/μL)	Volume (μL)	Total Mass (μg)	RIN	28S/18S	Library Type	Test Result	Remark
1	MR HP WSSV TRANSCRIPTOME AND MIRNA	0130418112	1	2328	38	88.46	6.6	0.0	Small RNA	Level A	
									RNA-Seq (Transcriptome)	Level A	

5	MR HP VIBRIO TRANSCRIPTOME	0130305207	1	714	44	31.42	7.2	0.0	Small RNA	Level A	
									RNA-Seq (Transcriptome)	Level A	

APPENDIX B: OVERVIEW OF THE CONTROL AND *V. PARAHAEMOLYTICUS* INFECTED TRANSCRIPTOME ASSEMBLY

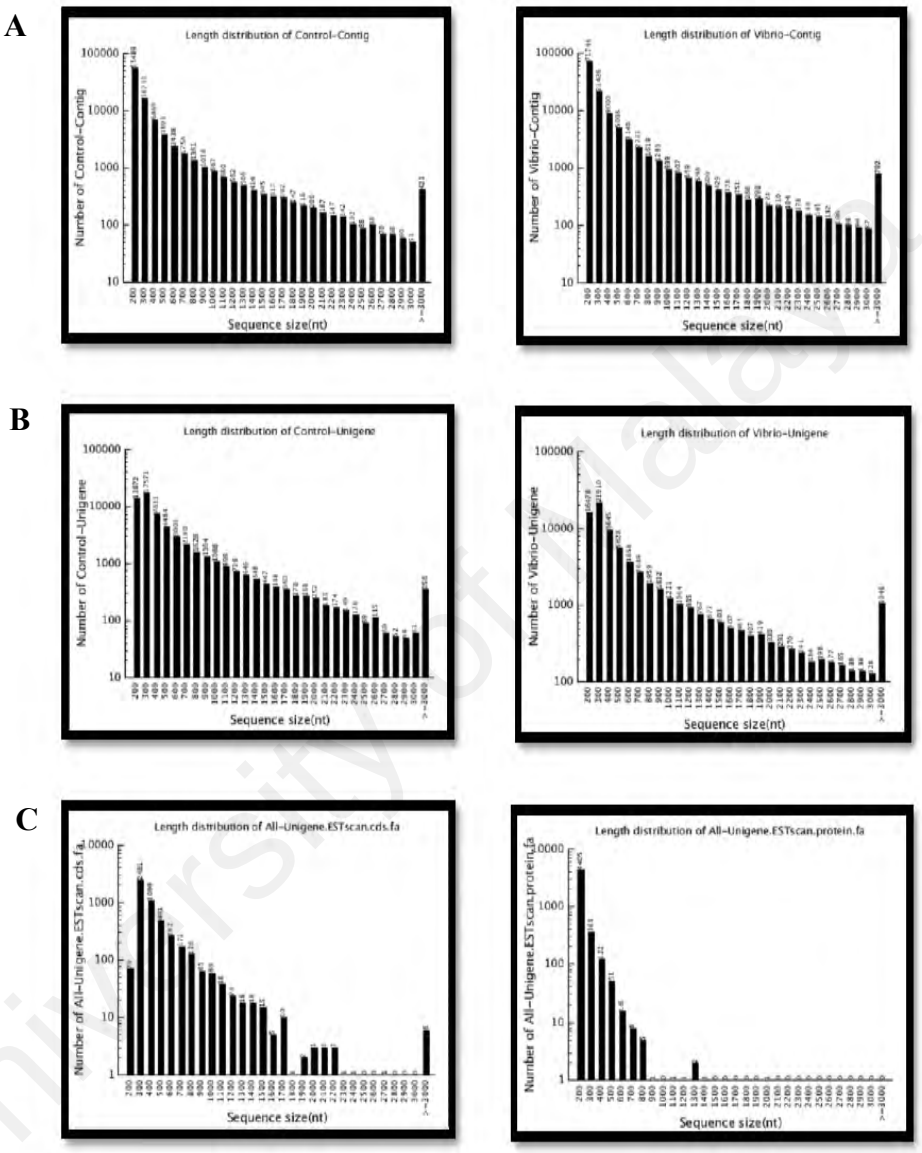


Figure B: (A) The size distribution of the contigs obtained from our denovo assembly of high-quality clean reads. (B) The size distribution of the unigenes produced from further assembly of contigs (i.e., contig joining, gap filling, and scaffold clustering) and (C) Size

distributions of the ESTs and proteins obtained from the ESTScan results. For unigene CDS that had no hits in the databases (Nr, SwissProt, KEGG and COG), the BLAST results were subjected to ESTScans and then translated into peptide sequences.

APPENDIX C: OVERVIEW OF THE CONTROL AND WSSV INFECTED TRANSCRIPTOME ASSEMBLY

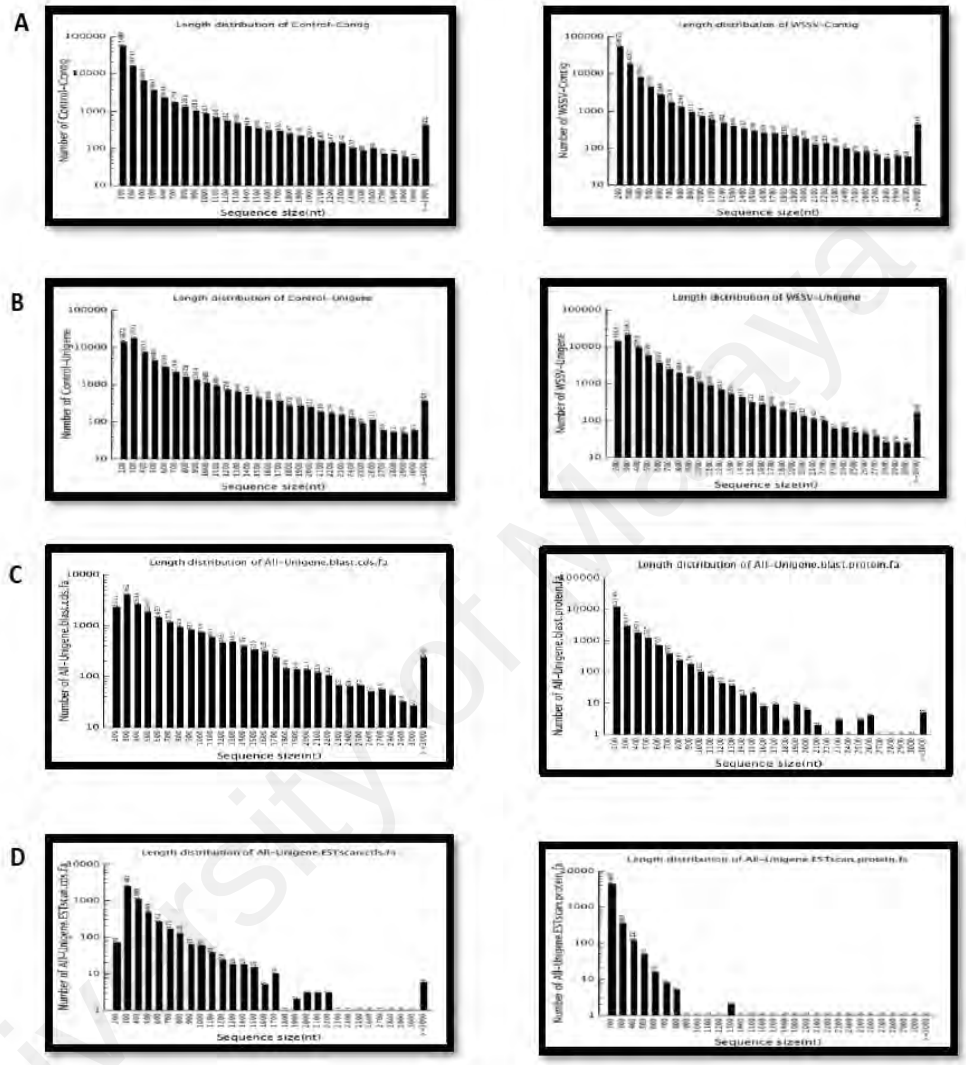
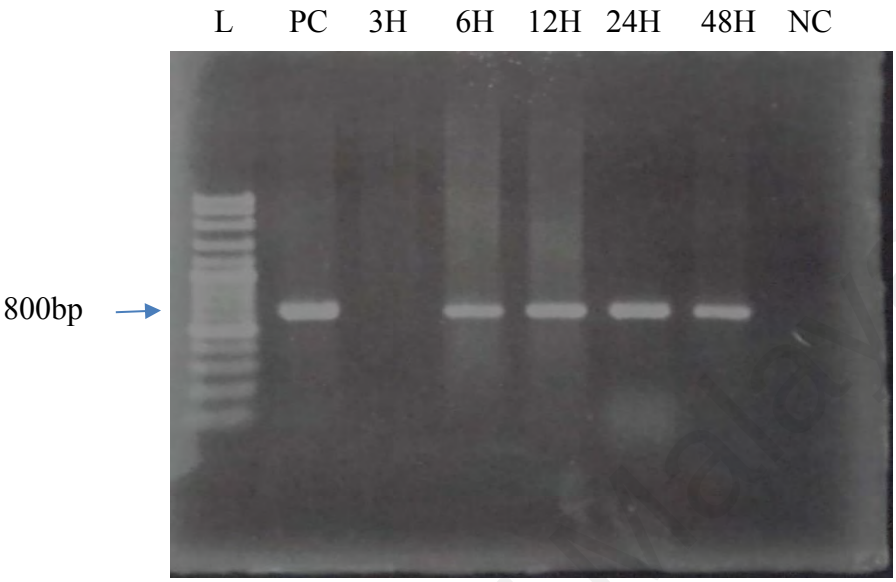


Figure C: (A) The size distribution of the contigs obtained from our *denovo a*ssembly of high-quality clean reads. (B) The size distribution of the unigenes produced from further assembly of contigs (i.e., contig joining, gap filling, and scaffold clustering). (C) Size distributions of the coding sequences (CDS) and identified proteins and (D) Size distributions of the ESTs and proteins obtained from the ESTScan results. For unigene

CDS that had no hits in the databases (Nr, SwissProt, KEGG and COG), the BLAST results were subjected to ESTScans and then translated into peptide sequences.

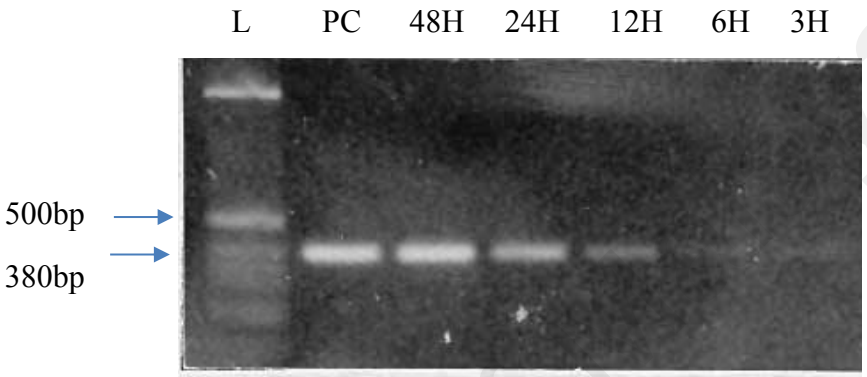
APPENDIX D: WSSV POSITIVE TESTING IN WSSV CHALLENGED *M. ROSENBERGII*



Forward Primer	Reverse Primer	Product Size (bp)
5'-GATGATTCTGTAGATGATAC-3'	5'-CGCATATGTCCTATTAC-3'	800bp

Figure D: Figure represents 1% (w/v) agarose gel of WSSV-positive testing randomly in WSSV challenged *M. rosenbergii* at time point: - Lane L: Promega 1 Kb ladder, Lane PC: positive control of WSSV (800 bp); Lane 1: 3h, Lane 2: 6 hour, Lane 3: 12 hour, Lane 4: 24 hour, Lane 5: 48 hour and NC: negative control.

APPENDIX E: *V. PARAHAEMOLYTICUS* POSITIVE TESTING IN *V. PARAHEMOLYTICUS* CHALLENGED *M. ROSENBERGII*



Forward Primer	Reverse Primer	Product Size (bp)
5'-TGCGAATTCGATAGG GTGTTAACC-3'	5'-CGAATCCTTGAACATACGCA GC-3'	380bp

Figure E: Figure represents 1% (w/v) agarose gel of *V. parahaemolyticus*-positive testing randomly in *V. parahaemolyticus* challenged *M. rosenbergii* at time point: - Lane L: Promega 1 Kb ladder, Lane PC: positive control of V. parahaemolyticus (380 bp); Lane 1: 48h, Lane 2: 24 hour, Lane 3: 12 hour, Lane 4: 6 hour and Lane 5: 3 hour.

APPENDIX F: QRT-PCR ($2^{-\Delta\Delta CT}$) FOR TEN IMMUNE RELATED UNIGENES USED TO EVALUATE THE WSSV TRANCSRIPTOME SEQUENCING DATA.

	Ak1		Alf		Apoptosis inhibitor		Caspace		Chaperone		Hsp21		Lectin 1		Mnsod		Nfk		PRX	
	control	infected	control	infected	control	infected	control	infected	control	infected	control	infected	control	infected	control	infected	control	infected	control	infected
Ct1	35	33.181	30.509	32.572	31.683	31.393	33.771	33.385	30.529	29.791	34	32.392	29.376	28.778	26.66	25.804	33.281	32.522	31.271	35.464
Ct2	34.943	33.894	30.371	33.413	32.449	31.224	33.323	33.11	30.339	29.995	33.625	32.237	29.141	28.778	26.491	25.604	32.715	32.487	30.845	34.544
Ct3	35.31	33.6	30.629	32.909	32	30.88	33.533	32.506	30.497	30.157	33.18	31.729	28.866	28.691	26.755	25.626	32.941	32.875	31.231	35.181
Average	35.1265	33.55833	30.503	32.96467	32.044	31.16567	33.54233	33.00033	30.455	29.981	33.60167	32.11933	29.12767	28.749	26.63533	25.678	32.979	32.628	31.11567	35.063
	Housekeeping gene																			
Ct1	26.172	23.156	26.157	26.157	26.411	25.311	27.421	26.421	26.308	27.308	27.626	24.626	26.304	24.304	25.311	23.311	27.552	26.552	26.514	28.677
Ct2	26.15	24	26.444	26.444	26.615	25.415	28	26	26.621	27.621	27.831	24.831	26.795	24.795	25.521	23.521	27.892	26.892	26.665	28
Ct3	26.8	23.657	27	27	27	25	27.976	26.976	27	28	28	25	27	25	26	23	28	26	27	28.536
	-8.828	-10.025	-4.352	-6.415	-5.272	-6.082	-6.35	-6.964	-4.221	-2.483	-6.374	-7.766	-3.072	-4.474	-1.349	-2.493	-5.729	-5.97	-4.757	-6.787
	-8.793	-9.894	-3.927	-6.969	-5.834	-5.809	-5.323	-7.11	-3.718	-2.374	-5.794	-7.406	-2.346	-3.983	-0.97	-2.083	-4.823	-5.595	-4.18	-6.544
	-8.51	-9.943	-3.629	-5.909	-5	-5.88	-5.557	-5.53	-3.497	-2.157	-5.18	-6.729	-1.866	-3.691	-0.755	-2.626	-4.941	-6.875	-4.231	-6.645
	454.457	1041.899	20.42126	85.33112	38.63938	67.743	81.57188	124.8455	18.64866	5.590588	82.94022	217.6702	8.409383	22.22328	2.547355	5.629474	53.03967	62.6829	27.03956	110.4309
	443.56447	951.4605	15.21055	125.2789	57.04387	56.06389	40.02973	138.1412	13.1592	5.183764	55.484	169.6009	5.084127	15.81257	1.958841	4.236873	28.30529	48.33512	18.12614	93.3126
	364.55685	984.3312	12.37194	60.08779	32	58.89201	47.07862	46.20573	11.29021	4.459865	36.25228	106.0793	3.645205	12.91522	1.687632	6.173121	30.71774	117.3765	18.77837	100.0793
Average	420.85944	992.5636	16.00125	90.23261	42.56108	60.89963	56.22674	103.0641	14.36602	5.078072	58.2255	164.4501	5.712905	16.98369	2.064609	5.346489	37.35423	76.13151	21.31469	101.2743
STDEV	49.062697	45.77792	4.082498	32.8708	12.97435	6.092891	22.23069	49.68756	3.82479	0.572723	23.46439	55.97344	2.443536	4.763261	0.439513	0.998661	13.63744	36.43251	4.968598	8.62148

APPENDIX G: QRT-PCR ($2^{-\Delta\Delta CT}$) FOR SEVEN IMMUNE RELATED UNIGENES USED TO EVALUATE THE *V. PARAHAEMOLYTICUS* TRANCSRIPTOME SEQUENCING DATA.

	Ak1		Alf		Apoptosis inhibitor		Caspace		Hsp21		Lectin 1		Nfk	
	control	infected	control	infected	control	infected	control	infected	control	infected	control	infected	control	infected
Ct1	35	34.768	30.509	31.799	32.449	30.312	33.771	30.737	34	30.716	29.376	30.019	33.281	30.632
Ct2	34.943	34.722	30.371	31.545	32	30.646	33.323	31	33.625	30.893	29.141	30.012	32.715	31.095
Ct3	35.31	34.301	30.629	31.909	31.683	30.823	33.533	30.931	33.18	30.422	28.866	29.76	32.941	31
Average	35.08433	34.597	30.503	31.751	32.044	30.59367	33.54233	30.88933	33.60167	30.677	29.12767	29.93033	32.979	30.909
Housekeeping gene														
Ct1	26.172	24.223	26.157	24	26.411	23.556	27.421	23.564	27.626	22.687	26.304	25.657	27.552	23.665
Ct2	26.15	24.356	26.444	24.211	26.615	24.11	28	23.231	27.831	22.135	26.795	25.342	27.892	23.211
Ct3	26.8	24.551	27	24.324	27	23.877	27.976	23.766	28	22.451	27	25.774	28	23.434
	-8.828	-10.545	-4.352	-7.799	-6.038	-6.756	-6.35	-7.173	-6.374	-8.029	-3.072	-4.362	-5.729	-6.967
	-8.793	-10.366	-3.927	-7.334	-5.385	-6.536	-5.323	-7.769	-5.794	-8.758	-2.346	-4.67	-4.823	-7.884
	-8.51	-9.75	-3.629	-7.585	-4.683	-6.946	-5.557	-7.165	-5.18	-7.971	-1.866	-3.986	-4.941	-7.566
	454.457	1494.037	20.42126	222.7065	65.70813	108.0833	81.57188	144.3073	82.94022	261.198	8.409383	20.5633	53.03967	125.1054
	443.5645	1319.705	15.21055	161.3444	41.78751	92.7966	40.02973	218.1233	55.484	432.933	5.084127	25.45717	28.30529	236.2221
	364.5569	861.0779	12.37194	192.005	25.6876	123.2975	47.07862	143.5093	36.25228	250.9055	3.645205	15.84549	30.71774	189.4929
Average	420.8594	1224.94	16.00125	192.0186	44.39441	108.0591	56.22674	168.6466	58.2255	315.0122	5.712905	20.62198	37.35423	183.6068
STDEV	49.0627	326.9474	4.082498	30.68105	20.13722	15.25048	22.23069	42.84993	23.46439	102.252	2.443536	4.806109	13.63744	55.79171

APPENDIX H: QRT-PCR ($2^{-\Delta\Delta CT}$) FOR TISSUE DISTRIBUTION AND ONE-WAY ANOVA WITH POST-HOC TUKEY HSD

	Eye Stalk	Gills	pleopod	stomach	hepato	muscle	intestine	haemocyte
CT1	34.143	37.84	29.755	34.806	36.953	34.477	35.804	28.946
CT2	33.954	39.523	30.188	34.618	38.064	34.601	35.143	28.999
CT3	34.176	38.144	30.122	34.669	36.842	34.928	34.613	29.026
Average	**34.091**	**38.50233**	**30.02167**	**34.69767**	**37.28633**	**34.66867**	**35.18667**	**28.9903333**
	Housekeeping gene							
CT1	32.732	27.181	24.647	25.296	28.503	25.524	29.568	21.403
CT2	33.053	27.109	24.877	25.48	28.446	25.613	29.468	21.511
CT3	33.285	27.057	24.676	25.512	28.337	25.451	29.017	21.763
	1.359	11.32133	5.374667	9.401667	8.783333	9.144667	5.618667	7.58733333
	1.038	11.39333	5.144667	9.217667	8.840333	9.055667	5.718667	7.47933333
	0.806	11.44533	5.345667	9.185667	8.949333	9.217667	6.169667	7.22733333
	1.067666667	**1.067667**	**1.067667**	**1.067667**	**1.067667**	**1.067667**	**1.067667**	**1.06766667**
	-0.291333333	-10.2537	-4.307	-8.334	-7.71567	-8.077	-4.551	-6.51966667
	0.029666667	-10.3257	-4.077	-8.15	-7.77267	-7.988	-4.651	-6.41166667
	0.261666667	-10.3777	-4.278	-8.118	-7.88167	-8.15	-5.102	-6.15966667
	1.22377076	1220.847	19.79412	322.6889	210.207	270.0345	23.44161	91.7519342
	0.979646618	1283.321	16.87716	284.0498	218.6784	253.8795	25.1241	85.1341862
	0.834123745	1330.421	19.40021	277.8187	235.8403	284.0498	34.34433	71.4898573
Average	**1.012513707**	**1278.196**	**18.69049**	**294.8525**	**221.5752**	**269.3213**	**27.63668**	**82.7919926**
STDEV	0.196891806	54.96628	1.582699	24.30753	13.05991	15.09779	5.869591	10.3321026

Eye Stalk

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	99.0700	102.2730	201.3430
mean $\bar{x}\bar{x}$	33.0233	34.0910	33.5572
sum of squares ∑x2i∑xi2	3,271.7759	3,486.6175	6,758.3934
sample variance s2s2	0.0771	0.0143	0.3786
sample std. dev. ss	0.2777	0.1198	0.6153
std. dev. of mean $SE\bar{x}SE\bar{x}$	0.1603	0.0692	0.2512

One-way ANOVA of kk-2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value

treatment	1.7099	1	1.7099	37.3900	0.0036
error	0.1829	4	0.0457		
total	1.8928	5			

Treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	8.6475	0.0036212	** p<0.01

Gills

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	81.3470	115.5070	196.8540
mean x¯x¯	27.1157	38.5023	32.8090
sum of squares ∑x2i∑xi2	2,205.7859	4,448.8979	6,654.6838
sample variance s2s2	0.0039	0.8044	39.2202
sample std. dev. ss	0.0623	0.8969	6.2626
std. dev. of mean SEx¯SEx¯	0.0360	0.5178	2.5567

One-way ANOVA of kk-2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	194.4843	1	194.4843	481.2170	$2.5555e^{-05}$
error	1.6166	4	0.4042		
total	196.1009	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	31.0231	0.0010053	** p<0.01

Pleopod

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	74.2000	90.0650	164.2650
mean x¯x¯	24.7333	30.0217	27.3775
sum of squares ∑x2i∑xi2	1,835.2447	2,704.0103	4,539.2550
sample variance s2s2	0.0157	0.0544	8.4180

sample std. dev. ss	0.1253	0.2333	2.9014
std. dev. of mean SEx¯SEx¯	0.0723	0.1347	1.1845

One-way ANOVA of kk-2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	41.9497	1	41.9497	1,196.6370	$4.1669e^{-06}$
error	0.1402	4	0.0351		
total	42.0899	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	48.9211	0.0010053	** p<0.01

Stomach

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	76.2880	104.0930	180.3810
mean x¯x¯	25.4293	34.6977	30.0635
sum of squares ∑x2i∑xi2	1,939.9802	3,611.8031	5,551.7833
sample variance s2s2	0.0136	0.0095	25.7798
sample std. dev. ss	0.1166	0.0972	5.0774
std. dev. of mean SEx¯SEx¯	0.0673	0.0561	2.0728

One-way ANOVA of kk-2 independent treatments:

source	Sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	128.8530	1	128.8530	11,184.3476	$4.7937e^{-08}$
error	0.0461	4	0.0115		
total	128.8991	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	149.5617	0.0010053	** p<0.01

Hepatopancreas

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑Xi∑xi	85.2860	111.8590	197.1450
mean X¯x¯	28.4287	37.2863	32.8575
sum of squares ∑X2i∑xi2	2,424.5815	4,171.7253	6,596.3068
sample variance S2s2	0.0071	0.4567	23.7230
sample std. dev. Ss	0.0843	0.6758	4.8706
std. dev. of mean SEx¯SEx¯	0.0487	0.3902	1.9884

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	117.6874	1	117.6874	507.5263	$2.2991e^{-05}$
error	0.9275	4	0.2319		
total	118.6149	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	31.8599	0.0010053	** p<0.01

Muscle

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	76.5880	104.0060	180.5940
mean x¯x¯	25.5293	34.6687	30.0990
sum of squares ∑x2i∑xi2	1,955.2537	3,605.8579	5,561.1117
sample variance s2s2	0.0066	0.0543	25.0826
sample std. dev. ss	0.0811	0.2330	5.0083
std. dev. of mean SEx¯SEx¯	0.0468	0.1345	2.0446

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	125.2911	1	125.2911	4,116.9043	$3.5343e^{-07}$
error	0.1217	4	0.0304		
total	125.4129				

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	90.7403	0.0010053	** p<0.01

Intestine

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	88.0530	105.5600	193.6130
mean $\bar{x}\bar{x}$	29.3510	35.1867	32.2688
sum of squares ∑x2i∑xi2	2,584.6159	3,715.0166	6,299.6326
sample variance s2s2	0.0862	0.3561	10.3934
sample std. dev. ss	0.2935	0.5967	3.2239
std. dev. of mean $SE\bar{x}SE\bar{x}$	0.1695	0.3445	1.3161

One-way ANOVA of your kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	51.0825	1	51.0825	231.0290	0.0001
error	0.8844	4	0.2211		
total	51.9669	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	21.4955	0.0010053	** p<0.01

Haemocyte

Treatment →	A	B	Pooled Total
observations N	3	3	6

sum ∑xi∑xi	64.6770	86.9710	151.6480
mean x¯x¯	21.5590	28.9903	25.2747
sum of squares ∑x2i∑xi2	1,394.4397	2,521.3216	3,915.7613
sample variance s2s2	0.0341	0.0017	16.5817
sample std. dev. ss	0.1847	0.0407	4.0721
std. dev. of mean SEx¯SEx¯	0.1067	0.0235	1.6624

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	82.8371	1	82.8371	4,629.7955	$2.7951e^{-07}$
error	0.0716	4	0.0179		
total	82.9086				

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	22.6800	0.0010053	** p<0.01

APPENDIX I: qRT-PCR ($2^{-\Delta\Delta CT}$) FOR WSSV CHALLENGED *M. ROSENBERGII* AND ONE-WAY ANOVA WITH POST-HOC TUKEY HSD

	control	0 hour	control	3 hour	control	6 hour	control	12 hour	control	24 hour	control	48 hour
CT1	32.164	32.272	32.855	29.518	32.955	37.518	31.91	32.796	31.845	32.577	31.798	37.091
CT2	32.001	32.193	32.051	29.841	32.361	33.894	32.341	34.027	32.11	32.64	32.43	37.876
CT3	32.099	32.324	32.089	29.751	32.199	36	32.291	33.105	32.331	34.436	32.599	37.8
Average	32.088	32.263	32.33167	29.70333	32.505	35.804	32.18067	33.30933	32.09533	33.21767	32.27567	37.589
	Housekeeping gene											
CT1	26.971	27.055	26.861	23.232	26.862	28.451	26.971	27.344	26.932	29.149	26.988	33.498
CT2	27	27.068	27	23.204	27.111	28.266	27.21	27.234	27.043	29.217	27.003	33.569
CT3	27.005	26.929	27.105	23.298	27.225	28.721	27.035	27.242	27.115	29.124	27.415	33.25
	-5.117	-5.208	-5.47067	-6.47133	-5.643	-7.353	-5.20967	-5.96533	-5.16333	-4.06867	-5.28767	-4.091
	-5.088	-5.195	-5.33167	-6.49933	-5.394	-7.538	-4.97067	-6.07533	-5.05233	-4.00067	-5.27267	-4.02
	-5.083	-5.334	-5.22667	-6.40533	-5.28	-7.083	-5.14567	-6.06733	-4.98033	-4.09367	-4.86067	-4.339
	34.70328	36.96275	44.34399	88.72897	49.97034	163.4834	37.00547	62.48047	35.83589	16.77995	39.06126	17.04173
	34.01266	36.63117	40.27092	90.46785	42.04901	185.8507	31.35594	67.43068	33.1821	16.0074	38.65724	16.22335
	33.89499	40.33611	37.4441	84.76127	38.85424	135.5799	35.39974	67.0578	31.56674	17.07326	29.05404	20.23807
Average	34.20364	37.97668	40.68634	87.98603	43.62453	161.638	34.58705	65.65632	33.52824	16.6202	35.59085	17.83439
STDEV	0.416093	2.882658	3.46865	2.924934	5.723076	25.18611	2.911127	2.75668	2.155522	0.550597	5.664646	2.12149

W SSV 0H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	102.6109	113.9300	216.5410
mean $x^{-}x^{-}$	34.2036	37.9767	36.0902
sum of squares ∑x2i∑xi2	3,510.0487	4,335.0890	7,845.1377
sample variance s2s2	0.1907	4.2027	6.0281
sample std. dev. ss	0.4367	2.0500	2.4552
std. dev. of mean $SEx^{-}SEx^{-}$	0.2521	1.1836	1.0023

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	21.3537	1	21.3537	9.7209	0.0356
error	8.7867	4	2.1967		

total	30.1404	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	4.4093	0.0356017	* p<0.05

WSSV 3H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	122.0590	263.9581	386.0171
mean x¯x¯	40.6863	87.9860	64.3362
sum of squares ∑x2i∑xi2	4,990.1977	23,241.7360	28,231.9337
sample variance s2s2	12.0315	8.5552	679.4130
sample std. dev. ss	3.4686	2.9249	26.0656
std. dev. of mean SEx¯SEx¯	2.0026	1.6887	10.6412

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	3,355.8914	1	3,355.8914	326.0241	$5.5312e^{-05}$
error	41.1735	4	10.2934		
total	3,397.0650				

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	25.5352	0.0010053	** p<0.01

WSSV 6H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	130.8736	484.9140	615.7876
mean x¯x¯	43.6245	161.6380	102.6313
sum of squares ∑x2i∑xi2	5,774.8055	79,649.1991	85,424.0046
sample variance s2s2	32.7536	634.3402	4,444.9906
sample std. dev. ss	5.7231	25.1861	66.6708
std. dev. of mean SEx¯SEx¯	3.3042	14.5412	27.2182

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	20,890.7656	1	20,890.7656	62.6322	0.0014
error	1,334.1876	4	333.5469		
total	22,224.9532	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	11.1922	0.0013788	** p<0.01

WSSV 12H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	103.7611	196.9690	300.7301
mean x¯x¯	34.5870	65.6563	50.1217
sum of squares ∑x2i∑xi2	3,605.7408	12,947.4550	16,553.1959
sample variance s2s2	8.4747	7.5993	296.0195
sample std. dev. ss	2.9111	2.7567	17.2052
std. dev. of mean SEx¯SEx¯	1.6807	1.5916	7.0240

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	1,447.9494	1	1,447.9494	180.1611	0.0002
error	32.1479	4	8.0370		
total	1,480.0973				

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	18.9822	0.0010053	** p<0.01

WSSV 24H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	100.5847	49.8606	150.4453
mean x¯x¯	33.5282	16.6202	25.0742
sum of squares ∑x2i∑xi2	3,381.7220	829.2997	4,211.0217
sample variance s2s2	4.6463	0.3032	87.7443
sample std. dev. ss	2.1555	0.5506	9.3672
std. dev. of mean SEx¯SEx¯	1.2445	0.3179	3.8241

One-way ANOVA of kk**=2 independent treatments:**

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	428.8228	1	428.8228	173.2816	0.0002
error	9.8989	4	2.4747		
total	438.7217	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	18.6162	0.0010053	** p<0.01

WSSV 48H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	106.7725	53.5032	160.2757
mean x¯x¯	35.5908	17.8344	26.7126
sum of squares ∑x2i∑xi2	3,864.3013	963.1973	4,827.4986
sample variance s2s2	32.0882	4.5007	109.2231
sample std. dev. ss	5.6646	2.1215	10.4510
std. dev. of mean SEx¯SEx¯	3.2705	1.2248	4.2666

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	472.9378	1	472.9378	25.8514	0.0071

error	73.1779	4	18.2945		
total	546.1157	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	7.1905	0.0070582	** p<0.01

APPENDIX J: qRT-PCR ($2^{-\Delta\Delta CT}$) FOR *V. PARAHAEMOLYTICUS* CHALLENGED *M. ROSENBERGII* AND ONE-WAY ANOVA WITH POST-HOC TUKEY HSD

	control	0 hour	control	3 hour	control	6 hour	control	12 hour	control	24 hour	control	48 hour
CT1	30.879	34.458	30.744	24.939	30.844	26.142	30.944	26.691	30.754	27.671	30.742	25.846
CT2	30.789	36	30.799	25.003	30.91	26.562	30.599	27.582	30.815	27.995	30.797	25.783
CT3	30.81	34.553	30.698	24.413	30.898	26.483	30.713	27.57	30.798	27.999	30.696	25.688
Average	30.826	35.00367	30.747	24.785	30.884	26.39567	30.752	27.281	30.789	27.88833	30.745	25.77233
	Housekeeping gene											
CT1	25.444	29.32	25.744	20.956	25.678	20.584	25.754	20.556	25.644	20.156	25.734	20.579
CT2	25.408	30.81	25.808	20.455	25.758	20.197	25.708	20.621	25.728	20.297	25.828	20.528
CT3	25.431	29.6	25.531	21.026	25.433	20.279	25.532	20.516	25.529	20.443	25.511	20.561
	-5.435	-5.138	-5	-3.983	-5.166	-5.558	-5.19	-6.135	-5.11	-7.515	-5.008	-5.267
	-5.381	-5.19	-4.991	-4.548	-5.152	-6.365	-4.891	-6.961	-5.087	-7.698	-4.969	-5.255
	-5.379	-4.953	-5.167	-3.387	-5.465	-6.204	-5.181	-7.054	-5.269	-7.556	-5.185	-5.127
	43.26115	35.21212	32	15.81257	35.90219	47.11126	36.50444	70.27794	34.5353	182.9112	32.17794	38.5057
	41.67181	36.50444	31.80099	23.39292	35.55548	82.42442	29.67138	124.5862	33.98909	207.6486	31.31973	38.18674
	41.61408	30.9743	35.92709	10.46137	44.17016	73.72081	36.27742	132.8818	38.55911	188.184	36.37814	34.94466
Average	42.18235	34.23029	33.24269	16.55562	38.54261	67.75216	34.15108	109.2486	35.6945	192.9146	33.29194	37.21237
STDEV	0.934712	2.892851	2.32688	6.497717	4.876681	18.39765	3.881196	34.00355	2.495812	13.02949	2.706959	1.970358

Vibrio 0H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	126.5470	102.6909	229.2379
mean $\bar{x}\bar{x}$	42.1823	34.2303	38.2063
sum of squares ∑x2i∑xi2	5,339.7989	3,531.8747	8,871.6736
sample variance s2s2	0.8737	8.3686	22.6675
sample std. dev. ss	0.9347	2.8929	4.7610
std. dev. of mean $SE\bar{x}SE\bar{x}$	0.5397	1.6702	1.9437

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value

treatment	94.8529	1	94.8529	20.5259	0.0106
error	18.4845	4	4.6211		
total	113.3375	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	6.4072	0.0105714	* p<0.05

Vibrio 3H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	99.7281	49.6669	149.3949
mean x¯x¯	33.2427	16.5556	24.8992
sum of squares ∑x2i∑xi2	3,326.0588	906.7063	4,232.7651
sample variance s2s2	5.4144	42.2203	102.5914
sample std. dev. ss	2.3269	6.4977	10.1287
std. dev. of mean SEx¯SEx¯	1.3434	3.7515	4.1350

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	417.6876	1	417.6876	17.5371	0.0138
error	95.2694	4	23.8173		
total	512.9570	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	5.9224	0.0138330	* p<0.05

Vibrio 6H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	115.6278	203.2565	318.8843
mean x¯x¯	38.5426	67.7522	53.1474
sum of squares ∑x2i∑xi2	4,504.1621	14,448.0142	18,952.1763
sample variance s2s2	23.7820	338.4734	400.8616
sample std. dev. ss	4.8767	18.3976	20.0215

std. dev. of mean SEx¯SEx¯	2.8156	10.6219	8.1738

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	1,279.7972	1	1,279.7972	7.0657	0.0565
error	724.5109	4	181.1277		
total	2,004.3081	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	3.7592	0.0565011	insignificant

Vibrio 12H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	102.4532	327.7459	430.1992
mean x¯x¯	34.1511	109.2486	71.6999
sum of squares ∑x2i∑xi2	3,529.0160	38,118.2799	41,647.2959
sample variance s2s2	15.0637	1,156.2412	2,160.4152
sample std. dev. ss	3.8812	34.0035	46.4803
std. dev. of mean SEx¯SEx¯	2.2408	19.6320	18.9755

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	8,459.4660	1	8,459.4660	14.4445	0.0191
error	2,342.6098	4	585.6525		
total	10,802.0758				

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	5.3749	0.0190923	* p<0.05

Vibrio 24H

Treatment →	A	B	Pooled Total
observations N	3	3	6

sum ∑xi∑xi	107.0835	578.7438	685.8273
mean x¯x¯	35.6945	192.9146	114.3045
sum of squares ∑x2i∑xi2	3,834.7510	111,987.6564	115,822.4074
sample variance s2s2	6.2291	169.7676	7,485.8456
sample std. dev. ss	2.4958	13.0295	86.5208
std. dev. of mean SEx¯SEx¯	1.4410	7.5226	35.3220

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	37,077.2349	1	37,077.2349	421.3403	$3.3269e^{-05}$
error	351.9933	4	87.9983		
total	37,429.2282	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	29.0290	0.0010053	** p<0.01

Vibrio 48H

Treatment →	A	B	Pooled Total
observations N	3	3	6
sum ∑xi∑xi	99.8758	111.6371	211.5129
mean x¯x¯	33.2919	37.2124	35.2522
sum of squares ∑x2i∑xi2	3,339.7147	4,162.0452	7,501.7599
sample variance s2s2	7.3276	3.8823	9.0949
sample std. dev. ss	2.7070	1.9704	3.0158
std. dev. of mean SEx¯SEx¯	1.5629	1.1376	1.2312

One-way ANOVA of kk=2 independent treatments:

source	sum of squares SS	degrees of freedom νν	mean square MS	F statistic	p-value
treatment	23.0546	1	23.0546	4.1133	0.1125
error	22.4199	4	5.6050		
total	45.4745	5			

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
A vs B	2.8682	0.1124527	insignificant