

Research Article

A preliminary study on differential gene expression in hemocyte of the mantis shrimp *Harpiosquilla raphidea* by next generation sequencing technique

Naiyasit Sodsee¹, Sarawut Wongphayak², Hidehiro Kondo³, Ikuo Hirono³ and Rachanimuk Hiransuchalert^{1*}

¹Faculty of Marine Technology, Burapha University, Chanthaburi Campus, Chanthaburi 22170, Thailand

²Vishuo Biomedical (Thailand) Ltd. 17th Floor Alma Link Building, Bangkok 10330 Thailand

³Laboratory of Genome Science, Tokyo University of Marine Science and Technology, Tokyo, Japan

*E-mail: rachanimuk@buu.ac.th

Abstract

This study aims to determine gene expression profiling in hemocyte of *Harpiosquilla raphidea* by using bioinformatics technology and Next Generation Sequencing (NGS) technique. The *De novo* RNA-sequencing was performed using Illumina HiSeq2000. In total, 93,344 assembly consensus sequences were identified by using Trinity software. These 93,344 transcripts were estimated the expression level of each transcript in the hemocyte of *H. raphidea* (both infected by *Vibrio parahemolyticus* for 24 hours and normal condition) in terms of FPKM (Frackments Per Kilobase of target transcript length per Million reads mapped). The 219 differential expression of transcripts were classified into 4 clusters, i.e., down regulated expression levels (cluster 1/1 and 2/1) and up regulated expression levels (cluster 1/2 and 2/2). Clusters were further functional and annotation analyzed by using Blast2GO software. The important genes that relevant in the immune system of *H. raphidea* were found this study. These included histone H1-delta-like, Cyclooxygenases, 5-aminolevulinat erythroid-mitochondrial, electron transfer flavo-ubiquinone mitochondrial and CCR4-NOT transcription complex subunit 7 isoform X1 gene. The basic information related to these immune genes in *H. raphidea* was investigated in this study and will be the guidance for further studies on the important process in the immune system of *H. raphidea* and for its commercial culture in the future.

Keywords: *Harpiosquilla raphidea*, NGS, RNA-seq, *de novo*, transcriptome assembly

Introduction

Recently, mantis shrimps are popular for food consumption especially in Asia. *Harpiosquilla raphidea* is one of the commercial mantis shrimp species. *H. raphidea* become more desirable food for customers (Tangkrock-olan & Ruksar, 2012), however, most of *H. raphidea* is not come from aquaculture. These affect natural populations of *H. raphidea* because of the over fishing and ore potential fishing gear were used.

Therefore, Department of Fisheries tries to culture mantis shrimps aiming to increase the natural population and commercially culture of mantis shrimp (Srimukda & Champasri, 1994). Unfortunately, mantis shrimps culture still unsuccessful cause of lacking in some biological knowledge, i.e., immune system and reproductive system.

Next Generation Sequencing (NGS) technology enables us to identify known as well as novel gene (Buermans & den Dunnen, 2014) and multiomics approaches including interactome sequencing (Ohashi et al., 2015). We have demonstrated the NGS technology for *de novo* assembly and annotation of genes without prior genome information, as well as its reliability and complementarily by alignment with sequence database. Moreover, it also can apply this technique for the study of gene expression level in transcription level under physical state or diseases. RNA-seq maybe covered part of mRNA in genome for gene expression analysis (Wilantho et al., 2012). NGS technology can investigate gene structure and gene function by complementary DNA (cDNA), cDNA synthesis from RNA that call genetic data is RNA-seq data, therefore the study of *H. raphidea* in reproductive and immune organ by bioinformatics tools to know novel transcriptome of *H. raphidea* for gene expression in importance organ. These will be the useful information for commercial culture and increase nature population of *H. raphidea* in the future.

Materials and methods

Samples

The wild *H. raphidea* were derived from Trad province, Thailand and separated by control (normal condition; N=2) and treatment (infection by 10^4 CFU of *Vibrio parahemolyticus* at 24 hours; N=2). The hemolymphs were collected from all samples and centrifuge at 3000g for 10 minutes. The hemocytes were kept to 10 mL of RNA later by incubated in 4 °C for 12 hours after that incubated in -80 °C until used.

Total RNA extraction

All hemocytes samples (approximately 0.5 g per each sample) were transferred to a microcentrifuge tube containing 500 μ L of TRI REAGENT[®] and homogenized. Additional 500 μ l of TRI REAGENT[®] were added. The homogenate were left for 5 minutes, before adding 0.2 ml of chloroform. The homogenate was vortexed for 15 seconds and left at room temperature for 2 – 15 minutes and centrifuged at 12000 g for 15 minutes at 4 °C. RNA was precipitated by an addition of 0.5 ml of isopropanol and washed with 1 ml of 75% ethanol. RNA was dissolved in DEPC-treated H₂O for immediately used. Alternatively, the RNA pellet was kept under absolute ethanol in a -80 °C freezer for long storage.

Library construction and sequencing

RNA-seq library preparation and sequencing was carried out by Professor Dr. Hidehiro Kondo and Professor Dr. Ikuo Hirono, Laboratory of Genome Science Tokyo University of Marine Science and Technology, Japan. Total RNA of each sample was quantified and qualified by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), NanoDrop (Thermo Fisher Scientific Inc.) and 1% agarose gel. 1 μ g total RNA with RIN value above 7 was used for following library preparation.

Next generation sequencing library preparations were constructed according to the manufacturer's protocol (NEBNext® Ultra™ RNA Library Prep Kit for Illumina®). The poly(A) mRNA isolation was performed using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). The mRNA fragmentation and priming was performed using NEBNext First Strand Synthesis Reaction Buffer and NEBNext Random Primers. First strand cDNA was synthesized using ProtoScript II Reverse Transcriptase and the second-strand cDNA was synthesized using Second Strand Synthesis Enzyme Mix. The purified (by AxyPrep Mag PCR Clean-up (Axygen)) double-stranded cDNA was then treated with End Prep Enzyme Mix to repair both ends and add a dA-tailing in one reaction, followed by a T-A ligation to add adaptors to both ends. Size selection of Adaptor-ligated DNA was then performed using AxyPrep Mag PCR Clean-up (Axygen), and fragments of ~360 bp (with the approximate insert size of 300 bp) were recovered. Each sample was then amplified by PCR for 11 cycles using P5 and P7 primers, with both primers carrying sequences which can anneal with flow cell to perform bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR products were cleaned up using AxyPrep Mag PCR Clean-up (Axygen), validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) , and quantified by Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA).

Then libraries with different indices were multiplexed and loaded on an Illumina HiSeq instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2x150bp paired-end (PE) and 10 million reads per sample configuration; image analysis and base calling were conducted by the HiSeq Control Software (HCS) + RTA 2.7 (Illumina) on the HiSeq2000 instrument.

Gene differential expression

Illumina build RNA-seq data like a mass nucleotide reads that made a lot of same length reads. *De novo* transcriptome assembly made by include RNA-seq data in to Trinity software that contain 3 step first Inchworm for connect reads to contig that can reconstructs linear transcripts, Chrysalis builds de Bruijn transcript graphs of cluster minimally-overlapping Inchworm contigs into sets of connected components, and constructs complete de Bruijn graphs for each component, Butterfly reconstructs plausible full-length, linear transcripts by reconciling the individual de Bruijn graphs generated by Chrysalis with the original reads and paired-ends.

The expression levels (FPKM, Frackments Per Kilobase of target transcript length per Million reads mapped) were compared in each transcript by R studio software that generated gene expression profile, first expunge the transcripts that have FPKM value = 0 in treatment ro control, convert FPKM values to \log_{10} format, estimate standard deviation of each transcript between treatment and control, expunge the transcripts that have standard deviation of each transcript between treatment and control less than 2, cluster transcripts depend on expression structure by gPlots package (function heatmap = 2).

Transcripts were cluster by R studio software not contain nucleotide in each transcript. Before gene annotation analysis transcripts were match with nucleotide sequences and saved with another transcript in same cluster in fasta format.

Function and annotation of transcripts made in each cluster by blast2go software (default setting), follow by 1.) BlastX towards transcripts without of intron, 2.) Map transcripts with all kinds of protein database in blast2go, 3.) Gene ontology with GO database and 4.) Pathway identification from Kyoto Encyclopedia of Genes and Genomes Pathway database (KEGG).

Results

Gene expression profile

In total, 93,344 transcript sequences were identified by *De novo* transcriptome assembly, after removing disreputable data left 28,323 sequences, after compare standard deviation in log₁₀ format between normal and 24 hours infection of *Vibrio parahaemolyticus* get 214 sequences that have standard deviation more than 2. Gene expression profile displayed in heatmap that contain two main clusters and each cluster contain two sub clusters. The difference of the main cluster 1 and 2 are that cluster 1 contained a low differentially level between normal and *Vibrio parahaemolyticus*. Whereas, cluster 2 contained a high differentially level between normal and *Vibrio parahaemolyticus* infection. By each sub cluster, they contained down regulated cluster (/1) and up regulated cluster (/2), respectively (Figure 1).

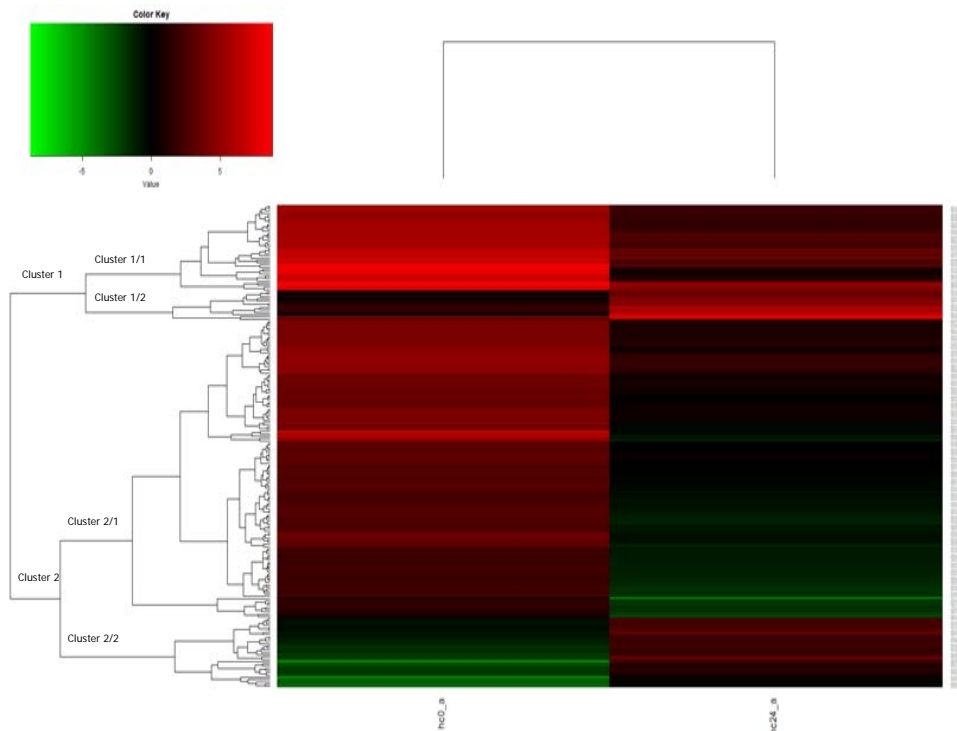


Figure 1. Gene expression profile of *H. raphidea* hemocyte between normal condition (hc0_a) and 24 hours infection of *Vibrio parahaemolyticus* (hc24_a).

Functional annotation of genes

All of 219 transcripts were divided to two main clusters, one main cluster contains two subcluster that up regulated and down regulated expression. After blast to GenBank using BlastX, 68 transcripts were identified as known transcript. Moreover, each cluster was specified the gene functional and annotation in biological process, molecular function and cellular component. Gene function and annotation was specified by gene ontology format from GO database. Down regulated expression while infection of *Vibrio parvhemolyticus* condition classified in cluster 1/1 and cluster 2/1. Whereas, up regulated expression while infection of *Vibrio parvhemolyticus* condition classified in cluster 1/2 and cluster 2/2 (Table 3).

Table 1. Gene function of down regulated expression transcripts in clusters 1/1 and 2/1.

description	biological process (level 2)	molecular function (level 2)	cellular component (level 2)
cluster 1/1			
1. histone H1-delta-like	cellular process, cellular component	binding	macromolecular complex, organelle, cell part, cell, organelle part
2. histone H1-delta-like	cellular process, cellular component	binding	macromolecular complex, organelle, cell part, cell, organelle part
3. venom toxin	regulation of biological process, biological regulation, growth, cellular process, cellular component, single-organism process	binding	macromolecular complex, extracellular region part, extracellular region
4. 40S ribosomal S30	cellular process, cellular component, metabolic process (translation)	binding, structural molecule activity	macromolecular complex, organelle, cell part, cell
5. histone H1-delta-like	cellular process, cellular component	binding	macromolecular complex, organelle, cell part, cell, organelle part
cluster 2/1			
1. crustacean hyperglycemic hormone	cellular process, single-organism process, regulation of biological process, biological regulation, signaling, response to stimulus	binding	extracellular region
2. 5-aminolevulinate erythroid-mitochondrial	metabolic process, cellular process, single-organism process	binding, catalytic activity	membrane-enclosed lumen, cell part, cell, organelle part, organelle
3. venom toxin	growth, cellular component, cellular process, single-organism process, regulation of biological process, biological regulation	binding	macromolecular complex, extracellular region part, extracellular region
4. bumetanide-sensitive sodium-(potassium)-chloride cotransporter isoform X1	localization	-	membrane
5. tyrosine decarboxylase	metabolic process, cellular process, single-organism process	binding, catalytic activity	-
6. 5-aminolevulinate erythroid-mitochondrial	metabolic process, cellular process, single-organism process	binding, catalytic activity	membrane-enclosed lumen, cell part, cell, organelle part, organelle
7. low-density lipo receptor	cellular process, single-organism process, regulation of biological process, biological regulation, signaling, response to stimulus	binding	-
8. 70 kDa heat shock	-	binding	-
9. crustin-like peptide type	-	molecular function regulator	extracellular region
10. Crustin antimicrobial peptide	-	molecular function regulator	extracellular region
11. cryptochrome-1-like isoform X2	-	catalytic activity	-
12. programmed cell death 5	-	binding	-
13. Ovalbumin-related Y	-	-	extracellular region part, extracellular region

Table 2. Gene function of up regulated expression transcripts in clusters 1/2 and 2/2.

description	biological process (level 2)	molecular function (level 2)	cellular component (level 2)
cluster 1/2			
1. ccaat enhancer-binding	regulation of biological process, biological regulation, metabolic process, cellular process	nucleic acid binding, binding	macromolecular complex, cell part, cell
2. cyclooxygenase	detoxification, multicellular organismal process, single-organism process, biological regulation, metabolic process, cellular process response to stimulus	antioxidant activity, binding, catalytic activity	membrane part, membrane
3. phosphoenolpyruvate carboxykinase	single-organism process, metabolic process	binding, catalytic activity	-
4. MAP kinase-interacting serine threonine- kinase 1 isoform X1	single-organism process, metabolic process, cellular process	binding, catalytic activity	-
5. elongation factor 1- alpha	growth, locomotion, multicellular organismal process, single-organism process, developmental process, regulation of biological process, biological regulation, metabolic process, cellular process, reproduction	binding, catalytic activity	macromolecular complex, cell part, cell, organelle
6. Phosphoenolpyruvate carboxykinase	single-organism process	binding, catalytic activity	-
7. prophenoxidase	single-organism process, metabolic process	binding, catalytic activity	-
8. serum amyloid A -like	-	-	extracellular region
cluster 2/2			
1. CCR4-NOT transcription complex subunit 7 isoform X1	biological regulation, metabolic process, cellular process, regulation of biological process	catalytic activity, binding	-
2. 39S ribosomal mitochondrial-like	metabolic process, cellular process, cellular component	structural molecule activity	cell part, cell, macromolecular complex, organelle
3. transforming growth factor-beta receptor-associated 1-like	biological regulation, localization, cellular process, single-organism process, regulation of biological process	binding	cell part, cell
4. hypothetical protein Y032_0655g1201	biological regulation, metabolic process, cellular process, regulation of biological process	binding	-
5. 5-3 exoribonuclease 2 homolog	biological regulation, metabolic process, cellular process, regulation of biological process	catalytic activity, binding	cell part, cell, organelle
6. electron transfer flavo - ubiquinone mitochondrial	metabolic process, cellular process, single-organism process, response to stimulus	catalytic activity, electron carrier activity, binding	membrane, organelle part, cell part, cell, organelle, membrane part
7. PREDICTED: uncharacterized protein KIAA2013-like	-	-	membrane, membrane part
8. serine ase inhibitor	-	-	extracellular region part, extracellular region

Table 3. The summary of transcripts by step after functional annotation of genes using blast2go software.

Clusters	Total (transcripts)	BlastX (transcripts)	Unknown / <i>De novo</i> (transcripts)	Annotated (transcripts)	Pathway (transcripts)
Cluster 1/1	18	9	4	5	-
Cluster 1/2	34	9	1	8	2
Cluster 2/1	135	34	21	13	2
Cluster 2/2	32	16	8	8	-
Total	219	68	34	34	4

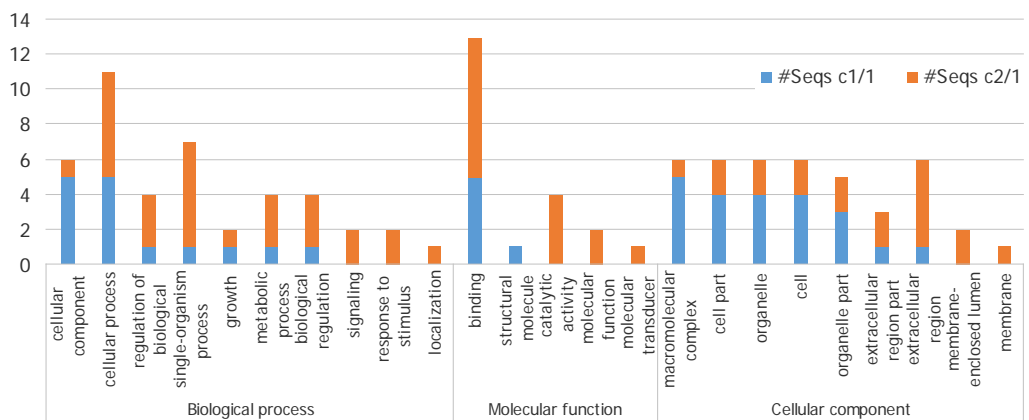


Figure 2. Gene ontology and annotation result (level 2) of Cluster 1/1 and 2/1 showing the number of down regulated expression transcripts while infection of *Vibrio parhemyticus* at 24 hours.

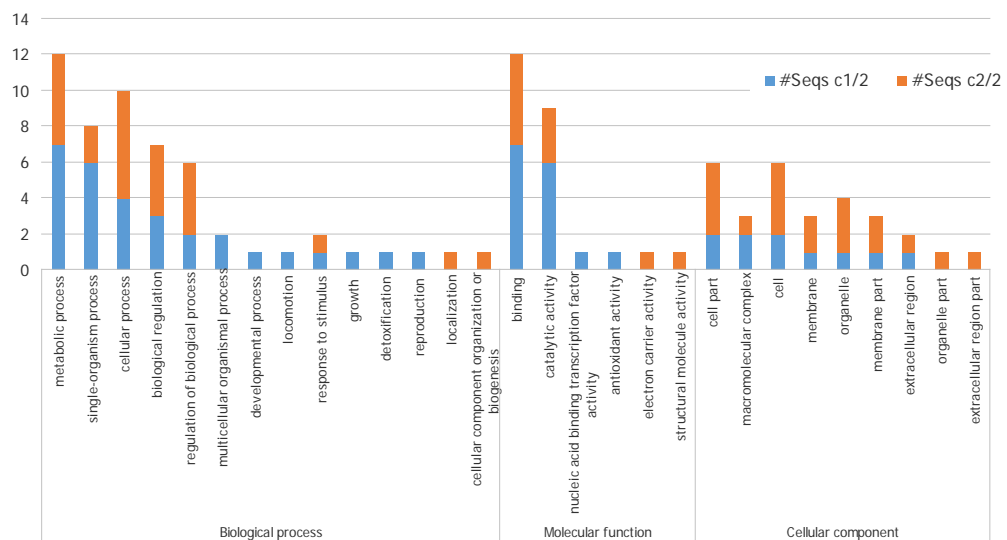


Figure 3. Gene ontology and annotation result (level 2) of Cluster 1/2 and 2/2 showing the number of up regulated expression transcripts while infection of *Vibrio parhemyticus* at 24 hours.

Pathway analysis

Six pathways were identified from cluster 1/2 (4 pathways) and cluster 2/1 (2 pathways). Cluster 1/2 contained Arachidonic acid metabolism, Phenylpropanoid biosynthesis, Thiamine metabolism, Purine metabolism. While Cluster 2/1 contained Glycine, serine and threonine metabolism and Porphyrin and chlorophyll metabolism (table 4).

Table 4. The metabolism pathway identified from the up regulation transcripts (cluster 1/2) and the down regulation transcripts (cluster 2/1).

Seq	Pathway	Pathway ID	#Enzs in Pathway	Enzyme
Cluster 1/2				
TR33858 c0_g1_i1	Arachidonic acid metabolism, Phenylpropanoid biosynthesis	map00590, map00940	1, 1	ec:1.14.99.1 - synthase, ec:1.11.1.7 - lactoperoxidase
TR4161 c0_g1_i1	Thiamine metabolism, Purine metabolism	map00730, map00230	1, 1	ec:3.6.1.15 - phosphatase, ec:3.6.1.15 - phosphatase
Cluster 2/1				
TR31322 c0_g1_i1	Glycine, serine and threonine metabolism, Porphyrin and chlorophyll metabolism	map00260, map00860	1, 1	ec:2.3.1.37 - synthase, ec:2.3.1.37 - synthase
TR31322 c0_g2_i1	Glycine, serine and threonine metabolism, Porphyrin and chlorophyll metabolism	map00260, map00860	1, 1	ec:2.3.1.37 - synthase, ec:2.3.1.37 - synthase

Discussion

Nucleotide data analysis in hemocyte of *H. raphidea* by Next Generation Sequencing (NGS) technology and bioinformatic tools study in different condition between normal *H. raphidea* and 24 hours infected *Vibrio parahaemolyticus* that reconstruct transcripts of *H. raphidea* without referent genome to know more biological data of *H. raphidea* that importance for commercial culture of *H. raphidea* in future. This study found importance gene that concerned in immune system of *H. raphidea*.

Histone H1-delta-like transcript was found in cluster 1/1 down regulated while infection of *Vibrio parahaemolyticus* condition Transcripts (TR8251|c0_g3_i1, TR51934|c0_g1_i1 and TR30957|c0_g1_i1). Histone is the protein in chromatin structure that contains core histones (H2A, H2B, H3, H4) and linker histones (H1, H5). Linker histone H1 link DNA between nucleosome and compact chromatin (Hergeth & Schneider, 2015). Accordingly, Nam et al. (2012) studied on antimicrobial function of histone H1-like protein in olive flounder (*Paralichthys olivaceus*), fish lived in microbe-rich environment and exposed to high concentration of pathogenic microbes. To protect the pathogenic microbes, most of fish can protect them from antimicrobial peptides (AMPs). AMPs are key mediator of innate immunity particularly from primary

barriers such as the skin, mucus, gills, and intestinal mucus. These include pardaxin, pleurocidin, piscidin, misgurin, bass hepcidin, myxinidin. All of AMPs are peptides. In addition, some small protein has antimicrobial property such as histone-like proteins (HLPs). That can specify by N- or C- terminus (fish histone). From Nam et al. (2012) study by compare between antimicrobial proteins and histone H1 that represent antimicrobial property, positive gram contain *Bacillus subtilis*, *Staphylococcus aureus*, *Streptococcus iniae*, negative gram contain *Aeromonas hydrophila*, *Escherichia coli* D31, *Vibrio parahaemolyticus*, *Candida albicans*. cDNA cloning and tissue distribution of histone H1 from olive flounder expressed depend on development stage so anyway without histone H1 in other immunity organ such as kidney and spleen and represent histone H1 importance play in innate immunity and reproduction stage.

Cyclooxygenase transcript was found in cluster 1/2 up regulated while infection of *Vibrio parahaemolyticus* condition. Varvas et al. (2009) studied on cyclooxygenase pathway of prostaglandin synthesis in arthropods. The cloning of cyclooxygenase (COX) gene from *Gammarus* spp. and *Caprella* spp., amphipod organisms, represent COX protein from amphipods role are N-glycosylate and same thin position in endoplasmic reticulum and nuclear envelope of vertebrate organism. COX activate arachidonic acid to prostaglandin are lipid mediator in vertebrate organism and play in reproductive system, immune system and ion transport of arthropods including crustacean *Daphnia* (Heckmann et al., 2008).

Conclusion

The basic information related to these immune genes in *H. raphidea* was investigated in this study. These transcripts had never been reported and studied. A better understanding of their biochemistry and genetic control in *H. raphidea* will help to improve our understanding of their information for commercial culture and increase nature population of these organisms in the future.

Gene expression value file and nucleotide sequences file are separated. After clustering gene expression profiles in each cluster was merge with nucleotide sequences by developed GUI software for this thing. Those develop by Mr. Suphachip boonta, a bachelor degree student in major information technology, faculty of science and arts, Burapha University in visual studio platform. Mention in searching principle between transcript names in each cluster and transcript names in nucleotide sequences file (fasta files) and export files in fasta format. That working well than searching by "VLOOKUP" function in Microsoft excel because of too more nucleotide sequences (93,344 row).

Acknowledgements

A research grant was supported by National Research Council of Thailand (research ID 10/2560). A travel grant was supported by Faculty of Marine Technology, Burapha University, Chanthaburi Campus.

References

- Buermans, H. P. & den Dunnen, J. T. (2014). Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta*, 1842, 1932–1941. doi: 10.1016/j.bbadis.2014.06.015

- Collart, M. A. & Panasenko, O. O. (2012). The Ccr4-Not complex. *Gene*, 492, 42-53. doi: 10.1016/j.gene.2011.09.033
- Heckmann, L. H., Sibly, R. M., Timmermans, M. & Callaghan, A. (2008). Outlining eicosanoid biosynthesis in the crustacean *Daphnia*. *Frontiers in Zoology*, 5(11), 1-9. doi: 10.1186/1742-9994-5-11
- Hergeth, S. P. & Schneider, R. (2015). The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Reports*, 16(11), 1439-1453. doi: 10.15252/embr.201540749
- Hunter, G. A. & Ferreira, G. C. (2011). Molecular enzymology of 5-Aminolevulinic synthase, the gatekeeper of heme biosynthesis. *Biochimica et Biophysica Acta*, 1814, 1467-1473. doi: 10.1016/j.bbapap.2010.12.015
- Nam, B. H., Seo, J. K., Go, H. J., Lee, M. J., Kim, Y. O., Kim, D. G., et al. (2012). Purification and characterization of an antimicrobial histone H1 like protein and its gene from the testes of olive flounder, *Paralichthys olivaceus*. *Fish and Shellfish Immunology*, 33, 92-98. doi: 10.1016/j.fsi.2012.04.006
- Ohashi, H., Hasegawa, M., Wakimoto, K. & Miyamoto-Sato, E. (2015). Next-generation sequencing technologies for multiomics approaches including interactome sequencing. *BioMed Research International*, Article ID 104209, 9 pp. <http://dx.doi.org/10.1155/2015/104209>
- Srimukda, B. & Champasri, S. (1994). *Preliminary study on breeding and nursing of mantis shrimp, Harpiosquilla raphidea (Fabricius, 1798)*. pp. 453-459. In The 32nd Kasetsart University Annual Conference: Animal Science, Veterinary Science and Fisheries, 3-5 February 1994, Kasetsart University, Bangkok. (In Thai)
- Tangkrock-olan, N. & Ruksar, P. (2012). Effect of salinity on blood osmolality of mantis shrimp, *Miyakea nepa* (Latreille, 1828). *Burapha Science Journal*, 17(1), 50-58. (In Thai)
- Varvas, K., Kurg, R., Hansen, K., Järving, R., Järving, L., Valmsen, K., et al. (2009). Direct evidence of the cyclooxygenase pathway of prostaglandin synthesis in arthropods: Genetic and biochemical characterization of two crustacean cyclooxygenase. *Insect Biochemistry and Molecular biology*, 39, 851-860. doi: 10.1016/j.ibmb.2009.10.002
- Watmough, N. J. & Frerman, F. E. (2010). The electron transfer flavoprotein: Ubiquinone oxidoreductases. *Biochimica et Biophysica Acta*, 1797, 1910-1916. <https://doi.org/10.1016/j.bbabi.2010.10.007>
- Waiho, K., Fazhan, H., Shahreza, M. S., Moh, J. H. Z., Noorbaiduri, S., Wong, L. L., et al. (2017). Transcriptomy analysis and differential gene expression on the testis of organ mud crab, *Scylla olivacea*, during sexual maturation. *PLOS ONE*, 12(1), e0171095. doi: 10.1371/journal.pone.0171095
- Wilantho, A., Praditsup, O., Charoenchim, W., Kulawonganchai, S., Assawamakin, A. & Tongsimma, S. (2012). Next generation sequencing (NGS) technologies and their applications in omics-research. *Thai Journal of Genetics*, 5(2), 104-129. (In Thai)