Identification of Long Noncoding RNAs Associated to White Spot Syndrome Virus Infection in White Shrimp, *Litopenaeus vannamei*

Ifwa Wirasit[1], Apinunt Udomkit[2], and Ponsit Sathapondecha[1, *]

[1]Center for Genomics and Bioinformatics Research, Division of Biological Science, Faculty of Science, Prince of Songkla University, Songkla, Thailand
[2]Institute of Molecular Bioscience, Mahidol University, Salaya Campus, Nakhon Pathom, Thailand
[*]Corresponding author, E-mail: ponsit.sat@gmail.com

Abstract

Long non-coding RNA (lncRNA) is a noncoding RNA that lacks of the capability for protein coding, usually over 200 nucleotides in length. It is reported to play significant roles in various physiological processes, including immune regulation, reproduction and development. Although several transcriptomes have been studies in response to viral infections in many organisms, the role of lncRNAs in viral responses has not been elucidated in shrimps. Therefore, this study aimed to identify and classify putative lncRNAs related to white spot syndrome virus (WSSV) infection in white shrimp. The RNA sequencing data of gills from WSSV infection experiment were used to identify lncRNAs. Among 226,797 of *de novo* assembled transcripts, 3,235 transcripts were differentially expressed upon WSSV infection, and 1,628 of them were identified as putative lncRNAs. At 24 h post-WSSV infection, 692 putative lncRNAs were highly expressed, while expressions of the rest 936 lncRNAs were decreased in gills of *L. vannamei*. Furthermore, the differentially expressed lncRNAs were classified based on their location on genome. The 860 and 191 of putative lncRNAs were sense and antisense lncRNAs, respectively, while 15 and 144 of them were promoter-and 3′ UTR-associated lncRNAs, respectively. Other 336 and 82 lncRNAs were classified as intronic and intergenic lncRNAs, respectively. The findings revealed an association between *L. vannamei* lncRNAs and WSSV infection.

Keywords: *long noncoding RNA, Litopenaeus vannamei, White spot syndrome virus*

1. Introduction

Long noncoding RNAs (lncRNAs) are transcripts over 200 nucleotides in length, which are a large and functionally diverse class among noncoding RNAs (ncRNAs) or RNAs that lack of the capability to encode for proteins (Aliperti et al., 2021; Bridges et al., 2021; Kazimierczyk et al., 2020). The significance of lncRNAs as a new biomolecule become greater as there are many reports about the lncRNAs engage in many biological processes and play regulatory functions in transcription, translation, splicing, protein localization, imprinting, cellular structure integrity, cell cycle and apoptosis, stem cell pluripotency and reprogramming, and heat shock response (Aliperti et al., 2021; Ma et al., 2013). LncRNAs have been found to play important roles in immune system of organisms. For instance, in human, the lncRNA *Morrbid*, *HOTAIRM1* or *lnc-DC* involved in the development and differentiation of dendritic and myeloid cells8 (Kotzin et al., 2016; Zhuang et al., 2018). Nevertheless, studies of immune-related lncRNAs in invertebrate species are still scarce.

White shrimp (*Litopenaeus vannamei*), one of most economically important crustaceans has been facing to deadly pathogens so far (Gucic et al., 2013). White spot syndrome virus (WSSV) is a double-stranded DNA virus causing white spot disease in a wide range of crustaceans including white shrimp, and it can spread very quickly in saline waters which tends to lead the 100% mortality in a shrimp pond (Leu et al., 2008; Santos et al., 2018). In white shrimp, the number of transcriptome study has been gradually increased as it is an effective way to analyze number of genes expressed in organisms under certain conditions (Casamassimi et al., 2017; Lowe et al., 2017). In particular, identification and functional study of genes involved in immune system of the host against the infection of pathogens could be beneficial to provide the key for accurate therapeutic treatments or any other possible solutions. In recent, transcriptome

Published online: Copyright © 2016-2023 Rangsit University

studies have uncovered the alteration of gene expression patterns in *L. vannamei* upon WSSV infection in various conditions. Millard et al. (2021) discovered amounts of mRNAs and miRNAs which were differentially expressed following WSSV infection. However, the lncRNA transcript profiles induced by WSSV infection have not been characterized so far.

In this study, identification of lncRNAs associated with WSSV infection in white shrimps was performed, and differential expressions of lncRNAs were investigated. Herein, this research also classified types of the putative lncRNAs depending on their localization on a genome. This study provided novel insights into lncRNAs related to antiviral immune defense of invertebrate.

2. Objectives

To identify lncRNAs differentially expressed in *L. vannamei* against WSSV infection

3. Materials and Methods

3.1 Transcriptome assembly and gene differential expression analysis

The RNA seq data derived from gills of WSSV-challenged and control shrimps at 24 h post-infection was retrieved from NCBI database (BioProject No. PRJNA716175; Millard et al., 2021). The raw reads were examined for their qualities by FastQC program, and their adaptors and low quality sequence were removed using Trimmomatic program (Bolger et al., 2014). The clean reads were used to generate *de novo* assembled transcriptome using Trinity package (version 2.9.1) with default parameters (Grabherr et al., 2011). After using CD-HIT-EST program to remove duplicate sequences with 95% identity, the quality of the assembled transcriptome was determined by TrinityStat and BUSCO program. The read sets were analyzed for differentially expressed gene (DEG) by mapping to the transcriptome and estimating transcript abundance using Bowtie2 (Langmead & Salzberg, 2012) and RSEM (Li & Dewey, 2011), respectively. Next, DESeq2 package was used for identifying DE features with default parameters of FDR < 0.01, p-value < 0.05, and $|\log2(\text{fold-change})| > 1$ (Love et al., 2014).

3.2 *In silico* identification and classification of long noncoding RNA related to WSSV infection in *L.vannamei* gills

To identify putative lncRNAs, the DEG was in-house analyzed by BLASTx and BLASTn against the arthropod database from NCBI with an e-value of 1e-20, and all positive hit was discarded. Then, filtered transcripts were analyzed with other non-coding RNAs databases to remove non-related non-coding RNAs and pseudogenes (Figure 1). Finally, the remaining transcripts were predicted for putative lncRNAs by coding potential assessment tools (CPAT; http://lilab.research.bcm.edu/cpat/). The differentially expressed lncRNAs were predicted as sense and antisense lncRNAs, promoter-associated lncRNAs, 3′ UTR-associated lncRNAs, intronic lncRNAs, and intergenic lncRNAs during alignment and coding potential prediction steps as shown in Figure 1.

Published online: Copyright © 2016-2023 Rangsit University

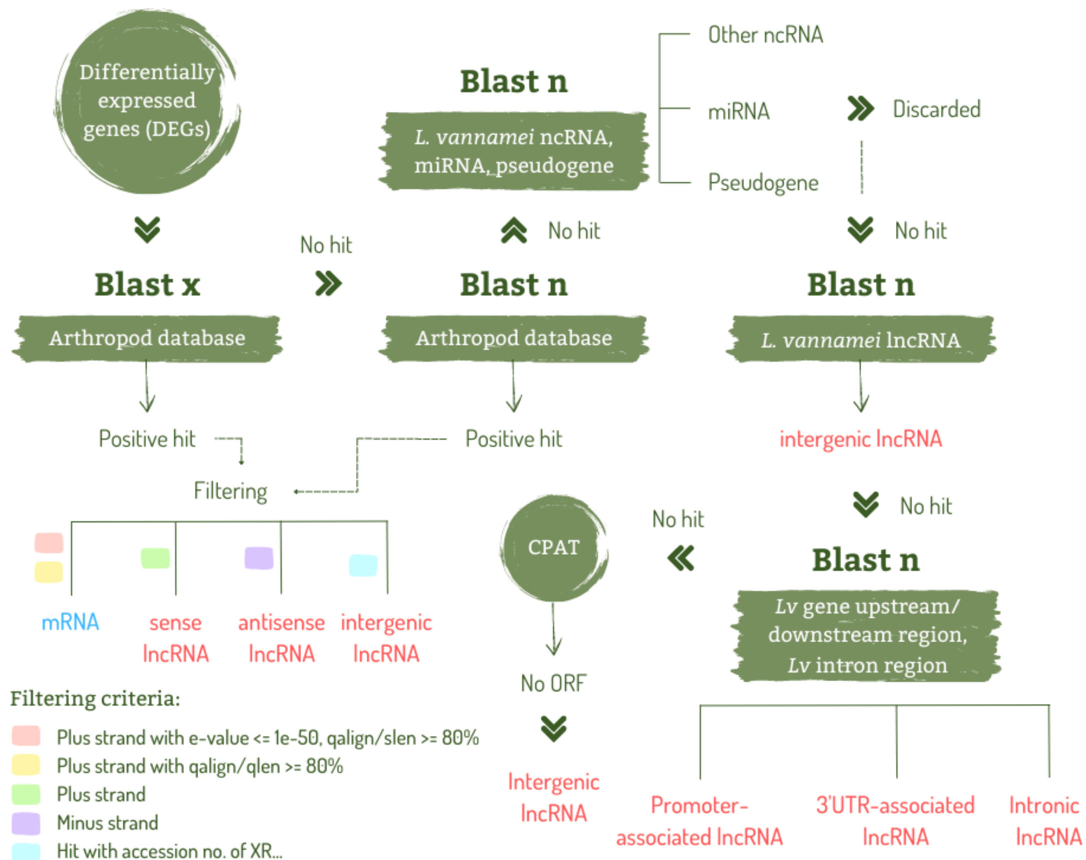

Figure 1 Schematic diagram of identification of putative lncRNAs

4. Results and Discussion

4.1 *In silico* identification of lncRNAs involved in WSSV infection in *L. vannamei* gills

Using RNA sequencing and miRNA sequencing, the transcription of mRNAs and miRNAs in *L. vannamei* gills was investigated over 36 h following WSSV infection (Millard et al., 2021). The research discovered a total of 6,192 mRNA transcripts and 27 miRNAs which were significantly induced by WSSV infection. Silencing of an immune priming gene Dscam by a novel shrimp miRNA (Pva-pmiR-78) might prevent the WSSV invasion. Nonetheless, studies about lncRNA expression profiles upon WSSV infection remain elucidated.

By using the same RNA-seq data provided by Millard et al. (2021), a total of 226,797 unigenes of *de novo* assembled transcriptome were obtained with a GC content of 43.64%, N50 of 929 bp and an assembled completeness of 96.8% (Table 1). After statistically analyzing the expression profile between control and WSSV-infected group, a total of 3,235 transcripts were significantly induced after WSSV infection (Table 1).

Notably, a total of 1,232 transcripts which made up a 38.1% of the total DEGs were mRNA. After filtering the transcripts subjected to protein coding sequences, pseudogenes, and other noncoding RNAs, the remaining 1,628 transcripts (50.3%) were identified as putative lncRNAs (Table 1). Additionally, of the 1,628 differentially expressed putative lncRNAs, 692 and 936 were up- and down-regulated at 24 h post-WSSV infection, respectively (Figure 2).

Many transcriptome studies focus on lncRNAs as important regulators in viral infections regarding their differential expressions upon viral infection (W. Liu & Ding, 2017; Peng et al., 2010). In WSSV

Published online: Copyright © 2016-2023 Rangsit University

infection, 6,544 lncRNAs were identified from hepatopancreas of kuruma shrimp, *Penaeus japonicas*, and 457 from them were significantly induced by WSSV infection (Zhang et al., 2022). Studies on shrimp lncRNAs are still limited to the identification and differential expression of lncRNAs and their responsible genes so far. Thus further studies should focus on the regulatory mechanisms of lncRNAs, especially in the host immune defense system and viral-host interaction.

Table 1 Statistics of *de novo* assembled transcriptome of shrimp during 24 h WSSV infection in gills

General information	
SRA accessions for control group	SRR14027734- SRR14027737
SRA accessions for WSSV-infected group	SRR14027701-SRR14027702, SRR14027704-SRR14027705
Total assembled bases	139,979,357
Number of transcripts	226,797
GC content (%)	43.64%
Average contig length (bp)	617.20
Contig N50 (bp)	692
Assemble completeness (%)	96.8
Differential expression of genes	
No. of differentially expressed transcripts	3,235
No. of differentially expressed mRNA	1,232
No. of differentially expressed lncRNA	1,628
lncRNA up-regulated during 24 h WSSV infection	692
lncRNA down-regulated during 24 h WSSV infection	936

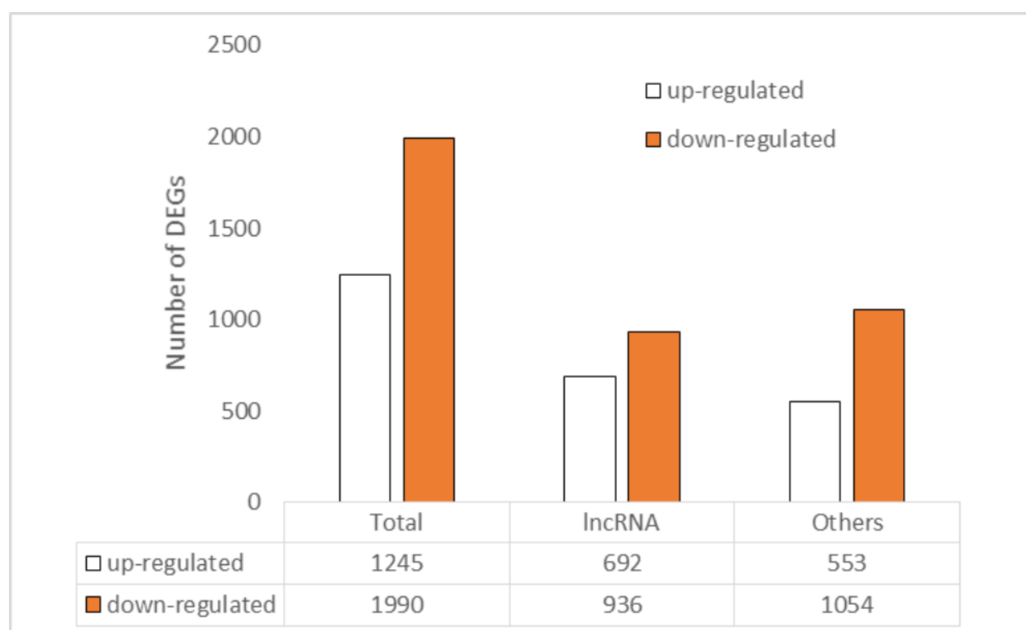

Figure 2 *In silico* analysis of differentially expressed lncRNAs. *De novo* assembled transcriptome was generated and DEG was analysis. Number of DEG composing of total transcripts, putative lncRNA, and others were shown

4.2 Classification of DE lncRNAs upon WSSV infection

The differentially expressed lncRNAs were classified based on their location on genome (Table 2). Among 1,628 differentially expressed lncRNAs, 860 and 191 were found to be transcribed from the same or

Published online: Copyright © 2016-2023 Rangsit University

opposite strand of protein coding genes, known as sense and antisense lncRNA, respectively. Apart from that, 15 and 144 lncRNAs were located within 1,000 upstream and downstream of coding genes, which were classified as promoter-associated and 3′ UTR-associated lncRNAs, respectively. The remaining lncRNAs were mapped with the intron region, returning 336 lncRNAs classified as intronic lncRNA. Apart from that, the remaining 82 transcripts matched with the lncRNAs on NCBI database, or *L. vannamei* lncRNA genome database, or they contained no ORF predicted by CPAT subjected to intergenic lncRNAs.

Based on the results, lncRNAs involving in WSSV infection were identified, and the classification was demonstrated. However, more understanding about the mechanism of lncRNAs regulate genes involved in host immune defense against infection remains elucidated. Understanding the genomic location of lncRNAs is relevant to the prediction of the possibility of lncRNA regulatory functions. Sense lncRNAs have been reported to comprise the majority of the lncRNAs, albiet poorly studied for their function regarding the difficulty in differentiating their sequences from their overlapping coding genes (Perez et al., 2021). In, a small planktonic crustacean, *Daphnia magna*, the sense lncRNA *DAPALR* reactivated the transcription of mRNA by canceling the suppression caused by a specific repressor (Perez et al., 2021). In this study, most of the identified lncRNAs involving viral infection belonged to sense lncRNA subclass, which took up 52.8% of the total lncRNAs.

In human, the cyclin D1 promoter-associated lncRNA represses transcription through an RNA binding protein TLS (Kurokawa 2011). Apart from that, a large proportion of antisense lncRNAs have been reported in many research studiesto have many regulatory functions over specific genes to modulate different signaling pathways (B. Liu et al., 2021). Intronic lncRNAs has been explored only in a small portion, regarding their function (Ma et al., 2013). A large number of lincRNAs have been identified with various regulation functions such as transcriptional regulation, translational control, splicing regulation, other post-transcriptional regulation, etc. (Ma et al., 2013). In this research, 82 lincRNAs which might regulate the function of their neighboring genes were found. In addition, these findings visualized the function of lncRNAs in WSSV infection, which might be mostly associated with transcriptional modulation which up- or down-regulated their adjacent or their co-expressed genes.

Table 2 Classification of lncRNAs induced by WSSV infection

lncRNA subclasses	No. of lncRNAs
Sense lncRNAs	860
Antisense lncRNAs	191
Promoter-associated lncRNAs	15
3′ UTR-associated lncRNAs	144
Intronic lncRNAs	336
Intergenic lncRNAs	82

5. Conclusion

From 3,235 DEGs, 1,628 lncRNAs have been identified in gills of *L. vannmei* infected with WSSV. The classification of putative lncRNAs returned sense lncRNAs as the major subclass, made up to 52.8% of the total putative lncRNAs, reflecting their regulatory function on expression of their adjacent genes. Further studies on lncRNAs should emphasize their functions in immune defense of host against viral infection and the mechanisms of lncRNAs regulating other coding genes should be clarified. Nevertheless, the research successfully discovered a number of lncRNAs which somehow might be associated with WSSV infection in *L. vannamei*.

6. Acknowledgements

This project was funded by National Research Council of Thailand (NRCT) and Prince of Songkla University (Grant No. N42A650383), and IW was supported by Graduate Fellowship (Research Assistant), Faculty of Science, Prince of Songkla University, Contract No. 1-2564-02-008.

Published online: Copyright © 2016-2023 Rangsit University

7. References

Aliperti, V., Skonieczna, J., & Cerase, A. (2021). Long non-coding rna (Lncrna) roles in cell biology, neurodevelopment and neurological disorders. *Non-Coding RNA*, *7*(2). https://doi.org/10.3390/ncrna7020036

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England), 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bridges, M. C., Daulagala, A. C., & Kourtidis, A. (2021). LNCcation: lncRNA localization and function. *Journal of Cell Biology*, *220*(2), 1–17. https://doi.org/10.1083/JCB.202009045

Casamassimi, A., Federico, A., Rienzo, M., Esposito, S., & Ciccodicola, A. (2017). Transcriptome Profiling in Human Diseases: New Advances and Perspectives. *International Journal of Molecular Sciences 2017, Vol. 18, Page 1652*, *18*(8), 1652. https://doi.org/10.3390/IJMS18081652

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29(7), 644-652

Gucic, M., Cortés-Jacinto, E., Civera-Cerecedo, R., Ricque-Marie, D., & Martínez-Córdova, L. R. (2013). Apparent carbohydrate and lipid digestibility of feeds for whiteleg shrimp, Litopenaeus vannamei (Decapoda: Penaeidae), cultivated at different salinities. *Revista de Biologia Tropical*, *61*(3). https://doi.org/10.15517/rbt.v61i3.11935

Kazimierczyk, M., Kasprowicz, M. K., Kasprzyk, M. E., & Wrzesinski, J. (2020). Human long noncoding RNA interactome: Detection, characterization and function. *International Journal of Molecular Sciences*, *21*(3). https://doi.org/10.3390/ijms21031027

Kotzin, J. J., Spencer, S. P., McCright, S. J., Kumar, D. B. U., Collet, M. A., Mowel, W. K., Elliott, E. N., Uyar, A., Makiya, M. A., Dunagin, M. C., Harman, C. C. D., Virtue, A. T., Zhu, S., Bailis, W., Stein, J., Hughes, C., Raj, A., Wherry, E. J., Goff, L. A., … Henao-Mejia, J. (2016). The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature 2016 537:7619*, *537*(7619), 239–243. https://doi.org/10.1038/nature19346

Kurokawa R. (2011). Promoter-associated long noncoding RNAs repress transcription through a RNA binding protein TLS. Advances in experimental medicine and biology, 722, 196–208. https://doi.org/10.1007/978-1-4614-0332-6_12

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), 357-359

Leu, J. H., Tsai, J. M., & Lo, C. F. (2008). White Spot Syndrome Virus. *Encyclopedia of Virology*, 450–459. https://doi.org/10.1016/B978-012374410-4.00776-7

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics, 12, 1-16

Liu, B., Xiang, W., Liu, J., Tang, J., Wang, J., Liu, B., Long, Z., Wang, L., Yin, G., & Liu, J. (2021). The regulatory role of antisense lncRNAs in cancer. *Cancer Cell International*, *21*(1). https://doi.org/10.1186/s12935-021-02168-4

Liu, W., & Ding, C. (2017). Roles of LncRNAs in viral infections. In *Frontiers in Cellular and Infection Microbiology* (Vol. 7, Issue MAY). https://doi.org/10.3389/fcimb.2017.00205

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, *13*(5). https://doi.org/10.1371/JOURNAL.PCBI.1005457

Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. In *RNA Biology* (Vol. 10, Issue 6). https://doi.org/10.4161/rna.24604

Published online: Copyright © 2016-2023 Rangsit University

Millard, R. S., Bickley, L. K., Bateman, K. S., Farbos, A., Minardi, D., Moore, K., Ross, S. H., Stentiford, G. D., Tyler, C. R., van Aerle, R., & Santos, E. M. (2021). Global mRNA and miRNA analysis reveal key processes in the initial response to infection with WSSV in the pacific whiteleg shrimp. *Viruses*, *13*(6). https://doi.org/10.3390/v13061140

Peng, X., Gralinski, L., Armour, C. D., Ferris, M. T., Thomas, M. J., Proll, S., Bradel-Tretheway, B. G., Korth, M. J., Castle, J. C., Biery, M. C., Bouzek, H. K., Haynor, D. R., Frieman, M. B., Heise, M., Raymond, C. K., Baric, R. S., & Katze, M. G. (2010). Unique signatures Of long noncoding RNA expression in response to virus infection And altered innate immune signaling. *MBio*, *1*(5). https://doi.org/10.1128/mBio.00206-10

Perez, C. A. G., Adachi, S., Nong, Q. D., Adhitama, N., Matsuura, T., Natsume, T., Wada, T., Kato, Y., & Watanabe, H. (2021). Sense-overlapping lncRNA as a decoy of translational repressor protein for dimorphic gene expression. *PLoS Genetics*, *17*(7), 1–18. https://doi.org/10.1371/journal.pgen.1009683

Santos, C. A., Andrade, S. C. S., Teixeira, A. K., Farias, F., Kurkjian, K., Guerrelhas, A. C., Rocha, J. L., Galetti, P. M., & Freitas, P. D. (2018). Litopenaeus vannamei Transcriptome profile of populations evaluated for growth performance and exposed to White Spot Syndrome Virus (WSSV). *Frontiers in Genetics*, *9*(APR), 1–6. https://doi.org/10.3389/fgene.2018.00120

Zhang, Y., Yao, N., Zhang, C., Sun, X., Huang, J., Zhao, B., & Li, H. (2022). LncRNA-mRNA integrated profiling analysis in response to white spot syndrome virus in hepatopancreas in Penaeus japonicus. *Fish and Shellfish Immunology*, *129*(August), 251–262. https://doi.org/10.1016/j.fsi.2022.08.061

Zhuang, L., Tian, J., Zhang, X., Wang, H., & Huang, C. (2018). Lnc-DC regulates cellular turnover and the HBV-induced immune response by TLR9/STAT3 signaling in dendritic cells. *Cellular & Molecular Biology Letters*, *23*(1). https://doi.org/10.1186/S11658-018-0108-Y

Published online: Copyright © 2016-2023 Rangsit University