- Note -

Marine Metatranscriptome Profiling in the Sea Adjacent to Jeju Island, Korea, by RNA-sequencing

Jinik Hwang[1,2], Mingyeong Kang[1,3], Kang Eun Kim[1,3], Seung Won Jung[4] and Taek-Kyun Lee[1,3*]

[1]*South Sea Environment Research Department, Korea Institute Ocean Science and Technology, Geoje 656-830, Korea*
[2]*Marine Ecology and Resource Convergence Center, Advanced Institutes of Convergence Technology, Suwon 443-270, Korea*
[3]*Ocean Science, University of Science and Technology, Daejeon 34113, Korea*
[4]*Library of Marine Samples, Korea Institute of Ocean Science & Technology, Geoje 53201, Korea*

The Ocean is a rich source of diverse living organisms include viruses. In this study, we examined the microbial communities in the sea adjacent to Jeju Island in two seasons by metatranscriptomics. We collected and extracted total RNA, and, using the next-generation sequencing HiSeq 2000 and *de novo* transcriptome assembly, we identified 652,984 and 163,759 transcripts from the March and December samples, respectively. The most abundant organisms in March were bacteria, while eukaryotes were dominant in the December sample. The bacterial communities differed between the two samples, suggesting seasonal change. To identify the viruses, we searched the transcripts against a viral reference database using MegaBLAST with the most identified being bacteriophages infecting the marine bacteria. However, we also revealed an abundance of transcripts associated with diverse herpesviruses in the two transcriptomes, indicating the presence or possible threat of infection of fish in the sea around Jeju Island. This data is valuable for the study of marine microbial communities and for identifying possible viral pathogens.

Key words : Metatranscriptome, NGS, microbial community, Jeju Island, viral pathogens

Introduction

Oceans are reservoirs of diverse organisms, including the three domains of archaea, bacteria, and eukaryotes. Of them, marine microorganisms, including bacteria as well as viruses, are the most dominant organisms in the seawater [1]. In particular, diverse viruses, including numerous unknown species, are abundantly present and their infection in a range of living organisms causes viral disease [2]. The number of known marine microorganisms is very small compared to that of unknown microorganisms, indicating that living microorganisms in the oceans should be explored in greater depth [3].

Similar to metagenomics [4], metatranscriptomics reveals mRNA information from target environments that contain an array of diverse living organisms without any prior knowledge of the environmental community [5]. In general,

there are many technical obstacles to isolating mRNAs from environments due to the short half-lives of mRNAs and absence of poly (A) tails of eukaryotic mRNAs [6]. However, with the advance of library construction as well as next-generation sequencing, it is now possible to reveal populations of marine microorganisms with unknown functions [7], the microbial metabolic process [8], and to monitor emerging marine pathogens causing the massive mortality of marine organisms [9].

Korea is a peninsular surrounded by the sea; in particular, the South Sea is an optimal place for the cultivation of marine products, including fishes, shellfishes, and many kinds of seaweeds [10]. Jeju Island is the largest island in Korea. It is located in the South Sea, South Pacific Ocean, and is famous as a tourist site and for its fishing industry. However, the fishing industry in Jeju Island has been threatened by unknown emerging pathogens for many years. It seems several pathogenic bacteria and viruses are possible causes; however, the information associated with pathogenic marine microorganisms is currently limited. Furthermore, many unknown living organisms might be involved in the perishing of fishes and red tide. In this study, we examined marine communities through metatranscriptomics in order to identify new emerging marine pathogens, including bacteria and

*Corresponding author
Tel : +82-55-639-8530, Fax : +82-55-639-8539
E-mail : tklee@kiost.ac.kr

viruses in the sea adjacent to Jeju Island.

Materials and Methods

For metatranscriptomics, we collected seawater samples from the sea (33.2275 N 126.6006 E) close to Jeju Island, Korea, in two different seasons in 2014, March and December (Fig. 1). To maximize the number of identified marine organisms, we have collected 10 l of ambient seawater derived from three different depths (1-2 m, 10 m, and 20 m) using sterile plastic bottles and pooled. The main subject of our project is marine viruses. Therefore, we conducted $FeCl_3$-mediated flocculation to enrich small-sized viral particles efficiently, as described previously [11, 12]. After flocculation, the incubated seawater was filtered using a polycarbonate filter membrane (0.22 μm Millipore) (Billerica, MA, USA) and kept at 4°C for further experiments. We extracted the total RNA from the filtered membranes possessing viral particles as well as microbial cells using the QIAeasy Plus Viral DNA/RNA Extraction Kit (Intron, Seongnam, Korea) according to the manufacturer's instructions. The extracted total RNA was treated with DNase I. The RNA library was generated using the TruSeq Nano RNA library preparation kit according to the manufacturer's manual (Illumina, San Diego, USA). We sequenced two libraries for paired-end sequencing using the HiSeq2000 platform, which was performed by Macrogen, Seoul, Korea. To assign the taxonomy of assembled transcripts, we performed BLAST against NCBI's microbial and viral database (http://www.ncbi.nlm.nih.gov/genome/ Refseq microbial and viruses) [13]. To classify the transcripts according to taxonomy, the obtained XML file was imported into the MEGAN5 program [14].
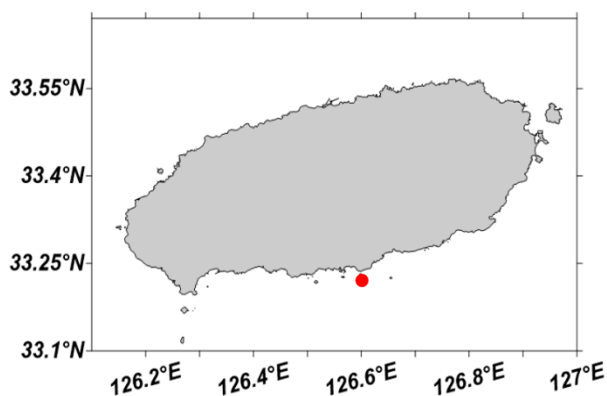
Results and Discussion

Transcriptome sequencing yielded 11.33 GB (Gigabytes) and 8.39 GB of raw data for the March and December samples, respectively. The raw data were subjected to *de novo* transcriptome assembly using the Trinity program [15]. We identified 652,984 and 163,759 transcripts from the March and December libraries, respectively (Table 1). The number of transcripts from the March library produced 4.5 times more than that of the December library. The GC percentages for March and December were 58.86% and 39.91%, respectively. The N50 values of the assembled contigs for March and December were 477 bp and 291 bp, respectively.

Through BLAST the transcripts, biological community was identified in March and December. In March, most transcripts were derived from bacteria (35.36%), followed by eukaryota (3.47%), viruses (0.04%) and archaea (0.02%) (Fig. 2A). In December, many transcripts were assigned to eukaryota (11.00%), followed by bacteria (3.54%) and viruses (0.15%) (Fig. 2B). As we expected, many transcripts, such as 60.32% (March) and 84.00% (December), did not show any homology to known sequences in the nt database. From the March sample, we identified diverse organisms that were assigned in 55 phyla, 121 classes, 333 orders, 576 families, 1,128 genera and 2,124 species, and 28 phyla, 57 classes, 146 orders, 217 families, 354 genera and 471 species were identified in the December sample (Fig. 3). We compared commonly identified marine organisms between the March and December libraries based on species. We found that 326 species were commonly identified in both seasons, and 1,798 species and 145 species were specific to March and December.

The BLAST search revealed that the most abundant organisms were derived from the bacteria in the March sample, while eukaryota were dominant in the December sample. The most frequently identified species in March were *Halomonas campaniensis, Halomonas sp. KO116, Marinobacter*

Fig. 1. Maps for the sampling location near Jeju Island of Korea.

Table 1. Result of transcriptome assembly

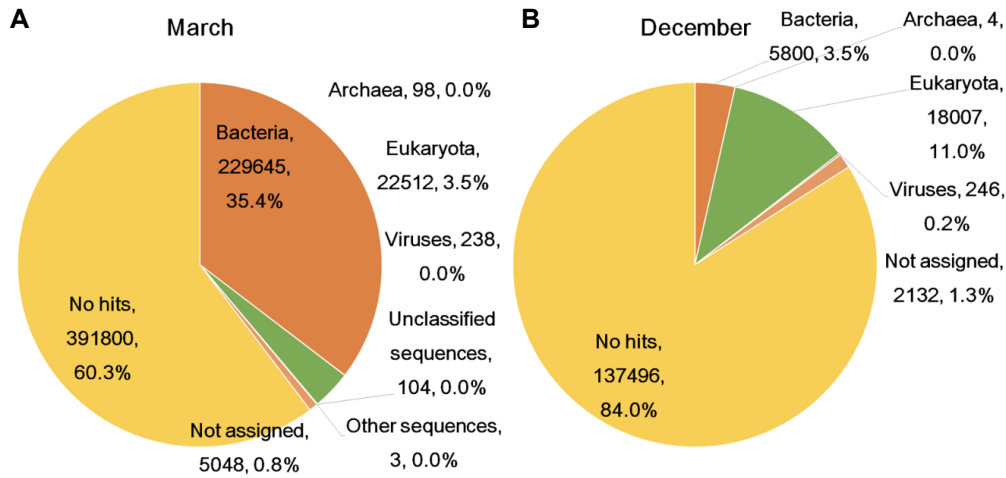|	March	December
Number of total genes	615,373	136,249
Number of total transcripts	652,984	163,759
GC (%)	58.86	39.91
N50	477	291
Median contig length (bp)	333	268
Average contig length (bp)	468.42	316.99
Total assembled bases (bp)	305,872,232	51,909,863

Fig. 2. The distribution of blast results in March (A) and December (B) samples based on taxonomy binning using MEGAN program.
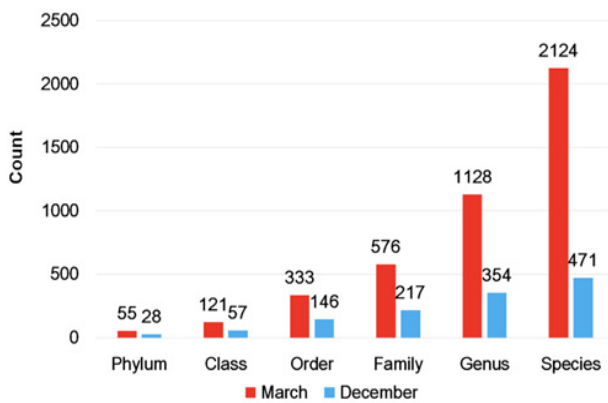

Fig. 3. The number of transcripts and percentages of March (A) and December (B) samples.

salarius and *Alcanivorax dieselolei*, all of which belong to the gammaproteobacteria (Fig. 4A). In December, three bacteria, *Zunongwangia profunda* (Bacteroidetes), *Pseudoalteromonas sp. SM9913* (ɣ-proteobacteria) and *Oleispira antarctica* (ɣ-proteobacteria) (Fig. 4B).

To identify marine viruses in two metatranscriptomes, we blasted all transcripts against the viral reference database using Megablast with the *E*-value $1 \times e^{-5}$ as a cutoff. As a result, we identified 215 and 66 viruses from the March and December samples, respectively (Fig. 5A). Of the two samples, 39 viruses were commonly identified. The most identified viruses in the two transcriptomes were bacteriophages infecting marine bacteria (72%) (Fig. 5B). Based on viral ge-
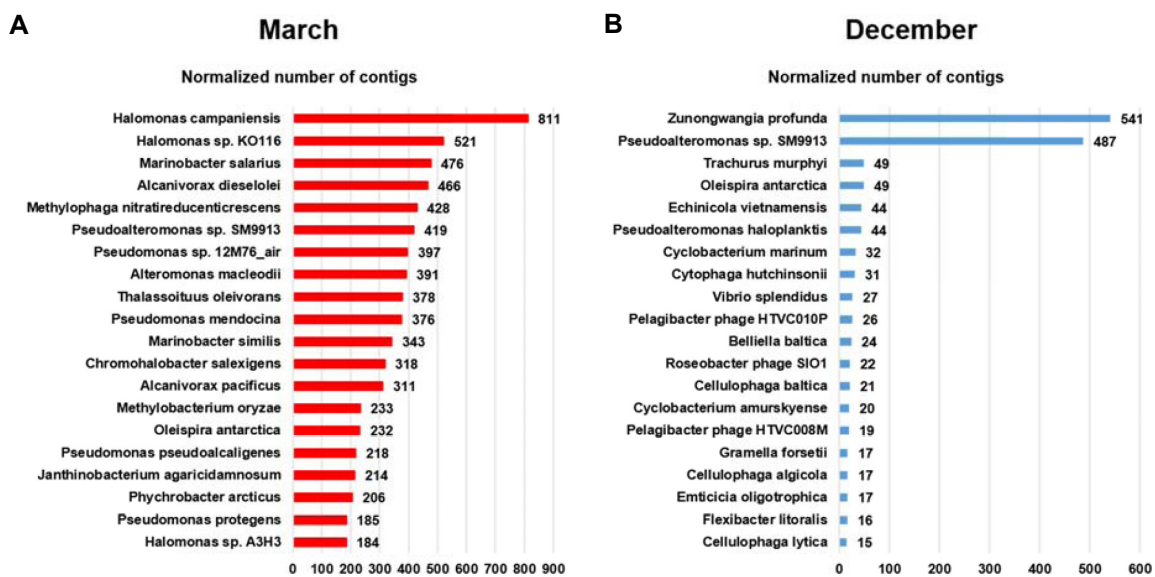

Fig. 4. The numbers of assigned taxonomies in March and December samples. The top 20 species abundantly present in March (A) and December (B) samples based on normalized number of transcripts.
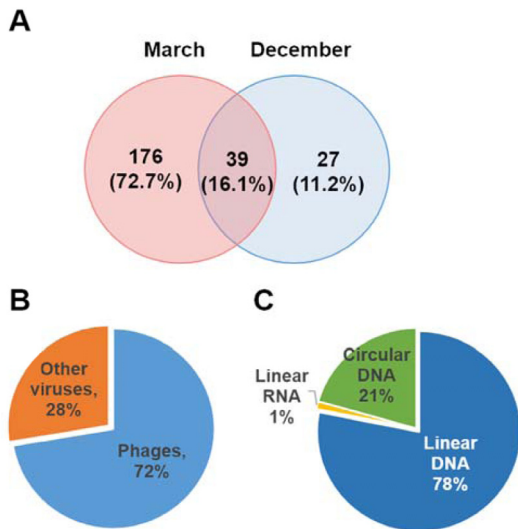
Fig. 5. Comparison of identified virus species between March and December samples (A). Distribution of identified viruses in both samples according to the host range (B) and viral genome (C).
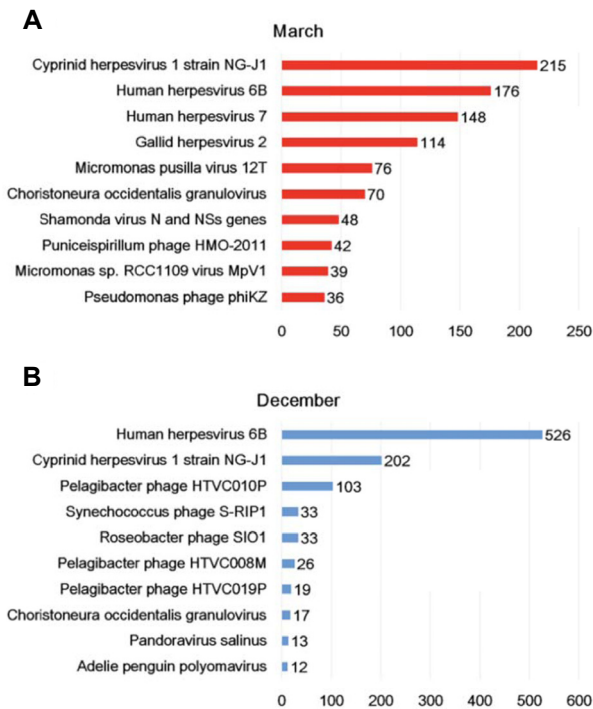

Fig. 6. The top ten viruses which were dominantly present in March (A) and December (B) samples.

nome types, the majority of viruses were composed of linear DNA (78%), followed by circular DNA (21%) and linear RNA (1%) (Fig. 5C).

Based on the number of assembled transcripts, the most frequently identified viruses in the March transcriptome were homologous to the four herpesviruses, *Cyprinid herpes-*

virus 1 strain NG-J1 (215 transcripts), *Human herpesvirus 6B* (176 transcripts), *Human herpesvirus 7* (148 transcripts) and *Gallid herpesvirus 2* (114 transcripts). In addition, many transcripts were matched to the known two algae viruses, *Micromonas pusila virus 12T* (76 transcripts) and *Micromonas sp. RCC1108 virus* (39 transcripts) (Fig. 6A). In the December transcriptome, many transcripts were homologous to the two known herpesviruses, *Human herpesvirus 6B* and *Cyprinid herpesvirus 1* strain NG-J1 (Fig. 6B). We found the presence of high levels of herpesviruses in both the March and December transcriptomes. The herpesviruses are important pathogens in fish as well as in many mammals and birds [16]. In general, infection with most herpesviruses causes mild disease symptoms or is non-symptomatic; however, an abnormal host or environmental changes can affect the pathogenicity of the herpesvirus infection [17-19]. The abundance of transcripts associated with several herpesviruses indicates the presence or potential risks of marine viral diseases caused by herpesviruses in the sea adjacent to Jeju Island.

Taken together, we conducted metatranscriptome profiling for the first time in the sea adjacent to Jeju Island, Korea, at two different time points by RNA-Sequencing. Our results showed specificity and commonality in marine communities between the March and December samples. In general, marine bacteria were dominantly present in both samples; however, the detailed bacterial communities between the two samples were different from each other, suggesting changes of bacterial communities during seasonal changes. To be specific, we revealed the abundance of transcripts associated with diverse herpesviruses in the two transcriptomes, indicating the presence or possible threat of herpesviruses infecting fishes in the sea adjacent to Jeju Island, Korea.

Acknowledgements

The Conflict of Interest Statement

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Kato, C., Inoue, A. and Horikoshi, K. 1996. Isolating and characterizing deep-sea marine microorganisms. *Trends Biotechnol.* **14,** 6-12.

2. Suttle, C. A. 2007. Marine viruses-major players in the global ecosystem. *Nat. Rev. Microbiol.* **5,** 801-812.

3. Handelsman, J. 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68,** 669-685.

4. Gilbert, J. A. and Dupont, C. L. 2011. Microbial metagenomics: beyond the genome. *Annu. Rev. Mar. Sci.* **3,** 347-371.

5. Poretsky, R. S., Gifford, S., Rinta-Kanto, J., Vila-Costa, M. and Moran, M. A. 2009. Analyzing gene expression from marine microbial communities using environmental transcriptomics. *J. Vis. Exp.* **24,** 1086.

6. Sorek, R. and Cossart, P. 2010. Prokaryotic transcriptomics: a new view on regulation, *physiology* and pathogenicity. *Nat. Rev. Genet.* **11,** 9-16.

7. Baker, B. J., Lesniewski, R. A. and Dick, G. J. 2012. Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J.* **6,** 2269-2279.

8. Kujawinski, E. B. 2011. The impact of microbial metabolism on marine dissolved organic matter. *Annu. Rev. Mar. Sci.* **3,** 567-599.

9. Segarra, A., Mauduit, F., Faury, N., Trancart, S., Dégremont, L., Tourbiez, D., Haffner, P., Barbosa-Solomieu, V., Pépin, J. F. and Travers, M. A. 2014. Dual transcriptomics of virus-host interactions: comparing two Pacific oyster families presenting contrasted susceptibility to ostreid herpesvirus 1. *BMC Genomics* **15,** 580.

10. Chung, I. K., Kang, Y. H., Yarish, C., Kraemer, G. P. and Lee, J. A. 2002. Application of seaweed cultivation to the bioremediation of nutrient-rich effluent. *Algae* **17,** 187-194.

11. Zhu, B., Clifford, D. A. and Chellam, S. 2005. Virus removal by iron coagulation‐microfiltration. *Water Res.* **39,** 5153-5161.

12. Hwang, J., Park, S. Y., Park, M., Lee, S. and Lee, T. K. 2017. Seasonal dynamics and metagenomic characterization of marine viruses in Goseong Bay, Korea. *PLos One* **12,** e0169841.

13. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421.

14. Huson, D. H., Auch, A. F., Qi, J. and Schuster, S. C. 2007. MEGAN analysis of metagenomic data. *Genome Res.* **17,** 377-386.

15. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. and Zeng, Q. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29,** 644-652.

16. Hanson, L., Dishon, A. and Kotler, M. 2011. Herpesviruses that infect fish. *Viruses* **3,** 2160-2191.

17. Woźniakowski, G. and Samorek-Salamonowicz, E. 2015. Animal herpesviruses and their zoonotic potential for cross-species infection. *Ann. Agric. Environ. Med.* **22,** 191-194.

18. Sehrawat, S., Kumar, D. and Rouse, B. T. 2018. Herpesviruses: Harmonious pathogens but relevant cofactors in other diseases? *Front. Cell Infect. Microbiol.* **8,** 177.

19. Bento, M. C., Canha, R., Eira, C. B., Vingada, J., Nicolau, L., Ferreira, M., Domingo, M., Tavares, L. and Duarte, A. 2019. Herpesvirus infection in marine mammals: A retrospective molecular survey of stranded cetaceans in the Portuguese coastline. *Infect. Genet. Evol.* **67,** 222-233.

초록 : RNA-sequencing을 이용한 제주도 인접 바다의 메타전사체 프로파일링

황진익[1,2] · 강민경[1,3] · 김강은[1,3] · 정승원[4] · 이택견[1,3*]

([1]한국해양과학기술원 생태위해성연구부, [2]차세대융합기술연구원 환경자원융합센터, [3]과학기술연합대학원 대학교 해양과학과, [4]한국해양과학기술원 시료도서관)

바다는 바이러스를 포함하는 다양한 생물체의 풍부한 자원을 제공한다. 본 연구에서는 계절에 따른 제주 바다의 해양 미생물 군집을 확인하기 위해 3월과 12월에 해수 샘플을 수집하여 total RNA를 추출, HiSeq2000 및 *de novo* 전사체 어셈블리를 사용한 NGS를 실시하였다. 그 결과, 3월 및 12월 시료에서 각각 652,984 및 163,759 개의 전사체를 확인하였다. 3월 샘플에서는 해양 박테리아가 우점하였으나 12월 샘플에서는 진핵생물이 우점하였다. 박테리아 군집은 두 샘플간에 상이하였으며, 이는 계절 변화 동안 박테리아 군집이 변화하였음을 보여주었다. 또한, 해양바이러스를 확인하기 위하여, Megablast를 사용하여 바이러스 참조 데이터베이스에 전사체를 검색하였다. 해양박테리아를 감염시키는 박테리오파지가 두 샘플에서 우점하는 것을 확인하였다. 그러나, 우리는 두 개의 전사체에서 다양한 헤르페스바이러스와 관련된 transcripts가 풍부함을 확인하였으며, 이는 제주도 인근 바다에서 물고기를 감염시키는 헤르페스바이러스의 위협 가능성을 나타낸다. 종합하면, 우리의 데이터는 해양 커뮤니티 연구 및 가능한 해양 바이러스 병원체를 식별하는 데 유용할 것이다.