Thesis for the Degree of Master of Science

Diversity of bacteria and viruses in sea cucumbers and shrimps

by

Eun Soo Noh

Department of Microbiology

The Graduate School

Pukyong National University

February 2013

Diversity of bacteria and viruses in sea cucumbers and shrimps
(해삼과 새우에 존재하는 세균과 바이러스의 다양성)

Advisor: Prof. Gun Do Kim

by

Eun Soo Noh

A thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Science

In Department of Microbiology, The Graduate School,

Pukyong National University

February 2013

Diversity of bacteria and viruses in sea cucumbers and shrimps

A dissertation

by

Eun Soo Noh

Approved by:

(Chairman) Kyoung Ho Kim

(Member) Tae Jin Choi

(Member) Gun Do Kim

February 22, 2013

CONTENTS

Ⅰ. Introduction ---------- 1

Ⅱ. Materials and Methods ---------- 6

2.1 Sample preparation ---------- 6
2.2 Bacterial DNA extraction and 16S rRNA gene amplification ----- 8
2.3 Analysis of bacterial 16S rRNA gene sequences ---------- 8
2.4 Viral purification and concentration ---------- 9
2.5 Counting of viral particles using epi-fluorescence microscopy ---- 10
2.6 DNA extraction and multiple displacement amplification (MDA) method -- 10
2.7 Analysis of viral metagenomic sequences ---------- 11

Ⅲ. Results ---------- 13

3.1 Bacterial metagenomic analysis ---------- 13
3.2 Phylogenetic relationship of bacterial samples ---------- 13
3.3 Abundance of viruses from sea cucumbers and shrimps --------- 18
3.4 Pyrosequencing of viral metagenomes ---------- 20
3.5 Comparison with viromes from different environments --------- 23
3.6 Virus composition based on the database comparison ---------- 25
3.7 Phylogenetic diversity of viral genes ---------- 32

Ⅳ. Discussion ---------- 35

Ⅴ. 국문초록 ---------- 41

Ⅵ. Acknowledgement ---------- 43

Ⅶ. References ---------- 44

Diversity of bacteria and viruses in sea cucumbers and shrimps

Eun Soo Noh

Department of Microbiology, The Graduate School
Pukyong National University

ABSTRACT

Various studies have been conducted about the diversity of both bacteria and viruses from marine environment. However, diversity of microbes in marine invertebrates has not been greatly explored. Being genetically diverse, both bacteria and viruses from marine invertebrates can be promising biological resources. In this study, the diversity and genetic resources of such microbes in intestine samples of sea cucumbers and shrimps were investigated by metagenomics. Bacterial DNA was isolated using the Power Soil DNA Extraction Kit (MOBIO) and the extracted DNA was amplified using barcoded primers (27F and 518R). PCR products were subjected to 454 pyrosequencing and the resulting sequences were analyzed using QIIME. Various analyses revealed that sea cucumbers and shrimps have different bacterial composition. Viruses were collected through sequential filtration and their abundance ranged from 10^8 to 10^9 per gram of intestine. Viral DNA was extracted and was amplified via the multiple displacement amplification (MDA) method using the phi29 polymerase, and was analyzed by 454 pyrosequencing. Results showed that three viral families with double stranded DNA genomes including *Siphoviridae* (14.2%, 33.2%), *Podoviridae* (11.3%, 5.2%) and *Myoviridae* (8.9%, 9%), and three viral families with single stranded DNA genomes including *Microviridae* (52.5%, 29.7%), *Circoviridae* (6.3%, 10.2%), *Nanoviridae* (3.4%, 9.2%)

dominated in sea cucumbers and shrimps, respectively. Phylogenetic analyses revealed that a huge part of viromes are novel sequences and that most of the isolated viromes are distantly related to the previously known viruses. Thus, this study showed that the microflora in the intestine of both sea cucumbers and shrimps consisted of unknown bacteria and viruses.

INTRODUCTION

It is well known that oceans occupy more than 70% of the Earth (28). However, the fact that almost 99% of organisms inhabit oceans is the most important (6). The reason for this is the presence of various life forms extending towards the deep portions of the sea, while on land, organisms like plants and animals only inhabit a shallow vertical zone way below the land surface. The marine system is an extremely complex environment having an enormous diversity of life assemblages. Thus, marine living resources are very important for genetic resource studies (19).

The marine environment is immensely diverse, considering viruses as one of the most important sources for genetic studies (43). Marine microorganisms play a key role in marine food webs and biogeochemical cycling in marine ecosystems. This makes them an interesting subject in the field of biotechnology as they are also found out to be potential and significant sources of novel enzymes and metabolites, important in biotechnological applications (43).

Bacteria are essential organisms for mankind. They take part in various biological cycles, cause infectious diseases and play significant roles in biotechnological production. Thus, understanding bacterial diversity is important (10). Most bacteria and archaea from the environment are found to be unculturable and a large part of thembelong to previously unidentified groups of both Bacteria and Archaea domains. These unculturable bacteria have some ecological functions and biotechnological applications unidentified so far (31).

Microbes are known to be ubiquitous, hence, the marine environment, as well, possesses a number of microorganisms that are possible sources of bioactive compounds. Some have bioremediation abilities, and some contribute to organic matter decomposition and nutrient cycling.

They also serve as food sources to other marine organisms such as marine invertebrates (40). As the oceanic biomass is mostly made up of a huge number of bacteria, archaea, protists and unicellular fungi, the occurrence of mutations has been rampant, leading to very high levels of genetic and phenotypic variations (37). With this, molecular ecology methods have been proposed for the culture-independent characterization of complex bacterial communities associated with various habitats, revealing the extent of their diversity (38).

Metagenomics is a new approach used to discover novel microbial products from unculturable microorganisms (21). Complex microbial communities inhabit the surfaces and gastrointestinal tract of mammals including humans. The intestine has a high microbial diversity and the human intestinal microbiota is known to function for host nutrition, epithelial development regulation, intestinal angiogenesis stimulation, inflammatory immune responses and pathogen resistance (44).

This complex microbiota is considered to be critically important for various gut functions, such as host nutrition, regulation of epithelial development (3), regulation of host fat storage (4), stimulation of intestinal angiogenesis (5), inflammatory immune responses (6) and pathogen resistance

As for marine microorganisms, the gastrointestinal (GI) microbiota also participates in several physiological functions of their host. Thus, studying the microbial diversity of the GI microbiota might lead to the understanding of the functional mechanisms of their host (36).

As it has been believed that microbes run the world and that viruses control other microbes most of the time, it can then be assumed that viruses run the world (24). Viruses are considered to be the most abundant biological entities on Earth, having an estimated number of 10^{31}(29), with the bacteriophages alone at an estimated total abundance of 10^{30} (24). The same thing goes for the marine ecosystems as they were also dominated by viruses, serving as an important part of the marine microbial loop (3). In

oceans, marine viruses, bacteriophages, in particular, killing both heterotrophic and autotrophic microbes are present in most number (1). Viruses also are copious and significant members of marine habitats, which control microbial abundance, influence community composition through lysis of specific host organisms and are a critical link in global biogeochemical cycles. Through transduction and lysogenic conversion, phages contribute to bacterial diversity (43). Phages and nanoflagellates, a group of marine microbes that prey on other microbes, regulate the number of marine microbes in oceans (1). It has been known that virus-like particles are dominated by double-stranded DNA (dsDNA) viruses that contain genomes ranging from 25 to 70 kilobases (kb) in length, through direct counts with epifluorescence microscopy and flow cytometry (43). Previous studies using electron and fluorescence microscopy regarding marine viruses showed an unexpected number of viral particles with values of 10^6 to 10^9 per milliliter of seawater. This reveals that viruses are the most abundant microbes not only in marine environments, but also in the whole biosphere (41).

The immense abundance of viruses is known to be a significant cause of microbial mortality in oceans. Some viruses are pathogens that infect heterotrophic bacteria and cyanobacteria, as well as photosynthetic and non-photosynthetic protists. It has been known that marine viral communities are comprised almost entirely of double-stranded DNA viruses. Accordingly, natural communities of RNA viruses have not been greatly explored (7).

Viruses have no inherent metabolic activities, requiring them to interact with the replication machinery of their host organisms. Viruses also serve as by-products of intimate intracellular interactions, making them major drivers of evolutionary changes for cellular life. Access to sequence information harbored in environmental viral assemblages has become of interest, because it provides insight into the types of viruses present in different habitats, and reveals the wealth of extracellular genetic information

with which planktonic organisms are in constant communication (39).

The development of advanced molecular techniques provides additional information regarding the diversity of viruses. Viral studies have gained focus in the recent years due to the innovations associated with them. Multiple displacement amplification (MDA) with phi29 polymerase has been widely used to study single stranded DNA viruses (ssDNA), especially microphages. It has also given way to study ssDNA viruses by making use of random-priming and preferential amplification of circular genomes. With this, investigations on the genotypic diversity of ssDNA viruses have been conducted from environments such as paddy soil (19), the sea (1), fresh water (33), Antarctic lake (23) and microbialites (9).

Invertebrates such as sea cucumbers were discovered to be significant sources of bioactive compounds. One of which is lectin (HSL) from *Holuthuriascabra*, which has a strong broad spectrum of antimicrobial activity against bacteria. Aside from being edible food sources, sea cucumbers and their commensals were found to also exhibit medicinal properties (13).

Other invertebrates like sea squirts were also known sources of biologically active compounds such as didemnin B, a metabolite which functions as a strong antiviral agent and displays cytotoxic effects on leukemic cells. It has also demonstrated antineoplastic activity against a variety of murine tumor models (17).

Halichondrin B is also a known bioactive from invertebrates, mostly sponges. This compound was proven to have anti-cancer effects (16). The red alga, *Portieria homemannii*, was also found to contain a bioactive compound called halomon that is against tumors(35).

The culturing of invertebrates has been a worldwide economic activity specifically in intertropical developed and developing countries. However, along with the intensification of farming comes the development of many infectious diseases, mostly of viral origin, causing a decrease in

growth of production, eventually resulting in vast economic losses (29). Likewise, considering how rich and diverse the marine environment is and as viruses are more diverse than bacteria, only a few studies have been conducted about the viral diversity in oceans. Herewith, this study was conducted and aims to determine the diversity and abundance of sea cucumber and shrimp samples since this field of study remained greatly unexplored.

Aquatic invertebrates were known to possess gut bacteria that gives mutual benefits from the association of the host with these microbes. Such bacteria, as ingested, can provide food or enzymes to their host, supplying essential nutrients that the host is lacking. Gut bacteria can also aid on the process of ion transport across the gut wall of the host, enriching its resistance to toxic chemicals and out competing opportunistic pathogens. The abundance and species composition of gut bacteria inhabiting aquatic invertebrates are influenced by several factors, relating to the anatomy and physiology of the invertebrate host. Conditions of the external environment such as water salinity and temperature, the presence of toxicants in the water column, and food availability have also great effects on gut bacteria (26).

MATERIALS AND METHODS

Sample preparation

Sea cucumber (*Apostichopus japonicus*) sampling was performed twice in the same area in Busan, South Korea and samples were taken from the wild. Fresh samples were processed accordingly and their intestines were used for further processing. As for shrimps (*Litopenaeus vannamei*), samples were collected from six different shrimp farms in close areas located in Sinan-gun, Jeollanam-do, South Korea. Intestine samples were taken from deep-frozen shrimps collected previously. The size of each sample was then recorded in table 1.

Table 1. Sampling details.

Sample	Target	Sample name	Size (mm)	Data
Sea cucumber	Bacteria	BSC1	8.5	10.14.2011
		BSC2	8.0	
		BSC3	7.8	
		BSC4	14.0	11.08.2011
		BSC5	9.0	
		BSC6	10.5	
	Viruses	VSC1	9.5	01.13.2012
		VSC2	9.0	
		VSC3	10.5	
Shrimp	Bacteria and Viruses	SH2	13.0	10.10.2011
		SH4	15.0	
		SH7	14.5	
		SH12	14.0	
		SH16	15.0	
		SH18	14.0	

Bacterial DNA extraction and 16S rRNA gene amplification

Extraction of bacterial DNA from 500 mg of wet intestine samples from each of the six sea cucumbers and six shrimps was performed using the Power Soil DNA Extraction Kit (MOBIO). Extracted DNA was amplified using barcoded primers. The forward primer 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and the reverse primer 518R (5'-WTTACCGCGGCTGCTGG-3') were utilized for amplification. A 20 µl PCR mix (Accupower PCR premix, Bioneer) was used with each barcoded primer set and 1 µl of the extracted DNA template was added. The following conditions were employed for PCR: 94°C for 7 min, 32 cycles; 94°C for 30 s, 60°C for 30 s, 72°C for 45 s, final extension at72°C for 10 min. PCR products were subjected to 454 pyrosequencing using Genome Sequencer FLX titanium.

Analysis of bacterial 16s rRNA gene sequences

The QIIME (Quantitative Insights Into Microbial Ecology) pipeline was used to process data from the high-throughput 16S rRNA sequencing studies (reference for QIIME). The DNA sequence reads were filtered for quality and multiplexed reads were assigned to starting samples by nucleotide barcode. Operational Taxonomic Units (OTUs) were picked based on sequence similarity within the reads and a representative sequence from each OTU was picked. The OTU was then assigned to a taxonomic identity based on the Silva databases with BLAST (E value, $<10^{-30}$). The OTU sequences were aligned and a phylogenetic tree was created. Diversity metrics were calculated for each sample and the types of communities were compared using the taxonomic and phylogenetic assignments. PCoA analysis was performed to visually depict the differences between the samples using the Fast Unifrac web server. (reference for Fast UniFrac) Publication quality

figures were then generated by dynamically working with such graphs.

The QIIME pipeline was used to generate images of the OTU table. The OTU table was utilized to make a heatmap. A graphic was generated and for each sample, the number of times each OTU was found in that sample was shown in the heatmap. OTU counts were filtered by 49, representing 0.1% of the total count. Thus, counts were adjusted to percentage values to show the corresponding ratio and proportion.

In this study, the beta diversity was computed using the default beta diversity metrics of weighted and unweighted unifrac, which are phylogenetic measures used extensively in recent microbial community sequencing projects. Beta diversity plots were also generated. In performing the analysis, the OTU table was rarified to remove sample heterogeneity. Preferences were made to define the colors for each of the samples or for a particular category within a mapping column. Beta diversity was then computed to assess the differences between microbial communities. Principal coordinates were generated as 2D PCoA plots. Distance histograms were then made to compare samples from different categories and see which categories tend to have larger or smaller beta diversity than others.

Viral purification and concentration

Viruses were extracted from intestines of three sea cucumber samples (Table 1). Isolation of viruses was conducted individually for each of the three samples. As for shrimps, six intestine samples were combined prior to the extraction. Every sample was mixed with 10 ml SM buffer containing 0.1 M NaCl, 50 mM Tris-HCl [pH 8.0] and 10 mM $MgSO_4$ for resuspension. Viruses were extracted from intestines by vortexing twice for 10 min. The suspension was centrifuged at 10,000 X g for 10 min to remove unnecessary sedimentary particles. The suspension was transferred to a

clean tube and was centrifuged at 10,000 X g for 10 min prior to a series of filtration. The suspension was passed through a 0.45 μm (Pall, USA) membrane filter paper disk, respectively. Samples were then resuspended into a fresh eluent and the extraction was done twice by filtration through a 0.20 μm syringe filter (Millipore, USA) to remove bacteria and other small particles. Intestine pellets were resuspended in SM buffer and the extraction procedure was repeated one more time. The filtrate was concentrated using a centrifugal concentration filter (Vivaspin-20, 50 kDa, Biotech) and the supernatant was then concentrated 20ml to 1ml for each sample. The filtrate was treated with DNase I (with a final concentration of 0.1 μg/ml) at 37°C for 30 min and was incubated at 80°C for 10 min for the inactivation of DNase I, prior to the extraction of nucleic acids.

Counting of viral particles using epi-fluorescence microscopy

Quantitative analysis of viruses was done by epifluorescence microscopy (EFM) with SYBR Gold (Molecular Probes) staining. An Anodisc filter (0.02 μm) was placed over a pre-wetted 0.45 μm Millipore filter. Two hundred microliter of concentrate sample was passed through an Anodisc filter with vacuum pressure and 1X SYBR Gold staining solution was dropped to the filter for 10 min in the dark. After staining, the excess SYBR Gold staining solution was removed. A drop of ProLong Gold Antifade was placed on the filter and a piece of portion of the filter was put on a glass slide. Then, quantitative analysis was performed using epifluorescence microscopy at 480-495nm.

DNA extraction and multiple displacement amplification (MDA) method

Purified and concentrated viruses from three individual sea cucumber samples and one combined sample from shrimps were used to extract DNA. Viral DNA extraction for both sea cucumber and shrimp

samples was then conducted. Two hundred microliter of concentrated viruses from each of the samples was subjected to viral DNA extraction using the QIAamp MinElute Virus Spin Kit (QIAGEN) according to the manufacturer's instructions. After extraction, the DNA extracted from the combined shrimp sample was divided into three aliquots prior to the process of viral DNA amplification. To amplify DNA, the Genomiphi V2 DNA amplification kit (GE Healthcare, Piscataway, NJ) was used. From the kit, 9 µl of the sample buffer was added to 1µl of the DNA and the mixture was heated for 3 min at 95 °C. As also provided in the kit, 9 µl of the reaction buffer and 2 µl of phi 29 DNA polymerase were added to the previous mixture. The resulting mixture was incubated for 1.5 h at 30 °C. The mixture was further incubated for 10 min at 65°C to inactivate polymerase (17,21,36). After the amplification process, three sea cucumber and three shrimp amplified DNA samples adjusted to have same concentration were combined, respectively. DNA amplified through this method was further digested with 5U/µl S1 nuclease in 1x buffer for 1 h at 30°C prior to 454 pyrosequencing using Genome Sequencer FLX titanium.

Analysis of viral metagenomic sequences

To analyze the viral metagenomic sequences, CAMERA 2 web server (http://portal.camera.calit2.net), specifically the 454 Duplicate Clustering with the CH-HIT-454 4.5.3 program, was used (11, 27). Sequence identity and alignment coverage were set to 0.96 and 0.9, respectively. To check for sequence anomalies, quality control, sequence manipulation and the Galaxy (https://main.g2.bx.psu.edu) web server were used. The minimal quality score were set to 10, while the minimal length of contiguous segment was set to 100 (5, 20). Resulting sequences after data deduplication and quality control processes were compared to that of the

previously recorded databases such as the NCBI non-redundant database and the CAMERA viral protein database. The Metavir web server was used to analyze the obtained viral metagenomes from both sea cucumber and shrimp samples. Metagenomic reads were inserted in phylogenetic trees containing reference sequence marker genes. As for the Rep marker gene, the viral diversity of Lake Pavin freshwater was compared to that of sea cucumbers and shrimps. While for VP1 – a type of capsid protein – comparison of both samples was done by comparing their viral diversity to that of other communities – Lake Pavin freshwater, British Columbia seawater and the microbialites (34).

RESULTS

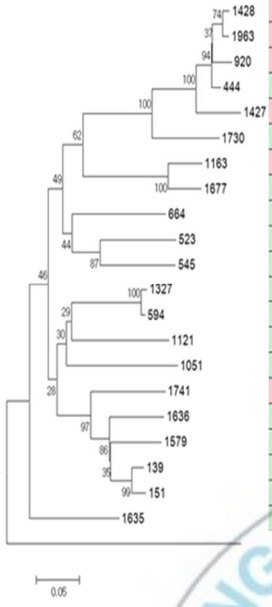
Bacterial metagenomic analysis

Bacterial metagenomic sequences from both sea cucumber and shrimp samples were analyzed to check for sequence anomalies using Silva RDP database (Table 2). After the quality check, 539 OTU groups were found to be of high quality. Only OTU groups with more than 5% dominance were used for further analysis. Results then revealed a total of 21 OTU groups after the final checking (Figure 1).

Each type of samples is known to have a different bacterial composition as revealed by the heatmap analysis results. In the phylogenetic tree, red and green were used to represent sea cucumber and shrimp samples, respectively. Sea cucumbers are dominated by *Propionigenium* except in BSC4, wherein *Bacteroidetes* are much greatly found. Great dominance is denoted by values highlighted in red and slight dominance in yellow (Figure 1).

Table 2. The number of metagenomic reads analyzed using QIIME.

	Sample	Reads
Sea cucumber	BSC1	5014
	BSC2	5289
	BSC3	7285
	BSC4	4745
	BSC5	4001
	BSC6	1950
Shrimp	SH02	2443
	SH04	2129
	SH06	2301
	SH12	3310
	SH17	10723
	SH18	2550
	Total	51740



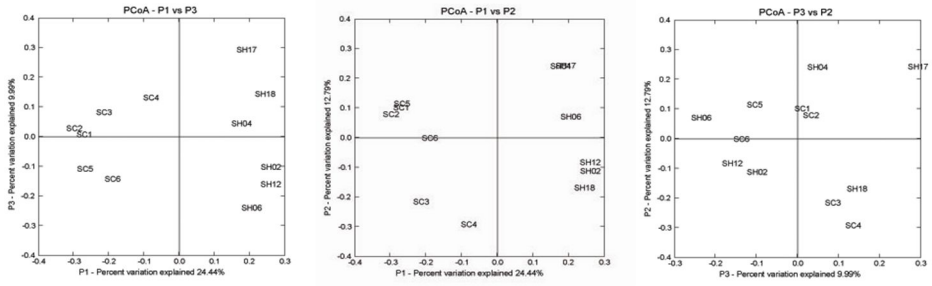
OTU	Genus	SC1	SC2	SC3	SC4	SC5	SC6	SH02	SH04	SH06	SH12	SH17	SH18
1428	Bd2-2	0.5%	1.3%	6.0%	38.0%	2.6%	6.3%	0.6%	0.0%	0.1%	0.2%	0.0%	0.0%
1963	Bd2-2	0.3%	0.6%	2.3%	14.2%	1.2%	0.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
920	Uncultured	0.2%	0.2%	1.5%	5.6%	0.2%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
444	Uncultured	0.0%	0.0%	0.1%	0.9%	0.2%	0.4%	4.0%	0.5%	4.0%	8.9%	0.1%	1.8%
1427	Lutibacter	8.0%	14.0%	10.3%	0.0%	0.3%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%
1730	Ns10_marine_group	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	2.2%	3.6%	0.0%	9.4%	0.5%
1163	Propionigenium	64.2%	65.3%	36.9%	0.4%	76.4%	52.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1677	Uncultured	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.8%	0.1%
664	Lactococcus	0.1%	0.0%	0.2%	0.5%	0.1%	0.0%	6.6%	1.9%	3.3%	4.0%	16.6%	9.0%
523	Candidatus_bacilloplasma	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.8%	0.0%	0.2%	28.7%	4.9%
545	Fusobacterium	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.4%	55.3%	0.0%	0.2%	3.4%	3.0%
1327	Vibrio	0.0%	0.2%	0.1%	0.0%	0.1%	0.0%	0.1%	0.4%	0.1%	0.2%	10.7%	8.4%
594	Vibrio	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	2.7%	5.9%
1121	Marinicella	0.0%	0.0%	0.1%	0.2%	0.1%	0.2%	3.7%	0.2%	1.8%	7.1%	0.1%	2.0%
1051	Thalassobius	1.0%	0.1%	0.6%	0.7%	0.2%	1.0%	1.6%	0.1%	0.4%	5.8%	0.0%	1.8%
1741	Crenothrix	4.6%	2.8%	12.0%	0.2%	3.2%	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
1636	Subsectioniii	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	12.4%	0.0%	0.0%	0.0%
1579	Merismopedia	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	0.0%	8.4%	0.0%	0.4%	0.6%
139	Prochlorococcus	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.9%	0.1%	8.2%	0.2%	0.2%	0.0%
151	Ta18	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	38.6%	25.4%	8.1%	0.2%	2.6%	2.0%
1635	Sulfurovum	0.0%	0.0%	0.1%	0.7%	0.0%	0.2%	0.0%	0.0%	2.7%	11.7%	0.0%	1.0%
		78.8%	84.5%	70.0%	61.4%	84.4%	62.3%	59.7%	87.1%	53.1%	38.8%	83.6%	40.9%

Figure 1. The bacterial composition from the intestine of sea cucumber and shrimp samples. Sea cucumber is in red and shrimp is in green.

Phylogenetic relationship of bacterial samples

The phylogenetic relationship of both samples to determine the bacterial diversity was exhibited in Figure 2. The abundance and type of bacteria were considered in weighted unifrac, while only the type of bacteria was considered in unweighted unifrac. Bacterial clusters were formed in sea cucumber samples, except for BSC4, and the relationship of bacteria among shrimp samples was found be distant from each other.

(a)



(b)

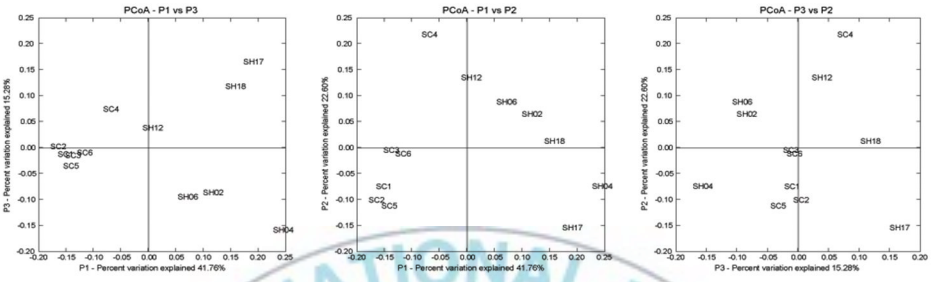
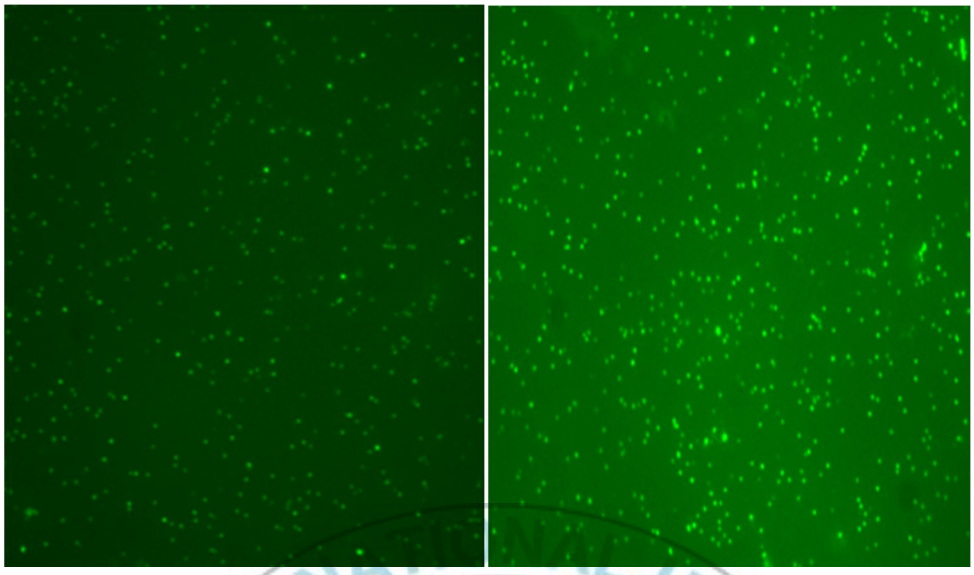


Figure 2. The unweighted unifrac is exhibited in (a), showing only the types of bacteria, while the unweighted unifrac is displayed in (b), considering both the abundance and types of bacteria.

Abundance of viruses from sea cucumbers and shrimps

Direct counting of virus-like particles was done using epifluorescence microcopy (EFM) after sequential filtration to estimate the abundance of viruses in sea cucumber and shrimp samples. About 4.08 $X10^9$ and 2.7 X 10^8 viruses per 1 g (wet weight) of samples were obtained, respectively (Figure 3).



(a) (b)

Figure 3. Viral particles (VLPs) stained with SYBR Gold and observed under an epifluorescence microscope (1000x). Viral particles from sea cucumbers are shown in (a) and viral particles from shrimps are shown in (b).

Pyrosequencing of viral metagenomes

Initially, a total of 106,034 reads for sea cucumber and 128,800 reads for shrimp were obtained from MDA libraries. After which, the pre-processing of the sequences such as data deduplication and quality control, was done using CAMERA and GALAXY web servers, respectively. Results show that a total of 84,620 and 76,340 reads, respectively, remained in MDA libraries (Table 3). Also, based on the NCBI non-redundant database, sea cucumber and shrimp sequences, respectively, were found to be 29.7% and 23.4% similar (Fig 4a). In this partition, 0.44% and 2.58% similarities were observed for the domain Archaea, 37.2% and 52.3% for the domain Bacteria, and 0.74% and 0.68% for the domain Eukarya, with regards to the sea cucumber and shrimp sequences, respectively (Fig 4b). Additionally, the CAMERA viral protein database was also used in comparing viral sequences from sea cucumber and shrimp sequences. Results reveal that the obtained viral sequences were classified into about 23 to 27 viral families, depending on the type of sample. However, in sea cucumber, most sequences were concerted only in seven families. dsDNA viruses from sea cucumber samples belong mostly to *Siphoviridae* (14.22%), *Podoviridae* (11.37%) and *Myoviridae* (8.88%) families, while ssDNA viruses belong mostly to *Microviridae* (52.52%), *Circoviridae* (6.30%), *Nanoviridae* (3.43%) and *Geminiviridae* (1.20%). As for shrimp samples, most sequences were concerted only in six families. dsDNA viruses belong to *Siphoviridae* (33.17%), *Myoviridae* (9.04%) and *Podoviridae* (5.17%), while ssDNA viruses belong to *Microviridae* (29.67%), *Circoviridae* (10.20%) and *Nanoviridae* (9.23%) (Fig 4c).

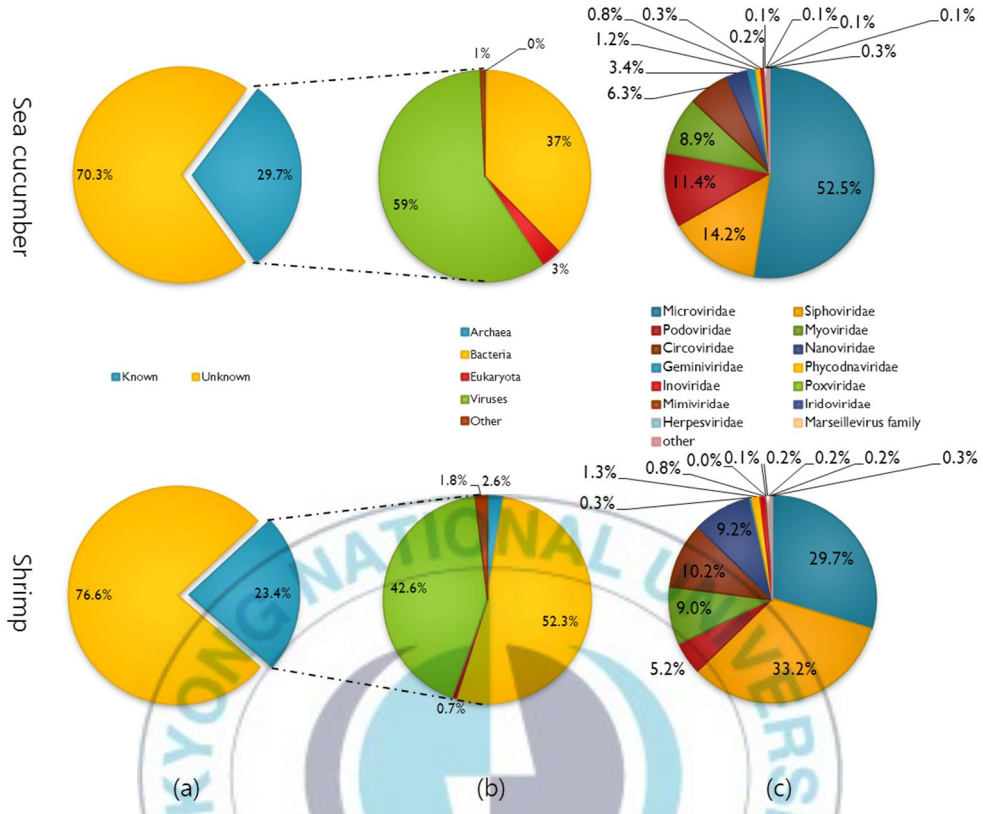


Figure 4. The composition and taxonomic affiliations of sea cucumber and shrimp virome reads. (a) and (b) were analyzed with NCBI non-redundant database, showing that more than half of the sequences are unknown sequences. (c) was analyzed using the CAMERA viral protein database.

Table 3. The number of reads obtained from MDA libraries after 454 pyrosequencing.

Processing	SC		SH	
Initial reads	106,034		128,800	
After deduplication	96,290	90.8%	84,669	65.7%
QC_q10 longest	91,479	86.3%	80,459	62.5%
QC_q10 longest longer than 100	84,620	79.8%	76,340	59.3%

Comparison with viromes from different environment

The viromes of both sea cucumber and shrimp samples were compared to the viral diversity of other communities such as Antarctic Lake freshwater and microbialites using Metavir (http://metavir-meb.univ-bpclermont.fr). These communities are divided into five groups: freshwater, seawater, two eukaryotes and this study group (34). Sea cucumber and shrimp samples have a slightly close relationship with microbialites and the Antarctic Lake(Figure 5). However, they still exhibit distant relationship from each other.

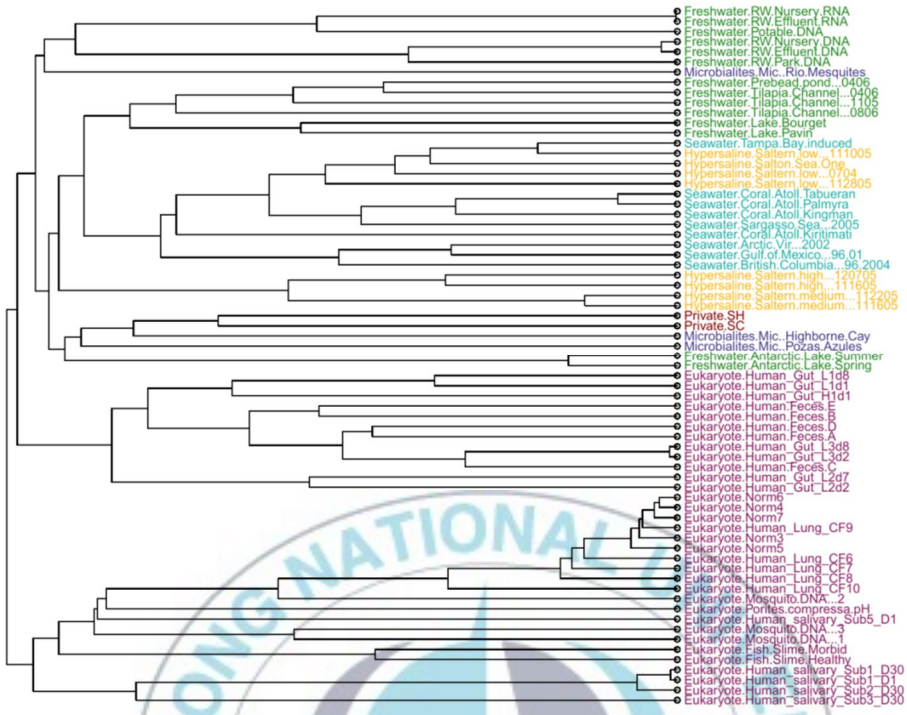


Figure 5. A constructed phylogenetic tree using the Metavir web server. The figure shows the comparison of viruses from marine invertebrates from that of the database.

Virus composition based on the database comparison

The viral composition of samples from microbialites, freshwater, seawater, and sea cucumber and shrimp samples was shown in Figure 6. It could then be observed that all samples possess both double stranded and single stranded DNA viruses. However, most part is dominated by single stranded DNA viruses. The individual composition of dsDNA and ssDNA viruses in both sea cucumber and shrimp samples were also exhibited in Figures 7 and 8, respectively.

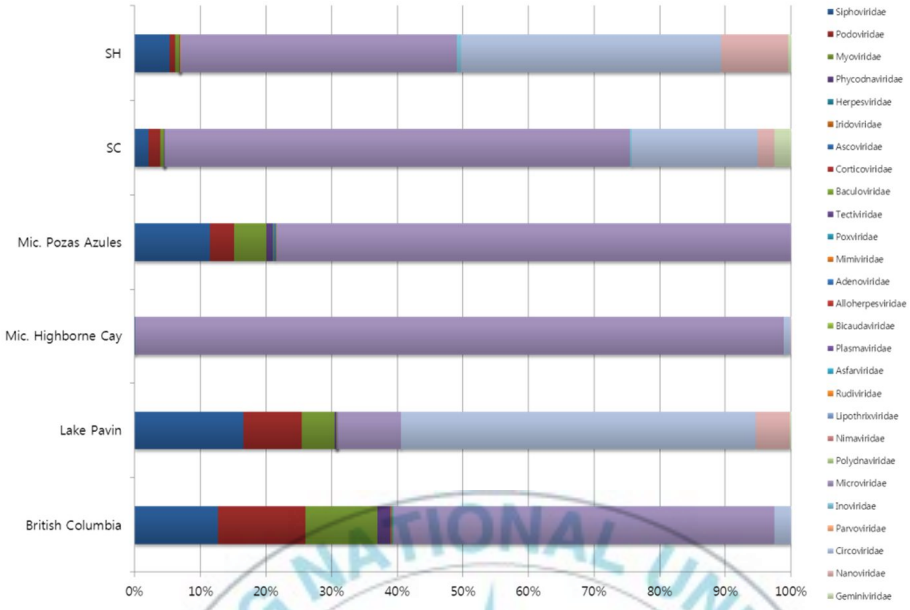


Figure 6. The composition of both dsDNA and ssDNA viruses in sea cucumber and shrimp samples. The viral composition of samples from microbialites, freshwater, seawater, and sea cucumber and shrimp samples was shown.

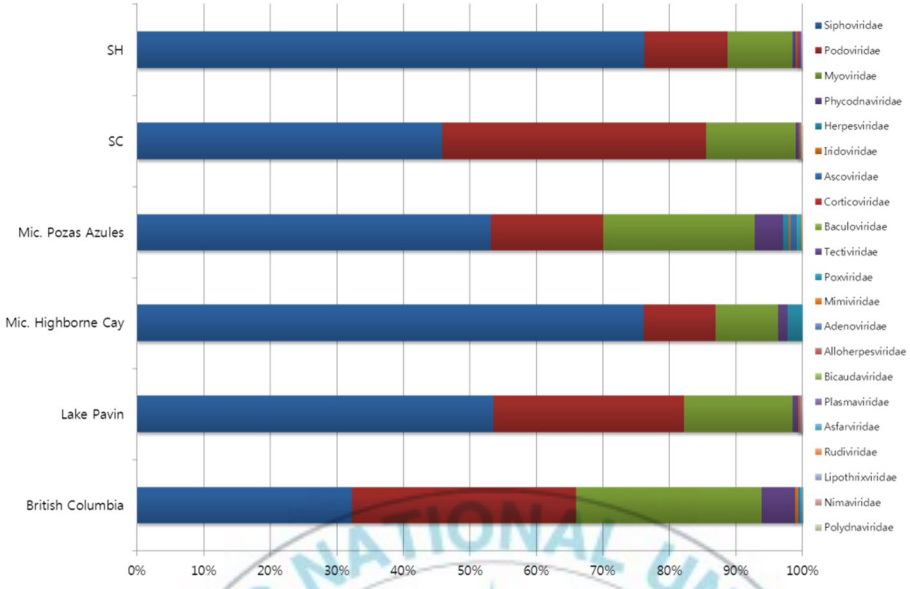


Figure 7. The composition of dsDNA viruses in sea cucumber and shrimp samples compared to related viromes.

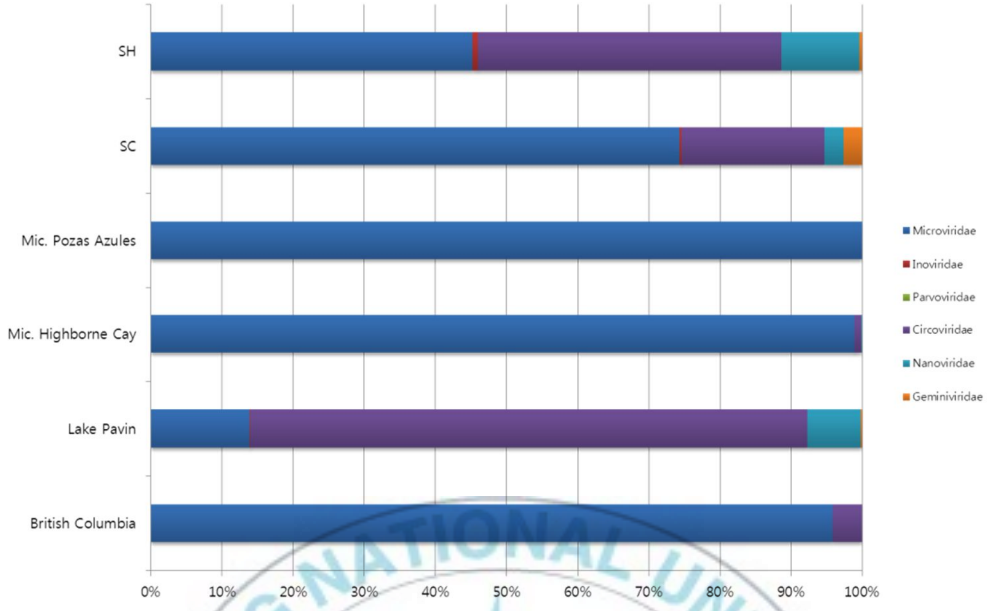


Figure 8. The composition of ssDNA viruses in sea cucumber and shrimp samples compared to related viromes.

The viral composition of sea cucumber and shrimp samples was compared using the MEGAN software. The blue color represents sea cucumber, while the red color represents shrimp. Figure 9 shows the distribution of single stranded DNA viruses between sea cucumber and shrimp samples. Four viral families dominate both samples, including *Circoviridae*, *Geminiviridae*, *Microviridae* and *Nanoviridae*. It could be seen from the figure that a greater percentage of single stranded DNA viruses is from shrimp samples, as dominated by the family *Circoviridae* and unclassified ssDNA viruses. However, circoviruses were found mostly in sea cucumber samples and not in shrimp samples, while the other members of the family *Circoviridae* dominating in shrimp samples remain unclassified. As for the family *Geminiviridae*, viruses from sea cucumber samples exhibit a larger part than shrimp samples. Same goes for the viral families *Microviridae* and *Nanoviridae*. The figure also shows that the unassigned members of family *Nanoviridae* were found only in shrimp samples, while *Densovirinae* members were found only in sea cucumber samples. As for double stranded DNA viruses, it could be noticed from figure 10 that most viruses found in both sea cucumber and shrimp samples belong to the order *Caudovirales*, mostly from the viral families *Myoviridae*, *Podoviridae* and *Siphoviridae*. Viruses from *Myoviridae* and *Podoviridae* were dominating in sea cucumber samples. As for the family *Siphoviridae*, almost equal distribution for both samples could be observed. But most of virus was present in the shrimps except unclassified *Siphoviridae*. Additionally, most of virus occupied the unassigned and unclassified members of the family.

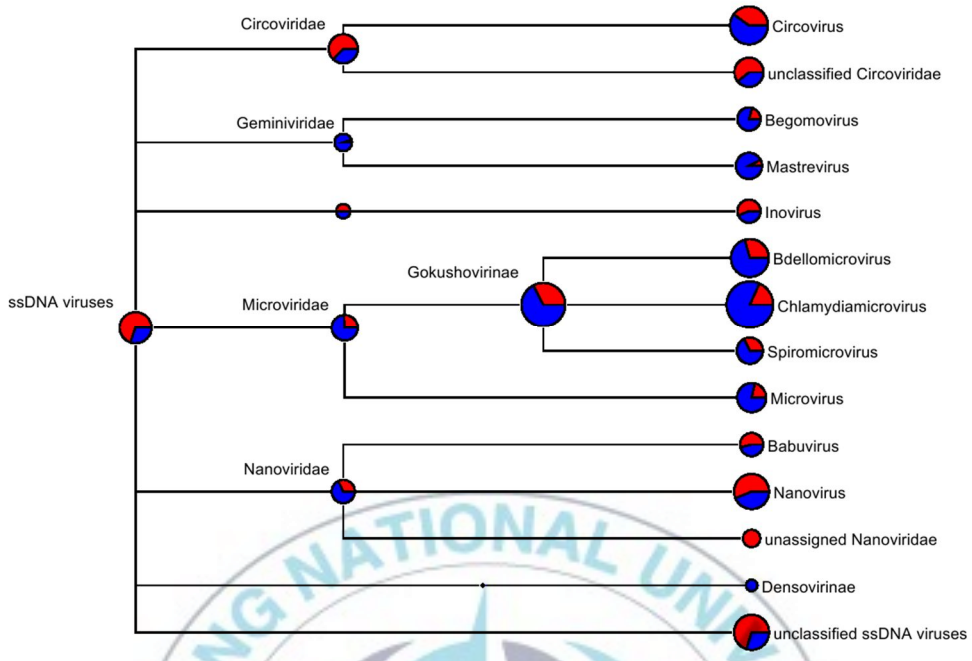


Figure 9. The distribution of single stranded DNA viruses in sea cucumber and shrimp samples.

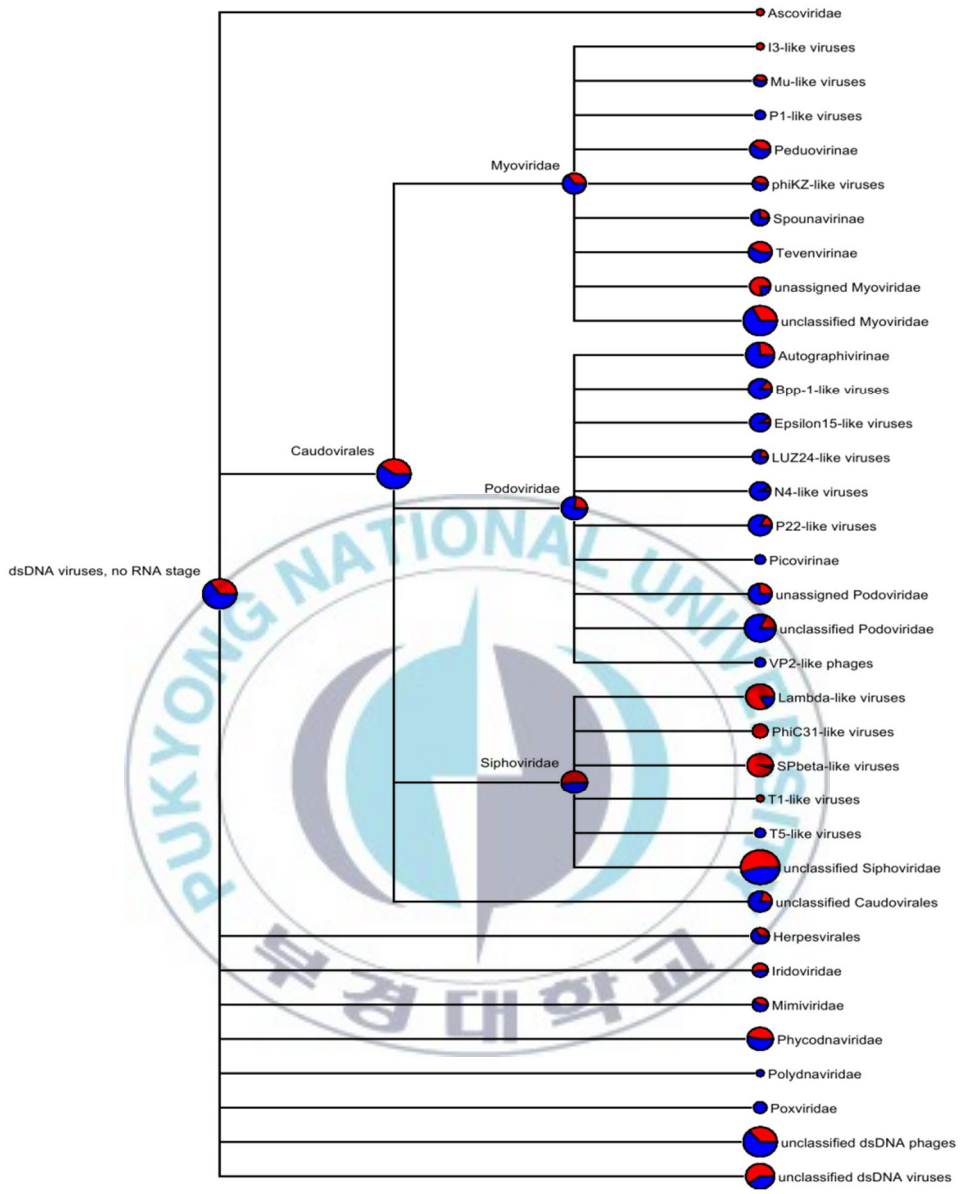


Figure 10. The distribution of double stranded DNA viruses in sea cucumber and shrimp samples.

Phylogenetic diversity of viral genes

Different marker genes were used to determine the phylogenetic relationship of the viral diversity observed in both sea cucumber and shrimp samples. Analysis of the phylogenetic diversity of viruses using Metavir revealed distant relationship among viruses. The viral diversity in Lake Pavin freshwater was used for comparison since the composition of both British Columbia seawater and microbialites has less double stranded DNA viruses, which is contradicting to that of Lake Pavin freshwater. It was known that the viral families Circoviridae and *Nanoviridae* were not reported in British Columbia seawater and microbialites. Thus, only Lake Pavin freshwater was available for Rep like gene comparison. The VP1 marker gene was used to determine the phylogenetic relationship of both sea cucumber and shrimp samples, Lake Pavin, British Columbia, Mic.PozasAzules and Mic.Highborne Cay with the other previously known viral groups (Fig 12). The blue color represents sea cucumber and the red color represents shrimp. The VP1 marker gene is basically for the family *Microviridae*. With this, it could be observed that all viruses in the phylogenetic tree belong to the *Microviridae* family. Figure 9 represents the phylogenetic relationship of viruses from sea cucumber, shrimp and Lake Pavin with previously known viruses using the Rep-like marker gene. The Rep marker gene can detect viruses belonging to three viral families including *Circoviridae*, *Nanoviridae* and *Geminiviridae*.

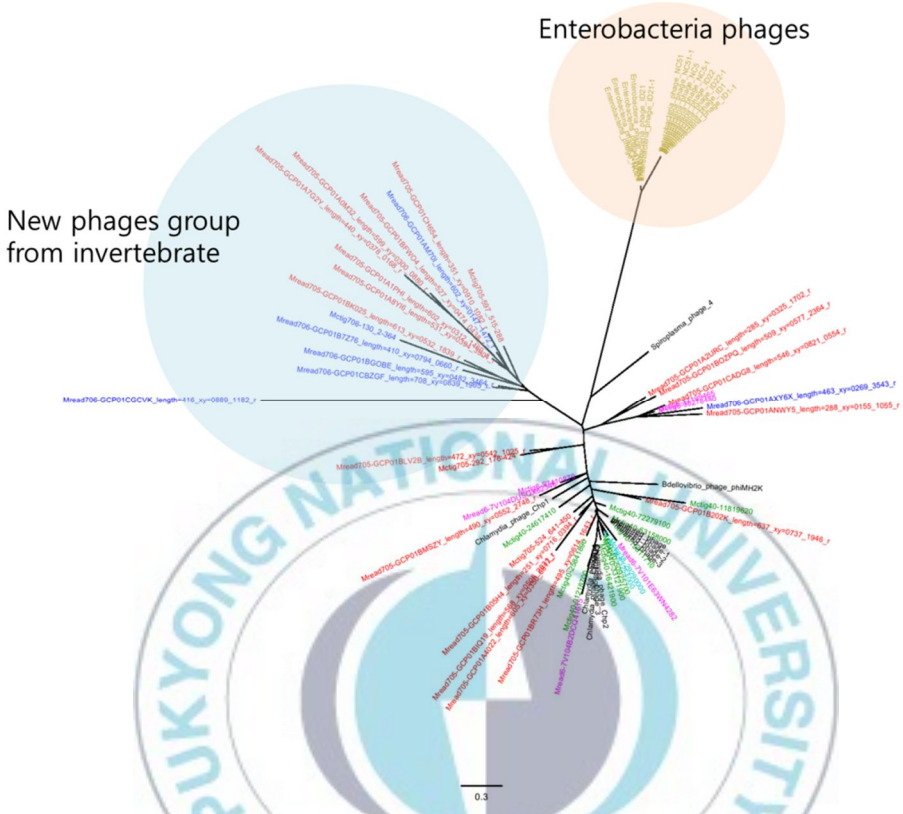


Figure 8. The phylogenetic relationship of the viral diversity in sea cucumber and shrimp samples using the VP1 marker gene.

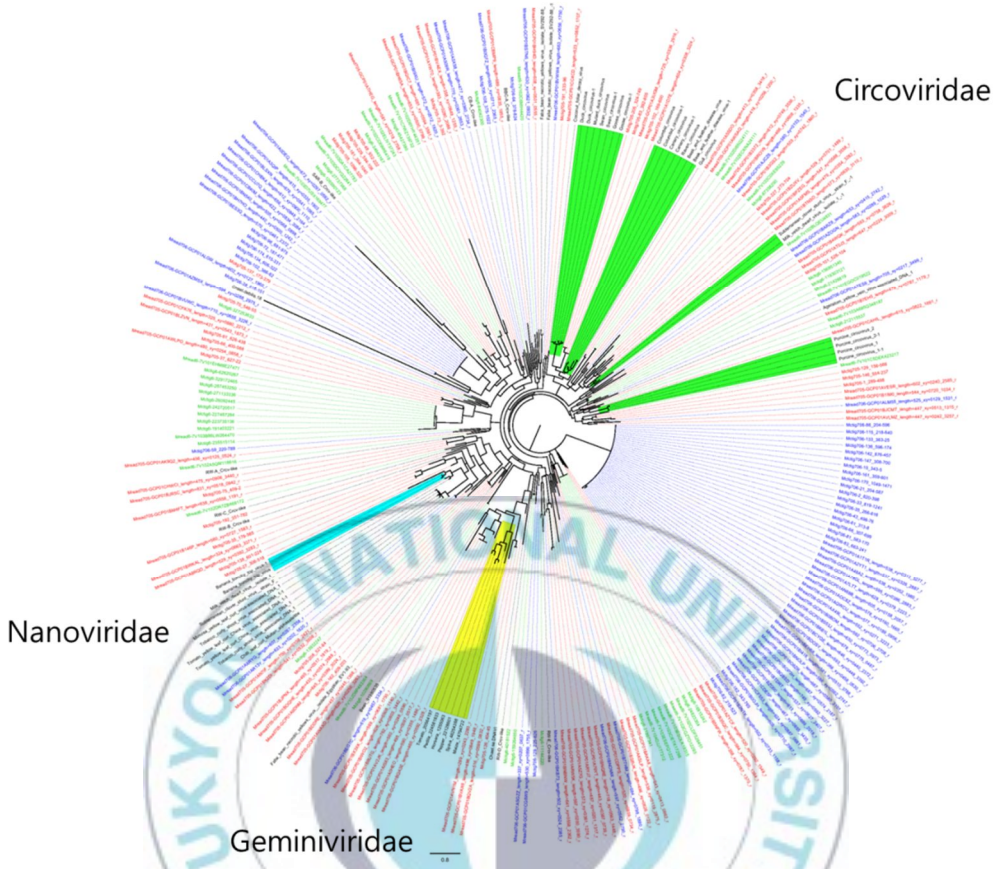


Figure 13. The phylogenetic relationship of the viral diversity in sea cucumber and shrimp samples using the Rep marker gene.

DISCUSSION

The diversity of both bacteria and viruses in marine invertebrates, particularly sea cucumbers and shrimps, were studied using metagenomic analyses. Much information remains unknown about the composition and diversity of both bacteria and viruses in marine invertebrates, thus, this study was conducted.

Sea cucumbers were found to be dominated by *Propionigenium*, excluding BSC4 which was dominated by *Bacteroidetes*. It can be seen from Table 1 that BSC4 has a much bigger size than the other sea cucumber samples. As BSC4 has a different bacterial composition compared to the other sea cucumber samples, it can be assumed that the sample size greatly contributes to the differences in bacterial composition and diversity among the samples. Moreover, as the size of the sample could be a great factor to the composition and diversity of bacteria, this assumption is not limited to marine invertebrates and may be applied to other organisms, as well. According to previous researches, some members of the bacterial genus *Propionigenium* were found in mud crabs (HE611102, NCBI), lugworms, marine sponges and sea urchins (2, 22, 25). Also, an uncultured bacterium with the strain name Bd2-2 dwelt in marine sponges and sediments (18, 22, 42).Strain Ta18, another uncultured bacterium, was known to inhabit in mud crab (HE186766) and a marine bacterioplanktons (GU170743). Furthermore, the bacterial genus *Fusobacterium* was isolated from white shrimps (EF186766). One of the samples used in this study is white shrimp, suggesting that the results obtained from this study are, by some means, matching to that of the previous studies.

The default beta diversity metrics of weighted and unweighted unifrac are phylogenetic measures used extensively in recent microbial community sequencing projects. The abundance and type of bacteria were

considered in weighted unifrac, while only the type of bacteria was considered in unweighted unifrac. In sea cucumber samples, bacterial clusters were formed, except in BSC4. In unweighted unifrac, sea cucumber shows differences in bacterial composition, while in weighted unifrac, close relationship was observed in connection to both bacterial diversity and abundance. This can be attributed to the fact that these samples were taken from the wild and they all belonged to the same environment. As for the relationship of bacteria among shrimp samples, it was found that bacteria isolated from shrimps form distinct assemblages based from the unweighted unifrac. For the weighted unifrac, in relation to bacterial diversity and abundance, it was found that bacteria from shrimp samples are distant from each other. This can be accounted to the fact that shrimp samples were taken from six different shrimp farms. Thus, the various environments they came from also differ from each other. Also, the growth conditions in these environments may also vary, mostly depending on the feeds that the shrimps rely on for food sources.

SYBR Gold staining was performed to count viral particles (VLPs) directly. The stained viruses were observed under epifluorescence microscopy (EFM) at 480-495nm. Around 4.08 $X10^9$ and 2.7 X 10^8 viruses per g (wet weight) of samples were obtained, respectively (Figure 1). Different studies have previously reported about the abundance of viruses from other communities. In aquatic environments, around 2.5 X 10^8 VLPs per milliliter of natural waters was counted (4). Consequently, around 1.2 X 10^6 VLPs per milliliter of seawater was observed in marine waters (15). Successively, in human feces, about 1.2 to 3.5 X 10^6 VLPs per gram of wet sample was noted (20).

With this, it can be observed that more viruses were isolated from marine invertebrates than in humans, considering that the intestine of marine invertebrates is much smaller than that of humans. In general, sea cucumbers are scavengers which feed on debris in the benthic zone of the

ocean. The diet of most sea cucumbers typically includes planktons, and dead and decaying organic matter found in the ocean. On the other hand, most shrimps are omnivorous which consumes a variety of materials as significant food sources in their natural diet. Also, some filter feeding shrimps use their setose (bristly) legs to gather phytoplanktons, while some scrape algae from rocks. With this, as sea cucumbers and shrimps are known to be filter feeders that sieve suspended particles matter and food particles from water (usually by passing the water over a specialized filtering structure), and as they inhabit the marine environment which is an open system consisting of a wide range of species, they eventually take in the species richness of this ecosystem.

Comparison of the sample sequences to that of NCBI non-redundant database revealed that sea cucumber and shrimp sequences, respectively, has a similarity of 29.7% and 23.4% (Figure 4a). In addition, more than half of sequences are unknown like in previous studies of viral metagenomics (19, 20, 32). Also, these results show that more research is needed to understand the problem in the partition of unknown sequences. In this partition, 0.44% and 2.58% similarities were observed for the domain Archaea, 37.2% and 52.3% for the domain Bacteria, and 0.74% and 0.68% for the domain Eukarya, with regards to the sea cucumber and shrimp sequences, respectively (Figure 4b). This result shows consistency with previous studies as it points out a strong implication about the lack of viral gene annotation and horizontal gene transfer between viruses and host genomes (8, 30).

Likewise, the CAMERA viral protein database was also used to compare sample sequences, showing that the obtained viral sequences belong to about 23 to 27 viral families (Fig 4c). Based on results, viral families isolated from invertebrates include *Iridoviridae*, *Parvoviridae*, *Polydnaviridae* and *Baculoviridae* for both sea cucumber and shrimp sequences. However, sequences happen to appear only in small counts since MDA generally provides an even representation of genomes except at the

ends certain genomes such as those which are small and circular, or large and linear. With this, sequences may be preferentially amplified (30, 45).

Comparison was done among the viromes of samples to that of other communities (Antarctic Lake and microbialites) using Metavir. The samples how a slightly close relationship with microbialites and the Antarctic Lake (Fig 5), but still, exhibit distant relationship from each other. Moreover, it could be seen that viruses in hypersaline salterns and seawater somehow possess similarity to the samples. Hypersaline salterns and the Antarctic lake freshwater are also known to be extreme environments, which is contradicting to the environments where sea cucumber and shrimp samples were taken from. Antarctic lake freshwater is known to have an extreme cold temperature, while hypersaline salterns contain high concentration of sodium chloride or mineral salts, surpassing that of the ocean water (12, 14). Therefore, Lake Pavin and British Columbia were used to compare the diversity of viruses from freshwater and seawater environments, respectively.

The viral composition of the samples was compared to that of the microbialites, freshwater, and seawater (Fig 6). All samples are composed of both double stranded and single stranded DNA viruses, as dominated by single stranded DNA viruses. Figure 7 shows the distribution of double stranded DNA viruses found in different communities. Viral families *Siphoviridae*, *Podoviridae* and *Myoviridae*, respectively, seem to occupy a large part in the viral diversity of all samples, exhibiting almost similar proportions. As for single stranded DNA viruses in Figure 8, the viral composition of microbialites and British Columbia seawater seem to be similar, and the viral composition of Lake Pavin freshwater tends to be different since microbialites and British Columbia seawater are both marine related environments, while Lake Pavin is a freshwater environment, which is not of the environment as the other two. With this, both sea cucumber and shrimp samples exhibit distant similarity with the viral composition of the

microbialites and British Columbia seawater since they are all related to the marine ecosystem.

The distribution of single stranded DNA viruses between sea cucumber and shrimp samples were analyzed using the MEGAN software (Figure 9). However, this result does not represent the exact distribution of viruses in both samples, suggesting that there could be a probable presence of host specific virus in sea cucumbers and shrimp samples, respectively. Four viral families dominate both samples, including *Circoviridae*, *Geminiviridae*, *Microviridae* and *Nanoviridae*. It could be seen from the figure that a greater percentage of single stranded DNA viruses is from shrimp samples, as dominated by the Family *Circoviridae* and unclassified ssDNA viruses. However, circoviruses were found mostly in sea cucumber samples and not in shrimp samples, while the other members of the Family *Circoviridae* dominating in shrimp samples remain unclassified. As for Family *Geminiviridae*, viruses from sea cucumber samples exhibit a larger part than shrimp samples. Same goes for the viral families *Microviridae* and *Nanoviridae*. The figure also shows that the unassigned members of Family *Nanoviridae* were found only in shrimp samples, while *Densovirinae* members were found only in sea cucumber samples. However, this result does not represent the exact distribution of viruses in both samples, suggesting that there could be a probable presence of host specific virus in sea cucumbers and shrimp samples, respectively. As for double stranded DNA viruses, it could be noticed from Figure 10 that most viruses found in both sea cucumber and shrimp samples belong to Order *Caudovirales*, mostly from the viral families *Myoviridae*, *Podoviridae* and *Siphoviridae*. Viruses from *Myoviridae* and *Podoviridae* were dominating in sea cucumber samples. As for Family *Siphoviridae*, almost equal distribution for both samples could be observed. But most of virus was present in the shrimps except unclassified *Siphoviridae*. Additionally, most of virus occupied the unassigned and unclassified members of Family.

To determine the phylogenetic diversity of viruses from both samples, different marker genes were used using Metavir, exhibiting distant relationship among viruses. The VP1 marker gene is used to study Family *Microviridae* (Fig 12). It can be noticed from the phylogenetic tree that all viruses are members of the *Microviridae* family. It could be noted how diverse viruses are, considering a small amount of samples examined in this study. Also, the phylogenetic tree shows that the relationship of viruses is not close, although they have almost similar characteristics. Additionally, it could be stated that viruses isolated from both sea cucumber and shrimp samples are suspected to be new viral species. Likewise, the Rep marker gene was used to determine the phylogenetic relationship of viruses from sea cucumber, shrimp and Lake Pavin. Three viral families (*Circoviridae*, *Nanoviridae* and *Geminiviridae*) are detected using the Rep gene marker (Fig 9). Only small amount of samples were used in this study. However, diversity of viruses based from the results is unexpected, such as in VP1 and as exhibited in the figure. In addition, it can be noticed that some viruses isolated from the samples do not belong to the three existing families, suggesting that these viruses could be under new groups, which have not been studied previously. Thus, it can be assumed that marine invertebrates can have a great value as genetic resources that can be used in further studies.

국문 초록

현재까지 해양 환경에서 세균과 바이러스의 다양성에 대한 많은 연구가 진행이 되었다. 하지만 해양무척추동물과 공생하는 미생물의 다양성에 대한 연구는 아직 미비하였다.

본 연구에서는 metagenomics 를 통해 해삼과 새우의 내장에 존재하는 미생물의 다양성을 연구하였다. Power Soil DNA Extraction Kit (MOBIO)를 이용하여 세균의 DNA 를 추출하였으며, 추출된 DNA 를 barcoded primers (27F 와 518R)를 이용하여 PCR 을 진행하였다. 얻어진 PCR 산물을 차세대 서열분석 방법인 454 pyrosequencing 을 진행하였고, 얻어진 결과를 QIIME 프로그램을 이용하여 분석하였다. 여러 결과들을 통해 해삼과 새우와 공생하는 주요 세균들의 구성이 다른 것을 확인할 수 있었다. 해삼에는 *Propionigenium* 이, 새우에는 *Lactococcus* 와 *Fusobacterium* 이 대부분을 차지했다.

순수한 바이러스는 연속적인 여과방법을 통해 분리되었으며, SYBR gold 염색법을 이용해 샘플 그람 당 약 108~109 개의 바이러스가 존재하는 것을 확인하였다. 순수하게 분리된 바이러스로부터 DNA 를 추출하였고 multiple displacement amplification 방법을 통해 DNA 를 증폭시킨 후 454 pyrosequencing 을 진행하였다. 그 결과, 해삼과 새우에 각각 Double stranded DNA 바이러스는 *Siphoviridae* (14.2%, 33.2%), *Podoviridae* (11.3%, 5.2%) 그리고 *Myoviridae* (8.9%, 9%), single stranded DNA 바이러스는 *Microviridae* (52.5%, 29.7%), *Circoviridae* (6.3%, 10.2%), *Nanoviridae* (3.4%, 9.2%) 가 대부분을 차지하는 것을 확인할 수 있었다. 계통학적 분석을 통해 확인해본 결과 이들은 대부분이 알려지지 않은 바이러스였으며, 기존에 연구된 바이러스들과의 관계가 상당히 먼 것을 확인할 수 있었다. 또한, 새우에만 공생하는 새로운 바이러스 그룹과 해삼과 새우 모두에 공생하는 바이러스 그룹을 확인할 수 있었으며, 이들은 해양무척추동물과 공생하는 바이러스라고 생각된다. 바이러스의 유전적 다양성을 고려할 때 이들이 가진 유전자는 유전적 자원으로서 높은 가치가 있을 것이라 생각된다.

장내에 공생하는 미생물은 숙주의 건강 및 성장과의 관련이 깊기 때문에 이들을 이해하는 것은 프로바이오틱스 연구의 기초자료가 되어 양식업에 많은 도움이 될 수 있을 것이라 생각된다.

ACKNOWLEDGMENTS

어느덧 학창시절의 마지막 겨울이 찾아 왔습니다. 여러 사람들이 말했던 것처럼 2 년동안의 대학원 생활이 저 역시도 정말 짧게 느껴지는 것 같습니다. 더 하고 싶은 것이 많았지만 다 할 수 없어 아쉬움이 많이 남습니다. 많이 부족한 저를 이렇게 논문을 쓰고, 졸업을 할 수 있도록 항상 옆에서 이끌어주신 김경호 교수님께 진심으로 감사의 말씀을 드립니다. 그리고 제가 이렇게 계속 공부를 할 수 있도록 인도해주신 이훈구 교수님과 김군도 교수님께도 깊이 감사 드립니다. 바쁘신 와중에도 이 논문을 완성할 수 있게 많은 도움을 주신 최태진 교수님께도 감사 드리며, 항상 웃으며 저를 맞아주시는 이명숙 교수님, 많은 가르침을 주신 송영환 교수님, 김영태 교수님께도 감사 드립니다.

6 년이라는 학교 생활을 하면서 정말 소중한 추억을 많이 만들었습니다. 3 년이란 시간을 같이 보낸 최고의 파트너 영삼이, 혼자 외로웠던 나에게 힘이 되어준 동현이, 선배 잘못 만나 많은 눈물을 흘렸던 첫 제자 Coy, 보다 일찍 친해졌어야 했는데 언어 때문에 다가가기 힘들었던 Dr. Kasin, 항상 열심히 하는 모습이 멋있었던 진수형, 매일 늦은 시간까지 남아 공부하는 열혈청년 우석이형, 2 년동안 많이 챙겨준 순진이, 점점 예뻐지는 미정이, 시답잖은 개그에도 빵빵 터져주는 덕현이, 털털하면서도 귀여운 초원이, 멍한게 매력인 난희, 바른생활 사나이 창원이, 항상 적극적이고 열심히인 민석이, 자주 찾아가서 이야기도하고 같이 놀면서 친해졌어야 했는데 그러지 못해서 아쉬운 종용이와 지혜, 모르는게 많아 귀찮게 만들었던 학과 조교 미진이와 수경이, 학교 입학때 부터 같이 학교생활을 해온 순수남 한성이, 대학원생활에 많은 즐거움을 만들어준 현철이와 현우, 또 학과 생활에 있어 많은 격려와 힘이 되어준 선배, 후배님들에게도 감사의 말씀을 전합니다. 바쁘다는 핑계로 자주 만나지 못하지만 그래도 항상 힘이 되어주는 친구들 모두들 고맙다.

마지막으로 사랑하는 우리 가족. 항상 저를 믿고 사랑해주신 아버지, 어머니 죄송하고, 감사드립니다. 항상 아버지와 어머니를 웃게 만들어 드릴 효자가 되도록 하겠습니다. 사랑합니다. 철없는 동생 챙긴다고 일찍부터 고생이 많았던 누나 너무 고맙고 이 은혜는 또미에게 갚을께. 사랑해.

제가 이렇게 글을 쓸 수 있게 도와주신 모든 여러분께 감사의 말씀을 드립니다. 감사합니다.

REFERENCES

1. Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. A. Suttle, and F. Rohwer. 2006. The marine viromes of four oceanic regions. PLoS Biol 4:e368.
2. Ashforth, E. 2009. Bacterial diversity in an arenicola marina L. bioturbated mesocosm.
3. Bench, S. R., T. E. Hanson, K. E. Williamson, D. Ghosh, M. Radosovich, K. Wang, and K. E. Wommack. 2007. Metagenomic characterization of Chesapeake Bay virioplankton. Appl Environ Microbiol 73:7629-41.
4. Bergh, Ø., K. Y. BØrsheim, G. Bratbak, and M. Helda. 1989. High abundance of viruses found in aquatic environments. Nature 340:467-468.
5. Blankenberg, D., A. Gordon, G. Von Kuster, N. Coraor, J. Taylor, and A. Nekrutenko. 2010. Manipulation of FASTQ data with Galaxy. Bioinformatics 26:1783-5.
6. Costanza, R. 1999. The ecological, economic, and social importance of the oceans. Ecological Economics 31:199-213.
7. Culley, A. I., A. S. Lang, and C. A. Suttle. 2003. High diversity of unknown picorna-like viruses in the sea. Nature 424:1054-7.
8. Delwart, E. L. 2007. Viral metagenomics. Rev Med Virol 17:115-31.
9. Desnues, C., B. Rodriguez-Brito, S. Rayhawk, S. Kelley, T. Tran, M. Haynes, H. Liu, M. Furlan, L. Wegley, B. Chau, Y. Ruan, D. Hall, F. E. Angly, R. A. Edwards, L. Li, R. V. Thurber, R. P. Reid, J. Siefert, V. Souza, D. L. Valentine, B. K. Swan, M. Breitbart, and F. Rohwer. 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature 452:340-3.
10. Fry, J. 2000. Bacterial diversity and 'unculturables'. Microbiology today 27.
11. Gomez-Alvarez, V., T. K. Teal, and T. M. Schmidt. 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J 3:1314-7.
12. Gordon, D. A., J. Priscu, and S. Giovannoni. 2000. Origin and Phylogeny of Microbes Living in Permanent Antarctic Lake Ice. Microb Ecol 39:197-202.
13. Gowda, N. M., U. Goswami, and M. I. Khan. 2008. T-antigenbindinglectin

with antibacterialactivity from marineinvertebrate, seacucumber (Holothuria scabra): Possibleinvolvement in differentialrecognition of bacteria. Journal of Invertebrate Pathology 99:141-145.

14. Gunde-Cimermana, N., P. Zalarb, S. de Hoogc, and A. Plemenitasd. 2000. Hypersaline waters in salterns - natural ecological niches for halophilic black yeasts. FEMS Microbiol Ecol 32:235-240.
15. Hara, S., K. Terauchi, and I. Koike. 1991. Abundance of viruses in marine waters: assessment by epifluorescence and transmission electron microscopy. Appl Environ Microbiol 57:2731-4.
16. Hirata, Y., and D. lJemura. 1986. Halichondrins-antitumor polyether macrolides from a marine sponge. Pure and Applied Chemistry 58:701-710.
17. Jones, D. V., Jr., J. A. Ajani, R. Blackburn, K. Daugherty, B. Levin, Y. Z. Patt, and J. L. Abbruzzese. 1992. Phase II study of didemnin B in advanced colorectal cancer. Invest New Drugs 10:211-3.
18. Kennedy, J., C. E. Codling, B. V. Jones, A. D. Dobson, and J. R. Marchesi. 2008. Diversity of microbes associated with the marine sponge, Haliclona simulans, isolated from Irish waters and identification of polyketide synthase genes from the sponge metagenome. Environ Microbiol 10:1888-902.
19. Kim, K. H., H. W. Chang, Y. D. Nam, S. W. Roh, M. S. Kim, Y. Sung, C. O. Jeon, H. M. Oh, and J. W. Bae. 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. Appl Environ Microbiol 74:5975-85.
20. Kim, M. S., E. J. Park, S. W. Roh, and J. W. Bae. 2011. Diversity and abundance of single-stranded DNA viruses in human feces. Appl Environ Microbiol 77:8062-70.
21. Lee, S.-W. 2005. Metagenome, the Untapped Microbial Genome, toward Discovery of Novel Microbial Resources and Application into the Plant Pathology. The Plant Pathology Journal 21.
22. Li, C. Q., W. C. Liu, P. Zhu, J. L. Yang, and K. D. Cheng. 2011. Phylogenetic diversity of bacteria associated with the marine sponge Gelliodes carnosa collected from the Hainan Island coastal waters of the South China Sea.

Microb Ecol 62:800-12.

23. Lopez-Bueno, A., J. Tamames, D. Velazquez, A. Moya, A. Quesada, and A. Alcami. 2009. High diversity of the viral community from an Antarctic lake. Science 326:858-61.

24. McDaniel, L., M. Breitbart, J. Mobberley, A. Long, M. Haynes, F. Rohwer, and J. H. Paul. 2008. Metagenomic analysis of lysogeny in Tampa Bay: implications for prophage gene expression. PLoS One 3:e3263.

25. Meziti, A., K. A. Kormas, M.-A. Pancucci-Papadopoulou, and M. Thessalou-Legaki. 2007. Bacterial phylotypes associated with the digestive tract of the sea urchin Paracentrotus lividus and the ascidian Microcosmus sp. *Russian Journal of Marine Biology* 33:84-91.

26. Moss, S. M., B. R. LeaMaster, and J. N. Sweeney. 2000. Relative Abundance and Species Composition of Gram-Negative, Aerobic Bacteria Associated with the Gut of Juvenile White Shrimp Litopenaeus vannamei Reared in Oligotrophic Well Water and Eutrophic Pond Water. Journal of the World Aquaculture Society 31:255-263.

27. Niu, B., L. Fu, S. Sun, and W. Li. 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. BMC Bioinformatics 11:187.

28. NL Thakur, A. T., WEG Muller 2005. Marine natural products in drug discovery.

29. Park, E. J., K. H. Kim, G. C. Abell, M. S. Kim, S. W. Roh, and J. W. Bae. 2010. Metagenomic analysis of the viral communities in fermented foods. Appl Environ Microbiol 77:1284-91.

30. Pinard, R., A. d. Winter, G. J. Sarkis, M. B. Gerstein, K. R. Tartaro, R. N. Plant, M. Egholm, J. M. Rothberg, and J. H. Leamon. 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Bioinformatics 7:216.

31. Rakesh Sharma, R. R., Raj Kishor Kapardar and Amit Grover. 2005. Unculturable'bacterial diversity: an untapped resource. CURRENT SCIENCE 89.

32. Roux S, E. F., Robin A, Ravet V, Personnic S, et al. 2012. Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLoS One 7:e33641.

33. Roux, S., F. Enault, A. Robin, V. Ravet, S. Personnic, S. Theil, J. Colombet, T. Sime-Ngando, and D. Debroas. 2012. Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLoS One 7:e33641.
34. Roux, S., M. Faubladier, A. Mahul, N. Paulhe, A. Bernard, D. Debroas, and F. Enault. 2011. Metavir: a web server dedicated to virome analysis. Bioinformatics 27:3074-5.
35. Schwartsmann, G., A. B. d. Rocha, R. G. Berlinck, and J. Jimeno. 2001. Marineorganisms as asource of newanticanceragents. Elsevier 2:221–225.
36. Shaofeng Han, Y. L., Zhigang Zhou, Suxu He,Yanan Cao, Pengjun Shi, BinYao & Einar Ringo. 2010. Analysis of bacterial diversity in the intestine of grass carp (Ctenopharyngodon idellus) based on 16S rDNA gene sequences. Aquaculture Research 42:47-56.
37. Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A 103:12115-20.
38. Sonia R. Vartoukian, R. M. P. W. G. W. 2010. Strategies for culture of 'unculturable' bacteria. FEMS Microbiol Lett 309.
39. Steward, G. F., and C. M. Preston. 2011. Analysis of a viral metagenomic library from 200 m depth in Monterey Bay, California constructed by direct shotgun cloning. Virol J 8:287.
40. *Surajit Das, P. S. L. a. S. A. K.* 2006. Marine microbial diversity and ecology: importance and future perspectives. CURRENT SCIENCE, 90.
41. Suttle, C. A., A. M. Chan, and M. T. Cottrell. 1991. Use of Ultrafiltration To Isolate Viruses from Seawater Which Are Pathogens of Marine Phytoplankton. Appl Environ Microbiol 57:721.
42. Tamminena, M., A. Karkmana, J. Coranderb, L. Paulinc, and M. Virta. 2011. Differences in bacterial community composition in Baltic Sea sediment in response to fish farming. Elsevier 313:15-23.
43. Tucker, K. P., R. Parsons, E. M. Symonds, and M. Breitbart. 2010. Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. ISME J 5:822-30.

44. Ventura, M., F. Turroni, C. Canchaya, E. E. Vaughan, P. W. O'Toole, and D. van Sinderen. 2009. Microbial diversity in the human intestine and novel insights from metagenomics. Front Biosci 14:3214-21.

45. Willner, D., M. Furlan, M. Haynes, R. Schmieder, F. E. Angly, J. Silva, S. Tammadoni, B. Nosrat, D. Conrad, and F. Rohwer. 2009. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One 4:e7370.