



High Diversity of Myocyanophage in Various Aquatic Environments Revealed by High-Throughput Sequencing of Major Capsid Protein Gene With a New Set of Primers

Weiguo Hou^{1*}, Shang Wang², Brandon R. Briggs³, Gaoyuan Li¹, Wei Xie⁴ and Hailiang Dong^{1,5*}

¹ State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing, China, ² CAS Key Laboratory of Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences (CAS), Beijing, China, ³ Department of Biological Sciences, University of Alaska Anchorage, Anchorage, AK, United States, ⁴ State Key Laboratory of Marine Geology, Tongji University, Shanghai, China, ⁵ Department of Geology and Environmental Earth Science, Miami University, Oxford, OH, United States

OPEN ACCESS

Edited by:

Hongyue Dang,
Xiamen University, China

Reviewed by:

Rui Zhang,
Xiamen University, China
Meng Li,
Shenzhen University, China

*Correspondence:

Weiguo Hou
weiguohou@cugb.edu.cn;
hail@126.com
Hailiang Dong
dongh@cugb.edu.cn

Specialty section:

This article was submitted to
Aquatic Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 20 January 2018

Accepted: 18 April 2018

Published: 03 May 2018

Citation:

Hou W, Wang S, Briggs BR, Li G,
Xie W and Dong H (2018) High
Diversity of Myocyanophage
in Various Aquatic Environments
Revealed by High-Throughput
Sequencing of Major Capsid Protein
Gene With a New Set of Primers.
Front. Microbiol. 9:887.
doi: 10.3389/fmicb.2018.00887

Myocyanophages, a group of viruses infecting cyanobacteria, are abundant and play important roles in elemental cycling. Here we investigated the particle-associated viral communities retained on 0.2 μm filters and in sediment samples (representing ancient cyanophage communities) from four ocean and three lake locations, using high-throughput sequencing and a newly designed primer pair targeting a gene fragment (~ 145 -bp in length) encoding the cyanophage gp23 major capsid protein (MCP). Diverse viral communities were detected in all samples. The fragments of 142-, 145-, and 148-bp in length were most abundant in the amplicons, and most sequences ($>92\%$) belonged to cyanophages. Additionally, different sequencing depths resulted in different diversity estimates of the viral community. Operational taxonomic units obtained from deep sequencing of the MCP gene covered the majority of those obtained from shallow sequencing, suggesting that deep sequencing exhibited a more complete picture of cyanophage community than shallow sequencing. Our results also revealed a wide geographic distribution of marine myocyanophages, i.e., higher dissimilarities of the myocyanophage communities corresponded with the larger distances between the sampling sites. Collectively, this study suggests that the newly designed primer pair can be effectively used to study the community and diversity of myocyanophage from different environments, and the high-throughput sequencing represents a good method to understand viral diversity.

Keywords: myocyanophage, MCP gene, primer, diversity, high-throughput sequencing

INTRODUCTION

T4virus is a genus of viruses that belongs to the order Caudovirales, the family Myoviridae, and the subfamily Tevenvirinae. T4 phages are the archetype of this genus, which efficiently infects *Escherichia coli*. Within this genus are also myocyanophages that infect cyanobacteria and have attracted special attention for their critical role in regulating cyanobacterial population

structure, mortality, and evolution (Clokie and Mann, 2006). Myocyanophages are also called T4-like cyanophages for their morphological resemblance to T4 phages. However, phylogenetic analyses based on DNA or amino acid sequences revealed that myocyanophages are distinct from T4 phages or other T4-like viruses (Hambly et al., 2001; Filee et al., 2005; Sullivan et al., 2010; Ignacio-Espinoza and Sullivan, 2012).

Ecologically, myocyanophages are highly diverse and abundant in open-ocean, coastal and estuarine environments, fresh water lakes, and paddy soils (Filee et al., 2005; Comeau and Krisch, 2008; Hellweger, 2009; Sullivan et al., 2010; Wang K. et al., 2011; Wang et al., 2016; Chow and Fuhrman, 2012; Ignacio-Espinoza and Sullivan, 2012; Marston et al., 2013). These viruses play a key role in geochemical cycling of carbon and other elements through their interactions with host bacteria. For example, infection by cyanophages inhibits CO₂ fixation in cyanobacteria (Bailey et al., 2004; Puxty et al., 2016); viral lysis of cyanobacteria releases cell contents to the organic carbon pool and accelerates carbon cycling in aquatic ecosystems (Weitz and Wilhelm, 2012). However, the photosynthetic genes (*PsbA*, *PsbD*) present in cyanophages can partially compensate for the loss of the carbon fixation ability of cyanobacteria and protect their hosts against photo-inhibition (Bailey et al., 2004; Hellweger, 2009; Rohwer and Thurber, 2009). Viruses also drive evolution of bacteria by their antagonistic interactions or by introducing new genetic information to bacteria (Lindell et al., 2004; Suttle, 2007; Ignacio-Espinoza and Sullivan, 2012; Marston et al., 2012).

Research in myocyanophage diversity will facilitate a better understanding of their ecological importance. Marker gene and metagenomic analyses are some of the major methods to reveal the viral diversity in various environments (Chow and Fuhrman, 2012; Mizuno et al., 2013; Ma et al., 2014; Labonté et al., 2015). Variable sequence regions (~400 to ~800 bp) within those genes encoding g23 major capsid protein (MCP), g20 portal protein, and g43 DNA polymerase gene, have been used in studies of environmental myocyanophage or other *T4virus* diversity (Mann, 2003; Filee et al., 2005; Dekel-Bird et al., 2013; Marston et al., 2013; Needham et al., 2013). However, comparatively long lengths of these genes are not appropriate for sedimentary DNA-based diversity study (ancient DNA) (Boere et al., 2011).

Ancient DNA obtained from sediments represents an important source of information on past biodiversity (Pedersen et al., 2015). Short, but variable, DNA markers (<200 bp)

were recommended to study past biodiversity (Sønstebo et al., 2010). For example, 60–84-bp length of mammal gene fragments and 10–146 bp length of plant gene fragments were amplified from sediment DNA to reconstruct the livestock farming and landscape histories, respectively (Giguët-Covex et al., 2014). In order to investigate evolution of myocyanophage communities from ancient DNA, we designed a pair of degenerate PCR primers targeting the MCP gene fragment with comparatively short amplicons (~145-bp). The primer pair was tested with a series of lacustrine and marine samples (Table 1 and Figure 1). In order to assess the effects of sequencing depth on viral diversity estimates, three sequencing efforts with different depths, including clone library sequencing, shallow Illumina sequencing, and deep Illumina sequencing, were conducted, which produced about 100, 1,000–2,000, and more than 50,000 reads, respectively. Thus, the aims of this study were to (1) evaluate the efficiency of the newly designed primers in measuring myocyanophage diversity in various aquatic environments (especially a sediment sample to reveal cyanophage diversity from more than 2,000 years ago), (2) assess the effects of sequencing depth of MCP gene on myocyanophage diversity estimates, and (3) compare the different myocyanophage communities from various aquatic environments.

MATERIALS AND METHODS

Primer Design

Multiple degenerate primer sets were designed using BioEdit v7.0.1 package¹ to amplify the myocyanophage g23 MCP gene fragments by identifying conserved sequences after aligning 20 gene sequences retrieved from GenBank. These MCP gene sequences are from 17 representative genomes, which are assigned to CLUSTER A and CLUSTER B (Ignacio-Espinoza and Sullivan, 2012), one freshwater myocyanophage genome, one environmental meta-genome, and one uncultured Mediterranean viral clone sequence. Hairpin check and melting temperature assessment were conducted using Oligo Analyzer 3.1². The primers were then used to amplify myocyanophage MCP genes from water and sediment samples described

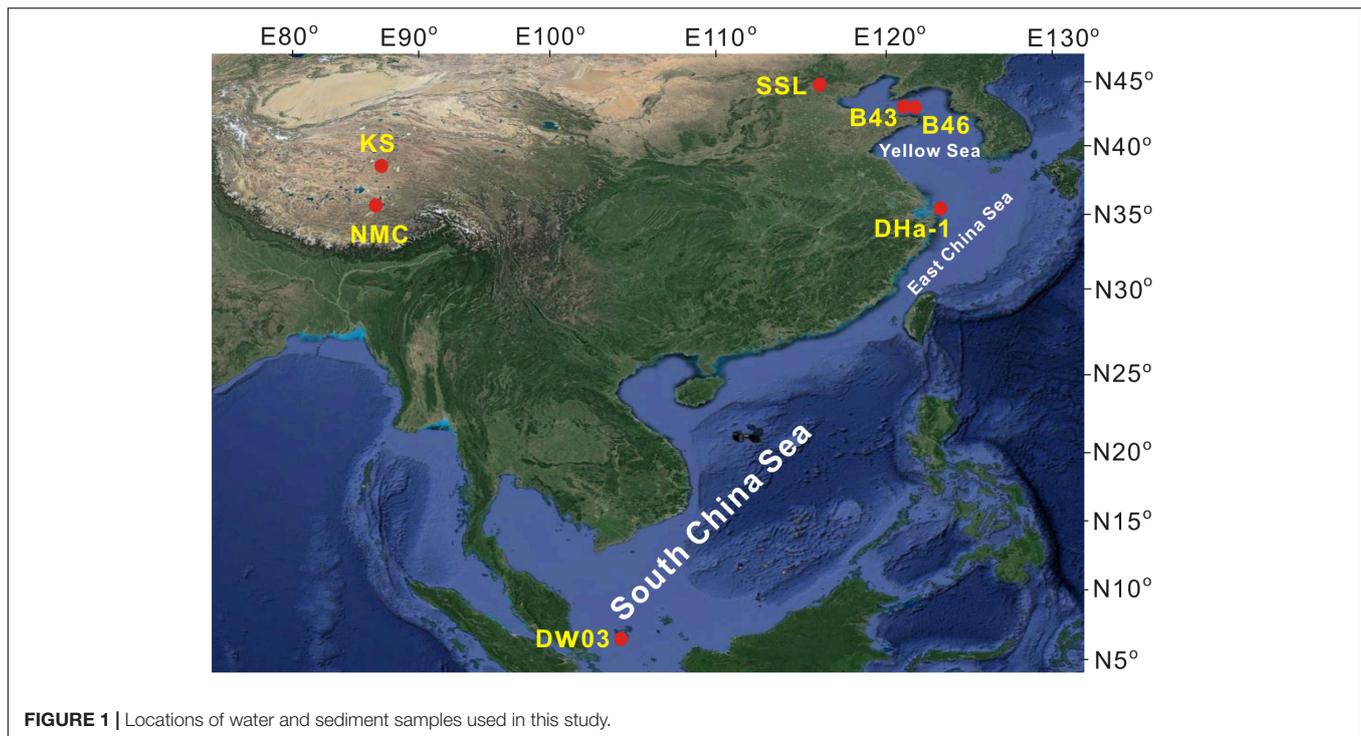
¹<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

²<http://sg.idtdna.com/calc/analyzer>

TABLE 1 | Sample location and description.

| Sample ID | Name or location | Latitude | Longitude | Sample type | Water type | Date collected |
|-----------|---------------------------------|----------|-----------|-------------------------------|-------------|------------------|
| NMC | Namco Lake on Tibetan Plateau | 30.51°N | 90.50°E | Water sample | Oligosaline | October 13, 2013 |
| KS1 | Kusai Lake on Tibetan Plateau | 35.69°N | 93.00°E | Surface sediment | Saline | October 28, 2010 |
| KS2* | Kusai Lake on Tibetan Plateau | 35.69°N | 93.00°E | 3.85 m below sediment surface | Saline | October 28, 2010 |
| SSL | Shisanling Reservoir in Beijing | 40.26°N | 116.25°E | Surface sediment | Freshwater | October 1, 2012 |
| B43 | Yellow Sea | 38.10°N | 122.00°E | Surface sediment | Seawater | August 12, 2011 |
| B46 | Yellow Sea | 37.90°N | 122.50°E | Surface sediment | Seawater | August 12, 2011 |
| DHa-1 | East China Sea | 30.50°N | 122.50°E | Surface sediment | Seawater | August 14, 2011 |
| DW03 | South China Sea | 2.95°N | 105.84°E | Water sample | Seawater | July 13, 2012 |

*The sample was obtained from a sediment core in October, 2010, with a ¹⁴C calibrated age of ~2,250 years BP. (Hou et al., 2014).



below (Table 1). The amplification efficiencies were assessed according to the brightness of the amplicon bands in agar gel after PCR and electrophoresis. The primer set mcp-821F (CTKGC DGARATYAACMGIGAART) and mcp-966R (ADDAG WCCYTTGAAYTTYTCAAC) targeting partial MCP gene was the most efficient set to amplify the gene (Supplementary Figure S1). On the MCP gene of *Synechococcus* cyanophage S-PM2, primer set mcp-821F-mcp-966R embraces the position from 748 to 893.

Sample Collection

Surface sediment and water samples were collected from diverse aquatic environments (Figure 1 and Table 1). The list included three lake samples from Tibetan Plateau (one water sample from NMC, two sediments from a KS sediment core, i.e., KS1 and KS2), one lake sediment from Beijing (SSL), two marine sediments from Yellow Sea (B43 and B46), one marine sediment from East China Sea (DHa-1), and one water sample from South China Sea (DW03). Surface sediment samples were collected with a gravity sediment sampler. Planktonic microorganisms and suspension particles were collected onto 0.2 μm membrane filters (Supor-200, Pall Life Sciences) by filtering through about 200 mL water. The above samples were used to analyze viral community (Table 1).

DNA Extraction, Sequencing Strategy

DNA from biomass-containing filters and sediment samples were extracted with modified methods by using MP Biomedical Fast DNA Spin kits (Hou et al., 2013). The polymerase chain reaction (PCR) mix contained 10 ng of template DNA, 2.5 μL rTaq reaction buffer, 400 nM of each primer, 200 mM dNTPs,

and 0.3 unit of rTaq polymerase (Takara, Dalian, China) in a 25 μL reaction system. The amplification procedure was as follows: an initial denaturation step at 95°C for 5 min, and 30 cycles of denaturing at 94°C for 30 s, annealing at 54°C for 30 s, and extension at 72°C for 30 s, followed by a final extension step at 72°C for 5 min. PCR products were purified with the Qiagen gel extraction kit (Qiagen, United States) and quantified on a QuBit 2.0 fluorometer (Invitrogen Corp., United States).

Major capsid protein fragments were sequenced with three strategies, i.e., clone-library Sanger sequencing (about 100 reads), shallow Illumina sequencing (1,000–2,000 reads), and deep Illumina sequencing (more than 100,000 reads). Clone sequencing was conducted by cloning the PCR products into pGEM®-T Easy Vector Systems (Promega Corporation). One hundred clones for each sample were sequenced with a BigDye Terminator v3.1 sequencing kit (Thermo Fisher Scientific Corporation). In order to sequence the MCP gene on Illumina Miseq platform, Illumina adapter, primer pad, linker, and barcodes were added to the 5' ends of primer pairs according to a previous publication (Caporaso et al., 2012). Equal amounts of PCR products from all samples were added to the sequencing pool, and the amounts for the shallow Illumina sequencing were reduced by 100 times relative to those used for deep Illumina sequencing. Deep and shallow Illumina sequencing experiments were both conducted with a Miseq V2 (300 cycles) kit by extending 200 cycles, with one primer consisting of the reverse complemented sequence of 3' Illumina adapter, reverse primer pad, and reverse primer linker, i.e., "CAAGCAGAAGACGGCATAACGAGAT AGTCAGCCAG CC" (Caporaso et al., 2012). Using this method, the sequencing reads covered the MCP gene fragments, the forward primer,

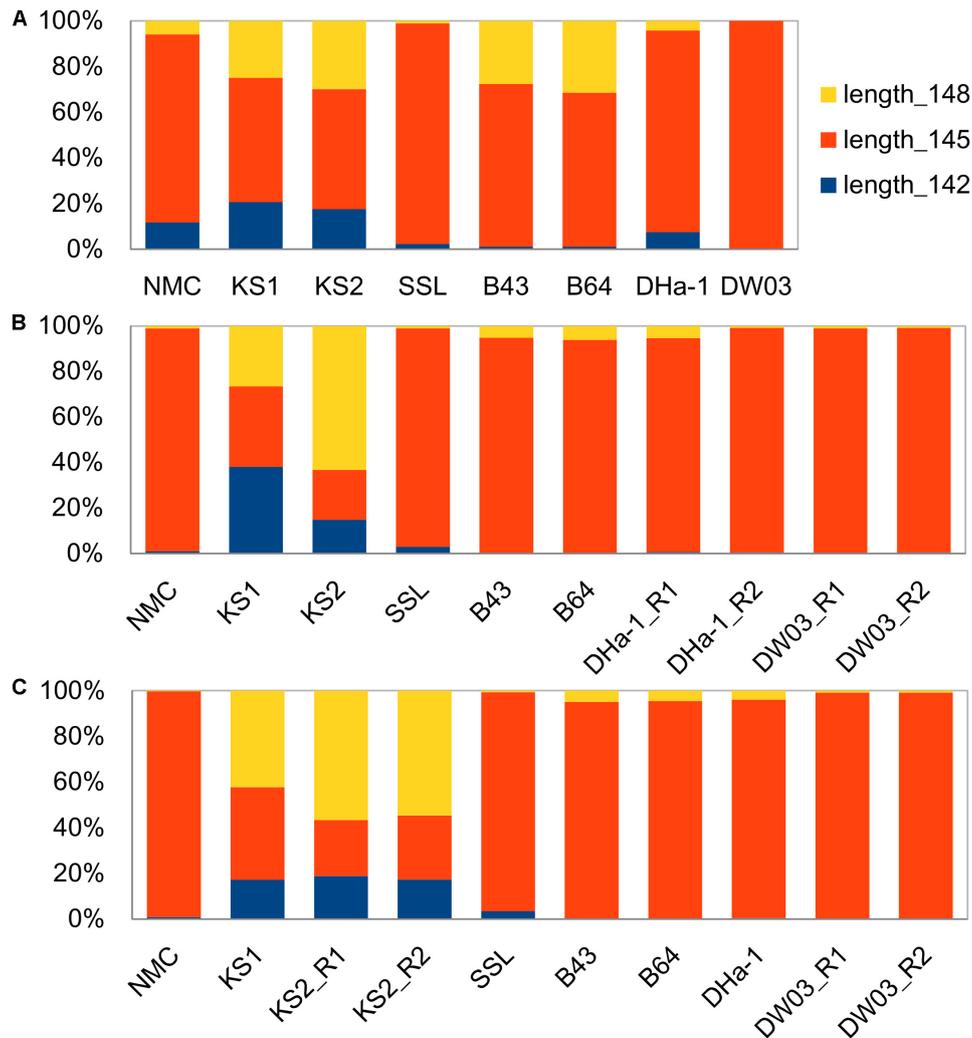


FIGURE 2 | Amplicon length distributions of MCP gene fragments from various samples with clone (A), shallow Illumina sequencing (B), and deep Illumina sequencing (C) methods. “R1” and “R2” for some samples refer to two sequencing replicates with different barcodes. Three bar-graphs share the same legends. The Y-axis is relative proportion of each length fragment.

forward primer linker, forward primer pad, barcode, and 5' Illumina adapter.

In the case of shallow Illumina sequencing strategy, the MCP gene was sequenced with a thousand sequences on a MiSeq platform. The MCP gene sequences obtained with clone sequencing were deposited in GenBank with accession numbers MG267122 – MG267315. High-throughput sequencing data could be found under a BioProject No. PRJNA393230.

Data Analysis

Illumina-sequencing errors were minimized by removing reads with low quality scores (<25). Cutadapt 1.10 was used to demultiplex raw sequences and to remove clone-vector and adapter sequences (Matin, 2011). CLC Sequence Viewer 7 (CLC Bio Qiagen, Aarhus, Denmark) was used to translate DNA sequences into amino acid sequences, which was then used

to verify the MCP gene. The following steps, including OTU assignment, alpha diversity, and beta diversity calculations, were subsequently completed using the QIIME software package (Caporaso et al., 2010). OTUs were defined at the identity levels of 80, 90, 95, and 97% using UCLUST (Edgar, 2010) based on nucleotide sequences. The first sequence within each OTU cluster was picked as a representative sequence. The representative sequences were aligned with MUSCLE and used to construct a tree using FastTree (Edgar, 2004). Representative sequences were checked by translating into amino acid sequences with CLC Sequence Viewer 7, and further confirmed by protein annotation with NCBI's Conserved Domain Database (Marchler-Bauer et al., 2017). An amino acid sequence-based maximum-likelihood tree was constructed to display the phylogenetic relations between these dominant representative sequences (representing top-50 abundant OTUs) and reference sequences retrieved from NCBI database using

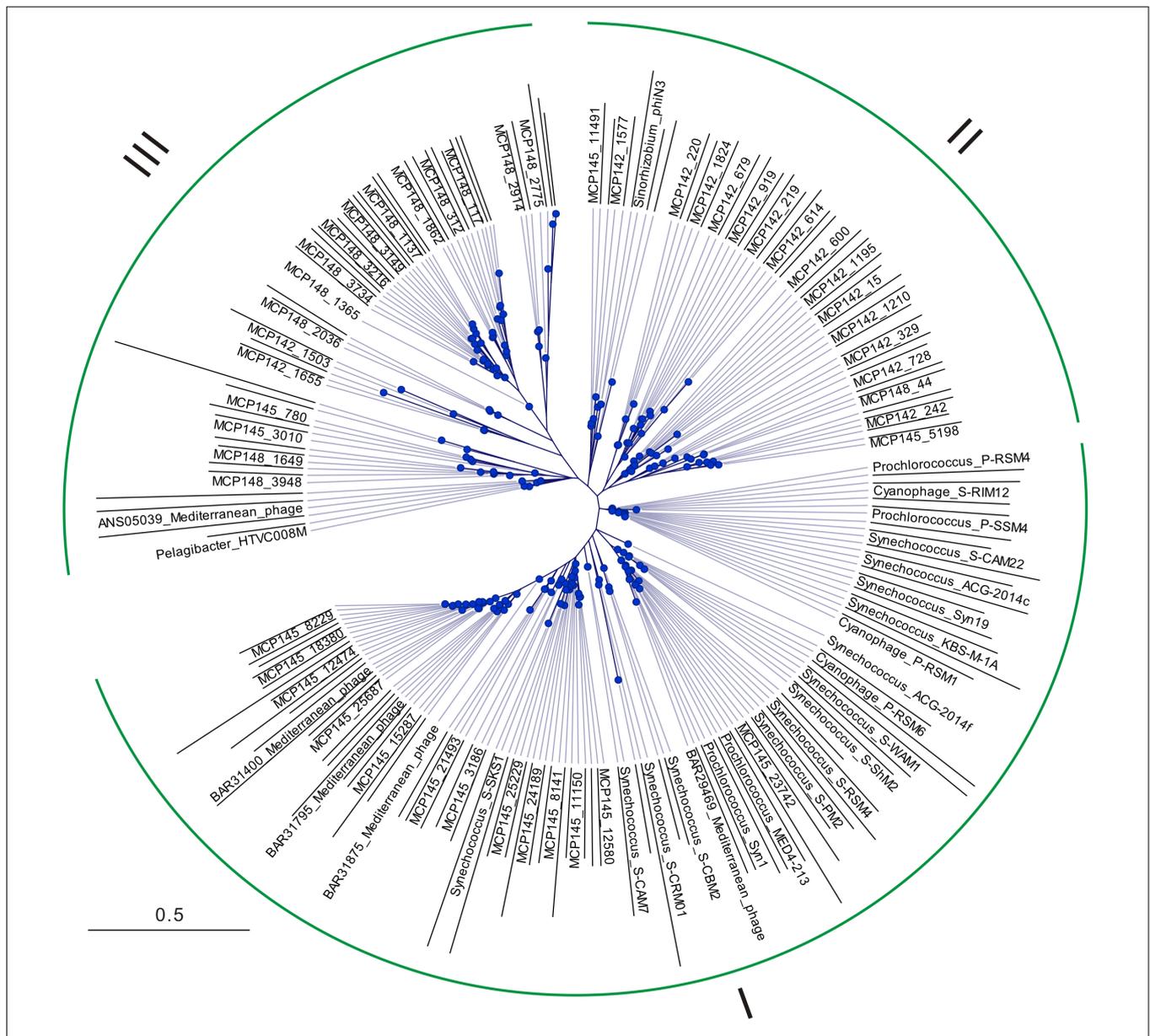


FIGURE 3 | Phylogenetic tree generated with the Maximum-Likelihood method from amino acid sequences derived from partial MCP gene sequences. The tree included representative sequences of top-50 abundant OTUs of 142-, 145-, and 148-bp lengths, as well as their relative sequences retrieved from GenBank. Groups I, II, and III approximately divided by gene fragment lengths 145-, 148-, and 142-bp, respectively. The scale bar refers to evolutionary distance inferred by the Jones–Taylor–Thorn algorithm.

the Jones–Taylor–Thornton evolutionary substitute model (Jones et al., 1992).

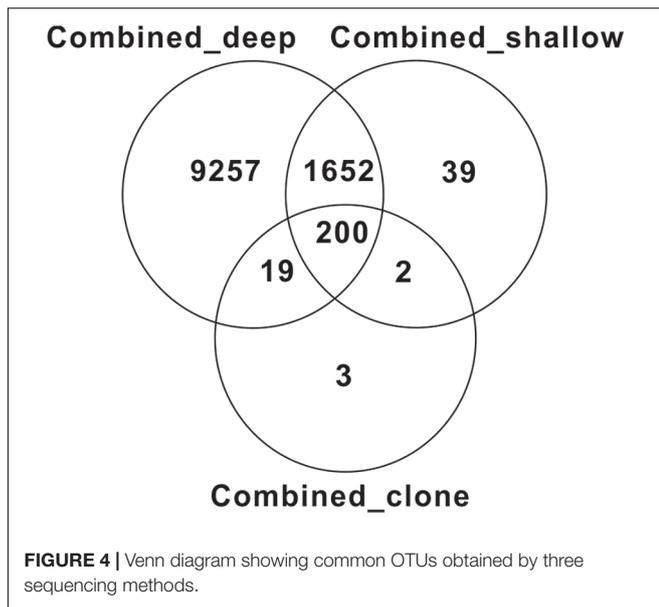
Rarefaction curves, used to estimate the myocyanophage richness, were generated with QIIME. The Good’s coverage was used to evaluate sequencing completeness by calculating the probability that a given sequence was chosen from a library of sequencing reads (Good, 1953). Paired sample *T*-test was conducted by using IBM SPSS Statistics 19 to reveal any difference in sequencing Good’s coverage between different sequencing methods. Principal coordinate analyses (PCoAs) were performed to assess the degree of variation of myocyanophage

community based on both weighted and unweighted UniFrac distances (Lozupone and Knight, 2005) by using the QIIME software package.

RESULTS

Sequencing Profiles

With the newly designed primer, MCP gene was successfully amplified with water (i.e., suspended particles in water obtained by going through 0.2 μm filters) and sediment samples. A total

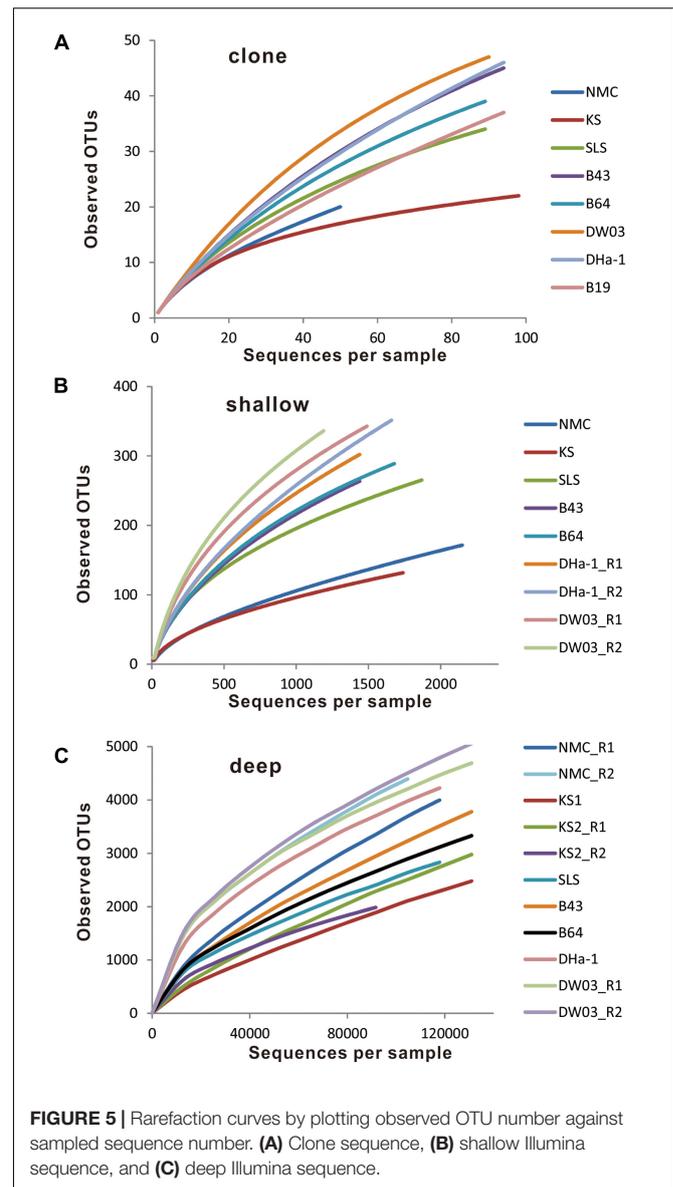


of 746 clone sequences were obtained for eight samples, with an average of 93 ± 4 sequences per sample (Supplementary Table S1). The overall quality scores of deep Illumina sequencing were higher than 25, with an average quality score of 30.8 (Supplementary Figure S2). By filtering out those reads with low quality scores (<25), average number of sequences per sample was $1,471 \pm 403$ and $144,254 \pm 31,878$ obtained by shallow, and deep Illumina sequencing, respectively.

Length of MCP Gene Fragment Amplicons

The deep Illumina sequencing reads were variable in length, ranging from < 136 -bp to > 154 -bp (Supplementary Figure S3). The MCP gene encodes a protein, thus the reads with lengths that did not have a complete codon (i.e., $145 \pm 3n$, where n is any integer) or contained stop codons were discarded. The discarded sequences accounted for an average of 4.6% of total reads in each sample. Overall, the sequences with lengths of 142-, 145-, and 148-bp dominated all samples (97.2–100% after removing non-protein coding sequences; Supplementary Figure S3). However, different samples exhibited different distribution patterns of these three major fragments (**Figure 2**). In two lacustrine (NMC and SSL) and four oceanic (B43, B64, DHa-1, and DW03) samples, 145-bp fragment was predominant (94.8–98.9% in relative percentages, with an average of $97.0\% \pm 1.8\%$). In the KS samples, the 148-bp length was most abundant (41.2–56.7% in relative percentages), followed by 145-bp (around 30.1%) and 142-bp (around 17.6%). Thus, the subsequent diversity and community analyses were carried out for these three lengths.

The length distribution of amplified MCP gene fragments based on clone sequencing (**Figure 2A**) was slightly different from those obtained from deep and shallow Illumina sequencing (**Figures 2B,C**). The relative abundance of the 148-bp fragment for B43 and B64 obtained with clone sequencing was higher than



that from deep Illumina sequencing. Likewise, for sample NMC, the relative abundances of the 142- and 148-bp fragments from clone sequencing were higher than those from deep Illumina sequencing (**Figure 2**). In sample KS2, the 148-bp fragment was more abundant than the other two. The length distribution of amplified MCP gene fragments based on shallow Illumina sequencing method (**Figure 2B**) was slightly different from both deep Illumina sequencing and clone library methods in that the 148-bp fragment was predominant in the two KS samples. In the other samples, the 145-bp fragment was by far predominant.

Nucleotide and predicted amino acid sequences were identified by comparing them to reference sequences retrieved from GenBank using BlastN and BlastP, and further confirmed by functional protein annotation through NCBI's Conserved Domain Database (Marchler-Bauer et al., 2017). The results showed that cyanophage MCP gene fragments

were predominant (>92%) in the top-200 hit list with a length of 145-bp. Some 142-bp, 148-bp fragments were related to non-cyanophages, including *Sinorhizobium* phages phiN3 and phiM12, *Pseudomonas* phage pf16, *Acidovorax* phage ACP17, and *Ralstonia* phage RSP15, a freshwater phage from Baikal Lake, which is related to *Polynucleobacter* (Cabello-Yeves et al., 2017).

Maximum likelihood phylogenetic tree based on amino acid sequences (Figure 3) roughly supported the length-depended lineages with the sequences divided into three major groups. The sequences with 145-bp nucleotide length, mainly distributed in several deep clades (Group I in Figure 3), shared close relationships with cyanophage MCP gene fragments and uncultured Mediterranean phage sequences. The majority of 142-bp sequences were divided into two clusters, within Group II (Figure 3). In this group, some 142-bp sequences and 145-bp sequences were closely related to *Sinorhizobium* phages. The majority of 148-bp sequences formed Group III (Figure 3). In this group, some sequences were closely related to *Pelagibacter* phages and uncultured Mediterranean phages. Other 148-bp sequences (on top of Figure 3) were closely related to *Pseudomonas* phage and *Acidovorax* phage.

Alpha Diversity of MCP Gene Fragment

Operational taxonomic units (OTUs) from all samples were obtained with UCULST method at 80, 90, 95, and 97% similarity levels with deep Illumina sequencing data (Supplementary Figure S4). Around 1,000, 4,000, 20,000, and 45,000 OTUs were obtained at four similarity levels, respectively. For the tractability of OTU numbers, the OTU table at similarity 90% was chosen hereafter. At 90% similarity, the highest numbers of OTUs after removing singletons were observed in Numco Lake water sample and the South China Sea water sample DW03 in the deep Illumina sequences (Supplementary Figure S5).

Venn diagrams showed that most OTUs from the deep sequencing effort were also captured by the shallow sequencing (Figure 4). For example, most OTUs from a combined clone library for all samples were detected by both shallow and deep Illumina sequencing methods, and most OTUs from shallow Illumina sequencing were also detected by deep Illumina sequencing (Figure 4). Similar overlapping pattern also occurred at the individual sample level (Supplementary Figure S6). However, rarefaction curves did not reach plateaus for all samples with all three sequencing methods (Figure 5). As expected, the coverage value for full-length MCP gene fragments was dependent on sequencing depth, i.e., deeper sequencing resulted in a higher coverage (Table 2).

Beta Diversity

Principal coordinate analysis based on deep sequencing results revealed that, within the marine samples, the communities displayed significant variations. As the distance between the samples increased, myocyanophage community became increasingly dissimilar (Supplementary Figure S7). However, such effect was not observed for the lake samples (data not shown).

TABLE 2 | Good's coverage values for the samples with different sequencing depths and lengths.

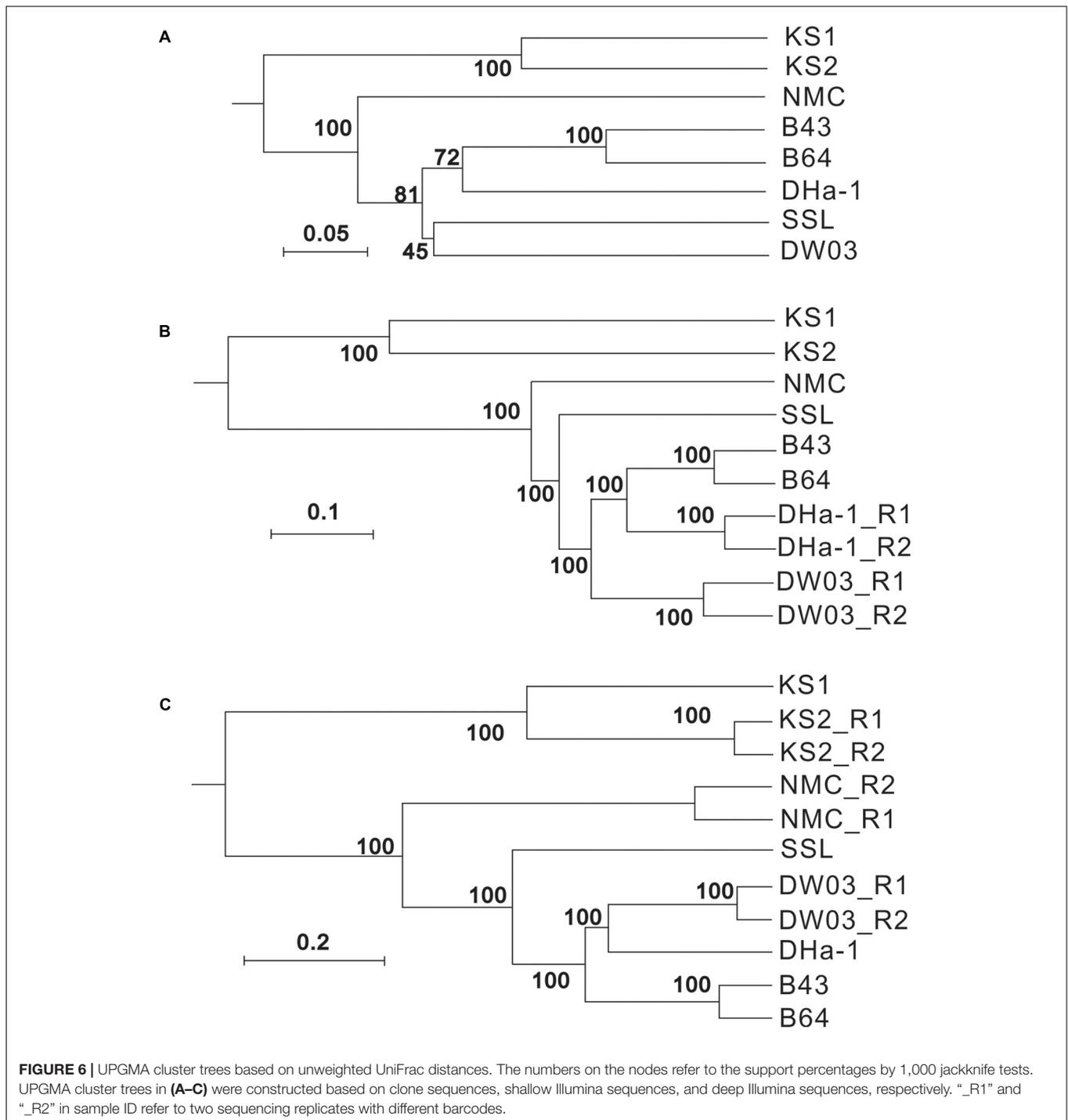
| Sample ID | Clone sequencing with 50 sub-sampling depth | Shallow sequencing with 1,600 sub-sampling depth | Deep sequencing with 38,000 sub-sampling depth | Deep sequencing with 130,000 sub-sampling depth |
|-----------|---|--|--|---|
| NMC | 0.740 | 0.944 | 0.968 | 0.976 |
| KS1 | 0.910 | 0.980 | 0.981 | 0.986 |
| KS2 | 0.918 | 0.953 | 0.979 | 0.985 |
| SSL | 0.811 | 0.926 | 0.978 | 0.985 |
| B43 | 0.716 | 0.900 | 0.972 | 0.980 |
| B64 | 0.737 | 0.916 | 0.970 | 0.984 |
| DHa-1 | 0.684 | 0.875 | 0.968 | 0.981 |
| DW03 | 0.744 | 0.865 | 0.966 | 0.982 |

Different sequencing methods yielded similar myocyanophage community groupings across different samples (Figure 6). The KS samples were the most different from others and they formed a separate cluster, likely because of higher abundances of 142- and 148-bp sequences relative to other samples. The composition of dominant OTUs (relative abundances >2%), based on deep Illumina sequencing, was different across different samples (Figure 7). Specifically, oceanic (B43, B64, and DHa-1, and DW03) and lacustrine samples did not share any dominant OTUs (0/65). One common OTU (1/30) was observed across all marine samples, 11 OTUs (11/30) were common in two Yellow Sea samples (B43 and B64), and 2 OTUs (2/17) were common to DHa-1 and DW03. Dominant OTUs were not shared between any lakes. Within the same lake (e.g., KS), 6 dominant OTUs (6/14) were shared between the two KS sediment samples.

DISCUSSION

Sensitivity and Specificity the Newly Designed Primer

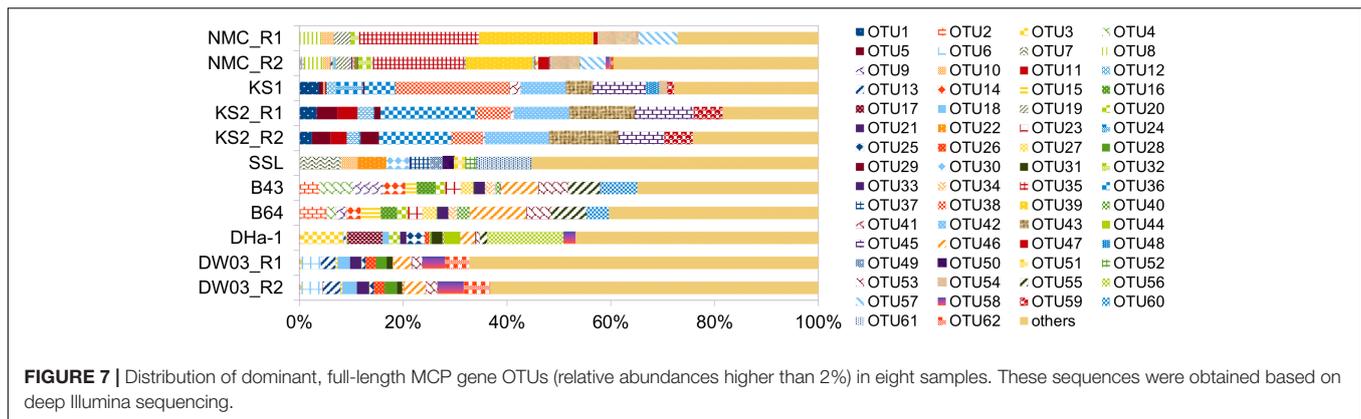
In this study, a MCP gene fragment was successfully amplified from all studied aquatic samples (DNA extracted from suspended particles retained on filters and sediment samples) by using a newly designed primer. A large range of the amplicon length (from 136-bp to 154-bp) suggested some non-specific amplification. However, other than the Kusai Lake samples, the dominance of 145-bp amplicon length (higher than 92.0%, represented myocyanophage) in most samples suggested a high specificity of the newly designed primer set. In the Kusai Lake samples, three amplicon lengths, i.e., 142-, 145-, and 148-bp, all occurred and the 142- and 148-bp amplicons represented non-cyanophages. The difference in the viral communities between Kusai Lake and other aquatic environments may be caused by the presence of both myocyanophages and non-myocyanophages in the Kusai Lake, whereas other lakes or oceans were dominated by myocyanophages. According to the extent of degeneration, the primer set designed in this study may capture more myocyanophages than MZIA1bis-MZIA6



primer set (Filee et al., 2005), but less myocyanophages than T4superF1-T3superR1 primer set (Chow and Fuhrman, 2012).

Comparatively, a high percentage of non-myocyanophage was detected in lacustrine samples, especially for the Kusai Lake samples. The hosts of some non-cyanophages appear to be soil bacteria. For example, species of *Sinorhizobium* are soil bacteria capable of nodulating leguminous

plants (Johnson et al., 2015); species of *Ralstonia* are soil phyto-pathogens (Fujiwara et al., 2008). Thus, these viruses with amplicon lengths of 142- and 148-bp may represent terrestrial input to lakes, and it is reasonable to observe that the lake samples contained higher percentage of 142- and 148-bp MCP fragments than the oceanic samples, because the lakes in this study have higher terrestrial input than the oceans.



Sample Types in This Study

In this study, sediment particles and microbial cells retained on 0.2 μm membrane filters were used in viral community analyses, which is different from previous studies using ultracentrifugation, tangential flow filtration, or ultrafiltration (Chow and Fuhrman, 2012; Chenard et al., 2015; Cai et al., 2016). Therefore, the viral communities determined in this study should represent bacteria and sediment-attached particle rather than the free-living viral particles.

Another sample type used in this study was sediment. Myocyanophage MCP genes detected in these sediment samples should have been derived from water column but settled into sediments. Indeed, our previous study detected cyanobacteria genes in Kusai Lake sediment (Hou et al., 2014). A similar observation was also made in the subtropical Pearl River (He et al., 2017), where cyanophage sequences were detected in estuarine sediments.

Myocyanophage Diversity Estimate in Various Environments

T4-like phages, including cyanophages, belong to a superfamily (Filee et al., 2005; Comeau and Krisch, 2008), and are widespread in various environments (Comeau and Krisch, 2008). The MCP gene is one of the quickly evolving genes involved in the interaction between viruses and their hosts (Marston et al., 2012). As a result, it is not surprising that extremely diverse viruses were detected from various water and sediment samples. Myocyanophage communities have been extensively studied based on diversity surveys of MCP gene using various primers such as MZIA1bis and MZIA6 (Filee et al., 2005) or T4superF1 and T3superR1 (Chow and Fuhrman, 2012), which amplify the similar region of MCP gene fragments, from position ~ 300 to ~ 760 on *Synechococcus* cyanophage S-PM2. These primer pairs amplify the MCP gene fragments of ~ 400 – ~ 500 bp in length. The short primer pair designed in this study, amplifying a different region, was initially intended to study myocyanophage community variation based on ancient DNA preserved in lake sediments in response to environmental change. A comparison between our results and those from a previous study (Chow and Fuhrman, 2012) revealed that when the sequencing depth is similar and percent identity is at 90%, shorter sequences

(e.g., 145-bp obtained in this study) resulted in more OTUs than longer sequences (~ 400 to ~ 500 bp as obtained by Chow and Fuhrman, 2012), which may be caused by location of the 145-bp amplicon in a more variable region of the MCP gene.

Variations in Myocyanophage Community

Overall, the significant variations of myocyanophage community from different samples observed in this study may be collectively caused by different distributions of host cyanobacteria in different environmental settings. Phylogenetic analysis suggested that cyanobacteria from Tibetan lakes formed different clusters, separated from other habitats, including other freshwater lakes and marine samples (Wu et al., 2010), which may be the fundamental reason for different myocyanophage community in the different environments (Figures 6, 7). Wang G. et al. (2011) also attributed their observed variation of viral community to different distribution of host cyanobacteria. Two previous studies on estuary viral communities identified marine and freshwater viral biomes in the mixing zone (Cai et al., 2016; He et al., 2017) which is also the mixing zone of the marine and freshwater viral hosts. These results collectively suggested that variation of hosts affect the T4-like viral community composition. Although the Kusai Lake and Namuco Lake are both located on the Tibetan Plateau, but the myocyanophage communities were very different, which may be caused by geographic separation of both cyanophages and their hosts. For the marine samples, the circulating ocean current may have led to transport of both microbes and viruses across different oceans. As a result, more common myocyanophage groups or OTUs were found within marine samples, and the communities were more similar if the geographic distances were close (Supplementary Figure S7). Substantial difference in the myocyanophage communities between the surface sediment and deep sediment in the Kusai Lake samples (Figure 6) may be caused by temporal variations in hydrological and climate conditions, because these two samples should have been deposited at different times. The viral community in the surface sediment of Kusai Lake represents a modern community, and the community from the 3.85 m depth represents an ancient community from 2,250 years cal. before present (BP) (Hou et al., 2014). According to the result from the

deep sequencing effort, the read ratio of 142-/145-bp, an estimate of terrestrial input relative to lacustrine production, was higher for the deeper sample (KS2) than that for the surface sediment sample (KS1), suggesting a higher terrestrial input 2,250 years cal. BP. This result is consistent with our earlier ancient DNA based study (Hou et al., 2014) and supports a strong summer monsoon scenario during that time (Liu et al., 2009). In summary, the newly designed primer was effective in studying the diversity of myocyanophages in various aquatic samples, including sediment samples representing cyanophage diversity more than 2,000 years ago. Due to high diversity, deep sequencing presents a good method to understand myocyanophage community. The amplicons were mainly composed of 142-, 145-, and 148-bp fragments. These different lengths of the gene fragments may represent source variations, terrestrial source or aquatic sources. The myocyanophage communities were also more heterogeneous in lakes than in oceans. Due to connection via oceanic currents, the marine viral communities displayed a distance effect, i.e., with increased geographic distances, the oceanic viral communities became increasingly dissimilar.

AUTHOR CONTRIBUTIONS

WH and HD conceived and designed the experiments. WH and SW performed the experiments. WX contributed

the sampling vessel and tools. WH, BB, and GL analyzed the data. WH, SW, BB, GL, WX, and HD wrote the paper.

FUNDING

This research was supported by a grant from the National Natural Science Foundation of China (No. 41302022) and China Scholarship Council (No. 201406405005).

ACKNOWLEDGMENTS

We thank James Ford from the Center for Genomics and Bioinformatics, Indiana University, for his help with Illumina sequencing. Oceanic samples were obtained by the voyage MD190-CIRCEA. We are grateful to the two reviewers whose comments improved the quality of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00887/full#supplementary-material>

REFERENCES

- Bailey, S., Clokie, M. R., Millard, A., and Mann, N. H. (2004). Cyanophage infection and photoinhibition in marine cyanobacteria. *Res. Microbiol.* 155, 720–725. doi: 10.1016/j.resmic.2004.06.002
- Boere, A. C., Rijpstra, W. I., De Lange, G. J., Sinninghe Damste, J. S., and Coolen, M. J. (2011). Preservation potential of ancient plankton DNA in Pleistocene marine sediments. *Geobiology* 9, 377–393. doi: 10.1111/j.1472-4669.2011.00290.x
- Cabello-Yeves, P. J., Zemskaya, T. I., Rosselli, R., Coutinho, F. H., Zakharenko, A. S., Blinov, V. V., et al. (2017). Genomes of novel microbial lineages assembled from the sub-ice waters of Lake Baikal. *Appl. Environ. Microbiol.* 84:e02132-17. doi: 10.1128/AEM.02132-17
- Cai, L., Zhang, R., He, Y., Feng, X., and Jiao, N. (2016). Metagenomic analysis of viroplankton of the subtropical Jiulong River estuary, China. *Viruses* 8, 35–47. doi: 10.3390/v8020035
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624. doi: 10.1038/ismej.2012.8
- Chenard, C., Chan, A. M., Vincent, W. F., and Suttle, C. A. (2015). Polar freshwater cyanophage S-EIV1 represents a new widespread evolutionary lineage of phages. *ISME J.* 9, 2046–2058. doi: 10.1038/ismej.2015.24
- Chow, C. E., and Fuhrman, J. A. (2012). Seasonality and monthly dynamics of marine myovirus communities. *Environ. Microbiol.* 14, 2171–2183. doi: 10.1111/j.1462-2920.2012.02744.x
- Clokie, M. R., and Mann, N. H. (2006). Marine cyanophages and light. *Environ. Microbiol.* 8, 2074–2082. doi: 10.1111/j.1462-2920.2006.01171.x
- Comeau, A. M., and Krisch, H. M. (2008). The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. *Mol. Biol. Evol.* 25, 1321–1332. doi: 10.1093/molbev/msn080
- Dekel-Bird, N. P., Avrani, S., Sabehi, G., Pekarsky, I., Marston, M. F., Kirzner, S., et al. (2013). Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. *Environ. Microbiol.* 15, 1476–1491. doi: 10.1111/1462-2920.12103
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Filee, J., Tetart, F., Suttle, C. A., and Krisch, H. M. (2005). Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12471–12476. doi: 10.1073/pnas.0503404102
- Fujiwara, A., Kawasaki, T., Usami, S., Fujie, M., and Yamada, T. (2008). Genomic characterization of *Ralstonia solanacearum* phage ϕ RSA1 and its related prophage (ϕ RXS) in strain GMI1000. *J. Bacteriol.* 190, 143–156. doi: 10.1128/JB.01158-07
- Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P. J., Griggo, C., Gielly, L., et al. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nat. Commun.* 5:3211. doi: 10.1038/ncomms4211
- Good, I. J. (1953). On population frequencies of species and the estimation of population parameters. *Biometrika* 40, 237–264. doi: 10.1093/biomet/40.3-4.237
- Hambly, E., Tetart, F., Desplats, C., Wilson, W. H., Krisch, H. M., and Mann, N. H. (2001). A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11411–11416. doi: 10.1073/pnas.191174498
- He, M., Cai, L., Zhang, C., Jiao, N., and Zhang, R. (2017). Phylogenetic diversity of T4-type phages in sediments from the subtropical pearl river estuary. *Front. Microbiol.* 8:897. doi: 10.3389/fmicb.2017.00897
- Hellweger, F. L. (2009). Carrying photosynthesis genes increases ecological fitness of cyanophage in silico. *Environ. Microbiol.* 11, 1386–1394. doi: 10.1111/j.1462-2920.2009.01866.x
- Hou, W., Dong, H., Li, G., Yang, J., Coolen, M. J., Liu, X., et al. (2014). Identification of photosynthetic plankton communities using sedimentary ancient DNA and

- their response to late-Holocene climate change on the Tibetan Plateau. *Sci. Rep.* 4:6648. doi: 10.1038/srep06648
- Hou, W. G., Wang, S., Dong, H. L., Jiang, H. C., Briggs, B. R., Peacock, J. P., et al. (2013). A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. *PLoS One* 8:e53350. doi: 10.1371/journal.pone.0053350
- Ignacio-Espinoza, J. C., and Sullivan, M. B. (2012). Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ. Microbiol.* 14, 2113–2126. doi: 10.1111/j.1462-2920.2012.02704.x
- Johnson, M. C., Tatum, K. B., Lynn, J. S., Brewer, T. E., Lu, S., Washburn, B. K., et al. (2015). *Sinorhizobium meliloti* phage Φ m9 defines a new group of T4 superfamily phages with unusual genomic features but a common T = 16 capsid. *J. Virol.* 89, 10945–10958. doi: 10.1128/JVI.01353-15
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Labonté, J. M., Swan, B. K., Poulos, B., Luo, H. W., Koren, S., Hallam, S. J., et al. (2015). Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* 9, 2386–2399. doi: 10.1038/ismej.2015.48
- Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11013–11018. doi: 10.1073/pnas.0401526101
- Liu, X., Dong, H., Yang, X., Herzschuh, U., Zhang, E., Stuetz, J. B., et al. (2009). Late Holocene forcing of the Asian winter and summer monsoon as evidenced by proxy records from the northern Qinghai-Tibetan Plateau. *Earth Planet. Sci. Lett.* 280, 276–284. doi: 10.1016/j.epsl.2009.01.041
- Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005
- Ma, Y. F., Allen, L. Z., and Palenik, B. (2014). Diversity and genome dynamics of marine cyanophages using metagenomic analyses. *Environ. Microbiol. Rep.* 6, 583–594. doi: 10.1111/1758-2229.12160
- Mann, N. H. (2003). Phages of the marine cyanobacterial picophytoplankton. *FEMS Microbiol. Rev.* 27, 17–34. doi: 10.1016/S0168-6445(03)00016-0
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucl. Acids Res.* 45, D200–D203. doi: 10.1093/nar/gkw1129
- Marston, M. F., Pierciey, F. J. Jr., Shepard, A., Gearin, G., Qi, J., Yandava, C., et al. (2012). Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4544–4549. doi: 10.1073/pnas.1120310109
- Marston, M. F., Taylor, S., Sme, N., Parsons, R. J., Noyes, T. J., and Martiny, J. B. (2013). Marine cyanophages exhibit local and regional biogeography. *Environ. Microbiol.* 15, 1452–1463. doi: 10.1111/1462-2920.12062
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E., and Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genet.* 9:e1003987. doi: 10.1371/journal.pgen.1003987
- Needham, D. M., Chow, C. E., Cram, J. A., Sachdeva, R., Parada, A., and Fuhrman, J. A. (2013). Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. *ISME J.* 7, 1274–1285. doi: 10.1038/ismej.2013.19
- Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Sarkissian, C. D., Haile, J., Hellstrom, M., et al. (2015). Ancient and modern environmental DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130383. doi: 10.1098/rstb.2013.0383
- Puxty, R. J., Millard, A. D., Evans, D. J., and Scanlan, D. J. (2016). Viruses Inhibit CO₂ fixation in the most abundant phototrophs on Earth. *Curr. Biol.* 26, 1585–1589. doi: 10.1016/j.cub.2016.04.036
- Rohwer, F., and Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature* 459, 207–212. doi: 10.1038/nature08060
- Sønstebø, J. H., Gjelly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., et al. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol. Ecol. Resour.* 10, 1009–1018. doi: 10.1111/j.1755-0998.2010.02855.x
- Sullivan, M. B., Huang, K. H., Ignacio-Espinoza, J. C., Berlin, A. M., Kelly, L., Weigle, P. R., et al. (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* 12, 3035–3056. doi: 10.1111/j.1462-2920.2010.02280.x
- Suttle, C. A. (2007). Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750
- Wang, G., Asakawa, S., and Kimura, M. (2011). Spatial and temporal changes of cyanophage communities in paddy field soils as revealed by the capsid assembly protein gene g20. *FEMS Microbiol. Ecol.* 76, 352–359. doi: 10.1111/j.1574-6941.2011.01052.x
- Wang, K., Wommack, K. E., and Chen, F. (2011). Abundance and distribution of *Synechococcus* spp. and Cyanophages in the Chesapeake Bay. *Appl. Environ. Microbiol.* 77, 7459–7468. doi: 10.1128/AEM.00267-11
- Wang, X. Z., Jing, R. Y., Liu, J. J., Yu, Z. H., Jin, J., Liu, X. B., et al. (2016). Narrow distribution of cyanophage psbA genes observed in two paddy waters of Northeast China by an incubation experiment. *Virol. Sin.* 31, 188–191. doi: 10.1007/s12250-015-3673-5
- Weitz, J. S., and Wilhelm, S. W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* 4:17. doi: 10.3410/B4-17
- Wu, Q. L., Xing, P., and Liu, W. T. (2010). East Tibetan lakes harbour novel clusters of picocyanobacteria as inferred from the 16S-23S rRNA internal transcribed spacer sequences. *Microb. Ecol.* 59, 614–622. doi: 10.1007/s00248-009-9603-z

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RZ and the handling Editor declared their shared affiliation.

Copyright © 2018 Hou, Wang, Briggs, Li, Xie and Dong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.