



# Metagenomic Analysis of the Diversity of DNA Viruses in the Surface and Deep Sea of the South China Sea

Yantao Liang<sup>1,2,3†</sup>, Long Wang<sup>1†</sup>, Zengmeng Wang<sup>1</sup>, Jiulong Zhao<sup>1</sup>, Qingwei Yang<sup>2</sup>, Min Wang<sup>2</sup>, Kaiguang Yang<sup>4</sup>, Lihua Zhang<sup>4</sup>, Nianzhi Jiao<sup>3</sup> and Yongyu Zhang<sup>1\*</sup>

<sup>1</sup> Key Laboratory of Biofuels, Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China, <sup>2</sup> College of Marine Life Sciences, Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao, China, <sup>3</sup> State Key Laboratory of Marine Environmental Science, Institute of Marine Microbes and Ecospheres, Xiamen University, Xiamen, China, <sup>4</sup> Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China

## OPEN ACCESS

### Edited by:

Jie Xu,  
South China Sea Institute  
of Oceanology (CAS), China

### Reviewed by:

Huiluo Cao,  
The University of Hong Kong,  
Hong Kong  
Alejandra Prieto-Davó,  
Universidad Nacional Autónoma  
de México, Mexico

### \*Correspondence:

Yongyu Zhang  
zhangyy@qibebt.ac.cn

† These authors have contributed  
equally to this work

### Specialty section:

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

Received: 25 April 2019

Accepted: 08 August 2019

Published: 23 August 2019

### Citation:

Liang Y, Wang L, Wang Z, Zhao J,  
Yang Q, Wang M, Yang K, Zhang L,  
Jiao N and Zhang Y (2019)  
Metagenomic Analysis of the Diversity  
of DNA Viruses in the Surface  
and Deep Sea of the South China  
Sea. *Front. Microbiol.* 10:1951.  
doi: 10.3389/fmicb.2019.01951

A metagenomic analysis of the viral community from five surface and five deep sea water (>2000 m below the surface, mbs) samples collected from the central basin of the South China Sea and adjacent Northwest Pacific Ocean during July–August 2017 was conducted herein. We build up a South China Sea DNA virome (SCSV) dataset of 29,967 viral Operational Taxonomic Units (vOTUs), which is comparable to the viral populations from the original Tara Ocean and Malaspina expeditions. The most abundant and widespread viral populations were from the uncultivated viruses annotated from the viral metagenomics. Only 74 and 37 vOTUs have similarity with the reported genomes from the cultivated viruses and the single-virus genomics, respectively. The community structures of deep sea viromes in the SCSV were generally different from the surface viromes. The carbon flux and nutrients (PO<sub>4</sub> and NO<sub>x</sub>) were related to the surface and deep sea viromes in the SCSV, respectively. In the SCSV, the annotated vOTUs could be affiliated to the cultivated viruses mainly including *Pelagibacter* (SAR11) phage HTVC010P, *Prochlorococcus* phages (P-GSP1, P-SSM4, and P-TIM68), Cyanophages (MED4-184 and MED4-117) and *Mycobacterium* phages (Sparky and Squirry). It indicated that phage infection to the SAR11 cluster may occur ubiquitously and has significant impacts on bathypelagic SAR11 communities in the deep sea. Meanwhile, as *Prochlorococcus* is prominently distributed in the euphotic ocean, the existence of their potential phages in the deep sea suggested the sedimentation mechanism might contribute to the formation of the deep sea viromes. Intriguingly, the presence of *Mycobacterium* phages only in the deep sea viromes, suggests inhabitation of endemic viral populations in the deep sea viromes in the SCSV. This study provided an insight of the viral community in the South China Sea and for the first time uncovered the deep sea viral diversity in the central basin of the South China Sea.

**Keywords:** marine viruses, South China Sea, metagenomics, community structure, deep sea

## INTRODUCTION

Viruses are the most numerous, ubiquitous and diverse organisms in the aquatic environment. They play an essential role in regulating microbial fatality, community, and functions, and impact the microbial genetic diversity through the horizontal gene transfer (Wommack and Colwell, 2000; Weinbauer, 2004; Suttle, 2005, 2007). By controlling microbial community and inducing the release of organic matters from host cells, viruses significantly impact the marine biogeochemical cycles (Suttle, 2005) and contribute to the sequestration of CO<sub>2</sub> through “microbial carbon pump” and “biological pump” (Jiao et al., 2010; Guidi et al., 2016).

Though viruses are of great significance in the marine ecosystems, their diversity and variability are still not clearly understood as compared to that of the bacterial communities. This is mainly because of lacking universal gene markers for investigating viral communities and the relatively recent development and application of culture-independent high-throughput sequencing methods (Breitbart et al., 2002; Brum et al., 2015; Aylward et al., 2017). With the rapid progress of the high-throughput sequencing technology and bioinformatic analysis, metagenomics has become an essential and powerful method to better understand viral community structure, diversity, and variability than the original culture-dependent and gene markers dependent methods. Mainly it can be easily used to compile and annotate the complex, pristine ecological viral genomes (Angly et al., 2006; Culley et al., 2006; Hurwitz and Sullivan, 2013; Paez-Espino et al., 2018; Roux et al., 2018). During the last two decades, the viral communities have been resolved from more and more different marine environments through the metagenomics, particularly through several great expedition and datasets like the Global Ocean Sampling Expedition, Pacific Ocean Virome, Tara Ocean Expedition, Malaspina expedition, Tara Oceans Polar Circle (TOPC) expedition and so on (Angly et al., 2006; Steward and Preston, 2011; Williamson et al., 2012; Hurwitz and Sullivan, 2013; Brum et al., 2015; Roux et al., 2016; Gregory et al., 2019). Recently, the Global Ocean Viromes 2.0 (GOV 2.0) dataset from 145 marine viral metagenomic samples identified a total of 195,728 viral populations, which is about 12-fold of viral populations from the original Tara Ocean and Malaspina expeditions (Brum et al., 2015; Roux et al., 2016; Gregory et al., 2019). However, comparing with the virome information in the euphotic zone, our knowledge about the deep seas is still limited, especially in the Northwest Pacific Ocean (Winter et al., 2014; Mizuno et al., 2016; Gregory et al., 2019).

The South China Sea (SCS) is an immense marginal sea in the Northwest Pacific Ocean and connects the Pacific Ocean and the Indian Ocean through the complex western boundary currents systems (Hu et al., 2015). The SCS located in the Western Pacific Warm Pool that potentially influence global marine ecosystems and climate change. Nonetheless, there are still few reports on viral diversity in the South China Sea. Till date, there have only been two studies about the viral diversity using the gene markers (DNA polymerase gene and g23 gene, respectively) in the SCS (Huang et al., 2010; He et al., 2017). The viral community structure of cyanophage in one surface sample

of the central SCS was distantly discriminated from the samples from other oceans (Huang et al., 2010), suggested that the viral community structure and diversity of SCS might contain unique features. Currently, there is still not metagenomic report of the viral community structure and diversity in the South China Sea, especially in the deep sea (>1,000 m below the surface, mbs), which is awaiting in-depth investigation to unveil their unknown ecological characteristics.

To illustrate the viral community and diversity in the deep sea of South China Sea and to compare the differences of virome information between surface and deep sea, we present a DNA viral metagenomic dataset, including 10 seawater samples from the surface and deep sea layers (from about 2,000 to 3,500 mbs) of five stations in the central South China Sea and adjacent Northwest Pacific Ocean. This study will present the first metagenomic insight on the viral diversity, community structure and the differences among surface and deep sea seawater samples in the central South China Sea.

## MATERIALS AND METHODS

### Sampling and Analysis of Environmental Factors

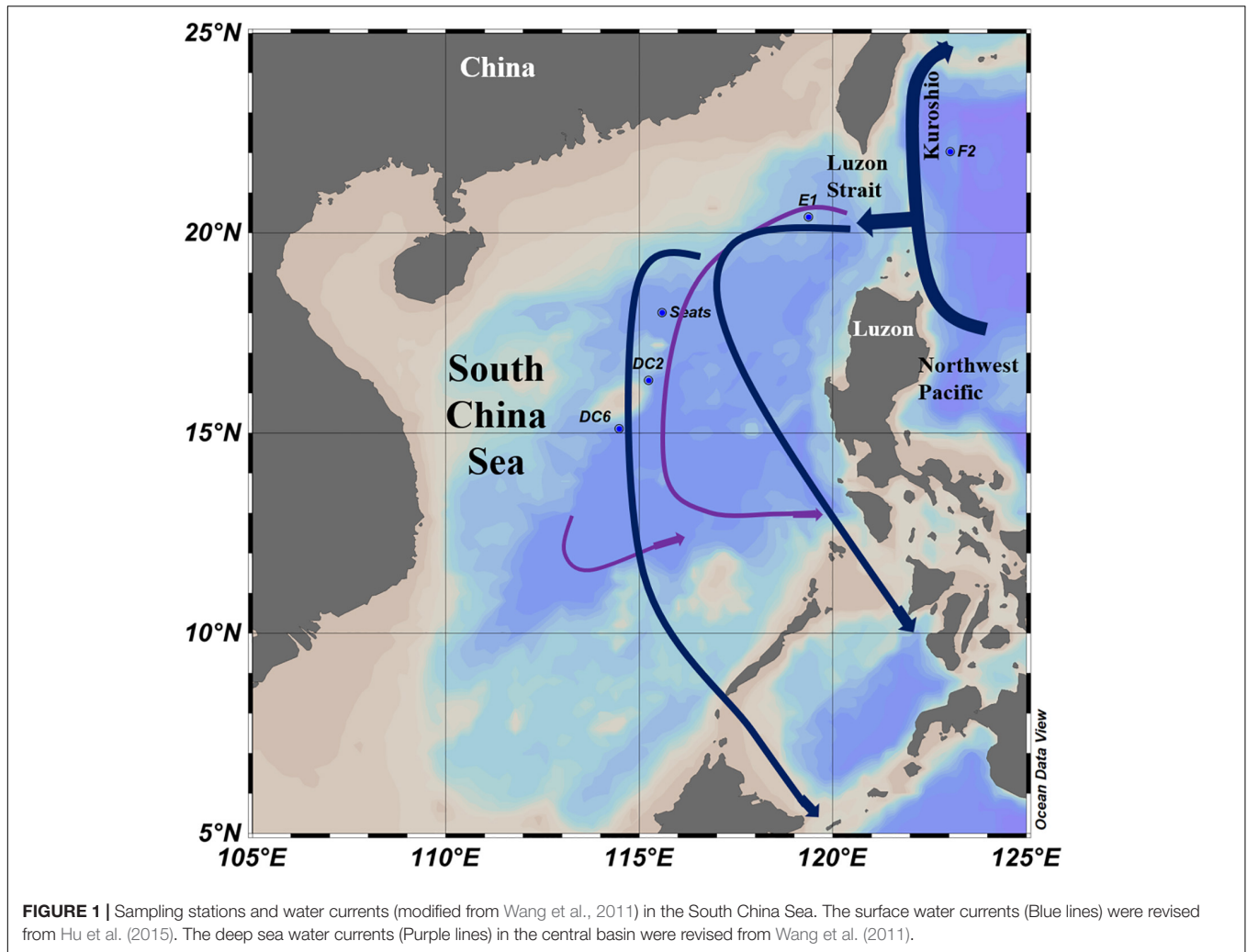
Ten viral metagenomic seawater samples (five surface and five deep sea deeper than 2,000 m) were collected from four stations (E1, SEATS, DC2, DC6) in the central South China Sea and one station (F2) in the Northwest Pacific Ocean adjacent to the South China Sea during July–August 2017 (**Figure 1**). Seawater samples (240 L) were collected using Niskin bottles fitted on a rosette frame which also equipped the SBE-9 plus CTD sensors (SBE 911; Sea-Bird Electronics) for temperature, salinity, and depth. Hundred milliliter sub-samples were collected from each site to determine concentrations of nutrient and dissolved organic carbon (DOC). Nutrient concentrations PO<sub>4</sub>, NO<sub>2</sub>, and NO<sub>x</sub> (NO<sub>2</sub> + NO<sub>3</sub>) were investigated using an onboard nutrient auto-analyzer (SKALAR SAN plus, Netherlands). The detection limits for PO<sub>4</sub>, NO<sub>2</sub>, and NO<sub>x</sub> were 0.03, 0.01, and 0.03 μM, respectively (Han et al., 2012). Unfiltered DOC samples were stored in precombusted EPA vials at –20°C until further analysis. The TOC concentration was measured using a Shimadzu TOC-V analyzer according to Wu et al. (2015).

The carbon fluxes of each sample were collected from the primary production for the surface samples and the vertical flux ( $F_z$ ; g C m<sup>-2</sup> yr<sup>-1</sup>) of particulate organic carbon (POC) for the deep sea samples. The  $F_z$  was estimated according to Antia et al. (2001) and Yokokawa et al. (2013):

$$F_z = 0.1 \times (PP)^{1.77} \times (Z)^{-0.68} \quad (1)$$

$F_z$  related to primary production (PP; g C m<sup>-2</sup> yr<sup>-1</sup>) for depth (Z; m) greater than 100 m. For each station, the PP during July 2017 was acquired from the custom products of the Ocean Productivity website<sup>1</sup> based on the Epply Vertically Generalized Production Model from MODIS/Aquar data (Behrenfeld and Falkowski, 1997).

<sup>1</sup><http://www.science.oregonstate.edu/ocean.productivity/>



## Preparation of Viral Concentrates, DNA Extraction, and Sequencing

The seawater samples (240 L) were immediately filtered through a 300 mm diameter cellulose membrane with a 3  $\mu\text{m}$  pore size, and then re-filtered the filtrate through a 0.22  $\mu\text{m}$  membrane, to remove the bigger organisms, such as zooplankton, phytoplankton, and bacteria. Free viruses in the final filtrate were concentrated to a volume of 500 mL by the large-scale Tangential Flow Filtration (membrane package with a total surface area of 0.5  $\text{m}^2$ : Pellicon<sup>®</sup> 2 Cassette, Biomax<sup>®</sup> 50 kDa; polyethersulfone) (Sun et al., 2014). After concentrating sample each time, the filtration membrane cassettes were cleaned by rinsing with ample virus-free deionized water, followed by cleaning with 0.1 N NaOH for at least 30 min. The concentrated viral samples were stored in the black polycarbonate bottle and then stored at 4°C until being further processed (Angly et al., 2006; Thurber et al., 2009; Cai et al., 2016).

Just before the analysis, the viral concentrates were re-filtered through a 0.22  $\mu\text{m}$  filter to take away any remaining cellular microorganisms or aggregation. The filtrates were treated with 2  $\text{ng} \cdot \text{L}^{-1}$  DNase I at room temperature for 1 h to remove the free

DNA. After digestion, NaCl (final concentration: 1 M) was added and incubated at 4°C for 1 h. The filtrates were concentrated using polyethylene glycol (PEG-8000) precipitation (10% w/v) and incubated at 4°C in the dark for 24 h. The mixed samples were centrifuged at 10,000  $\times g$  for 60 min at 4°C, and the pellets were resuspended in SM buffer. Afterward, the virome samples were purified by CsCl gradient ultra-centrifugation (gradient-density: 1.5  $\text{g} \cdot \text{mL}^{-1}$ , 200,000  $\times g$ , 8 h, 4°C; CP-100WX). The purified viral particles were collected and dialyzed three times in SM buffer using 30 kDa super-filters (UFC5030) for DNA extraction. The viral DNA was obtained using the phenol/chloroform/isoamylol method and stored at -80°C until sequencing (Thurber et al., 2009). High-throughput sequencing of the original viral DNA was carried out by Novogene (Beijing, China) using Illumina NovaSeq 6000 (pair-end sequencing, 2  $\times$  150 bp).

## Metagenomic, Genomic, and Function Analyses

Only the high-quality reads were picked from the raw reads giving 30–51 million (clean data rate >0.90) 150 bp paired-end

reads. The paired-end reads were then segregated by implying the following conditions: (1) having more than 10% N; (2) were of low quality (40% reads length,  $Q \leq 5$ ); (3) with the adapter. Quality-filtered reads were assembled using metaSPAdes (version 3.12.0) (Nurk et al., 2017). The contigs with lengths of less than 300 bp were filtered out. The assembled contig data were analyzed using VirSorter V2 for the viral predictions (Roux et al., 2015). Only the sure (category 1) and somewhat sure (category 2) phage contigs and prophages (category 4 and category 5) were considered as viral contigs and used for further analysis. A short sketch of the 10 viromes is shown in **Table 1**. The average abundance was calculated as the transcripts (gene) per million reads mapped (TPM) using Salmon (Patro et al., 2017).

The taxonomic annotation and potential hosts of the viral contigs from the VirSorter 2.0 were blasted to the Integrated Microbial Genome/Virus (IMG/VR) system v.2.0<sup>2</sup> dataset using the routine parameters (Paez-Espino et al., 2018). The open reading frames (ORFs) were predicted for each viral contig through Prodigal (Noguchi et al., 2008).

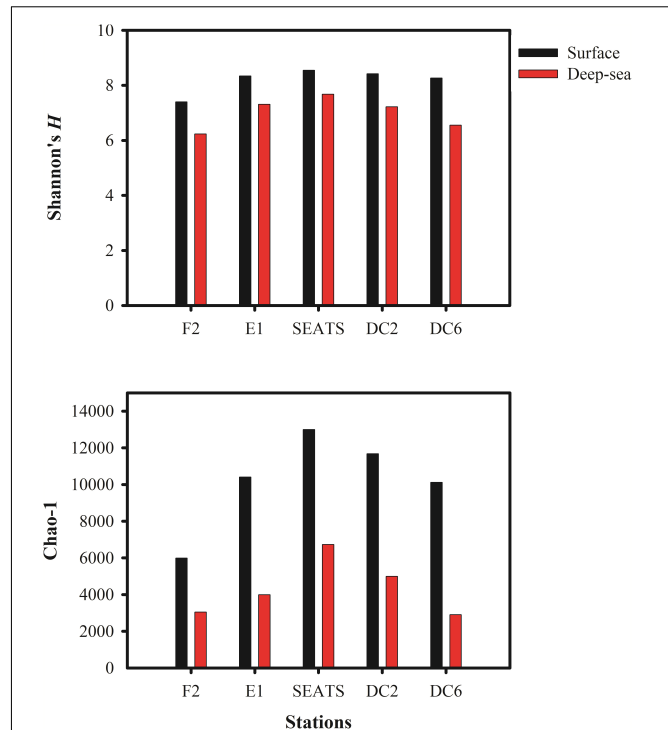
The functional contents of the 10 SCSV samples and the deep sea unique viral contigs were further characterized using Meta Genome Rapid Annotation using Subsystem Technology (MG-RAST) (Meyer et al., 2008) (with MG-RAST accession number 4839906.3, 4839904.3, 4839922.3, 4839908.3, 4839912.3, 4839916.3, 4839914.3, 4839910.3, 4839920.3, 4839918.3, and 4840005.3, respectively), an online metagenome annotation service<sup>3</sup>. The viral contigs processed by MG-RAST were compared to the SEED Subsystems database using a maximum *E*-value of  $10^{-5}$ , a minimum alignment length of 15, and a minimum identity of 60%.

## Phylogenetic Analysis

The phage terminase large-subunit domain (*TerL*) and family B DNA polymerase (DNA polB), which were present in phages of the order Caudovirales (*Terminase\_6*, PF03237) and eukaryotic and cyanobacteria viruses respectively (Chen and Suttle, 1995; Koonin et al., 2015; Roux et al., 2017), were utilized to establish the phylogenetic tree. The *TerL* and DNA polB sequences were dereplicated at the 97%

<sup>2</sup><https://img.jgi.doe.gov/vr/>

<sup>3</sup><http://www.mg-rast.org/>



**FIGURE 2** | The Alpha diversity indexes in South China Sea viromes (SCSV) data set. **(A)** Shannon's diversity  $H'$  calculated from the relative abundance of viral Operational Taxonomic Units (vOTUs) for each sample. **(B)** Chao 1 diversity calculated from relative abundances of vOTUs for each sample. The relative abundance is computed for each sample as the number of the transcripts per million reads mapped (TPM).

nucleotide level using cd-hit (Li and Godzik, 2006). The *TerL* and DNA polB sequences from the SCSV virome genes were screened by the DOE-JGI Metagenome Annotation Pipeline and compared to the viral RefSeq database using BLASTP (*E*-value  $< 10^{-5}$ ) to recruit relevant reference sequences. All sequences were aligned at the amino acid level using MUSCLE (Edgar, 2004) (using default parameters), manually inspected and trimmed as necessary. The maximum likelihood (ML) tree with 1000 bootstraps was constructed using the program FastTree (v2.1.10) (Price et al., 2010) using a JTT + CAT model and an

**TABLE 1** | Data summary of 10 viral metagenomic samples in the South China Sea and adjacent Northwest Pacific Ocean.

Sample name	Raw reads	Clean bases (G)	Q20 (%)	Q30 (%)	GC content (%)	Viral contigs (>300 bp)	vOTUs
F2_5	31,114,321	9.32	98.15	94.17	42.95	10,592	5,987
F2_3500	46,752,263	14.02	96.80	91.35	54.08	5,803	3,051
E1_5	30,307,542	9.05	96.87	91.05	39.35	18,395	10,407
E1_3000	39,497,164	11.83	96.75	91.05	51.86	6,997	3,988
SEATS_5	36,978,668	11.08	97.61	92.71	37.37	23,257	13,001
SEATS_3000	51,175,713	15.34	97.49	92.84	48.92	11,150	6,732
DC2_5	37,768,566	11.31	96.66	90.62	37.25	20,810	11,677
DC2_2000	47,358,488	14.20	97.66	93.27	50.79	8,333	4,994
DC6_5	33,065,273	9.88	97.07	92.15	39.36	10,989	10,124
DC6_2000	54,758,166	16.41	98.19	94.36	52.28	5,626	2,907

estimation of the gamma parameter. The phylogenetic tree was visualized and displayed using iTOL (Interactive Tree of Life) (Letunic and Bork, 2016).

## Statistical Analyses

The cluster analysis and non-metric multidimensional scaling (NMDS) analysis of the DNA viral communities were performed using PRIMER v7 (PRIMER E, Ltd., United Kingdom) (Relative abundance of viruses in each sample represent the biont number). The difference analysis, venn diagram and heatmap of the viral clusters between the surface and deep sea were created using the free online platform of Majorbio I-Sanger Cloud Platform<sup>4</sup>. Canonical correspondence analysis (CCA) was performed in R v. 3.5.1 (R Development Core Team) using CCA and RDA functions from the “vegan” package v2.5-2 (Oksanen et al., 2018) to interrogate the relationships between viral clusters and environmental variables. A matrix of the total viral Operational Taxonomic Units (29,967 vOTUs) was processed for factor analysis. A total of nine environmental variables were used to assess the variation of viral species, including longitude, latitude, depth, temperature, salinity, NO<sub>x</sub>, PO<sub>4</sub>-P, total organic carbon

(TOC) and carbon flux. All variables were logarithmically (base 10) converted before CCA to reduce the influence of extreme values on ordination scores and to normalize data distribution.

## Accession Number

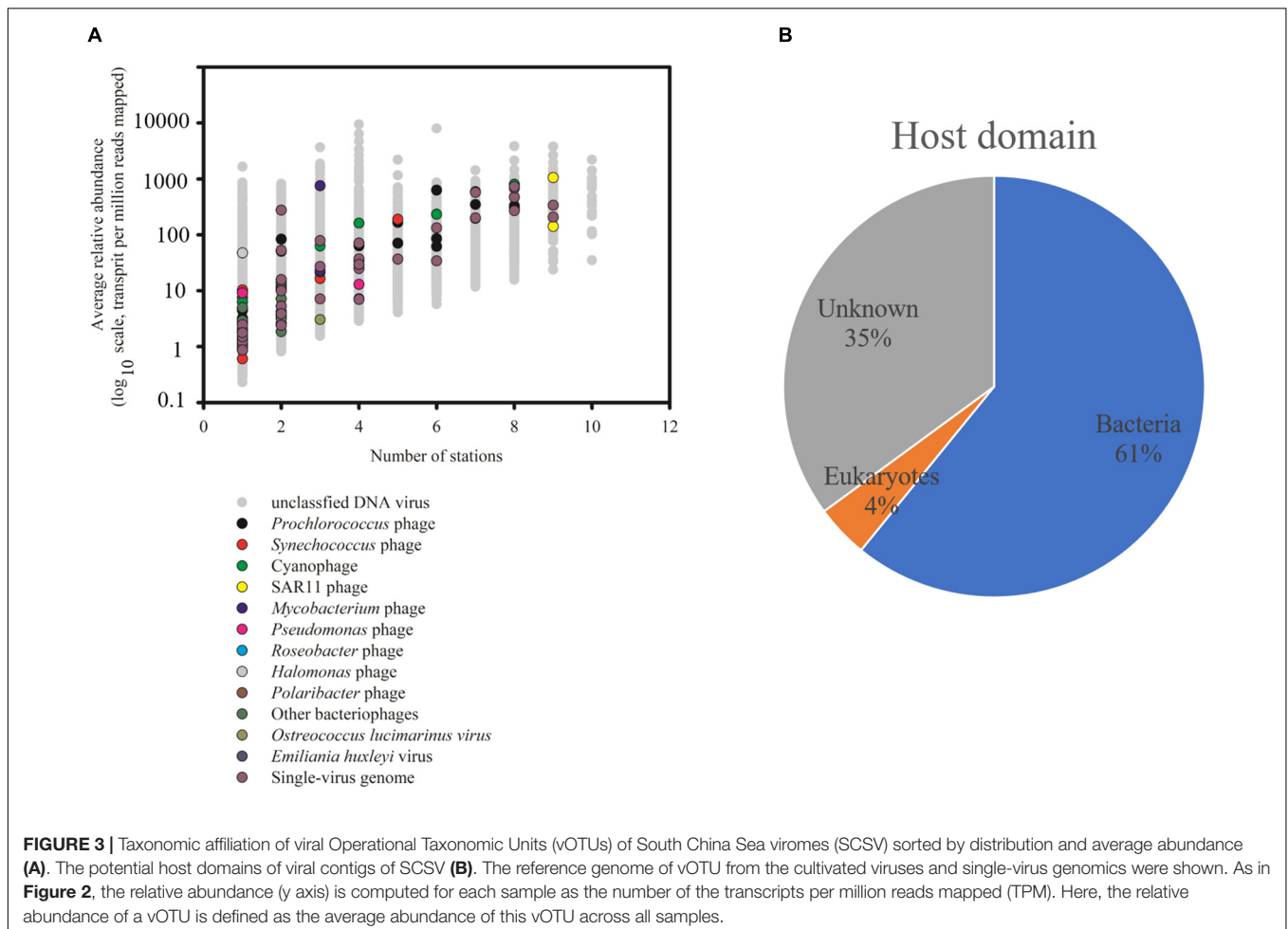
All the viral reads data in this study were submitted to the NCBI Sequence Read Archive (SRA). The SRA accession number: PRJNA535364.

## RESULTS AND DISCUSSION

### The South China Sea Viromes (SCSV) Dataset and Contig Assembly

The 10 South China Sea Viromes (SCSV) data set contains 122 Gb of sequences from five surface and five deep sea (>2000 m) samples in the central basin of the South China Sea and the adjacent Northwest Pacific Ocean (Figure 1 and Table 1). The SCSV data set offers the first glimpse of surface and deep sea viral communities of the South China Sea, which is the largest marginal sea located at the Northwest Pacific Ocean. The SCSV samples were filtered through 0.22 μm pore-size filters, and then the viruses were concentrated using the TFF method. These steps

<sup>4</sup>www.i-sanger.com



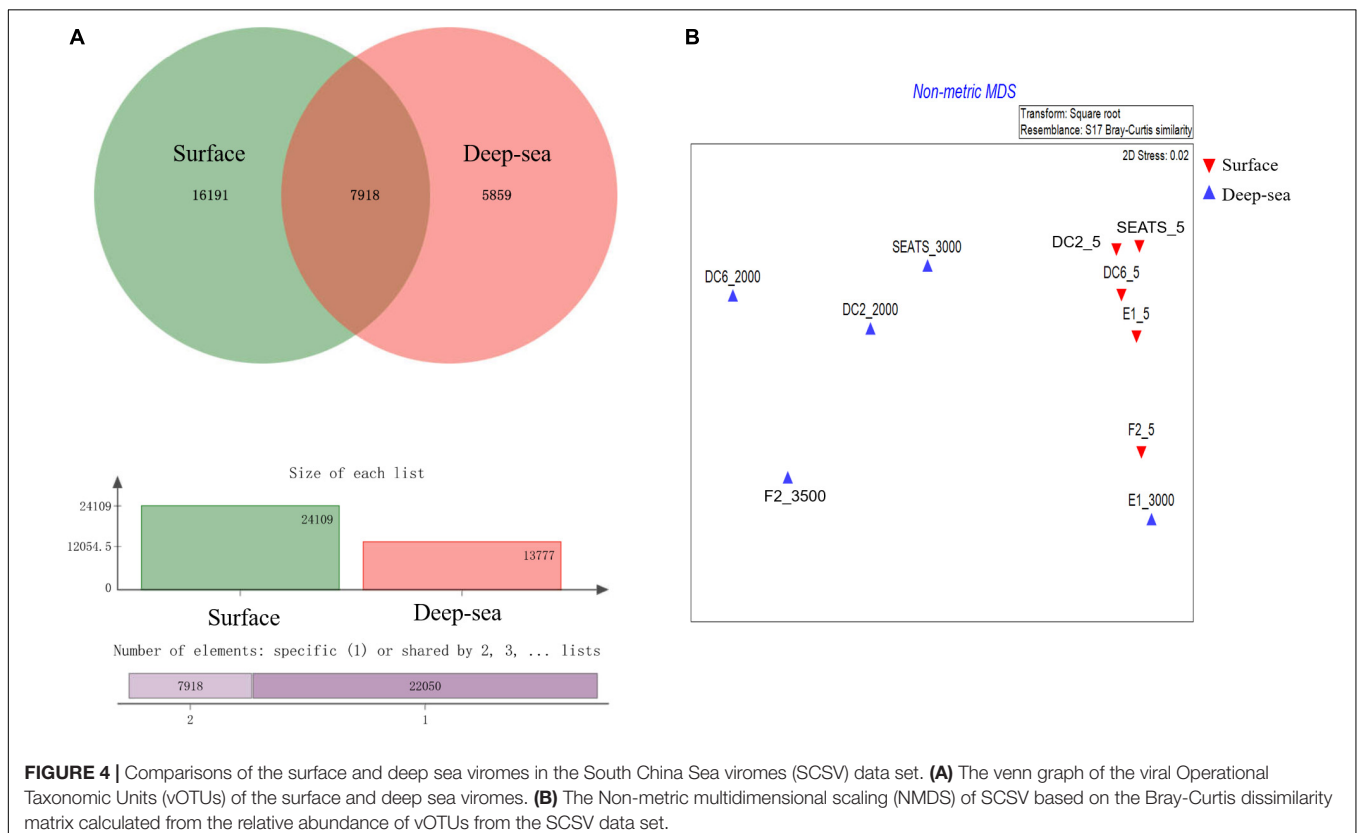
are one of the conventional techniques to collect the virome samples, which would have removed most of the prokaryotic and eukaryotic cells, and would also have excluded most of the viruses larger than 0.22  $\mu\text{m}$  (Brum et al., 2015; Gong et al., 2018). The SCSV viral concentrates were purified using DNase digestion and CsCl density gradients to reduce contamination by non-viral DNA (Hurwitz and Sullivan, 2013). The extracted viral DNA was sequenced directly without the amplification which excludes the bias against single-stranded DNA (ssDNA) viruses caused by the amplification treatment (Duhaime and Sullivan, 2012; Marine et al., 2014).

In the SCSV data set, the assembled contig data was analyzed for the viral predictions using VirSorter V2. A total of 121,952 viral contigs longer than 300 bp (of which 99,791 viral contigs were  $\geq 1,500$  bp) were predicted and ranged from 5626 at DC6\_2000 to 23,257 at SEATS\_5. The viral contigs were blasted against the IMG/VR 2.0 dataset to classify the viral contigs as the vOTUs and then classify vOTUs that corresponding to the viral populations in the GOV 2.0 data set (Gregory et al., 2019). A total of 29,967 vOTUs were predicted (of which 27,249 vOTUs were  $\geq 1,500$  bp, and 5,857 vOTUs were  $\geq 10$  kb) and ranged from 2,907 vOTUs at DC6\_2000m to 13,001 vOTUs at SEATS\_5m, which is comparable to the viral populations from the original Tara Ocean and Malaspina expeditions (15,280 viral populations which are defined as viral contigs have  $\geq 95\%$  ANI across its members) (Roux et al., 2016). The taxonomic analysis of the viral populations based on virome data remains arduous (Brum and Sullivan, 2015; Brum et al., 2015). Only recently, the

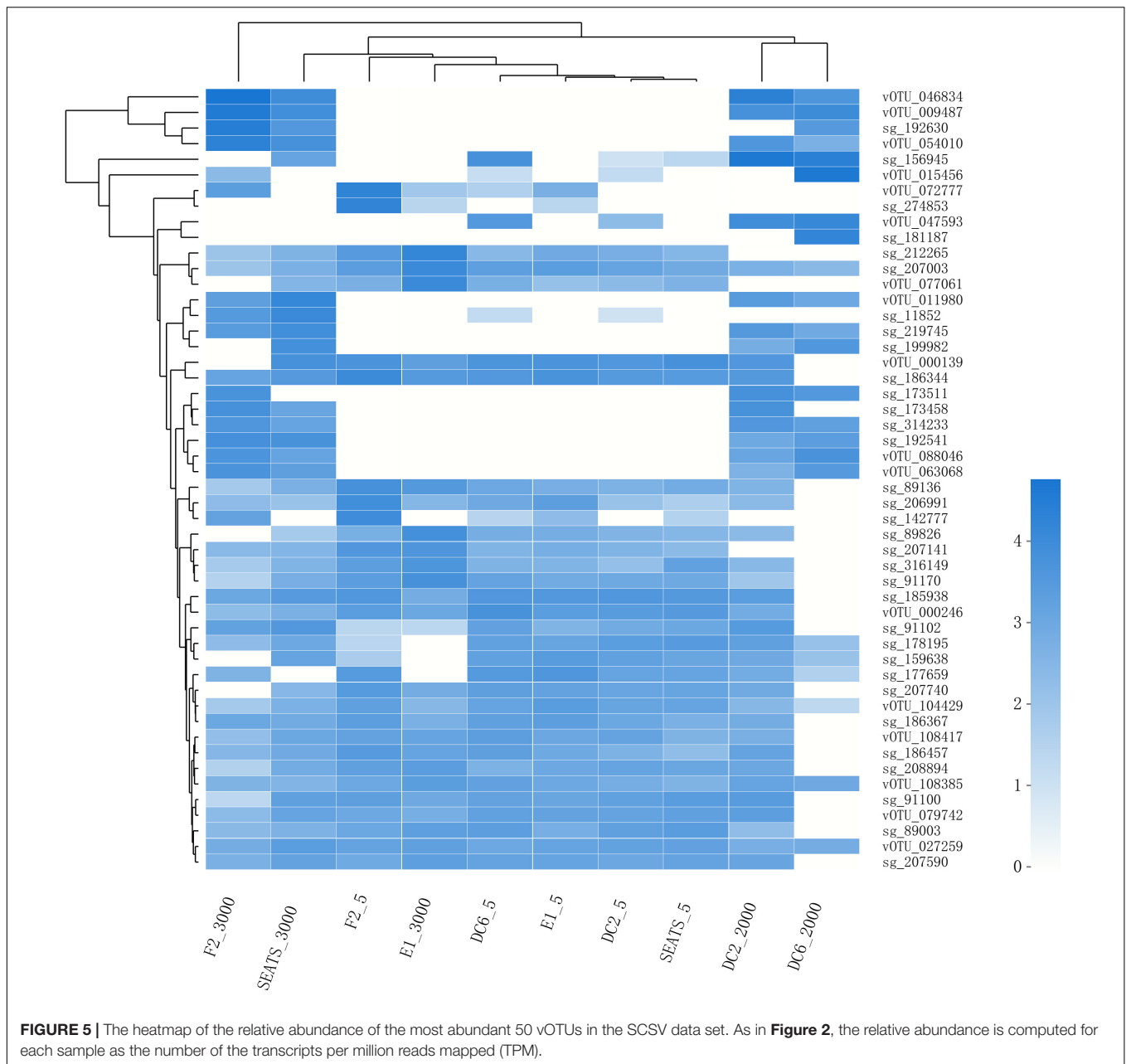
establishment and combination usages of the VirSorter 2.0 and IMG/VR 2.0 data sets have greatly improved the assembly of large contigs (up to 100 kb) and aided the taxonomic analysis of viral populations at the genome-level and viral-cluster-level (Roux et al., 2015; Paez-Espino et al., 2018). Most of the viral contigs of SCSV (117,161 viral contigs, 90.9% of total viral contigs) predicted by the VirSorter 2.0 could be taxonomically classified as vOTUs using the IMG/VR 2.0 data set, indicating the feasibility and high-efficiency of the taxonomic classification using the IMG/VR 2.0 data set.

## The Genetic Diversity of vOTUs in the SCSV

The Shannon's  $H'$  and Chao I diversity indexes of the vOTUs in each sample were shown in **Figure 2**. The deep sea viromes were less diverse (the Shannon's  $H'$  ranged from 6.23 at F2\_3500 to 7.68 at SEATS\_3000, and Chao I ranged from 2907 at DC6\_2000 to 6732 at SEATS\_3000) than the surface viromes (the Shannon's  $H'$  ranged from 7.40 at F2\_5 to 8.54 SEATS\_5 and Chao I ranged from 5987 at F2\_5 to 13,000 at SEATS\_5, respectively). The diversity of vOTUs in the F2 station located at the adjacent Northwestern Pacific Ocean was lower than that in the stations within the South China Sea (**Figure 2**). The high primary production (PP) at the SEATS station ( $473.4 \text{ mg C m}^{-2} \text{ day}^{-1}$ ) and the low PP at the F2 station ( $387.6 \text{ mg C m}^{-2} \text{ day}^{-1}$ ) might be the reason of the diversity patterns of vOTUs in the SCSV data set.



**FIGURE 4 |** Comparisons of the surface and deep sea viromes in the South China Sea viromes (SCSV) data set. **(A)** The venn graph of the viral Operational Taxonomic Units (vOTUs) of the surface and deep sea viromes. **(B)** The Non-metric multidimensional scaling (NMDS) of SCSV based on the Bray-Curtis dissimilarity matrix calculated from the relative abundance of vOTUs from the SCSV data set.



## Taxonomic Composition and Potential Hosts of Viral Communities in the SCSV

The relative abundance of 29,967 vOTUs in the SCSV was estimated using the IMG/VR 2.0. Only 20 vOTUs were observed across all the 10 samples, and 3,893 vOTUs (13.0%) were commonly observed across > five samples. 13,001 vOTUs (43.4%) were endemic to one sample, which is different from the results of Tara Ocean expedition (15% of viral populations were observed at only one station) (Brum et al., 2015). The high percent of the endemic vOTUs in the SCSV might reflect the high diversity of viral communities in the South China Sea. Only 74 of 29,967 vOTUs (0.25%) from IMG/VR 2.0 could be affiliated to cultivated reference viruses,

which reflects the shortage of reference viral genomes in the databases (Brum et al., 2015; Roux et al., 2016). These cultivated viruses include the viruses infecting the dominant and widespread hosts SAR11, *Prochlorococcus*, *Synechococcus*, *Roseobacter*, *Pseudomonas*, *Pseudoalteromonas*, *Mycobacterium*, *Halomonas*, *Polaribacter*, *Ostreococcus lucimarinus*, *Emiliania huxleyi*, etc. (**Figure 3A**; Kang et al., 2013; Zhao et al., 2013; Huang et al., 2015; Liang et al., 2016). Additionally, there are 37 vOTUs could be affiliated to the uncultured viral genomes from the single-virus (35 vOTUs) or single-cell (2 vOTUs) genomics (**Figure 3A**; Martinez-Hernandez et al., 2017; Berube et al., 2018). However, the most abundant and widespread vOTUs observed in the SCSV were from the uncultivated viruses annotated from

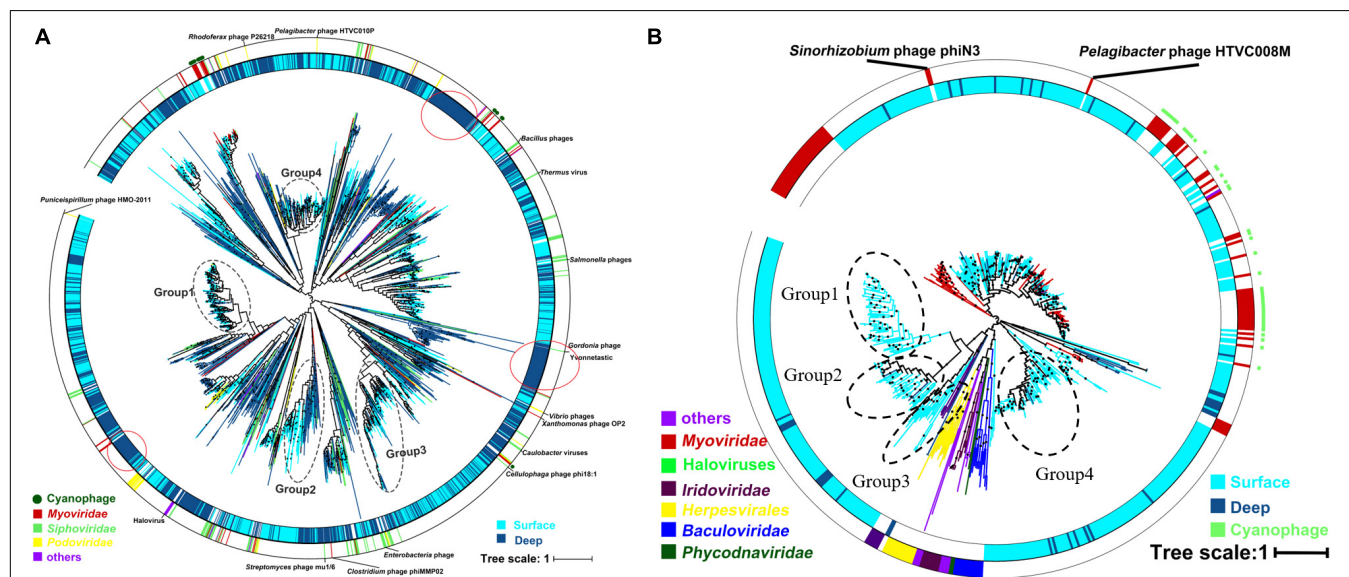
the viral metagenomics (Figure 3A). This indicates that most of the marine viruses are yet to be characterized even though more than 8,000 cultivated viral and single-virus genomes had been reported and deposited in the GenBank and represented several groups of the dominant microbial hosts (Kang et al., 2013; Zhao et al., 2013; Huang et al., 2015; Liang et al., 2016; Martinez-Hernandez et al., 2017; Berube et al., 2018). As most of the vOTUS affiliated to cultivated reference viruses are phages (Figure 3A) and the prominent predicted host domains using the IMG/VR 2.0 datasets are bacteria (62%, Figure 3B), the uncultured phages infecting prokaryotes might dominate the viromes in the SCSV.

## Comparison Between the Surface and Deep Sea Viromes in the SCSV

The result of the venn analysis showed that 7918 vOTUs (26.4%) were observed at both the surface and deep sea viromes, while 16,191 (54.0%) and 5,859 (19.6%) vOTUs were endemic to the surface and deep sea viromes, respectively (Figure 4A). According to the results of NMDS assessed using Bray-Curtis dissimilarity distances, the surface and deep sea viromes (except the E1\_3000m) could be generally divided into two groups, and the distances among the surface viromes were shorter than that of the deep sea viromes (Figure 4B). The heatmap of the relative abundance of most abundant 50 vOTUs showed that the distribution patterns of vOTUs in the deep sea viromes are different from that in the surface viromes (Figure 5). And the phylogenetic tree of the phage terminase large-subunit domains showed the presence of the several new viral groups in the SCSV (Groups 1–4) and the unique viral clusters in the deep sea viromes

(Figure 6A). The result is comparable to the results from the GOV 2.0, which divided the surface and deep sea viromes in the tropical oceans into the bathypelagic (>2,000 m) and temperate and tropical epipelagic (0–150 m) ecological zones (Gregory et al., 2019). For the phylogenetic tree of the DNA polB sequences, several new viral groups in the SCSV were observed (Figure 6B), which is similar with the results from the TerL sequences (Figure 6A). However, the novel viral groups from the DNA polB in the deep-sea viromes were fewer than that from the TerL sequences (Figure 6), which might indicate the smaller proportion of eukaryotic and cyanobacterial viruses than bacteriophages in the deep-sea viromes of SCSV (Chen and Suttle, 1995; Koonin et al., 2015; Chénard et al., 2016; Roux et al., 2017). The results from the phylogenetic analysis using only one marker might not represent the actual phylogenetic relationships among different viruses. In the future, the phylogenetic analysis using the whole viral genomes could provide more solid phylogenetic relationships among different viruses (Castelan-Sanchez et al., 2019).

Interestingly, the viral community in the deep sea virome of the E1 station (E1\_3000m) was quite different from those of other deep sea viromes in the SCSV and closed to the surface water viromes (F2\_5) in the NMDS biplot and heatmap (Figures 4B, 5). The E1 station was at the west side of the Luzon Strait where the horizontal and vertical physical transportation is robust. The strong western boundary current (Kuroshio) transport seawater from the Northwest Pacific Ocean (around F2 station) into the northeast of the South China Sea (around E1 station) (Figure 1; Wang et al., 2011; Hu et al., 2015). The average velocity of flow across interfacial Ekman transport along isopycnal surfaces at their edges was frequently above  $0.5 \text{ m s}^{-1}$ ,

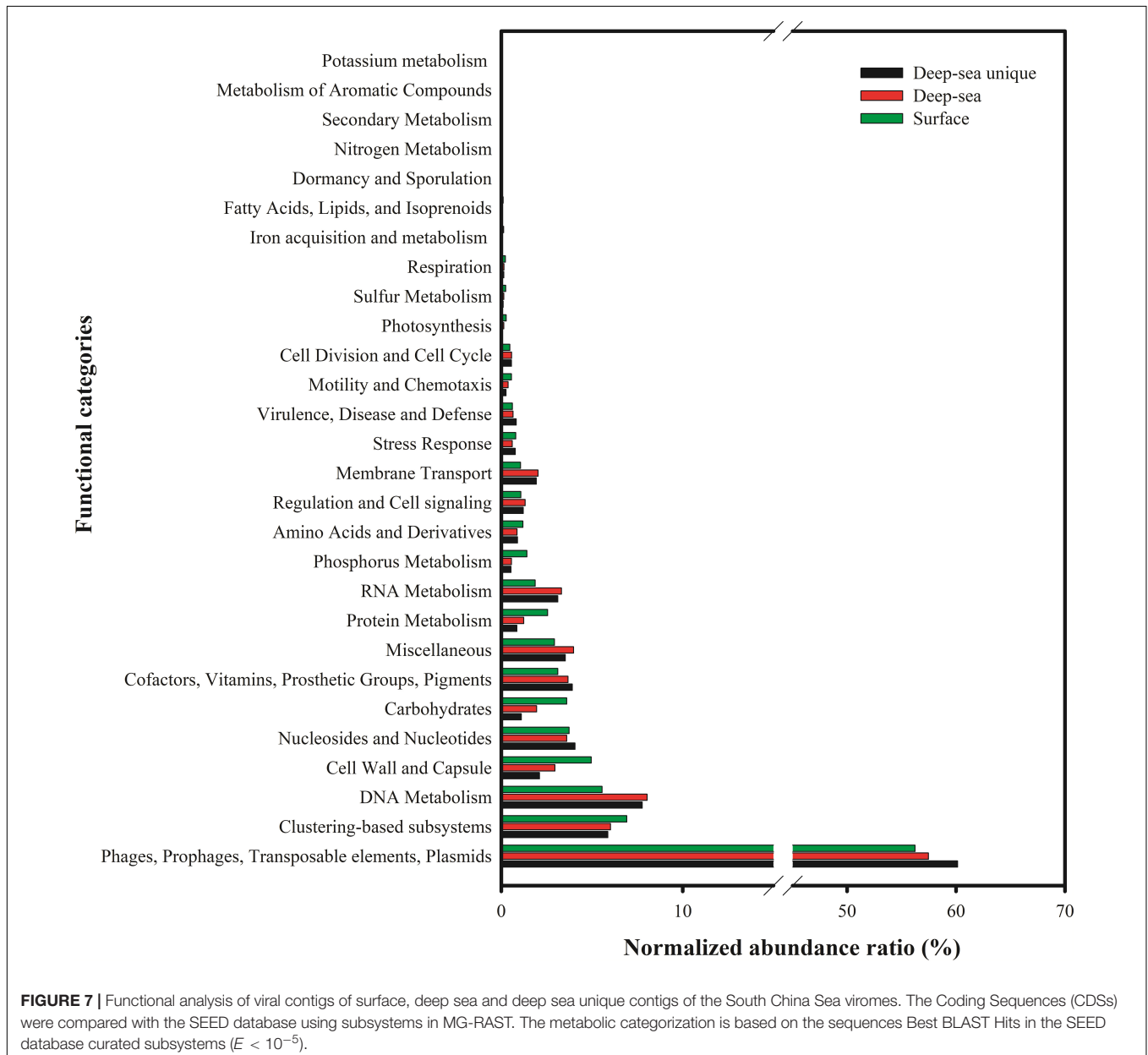


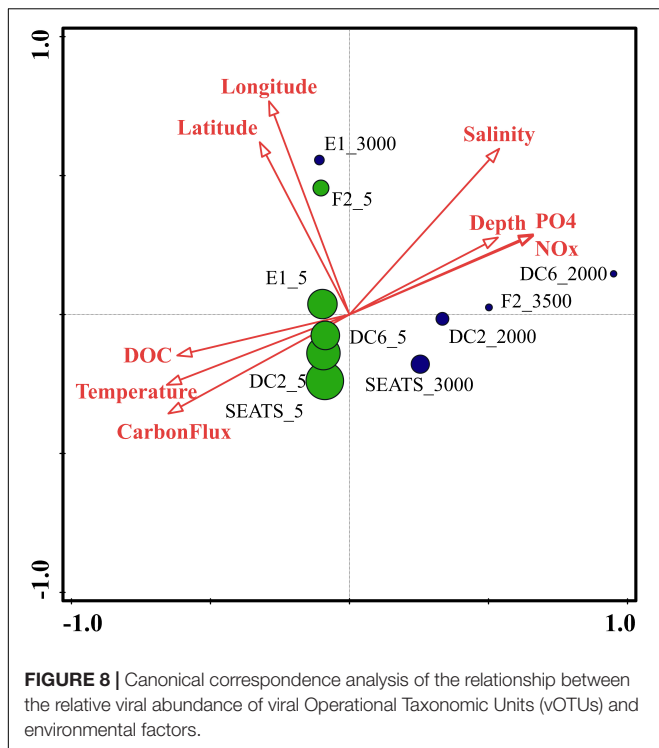
**FIGURE 6 |** Phylogenetic tree of the viral contigs based on the terminase large-subunit domains (A) and family B DNA polymerase (B) in the South China Sea viromes (SCSV) data set. The maximum likelihood phylogenetic trees of the *Caudovirals* terminase large-subunit domains (PF03237, A) and family B DNA polymerase (B) are shown (1000 iterations, JTT + G model). Only bootstrap values >50% are indicated at the nodes of the tree, and bootstrap scores greater than 90% are indicated with a black dot. The original terminase and family B DNA polymerase sequences from the surface viromes in the SCSV are indicated by wathet blue, and the original sequences from the deep sea viromes in the SCSV are in dark blue. Reference sequences are marked (see color legend at the bottom).



which could transfer the surface water to the depth of  $\sim 1,000$  m within a few days (Yuan, 2002; Tian et al., 2009). Other vertical physical transportation like mesoscale eddies and solitons were active at the Luzon Strait (Yuan, 2002; Warnvamas et al., 2010). The combination of the vertical physical transportations could probably transport the *Prochlorococcus* from surface waters into the mesopelagic zones (800 m) of the Luzon Strait (Jiao et al., 2014). We assumed that the strong horizontal and vertical physical transportation might be responsible for the existence of surface viruses of the adjacent Northwest Pacific Ocean (around F2 station) in the deep sea viromes of Northeast South China Sea (around E1 station). This is supported by the transportation of the surface viromes by the surface water currents in the Tara Ocean expedition (Brum et al., 2015).

The putative functions of the viral contigs from the five surface viromes, five deep sea viromes and the deep sea unique viral contigs (5,859 contigs) were predicted using MG-RAST. Using the subsystems approach, nearly 58% (56.2–60.1%) of the annotated proteins were classified as “Phage, Prophage, Transposable elements, or Plasmids” (Figure 7). “DNA metabolism” (5.6–8.0%), “nucleosides/nucleotides” (3.6–4.1%), “Cofactors, Vitamins, Prosthetic groups, Pigment” (3.1–3.9%), “Miscellaneous” (2.9–4.0%), “Cell Wall and Capsule” (2.1–5.0%) were most commonly identified ( $> 3\%$ ) and 5.9–6.9% of them were classified into “Clustering-based subsystems.” The deep sea unique viral contigs contain the highest proportion of functional categories of “Phage, Prophage, Transposable elements, or Plasmids” and the lowest proportions of “Cell





Wall and Capsule,” “Carbohydrates,” “Protein metabolism,” and “Phosphorus metabolism” (Figure 7), which indicated the unique characteristics of deep sea viruses. It is likely that the deep sea viruses have less auxiliary metabolic activities compared to the viruses in the euphotic zone. As the functional analysis was limited to MGRAST, which is mainly limited to bacterial gene data rather than the viral and eukaryotic database, the functional analysis in the study might not represent the total functional properties of the viral communities in the SCSV. The usages of other platforms and strategies to analyze the virome (i.e., Virome, ViromeScan, Virus seeker, etc.) could improve the functional prediction of the SCSV in the future.

## Relationship Between Viral Community Structure and Environmental Factors

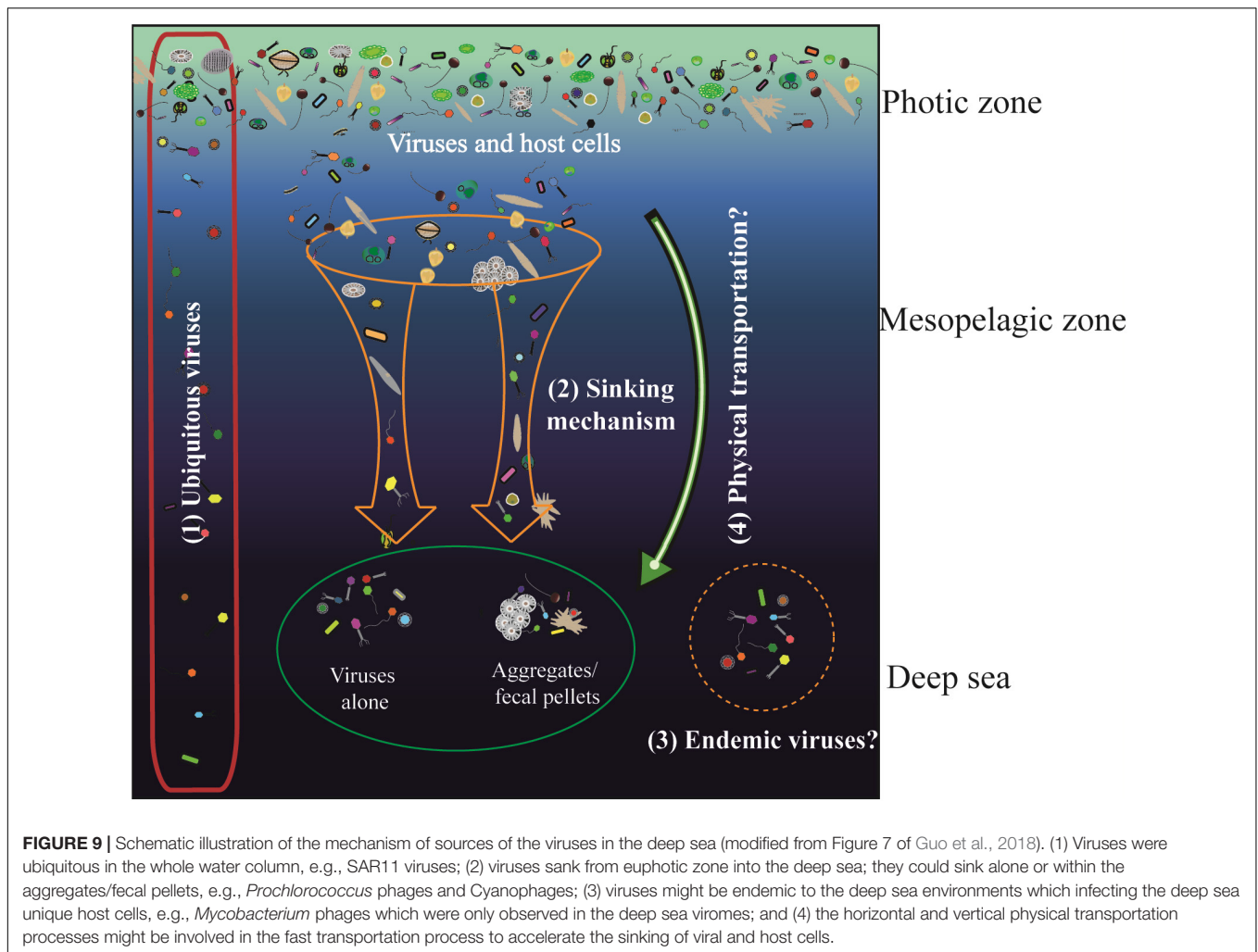
To identify the best predictor variables to explain the variation of the DNA viral community framework in the surface and deep sea of the South China Sea, multivariate regression analysis was used (Figure 8). The first CCA axis interpreted 31% of the total variability in the vOTUs, and the first two axes explained 48% of the total variability. The CCA demonstrated three clear groups of DNA viromes, including the surface viromes, the deep sea viromes and the Luzon physical transport viromes (F2\_5 and E1\_3000). Most of surface DNA viromes (E1\_5, DC6\_5, DC2\_5, and SEATS\_5) were firmly related to the carbon flux, temperature, and DOC. The deep sea DNA viromes (DC6\_2000, F2\_3500, DC2\_2000, and SEATS\_3000) were related to PO<sub>4</sub>, NO<sub>x</sub>, depth and salinity, while the Luzon vertical physical transport DNA viromes were most closely related to the longitude and latitude (Figure 8). The results suggest that the biogeography

of surface and deep sea viral communities in the SCSV is structured by the environmental factors, which is similar with the results from the Tara Ocean, GOV 2.0 and Antarctic viromes (Brum et al., 2015; Roux et al., 2016; Gong et al., 2018; Gregory et al., 2019; Yang et al., 2019).

As the viruses require host cells to replicate, the biogeography trends of viruses usually follow the trends of the host community, which is mostly shaped by the environmental factors (Brum et al., 2015; Liang et al., 2017; Gregory et al., 2019). Hence, the relationship between viral communities and environmental factors in the SCSV might be the reflections of the relationships between host and viruses and between host communities and environmental factors. Interestingly, the close relationship between the surface virome of Northwest Pacific Ocean (F2\_5) and deep sea virome of the Northeast South China Sea (E1\_3000) might suggest the passive transport of viruses by not only the horizontal water currents which is testified by the Tara Ocean expedition (Brum et al., 2015), but also the vertical physical transportations (Figures 1, 4B, 5).

## The Possible Sources of the Deep Sea Viruses in the SCSV

According to the community structures and comparisons of the surface and deep sea viromes in the SCSV, we hypothesized several potential sources of the deep sea viromes (Figure 9). Firstly, some of the viruses are ubiquitous in the whole water column, which is coincident with their host organisms. For example, the SAR11 viruses (*Pelagibacter* phage HTVC010P) and SAR11 were abundant in both the surface and deep sea waters (Figures 3A, 5, 9; Zhao et al., 2013; Nunoura et al., 2015; Zhang et al., 2016). Secondly, the sedimentation mechanism might contribute to the formation of the deep sea viromes (Guidi et al., 2016). Several potential phages of autotrophic cyanobacteria were detected in the deep sea viromes, e.g., *Prochlorococcus* phages (P-GSP1, P-SSM4, and P-TIM68) and Cyanophages (MED4-184 and MED4-117) (Figures 3A, 5, 9). As *Prochlorococcus* are prominently distributed in the euphotic ocean (0–200 m), the existence of cyanophages in the deep sea viromes could be explained by the sedimentation mechanism, which is the most common and widely acknowledged theory accounting for the life of phytoplankton cells in the deep sea (Jiao et al., 2014; Agusti et al., 2015; Guo et al., 2018). The co-sinking of cyanophages with their host cells might also contribute to the existence of cyanophages in the viromes of the deep sea and sub-seafloor sediments (Guidi et al., 2016; Berube et al., 2018; Cai et al., 2019). Thirdly, a significant fraction of viruses is endemic to the deep sea viromes (Figures 4A, 5, 9), which might infect the deep sea dominant host organisms. For example, several vOTUs related to the *Mycobacterium* phages (Sparky and Squirty) were detected in the deep sea viromes and could probably infect the *Mycobacterium*, which was commonly detected in the deep sea waters and sediments (Lu et al., 2011). Fourthly, the similarity of viromes between E1\_3000 and F2\_5 might suggest that the passive transport of viruses by the horizontal and vertical physical transportations contributed to the formations of the deep sea viromes (Figures 1, 4B, 5, 9). In the future, intensive sampling



**FIGURE 9** | Schematic illustration of the mechanism of sources of the viruses in the deep sea (modified from Figure 7 of Guo et al., 2018). (1) Viruses were ubiquitous in the whole water column, e.g., SAR11 viruses; (2) viruses sank from euphotic zone into the deep sea; they could sink alone or within the aggregates/fecal pellets, e.g., *Prochlorococcus* phages and Cyanophages; (3) viruses might be endemic to the deep sea environments which infecting the deep sea unique host cells, e.g., *Mycobacterium* phages which were only observed in the deep sea viromes; and (4) the horizontal and vertical physical transportation processes might be involved in the fast transportation process to accelerate the sinking of viral and host cells.

and studies of viromes in the areas between Northwestern Pacific Ocean and northeast of South China Sea could give robust verification about the hypothesis that viral communities might be used to trace water sources.

## CONCLUSION

This is the first study to explore and reveal the main characteristics of both the surface and deep sea viromes in the South China Sea using the metagenomic analysis. The most dominant and widely spread viral populations in the SCSV were not similar with the reported genomes from the cultivated viruses and the single-virus genomics, which reflected the superiority of viral metagenomics in revealing the biogeography patterns of natural viromes and their relationship with environmental factors. The community structures of deep sea viromes were different from the surface viromes, though one deep sea virome (E1\_3000) in the Northeast South China Sea is similar with one surface virome in the Northwest Pacific Ocean, which suggested the possible passive transport of marine viruses by the horizontal and vertical physical transportations. The presence of viruses in

the deep sea waters may benefit from ubiquitous viruses, sinking mechanism, endemic living on the deep sea host communities and passive transport through physical transportation. In the future, it will be obligatory to evaluate the host community structure simultaneously, to verify the notionally theorized linkage among viruses, host communities and environmental conditions. The comparisons between SCSV and other studies, such as Tara Ocean, Malaspina and etc. could illustrate the unique feature of SCSV in the future.

## DATA AVAILABILITY

The datasets generated for this study can be found in NCBI, PRJNA535364.

## AUTHOR CONTRIBUTIONS

YL, YZ, and NJ designed this study. YL, ZW, and JZ performed the experiments. Data were analyzed by YL, LW, and QY. YL and YZ wrote the manuscript. LW, MW, KY, LZ, and

NJ contributed to writing by providing suggestions and helping with the revisions. All authors reviewed and approved the final version of the manuscript.

## FUNDING

This work was supported by the open task of Qingdao National Laboratory for Marine Science and Technology (QNL2016ORP0311), the National Key Research and Development Program of China (2016YFA0601402), the Senior User Project of RV KEXUE (KEXUE2019GZ03 and KEXUE2018G02) supported by Center for Ocean Mega-Science, Chinese Academy of Sciences, NSFC projects (41876174, 41976117, and 41606153), the DICP&QIBEBT (DICP&QIBEBT UN201803), the QIBEBT (QIBEBT ZZBS 201805), Dalian National Laboratory For Clean Energy (DNL), CAS, the Key R&D project in Shandong Province (2019GHY112079), and the

MEL Visiting Fellowship Program (MELRS1511). This study was a contribution to the international IMBER project.

## ACKNOWLEDGMENTS

We sincerely thank the captain and crews of the R/V Dongfanghong 2, chief scientist Dr. Xiaolin Li for their support on board the ship and providing the temperature, salinity and total organic carbon data. We are grateful to Dr. Minhan Dai for providing the nutrient data. We thank Drs. Curtis Suttle, Huiluo Cao, and Alejandra Prieto-Dav for their kind suggestions which improved the quality of the study. We thank Mr. Shailesh S. Nair for helping with the proofreading. Data and samples were collected onboard of R/V Dongfanghong 2 implementing the open research cruise NORC2017-05 supported by NSFC ShipTime Sharing Project.

## REFERENCES

- Agusti, S., Gonzalez-Gordillo, J. I., Vaque, D., Estrada, M., Cerezo, M. I., Salazar, G., et al. (2015). Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nat. Commun.* 6:7608. doi: 10.1038/ncomms8608
- Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., et al. (2006). The marine viromes of four oceanic regions. *PLoS Biol.* 4:e368. doi: 10.1371/journal.pbio.0040368
- Antia, A. N., Koeve, W., Fischer, G., Blanz, T., Schulz-Bull, D., Scholten, J., et al. (2001). Basin-wide particulate carbon flux in the Atlantic Ocean: regional export patterns and potential for atmospheric CO<sub>2</sub> sequestration. *Glob. Biogeochem. Cycl.* 15, 845–862.
- Aylward, F. O., Boeuf, D., Mende, D. R., Wood-Charlson, E. M., Vislova, A., Eppley, J. M., et al. (2017). Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl. Acad. Sci. U.S.A.* 114, 11446–11451. doi: 10.1073/pnas.1714821114
- Behrenfeld, M. J., and Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol. Oceanogr.* 42, 1–20.
- Berube, P. M., Biller, S. J., Hackl, T., Hogle, S. L., Satinsky, B. M., Becker, J. W., et al. (2018). Single cell genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine environments. *Sci. Data* 5, 180154. doi: 10.1038/sdata.2018.154
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., et al. (2002). Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. U.S.A.* 99:14250.
- Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doucier, G., Acinas, S. G., Alberti, A., et al. (2015). Ocean plankton. Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. doi: 10.1126/science.1261498
- Brum, J. R., and Sullivan, M. B. (2015). Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* 13, 147–159. doi: 10.1038/nrmicro3404
- Cai, L., Jorgensen, B. B., Suttle, C. A., He, M., Cragg, B. A., Jiao, N., et al. (2019). Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years. *ISME J.* 13, 1857–1864. doi: 10.1038/s41396-019-0397-9
- 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:E35. doi: 10.3390/v8020035
- Castelan-Sanchez, H. G., Lopez-Rosas, I., Garcia-Suastegui, W. A., Peralta, R., Dobson, A. D. W., Batista-Garcia, R. A., et al. (2019). Extremophile deep-sea viral communities from hydrothermal vents: structural and functional analysis. *Mar. Genomics* 46, 16–28. doi: 10.1016/j.margen.2019.03.001
- Chen, F., and Suttle, C. A. (1995). Amplification of DNA polymerase gene fragments from viruses infecting microalgae. *Appl. Environ. Microbiol.* 61, 1274–1278.
- Chénard, C., Wirth, J. F., and Suttle, C. A. (2016). Viruses infecting a freshwater filamentous cyanobacterium (*Nostoc* sp.) Encode a functional CRISPR array and a proteobacterial DNA polymerase B. *mBio* 7:e00667-16. doi: 10.1128/mBio.00667-16
- Culley, A. I., Lang, A. S., and Suttle, C. A. (2006). Metagenomic analysis of coastal RNA virus communities. *Science* 312, 1795–1798.
- Duhaime, M. B., and Sullivan, M. B. (2012). Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* 434, 181–186. doi: 10.1016/j.virol.2012.09.036
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Gong, Z., Liang, Y., Wang, M., Jiang, Y., Yang, Q., Xia, J., et al. (2018). Viral diversity and its relationship with environmental factors at the surface and deep sea of Prydz Bay, Antarctica. *Front. Microbiol.* 9:2981. doi: 10.3389/fmicb.2018.02981
- Gregory, A. C., Zayed, A. A., Conceicao-Neto, N., Temperton, B., Bolduc, B., Alberti, A., et al. (2019). Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177, 1109.e14–1123.e14. doi: 10.1016/j.cell.2019.03.040
- Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532, 465–470. doi: 10.1038/nature16942
- Guo, R., Liang, Y., Xin, Y., Wang, L., Mou, S., Cao, C., et al. (2018). Insight into the pico- and nano-phytoplankton communities in the deepest biosphere, the Mariana Trench. *Front. Microbiol.* 9:2289. doi: 10.3389/fmicb.2018.02289
- Han, A. Q., Dai, M. H., Kao, S. J., Gan, J. P., Li, Q., Wang, L. F., et al. (2012). Nutrient dynamics and biological consumption in a large continental shelf system under the influence of both a river plume and coastal upwelling. *Limnol. Oceanogr.* 57, 486–502.
- 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
- Hu, D., Wu, L., Cai, W., Gupta, A. S., Ganachaud, A., Qiu, B., et al. (2015). Pacific western boundary currents and their roles in climate. *Nature* 522, 299–308. doi: 10.1038/nature14504
- Huang, S., Wilhelm, S. W., Jiao, N., and Chen, F. (2010). Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *ISME J.* 4, 1243–1251. doi: 10.1038/ismej.2010.56
- Huang, S., Zhang, S., Jiao, N., and Chen, F. (2015). Marine cyanophages demonstrate biogeographic patterns throughout the global ocean.

- Appl. Environ. Microbiol.* 81, 441–452. doi: 10.1128/AEM.02483-14
- Hurwitz, B. L., and Sullivan, M. B. (2013). The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* 8:e57355. doi: 10.1371/journal.pone.0057355
- Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., et al. (2010). Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.* 8, 593–599. doi: 10.1038/nrmicro2386
- Jiao, N., Luo, T., Zhang, R., Yan, W., Lin, Y., Johnson, Z. I., et al. (2014). Presence of *Prochlorococcus* in the aphotic waters of the western Pacific Ocean. *Biogeosciences* 11, 2391–2400.
- Kang, I., Oh, H. M., Kang, D., and Cho, J. C. (2013). Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12343–12348. doi: 10.1073/pnas.1219930110
- Koonin, E. V., Krupovic, M., and Yutin, N. (2015). Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. *Ann. N. Y. Acad. Sci.* 1341, 10–24. doi: 10.1111/nyas.12728
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659.
- Liang, Y., Zhang, Y., Zhang, Y., Luo, T., Rivkin, R. B., and Jiao, N. (2017). Distributions and relationships of virio- and pico-plankton in the epi-, meso- and bathypelagic zones of the Western Pacific Ocean. *FEMS Microbiol. Ecol.* 93:fiw238.
- Liang, Y., Zhang, Y., Zhou, C., Chen, Z., Yang, S., Yan, C., et al. (2016). Complete genome sequence of the siphovirus roseophage RDJLPhi 2 infecting *Roseobacter denitrificans* OCh114. *Mar. Genomics* 25, 17–19. doi: 10.1016/j.margen.2015.10.009
- Lu, Z., Deng, Y., Van Nostrand, J. D., He, Z., Voordeckers, J., Zhou, A., et al. (2011). Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME J.* 6, 451–460. doi: 10.1038/ismej.2011.91
- Marine, R., Mccarren, C., Vorrasane, V., Dan, N., Crowgey, E., Polson, S. W., et al. (2014). Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. *Microbiome* 2, 3. doi: 10.1186/2049-2618-2-3
- Martinez-Hernandez, F., Fornas, O., Lluésma Gomez, M., Bolduc, B., De La Cruz Pena, M. J., Martínez, J. M., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* 8:15892. doi: 10.1038/ncomms15892
- Meyer, F., Paarmann, D., D'souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.* 9:386. doi: 10.1186/1471-2105-9-386
- Mizuno, C. M., Ghai, R., Saghai, A., Lopez-Garcia, P., and Rodriguez-Valera, F. (2016). Genomes of Abundant and widespread viruses from the Deep Ocean. *mBio* 7:e00805-16. doi: 10.1128/mBio.00805-16
- Noguchi, H., Taniguchi, T., and Itoh, T. (2008). MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* 15, 387–396. doi: 10.1093/dnares/dsn027
- Nunoura, T., Takaki, Y., Hirai, M., Shimamura, S., Makabe, A., Koide, O., et al. (2015). Hadal biosphere: insight into the microbial ecosystem in the deepest ocean on Earth. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1230–E1236. doi: 10.1073/pnas.1421816112
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., et al. (2018). *vegan: Community Ecology Package (R Package Version 2.5.2)*. Available at: <https://CRAN.R-project.org/package=vegan>
- Paez-Espino, D., Roux, S., Chen, I. A., Palaniappan, K., Ratner, A., Chu, K., et al. (2018). IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res.* 47, D678–D686. doi: 10.1093/nar/gky1127
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., et al. (2018). Minimum information about an uncultivated virus genome (MIUViG). *Nat. Biotechnol.* 37, 29–37. doi: 10.1038/nbt.4306
- Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537, 689–693. doi: 10.1038/nature19366
- Roux, S., Chan, L. K., Egan, R., Malmstrom, R. R., McMahon, K. D., and Sullivan, M. B. (2017). Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nat. Commun.* 8:858. doi: 10.1038/s41467-017-01086-2
- Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. doi: 10.7717/peerj.985
- Steward, G. F., and Preston, C. M. (2011). Analysis of a viral metagenomic library from 200 m depth in Monterey Bay, California constructed by direct shotgun cloning. *Virology* 437, 8287. doi: 10.1016/j.virus.2011.08.028
- Sun, G., Xiao, J., Wang, H., Gong, C., Pan, Y., Yan, S., et al. (2014). Efficient purification and concentration of viruses from a large body of high turbidity seawater. *MethodsX* 1, 197–206. doi: 10.1016/j.mex.2014.09.001
- Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361.
- Suttle, C. A. (2007). Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812.
- Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4, 470–483. doi: 10.1038/nprot.2009.10
- Tian, J., Yang, Q., and Zhao, W. (2009). Enhanced diapycnal mixing in the South China Sea. *J. Phys. Oceanogr.* 39, 3191–3203.
- Wang, G., Xie, S.-P., Qu, T., and Huang, R. X. (2011). Deep South China Sea Circulation. *Geophys. Res. Lett.* 38:L05601.
- Warnvamas, A., Hawkins, J., Lamb, K. G., Piacsek, S., Chinbing, S., King, D., et al. (2010). Solitary wave generation dynamics at Luzon Strait. *Ocean Model.* 31, 9–27.
- Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* 28, 127–181.
- Williamson, S. J., Allen, L. Z., Lorenzi, H. A., Fadrosch, D. W., Bami, D., Thiagarajan, M., et al. (2012). Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS One* 7:e42047. doi: 10.1371/journal.pone.0042047
- Winter, C., Garcia, J. A., Weinbauer, M. G., Dubow, M. S., and Herndl, G. J. (2014). Comparison of deep-water viromes from the Atlantic Ocean and the Mediterranean Sea. *PLoS One* 9:e100600. doi: 10.1371/journal.pone.0100600
- Wommack, K. E., and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114.
- Wu, K., Dai, M., Chen, J., Meng, F., Li, X., Liu, Z., et al. (2015). Dissolved organic carbon in the South China Sea and its exchange with the Western Pacific Ocean. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 122, 41–51.
- Yang, Q., Gao, C., Jiang, Y., Wang, M., Zhou, X., Shao, H., et al. (2019). Metagenomic characterization of the viral community of the South Scotia Ridge. *Viruses* 11:E95. doi: 10.3390/v11020095
- Yokokawa, T., Yang, Y., Motegi, C., and Nagata, T. (2013). Large-scale geographical variation in prokaryotic abundance and production in meso- and bathypelagic zones of the central Pacific and Southern Ocean. *Limnol. Oceanogr.* 58, 61–73.

- Yuan, D. (2002). A numerical study of the South China Sea deep circulation and its relation to the Luzon Strait transport. *Acta Oceanol. Sin.* 21, 187–202.
- Zhang, Y., Xiao, W., and Jiao, N. (2016). Linking biochemical properties of particles to particle-attached and free-living bacterial community structure along the particle density gradient from freshwater to open ocean. *J. Geophys. Res. Biogeosci.* 121, 2261–2274.
- Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., et al. (2013). Abundant SAR11 viruses in the ocean. *Nature* 494, 357–360. doi: 10.1038/nature11921

**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.

Copyright © 2019 Liang, Wang, Wang, Zhao, Yang, Wang, Yang, Zhang, Jiao and Zhang. 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(s) 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.