


Special Issue Article

T4-like myovirus community shaped by dispersal and deterministic processes in the South China Sea

Huifang Li,¹ Lu Liu,¹ Yu Wang ¹, Lanlan Cai,² Maoqiu He,³ Long Wang,¹ Chen Hu,¹ Nianzhi Jiao¹ and Rui Zhang ^{1*}

¹State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Fujian Key Laboratory of Marine Carbon Sequestration, Xiamen University (Xiang'an), Xiamen, Fujian, China.

²Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong, China.

³State Key Laboratory of Trophic Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China.

Summary

As the most abundant and genetically diverse biological entities, viruses significantly influence ecological, biogeographical and evolutionary processes in the ocean. However, the biogeography of marine viruses and the drivers shaping viral community are unclear. Here, the biogeographic patterns of T4-like viruses and the relative impacts of deterministic (environmental selection) and dispersal (spatial distance) processes were investigated in the northern South China Sea. The dominant viral operational taxonomic units were affiliated with previously defined Marine, Estuary, Lake and Paddy Groups. A clear viral biogeographic pattern was observed along the environmental gradient from the estuary to open sea. Marine Groups I and IV had a wide geographical distribution, whereas Marine Groups II, III and V were abundant in lower-salinity continental or eutrophic environments. A significant distance-decay pattern was noted for the T4-like viral community, especially for those infecting cyanobacteria. Both deterministic and dispersal processes influenced viral community

assembly, although environmental selection (e.g. temperature, salinity, bacterial abundance and community, etc.) had a greater impact than spatial distance. Network analysis confirmed the strong association between viral and bacterial community composition, and suggested a diverse ecological relationship (e.g. lysis, co-infection or mutualistic) between and within viruses and their potential bacterial hosts.

Introduction

Viruses are the most abundant and diverse entities in marine ecosystems and considerably impact ecological, biogeographical and evolutionary processes in the ocean. Viruses could be responsible for reducing primary production in the oceans by up to 78%, and for an important percentage (~20%) of the mortality of marine heterotrophic bacteria (Suttle *et al.*, 1990; Suttle, 1994). Therefore, it is becoming increasingly clear that we need to incorporate viruses and virus-mediated processes into our perception of marine ecology and oceanography (Suttle, 2007). Understanding the occurrence and mechanisms of the biogeographic patterns of marine viruses will help in predicting their influence on host populations and microbial community composition, and thus on large-scale oceanic processes (Stegen *et al.*, 2012; Marston *et al.*, 2013; Chow and Suttle, 2015; Gregory *et al.*, 2019).

Recently, researchers have observed the existence of biogeographic patterns, often defined as the presence of spatial patterns of biodiversity resulting from deterministic (selection) and dispersal (spatial distance) processes, among a wide range of microorganisms (Hanson *et al.*, 2012; Chow and Suttle, 2015). Selection varies the relative abundance of microbial species based on their ability to survive and reproduce. Potential variables influencing selection include the physical, chemical and biotic factors of a given environment. Dispersal refers to the movement, and subsequent successful establishment, of a microbial species to a new location (Hanson *et al.*, 2012). Williamson *et al.* (2008) implicated that

Received 14 October, 2020; accepted 19 October, 2020. *For correspondence. E-mail ruizhang@xmu.edu.cn; Tel. 86-592-2880152; Fax +86-592-2185375.

marine viral communities display biogeographic patterns that are widely dispersed and influenced by local environmental selective pressures. Several other surveys also found a significant correlation between viral community structure and environmental parameters (Desnues *et al.*, 2008; Marston *et al.*, 2013). For instance, Angly *et al.* (2006) showed that salinity, temperature and oxygen concentration shape viral communities in the Gulf of Mexico, the Sargasso Sea and the Arctic Ocean. Frederickson *et al.* (2003) found that the physical structure of the water column influenced the structure of cyanophage communities in inlets in Canada. Studies on the influence of microbes on viral diversity and abundance both in coastal waters and the open ocean have shown that viral community structure may be strongly influenced by their hosts' communities (Fuhrman, 1999; Short and Suttle, 2002; Winter *et al.*, 2005; Sandaa *et al.*, 2007; Suttle, 2007; Parada *et al.*, 2008; Short *et al.*, 2011; Breitbart, 2012; Chow and Fuhrman, 2012; De Corte *et al.*, 2016). While several studies have suggested that viral community diversity is mainly influenced by dispersal processes (Snyder *et al.*, 2007; Winter *et al.*, 2013), yet the influence of abiotic environmental variables, potential host community and spatial parameters are seldom investigated simultaneously in viral biogeographic studies.

The South China Sea (SCS), which is connected to the Pacific Ocean by the Luzon Strait, is one of the largest semi-enclosed marginal seas in the world. It has a wide continental shelf in the northwest area, which receives high volumes of runoff ($3 \times 10^{11} \text{ m}^3 \text{ v}^{-1}$) from the Pearl River, as well as a 4700-m-deep basin (Shaw and Chao, 1994; Zhang *et al.*, 1999; Chen *et al.*, 2001). The northern South China Sea (nSCS) is separated into three distinct regions based on bathymetric features: estuarine coastal plume (mostly eutrophic coastal region), continental shelf (transitional region) and open ocean (oligotrophic region) (He *et al.*, 2009). Environmental variables such as salinity and nutrient availability often form a gradient in estuarine environments, producing a distance effect on spatial changes in microbial composition. The SCS is thus an ideal environment for investigating how viral community composition (VCC) and its control mechanisms alter with regard to distance.

T4-like phages are a subset of tailed phages that make up a significant fraction of the viral communities (Ackermann and Krisch, 1997; Breitbart *et al.*, 2002; Brum *et al.*, 2013). Here, we investigated the impact of environmental selection and dispersal processes on T4-like viral community assembly along an ~500-km transect with an environmental gradient from the estuary to open sea. Because most viruses in the ocean infect prokaryotes (Weinbauer, 2004; Baudoux *et al.*, 2007; Payet and Suttle, 2008), we hypothesized that bacterial

community composition (BCC) acts as an important environmental selection parameter in shaping VCC. To test this hypothesis, we used high-throughput sequencing to examine the composition and diversity of T4-like viruses and their potential bacterial hosts in the northern SCS.

Results

Samples were collected in the current study from the nSCS along a gradient extending from an estuarine environment to the continental shelf and on to the open sea (Fig. S1). Overall, three clusters of water samples were recognized based on temperature and salinity (Fig. S2, Table S1). Cluster 1 (sample A9-5 m) could be identified as the estuarine water mass (EWM) with relatively high temperature (28.1°C) and low salinity (31.19), whereas cluster 3 (comprising samples J2-50 m, I1-75 m, K3-75 m, K4-75 m and SEATS-75 m) could be identified as the sub-surface open sea water mass (ssOWM) and had the lowest water temperature (21.9–23.8°C) and highest salinity (34.22–34.36) among all samples. Cluster 2 samples showed a wider range of salinities (33.04–33.70) and temperatures (28.6–30.0°C) and could be separated into two sub-clusters based on trophic features. The first sub-cluster contained samples with a salinity <33.12 (samples J1-5 m, J2-5 m and J3-5 m), representative of the continental water mass (CWM; transitional region between eutrophic and oligotrophic zones), while the second sub-cluster consisted of samples from the surface open sea water mass (sOWM; oligotrophic region; samples J4-5 m, J4-25 m, I1-5 m, D-5 m, K2-5 m, K3-5 m, K4-5 m, and SEATS-5 m), which showed higher salinity (33.35–33.70).

The abundance of heterotrophic bacteria was highest in the eutrophic EWM sample, followed by the continental and open sea environment samples (Table S1). Among the autotrophic cyanobacteria, the highest concentration of *Prochlorococcus* was observed in the sOWM samples, followed by the ssOWM and coastal CWM and EWM samples. *Synechococcus* were most abundant in samples from the continental and open sea environments but were less abundant than *Prochlorococcus* in the sOWM samples (Table S1). In addition, the abundance of picoeukaryotes was higher in ssOWM and EWM samples (Table S1).

Diversity among viral and bacterial communities

In total, 122 969 viral sequences and 750 658 bacterial sequences were generated after quality control in this study. We recovered 824 viral operational taxonomic units (vOTUs) and 1111 bacterial OTUs (bOTUs). Good's coverage values from T4-like viral and bacterial sequences varied from 96.3% to 99.0% and 99.7% to 99.8% (Table S2) respectively. The rarefaction curve generated from the sequence data approached saturation

(Fig. S3). Those indicated that the diversity of T4-like viruses and bacteria in the library was most covered. A clear inverse correlation was displayed between species richness (number of OTUs) and frequency of occurrences for both vOTUs and bOTUs (Fig. S4). The majority of the vOTUs and bOTUs appeared in only a few samples, with few OTUs occurring widely across samples. Moreover, there was a positive relationship between the average contribution of an OTU to the total community and its occurrence (Fig. S4). For example, vOTU and bOTU with the highest average contribution (12.6% and 6.28% of the community respectively) were observed in all samples. The rank-abundance plot of the average contribution per OTU had a clear inverted exponential curve (Fig. S4B and D), showing that the viral and bacterial community structures were uneven.

Phylogeny of T4-like viruses and bacteria

Basic local alignment search tool analysis of nucleotide sequences indicated that the majority of the dominant vOTUs (relative abundance >1%, containing 93 OTUs and 37 900 sequences) in the nSCS samples were highly similar to those uncultured T4-like viruses. The sequences were affiliated with previously defined Marine Group, Estuary Group, Lake Group and Paddy Group T4-like viruses (Fig. 1). The most abundant myovirus genotypes were Marine Groups, accounting for 47.31% of richness and 49.06% of abundance (Fig. 2). The Marine Group can be subdivided into Marine Groups I–V, which are usually recovered from environmental sequences, and the Exo T-evens, which infect cyanobacteria (Filée *et al.*, 2005). Exo T-evens group was the most diverse and abundant subgroup in the current study (25.30% of abundance, 21.50% of richness), and could be divided into two clusters. The Exo T-evens-I sequences were mainly similar to *g23* sequences from *Prochlorococcus* phages (e.g. P-SSM4 and P-SSM2) and *Synechococcus* phages (e.g. S-SSM5 and S-SM1), whereas the Exo T-evens-II sequences were most similar to environmental viral *g23* sequences mostly recovered from the Pacific Ocean. Sequences belonging to Marine Groups I (8 OTUs, 12.96% of sequences) and IV (10 OTUs, 8.27% of sequences) were recovered from all four water masses in the nSCS. Spatially, Marine Group IV was most abundant in samples from the CWM and sOWM, while Marine group I was most prevalent in the sOWM and ssOWM samples (Fig. 2). In addition, Marine Groups II and III were mainly present in the CWM samples, while Marine Group V was mostly recovered from the EWM (Fig. 2). In the nSCS, 0.98% of *g23* sequence (2.15% of richness) fell within the Estuary Groups, while 0.19% of sequence (2.15% of richness) were affiliated with the Paddy and Lake Groups (Fig. 2). None of the

sequences were grouped with phages belonging to the T-even and Pseudo T-even groups, whose hosts primarily inhabit animal guts and are considered essentially absent from the open sea (Fuhrman *et al.*, 1993; Bano *et al.*, 2004), despite being observed in the marine environment in previous studies (Filée *et al.*, 2005).

In total, 17 bacterial phyla were identified across all samples. *Proteobacteria* was the most abundant phylum across all stations with the exception of sample J2-5m. *Bacteroidetes* and *Cyanobacteria* were highly abundant in samples from the open sea (Fig. S5A). At the class level, α - and γ -*Proteobacteria* and *Bacteroidia* were highly abundant in all samples. β -*Proteobacteria* were more abundant in samples from the estuarine environment than in any other samples. Samples collected from the open sea had a high abundance of *Oxyphotobacteria*, *Rhodothermia* and *Acidimicrobiia* (Fig. S5B), while *Phycisphaerae*, *Actinobacteria* and *Oxyphotobacteria* were enriched in CWM and EWM samples.

Geographic patterns

There was an obvious separation of viral communities among the samples, with community distribution corresponding to the water mass clusters (Fig. 3, Fig. S2). This finding was supported by the analysis of similarity (ANOSIM) global test (Global $R = 0.909$, $P = 0.001$). In addition, a vertical stratification between surface (sOWM) and sub-surface (ssOWM) communities was observed (Fig. 3). ANOSIM analysis showed that sample pairings CWM and sOWM, CWM and ssOWM and sOWM and ssOWM exhibited significant differences in viral community structure ($P < 0.05$) (Table 1). Furthermore, our results revealed a clear distance-decay pattern between viral community compositional similarity and geographic distance (decreased community similarity with increasing geographic distance) (Fig. 4A). This pattern was observed for the most abundant T4-like viral groups, i.e. Exo T-evens, Marine Group I and Marine Group IV (Fig. 4B–D). Exo T-evens and Marine Group I showed a significant relationship between community similarity and geographic distance, with the Exo T-evens group exhibiting the highest correlation, indicative of the strongest distance-decay pattern. Moreover, the Exo T-evens group made the greatest contribution to the dissimilarity between the CWM and sOWM, CWM and ssOWM and sOWM and ssOWM communities. Marine Groups I and IV were mainly responsible for the dissimilarity between the sOWM and ssOWM communities, as well as between the CWM and sOWM communities, while Marine Group IV made a smaller contribution to the dissimilarity between the EWM and CWM communities (Fig. S6).

Bacterial community distribution was also corresponding to the water mass clusters (ANOSIM global test: global

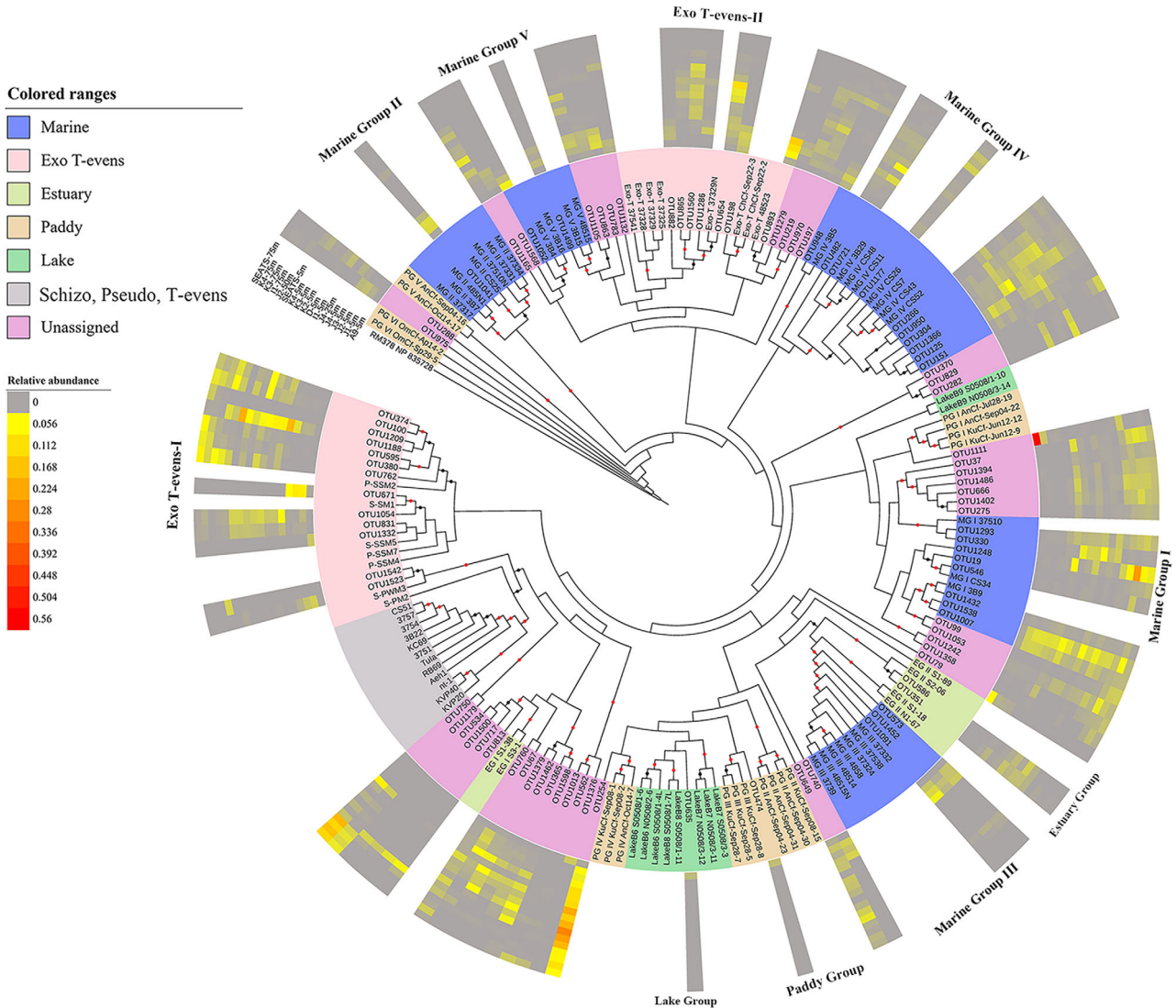


Fig. 1. Maximum-likelihood phylogenetic analysis based on amino acid sequence of the obtained *g23* OTUs. OTUs with a relative abundance of >1% in each sample (a total of 93 OTUs) from the nSCS were selected. Black and red dots show internal nodes with >50% and >80% bootstrap (1000-fold replicated) support respectively. Different coloured ranges indicate *g23* sequences from different groups or origins. The outer coloured rings indicate the relative abundance of each OTU in each sample. Light grey indicates sequences not detected in the samples (relative abundance of 0%). [Color figure can be viewed at wileyonlinelibrary.com]

$R = 0.498$, $P = 0.001$) (Fig. 3, Fig. S2). The vertical stratification of the BCC was less obvious than that of the VCC. In addition, the composition of the CWM communities differed significantly with the sOWM communities (ANOSIM pairwise tests: $R = 0.516$, $P < 0.05$) and ssOWM communities ($R = 0.538$, $P < 0.05$) (Table 1).

Environmental selection and dispersal processes associated with the patterns of viral community

Our results showed a significant correlation between the VCC and BCC (Spearman's correlation, $P < 0.01$)

(Table 2). Abiotic and abundance variables and geographic distance showed a significantly negative relationship with the T4-like viral community and with the Exo T-evens and Marine Group I sub-groups ($P < 0.01$) (Fig. 4; Table 2). Abiotic and abundance variables and the BCC had similar degrees of correlation (almost equivalent r -values) with the VCC, while geographic distance showed a relatively weak correlation (Table 2). Moreover, abiotic and abundance variables showed a significant correlation with the VCC of sOWM ($\rho = -0.481$, $P < 0.05$), while geographic distance was significantly correlated with the VCC of ssOWM ($\rho = -0.915$, $P < 0.001$) (Table 2).

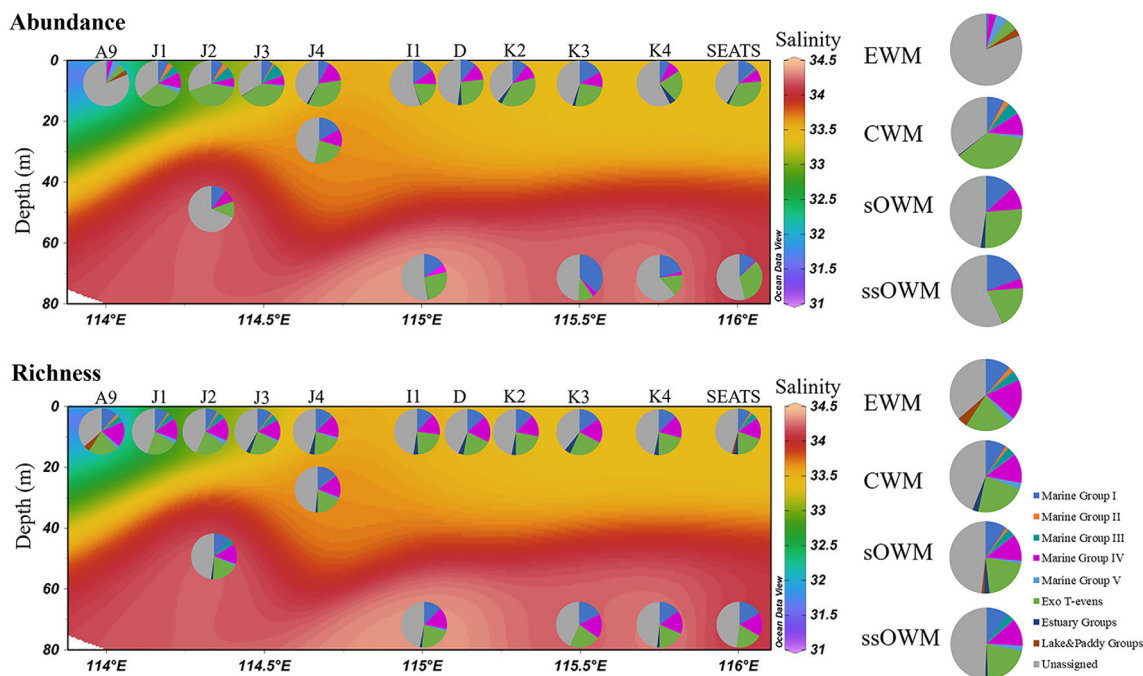


Fig. 2. Comparison of the relative distribution of each group of T4-like viruses obtained from the transect of the nSCS. [Color figure can be viewed at wileyonlinelibrary.com]

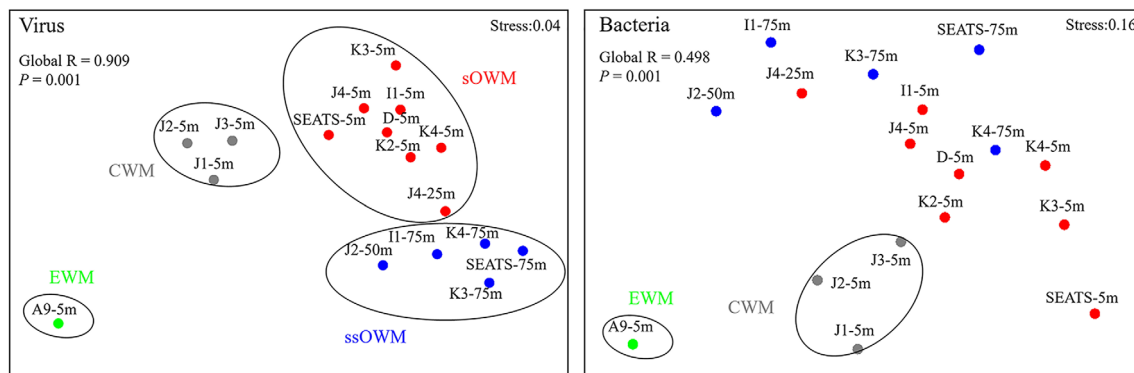


Fig. 3. NMDS of T4-like virus and bacteria in the nSCS based on Bray–Curtis similarity. Different oceanic water masses were marked with green (EWM), grey (CWM), red (sOWM) and blue (ssOWM) spots. [Color figure can be viewed at wileyonlinelibrary.com]

Table 1. Community comparison of T4-like virus and bacteria based on ANOSIM analysis with 999 permutations.

		EWM	CWM	sOWM	ssOWM
Virus	EWM	-	-	-	-
	CWM	1	-	-	-
	sOWM	1	0.987**	-	-
	ssOWM	1	1*	0.775**	-
Bacteria	EWM	-	-	-	-
	CWM	1	-	-	-
	sOWM	0.938	0.516*	-	-
	ssOWM	1	0.538*	0.287*	-

* $P < 0.05$; ** $P < 0.01$ and *** $P < 0.001$.

To further explore the mechanisms shaping the viral biogeographic patterns, the relative influence of selective factors and spatial variables in forming viral community structure were analysed. Three abiotic (temperature, salinity and $\text{NO}_3^- + \text{NO}_2^-$) and one abundance (heterotrophic bacterial abundance) environmental factors and two spatial factors (PCNM 1 and 2) were shown to make a significant contribution to VCC by canonical correlation analysis (CCA) (Fig. 5). Mantel tests further indicated significant relationships ($P < 0.05$) between the above four environmental variables and the VCC (Table S3). Pure abiotic and

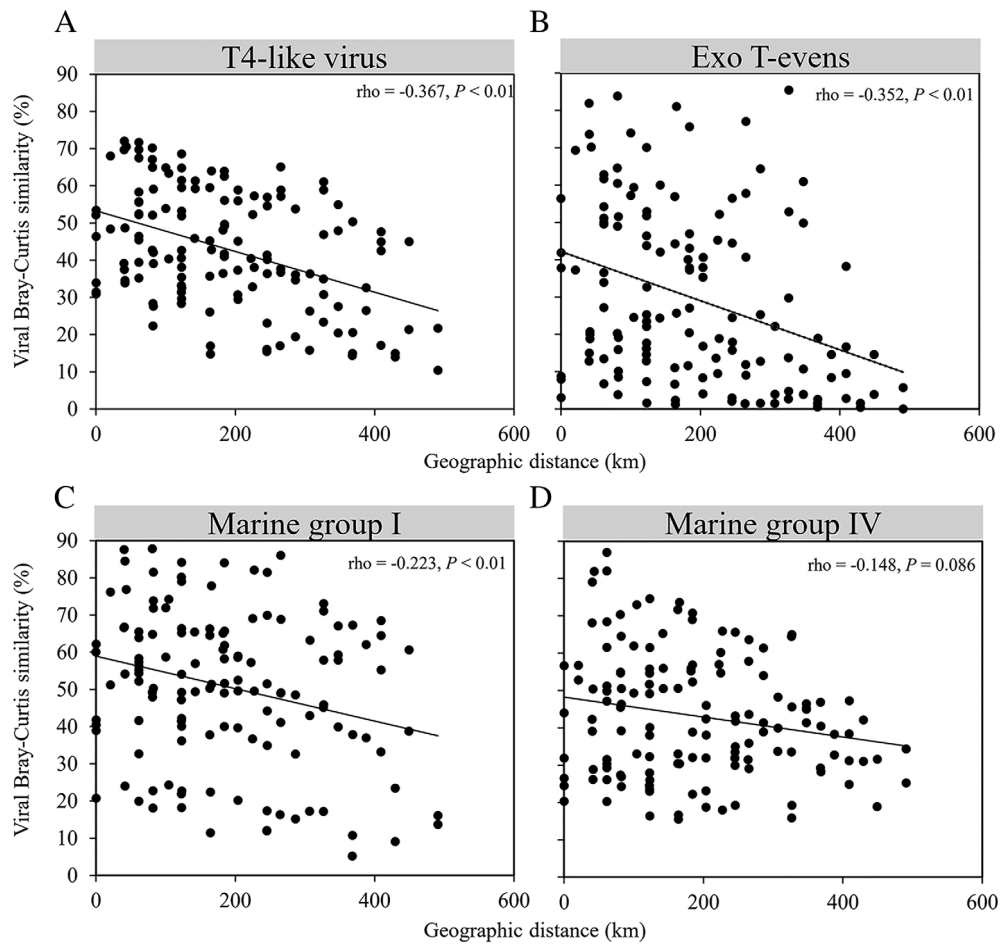


Fig. 4. Distance-decay patterns of T4-like viral community (A) and major viral groups (B–D). Spearman's correlation (ρ) and the significance (P) values are indicated.

Table 2. Spearman's correlation analysis of relationships between VCC, BCC, geographic distance and Euclidean distance of all the abiotic and abundance variables.

	Euclidean distance		Bacterial community		Geographic distance	
	ρ	P	ρ	P	ρ	P
T4-like viral community	-0.472	<0.01	0.625	<0.01	-0.367	<0.01
CWM	-0.500	0.667	0.500	0.667	0.500	0.667
sOWM	-0.481	<0.05	0.297	0.190	-0.264	0.174
ssOWM	-0.212	0.556	0.818	<0.01	-0.915	<0.01
Exo T-evens	-0.447	<0.01	0.526	<0.01	-0.352	<0.01
Marine Group I	-0.428	<0.01	0.426	<0.01	-0.223	<0.01
Marine Group IV	-0.422	<0.01	0.332	<0.01	-0.148	0.086

abundance factors ($(E|(S\&B))$, 9%) explained the same fraction of the variations as pure BCC ($B|(E\&S)$, 9.0%), and appeared to play a greater role than pure spatial factors ($(S|(E\&B))$, 2.5%) (Fig. 6A and B). Shared environmental (abiotic and abundance and BCC) and dispersal (spatial parameters) processes ($S\cap B\cap E$) explained 2.5% of the variation in the VCC. In addition,

shared abiotic and abundance parameters together with the BCC ($(E\cap B)|S$) explained a large proportion of the variation (35.1%) in viral communities. Partial Mantel tests showed that changes in the VCC were significantly correlated with the BCC, abiotic and abundance and spatial variables in the environmental gradient from the estuary to the open sea (Fig. 6C).

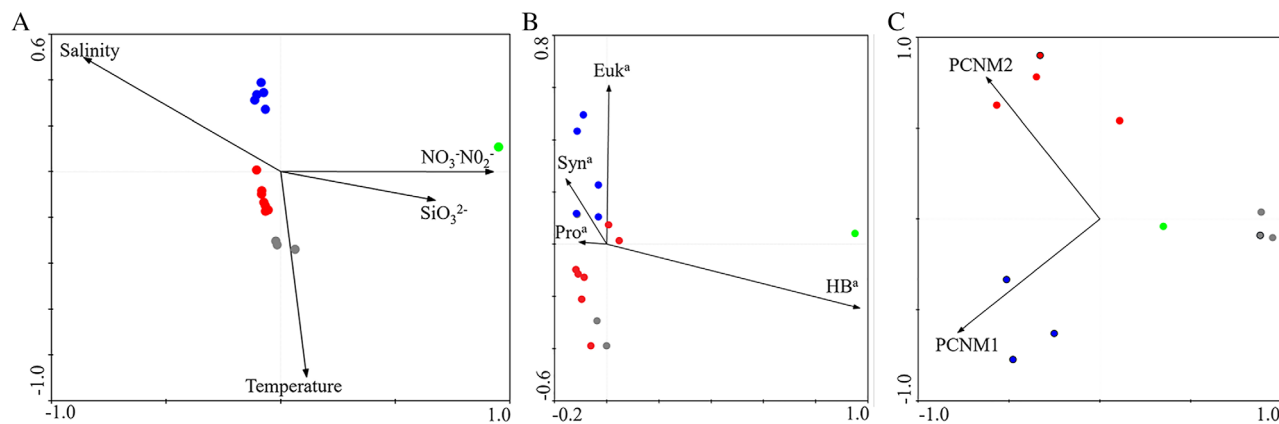


Fig. 5. CCA was conducted to show abiotic (A) and abundance (B) and spatial (C) variables in governing the assembly of viral communities. HB^a, Heterotrophic bacterial abundance; Pro^a, *Prochlorococcus* abundance; Syn^a, *Synechococcus* abundance; Euk^a, picoeukaryotes abundance. [Color figure can be viewed at wileyonlinelibrary.com]

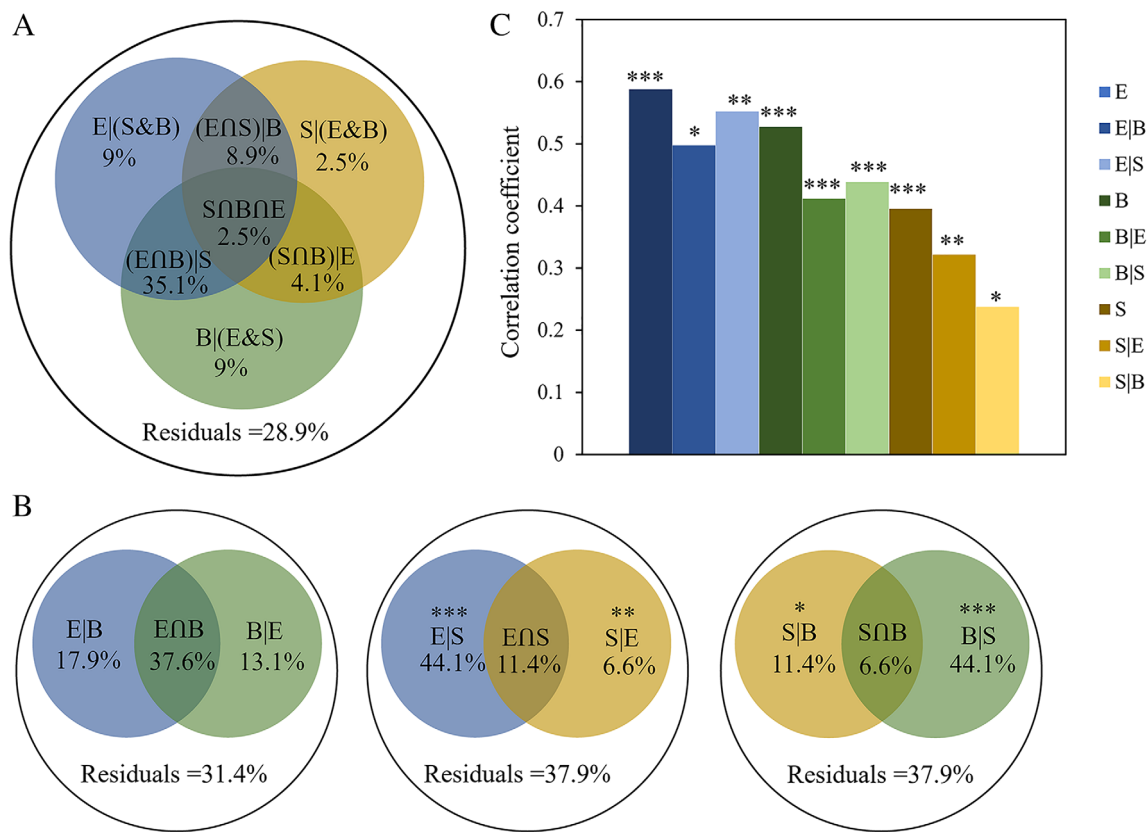


Fig. 6. Variation in the viral community explained by abiotic and abundance (E), bacterial community (B) and spatial variables (S). (A) Variation partitioning analysis of viral community composition between bacterial community composition, abiotic and abundance and spatial variables. (B) Variation partitioning analysis of viral community composition between any two factors among bacterial community composition, abiotic and abundance and spatial variables. The variation explained by pure bacterial community composition, abiotic and abundance and spatial factors corresponds to the viral community without the effect of others by ANOVA permutation tests. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001. The out circles represent 100% of the variation. (C) Mantel and partial Mantel tests to identify the correlations between the viral community structure and bacterial community composition and abiotic and abundance and spatial variables using Pearson's coefficient. '|' indicates partial mantel test. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001. [Color figure can be viewed at wileyonlinelibrary.com]

Virus–bacterium interaction

Co-occurrence network based on Spearman's correlation coefficient analysis was constructed to deconvolute relationships between the predominant vOTUs and bOTUs (Fig. 7). The average path distance was 3.062, while the average clustering coefficient was 0.396. Most of the links in the network were virus-virus associations (77.73% of 494 links), with 14.78% of the links identified as virus–bacterium. Notably, positive connections dominated the interactions between vOTUs. All of the links among the Exo T-evens (22 links) and Marine Group I (3 links) and most of the links within Marine Group IV (8/9 links) were positive connections. There were no inter- and intra-links among the Estuary, Paddy and Lake Groups.

A total of 73 links were identified between virus and bacterium, including 50 positive correlations and

23 negative correlations. Most of the linked bOTUs were identified as α -Proteobacteria. In addition, there were 10 positive links between cyanobacteria and virus. A vOTU was frequently correlated to multiple bOTUs, which was often correlated with two or more vOTUs, resulting in one large interconnected cluster. There also had the correlation between one bOTU and several vOTUs. For example, one bOTU that was affiliated with the abundant family *Flavobacteriaceae* was both positively and negatively correlated with multiple vOTUs.

Discussion

Although the spatial distribution of viruses has been investigated in some marine environments, the mechanisms underlying the development of the complicated biogeographic patterns of viral community structure are still unclear (Chow and Suttle, 2015; Le Moine Bauer

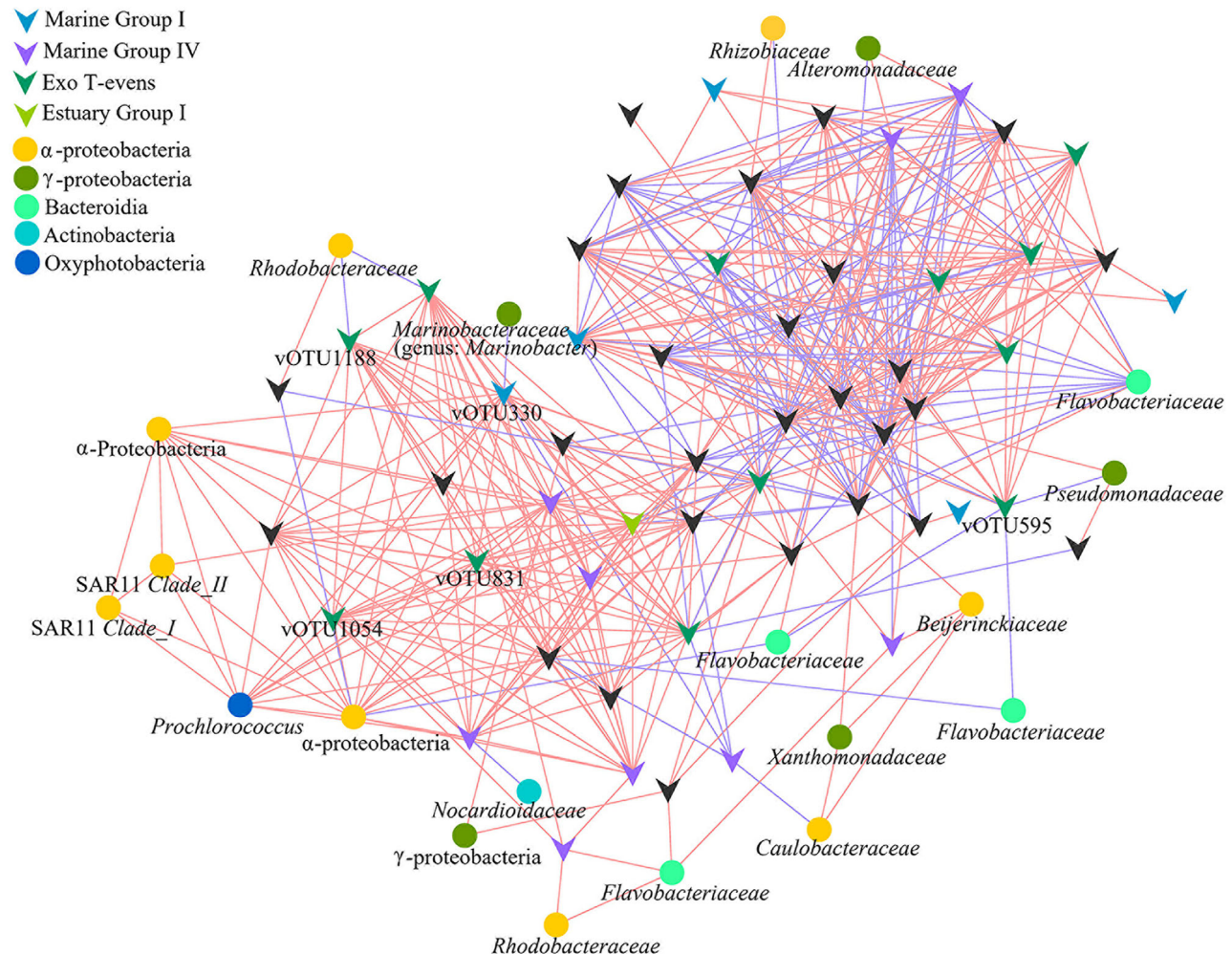


Fig. 7. Association network comprised of dominant vOTUs and the top 10 bOTUs per sample. Node degree with equal or less than two was omitted. The positive or negative linkages were determined based on Spearman's correlation analysis between any pairs of nodes. Positive linkages are shown in red, while negative linkages are shown in blue. [Color figure can be viewed at wileyonlinelibrary.com]

et al., 2018). In this present study, we investigated the biogeographic patterns of T4-like myoviruses in an ~500-km long subtropical estuary-sea environment and determined the relative importance of local environmental and dispersal processes on viral community assembly.

Contrasting biogeography of the T4-like myoviruses in the nSCS

We found a clear biogeographic distribution of T4-like viral groups along the transect from the estuary to open sea. Both Marine Groups I and IV showed a wide geographical distribution in the nSCS, supporting previous studies showing that these groups are abundant in various marine environments, such as the Arctic glaciers, the northeastern Gulf of Mexico and the Pacific Ocean (Filée et al., 2005; Bellas and Anesio, 2013). Marine Group I may prefer colder water with a higher salt content, as the sequences were more abundant in the sOWM and ssOWM samples than in other samples. Marine Group III was obtained in water off the coast of British Columbia and Jericho Pier in previous study (Filée et al., 2005). However, our survey showed that Marine Group III was most abundant in the transitional continental shelf region than in coastal areas in the nSCS. The distribution of Marine Group II was similar to that of Marine Group III, suggesting that both sub-groups or their hosts may be adapted to coastal shelf environments rather than the relatively high-salinity oligotrophic open ocean. Marine Group V was mainly found in the estuary water mass samples and was rare in sOWM and ssOWM samples, suggesting that Marine Group V is well adapted to the eutrophic environments (Filée et al., 2005; Liu et al., 2017). Although the majority of the OTUs belonged to the Marine Groups, several *g23* OTUs from the nSCS clustered with Estuary, Paddy and Lake Groups (Wang et al., 2009; Butina et al., 2010; Liu et al., 2017). This is consistent with the findings of Sano et al. (2004), who showed that viruses, in particular phages, can move between different biomes (e.g. soil and seawater). The broad host range probably gives them a better chance of survival when moving between different environments (Jensen et al., 1998; Sano et al., 2004; Breitbart and Rohwer, 2005).

The T4-like viral biogeography in the nSCS showed a clear distance-decay pattern (Fig. 4). The higher decay rates with increasing geographic distance suggest there may be a higher dispersal limitation for viruses in the nSCS. The Exo T-evens group was the most abundant group across all samples and showed the highest decay rate (Fig. 4), which implies the highest level of variation occurs within this group of the viral community. Despite the increased likelihood of dispersal of the highly abundant taxa, dispersal ability and establishment may also

depend on the species survival strategy, the limited adaptive ability in the environment and host interactions (Östman et al., 2010). In the nSCS, the community composition of cyanobacteria, the major hosts of Exo T-evens phages, showed clear spatial variation, with *Synechococcus* most common in coastal sites but *Prochlorococcus* more abundant in the open sea (Huang et al., 2012). This was supported by Spearman's analysis (Table 2) and the partial Mantel test (Table S4). It is worth pointing out that the strong correlation between spatial distance and the Exo T-evens group, revealed by Spearman's analysis and Mantel test (Table 2, S4), may be attributed to the impact of spatial distance on bacterial community structure. After controlling for the bacterial community using the partial Mantel test, the significant correlation between spatial distance and the Exo T-evens group decreased markedly (*r*-value from 0.3687 to 0.1945) (Table S4). Therefore, we propose that the high distance-decay rate observed for Exo T-evens phages in our study results from the adaptive abilities of their hosts to the environment and a close interaction between phages and hosts.

Both environmental selection and dispersal processes influenced viral communities

We found that T4-like viral community assembly was significantly associated with local environmental factors and spatial variables (Fig. 6; Table 2). On the one hand, the environmental variables (temperature, salinity, $\text{NO}_3^- + \text{NO}_2^-$, heterotrophic bacterial abundance and bacterial community structure) strongly explained the cluster patterns in the viral communities (Fig. 5; Table 2, S3). Previous analysis has shown that temperature and salinity strongly affect VCC because they play an important role in the growth and survival of viruses (Guixa-Boixareu et al., 1996; Jiang et al., 2004; Mojica and Brussaard, 2014). Temperature can regulate infection dynamics and the sensitivity of viral lipid membranes or capsid protein, while the adsorption of viruses to host cells can be affected by salt concentration (Kukkaro and Bamford, 2009; Mojica and Brussaard, 2014). Surveys from the *Tara* Oceans expedition also demonstrated that viral community structure varied consistently with temperature and salinity on a global scale (Brum et al., 2013). Moreover, trophic status might be of greater importance to viral geography because viral population size and activity are highly dependent on host population size and activity (Middelboe, 2000; Suttle, 2000; Winter et al., 2004). This was supported by our findings showing that relatively high salinity and oligotrophic OWM and CWM samples grouped together but away from the low-salinity and eutrophic EWM samples (Fig. 3). In addition, BCC has been suggested as to be the most important

determining factor for the viral community as a result of specific virus–host interactions such as infection (Winter *et al.*, 2004, 2005; Parada *et al.*, 2008; Chow *et al.*, 2014) (see below for detailed discussion).

On the other hand, spatial factors (dispersal processes) shape the viral community. In oceans, one of the dominant dispersal mechanisms for viruses is the movement of water masses (Frederickson *et al.*, 2003; Winter *et al.*, 2013; Chow and Suttle, 2015; Le Moine Bauer *et al.*, 2018). Water mass transport usually contributes to a negative relationship between microbial compositional similarity and geographic distance (the distance-decay relationship) (Green and Bohannan, 2006; Hanson *et al.*, 2012). In the nSCS, spatial variables appear to contribute significantly, although to a lesser extent than environmental selection, to viral community structure. Dispersal includes establishment, meaning that at least some reproduction in the new location rather than simply just being present (Hanson *et al.*, 2012). Given the small size of viruses, there are few limitations to their global dispersal, yet host availability and environmental factors may increase limitations, either directly or indirectly, that restrict geographic range (Chow and Suttle, 2015).

The partial Mantel test corroborated this result by showing that environmental selection had a great effect than dispersal processes. This result was consistent with studies conducted in the Labrador Sea (Winter *et al.*, 2013). However, viral communities in the Arctic Archipelago (~1000-km) showed a significant response to spatial factors rather than environmental variables (Winter *et al.*, 2013). These differences may result from differing degrees of environmental heterogeneity and/or geographic scale (Chow and Suttle, 2015). Although samples collected in the current study covered a ~500-km transect, the degree of environmental heterogeneity in the nSCS was higher than that in the Arctic Archipelago, meaning that large environmental gradients can increase the influence of environmental selection on viral community assembly. This observation agrees with research carried out on marine bacteria, suggesting that the relative influence of dispersal processes might shift in response to the degree of environmental heterogeneity in community assembly (Logares *et al.*, 2013; Mo *et al.*, 2018).

Ecological role of the host in viral community assembly

In general, co-occurrence correlations between microbial taxa can be explained as competition or cooperation for nutrients, material and space (Deng *et al.*, 2012). In our study, co-occurrence networks suggested complex interactions within and between viral and bacterial communities. Virus propagation is host-dependent and infection is usually species- or even strain-specific. Negative

correlations between virus and bacteria indicate the presence of viral infection and lysis (Chow *et al.*, 2014). Therefore, the host range of a vOTU may be suggested by negative correlations with bOTUs, with a greater number of negative connections indicative of a broad host range (such as the negative correlation between vOTU595 and three bOTUs). In the current study, bOTUs that were negatively correlated with T4-like viruses were affiliated with α - and β -*Proteobacteria*, *Cyanobacteria*, *Bacteroidia* and *Actinobacteria*, indicating that bacteria belonging to these taxa are the likely hosts of T4-like viruses in the nSCS. Positive correlations suggest co-occurrence as a result of similar preferred conditions or a commensal or a mutualistic relationship between organisms cooperating within the same niche (Barberan *et al.*, 2012; Faust and Raes, 2012; Chow *et al.*, 2014). For example, the positive correlation between vOTU330 and *Prochlorococcus* might be a result of the viral lysis of bOTU9334 (negative correlation with vOTU330), since bOTU9334 showed 100% nucleotide sequence identity to *Marinobacter* sp. HOT4B5, which inhibits the growth of *Prochlorococcus* MIT9313 (Sher *et al.*, 2011). An alternative explanation is that host lysis by viruses releases nutrients and stimulates the growth of uninfected hosts, including cyanobacteria and heterotrophic bacteria (Sher *et al.*, 2011; Shelford and Suttle, 2018).

The positive associations (57.89%) within the viral community may be the result of mutualism between viruses in long-term co-evolutionary processes (Chow *et al.*, 2014). Almost all of the intra-links for the Exo T-evens, Marine Group I and Marine Group IV were positive, showing that T4-like viruses affiliated with these groups may have a co-habitat or co-infection (Needham *et al.*, 2013; Zhang *et al.*, 2017). The vOTU1054 and vOTU1188 of Exo T-evens group, which were affiliated with *Synechococcus* phages S-CAM8 and S-SM2 respectively, were positively linked. Sullivan *et al.* (2003) demonstrated that phage S-SM2, which was isolated from *Synechococcus* WH7803, has a broad host range and can infect the original host of phage S-CAM8 (*Synechococcus* WH8017) (Sullivan *et al.*, 2003). *Synechococcus* strains WH7803 and WH8017 are closely related and have similar growth conditions and ecological niches (Sullivan *et al.*, 2006). These likely explain the positive link between vOTU1054 and vOTU1188 in our study. In comparison, negative correlations between viruses might indicate a competition or non-overlapping niches between viruses or their hosts (Barberan *et al.*, 2012; Faust and Raes, 2012; Chow *et al.*, 2014). For example, vOTU219 and vOTU380 were affiliated with cyanophages S-TIM4 and P-SSM5 respectively, whose original hosts were *Synechococcus* WH8102 and *Prochlorococcus* NATL2A respectively. The negative correlation between vOTU219 and vOTU380 in our study

may correspond with the different niche preference of *Synechococcus* WH8102 and *Prochlorococcus* NATL2A.

Conclusions

The present study demonstrated the biogeographic and significant distance-decay patterns of the T4-like viral community in the nSCS, which is a good representative of an estuary–sea environment. We determined that T4-like viruses belonging to Marine Groups I and IV had a wide geographical distribution, those belonging to Marine Groups II and III were mainly present in lower-salinity continental environments, and that Marine Group V viruses preferred eutrophic coastal estuarine environments. Compared with spatial distance, environmental selection (e.g. temperature, salinity, nitrogen nutrient, bacterial abundance and community composition) had a greater impact on viral community structure. Among environmental selection factors, the bacterial community structure alone exerted almost the same degree of selective pressure as other abiotic and biotic variables in shaping T4-like viral community assembly. In addition, network analysis suggested that BCC may impact on viral community assembly and biogeography through diverse ecological interactions, such as lysis, co-infection and mutualism. Our data provide important ecological insights into marine viral biogeography, including community assembly mechanisms and interactions with host communities.

Experimental procedures

Viral and bacterial sample collection and DNA extraction

Eleven stations were visited during the cruise of the nSCS in September 2014. Water samples were retrieved from the surface (5 and 25 m) and subsurface (50 and 75 m) layers (Fig. S1). After collection, 5-L water samples were pre-filtered through a 20- μ m sieve and then subjected to sequential tangential-flow filtration using a 0.22- μ m-pore-size filter (Millipore Corp.) to obtain bacterial concentrates (50 ml) and a 30-kDa cartridge (Millipore Corp.) to concentrate viruses (50 ml). Total DNA was extracted from the viral concentrate using the phenol/chloroform/isoamylol method (Liu *et al.*, 2017). Total DNA was extracted from the bacterial concentrate as described previously (Kan *et al.*, 2008).

Environmental parameters

Temperature, salinity and depth were measured *in situ* using conductivity-temperature-depth oceanic profilers (SBE9/11 plus; Sea-Bird, USA). Temperature and salinity

were used to cluster water masses using the *pamk* function of package *fpc* in R (Fig. S2) (R Development Core Team, 2015 (Li *et al.*, 2018a)). $\text{NO}_3^- + \text{NO}_2^-$, SiO_3^{2-} and PO_4^{3-} , were obtained by filtration using 0.45- μ m cellulose acetate filters and measured using a colorimetric method. To estimate the abundance of microbes, replicate samples (1.98 ml) were collected and fixed using glutaraldehyde (0.5% (vol./vol.) final concentration) at 4°C for 15 min in the dark, before being frozen in liquid nitrogen. The abundance of autotrophic picoeukaryotes, *Synechococcus* and *Prochlorococcus* was determined by direct counting without staining using flow cytometer (Accuri C6; BD Biosciences) (Jiao *et al.*, 2002). Heterotrophic bacterial abundance was also determined by flow cytometer (Accuri C6) after staining with SYBR Green I (Table S1) (Marie *et al.*, 1999). All data analyses were performed using FlowJo vX.0.7 software (Tree Star, USA).

Polymerase chain reactions, sequencing and phylogenetic analysis

A fragment of the major capsid protein-encoding gene *g23* was amplified from T4-type phages using degenerate primers MZIA1 bis and MZIA6, as described by Filée *et al.* (2005). The amplicons were sequenced using the 454 GS FLX platform. The sequences were quality screened using MOTHUR (v.1.41.1), and chimeric sequences were discarded (Schloss *et al.*, 2009; Edgar, 2010). OTUs were identified with a 97% sequence similarity. Singletons (OTUs that occurred only once) identified by *filter_otus_from_otu_table.py* command within QIIME were removed and samples were subsampled to 2709 (minimum sample size) from the OTU table for downstream analyses. The most prevalent *g23* OTUs, with a relative abundance >1% per sample (a total of 93 OTUs), were translated into amino acid sequences and aligned using MEGA7 (Kumar *et al.*, 2016). A phylogenetic tree was constructed using the maximum likelihood method in RAXML, with 1000 bootstrap replicates (Stamatakis, 2014). The tree was viewed and graphically edited using iTOL (Letunic and Bork, 2016). Viral *g23* sequences have been deposited in the National Centre for Biotechnology Information Sequence Read Archive (NCBI SRA) database under the accession numbers SRR11786657–SRR11786673 (BioProject accession number PRJNA632563; BioSample accession numbers SAMN14912226–SAMN14912242).

The conserved V3–V4 region of the bacterial 16S rRNA gene was amplified using primers 338F and 806R (Mori *et al.*, 2014; Li *et al.*, 2018b). Resulting amplicons were sequenced using the Illumina MiSeq PE300 platform. Raw reads were trimmed using Trimmomatic and merged using FLASH (Magoc and Salzberg, 2011; Bolger *et al.*, 2014). Thereafter, *usearch61* was used to

identify and remove chimeric sequences (Edgar, 2010). The remaining sequences were clustered into OTUs based on 97% sequence similarity. bOTUs with a total community abundance of >0.001% were then selected for further analysis. The number of sequences from each sample was subsampled to the same amount (30 000). Classification was carried out using MOTHUR (v.1.41.1) with SILVA (v132) reference sequences and taxonomic outline. Bacterial 16S rRNA gene sequences have been submitted to the NCBI SRA database under the accession numbers SRR11794483–SRR11794499 (BioProject accession number PRJNA632681, BioSample accession numbers SAMN14917627–SAMN14917643). Details on the primers and sequence processing of viral and bacterial sequences are supplied in the Supporting Information.

Community diversity and structure

Rarefaction curves based on the identified vOTUs and bOTUs were estimated by PAST (v.3.18). The rarefied sequences were then used for Good's coverage and alpha diversity indices in QIIME. Non-metric multi-dimensional scaling (NMDS) analysis based on the Bray–Curtis similarity of VCC and BCC was performed to disentangle the difference between the samples. ANOSIM was used to test significant difference in bacterial and viral communities. Similarity percentage (SIMPER) analysis was used to conclude the contribution of major vOTUs to the observed dissimilarity between different oceanic water masses. NMDS, ANOSIM and SIMPER analyses were performed in PRIMER 6.0 (Clarke and Gorley, 2006).

Relationships between VCC and deterministic and dispersal processes

The Bray–Curtis similarity results for VCC and BCC as well as the Euclidean distance matrices for the abiotic and abundance parameters were transformed using algorithms of $\log_{(x+1)}$ in PRIMER 6.0. Geographic distance was calculated based on the coordinates of the sampling stations using a spheroidal model of Earth (Winter *et al.*, 2013). The relationships among geographic distance, Euclidean distance and Bray–Curtis similarity were analysed using Spearman's rank correlations. CCA was used to connect the structure of viral communities with abiotic, abundance and spatial variables. The longest gradient lengths in the detrended correspondence analysis were >4, indicating that CCA is suitable for examining VCC. Prior to CCA analysis, the variables with a high variance inflation factor (>10) were eliminated to avoid collinearity among factors.

Variation partitioning analysis (VPA) was used to assess the relative importance of environmental and spatial parameters in forming viral community, which was widely used in ecological research to determine the relative importance of environmental selection versus dispersal processes for community structure (Legendre and Legendre, 2012; Winter *et al.*, 2013; Mo *et al.*, 2018). Bray–Curtis similarity is one of the most commonly used similarity quantification methods when it comes to ecological abundance data collected at different sampling locations (Legendre and Legendre, 2012; Ricotta and Podani, 2017). Principal coordinates analysis explores similarities or dissimilarities, such as Bray–Curtis similarity, and takes species identity into account, which is a better method of analysing bacterial community structure data for VPA (Mohammadi and Prasanna, 2003; Legendre and Legendre, 2012; Hamdan *et al.*, 2013; Nagaraj *et al.*, 2017; Hu *et al.*, 2020). To avoid collinearity in the statistical calculation, principal coordinates axes with a cumulative 73.9% variation based on Bray–Curtis similarity were used to summarize the bacterial community structure (relative abundance data) for VPA. Abiotic and abundance variables, which were significantly (Mantel test, $P < 0.05$) correlated with viral community structure, were used. Spatial variables were generated using the principal coordinates of neighbour matrices analysis. The relative importance of all components was explained by pure abiotic and abundance environmental variables (E|S&B) (i.e. the exclusive abiotic and abundance variables excluding spatial variables and BCC), pure spatial variable (S|(E&B)), pure BCC (B|(E&S)) and the combined effects of spatial variables and BCC ((S∩B)|E) (i.e. interaction component of spatial variables and BCC excluding abiotic and abundance variables), spatial and abiotic and abundance variables ((S∩E)|B), BCC and abiotic and abundance variables ((B∩E)|S), and the combined effects of spatial, BCC and abiotic and abundance variables (S∩B∩E) respectively. The Mantel and partial Mantel tests were conducted to verify the results obtained from VPA. All these statistical analyses were performed in R.

Network analysis

Potential interactions between viral and bacterial taxa were determined by modelling the community in network structure. Association networks were constructed using an online Molecular Ecological Network Analyses Pipeline (<http://ieg4.rccc.ou.edu/MENA>) (Deng *et al.*, 2012). Pairwise correlation matrices were generated for the viral and bOTUs. Similarity matrices were calculated based on Spearman's rank correlation. Random matrix theory-based algorithm was used to determine the threshold of network (Deng *et al.*, 2012; Zhang *et al.*, 2017).

Cytoscape 3.2.1 was used to describe the network, showing nodes (vOTUs and bOTUs) linked by lines that denoted positive or negative correlations.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (91951209, 41706154, 41861144018). R.Z. was partially supported by the Fundamental Research Funds for the Central Universities (20720200027, 20720170107) and Science and Technology Program of Guangzhou, China (201904020029). We sincerely thank Jia Sun and Jianning Wang for providing the nutrient and CTD data.

References

- Ackermann, H.-W., and Krisch, H. (1997) A catalogue of T4-type bacteriophages. *Arch Virol* **142**: 2329–2345.
- Angly, F., 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.
- Bano, N., Ruffin, S., Ransom, B., and Hollibaugh, J.T. (2004) Phylogenetic composition of Arctic Ocean archaeal assemblages and comparison with Antarctic assemblages. *Appl Environ Microbiol* **70**: 781–789.
- Barberan, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**: 343–351.
- Baudoux, A.C., Veldhuis, M.J.W., Witte, H., and Brussaard, C.P. (2007) Viruses as mortality agents of Picophytoplankton in the deep chlorophyll maximum layer during IRONAGES III. *Limnol Oceanogr* **52**: 2519–2529.
- Bellas, C.M., and Anesio, A.M. (2013) High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. *Extremophiles* **17**: 861–870.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Breitbart, M. (2012) Marine viruses: truth or dare. *Annu Rev Mar Sci* **4**: 425–448.
- Breitbart, M., and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* **13**: 278–284.
- 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**: 14250–14255.
- Brum, J.R., Schenck, R.O., and Sullivan, M.B. (2013) Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *ISME J* **7**: 1738–1751.
- Butina, T.V., Belykh, O.I., Maksimenko, S.Y., and Belikov, S. I. (2010) Phylogenetic diversity of T4-like bacteriophages in Lake Baikal, East Siberia. *FEMS Microbiol Lett* **309**: 122–129.
- Chen, C.-T.A., Wang, S.-L., Wang, B.-J., and Pai, S.-C. (2001) Nutrient budgets for the South China Sea basin. *Mar Chem* **75**: 281–300.
- Chow, C.E., and Fuhrman, J.A. (2012) Seasonality and monthly dynamics of marine myovirus communities. *Environ Microbiol* **14**: 2171–2183.
- Chow, C.E., and Suttle, C.A. (2015) Biogeography of viruses in the sea. *Annu Rev Virol* **2**: 41–66.
- Chow, C.-E.T., Kim, D.Y., Sachdeva, R., Caron, D.A., and Fuhrman, J.A. (2014) Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. *ISME J* **8**: 816–829.
- Clarke, K.R., and Gorley, R.N. (2006) *PRIMER v6: User Manual/Tutorial*. Plymouth, UK: PRIMER-E, p. 57–141.
- De Corte, D., Sintes, E., Yokokawa, T., Lekunberri, I., and Herndl, G.J. (2016) Large-scale distribution of microbial and viral populations in the South Atlantic Ocean. *Environ Microbiol Rep* **8**: 305–315.
- Deng, Y., Jiang, Y., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012) Molecular ecological network analyses. *Bioinformatics* **13**: 113.
- Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., et al. (2008) Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature* **452**: 340–343.
- Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.
- Faust, K., and Raes, J. (2012) Microbial interactions: from networks to models. *Nat Rev Microbiol* **10**: 538–550.
- Filée, J., Tétart, F., Suttle, C.A., and Krisch, H. (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.
- Frederickson, C., Short, S., and Suttle, C. (2003) The physical environment affects cyanophage communities in British Columbia inlets. *Microb Ecol* **46**: 348–357.
- Fuhrman, J., McCallum, K., and Davis, A. (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific oceans. *Appl Environ Microbiol* **59**: 1294–1302.
- Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541–548.
- Green, J., and Bohannan, B.J.M. (2006) Spatial scaling of microbial biodiversity. *Trends Ecol Evol* **21**: 501–507.
- 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**: 1109–1123.e14.
- Guixa-Boixareu, N., Calderón-Paz, J.I., Haldal, M., Bratbak, G., and Pedrós-Alió, C. (1996) Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. *Aquat Microb Ecol* **11**: 215–227.
- Hamdan, L.J., Coffin, R.B., Sikaroodi, M., Greinert, J., Treude, T., and Gillevet, P.M. (2013) Ocean currents shape the microbiome of Arctic marine sediments. *ISME J* **7**: 685–696.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- He, L., Yin, K., Yuan, X., Li, D., Zhang, D., and Harrison, P. J. (2009) Spatial distribution of viruses, bacteria and chlorophyll in the northern South China Sea. *Aquat Microb Ecol* **54**: 153–162.

- Hu, H., He, J., Yan, H., Hou, D., Zhang, D., Liu, L., and Wang, K. (2020) Seasonality in spatial turnover of Bacterioplankton along an ecological gradient in the East China Sea: biogeographic patterns, processes and drivers. *Microorganisms* **8**:1484.
- Huang, S., Wilhelm, S.W., Harvey, H.R., Taylor, K., Jiao, N., and Chen, F. (2012) Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6**: 285–297.
- Jensen, E.C., Schrader, H.S., Rieland, B., Thompson, T.L., Lee, K.W., Nickerson, K.W., and Kokjohn, T.A. (1998) Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. *Appl Environ Microbiol* **64**: 575–580.
- Jiang, S., Steward, G., Jellison, R., Chu, W., and Choi, S. (2004) Abundance, distribution, and diversity of viruses in alkaline, hypersaline mono Lake, California. *Microb Ecol* **47**: 9–17.
- Jiao, N., Yang, Y., Koshikawa, H., and Watanabe, M. (2002) Influence of hydrographic conditions on picoplankton distribution in the East China Sea. *Aquat Microb Ecol* **30**: 37–48.
- Kan, J., Evans, S.E., Chen, F., and Suzuki, M.T. (2008) Novel estuarine bacterioplankton in rRNA operon libraries from the Chesapeake Bay. *Aquat Microb Ecol* **51**: 55–66.
- Kukkaro, P., and Bamford, D.H. (2009) Virus-host interactions in environments with a wide range of ionic strengths. *Environ Microbiol Rep* **1**: 71–77.
- Kumar, S., Stecher, G., and Tamura, K. (2016). Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* **33**: 1870–1874.
- Le Moine Bauer, S., Stensland, A., Daae, F.L., Sandaa, R. A., Thorseth, I.H., Steen, I.H., and Dahle, H. (2018) Water masses and depth structure prokaryotic and T4-like viral communities around hydrothermal Systems of the Nordic Seas. *Front Microbiol* **9**: 1002.
- Legendre, P., and Legendre, L. (2012) *Numerical Ecology*, 3rd edn. Amsterdam, Netherlands: Elsevier.
- 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.
- Li, W., Wang, M., Pan, H., Burgaud, G., Liang, S., Guo, J., et al. (2018a) Highlighting patterns of fungal diversity and composition shaped by ocean currents using the East China Sea as a model. *Mol Ecol* **27**: 564–576.
- Li, Y., Sun, L.L., Sun, M.L., Su, H.N., Zhang, X.Y., Xie, B.B., et al. (2018b) Vertical and horizontal biogeographic patterns and major factors affecting bacterial communities in the open South China Sea. *Sci Rep* **8**: 8800.
- Liu, L., Cai, L., and Zhang, R. (2017) Co-existence of freshwater and marine T4-like myoviruses in a typical subtropical estuary. *FEMS Microbiol Ecol* **93**: fix119.
- Logares, R., Lindström, E.S., Langenheder, S., Logue, J.B., Paterson, H., Laybourn-Parry, J., et al. (2013) Biogeography of bacterial communities exposed to progressive long-term environmental change. *ISME J* **7**: 937–948.
- Magoc, T., and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.
- Marie, D., Partensky, F., Vaultot, D., and Brussaard, C. (1999) Enumeration of phytoplankton, bacteria, and viruses in marine samples. *Curr Protoc Cytom* **10**: 11.11.1–11.11.15.
- 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.
- Middelboe, M. (2000) Bacterial growth rate and marine virus-host dynamics. *Microb Ecol* **40**: 114–124.
- Mo, Y., Zhang, W., Yang, J., Lin, Y., Yu, Z., and Lin, S. (2018) Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME J* **12**: 2198–2210.
- Mohammadi, S.A., and Prasanna, B.M. (2003) Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci* **43**: 1235–1248.
- Mojica, K.D., and Brussaard, C.P. (2014) Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *FEMS Microbiol Ecol* **89**: 495–515.
- Mori, H., Maruyama, F., Kato, H., Toyoda, A., Dozono, A., Ohtsubo, Y., et al. (2014) Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. *DNA Res* **21**: 217–227.
- Nagaraj, V., Skillman, L., Ho, G., Li, D., and Gofton, A. (2017) Characterisation and comparison of bacterial communities on reverse osmosis membranes of a full-scale desalination plant by bacterial 16S rRNA gene metabarcoding. *NPJ Biofilms Microbiomes* **3**: 13.
- Needham, D.M., Chow, C.-E.T., 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.
- Östman, O., Drakare, S., Kritzberg, E.S., Langenheder, S., Logue, J.B., and Lindström, E.S. (2010) Regional invariance among microbial communities. *Ecol Lett* **13**: 118–127.
- Parada, V., Baudoux, A.C., Sintes, E., Weinbauer, M.G., and Herndl, G.J. (2008) Dynamics and diversity of newly produced viroplankton in the North Sea. *ISME J* **2**: 924–936.
- Payet, J.P., and Suttle, C.A. (2008) Physical and biological correlates of virus dynamics in the southern Beaufort Sea and Amundsen gulf. *J Mar Syst* **74**: 933–945.
- R Development Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation Statistical Computing.
- Ricotta, C., and Podani, J. (2017) On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecol Complex* **31**: 201–205.
- Sandaa, R.A., Clokie, M., and Mann, N.H. (2007) Photosynthetic genes in viral populations with a large genomic size range from Norwegian coastal waters. *FEMS Microbiol Ecol* **63**: 2–11.
- Sano, E., Carlson, S., Wegley, L., and Rohwer, F. (2004) Movement of viruses between biomes. *Appl Environ Microbiol* **70**: 5842–5846.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Shaw, P.-T., and Chao, S.-Y. (1994) Surface circulation in the South China Sea. *Oceanogr Res Pap* **41**: 1663–1683.

- Shelford, E.J., and Suttle, C.A. (2018) Virus-mediated transfer of nitrogen from heterotrophic bacteria to phytoplankton. *Biogeosciences* **15**: 809–819.
- Sher, D., Thompson, J.W., Kashtan, N., Croal, L., and Chisholm, S.W. (2011) Response of *Prochlorococcus* ecotypes to co-culture with diverse marine bacteria. *ISME J* **5**: 1125–1132.
- Short, C.M., Rusanova, O., and Short, S.M. (2011) Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. *ISME J* **5**: 810–821.
- Short, S.M., and Suttle, C.A. (2002) Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl Environ Microbiol* **68**: 1290–1296.
- Snyder, J.C., Wiedenheft, B., Lavin, M., Roberto, F.F., Spuhler, J., Ortmann, A.C., et al. (2007) Virus movement maintains local virus population diversity. *Proc Natl Acad Sci U S A* **104**: 19102–19107.
- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Stegen, J.C., Lin, X., Konopka, A.E., and Fredrickson, J.K. (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* **6**: 1653–1664.
- Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., and Chisholm, S.W. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**: e234.
- Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**: 1047–1051.
- Suttle, C.A. (1994) The significance of viruses to mortality in aquatic microbial communities. *Microb Ecol* **28**: 237–243.
- Suttle, C.A. (2000) Cyanophages and their role in the ecology of cyanobacteria. In *Encyclopedia of Life Sciences*, Whitton, B., and Potts, M. (eds). The Netherlands: Kluwer Academic Publishers, pp. 563–589.
- Suttle, C.A. (2007) Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.
- Suttle, C.A., Chan, A.M., and Cottrell, M.T. (1990) Infection of phytoplankton by viruses and reduction of primary productivity. *Nature* **347**: 467–469.
- Wang, G., Murase, J., Taki, K., Ohashi, Y., Yoshikawa, N., Asakawa, S., and Kimura, M. (2009) Changes in major capsid genes (*g23*) of T4-type bacteriophages with soil depth in two Japanese rice fields. *Biol Fertil Soils* **45**: 521–529.
- Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**: 127–181.
- Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., et al. (2008). The sorcerer II global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* **3**: e1456.
- Winter, C., Matthews, B., and Suttle, C.A. (2013) Effects of environmental variation and spatial distance on bacteria, archaea and viruses in sub-polar and arctic waters. *ISME J* **7**: 1507–1518.
- Winter, C., Smit, A., Herndl, G.J., and Weinbauer, M.G. (2004) Impact of Virioplankton on archaeal and bacterial community richness as assessed in seawater batch cultures. *Appl Environ Microbiol* **70**: 804–813.
- Winter, C., Smit, A., Herndl, G.J., and Weinbauer, M.G. (2005) Linking bacterial richness with viral abundance and prokaryotic activity. *Limnol Oceanogr* **50**: 968–977.
- Zhang, J., Gao, Q., Zhang, Q., Wang, T., Yue, H., Wu, L., et al. (2017) Bacteriophage–prokaryote dynamics and interaction within anaerobic digestion processes across time and space. *Microbiome* **5**: 57.
- Zhang, J., Yu, Z.G., Wang, J.T., Ren, J.L., Chen, H.T., Xiong, H., et al. (1999) The subtropical Zhujiang (Pearl River) estuary nutrient, trace species and their relationship to photosynthesis. *Estuar Coast Shelf Sci* **49**: 385–400.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Location of the 11 sampling stations in the nSCS. Water samples were collected from surface and subsurface layers.

Fig. S2. (A) Distribution of salinity and temperature among the 17 samples in the nSCS. (B) Three clusters of water samples were identified based on the temperature and salinity. The two components (i.e. components 1 and 2) explained 100% of the point variability. Different water masses are indicated in green (EWM), grey (CWM), red (sOWM) and blue (ssOWM).

Fig. S3. Rarefaction curve of similarity-based operational taxonomic units (OTUs) at 97% similarity level. (A) Viral *g23* and (B) bacterial 16S rRNA gene libraries from the nSCS.

Fig. S4. Frequency and abundance rank plot for viral (A–B) and bacterial (C–D) OTUs. A spline curve and scatter diagram were used to express the relationship between the frequency of OTUs occurrence and the number of OTUs and average contribution of OTUs to the community respectively. Each black dot represents an individual OTU.

Fig. S5. Bacterial community biodiversity and composition at the phylum (A) and class (B) level.

Fig. S6. SIMPER analysis of the dissimilarity among four viral oceanic water masses (cut-off of low contribution: 90%). Only dominant *g23* OTUs were used.

Table S1. Location and environmental parameters of the 17 samples from the nSCS.

Table S2. Good's coverage, richness (Chao I) and diversity (Shannon and Simpson) indices across all samples at the 97% similarity level.

Table S3. Mantel test for the correlation between viral community composition and abiotic and abundance environmental variables using Pearson's coefficient. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$.

Table S4. Mantel and partial Mantel tests for the correlation between the Exo T-even viral community and other variables. ']' indicates partial mantel test.