







Viral Characteristics of the Warm Atlantic and Cold Arctic Water Masses in the Nordic Seas

Chen Gao,^{a,b} Jun Xia,^{a*} Xinhao Zhou,^a  Yantao Liang,^{a,b} Yong Jiang,^{a,b} Meiwen Wang,^a Hongbing Shao,^{a,b} Xiaochong Shi,^a Cui Guo,^{a,b} Hui He,^{a,b} Hualong Wang,^{a,b} Jianfeng He,^c Denghui Hu,^d Xiaoyu Wang,^a Jinping Zhao,^e  Yu-Zhong Zhang,^{f,g} Yeong Yik Sung,^{b,h} Wen Jye Mok,^{b,h} Li Lian Wong,^{b,h}  Andrew McMinn,^{a,i}  Curtis A. Suttle,^j Min Wang^{a,b,k}

^aCollege of Marine Life Sciences, Key Lab of Polar Oceanography and Global Ocean Change, Institute of Evolution and Marine Biodiversity, and Frontiers Science Center for Deep Ocean Multispheres and Earth System, Ocean University of China, Qingdao, China

^bUMT-OUC Joint Centre for Marine Studies, Qingdao, China

^cSOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai, China

^dCAS Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

^eKey Lab of Polar Oceanography and Global Ocean Change, Ocean University of China, Qingdao, China

^fCollege of Marine Life Sciences and Frontiers Science Center for Deep Ocean Multispheres and Earth System, Ocean University of China, Qingdao, China

^gState Key Laboratory of Microbial Technology, Marine Biotechnology Research Center, Shandong University, Qingdao, China

^hInstitute of Marine Biotechnology, Universiti Malaysia Terengganu, Kuala Nerus, Malaysia

ⁱInstitute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia

^jDepartments of Earth, Ocean and Atmospheric Sciences, Microbiology and Immunology, and Botany and Institute for the Oceans and Fisheries, The University of British Columbia, Vancouver, British Columbia, Canada

^kThe Affiliated Hospital of Qingdao University, Qingdao, China

Chen Gao, Jun Xia, and Xinhao Zhou contributed equally to this article. Author order was determined alphabetically.

ABSTRACT Nordic Seas are the subarctic seas connecting the Arctic Ocean and North Atlantic Ocean with complex water masses, experiencing an abrupt climate change. Though knowledge of the marine virosphere has expanded rapidly, the diversity of viruses and their relationships with host cells and water masses in the Nordic Seas remain to be fully revealed. Here, we establish the Nordic Sea DNA virome (NSV) data set of 55,315 viral contigs including 1,478 unique viral populations from seven stations influenced by both the warm Atlantic and cold Arctic water masses. *Caudovirales* dominated in the seven NSVs, especially in the warm Atlantic waters. The major giant nucleocytoplasmic large DNA viruses (NCLDVs) contributed a significant proportion of the classified viral contigs in the NSVs (32.2%), especially in the cold Arctic waters (44.9%). The distribution patterns of *Caudovirales* and NCLDVs were a reflection of the community structure of their hosts in the corresponding water masses and currents. Latitude, pH, and flow speed were found to be key factors influencing the microbial communities and coinfluencing the variation of viral communities. Network analysis illustrated the tight coupling between the variation of viral communities and microbial communities in the Nordic Seas. This study suggests a probable linkage between viromes, host cells, and surface water masses from both the cool Arctic and warm Atlantic Oceans.

IMPORTANCE This is a systematic study of Nordic Sea viromes using metagenomic analysis. The viral diversity, community structure, and their relationships with host cells and the complex water masses from both the cool Arctic and the warm Atlantic oceans were illustrated. The NCLDVs and *Caudovirales* are proposed as the viral characteristics of the cold Arctic and warm Atlantic waters, respectively. This study provides an important background for the viromes in the subarctic seas connecting the Arctic Ocean and North Atlantic Ocean and sheds light on their responses to abrupt climate change in the future.

Citation Gao C, Xia J, Zhou X, Liang Y, Jiang Y, Wang M, Shao H, Shi X, Guo C, He H, Wang H, He J, Hu D, Wang X, Zhao J, Zhang Y-Z, Sung YY, Mok WJ, Wong LL, McMinn A, Suttle CA, Wang M. 2021. Viral characteristics of the warm Atlantic and cold Arctic water masses in the Nordic Seas. *Appl Environ Microbiol* 87:e01160-21. <https://doi.org/10.1128/AEM.01160-21>.

Editor Knut Rudi, Norwegian University of Life Sciences

Copyright © 2021 American Society for Microbiology. All Rights Reserved.

Address correspondence to Yantao Liang, liangyantao@ouc.edu.cn, or Min Wang, mingwang@ouc.edu.cn.

* Present address: Jun Xia, Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto, Japan.

Received 14 June 2021

Accepted 23 August 2021

Accepted manuscript posted online

1 September 2021

Published 28 October 2021

KEYWORDS DNA virome, Nordic Seas, microbial community, metagenomics, water masses and currents

Viruses are the most abundant acellular biological entities on earth and are a large reservoir of genetic diversity in marine ecosystems (1, 2), with an abundance reaching approximately 4×10^{30} virus-like particles in the global ocean (2). Marine viruses play an important role in the marine microbial loop and biogeochemical cycles through viral lysing, which controls the microbial abundance and community structure (3). In addition, viruses induce microbial genetic diversity and evolution through viral infection and horizontal gene transfer (4, 5) and contribute to ocean carbon sequestration through the “microbial carbon pump” and “biological pump” (6–8).

First applied to marine viral communities in 2002 (9), metagenomics has greatly enhanced the understanding of viral community structure, diversity, and distribution patterns (10–14). As a result of progress in DNA extraction, sequencing, assembly, annotation, and other bioinformatic analyses, viral metagenomics has now been applied to many different environments and has become the most powerful method for studying viral community structure, evolution, and relationships with environmental factors at the gene and genome levels (10–13). Through bioinformatic analysis of the metagenome data, our current knowledge of giant nucleocytoplasmic large DNA viruses (NCLDVs) has been greatly expanded (15, 16), and relationships between viral distribution patterns and community structure and oceanic water masses and currents have been proposed (17–19). Although this understanding of the diversity and distribution patterns of marine viromes has been greatly expanded over the last decade, there are still very few reports of viromes from the Arctic Ocean and adjacent regions (10, 20). Recently, the *Tara* Oceans Polar Circle (TOPC) expedition studied the viromes of the Arctic Ocean through the sequencing of 41 samples from 20 different sites and showed that the Arctic Ocean is a unique viral ecological zone with high viral diversity (10). However, understanding the diversity of viromes in the Arctic Ocean and adjacent oceans remains limited. In particular, no study identifying the relationship between DNA viromes, microbial community structure, and water masses and currents in areas with strong intermixing between different oceans, such as in the Nordic Seas, has been performed.

The “Nordic Seas” are the subarctic seas connecting the Arctic Ocean and the North Atlantic Ocean, located at the southern end of the Fram Strait, are the northern part of the Greenland-Scotland Ridge and are experiencing an abrupt climate change (21). This region facilitates the passage of heat and water mass between the Arctic Ocean and North Atlantic Ocean (22). In summer, the surface waters of the Nordic Seas are affected by both the warm Norwegian Atlantic Current, which flows along the Norwegian continental slope from south to the north, and the cold East Greenland Current, which carries the outflowing Arctic water from the western part of the Fram Strait to the south (Fig. 1) (23–26). The complex pattern of surface water currents in the Nordic Seas provides a unique environment to study the relationship between marine viromes and water masses and currents. In this study, we present a Nordic Sea Virome (NSV) data set from seven surface seawater samples and characterize the community structure and diversity of viruses and their relationship with host cells, water masses, and currents in the Nordic Seas.

RESULTS

Hydrological characteristics. Seven viral DNA samples were collected from the surface waters of different water masses of the Nordic Seas during a cruise from 2 to 26 June 2015. The detailed information of environmental factors was shown in Table S1 in the supplemental material. Temperature was the most variable parameter and the surface water temperatures conformed to the hydrological characteristics (Fig. 1A). The seven stations can be categorized into two groups based on temperature. Four stations (WB1, WB2, WC1, and WC2) were located in warmer water masses (average temperature was $6.99 \pm 1.05^\circ\text{C}$) with low chlorophyll *a* (Chl *a*) (average value was $0.84 \pm 0.44 \text{ mg m}^{-3}$) and were mainly influenced by the warm and northward Atlantic currents. Three stations

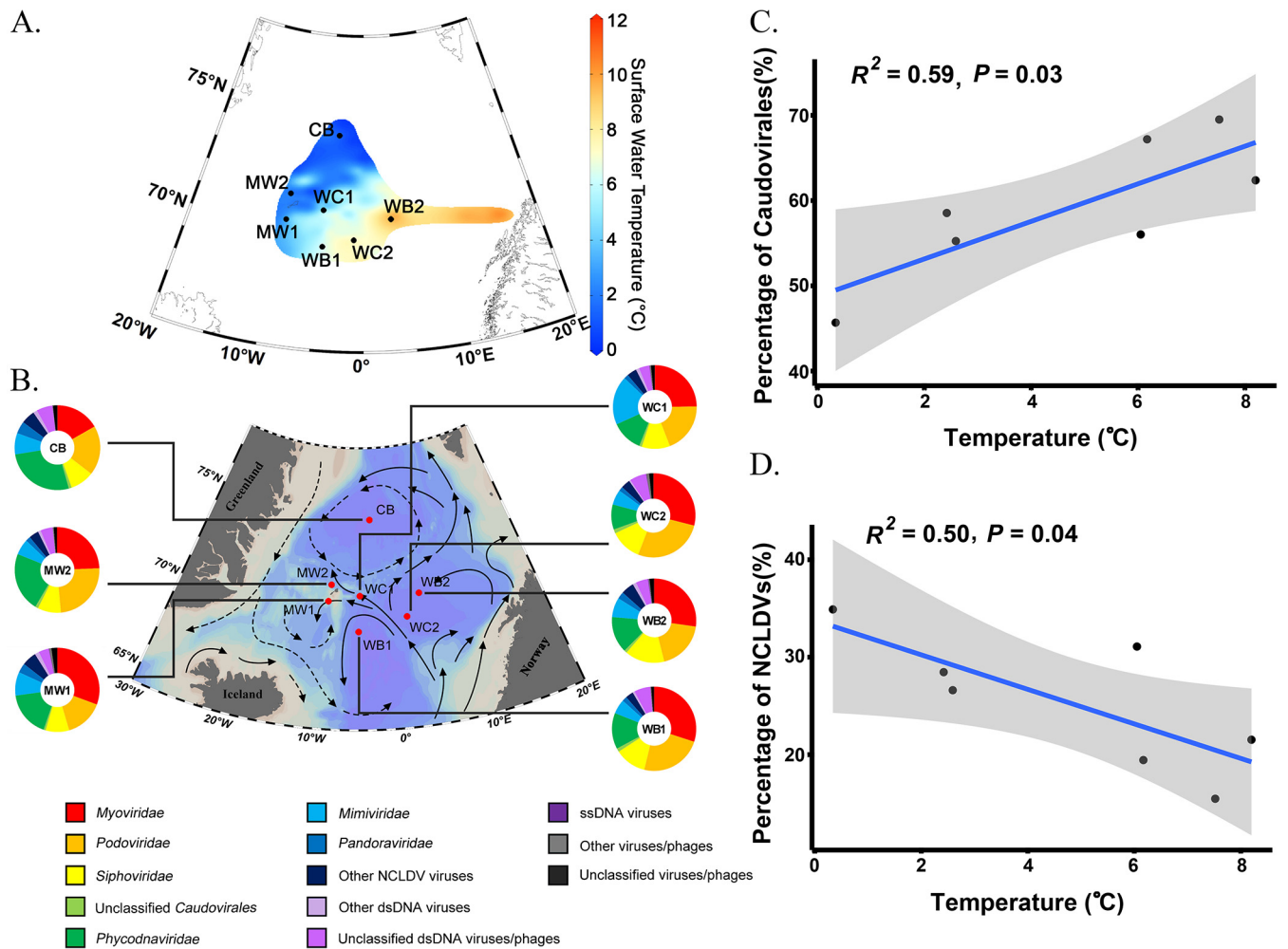


FIG 1 Nordic Sea Viromes (NSVs) and relationships with temperature. (A) Map of sampling stations with surface hydrological characteristics information. (B) Percentages of viral families of each metagenomic site in the Nordic Sea. The direction of flow velocity (arrow), warm current (solid line), cold current (dashed line), and locations of each sampling site (red point) are shown. Surface seawater samples (150 liters) were collected for each viral metagenomic sample during the cruise in June 2015 in Nordic Seas. The information of hydrological characteristics is revised from Wang et al. (27). Different colors represent the following: deep gray, land and light gray to light blue to deep blue (the depths from shallow to deep sea). All contigs were aligned with the viral reference protein sequence database (BLASTx, E value < 10⁻³, the best BLAST hit number) on MetaVir. The percentage refers to the proportion of relative abundances of all viral populations in the seven samples. Different colors refer to different viral groups shown in the legend on the right side. (Map created using Ocean Data View software [76].) (C and D) Relationships between surface water temperatures and the percentages of *Caudovirales* (C) and NCLDVs (D). Gray shading indicates the 95% confidence interval.

(MW1, MW2, and CB) were located in the colder water masses (average temperature, 1.78 ± 1.25°C) with high chlorophyll *a* (average, 1.79 ± 1.28 mg m⁻³) and were mainly influenced by the cold and southward Arctic currents (Fig. 1; see also Table S1).

According to the surface water current characteristics (Fig. 1B) (27), WB1 and WB2 were located in the stable warm waters (higher temperature and low flow speed), and WC2 and WC1 were located on the same warm current (higher temperature, higher flow speed). MW1 and MW2 were located in mixed areas of warm and cold currents. A warm current flowing around Jan Mayen Island split into two and merged separately with two cold currents. CB was located in the stable cold water (lower temperature and lower flow speed) in the northernmost basin.

For the three stations (MW1, MW2, and CB) located in the colder water masses, high concentrations of Chl *a* (3.15 and 1.62 mg m⁻³, respectively) and primary production (1,999 and 1,237 mg C m⁻² day⁻¹, respectively) were detected in the mixed water bodies because the merging can accelerate vertical convection (*Z_{eu}* values [i.e., the depth of the euphotic zone] were 20.2 and 26.6 m, respectively) (28). Low concentrations of Chl *a*

TABLE 1 Assembly information of seven samples of Nordic Sea Viromes^a

Site	Total length (Mbp)	GC content (%)	No. of contigs	Avg length (bp)	N ₅₀	Length (bp)		Read ratio ^b (%)
						Longest	Shortest	
CB	71.36	46.06	31,626	2,218	2,489	33,618	1,000	19.51
MW1	42.78	48.77	24,476	1,718	1,686	25,951	1,000	7.62
MW2	51.07	51.37	28,027	1,791	1,780	40,991	1,000	10.83
WB1	45.48	43.81	21,023	2,127	2,240	54,441	1,000	9.82
WB2	63.25	42.34	31,341	1,984	2,027	82,375	1,000	14.58
WC1	60.76	45.56	27,406	2,180	2,311	92,674	1,000	12.36
WC2	40.59	40.51	18,660	2,139	2,267	50,916	1,000	8.49

^aAbbreviations: bp, base pair; Mbp, million base pairs; GC, guanine and cytosine.

^bRead ratio refers to the percentage of the reads used for assembly.

(0.61 mg m⁻³) and primary production (587 mg C m⁻² day⁻¹) were detected in the northernmost CB station, with the lowest temperature (0.34°C) and deepest depth of the euphotic zone (40.4 m) (see Table S1).

Data set for NSVs. A total of 182,559 contigs (longer than 1,000 bp) were assembled from 35.0 Gb of sequencing data across the seven surface water samples. The total length of all the assembled contigs was 375.29 Mbp. Other statistical data of the assembly results were shown in the Table 1. A total of 453,556 open reading frames (ORFs) were predicted from the assembled contigs, with the total length of 93.79 million amino acids and an average length of 207 amino acids (see Table S2). Detailed information for the predicted ORFs in each sample is shown in Table S2.

After BLAST with the reference viral genome sequence data set, a total of 55,315 contigs (30.3% of total contigs) were hit with specific viral genome sequences. Interestingly, the BLAST hit percentages of contigs to viral genome sequences in warmer water (WC1, 37.56%; WB1, 39.36%; WC2, 39.66%; WB2, 29.39%) were higher than that in colder water (MW1, 23.73%; CB, 23.80%; MW2, 24.24%) (analysis of variance [ANOVA], $P < 0.01$; see Fig. S1). This percentage was related to the temperature (linear regression, $R^2 = 0.58$; see Fig. S1).

Taxonomic composition and diversity of DNA viral communities in NSVs. The taxonomic analysis showed that most of the viral populations predicted were double-stranded DNA (dsDNA) viruses (97.84%); only 0.16% single-stranded DNA (ssDNA) viruses were detected. Most of the viral populations in the NSVs were classified into the order *Caudovirales* (average, 59.22% [ranging from 45.68 to 69.51%]), which included the families *Myoviridae* (26.11%), *Podoviridae* (20.94%), and *Siphoviridae* (11.07%). Interestingly, the *Phycodnaviridae* (16.65%), *Mimiviridae* (8.71%), and *Pandoraviridae* (2.60%) families, which belong to the NCLDVs and contain giant viruses infecting eukaryotes (29), were much more abundant than in other oceanic viromes (10, 11). As with the viral hit percentage (see Fig. S1), the abundance of *Caudovirales* was significantly correlated with a decline in temperature (Fig. 1C, linear regression, $R^2 = 0.59$, $P = 0.03$). However, the reverse trend was seen in NCLDVs, which declined in abundance in warmer water (Fig. 1B). The abundance of the NCLDV taxa at each site was negatively correlated with water temperature (linear regression, $R^2 = 0.50$, $P = 0.04$; Fig. 1D).

The 10 most abundant viral populations are shown in Fig. 2A. The most abundant, with a greatest abundance in WB1 (7.95%), was predicted to be *Pelagibacter* phage HTVC008M (5.45%). The top three abundant viral populations—HTVC008M, *Puniceispirillum* phage HMO-2011 (rank 2, 5.03%), and *Cellulophaga* phage phi38:1 (rank 3, 4.28%)—all belong to the *Caudovirales*, as well as *Bacillus* phage G (rank 6, 2.67%), *Prochlorococcus* phage P-TIM68 (rank 9, 1.95%), and *Cellulophaga* phage phi14:2 (rank 10, 1.86%). *Micromonas* sp. strain RCC1109 virus MpV1 (rank 4, 3.21%) and *Chrysochromulina ericina* virus (CeV; rank 7, 2.03%) were classified into *Phycodnaviridae*. For the *Mimiviridae*, *Megavirus chiliensis* (2.84%) and *Cafeteria roenbergensis* virus BV-PW1 (2.01%) were ranked 5 and 8, respectively. The top 10 viral populations were thus all classified into the *Myoviridae*, *Podoviridae*, *Phycodnaviridae*, or *Mimiviridae*. Correlation analysis between the most abundant 10 viral populations and environmental factors indicated that *Micromonas* sp. strain RCC1109 virus (MpV1) was significantly positively correlated with latitude ($R = 0.83$, $P < 0.05$) (Fig. 2B). *Cafeteria roenbergensis* virus BV-PW1 was significantly negatively correlated with longitude ($R = -0.88$, $P < 0.01$), and

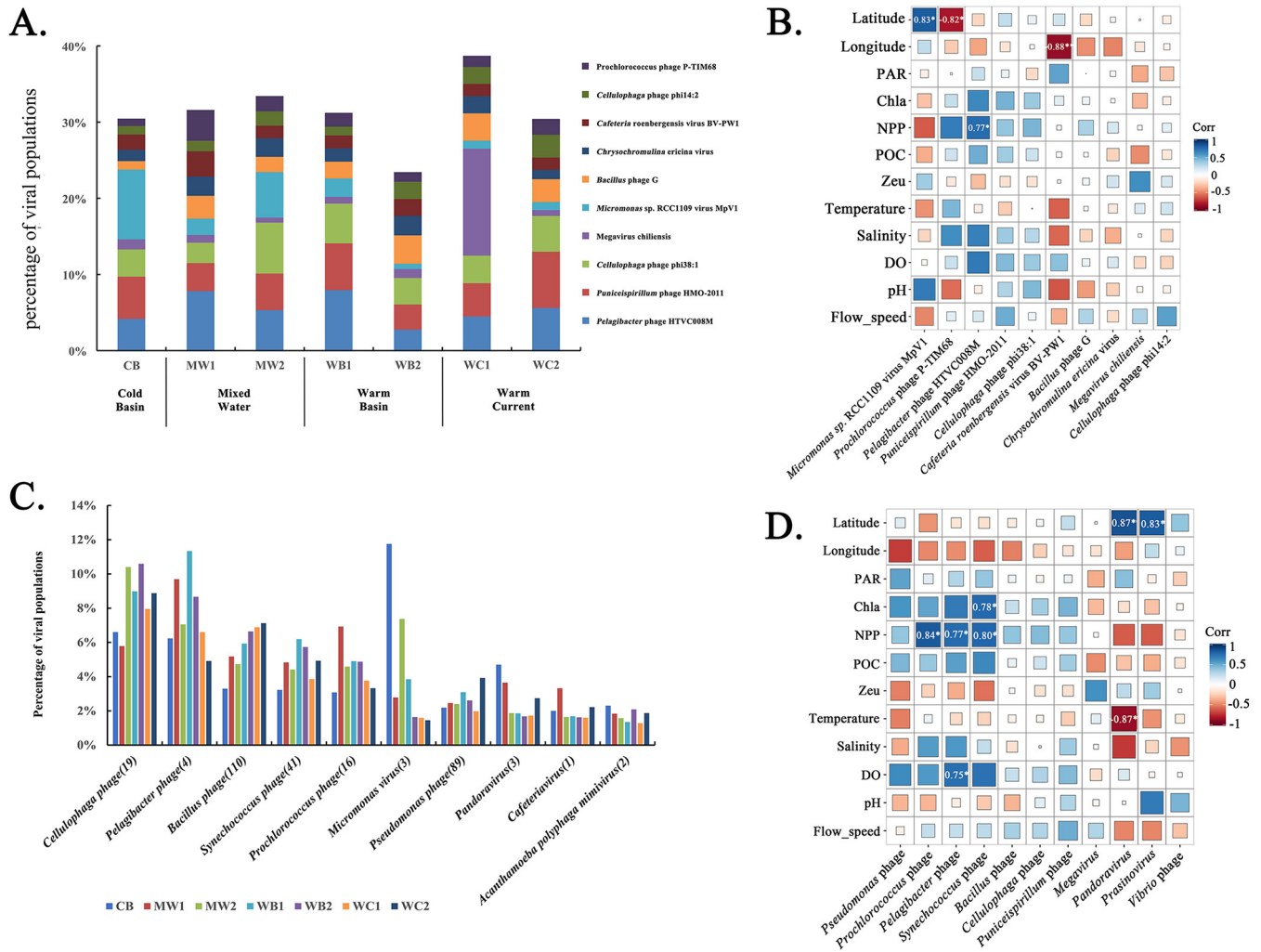


FIG 2 Abundant viral populations in the NSVs and their correlations with environmental factors. (A and B) Percentage of relative abundance of top 10 viral populations to the total viral populations in each sample (A) and their correlations with environmental factors (B). (C) Percentage of total abundance of specific viral populations classified by the classification of hosts. (D) Correlation of top 11 most abundant viral populations (genus level) with environmental factors. The numbers in brackets for specific viral populations indicate the total viral population numbers found in all stations. Stations are ranged from cold water (left) to warm water (right) in order. All the contigs were aligned with a viral reference protein sequence database (BLASTx, E value < 10⁻³, best BLAST hit number) on MetaVir. Different colors refer to different viral populations in panels A and C. The most abundant populations are ordered by the percentage of total populations in all stations and are not represented in the order of the populations for each sample. The most abundant populations to the tenth were listed from the bottom to peak of the columns. Different colors refer to the different stations indicated in the legend at the bottom. The colors and sizes of the boxes in panels B and D represent the correlation values.

Prochlorococcus phage P-TIM68 were significantly positively correlated with net primary productivity (NPP) ($R = 0.77, P < 0.05$) (Fig. 2B).

For the viral populations infecting the specific hosts, *Cellulophaga* phages (19 populations; average, 8.46%) and *Pelagibacter* phages (four populations; average, 7.79%) were the most abundant (Fig. 2C). In addition, *Bacillus* phages (110 populations; average, 5.69%), *Synechococcus* phages (41 populations; average, 4.74%), and *Prochlorococcus* phages (16 populations; average, 4.49%) were also abundant. It is worth noting that *Micromonas* viruses (3 populations, 4.35%), *Pandoravirus* (3 populations, 2.60%), *Cafeteria* virus (1 population, 2.01%), and *Acanthamoeba polyphaga* mimiviruses (2 populations; average, 1.76%), which belong to NCLDVs, also accounted for a large percentage (total 10.71%). *Micromonas* viruses and pandoraviruses, were significantly more abundant in colder water than in warmer water (ANOVA, $P < 0.05$). This was especially clear at the coldest water site near the Arctic Ocean (CB in the Greenland Basin) (Fig. 2C). This pattern is similar to that of *Phycodnaviridae*. The brown tide algae viruses (2.95%), including *Aureococcus anophagefferens* virus, *Ectocarpus siliculosus* virus, and *Feldmannia* virus, were unexpectedly abundant. Other viral taxonomic groups, such as *Chlorella* viruses, *Ostreococcus* viruses, *Vibrio* phages, and archaeal viruses,

were also abundant. Interestingly, NCLDV-associated virophages (6 populations, including *Phaeocystis globosa* virus virophage, Sputnik virophage, Yellowstone Lake virophages 5 to 7, and Zamilon virophage; average, 0.12%) were found in the NSVs. *Pithovirus sibericum*, an “ancient NCLDV” that was first discovered in an ice core more than 30,000 years old in Siberian permafrost (30), was detected in each of the NSV samples. *Pithovirus sibericum* was more abundant in the colder Arctic waters (0.70%) than in the warmer Atlantic waters (0.65%). Correlation analysis between the most abundant viral populations infecting the specific host taxonomies and environmental factors showed that *Micromonas* virus (MpV1 and *Prasinovirus*) was significantly positively correlated with latitude ($R = 0.83$, $P < 0.05$) (Fig. 2B and D). *Cafeteriavirus* was significantly negatively correlated with longitude ($R = -0.88$, $P < 0.01$) (Fig. 2B), and *Pandoravirus* was significantly negatively correlated with temperature ($R = -0.87$, $P < 0.05$) and positively correlated with latitude ($R = 0.87$, $P < 0.05$) (Fig. 2D).

The viral diversity (Shannon-Weaver biodiversity index and Pielou's J' evenness index) of each sample was calculated from the annotated viral populations (see Table S3). A total of 1,478 viral populations from the MetaVir database were detected; the populations were greatest at WC1 ($n = 982$) and lowest at MW1 ($n = 703$). The Shannon H' index was highest at WC2 (5.591) and lowest at WC1 (5.33), which is similar to the results from the Northeast Atlantic section of the Arctic Ocean but lower than for other regions of the Arctic Ocean (10). Combined with the Shannon H' diversity of surface virome samples in the GOV 2.0, the Arctic Ocean could be divided into areas of high diversity (ARC-H, mainly located in the Pacific Arctic region, the Arctic Archipelago, and the Davis-Baffin Bay) and low diversity (ARC-L, mainly located in the Kara-Laptev Sea, the Barents Sea, and subarctic areas). The Nordic Seas were classified into the ARC-L region of Arctic Ocean (Fig. 3A). If only the ARC-L region is considered, a significant latitudinal pattern in diversity is observed (Cubic regression, $R^2 = 0.72$; Fig. 3B). If the ARC-L region is excluded, the Arctic Ocean diversity increases (cubic regression, $R^2 = 0.34$, Fig. 3C), which is similar to the results of GOV 2.0 (10).

Phylogenetic analysis of caudoviruses and NCLDVs. The terminase large subunit (*terL*, PF03237) and major capsid protein (*mcp*, PF04451) were used to construct the maximum likelihood phylogenetic tree of viral contigs in the NSVs. The phylogenetic tree of *terL* (Fig. 4A) showed that the *Caudovirales* contigs from the NSVs were more likely related to myoviruses and podoviruses. A new viral cluster (cluster I, in light yellow) was found in the study. The evolutionary position of cluster I was between two groups of podoviruses; *Caudovirales* was close to three *Pseudomonas* phages (Fig. 4A). For NCLDVs, the phylogenetic tree of *mcp* (Fig. 4B) showed that the NSV sequences were closely related to *Prasinovirus* (*Phycodnaviridae*) infecting the *Prasinophyceae* *Micromonas pusilla*, *Micromonas* sp., *Osterococcus lucimarinus*, and *Osterococcus tauri*, and the unclassified viral cluster (deep blue) was below one of the clusters of the prasinoviruses. Two new NCLDV viral clusters (clusters II and III) were found located close to the root of the tree.

Comparison of viromes and relationships with environmental factors. According to the results of hierarchical clustering of Bray-Curtis dissimilarity distances (Fig. 4C), the seven viromes in the NSVs could be robustly assigned to two groups: (i) MW1, MW2, and CB and (ii) WC1, WC2, WB1, and WB2, which is consistent with the two different water masses (Fig. 1A and B). To assess the relationship between viral community composition and environmental factors, a Bioenv test with Spearman rank correlations was applied to the viral population profile (96, 102). This Bioenv analysis showed the variation of viral populations was mainly explained by latitude and pH (best variable combination: $R = 0.6234$, $P < 0.001$), while the R value between viral populations and latitude, pH, and flow speed was 0.6156 ($P < 0.001$; Table 2).

Taxonomic composition and diversity of bacterial and microbial eukaryotic communities in Nordic Seas. A total of 427 bacterial operational taxonomic units (OTUs) were assigned from the seven samples at 3% dissimilarity threshold. These OTUs were classified into 21 phyla, 28 classes, 84 orders, 139 families, and 200 genera. The dominant phyla were *Proteobacteria* (64.68%, mainly including classes *Gammaproteobacteria* at 38.55% and *Alphaproteobacteria* at 26.13%), *Bacteroidota* (22.31%), and *Cyanobacteria* (10.31%). The dominant genera with relative abundances higher than 5% were *Psychrobacter* (7.8%) and *SUP05_cluster* (6.5%) (Fig. 5A). The

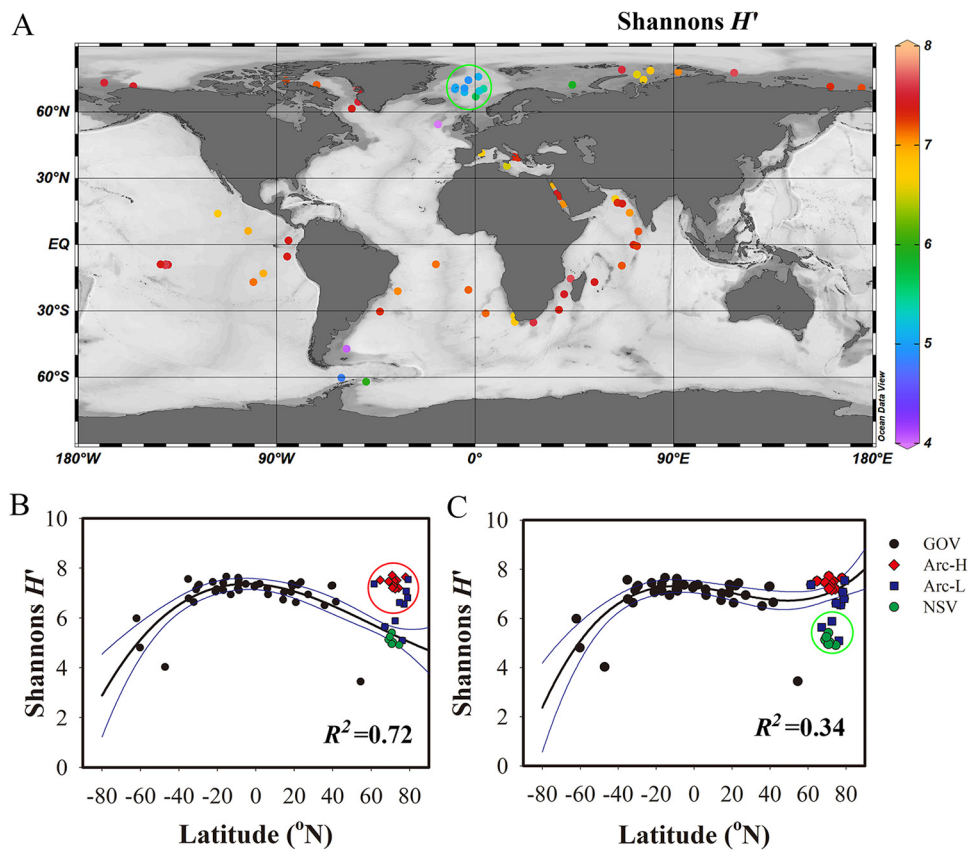


FIG 3 Different latitudinal patterns of surface viral diversity in the global ocean. (A) Global distribution of surface viral diversity in the ocean. The green oval represents the low-diversity region of the Arctic Ocean. (Map created using Ocean Data View software [76].) (B and C) For the relationship between viral diversity and latitude, only the low-diversity regions of the Arctic Ocean were considered (B), and only the high-diversity regions of the Arctic Ocean were considered (C). To distinguish the results of the low- and high-diversity regions of the Arctic Ocean separately, the data in the ovals of panels B and C were excluded from the cubic regression analysis. Abbreviations: GOV, Global Ocean Viromes; Arc-H, high-diversity regions in the *Tara* Oceans Polar Circle; Arc-L, low-diversity regions in the *Tara* Oceans Polar Circle; NSVs, Nordic Sea Viromes.

abundance of *Psychrobacter* was low in CB and MW samples. *Nitrincolaceae* were less abundant in the WB samples than in other samples.

After eliminating 22 animal OTUs, 232 microbial eukaryotic OTUs were identified from the seven samples. The OTUs were classified into 11 phyla, 37 classes, 55 orders, 71 families, and 108 genera. The dominant classes or phyla were *Intramacronucleata* (27.87%), *Mamiellophyceae* (25.16%), *Prymnesiophyceae* (2.99%), *Dinophyceae* (2.89%), and the SAR supergroup (i.e., Stramenopiles, Alveolata, and Rhizaria). *Intramacronucleata* and *Mamiellophyceae* were least abundant in MW2. The dominant genera were *Micromonas* (33.41%) and *Strombidium* (26.48%). The abundance of *Micromonas* spp. varied slightly in different samples. The abundance of *Strombidium* spp. was slightly lower in MW2 than in other samples (Fig. 5B).

Correlation analysis between the most abundant bacterial genera, microbial eukaryotic genera, and environmental factors indicated that the relative abundances of eukaryotic genera were significantly positively correlated with flow speed ($R = 0.82\sim 0.98$, $P < 0.05$), while the results for bacterial genera were more complicated (Fig. 5C and D). NS5 marine group and *Cryomorphaceae* were positively correlated with flow speed ($R = 0.85$ and 0.89 , $P < 0.05$). *Planktomarina* and *Aurantivirga* were negatively correlated with temperature ($R = -0.87$ and 0.91 , $P < 0.01$) and positively correlated with latitude ($R = 0.81$ and 0.86 , $P < 0.05$). *Flavobacteriaceae* and OM182_clade were significantly positively correlated with Chl *a*, NPP, particular organic carbon (POC), and dissolved oxygen (DO) ($R = 0.76\sim 0.96$, $P < 0.05$).

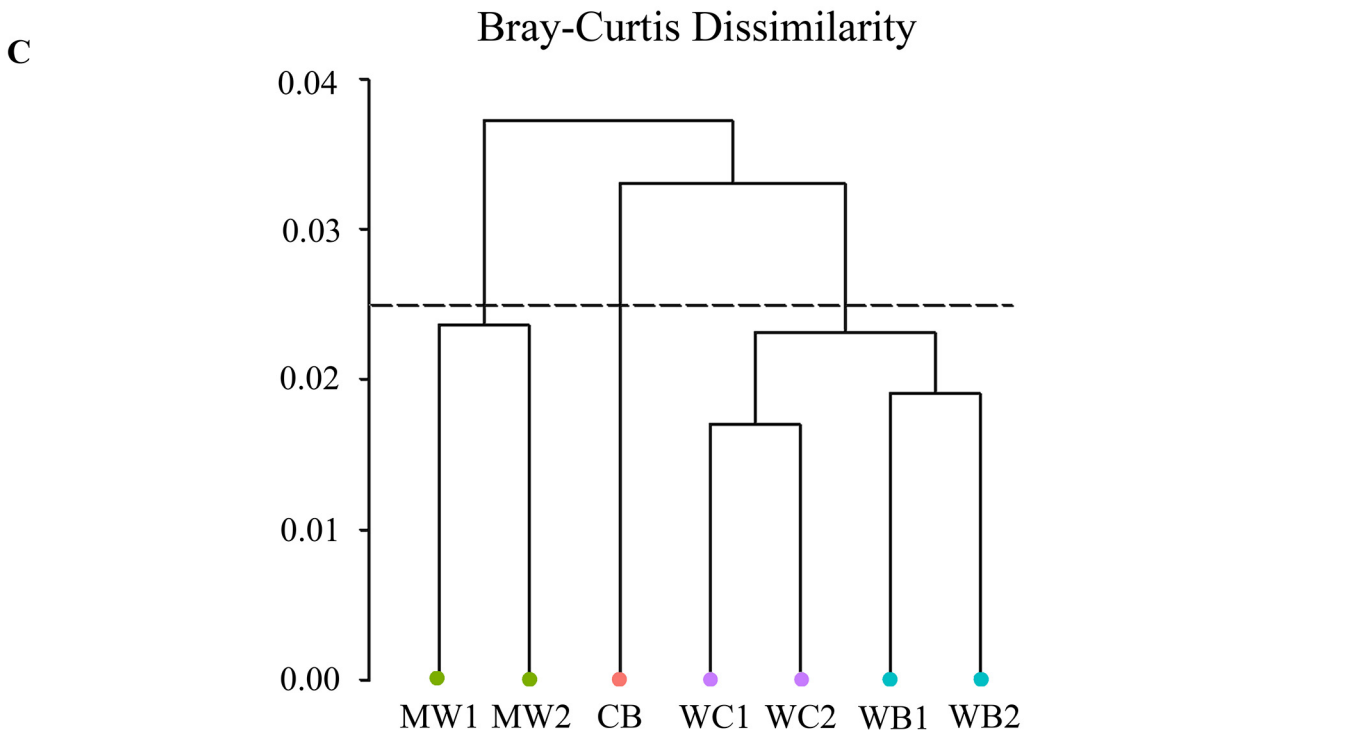
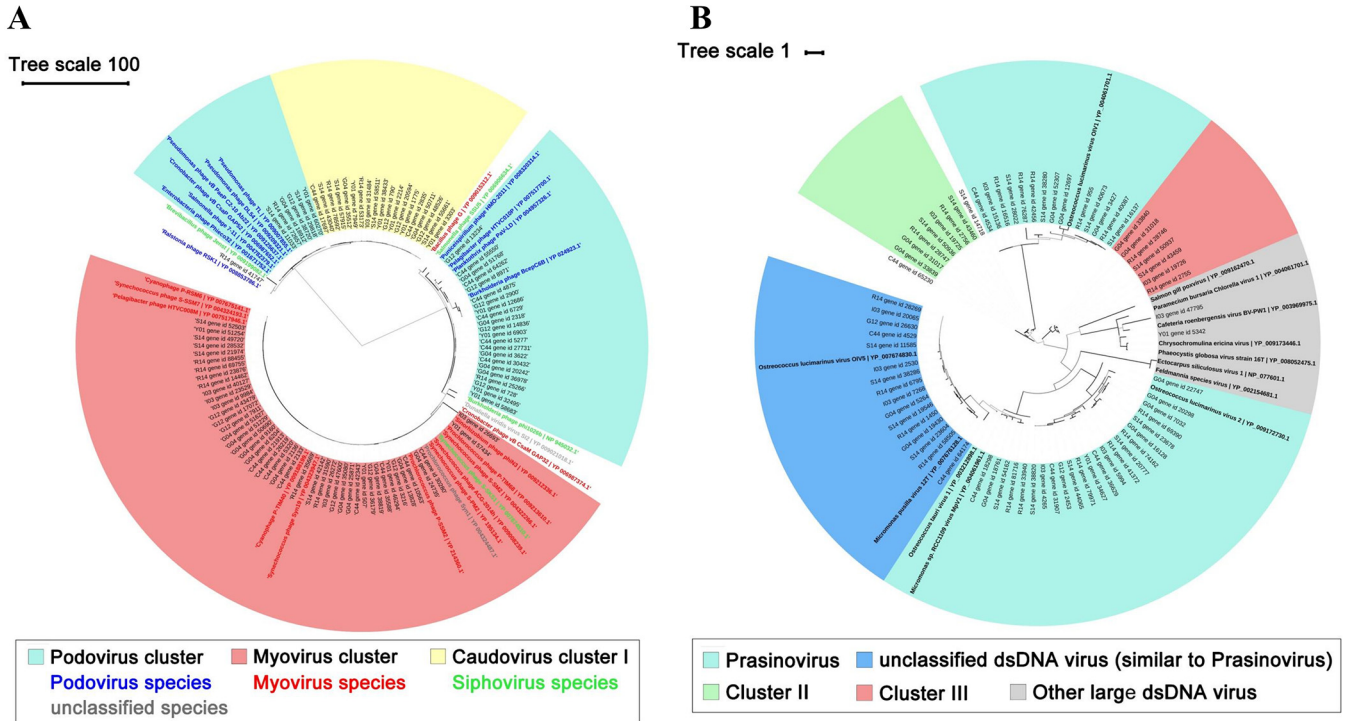


FIG 4 Maximum-likelihood phylogenetic tree of viral contigs in NSVs and hierarchical clustering of NSVs based on amino acid sequences hit with the terminase large subunit (*TerL*, PF03237) (A) and the major capsid protein (*mcp*, PF04451) (B). All ORFs were predicted from contigs of each virome and aligned with the viral marker genes *TerL* and *mcp* based on BLASTp (E value < 10⁻³). ORFs with highly aligned result (score > 300 and length > 150 amino acids for *TerL*; score > 150 and length > 150 amino acids for *mcp*) were selected to analyze phylogenetic relations. The text of reference sequences is indicated in boldface and highlighted in different colors. The bootstrap values are presented from minimum (light gray) to maximum (black) on the branches. All *TerL* reference sequences belonged to the complete genome of *Caudovirales*, and all *mcp* reference sequences belonged to the complete genome of NCLDVs. (C) Distance based on the Bray-Curtis dissimilarity was scored in a matrix based on the dissimilarity frequency of the tetranucleotide. A dashed line divides the seven viromes into several groups based on 0.025 dissimilarity. The colors of each site represent different waters as follows: blue point, warm basin; purple point, warm current; green point, mixed water; and red point, cold basin.

TABLE 2 Bioenv test results, along with Spearman rank correlations between viral populations and environmental factors

Rank	R	Environmental variable(s)	P
1	0.6234	Latitude, pH	0.001
2	0.6156	Latitude, pH, flow speed	0.001
3	0.5649	Latitude, NPP, pH, flow speed	0.001
4	0.4571	Latitude	0.001
5	0.4416	Latitude, NPP, POC, pH, flow speed	0.001
6	0.3299	Latitude, longitude, NPP, temp, pH, flow speed	0.001
7	0.2701	Latitude, longitude, NPP, temp, salinity, pH, flow speed	0.001
8	0.1987	Latitude, Chl <i>a</i> , NPP, POC, Zeu, temp, pH, flow speed	0.001
9	0.1065	Latitude, Chl <i>a</i> , NPP, POC, Zeu, temp, DO, pH, flow speed	0.001
10	0.0143	Latitude, PAR, Chl <i>a</i> , NPP, POC, Zeu, temp, DO, pH, flow speed	0.001

Network analysis of the relationship between viruses and hosts. The ecological network analysis, based on the Pearson index, was applied to determine the relationship between viruses and microbial eukaryotic and bacterial communities in the Nordic Seas. A total of 16,157 correlations were detected. Of these, 11,936 virus-host linkages were positively related, which indicated that the covariations of most viruses with their hosts (Fig. 6A). The most abundant viral populations were linked to *Micromonas* and *Calanoida* (Fig. 6B). *Siphoviridae*, *Myoviridae*, and *Podoviridae* comprised most of the viruses in this network. The relationship between viruses and their hosts, with relative abundances higher than 0.14%, are shown in Fig. 6B.

The ecological network analysis was applied to the total (see Fig. S2A) and to the most abundant viral populations (see Fig. S2B). All curves of network connectivity were fitted with the power-law model ($R > 0.9$). As with the relative abundance of viral populations in the NSVs (Fig. 1B), the viral populations from *Myoviridae*, *Podoviridae*, *Phycodnaviridae*, and *Mimiviridae* were found to comprise the hub nodes in the networks (see Fig. S2). Positive connections dominated the interactions between viral populations (see Fig. S2), suggesting that the effects of mutualism might dominate viral networks in the Nordic Seas.

DISCUSSION

Abundant NCLDV in NSVs might be the viral characteristic of cold Arctic water mass. The NCLDVs (the *Phycodnaviridae*, *Mimiviridae*, and *Pandoraviridae* families) were abundant in the Nordic Seas viromes (Fig. 1 and 2), especially in the cold Arctic waters (44.9%) (Fig. 1B); this distribution is similar to that of other viromes in Arctic (Arctic 2002 virome on MetaVir) and Antarctic waters (31). NCLDVs were also present as hub nodes in the ecological network (see Fig. S2). The higher GC contents and longer lengths of contigs in colder Arctic waters (CB; Table 1) might be a result of the abundance of NCLDV sequences, which have both high GC content (20, 32–34) and large genomes (35); this is confirmed by the higher proportion of NCLDVs in the CB in the NSV data sets (Fig. 1B and Fig. 2) and in the Arctic Ocean near Svalbard (34). However, the lowest BLAST percentage (23.80%) also occurred in CB, which suggests that a large proportion of unknown viruses was present in the Arctic, with a smaller proportion in the East Greenland Current. These results indicate that Nordic Seas might be a previously unrecognized hot spot of giant viruses.

Micromonas sp. RCC1109 virus MpV1 and *Chrysochromulina ericina* virus (CeV) were the most abundant *Phycodnaviridae* taxa in the NSVs (Fig. 2). This is possibly explained by the cooccurrences of these viruses with their potential hosts (Fig. 6B). Both *Micromonas* and *Haptolina* (*Chrysochromulina*) are common and abundant phytoplankton genera in Norwegian coastal waters (34, 36) and the Nordic Seas (Fig. 5B) (37). *Micromonas pusilla*, a cold-adapted alga with an optimum growth temperature of 2 to 6°C, is more abundant in the Arctic due to climate-induced changes (38–40). The mixing of warm Norwegian Atlantic Current and cold East Greenland Current probably provides a suitable environment for *Micromonas* and its associated viruses (Fig. 5B).

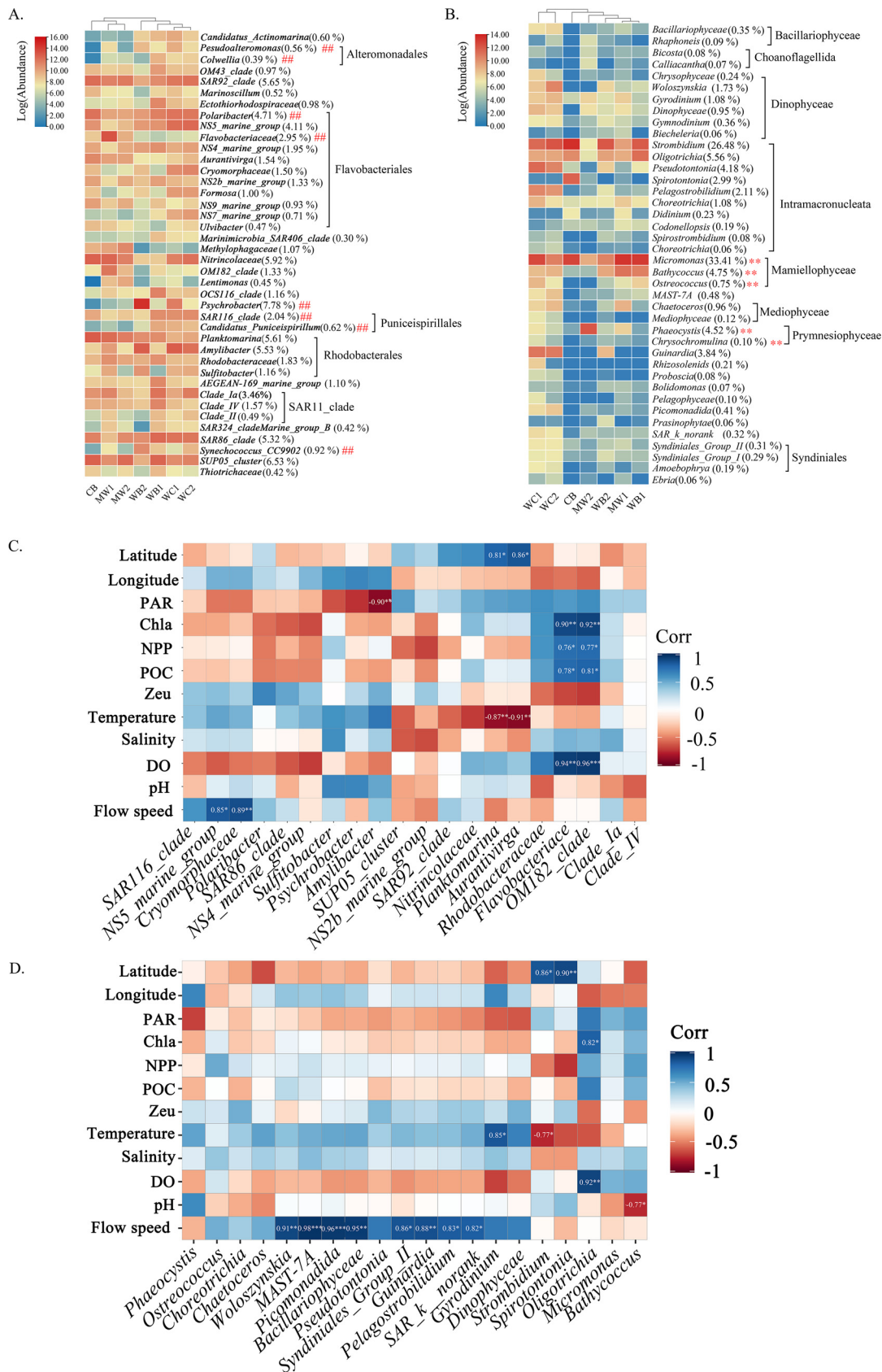


FIG 5 Abundance of bacterial and microbial eukaryotic community in the NSVs and their correlations with environmental factors. (A) Heatmap of top 40 abundant bacterial taxonomies at the genus level. The color of the block represents the (Continued on next page)

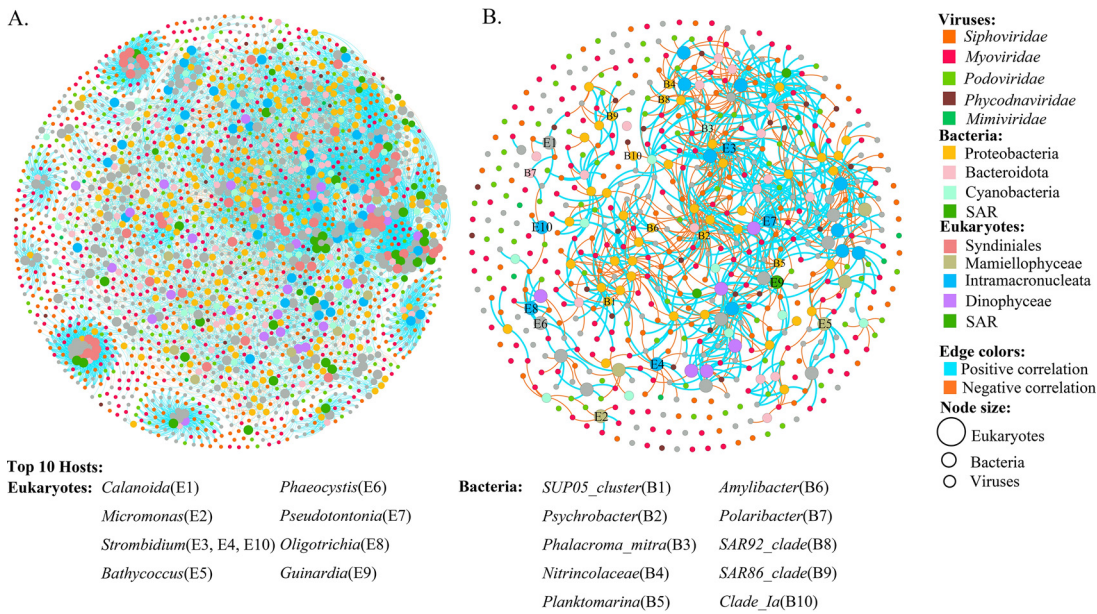


FIG 6 Correlation of viruses and hosts (bacterial and microbial eukaryotic community). (A) Correlation network of viruses and hosts. The blue edge means that two nodes are positively correlated, whereas the orange edge means that two nodes are negatively correlated. (B) Subnetwork that only contains nodes whose abundance is >0.14%. The node size represents the taxonomy. Top 10 abundant hosts for eukaryotes and bacteria, respectively, are illustrated.

Megavirus chiliensis and *Cafeteria roenbergensis* virus BV-PW1 (CroV) within the family *Mimiviridae* were abundant in the NSVs (Fig. 2) but not in the waters around Svalbard (79 to 82.6°N, 1 to 22°E) and outer Oslofjorden (59.17°N, 10.69°E) (34, 36). CroV was first isolated from the marine environment (41); using bioinformatics analysis, it was detected in Global Ocean Sampling (GOS) marine metagenomes and predicted to be an abundant viral group (42–44). CroV has a large genome (~730 kb of dsDNA), which could infect a major marine microflagellate grazer (41) and be infected by its virophage *Mavirus* (45), which was also detected in the NSVs with low relative abundance.

Several “ancient NCLDVs” infecting amoebas, such as *Pithovirus sibericum* (30), *Acanthamoeba polyphaga* moulmouvirus, and *Acanthamoeba polyphaga* mimivirus, were detected in each of the NSV samples (Fig. 2). This may be because their hosts—amoebas—are present in most marine environments and because they have a broad host range, being able to infect different species of amoebas from many different areas (15, 46).

Two new NCLDV viral clusters were detected at the root of the phylogenetic tree of *mcp* (Fig. 4B). This might indicate that the two clusters are from very old viruses and that the diversity of the NCLDVs was high. Most of the *mcp* sequences (Fig. 4B) were related to prasinoviruses infecting the *Prasinophyceae*, which are one of the most abundant picoeukaryotes in Nordic Seas (Fig. 5B) (37, 47). Using metabarcoding analysis, prasinoviruses were reported as the dominant *Phycodnaviridae* in the Arctic Ocean near Svalbard (34), and the results of the phylogenetic analysis suggest a close co-occurrence of viruses and their host cells in Nordic Seas (Fig. 2, 5, and 6).

The Nordic Seas is an area of mixing between the Arctic and Atlantic Oceans (Fig. 1). The cold-water mass is from the Arctic, while the warm water mass represents the waters of the Atlantic. The NCLDVs groups might be an indicator of cold Arctic water viral groups (Fig. 1 and 2). The proportion of NCLDVs (*Phycodnaviridae*, *Mimiviridae*, and *Pandoraviridae* families) decreased from the colder Arctic waters to the warmer Atlantic waters and was significantly

FIG 5 Legend (Continued)

abundance. (B) Heatmap of the top 40 abundant microbial eukaryotic taxonomies at the genus level. (C) Correlation of bacterial community with environmental factors. (D) Correlation of microbial eukaryotic with environmental factors. “##” and “***” represent the corresponding viruses of these bacteria and the microbial eukaryotic genera, respectively, detected in the NSVs.

correlated with temperature (Fig. 1B and D), which reflects the abundance of NCLDV sequences in cold Arctic waters (32, 33). For the most abundant viral populations, the percentage of *Micromonas* virus showed similar patterns to the total NCLDVs (Fig. 2A and C) and was positively correlated with latitude (Fig. 2B and D). This pattern could be a reflection of the distribution of their host cells (Fig. 5B) (37, 47). CeV and CroV were positively correlated with flow speed (Fig. 2B and D), which suggests that the two NCLDVs might be characteristic viral groups for water currents.

Caudovirales in NSVs might be characteristic of warm Atlantic water masses.

Caudovirales were the dominant dsDNA viruses in the NSVs (Fig. 1B), which is consistent with those found in the Global Ocean Virome 2.0 for the Indian Ocean and Monterey Bay (10, 48, 49). The lower abundance of *Caudovirales* in the Nordic Seas compared to the viromes in tropical and subtropical waters might be explained by a higher abundance of NCLDVs (Fig. 1B).

The dominant viruses in the NSVs were the viruses infecting the most abundant bacterial groups (Fig. 2), such as *Pelagibacter* phage HTVC008M infecting SAR11 clade (50–52), *Puniceispirillum* phage HMO-2011 infecting SAR116 clade (53, 54), *Cellulophaga* phage infecting *Bacteroidetes* (55), and cyanophages infecting cyanobacteria (Fig. 2 and Fig. 5A). This is similar to the results from other marine viromes (10–12, 18, 20, 31, 46). Since SAR11 and SAR116 clades contain a number of genes associated with dimethylsulfide production and dimethylsulfoniopropionate degradation and play important roles in the global sulfur cycle (54, 56), their phages might indirectly influence the sulfur cycle by infecting and lysing both clades. *Cellulophaga* phages, which correspond to viral cluster 5 (VC_5) and VC_47 in the Tara Ocean expedition, were recently reported as diverse, ubiquitous, and abundant phage genera in global oceans (11, 55, 57). Cyanophages are unexpectedly abundant in the NSV data set (Fig. 1B and Fig. 2), especially for the cyanophages infecting *Prochlorococcus*. *Prochlorococcus*, although the most abundant autotrophic bacterial group in the tropical and subtropical oceans, is uncommon or absent from polar regions (58), although it has been identified as far north as 60°N in the North Atlantic Ocean (59, 60). Here, all seven stations were located farther north than 69°N (Fig. 1). Although *Prochlorococcus* was not abundant in the study area (Fig. 5A), *Prochlorococcus* phages were both detected and abundant (Fig. 2). Two possible hypotheses are suggested for this observation: (i) *Prochlorococcus* phages can be transported from the midlatitude North Atlantic by the warm North Atlantic Drift, and these could persist in the Nordic Seas without their corresponding host cells; and (ii) *Prochlorococcus* phages (especially myoviruses) might have a broader host range and be able to infect other groups of cyanobacteria, such as *Synechococcus*. Large abundances of *Synechococcus* CC9902, which is a *Parasynechococcus* closed toward *Prochlorococcus*, have been detected in the Nordic Seas (Fig. 5B), although it is uncommon in the Arctic Ocean (34, 61).

Compared to the NCLDVs, the proportion of *Caudovirales* (*Myoviridae*, *Podoviridae*, and *Siphoviridae*) in the NSV data set significantly increased from the colder Arctic waters to the warmer Atlantic waters (Fig. 4 and Fig. 1C). This pattern might reflect the distribution and abundance of bacteriophage sequences in the warm Atlantic water (10, 17). Viruses infecting specific host groups, e.g., *Bacillus* phages, *Synechococcus* phages, *Cellulophaga* viruses, and *Pseudomonas* phages, increased from the colder Arctic waters to the warmer Atlantic waters (Fig. 2C). This might be the result of the dominance of the corresponding bacterial groups in the warmer Atlantic waters (Fig. 5A) (62, 63). These *Caudovirales* groups might thus be a characteristic of viral groups associated with the warmer Atlantic waters (Fig. 1 and 2).

Different latitudinal patterns of surface viral diversity in the global ocean. To compare the viral diversity of NSVs with the surface viromes of GOV2.0, two different latitudinal patterns of surface water viral diversity were observed in the global ocean (Fig. 3). In the open ocean areas of the global ocean, the Shannon H' diversity of marine viruses was highest at lower latitudes and decreased poleward (Fig. 3B), a pattern similar to that for most flora and fauna from both terrestrial and marine environments (64, 65) but different from the viral diversity of the TOPC (10). Gregory et al. reported that viral diversity increased toward the Arctic Ocean and the Arctic Ocean could be divided into areas of high diversity (ARC-H, mainly located in the Pacific Arctic region, the Arctic Archipelago, and the Davis-Baffin Bay) and low diversity (ARC-L, mainly located in the Kara-Laptev Sea, the Barents Sea, and sub-arctic areas, including Nordic Seas; Fig. 3) (10). This is consistent with their biogeography and

the observed Redfield ratio (N^*), which reflects deficiency in dissolved inorganic nitrogen and acts as a geochemical tracer of Pacific and Atlantic water masses (10). These authors hypothesized that the strong denitrification in the Bering Strait (66) and the oligotrophy in the Beaufort Gyre might increase the host diversity (67) and indirectly cause the high viral diversity. However, we hypothesize here that the water depth, water currents, and primary production might be additional parameters influencing the viral diversity in the Arctic Ocean. Most of the low-diversity regions are located above deep open ocean areas (Fig. 3A) and are associated with the warm Norwegian Atlantic Current (25) and high primary production (see Table S1) (68). The warm Norwegian Atlantic current might raise primary production levels in summer in the Nordic Seas and Barents Sea (68). High primary production might decrease phytoplankton diversity (69) and indirectly cause the low viral diversity in the summer (Fig. 3A).

Methodological considerations. Almost all of the contigs in the NSV data set belonged to dsDNA viruses. This is reasonable since the extracted viral DNA was sequenced directly without amplification (MDA/LASL), which excludes the bias of preferential amplification of ssDNA viruses (70, 71). The NSV samples were concentrated using the routine method of TFF method after filtering through 0.22- μm -pore-size filters, which excludes most of the prokaryotic and eukaryotic cells and viruses larger than 0.22 μm (11, 17, 31); this might underestimate the proportion of NCLDVs in the NSV data set (31, 72, 73).

Conclusion. The viral community of the Nordic Seas had specific features that were closely associated with the hydrological characteristics. Although *Caudovirales* was the dominant viral group, NCLDVs also comprised a large component of the known sequences, which was consistent with other polar viromes. Several viral groups, such as NCLDVs, showed significant differences between water masses. However, unknown viruses are still abundant in NSVs, especially in the water near the Arctic cold current. Temperature, latitude, and flow speed were the main drivers of microbial community structure and codrivers of the viral community structure of NSVs. Correlation of viruses and hosts indicated that abundance of viruses varied with the growth and decline of hosts, reflecting the variation of water masses and currents. Future studies of both viral and host community structures at a large number stations of Nordic Seas, especially from the cold waters near Arctic Ocean, could give us a clearer linkage between viruses, host cells, and water masses and currents.

MATERIALS AND METHODS

Sampling. Sampling stations are shown in Fig. 1. These stations include (i) warm basin (WB), with WB1 (Norwegian Basin) and WB2 (Lofton Basin), which are surrounded by the warm Faroe Current; (ii) warm current (WC), with WC1 and WC2, which are located on a branch of the Faroe Current (warm current); (iii) mixed water (MW), with MW1 and MW2, which are comprised of mixed water with both warm current and cold current influences near Jan Mayen Island; and (iv) cold basin (CB), which is surrounded by a cold current (Greenland Basin). Portions (150 liters) of surface water from each site were sampled by using a submerged pump in June 2015. The samples were immediately filtered through a 300-mm-diameter cellulose membrane filter with a 3- μm pore size; this was followed by filtration through a 0.22- μm membrane filter to remove the bigger cellular organisms, such as zooplankton, phytoplankton, and bacteria. Free viruses in the final filtrate were concentrated into 500 ml with a middle-size tangential flow filtration unit (membrane package: Pellicon 2 Cassette, Biomax 50 kDa; polyethersulfone) and then concentrated further into 10-ml plastic microtubes by small-size tangential flow filtration (membrane package: Pellicon XL Cassette, Biomax 50 kDa; polyethersulfone) (31, 74). The highly concentrated samples were instantly frozen in liquid nitrogen and stored at -80°C for further DNA extraction (31, 75). It took about 4 to 5 h from collecting the sample to freezing the sample.

Environmental factors. Temperature, conductivity, salinity, dissolved oxygen, and pH were measured using a CTD (YSI QUATRO, Professional). The site map and surface temperatures were plotted by Ocean Data View (version 4.6.4) (76). Chlorophyll *a*, photosynthetically active radiation (PAR), POC, and the depth of the euphotic zone (Z_{eu}) of each station, during the sampling month and year, were retrieved from the monthly averaged MODIS/Aqua level 3 data with a resolution ratio of 4 km \times 4 km (<http://oceansatellite.nasa.gov>). The primary production of each station during the sampling month and year was retrieved from the standard product of the Ocean Productivity website (<http://www.science.oregonstate.edu/ocean.productivity/>), which determined estimates using the vertically generalized production model (77).

Flow speed between two stations (Geostrophic current flow velocity) was calculated by the dynamic height method, according the following equation (78):

$$V_1 - V_0 = \frac{1}{fL} \Delta D \approx \frac{1}{fL} \left(\sum_{p_1}^{p_0} \bar{\alpha}_B \Delta p - \sum_{p_1}^{p_0} \bar{\alpha}_A \Delta p \right)$$

where V is the flow speed, f is the Coriolis parameter, L is the distance between two stations, ΔD is the

dynamic height between the two isobaric surfaces, Δp is the pressure difference, and $\bar{\alpha}$ is the average value of specific volume in scale Δp .

DNA extraction. Genomic DNA of the bacterial and microbial eukaryotic community was extracted using an E.Z.N.A. soil DNA kit (Omega Bio-Tek, Norcross, GA) from the 0.2- μm cellulose membrane filter. The agarose gel (1%) was then used to check the extract. DNA concentration and purity were determined with NanoDrop 2000 UV-vis spectrophotometer (Thermo Scientific, Wilmington, DE). The DNA solution of the microbial and microbial eukaryotic samples was stored at -80°C .

The frozen viral metagenomic samples were bathed at 37°C to melt them. Polyethylene glycol (PEG; 10% final concentration) and NaCl (0.6% final concentration) were added to the viral subsample, followed by incubation at 4°C in the dark for 24 h. The mixed sample was centrifuged at 10,500 r/min ($11,827 \times g$) for 40 min at 4°C , using an H3-20KR with rotor 1 (20,500 rpm) and rotor 9 (12,000 rpm), and then resuspended in 300 μl of SM buffer. Next, 100 μl of KCl (1 M) solution was added, followed by incubation on ice for 30 min. The samples were then centrifuged at 12,000 rpm ($15,450 \times g$) for 10 min at 4°C . Then, 10 μl of proteinase K and 20 μl of 10% SDS solution were added, and the samples were placed in a 56°C bath for 1 h. DNA was extracted using phenol-chloroform and precipitated with ethanol. The viral DNA solution was stored at -80°C before sequencing (79). The DNA of cellular microorganisms in the 0.22- μm membrane was extracted using a FastDNA spin kit for soil for 16S and 18S sequencing.

PCR amplification, sequencing, and sequence processing of 16S/18S rRNA genes. Triplicate samples of the hypervariable region V3-V4 of the bacterial 16S rRNA genes were amplified using the primer pairs 338F (5'-ACTCTACTACGGGAGGAGCAG-3') and 806R (5'-GGACTACHVGGGTWCTAAT-3'). Triplicate samples of the hypervariable region V2-V3 of the microbial eukaryotic 18S rRNA genes were amplified with 82F (5'-GAAACTGCGAATGGCTC-3') and 516R (5'-ACCAGACTTGCCCTCC-3'). The PCRs were performed by using an ABI GeneAmp 9700 PCR thermocycler (ABI, Oakland, CA), and the purified PCR products were then sequenced using an Illumina MiSeq PE300 platform (Illumina, San Diego, CA).

The raw 16S and 18S rRNA gene reads were quality controlled by fastp (version 0.20.0) and merged using FLASH software (version 1.2.7) (80). Low-quality reads that contained unknown nucleotides (Ns) and reads that were truncated shorter than 50 bp with a 50-bp sliding window were discarded. When their overlapping sequences were longer than 10 bp, the reads were assembled producing a mismatch ratio of the overlap region that was lower than 0.2. Quality-controlled sequences were analyzed using UPARSE (version 7.1) (81) according to the standard operating procedure. The chimeric sequences were filtered. Sequences were clustered in OTUs at 97% sequence divergence using the furthest neighbor algorithm and classified with the SILVA 138 SSU Ref NR database (<https://www.arb-silva.de/>).

Virome library construction, sequencing, and assembly. Library construction and sequencing were implemented by Annoroad Gene Technology (Beijing) Co., Ltd. DNA samples were inspected by performing the following steps: (i) 1% agarose gel detection of degradation and contamination, (ii) nano-photometer spectrophotometer DNA purification detection, and (iii) Qubit 2.0 fluorometer DNA concentration detection. An ultrasonic processor was used to carry out qualified DNA fragmentation. The mean length of the inserted DNA fragments was ~ 500 bp. Terminal repair, adenine addition, sequence adapter addition, purification, and PCR amplification were then completed for library preparation. The library concentration was diluted to 1 $\text{ng}/\mu\text{l}$. The insert size of the library was verified by using an Agilent 2100. Finally, Q-PCR analyses were undertaken to ensure the effective quantitative concentrations of the library. High-throughput sequencing was performed using an Illumina HiSeq 2500 (paired end sequencing, 2×250 bp). High-quality reads were selected from raw reads (clean data rate > 0.90). Paired-end reads were removed based on one or more of the following conditions: (i) the reads contained $> 5\%$ N, (ii) the reads included low-quality reads (50% read length, $Q \leq 5$), and (iii) there was an adapter present.

Assembly was performed using the Velvet program (version 1.2.10); the parameters were set as "velvet file_folder 45 -shortPaired -fastq -separate *_1.fq *_2.fq, velvetg file_folder -cov_cutoff 3 -ins_length 450 -exp_cov auto" (82, 83). After assembly, contigs shorter than 1,000 bp were filtered out. The sequence information was extracted by a program written in Perl and Python.

Taxonomic composition analysis. The assembled sequences were aligned with the viral reference sequence protein database (84) on the MetaVir website by using BLASTx (E value $< 10^{-3}$; the composition type was the best BLAST hit number) (85). Viral taxonomic composition analysis was computed by GAAS (86). The average abundance was calculated as DNA reads per kilobase of the transcript (gene) per million reads mapped (DNA-RPKM; this is equal to the number of reads mapped to the contig and normalized by the contig length and per million mappable reads) (87) using Bowtie2 (version 2.1.0) (88) and SAMtools (version 1.1) (89). The abundances of all contigs annotated into one viral population were added as the viral population's abundance. The 10 most abundant viral populations and specific viral groups were ranked by their total relative abundance of all seven stations. Specific viral groups were clustered by the classification of their corresponding hosts.

Functional analysis and phylogenetic analysis. ORF prediction was completed using MetaGeneMark (version 2.10) (90). The ORFs were aligned to the COG data set for functional gene annotation based on BLASTp (E value $< 10^{-3}$) on the WebMGA website (91). The cd-hit program (version 4.6) was used to remove redundant ORFs (95% identity) (10, 92, 93). Two conserved domains, viral terminase large subunit (*terL*, PF03237) for *Caudovirales* and the major capsid protein (*mcp*, PF04451) for the NCLDVs, were searched in nonredundant ORFs by BLASTp (E value $< 10^{-3}$). The aligned ORFs with fewer than 150 amino acids or a total score of < 300 (*terL*)/150 (*mcp*) were filtered out. Maximum-likelihood phylogenetic trees were constructed using the Jones-Taylor-Thornton substitution model and modified by using MEGA (version 6.06) (94). The bootstrap (100 iterations) method was used to test phylogeny.

Comparison of viromes and statistical analysis. Viromes were compared to each other on MetaVir (32, 86). The dissimilarity matrix was based on the tetranucleotide frequency variation (32). The Bray-Curtis distance was then calculated using R software (version 3.3.0). A dendrogram was plotted using the CLUSTER package (95).

To compare the alpha-diversity of seven NSVs to other marine viromes, the Shannon H' diversity of the surface viromes (61 samples) of the GOV2.0 data set was collected from the iVirus, which is a sub-project of iMicrobe (<http://data.imicrobe.us/>).

One-way ANOVA was used to analyze the differences among different groups of factors (SPSS). Linear regression and Pearson correlation analysis were used to analyze the relationship between viruses and environmental factors. Cubic regression was used to analyze the relationship between the Shannon H' diversity of the 68 surface viromes and latitude. Bioenv analysis and the Mantel test were performed using the VEGAN package (96).

Network analysis. Potential interactions between viral populations and hosts were determined through the modeling of the viral and host community in a network structure (97–99). All possible pairwise correlations (Spearman) between those viral populations and host OTUs were calculated using the “psych” package in R software (100). Only robust ($R > 0.8$ or $R < -0.8$) and statistically significant ($P < 0.01$) correlations were included into the network analysis. Network visualization was performed with Gephi version 0.9.2 (101). In a network graph, each node represents a viral population or host OTU indicating an individual taxon. The edge of two nodes represents positive or negative correlations between the viral populations and host OTUs.

Data availability. All reads in this study were submitted to NCBI Sequence Read Archive (SRA). The SRA accession numbers of viromes are [SRR4131952](https://www.ncbi.nlm.nih.gov/sra/SRR4131952), [SRR4131953](https://www.ncbi.nlm.nih.gov/sra/SRR4131953), [SRR4131954](https://www.ncbi.nlm.nih.gov/sra/SRR4131954), [SRR4131955](https://www.ncbi.nlm.nih.gov/sra/SRR4131955), [SRR4131956](https://www.ncbi.nlm.nih.gov/sra/SRR4131956), [SRR4131957](https://www.ncbi.nlm.nih.gov/sra/SRR4131957), and [SRR4131958](https://www.ncbi.nlm.nih.gov/sra/SRR4131958). The project accession number of 18S rRNA genes is [PRJNA703440](https://www.ncbi.nlm.nih.gov/sra/PRJNA703440). The project accession number of 16S rRNA genes is [PRJNA703722](https://www.ncbi.nlm.nih.gov/sra/PRJNA703722).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.4 MB.

ACKNOWLEDGMENTS

We sincerely appreciate the captain and crews of STALBAS during the summer 2015 Nordic cruise. We thank the onboard scientific researchers from Polar Oceanic Laboratory of Ocean University of China and the Collaborative Innovation Center for Distant-water Fisheries of Shanghai Ocean University. We are grateful for the support of the high-performance server of the Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao). We appreciate the computing resources provided by IEMB-1, a high-performance computing cluster operated by the Institute of Evolution and Marine Biodiversity. We thank the three anonymous reviewers for their constructive comments and suggestions.

We declare that we have no conflicts of interest regarding this study.

This study was supported by the National Key Research and Development Program of China (2018YFC1406704 and 2017YFA0603200), the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao, China; 2018SDKJ0406-6), the Natural Science Foundation of China (41976117, 42120104006, 42176111, 41676178, 41906126, 42176149, and 41806131), the Fundamental Research Funds for the Central Universities (201812002 and 202072002), the response and feedback of the Southern Ocean to climate change (RFSOCC2020-2025), the Global Change and Air-Sea Interface Project supported by the State Oceanic Administration, People's Republic of China, cruises for an oceanographic survey of South-central Western Pacific in Autumn (GASI-02-PAC-ST-Msaut), the Scientific and Technological Innovation Project Financially Supported by Pilot Qingdao National Laboratory for Marine Science and Technology (2016ASKJ14), the Center for High Performance Computing and System Simulation, the Pilot National Laboratory for Marine Science and Technology (Qingdao), a Key R&D Project in Shandong Province (2019GHY112079), and the MEL Visiting Fellowship Program (MELRS1511).

REFERENCES

1. Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* 64:69–114. <https://doi.org/10.1128/MMBR.64.1.69-114.2000>.
2. Suttle CA. 2005. Viruses in the sea. *Nature* 437:356–361. <https://doi.org/10.1038/nature04160>.
3. Suttle CA. 2007. Marine viruses: major players in the global ecosystem. *Nat Rev Microbiol* 5:801–812. <https://doi.org/10.1038/nrmicro1750>.
4. Brum JR, Sullivan MB. 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat Rev Microbiol* 13:147–159. <https://doi.org/10.1038/nrmicro3404>.

5. Suttle CA. 2016. Environmental microbiology: viral diversity on the global stage. *Nat Microbiol* 1:16205–16202. <https://doi.org/10.1038/nrmicrobiol.2016.205>.
6. Jiao N, Herndl GJ, Hansell DA, Benner R, Kattner G, Wilhelm SW, Kirchman DL, Weinbauer MG, Luo T, Chen F, Azam F. 2010. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat Rev Microbiol* 8:593–599. <https://doi.org/10.1038/nrmicro2386>.
7. Jiao N, Cai R, Zheng Q, Tang K, Liu J, Jiao F, Wallace D, Chen F, Li C, Amann R, Benner R, Azam F. 2018. Unveiling the enigma of refractory carbon in the ocean. *National Science Rev* 5:459–463. <https://doi.org/10.1093/nsr/nwy020>.
8. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR, Coelho LP, Espinoza JC, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Tara Oceans Consortium Coordinators, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532:465–470. <https://doi.org/10.1038/nature16942>.
9. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. 2002. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 99:14250–14255. <https://doi.org/10.1073/pnas.202488399>.
10. Gregory AC, Zayed AA, Conceicao-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Dominguez-Huerta G, Ferland J, Kandels S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay JE, Vik D, Tara Oceans Coordinators, Babin M, Bowler C, Culley AI, de Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB. 2019. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177:1109–1123. <https://doi.org/10.1016/j.cell.2019.03.040>.
11. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vague D, Tara Oceans Coordinators, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693. <https://doi.org/10.1038/nature19366>.
12. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. *Nature* 536:425–430. <https://doi.org/10.1038/nature19094>.
13. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, et al. 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat Biotechnol* 37:29–37. <https://doi.org/10.1038/nbt.4306>.
14. Roux S, Paez-Espino D, Chen IA, Palaniappan K, Ratner A, Chu K, Reddy TBK, Nayfach S, Schulz F, Call L, Neches RY, Woyke T, Ivanova NN, Eloe-Fadrosh EA, Kyrpides NC. 2021. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res* 49:D764–D775. <https://doi.org/10.1093/nar/gkaa946>.
15. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denev VJ, McMahon KD, Konstantinidis KT, Eloe-Fadrosh EA, Kyrpides NC, Woyke T. 2020. Giant virus diversity and host interactions through global metagenomics. *Nature* 578:432–436. <https://doi.org/10.1038/s41586-020-1957-x>.
16. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. 2020. Dynamic genome evolution and complex Virocell metabolism of globally-distributed giant viruses. *Nat Commun* 11:1710. <https://doi.org/10.1038/s41467-020-15507-2>.
17. Brum JR, Ignacio-Espinoza JC, Roux S, Doucier G, Acinas SG, Alberti A, Chaffron S, Cruaud C, de Vargas C, Gasol JM, Gorsky G, Gregory AC, Guidi L, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Poulos BT, Schwenck SM, Speich S, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Tara Oceans Coordinators, Bork P, Bowler C, Sunagawa S, Wincker P, Karsenti E, Sullivan MB. 2015. Ocean plankton: patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. <https://doi.org/10.1126/science.1261498>.
18. Liang Y, Wang L, Wang Z, Zhao J, Yang Q, Wang M, Yang K, Zhang L, Jiao N, 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. <https://doi.org/10.3389/fmicb.2019.01951>.
19. De Corte D, Martinez JM, Cretoiu MS, Takaki Y, Nunoura T, Sintez E, Herndl GJ, Yokokawa T. 2019. Viral communities in the global deep ocean conveyor belt assessed by targeted viromics. *Front Microbiol* 10:1801. <https://doi.org/10.3389/fmicb.2019.01801>.
20. Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle CA, Rohwer F. 2006. The marine viromes of four oceanic regions. *PLoS Biol* 4:e368. <https://doi.org/10.1371/journal.pbio.0040368>.
21. Jansen E, Christensen JH, Dokken T, Nisancioglu KH, Vinther BM, Capron E, Guo C, Jensen MF, Langen PL, Pedersen RA, Yang S, Bentsen M, Kjær HA, Sadatzki H, Sessford E, Stendel M. 2020. Past perspectives on the present era of abrupt Arctic climate change. *Nat Clim Chang* 10:714–721. <https://doi.org/10.1038/s41558-020-0860-7>.
22. Tsubouchi T, Våge K, Hansen B, Larsen KMH, Østerhus S, Johnson C, Jónsson S, Valdimarsson H. 2021. Increased ocean heat transport into the Nordic Seas and Arctic Ocean over the period 1993–2016. *Nat Clim Chang* 11:21–26. <https://doi.org/10.1038/s41558-020-00941-3>.
23. Swift JH, Aagaard K. 1981. Seasonal transitions and water mass formation in the Iceland and Greenland seas. *Deep Sea Res Part A Oceanogr Res Pap* 28:1107–1129. [https://doi.org/10.1016/0198-0149\(81\)90050-9](https://doi.org/10.1016/0198-0149(81)90050-9).
24. Hansen B, Østerhus S. 2000. North Atlantic–Nordic seas exchanges. *Progress in Oceanography* 45:109–208. [https://doi.org/10.1016/S0079-6611\(99\)00052-X](https://doi.org/10.1016/S0079-6611(99)00052-X).
25. Rossby T, Prater M, Søiland H. 2009. Pathways of inflow and dispersion of warm waters in the Nordic seas. *J Geophys Res* 114. <https://doi.org/10.1029/2008JC005073>.
26. He Y, Zhao J. 2011. Distributions and seasonal variations of fronts in GIN Seas. *Adv Earth Sci* 26:1–79.
27. Wang X, Zhao J, Li T, Zhong W, Jiao Y. 2015. Hydrographic features of the Norwegian Sea and the Greenland Sea in summer 2012. *Advances in Earth Science* 30:346–356.
28. Børsheim KY, Milutinović S, Drinkwater KF. 2014. TOC and satellite-sensed chlorophyll and primary production at the Arctic Front in the Nordic Seas. *J Mar Syst* 139:373–382. <https://doi.org/10.1016/j.jmarsys.2014.07.012>.
29. Koonin EV, Yutin N. 2019. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv Virus Res* 103:167–202. <https://doi.org/10.1016/bs.aivir.2018.09.002>.
30. Abergel C, Claverie J-M. 2014. *Pithovirus sibericum*: awakening of a giant virus of more than 30,000 years. *Med Sci (Paris)* 30:329–331. <https://doi.org/10.1051/medsci/20143003022>.
31. Gong Z, Liang Y, Wang M, Jiang Y, Yang Q, Xia J, Zhou X, You S, Gao C, Wang J, He J, Shao H, McMinn A. 2018. Viral diversity and its relationship with environmental factors at the surface and deep sea of Prydz Bay, Antarctica. *Front Microbiol* 9:2981. <https://doi.org/10.3389/fmicb.2018.02981>.
32. Willner D, Thurber RV, Rohwer F. 2009. Metagenomic signatures of 86 microbial and viral metagenomes. *Environ Microbiol* 11:1752–1766. <https://doi.org/10.1111/j.1462-2920.2009.01901.x>.
33. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie J-M, Abergel C. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. *Science* 341:281–286. <https://doi.org/10.1126/science.1239181>.
34. Sandaa RA, J ES, Olesin E, L Paulsen M, Larsen A, Bratbak G, Ray JL. 2018. Seasonality drives microbial community structure, shaping both eukaryotic and prokaryotic host(–)viral relationships in an arctic marine ecosystem. *Viruses* 10:715. <https://doi.org/10.3390/v10120715>.
35. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Couté Y, Rivkina E, Abergel C, Claverie J-M. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proc Natl Acad Sci U S A* 111:4274–4279. <https://doi.org/10.1073/pnas.1320670111>.
36. Gran-Stadniczenko S, Krabberod AK, Sandaa RA, Yau S, Egge E, Edvardsen B. 2019. Seasonal dynamics of algae-infecting viruses and their inferred interactions with protists. *Viruses* 11:1043. <https://doi.org/10.3390/v11111043>.
37. Liu Q, Jiang Y, Wang Q, Wang M, McMinn A, Zhao Q, Xue C, Wang X, Dong J, Yu Y, Han Y, Zhao J. 2019. Using picoeukaryote communities to indicate the spatial heterogeneity of the Nordic Seas. *Ecological Indicators* 107:105582. <https://doi.org/10.1016/j.ecolind.2019.105582>.
38. Hoppe CJM, Flintrop CM, Rost B. 2018. The Arctic picoeukaryote *Micromonas pusilla* benefits synergistically from warming and ocean acidification. *Biogeosciences* 15:4353–4365. <https://doi.org/10.5194/bg-15-4353-2018>.
39. van Baren MJ, Bachy C, Reistetter EN, Purvine SO, Grimwood J, Sudek S, Yu H, Poirot C, Deerinck TJ, Kuo A, Grigoriev IV, Wong C-H, Smith RD, Callister SJ, Wei C-L, Schmutz J, Worden AZ. 2016. Evidence-based green algal genomics

- reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics* 17:267. <https://doi.org/10.1186/s12864-016-2585-6>.
40. Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau M-J, Terrado R, Potvin M, Massana R, Pedrós-Alió C. 2007. Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic Seas. *J Phycology* 43:78–89. <https://doi.org/10.1111/j.1529-8817.2006.00310.x>.
 41. Fischer MG, Allen MJ, Wilson WH, Suttle CA. 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc Natl Acad Sci U S A* 107:19508–19513. <https://doi.org/10.1073/pnas.1007615107>.
 42. Claverie J-M, Abergel C. 2018. *Mimiviridae*: an expanding family of highly diverse large dsDNA viruses infecting a wide phylogenetic range of aquatic eukaryotes. *Viruses* 10:506. <https://doi.org/10.3390/v10090506>.
 43. Monier A, Claverie J-M, Ogata H. 2008. Taxonomic distribution of large DNA viruses in the sea. *Genome Biol* 9:R106. <https://doi.org/10.1186/gb-2008-9-7-r106>.
 44. Williamson SJ, Rusch DB, Yooseph S, Halpern AL, Heidelberg KB, Glass JJ, Andrews-Pfannkoch C, Fadrosch D, Miller CS, Sutton G, Frazier M, Venter JC. 2008. The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* 3:e1456. <https://doi.org/10.1371/journal.pone.0001456>.
 45. Fischer MG, Hackl T. 2016. Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature* 540:288–291. <https://doi.org/10.1038/nature20593>.
 46. Yang Q, Gao C, Jiang Y, Wang M, Zhou X, Shao H, Gong Z, McMinn A. 2019. Metagenomic characterization of the viral community of the South Scotia Ridge. *Viruses* 11:95. <https://doi.org/10.3390/v11020095>.
 47. Moreau H, Piganeau G, Desdevises Y, Cooke R, Derelle E, Grimsley N. 2010. Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *J Virol* 84:12555–12563. <https://doi.org/10.1128/JVI.01123-10>.
 48. Williamson SJ, Allen LZ, Lorenzi HA, Fadrosch DW, Bami D, Thiagarajan M, McCrow JP, Tovchigrechko A, Yooseph S, Venter JC. 2012. Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS One* 7:e42047. <https://doi.org/10.1371/journal.pone.0042047>.
 49. Steward GF, Preston CM. 2011. Analysis of a viral metagenomic library from 200 m depth in Monterey Bay, California, constructed by direct shotgun cloning. *Virol J* 8:287–287. <https://doi.org/10.1186/1743-422X-8-287>.
 50. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M, Deerinck T, Sullivan MB, Giovannoni SJ. 2013. Abundant SAR11 viruses in the ocean. *Nature* 494:357–360. <https://doi.org/10.1038/nature11921>.
 51. Brown MV, Lauro FM, DeMaere MZ, Muir L, Wilkins D, Thomas T, Riddle MJ, Fuhrman JA, Andrews-Pfannkoch C, Hoffman JM, McQuaid JB, Allen A, Rintoul SR, Cavicchioli R. 2012. Global biogeography of SAR11 marine bacteria. *Mol Syst Biol* 8:595. <https://doi.org/10.1038/msb.2012.28>.
 52. Kraemer S, Ramachandran A, Colatristo D, Lovejoy C, Walsh DA. 2020. Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *ISME J* 14:79–90. <https://doi.org/10.1038/s41396-019-0499-4>.
 53. Kang I, Oh HM, Kang D, Cho JC. 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. <https://doi.org/10.1073/pnas.1219930110>.
 54. Zeng Y-X, Qiao Z-Y, Yu Y, Li H-R, Luo W. 2016. Diversity of bacterial dimethylsulfoniopropionate degradation genes in surface seawater of Arctic Kongsfjorden. *Sci Rep* 6:33031–33039. <https://doi.org/10.1038/srep33031>.
 55. Holmfeldt K, Solonenko N, Howard-Varona C, Moreno M, Malmstrom RR, Blow MJ, Sullivan MB. 2016. Large-scale maps of variable infection efficiencies in aquatic *Bacteroidetes* phage-host model systems. *Environ Microbiol* 18:3949–3961. <https://doi.org/10.1111/1462-2920.13392>.
 56. Howard EC, Sun S, Biers EJ, Moran MA. 2008. Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ Microbiol* 10:2397–2410. <https://doi.org/10.1111/j.1462-2920.2008.01665.x>.
 57. Holmfeldt K, Solonenko N, Shah M, Corrier K, Riemann L, Verberkmoes NC, Sullivan MB. 2013. Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci U S A* 110:12798–12803. <https://doi.org/10.1073/pnas.1305956110>.
 58. Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao N, Karl DM, Li WK, Lomas MW, Veneziano D, Vera CS, Vrugt JA, Martiny AC. 2013. Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* 110:9824–9829. <https://doi.org/10.1073/pnas.1307701110>.
 59. Partensky F, Garczarek L. 2010. *Prochlorococcus*: advantages and limits of minimalism. *Annu Rev Mar Sci* 2:305–331. <https://doi.org/10.1146/annurev-marine-120308-081034>.
 60. Buck KR, Chavez FP, Campbell L. 1996. Basin-wide distributions of living carbon components and the inverted trophic pyramid of the central gyre of the North Atlantic Ocean, summer 1993. *Aquat Microb Ecol* 10:283–298. <https://doi.org/10.3354/ame010283>.
 61. Cottrell MT, Kirchman DL. 2009. Photoheterotrophic microbes in the Arctic Ocean in summer and winter. *Appl Environ Microbiol* 75:4958–4966. <https://doi.org/10.1128/AEM.00117-09>.
 62. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Tara Oceans coordinators. 2015. Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348:1261359. <https://doi.org/10.1126/science.1261359>.
 63. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh HJ, Cuenca M, Field CM, Coelho LP, Cruaud C, Engelen S, Gregory AC, Labadie K, Marec C, Pelletier E, Royo-Llonch M, Roux S, Sanchez P, Uehara H, Zayed AA, Zeller G, Carmichael M, Dimier C, Ferland J, Kandels S, Picheral M, Pisarev S, Poulain J, Tara Oceans Coordinators, Acinas SG, Babin M, Bork P, Bowler C, de Vargas C, Guidi L, Hingamp P, Iudicone D, Karp-Boss L, Karsenti E, Ogata H, Pesant S, Speich S, Sullivan MB, Wincker P, Sunagawa S. 2019. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* 179:1068–1083. <https://doi.org/10.1016/j.cell.2019.10.014>.
 64. Mannion PD, Upchurch P, Benson RBJ, Goswami A. 2014. The latitudinal biodiversity gradient through deep time. *Trends Ecol Evol* 29:42–50. <https://doi.org/10.1016/j.tree.2013.09.012>.
 65. Miraldo A, Li S, Borregaard MK, Flórez-Rodríguez A, Gopalakrishnan S, Rizvanovic M, Wang Z, Rahbek C, Marske KA, Nogués-Bravo D. 2016. An Anthropocene map of genetic diversity. *Science* 353:1532–1535. <https://doi.org/10.1126/science.aaf4381>.
 66. Devol AH, Codispoti LA, Christensen JP. 1997. Summer and winter denitrification rates in western Arctic shelf sediments. *Continental Shelf Res* 17:1029–1033. [https://doi.org/10.1016/S0278-4343\(97\)00003-4](https://doi.org/10.1016/S0278-4343(97)00003-4).
 67. Li WKW, McLaughlin FA, Lovejoy C, Carmack EC. 2009. Smallest algae thrive as the Arctic Ocean freshens. *Science* 326:539–539. <https://doi.org/10.1126/science.1179798>.
 68. Pabi S, van Dijken GL, Arrigo KR. 2008. Primary production in the Arctic Ocean, 1998–2006. *J Geophys Res* 113. <https://doi.org/10.1029/2007JC004578>.
 69. Chen B, Smith SL, Wirtz KW. 2019. Effect of phytoplankton size diversity on primary productivity in the North Pacific: trait distributions under environmental variability. *Ecol Lett* 22:56–66. <https://doi.org/10.1111/ele.13167>.
 70. Marine R, McCarren C, Vorrasane V, Nasko D, Crowgwey E, Polson SW, Wommack KE. 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. <https://doi.org/10.1186/2049-2618-2-3>.
 71. Duhaime MB, Sullivan MB. 2012. Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* 434:181–186. <https://doi.org/10.1016/j.virol.2012.09.036>.
 72. Hurwitz BL, Sullivan MB. 2013. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS One* 8:e57355. <https://doi.org/10.1371/journal.pone.0057355>.
 73. Winter C, Garcia JA, Weinbauer MG, DuBow MS, Herndl GJ. 2014. Comparison of deep-water viromes from the Atlantic Ocean and the Mediterranean Sea. *PLoS One* 9:e100600. <https://doi.org/10.1371/journal.pone.0100600>.
 74. Sun G, Xiao J, Wang H, Gong C, Pan Y, Yan S, Wang Y. 2014. Efficient purification and concentration of viruses from a large body of high turbidity seawater. *MethodsX* 1:197–206. <https://doi.org/10.1016/j.mex.2014.09.001>.
 75. Cai L, Yang Y, Jiao N, Zhang R. 2015. Evaluation of tangential flow filtration for the concentration and separation of bacteria and viruses in contrasting marine environments. *PLoS One* 10:e0136741. <https://doi.org/10.1371/journal.pone.0136741>.
 76. Schlitzer R. 2021. Ocean Data View. <https://odv.awi.de>.
 77. Behrenfeld MJ, Falkowski PG. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol Oceanogr* 42:1–20. <https://doi.org/10.4319/lo.1997.42.1.0001>.
 78. Ye A, Li F. 1992. Physical oceanography. Qingdao Ocean University Press, Qingdao, China.

79. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. 2009. Laboratory procedures to generate viral metagenomes. *Nat Protoc* 4:470–483. <https://doi.org/10.1038/nprot.2009.10>.
80. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
81. Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10:996–998. <https://doi.org/10.1038/nmeth.2604>.
82. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
83. Namiki T, Hachiya T, Tanaka H, Sakakibara Y. 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* 40:e155. <https://doi.org/10.1093/nar/gks678>.
84. O'Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44:D733–D745. <https://doi.org/10.1093/nar/gkv1189>.
85. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
86. Roux S, Tournayre J, Mahul A, Debroas D, Enault F. 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics* 15:76. <https://doi.org/10.1186/1471-2105-15-76>.
87. Calusinska M, Marynowska M, Goux X, Lentzen E, Delfosse P. 2016. Analysis of ds DNA and RNA viromes in methanogenic digesters reveals novel viral genetic diversity. *Environ Microbiol* 18:1162–1175. <https://doi.org/10.1111/1462-2920.13127>.
88. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
89. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
90. Zhu W, Lomsadze A, Borodovsky M. 2010. *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Res* 38:e132. <https://doi.org/10.1093/nar/gkq275>.
91. Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12:444. <https://doi.org/10.1186/1471-2164-12-444>.
92. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
93. Paez-Espino D, Roux S, Chen IA, Palaniappan K, Ratner A, Chu K, Huntemann M, Reddy TBK, Pons JC, Llabres M, Eloë-Fadrosch EA, Ivanova NN, Kyrpides NC. 2019. IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res* 47:D678–D686. <https://doi.org/10.1093/nar/gky1127>.
94. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>.
95. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. 2015. Cluster analysis basics and extensions. R package version 2.0.1. <https://CRAN.R-project.org/view=Cluster>.
96. Dixon PJ. 2003. VEGAN, a package of R functions for community ecology. *J Vegetation Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.
97. Zhou J, Deng Y, Luo F, He Z, Yang Y. 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO₂. *mBio* 2:e00122-11. <https://doi.org/10.1128/mBio.00122-11>.
98. Zhou J, Deng Y, Luo F, He Z, Tu Q, Zhi X. 2010. Functional molecular ecological networks. *mBio* 1:e00169-10. <https://doi.org/10.1128/mBio.00169-10>.
99. Deng Y, Jiang Y-H, Yang Y, He Z, Luo F, Zhou J. 2012. Molecular ecological network analyses. *BMC Bioinformatics* 13:113. <https://doi.org/10.1186/1471-2105-13-113>.
100. Revelle W. 2013. psych: procedures for psychological, psychometric, and personality research. R package version 2.1.6. Northwestern University, Evanston, IL. <https://CRAN.R-project.org/package=psych>.
101. Mathieu B, Sebastien H, Mathieu J. 2009. Gephi: an open-source software for exploring and manipulating networks. Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM 2009, San Jose, CA.
102. Clarke KR, Ainsworth M. 1993. A method of linking multivariate community structure to environmental variables. *Mar Ecol Progr Ser* 92:205–219.