Genomic Characterization and Distribution Pattern of a Novel Marine OM43 Phage

Mingyu Yang, Qian Xia†, Sen Du, Zefeng Zhang, Fang Qin and Yanlin Zhao*

Fujian Provincial Key Laboratory of Agroecological Processing and Safety Monitoring, College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, China

Bacteriophages have a significant impact on the structure and function of marine microbial communities. Phages of some major bacterial lineages have recently been shown to dominate the marine viral communities. However, phages that infect many important bacterial clades still remained unexplored. Members of the marine OM43 clade are methylotrophs that play important roles in C1 metabolism. OM43 phages (phages that infect the OM43 bacteria) represent an understudied viral group with only one known isolate. In this study, we describe the genomic characterization and biogeography of an OM43 phage that infects the strain HTCC2181, designated MEP301. MEP301 has a genome size of 34,774 bp. We found that MEP301 is genetically distinct from other known phage isolates and only displays significant sequence similarity with some metagenomic viral genomes (MVGs). A total of 12 MEP301-type MVGs were identified from metagenomic datasets. Comparative genomic and phylogenetic analyses revealed that MEP301-type phages can be separated into two subgroups (subgroup I and subgroup II). We also performed a metagenomic recruitment analysis to determine the relative abundance of reads mapped to these MEP301-type phages, which suggested that subgroup I MEP301-type phages are present predominantly in the cold upper waters with lower salinity. Notably, subgroup II phages have an inverse different distribution pattern, implying that they may infect hosts from a distinct OM43 subcluster. Our study has expanded the knowledge about the genomic diversity of marine OM43 phages and identified a new phage group that is widespread in the ocean.

Keywords: OM43, OM43 phage, genomics, metagenomic viral genomes, distribution patterns

INTRODUCTION

Viruses play an important role in the marine microbial loop and biogeochemical cycles (Fuhrman, 1999; Suttle, 2005, 2007). They are the most abundant and genetically diverse biological entities in the ocean, with average abundance of an order of magnitude more than that of bacteria (Suttle, 2007; Güemes et al., 2016). Most marine viruses are bacteriophages that infect bacteria (Fuhrman, 1999; Wommack and Colwell, 2000). Bacteriophages can modulate the structure and function of the bacterioplankton communities since they serve as an important driver of bacterial mortality (Breitbart, 2012; Thingstad et al., 2014). Bacteriophages can also drive the

evolution of bacteria *via* selective pressure (Marston et al., 2012; Martiny et al., 2014). Owing to the tremendous advance in cultured-independent technologies, such as metagenomics (Hurwitz and Sullivan, 2013; Mizuno et al., 2013; Brum et al., 2015; Paez-Espino et al., 2016; Gregory et al., 2019; Beaulaurier et al., 2020; Luo et al., 2020) and single-cell genomics (Roux et al., 2014; Labonté et al., 2015; Martinez-Hernandez et al., 2017), a comprehensive understanding of the viral community structure and diversity has emerged and many viral genome fragments have been obtained from these analyses. Compared to the thriving metagenomic studies, fewer studies describe the isolation and study of new marine phages. Therefore, most metagenomic viral sequences lack cultured counterparts that hamper the understanding of the ecological roles and biological traits of marine viruses. In recent years, bacteriophages that infect ecologically important and abundant marine bacterioplankton have received increasing attention. For example, pelagiphages, RCA phages and SAR116 phages have been shown to be diverse and dominate the viral communities (Kang et al., 2013; Zhao et al., 2013, 2019; Zhang et al., 2019, 2021; Buchholz et al., 2021). Other than the above-mentioned phages, phages infecting many important marine bacteria still remain poorly investigated.

The OM43 clade affiliated to the Type I methylotrophs of the family *Methlophilaceae* is known as an important component of marine coastal bacterial communities (Rappé et al., 2000; Suzuki et al., 2004), where it is commonly associated with phytoplankton blooms (Morris et al., 2006). OM43 clade has diverse metabolic profiles and plays an important role in the metabolism of the C1 compounds in the marine environments (Giovannoni et al., 2008; Halsey et al., 2012; Jimenez-Infante et al., 2016). Bacteria of this clade are difficult to culture in laboratories, owing to their sensitivity to slight biochemical variations of the seawater (Giovannoni et al., 2008), for this reason, there are limited OM43 strains have been isolated and genomically characterized so far. (Giovannoni et al., 2008; Huggett et al., 2012; Jimenez-Infante et al., 2016). To date, only one OM43 phage has been reported, limiting our understanding on their diversity, evolution, ecology, and impacts. Venkman, the first OM43 phage isolated on the OM43 strain H5P1, was isolated from the Western English Channel (Buchholz et al., 2021). Venkman exhibits significant genomic similarity with freshwater *Methylophilales* LD28 phage P19250A (Buchholz et al., 2021). HTCC2181 was the first isolate from the OM43 clade, which served as a good example of bacterioplanktons with reduced genome size to adapt to the nutrition limited environments (Giovannoni et al., 2008; Halsey et al., 2012). Although it is one of the most important model strain of the OM43 clade, HTCC2181 has no reported phages to date.

In this study, we describe the isolation, genomic characterization, and global distribution pattern of MEP301, a phage that infect OM43 strain HTCC2181. Genomic analyses reveal that MEP301 belongs to a novel phage group with all known relatives from phages assembled from metagenomic studies. We also identified some MEP301-type genomes from marine metagenomic datasets and conducted a metagenomic recruitment analysis to illustrate the global distribution patterns of MEP301-type phages.

MATERIALS AND METHODS

Cultivation of OM43 strain HTCC2181

The axenic OM43 strain HTCC2181 was first isolated from the Oregon coast (Giovannoni et al., 2008) and was kindly provided by Prof. Stephen Giovannoni, Oregon State University. HTCC2181 was grown in seawater-based medium supplemented with 100 µM methanol, 1 mM NH_4Cl, 100 µM KH_2PO_4, 1 µM $FeCl_3$, and excess vitamins (Giovannoni et al., 2008). HTCC2181 cultures were incubated at 23°C in the dark without shaking.

Source Waters and Phage Isolation

The water sample used for phage isolation was collected at Yantai coast, Bohai sea in China (N37°28, E121°28'). The seawater sample was filtered through 0.1-µm-pore-size filters to remove non-viral components. This cell-free sample was then stored at 4°C for further experiments. The details of phage isolation have been described in previous studies (Zhao et al., 2013; Zhang et al., 2019). Briefly, the filtered seawater samples were inoculated with exponential phase HTCC2181 culture and a Guava EasyCyte flow Cytometer was used to monitor the lysis of cell culture (Merck Millipore, Billerica, MA, United States). The presence of phage particles was confirmed by epifluorescence microscopy after cell lysis was detected (Suttle and Fuhrman, 2010). Purified phage clones were obtained by the dilution-to-extinction method (Zhao et al., 2013; Zhang et al., 2019), and the purity of the phage was verified by genome sequencing.

Transmission Electron Microscopy

The lysate of MEP301 was filtered through a 0.1-µm filter and then centrifuged by ultracentrifugation (Beckman Coulter, United States) at 50,000 *g* for 2 h. A drop of the concentrated phage sample was then placed on a copper grid and subsequently dried in the air. The grid was stained for 2 min in 2% uranyl acetate and them observed by a Hitachi transmission electron microscope at a voltage of 80 kV.

Phage DNA Preparation, Genome Sequencing, and Annotation

The phage lysate was filtered through 0.1-µm filters (Pall Life Sciences) to remove cell debris. Filtered phage lysate was concentrated to approximately 300 µl by using Amicon Ultra Centrifugal Filters (30 kDa, Merck Millipore) and Nanosep 10 K centrifugal tubes Ultra-centrifugal tubes (30 kDa, Pall Life Sciences). Phage DNA was extracted by the method of phenol-chloroform extraction (Sambrook et al., 1989). For Illumina sequencing, the DNA sample was sheared by Covaris M220 (Covaris, United States), and the DNA library was prepared using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, United States). Approximately 10 ng of the DNA

sequencing library was used to generate a cluster in cBot using a TruSeq PE Cluster Kit (Illumina, United States) and whole genome sequencing of MEP301 was conducted using Illumina HiSeq™ 4,000 plateform (paired-end technology 2 × 150 bp) at Shanghai Hanyu Bio-Tech (Shanghai, China). The CLC Genomic Workbench 11.0.1 software (Qiagen, Hilden, Germany) was used to perform the quality control, trimming and *de novo* genome assembly with default parameters. To complete the whole genome, PCR amplification and Sanger sequencing were used to close gaps. The genome sequence of MEP301 has been deposited in the GenBank database under the accession number MW452941.

Putative ORFs in the MEP301 genome were identified using GeneMark (Besemer and Borodovsky, 1999). The translated ORFs were used as BLASTP queries to search against the NCBI database (e-value ≤ 1e-3, alignment coverage ≥50%, ≥25% amino acid identity). Putative functions were assigned to ORFs based on their homology to proteins of known function. The conserved domains of proteins were predicted by searching the PFAM database with Hmmer web server (Potter et al., 2018). Distant protein homologs were predicted by HHpred server (Söding et al., 2005). tRNA genes were searched by the tRNAscan-SE software (Lowe and Eddy, 1997).

Retrieval of MEP301-Type Metagenomic Viral Genomes (MVGs)

To recover MEP301-type MVGs, the amino acid sequences of MEP301 ORFs were searched against 515,588 MVG sequences from Global Ocean Viromes (GOV and GOV2.0) (Roux et al., 2016; Gregory et al., 2019), the MedDCM fosmid library (Mizuno et al., 2013) and Station ALOHA (Beaulaurier et al., 2020) using BLASTP (e-value ≤ 10⁻³, ≥25% amino acid identity). Orthologous groups were determined using OrthoMCLv2.0 (Fischer et al., 2011). The contigs that share ≥40% genes with MEP301 and have a size ≥25 kb were designated as MEP301-type MVGs.

Phylogenetic Analysis and Whole-Genome Phylogeny Analysis

A maximum likelihood phylogenetic tree of TerL protein sequences was constructed. Amino acid sequence alignment and editing were performed using Mafft and (Katoh et al., 2009) trimAl (Capella-Gutierrez et al., 2009), respectively. The phylogenetic tree was constructed using IQ-TREE v1.6.12 with 1,000 bootstrap replicates. Whole-genome based phylogeny based on amino acid sequences was built using VICTOR (Meier-Kolthoff and Göker, 2017).[1] Genome-based classification at the genus and family level was performed using the OPTSIL program (Göker et al., 2009).

Metagenomic Read Mapping

Marine viromic datasets that were downloaded to evaluate the relative abundance of MEP301-type sequences include Pacific Ocean Virome (POV) (Hurwitz and Sullivan, 2013), Malaspina Expedition Virome (MEV) (Roux et al., 2016), GOV

---

[1]https://ggdc.dsmz.de/victor.php

(Roux et al., 2016), and GOV2.0 (Gregory et al., 2019). The MEP301-type sequences were first clustered at 70% average nucleotide identity and only the longest MVGs within a cluster was retained for recruitment analysis. Viromic reads (≥50 bp) were recruited using BLASTn as described (Mizuno et al., 2016; Martinez-Hernandez et al., 2017; Kim et al., 2019; López-Pérez et al., 2019; Zaragoza-Solas et al., 2020), with ≥70% identity and ≥90% read coverage. The relative abundances of phage genomes were normalized by total recruited nucleotides (kb) per kb of genome per gigabase of metagenome (KPKG). MEP301-type genomes for which <40% of the genomes was covered by recruited reads in a given viromic dataset were given a RPKG value of 0 (Buchholz et al., 2021).

RESULTS AND DISCUSSION

Isolation and Morphology of MEP301

OM43 phage MEP301 was isolated from coastal surface water of Yantai, Bohai sea, China (N37°28', E121°28'). The TEM image shows that MEP301 is a phage with an icosahedral capsid of 59 ± 2 nm in diameter and a short tail (about 7 nm in length; **Figure 1A**). The morphological characteristics of MEP301 suggest that it belongs to a yet unassigned family of the *Caudovirales* order.

General Genome Characteristics of MEP301

The genome of MEP301 was assembled into a single circular contig. This indicated that the sequence obtained represents a complete genome and suggested that its linear genome is either circularly permuted or has terminal repeats. MEP301 has a dsDNA genome of 34,774 bp in size. The G+C content of MEP301 genome is 40.4%, which was similar to that of its host HTCC2181 (37.9%) but significantly higher than that of the first OM43 phage Venkman (31.9%). A total of 55 open reading frames (ORFs) were identified in the MEP301 genome (**Figure 1B**). The BLASTP analysis revealed that the majority of the ORFs in MEP301 are most similar to genes found in metagenomic sequences. In addition, MEP301 does not display substantial genome similarity with any cultured phages. Overall, the above analyses suggest that MEP301 represents a new group of phages.

Approximately half of the ORFs in MEP301 can be assigned with predicted biological functions. These ORFs encode proteins associated with the DNA metabolism, virion structure, DNA packaging, and host lysis. No tRNA was identified in MEP301. Despite the universal metabolisms reprograming in host-phage interactions, it remains to be verified whether MEP301 rewires the host cellular metabolisms to utilize host specific tRNAs to synthetize viral proteins during infection.

DNA Replication and Metabolism Related Genes

A total of five genes involved in DNA replication and metabolism have been identified from the MEP301 genome (**Figure 1B**).
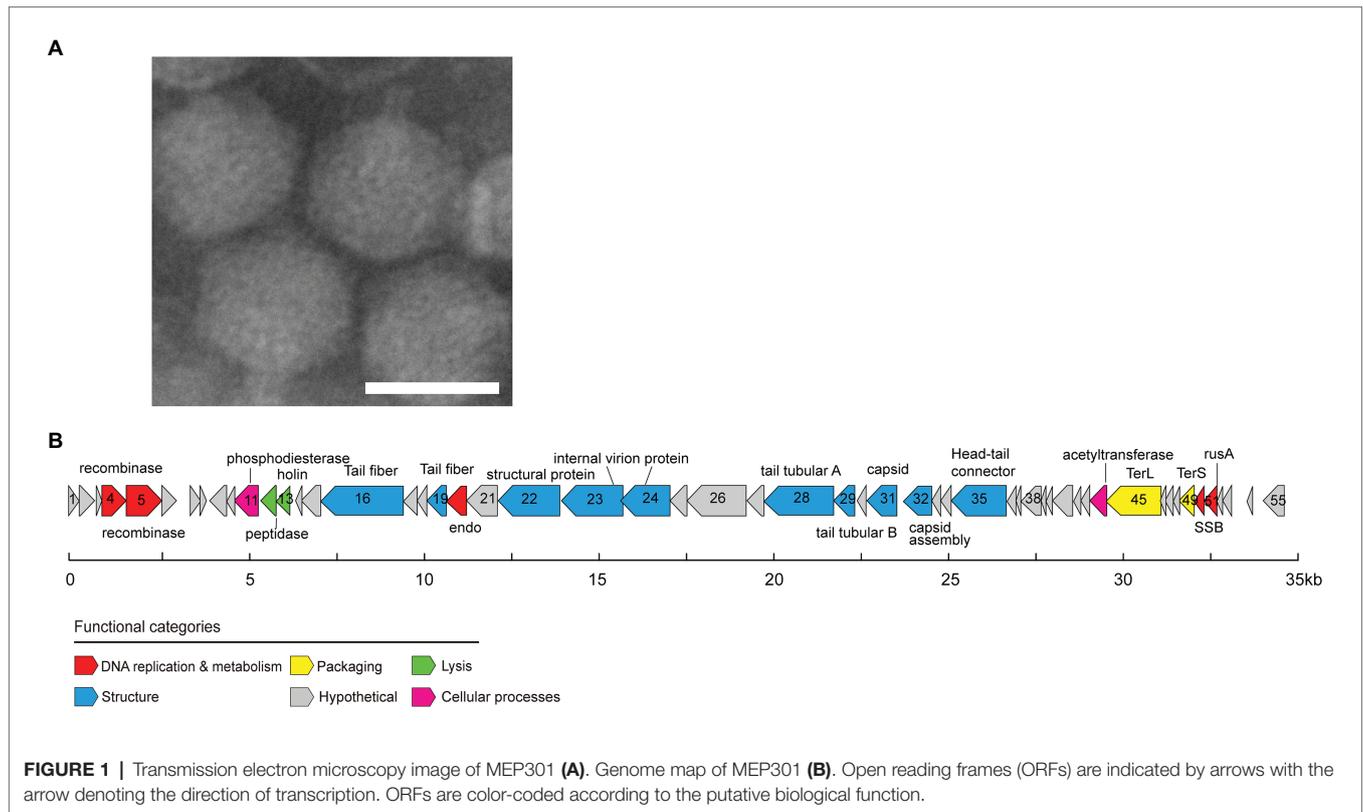
FIGURE 1 | Transmission electron microscopy image of MEP301 **(A)**. Genome map of MEP301 **(B)**. Open reading frames (ORFs) are indicated by arrows with the arrow denoting the direction of transcription. ORFs are color-coded according to the putative biological function.

TABLE 1 | Genomic features of 12 nearly full-length MEP301-type MVGs.

MEP301-type MVGs	Length (bp)	G+C%	Source
Station85_DCM_COMBINED_FINAL_NODE_755_length_35564_cov_5.980596	35,564	36.6	Antarctic province
Station189_DCM_ALL_assembly_NODE_895_length_35158_cov_59.730052	35,158	34.9	Boreal polar province
Station201_SUR_ALL_assembly_NODE_1343_length_33765_cov_9.933136	33,765	36.7	Boreal polar province
Station196_SUR_ALL_assembly_NODE_1083_length_33375_cov_75.986705	33,375	37.6	Boreal polar province
Station102_SUR_ALL_assembly_NODE_387_length_33295_cov_11.029693	33,295	37.8	Pacific equatorial divergence province
Station82_DCM_COMBINED_FINAL_NODE_1219_length_27352_cov_13.359856	27,352	33.5	Southwest Atlantic shelves province
Station180_SUR_ALL_assembly_NODE_3819_length_27194_cov_85.738937	27,194	38.6	Boreal polar province
Station173_DCM_ALL_assembly_NODE_3468_length_26058_cov_29.993501	26,058	35.0	Boreal polar province
Tp1_102_SUR_0-0d2_scaffold65748_2	33,618	37.6	Pacific equatorial divergence province
uvMED-CGR-U-MedDCM-OCT-S25-C65	32,359	37.1	Mediterranean Sea deep chlorophyll maximum
KX158644.1 Uncultured bacterium clone VISS3_088	34,150	36.4	East Greenland current (station 3) of the boreal polar region
NYTA01000058 MHASMcontig_719174138	33,650	37.8	Indian monsoon gyres province

This includes genes encoding the N terminal resolvase domain protein (ORF4), YqaJ viral recombinase family protein (ORF5), single-stranded DNA (ssDNA) binding protein (SSB, ORF50), crossover junction endodeoxyribonuclease RusA superfamily protein (ORF51), and endonuclease (ORF20). Neither RNA polymerase nor DNA polymerase was identified in the MEP301 genome, suggesting that MEP301 is highly dependent on the host's machinery for transcription and DNA replication.

ORF4 encodes a protein containing a N terminal resolvase domain (PF00239). In serine integrase, this domain is always associated with a catalytic recombinase PF07508 domain (Smith et al., 2010; Ambroset et al., 2016). However, no protein containing the PF07508 domain was detected in MEP301, indicating that MEP301 may not contain a functional serine recombinase.

ORF5 encodes a YqaJ viral recombinase family protein (PF09588). This protein shares weak sequence identity with *Geobacillus* phage GBK2 and *Bacillus* phage SPP1 encoded exonuclease/ recombinase (23 and 22% amino acid identity, respectively). This protein in phage SPP1 was reported to function as an alkaline exonuclease (Vellani and Myers, 2003). The SSB gene (ORF50) in the MEP301 genome is most similar to SSB genes from uncultured phage sequences, and also has similarity to those of some siphoviruses. For example, it shares 32 and 33% amino acid identity with those in *Acinetobacter* phage SH-Ab and DMU1, respectively. SSB proteins are ubiquitous in phage genomes. SSB proteins prevent ssDNA forming dsDNA by binding to the ssDNA and are essential for numerous DNA metabolic processes (Shereda et al., 2008).
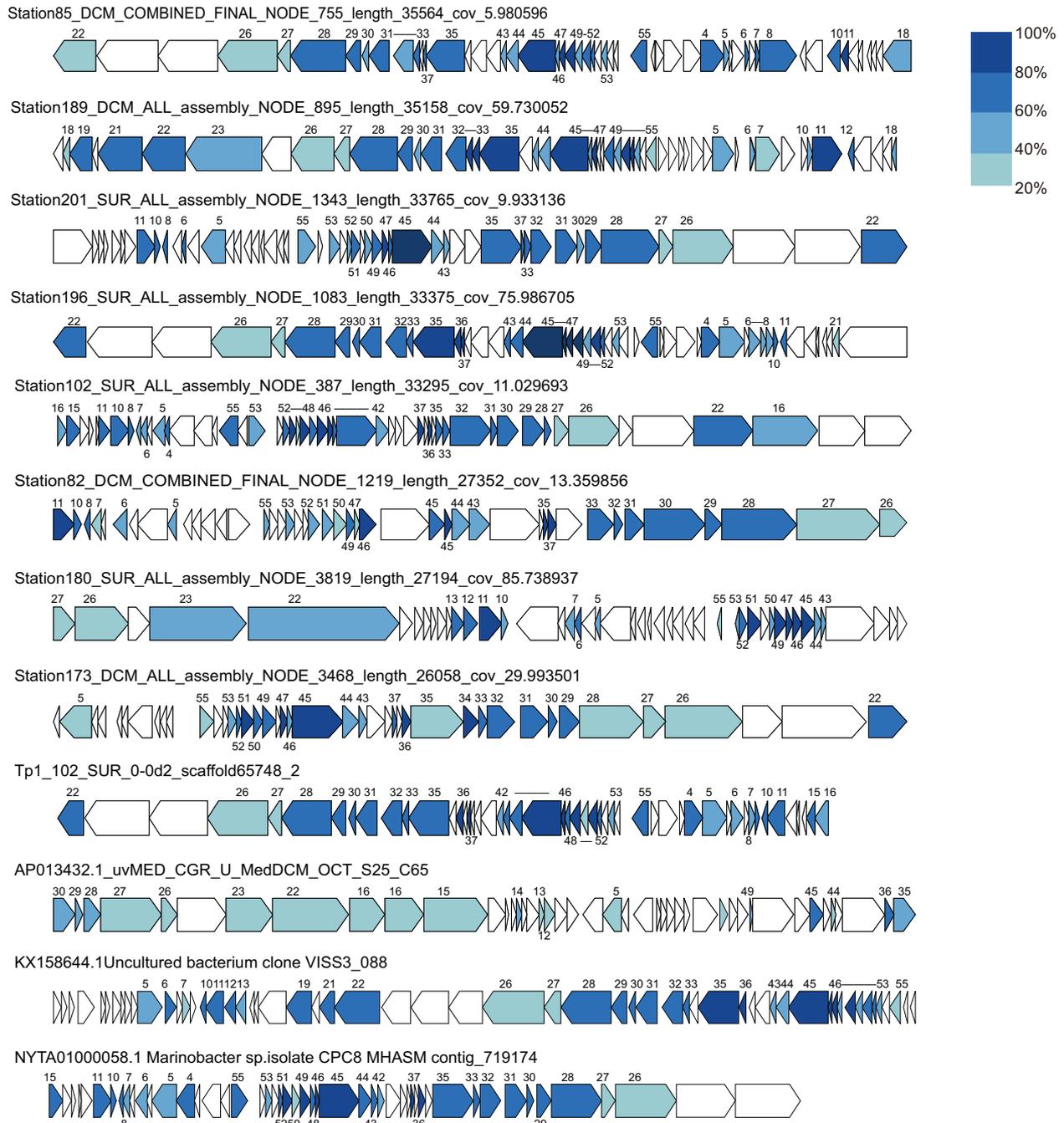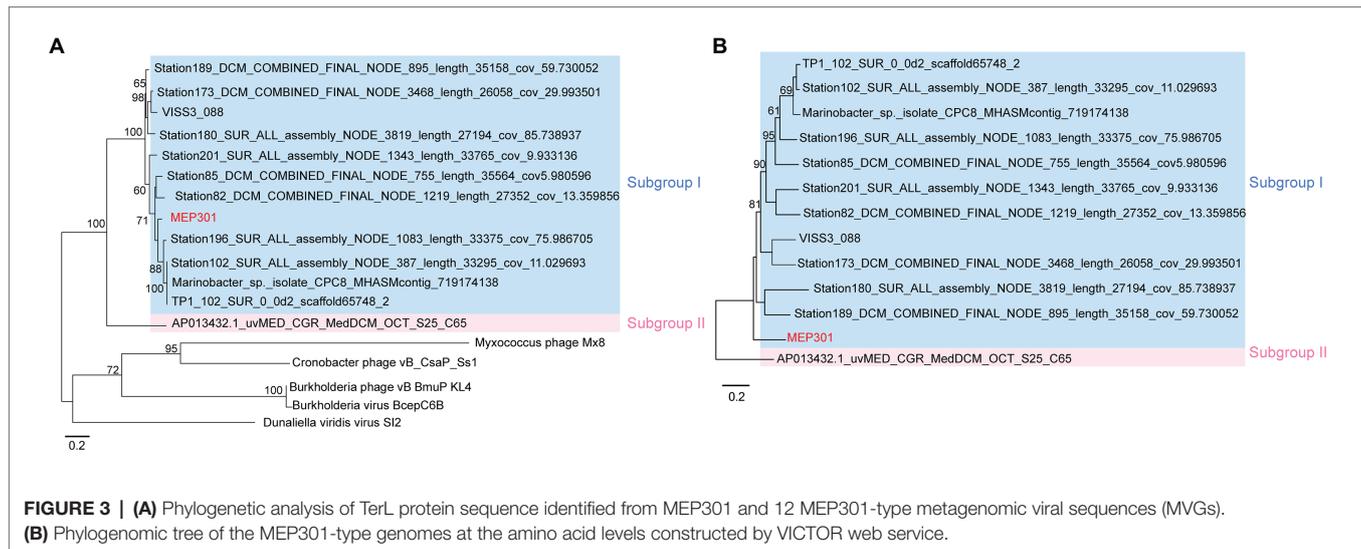
FIGURE 2 | Genome organization of 12 metagenomic viral genomes (MVGs) related to MEP301. Open reading frames (ORFs) are colored according to the degree of amino acid sequence identity to the genes in MEP301. The number of MEP301 homologous ORFs are indicated above the frames.

Two genes involved in DNA cleavage have been identified including the crossover junction endodeoxyribonuclease RusA gene (ORF51, PF05866) and the predicted endonuclease gene (ORF20). Both gene products are predicted to cut the phosphodiester bonds between nucleotides in the interior DNA helix, resulting in the production of nucleotide monomers.

MEP301 RusA shares 37.4% amino acid identity with that in *Shigella* phage 75/02 Stx. The RusA protein is an endonuclease that functions as a holliday junction resolvase and has been found to mediate the genetic recombination and repair in the prophage of *Escherichia coli* K-12 (Bolt et al., 1999). The predicted endonuclease shares 39.0% amino acid identity with

FIGURE 3 | (A) Phylogenetic analysis of TerL protein sequence identified from MEP301 and 12 MEP301-type metagenomic viral sequences (MVGs). **(B)** Phylogenomic tree of the MEP301-type genomes at the amino acid levels constructed by VICTOR web service.

that in *Pseudomonas* phage JG004. Thus, these two genes could presumably be involved in degrading host genomes to obtain sufficient nucleotide monomers for phage DNA synthesis, which is fundamental for phage reproduction.

Structural Genes and DNA Packaging Genes

A set of structural genes, including genes encoding phage head-tail connector, capsid assembly protein, major capsid protein, internal virion proteins, and some tail related proteins, were predicted from the MEP301 genome (**Figure 1B**). Proteins encoded by these genes play important roles in phage morphogenesis. Most of these structural genes show higher sequence similarities with uncultured environmental viral genomes than to known phages, highlighting the novelty of MEP301. Among all cultured phages, some structural genes in MEP301 are most related to marine pelagiphages or SAR116 phage HMO-2011 with limited sequence identity. For example, the tail tubular protein B (ORF28) shares 27.6% amino acid identity with that in pelagiphage HTVC010P. The head-tail connector (ORF35) shares 27.0% amino acid identity with that in pelagiphage HTVC010P. The putative major capsid protein (ORF31) shares 30.7% amino acid identity with that in pelagiphage HTVC010P. The tail fiber protein (ORF19) shares 35.3% amino acid identity with SAR116 phage HMO-2011. The internal virion protein (ORF23) shares 30.3% amino acid identity with that in pelagiphage HTVC011P.

Both the large and small subunits of terminase (TerL and TerS) genes were identified in MEP301. These two subunits are indispensable for packaging the genome of most tailed bacteriophages (Sun et al., 2012). The terminases recognize the DNA for packaging and have a nuclease activity that is responsible for creating the ends of the virion chromosome (Casjens et al., 2005). The most closely related relatives of MEP301 TerL gene in cultured phage genomes are from *Cronobacter* phage vB_CsaP_Ss1, *Dunaliella viridis* virus SI2, *Burkholderia* phage vB_BmuP_KL4, and BcepC6B, with approximately 40% amino acid identity,

suggesting that MEP301 may use a similar DNA packaging strategy. The TerS gene (ORF49; PF03592) has also been identified but does not share homology with any known TerS gene.

Lysis Gene

The peptidase M15A (Peptidase_M15_3; PF08291), encoded by ORF12, shares 55% amino acid identity with that in an unfinished *Methylophilales* phage HIM624-A (accession number: AFB70783.1). HIM624-A, which infects *Methylophilacea* HIMB624, is also an OM43 phage, but its genome is not completed yet.

Genes Likely Involved in Some Cellular Processes

The MEP301 ORF11 is predicted to belong to the metallophosphoesterase superfamily (PF00149) and shares homology with the phosphoesterase gene in *Vibrio* phage Vp670 (47.3% amino acid identity). Calcineurin-like phosphoesterases are common in bacteria and archaea genomes and have also been identified in many phage genomes. Calcineurin-like phosphoesterases have hydrolase activity against diverse phosphorylated substrates. The function of MEP301-encoded phosphodiesterases is still unclear. The MEP301 ORF44 encodes an acetyltransferase. Acetyltransferases are involved in the acetylation of a wide variety of substrates and play important roles in a large number of bacterial biological functions. Glycosyltransferases have been identified in many phage genomes and presumed to confer protection to phage from host restriction endonucleases and play important roles in temperate phage-mediated host immunity and host serotype conversion (Markine-Goriaynoff et al., 2004). The specific role of MEP301 encoded acetyltransferase, however, remains unclear.

MEP301-Type Metagenomic Viral Genome

A search was performed to retrieve MEP301-type MVGs from environmental metagenomes. A total of 12 MEP301-type MVGs that share ≥40% genes with MEP301 were retrieved from

FIGURE 4 | Heatmap displaying the relative abundance of each MEP301-type phage in different marine viromic datasets. Normalized relative abundance is depicted as total mapped nucleotides (kb) per kb of genome per gigabase of metagenome (KPKG). The number (1–10) on the x-axis represent the MEP301-type phage: 1. KX158644.1Uncultured_bacterium_clone_VISS3_088, 2. Station102_SUR_ALL_assembly_NODE_387_length_33295_cov_11.029693, 3. Station173_DCM_ALL_assembly_NODE_3468_length_26058_cov_29.993501, 4. Station180_SUR_ALL_assembly_NODE_3819_length_27194_cov_85.738937, 5. Station189_DCM_ALL_assembly_NODE_895_length_35158_cov_59.730052, 6. Station196_SUR_ALL_assembly_NODE_1083_length_33375_cov_75.986705, 7. Station201_SUR_ALL_assembly_NODE_1343_length_33765_cov_9.933136, 8. Station82_DCM_COMBINED_FINAL_NODE_1219_length_27352_cov_13.359856, 9. Station85_DCM_COMBINED_FINAL_NODE_755_length_35564_cov_5.980596, 10. uvMED−CGR−U−MedDCM−OCT−S25−C65, respectively.

various marine viromic datasets. These MEP301-type MVGs range in size from 26 to 35 kb, encoding 42 to 57 ORFs (**Table 1**). The G+C content of these MVGs range from 33.5 to 37.8%, similar to that of MEP301. Genomic comparison indicated that these 12 MVGs phages display obvious relationships with the MEP301 with sequence similarities and overall conservation of genome architectures (**Figure 2**). We classified these MVGs into the MEP301-type phage group. Compared to genes in other MVGs, corresponding genes in an uncultured Mediterranean phage fosmid sequence uvMED-CGR-U-MedDCM-OCT-S25-C65 have lower sequence similarity to MEP301 (23–65%, average 39.3%). To investigate the evolutionary relationships among these MEP301-type phage genomes, TerL protein were used for phylogenetic analysis. Based on the TerL phylogeny, these 12 MVGs are clustered with MEP301 and form a separate branch, indicating that MEP301-type is a novel phage group. All MEP301-type genomes were further separated into two major subgroups with subgroup I containing 11 genomes and subgroup II containing uvMED-CGR-U-MedDCM-OCT-S25-C65 (**Figure 3A**). The whole genome-based VICTOR phylogeny also shows a similar topology (**Figure 3B**). The OPTSIL taxon prediction suggests that these MEP301-type phages belong to a subfamily-level group and can be separated into two genus-level subgroups. Collectively, these results confirm that these MVGs belong to the same phage group with MEP301, and MEP301-type phage group containing at least two closely-related subgroups.

Global Distribution of MEP301-Type Phages

We next queried for the presence of MEP301-type phages in publicly available marine metagenomes by mapping viromic reads to MEP301-type genomes. Viromic read mapping (≥70% nucleotide identity) revealed that these MEP301-type phages can be detected in many oceanic stations (**Figure 4**), but their relative abundance was significantly low compared to the relative abundance of previously reported pelagiphages (Zhang et al., 2019). These phages were found to be distributed mainly in surface ocean waters (>200 m). In surface waters, all subgroup I genomes were found predominantly in the cold waters of Arctic Ocean and North Atlantic Ocean, where both temperature and salinity were lower (**Figure 4**). These oceanic regions also displayed higher chlorophyll values. They were also detected in some stations of South Atlantic, Pacific Ocean, and Southern Ocean. We also noticed that the first sequenced OM43 phage Venkman had a similar distribution pattern with subgroup I MEP301-type phages. Interestingly, subgroup II genome uvMED_CGR_U_MedDCM_OCT_S25_C65 displayed a distinct distribution pattern. It was not detected in Arctic Ocean stations but was detected in other analyzed oceanic regions, where temperature and salinity were higher. Phylogenetically, uvMED_CGR_U_MedDCM_OCT_S25_C65 is more distantly related to other MEP301-type phages. Their different distribution pattern may be due to differences in host species. The OM43 clade can be separated into two divergent clusters with different distribution pattern (Jimenez-Infante et al., 2016). HTCC2181-cluster has high abundance in high chlorophyll content but lower temperature waters, while the H-RS cluster is more abundant in

warmer and high salinity waters (Jimenez-Infante et al., 2016). Generally, the distribution of a phage is broadly correlated to the distribution of its hosts. Based on these observations, it is likely that the subgroup II phage infect hosts from the H-RS cluster.

CONCLUSION

In this study, we sequenced and analyzed the genome of a new OM43 phage MEP301 and obtained related phage sequences from viromic datasets, providing new insights into the diversity and evolution of marine OM43 phages. We also provided new insights into the distribution patterns of this important phage and highlight their ecological and evolutionary relevance. We showed that MEP301-type phages are widely distributed in marine environments although they were not in high abundance. Our results raise questions for future studies on the influence of MEP301-type phages on OM43 clade physiology and diversification. Phage MEP301 can serve as a model system for the study of OM43 phage ecological roles and phage-host interactions.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MY and QX performed the genomic analyses and prepared the manuscript. SD performed the metagenomic analysis. ZZ contributed in bioinformatics analysis. FQ isolated the phage. YZ designed the study and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.651326/full#supplementary-material

REFERENCES

Ambroset, C., Coluzzi, C., Guedon, G., Devignes, M. D., Loux, V., Lacroix, T., et al. (2016). New insights into the classification and integration specificity of *streptococcus* integrative conjugative elements through extensive genome exploration. *Front. Microbiol.* 6:1483. doi: 10.3389/fmicb.2015.01483

Beaulaurier, J., Luo, E., Eppley, J. M., Uyl, P. D., and Dai, X. G. (2020). Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res.* 30, 437–446. doi: 10.1101/gr.251686.119

Besemer, J., and Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res.* 27, 3911–3920. doi: 10.1093/nar/27.19.3911

Bolt, E. L., Sharples, G. J., and Lloyd, R. G. (1999). Identification of three aspartic acid residues essential for catalysis by the RusA Holliday junction resolvase. *J. Mol. Biol.* 286, 403–415. doi: 10.1006/jmbi.1998.2499

Breitbart, M. (2012). Marine viruses: truth or dare. *Annu. Rev. Mar. Sci.* 4, 425–448. doi: 10.1146/annurev-marine-120709-142805

Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., et al. (2015). Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498. doi: 10.1126/science.1261498

Buchholz, H. H., Michelsen, M. L., Bolaños, L. M., Browne, E., Allen, M. J., and Temperton, B. (2021). Efficient dilution-to-extinction isolation of novel virus-host model systems for fastidious heterotrophic bacteria. *ISME J.* doi: 10.1038/s41396-020-00872-z [Epub ahead of print]

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Casjens, S. R., Gilcrease, E. B., Winn-Stapley, D. A., Schicklmaier, P., Schmieger, H., Pedulla, M. L., et al. (2005). The generalized transducing salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J. Bacteriol.* 187, 1091–1104. doi: 10.1128/JB.187.3.1091-1104.2005

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., et al. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr. Protoc. Bioinformatics* 35, 6–12. doi: 10.1002/0471250953.bi0612s35

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548. doi: 10.1038/21119

Giovannoni, S. J., Hayakawa, D. H., Tripp, H. J., Stingl, U., Givan, S. A., Cho, J. C., et al. (2008). The small genome of an abundant coastal ocean methylotroph. *Environ. Microbiol.* 10, 1771–1782. doi: 10.1111/j.1462-2920.2008.01598.x

Göker, M., García-Blázquez, G., Voglmayr, H., Tellería, M. T., and Martín, M. P. (2009). Molecular taxonomy of phytopathogenic fungi: a case study in Peronospora. *PLoS ONE* 4:6319e. doi: 10.1371/journal.pone.0006319

Gregory, A. C., Zayed, A. A., Conceição-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. doi: 10.1016/j.cell.2019.03.040

Güemes, A. G. C., Youle, M., Cantú, V. A., Felts, B., Nulton, J., and Rohwer, F. (2016). Viruses as winners in the game of life. *Annu. Rev. Virol.* 3, 197–214. doi: 10.1146/annurev-virology-100114-054952

Halsey, K. H., Carter, A. E., and Giovannoni, S. J. (2012). Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. *Environ. Microbiol.* 14, 630–640. doi: 10.1111/j.1462-2920.2011.02605.x

Huggett, M. J., Hayakawa, D. H., and Rappé, M. S. (2012). Genome sequence of strain HIMB624, a cultured representative from the OM43 clade of marine Betaproteobacteria. *Stand. Genomic Sci.* 6, 11–20. doi: 10.4056/sigs.2305090

Hurwitz, B. L., and Sullivan, M. B. (2013). The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* 8:57355e. doi: 10.1371/journal.pone.0057355

Jimenez-Infante, F., Ngugi, D. K., Vinu, M., Alam, I., Kamau, A. A., Blom, J., et al. (2016). Comprehensive genomic analyses of the OM43 Clade, including a novel species from the Red Sea, indicate ecotype differentiation among marine methylotrophs. *Appl. Environ. Microbiol.* 82, 1215–1226. doi: 10.1128/AEM.02852-15

Kang, I., Oh, H. M., Kang, D., and Cho, J. C. (2013). Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12343–12348. doi: 10.1073/pnas.1219930110

Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 537, 39–64. doi: 10.1007/978-1-59745-251-9_3

Kim, J. G., Kim, S. J., Cvirkaite-Krupovic, V., Yu, W. J., Gwak, J. H., López-Pérez, M., et al. (2019). Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15645–15650. doi: 10.1073/pnas.1905682116

Labonté, J. M., Swan, B. K., Poulos, B., Luo, H. W., Koren, S., Hallam, S. J., et al. (2015). Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* 9, 2386–2399. doi: 10.1038/ismej.2015.48

López-Pérez, M., Haro-Moreno, J. M., Torre, J. R. D. L., and Rodriguez-Valera, F. (2019). Novel Caudovirales associated with Marine Group I Thaumarchaeota assembled from metagenomes. *Environ. Microbiol.* 21, 1980–1988. doi: 10.1111/1462-2920.14462

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Luo, E., Eppley, J. M., Romano, A. E., Mende, D. R., and Delong, E. F. (2020). Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J.* 14, 1304–1315. doi: 10.1038/s41396-020-0604-8

Markine-Goriaynoff, N., Gillet, L., Etten, J. L. V., Korres, H., Verma, N., and Vanderplasschen, A. (2004). Glycosyltransferases encoded by viruses. *J. Gen. Virol.* 85, 2741–2754. doi: 10.1099/vir.0.80320-0

Marston, F. M., Pierciey, F. J. Jr., Shepard, A., Gearin, G., Qi, J., Yandava, C., et al. (2012). Rapid diversification of coevolving marine Synechococcus and a virus. *Proc. Natl. Acad. Sci. U. S. A.* 109, 4544–4549. doi: 10.1073/pnas.1120310109

Martinez-Hernandez, F., Fornas, O., Gomez, M. L., Bolduc, B., de La Cruz Pena, M. J., Martínez, J. M., et al. (2017). Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* 8:15892. doi: 10.1038/ncomms15892

Martiny, J. B. H., Riemann, L., Marston, M. F., and Middelboe, M. (2014). Antagonistic coevolution of marine planktonic viruses and their hosts. *Annu. Rev. Mar. Sci.* 6, 393–414. doi: 10.1146/annurev-marine-010213-135108

Meier-Kolthoff, J. P., and Göker, M. (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* 33, 3396–3404. doi: 10.1093/bioinformatics/btx440

Mizuno, C. M., Ghai, R., Saghaï, A., López-García, P., and Rodriguez-Valera, F. (2016). Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7, 805–816e. doi: 10.1128/mBio.00805-16

Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E., and Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genet.* 9:1003987e. doi: 10.1371/journal.pgen.1003987

Morris, R. M., Longnecker, K., and Giovannoni, S. J. (2006). *Pirellula* and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ. Microbiol.* 8, 1361–1370. doi: 10.1111/j.1462-2920.2006.01029.x

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., et al. (2016). Uncovering Earth's virome. *Nature* 536, 425–430. doi: 10.1038/nature19094

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448

Rappé, M. S., Vergin, K., and Giovannoni, S. J. (2000). Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microb. Ecol.* 33, 219–232. doi: 10.1111/j.1574-6941.2000.tb00744.x

Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537, 689–693. doi: 10.1038/nature19366

Roux, S., Hawley, A. K., Beltran, M. T., Scofield, M., Schwientek, P., Stepanauskas, R., et al. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *Elife* 3:3125e. doi: 10.7554/eLife.03125

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning: A laboratory manual. 2nd Edn*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M., and Keck, J. L. (2008). SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol.* 43, 289–318. doi: 10.1080/10409230802241296

Smith, M. C. M., Brown, W. R. A., McEwan, A. R., and Rowley, P. A. (2010). Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochem. Soc. Trans.* 38, 388–394. doi: 10.1042/BST0380388

Söding, J., Biegert, A., and Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33, W244–W248. doi: 10.1093/nar/gki408

Sun, S., Gao, S., Kondabagil, K., Xiang, Y., Rossmann, M. G., and Rao, V. B. (2012). Structure and function of the small terminase component of the DNA packaging machine in T4-like bacteriophages. *Proc. Natl. Acad. Sci. U. S. A.* 109, 817–822. doi: 10.1073/pnas.1110224109

Suttle, C. A. (2005). Viruses in the sea. *Nature* 437, 356–361. doi: 10.1038/nature04160

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750

Suttle, C. A., and Fuhrman, J. A. (2010). "Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy" in *Manual of aquatic viral ecology*. eds. S. Wilhelm, M. Weinbauer and C. Suttle (Waco, TX: American Society of Limnology and Oceanography), 145–153.

Suzuki, M. T., Preston, C. M., Béjà, O., de la Torre, J. R., Steward, G. F., and DeLong, E. F. (2004). Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. *Microb. Ecol.* 48, 473–488. doi: 10.1007/s00248-004-0213-5

Thingstad, T. F., Vage, S., Storesund, J. E., Sandaa, R. A., and Giske, J. (2014). A theoretical analysis of how strain-specific viruses can control microbial species diversity. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7813–7818. doi: 10.1073/pnas.1400909111

Vellani, T. S., and Myers, R. S. (2003). Bacteriophage SPP1 chu is an alkaline exonuclease in the synexo family of viral two-component recombinases. *J. Bacteriol.* 185, 2465–2474. doi: 10.1128/JB.185.8.2465-2474.2003

Wommack, K. E., and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114. doi: 10.1128/MMBR.64.1.69-114.2000

Zaragoza-Solas, A., Rodriguez-Valera, F., and López-Pérez, M. (2020). Metagenome mining reveals hidden genomic diversity of pelagimyophages in aquatic environments. *mSystems* 5, 905–919e. doi: 10.1128/mSystems.00905-19

Zhang, Z., Chen, F., Chu, X., Zhang, H., Luo, H., Qin, F., et al. (2019). Diverse, abundant, and novel viruses infecting the marine Roseobacter RCA lineage. *mSystems* 4, 494–419e. doi: 10.1128/mSystems.00494-19

Zhang, Z., Qin, F., Chen, F., Chu, X., Luo, H., Zhang, R., et al. (2021). Culturing novel and abundant pelagiphages in the ocean. *Environ. Microbiol.* 23, 1145–1161. doi: 10.1111/1462-2920.15272

Zhao, Y., Qin, F., Zhang, R., Giovannoni, S. J., Zhang, Z., Sun, J., et al. (2019). Pelagiphages in the Podoviridae family integrate into host genomes. *Environ. Microbiol.* 21, 1989–2001. doi: 10.1111/1462-2920.14487

Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., et al. (2013). Abundant SAR11 viruses in the ocean. *Nature* 494, 357–360. doi: 10.1038/nature11921