



# Diversity and distribution of viruses inhabiting the deepest ocean on Earth

Huahua Jian<sup>1,2</sup> · Yi Yi<sup>1</sup> · Jiahua Wang<sup>1</sup> · Yali Hao<sup>1</sup> · Mujie Zhang<sup>1</sup> · Siyuan Wang<sup>1</sup> · Canxing Meng<sup>1</sup> · Yue Zhang<sup>1</sup> · Hongmei Jing<sup>3</sup> · Yinzao Wang<sup>1,2</sup> · Xiang Xiao<sup>1,2</sup>

Received: 9 December 2020 / Revised: 8 April 2021 / Accepted: 20 April 2021 / Published online: 10 May 2021  
© The Author(s), under exclusive licence to International Society for Microbial Ecology 2021

## Abstract

As the most abundant biological entities on the planet, viruses significantly influence the overall functioning of marine ecosystems. The abundance, distribution, and biodiversity of viral communities in the upper ocean have been relatively well studied, but our understanding of viruses in the hadal biosphere remains poor. Here, we established the oceanic trench viral genome dataset (OTVGD) by analysing 19 microbial metagenomes derived from seawater and sediment samples of the Mariana, Yap, and Kermadec Trenches. The trench viral communities harbored remarkably high novelty, and they were predicted to infect ecologically important microbial clades, including Thaumarchaeota and *Oleibacter*. Significant inter-trench and intra-trench exchange of viral communities was proposed. Moreover, viral communities in different habitats (seawater/sediment and depth-stratified ocean zones) exhibited distinct niche-dependent distribution patterns and genomic properties. Notably, microbes and viruses in the hadopelagic seawater seemed to preferably adopt lysogenic lifestyles compared to those in the upper ocean. Furthermore, niche-specific auxiliary metabolic genes were identified in the hadal viral genomes, and a novel viral D-amino acid oxidase was functionally and phylogenetically characterized, suggesting the contribution of these genes in the utilization of refractory organic matter. Together, these findings highlight the genomic novelty, dynamic movement, and environment-driven diversification of viral communities in oceanic trenches, and suggest that viruses may influence the hadal ecosystem by reprogramming the metabolism of their hosts and modulating the community of keystone microbes.

## Introduction

Believed to be the most abundant ( $\sim 10^{30}$ ) biological agents in the ocean, viruses play an important roles in microbial

metabolism, mortality, and nutrient recycling in marine ecosystems, thereby significantly influencing geochemical cycles on our planet [1–3]. Leveraged by global oceanographic expedition sampling and constantly updated viromic methods, the diversity, distribution, and ecological functions of viruses in the upper ocean have been relatively well characterized [4–7]. Notably, viruses have been demonstrated to play an essential roles in controlling prokaryotic mortality, releasing labile organic matter, and compensating for host metabolic pathways in deep-sea ecosystems [8–13], and the decomposition of viral particles provides an unignorable contribution to nutrient cycling in the benthic environment [14]. Furthermore, the abundance, taxonomic composition, living strategy, and genomic features of benthic viruses have been successively revealed [8, 9, 15–21]. Despite this progress, the viral community in the hadal zone (with a water depth exceeding 6000 m), which represents the deepest part of the ocean, is largely unexplored.

Exclusively comprising oceanic trenches, the hadal zone, which covers only 1–2% of the global deep ocean floor but

---

These authors contributed equally: Huahua Jian, Yi Yi, Jiahua Wang

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41396-021-00994-y>.

✉ Xiang Xiao  
zjxiao2018@sjtu.edu.cn

- <sup>1</sup> State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China
- <sup>2</sup> Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China
- <sup>3</sup> CAS Key Laboratory for Experimental Study under Deep-sea Extreme Conditions, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, China

accounts for up to 45% of the vertical depth of the ocean, represents one of the least studied environments on the planet [22, 23]. The hadal zone is characterized by multiple extreme environmental conditions, such as high hydrostatic pressure, low temperature, and deficiency of labile nutrients [23, 24]. Nevertheless, recent studies have shown that the abundance and activity of microorganisms in this ecosystem are unexpectedly high compared with those in the abyssal plain, indicating a considerable role of hadal microbes in the biogeochemical processes of the ocean [25]. Facilitated by their unique V-shaped trench topography and frequent earthquakes, hadal trenches are believed to be trappers of sedimentary organic matter from the upper ocean and neighboring abyssal areas, thus supporting endogenous cycles of organic carbon and nitrogen [26]. Recently, hydrocarbon-degrading bacteria and ammonia-oxidizing Thaumarchaeota have been revealed as predominant microbial clades in the Mariana Trench, and are thought to play important roles in this extreme habitat [27, 28]. Although diverse novel bacterial species have been isolated and characterized from the hadal biosphere [29–37], only one virus, PstS-1, which was integrated into the genome of the psychrotolerant bacterium *Pseudomonas stutzeri* 1–1–1b, has been isolated from this environment to date [38]. Therefore, the scarcity of viral genomes has severely hindered our understanding of the ecological function of viruses in the hadal biosphere.

In pioneering work on the viral communities in hadal pelagic sediment, ssDNA viruses were demonstrated to be the dominant group among the taxonomically assigned viral populations [39]. However, these results were relatively inconclusive, and the analysis was limited by methodological biases, scarcity of viromic data and the short lengths of the assembled viral contigs [39]. Recently, hadal trenches were found to probably feature highly dynamic viral infections, which can boost prokaryotic biomass production and organic material cycling in these remote and extreme environments [40]. Prompted by these clues implying a high level of undiscovered diversity and potential ecological functions of hadal viruses, we performed a comprehensive analysis of both seawater and sediment viral communities in oceanic trenches and aimed to characterize the distribution, genomic features, lifestyle, and ecological function of viruses inhabiting the hadal biosphere.

## Materials and methods

### Oceanic trench microbial metagenomes and genome collection

A total of 19 publicly available oceanic trench metagenomes were retrieved from the NCBI SRA database in June

2019. The 19 metagenomic samples were derived from the Mariana, Yap, and Kermadec Trenches in the Pacific Ocean, covering water depths from 0 to 10,500 m and including seawater (free-living and particle-associated) and sediment samples (Supplementary Table S1). For the seawater samples from Mariana Trench, the free-living and particle-associated fractions were collected on the 3- $\mu$ m and 0.22- $\mu$ m filters, respectively [28]. The 0.22- $\mu$ m filters were also utilized for the seawater samples from the Yap Trench [41]. For the sediment samples from these three trenches, the total DNA was extracted without pre-filtration and thus considered to retain both microbial and viral DNA [41–43]. All the metagenomes were sequenced on the Illumina HiSeq platform, and each one had a raw data size of >40 Gb, yielding a total size of 1.9 Tb. Additionally, a total of 108 archaeal and bacterial genomes, including microbial isolate genomes, metagenomic bins, and single-cell genomes, derived from hadal seawater, sediment and amphipod samples from depths  $\geq$  6000 m, were retrieved from the NCBI SRA database in March 2020 (Supplementary Table S2).

### Metagenomic assembly

Metagenomic raw reads were trimmed using Sickle (v1.33) with the default parameters [44]. We assembled each sample metagenome using IDBA\_UD [45] with k-mer-min 20, k-mer-max 150, and k-mer-step 6. Because the quality of assembled contigs significantly influenced the accuracy of subsequent viral recovery [46–48], we endeavored to optimize this procedure by a multiple-assembly strategy. Specifically, the “--min-count” parameter was set to 2, 5, 10, 15, and 20; that is, the raw reads of each metagenome were assembled five times with five different minimum coverage thresholds. Using this strategy, contigs from different communities with different coverage levels could be assembled with a low level of interference, and abundant populations could be better recruited. Indeed, the optimized assembly procedure significantly increased the number (57–1669) and total length (0.96–16 Gb) of retained dereplicated viral contigs and generally improved their N50 ( $\approx$ 1.7 kb for each sample) (Supplementary Table S3).

### Viral contig identification, dereplication, and virus operational taxonomic unit (vOTU) clustering

All assembled contigs longer than 5 kb were first dereplicated by CD-HIT (v4.8.1) [49] with the parameters “-aS 0.95 -c 0.95”. The open reading frames (ORFs) of non-redundant contigs were predicted by Prodigal (v2.6.3) [50] with “-p meta”. To thoroughly recover the putative viral contigs, three commonly recognized tools

and pipelines, namely, VirSorter (v1.0.5) [46], VirFinder (v1.1) [47], and the JGI Pipeline [51], were used in this study. Contigs sorted into categories 1, 2, 4, and 5 by VirSorter and with scores  $\geq 0.7$  and  $p < 0.05$  by VirFinder were selected for further curation, based on previous benchmarking evaluation of these two tools [46, 47]. For the JGI Pipeline, the default three filters (five hits to viral protein families combined with other criteria) were chosen, because they have previously been shown to have a high precision (99.6%) for viral contig identification [52]. In addition to its original workflow, we added a detection procedure to scan each local region of contigs harboring ten ORFs and select putative viral regions with the same criteria as previously described [51]. This procedure (the optimized JGI Pipeline) was designed to increase sensitivity to proviruses and contributed to the recovery of an additional 175 vOTUs (10.3% of proviral OTUs) with an average length of 25.2 kb. For viral prediction from 108 hadal microbial genomes, besides the above pipelines, Prophage Hunter was performed in default parameters, and prophages with scoring  $> 0.5$  were selected according to the previous performance evaluation [53]. In summary, by combining the use of these tools and pipelines (Supplementary Fig. S1), the number and length range of the recovered viral contigs were significantly increased (Supplementary Fig. S2).

The ORFs in the resultant contigs were annotated by multiple databases and bioinformatics tools as follows: RefSeq Virus database (v95, July 8, 2019) [54] using DIAMOND (v0.9.25.126) [55] with an  $e$ -value of  $1e^{-5}$ , identity of 30% and coverage of 50%; eggNOG (v5.0) [56] using emapper [57] with “-m diamond --query-cover 50”; pVOGs [58] using HMMER (v3.2.1) [59] with an  $e$ -value of  $1e^{-5}$ ; Kyoto Encyclopedia of Genes and Genomes (KEGG) database [60] using online GhostKOALA [61]; and Pfam [62] using RPS-BLAST with an  $e$ -value of  $1e^{-5}$ . All candidate viral contigs, especially the “low-confidence” contigs that were located in categories 2 and 5 by VirSorter and scored  $< 0.9$  by VirFinder, were manually curated. For contigs predicted to be proviruses, the boundaries were carefully evaluated, and only the proviral regions were retained. Consequently, 55.5% of the “low-confidence” contigs were discarded from the total candidate viral contigs (57.4% per sample on average) after manual curation (Supplementary Table S4).

All curated viral contigs sharing identity  $\geq 95\%$  and coverage  $\geq 85\%$  were dereplicated by CD-HIT [49]. Then, these viral contigs were further grouped into vOTUs as previously proposed [63, 64] by pyani (v0.2.7) [65] with the following parameters: at least 95% mummer-based average nucleotide identity (ANIm) over at least 85% of the length of the shorter sequence. Finally, we obtained 12,710 vOTUs that constituted the oceanic trench virome (OTVGD),

among which 4938 vOTUs contained contigs derived from hadal seawater and sediment samples (Supplementary Table S5). The longest contig of each OTU was regarded as the representative contig and was used for subsequent vOTU-related analysis. The circular viral contigs were identified by ViRal and Circular content from metAgenomes (VRCA) as previously described [66] (Supplementary Table S6).

### Viral taxonomic assignment and network analysis

A majority-rules approach was adopted to assign viral taxonomy as previously described [5]. In brief, all proteins from viral contigs were subjected to BLASTp alignment against RefSeq Virus, and a viral contig was considered to belong to a viral family if  $\geq 50\%$  of the proteins were assigned to that family with a bitscore  $\geq 50$ . Additionally, ViralRecall [67] was used to determine the taxonomy of nucleocytoplasmic large DNA viruses (NCLDVs) with a score threshold of 1.5. Protein-sharing network analysis of viral populations was performed by vConTACT (v2.0) [68]. Briefly, the protein sequences of the vOTUs were grouped into PCs via all-to-all BLASTp with the default parameters of vConTACT (v2.0) [68]. The degree of similarity between the vOTUs was calculated based on the number of shared PCs. Then, pairs of closely related vOTUs with a similarity score of  $\geq 1$  were grouped into viral clusters. The networks were visualized by Cytoscape (v3.7.0) [69] using an edge-weighted spring-embedded model.

### Protein clustering and comparison with public viral databases

All the viral proteins were clustered by CD-HIT [49] using the same parameters as the Pacific Ocean Virome (POV) [70]: 60% identity, 80% coverage and “-g 1 -n 4 -d 0”. In total, 159,855 PCs were obtained for OTVGD. To describe the novelty of the OTVGD, the vOTUs and proteins were compared against three public viral databases, including the RefSeq Virus database [54], Global Ocean Viromes 2.0 (GOV 2.0) [5] and IMG/VR v2.0 [64]. For vOTU comparison, pyani (v0.2.7) [65] was used with parameters of 60% identity and 80% coverage. For protein comparison, BLASTp was performed by DIAMOND [55] with an  $e$ -value threshold of  $1e^{-5}$ , identity of 30% and coverage of 50%.

For the recruitment analysis, virome reads for the Pacific Ocean seawater (POV) [70], the North Atlantic Deep Water (NADW) and Pacific Antarctic Bottom Water (PABW) [71] and for the sediment of the Izu-Ogasawara and Mariana Trench [39] were recruited against OTVGD using Blastn, and the hits with an  $e$ -value  $< 10^{-5}$ , identity  $> 95\%$  and hit

length >50 bp were retained for reads recruited per kb of genome per Gb of metagenome (RPKG) calculation as previously described [72].

### Calculating the relative abundances, diversity, GC content, and N/C-ARSC of vOTUs

The trimmed reads of each sample were mapped to the representative OTV contigs using Bowtie 2 (v2.3.4.1) [73] with default parameters. A BamM (v.17.3, <https://github.com/ecogenomics/BamM>) filter was used to screen the reads mapped to the contigs with coverage  $\geq 90\%$  and identity  $\geq 95\%$ . The viral contigs with  $\geq 70\%$  of their length covered by the reads were selected by genomeCoverageBed in BEDTools (v2.26.0) [74] and passed to the next step. The average per-base-pair coverage of the contigs was calculated by BamM “parse” with the parameter “tpmean” to remove the top and bottom 10% coverage regions. We then performed normalization among the samples as previously described [75]. Briefly, the average coverage of each contig was first divided by the total trimmed read number of the sample from which it was derived, and then multiplied by the average read number of all 19 samples. The normalized average coverage of the viral contigs was used to generate a matrix of relative vOTU abundance (Supplementary Table S7). This matrix was clustered and visualized by the heatmap R package with the default parameters.

Alpha diversity (Shannon’s H) analysis of each samples was performed by VEGAN in R package [76]. Based on Bray–Curtis dissimilarity matrices, which were calculated using the VEGAN function `vegdist`, principal coordinate analysis (PCoA) was performed using the `pcoa` function in the APE package, and ANOSIM using the VEGAN function `anosim` was performed to test the significance of dissimilarity between groups, as previously described [5]. The python script “get\_gc\_and\_narsc.py” (<https://github.com/fayward/pangenomics/>) was used to calculate GC content and nitrogen/carbon atoms per residue side chain (N/C-ARSC) of the predicted viral proteins in OTVGD using previously described methods [77].

### Quantification of viral movement

Viral movement between neighboring samples was calculated according to a previously described method [7] with a minor modification. Because of our optimized assembly method, which markedly increased the length of the viral contigs, the longest contig of each vOTU did not necessarily indicate its original sample. Considering that Brum’s “origin assumption” was significantly supported by the result that viral populations were most abundant in their original samples [7], the sample in which a vOTU showed

the highest abundance was considered its original sample in the present study.

### Metagenomic binning and microbial OTU grouping

To strictly avoid the shifting of microbial tetranucleotide frequencies and/or the calculation of coverages influenced by exogenous or exceptional sequences, the predicted viral sequences, rRNA genes, and CRISPR segments were temporarily removed from their derived contigs (which might have different nucleotide compositions from their derived genomes and/or occur in multiple copy numbers). Pre-processed contigs longer than 2.5 kb were used to calculate contig coverages and perform binning. The average coverage of microbial contigs were calculated by the same method used for viral contigs. The t-SNE algorithm-based dimension reduction from the tetranucleotide frequency matrix and visualization were performed using the R package `mmgenome2` (<https://github.com/KasperSkytte/mmgenome2>) [78], and manual binning was adopted to carefully divide the boundaries between metagenome-assembled genomes (MAGs). Rebinning was performed by plotting all the contigs from each bin based on their abundances plus GC content, and contigs with inconsistent coverage were manually deleted. Notably, the removed viral, rRNA and CRISPR sequences were restored to their original contigs after binning. The completeness and contamination of the MAGs were estimated using CheckM (v1.1.2) [79], and MAGs with <30% completeness or >10% contamination were removed. The average amino acid identities (AAI) among all the bins were calculated using `pyani` (v0.2.7) [65], and the MAGs sharing a >95% AAI and 30% alignment length to their query sequence were grouped into microbial OTUs (mOTUs). In each sample, the MAG with the highest value of completeness-4 $\times$  contamination in each mOTU was chosen as the representative (Supplementary Table S8), for which the total average per-base-pair coverage of all binned contigs, notably not including provirus contigs, was used as a proxy of mOTU abundance.

### Host prediction

Two pipelines were used to identify the “ex situ” and “in situ” hosts for the viral contigs. The former refers to a method of predicting putative hosts in public databases that has been widely used in previous studies [4, 6, 52]. The latter refers to matching viral contigs to metagenomic bins derived from the same sample, thus linking each virus to a putative “in situ” host. This strategy was utilized by Emerson et al. in a study on soil viromes, and has been confirmed to significantly enhance the efficiency of host prediction [75]. Both pipelines involved multiple *in silico*



approaches, including tRNA matching, CRISPR matching, BLASTn-based genomic alignment and *k*-mer-based genomic matching. Specifically, (1) the tRNA genes of vOTUs and MAGs were identified by tRNAscan-SE (v2.0.3) [80] with the parameters “-A” and “-B”, and the viral tRNA genes were aligned with both MAG-derived and tRNAviz database [81]-derived sequences using BLASTn. Then, perfect matches for non-gammaproteobacterial hosts were selected [52]. (2) The CRISPR clusters of MAGs in situ and all the bacterial and archaeal genomes in the RefSeq database (v95, July 8, 2019) were predicted using the CRISPR Recognition Tool (CRT) [82] with optimized parameters as previously described [52], and CRISPR clusters with less than three spacers were ignored. The retained CRISPR spacers were aligned with the viral contigs using BLASTn to identify protospacers in the viral contigs, and matches satisfying the thresholds of  $\geq 95\%$  identity and  $\leq 2$  SNPs were selected [52]. (3) For direct genomic BLASTn matches, the viral sequences were aligned with in situ metagenomic bins and all RefSeq microbial genomes using BLASTn with the following parameters: bitscore  $\geq 50$ , *e*-value  $\geq 10^{-3}$ , identity  $\geq 70\%$ , and matching length  $\geq 2500$  bp, as previously described [75, 83]. (4) Then, *k*-mer-based host prediction was performed using VirHostMatcher (VHM) [48]. For “ex situ” host prediction, the VHM  $d_2^*$  values of each viral sequence versus 245,723 bacterial and 2837 archaeal genomes in GenBank (v234, October 15, 2019) were calculated, and the most frequent genus among the top 30 hits with  $d_2^*$  values  $< 0.25$  was considered the genus of the microbial host, as previously suggested [48]. For “in situ” host prediction, the VHM  $d_2^*$  values of each viral sequence were further compared with those of MAGs from the same sample. The MAG was considered the “in situ” host if it ranked among the top ten hits with  $d_2^*$  values  $< 0.25$ .

Additionally, we developed a “flanking sequence binning” approach for the “in situ” host prediction of proviruses. Specifically, the flanking sequences of proviruses were reserved for binning processes, and proviruses and hosts matched when they were binned together. This approach was designed to avoid the effects of proviral fragments on the tetranucleotide frequencies, GC% and coverage calculation of the derived microbial contigs, and it consequently contributed to the identification of 201 pairs (29.3% of total assignments) of in situ virus–host matches (Supplementary Fig. S3). Finally, a total of 685 “in situ” and 2320 “ex situ” hosts were identified by these two pipelines (Supplementary Table S9). It should be noted that all taxonomic classifications of the “in situ” and “ex situ” hosts were unified by GTDB-Tk [84] based on 120 and 122 proteins concentrated in the bacterial and archaeal genomes, respectively, to avoid taxonomic inconsistency between NCBI and GTDB. The aligned concentrated proteins were

used to construct a host phylogenetic tree in FastTree 2 (v2.1.10) [85] based on the maximum-likelihood algorithm, and the tree was visualized with MEGA X [86]. For those vOTUs ( $n = 424$ ) with multiple predicted hosts resulted from different predication approaches, the host assignments had a high consistency of 90.3% and 88.9% at the phylum and class level, respectively (Supplementary Fig. S3). Additionally, as viral generalists that infect hosts across different phyla have been suggested to exist [52, 87], the vOTUs ( $n = 47$ ) with different predicted host assignments at the phylum or class level were retained (Supplementary Table S9).

### Identification of (putatively active) proviruses and lysogens

To detect proviruses in the OTVGD, we established two criteria: (1) viral contigs originating from contigs with non-viral (host) flanking sequences, or (2) viral contigs harboring lysogenic marker proteins, which include integrase, invertase, serine recombinase, and CI/Cro repressor, as previously proposed [88]. Finally, a total of 1704 proviruses in the OTVGD were identified by this screening. Accordingly, the microbial OTUs that harbored proviruses were regarded as lysogens. Active proviruses were defined by a higher abundance of provirus sequences than the microbial host ( $VHR > 1$ ), indicating active DNA replication of the provirus [89]. Likewise, the microbial OTUs that harbored active proviruses were regarded as active lysogens.

### Viral auxiliary metabolic gene (vAMG) identification and analysis

A comprehensive annotation of ORFs in the viral contigs was performed based on the RefSeq Virus database [54], KEGG [60], and eggNOG [56]. All the proteins were searched against the Prokaryotic Virus Orthologous Groups (pVOGs) database [58] by HMMER (v3.2.1) [59], and those with scores  $\geq 50$  were marked as conserved viral proteins, as previously suggested [63]. In addition, those proteins annotated as capsid, tail, lysozyme, and terminase were also set as conserved viral proteins [90]. For proviruses, two marker proteins (integrase and invertase) were set as boundary signals in the contigs. Subsequently, the conserved viral regions for vAMG selection were identified based on two criteria: (1) both the start and end genes were annotated as conserved viral proteins or boundary signals; (2) the viral region harbored at least one conserved viral protein or more than 30% proteins had the highest similarity to viral proteins according to the annotation. The selected conserved viral regions were further manually curated. All the metabolic proteins from these regions were considered to be putative vAMGs, and a final list was determined by

manual curation. Furthermore, type I and type II vAMGs were classified according to the previous definitions [13, 91] (Supplementary Tables S10 and S11). For vAMG abundance calculation, the sum of the abundance of AMG-derived viral contigs rather than the reads mapping to these genes was regarded as the AMG abundance to ensure that the resultant abundance value was unambiguously derived from viruses.

For the phylogenetic analysis of D-amino acid oxidase (DAO), homologs of DAO were retrieved from in situ MAGs, the KEGG database and Swiss-Prot. Clustal Omega (v1.2.4) [92] was used to perform multiple sequence alignment, and the aligned concentrated proteins were used to construct a phylogenetic tree with FastTree 2 (v2.1.10) [85] based on the maximum-likelihood algorithm. The tree was visualized by MEGA X [86]. To detect the abundance of DAO-encoding genes (*dao*) in oceanic trenches, the *dao* genes were retrieved from metagenomes by BLASTp with the following parameters: identity  $\geq 30\%$ , coverage  $\geq 50\%$ , *e*-value  $\geq 10^{-3}$ . The percentage of mapped reads to the *dao* genes among the total reads in each sample was calculated to determine their relative abundances.

### Functionally characterization of viral D-amino acid oxidase (vDAO)

The gene encoding for vDAO (*daoV*) from the hadal virus OTV\_OTU6572 was synthesized and cloned into the pET28a plasmid (Personalbio Technology, Shanghai, China). Then, the *daoV* gene was amplified by PCR and inserted into the pCold TF vector (Takara Bio, Dalian, China) using the ClonExpress II One Step Cloning Kit (Vazyme Biotech, Nanjing, China). *E. coli* BL21 (DE3) cells were transformed with this recombinant plasmid. The transformant was selected on LB medium containing ampicillin and confirmed by PCR and DNA sequencing. *E. coli* BL21 (DE3) cells containing the *daoV* expression vector were grown in LB broth with 50  $\mu\text{g}/\text{ml}$  ampicillin at 37 °C. vDAO expression was induced by the addition of IPTG (0.4 mM) when the OD<sub>600</sub> reached 0.6, and the culture was then incubated at 15 °C for 16 h. The cells were collected by centrifugation, resuspended in binding buffer [500 mM NaCl, 10% glycerol, and 20 mM Tris-HCl (pH 8.0)], and sonicated on ice. The cell extract was clarified by centrifugation at 10,000  $\times g$  for 40 min at 4 °C. Ni Sepharose High Performance resin (GE Healthcare, Milwaukee, WI, USA) was used to purify the TF-tagged vDAO according to the manufacturer's instructions. The protein was eluted in elution buffer [500 mM NaCl, 10% (v/v) glycerol, 300 mM imidazole, and 20 mM Tris-HCl (pH 8.0)]. The purity of protein was confirmed by SDS-PAGE (12% acrylamide) with visualization using

Coomassie Brilliant Blue R-250, and the concentration of the purified protein was determined using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, USA). The enzyme activity of D-amino acid oxidation was determined using a coupled *o*-dianisidine/peroxidase method, as previously described [93]. Briefly, the reaction mixture contained 1.3 mM *o*-dianisidine, 150 U/ml peroxidase, 40  $\mu\text{M}$  FAD, and 40 mM D-amino acid (D-Met, D-His, D-Ala, and D-Glu respectively) in 50 mM potassium phosphate buffer (pH 7.5) at 25 °C. Enzyme activity was determined colorimetrically at 430 nm using a Synergy H1 microplate reader (BioTek, Winooski, USA). The H<sub>2</sub>O<sub>2</sub> and purified TF-Tag protein were used in the same reaction mixture as the positive and negative controls, respectively.

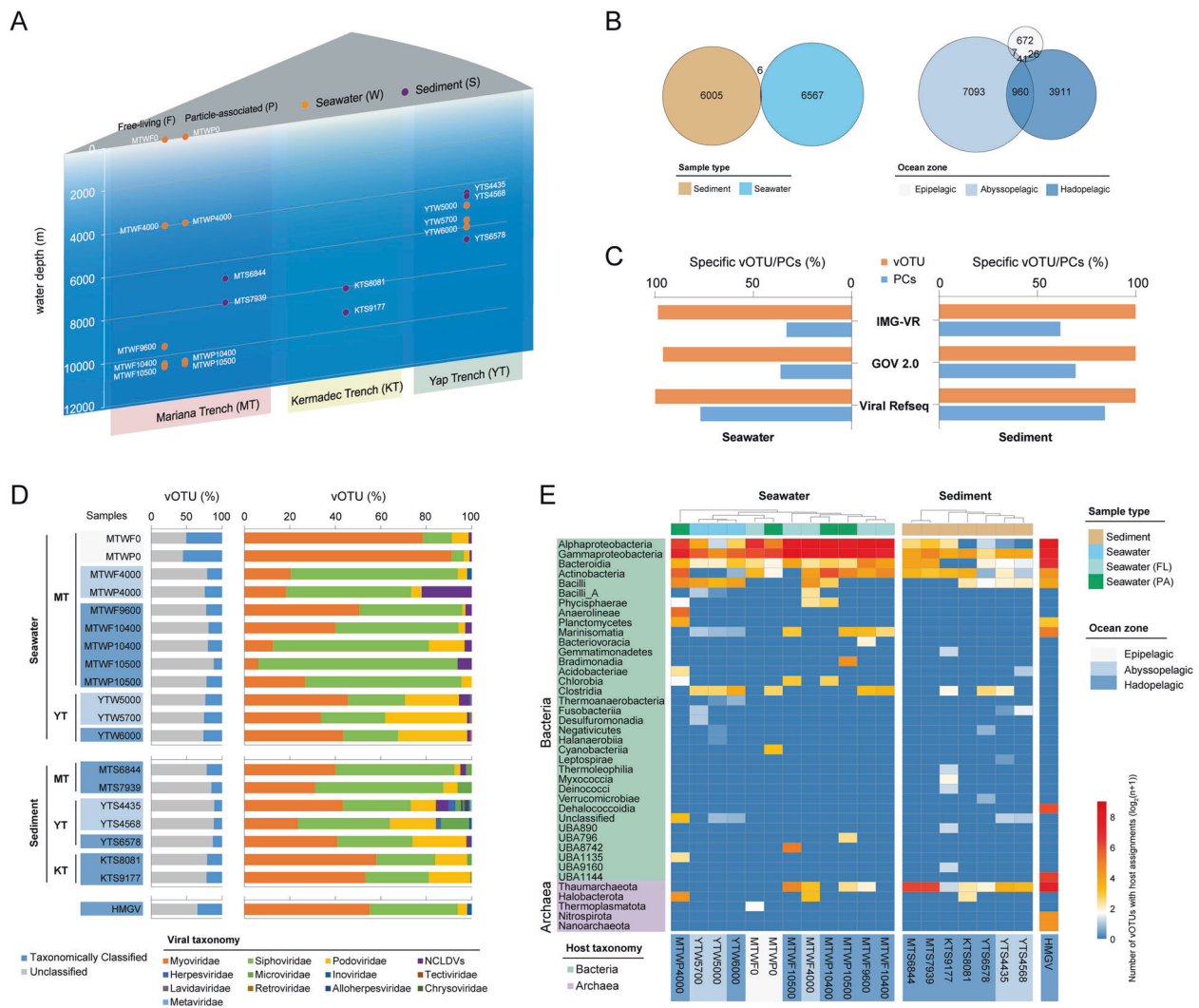
### Availability of data

The data sets that were used for analysis in this study are publicly available in the NCBI repository at <https://www.ncbi.nlm.nih.gov/>. The sequences of the vOTUs generated from the current study have been deposited in the National Omics Data Encyclopedia (NODE) database under the project IDs OEP001086 and OEP001087.

## Results and discussion

### Construction of the oceanic trench viral genome dataset (OTVGD)

Initially, the raw reads of 19 publicly available metagenomes, derived from seawater and sediment samples of three oceanic trenches (the Mariana, Kermadec and Yap Trenches) (Fig. 1A), were collected from the NCBI Sequence Read Archive (SRA) (Supplementary Table S1). Samples obtained from the Mariana Trench covered seawater (free-living and particle-associated) and sediment environments, from 0 to 10,500 m. For the Yap Trench, the depth of the seawater and sediment samples ranged from 4435 to 6578 m, while both sediment samples from the Kermadec Trench were from deeper than 6000 m. Reads were trimmed and assembled to obtain total contigs, and then, viral contigs were identified using multiple tools, such as VirSorter [46], VirFinder [47], and JGI Pipeline [51], as well as manual curation (Supplementary Figs. S1, S2 and Tables S3, S4). Then, we successfully recovered 17,063 viral contigs that were  $\geq 5$  kb in size. Additionally, 141 viral contigs were identified from 108 (complete/partial) public hadal microbial genomes (Supplementary Table S2). They were combined to define a final set of 12,710 virus operational taxonomic units (vOTUs), which represented a species-level



**Fig. 1 Overview of the oceanic trench viral genome dataset (OTVGD).** **A** Schematic diagram showing the sampling locations and sample types of the metagenomes used in this study. **B** Venn diagram displaying the number of vOTUs composing the OTVGD. **C** Percentages of specific vOTUs and protein clusters within the OTVGD by comparison with three public datasets: NCBI Viral RefSeq, IMG-VR, and GOV 2.0. **D** Taxonomic compositions of the trench viruses grouped by sample types, ocean zones, and trenches. The bar charts on the left indicate the abundance (%) of vOTUs with predicted taxonomy (blue), and the bar graphs on the right show the abundance (%) of viral

taxa at the family level in viruses with predicted taxonomy. The bar at the bottom corresponds to the vOTUs that were derived from hadal microbial genomes (HMGV). **E** Distribution of the predicted bacterial and archaeal hosts of the trench viruses. The heatmaps show the number of vOTUs with host assignments, which are hierarchically clustered by samples. The bacterial and archaeal host taxonomy shown are in the class and phylum level, respectively. The bars on the top and bottom of the figure indicate the type and ocean zone of each sample, respectively. FL free living, PA particle associated.

taxonomy as previously proposed for uncultivated virus genomes [63], and designated the oceanic trench viral genome dataset (OTVGD) (Supplementary Table S5). Notably, among all the vOTUs in OTVGD, a set of 6011 and 6573 vOTUs was recovered from seawater and sediment, respectively, and only six vOTUs were concurrently present in both sample types (Fig. 1B), indicating distinct viral communities in the oceanic water column and sedimentary environments. Interestingly, 1027 vOTUs, which accounted for 20.8% of the total set of 4938 vOTUs recovered from hadopelagic

metagenomes and microbial genomes, were also present in non-hadal environments, suggesting a close connection between the hadal viral populations and those in the upper ocean (Fig. 1B).

In comparison with the previously reported hadal trench virome [39], this dataset increased the total and average genome size of oceanic trench viruses by 25.4-fold and 37.1-fold, respectively (Supplementary Fig. S4A), providing a reliable basis for the detailed subsequent analyses. Read recruitment of OTVGD was performed against the previously published viromes derived from the deep

seawater of the Pacific Ocean [70, 71]. Additionally, similar recruitment ratios were observed between the neighboring seawater samples, while the ratio was significantly divergent among different sediment samples (Supplementary Fig. S4B), implying distinct levels of heterogeneity of viral communities in the seawater and sediments of oceanic trenches.

### Novel viruses are enriched in oceanic trenches

The similarity of the oceanic trench viral populations to NCBI Viral RefSeq [94], GOV 2.0 [5], and IMG/VR v.2.0 [64] was investigated at the OTU and protein levels. Remarkably, over 99% of the vOTUs of the OTVGD were distinct from those of these three datasets (Fig. 1c). Moreover, with a higher proportion of previously undiscovered protein clusters (PCs) in sediment than in seawater, the majority of the OTVGD PCs (159,496, 54.15%) were demonstrated to be specific by comparison with GOV 2.0, which is by far the largest marine virus database, harboring 823,193 PCs [5]. Moreover, the vOTU accumulation rate analysis showed that trench viruses were far from well sampled and remained largely uncharacterized (Supplementary Fig. S5). Taken together, these results indicated that remarkably novel viruses were enriched in the oceanic trenches.

In agreement with previous marine viromic studies, the majority of the seawater (73.84%)-derived and sediment (84.82%)-derived vOTUs could not be taxonomically classified into any known viral family. Among the taxonomically classified vOTUs, dsDNA viral families such as *Myoviridae*, *Siphoviridae*, and *Podoviridae* were predominant, accounting for 89.85% and 62.87% of the seawater and sediment-derived vOTUs, respectively (Fig. 1D). Intriguingly, 77 vOTUs could be classified as nucleocytoplasmic large DNA viruses (NCLDVs) (Supplementary Table S6), and these vOTUs were enriched in the abyssal and hadal samples of the Mariana and Yap Trenches (Fig. 1D). Although we could not rule out the possibility that these giant viruses originated from the ocean surface through biomass sinking, the data strongly indicated a distribution of NCLDVs in the hadal biosphere. Interestingly, we also recovered two vOTUs with genomes of more than 200 kb (Supplementary Fig. S6), which could be classified into the recently revealed “huge phage” clades [95].

In total, 306 vOTUs were identified to contain circular viral contigs and, therefore, assumed to be complete (Supplementary Table S5). Among them, 49 vOTUs with putative complete genomes were detected exclusively in the hadopelagic zone. Notably, OTV\_OTU10144 (represented by four circular genomes, 6681–6901 bp) was classified as an inovirus. It was predicted to infect Alteromonadaceae strains according to CRISPR matching, and it was

exclusively identified in all five hadal seawater samples of the Mariana Trench. The phylogenetic analysis showed that it was located in a separate branch compared to all the currently known inoviruses (Supplementary Fig. S7), thus representing a novel inovirus population inhabiting hadal zones.

### Viruses target keystone species of the microbial community in the hadal biosphere

To investigate how these viruses affect hadal ecosystem function by infecting microbial hosts that participate in biogeochemical cycles, we next sought to identify viral hosts by screening all the prokaryotic genomes in GenBank (v234) using multiple in silico methods. To maximize host assignment and to link the viruses to hosts under natural circumstances, we further searched against a database composed of 379 bacterial and archaeal OTUs (consisting of 1959 genomic bins) recovered from the same metagenomes in which the viral contigs were identified (Supplementary Table S8). The hosts assigned through these two methods were designated “ex situ” and “in situ” hosts, respectively (Supplementary Fig. S7). Overall, the hosts could be predicted for 1229 (9.67%) of the total vOTUs of the OTVGD (Supplementary Table S9). Notably, the abundance of the viruses and their hosts was highly correlated (Pearson’s coefficient = 0.619,  $p < 2.2e-16$ ) (Supplementary Fig. S8), suggesting an underlying function of these viruses in regulating the microbial communities in oceanic trenches.

Among the predicted microbial hosts, Gammaproteobacteria and Thaumarchaeota were the dominant bacteria and archaea, respectively (Fig. 1E). Additionally, Alphaproteobacteria, Bacterioidia, and Actinobacteria were highly enriched according to the host prediction, especially in the hadal seawater and sediment samples. These results were consistent with the previously reported dominant microbial groups in the hadal seawater and sediment [41, 42]. Specifically, OTV\_OTV6518, represented by three putative complete genomes (40.24–40.71 kb), was predicted to infect *Oleibacter*, which has been shown to be the most abundant bacterial genus (19.5%) in the near-bottom water of the Mariana Trench samples [28]. Consistent with the host prevalence, this population showed a high abundance in the deepest hadopelagic samples (Supplementary Fig. S9). Notably, 27 vOTUs from both abyssal and hadal samples were predicted to be thaumarchaeal viruses, which exhibited a higher abundance in sediment than in seawater of trenches (Supplementary Fig. S10A, C). A protein-sharing network analysis indicated that these vOTUs were related to published marine Thaumarchaeota viral contigs [96], while they were distinct from the three isolated *Nitrosopumilus* spindle-shaped viruses [97]. Moreover, the



genomic comparison showed that the putative thaumarchaeal viruses from the OTV shared no similarity with the previously reported ones in terms of viral structural proteins (Supplementary Fig. S10B), implying their novelty and the unrevealed diversity of viruses infecting Thaumarchaeota, which is the dominant archaeal phylum in the hadal environment and plays a prominent role in carbon and nitrogen cycling by contributing to ammonia oxidization [26, 27, 98].

### Inter-trench and intra-trench exchange of viral communities

Interestingly, based on the calculation of the vOTU relative abundance for all samples, a large number of viral populations were observed to be concurrently present in different habitat samples, and we considered them to be exchange viral populations (Fig. 2 and Supplementary Table S7). Specifically, 815 hadal vOTUs were present in multiple trenches, and the co-occurred viral populations were accounted for 43.54 and 21.92% of the hadal vOTUs derived from the Mariana and Kermadec Trench, respectively (Fig. 2A), suggesting considerable occurrence of viral population exchange between different hadal trenches. To further support this hypothesis, we performed a protein-sharing network analysis of all the trench viruses. Surprisingly, they did not tend to cluster by trench, but in contrast, pervasive connectivity was observed among viral populations from different trenches (Fig. 2B). Additionally, a large number of unclustered vOTUs presented as singletons (Supplementary Fig. S11), further indicating that a large amount of viral diversity remains to be uncovered in oceanic trenches. Specifically, OTV\_OTU10869 (40.22 kb) with a complete genome was identified in both the Yap and Kermadec Trenches (Supplementary Fig. S12). These results indicated considerable widespread inter-trench connectivity of viral populations, despite the distinct geographical isolation.

To further demonstrate intra-trench exchange of viral communities, we separately conducted occurrence and network analysis of the vOTUs in the Mariana and Yap Trenches (Fig. 2C–F). In the Mariana Trench, a large group of viruses (751 vOTUs) was present in both hadal (free-living and particle-associated) and epi-/abyssopelagic seawater (Fig. 2C), and those at different depths were highly connected in terms of protein content (Fig. 2D), suggesting continuous genetic communication of viruses among the depth-stratified oceanic zones. Accordingly, host predictions across samples indicated that a large number of viruses could infect microorganisms inhabiting different water layers in the Mariana Trench (Supplementary Fig. S13). Similarly, the vOTUs that co-occurred in the abyssal and hadal samples dominated the viral

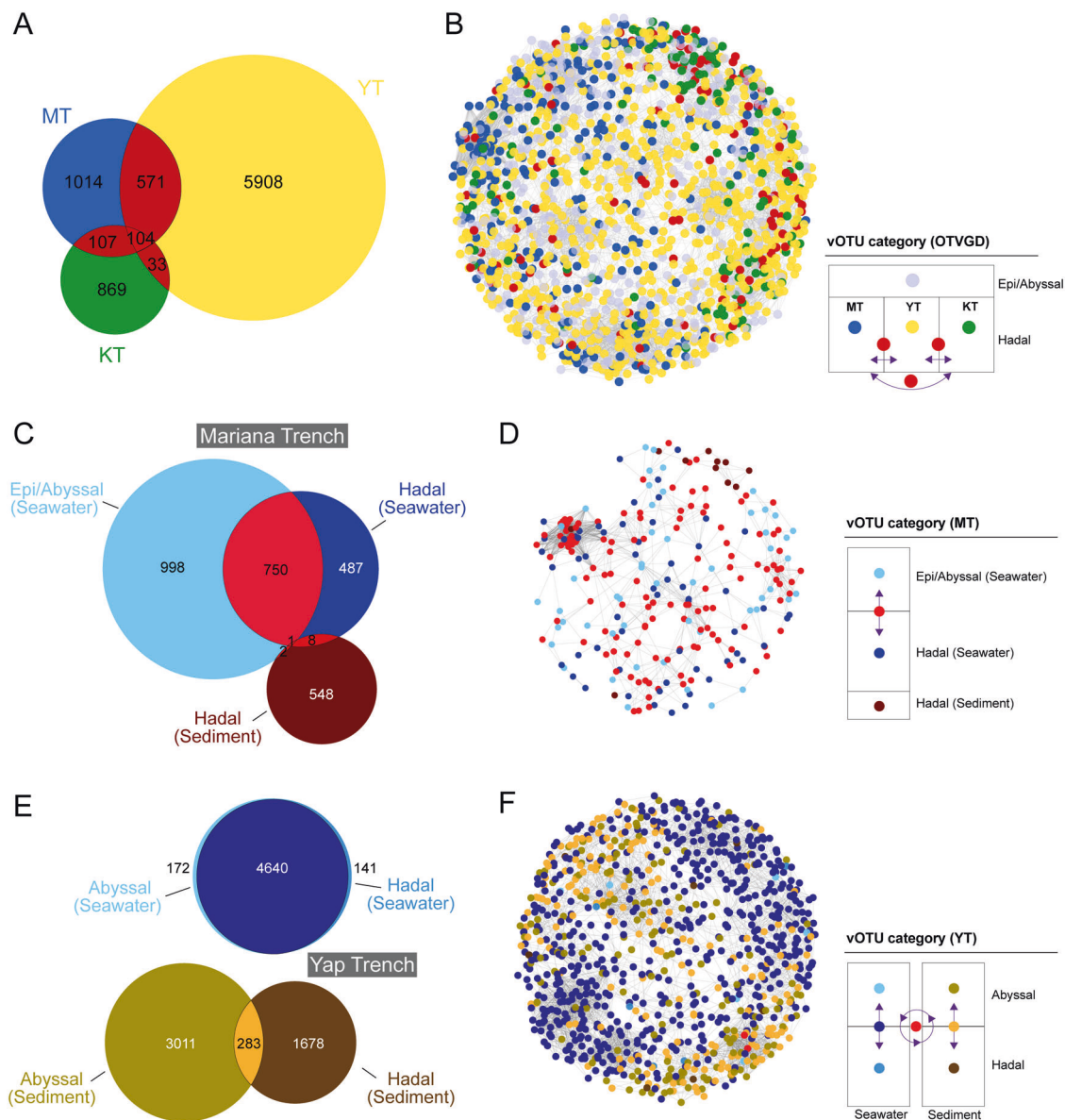
populations of the Yap Trench (Fig. 2E). Moreover, exchange of viruses in seawater seemed to occur significantly more than that in sediment and between seawater and sediment (Fig. 2E), and the sediment and seawater viral populations tended to cluster by themselves in terms of the protein content (Fig. 2F), implying the influence of a natural barrier on the connectivity of the two types of viral communities.

To further investigate to what extent viral communities were probably exchanged throughout the oceanic trenches, we compared the vOTU abundances of neighboring samples to quantitatively assess bidirectional exchange between viral populations (Supplementary Fig. S14). Intriguingly, this analysis demonstrated unexpectedly similar levels of viral population exchange in both directions, except between the sediments of the Mariana Trench, which might be due to frequent earthquakes and sediment slide in this hadal trench [99]. Overall, these data strongly indicate that inter-trench and intra-trench exchange of viral communities were considerably pervasive. Meanwhile, it should also be noted that this conclusion needs to be further supported by analyzing more viromic data derived from different trench samples, given the currently inadequate sampling of trench viruses (Supplementary Fig. S4).

Previously, oceanic current-driven movement of viruses and vertical viral transport from the marine surface to the deep chlorophyll maximum (DCM) zone have been observed in the upper ocean [7], and potential connectivity of microbial communities between geographically separated hadal sediment has been proposed [42]. In the present study, inter-trench and intra-trench exchange of viral communities were suggested to be pervasive, indicating that the transport of viruses also frequently occurred in the deep ocean, probably due to the downward flux of particulate organic matter [100, 101], upwelling of deep waters to the sea surface [102], and current flow and water mass circulation in the deep ocean [103, 104].

### Niche-dependent distribution, genomic properties, and lifestyle of viral communities in the oceanic trenches

The significant habitat heterogeneity, including degrees of geographical isolation, biochemical parameters, substrate compositions and topographical features, has been previously revealed at inter-trench and intra-trench levels [105]. Therefore, it is intriguing to explore the distribution profile of viruses among different oceanic trenches and in different biozones within trenches and to examine, whether the habitat heterogeneity influenced the genomic properties and lifestyle of the viruses. To explore these questions, the abundance patterns of vOTUs were first analysed for all samples (Supplementary Table S7). The viral communities



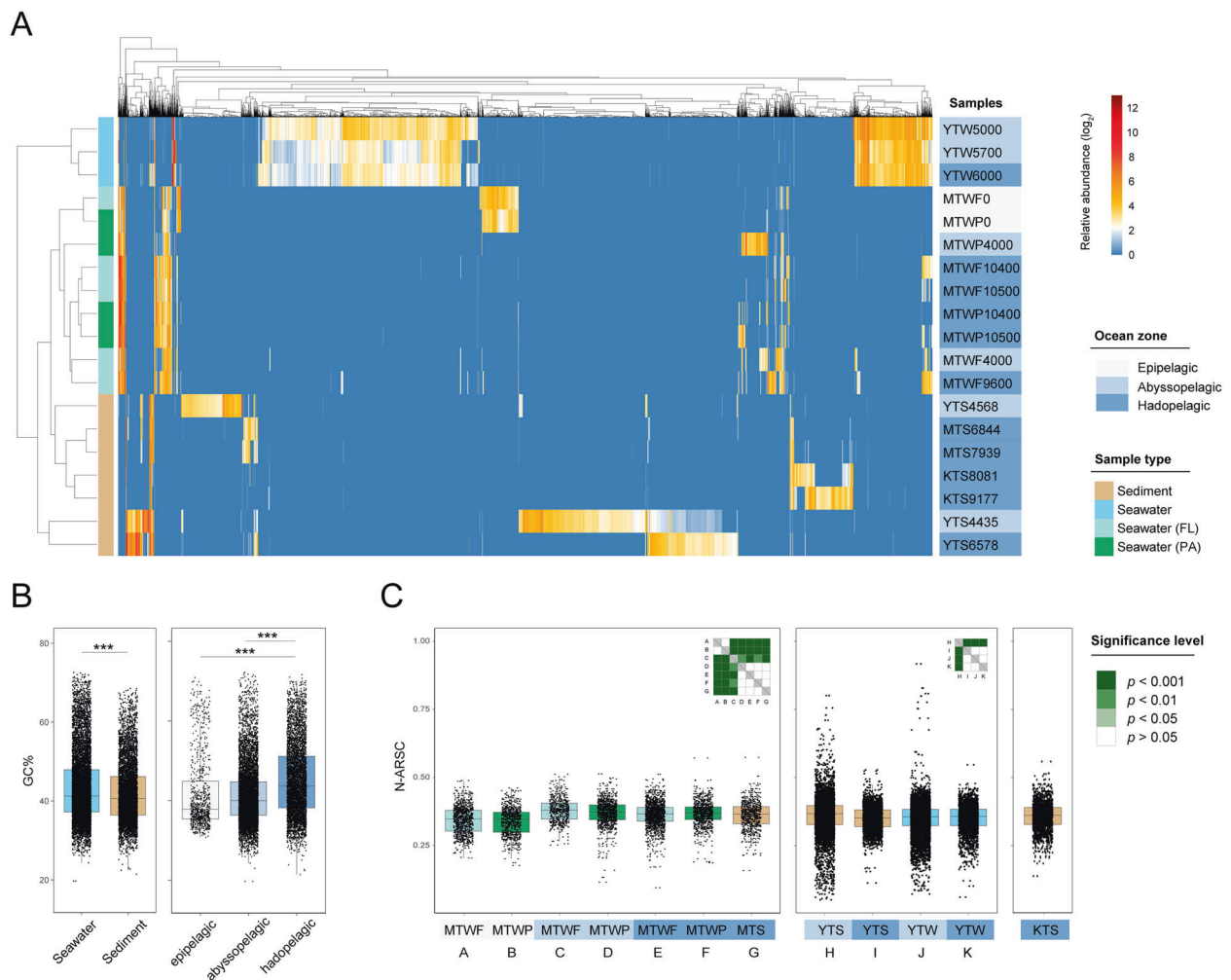
**Fig. 2** Connectivity of viruses in the oceanic trenches. **A, C, E** Venn diagrams displaying the inter-trench and intra-trench occurrence of vOTUs. The existence of viral populations in multiple trenches (**A**), the Mariana Trench (**C**), and the Yap Trench (**E**) were based on the relative abundance calculation of vOTUs in each sample. The number of vOTUs are indicated on different portion of each diagram. **B, D, F** Protein-sharing network of the vOTUs in OTVGD using vConTACT v2.0. The nodes and the connecting

edges represent vOTUs and their shared protein content, respectively. Nodes are depicted in different colors representing vOTUs derived from different trenches (**B**) and different ocean zones/sample types of the Mariana Trench (**D**) and Yap Trench (**F**). The co-occurrence of vOTUs in multiple habitats are indicated by purple arrows in the legend and nodes with distinct colors in the network. For clarity, only the major viral clusters are shown.

were primarily clustered by the sample types and secondarily by the trenches and ocean zones (Fig. 3A). According to the principal coordinate analysis (PCoA), the most significant difference in viral communities was observed between seawater and sediment (ANOSIM,  $r = 0.971$ ,  $p = 0.002$ ), followed by different trenches ( $r = 0.4715$ ,  $p = 0.003$ ) (Supplementary Fig. S15). Moreover, when considering only the seawater viruses in the Mariana Trench, depth-stratified ocean zones were the significant

differentiation factor ( $r = 0.9583$ ,  $p = 0.003$ ). Moreover, Shannon's H index showed that viral communities in the seawater of the Yap Trench exhibited the highest diversity, followed by those in the sediment of the Yap Trench, whereas those in the Mariana and Kermadec trenches showed a lower diversity at similar levels (Supplementary Fig. S15).

Since the genomic GC content of marine bacteria and viruses has been revealed to be significantly correlated

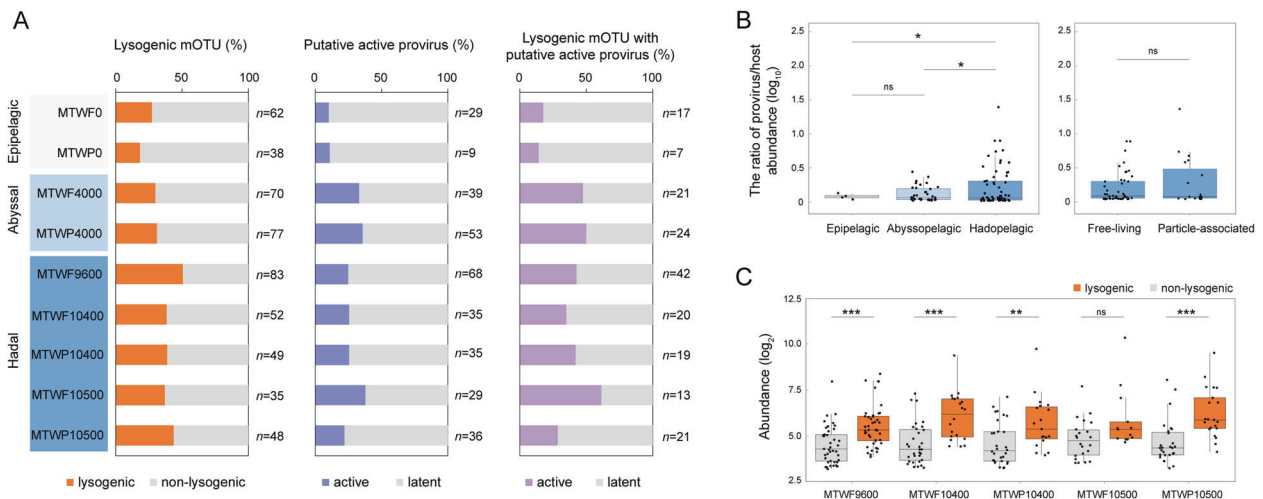


**Fig. 3** Abundance profiles and genomic features of oceanic trench viruses. **A** Distribution patterns of all vOTUs in the OTVGD. The heatmap displays the normalized average coverage of each vOTU (x-axis) in each sample (y-axis) on the log<sub>2</sub> scale. The vOTUs are hierarchically clustered by samples. The vertical bars on the left and right sides indicate the type and ocean zone of each sample, respectively. **B** GC content of viral communities in different sample types and ocean zones. The values were analysed by using a two-tailed Student's *t*-test,

\*\*\* $p < 0.001$ . **C** The nitrogen atoms per residue side chain (N-ARSC) analysis of viral genomes in OTVGD. All representative viral contigs in each sample were used for the calculation. Heatmaps in the top right corner of frames show significance levels of differences between all pairs of sample group calculated by the two-tailed Student's *t*-test. The bottom letters representing sample groups corresponded to coordinates of significance heatmaps.

with their habitats [16, 106], we then calculated the GC content of all vOTUs, and the results showed a significantly higher GC content ( $p < 0.001$ ) of vOTUs derived from seawater (average = 43.3%) than of those derived from sediment (average = 41.9%) (Fig. 3B). With respect to seawater samples, the GC contents of the viral communities from the hadal biosphere (average = 45.08%) was significantly higher ( $p < 0.001$ ) than that from epi- (average = 41.75%) and abyssopelagic zones (average = 41.31%) (Fig. 3B). Moreover, the viral GC distribution pattern in hadal seawater and sediment were significantly different from that in the upper ocean (Supplementary Fig. S16), as shown by comparison with the GOV 2.0 datasets [5], suggesting depth-dependent

evolutionary differentiation of viral genomes. Considering that GC pairs use more nitrogen than AT pairs [107] and the higher nitrogen availabilities in deeper waters in the ocean [106], these differences were consistent with the nitrogen atoms per residue side chain (N-ARSC) analysis, which showed a significantly higher N-ARSC of viral genomes from abyssopelagic and hadal zones than of those from the trench surfaces (Fig. 3C). In addition, the carbon atoms per residue side chain (C-ARSC) of sediment viruses were generally higher than those of viruses from neighboring seawater (Supplementary Fig. S17), reflecting the higher carbon availabilities in sediment habitats. Moreover, viruses from different trenches showed significantly different C-ARSC levels, and the



**Fig. 4** Lysogenic lifestyle of the viruses in seawater of the Mariana Trench. **A** Overview of lysogenic mOTUs and proviruses in the Mariana Trench seawater. The total number of mOTUs, lysogenic mOTUs, and proviruses in each sample are indicated on the right side of each figure. **B** Provirus/host abundance ratios of the Mariana Trench grouped by ocean zone (epi-/abyssopelagic/hadopelagic) and sample type (free-living/particle-associated). The ratios were calculated by the

average per-base-pair coverage of the provirus and host contigs, respectively. **C** Comparison of lysogenic and non-lysogenic mOTUs abundances in the hadopelagic zone of the Mariana Trench. The abundances were calculated by the average per-base-pair coverage of the mOTUs. The values were analysed by a two-tailed Student's *t*-test. \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; ns not significantly different.

C-ARSC levels of viruses derived from free-living and particle-associated seawater in the Mariana Trench were positively and negatively correlated with the increasing water depth, respectively, suggesting their adaptation to different carbon nutrients in different trenches and size-fractionated microbial assemblages [105, 108]. Taken together, these data suggested that the demand and availability of nutrients might profoundly influence viral genome evolution and, thus, lead to the niche-dependent distribution of viral communities in oceanic trenches.

To explore the lifestyles of hadal viruses, we identified 1704 proviruses in the OTVGD and found that more than half of them ( $n = 905$ ) were derived from the hadopelagic samples (Supplementary Table S5). This high proportion of proviruses in the hadal samples enabled us to further analyse the distribution of lysogens throughout the water column in the oceanic trench by using the metagenome of Mariana Trench seawater, where adequate microbial genomes were obtained by binning. It is worth-noting that, considering the bias of metagenome-derived viral database to proviruses, we chose the numbers of proviruses and hosts as denominators for all percentage calculations to eliminate bias. Notably, the proportion of lysogenic microbial OTUs (mOTUs) in the hadal seawater (average = 41.75%,  $n = 267$ ) was significantly higher than that in the epipelagic (average = 22.92%,  $n = 100$ ) and abyssopelagic (average = 30.58%,  $n = 147$ ) zones (Fig. 4A). By calculating the provirus-host abundance ratio (PVHR), we demonstrated that a considerable proportion of proviruses in hadal seawater (average = 27.32%,

$n = 203$ ) were potentially active (see “Materials and methods” section for the definition of active provirus), and 42.01% of lysogenic mOTUs ( $n = 115$ ), harbored active proviruses (Fig. 3A). Further analysis focusing on these putative active proviruses revealed that the PVHR was significantly higher in the hadopelagic zone (average = 2.28,  $n = 63$ ) than in the epipelagic (average = 1.17,  $n = 4$ ) and abyssopelagic (average = 1.32,  $n = 36$ ) zones (Fig. 4B), implying a high production rate or burst size of proviruses in the hadal seawater. In particular, OTV\_OTU2258 (43.01 kb) that was predicted to integrate into the genome of *Pseudoceanicola*, which generally showed a high abundance in hadal seawater in the Mariana Trench, was highly active in four of the five hadopelagic seawater samples (Supplementary Fig. S18), and therefore might play an important role in regulating the abyssal microbial communities. However, there was no significant difference in PVHR between these proviruses from free-living and particle-associated seawater samples (Fig. 4B). Next, a significantly higher abundance of lysogenic mOTUs was observed in four of the five hadal seawater samples compared with non-lysogenic ones (Fig. 4C), suggesting that proviruses might confer growth and survival advantages to their hosts and stimulate the activity of microbial communities in the hadal ecosystem. Additionally, an analysis of published hadal microorganisms with complete genome sequences revealed that most of them (6 out of 11) were lysogens (Supplementary Table S4), which is consistent with our recent investigation showing that seven out of the ten



bacteria isolated from hadal sediment were lysogenized by 12 prophages (data not shown). Collectively, the above results indicated that a lysogenic lifestyle seemed to be preferably adopted by hadal microbes and proviruses, which could be more active than their counterparts in the upper ocean.

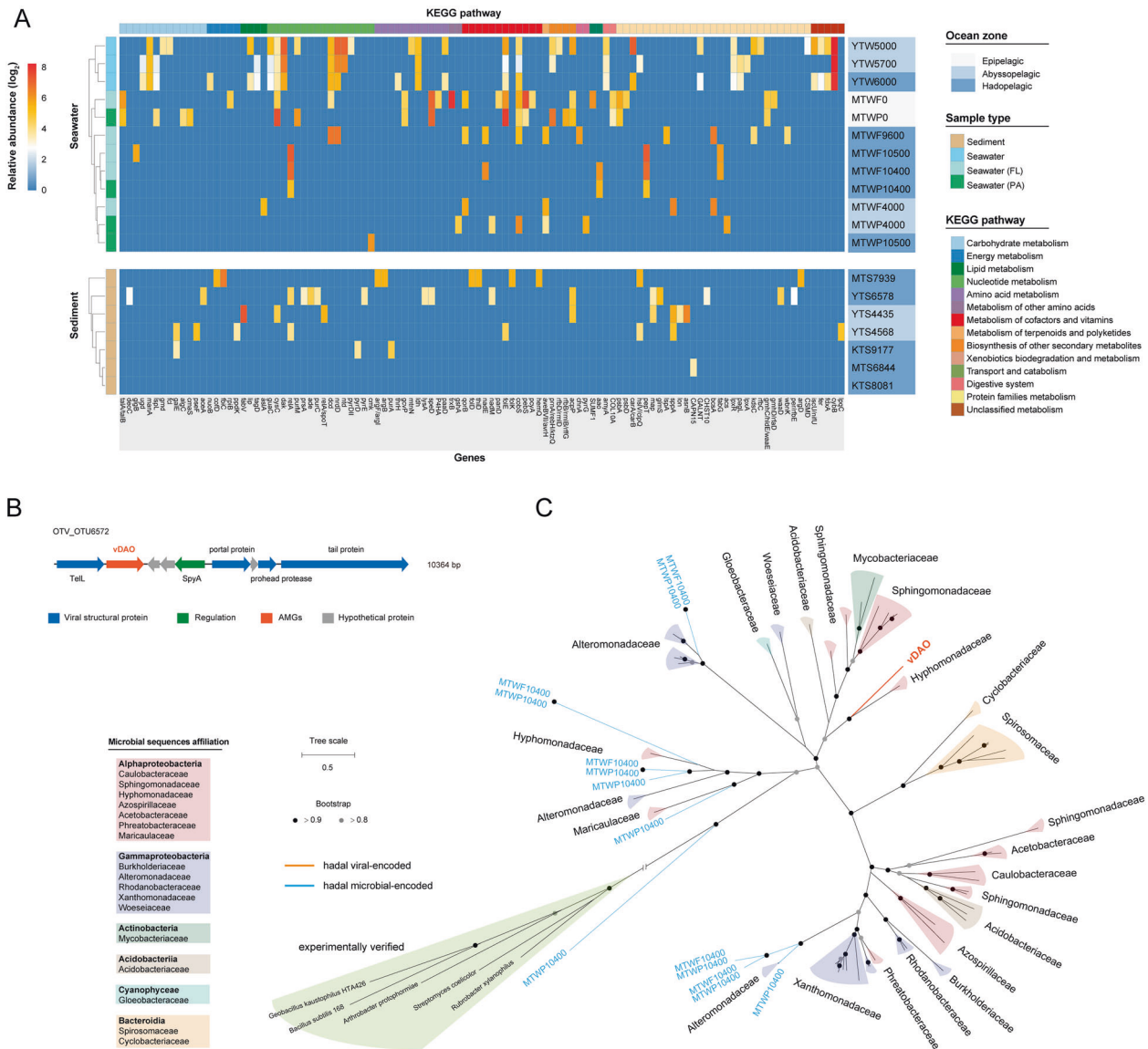
### Niche-specific auxiliary metabolic genes (AMGs) harbored in hadal viruses

Numerous marine viruses possess AMGs that participate in host metabolism, thus facilitating host adaptation to the environment [91, 109, 110]. Strikingly, the depth-stratified distribution of AMGs has been revealed in the Pacific Ocean Virome (POV) [13]. To achieve a more comprehensive understanding of the niche specialization and ecological functions of viruses in oceanic trenches, we screened and identified AMGs in the OTVGD viral genomes and calculated their relative abundances. In total, 34 and 42 types of hadal-specific AMGs (designated as those exclusively found in hadal vOTUs) with diverse functions were identified and could be classified as class I and class II AMGs (Supplementary Tables S10 and S11), respectively, according to a previous definition [91]. The composition and abundance patterns indicated that the AMGs were largely structured by geographic region (Fig. 5A and Supplementary Fig. S19), and similar distribution profiles were observed for AMGs from neighboring samples, suggesting a correlation between AMGs and environmental factors. Notably, the hadal-specific class I AMGs were associated with nucleotide metabolism (*prsA*, *purC*, *pyrD*, *pyrE*), amino acid metabolism (*purA*, *lysA*), metabolism of cofactors and vitamins (*nadE*, *folK*, *hemQ*, *thiD*), and metabolism-related protein families (*pepT*, *fabG*, *aao*, *argD*), according to KEGG pathway annotation (Fig. 5A). For class II AMGs, protein families involved in genetic information processing and signaling/cellular processes, such as the phosphate regulon response regulator PhoB, cell cycle response regulator CtrA, and glycine cleavage system transcriptional activator GcvA, were significantly enriched in hadal viruses (Supplementary Fig. S19), suggesting an important role of viruses in host regulatory networks under extreme environmental conditions.

Intriguingly, an AMG encoding a D-amino acid oxidase (vDAO) was identified in a hadal virus genome (OTV\_OTU6572) (Fig. 5B). DAO catalyzes the oxidative deamination of D-amino acids [111], which are produced by many marine microbes and are the major constituents of the organic carbon and nitrogen pools in the ocean [112, 113]. Although this vDAO shared low amino acid sequence similarity (10.24–14.78%) with the previously reported microbial DAOs, they had significantly similar tertiary structures (Supplementary Fig. S20). To verify whether the

identified vDAO was functionally active, we expressed the protein and examined its activity. As a result, the oxidation activity was confirmed for several D-amino acids (Supplementary Fig. S21), especially the D-alanine and D-glutamic acid, which have been detected as the most abundant D-amino acids in the ocean [112, 113]. Phylogenetic analysis showed that the vDAO sequence was evolutionarily distinct from those of the microbial DAOs with experimentally verified functions; therefore, it might represent a novel DAO encoded by viruses (Fig. 5C). Moreover, the phylogeny of DAO did not correlate well with the taxonomic classification of the microbial species, indicating that horizontal gene transfer (HGT) had significantly influenced its distribution. OTV\_OTU6572 was predicted to infect *Hyphomonadaceae* strains (Supplementary Table S9), and the vDAO was evolutionarily similar to the DAO of this family, suggesting that the vDAO might have originated via HGT from the microbial host infected by the virus. Additionally, read mapping analysis showed that the *dao* gene was generally more abundant in the seawater of the Mariana Trench than in that of other trenches (Supplementary Fig. S22), which implied that D-amino acids might represent an important nutrient source for microorganisms in this harsh environment. Considering that recalcitrant material-degrading microbial taxa were enriched in the sediment of the Mariana Trench [42], viruses likely contribute to the cycle of refractory nutrients in the hadal ecosystem by reprogramming the carbon and nitrogen metabolism of their hosts.

To the best of our knowledge, this is the first study to holistically explore the viral community across multiple oceanic trenches, especially the hadal biosphere. The results provide unprecedented insight into the diversity, distribution, metabolic potential, and ecological functions of viruses in these extreme habitats. In particular, the hadal viruses identified herein featured remarkably high genetic novelty and infected several ecologically important microbial clades (i.e., Thaumarchaeota and *Oleibacter*) in hadal ecosystems. The inter-trench and intra-trench communication between viral communities was predicted to be pervasive, probably resulting from multiple oceanographic processes. Moreover, the variations in topographical features, geochemical parameters and microbial composition between different trenches, sample types, and ocean zones could profoundly influence the distribution, genomic features, and lifestyle of viral communities in oceanic trenches. Moreover, the hadal viral genomes harbored a variety of AMGs with diverse functions (i.e., D-amino acid oxidation, fatty acid and NAD<sup>+</sup> synthesis, transcriptional regulation), and thereby might contribute to the microbial metabolism and biogeochemical cycles. These findings strongly indicate that viruses perform important ecological functions in the hadal biosphere. With the continuous improvement of sampling



**Fig. 5 Characterization of auxiliary metabolic genes (AMGs) in hadal viruses. A** Abundance patterns of class I AMGs in the OTVGD. The heatmap displays the relative abundance of each viral AMG (x-axis) in each sample (y-axis). Total normalized coverage of vOTUs harboring the corresponding AMGs is depicted as a heatmap of relative abundances on the  $\log_2$  scale. The AMGs are hierarchically clustered by samples. The vertical bars on the left and right sides indicate the type and ocean zone of each sample, respectively. **B** Genomic map of the hadal virus harboring D-amino acid oxidase (DAO). The arrows depict the location and direction of predicted proteins on the viral genomes, and the fill colors indicate different functional categories of genes, as indicated in the legend. The viral

AMGs are shown in orange-red. **C** Unrooted phylogenetic tree of DAO. The trees were built by the maximum-likelihood method with 1000 bootstrap replicates. Nodes with bootstrap support values greater than 0.9 and 0.8 are marked with black and gray circles, respectively. The distinct lineages of the DAO proteins from bacteria are assigned different background colors according to their taxonomic affiliations, as indicated in the legend. Orange and blue branches denote sequences obtained from hadal viral and metagenome contigs in this study, respectively. The microbial DAO proteins that have been experimentally verified in previous publications are shaded in green, and this branch was truncated for display.

technologies, increasing numbers of scientific cruises, and new research projects focusing on oceanic trenches, a more comprehensive understanding of hadal viruses is expected in the near future.

**Acknowledgements** This work was financially supported by the National Key R&D Program of China (2018YFC0309800), and the

National Natural Science Foundation of China (grants 91851113, 41921006, 41776173, and 41676118).

**Author contributions** H.J. and X.X. conceived and designed the study. Y.Y., J.W., and H.J. conducted the bioinformatic analyses and data visualization. Y.H., S.W., and C.M. performed functional analysis of AMG. Y.H., and M.Z. helped in collecting hadal microbial genome data. Y.Z. participated in statistical analysis. H.J. and Y.Y. interpreted

the data and drafted the manuscript. Y.W. and H.J. provided useful suggestions to improve the manuscript. X.X. supervised the project. All the authors reviewed the results and approved the manuscript.

## Compliance with ethical standards

**Conflict of interest** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

- Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol.* 2007;5:801–12.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nat Microbiol.* 2018;3:754–66.
- Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, et al. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol.* 2020;18:21–34.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 2016;537:689–93.
- Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 2019;177:1109–23.
- Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat Commun.* 2017;8:15955.
- Brum JR, Ignacio-Espinoza JC, Roux S, Doucier G, Acinas SG, Alberti A, et al. Patterns and ecological drivers of ocean viral communities. *Science* 2015;348:1261498.
- Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C, et al. Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* 2008;454:1084–7.
- Danovaro R, Dell'Anno A, Corinaldesi C, EugenioRastelli, Cavicchioli R, Krupovic M, et al. Virus-mediated archaeal hecatomb in the deep seafloor. *Sci Adv* 2016;2:e1600492.
- Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. Sulfur oxidation genes in diverse deep-sea viruses. *Science* 2014;344:757–60.
- He T, Li H, Zhang X. Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus–host interactions. *mBio* 2017;8:e00893–17.
- Hurwitz BL, Hallam SJ, Sullivan MB. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol.* 2013;14:R123.
- Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific ocean virome. *ISME J.* 2015;9:472–84.
- Dell'Anno A, Corinaldesi C, Danovaro R. Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *Proc Natl Acad Sci USA.* 2015;112:E2014–E9.
- Lara E, Vaqué D, Sà EL, Boras JA, Gomes A, Borrull E, et al. Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci Adv.* 2017;3:e1602565.
- Mizuno CM, Ghai R, Saghai A, López-García P, Rodríguez-Valera F. Genomes of abundant and widespread viruses from the deep ocean. *mBio.* 2016;7:e00805–16.
- Tangherlini M, Dell'Anno A, Allen LZ, Riccioni G, Corinaldesi C. Assessing viral taxonomic composition in benthic marine ecosystems: reliability and efficiency of different bioinformatic tools for viral metagenomic analyses. *Sci Rep.* 2016;6:28428.
- Yang Y, Yokokawa T, Motegi C, Nagata T. Large-scale distribution of viruses in deep waters of the Pacific and Southern Oceans. *Aquat Micro Ecol.* 2014;71:193–202.
- Tang K, Lin D, Zheng Q, Liu K, Yang Y, Han Y, et al. Genomic, proteomic and bioinformatic analysis of two temperate phages in *Roseobacter* clade bacteria isolated from the deep-sea water. *BMC Genomics.* 2017;18:485.
- Corte DD, Sintès E, Yokokawa T, Reinthaler T, Herndl GJ. Links between viruses and prokaryotes throughout the water column along a North Atlantic latitudinal transect. *ISME J.* 2012;6:1566–77.
- Corte DD, Sintès E, Winter C, Yokokawa T, Reinthaler T, Herndl GJ. Links between viral and prokaryotic communities throughout the water column in the (sub)tropical Atlantic Ocean. *ISME J.* 2010;4:1431–42.
- Blankenship-Williams LE, Levin LA. Living deep: a synopsis of hadal trench ecology. *Mar Technol Soc J.* 2009;43:137–43.
- Jamieson AJ, Fujii T, Mayor DJ, Solan M, Priede IG. Hadal trenches: the ecology of the deepest places on Earth. *Trends Ecol Evol.* 2010;25:190–7.
- Jamieson AJ. Ecology of deep oceans: hadal trenches. eLS. Chichester: Wiley; 2001.
- Glud RN, Wenzhöfer F, Middelboe M, Oguri K, Turnewitsch R, Canfield DE, et al. High rates of microbial carbon turnover in sediments in the deepest oceanic trench on Earth. *Nat Geosci.* 2013;6:284–8.
- Nunoura T, Takaki Y, Hirai M, Shimamura S, Makabe A, Koide O, et al. Hadal biosphere: insight into the microbial ecosystem in the deepest ocean on Earth. *Proc Natl Acad Sci USA.* 2015;112: E1230–E6.
- Zhong H, Lehtovirta-Morley L, Liu J, Zheng Y, Lin H, Song D, et al. Novel insights into the Thaumarchaeota in the deepest oceans: their metabolism and potential adaptation mechanisms. *Microbiome* 2020;8:78.
- Liu J, Zheng Y, Lin H, Wang X, Li M, Liu Y, et al. Proliferation of hydrocarbon-degrading microbes at the bottom of the Mariana Trench. *Microbiome* 2019;7:47.
- Tamegai H, Li L, Nasui N, Kato C. A denitrifying bacterium from the deep sea at 11000-m depth. *Extremophiles* 1997;1:207–11.
- Kato C, Li L, Nogi Y, Nakamura Y, Tamaoka J, Horikoshi K. Extremely barophilic bacteria isolated from the Mariana Trench, challenger deep, at a depth of 11,000 meters. *Appl Environ Microbiol.* 1998;64:1510–3.
- Nogi Y, Kato C. Taxonomic studies of extremely barophilic bacteria isolated from the Mariana Trench and description of *Moritella yayanosii* sp. nov., a new barophilic bacterial isolate. *Extremophiles* 1999;3:71–7.
- Nogi Y, Hosoya S, Kato C, Horikoshi K. *Colwellia piezophila* sp. nov., a novel piezophilic species from deep-sea sediments of the Japan Trench. *Int J Syst Evol Microbiol.* 2004;54:1627–31.
- Nogi Y, Hosoya S, Kato C, Horikoshi K. *Psychromonas hadalis* sp. nov., a novel piezophilic bacterium isolated from the bottom of the Japan Trench. *Int J Syst Evol Microbiol.* 2007;57:1360–4.
- Kusube M, Kyaw TS, Tanikawa K, Chastain RA, Hardy KM, Cameron J, et al. *Colwellia marinimaniae* sp. nov., a hyperpiezophilic species isolated from an amphipod within the Challenger Deep, Mariana Trench. *Int J Syst Evol Microbiol.* 2017;67:824–31.
- Wei Y, Fang J, Xu Y, Zhao W, Cao J. *Corynebacterium hadale* sp. nov. isolated from hadopelagic water of the New Britain Trench. *Int J Syst Evol Microbiol.* 2018;68:1474–8.
- Ahmad W, Zheng Y, Li Y, Sun W, Hu Y, He X, et al. *Marinobacter salinexigens* sp. nov., a marine bacterium isolated from

- hadal seawater of the Mariana Trench. *Int J Syst Evol Microbiol.* 2020;70:3794–800.
37. Zhao X, Liu J, Zhou S, Zheng Y, Wu Y, Kogure K, et al. Diversity of culturable heterotrophic bacteria from the Mariana Trench and their ability to degrade macromolecules. *Mar Life Sci Technol.* 2020;2:181–93.
  38. Yoshida M, Yoshida-Takashima Y, Nunoura T, Takai K. Identification and genomic analysis of temperate *Pseudomonas* bacteriophage PstS-1 from the Japan trench at a depth of 7000 m. *Res Microbiol.* 2015;166:668–76.
  39. Yoshida M, Takaki Y, Eitoku M, Nunoura T, Takai K. Metagenomic analysis of viral communities in (Hado)pelagic sediments. *PLoS ONE.* 2013;8:e57271.
  40. Manea E, Dell'Anno A, Rastelli E, Tangherlini M, Nunoura T, Nomaki H, et al. Viral infections boost prokaryotic biomass production and organic C cycling in Hadal Trench sediments. *Front Microbiol.* 2019;10:1952.
  41. Zhang X, Xu W, Liu Y, Cai M, Luo Z, Li M. Metagenomics reveals microbial diversity and metabolic potentials of seawater and surface sediment from a Hadal biosphere at the Yap Trench. *Front Microbiol.* 2018;9:2402.
  42. Peoples LM, Grammatopoulou E, Pombrol M, Xu X, Osuntokun O, Blanton J, et al. Microbial community diversity within sediments from two geographically separated Hadal Trenches. *Front Microbiol.* 2019;10:347.
  43. Peoples LM, Donaldson S, Osuntokun O, Xia Q, Nelson A, Blanton J, et al. Vertically distinct microbial communities in the Mariana and Kermadec trenches. *PLoS ONE.* 2018;13:e0195102.
  44. NA J, JN F. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). <https://github.com/najoshi/sickle2011>.
  45. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;28:1420–8.
  46. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 2015;3:e985.
  47. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* 2017;5:69.
  48. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free  $d_2^*$  oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* 2017;45:39–53.
  49. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;28:3150–2.
  50. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 2010;11:119.
  51. Paez-Espino D, Pavlopoulos GA, Ivanova NN, Kyrpidis NC. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. *Nat Protoc.* 2017;12:1673–82.
  52. Paez-Espino D, Eloie-Fadrosch EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature* 2016;536:425–30.
  53. Song W, Sun H-X, Zhang C, Cheng L, Peng Y, Deng Z, et al. Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.* 2019;47:W74–W80.
  54. O'Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44:D733–D45.
  55. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12:59–60.
  56. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 2019;47:D309–D14.
  57. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Mering CV, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol.* 2017;34:2115–22.
  58. Graziotin AL, Koonin EV, Kristensen DM. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res.* 2017;45:D491–D8.
  59. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res.* 2018;46:W200–W4.
  60. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2014;42:D199–D205.
  61. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol.* 2016;428:726–31.
  62. El-Gebali S, Mistry J, Bateman A, Eddy SR, Ae L, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47:D427–D32.
  63. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et al. Minimum information about an uncultivated virus genome (MIUViG). *Nat Biotechnol.* 2019;37:29–37.
  64. Paez-Espino D, Roux S, Chen I-MA, Palaniappan K, Ratner A, Chu K, et al. IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res.* 2019;47:D678–D86.
  65. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods.* 2016;8:12–24.
  66. Crits-Christoph A, Gelsinger DR, Ma B, Wierzbos J, Ravel J, Davila A, et al. Functional interactions of archaea, bacteria and viruses in a hypersaline endolithic community. *Environ Microbiol.* 2016;18:2064–77.
  67. Aylward FO, Moniruzzaman M. ViralRecall—a flexible command-line tool for the detection of giant virus signatures in Omic Data. *Viruses* 2021;13:150.
  68. Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol.* 2019;37:632–9.
  69. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
  70. Hurwitz BL, Sullivan MB. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE.* 2013;8:e57355.
  71. Corte DD, Martínez JM, Cretoiu MS, Takaki Y, Nunoura T, Sintés E, et al. Viral communities in the global deep ocean conveyor belt assessed by targeted viromics. *Front Microbiol.* 2019;10:1801.
  72. Ghai R, Mehrshad M, Mizuno CM, Rodriguez-Valera F. Metagenomic recovery of phage genomes of uncultured freshwater actinobacteria. *ISME J.* 2017;11:304–8.
  73. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
  74. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.



75. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft BJ, Jang HB, et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol.* 2018;3:870–80.
76. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
77. Getz EW, Tithi SS, Zhang L, Aylward FO. Parallel evolution of genome streamlining and cellular bioenergetics across the marine radiation of a bacterial phylum. *mBio* 2018;9:e01089–18.
78. Karst SM, Kirkegaard RH, Albertsen M. mmgenome: a toolbox for reproducible genome extraction from metagenomes. Preprint at bioRxiv. 2016.
79. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
80. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25:955–64.
81. Lin BY, Chan PP, Lowe TM. tRNAviz: explore and visualize tRNA sequence features. *Nucleic Acids Res.* 2019;47:W542–W7.
82. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinform.* 2007;8:209.
83. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol Rev.* 2016;40:258–72.
84. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2020;36:1925–7.
85. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5:e9490.
86. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547–9.
87. Malki K, Kula A, Bruder K, Sible E, Hatzopoulos T, Steidel S, et al. Bacteriophages isolated from Lake Michigan demonstrate broad host-range across several bacterial phyla. *Virology.* 2015;12:164.
88. Touchon M, Bernheim A, Rocha EP. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* 2016;10:2744–54.
89. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J* 2017;11:1511–20.
90. Hulo C, Castro ED, Masson P, Bougueleret L, Bairoch A, Xenarios I, et al. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* 2011;39:D576–82.
91. Hurwitz BL, U'Ren JM. Viral metabolic reprogramming in marine ecosystems. *Curr Opin Microbiol.* 2016;31:161–8.
92. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 2018;27:135–45.
93. Takahashi S, Furukawara M, Omae K, Tadokoro N, Saito Y, Abe K, et al. A highly stable D-amino acid oxidase of the thermophilic bacterium *Rubrobacter xylanophilus*. *Appl Environ Microbiol.* 2014;80:7219–29.
94. Brister JR, Ako-adjei D, Bao Y, Blinkova O. NCBI viral genomes resource. *Nucleic Acids Res.* 2015;43:D571–D7.
95. Al-Shayeb B, Sachdeva R, Chen L-X, Ward F, Munk P, Devoto A, et al. Clades of huge phages from across Earth's ecosystems. *Nature* 2020;578:425–31.
96. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine *Thaumarchaeota* viruses that encode *amoC* nitrification genes. *ISME J* 2019;13:618–31.
97. Kim J-G, Kim S-J, Cvirkaite-Krupovic V, Yu W-J, Gwak J-H, López-Pérez M, et al. Spindle-shaped viruses infect marine ammonia-oxidizing thaumarchaea. *Proc Natl Acad Sci USA.* 2019;116:15645–50.
98. Hiraoka S, Hirai M, Matsui Y, Makabe A, Minegishi H, Tsuda M, et al. Microbial community and geochemical analyses of trans-trench sediments for understanding the roles of hadal environments. *ISME J.* 2020;14:740–56.
99. Kioka A, Schwestermann T, Moernaut J, Ikehara K, Kanamatsu T, McHugh CM, et al. Megathrust earthquake drives drastic organic carbon supply to the hadal trench. *Sci Rep.* 2019;9:1553.
100. Mestrea M, Ruiz-González C, Logares R, Duarte CM, Gasol JM, Sala MM. Sinking particles promote vertical connectivity in the ocean microbiome. *Proc Natl Acad Sci USA.* 2018;115:E6799–E807.
101. Boeuf D, Edwards BR, Eppley JM, Hu SK, Poff KE, Romano AE, et al. Biological composition and microbial dynamics of sinking particulate organic matter at abyssal ocean. *Proc Natl Acad Sci USA.* 2019;116:11824–32.
102. Tamsitt V, Drake HF, Morrison AK, Talley LD, Dufour CO, Gray AR, et al. Spiraling pathways of global deep waters to the surface of the Southern Ocean. *Nat Commun.* 2017;8:172.
103. Fukamachi Y, Rintoul SR, Church JA, Aoki S, Sokolov S, Rosenberg MA, et al. Strong export of Antarctic bottom water east of the Kerguelen plateau. *Nat Geosci.* 2010;3:327–31.
104. Garabato ACN, Frajka-Williams EE, Spingys CP, Legg S, Polzin KL, Forryan A, et al. Rapid mixing and exchange of deep-ocean waters in an abyssal boundary current. *Proc Natl Acad Sci USA.* 2019;116:13233–8.
105. Stewart HA, Jamieson AJ. Habitat heterogeneity of hadal trenches: considerations and implications for future studies. *Prog Oceanogr.* 2018;161:47–65.
106. Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, et al. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol.* 2017;2:1367–73.
107. Giovannoni SJ, Thrash JC, Temperton B. Implications of streamlining theory for microbial ecology. *ISME J.* 2014;8:1553–65.
108. Xu Y, Ge H, Fang J. Biogeochemistry of hadal trenches: recent developments and future perspectives. *Deep Sea Res Part II.* 2018;155:19–26.
109. Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismaeel N, et al. Photosystem I gene cassettes are present in marine virus genomes. *Nature* 2009;461:258–62.
110. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. *Nature* 2003;424:741.
111. Takahashi S, Abe K, Kera Y. Bacterial d-amino acid oxidases: recent findings and future perspectives. *Bioengineered* 2015;6:237–41.
112. Kaiser K, Benner R. Major bacterial contribution to the ocean reservoir of detrital organic carbon and nitrogen. *Limnol Oceanogr.* 2008;53:99–112.
113. Zhang Z, Zheng Q, Jiao N. Microbial D-amino acids and marine carbon storage. *Sci China Earth Sci.* 2016;59:17–24.