

Diversity and ecological potentials of viral assemblages from the seamount sediments of the Northwest Pacific Ocean

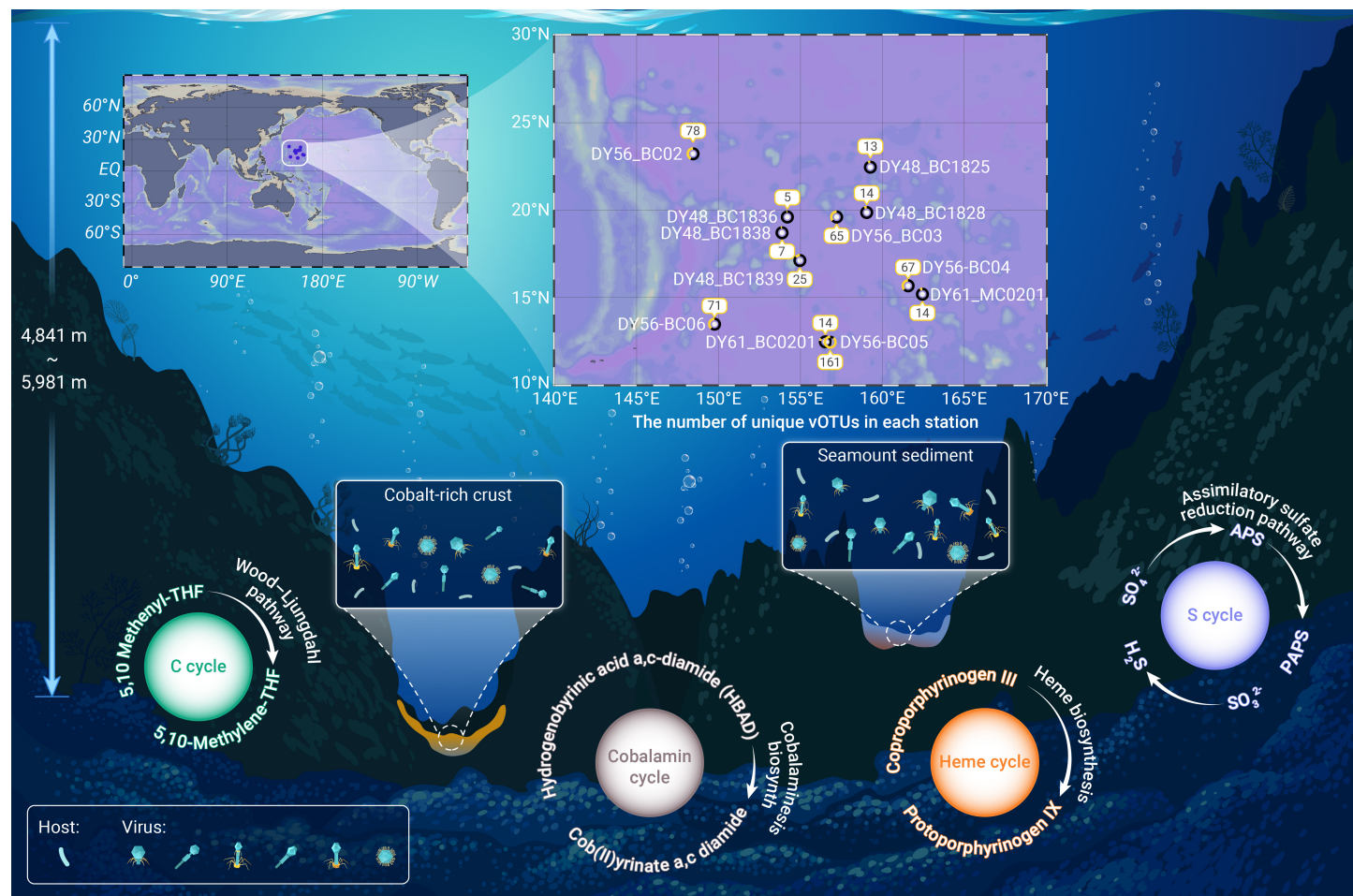
Ying Chen,^{1,7} Chen Gao,^{2,7} Qian Liu,^{3,*} Yantao Liang,^{1,4,*} Mingyan Lai,³ Fuyue Ge,¹ Kaiyang Zheng,¹ Hao Yu,¹ Hongbing Shao,^{1,4} Andrew McMinn,^{1,5} and Min Wang^{1,4,2,6,*}

*Correspondence: liuqian@sio.org.cn (Q.L.); liangyantao@ouc.edu.cn (Y.L.); mingwang@ouc.edu.cn (M.W.)

Received: November 8, 2023; Accepted: July 13, 2024; Published Online: July 28, 2024; <https://doi.org/10.59717/j.xinn-geo.2024.100088>

© 2024 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Seamount sediments harbor diverse viruses, cobalt-rich crust influent the virus communities.
- Nitrososphaeria* was one of the major hosts of seamount sediment viruses.
- Viruses from cobalt-rich crusts are enriched with auxiliary metabolic genes for heme and cobalamin metabolism.

Diversity and ecological potentials of viral assemblages from the seamount sediments of the Northwest Pacific Ocean

Ying Chen,^{1,7} Chen Gao,^{2,7} Qian Liu,^{3,*} Yantao Liang,^{1,4,*} Mingyan Lai,³ Fuyue Ge,¹ Kaiyang Zheng,¹ Hao Yu,¹ Hongbing Shao,^{1,4} Andrew McMinn,^{1,5} and Min Wang^{1,4,2,6,*}

¹Frontiers Science Center for Deep Ocean Multispheres and Earth System, College of Marine Life Sciences, Institute of Evolution and Marine Biodiversity, MOE Key Laboratory of Evolution and Marine Biodiversity, Center for Ocean Carbon Neutrality, Ocean University of China, Qingdao 266100, China

²Haide College, Ocean University of China, Qingdao 266100, China

³Key Laboratory of Marine Ecosystem Dynamics Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

⁴UMT-OUC Joint Academic Centre for Marine Studies, Qingdao 266003, China

⁵Institute for Marine and Antarctic Studies, University of Tasmania, Hobart 7001, Australia

⁶The Affiliated Hospital of Qingdao University, Qingdao 266000, China

⁷These authors contributed equally

*Correspondence: liuqian@sio.org.cn (Q.L.); liangyantao@ouc.edu.cn (Y.L.); mingwang@ouc.edu.cn (M.W.)

Received: November 8, 2023; Accepted: July 13, 2024; Published Online: July 28, 2024; <https://doi.org/10.59717/j.xinn-geo.2024.100088>

© 2024 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: Chen Y., Gao C., Liu Q., et al., (2024). Diversity and ecological potentials of viral assemblages from the seamount sediments of the Northwest Pacific Ocean. *The Innovation Geoscience* 2(3): 100088.

Viruses are the most abundant life forms in the sea, influencing the community structure and metabolism of host cells and biogeochemical cycles. However, the diversity of viruses and their ecological roles within seamount sediments ecosystems, natural microbiota havens characterized by high biodiversity, remain largely unknown. Here, the seamount viral genome (SMVG) dataset, based on a metagenomic analysis of twelve seamount sediment samples collected from the seamount regions of the Northwest Pacific Ocean, was established. A total of 78,069 viral operational taxonomic units (vOTUs) were found, spanning 18 viral classes and 63 viral families. The detection of sixteen viral auxiliary metabolic genes (AMGs) suggests that viruses may participate in the metabolic processes associated with sediment microbial communities and biogeochemical cycles, including carbon, sulfur, metal, heme, and cobalamin cycling. AMGs involved in the metabolism of heme, cobalamin, and metals were more often detected in seamount sediments than in trenches, cool seeps, and hydrothermal vents. This investigation of the viral communities in these seamount sediments provides new insights into the high diversity and ecological potential of the viruses and establishes a foundation for the future study of benthic viruses from deep-sea seamounts.

INTRODUCTION

Many deep-sea habitats, such as sea mountains, cold seeps and hydrothermal vents, have exceptionally high biodiversity. Within these habitats, viruses might have evolved numerous strategies to ensure their successful reproduction, which includes their ability to reprogram the hosts' metabolism via the viral auxiliary metabolic genes (AMGs), horizontal gene transfer, etc.¹⁻⁴ As critical mortality agents of both prokaryotes and eukaryotes, they can also affect biogeochemical cycles and ecosystem dynamics.^{5,6} Benefiting from rapid advances in metagenomics, viruses in bulk metagenomic data present an opportunity for large-scale identification of unclassified viruses from the deep ocean, from areas such as trenches, cold seeps, and hydrothermal vents.⁷⁻⁹

Seamounts are isolated topographic features with summit heights at least 100 m above the seafloor but not reaching the sea surface.¹⁰⁻¹² Seamounts are generally volcanic in origin and worldwide are recognized as important offshore deep ecosystems.¹³ These structures frequently form on the oceanic crust and transitional crust and interact with ocean currents, triggering complex oceanographic processes, such as the "Taylor column," turbulent mixing, and benthic boundary effects.^{14,15} Although there are around 33,452 identified seamounts and 138,412 knolls in the global ocean, only 1.5% of seamounts and 0.7% of knolls are located within marine protected areas listed on the World Database of Protected Areas (WDPA).¹⁶ Less than 300 seamounts have been explored,¹⁷ primarily in the Northwest Pacific, leaving a dearth of information from other areas.

Recent investigations of seamount ecology have shown that their

geographical features and hydrological conditions allow for the existence of abundant prokaryotes and their predators, including pelagic marine predators and viruses, which then release labile organic matter into the surrounding seawater and sediments.^{8,14} Biological and ecological studies have shown that prokaryotic abundance in sediments surrounding seamounts is typically 2-3 times higher than in non-seamount sediments.¹⁴ These bacteria and archaea, which use sulfur, nitrate, and heavy metals as energy sources,^{8,18-20} may have developed morphological adaptations to the seamount environment.²¹ Notably, viruses have been demonstrated to manipulate the mortality of microorganisms, releasing labile organic matter and expressing AMGs that can rewire host metabolism.^{6,8,22} However, the community composition, diversity, and ecological roles of viral assemblages inhabiting seamount sediments still need to be discovered.

In this study, viral communities from Northwest Pacific Ocean seamount sediments, together with their metabolic potentials, were investigated. The seamount viral genome (SMVG) dataset was established, comprehensively analyzing virus community structure, genomic diversity, distribution, lifestyle, and ecological potentials on the seamount sediment surface.

MATERIALS AND METHODS

Sample collection

Twelve surface sediment samples were collected from Northwest Pacific Ocean seamounts within a geographic range of 148.43942°E to 156.522634°E and 12.429233°N to 23.23131°N over three years (Supplemental Table S1). The sediment samples were collected from the top 1 cm of the surface layer using a box corer (DY61_MC0201 using multi-corer). The collected samples were immediately frozen at -80°C until further processing.

DNA extraction and sequencing

The total DNA from sediment samples (0.5 g) from cruise DY48 was extracted by Novogene (Nanjing, China), DNA from cruises DY56 and DY61 was extracted by Guangdong Magigene Biotechnology Co., Ltd. (Guangzhou, China), using ALFA-SEO Magnetic Soil DNA Kit according to the manufacturer's instructions, respectively. DNA integrity and purity were monitored on 1% agarose gels. DNA concentration and purity were measured using Qubit 3.0 (Thermo Fisher Scientific, Waltham, USA) and Nanodrop One (Thermo Fisher Scientific, Waltham, USA) at the same time. Subsequently, the total DNA of each sample (100–200 ng DNA as a template) was amplified with QIAGEN REPLI-g Mini Kit (phi-29 DNA polymerase), which used the whole genome multiple displacement amplification (MDA) method (details see Supplemental Information).^{23,24} Library preparation and high-throughput sequencing of each sample were sequenced on the Illumina platform, and 150 bp paired-end reads at the Novogene (Nanjing, China) or using NEB Next® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, MA, USA) following the manufacturer's recommendations and index were added at the Guangdong Magigene Biotechnology Co., Ltd. (Guangzhou, China).

Quality control and assembly and identification

Raw reads underwent quality control by removing adapters using Fastp (n 0).²⁵ Subsequently, high-quality paired-end reads were filtered by Perl scripts, ensuring the following criteria: (1) absence of N; (2) no more than 20% of bases with a quality score below 20; (3) no more than 30% of bases with a quality score below 30.^{26,27} Quality-controlled reads from each sample were assembled using MEGAHIT v1.2.9,²⁸ with the removal of contigs $\leq 1,500$ bp before viral identification,^{29,30} then metagenomic assemblies were evaluated by metaQUAST (default mode).³¹ Metagenomic contigs ($> 1,500$ bp) from each site were subjected to viral identification tools: VirSorter v1.0.6 and VirFinder. The following criteria were applied for viral contig identification: (1) contigs sorted by VirFinder with a score ≥ 0.9 and $p < 0.05$; (2) contigs sorted by VirSorter in categories 1 and 2; (3) contigs sorted by VirFinder with a score ≥ 0.7 and $p < 0.05$ and VirSorter categories 1-6.^{32,33} CheckV v0.7.0. was used to remove non-viral fragments and mitigate the effects of microbial contamination.³⁴

Dereplication, relative abundances, and taxonomic profiling

Viral contigs from each sample were dereplicated using CD-HIT v4.8.1 (parameters: -c 0.95 -G 0 -M 0 -aS 0.8 -T 8 -n 9), meaning that viral contigs met the condition that $\geq 95\%$ nucleotide identity across $\geq 80\%$ of shorter viral contigs were considered to be the same viral population and were removed.³⁵ As a result, 78,069 viral operational taxonomic units (vOTUs) were retained. TPM (Transcripts per million) values were used to represent the relative abundance of viruses. Quality-controlled contigs from each sample were analyzed using CoverM v0.6.1 to count coverage profiles across samples (parameters: -p bwa-mem --min-read-percent-identity 0.95 --min-read-aligned-percent 0.75 -m tpm -t 20).^{36,37}

To identify the taxonomic affiliation of the vOTUs, a method that was based on the fitting of sequence similarity and the classification extension based on the bipartite graph were performed using VITAP (<https://github.com/DrKaiyangZheng/VITAP>). ORFs were aligned against the NCBI RefSeq (NR) database (2023-07-20) with an e-value threshold of $1e-5$ and reporting the top 1000 hits. Subsequently, ORFs were mapped against the GenBank taxonomy database: the ratio of homologous ORFs within a genome assigned to a family compared to the total number of ORFs; the highest ratio among the consistent results was then predicted as the most suitable taxonomic affiliation. Finally, the NCBI viral_Refseq database validated the taxonomic affiliation results using CAT v5.0.3.³⁸ The lifestyle of viruses was detected by VIBRANT v1.2.1 with the default parameters (except -virome).³⁹

Phylogenetic analysis

The viral terminase large subunit (Terminase_1, Terminase_3, Terminase_6N, Terminase_Gpa) and the NCLDV-derived family B DNA polymerase (polB) protein were used as the conserved marker genes to construct phylogenetic trees of *Caudoviricetes* and Nucleocytoplasmic large DNA viruses (NCLDVs).⁴⁰ For the phylogenetic trees of viral terminase, a total of 990 viral terminase large subunits (TerL) were identified from the samples by PfamScan. Then, a total of 792 reference sequences were downloaded from the NCBI virus database (NCBI Virus (nih.gov)), and a dereplication was performed (Supplemental Table S3). The NCLDV polB proteins were searched against a hidden Markov model of NCLDV polB sequences using hmmsearch (-E $1e-10$). Reference sequences of the polB protein were obtained from a previous study.⁴¹ Sequences with less than 300 amino acids were removed. The 36 dereplicated sequences, together with 500 reference sequences, were aligned using MAFFT v7.490 (auto mode), then trimmed using TrimAL (v1.4. rev15) to remove position with gaps over 90%.^{42,43} The maximum-likelihood phylogenetic trees were constructed using IQ-TREE v2.1.4 (--safe -m MFP -B 1000)⁴⁴ and visualized using iTOL (version 6).⁴⁵

Comparison with viruses from other habitats.

To understand the similarities and differences between viruses from seamounts sediments and other environments (including oceanic, deep subsurface, non-marine saline and alkaline environments, and extreme habitats), 2,890 viral contigs with lengths over 10 kb were selected from SMVG. Subsequently, 20,195 viral reference sequences were used to construct the dataset, including 4,509 high-quality viral genomes from oceanic habitats

(IMG/VR v4),⁴⁶ 4,060 high-quality viral genomes from the deep subsurface (IMG/VR v4), 3,379 high-quality viral genomes from non-marine saline and alkaline environments (IMG/VR v4) and 8,247 viral sequences from other extreme habitats. Specifically, the 'extreme habitats' consisted of 1,626 sequences from hydrothermal vents, 2,885 sequences from cool seeps,⁴⁷ and 3,736 sequences from Challenger Deep.⁸ For each viral contig, ORFs were predicted using metaProdigal v2.6.3,⁴⁸ then all-verses-all DIAMOND v0.9.29.130 BLASTP was used to compare the protein with parameters: e-value $\leq 1 \times 10^{-5}$, query coverage $\geq 50\%$, identity $\geq 25\%$.^{30,49} All sequences were analyzed using vConTACT 2 with the following parameters: -db 'None' -pcs-mode MCL -vcs-mode ClusterONE.⁵⁰ The network was illustrated using Cytoscape v3.9.1, and the number of VCs was plotted using Evenn.^{51,52}

Host binning and taxonomic profiling

Quality-controlled metagenomic sequences were binned (parameters: --metabat2 --maxbin2 --concoct) and refined (parameters: -c 70 -x 5) using metaWRAP v1.3.2.^{47,53} A total of 155 metagenome-assembled genomes (MAGs) were initially produced, and these were then de-replicated into 71 using dRep v3.2.0 (parameters: -sa 0.95 -nc 0.30 -p 24 -comp 70 -con 10).⁵⁴ Taxonomic annotation of the MAGs was carried out using GTDB-Tk v1.0.2,⁵⁵ and relative abundance was calculated by TPM using CoverM v0.6.1 (parameters: -p bwa-mem --min-read-percent-identity 0.95 --min-read-aligned-percent 0.75 -m tpm -t 20).

Virus-host prediction.

Four different *in silico* methods were employed to predict virus-host interactions.⁴⁷ (1) Nucleotide sequence homology; vOTUs were compared with sequences of MAGs using BLASTn with parameters: query coverage $\geq 75\%$, identity $\geq 70\%$, bit score ≥ 50 , and e-value $\leq 1e-3$; (2) Oligonucleotide frequency (ONF); oligonucleotide frequency and distance between vOTUs and MAG sequences were calculated using VirHostMatcher v1.0 with default parameters. Predictions with d2* values ≤ 0.2 were considered eligible for assigning virus-host linkages.⁵⁶ (3) Transfer RNA (tRNA) match; screening of tRNA genes from viral contigs and MAGs was performed with ARAGORN v1.2.38 with the parameters "t".⁵⁷ A match was identified when there was $\geq 90\%$ length identity in $\geq 90\%$ of the sequences by BLASTn.⁵⁸ (4) CRISPR spacer match; CRISPR arrays were assembled from clean reads of metagenomes using Crass v1.0.1 with default parameters.⁵⁹ The retained spacers from reads were then compared with vOTUs using BLASTn (e-value $\leq 10^{-5}$, mismatch ≤ 1).⁶⁰ Replicate comparisons were performed using BLASTn with the same parameters as prokaryotic MAGs for each matched CRISPR spacer.⁴⁷

Functional annotation and abundance of ORFs and putative AMGs.

The ORFs from viral contigs were identified using metaProdigal with default parameters. Comprehensive annotation of viral ORFs was performed based on four databases: CAZy, eggNOG, KEGG, and pVOG databases.⁶¹⁻⁶⁴ Firstly, ORFs were compared with the Carbohydrate-active enzymes (CAZy) database using DIAMOND (parameters: (e value 1×10^{-5} , $> 60\%$ identity)), to identify the carbohydrate-metabolism-related gene.⁶¹ Furthermore, all ORFs were annotated against the eggNOG v5.0 database to classify COG functional classifications by eggno-mapper with default parameters.⁶² Additionally, the ORFs were submitted to GhostKoala to be annotated by the KEGG database (<https://www.kegg.jp/ghostkoala/>).⁶³ To ensure that the AMGs were encoded by viruses, all the ORFs were compared against the Prokaryotic Virus orthologous groups (pVOGs) database using a hmmsearch search.⁶⁴ Viruses encoded AMGs were verified only if both flanking regions were annotated by the pVOG database and manually selected. The selection criteria included: more than 40% of the region exhibiting the highest similarity to viral proteins, and the viral region possessed more than 15% hallmark genes and viral-like. The relative abundance of putative AMGs in each site was calculated by CoverM v0.6.1 (parameters: -p bwa-mem --min-read-percent-identity 0.95 --min-read-aligned-percent 0.75 -m tpm -t 20). Heatmaps were generated and visualized using TBtool v1.108.⁶⁵

For the phylogenetic analysis of cobaltochelatase (*cobS* and *cobT*), reference protein sequences were retrieved from IMG/VR v4⁴⁶ and the UniProt databases. *cobS* and *cobT* reference sequences were retrieved from the

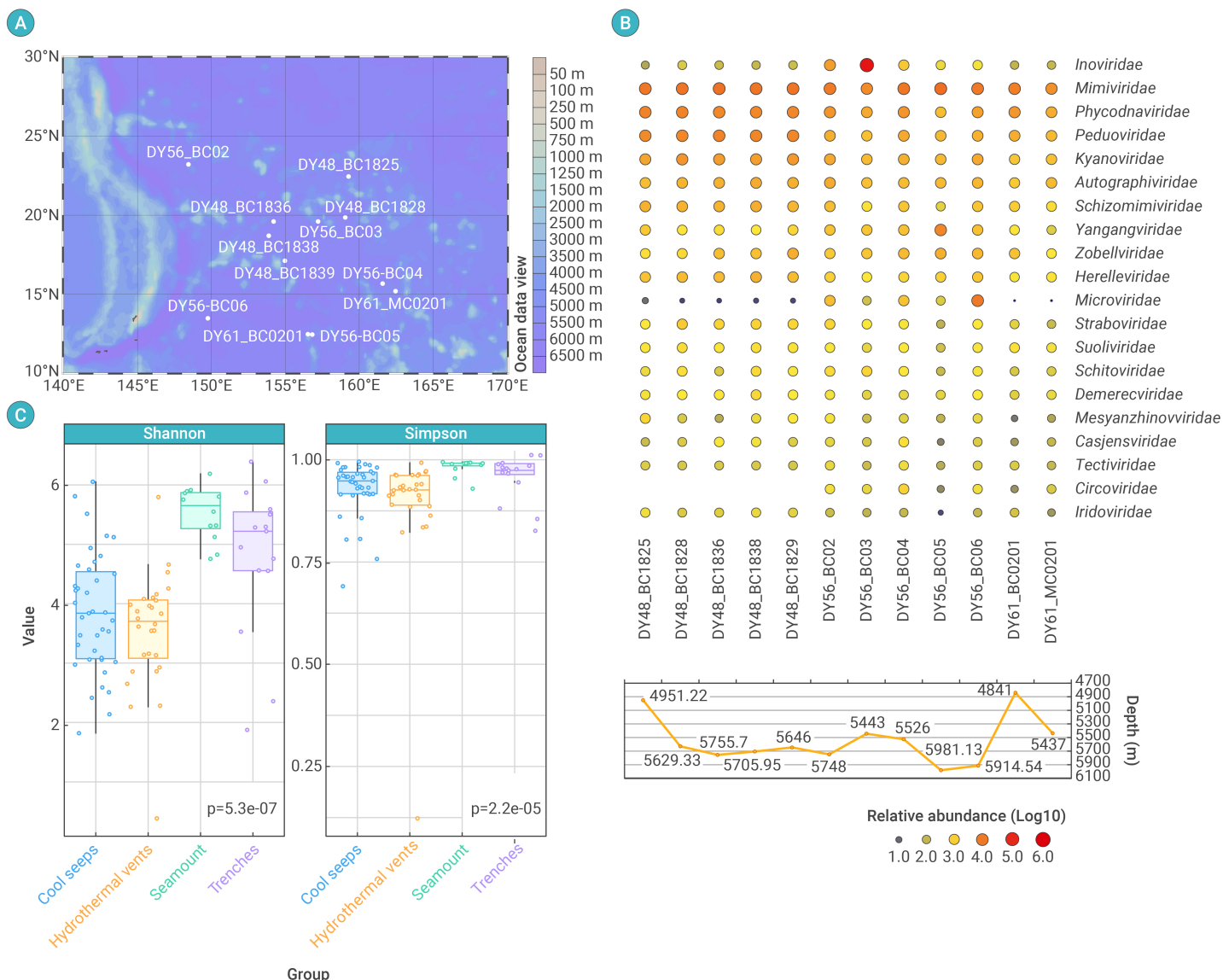


Figure 1. Overview of the seamount sediments viruses distribution and identified in this study (A) Map of sampling of the seamount sediment in the Northwest Pacific Ocean in this study. (B) Relative abundance of viral contigs at the top 20 family level (upper half) and the water depth of sampling stations. (lower half). (C) Shannon and Simpson indices of the viral community from different environments. Relative sequence abundance was calculated using CoverM v0.6.1.

IMG/VR v4 database by BLASTp with parameters: e-value $\geq 1e-10$, 244 and 475 high-quality sequences were retrieved, respectively. Subsequently, reference sequences from the IMG/VR database were de-replicated using CD-HIT v4.8.1 (parameters: -c 1 -G 0 -M 0 -aS 0.8 -T 8 -n 9), 119 and 138 viral reference protein sequences were retained from IMG/VR v4 database. Microorganism reference sequences were downloaded from the UniProtKB database, resulting in 247 and 4,766 reference sequences, respectively. Reference sequences from the UniProtKB database were de-replicated using CD-HIT v4.8.1, then 204 and 193 microorganism reference sequences were retained from the UniProtKB database. The construction of a phylogenetic tree was described in the "Phylogenetic analysis" section.

Comparison to specific AMGs from other habitats.

To compare SMVG viral heme, cobalamin, and metal AMGs with those from other habitats, three ecological environmental metagenomic data sets were used, including Treachs,⁸ hydrothermal vents,⁶⁶ cool seeps⁴⁷ (samples detail see Table S9), and data from this study comprising 79 samples in total. Contigs associated with heme, cobalamin, and metal functional annotations from the KEGG database were singled out. Then they passed to CoverM v0.6.1 to count coverage profiles across metagenomic data sets (parameters: -p bwa-mem --min-read-percent-identity 0.95 --min-read-aligned-percent 0.75 -m tpm -t 20).

Statistical Analysis

All statistical analyses were performed using R version 4.2.3. vegan v2.6.4 was used to calculate the alpha diversity based on the relative abundance of viral contigs.⁶⁷ To compare different environments (Table S9), i.e., seamount sediments, cold seeps, hydrothermal vents, and hadal trenches, all viral contigs with lengths over 10 kb were selected to calculate the Shannon and Simpson diversities.

RESULTS AND DISCUSSIONS

Most previous studies on seamounts and deep sea sediments have focused on components of microbial ecosystems.^{21,68,69} These studies were complemented here by studying the viral community of seamount sediments. Here, the distinct viral biogeographic distribution patterns, the relationship between viruses and their hosts, endemism, functional genes, and biosynthetic potentials of viral assemblages inhabiting the seamount sediments were explored.

Overview of viral community composition in the seamount sediments

This seamount viral genome (SMVG) dataset was assembled from 12 DNA virome libraries with 225.4 GB of sequence data across 12 seamount sediment samples. Database assembly yielded 29.97 GB of sequence data (Figure 1A). From this dataset, 81,418 viral contigs with a size ≥ 1.5 kb were

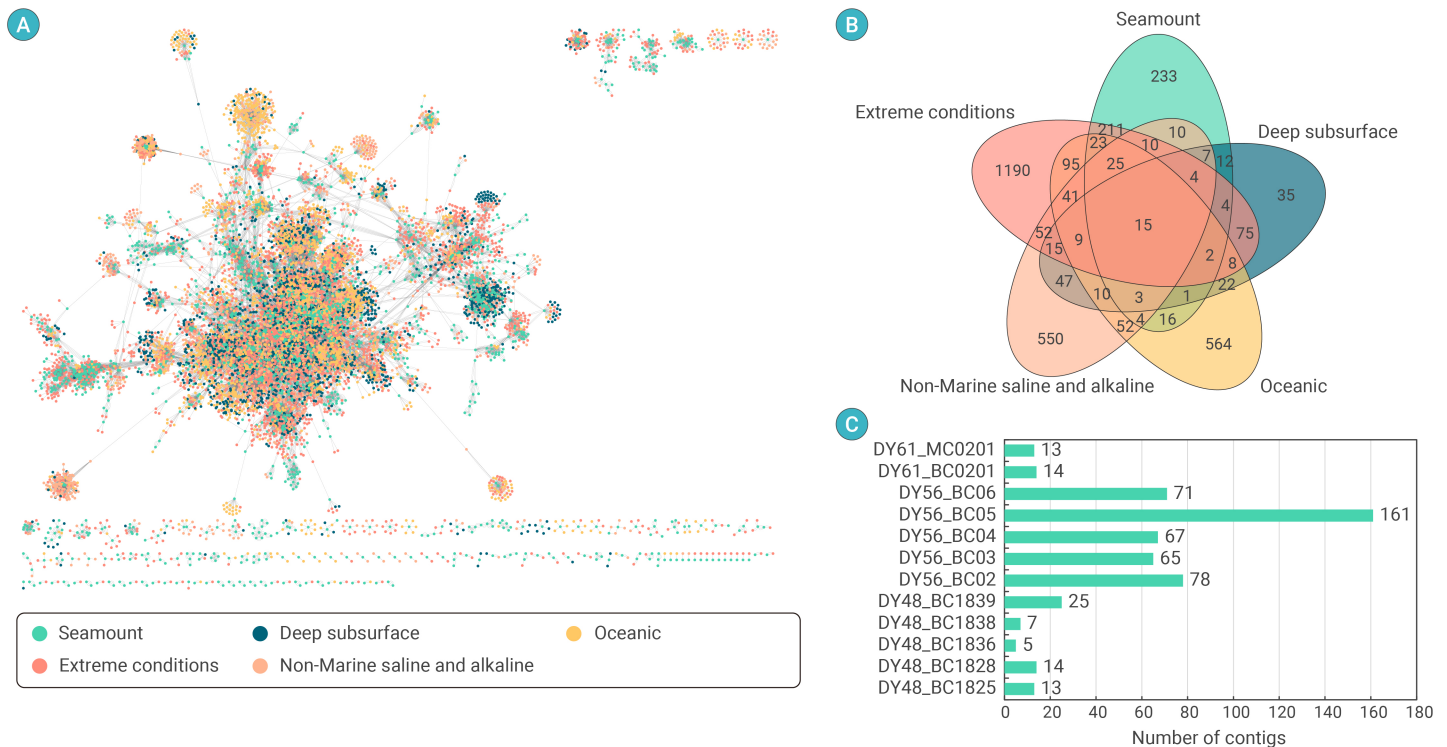


Figure 2. Comparison analysis between seamount sediments vOTUs and other environmental data sets (A) The gene-sharing network compares five environmental viral contigs, five habitats including seamount sediments, extreme conditions (i.e., challenger deep, cold seep, and hydrothermal vents), deep subsurface, non-marine saline and alkaline, and oceanic. (B) Venn diagram of shared viral clusters (VCs) among five environments. (C) Bar charts of specific viral contigs among different stations based on (B).

retrieved and classified into 78,069 viral Operational Taxonomic Units (vOTUs). Of these, 444 and 244 viral contigs were identified as complete or high-quality genome sequences, respectively (Supplemental Table S2). The most abundant vOTUs differed between the three cruises (Figure S1) and the five viromes of DY56 have the highest number ($n=55$). Based on the VIBRANT results, 35,510 of the 78,069 vOTUs (45.49%) were predicted to be lytic viruses, and 716 (0.91%) were predicted to be temperate viruses (Figure S2).

A total of 17,794 vOTUs (22.79%) were taxonomically assigned to 63 families and 18 to class level. *Inoviridae* (7.71%), *Mimiviridae* (5.42%), and *Phycodnaviridae* (2.65%) had the highest relative abundance (Figure 1B). Of the top 20 most abundant families, 12 belonged to class *Caudoviricetes*, with no noteworthy distinction between sites. (Figure 1B). By contrast, *Microviridae*, *Circoviridae*, *Smacoviridae* and *Vilyaviridae*, which are icosahedral viruses with a circular single-stranded DNA genome, were mainly detected in sediments of the DY56 cruise (Figure S3).

A comparison of Shannon diversity and Simpson diversity indices of the viral community revealed different diversities between the four environments: cold seeps (number of viral contigs, $n=2,885$), hydrothermal vents ($n=935$), seamounts ($n=2,883$), and trenches ($n=3,736$). Seamount sediment had a higher viral diversity than the other three environments (Figure 1C).

Phylogenetic diversity of *Caudoviricetes* and NCLDV includes novel and ancient viruses

The phylogenetic tree predominantly grouped the seamount sediment viruses as unclassified groups within the *Caudoviricetes* class (Figure S4). Different types of TerL domains were located on significantly different branches and the sequences in the samples were widely distributed among the various types of TerL domains. Interestingly, the terminase sequences formed a distinct and divergent group with long branches (Figure S4). Furthermore, the presence of specific terminase-based viral clades solely in the samples investigated here and located near the root of the phylogenetic tree implies that they could potentially represent endemic and ancient *Caudoviricetes* taxa in the seamount sediments.

Furthermore, nucleocytoplasmic large DNA viruses (NCLDVs) showed high relative abundance in the 12 samples. Thus, viral proteins were matched to the polB marker gene using a hmmsearch search, which detected 36 polB

sequences in the samples (Supplemental Table S4). Unlike the phylogenetic tree of the TerL domain, a large proportion of the polB sequences formed an independent branch, except for reference sequences, indicating that the majority of NCLDVs in the seamount sediments might have a similar phylogenetic status but were different from known NCLDVs (Figure S5). As reference sequences were collected from shallow or intermediate water depths (1–1000 m)⁷⁰ and 30 polB seamount sediment sequences have a close phylogenetic position with reference sequences, it was hypothesized that the deep-sea viruses could be linked to the surface ocean.^{71,72} However, this hypothesis remains unconfirmed.

Comparison of viral endemism between seamount sediments and other ecosystems

To investigate the relationship between seamount vOTUs and publicly available virus sequences from a vast number diversity of ecosystems, a gene-sharing network was constructed using vConTACT 2 (Figure 2A). Seamounts from this study, deep subsurface, oceanic, non-marine saline and alkaline and extreme ecosystem were grouped into 3,345 viral clusters (VCs). Surprisingly, only 15 VCs were shared amongst all ecosystems (Figure 2B). Different virus community structures were observed among the 11,982 vOTUs identified from the seamount sediments. 580 VCs were formed, of which 233 were exclusively found in this ecosystem and not detected elsewhere. This implies that 40.17% of total vOTUs in seamount sediments were absent from the other four habitats in the IMG/VR v4 dataset. In addition, the presence of novel and likely endemic viruses was evident in the seamount sediments, with each station containing distinctive vOTUs, with DY56_BC05 having the highest number of unique vOTUs (Figure 2C). Deep-sea seamounts are generally covered with ferromanganese (Fe–Mn) crusts.^{73–75} These Fe–Mn crusts and nodules support microbial communities and play a role in the cycling of Fe and Mn and the growth of Fe–Mn crusts.⁶⁸ Therefore, cobalt crusts identified on cruise DY56^{76–79} may have contributed to the endemism of viruses in the seamount sediments.

Diversity and virus-host linkages in seamount sediments

To explore the association between viruses and their potential hosts, an analysis involving sequence similarity comparison, oligonucleotide frequen-

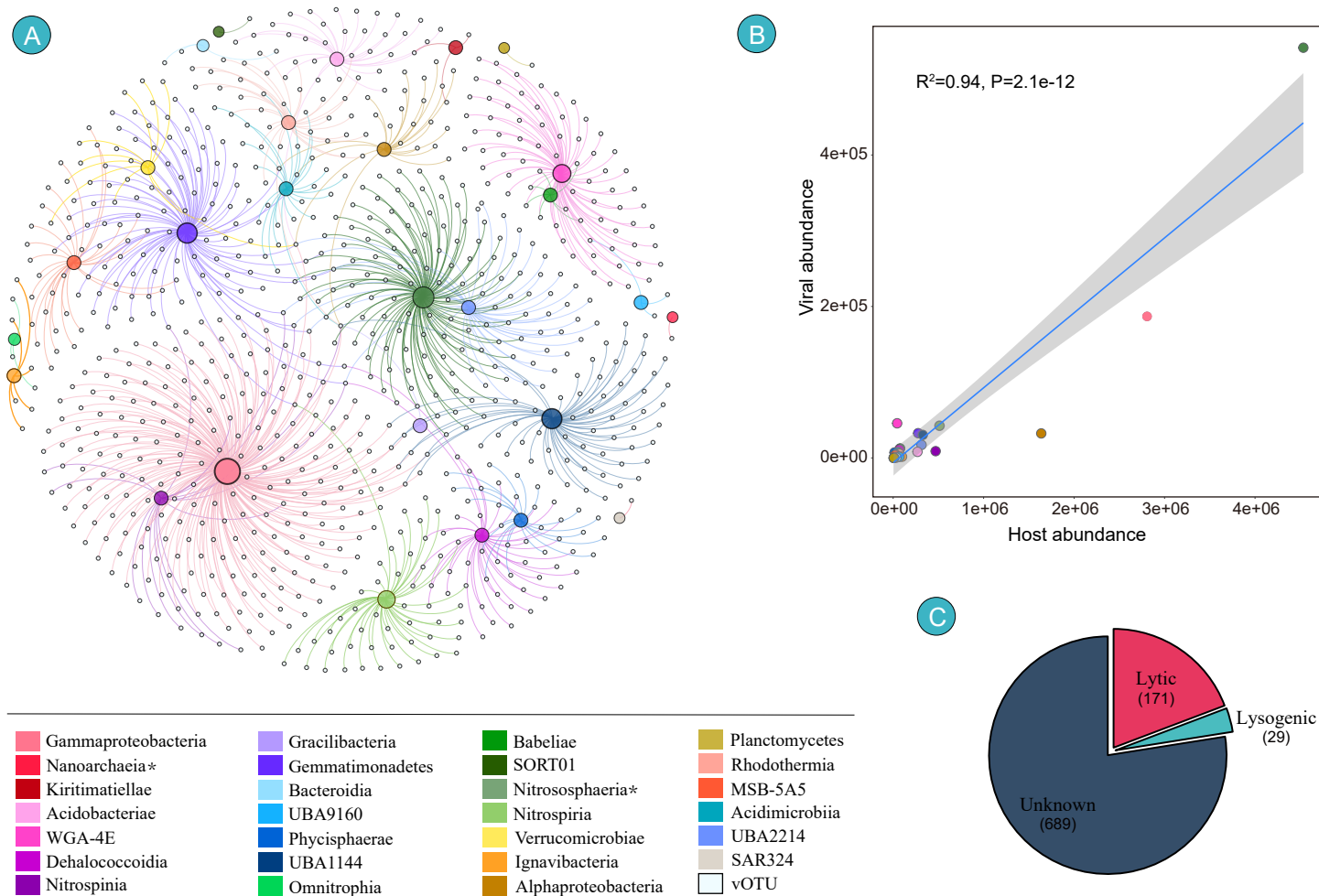


Figure 3. Viruses and their potential host linkage in seamount sediments (A) The connection between viruses and potential hosts was based on the Nucleotide sequence homology, Oligonucleotide frequency (ONF), Transfer RNA (tRNA) match, and CRISPR spacer match methods. The size of the host circle depended on the number of vOTU. (B) Significant Pearson correlation between relative abundances of viruses and their hosts. (C) The proportion of host-predicted viral lifestyles. Predicted hosts in (A) and (B) are indicated in the color bar on the bottom left.

cies, tRNA sequences and CRISPR spacers was performed.⁸⁰ However, most potential vOTU hosts could not be classified, and only 843 vOTUs in the seamount sediments (1.14% of the total vOTUs) were linked to twenty-six potential hosts (Table S6).

Overall, only 30 vOTUs predicted two or more potential hosts, with the majority (813) only predicting one potential host (Figure 3A), consistent with previous observations.^{47,81,82} Even though 98.86% of the contigs identified in this study could not be linked to a host, a significant positive correlation between viruses and hosts, based on the matching of the relative abundance of host-linked viruses ($P < 0.001$), was observed, supporting the previously predicted result that the composition of predicted microbial hosts aligns well with the composition of their viruses (Figure 3B). Additionally, when examining the proportion of host-linked viruses with lysogenic and lytic lifestyles using VIRBRANT, most of the viruses with potential hosts (171 vOTUs) were appear to have a lytic lifestyle (Figure 3C).

Predicted prokaryotic hosts of vOTUs encompassed two archaeal and twenty-four bacterial classes, representing two archaea and twenty bacterial phyla (Table S6). Of the predicted classes, *Gammaproteobacteria* and *Nitrososphaeria* were the most common hosts, accounting for 25.20% and 11.92% of virus-host linkages, respectively (Table S6). By comparing the data present here to that in the IMG/VR v4 database, two novel viral-bacterial linkages were identified, namely *SORT01* and *Planctomycetes* (Table S6), which implies that these two classes were viral putative hosts, in particular, *SORT01* belongs to *Marinisomatota*, which contains substantial proportions of predicted heme auxotrophs,⁸³ and *Planctomycetes* an important player in global carbon and nitrogen cycles.⁸⁴ In addition, some host-virus associations not frequently detected in previous studies were detected in the data

presented here, including *Nitrososphaeria*, *Nitrospina*, *Omnitrophia*, *UBA2214*, *MSB-5A5*, *UBA1144*, *WGA-4E* and *UBA9160*.^{47,85,86} These results confirm that seamount sediment ecosystems also contain diverse prokaryotes with a species complexity similar to shallower seamounts.⁸⁶ It also implies that seamount sediments were diverse in prokaryotes and had viruses actively infecting these microbes.

Potential viral auxiliary metabolic genes (AMGs) associated with biogeochemical cycling and cold shock of host cells

Benthic microorganisms on seamounts occupy diverse habitats and experience a range of environmental conditions.^{86,87} Deep seamounts are usually below the zone of influence of the deep-scattering layer and the biogeography at such depth tends to be less well-characterized. Several studies have highlighted the important contribution of viruses to the carbon, nitrogen, sulfur and phosphorus cycles in sediments.^{6,8,47}

The ORFs identified the Clusters of Orthologous Genes (COGs) function classification through the eggNOG pipeline. Based on the COG classification, 75,057 ORFs (78.90% of total ORFs) were clustered into 25 COG categories (Figure S6). The highest relative abundance of orthologs covered the categories of information storage and processing (LK), metabolism (CE), and cellular processes and signaling (O). Numerous ORFs categorized as having an unknown function (S) were detected at several stations, which implies that seamount sediments still have considerable exploration potential. In addition, according to the CAZyme database, the majority of these genes were associated with GlycosylTransferase (GTs) and Glycoside Hydrolases (GHs), which help assemble or break down complex polysaccharides (Figure S7).^{5,61}

Within the seamount sediment viromes, sixteen categories of AMGs

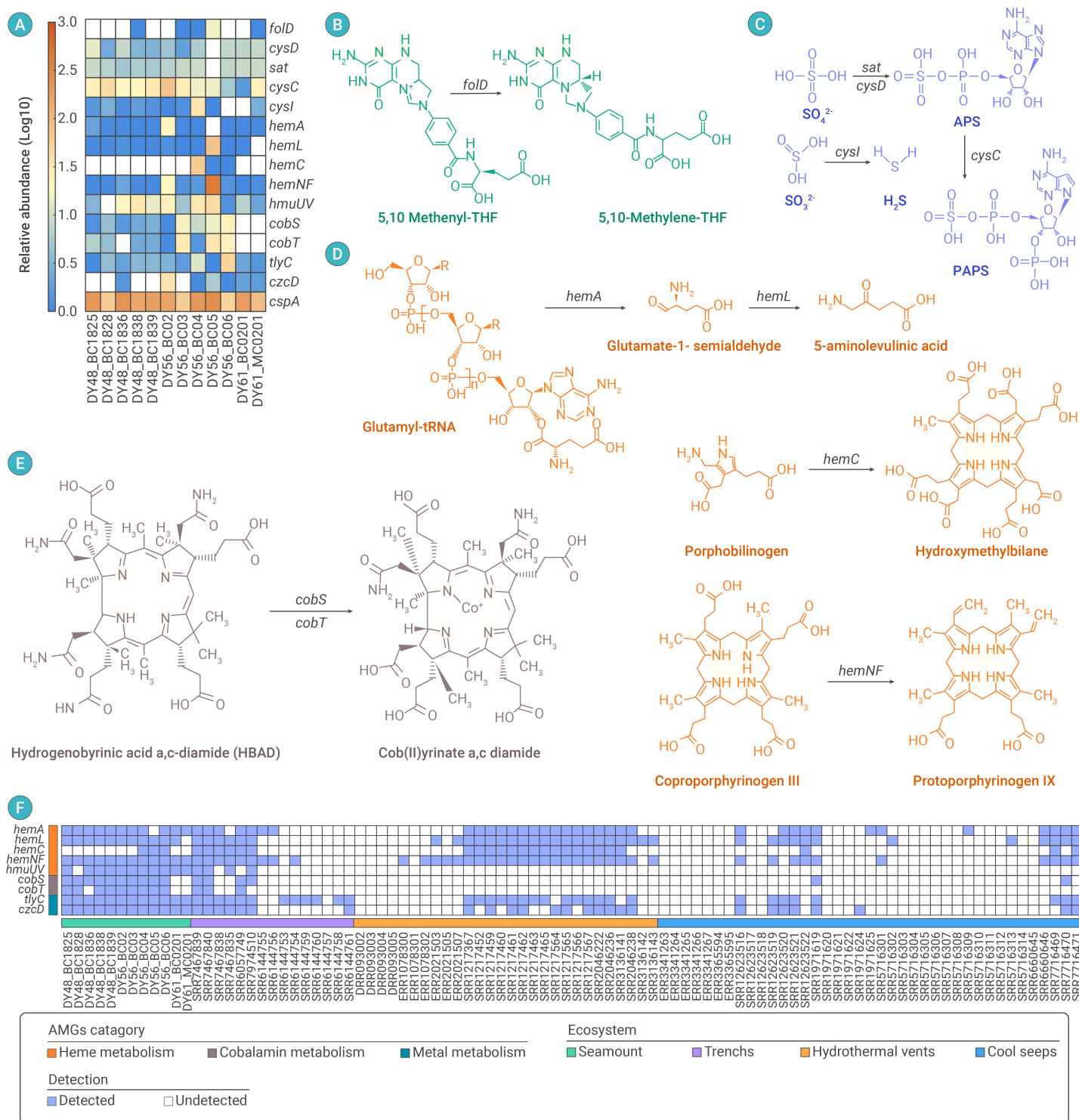


Figure 4. Viral auxiliary metabolic genes within the seamount viral genome (SMVG) dataset (A) Heatmap of the relative abundance for different metabolic genes, and results were log10 transformed for description. The cycling pathway of carbon (B) sulfur (C), heme biosynthesis (D), and cobalamin biosynthesis (E) in seamount sediments. APS: Adenylyl sulfate; PAPS: 3'-Phosphoadenylyl sulfate. (F) is the presence or absence of heme, cobalamin and metal-associated AMGs in four environments. Genes that were not detected in a given sample remained white.

involved in the carbon, sulfur, heme, and cobalamin cycle, and two AMGs associated with metal were detected based on the KEGG database (Table S7). For the carbon metabolism genes, only one AMGs, namely *folD*, which was involved in the conversion of the 5,10-Methenyl-THF to 5,10-Methylene-THF, was detected at five out of the twelve sites (Figure 4A). The Wood-Ljungdahl (WL) pathway was the only carbon fixation pathway in which AMGs were involved and many anaerobic prokaryotes use this pathway to convert one-carbon substrates (i.e. carbon dioxide or formate) into acetate (Figure 4B).⁸⁹⁻⁹⁰

Pathways for AMGs that encode sulfur metabolism-related genes were also detected (Figure 4A). It is noteworthy that most sites contained the function for the assimilatory sulfate reduction pathway (ASR) (Figure 4C), reflecting that in seamount sediments, viruses tend to complement the hosts' metabolism through the ASR pathway to accomplish the reduction of sulfate. Furthermore, it was found that *UBA1144* and *Nitrospira*, typical sulfate-reducing bacteria,⁹¹ were associated with 12.26% of the viral contigs, suggesting that the hosts could acquire the corresponding AMGs which would avoid the energetic costs of the ASR pathway, and the ASR pathway

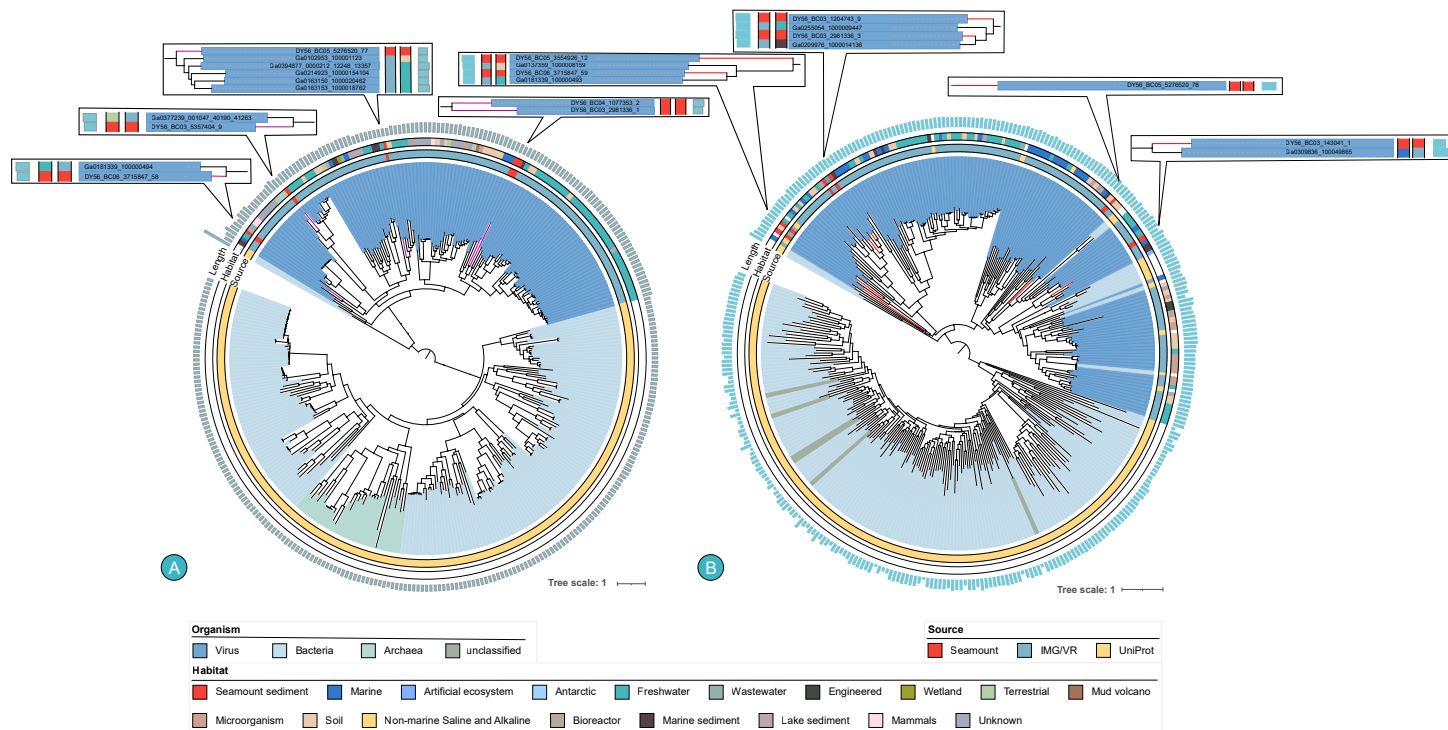


Figure 5. The maximum-likelihood phylogenetic tree of cobaltochelatase *cobS* (A) and cobaltochelatase *cobT* (B). *cobS* and *cobT* from seamount sediments were performed in different colors (*cobS*: purple, *cobT*: red). The inner circle indicated the sequence source and the habitat of sequences was indicated by the outer circle. The length of each *cobS* sequence ranges from 199 to 538 bp and the length of each *cobT* sequence ranges from 91 to 929 bp.

might primarily drove the sulfur cycles in seamount sediments.

An AMG encoding a cold shock protein (*cspA*) was also identified at all stations (Table S8). The *cspA* gene functions as an RNA chaperone by destabilizing secondary structures in mRNA, thereby maintaining the single-stranded state of mRNA and concentrating it on transcription and translation.⁹²⁻⁹⁴ The presence of this gene in these samples was consistent with previous studies,⁹⁵ indicating that viruses might help the host cell live in the cold deep oceans.

Comparisons between the seamount sediments and other habitats suggest seamount sediments are a hotspot of AMGs against heme, cobalt and metal metabolisms

Because deep-sea environments are characterized by high hydrostatic pressure, hypoxia and anoxia, extremely high/low temperatures, absence of photosynthesis and nutrient limitation, it is imperative for deep-sea organisms to find adaptations for their survival.^{96,97} Interestingly, two phyla lacking heme biosynthesis were detected in the samples, including *Dependentiae* (class *Babeliae*) and *Patescibacteria* (class *Gracilibacteria*).⁸³

Heme (iron-protoporphyrin IX) is an iron-containing tetrapyrrole that is essential for many biochemical reactions, including oxygenation reactions, oxygen binding and transport, oxidative stress responses, electron transport, oxygen metabolism, and xenobiotic detoxification.⁹⁸⁻¹⁰⁰ Heme is also used as a cofactor of hemoproteins. Some prokaryotes lack a heme biosynthesis pathway or contain substantial proportions of predicted heme auxotrophs (e.g., *Dependentiae*, *Patescibacteria*, and *Marinisomatota* in the SMVG data set), these thus require exogenous heme for their growth.⁸³

Five AMGs related to heme were detected (Figure 4A). While most AMGs for heme biosynthesis were detected in all twelve samples, the presence of *HemC* was limited to a few specific sites (DY56_BC04, DY56_BC05, and DY56_BC06), indicating that viral-mediated conversion of Porphobilinogen to hydroxymethylbilane might only occur at these three sites (Figure 4D). Generally, the formation of 5-aminolevulinic acid (ALA) was considered to be the rate-limiting step of heme biosynthesis, and *HemA* and *HemL* are associated with ALA synthesis (Figure 4D).^{91,101} These two genes encoding ALA were abundant at most sites, suggesting that AMGs can effectively assist in the synthesis of heme in host cells that lack the heme synthesis pathway. Although heme can efficiently help transport iron, heme toxicity is produced

when the bacteria's heme tolerance concentration is exceeded.^{99,102} Heme transport systems-related AMGs (*HmuUV*) were searched in seamount sediment samples (Figure 4A). The heatmap shows that the samples had a high relative abundance of *HmuUV*, the heme transport system for ATP-binding protein that powers substrate transports.¹⁰³ These results indicate that viruses could provide an additional source, e.g., AMGs, to supplement the heme biosynthetic pathway.

Nine noteworthy AMGs found in trenches, cool seeps, and hydrothermal vents environments were underrepresented (Figure 4F), including two AMGs related to the utilization and transport of heavy metals were detected in the SMVG dataset, including magnesium and cobalt exporter (*tlyC*) and cobalt-zinc-cadmium efflux system protein (*czcD*), which related to the cobalt metabolism (Figure 4A). Although AMGs associated with heme and metal metabolism were detected more frequently in the other three environments, primarily in sediments of trenches, plumes of hydrothermal vents and sediment of cool seeps, when compared to cobalamin-associated AMGs (Supplemental Table S8), there were significant differences between seamount and most other sites and this suggests that the above metabolisms were more critical biogeochemical processes in seamount sediments than in other environments (Figure 4F).

Intriguingly, two AMGs encoding cobaltochelatase subunit (*cobS* and *cobT*) were identified in primarily seamount sediments (Figure 4F). Cobalamin (Vitamin B₁₂), also known as "nature's most beautiful cofactor," is used as a cofactor in multiple biological processes, including amino acid metabolism and gene expressions, and is also required by many prokaryotes.¹⁰⁴⁻¹⁰⁷ *cobS* and *cobT* were the subunits that constituted the cobaltochelatase with *cobN*. The three-subunit composition helped cobalt to be inserted into hydrogenobyrinic acid a,c-diamide (HBAD) and form the cobalamin (Figure 4E).^{108,109} Therefore, *cobS* and *cobT* genes were analyzed at the amino acid level using IQ-TREE, respectively. Detailed reference sequences, including download information and protein names, can be found in Table S7 and Table S8. The two phylogenetic trees exhibit the same pattern tree. The *cobS* protein sequences of viruses appear to be evolutionarily distinct from the *cobS* protein sequences of microbes (Figure 5A); even though some *cobT* protein sequences of viral associated with bacteria, the majority of viral sequences were in the same monophyletic clade, similar to the *cobS* phylogenetic tree (Figure 5B). However, as all seamount sediment protein

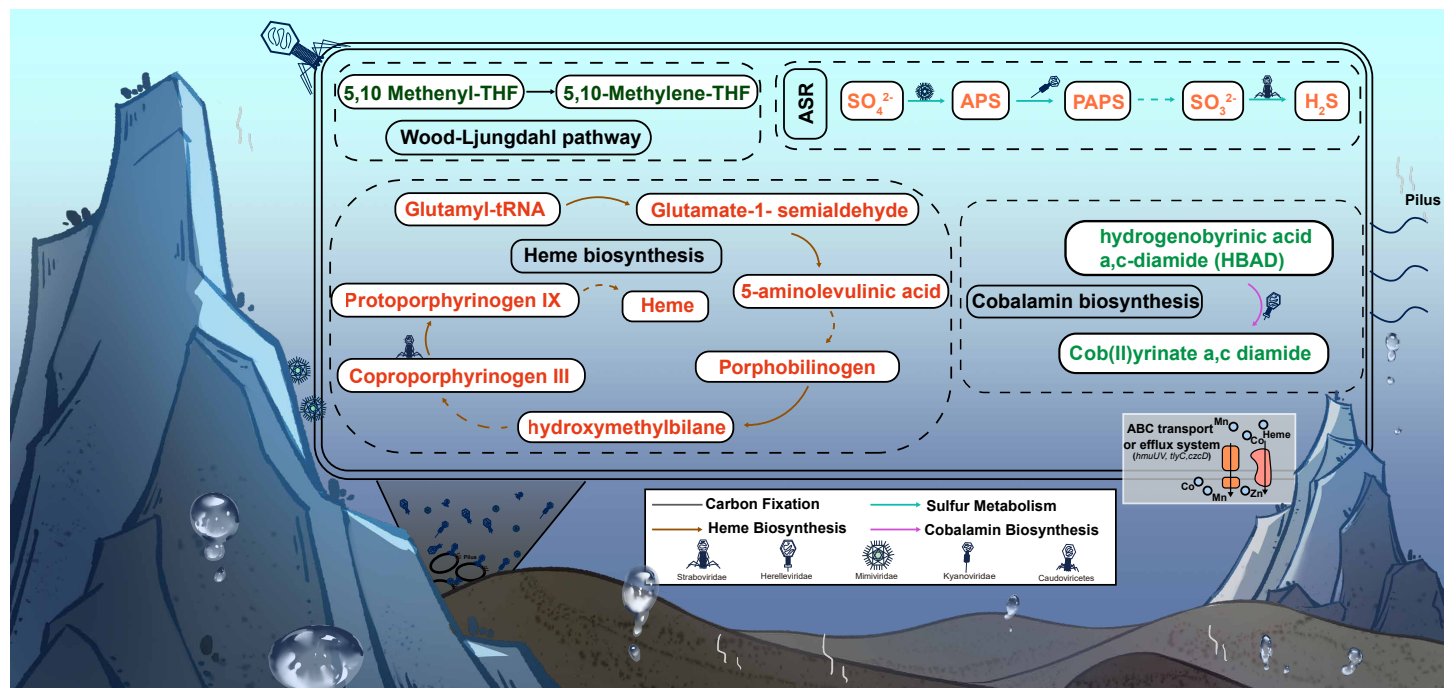


Figure 6. Proposed elemental cycling in seamount sediments involving viruses encoding AMGs The figure highlights the proposed elemental cycles of seamount sediments in which AMGs were involved. Elements, including Carbon, Sulfur, Heme, Cobalamin, and Metal, have different pathways indicated by different arrow colors. ABC and efflux system (orange and pink) transporters were involved. ASR, assimilatory sulfate reduction.

sequences cluster with IMG/VR sequences (viral protein sequences), it was concluded that while viral *cobS* and *cobT* originated from microbes. Subsequently, there was divergent evolution between viruses and microbes. The trend of virus-microbe co-evolution can be inferred in the phylogenetic tree of *cobT* with some bacteria sequences inserted in the viral clade, which indicates that horizontal gene transfer (HGT) may have occurred during the evolution of *cobT*.

Additionally, the distribution patterns of both AMGs showed that cobalamin-associated genes are more readily available in seamount sediment than in trenches, cool seeps, and hydrothermal vents environments. Furthermore, due to the chromium-rich nature of the sampling site, it was not surprising that the AMGs could assist the host in synthesizing cobalamin. Heme and cobalamin play crucial roles as nutrient carriers within deep-sea sediment environments. In conclusion, viruses likely contribute to cobalamin biosynthesis in their hosts since microorganisms are the only source of cobalamin in the ocean and are an essential vitamin for supporting productivity and interactions in the microbial community.^{105,110}

The AMGs observed in the seamount sediments indicate that viruses in the deep ocean contribute to prokaryotic metabolism and nutrient cycling and assist the host cells in adapting to nutrient-limiting and cold habitats (Figure 6).

CONCLUSION

Here, the diverse and unique viral assemblages inhabiting the seamount sediments were identified, most of which differed from other cultured and environmental viral communities. *Nitrososphaeria* was one of the most common hosts, and two novel virus-host linkages were identified from the SMVG dataset, e.g., *SORT01* and *Planctomycetes*. Viruses encoding AMGs, encompassing genes related to the metabolism of carbon, sulfur, heme, cobalamin, and heavy metals, suggest potentially important roles of viruses in the biogeochemical and heavy-metals of seamount sediments in the Northwest Pacific Ocean. This study sheds light on the diversity and ecological significance of benthic viruses and provides a baseline for incorporating viruses in a model of deep-sea seamount sediment ecosystems. With the continuous improvement of sampling and *in-situ* cultivation technologies and increasing numbers of scientific cruises focusing on deep-sea seamounts, a more comprehensive understanding of viruses in the deep-sea seamounts is expected shortly.

REFERENCES

- Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541–548. DOI: 10.1038/21119.
- Brum, J.R. and Sullivan, M.B. (2015). Rising to the challenge: Accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **13**: 147–159. DOI: 10.1038/nrmicro3404.
- He, T., Li, H., and Zhang, X. (2017). Deep-sea hydrothermal vent viruses compensate for microbial metabolism in virus-host interactions. *mBio* **8**:00893–17. DOI: 10.1128/mBio.00893-17.
- Zhao, K., Liu, A., and Xia, Y. (2020). Insights into hepatitis b virus DNA integration-55 years after virus discovery. *The Innovation* **1**: 100034. DOI: 10.1016/j.xinn.2020.100034.
- Brussaard, C.P., Wilhelm, S.W., Thingstad, F., et al. (2008). Global-scale processes with a nanoscale drive: The role of marine viruses. *ISME J.* **2**: 575–578. DOI: 10.1038/ismej.2008.31.
- Braga, L.P.P., Orland, C., Emilson, E.J.S., et al. (2022). Viruses direct carbon cycling in lake sediments under global change. *Proc. Natl. Acad. Sci. U.S.A.* **119**: e2202261119. DOI: 10.1073/pnas.2202261119.
- Yang, Q., Gao, C., Jiang, Y., et al. (2019). Metagenomic characterization of the viral community of the south scotia ridge. *Viruses* **11**: 95. DOI: 10.3390/v11020095.
- Jian, H., Yi, Y., Wang, J., et al. (2021). Diversity and distribution of viruses inhabiting the deepest ocean on earth. *ISME J.* **15**: 3094–3110. DOI: 10.1038/s41396-021-00994-y.
- Zhou, H., Chen, P., Zhang, M., et al. (2021). Revealing the viral community in the hadal sediment of the new britain trench. *Genes* **12**: 990. DOI: 10.3390/genes12070990.
- Rogers, A.D. (2018). The biology of seamounts: 25 years on. *Adv. Mar. Biol.* **79**: 137–224. DOI: 10.1016/bs.amb.2018.06.001.
- Rogers, A.D. (2019). Chapter 23 - threats to seamount ecosystems and their management. Sheppard, C. (ed). *World seas: An environmental evaluation* (second edition) (Academic Press), pp: 427–451. DOI: 10.1016/B978-0-12-805052-1.00018-8.
- Marzia, B., Martina, C., Federico, B., et al. (2020). Unveiling the deep biodiversity of the janua seamount (ligurian sea): First mediterranean sighting of the rare atlantic bamboo coral chelidonis aurantiaca studer, 1890. *Deep-Sea Res. I* **156**: 103186. DOI: 10.1016/j.dsr.2019.103186.
- Consalvey, M., Clark, M.R., Rowden, A.A., et al. (2010). Life on seamounts. *Life in the World's Oceans*. pp:123–139. DOI:10.1002/9781444325508.ch7.
- Danovaro, R., Corinaldesi, C., Luna, G.M., et al. (2009). Prokaryote diversity and viral production in deep-sea sediments and seamounts. *Deep-Sea Res. II* **56**: 738–747. DOI: 10.1016/j.dsr2.2008.10.011.
- Rivera, J., Canals, M., Lastras, G., et al. (2016). Morphometry of concepcion bank: Evidence of geological and biological processes on a large volcanic seamount of the canary islands seamount province. *PLOS One* **11**: e0156337. DOI: 10.1371/journal.

- pone.0156337.
16. Yesson, C., Clark, M.R., Taylor, M.L., et al. (2011). The global distribution of seamounts based on 30 arc seconds bathymetry data. *Deep-Sea Res. I* **58**: 442–453. DOI: 10.1016/j.dsr.2011.02.004.
 17. Rogers, A., Brierley, A., Croot, P., et al. (2015). Delving deeper: Critical challenges for 21st century deep-sea research. DOI: 10.13140/RG.2.1.1868.2327. https://www.marineboard.eu/sites/marineboard.eu/files/public/publication/EMB_P22_Web_v4.pdf.
 18. Thomas, E., Anderson, R.E., Li, V., et al. (2021). Diverse viruses in deep-sea hydrothermal vent fluids have restricted dispersal across ocean basins. *mSystems* **6**: e0006821. DOI: 10.1128/mSystems.00068-21.
 19. Dai, S., Zhao, Y., Li, X., et al. (2020). The seamount effect on phytoplankton in the tropical western pacific. *Mar. Environ. Res.* **162**: 105094. DOI: 10.1016/j.marenvres.2020.105094.
 20. McAllister, S.M., Vandzura, R., Keffer, J.L., et al. (2021). Aerobic and anaerobic iron oxidizers together drive denitrification and carbon cycling at marine iron-rich hydrothermal vents. *ISME J.* **15**: 1271–1286. DOI: 10.1038/s41396-020-00849-y.
 21. Liu, J., Zhang, W., Li, X., et al. (2017). Bacterial community structure and novel species of magnetotactic bacteria in sediments from a seamount in the mariana volcanic arc. *Sci. Rep.* **7**: 17964. DOI: 10.1038/s41598-017-17445-4.
 22. Gao, C., Liang, Y., Jiang, Y., et al. (2022). Virioplankton assemblages from challenger deep, the deepest place in the oceans. *iScience* **25**: 104680. DOI: 10.1016/j.isci.2022.104680.
 23. Dean, F.B., Hosono, S., Fang, L., et al. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U.S.A.* **99**: 5261–5266. DOI: 10.1073/pnas.082089499.
 24. Thoendel, M., Jeraldo, P., Greenwood-Quaintance, K.E., et al. (2017). Impact of contaminating DNA in whole-genome amplification kits used for metagenomic shotgun sequencing for infection diagnosis. *J. Clin. Microbiol.* **55**: 1789–1801. DOI: 10.1128/jcm.02402-16.
 25. Chen, S., Zhou, Y., Chen, Y., et al. (2018). Fastp: An ultra-fast all-in-one fastq preprocessor. *Bioinformatics (Oxford, England)* **34**: i884–i890. DOI: 10.1093/bioinformatics/bty560.
 26. Kechin, A., Boyarskikh, U., Kel, A., et al. (2017). Cutprimers: A new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J. Comput. Biol.* **24**: 1138–1143. DOI: 10.1089/cmb.2017.0096.
 27. Yang, Y., Liu, G., Ye, C., et al. (2019). Bacterial community and climate change implication affected the diversity and abundance of antibiotic resistance genes in wetlands on the qinghai-tibetan plateau. *J. Hazard. Mater.* **361**: 283–293. DOI: 10.1016/j.jhazmat.2018.09.002.
 28. Li, D., Luo, R., Liu, C.M., et al. (2016). Megahit v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods (San Diego, Calif.)* **102**: 3–11. DOI: 10.1016/j.jmeth.2016.02.020.
 29. Nurk, S., Meleshko, D., Korobeynikov, A., et al. (2017). Metaspades: A new versatile metagenomic assembler. *Genome Res.* **27**: 824–834. DOI: 10.1101/gr.213959.116.
 30. Gu, C., Liang, Y., Li, J., et al. (2021). Saline lakes on the qinghai-tibet plateau harbor unique viral assemblages mediating microbial environmental adaption. *iScience* **24**: 103439. DOI: 10.1016/j.isci.2021.103439.
 31. Mikheenko, A., Saveliev, V., and Gurevich, A. (2016). Metaquast: Evaluation of metagenome assemblies. *Bioinformatics (Oxford, England)* **32**: 1088–1090. DOI: 10.1093/bioinformatics/btv697.
 32. Roux, S., Enault, F., Hurwitz, B.L., et al. (2015). Virsorter: Mining viral signal from microbial genomic data. *PeerJ* **3**: e985. DOI: 10.7717/peerj.985.
 33. Ren, J., Ahlgren, N.A., Lu, Y.Y., et al. (2017). Virfinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**: 69. DOI: 10.1186/s40168-017-0283-5.
 34. Nayfach, S., Camargo, A.P., Schulz, F., et al. (2021). Checkv assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**: 578–585. DOI: 10.1038/s41587-020-00774-7.
 35. Fu, L., Niu, B., Zhu, Z., et al. (2012). Cd-hit: Accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)* **28**: 3150–3152. DOI: 10.1093/bioinformatics/bts565.
 36. Li, B., Ruotti, V., Stewart, R.M., et al. (2010). Rna-seq gene expression estimation with read mapping uncertainty. *Bioinformatics (Oxford, England)* **26**: 493–500. DOI: 10.1093/bioinformatics/btp692.
 37. Patro, R., Duggal, G., Love, M.I., et al. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**: 417–419. DOI: 10.1038/nmeth.4197.
 38. von Meijenfildt, F.A.B., Arkhipova, K., Cambuy, D.D., et al. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with cat and bat. *Genome Biol.* **20**: 217. DOI: 10.1186/s13059-019-1817-x.
 39. Kieft, K., Zhou, Z., and Anantharaman, K. (2020). Vibrant: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**: 90. DOI: 10.1186/s40168-020-00867-0.
 40. Esterman, E.S., Wolf, Y.I., Kogay, R., et al. (2021). Evolution of DNA packaging in gene transfer agents. *Virus Evol.* **7**: veab015. DOI: 10.1093/ve/veab015.
 41. Endo, H., Blanc-Mathieu, R., Li, Y., et al. (2020). Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat. Ecol. Evol.* **4**: 1639–1649. DOI: 10.1038/s41559-020-01288-w.
 42. Katoh, K. and Standley, D.M. (2016). A simple method to control over-alignment in the mafft multiple sequence alignment program. *Bioinformatics (Oxford, England)* **32**: 1933–1942. DOI: 10.1093/bioinformatics/btw108.
 43. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). Trimal: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)* **25**: 1972–1973. DOI: 10.1093/bioinformatics/btp348.
 44. Minh, B.Q., Schmidt, H.A., Chernomor, O., et al. (2020). Iq-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**: 1530–1534. DOI: 10.1093/molbev/msaa015.
 45. Letunic, I. and Bork, P. (2021). Interactive tree of life (itol) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**: w293–w296. DOI: 10.1093/nar/gkab301.
 46. Camargo, A.P., Nayfach, S., Chen, I.A., et al. (2023). Img/vr v4: An expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **51**: D733–d743. DOI: 10.1093/nar/gkac1037.
 47. Li, Z., Pan, D., Wei, G., et al. (2021). Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. *ISME J.* **15**: 2366–2378. DOI: 10.1038/s41396-021-00932-y.
 48. Hyatt, D., Chen, G.L., Locascio, P.F., et al. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**: 119. DOI: 10.1186/1471-2105-11-119.
 49. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using diamond. *Nat. Methods* **12**: 59–60. DOI: 10.1038/nmeth.3176.
 50. Bolduc, B., Jang, H.B., Doulcier, G., et al. (2017). Vcontact: An ivirus tool to classify double-stranded DNA viruses that infect archaea and bacteria. *PeerJ* **5**: e3243. DOI: 10.7717/peerj.3243.
 51. Shannon, P., Markiel, A., Ozier, O., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504. DOI: 10.1101/gr.1239303.
 52. Chen, T., Zhang, H., Liu, Y., et al. (2021). Evenn: Easy to create repeatable and editable venn diagrams and venn networks online. *J. Genet. Genomics* **48**: 863–866. DOI: 10.1016/j.jgg.2021.07.007.
 53. Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018). Metawrap-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**: 158. DOI: 10.1186/s40168-018-0541-1.
 54. Olm, M.R., Brown, C.T., Brooks, B., et al. (2017). Drep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME journal* **11**: 2864–2868. DOI: 10.1038/ismej.2017.126.
 55. Parks, D.H., Chuvochina, M., Waite, D.W., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**: 996–1004. DOI: 10.1038/nbt.4229.
 56. Ahlgren, N.A., Ren, J., Lu, Y.Y., et al. (2017). Alignment-free \$d_{2*}\$ oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* **45**: 39–53. DOI: 10.1093/nar/gkw1002.
 57. Laslett, D. and Canback, B. (2004). Aragorn, a program to detect trna genes and tmrna genes in nucleotide sequences. *Nucleic Acids Res.* **32**: 11–16. DOI: 10.1093/nar/gkh152.
 58. Coutinho, F.H., Silveira, C.B., Gregoracci, G.B., et al. (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* **8**: 15955. DOI: 10.1038/ncomms15955.
 59. Skennerton, C.T., Imelfort, M., and Tyson, G.W. (2013). Crass: Identification and reconstruction of crispr from unassembled metagenomic data. *Nucleic Acids Res.* **41**: e105. DOI: 10.1093/nar/gkt183.
 60. Emerson, J.B., Roux, S., Brum, J.R., et al. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**: 870–880. DOI: 10.1038/s41564-018-0190-y.
 61. Drula, E., Garron, M.L., Dogan, S., et al. (2022). The carbohydrate-active enzyme database: Functions and literature. *Nucleic Acids Res.* **50**: D571–d577. DOI: 10.1093/nar/gkab1045.
 62. Huerta-Cepas, J., Forslund, K., Coelho, L.P., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggno-mapper. *Mol. Biol. Evol.* **34**: 2115–2122. DOI: 10.1093/molbev/msx148.
 63. Kanehisa, M., Sato, Y., and Morishima, K. (2016). Blastkoala and ghostkoala: Kegg tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**: 726–731. DOI: 10.1016/j.jmb.2015.11.006.
 64. Graziotin, A.L., Koonin, E.V., and Kristensen, D.M. (2017). Prokaryotic virus orthologous groups (pvogs): A resource for comparative genomics and protein family annotation. *Nucleic Acids Res.* **45**: D491–d498. DOI: 10.1093/nar/gkw975.
 65. Chen, C., Chen, H., Zhang, Y., et al. (2020). Tltools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**: 1194–1202. DOI: 10.1016/j.molp.2020.06.009.

66. Cheng, R., Li, X., Jiang, L., et al. (2022). Virus diversity and interactions with hosts in deep-sea hydrothermal vents. *Microbiome* **10**: 235. DOI: 10.1186/s40168-022-01441-6.
67. Dixon, P. (2003). Vegan, a package of r functions for community ecology. *J. Veg. Sci.* **14**: 927–930. DOI: 10.1111/j.1654-1103.2003.tb02228.x.
68. Kato, S., Hirai, M., Ohkuma, M., et al. (2019). Microbial metabolisms in an abyssal ferromanganese crust from the takuyo-daigo seamount as revealed by metagenomics. *PLOS One* **14**: e0224888. DOI: 10.1371/journal.pone.0224888.
69. Dell'Anno, F., Rastelli, E., Tangherlini, M., et al. (2021). Highly contaminated marine sediments can host rare bacterial taxa potentially useful for bioremediation. *Front. Microbiol.* **12**: 584850. DOI: 10.3389/fmicb.2021.584850.
70. Gregory, A.C., Zayed, A.A., Conceição-Neto, N., 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.
71. Liang, Y., Wang, L., Wang, Z., et al. (2019). Metagenomic analysis of the diversity of DNA viruses in the surface and deep sea of the south china sea. *Front. Microbiol.* **10**: 1951. DOI: 10.3389/fmicb.2019.01951.
72. Meiaoxue, H., Shunan, C., Guangfu, L., et al. (2022). Distributions of virio- and picoplankton and their relationships with ice-melting and upwelling in the indian ocean sector of east antarctica. *Deep-Sea Res. II* **197**: 105044. DOI: 10.1016/j.dsr2.2022.105044.
73. Kato, S., Okumura, T., Uematsu, K., et al. (2018). Heterogeneity of microbial communities on deep-sea ferromanganese crusts in the takuyo-daigo seamount. *Microbes Environ.* **33**: 366–377. DOI: 10.1264/jsm2.ME18090.
74. Akira, U., Keisuke, N., Hisaaki, S., et al. (2017). Continuous growth of hydrogenetic ferromanganese crusts since 17myr ago on takuyo-daigo seamount, nw pacific, at water depths of 800–5500m. *Ore Geol. Rev.* **87**: 71–87. DOI: 10.1016/j.oregeorev.2016.09.032.
75. Kwan, Y.H., Mestre, N.C., Zhang, D., et al. (2023). Metal ecotoxicology: An essential component in environmental impact assessment of deep-sea mining. *The Innovation Geoscience* **1**: 100004. DOI: 10.59717/j.xinn-geo.2023.100004.
76. Zhang, W., Liu, Y., and Zhao, W. (2023). Occurrence and enrichment of cobalt in ferromanganese nodules from the western pacific. *Ore Geol. Rev.* **163**: 105758. DOI: 10.1016/j.oregeorev.2023.105758.
77. Liao, L., Xu, X.W., Jiang, X.W., et al. (2011). Microbial diversity in deep-sea sediment from the cobalt-rich crust deposit region in the pacific ocean. *FEMS Microbiol. Ecol.* **78**: 565–585. DOI: 10.1111/j.1574-6941.2011.01186.x.
78. Chen, W., Na, J., and Zhang, D. (2021). Description of three species of ophioplathacids, including a new species, from a deep seamount in the northwest pacific ocean. *PeerJ* **9**: e11566. DOI: 10.7717/peerj.11566.
79. Molodtsova, T.N., Opreko, D.M., and Wagner, D. (2022). Description of a new and widely distributed species of *bathypathes* (Cnidaria: Anthozoa: Antipatharia: Schizopathidae) previously misidentified as *bathypathes alternata* Brook, 1889. *Peer J.* **10**: e12638. DOI: 10.7717/peerj.12638.
80. Trubl, G., Jang, H.B., Roux, S., et al. (2018). Soil viruses are underexplored players in ecosystem carbon processing. *mSystems* **3**: 00076–18. DOI: 10.1128/mSystems.00076-18.
81. Paez-Espino, D., Elie-Fadrosch, E.A., Pavlopoulos, G.A., et al. (2016). Uncovering earth's virome. *Nature* **536**: 425–430. DOI: 10.1038/nature19094.
82. Roux, S., Hallam, S.J., Woyke, T., et al. (2015). Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *eLife* **4**: e08490. DOI: 10.7554/eLife.08490.
83. Kim, S., Kang, I., Lee, J.W., et al. (2021). Heme auxotrophy in abundant aquatic microbial lineages. *Proc. Natl. Acad. Sci. U.S.A.* **118**: e2102750118. DOI: 10.1073/pnas.2102750118.
84. Wiegand, S., Jogler, M., and Jogler, C. (2018). On the maverick planctomycetes. *FEMS Microbiol. Rev.* **42**: 739–760. DOI: 10.1093/femsre/fuy029.
85. Shao, Q., Sun, D., Fang, C., et al. (2022). Biodiversity and biogeography of abundant and rare microbial assemblages in the western subtropical pacific ocean. *Front. Microbiol.* **13**: 839562. DOI: 10.3389/fmicb.2022.839562.
86. Sun, Q., Song, J., Li, X., et al. (2020). Bacterial vertical and horizontal variability around a deep seamount in the tropical western pacific ocean. *Mar. Pollut. Bull.* **158**: 111419. DOI: 10.1016/j.marpolbul.2020.111419.
87. Ma, J., Song, J., Li, X., et al. (2019). Environmental characteristics in three seamount areas of the tropical western pacific ocean: Focusing on nutrients. *Mar. Pollut. Bull.* **143**: 163–174. DOI: 10.1016/j.marpolbul.2019.04.045.
88. Esposito, A., Tamburini, S., Triboli, L., et al. (2019). Insights into the genome structure of four acetogenic bacteria with specific reference to the wood-ljungdahl pathway. *MicrobiologyOpen* **8**: e938. DOI: 10.1002/mbo3.938.
89. Youssef, N.H., Farag, I.F., Rudy, S., et al. (2019). The wood-ljungdahl pathway as a key component of metabolic versatility in candidate phylum bipolaricaulota (acetothermia, op1). *Environ. Microbiol. Rep.* **11**: 538–547. DOI: 10.1111/1758-2229.12753.
90. Ragsdale, S.W. and Pierce, E. (2008). Acetogenesis and the wood-ljungdahl pathway of co(2) fixation. *Biochim. Biophys. Acta* **1784**: 1873–1898. DOI: 10.1016/j.bbapap.2008.08.012.
91. Nardella, C., Boi, D., di Salvo, M.L., et al. (2019). Isolation of a complex formed between acinetobacter baumannii hema and heml, key enzymes of tetrapyrroles biosynthesis. *Front. Mol. Biosci.* **6**: 6. DOI: 10.3389/fmolb.2019.00006.
92. Phadtare, S. (2004). Recent developments in bacterial cold-shock response. *Curr. Issues Mol. Biol.* **6**: 125–136. DOI: 10.1016/j.cimb.2004.03.001.
93. Jiang, W., Hou, Y., and Inouye, M. (1997). Cspa, the major cold-shock protein of *escherichia coli*, is an α chaperone. *J. Biol. Chem.* **272**: 196–202. DOI: 10.1074/jbc.272.1.196.
94. Behl, A., Kumar, V., Shevtsov, M., et al. (2020). Pleiotropic roles of cold shock proteins with special emphasis on unexplored cold shock protein member of plasmodium falciparum. *Malar. J.* **19**: 382. DOI: 10.1186/s12936-020-03448-6.
95. Giljan, G., Kamennaya, N.A., Otto, A., et al. (2020). Bacterioplankton reveal years-long retention of atlantic deep-ocean water by the tropic seamount. *Sci. Rep.* **10**: 4715. DOI: 10.1038/s41598-020-61417-0.
96. Liang, J., Feng, J.C., Zhang, S., et al. (2021). Role of deep-sea equipment in promoting the forefront of studies on life in extreme environments. *iScience* **24**: 103299. DOI: 10.1016/j.isci.2021.103299.
97. Feng, J.C., Liang, J., Cai, Y., et al. (2022). Deep-sea organisms research oriented by deep-sea technologies development. *Sci. Bull.* **67**: 1802–1816. DOI: 10.1016/j.scib.2022.07.016.
98. Dutt, S., Hamza, I., and Bartnikas, T.B. (2022). Molecular mechanisms of iron and heme metabolism. *Annu. Rev. Nutr.* **42**: 311–335. DOI: 10.1146/annurev-nutr-062320-112625.
99. Anzaldi, L.L. and Skaar, E.P. (2010). Overcoming the heme paradox: Heme toxicity and tolerance in bacterial pathogens. *Infect. Immun.* **78**: 4977–4989. DOI: 10.1128/iai.00613-10.
100. Dailey, H.A., Dailey, T.A., Gerdes, S., et al. (2017). Prokaryotic heme biosynthesis: Multiple pathways to a common essential product. *Microbiol. Mol. Biol.* **81**: 1128/mmbr.00048-16.
101. Brzezowski, P., Richter, A.S., and Grimm, B. (2015). Regulation and function of tetrapyrrole biosynthesis in plants and algae. *Biochim. Biophys. Acta* **1847**: 968–985. DOI: 10.1016/j.bbabi.2015.05.007.
102. Richard, K.L., Kelley, B.R., and Johnson, J.G. (2019). Heme uptake and utilization by gram-negative bacterial pathogens. *Front. Cell. Infect. Microbiol.* **9**: 81. DOI: 10.3389/fcimb.2019.00081.
103. Thompson, J.M., Jones, H.A., and Perry, R.D. (1999). Molecular characterization of the heme uptake locus (hmu) from *yersinia pestis* and analysis of hmu mutants for heme and hemoprotein utilization. *Infect. Immun.* **67**: 3879–3892. DOI: 10.1128/iai.67.8.3879-3892.1999.
104. Sañudo-Wilhelmy, S.A., Gómez-Consarnau, L., Suffridge, C., et al. (2014). The role of b vitamins in marine biogeochemistry. *Ann. Rev. Mar. Sci.* **6**: 339–367. DOI: 10.1146/annurev-marine-120710-100912.
105. Soto, M.A., Desai, D., Bannon, C., et al. (2023). Cobalamin producers and prokaryotic consumers in the northwest atlantic. *Environ. Microbiol.* DOI: 10.1111/1462-2920.16363.
106. Jin, X., Yang, Y., Cao, H., et al. (2022). Eco-phylogenetic analyses reveal divergent evolution of vitamin b(12) metabolism in the marine bacterial family 'psychromonadaceae'. *Environ. Microbiol. Rep.* **14**: 147–163. DOI: 10.1111/1758-2229.13036.
107. Nahvi, A., Barrick, J.E., and Breaker, R.R. (2004). Coenzyme b12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* **32**: 143–150. DOI: 10.1093/nar/gkh167.
108. Zhang, J.H., Yuan, H., Wang, X., et al. (2021). Crystal structure of the large subunit of cobaltochelatase from *mycobacterium tuberculosis*. *Proteins* **89**: 462–467. DOI: 10.1002/prot.26023.
109. Lundqvist, J., Elmlund, D., Heldt, D., et al. (2009). The AAA(+) motor complex of subunits CobS and CobT of cobaltochelatase visualized by single particle electron microscopy. *J. Struct. Biol.* **167**: 227–234. DOI: 10.1016/j.jsb.2009.06.013.
110. Warren, M.J., Raux, E., Schubert, H.L., et al. (2002). The biosynthesis of adenosylcobalamin (vitamin b12). *Nat. Prod. Rep.* **19**: 390–412. DOI: 10.1039/b108967f.
111. Chen, T., Chen, X., Zhang, S., et al. (2021). The genome sequence archive family: Toward explosive data growth and diverse data types. *Genom. Proteom. Bioinform.* **19**: 578–583. DOI: 10.1016/j.gpb.2021.08.001.
112. Database resources of the national genomics data center, china national center for bioinformatics in 2022. (2022). *Nucleic Acids Res.* **50**: D27–d38. DOI: 10.1093/nar/gkab951.

FUNDING AND ACKNOWLEDGMENTS

Grant support was provided by the high-performance server at the Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao, China), a high-performance computing cluster operated by the Institute of Evolution and Marine Biodiversity, and the high-performance servers of Marine Big Data Center of Institute for Advanced Ocean Study of the Ocean University of China. We thank the captain and crew of the cruises DY48, DY56, and DY61, for their assistance with the shipboard experiment. This study was supported by the Natural Science Foundation of China (42188102, 42120104006, 41976117, 42176111,

42176038), the Fundamental Research Funds for the Central Universities (202172002, 201812002, 201762017, 201562018), the Project of State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography (SOEDZZ2204), Science Foundation of Donghai Laboratory (DH-2022KF0211), and the Ocean Negative Carbon Emissions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Q.L., M.W., and Y.T.L. conceived and designed the experiment. Y.C. and C.G. methodology, formal analysis, writing, and original draft preparation. M.Y.L., and Y.C. DNA extraction and metagenomic dataset, Y.C. F.Y.G., and H.Y. participated in statistical analysis. Q.L., A.M., M.W., and Y.T.L. critically revised the manuscript. All authors revised and commented on the article and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AND CODE AVAILABILITY

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive ¹¹¹ in the National Genomics Data Center,¹¹² China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA012520), which is publicly accessible at <https://ngdc.cncb.ac.cn/gsa/s/89C84cZv>.

SUPPLEMENTAL INFORMATION

It can be found online at <https://doi.org/10.59717/j.xinn-geo.2024.100088>