Environmental Science & Technology

pubs.acs.org/est

Article

From Field Metagenomes to Mutant Genomes: Coevolution of Cyanophages and *Synechococcus* in Estuarine Ecosystems

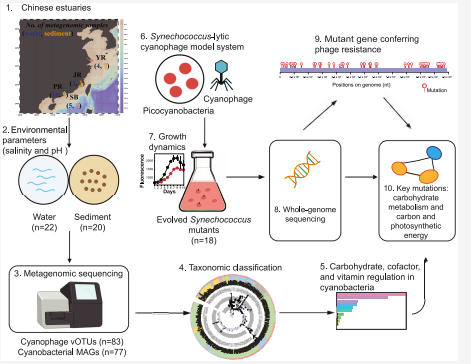
Yisha Tan, Dayu Zou, Can Ni, Qinglu Zeng,* and Meng Li*

Cite This: https://doi.org/10.1021/acs.est.5c12277

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Picocyanobacteria, represented by *Prochlorococcus* and *Synechococcus*, are major photosynthetic organisms in aquatic ecosystems, and their viruses (cyanophages) significantly impact cyanobacterial ecology and evolution. Here, we combined metagenomics of *Synechococcus* communities along four representative estuaries in China and whole-genome analyses of laboratory-evolved *Synechococcus* mutants to link viral diversity to host adaptation and evolution. We assembled 83 cyanophage genomes (mainly cyanomyoviruses), with expanded auxiliary metabolic genes encoding glycosyltransferases and radical S-adenosyl methionine proteins involved in amino acid and lipopolysaccharide metabolism. Metagenome-assembled cyanobacterial genomes revealed mutations predominantly in membrane-associated functions linked to phage infection. In parallel, we identified genetic pathways conferring phage resistance in 18 evolved *Synechococcus* mutant strains that are resistant to phage infection. Notably, mutations in carbohydrate (*rfbA*) and photosynthetic energy transfer (*cpeT*) of *Synechococcus* mutants recurred in both cultured isolates and recovered metagenomes. These results indicate that cyanophages in estuaries leverage broader metabolic toolkits, while *Synechococcus* repeatedly evolves resistance. Together, these findings outline a reciprocal adaptive landscape that helps explain the persistence and turnover of picocyanobacterial populations in estuarine environments.



KEYWORDS: *estuary, cyanophage, synechococcus, virus-host interactions, phage resistance*

Downloaded via HONG KONG UNIV SCIENCE TECHLGY on April 8, 2026 at 14:52:10 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

1. INTRODUCTION

The unicellular picocyanobacteria *Prochlorococcus* and *Synechococcus* are the smallest but the most abundant phototrophs on Earth.[1] They are responsible for approximately 25% of the oceanic net primary productivity, contributing greatly to the global carbon cycle.[2,3] *Prochlorococcus* is dominant in warm oligotrophic waters, whereas *Synechococcus* has a broader geographical distribution and is dominant in salt-wedge estuaries[4] and more temperate or mesotrophic open-ocean waters.[1,5,6] Cyanophages (viruses infecting cyanobacteria) are widely distributed in aquatic ecosystems,[2,7−10] lysing up to 40% of *Synechococcus* cells daily under certain conditions,[11] while more conservative estimates suggest that daily lysis rates typically range from 2% to 20% depending on environmental factors.[12,13] The high cyanophage activity observed in nutrient-rich estuarine environments highlights the important role of cyanophages in regulating cyanobacterial population abundance,[14,15] genetic diversity,[16] and metabolism and evolution,[17,18] although their regulatory effects on picocyanobacteria in estuaries remain less well constrained.

Cyanophages from the *Caudoviricetes* class belong to three morphological groups: *Myoviridae* (cyanomyovirus), *Podoviridae* (cyanopodovirus), and *Siphoviridae* (cyanosiphovirus).[19−24] Cyanomyoviruses are divided into T4-like or TIM5-like groups, with T4-like cyanomyoviruses being prevalent in marine ecosystems and having a broad host range.[24−26] Additionally, marine cyanophages carry numerous auxiliary metabolic genes (AMGs) that influence host photosynthesis, carbon metabolism, the pentose phosphate pathway (PPP), phosphorus utilization, and nucleic acid synthesis during infection.[27−32]

Experimental evolutional studies have been conducted to understand the coevolution of cyanophages and cyanobacteria,[14,21,23,33] such as the resistance of cyanobacteria to cooccurring cyanophages[16] and the cost of resistance to viral infection.[34] Regarding virus-host interactions, the "fluctuating selection" and "arms race" theories support the reciprocal evolution of host resistance and viral infectivity, as well as the antagonistic evolution of phage resistance in laboratory environments.[18,20,35,36] Comparative studies showed that a large number of evolutions in genes were associated with viral infection and classical restriction-modification systems among cyanobacterial genomes.[33,37,38] However, the genetic adapta-

Received: September 4, 2025
Revised: March 25, 2026
Accepted: March 26, 2026

© XXXX American Chemical Society

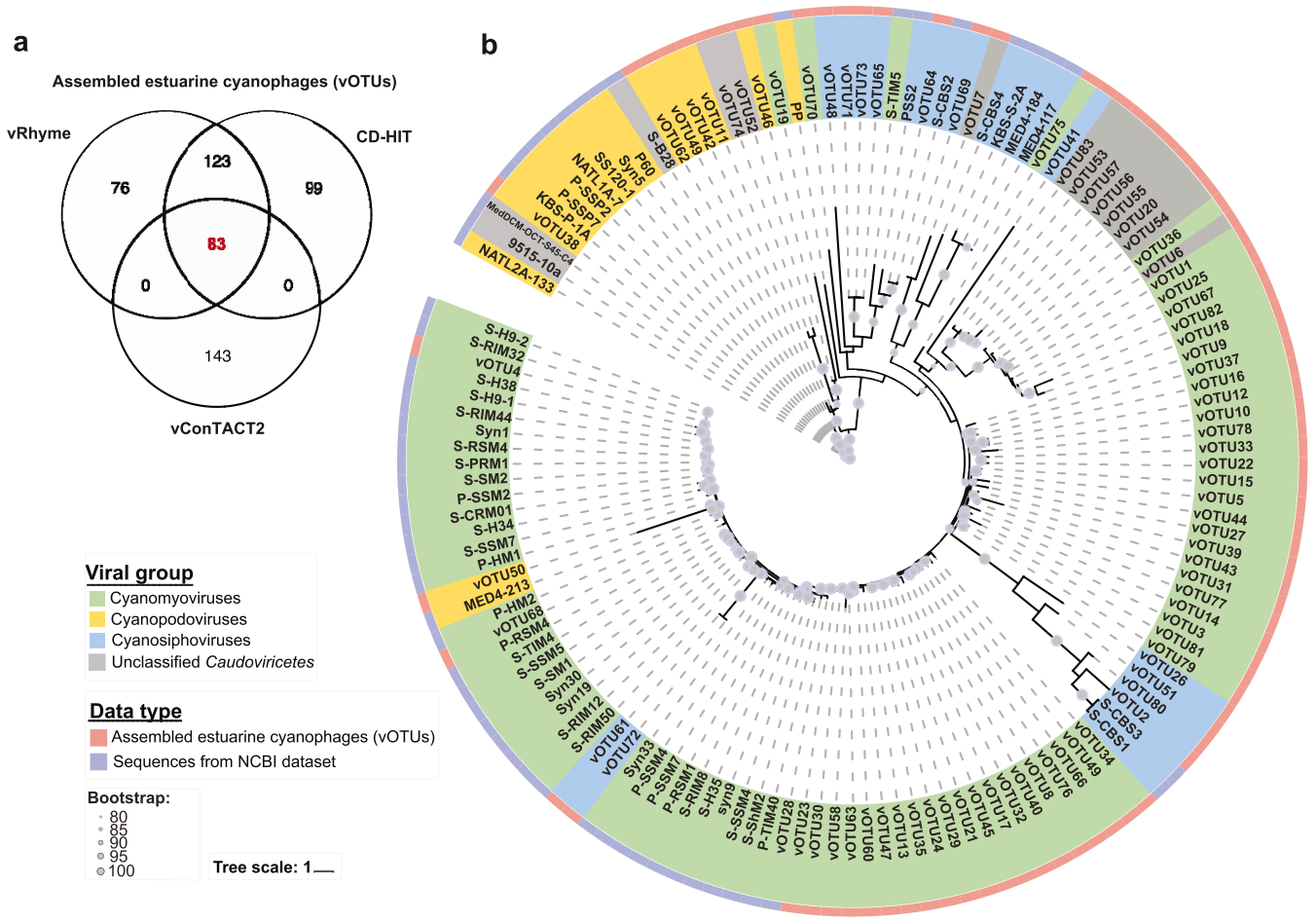


Figure 1. Evaluation of viral prediction approaches and phylogenetic analysis of cyanophages. (a) Predicted vOTUs by three methods. The Venn diagrams display the intertrench and intratrench occurrence of vOTUs. The numbers in the Venn indicate vOTUs were recovered by three different viral prediction approaches, including vRhyme, CD-HIT and vCONTACT. The plot shows shared cyanophage populations (n = 83) conducted by three methods. (b) The maximum-likelihood phylogenetic tree was generated by the concatenated amino acid sequences of the 5 ORFs with similar functions in cyanophage genomes. These data include assembled estuarine cyanophages in this study and reference cyanophages in previous studies. These 5 ORFs are DNA polymerase, DNA helicase, terminase large subunit, major capsid protein and structural components of the phage head and tail. It was estimated by IQ-TREE under best-fit model LG + F + R7. Nodes with bootstrap support values greater than 80 are marked with gray circles. The distinct lineages of the cyanophages are assigned different branch and leaf colors according to their taxonomic affiliations, as indicated in the legend. Cyanomyoviruses, cyanopodoviruses, cyanosiphoviruses and unclassified cyanophages are shown in green, yellow, blue and light gray colored background, respectively. The colors of outer ring indicate different data types, suggesting data associated with assembled estuarine cyanophages in this study (red) and sequenced reference cyanophages in the NCBI data set (purple).

tion mechanisms of cyanophages and cyanobacteria in estuaries remain mysterious.

Estuaries are partially enclosed coastal water bodies, where freshwater from rivers and streams intermingle with salt water from the ocean, forming a transition zone between freshwater and marine environments. Estuarine ecosystems, among the most productive ecosystems on Earth, possess and deliver organic matter and nutrients to both the water column and sediment.[39] Early estuarine work isolated picocyanobacteria and their cyanophages,[40−42] produced multiple estuarine cyanopodovirus and cyanosiphovirus genomes,[43,44] and enabled biogeographic comparisons showing virus-host correspondence and spatial co-occurrence between picocyanobacterial and cyanophage communities.[45] Estuaries contain unique genotypes of cyanobacteria[41,42] which can better tolerate changes in temperature, salinity, and heavy metals compared to their coastal and open-ocean counterparts.[46] Certain cyanopodovirus are enriched in estuaries but rare offshore, consistent with host distributions.[44] Inheritance of specific functional genes has been implicated in cyanophage evolutionary fitness,[44] revealing their ability to adapt to fluctuating selection pressures.[47] However, much of this evidence stems from marker-gene surveys that capture only subsets of cyanophage diversity, and few studies have directly connected viral genetic diversity to host adaptation and evolution in estuarine ecosystems.

To investigate the interactions between cyanophages and *Synechococcus* in estuaries, we coupled metagenomic analyses with whole-genome analyses of laboratory-evolved *Synechococcus* mutants resistant to phage infection. While pioneering works on cyanophage-cyanobacteria interactions have often focused on marine ecosystems and culture-based infection dynamics, we hypothesized that estuarine physicochemical gradients, particularly salinity, reshape virus-host encounters and favor selection on specific host resistance targets. Here we examined whether estuarine gradients are associated with cyanophage-*Synechococcus* co-occurrence patterns and whether genomic targets associated with phage resistance are detectable

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

in natural cyanobacterial populations. From the metagenomes, we reconstructed *Synechococcus* metagenome-assembled genomes (MAGs) and their cyanophage genomes, enabling community-scale inferences about viral taxonomy, distribution, and auxiliary metabolic genes (AMGs) under fluctuating estuarine conditions. We isolated and systematically evolved phage-resistant *Synechococcus* mutants under estuarine-like salinity regimes to ensure comparability with field conditions. By comparing genetic variations of these mutants with those of the wild-type genomes, we uncovered resistance pathways and identified recurrent *Synechococcus* mutations that also appear in metagenomes. Together, these results provide mechanistic evidence for the reciprocal adaptation between estuary cyanophages and *Synechococcus*.

2. MATERIALS AND METHODS

2.1. Environmental Sampling Area, DNA Extraction, and Metagenomic Sequencing

The field studies encompassing four representative estuaries along the coast of China were conducted (Figure S1a and Table S1): Pearl River Estuary (PR, July 2015), Shenzhen Bay (SB, July 2016), Yangtze River Estuary (YR, August 2016), and Jiulong River Estuary (JR, December 2018). The 42 metagenomic samples were derived from PR, SB, YR, and JR, covering water depths from 0 to 20 m and including water and sediment samples (Figure S1b and Table S1). Major chemical parameters and concentrations of dissolved nutrients, including nitrite, nitrate, ammonium, phosphate, silicate, pH, salinity, as well as dissolved oxygen (DO), and Chlorophyll (Chl) a were determined according to standard methods[48] and previously reported[49,50] (Table S1). The samples were sequentially filtered through 3-μm and 0.22-μm pore size filters for purification. DNA extraction was performed with a PowerMax DNA Isolation Kit (QIAGEN, Germany). For metagenomics, high-quality DNA libraries were sequenced on the Hiseq 2000 platform (Illumina) to yield 150 bp pair-end reads.

2.2. Viral Contig Identification

We employed a comprehensive workflow to identify and characterize estuarine cyanophage genomes from metagenomic assemblies (Figure S2). Initially, metagenomic raw reads from each sample were quality-trimmed using Sickle.[51] Viral contigs were then identified from trimmed metagenomic reads with default settings using VirSorter2,[52] retaining only contigs $\geq$ 1 kb with a minimum score of 0.5, resulting in 5753 viral contigs. To enhance confidence, we further analyzed these contigs with DeepVirFinder (v1.0),[53] applying stringent thresholds of scores $\geq$ 0.9, $P < 0.05$, and length $\geq$ 5 kb, yielding 4,525 identified viral contigs. CheckV[54] was used to estimate genome completeness and identify provirus-like contigs that include host-derived flanking regions. We retained only contigs classified by CheckV as medium-quality and high-quality and excluded any contigs labeled by CheckV as "provirus," which validated 3,958 viral contigs. Subsequently, viral contigs $\geq$ 10 kb that passed viral quality filters were first clustered within each sample at an average nucleotide identity $\geq$ 95% and alignment coverage $\geq$ 85% using CD-HIT[55] to define nonredundant vOTUs, resulting in preliminary 2,421 viral operational taxonomic units (vOTUs).

To identify cyanophage vOTUs, we employed three complementary methods (Figure S2). First, we compared the 2,421 viral contigs against the NCBI Viral RefSeq data set (v94) and a constructed reference cyanophage database (Table S2) using vConTACT2,[56] a network-based tool that classifies viruses into genus-level clusters based on protein similarity. This method identified 226 vOTUs belonging to cyanophage clusters. Second, vRhyme (v1.1.0)[57] was applied as a viral binner to aggregate fragmented contigs from the same viral population prior to vOTU definition, clustering contigs based on covariation in coverage profiles across samples and *k*-mer composition. This step helped aggregate fragmented contigs from the same underlying viral population prior to defining vOTUs, resulting in 203 vOTUs. Third, after combining candidate cyanophage contigs across all estuaries, we dereplicated the nucleotide sequences of vOTUs, using CD-HIT clustering at the same thresholds. This step helped remove redundancy between sites and obtain a nonredundant set of vOTUs used for downstream analyses, yielding 309 vOTUs. To ensure high confidence in our cyanophage vOTU identification, we considered only the intersection of results from these three methods, culminating in 83 high-confidence vOTUs representing assembled estuary cyanophage genomes (Figure 1a and Table S3). These genomes were subsequently used for downstream analyses, including assessments of viral diversity, host interactions, and functional genomic features.

2.3. Viral Taxonomic Assignment and Viral-Host Linkage Analyses

Open reading frames (ORFs) for each vOTU were predicted using Prodigal (version 2.6.3)[58] in meta mode. Then, predicted ORFs were annotated (Table S4) by searching against multiple databases, including KEGG,[59] eggNOG-Mapper,[60] Pfam,[61] and IMG/VR v3.[62] We constructed a reference cyanophage database (Table S2) to assign vOTUs to taxonomic groups by using vConTACT2.[56] To ensure host predictions for the vOTUs, we integrated multiple classification methods, including BLASTn, CRISPR (clustered regularly interspaced short palindromic repeats), matching, and tRNA (tRNA) matching. First matching viral genomes against the NCBI Viral RefSeq database (v94),[63] and nucleotide sequences of vOTUs were compared to bacterial MAGs using BLASTn, based on matches meeting the thresholds of a bit score $\geq$ 50, *e*-values $\leq 10^{-3}$, and $\geq$ 70% ANI across a minimum of 2,500 bp. Second, CRISPR spacers were identified in bacterial MAGs using the CRISPR Recognition Tool (CRT)[64] and matched to phage contigs using BLASTn, with host predictions requiring $\geq$ 95% identity and $\leq$ 2 single nucleotide polymorphisms were selected to serve as putative phage hosts. However, there was no result in cyanobacterial MAGs in this study when researching CRISPR spacers. Third, tRNA genes in phage contigs were predicted by tRNAscan-SE (version 2.0.12)[65] with default settings and blasted against the bacterial sequences using the BLASTn, retaining only the best hits with at least 95% sequence identity. The best hit by the highest bit-score was kept for each contig. Host lineages for assembled estuarine cyanophage genomes (Table S5) were inferred using a combination of results from these methods, with manually curated additional data from the Virus-Host Database (accessed 10th October 2023, https://www.genome.jp/virushostdb), allowing for a comprehensive analysis. The Virus-Host Database[66] offers experimentally validated host−virus pairs, which provide an independent validation of bioinformatics-based predictions. These approaches captured taxonomic signals that might be missed by vConTACT2, particularly for genomes with low-protein-level similarity to known viruses. The taxonomy details were based on International Committee on Taxonomy of Viruses (ICTV Release, October 2020) and further updated with the latest virus lineage information (Table S3).

2.4. Calculating vOTU Relative Abundances

To determine the relative abundances of vOTUs in each estuary sample, we calculated reads per kilobase per million mapped reads (RPKM) values (Table S7). Reads from each estuarine metagenomic sample were mapped to the raw metagenomic data using CoverM[67] at minimum percent identity of 95% and percent aligned bases of 75%. To explore the presence of vOTUs in other aquatic systems, we recruited reads from the Tara Oceans data set by using tBLASTx with an *e*-value of 10^{-5}. In cases where a read aligned to multiple regions of the genome, it was assigned to the region with the lowest *e*-value. To account for potential biases introduced by genome size, read counts were normalized based on the number of predicted ORFs for each vOTU.

2.5. Metagenomic Assembly and Genome Binning

Metagenomic raw reads from each sample were trimmed using Sickle,[51] removing reads with an average Phred score <20 or length <50 bp after trimming. Trimmed reads were assembled on a per-

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

sample basis using IDBA-UD (v1.1.1)[68] with fixed multiple *k*-mer lengths, increasing from 65 to 145 bp in steps of 10. This iterative approach optimizes assembly across a range of *k*-mer sizes to better capture genomic features. All of the generated contigs with sequence lengths of >2500 bp were binned using MetaBAT2 (v2.12.1)[69] to ensure high-confidence assembly and to exclude shorter, low-complexity sequences. The generated bins were refined using Das Tool (v1.0)[70] to obtain the MAGs, which integrate binning results from multiple algorithms to reduce contamination and improve completeness. The completeness and contamination of MAGs were determined by CheckM (v1.1.3).[71]

2.6. Taxonomic Profiling of Microbial Communities

The cyanobacterial composition of each estuarine sample was assessed by recruiting quality-controlled metagenomic reads to cyanobacterial full-length 16S rRNA gene sequences from SILVA (release 138) treated as reference contigs and calculating per-reference RPKM using CoverM (contig mode). The relative abundances of cyanobacterial community in each sample were calculated as RPKM to normalize the variations in gene length and data set size (Table S8). The taxonomic classification of the MAGs was performed using GTDB-Tk[72] (Table S9). ORFs of each MAG were translated into amino acid sequences using Prodigal (v2.6.3)[58] with meta mode. The amino acid sequences were aligned against the nonredundant protein database using BLASTp (*e*-values $<10^{-5}$) and assigned to existing clusters of orthologous genes by eggNOG-Mapper[60] with *e*-values $<10^{-3}$ (Figure S2).

2.7. Strains and Culturing Conditions

Synechococcus WH7803 strain was maintained in the Port Shelter (Hong Kong) seawater-based Pro99[73] at the continuous light level of 10−20 μmol photons m^{-2} s^{-1} with the temperature at 23 °C in the incubator. Growth of WH7803 cultures in liquid medium was followed using in vivo chlorophyll *a* autofluorescence values with a 1420-Multilabel Counter Victor3 plate reader (PerkinElmer) or a 10-AU Field Fluorometer (Turner Designs). The cell abundance of WH7803 was determined using a Flow Cytometer (BD AccuriC6TM Plus).

2.8. Cyanophage Isolates and Purification

The 8 cyanophage isolates used in this study were isolated from the Yongxing Harbor (16°50′14.14″ N, 112°19′34.30″ E) (Table S6). These cyanophages were isolated by individual plaques on overlays of host *Synechococcus* WH7803 in agar plates. The plaque morphologies of these cyanophages are similar, forming turbid plaques. Then well-isolated plaques were picked from the agar plates and dissolved in Pro99 medium. The infection was repeated three times to purify these cyanophages. These isolated cyanophages were named with the first character "S" standing for *Synechococcus*, "SCS" for South China Sea, and "M" for cyanomyovirus and number for phage phenotypes (listed in Table S6). For long-term storage, viral stocks were stored in tubes at 4 °C in the dark.

2.9. Phylogenetic Analyses

There is no universal marker gene (analogous to the 16S rRNA gene in microbes) that can provide taxonomic information for viruses. To unveil the phylogeny of assembled estuarine cyanophages in this study, we conducted phylogenetic analyses of cyanophages by comparing two encoded key components of DNA packaging machine (terminase large and small subunits), two encoded proteins involved in DNA metabolism and replication (DNA polymerase and DNA primase/helicase), and one encoded phage structural proteins (major capsid protein) with those from reference cyanophages in the NCBI database (Table S2). The concatenated protein sequences were aligned across these cyanophage genomes using MUSCLE (v7.245). The conserved positions in the alignments were selected with the default parameters of trimAl.[74] We constructed maximum-likelihood trees using IQTREE[75] under best-fit model LG + F + R7 and visualized them in iTOL[76] (Figure 1b).

2.10. Identification of Viral Auxiliary Metabolic Gene

We confirmed the presence of auxiliary metabolic genes (AMGs) through a refined analysis that utilized CheckV[54] and VIBRANT[77] to examine the integrity of the viral genome and the location of these AMGs relative to nearby viral genes. The AMGs were then grouped into KEGG orthologs (KO)[78] using KEGG Automatic Annotation Server (KAAS) for further comparisons of their diversities and compositions between different sites (YR, JR, PR, and SB), sample types (water and sediment), and origins of cyanophages (estuarine and marine environments) (Tables S10 and S11). We used the RPKM of a cyanophage genome that encoded an AMG to proxy the abundance of the AMG. Comparisons of the AMG diversities were performed at the KO level based on the KEGG database, and we calculated the frequencies of viruses that encoded AMGs belonging to specific functional pathways in different origins of cyanophages (estuarine cyanophages in this study and reference cyanophages in global databases).

2.11. Laboratory-Evolution Experiment: Selection of Phage-Resistant Mutants

Synechococcus WH7803 strains were plated on agarose plates and challenged individually with each cyanophage isolate by using 1% Invitrogen Ultra Pure Low Melting Point (LMP) agarose dissolved in Pro99 medium.[79] Each experimental group was a mixture of 1 mL of WH7803 culture (~10^8 cells ml^{-1}) and 1 mL of cyanophage lysate (~10^8 cyanophages ml^{-1}) whereas controls were in the absence of cyanophage. The parallel phage-free control groups were included to exclude nonspecific growth variations in the downstream whole-genome sequencing analyses. Replicate plates (n = 6) were used for each experimental group, with matched phage-free control plates processed in parallel. After 1 h of incubation, agarose Pro99 medium was added to the plate. Replicate plates were used for each phage-host strain combination. The plates were incubated at 23 °C under continuous light, with the colony number visually monitored. After one-three months, control colonies and potential phage-resistant colonies were visible, and individual colonies were removed from the agar by using sterile pipet tips and transferred into the liquid Pro99 medium. To confirm that the individual colonies were phage resistant, each one was tested for resistance to the cyanophage used in the initial selection. If the colony was able to grow in the presence of the cyanophage, then it was defined as phage-resistant mutants. These phage-resistant mutants were then used for further analyses. The growth dynamics of these phage-resistant mutants were measured by chlorophyll *a* fluorescence value, and maintained in the presence of initially selected cyanophages.

2.12. Genomic DNA Extraction, Whole-Genome Sequencing, and Genetic Characterization

Genomic DNA from both wild-type and resistant mutants was extracted using the TIANamp Bacteria DNA Extraction Kit (Tiangen), and the concentrations were determined by NanoDrop. Genomic DNA was sheared into small fragments of around 300 bp in length using the Covaris S220 Focused-ultrasonicator. DNA sequencing libraries were prepared using the NEBNext Ultra II DNA library prep kit, checked by fragment analyses, and sequenced using the HiSeq 4000 platform (Illumina). The sequencing data generated from each library contains approximately 2 Gb of 150 bp paired-end reads.

Sequences were filtered and trimmed to remove adapters and low-quality reads with Trimmomatic (v0.39).[80] All of the trimmed data were examined using MultiQC (v1.10.1).[81] The merged paired reads were aligned to reference genome WH7803 (GenBank accessions: NC_005070) with default parameters using Bowtie2.[82] Then, duplicate reads were marked by using PICARD (version 1.94). Mutation calling was conducted using Genome Analysis Toolkit Best Practices (GATK, v4.2.0.0).[83] Mutants and genotypes calling were performed by GATK HaplotypeCaller. The results were filtered by quality score recalibration (GATK BQSR module) to minimize sequencing errors and recalibrate the base qualities. We established Gaussian models and selected optimized tranche sensitivity to

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

proceed with mutation score recalibration (GATK VQSR module, truth-sensitivity filter level 90.0) to evaluate log odds ratios of false-positive mutants. Vcftools (v0.1.16)[84] was used to summarize statistics of variant call format (VCF) files.

We obtained VCF files from sequenced mutants (MT) and wild-type strains (WT) by applying the above mutation calling methods. After obtaining these quality VCF files, we took intersections for each MT and WT separately. The intersection of WT and MT is considered to be the common mutation site among them. The specific mutation sites, which are the complementary parts of MT, were determined. Filtering out common mutation sites among WT and MT, the remaining high-confidence specific mutation positions from mutants can be acquired.

2.13. Mutation and Functional Annotation

Mutation site annotations were counted and evaluated using SnpEff.[85] All of the mutation sites, including single nucleotide polymorphism (SNPs), insertion, or deletion of nucleotides into genomic DNA (INDELs) were defined. For metagenomes, we classified all of the SNPs as intergenic, intragenic (mutation sites located in the one kilobase upstream and downstream regions of the nearest gene), synonymous, or nonsynonymous as previously described.[86] Functional annotations of genes were assigned using HMMsearch[87] against HMM profiles generated from cyanophage vOTUs and isolated cyanophage genomes (*e*-values $\leq 10^{-3}$ and a bit score >10). We further complemented with information from operon clusters from BioCyc databases (https://biocyc.org), KEGG[59] and Cluster of Orthologous Groups (COG)[88] by eggNOG-Mapper,[60] and UniProt databases[89] via congruence in functional assignment. To better understand the variation in phage−bacterium interactions, DefenseFinder[90] was used to identify antiphage defense systems in bacteria and executed with default parameters. For isolated mutants, we estimated the relationships between identified SNPs in each mutant substrain and SNPs in other mutants. The key SNPs presented pertinent associations in regions based on correlations of genotypes.[91] Enrichment for mutant genes was calculated by calculating the hypergeometric probability for finding this number of genes in cyanobacterial genomes at random according to the length of the mutant genes in comparison to the rest of the genome.

2.14. Statistical Analyses

All statistical analyses were conducted in R version 4.2.2 (2022−10−31). Differences of GC contents and genome sizes among assembled estuarine cyanophages were accessed using two-tailed Student's *t*-tests. The Spearman correlation analysis was performed based on genome size and GC content of estuarine cyanophages, showing they were significantly negatively correlated in this study (Spearman correlation coefficient = −0.22, $p < 0.05$). For community composition analyses, we utilized the abundance of 83 vOTUs to represent cyanophage communities and 16S rRNA gene abundances of *Synechococcus* as proxies for cyanobacterial communities (Tables S7 and S8). Nonmetric multidimensional scaling (NMDS) analysis based on Bray−Curtis dissimilarity was used to cluster the vOTUs and their abundance profiles associated with different sites and different sample types. Principal coordinates analysis (PCoA) based on the Bray−Curtis distance was calculated to compare the differences in the *Synechococcus* community among sampling sites and sample types.

To relate microbial community variation to environmental gradients, Mantel tests were conducted using Spearman correlations between Bray−Curtis community dissimilarity matrices (cyanophages and *Synechococcus*) and Euclidean distance matrices of environmental variables. Environmental variables were centered and scaled prior to the distance calculations. To quantify the multivariate explanatory power of environmental drivers beyond Mantel associations, we additionally performed PERMANOVA (adonis2) to estimate R^2 contributions of salinity, pH, nutrients, dissolved oxygen, and spatial covariates, and distance-based redundancy analysis (dbRDA) to report adjusted R^2 and term significance. In addition, to evaluate potential nonlinear relationships and rank predictor importance, random forest models were used to evaluate nonlinear predictor importance of environmental drivers for community ordination scores, and variable importance was summarized using permutation-based importance.

To test links between environmental conditions and the cyanophage functional potential, AMG abundance profiles were summarized per sample at the KEGG pathway level (Table S10). Spearman correlations between environmental variables and AMG pathway abundances were then computed, and p-values were adjusted using the Benjamini-Hochberg false discovery rate procedure.

3. RESULTS

3.1. Assembled Cyanophage Genomes were Recovered from Metagenomic Samples across Estuaries

Metagenomic samples were collected from four major Chinese estuaries: Shenzhen Bay (SB), the Pearl River Estuary (PR), the Jiulong River Estuary (JR), and the Yangtze River Estuary (YR) during the summer and early winter seasons (Figure S1a and Table S1). In total, there were 42 samples, including 22 water and 20 sediment samples (Figure S1b and Table S1). From these samples, we recovered 3,958 viral contigs, which were clustered into virus operational taxonomic units (vOTUs), including 83 vOTUs of cyanophages with low contamination (by CheckV[54]) (Figure 1a). We used a database of reference cyanophage genomes (Table S2) to determine vOTUs by high-confidence completeness estimation methods (Figure S2). These 83 vOTUs (Table S3), represented the assembled estuarine cyanophage genomes, were of relatively high- or medium-quality (CheckV[54] estimated completeness ≥ 50% and contamination <1%), which allowed more reliable taxonomic, functional, and host assignment and fragment recruitment. The similarities of vOTUs were investigated by calculating the average nucleotide identity (ANI) among each pair of vOTUs. Overall, the ANI of every two vOTUs ranged from 30.20% to 99.99% (Figure S3), whose genome sizes ranged from 102 to 157 kb (Figure S4). Genomic comparisons revealed that more than 40% of the 12,805 open reading frames (ORFs) in the vOTUs do not share homologous genes with published reference cyanophage genomes (Table S4), suggesting undiscovered diversity within assembled estuarine cyanophages.

3.2. Cyanophage Communities Exhibit Taxonomic and Genetic Diversity in Estuarine Ecosystems

To further determine the estuarine cyanophage compositions, we carried out taxonomic classifications among them (Materials and Methods section). Specifically, 7 vOTUs were cyanopodoviruses, 13 vOTUs were cyanosiphoviruses, 52 vOTUs were cyanomyoviruses, and 11 vOTUs were unclassified cyanophages (Table S3). These vOTUs contained many genes related to DNA replication and recombination, nucleotide metabolism, virion formation (capsid assembly and tail attachment), and other fundamental processes of lytic infection (Table S5).

We performed phylogenetic analyses based on the concatenated five marker genes (Figure 1b) among cyanophages (Materials and Methods section and Table S5). Of the five marker genes, two encoded proteins are involved in DNA metabolism and replication (DNA polymerase and DNA primase/helicase), two encoded virion structural proteins (major capsid protein and tail tube proteins), and one encoded protein is involved in DNA packaging (terminase large/small subunits) (Figure S5). The phylogenetic analysis showed that some vOTUs, such as assembled cyanosiphoviruses, were distinct from the reference genomes and formed separate branches (Figure 1b). This finding aligned with the clustering

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

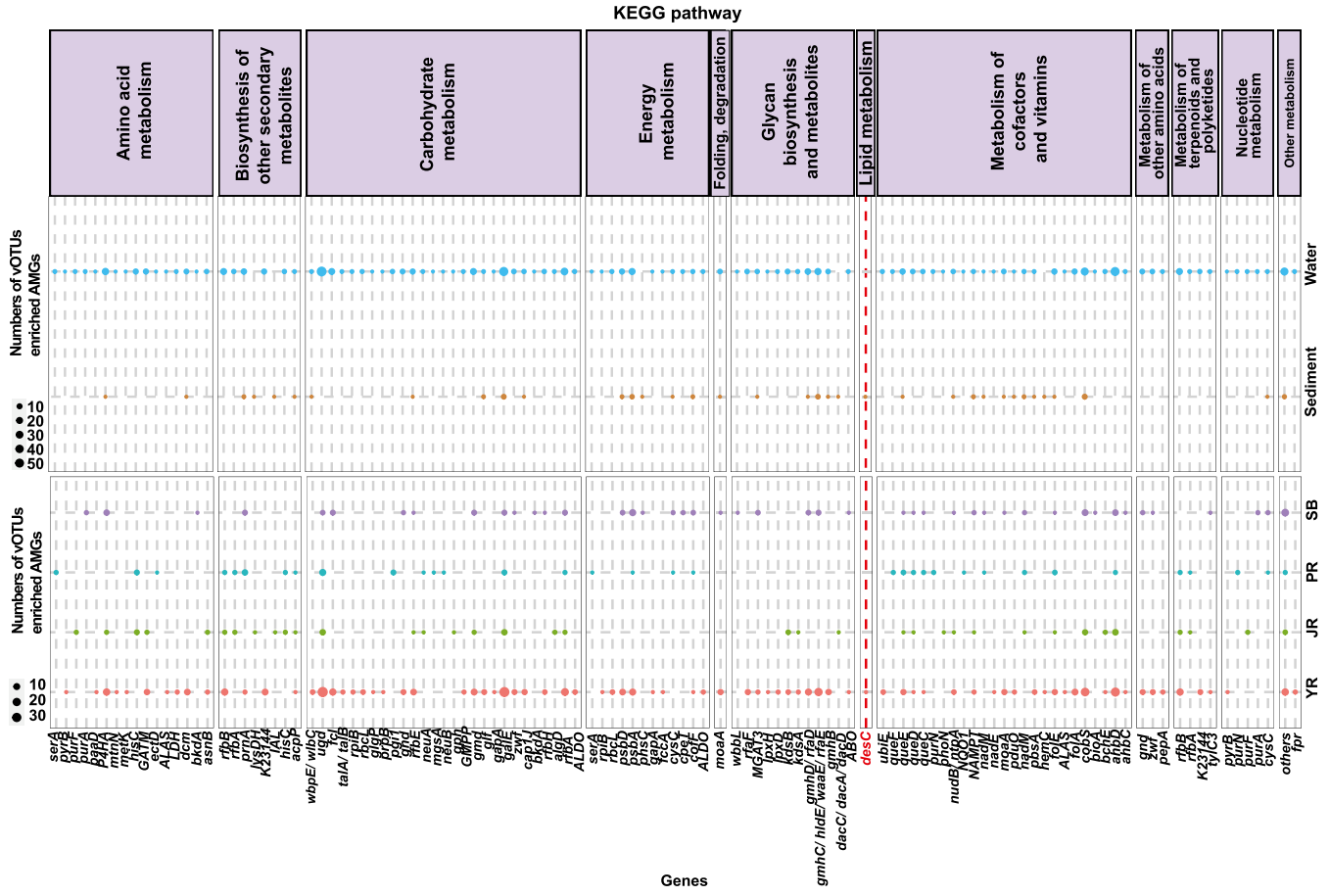


Figure 2. Compositions of virus-encoded auxiliary metabolic genes (AMGs) in the estuarine environment. Bubble plots (the upper layer) show the number of vOTUs enriched AMGs (*x*-axis) across different types of samples (*y*-axis), including sediment and water samples. The enriched AMGs were determined based on the total normalized coverage of cyanophage genomes harboring the corresponding AMGs. Bubble plots (the lower layer) display the number of vOTUs enriched AMGs (*x*-axis) across four estuarine environments (*y*-axis), including the Yangtze River Estuary (YR), the Jiulong River Estuary (JR), the Pearl River Estuary (PR), and Shenzhen Bay (SB). The purple blocks at each right side of bubble plots indicate the different metabolism pathways of each AMG based on KEGG database. The AMGs marked with red color are unique in estuarine cyanophages but not among marine cyanophages (shown in Figure S11) in this study.

of vOTUs based on the ANI (Figure S3). As phages with genomic identity above 95% are regarded as the same population,[92−94] the vOTUs were grouped into several clusters (an ANI below 90%), indicating their genetic novelty (Figure S3). Phylogenetic trees based on each of five genes also supported the separation of the assembled estuarine cyanophages from references (Figure S6). Most of the vOTUs belonging to cyanomyoviruses clustered separately from the vOTUs belonging to cyanopodoviruses (Figure 1b). Some vOTUs, such as vOTU2, vOTU26, vOTU51, and vOTU80, were defined as novel estuarine cyanosiphoviruses since they formed a monophyletic clade with the marine cyanosiphoviruses S-CBS3 and S-CBS1 (ANI < 90%) (Figure 1b).

The estuarine cyanophages showed varied distribution patterns in the four estuaries, with cyanomyoviruses being the dominant cyanophage family (Figure S1b). Cyanophages of SB had the highest total reads per kilobase per million mapped reads (RPKM) values (over 4200), nearly four times greater than those of the other sites (up to 1000) (Figure S7 and Table S7). The abundances of cyanosiphoviruses, such as P-SS2, KBS-S-2A, S-CBS4, and S-ESS1, were the highest in SB compared with other sites (Figure S8a). Specifically, cyanophage compositions also varied among the water and sediment samples, with RPKM values being greater in water than in sediment at most sites (Figure S7 and Table S7). In water, cyanomyoviruses (cyanophage Bellamy and S-SSM7) and unclassified viruses were highly abundant in the YR (Figure S8a and Table S7). Additionally, Syn1 was the most abundant genus in JR and PR and carried numerous domains of unknown function (Figure S8a and Table S7). PR and SB showed significantly different cyanopodovirus and cyanosiphovirus compositions relative to JR (Figure S8a and Table S7). The RPKM ratio of cyanosiphovirus: cyanopodoviruse: cyanomyovirus was roughly 5:7:35 in estuaries (Table S7). Nonmetric multidimensional scaling (NMDS) showed clear site-associated structuring of cyanophage communities (stress = 0.143), with YR separating from the other estuaries, whereas JR and PR were more similar to SB (Figure S8b). To formally test these patterns, we performed PERMANOVA, which indicated significant effects of both site and sample type (Site: R^2 = 0.2295, p = 0.001; Type: R^2 = 0.0986, p = 0.001; 999 permutations).

3.3. Prevalence of *Synechococcus* and Estuarine Cyanophages Show Distinct Spatial Abundance Patterns

Among the microbial taxonomic groups inferred from 16S rRNA gene read mapping, *Proteobacteria*, *Euryarchaea*, *Actino-*

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

bacteria, and *Cyanobacteria* were prevalent in all of the estuaries (Figure S9a). We observed that the normalized read abundance (RPKM) order of cyanobacteria in the four estuaries was consistent with that of cyanophages in the same estuaries (Tables S7 and S8). In particular, SB had the highest total cyanobacterial abundance (approximately 34,000 RPKM), nearly twice as high as that in the estuary with the second highest abundance (PR, approximately 17,000 RPKM). SB and PR are located in the Pearl River and flow into the South China Sea, where *Synechococcus* are highly abundant in estuarine water samples compared to sediment samples (Figure S9b).

To identify and track cyanobacterial communities across samples, a total of 77 cyanobacterial metagenome-assembled genomes (MAGs) were recovered from these estuarine sites (Table S9). Taxonomic assignment based on ANI indicated that these MAGs were most closely related to reference genomes from marine and coastal *Synechococcus* lineages (e.g., WH7803-, RCC307-, and CB0205-like genomes) (Table S9). A small number of *Planktothricoides* MAGs likely originated from filamentous cyanobacteria present in low abundance in the water column. Among them, *Synechococcus* were the dominant genus, representing 63 out of 83 predicted hosts of assembled estuarine cyanophages (as listed in Table S5). Based on the Virus-Host Database (accessed 10th October 2023, https://www.genome.jp/virushostdb), we inferred host lineages for vOTUs (Table S5). As shown in Table S7, cyanosiphovirus KBS-S-2A and cyanopodovirus KBS-P-1A were abundant in the YR samples. They can infect *Synechococcus*, suggesting that particular bacterial strains can remain dominant for extended periods despite high viral abundances.[95] Notably, 12 vOTUs were predicted to be cyanophages infecting WH7803 (Table S5), which exhibited greater abundance in water than in sediment from the estuarine environment.

Consistent with the cyanophage community ordination (Figure S8b), the PCoA of the *Synechococcus* community composition revealed an obvious site effect across the four estuaries (PCoA1−2 explained 53.8% of the variance), with the YR samples forming a distinct cluster relative to the other sites (Figure S8c). In contrast, *Synechococcus* communities from water and sediment overlapped substantially (PCoA1−2 explained 50.8% of the variance), suggesting that sampling site is a stronger determinant of *Synechococcus* community structure than sample type in this data set (Figure S8c). To explore how environmental factors shape cyanophage and *Synechococcus* community composition features, we performed Mantel tests to connect physicochemical parameters with cyanophage and *Synechococcus* community features in four estuaries in the summer or winter of 2015, 2016, and 2018 (Figure S9c and Tables S1, S7, and S8). The estuarine cyanophage community was correlated with various environmental factors, such as latitude, longitude, depth, dissolved oxygen (DO), salinity, nitrate, nitrite, and pH (Figure S9c). Cyanophage community dissimilarity showed significant associations with salinity and pH (Spearman; $p < 0.01$; Figure S9c), indicating that these gradients are key correlates of cyanophage community turnover across estuaries. A significant positive correlation was shown between cyanophage abundance and *Synechococcus* abundance (Spearman, $p < 0.01$; Figure S9c). PERMANOVA and random forest analyses further supported that cyanophage community gradients are substantially predictable from measured environmental factors (R^2 = 54.2%), with salinity emerging as the strongest predictor, while the host gradients show weak predictability (R^2 = 9.7%) (Figure S9d).

3.4. Unique Auxiliary Metabolic Genes in Estuarine Cyanophages Reveal Their Viral Adaptability

To explore how estuarine cyanophages adapt to their habitats, we identified auxiliary metabolic genes (AMGs) among estuarine cyanophages via a refined approach (Materials and Methods section). In total, 462 AMGs spanning 12 distinct metabolic pathways were detected across 83 vOTUs (Figure 2 and Table S10). The AMG distribution varied significantly among cyanophage groups: cyanomyoviruses contained the majority (348 AMGs), followed by unclassified cyanophages (108 AMGs), while relatively few AMGs were detected in cyanosiphoviruses (4 AMGs), and cyanopodoviruses (2 AMGs) (Table S10).

AMG composition varied across estuarine sites and sample types, and the highest AMG diversity was observed in cyanophages recovered from the Yangtze River Estuary (YR), particularly in sediment-derived samples (Figures 2 and S10a). Of particular ecological interest was the identification of the AMG (*desC*), associated with lipid desaturation and metabolism,[96] in vOTU25 derived from YR sediment samples. Given the role of desaturases in regulating membrane lipid composition and fluidity, the presence of *desC* suggests that estuarine cyanophages may modulate host membrane-associated physiology during infection, potentially contributing to host performance under fluctuating estuarine conditions (Figure S10a and Table S10).

Across all vOTUs, many AMGs were assigned to carbohydrate metabolism and cofactor and vitamin metabolism (Figure S10b), consistent with prior observations that cyanophages can augment host central metabolism and nutrient acquisition during infection.[97−99] To systematically evaluate whether estuarine AMG profiles differ from marine cyanophage, we compared our estuarine vOTUs with a curated set of sequenced marine cyanophage genomes from the NCBI Viral RefSeq database (v94) (Figure S11a,b and Table S11). Notably, in the marine reference genomes, only one glycosyltransferase (GT) family was detected (*ABO*) (Figure S11a and Table S11), whereas estuarine vOTUs encoded two GT families (*ABO* and *MGAT3*) (Figure 2 and Table S10). GTs have been implicated in glycosylation of viral structural proteins and sugar polymerization,[100] potentially enhancing phage adaptability and interaction with host cells. The broader GT repertoire observed in estuarine vOTUs may therefore reflect adaptation to the physicochemically variable estuarine environment.

To relate the AMG functional potential to estuarine physicochemical parameters, we summarized AMG abundances at the pathway level and examined among sample variation using Bray−Curtis dissimilarities and NMDS ordination (Figure S12a; stress = 0.167). Salinity was identified as the dominant gradient associated with cyanophage and *Synechococcus* community composition (r^2 = 0.745, p = 0.001) (Figure S12a). Environmental vector fitting (envfit) indicated that AMG pathway composition was most strongly aligned with nitrite (r^2 = 0.251, p = 0.058), with additional alignments with salinity (r^2 = 0.198, p = 0.108) and nitrate (r^2 = 0.183, p = 0.126) (Figure S12b). To identify which functional categories contributed to these patterns, we further performed a pathway-level Spearman correlation screen between environmental variables and AMG pathway abundances (Figure S12c). Along

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

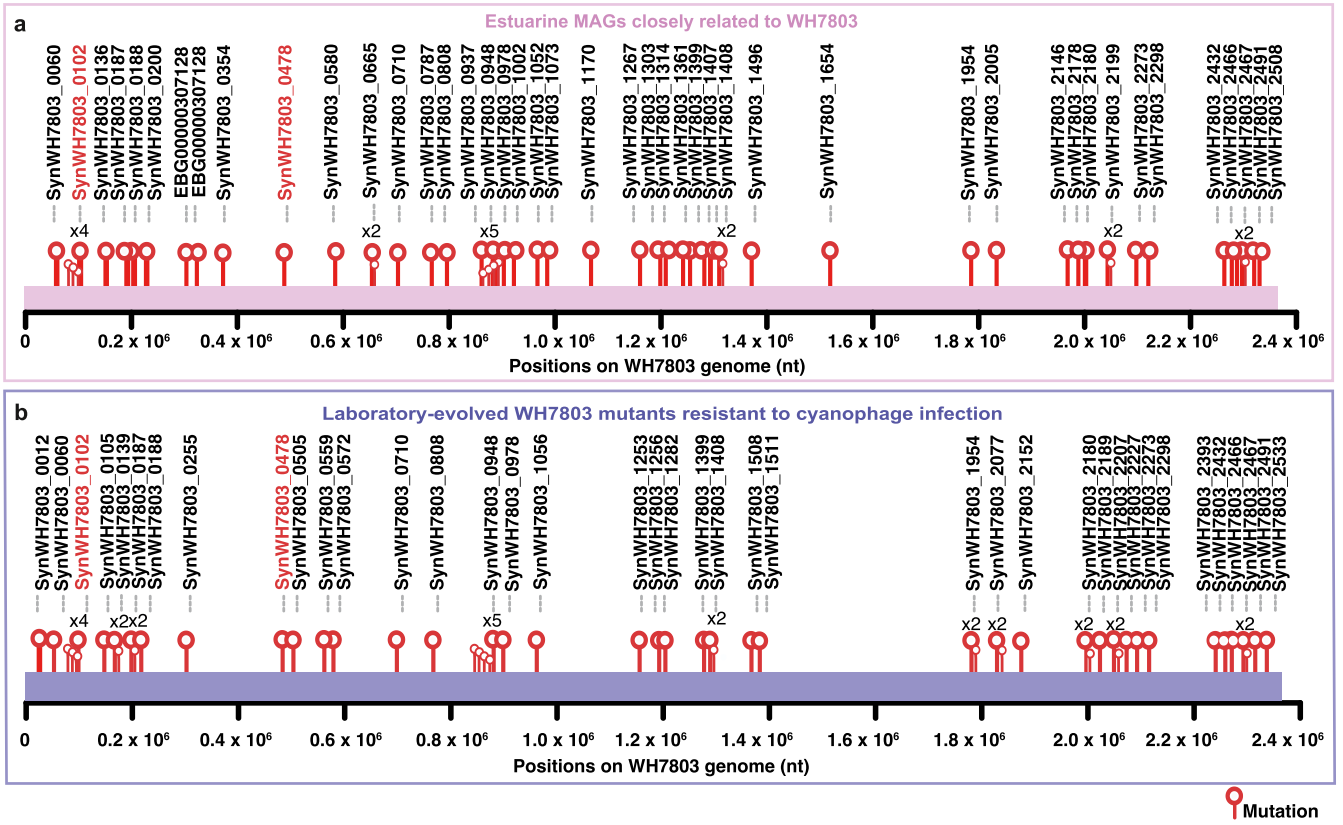


Figure 3. Mutant gene profiles in estuarine cyanobacterial metagenome-assembled genomes (MAGs) and laboratory-evolved cyanobacterial mutants resistant to cyanophage infection. This figure illustrates the distribution of potential mutant genes identified in both estuarine cyanobacterial MAGs and laboratory-evolved cyanobacterial mutants when compared to the reference genome of *Synechococcus* WH7803. Genes with mutations are indicated by red lines and circles. The number marked with "*x*" above the red circle represents the number of different kind mutations that occur on the mutant gene. The mutations SynWH7803_0102 (*rfbA*) and SynWH7803_0478 (*cpeT*), marked in red, are identified among these MAGs and mutants. The label "nt" refers to nucleotides. (a) Distribution of mutations within estuarine cyanobacterial MAGs ($n = 4$) that are closely related to WH7803. (b) Distribution of mutations in laboratory-evolved cyanobacterial mutants ($n = 18$) that are derived from WH7803 and resistant to cyanophage infection.

the estuarine-marine gradient, viral AMG repertoires show higher representation of pathways linked to carbon processing ($\rho = 0.528$, $p = 0.0096$), amino-acid related metabolism ($\rho = 0.511$, $p = 0.0128$), as well as cofactor and vitamin metabolism ($\rho = 0.478$, $p = 0.0211$) at higher salinity (Figure S12c). A notable negative association between nitrite and energy metabolism was shown ($\rho = -0.477$, $p = 0.0214$). Collectively, these analyses suggest that viral AMG repertoires vary along major estuarine gradients, particularly salinity and nitrogen chemistry (nitrite and nitrate), consistent with environmental modulation of viral functional strategies in estuarine ecosystems.

3.5. Genetic Variations Reveal Evolutionary Adaptations of Cyanobacteria under Cyanophage-Driven Selective Pressures

To gain deeper insights into the impacts of environmental selection on cyanobacterial genomic evolution, we first investigated genetic variations in four MAGs closely related to *Synechococcus* WH7803 from the YR samples, namely, YR_A3B_bin4.67, YR_J3B_bin7.192, YR_B8B_bin5.291 and YR_B8B_bin10.298 (Table S12). More than 14% of the SNPs identified among these MAGs represented nonsynonymous mutations compared to the *Synechococcus* WH7803 reference genome (Table S13), potentially altering protein sequences and impairing cell fitness.[101] Over 80% of these SNPs were associated with membrane transporters (Table S13), proteins situated on the host cell surface that are critical in the phage infection process.[102] Among these genetic variations (Figure 3a), approximately 7% of the nucleotide substitutions occurred in the SynWH7803_0102 (*rfbA*) gene. This gene encodes the first enzyme in the dTDP-L-rhamnose biosynthesis pathway, which is crucial for the biosynthesis of L-rhamnose, an important residue of the O-antigen of lipopolysaccharide (LPS) in Gram-negative bacteria.[103] LPS serves as a primary receptor for many cyanophages, which was also reported to be involved in bacteriophage attachment in WH7803.[104] Mutations in *rfbA* could lead to structural alterations or deficiencies in LPS, potentially preventing efficient phage adsorption or reducing the accessibility of phage-binding sites. This disruption in the initial attachment step of the phage infection cycle may explain why mutations in *rfbA* contribute to resistance. Additionally, we investigated bacterial defense systems within these MAGs using DefenseFinder[90] (Table S14). Our analysis revealed a diverse repertoire of defense system genes, including those in Cyclic-oligonucleotide-based Antiphage Signaling Systems (CBASS), Bacteriophage Exclusion (BREX), Defense Island System Associated with Restriction-Modification (DISARM), Restriction-Modification Systems (RM), and Toxin-Antitoxin Systems (ShosTA) (Table S14). Many identified defense-related genes involve nucleic acid modification or degradation (Table S14), such as

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

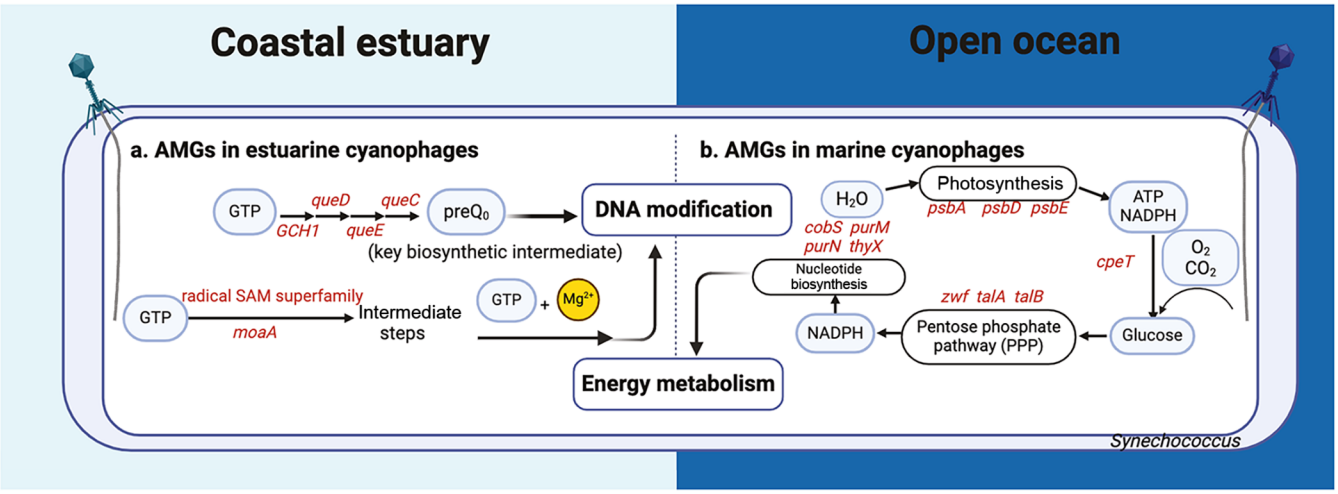


Figure 4. Virus-host interactions and environmental adaptation in the estuarine and marine *Synechococcus*. The figures show viral auxiliary metabolic genes (AMGs) belonging to specific functional pathways among estuarine (left) and marine (right) cyanophages in the light infection process. (a) AMGs in estuarine cyanophages. Related genes identified in cyanophages are marked in red. (b) AMGs in marine cyanophages. Related genes identified in cyanophages are marked in red. Detailed statistics are provided in Supporting Information Tables S10 and S11.

Dnd__DndA and *Dpd__QueE*, suggesting their role in resisting phage infection and thereby providing a broader context for understanding resistance strategies.

To validate and extend these observations, we conducted evolutionary experiments using a classical *Synechococcus*-cyanophage system (Figure S13). Eight cyanophages were isolated using WH7803 as a host, and of these were classified as T4-like cyanomyoviruses (Table S6). We selected 18 laboratory-evolved *Synechococcus* mutants in resistance to cyanophage infection (Figure S13) and performed whole-genome sequencing on these resistant mutants (Table S15). These mutants exhibited significantly lower growth rates compared to the wild-type strains (Figure S14), indicating potential fitness trade-offs associated with resistance. Among the mutant genomes, 128 mutations were identified (Figure 3b). Genome-wide linkage disequilibrium (LD) analysis revealed low LD (R^2 < 0.5) between neighboring SNPs in this study, suggesting that these mutations are independent and unique under cyanophage-driven selection. Mutations in the *metA* gene (SynWH7803_1408), encoding homoserine O-acetyltransferase, were detected in almost all isolates resistant to S-SCSM2b and in other isolates resistant to different cyanophages (Table S15). This enzyme plays a critical role in peptidoglycan synthesis and the biosynthesis of methionine and lysine,[105] potentially influencing the structural integrity of the cell wall and impacting phage infection. These findings suggest that *Synechococcus* can evolve resistance to multiple viruses through diverse mechanisms, as mutations in entirely different genes are capable of conferring resistance to the same virus (Table S15).

To decipher the diversification of evolved cyanobacteria in the natural environment, we compared mutations in MAGs with those in laboratory-evolved mutants. Notably, the same mutations in *rfbA* (SynWH7803_0102) and *cpeT* (SynWH7803_0478) were detected in both the MAGs and isolated mutants (Tables S13 and S15). Approximately 60% of the nucleotide substitutions in laboratory mutants were in the *rfbA* gene (Table S15), reinforcing its role in LPS biosynthesis and phage resistance. The *cpeT* gene, which encodes a chromophore lyase essential for photosynthesis, likely influences energy metabolism and intracellular conditions required for successful phage replication.[106] These results suggest that *rfbA* and *cpeT* are recurrent targets of cyanophage-driven selection in WH7803-like *Synechococcus* in estuaries and in a marine *Synechococcus* genetic background represented by WH7803. Mutations in genes associated with polysaccharide and LPS pathways were also detected (SynWH7803_1303 in Table S13 and SynWH7803_0139 in Table S15). These genes are predicted to participate in LPS and cell-surface glycoconjugate biosynthesis and, therefore, could regulate proteins exposed outside the cell wall to which viruses attach.[18] Together, host mutations in putative glycosylation-related genes (including GT-linked functions) and the broad repertoire of GTs encoded by cyanophage vOTUs support an evolutionary interplay between cyanophage infection and host surface modifications.

4. DISCUSSION

By integrating metagenomic reconstruction of cyanophages and cyanobacteria across four Chinese estuaries with laboratory evolution of phage-resistant *Synechococcus*, we connect environmental gradients, viral functions, and host resistance targets and experimentally verify the genetic basis of virus-host coevolution in natural environments. We identified an expanded set of AMGs and reciprocal evolutionary features that enhance viral adaptability, as shown in the virus-host interactions and environmental adaptation in estuarine and marine *Synechococcus* (Figure 4a,b).

Beyond light-dependent reactions of photosynthesis,[17,107] the pentose phosphate pathway (PPP), and regenerating NADPH[106] reported in marine systems (Figure 4b and Table S11), we identified genes in the radical S-adenosyl methionine (SAM) superfamily (*bchE*, *ahbC*, and *ahbD*) that may aid genome modification,[108] along with the *moaA* gene, which could facilitate cofactor utilization in nitrate-rich habitats (Table S10). Additionally, AMGs encoding GCHI and the Que superfamily (*queC*, *queD*, and *queE*) could drive the synthesis of GTP into the key biosynthetic intermediate 7-cyano-7-deazaguanine ($preQ_0$), thus protecting viral DNA from host restriction enzymes[109] (Table S10). The above AMGs are also reportedly encoded by viruses of deep-sea

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

origin[110] and freshwater origin[99] but have rarely been reported in estuarine environments, revealing the importance of these genes in estuarine environments. While prior work has examined the role of lipopolysaccharides in cyanophage infection,[16] we identified glycosyltransferases in estuarine cyanophages such as *ABO* and *MGAT3* (Table S10), which were uncommon in sequenced marine reference cyanophage genomes and may further reshape host surface structures, influencing attachment or evasion of host defenses. Disentangling selection from gene acquisition will require source-attribution analyses of AMG homologues, compositional signatures of recent gene gain, and expanded comparisons to metagenome-resolved marine cyanophage catalogs, coupled with infection assays across controlled salinity and nutrient regimes. Collectively, these findings imply that estuarine cyanophages may rely on a broader suite of metabolic functions and structural modifications to optimize infection and replication under the nutrient-variable and salinity-fluctuating conditions of brackish ecosystems.

In estuaries, dynamic environmental conditions such as salinity gradients and nutrient fluxes, likely amplify selective pressures on both *Synechococcus* and cyanophages, driving the coevolutionary arms race between *Synechococcus* and their viruses.[111] Coevolutionary dynamics between cyanophages and *Synechococcus* could influence the composition or functional capacity of cooccurring microbes through nutrient release and selective pressures.[112] Over time, frequent gene loss events occurred across *Synechococcus* lineages, reflecting adaptation to diverse environmental pressures. We predicted that *Synechococcus* can rapidly evolve in response to cyanophage infection, enhancing their survival. These observations are consistent with dynamic coevolutionary processes, wherein resistance emerges over tens to hundreds of generations within a relatively short period (weeks to months, based on the growth rates in Table S16), as *Synechococcus* populations adapt to fluctuating environmental conditions and viral pressures. The fundamental niche of *Synechococcus* in estuaries is primarily set by abiotic gradients, such as salinity, nutrients, light, and temperature, which can promote lineage turnover from freshwater to marine adapted populations and stratification patterns.[4] Spearman correlation analyses revealed that *Synechococcus* exhibits spatially distinct abundance and diversity patterns, transitioning from freshwater to marine lineages with increasing salinity and showing differences between surface and bottom layers, underscoring the estuary's role in supporting diverse *Synechococcus* across environmental gradients. In contrast, cyanophages occupy a narrower niche than their hosts because productive infection requires not only susceptible host presence but also permissive host physiological states.[113] The observation that multiple *Synechococcus* MAGs cluster with WH7803-like lineages and that assembled estuarine cyanophages are closely related to laboratory cyanomyoviruses suggests shared environmental filters across field and culture systems.

Concordantly, both mutations detected in field-derived *Synechococcus* MAGs (Table S13) and in laboratory-evolved mutants (Table S15) were predominantly associated with key processes, such as membrane-associated functions, host recognition and attachment, and energy metabolism. These patterns are compatible with arms-race and fluctuating-selection dynamics, where resistance and infectivity coevolve on short ecological time scales[32,102,114,115] as environmental conditions and host communities' shift. Experimental coevolution studies further demonstrated that genetic adaptations are specific to the coevolving cyanobacteria and cyanophages.[20] The enrichment of cyanophage auxiliary metabolic genes involved in redox regulation and envelope- and membrane-associated functions suggests that viral fitness may depend on specific combinations of salinity, nutrient supply, and light that maximize host metabolic capacity during infection, thereby shaping niche overlap and coevolution in estuaries. Salinity gradients in estuaries impose osmotic stress and can alter the composition and organization of the outer membrane glycoconjugates. Modifications in LPS and other surface glycans, such as those mediated by *rfbA* and GT-encoding genes, are therefore likely to be under strong selection as *Synechococcus* traverse brackish conditions. Although CRISPR-Cas can provide antiviral immunity, we did not recover CRISPR-based host evidence in our *Synechococcus* MAGs or detect significant spacer-protospacer matches, suggesting that CRISPR-mediated defense is rare in these lineages. This result is consistent with previous observations that a genome-wide survey showing that CRISPR-Cas systems,[116] while broadly distributed across *Cyanobacteria* but are largely absent in the marine *Synechococcus* and *Prochlorococcus* subclade. Even when CRISPR arrays are present, spacer homology to publicly available viral genomes are often extremely limited.[116] Cyanobacteria also deploy alternative defense portfolios, including restriction-modification (RM) systems, whose abundance varies markedly among cyanobacterial groups and is generally more expanded in filamentous than in unicellular lineages.[117] Thus, the occurrence of viral DNA-modification-linked AMGs in our cyanophages, such as SAM-superfamily and Que superfamily genes, may represent potential counter-defense strategies against nuclease-based immunity, complementing coevolutionary signatures observed in cell-surface and infection-related functions.

At the same time, temperature and seasonal stratification strongly influence *Synechococcus* growth rates, pigment composition, and membrane fluidity, and have been shown to correlate tightly with cyanophage dynamics in other systems.[118] We propose that in high-salinity and warm estuarine waters, salinity-driven changes in surface glycans and temperature-driven changes in physiology jointly shape the fitness landscape for cyanophage infection, with different combinations of *rfbA* and *cpeT* mutations and viral AMGs being favored under different seasonal regimes. We therefore anticipate that from winter to summer transitions and from low to high levels of salinity shifts will change infection networks, with high-salinity estuaries favoring selection on surface-glycan pathways. However, resolving seasonal impacts on virus-host coevolution will require paired, within-estuary longitudinal sampling across seasons and years, with comparable sequencing depth and standardized environmental measurements. Future work combining within-estuary time-series sampling with adsorption and fitness assays on defined LPS mutants and comparative infections using phages that differ in envelope and DNA modification AMG modules will be essential to directly test these proposed reciprocal mechanisms. Further experiments, such as targeted gene knockouts and infection assays, are necessary to determine whether these mutations in cyanobacteria directly influence cyanophage attachment or replication. Taken together, these patterns provide mechanistic evidence for coupled genomic and evolutionary dynamics of

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

cyanophages and hosts in estuaries, extending coevolutionary insights beyond laboratory systems to natural communities.

■ ASSOCIATED CONTENT

Data Availability Statement

Sequences associated with metagenome from environmental samples have been deposited in the NODE database under project accession no. OEP001524, OEP004266, and OEP004282. The data generated in this study can be accessed at Figshare repositories (10.6084/m9.figshare.28330520.v1 and 10.6084/m9.figshare.28330523.v1). The sources of the code, and programs used for analyses, are mentioned in *Materials and Methods*, and are also available at GitHub (https://github.com/SallyTam/Bioinformatic-codes-for-estuarine-cyanophages-and-cyanobacteria).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.5c12277.

Whole geographic distribution of collected samples and distribution patterns of cyanophage diversity in estuarine samples; the overall workflow of the bioinformatic analysis; pairwise comparisons of average nucleotide identity of shared genes in cyanophage genomes; Spearman's rank correlation coefficient between genome size and GC content in estuarine and marine cyanophages; genome maps of 83 assembled estuarine cyanophages; taxonomic evaluation of cyanophages; relative abundance of estuarine cyanophages in different estuarine environment; spatial patterns of cyanophage communities and comparative ordination of cyanophage and Synechococcus community composition across estuaries; relative abundance of host communities; function and abundance profiles of virus-encoded auxiliary metabolic genes (AMGs) in estuarine cyanophage genomes; function and abundance profiles of virus-encoded AMGs in sequenced genomes; environmental gradients associated with cyanophage community structure across estuarine samples; overview of the experimental design for selecting phage-resistant mutant; growth dynamics of mutant substrains compared with wild-type substrains (Figures S1−S14) (PDF)

Lists of metagenomic samples and environmental factors; general features of reference cyanophage genomes and their hosts based on Virus-Host Database; summary of assembled estuarine cyanophage genomes; predicted viral proteins of assembled estuarine cyanophage genomes based on KEGG, Pfam, and VOG databases; genomic profiles of assembled estuarine cyanophage genomes and their predicted hosts; summary of laboratory-isolated marine cyanophages; relative abundance of assembled estuarine cyanophages at species level calculated by RPKM values in different estuarine samples; normalized cyanobacterial read abundance (RPKM) at the species level based on 16S rRNA gene sequences across estuarine samples; summary of 77 cyanobacterial metagenomic assembly genomes (MAGs); the annotation of AMGs in assembled estuarine cyanophages; the annotation of AMGs in isolated marine cyanophage genomes; statistics of genetic variations detected in four cyanobacterial MAGs closely related to Synechococcus WH7803; mutations in MAGs closely related to WH7803; statistics of identified defense genes from MAGs closely related to WH7803; mutations in laboratory-evolved phage-resistant Synechococcus mutants; growth rates of phage-resistant Synechococcus mutants (Tables S1−S16) (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Qinglu Zeng − *HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen 518000, China; Department of Ocean Science, The Hong Kong University of Science and Technology, Kowloon, Hong Kong 999077, China*; Email: zeng@ust.hk

Meng Li − *SZU-HKUST Joint PhD Program in Marine Environmental Science, Shenzhen University, Shenzhen 518060, China; Archaeal Biology Centre, Synthetic Biology Research Center, Shenzhen Key Laboratory of Marine Microbiome Engineering, Key Laboratory of Marine Microbiome Engineering of Guangdong Higher Education Institutes, Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China*; orcid.org/0000-0001-8675-0758; Email: limeng848@szu.edu.cn

Authors

Yisha Tan − *SZU-HKUST Joint PhD Program in Marine Environmental Science, Shenzhen University, Shenzhen 518060, China; Archaeal Biology Centre, Synthetic Biology Research Center, Shenzhen Key Laboratory of Marine Microbiome Engineering, Key Laboratory of Marine Microbiome Engineering of Guangdong Higher Education Institutes, Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China; HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen 518000, China; Department of Ocean Science, The Hong Kong University of Science and Technology, Kowloon, Hong Kong 999077, China*

Dayu Zou − *Archaeal Biology Centre, Synthetic Biology Research Center, Shenzhen Key Laboratory of Marine Microbiome Engineering, Key Laboratory of Marine Microbiome Engineering of Guangdong Higher Education Institutes, Institute for Advanced Study, Shenzhen University, Shenzhen 518060, China*

Can Ni − *Department of Ocean Science, The Hong Kong University of Science and Technology, Kowloon, Hong Kong 999077, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.5c12277

Author Contributions

Y.T., M.L., and Q.Z. designed the study and wrote the manuscript. Y.T. and D.Z. conducted metagenomic sample preparations for sequencing, generated and analyzed the geochemical data. Y.T. performed the bioinformatic analyses related to the recovery and annotation of viruses from metagenomes and data visualization. Q.Z. collected virus samples. Y.T. purified the viruses and isolated phage-resistant mutants. Y.T. prepared DNA libraries of mutants for whole-genome sequencing, performed fragment analysis, and conducted the bioinformatic analyses. C.N. helped to investigate the data analyses. Y.T. drafted the manuscript, edited and finally reviewed by M.L. and Q.Z.

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Yuchen Xiao (City University of Hong Kong) for valuable discussions on graphs and the professional advice on MATLAB, Python, Shell, and R programming. This work was supported by grants from the National Key Research and Development Program of China (2022YFA0912200), the National Natural Science Foundation of China (92451302, 92251306, 32225003, 32393970, 32393971), the Guangdong Major Project of Basic and Applied Basic Research (2023B0303000017), the Shenzhen Science and Technology Program (JCYJ20200109105010363), Shenzhen University 2035 Program for Excellent Research (2022B002), and the Research Grants Council of Hong Kong Special Administrative Region, China (grant no. HKUST C6012-22G). This work was supported in part by the Project of Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone (HZQB-KCZYB-2020083).

■ REFERENCES

(1) Waterbury, J. B.; Watson, S. W.; Guillard, R. R. L.; Brand, L. E. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **1979**, *277* (5694), 293−294.

(2) Partensky, F.; Blanchot, J.; Vaulot, D. Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bull. Inst. Océanogr. Monaco* **1999**, *19*, 457−476.

(3) Cai, L.; Li, H.; Deng, J.; Zhou, R.; Zeng, Q. Biological interactions with Prochlorococcus: implications for the marine carbon cycle. *Trends Microbiol.* **2024**, *32* (3), 280−291.

(4) Xia, X.; Guo, W.; Tan, S.; Liu, H. Synechococcus assemblages across the salinity gradient in a salt wedge estuary. *Front. microbiol.* **2017**, *8*, No. 1254.

(5) Chisholm, S. W.; Olson, R. J.; Zettler, E. R.; Goericke, R.; Waterbury, J. B.; Welschmeyer, N. A. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **1988**, *334* (6180), 340−343.

(6) Flombaum, P.; Gallegos, J. L.; Gordillo, R. A.; Rincon, J.; Zabala, L. L.; Jiao, N. A. Z.; Karl, D. M.; Li, W. K. W.; Lomas, M. W.; Veneziano, D.; Vera, C. S.; Vrugt, J. A.; Martiny, A. C. Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (24), 9824−9829.

(7) Richardson, T. L.; Jackson, G. A. Small phytoplankton and carbon export from the surface ocean. *Science* **2007**, *315* (5813), 838−840.

(8) Zwirglmaier, K.; Jardillier, L.; Ostrowski, M.; Mazard, S.; Garczarek, L.; Vaulot, D.; Not, F.; Massana, R.; Ulloa, O.; Scanlan, D. J. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ. Microbiol.* **2008**, *10* (1), 147−161.

(9) Bergh, Ø.; Borsheim, K. Y.; Bratbak, G.; Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **1989**, *340* (6233), 467−468.

(10) Matteson, A. R.; Loar, S. N.; Bourbonniere, R. A.; Wilhelm, S. W. Molecular enumeration of an ecologically important cyanophage in a Laurentian Great Lake. *Appl. Environ. Microbiol.* **2011**, *77* (19), 6772−6779.

(11) Suttle, C. A. Viruses in the sea. *Nature* **2005**, *437* (7057), 356−361.

(12) Matteson, A. R.; Rowe, J. M.; Ponsero, A. J.; Pimentel, T. M.; Boyd, P. W.; Wilhelm, S. W. High abundances of cyanomyoviruses in marine ecosystems demonstrate ecological relevance. *FEMS Microbiol. Ecol.* **2013**, *84* (2), 223−234.

(13) Carlson, M. C. G.; Ribalet, F.; Maidanik, I.; Durham, B. P.; Hulata, Y.; Ferrón, S.; Weissenbach, J.; Shamir, N.; Goldin, S.; Baran, N.; et al. Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. *Nat. Microbiol.* **2022**, *7* (4), 570−580.

(14) Carlson, M. C. G.; Ribalet, F.; Maidanik, I.; Durham, B. P.; Hulata, Y.; Ferron, S.; Weissenbach, J.; Shamir, N.; Goldin, S.; Baran, N.; Cael, B. B.; Karl, D. M.; White, A. E.; Armbrust, E. V.; Lindell, D. Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. *Nat. Microbiol.* **2022**, *7* (4), 570−580.

(15) Demory, D.; Liu, R.; Chen, Y.; Zhao, F.; Coenen, A. R.; Zeng, Q.; Weitz, J. S. Linking light-dependent life history traits with population dynamics for *Prochlorococcus* and cyanophage. *mSystems* **2020**, *5* (2), No. e00586−19.

(16) Avrani, S.; Wurtzel, O.; Sharon, I.; Sorek, R.; Lindell, D. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **2011**, *474* (7353), 604−608.

(17) Lindell, D.; Jaffe, J. D.; Coleman, M. L.; Futschik, M. E.; Axmann, I. M.; Rector, T.; Kettler, G.; Sullivan, M. B.; Steen, R.; Hess, W. R.; Church, G. M.; Chisholm, S. W. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **2007**, *449* (7158), 83−86.

(18) Marston, M. F.; Pierciey, F. J.; Shepard, A.; Gearin, G.; Qi, J.; Yandava, C.; Schuster, S. C.; Henn, M. R.; Martiny, J. B. H. Rapid diversification of coevolving marine Synechococcus and a virus. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (12), 4544−4549.

(19) Suttle, C. A.; Chan, A. M. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar. Ecol.: Prog. Ser.* **1993**, *92* (1−2), 99−109.

(20) Enav, H.; Kirzner, S.; Lindell, D.; Mandel-Gutfreund, Y.; Beja, O. Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. *Nat. Commun.* **2018**, *9* (1), No. 4698.

(21) Waterbury, J. B.; Valois, F. W. Resistance to co-occurring phages enables marine synechococcus communities to coexist with cyanophages abundant in seawater. *Appl. Environ. Microbiol.* **1993**, *59* (10), 3393−3399.

(22) Ackermann, H. W. Bacteriophage observations and evolution. *Res. Microbiol.* **2003**, *154* (4), 245−251.

(23) Sullivan, M. B.; Waterbury, J. B.; Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **2003**, *424* (6952), 1047−1051.

(24) Cai, L.; Liu, H.; Zhang, W.; Xiao, S.; Zeng, Q.; Dang, S. Cryo-EM structure of cyanophage P-SCSP1u offers insights into DNA gating and evolution of T7-like viruses. *Nat. Commun.* **2023**, *14* (1), No. 6438.

(25) Doron, S.; Fedida, A.; Hernandez-Prieto, M. A.; Sabehi, G.; Karunker, I.; Stazic, D.; Feingersch, R.; Steglich, C.; Futschik, M.; Lindell, D.; Sorek, R. Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J.* **2016**, *10* (6), 1437−1455.

(26) McDaniel, L. D.; Young, E.; Delaney, J.; Ruhnau, F.; Ritchie, K. B.; Paul, J. H. High frequency of horizontal gene transfer in the oceans. *Science* **2010**, *330* (6000), No. 50.

(27) Breitbart, M.; Bonnain, C.; Malki, K.; Sawaya, N. A. Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* **2018**, *3* (7), 754−766.

(28) Enav, H.; Mandel-Gutfreund, Y.; Beja, O. Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome* **2014**, *2* (1), No. 9.

(29) Zhao, F.; Lin, X.; Cai, K.; Jiang, Y.; Ni, T.; Chen, Y.; Feng, J.; Dang, S.; Zhou, C. Z.; Zeng, Q. Biochemical and structural characterization of the cyanophage-encoded phosphate-binding protein: implications for enhanced phosphate uptake of infected cyanobacteria. *Environ. Microbiol.* **2022**, *24* (7), 3037−3050.

(30) Millard, A.; Clokie, M. R. J.; Shub, D. A.; Mann, N. H. Genetic organization of the psbAD region in phages infecting marine Synechococcus strains. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (30), 11007−11012.

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(31) Thompson, L. R.; Zeng, Q.; Kelly, L.; Huang, K. H.; Singer, A. U.; Stubbe, J.; Chisholm, S. W. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (39), E757−E764.
(32) Lindell, D.; Sullivan, M. B.; Johnson, Z. I.; Tolonen, A. C.; Rohwer, F.; Chisholm, S. W. Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (30), 11013−11018.
(33) Liu, R.; Liu, Y.; Chen, Y.; Zhan, Y.; Zeng, Q. Cyanobacterial viruses exhibit diurnal rhythms during infection. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (28), 14077−14082.
(34) Lennon, J. T.; Khatana, S. A. M.; Marston, M. F.; Martiny, J. B. Is there a cost of virus resistance in marine cyanobacteria? *ISME J.* **2007**, *1* (4), 300−312.
(35) Paterson, S.; Vogwill, T.; Buckling, A.; Benmayor, R.; Spiers, A. J.; Thomson, N. R.; Quail, M.; Smith, F.; Walker, D.; Libberton, B.; Fenton, A.; Hall, N.; Brockhurst, M. A. Antagonistic coevolution accelerates molecular evolution. *Nature* **2010**, *464* (7286), 275−278.
(36) Hampton, H. G.; Watson, B. N.; Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **2020**, *577* (7790), 327−336.
(37) Fucich, D.; Chen, F. Presence of toxin-antitoxin systems in picocyanobacteria and their ecological implications. *ISME J.* **2020**, *14* (11), 2843−2850.
(38) Cai, L.; Chen, Y.; Xiao, S.; Liu, R.; He, M.; Zhang, R.; Zeng, Q. Abundant and cosmopolitan lineage of cyanopodoviruses lacking a DNA polymerase gene. *ISME J.* **2023**, *17* (2), 252−262.
(39) Hopkinson, C. S.; Smith, E. M. Estuarine respiration: an overview of benthic, pelagic, and whole system respiration. *Respir. Aquat. Ecosys.* **2005**, 122−146.
(40) Wang, K.; Chen, F. Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ. Microbiol.* **2008**, *10* (2), 300−312.
(41) Wang, K.; Chen, F. Genetic diversity and population dynamics of cyanophage communities in the Chesapeake Bay. *Aquat. Microb. Ecol.* **2004**, *34* (2), 105−116.
(42) Chen, F.; Wang, K.; Kan, J.; Suzuki, M. T.; Wommack, K. E. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* **2006**, *72* (3), 2239−2243.
(43) Huang, S.; Wang, K.; Jiao, N.; Chen, F. Genome sequences of siphoviruses infecting marine Synechococcus unveil a diverse cyanophage group and extensive phage−host genetic exchanges. *Environ. Microbiol.* **2012**, *14* (2), 540−558.
(44) Huang, S.; Zhang, S.; Jiao, N.; Chen, F. Comparative genomic and phylogenomic analyses reveal a conserved core genome shared by estuarine and oceanic cyanopodoviruses. *PLoS One* **2015**, *10* (11), No. e0142962.
(45) Huang, S.; Zhang, S.; Jiao, N.; Chen, F. Marine cyanophages demonstrate biogeographic patterns throughout the global ocean. *Appl. Environ. Microbiol.* **2015**, *81* (1), 441−452.
(46) Chen, F. Picocyanobacteria in the Chesapeake Bay: isolation, diversity, and adaptation. *Mar. Life Sci. Technol.* **2025**, *7*, 434−449.
(47) Ruiz-Perez, C. A.; Tsementzi, D.; Hatt, J. K.; Sullivan, M. B.; Konstantinidis, K. T. Prevalence of viral photosynthesis genes along a freshwater to saltwater transect in Southeast USA. *Environ. Microbiol. Rep.* **2019**, *11* (5), 672−689.
(48) Association, A. P. H.; Association, A. W. W.; Federation, W.; Federation, W. *Standard Methods for the Examination of Water and Wastewater*; American Public Health Association, 1915; Vol. 2.
(49) Xie, W.; Luo, H.; Murugapiran, S. K.; Dodsworth, J. A.; Chen, S.; Sun, Y.; Hedlund, B. P.; Wang, P.; Fang, H.; Deng, M.; Zhang, C. L. Localized high abundance of Marine Group II archaea in the subtropical Pearl River Estuary: implications for their niche adaptation. *Environ. Microbiol.* **2018**, *20* (2), 734−754.
(50) Zou, D.; Wan, R.; Han, L.; Xu, M. N.; Liu, Y.; Liu, H.; Kao, S. J.; Li, M. Genomic characteristics of a novel species of ammonia-oxidizing archaea from the Jiulong River Estuary. *Appl. Environ. Microbiol.* **2020**, *86* (18), No. e00736−20.
(51) Joshi, N. A.; Fass, J. N.*Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files133*, In Software; GitHub, Inc., 2011.
(52) Guo, J.; Bolduc, B.; Zayed, A. A.; Varsani, A.; Dominguez-Huerta, G.; Delmont, T. O.; Pratama, A. A.; Gazitua, M. C.; Vik, D.; Sullivan, M. B.; Roux, S. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **2021**, *9* (1), No. 37.
(53) Ren, J.; Song, K.; Deng, C.; Ahlgren, N. A.; Fuhrman, J. A.; Li, Y.; Xie, X.; Poplin, R.; Sun, F. Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **2020**, *8* (1), 64−77.
(54) Nayfach, S.; Camargo, A. P.; Schulz, F.; Eloe-Fadrosh, E.; Roux, S.; Kyrpides, N. C. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **2021**, *39* (5), 578−585.
(55) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28* (23), 3150−3152.
(56) Bin Jang, H.; Bolduc, B.; Zablocki, O.; Kuhn, J. H.; Roux, S.; Adriaenssens, E. M.; Brister, J. R.; Kropinski, A. M.; Krupovic, M.; Lavigne, R.; Turner, D.; Sullivan, M. B. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **2019**, *37* (6), 632−639.
(57) Kieft, K.; Adams, A.; Salamzade, R.; Kalan, L.; Anantharaman, K. vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Res.* **2022**, *50* (14), No. e83.
(58) Hyatt, D.; Chen, G. L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* **2010**, *11* (1), No. 119.
(59) Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **1999**, *27* (1), 29−34.
(60) Huerta-Cepas, J.; Forslund, K.; Coelho, L. P.; Szklarczyk, D.; Jensen, L. J.; von Mering, C.; Bork, P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **2017**, *34* (8), 2115−2122.
(61) El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L. L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S. C. E.; Finn, R. D. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47* (D1), D427−D432.
(62) Roux, S.; Paez-Espino, D.; Chen, I. A.; Palaniappan, K.; Ratner, A.; Chu, K.; Reddy, T. B. K.; Nayfach, S.; Schulz, F.; Call, L.; Neches, R. Y.; Woyke, T.; Ivanova, N. N.; Eloe-Fadrosh, E. A.; Kyrpides, N. C. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res.* **2021**, *49* (D1), D764−D775.
(63) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2007**, *35* (Database issue), D61−D65.
(64) Bland, C.; Ramsey, T. L.; Sabree, F.; Lowe, M.; Brown, K.; Kyrpides, N. C.; Hugenholtz, P. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinf.* **2007**, *8*, No. 209.
(65) Chan, P. P.; Lin, B. Y.; Mak, A. J.; Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **2021**, *49* (16), 9077−9096.
(66) Mihara, T.; Nishimura, Y.; Shimizu, Y.; Nishiyama, H.; Yoshikawa, G.; Uehara, H.; Hingamp, P.; Goto, S.; Ogata, H. Linking virus genomes with host taxonomy. *Viruses* **2016**, *8* (3), No. 66.
(67) Aroney, S.; Newell, R.; Nissen, J.; Camargo, A.; Tyson, G.; Woodcroft, B. CoverM: Read coverage calculator for metagenomics *Zenodo* 2024.
(68) Peng, Y.; Leung, H. C.; Yiu, S. M.; Chin, F. Y. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **2012**, *28* (11), 1420−1428.
(69) Kang, D. D.; Li, F.; Kirton, E.; Thomas, A.; Egan, R.; An, H.; Wang, Z. MetaBAT 2: an adaptive binning algorithm for robust and

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

efficient genome reconstruction from metagenome assemblies. *PeerJ.* **2019**, *7*, No. e7359.
(70) Sieber, C. M. K.; Probst, A. J.; Sharrar, A.; Thomas, B. C.; Hess, M.; Tringe, S. G.; Banfield, J. F. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **2018**, *3* (7), 836−843.
(71) Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **2015**, *25* (7), 1043−1055.
(72) Chaumeil, P. A.; Mussig, A. J.; Hugenholtz, P.; Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **2020**, *36* (6), 1925−1927.
(73) Moore, L. R.; Post, A. F.; Rocap, G.; Chisholm, S. W. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* **2002**, *47* (4), 989−996.
(74) Capella-Gutiérrez, S.; Silla-Martinez, J. M.; Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25* (15), 1972−1973.
(75) Nguyen, L. T.; Schmidt, H. A.; von Haeseler, A.; Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32* (1), 268−274.
(76) Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49* (W1), W293−W296.
(77) Kieft, K.; Zhou, Z.; Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **2020**, *8* (1), No. 90.
(78) Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27−30.
(79) Morris, J. J.; Kirkegaard, R.; Szul, M. J.; Johnson, Z. I.; Zinser, E. R. Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by "helper" heterotrophic bacteria. *Appl. Environ. Microbiol.* **2008**, *74* (14), 4530−4534.
(80) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30* (15), 2114−2120.
(81) Ewels, P.; Magnusson, M.; Lundin, S.; Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32* (19), 3047−3048.
(82) Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9* (4), 357−359.
(83) McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; DePristo, M. A. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20* (9), 1297−1303.
(84) Danecek, P.; Auton, A.; Abecasis, G.; Albers, C. A.; Banks, E.; DePristo, M. A.; Handsaker, R. E.; Lunter, G.; Marth, G. T.; Sherry, S. T.; McVean, G.; Durbin, R. Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **2011**, *27* (15), 2156−2158.
(85) Cingolani, P.; Platts, A.; Wang le, L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S. J.; Lu, X.; Ruden, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6* (2), 80−92.
(86) Hua, Z. S.; Qu, Y. N.; Zhu, Q.; Zhou, E. M.; Qi, Y. L.; Yin, Y. R.; Rao, Y. Z.; Tian, Y.; Li, Y. X.; Liu, L.; Castelle, C. J.; Hedlund, B. P.; Shu, W. S.; Knight, R.; Li, W. J. Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* **2018**, *9* (1), No. 2832.
(87) Johnson, L. S.; Eddy, S. R.; Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinf.* **2010**, *11*, No. 431.
(88) Galperin, M. Y.; Wolf, Y. I.; Makarova, K. S.; Alvarez, R. V.; Landsman, D.; Koonin, E. V. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **2021**, *49* (D1), D274−D281.
(89) Martin, M. J.; Orchard, S.; Bateman, A.; The UniProt Consortium; et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D523−D531.
(90) Tesson, F.; Hervé, A.; Mordret, E.; Touchon, M.; D'humières, C.; Cury, J.; Bernheim, A. Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **2022**, *13* (1), No. 2561.
(91) Pritchard, J. K.; Przeworski, M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **2001**, *69* (1), 1−14.
(92) Roux, S.; Brum, J. R.; Dutilh, B. E.; Sunagawa, S.; Duhaime, M. B.; Loy, A.; Poulos, B. T.; Solonenko, N.; Lara, E.; Poulain, J.; Pesant, S.; Kandels-Lewis, S.; Dimier, C.; Picheral, M.; Searson, S.; Cruaud, C.; Alberti, A.; Duarte, C. M.; Gasol, J. M.; Vaque, D.; Tara Oceans, C.; Bork, P.; Acinas, S. G.; Wincker, P.; Sullivan, M. B. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **2016**, *537* (7622), 689−693.
(93) Gregory, A. C.; Zayed, A. A.; Conceicao-Neto, N.; Temperton, B.; Bolduc, B.; Alberti, A.; Ardyna, M.; Arkhipova, K.; Carmichael, M.; Cruaud, C.; Dimier, C.; Dominguez-Huerta, G.; Ferland, J.; Kandels, S.; Liu, Y.; Marec, C.; Pesant, S.; Picheral, M.; Pisarev, S.; Poulain, J.; Tremblay, J. E.; Vik, D.; Tara Oceans, C.; Babin, M.; Bowler, C.; Culley, A. I.; de Vargas, C.; Dutilh, B. E.; Iudicone, D.; Karp-Boss, L.; Roux, S.; Sunagawa, S.; Wincker, P.; Sullivan, M. B.; et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **2019**, *177* (5), 1109−1123.e14.
(94) Brum, J. R.; Ignacio-Espinoza, J. C.; Roux, S.; Doulcier, G.; Acinas, S. G.; Alberti, A.; Chaffron, S.; Cruaud, C.; de Vargas, C.; Gasol, J. M.; Gorsky, G.; Gregory, A. C.; Guidi, L.; Hingamp, P.; Iudicone, D.; Not, F.; Ogata, H.; Pesant, S.; Poulos, B. T.; Schwenck, S. M.; Speich, S.; Dimier, C.; Kandels-Lewis, S.; Picheral, M.; Searson, S.; Bork, P.; Bowler, C.; Sunagawa, S.; Wincker, P.; Karsenti, E.; Sullivan, M. B.; Tara Oceans Coordinators. Patterns and ecological drivers of ocean viral communities. *Science* **2015**, *348* (6237), No. 1261498.
(95) Fuhrman, J. A. Microbial community structure and its functional implications. *Nature* **2009**, *459* (7244), 193−199.
(96) Murata, N.; Siegenthaler, P. A. *Lipids in Photosynthesis: Structure, Function, and Genetics*; Kluwer Academic Publishers, 1998.
(97) Crummett, L. T.; Puxty, R. J.; Weihe, C.; Marston, M. F.; Martiny, J. B. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **2016**, *499*, 219−229.
(98) Wang, Q.; Cai, L.; Zhang, R.; Wei, S.; Li, F.; Liu, Y.; Xu, Y. A unique set of auxiliary metabolic genes found in an isolated cyanophage sheds new light on marine phage-host interactions. *Microbiol. spectr.* **2022**, *10* (5), No. e0236722.
(99) Ain, Q. u.; Wu, K.; Wu, X.; Bai, Q.; Li, Q.; Zhou, C.-Z.; Wu, Q. Cyanophage-encoded auxiliary metabolic genes in modulating cyanobacterial metabolism and algal bloom dynamics. *Front. virol.* **2024**, *4*, No. 1461375.
(100) Kashtan, N.; Roggensack, S. E.; Rodrigue, S.; Thompson, J. W.; Biller, S. J.; Coe, A.; Ding, H.; Marttinen, P.; Malmstrom, R. R.; Stocker, R.; Follows, M. J.; Stepanauskas, R.; Chisholm, S. W. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **2014**, *344* (6182), 416−420.
(101) Walsh, I. M.; Bowman, M. A.; Soto Santarriaga, I. F.; Rodriguez, A.; Clark, P. L. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (7), 3528−3534.
(102) Avrani, S.; Wurtzel, O.; Sharon, I.; Sorek, R.; Lindell, D. Genomic Island variability facilitates Prochlorococcus-virus coexistence. *Nature* **2011**, *474* (7353), 604−608.
(103) Blankenfeldt, W.; Asuncion, M.; Lam, J. S.; Naismith, J. H. The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). *EMBO J.* **2000**, *19* (24), 6652−6663.

https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(104) Zwirglmaier, K.; Spence, E.; Zubkov, M. V.; Scanlan, D. J.; Mann, N. H. Differential grazing of two heterotrophic nanoflagellates on marine *Synechococcus* strains. *Environ. Microbiol.* **2009**, *11* (7), 1767−1776.
(105) Viola, R.; Faehnle, C.; Blanco, J.; Moore, R.; Liu, X.; Arachea, B.; Pavlovsky, A. The catalytic machinery of a key enzyme in amino acid biosynthesis. *J. Amino Acids* **2011**, No. 352538, DOI: 10.4061/2011/352538.
(106) Zhao, C.; Hoppner, A.; Xu, Q. Z.; Gartner, W.; Scheer, H.; Zhou, M.; Zhao, K. H. Structures and enzymatic mechanisms of phycobiliprotein lyases CpcE/F and PecE/F. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (50), 13170−13175.
(107) Mann, N. H.; Cook, A.; Millard, A.; Bailey, S.; Clokie, M. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **2003**, *424* (6950), No. 741.
(108) Zhang, J.; Zheng, Y. G. SAM/SAH analogs as versatile tools for SAM-dependent methyltransferases. *ACS Chem. Biol.* **2016**, *11* (3), 583−597.
(109) Hutinet, G.; Kot, W.; Cui, L.; Hillebrand, R.; Balamkundu, S.; Gnanakalai, S.; Neelakandan, R.; Carstens, A. B.; Fa Lui, C.; Tremblay, D.; Jacobs-Sera, D.; Sassanfar, M.; Lee, Y. J.; Weigele, P.; Moineau, S.; Hatfull, G. F.; Dedon, P. C.; Hansen, L. H.; de Crecy-Lagard, V. 7-Deazaguanine modifications protect phage DNA from host restriction systems. *Nat. Commun.* **2019**, *10* (1), No. 5442.
(110) Peng, Y.; Lu, Z.; Pan, D.; Shi, L. D.; Zhao, Z.; Liu, Q.; Zhang, C.; Jia, K.; Li, J.; Hubert, C. R. J.; Dong, X. Viruses in deep-sea cold seep sediments harbor diverse survival mechanisms and remain genetically conserved within species. *ISME J.* **2023**, *17* (10), 1774−1784.
(111) Zimmerman, A. E.; Howard-Varona, C.; Needham, D. M.; John, S. G.; Worden, A. Z.; Sullivan, M. B.; Waldbauer, J. R.; Coleman, M. L. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* **2020**, *18* (1), 21−34.
(112) Wang, Y.; Pan, J.; Yang, J.; Zhou, Z.; Pan, Y.; Li, M. Patterns and processes of free-living and particle-associated bacterioplankton and archaeaplankton communities in a subtropical river-bay system in South China. *Limnol. Oceanogr.* **2020**, *65* (S1), S161−S179.
(113) Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **2007**, *5* (10), 801−812.
(114) Dart, E.; Fuhrman, J. A.; Ahlgren, N. A. Diverse marine T4-like cyanophage communities are primarily comprised of low-abundance species including species with distinct seasonal, persistent, occasional, or sporadic dynamics. *Viruses* **2023**, *15* (2), No. 581, DOI: 10.3390/v15020581.
(115) Hevroni, G.; Flores-Uribe, J.; Béjà, O.; Philosof, A. Seasonal and diel patterns of abundance and activity of viruses in the Red Sea. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (47), 29738−29747.
(116) Cai, F.; Axen, S. D.; Kerfeld, C. A. Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. *RNA biology* **2013**, *10* (5), 687−693.
(117) Zhao, F.; Zhang, X.; Liang, C.; Wu, J.; Bao, Q.; Qin, S. Genome-wide analysis of restriction-modification system in unicellular and filamentous cyanobacteria. *Physiol. Genomics* **2006**, *24* (3), 181−190.
(118) Mühling, M.; Fuller, N. J.; Millard, A.; Somerfield, P. J.; Marie, D.; Wilson, W. H.; Scanlan, D. J.; Post, A. F.; Joint, I.; Mann, N. H. Genetic diversity of marine Synechococcus and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environ. Microbiol.* **2005**, *7* (4), 499−508.



https://doi.org/10.1021/acs.est.5c12277
Environ. Sci. Technol. XXXX, XXX, XXX−XXX