



# Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters

M. Consuelo Gazitúa<sup>1,2</sup> · Dean R. Vik<sup>1</sup> · Simon Roux<sup>3</sup> · Ann C. Gregory<sup>1</sup> · Benjamin Bolduc<sup>1</sup> · Brittany Widner<sup>4,5</sup> · Margaret R. Mulholland<sup>4</sup> · Steven J. Hallam<sup>6</sup> · Osvaldo Ulloa<sup>7</sup> · Matthew B. Sullivan<sup>1,8</sup>

Received: 23 April 2020 / Revised: 30 September 2020 / Accepted: 27 October 2020 / Published online: 16 November 2020  
© The Author(s) 2020. This article is published with open access

## Abstract

Viruses play an important role in the ecology and biogeochemistry of marine ecosystems. Beyond mortality and gene transfer, viruses can reprogram microbial metabolism during infection by expressing auxiliary metabolic genes (AMGs) involved in photosynthesis, central carbon metabolism, and nutrient cycling. While previous studies have focused on AMG diversity in the sunlit and dark ocean, less is known about the role of viruses in shaping metabolic networks along redox gradients associated with marine oxygen minimum zones (OMZs). Here, we analyzed relatively quantitative viral metagenomic datasets that profiled the oxygen gradient across Eastern Tropical South Pacific (ETSP) OMZ waters, assessing whether OMZ viruses might impact nitrogen (N) cycling via AMGs. Identified viral genomes encoded six N-cycle AMGs associated with denitrification, nitrification, assimilatory nitrate reduction, and nitrite transport. The majority of these AMGs (80%) were identified in T4-like *Myoviridae* phages, predicted to infect *Cyanobacteria* and *Proteobacteria*, or in unclassified archaeal viruses predicted to infect *Thaumarchaeota*. Four AMGs were exclusive to anoxic waters and had distributions that paralleled homologous microbial genes. Together, these findings suggest viruses modulate N-cycling processes within the ETSP OMZ and may contribute to nitrogen loss throughout the global oceans thus providing a baseline for their inclusion in the ecosystem and geochemical models.

---

These authors contributed equally: M. Consuelo Gazitúa, Dean R. Vik

**Supplementary information** The online version of this article (<https://doi.org/10.1038/s41396-020-00825-6>) contains supplementary material, which is available to authorized users.

✉ Matthew B. Sullivan  
sullivan.948@osu.edu

- <sup>1</sup> Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA
- <sup>2</sup> Viromica Consulting, Santiago, Chile
- <sup>3</sup> DOE Joint Genome Institute, Berkeley, CA, USA
- <sup>4</sup> Department of Ocean, Earth and Atmospheric Sciences, Old Dominion University, Norfolk, VA, USA
- <sup>5</sup> Woods Hole Oceanographic Institution, Woods Hole, MA, USA
- <sup>6</sup> Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada
- <sup>7</sup> Departamento de Oceanografía & Instituto Milenio de Oceanografía, Universidad de Concepción, Concepción, Chile
- <sup>8</sup> Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA

## Introduction

Earth's biogeochemical cycles are driven by microbial interaction networks, with significant contributions from the oceans [1, 2]. These networks and the distribution of metabolic pathways within them are modulated by environmental factors, grazing, and viral infections. Ocean viruses are abundant, kill ~20–40% of microbial cells per day, transfer genes between microbial hosts, and can more subtly impact host physiology through temperate and inefficiently lytic infections [3–9]. Virus-infected cells, termed virocells, have dramatically altered metabolic output [10–12], but can also directly impact biogeochemical cycling through virus-encoded auxiliary metabolic genes (AMGs) [9, 13–20]. Culture-based studies have shown that cyanophages (viruses of photosynthetic cyanobacteria) routinely encode core and ancillary photosynthesis genes. Virus-encoded AMGs are widespread among cyanophages with long latent periods [13], and can be expressed during infection [14–19] to enable the infecting virus to tailor host metabolism to its own needs. AMGs have been transferred to and from

viruses over evolutionary timescales [13, 20], and likely provide a fitness advantage to the virus during infection [21, 22]. Metagenomic surveys using high-throughput sequencing platforms and advanced assembly methods now provide sufficient genomic coverage for taxonomic identification [23, 24], which helped identify AMGs associated with functions beyond photosynthesis, including nearly all of central carbon metabolism [25], phosphate scavenging [26, 27], sulfur cycling [28, 29], and, most recently, nitrogen (N) cycling genes including PII (a global N regulator) and *amoC* (ammonia monooxygenase subunit C) [30, 31].

Notably, however, AMG studies in the ocean have largely focused on surface ocean waters where photosynthesis is the dominant source of organic carbon. This leaves AMGs in oxygen minimum zones (OMZs), which have large impacts on climate active trace gas, nutrient cycling, and fisheries productivity [32], much less studied. Permanent OMZs make up ~8% of the total ocean volume, often have high concentrations of nitrate and nitrous oxide, and account for up to 50% of oceanic fixed-nitrogen loss—with OMZ expansion altering surface ocean primary production [32–37]. Under the suboxic conditions that characterize OMZs, nitrate serves as an alternative electron acceptor, starting the denitrification pathway that reduces nitrate to N<sub>2</sub> gas [38]. In fact, most N cycling occurs in the absence of oxygen, including assimilatory and dissimilatory nitrate reduction and anaerobic ammonium oxidation (anammox) [38]. Because microorganisms that inhabit OMZs rely upon chemoautotrophy for organic carbon production and redox-coupled metabolisms that link N and sulfur cycling [32, 39, 40], we hypothesized that OMZ viruses manipulate N cycling via AMGs in ways that differ from viruses in oxic waters. This hypothesis has some support, as genes involved in ammonia assimilation, nitrate and nitrite ammonification, nitric oxide synthesis, and denitrification have been previously identified in “gene-resolved viromes” [41]. However, technological limitations at the time prevented these genes from being confirmed as viral in origin, which is critical for AMG studies given that cellular DNA is routinely encountered in viromes and can lead to rampant false discovery (see re-analyses of prior work presented in [23, 24]).

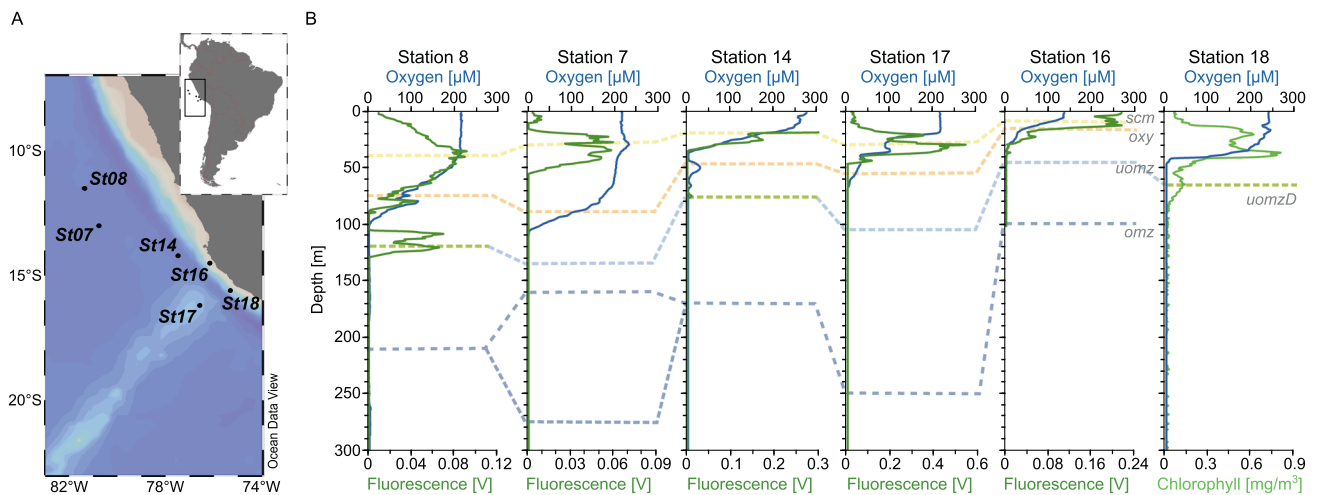
To further test the hypothesis that OMZ viruses manipulate N cycling via AMGs, we deeply-sequenced relatively quantitative viral particle metagenomes from the surface to OMZ waters of the Eastern Tropical South Pacific (ETSP) Ocean. Several N-cycle AMGs were identified, enabling us to contemplate the functional implications of viral reprogramming in relation to OMZ biogeochemistry.

## Materials and methods

### Sample collection

In total, 22 samples were collected from six stations in the ETSP OMZ region on December 31, 2014–January 22, 2015, during the R/V Atlantis cruise AT-2626, while traversing a transect from coastal to pelagic waters. A pump-profiling system was used to collect 20 L of seawater per sample. Oxygen concentrations per sample were measured with a nanoscale sensitive, STOX oxygen sensor. Samples for NO<sub>2</sub><sup>-</sup>, NO<sub>3</sub><sup>-</sup> + NO<sub>2</sub><sup>-</sup>, and NH<sub>4</sub><sup>+</sup> were collected with Niskin bottles and filtered using a 0.2-μm cartridge filter. The filtrate was collected into sterile Falcon™ tubes and stored upright at -20 °C until analysis. NO<sub>2</sub><sup>-</sup> and NO<sub>3</sub><sup>-</sup> + NO<sub>2</sub><sup>-</sup> concentrations were measured using an Astoria-Pacific autoanalyzer and standard colorimetric methods [42], and NH<sub>4</sub><sup>+</sup> was determined using fluorometric methods [43]. The limits of detection (LOD) for NO<sub>2</sub><sup>-</sup>, NO<sub>3</sub><sup>-</sup> + NO<sub>2</sub><sup>-</sup>, and NH<sub>4</sub><sup>+</sup> were 0.02 μM, 0.14 μM, and 10 nM (3σ, n = 7), respectively (Selden et al. submitted). Environmental features associated with each sample including oxygen, nutrient, and mineral concentrations can be found in Supplementary Table S1. Corresponding nitrate, nitrite, and ammonium concentrations were only available for half of the samples. The remaining concentrations were drawn from other samples within 10 m of the sampling depths from which our viromes were developed (denoted by an asterisk in Supplementary Table S1). Due to these sampling inconsistencies, the N-species measurements were only used qualitatively. The samples were selected corresponding to depth and habitat, including the surface chlorophyll maximum, suboxic upper oxycline, anoxic upper OMZ (with or without a deep chlorophyll maximum), and the core of the OMZ as indicated by measurements of oxygen and chlorophyll concentrations (Fig. 1 and Supplementary Table S1). A 0.2-μm filter (Millipore Express Plus; Millipore) was then used for each sample to remove cells and large debris. The filtrate from each sample was retained for subsequent viral concentration and DNA sequencing.

Iron chloride flocculation was used to concentrate the viral particles from each of the 22 samples [44, 45]. Viral particles were then resuspended in an ascorbic-EDTA buffer (0.1 M EDTA, 0.2 M MgCl<sub>2</sub>, 0.2 M ascorbic acid, pH 6.0). Free DNA was then removed from the viral concentrates using DNaseI at 100 U ml<sup>-1</sup> concentration [25]. A wizard DNA purification kit with 1 ml of resin to 0.5 ml of sample was then used to extract the concentrated viral DNA (Promega). A CsCl density gradient was used to further purify the viral DNA in samples with >1 μg DNA (surface chlorophyll maximum samples from stations 7, 8, 14, and 16, oxycline samples from stations 14 and 16, and the core



**Fig. 1** Map of the study area and vertical characterization of the sampling stations. **A** Location of stations 7, 8, 14, 16, 17, and 18, off the coast of Peru in the ETSP oxygen minimum zones (OMZ). The map was created with Ocean Data View (<http://odv.awi.de>). **B** Oxygen (solid blue) and fluorescence/chlorophyll (solid green, dark/light) depth profiles from each station. Fluorescence is reported instead of

chlorophyll for Station 18 due to differences in the sensors used during the collection of this sample. Sampling depths are indicated with dashed lines, connected by depth category: surface chlorophyll maximum (scm) in yellow, oxycline (oxy) in orange, upper OMZ with deep chlorophyll maximum (uomzD) in green, and without DCM (uomz) in light blue, and omz core (omz) in dark blue.

OMZ sample from station 16). Contigs derived from the CsCl-purified samples were only retained if they were the longest representative sequence of a population cluster [25]. The 22 DNase and CsCl-purified (where appropriate) samples were then used for downstream ecological analyses. Library preparation and sequencing were conducted at JGI using a Nextera kit and protocol, and an Illumina HiSeq 2000 platform.

### Assembly and processing

Data processing and metagenomic analyses were performed using high-memory computer nodes from the Ohio State Supercomputer Center [46]. Reads were split into paired and unpaired groups, adapter sequences were removed and low-quality sequencing regions below a Phred score threshold of 15, using a sliding window of four bases, were removed using a Trimmomatic version 0.33 [47], resulting in a mean read length of 149 bp. SPAdes version 3.11.1, using the `-meta -sc` and `-careful` option, was then used to assemble both paired and unpaired reads, from both samples purified with CsCl and those without, with k-mers of 21, 33, and 55 bases [48]. Population scale ecological groups were derived from these scaffolds using an in-house wrapper script for MUMmer with the nucmer package by clustering sequences at 95% ANI over 80% of the shorter sequence [49, 50]. Scaffolds that contained the same potential AMG sequence and shared an overlapping region >1 kb and 99% ANI were merged, and N-gaps in or nearby potential AMG genes were filled based on alignments against members of the same cluster. The edited scaffolds were mapped against

the reads from their corresponding sample and visualized with Integrative Genomics Viewer (IGV) [51] to address even distribution of the paired reads. The edited scaffolds were identified with the letter E at the end of their name.

### Viral identification and classification

Viruses were identified among the populations larger than 5 kbp or circular and larger than 1.5 kbp using the viral identification tools VirSorter and Virfinder [52, 53]. Populations with clearly identifiable viral or viral hallmark genes in VirSorter categories 1 and 2, and populations with a Virfinder score higher than 0.9 and a  $P$  value < 0.05 were considered to be viral. A final size threshold of 10 kbp was implemented. Taxonomic assignment of the AMG encoding contigs was established using vConTACT v2.0 with default settings on CyVerse [54] (<http://www.cyverse.org>).

### Viral relative abundance and distribution

The final selection of high-confidence viral populations larger than 10 kbp was concatenated into a single reference database, used with a custom wrapper script for bowtie2 to recruit quality trimmed reads from each sample, except for the CsCl treatment samples which were replaced with their corresponding non-CsCl counterparts to avoid treatment-specific effects [55]. An in-depth analysis of the viral community ecology from the same samples revealed no detectable correlation between the minimal variation of sequencing depth and the diversity in these samples [56]. The relative abundance of each population per sample was

then derived from the resulting bam files and converted into a relative abundance table using a custom wrapper script for BamM (<https://github.com/ecogenomics/BamM>). Coverage values as relative abundance proxies were calculated using the “tpmean” algorithm, normalized for the size of each metagenome in bases, and the length of each contig. Relative coverages were only reported for populations with read mapping coverage greater than 80% of contig length, and having at least 5× the coverage.

### AMG identification and annotation

Predicted genes were functionally annotated as done in Daly et al. [57]. Briefly, ORFs were predicted with prodigal v2.6.3 using the meta option [58]. Each predicted ORF was then screened against the KEGG, Uniref90, and InterPro database using USEARCH and Interproscan respectively [59–63]. The quality of annotation was then prescribed by a hierarchical ranking score from highest confidence to lowest confidence on a scale of A–E as follows. Annotations with a reciprocal blast hit (RBH) bitscore >350 to the KEGG database were given a score of A, RBH to the Uniref90 database with a bitscore >350 were given a score of B, the one-way blast hits to the Uniref90 database were given rank C, annotations with hits only to InterPro database were given a score of D, and those annotations with a bitscore of <60 to any database were given a rank of E [57]. Genes involved with nitrogen metabolism were then identified by manual curation of the functional annotations.

The nitrogen metabolism genes identified on viral populations were then submitted to a series of careful in silico validation steps in order to ensure that they were encoded on a bonafide viral sequence and that the functional annotations were correct. Only viral contigs larger than 10 kbp and encoding viral-like genes were reported. Where appropriate, BLAST-based homology searches and syntenic comparison with either viral or microbial references were conducted to identify the most related microbe and/or virus encoding similar metabolic and flanking genes. Genome maps and the systemic organization of related sequences were visualized using genbank files derived from NCBI for reference microbial and viral genomes, and PHANOTATE [64] for viral contigs, and the Easyfig package with tBLASTx (min. length 30 bp, max. *e*-value 0.001) [65].

Conserved residues and active sites for each predicted nitrogen metabolism-related AMG were identified using PROSITE (release 2019\_02 of 13-Feb-2019, default settings, <https://prosite.expasy.org/>) [66] and HHpred v2.0.13 (against PDB, default settings) [67]. Binding sites were checked manually when available, and promoter/terminator regions were predicted as done by Sullivan et al. [26], with some modifications. Briefly, ORF predictions were made using PHANOTATE [64], and manually refined where

alternate start sites were present, to maximize ORF size and coverage against reference genes. Bacterial sigma-70 promoters and terminators were predicted using BPROM (LDF > 2.75, Softberry, Mount Kisco, NY) [68] and TransTermHP (confidence score >90%) [69], respectively, using default parameters. Known cyanobacterial NtcA promoters were identified using the probabilistic model of NtcA-binding sites [70] that was more specifically adapted for use with marine cyanobacteria (5'-GTA-N8-TAC-3'; [71]). In addition to probability scoring cut-offs, all promoters or terminators also were required to be intergenic or within 10 bp of the start/stop of an ORF.

To avoid reporting erroneous functional annotations, based solely on sequence similarity searches, and to further support the possibility that the AMGs may be functional in the environment, we also predict the protein structure for each AMG. Secondary and tertiary structural homology searches were conducted for each AMG by first predicting the structure of the protein of interest with Phyre2 (version 2.0) [72] in expert batch submission mode. Predicted secondary structures with a 100% confidence score and alignment coverage above 70% were considered for further analyses. The most current version of SWISS-MODEL was then used to predict the quaternary structure of each protein with a Global Model Quality Estimation (GMQE) score above 0.5 [73]. Transmembrane domains were predicted with TMHMM [74].

Synonymous and non-synonymous mutations were calculated in order to determine the mode of selection acting on each protein. pN/pS values were calculated using the method from Schloissnig et al. [75], without reading coverage downsampling, and values of <0.3 were interpreted to indicate strong purifying selection.

### Phylogenetic tree generation

A phylogenetic analysis was used to further investigate the evolutionary origin of the AMGs. Sequences from this dataset were compared with the NCBI nr database [76] (blastp, the cutoff of 50 on bitscore, and 0.001 on *e*-value) to recruit closely related sequences and to add nonviral context to the phylogenetic trees. The best blast hits of each AMG, together with reference microbial sequences and previously described viral sequences (for *amoC* and *glnK*), were included in the final dataset for phylogenetic analysis. The multiple alignments and trees were built using ete3 toolkit v3.1.1 [77] with the eggno41 pipeline, i.e., multiple alignments computed with the built-in metaligner function, automatic alignment trimming with trimAL [78], automatic model selection with ProtTest [79], and tree built with Phym1 [80] with Chi<sup>2</sup>-based parametric branch supports. Trees were visualized using the ITOL (v3) online server [81].

## Results and discussion

### The ETSP virome dataset and overview of discovered N-related AMGs

A total of 29 viral metagenomes were sequenced from 22 samples that spanned oxic to anoxic waters across six depth profiles in the ETSP (Fig. 1A, B). The resultant 210 Gb dataset averaged ~49 M reads per virome and yielded 61,700 non-redundant, >5 kb scaffolds. Of these, 46,127 (75%) were identified as viral and 3,589 (6%) as microbial, while the remaining 11,984 scaffolds (19%) could not be identified confidently [56]. Collectively these pooled viruses recruit between 19 and 53% of the total reads, and the microbial contigs between 2 and 19%, with the rest representing unidentified contigs, or contigs smaller than 5 kb (Supplementary Fig. S1).

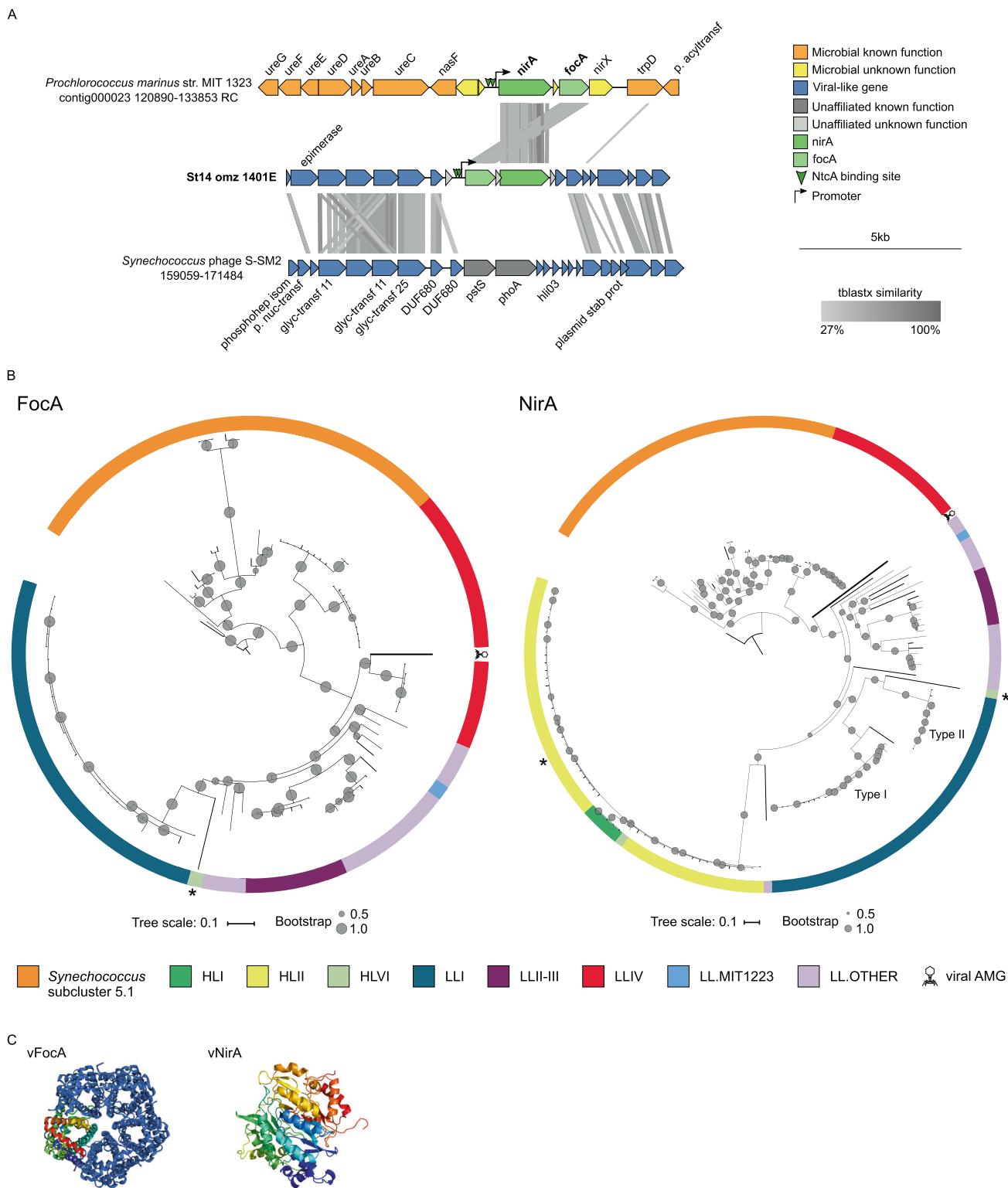
These high-confidence viral genomic scaffolds were screened for AMGs involved in N cycling and regulation. This screen was conducted using: (I) homology-based comparisons using USEARCH and Interproscan with reference databases (KEGG, Uniprot90, Interpro), (II) analyses for conserved residues using PROSITE and HHpred, and (III) structural modeling using PHYRE2 for tertiary structures and SWISS-MODEL for quaternary structures when applicable (details in “Materials and methods”). Together these analyses revealed six AMGs—*focA*, *nirA*, *nirK*, *norB*, *amoC*, and *glnK*—, of which only two, the nitrogen regulator PII gene encoded by *glnK* and the ammonia monooxygenase encoded by *amoC*, were previously known as AMGs. Below we describe the genomic contexts and taxonomic origins of these AMGs, assess their evolutionary histories and potential ecological roles and functionality, and map their distributions within the ETSP OMZ.

#### Ferredoxin-nitrite reductase and a nitrite transporter

The first and second of the six N-cycle AMGs included a nitrite transporter gene (*focA*) and a ferredoxin-nitrite reductase gene (*nirA*), neither of which had been previously detected in viral genomes. Both genes were colocalized on the viral scaffold St14\_omz\_1401E (11,826 bp) that contained 20 genes, 15 of which were viral-like according to VirSorter [52], and other features ordered as follows: two predicted NtcA-binding sites, a promoter, *focA*, an unknown gene and *nirA*, though with no terminator predicted (Fig. 2A). Three lines of evidence suggest this viral scaffold originated from a T4-like cyanomyophage. First, glycosyltransferases and hypothetical protein genes surrounding *focA* and *nirA* were most closely related to those from the T4-like cyanomyophage *Synechococcus* phage S-SM2, and 5 of these genes were enriched in *Synechococcus* and/or *Prochlorococcus* T4-like phages [26]

(Fig. 2A and Supplementary Table S2). Moreover, nine viral scaffolds with the most similar abundance profile as the *focA-nirA*-encoding scaffold, had a gene composition, including viral hallmark and structural genes, also associated to cyanophages, and synteny with *Synechococcus* phage S-SM2, suggesting that they all represent genome fragments of the same cyanophage population (Supplementary Fig. S2). Second, high-confidence gene-sharing network clustering (*sensu* vConTACT 2.0. [54]) placed this viral scaffold with the T4-like cyanophages (Supplementary Table S3). Consistent with this, the viral FocA and NirA were most similar to homologs in *Prochlorococcus* and *Synechococcus* (Fig. 2B, Supplementary Figs. S3 and S4), with identities of 86 and 67%, respectively, suggesting the AMGs were derived from and may function during infection in these hosts, as was observed for cyanophage photosynthesis AMGs (e.g., *hliP/psbA/psbD* [13, 14, 16]). Third, the putative regulation of these genes by NtcA is a common feature in cyanobacteria, where it activates genes involved in nitrogen transport and assimilation, including the *nirA* operon, heterocyst differentiation and acclimation to nitrogen starvation [82–86] (Fig. 2A).

Phylogenetic and comparative genomic analyses of these genes suggested a complicated evolutionary history. *focA* likely serves niche-defining functions in *Prochlorococcus* and *Synechococcus* as it was absent from most surface water metagenomes where nitrite is not abundant [87], and mostly absent in *Prochlorococcus* sps. from high-light (HL) clades [88]. Among cyanobacteria, regions encoding nitrate assimilation genes do not have conserved composition or order (synteny) [86]. However, the presence of these genes in multiple closely related cyanobacteria indicates that they can be horizontally transferred, perhaps even with viral versions being transferred back into the cyanobacterial genome, as observed for *psbA* [13]. These genes' mobility is also supported by the loss and subsequent gain of *nirA* in *Prochlorococcus* HLII strains (marked in yellow in Fig. 2B and Supplementary Fig. S4), where phages may have mediated these changes given the proximity of integrase genes to *nirA* in the genome of *Prochlorococcus* MIT0604 [87]. Furthermore, evidence for *focA* and *nirA* mobility has also been suggested for the high-light adapted AG-363-P06 single cell from the HLVI clade, which has recently acquired the nitrite assimilation cassette from a low-light adapted *Prochlorococcus* [88] (Fig. 2B). Only two (out of eleven) cyanobacterial nitrate assimilation gene arrangements encoded both *nirA* and *focA*, disposed in the opposite order from the OMZ viral genomic scaffold (Fig. 2A). While the variable synteny prevents the robust reconstruction of the HGT events that led to these genes being encoded by phages, it is possible that the different placement of these AMGs in single-gene trees reflects an independent acquisition of each gene, likely from *Prochlorococcus* sps. from LLIV clade (Fig. 2B).



**Fig. 2 Genomic context, diversity, and protein structure of viral *focA* and *nirA*.** **A** Genetic map of the scaffold encoding *nirA* and *focA* and its alignment to a reference cyanobacterial genome and a reference cyanophage genome. Detailed annotation of the ETSP viral contig can be found in Supplementary Table S2. **B** Maximum-likelihood trees from amino-acid alignments of the viral FocA or NirA found in ETSP and cyanobacterial sequences. Branches from viral AMGs found in this study are highlighted with thick lines. Internal nodes and SH-like

supports are represented by proportional circles (all nodes with support <0.50 were collapsed). Asterisks indicate *Prochlorococcus* sequences where horizontal gene transfer of *nirA* and *focA* (AG-363-P06 single cell (HLVI clade)) and *nirA* (*Prochlorococcus* MIT0604 (HLII clade)) have been proposed (from refs. [87, 88]). Colors represent *Synechococcus* subcluster 5.1, and *Prochlorococcus* high-light (HL) and low-light (LL) adapted clades. **C** Quaternary structure of viral FocA and tertiary structure of viral NirA.

Evidence of function for these viral AMGs stems from in silico observations of the proteins. For FocA, known features (six transmembrane domains) and a conserved and highly charged C-terminal regulatory region [89, 90] were identified (Supplementary Fig. S5A). The viral protein was similar (86% amino-acid identity) to that in *Prochlorococcus marinus* MIT1323 and *Prochlorococcus* sp. MIT0701 (LLIV) (Fig. 2B), and the structural model was predicted with 100% confidence to be a formate/nitrite transporter (Fig. 2C and Supplementary Table S4). For NirA, known features that were identified included two nitrite/sulfite reductase ferredoxin-like half domains and a nitrite/sulfite reductase 4Fe–4S-binding site [84] (Supplementary Fig. S5B). The viral protein was similar (67% amino-acid identity) to a functionally active *Synechococcus* NirA (Fig. 2B), and the structural model was again predicted with 100% confidence to be a NirA (Fig. 2C and Supplementary Table S4). In addition, the calculated ratio of non-synonymous to synonymous polymorphisms (pN/pS) for *focA* was 0, with seven single-nucleotide polymorphisms (SNPs) identified, and 0.25 for *nirA*, with 27 identified SNPs, indicating strong purifying selection (pN/pS ratios <0.3; Supplementary Table S5) as would be expected for a gene encoding a functional protein.

Functionally, it is plausible that the acquisition of *nirA* and *focA* genes could benefit OMZ viruses. NirA is involved in assimilatory nitrite reduction (nitrite reduction to ammonia; [91]), whereas FocA is a nitrite transporter from the formate/nitrite family [92]. NirA was initially described in cyanobacteria but is widely distributed among eukaryotic algae and vascular plants [91], as well as other bacteria and archaea, as observed in the NCBI database, while FocA is found in bacteria, predominantly Proteobacteria, archaea, fungi, algae, and parasites [93]. In OMZs, the lack of oxygen requires the use of alternative electron acceptors, with nitrate being the most energetically favorable. Nitrate is reduced by denitrifiers to nitrite, which in turn accumulates in OMZs and can fuel anaerobic ammonia oxidation [40]. Presumably then, if functional, viral FocA and NirA would be advantageous during infection by reducing their host's need to compete for limited nitrate and ammonia.

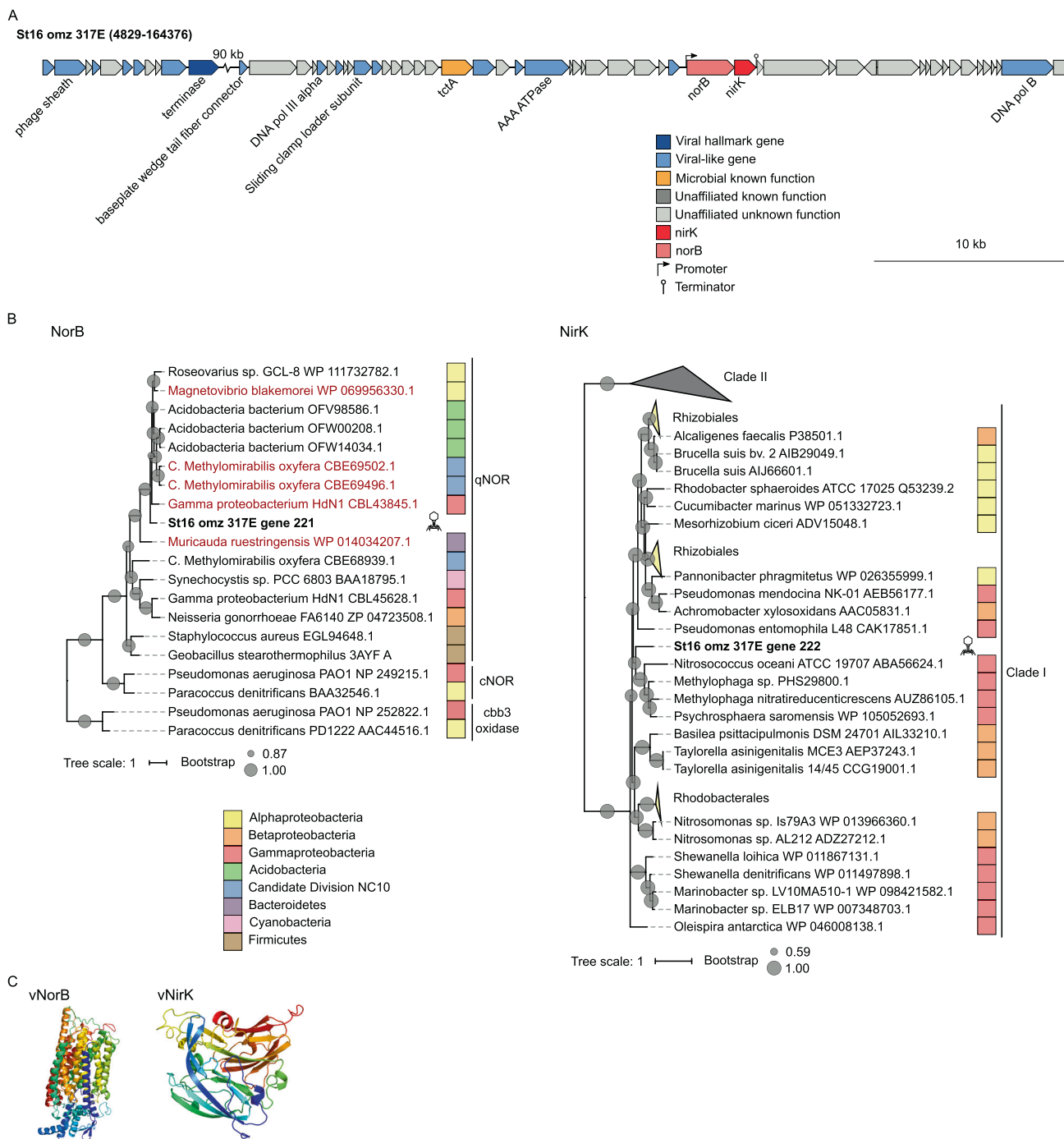
Finally, two additional features suggested that the viral *nirA* and *focA* are transcribed during host infection. First, both genes were in the same transcriptional unit, similar to what has been reported for *Prochlorococcus* MIT9313, where *nirA*, *focA*, and the uncharacterized gene between them are co-expressed [85]. Second, two NtcA-binding sites followed by sigma-70 promoters were observed upstream of these genes (Supplementary Table S6). In *Prochlorococcus*, these regulatory features were present and more numerous in this same region (triplicate in MIT9313 and duplicate in MIT1323, Supplementary Table S6), which suggests the downstream *nirA* and *focA* genes are likely to be active.

Similarly, we posit that the binding sites for NtcA and sigma 70 promote an efficient expression of viral *focA* and *nirA* during infection.

### Nitrite reductase and nitric oxide reductase

The third and fourth of the six N-cycle AMGs included a copper-containing nitrite reductase gene *nirK* and a nitric oxide reductase gene *norB*, neither of which had been previously detected in viral genomes. Similar to the *focA* and *nirA* AMGs, *nirK* and *norB* were contiguous, with *norB* immediately followed by *nirK*, though also with a predicted promoter and terminator (Fig. 3A), which might indicate the kind of transcriptionally autonomous unit (termed a “moron”) that is a hallmark of phage genome evolution [94]. The original viral scaffold containing these new AMGs, St16\_omz\_317 (52,903 bp), had 68 genes, 22 of them identified as viral-like by VirSorter. Two of these genes corresponded to T4 core genes (gp44 clamp loader subunit and gp43 DNA polymerase), and one to a non-cyanophage T4 core gene (gp9 baseplate wedge tail fiber connector) (Fig. 3A and Supplementary Table S2). Despite the low percent of genes annotated as viral in this scaffold, the high proportion of genes with no affiliation, viral or otherwise, together with the low rate of strand switching (i.e., change of coding strand between two consecutive genes), are also an indicator of a viral sequence, reinforcing the viral origin of this scaffold [52]. The *norB* / *nirK* region likely represents a viral genomic island where host DNA accumulates akin to that in T4-like cyanomyophages [26, 95] as it is surrounded by genes that encode proteins of unknown function, proteins involved in protein biosynthesis and modification, and in tricarboxylate transport (Supplementary Table S2). Taxonomic annotation of the viral-like genes suggested the virus is a Myovirus (Supplementary Table S2), though it is likely a new genus as it formed its own separate viral cluster in gene-sharing networks (Supplementary Table S3) that resolve genus-level taxonomy [54, 96].

The viral NirK protein clustered with microbial NirK from Clade I [97], being most closely related to *Gamma-proteobacteria* (67–70% protein identity; Fig. 3B), while NorB was most closely related to *Acidobacteria* (57–62% protein identity), *Magnetovibrio blakemorei* (62% protein identity), and the *Gammaproteobacteria* strain HdN1 (59.3% protein identity) (Fig. 3B). These genes could have been acquired by the virus together or, more likely, in separate events, as the order and orientation of the genes is different from that of any known microbial genome. Interestingly, mobility has been suggested for both of these genes, with *nirK* reported in plasmids from *Azospirillum brasilense* [98], and *norB* present in microbial contigs of four different phylotypes in metagenomes from ETNP



**Fig. 3 Genomic context, diversity, and protein structure of viral *norB* and *nirK*.** **A** Genetic map of the scaffold encoding *norB* and *nirK*. Detailed annotation of this contig can be found in Supplementary Table S2. **B** Maximum-likelihood trees from amino-acid alignments of viral NorB or NirK found in ETSP and reference microbial sequences. The first tree represents the heme-copper oxidases superfamily, including cytochrome C oxidase cbb3-type (cbb3 oxidase),

cytochrome c-dependent nitric oxide reductase (cNORs), and quinol-dependent nitric oxide reductases (qNORs) including the potential NO dismutases (in red) (from ref. [101]). Viral AMGs found in this study are highlighted in bold. Internal nodes and SH-like supports are represented by proportional circles (all nodes with support <0.50 were collapsed). **C** Tertiary structures of viral NorB and viral NirK.

OMZ, one of which contained plasmid competence genes upstream of *norB* [99]. The presence of these genes in viral genomes defines an adjunct agent for transferring niche-defining traits in the oxygen-deficient water column.

In silico, both genes appeared to be functional (Supplementary Table S4). Viral NirK contained two copper-binding motifs and key active site residues required for nitrite reducing activity [97] (Supplementary Fig. S6A) and



had a strongly supported (100% confidence) structural prediction as a copper-containing nitrite reductase (Fig. 3C and Supplementary Table S4). Viral NorB had similarity (59% protein identity) to a NOD-like NorB from *Gamma-proteobacteria* strain HdN1, which is a proposed protein family of length similar to that of the canonical NorB, with 12+ transmembrane domains, but with altered quinol binding and active sites [100, 101]. These sites in the viral “NorB” were also more similar to those from NOD-like NorB enzymes than the canonical NorB (Supplementary Fig. S6B). In contrast, the structural model prediction of the viral “NorB” was of high (100%) confidence to be a nitric oxide reductase (Fig. 3C and Supplementary Table S4), but the absence of a crystal structure of the NOD-like NorB enzyme hindered the discrimination between reductase and dismutase activity using this tool. Interestingly, numerous NOD-like NorB sequences were found in ETNP OMZ waters, with a maximum at the same depth as the second N<sub>2</sub>O peak (140 m) [99]. Since canonical NorB was absent at this same depth, it is suggested that some of these NOD-like NorB enzymes retain their function as nitric oxide reductases [99]. Thus, we interpret this viral NorB to be a NOD-like NorB, with its reductase and/or dismutase activity yet to be determined. Moreover, the pN/pS value for the viral *norB* gene was 0, with 2 SNPs identified, and no SNPs were identified for *nirK* (from one sample with minimal gene coverage >10×) (Supplementary Table S5), suggesting that both genes are likely functional and under strong purifying selection.

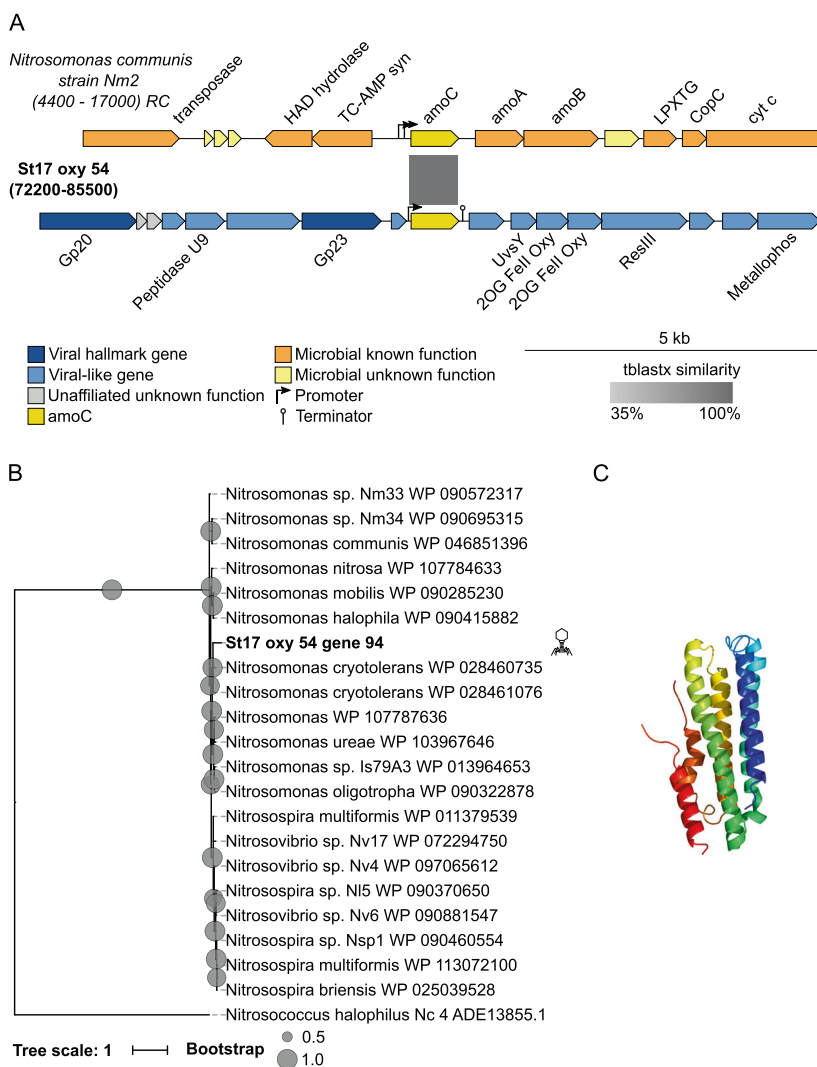
Functionally, both AMGs are denitrification genes that could be of benefit to the viruses, with the NirK protein reducing nitrite to nitric oxide, and NorB reducing nitric oxide to nitrous oxide (canonical NorB), or nitric oxide to nitrogen and oxygen (postulated nitric oxide dismutase, NOD-like NorB [101]). Denitrifying bacteria vary in the extent of denitrification genes they encode: some genomes contain all denitrifying genes to be able to reduce nitrate all the way to nitrogen gas, whereas others contain subsets of these genes that enable utilization of specific niches and interactions among partial denitrifiers in a community setting. Thermodynamically, complete and partial denitrification are both theoretically favorable, with Gibbs-free energy ( $\Delta G^\circ$ ) of  $-532$  kJ/mol for the complete pathway,  $-131$  kJ/mol for nitrate reduction to nitrite,  $-105$  kJ/mol for nitrite reduction to nitric oxide,  $-139$  kJ/mol for nitric oxide reduction to nitrous oxide, and  $-314$  kJ/mol for nitrous oxide reduction to nitrogen gas (calculated with pyruvate as an electron donor, standardized to the reduction of one mol of the electron acceptor, Supplementary Table S7). From a bioenergetic perspective, though, the proton-motive force generated from complete versus partial denitrification is the same, at most six protons per pair of electrons, since the nitrate reductase is the only denitrification module that

translocates protons [102]. However, if denitrifiers were able to make use of nitric oxide dismutation, the amount of energy conserved would improve, translocating 7.3 protons per electron pair, leading to 36.5% energy conservation [102]. Thus, we posit that viral NirK and NorB could benefit the host during infection through the exploitation of specific niche-defining genes or help to safeguard energetically vital pathways in low-redox environments, while the host genome is degraded over the course of infection, thus compensating for the likely elevated energetic cost associated viral production. Additionally, if the viral NorB functions as a nitric oxide dismutase, the increase in energy conservation associated with nitric oxide dismutation would significantly improve the host’s fitness. Future efforts focused on energy budgets and mining SAGs (Single-cell Amplified Genomes) could help resolve these and related hypotheses.

### Ammonia monooxygenase subunit C

The fifth N-related AMG identified was ammonia monooxygenase subunit C gene (*amoC*). In addition to an archaeal version that has been observed in the Tara Oceans Global Ocean Virome (GOV) dataset [30], a recent study described 15 new Thaumarchaeota virus populations that encode viral capsid and thaumarchaeal *amoC* genes. These are potentially tailed viruses that share a common ancestor with related marine Euryarchaeota viruses and are globally distributed in various marine habitats, including OMZs [31]. We found two archaeal-like *amoC* genes and one bacterial-like *amoC* in the ETSP OMZ dataset. These archaeal viruses, represented by scaffolds St14\_oxy\_254 (35,926 bp) and St17\_scm\_137 (42,459 bp), corresponded to species “C” and “G” of the previously mentioned 15 new *Thaumarchaeota* virus population [31], and thus will not be described in this section. The bacterial *amoC*-containing viral population, represented by the scaffold St17\_oxy\_54 (144,618 bp) appeared to be a complete, circularly permuted, T4-like phage genome of 144 kb that encoded 179 genes (Supplementary Table S2). Whole-genome alignments of this scaffold against its closest reference phage, *Pelagibacter* phage HTVC008M, revealed a high degree of homology and synteny between them (Supplementary Fig. S7). It also contained 52 of the 60 hypothesized “core T4” genes [26] (BLASTp *e*-value <0.001). As a reference, 50 hypothesized “core T4” genes were found in *Pelagibacter* phage HTVC008M, with the remaining 10 genes either absent or miss-annotated. The bacterial *amoC* was flanked by a promoter and a terminator and was also located next to the major capsid protein gene Gp23, which is a region known to be an AMG hotspot in T4-like phages (Fig. 4A) [26]. Not surprisingly, this viral contig was clustered with T4-like phages by the gene-sharing network

**Fig. 4 Genomic context, diversity, and protein structure of viral *amoC*.** **A** Genetic map of the viral scaffold encoding the bacterial-like *amoC*, and alignment to a reference microbial genome containing this gene. Detailed annotation of the viral scaffold can be found in Supplementary Table S2. **B** A maximum-likelihood tree from an amino-acid alignment of the bacterial-like viral AmoC found in ETSP and reference microbial sequences. The viral AMG found in this study is bolded. Internal nodes and SH-like supports are represented by proportional circles (all nodes with support <0.50 were collapsed). **C** Tertiary structure of the bacterial-like viral AmoC.



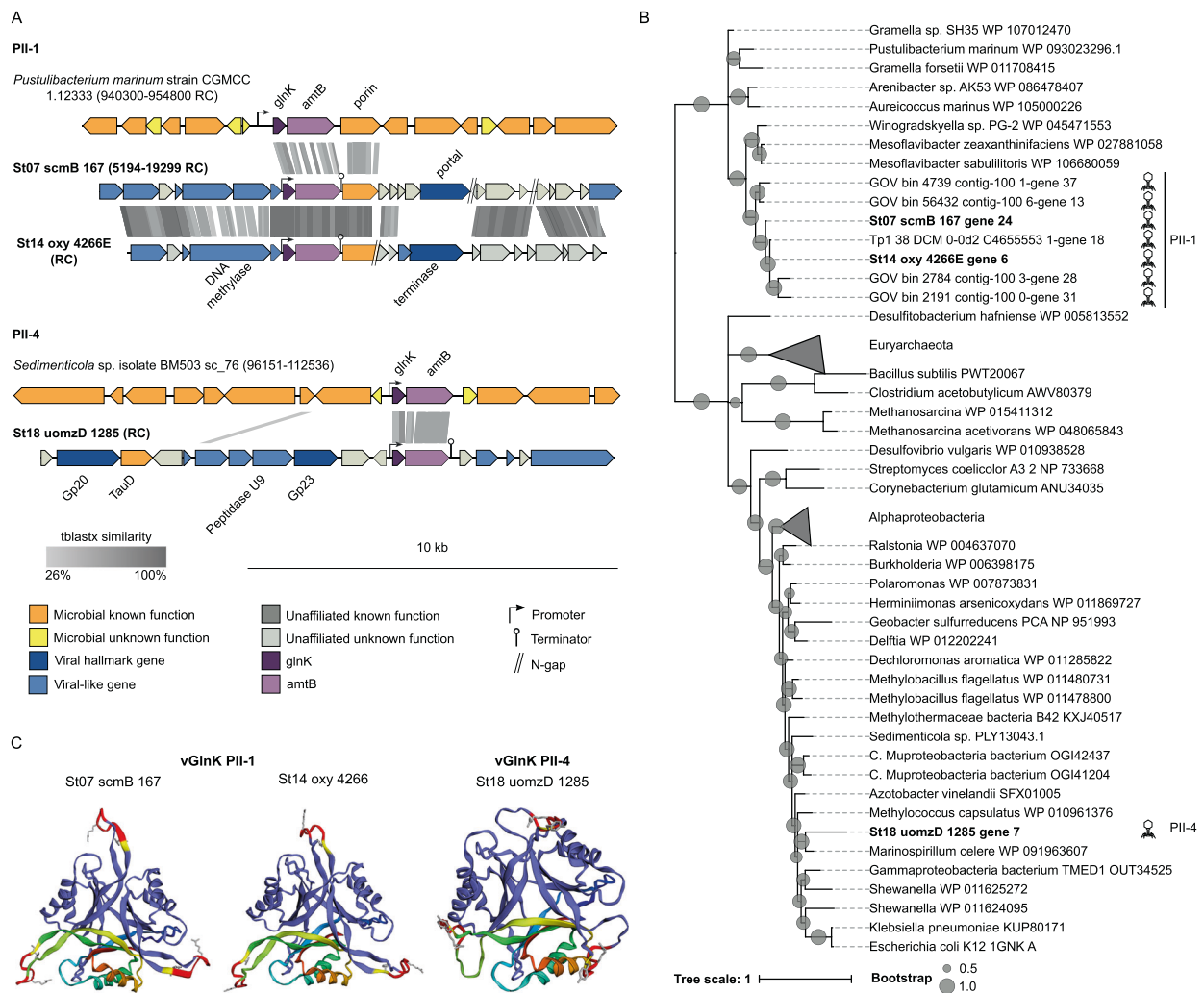
analyses (Supplementary Table S3). Evolutionarily, the bacterial-like *amoC* clusters within the *Nitrosomonas* clade (Fig. 4B), and was most closely related to AmoC from *Nitrosomonas communis*, with 86% amino-acid identity.

Functionally, as posited previously [30], *amoC* could benefit OMZ viruses by enabling access to energy liberated by ammonia oxidation. Across diverse archaea and bacteria, ammonia monooxygenase performs the first step and rate-limiting step in nitrification, aerobically oxidizing ammonia to either nitroxyl or hydroxylamine, respectively [103]. These intermediaries are further oxidized to nitrite, either spontaneously or with the help of additional oxidoreductases. While both ammonia-oxidizing bacteria and archaea are present in the ETSP OMZ [104–106], metatranscriptomic and metaproteomic data suggest that the ammonia-oxidizing archaea, mostly represented by *Nitrosopumilus maritimus*, dominate nitrification in the OMZ [106–108]. The resultant nitrite either continues through nitrification to nitrate in aerobic waters or is converted to

nitric oxide or nitrogen gas via denitrification or nitrogen gas via anammox pathways in deeper suboxic waters [33, 109]. While no conserved residues for this enzyme have been described, the high sequence similarity with microbial AmoC suggests that these viral AmoC are likely functional. The pN/pS ratio for the archaeal-like viral *amoC* was 0 for St17\_scm\_137, with two identified SNPs, and 0.13 for St14\_oxy\_254, with nine identified SNPs. No SNPs were identified for the bacterial-like *amoC* (from two samples with minimal gene coverage >10×) (Supplementary Table S5), showing a strong purifying selection of the three viral *amoC* genes.

### GlnK PII regulator

The last of the six N-related AMGs was GlnK, a PII signal transduction protein also previously observed in viruses in the GOV dataset [30]. There are three major subgroups of microbial PII proteins, based on gene linkage and similarity



**Fig. 5 Genomic context, diversity, and protein structure of viral *glnK*.** **A** Genetic map of the scaffolds encoding *glnK*, and alignment to reference microbial genomes containing this gene. Detailed annotation of the viral contigs can be found in Supplementary Table S2. **B** A maximum-likelihood tree from an amino-acid alignment of viral GlnK

found in ETSP and reference microbial sequences. Viral AMGs found in this study are highlighted in bold. Internal nodes and SH-like supports are represented by proportional circles (all nodes with support <0.50 were collapsed). **C** Tertiary structure of viral GlnK.

at the amino-acid sequence level: *glnB* linked to glutamine synthetase (*glnA*) or NAD synthetase (*nadE*), *glnK* linked to ammonia channel protein (*amtB*), and *nifI* linked to the nitrogenase operon (reviewed in refs. [110] and [111]). Phylogenies from the prior GOV work found viral representatives of the *glnK* and *glnB* subgroups (designated PII-1 and PII-2, respectively), and a functionally ambiguous PII protein that lacked a conserved motif (designated PII-3). PII-1 was widely distributed throughout mesopelagic waters, while PII-2 and PII-3 were geographically restricted to the North Pacific and the South Atlantic Oceans, respectively [30]. We identified three viral populations containing the *glnK* gene followed by the *amtB* gene in the ETSP OMZ dataset (Fig. 5A). The first two belonged to the previously described PII-1 group, and the third was a new viral *glnK* that we designated as PII-4 (Fig. 5B). The first

viral scaffold encoding *glnK*, St07\_scmB\_167 (35,108 bp), contained 66 genes, 26 of them considered viral-like by VirSorter, including one hallmark gene annotated as a portal protein. The second viral scaffold, St14\_oxy\_4266E (12,943 bp), had 19 genes, with 8 of them corresponding to viral-like genes, including one hallmark gene (terminase). In both cases, there was a predicted promoter upstream of *glnK* and a terminator downstream of the ammonium transporter gene (*amtB*), and a porin gene (*ompL*) immediately downstream of *amtB* (Fig. 5A). These viral contigs clustered together in the gene-sharing network analyses (Supplementary Table S3), but no reference viral genome was assigned to this cluster. The third viral population, represented by St18\_uomzD\_1285 (15,491 bp), had 18 genes, two of them hallmark viral genes (the major capsid protein Gp23 and the capsid assembly protein Gp20), and 9

viral-like genes. The *glnK* gene was close to the Gp23 gene, similar to the bacterial-like *amoC* viral contig, and a promoter and terminator flanking the *glnK* and *amtB* genes were also predicted. This contig formed a singleton in the gene-sharing network analysis, but the presence of the hallmark structural proteins previously mentioned strongly suggests that it's a T4-like phage.

Evolutionarily, the *glnK-amtB* genes have been captured by viruses at least two times, one from *Bacteroidetes*, where the new and previously described PII-1 viral *glnK* form a distinct clade, and the second time from *Gammaproteobacteria*, where PII-4 is located (Fig. 5B). The association of *glnK* and *amtB* has been reported as one of the most ancient in biological history [112], with an early lateral transfer event between bacteria and archaea, as proposed by the PII evolution model [111]. Despite the ancient origin of the PII-Amt association, found widespread in Bacteria and Archaea, the PII-Amt-OmpL-like association found in our PII-1 viruses have only been found in *Flavobacteriales* (Fig. 5A). This suggests that the viruses acquired these three genes from a member of *Bacteroidetes*, and it is most likely that their hosts belong to this phylum. The viral porin OmpL had a conserved beta-barrel structure, suggesting that it is functional, though its substrate is unknown. The *glnK* and *amtB* genes from PII-4 were most closely related to *Gammaproteobacteria*, with 68% protein identity, and in this case, no extra genes were incorporated by the phage (Fig. 5A).

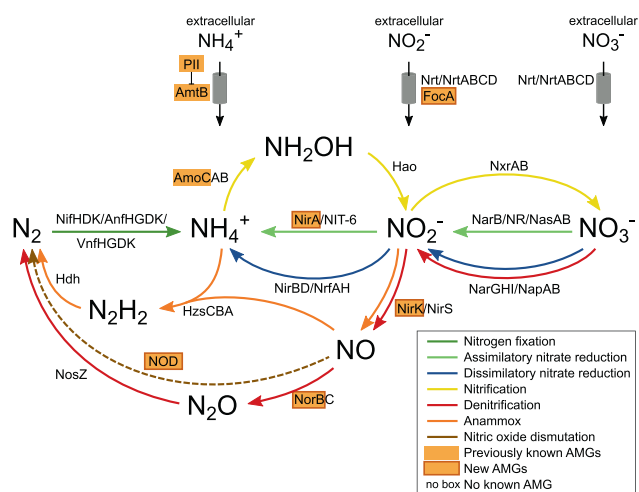
In silico, the *glnK* and *amtB* genes from the three viral contigs were identified as PII and Amt, respectively (Supplementary Table S2). The first two PII proteins (PII-1 group) were mostly related to the order *Flavobacteriales* PII, with 71–72% protein identity. The third PII protein (PII-4 group) was mostly related to *Gammaproteobacteria* PII, with 68% protein identity (Fig. 5B). Conserved in the viral GlnK proteins was the uridylylation residue Y51, located in the T-loop, (Fig. 5C and Supplementary Fig. S8), and the tertiary and quaternary structures, predicted with 100 confidence to be PII (Table S4). Unfortunately, pN/pS ratios could not be calculated for the *glnK* genes due to low gene coverage across the samples.

Functionally, OMZ viruses could benefit from the *glnK* and *amtB* genes by regulating the ammonia uptake of their host. AmtB has 11 transmembrane domains and an “N-terminus out” and “C-terminus in” topology [113, 114], and GlnK folds into a four-stranded  $\beta$ -sheet packed against two helices, forming three loops designated the B-, C-, and T-loops [115]. Both form homotrimeric structures, and together they form a complex where the GlnK trimer binds to the cytoplasmic face of AmtB by inserting the T-loop into the cytoplasmic pore exit of an adjacent AmtB subunit, therefore closing the channel [115]. Under nitrogen-limiting conditions, GlnK is covalently modified by uridylylation,

inhibiting the interaction with AmtB and therefore enabling the uptake of ammonium. Conversely, under conditions of fixed-nitrogen excess, GlnK is de-uridylylated, enabling its interaction with the ammonium transporter to inhibit ammonium uptake [116, 117]. Ammonium concentration in OMZ decreases with depth (Supplementary Table S1), due to its consumption for cell growth, aerobic oxidation, and anaerobic oxidation. A viral version of the transporter and its regulator could be beneficial, especially in the oxic-anoxic interface, by complementing or replacing the host's version, thereby increasing its nutrient uptake during infection, as reported in *Ostreococcus tauri* infected by the *O. tauri* virus RT-2011 [118]. The viral GlnK might also compete with the host's PII regulator on its union to the AmtB channel, preventing its closure and therefore ensuring ammonium uptake during infection.

### The biogeochemical and ecological context of virus-encoded N-cycling AMGs

The paradigm emerging from studies of viral AMGs is that viruses randomly sample host genetic material, but only a subset of these genes are retained in the viral genomes [119]. In the case of N-cycling AMGs, nitrification (*AmoC*) and ammonia regulation (*GlnK*) genes have been previously identified, and to these, we add genes encoding proteins mediating assimilatory nitrite reduction (*FocA* and *NirA*) and denitrification (*NirK*, *NorB* and/or NOD-like *NorB*). While these genes were scattered throughout known N-cycling pathways (Fig. 6), we posit that the specific AMGs observed, which are only a subset of known N-cycling genes, are those that represent nutrient or energetic bottlenecks during infection across myriad virus–host pairs in nature. For example, the highly elevated abundances of these genes (particularly *nirK*) in the dysoxic oxycline regions may be consistent with the functioning of these genes in microaerobic environments where some organisms have mechanisms to switch between aerobic respiration and denitrification pathways. The activation of each of these nitrite reduction genes requires elaborate transcriptional regulatory systems [120]. In the case of *nirA*, a nitrite assimilation regulatory gene, it may be that the specific virus–host interaction is augmented by the regulation of the host nitrite reduction pathways, while in the case of *nirK* the host regulatory pathways are most likely implemented for activation of nitrite reduction as no other regulatory regions were apparent in the virus. While many denitrification steps require specific electron carriers, such as ubiquinone, menaquinone, and cytochromes, none were observed in the phages. However, cyanophages that contain photosynthesis AMGs directly linked to electron flow often completely lack any of the related photosynthesis electron carriers [26], so this may simply represent parts of the metabolic machinery



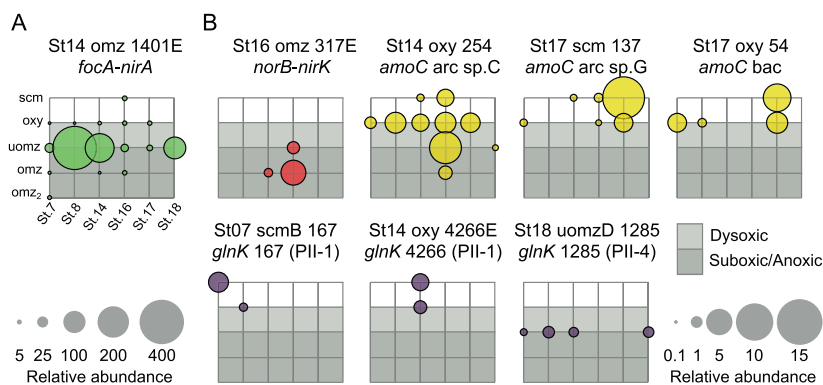
**Fig. 6 Potential contribution of viruses to nitrogen cycling and transport.** The schematic represents the main pathways that drive the nitrogen cycle and the participating enzymes and transporters. Proteins with a viral version are highlighted in orange boxes. Viral NorB and NOD correspond to the same protein that is homologous to both nitric oxide reductase and nitric oxide dismutase. Nitrogen uptake is performed by the ammonia transporter (AmtB, which is regulated by PII), the MFS-type nitrate/nitrite transporter (Nrt), the ABC-type nitrate transporter (NrtABCD), and the formate/nitrite transporter (FocA). The enzymes that transform nitrogen are molybdenum-iron nitrogenase (NifHDK), iron-iron nitrogenase (AnfHGDK), vanadium-iron nitrogenase (VnfHGDK), an ammonia monooxygenase (AmoCAB), hydroxylamine dehydrogenase (Hao), nitrite oxidoreductase (NxrAB), ferredoxin-nitrate reductase (NarB), nitrate reductase (NAD(P)H) (NR), assimilatory nitrate reductase (NasAB), ferredoxin-nitrite reductase (NirA), nitrite reductase (NAD(P)H) (NIT-6), membrane-bound nitrate reductase (NarGHI), periplasmic nitrate reductase (NapAB), nitrite reductase (NADH) (NirBD), cytochrome *c* nitrite reductase (NnfAH), copper-containing nitrite reductase (NirK), heme-containing nitrite reductase (NirS), nitric oxide reductase (NorBC), nitrous oxide reductase (NosZ), hydrazine synthase (HzsCBA), hydrazine dehydrogenase (Hdh), and nitric oxide dismutase (NOD).

that are not prone to turnover and remain intact during viral infection.

The N-related AMGs were distinctly distributed across the ETSP water column in patterns that mirror those of the key microbial N-metabolisms being modulated (Fig. 7). Some, like the *focA-nirA*-containing viral population, were present at nearly every depth and station, but considerably more abundant in the samples from the upper OMZ that had a deep chlorophyll maximum (normalized coverage of 97.7–382.8, compared to normalized coverage of 1.2–16.7 in the remaining samples with *focA-nirA*-containing viral population) (Fig. 7A). This viral population predicted to be a cyanophage (see above), has abundances notably coincident with those of novel and uncultivated *Prochlorococcus* cyanobacteria that predominate in the secondary chlorophyll maximum of anoxic oceanic OMZs (e.g., Arabian Sea, ETNP, and ETSP) [121–123]. Nitrite reduction is expected to occur in these upper OMZ samples, where nitrite concentration ranges from 0.45 to 11.36  $\mu\text{M}$

(Supplementary Table S1). The other AMG encoding viral populations were less abundant across the dataset and demonstrated strong depth preferences likely corresponding to specific N-related microbes and metabolisms (Fig. 7B). The *norB-nirK*-containing viral population was only found in anoxic waters from stations 14 and 16. As posited earlier, nitrite reduction is expected at these depths, where nitrite concentration varied from 0.45 to 11.36  $\mu\text{M}$  in upper OMZ samples, and from 0.27 to 9.02  $\mu\text{M}$  in OMZ core samples (Table S1), with most of these samples showing nitrite accumulation (>0.5  $\mu\text{M}$ ), which occurs only when oxygen falls below 50 nM [124]. Though denitrifiers and anammox bacteria are expected to be found at these depths [33, 40], they were not predicted as hosts for this viral population due to the scant relevant reference genomes for these groups. Similarly, *amoC*-containing viral populations were distributed mostly in surface and oxycline waters, which largely followed the distribution of aerobic ammonia oxidizers [106, 125] (Fig. 7B). Ammonium concentrations at these oxic depths support this first nitrification step, with concentrations that ranged from 75.79 to 779.68 nM in the surface samples, and from 3.48 to 891.01 nM in the oxycline samples. One of the archaeal-like *amoC*-containing viral populations, St14\_oxy\_254, was also present in OMZ samples from station 16, and the upper OMZ sample from station 18. Even though ammonium concentrations in these samples support ammonium oxidation (from 12.52 to 24.1 nM, Supplementary Table S1), this reaction is not expected in anoxic waters. However, these particular samples were collected at shallower depths (45 and 100 m in station 16, and 66 m in station 18), compared to the rest of the OMZ and upper OMZ samples (Supplementary Table S1), and might be prone to water mixing and intrusion of overlying oxygenated waters. Finally, the *glnK*-containing viral populations (Fig. 7B) were either present in surface and/or oxycline waters from a few stations (PII-1; stations 7 and 8 for St07\_scm\_167 and station 14 for St14\_oxy\_4226) or exclusively in upper OMZ waters (PII-4; stations 7, 8, 14 and 18 for St18\_uomzD\_1285). As posited earlier, ammonium is present in surface and oxycline samples, and in lower concentrations in upper OMZ samples, supporting the presence of PII-1 and PII-4 *glnK*-containing viral populations, respectively (Supplementary Table S1). Considering the wide distribution of the *glnK-amtB* genes among Bacteria and Archaea, a particular distribution of these genes across the oxygen gradient is not expected. However, the most closely related microbial *glnK* genes belonged to Proteobacteria (for PII-4) and Bacteroidetes (for PII-1), which represent the first and second most abundant phyla in the OMZs [32].

Stepping back, these findings may have implications for OMZs beyond viruses. First, reduced nitrogen assimilation is known in many cyanobacteria, with nearly all lineages of



**Fig. 7** Distribution of N-AMG-containing viral populations across the ETSP OMZ samples. Bubble plots representing the relative abundances, in terms of normalized coverage, of viral populations containing the *focA* and *nirA* genes (panel A, in green), *norB* and *nirK* genes (panel B, in red), *amoC* genes (panel B, in yellow) and the *glnK* genes (panel B, in purple). The *x* axis of each grid represents the stations (7, 8, 14, 16, 17, and 18), and the *y* axis represents the

sampling depths (from top to bottom: surface chlorophyll maximum (scm), oxycline (oxy), upper OMZ (uomz) and core of the OMZ (omz)). Station 7 had a second core OMZ sample (omz2) and station 18 was only sampled in the upper OMZ. Gray boxes represent the OMZ: light gray for dysoxic waters below the oxycline, and dark gray for suboxic and anoxic waters in the upper and core of the OMZ.

*Prochlorococcus* encoding *nirA*. The potentially high mobility of genes involved with nitrogen assimilation may be justified by the relatively high concentration and energetic favorability of nitrite (relative to ammonia and nitrate) in this system [126]. Here, we have further identified a cyanophage carrying *focA* and *nirA*, suggesting that N assimilation is also of value to the viruses infecting the novel cyanobacteria from the anoxic secondary chlorophyll maximum [121–123, 126]. Second, this viral manipulation of the denitrification pathway, via NirK, NorB, and/or NOD-like NorB, underlines the relevance of this pathway in anoxic waters. Though the host for the virus carrying these genes could not be predicted, it might be in future studies if parallel microbial metagenomes and single-cell amplified genomes (SAGs) were generated [29]. In this case, we speculate that *Gammaproteobacteria* or *Planctomycetes* might be the hosts, and if so, infection by viruses encoding nitrogen cycling AMGs would promote the denitrification pathway by the reduction of nitrite to nitrous oxide, or the anammox pathway by the reduction of nitrite to nitric oxide, respectively. In any case, infection with the *norB-nirK*-containing virus could potentially alter the nitrite, nitric oxide, and nitrous oxide levels within the OMZ, and would have important implications for those trying to assess climate change feedbacks resultant from changes to these ecosystems [127, 128].

In summary, understanding how viruses alter N-related biogeochemical cycling in OMZs is critical, considering the expansion of these suboxic and anoxic water masses and their effects in surface primary production, greenhouse gas emission, and fixed-nitrogen loss [32–34]. Our findings imply that OMZ viruses impact N cycling not only through lysis of key N-cycling microbes but also by modulating diverse N-metabolisms

during infection. Such infected “virocells” [10] would be drastically altered in their metabolic capacity and biogeochemical outputs as has been shown now in several environmental model virus–host systems [10, 12, 129]. With these N-related virus AMGs now uncovered, future OMZ virus work can evaluate virocell-impacted nitrogen cycling, as well as develop primer sets for “viral” vs “cellular” versions to differentially quantify the biogeochemical impacts of viruses in OMZ N-cycling genes and transcripts. As standardized practices emerge for viral ecogenomics [130–132], they are enabling the development of global maps of ocean viruses [30, 49, 133] that can be integrated into multi-organism ecological studies [134]. Together these efforts to understand virus-mediated nutrient cycling in climate-critical environments, along with parallel efforts on land (e.g., thawing permafrosts [135, 136]), are now providing quantitative information needed to incorporate viruses into predictive models [137].

## Data availability

All high-quality reads and assembled contigs are available on iVirus (CyVerse, <https://doi.org/10.25739/mmj5-kt58>). Requests for further information should be directed to Matthew B. Sullivan at [sullivan.948@osu.edu](mailto:sullivan.948@osu.edu).

**Acknowledgements** We thank Sullivan Lab members and Heather Maughan for comments on the paper, Bess Ward for her contribution in the N-cycle context of our story, Kurt Hanselmann for his assistance in the calculations of the Gibbs-free energies, and the scientific party and crew of the R/V Atlantis (grant OCE-1356056 to MRM) for the sampling opportunity and support at sea. This work was funded in part by awards from the Agouron Institute to OU and MBS, a Gordon and Betty Moore Foundation Investigator Award (#3790) and NSF

Biological Oceanography Awards (#1536989 and #1829831) to MBS, and the Millennium Science Initiative (grant ICN12\_019-IMO) to OU. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science*. 2008;320:1034–9.
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*. 1998;281:237–40.
- Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature*. 1999;399:541–8.
- Wommack KE, Colwell RR. Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev*. 2000;64:69–114.
- Suttle CA. Viruses in the sea. *Nature*. 2005;437:356–61.
- Suttle CA. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol*. 2007;5:801–12.
- Rohwer F, Thurber RV. Viruses manipulate the marine environment. *Nature*. 2009;459:207–12.
- Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J*. 2017;11:1511–20.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nat Microbiol*. 2018;3:754–66.
- Forterre P. The virocell concept and environmental microbiology. *ISME J*. 2013;7:233–6.
- Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, et al. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol*. 2020;18:21–34.
- Rosenwasser S, Ziv C, Creveld SG, van, Vardi A. Virocell metabolism: metabolic innovations during host–virus interactions in the ocean. *Trends Microbiol*. 2016;24:821–32.
- Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol*. 2006;4:e234.
- Clokier MRJ, Shan J, Bailey S, Jia Y, Krisch HM, West S, et al. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol*. 2006;8:827–35.
- Dammeyer T, Bagby SC, Sullivan MB, Chisholm SW, Frankenberg-Dinkel N. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr Biol*. 2008;18:442–8.
- Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*. 2005;438:86–9.
- Sharon I, Tzahor S, Williamson S, Shmoish M, Man-Aharonovich D, Rusch DB, et al. Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J*. 2007;1:492–501.
- Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci*. 2011;108:E757–64.
- Fridman S, Flores-Urbe J, Larom S, Alalouf O, Liran O, Yacoby I, et al. A myovirus encoding both photosystem i and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nat Microbiol*. 2017;2:1350–7.
- Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA*. 2004;101:11013–8.
- Bragg JG, Chisholm SW. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLoS ONE*. 2008;3:e3550.
- Hellweger FL. Carrying photosynthesis genes increases ecological fitness of cyanophage in silico. *Environ Microbiol*. 2009;11:1386–94.
- Roux S, Krupovic M, Debroas D, Forterre P, Enault F. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open Biol*. 2013;3:130160.
- Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit M-A. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *ISME J*. 2017;11:237–47.
- Hurwitz BL, Hallam SJ, Sullivan MB. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol*. 2013;14:R123.
- Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol*. 2010;12:3035–56.
- Zeng Q, Chisholm SW. Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Curr Biol*. 2012;22:124–8.
- Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. Sulfur oxidation genes in diverse deep-sea viruses. *Science*. 2014;344:757–60.
- Roux S, Hawley AK, Torres Beltran M, Scofield M, Schwientek P, Stepanauskas R, et al. Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell and meta-genomics. *eLife*. 2014;3:e03125.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*. 2016;537:689–93.
- Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. *ISME J*. 2019;13:618–31.

32. Wright JJ, Konwar KM, Hallam SJ. Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol*. 2012;10:381–94.
33. Lam P, Lavik G, Jensen MM, van de Vossenberg J, Schmid M, Wobken D, et al. Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci USA*. 2009;106:4752–7.
34. Ward BB, Devol AH, Rich JJ, Chang BX, Bulow SE, Naik H, et al. Denitrification as the dominant nitrogen loss process in the Arabian Sea. *Nature*. 2009;461:78–81.
35. Paulmier A, Ruiz-Pino D. Oxygen minimum zones (OMZs) in the modern ocean. *Prog Oceanogr*. 2009;80:113–28.
36. Kock A, Arevalo-Martinez DL, Loscher CR, Bange HW. Extreme N<sub>2</sub>O accumulation in the coastal oxygen minimum zone off Peru. *Biogeosciences*. 2016;13:827–40.
37. Stramma L, Johnson GC, Sprintall J, Mohrholz V. Expanding oxygen-minimum zones in the tropical oceans. *Science*. 2008;320:655–8.
38. Lam P, Kuypers MMM. Microbial nitrogen cycling processes in oxygen minimum zones. *Ann Rev Mar Sci*. 2011;3:317–45.
39. Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF, et al. A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science*. 2010;330:1375–8.
40. Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ. Microbial oceanography of anoxic oxygen minimum zones. *Proc Natl Acad Sci USA*. 2012;109:15996–6003.
41. Cassman N, Prieto-davó A, Walsh K, Silva GGZ, Angly F, Akhter S, et al. Oxygen minimum zones harbour novel viral communities with low diversity. *Environ Microbiol*. 2012;14:3043–65.
42. Parsons TR, Maita Y, Lalli CM. A manual of biological and chemical methods for seawater analysis. Oxford: Pergamon Press; 1984.
43. Holmes RM, Aminot A, Kérouel R, Hooker BA, Peterson BJ. A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can J Fish Aquat Sci*. 1999;56:1801–8.
44. John SG, Mendez CB, Deng L, Poulos B, Kauffman AKM, Kern S, et al. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep*. 2011;3:195–202.
45. Duhaime MB, Sullivan MB. Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology*. 2012;434:181–6.
46. Ohio Supercomputer Center. Ohio Supercomputer Center. Columbus OH: Ohio Supercomputer Center; 1987.
47. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
48. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. 2017;27:824–34.
49. Brum JR, Ignacio-Espinoza JC, Roux S, Doucier G, Acinas SG, Alberti A, et al. Patterns and ecological drivers of ocean viral communities. *Science*. 2015;348:1261498.
50. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5:R12.
51. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.
52. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ*. 2015;3:e985.
53. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*. 2017;5:69.
54. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol*. 2019;37:632–9.
55. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
56. Vik D, Gazitúa MC, Sun CL, Aldunate M, Mulholland MR, Ulloa O, et al. Genome-resolved viral ecology in a marine oxygen minimum zone (OMZ). *Environ Microbiol*. 2020; in press.
57. Daly RA, Borton MA, Wilkins MJ, Hoyt DW, Kountz DJ, Wolfe RA, et al. Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nat Microbiol*. 2016;1:1–9.
58. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma*. 2010;11:119.
59. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30.
60. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*. 2015;31:926–32.
61. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47:D351–60.
62. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
63. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics*. 2014;30:1236–40.
64. Mcnair K, Zhou C, Dinsdale EA, Souza B, Edwards RA. PHANOTATE: a novel approach to gene identification in phage genomes. *Bioinformatics*. 2019;35:4537–42.
65. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics*. 2011;27:1009–10.
66. Sigrist CJA, De Castro E, Cerutti L, Cuče BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. *Nucleic Acids Res*. 2013;41:D344–7.
67. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*. 2018;430:2237–43.
68. Solovyev V, Salamov A. Automatic annotation of microbial genomes and metagenomic sequences. In: Li RW, editor. *Metagenomics and its applications in agriculture, biomedicine, and environmental studies*. Hauppauge, NY: Nova Science Publisher's; 2011. p. 61–78.
69. Kingsford CL, Ayanbule K, Salzberg SL. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol*. 2007;8:R22.
70. Su Z, Olman V, Mao F, Xu Y. Comparative genomics analysis of NtcA regulons in cyanobacteria: Regulation of nitrogen assimilation and its coupling to photosynthesis. *Nucleic Acids Res*. 2005;33:5156–71.
71. Su Z, Mao F, Dam P, Wu H, Olman V, Paulsen IT, et al. Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium *Synechococcus* sp. WH 8102. *Nucleic Acids Res*. 2006;34:1050–65.
72. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–58.
73. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46:W296–303.



74. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol.* 2001;305:567–80.
75. Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A, et al. Genomic variation landscape of the human gut microbiome. *Nature.* 2013;493:45–50.
76. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2018;46:D8–13.
77. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 2016;33:1635–8.
78. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–3.
79. Cuff JA, Birney E, Clamp ME, Barton GJ. ProtEST: protein multiple sequence alignments from expressed sequence tags. *Bioinformatics.* 2000;16:111–6.
80. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307–21.
81. Letunic I, Bork P. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics.* 2007;23:127–8.
82. Luque I, Flores E, Herrero A. Molecular mechanism for the operation of nitrogen control in cyanobacteria. *EMBO J.* 1994;13:2862–9.
83. Herrero A, Muro-Pastor AM, Flores E. Nitrogen control in cyanobacteria. *J Bacteriol.* 2001;183:411–25.
84. Bird C, Wyman M. Nitrate/nitrite assimilation system of the marine picoplanktonic cyanobacterium *Synechococcus* sp. strain WH 8103: effect of nitrogen source and availability on gene expression. *Appl Environ Microbiol.* 2003;69:7009–18.
85. Tolonen AC, Aach J, Lindell D, Johnson ZI, Rector T, Steen R, et al. Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol Syst Biol.* 2006;2:53.
86. Ohashi Y, Shi W, Takatani N, Aichi M, Maeda SI, Watanabe S, et al. Regulation of nitrate assimilation in cyanobacteria. *J Exp Bot.* 2011;62:1411–24.
87. Berube PM, Biller SJ, Kent AG, Berta-Thompson JW, Roggensack SE, Roache-Johnson KH, et al. Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *ISME J.* 2015;9:1195–207.
88. Berube PM, Rasmussen A, Braakman R, Stepanauskas R, Chisholm SW. Emergence of trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. *eLife.* 2019;8:e41043.
89. Maeda S, Murakami A, Ito H, Tanaka A, Omata T. Functional characterization of the FNT family nitrite transporter of marine picocyanobacteria. *Life.* 2015;5:432–46.
90. Hunger D, Röcker M, Falke D, Lilie H, Sawers RG. The C-terminal six amino acids of the FNT channel FocA are required for formate translocation but not homopentamer integrity. *Front Microbiol.* 2017;8:1616.
91. Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol.* 1999;181:6573–84.
92. Lü W, Du J, Schwarzer NJ, Wacker T, Andrade SLA, Einsle O. The formate/nitrite transporter family of anion channels. *Biol Chem.* 2013;394:715–27.
93. Wang Y, Huang Y, Wang J, Cheng C, Huang W, Lu P, et al. Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature.* 2009;462:467–72.
94. Juhala RJ, Ford ME, Duda RL, Youton A, Hatfull GF, Hendrix RW. Genomic sequences of bacteriophages HK97 and HK022: Pervasive genetic mosaicism in the lambdaoid bacteriophages. *J Mol Biol.* 2000;299:27–51.
95. Perez Sepulveda B, Redgwell T, Rihtman B, Pitt F, Scanlan DJ, Millard A. Marine phage genomics: the tip of the iceberg. *FEMS Microbiol Lett.* 2016;363:fnw158.
96. Bolduc B, Jang HBin, Doucier G, You Z-QQ, Roux S, Sullivan MB. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ.* 2017;5:e3243.
97. Helen D, Kim H, Tytgat B, Anne W. Highly diverse nirK genes comprise two major clades that harbour ammonium-producing denitrifiers. *BMC Genomics.* 2016;17:155.
98. Pothier JF, Prigent-Combaret C, Haurat J, Moëgne-Loccoz Y, Wisniewski-Dyé F. Duplication of plasmid-borne nitrite reductase gene nirK in the wheat-associated plant growth-promoting rhizobacterium *Azospirillum brasilense* Sp245. *Mol Plant Microbe Interact.* 2008;21:831–42.
99. Fuchsman CA, Devol AH, Saunders JK, McKay C, Rocap G. Niche partitioning of the N cycling microbial community of an offshore oxygen deficient zone. *Front Microbiol.* 2017;8:2384.
100. Ehrenreich P, Behrends A, Harder J, Widdel F. Anaerobic oxidation of alkanes by newly isolated denitrifying bacteria. *Arch Microbiol.* 2000;173:58–64.
101. Ettwig KF, Speth DR, Reimann J, Wu ML, Jetten MSM, Keltjens JT. Bacterial oxygen production in the dark. *Front Microbiol.* 2012;3:273.
102. Chen J, Strous M. Denitrification and aerobic respiration, hybrid electron transport chains and co-evolution. *Biochim Biophys Acta - Bioenerg.* 2013;1827:136–44.
103. Hooper AB, Vannelli T, Bergmann DJ, Arciero DM. Enzymology of the oxidation of ammonia to nitrite by bacteria. *Antonie Van Leeuwenhoek.* 1997;71:59–67.
104. Molina V, Ulloa O, Farias L, Urrutia H, Ramirez S, Junier P, et al. Ammonia-oxidizing Beta-proteobacteria from the oxygen minimum zone off Northern Chile. *Appl Environ Microbiol.* 2007;73:3547–55.
105. Molina V, Belmar L, Ulloa O. High diversity of ammonia-oxidizing archaea in permanent and seasonal oxygen-deficient waters of the eastern South Pacific. *Environ Microbiol.* 2010;12:2450–65.
106. Stewart FJ, Ulloa O, Delong EF. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol.* 2012;14:23–40.
107. Hawley AK, Brewer HB, Norbeck AD, Paša-Tolic L, Hallam SJ. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc Natl Acad Sci.* 2014;111:11395–400.
108. Ganesh S, Bristow LA, Larsen M, Sarode N, Thamdrup B, Stewart FJ. Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone. *ISME J.* 2015;9:2682–961.
109. De Brabandere L, Canfield DE, Dalsgaard T, Friederich GE, Revsbech NP, Ulloa O, et al. Vertical partitioning of nitrogen-loss processes across the oxic-anoxic interface of an oceanic oxygen minimum zone. *Environ Microbiol.* 2013;16:3041–54.
110. Leigh JA, Dodsworth JA. Nitrogen regulation in bacteria and archaea. *Annu Rev Microbiol.* 2007;61:349–77.
111. Sant’Anna FH, Trentini DB, De Souto Weber S, Cecagno R, Da Silva SC, Schrank IS. The PII superfamily revised: a novel group and evolutionary insights. *J Mol Evol.* 2009;68:322–36.
112. Thomas G, Coutts G, Merrick M. The glnKamTB operon: a conserved gene pair in prokaryotes. *Trends Genet.* 2000;16:11–4.
113. Khademi S, O’Connell IIIJ, Remis J, Robles-Colmenares Y, Miercke LJ, Stroud RM. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. *Science.* 2004;305:1587–94.

114. Zheng L, Kostrewa D, Berneche S, Winkler FK, Li X-D. The mechanism of ammonia transport based on the crystal structure of AmtB of *Escherichia coli*. *Proc Natl Acad Sci USA*. 2004;101:17090–5.
115. Conroy MJ, Durand A, Lupo D, Li X-D, Bullough PA, Winkler FK, et al. The crystal structure of the *Escherichia coli* AmtB-GlnK complex reveals how GlnK regulates the ammonia channel. *Proc Natl Acad Sci USA*. 2007;104:1213–8.
116. Van Heeswijk WC, Hoving S, Molenaar D, Stegeman B, Kahn D, Westerhoff HV. An alternative PII protein in the regulation of glutamine synthetase in *Escherichia coli*. *Mol Microbiol*. 1996;21:133–46.
117. Atkinson MR, Ninfa AJ. Characterization of the GlnK protein of *Escherichia coli*. *Mol Microbiol*. 1999;32:301–13.
118. Monier A, Chambouvet A, Milner DS, Attah V, Terrado R, Lovejoy C, et al. Host-derived viral transporter protein for nitrogen uptake in infected marine phytoplankton. *Proc Natl Acad Sci USA*. 2017;114:E7489–98.
119. Hurwitz BL, U'Ren JM. Viral metabolic reprogramming in marine ecosystems. *Curr Opin Microbiol*. 2016;31:161–8.
120. Hartsock A, Shapleigh JP. Mechanisms of oxygen inhibition of nirK expression in *Rhodobacter sphaeroides*. *Microbiology*. 2010;156:3158–65.
121. Lavin P, González B, Santibáñez JF, Scanlan DJ, Ulloa O. Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol Rep*. 2010;2:728–38.
122. Goericke R, Olson RJ, Shalapyonok A. A novel niche for *Prochlorococcus* sp. in low-light suboxic environments in the Arabian Sea and the Eastern Tropical North Pacific. *Deep Sea Res Part I Oceanogr Res Pap*. 2000;47:1183–205.
123. Cepeda-Morales J, Beier E, Gaxiola-Castro G, Lavín M, Godínez VM. Effect of the oxygen minimum zone on the second chlorophyll maximum. *Ciencias*. 2009;35:389–403.
124. Thamdrup B, Dalsgaard T, Revsbech NP. Widespread functional anoxia in the oxygen minimum zone of the Eastern South Pacific. *Deep Res Part I Oceanogr Res Pap*. 2012;65:36–45.
125. Ulloa O, Wright JJ, Belmar L, Hallam SJ. Pelagic oxygen minimum zone microbial communities. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. *The prokaryotes*. Berlin, Heidelberg: Springer; 2013. p. 113–22.
126. Aldunate M, Henríquez-Castillo C, Ji Q, Lueders-Dumont J, Mulholland MR, Ward BB, et al. Nitrogen assimilation in picocyanobacteria inhabiting the oxygen-deficient waters of the eastern tropical North and South Pacific. *Limnol Oceanogr*. 2020;65:437–53.
127. Zakem EJ, Al-Haj A, Church MJ, van Dijken GL, Dutkiewicz S, Foster SQ, et al. Ecological control of nitrite in the upper ocean. *Nat Commun*. 2018;9:1206.
128. Penn J, Weber T, Deutsch C. Microbial functional diversity alters the structure and sensitivity of oxygen deficient zones. *Geophys Res Lett*. 2016;43:9773–80.
129. Howard-Varona C, Lindback MM, Bastien GE, Solonenko N, Zayed AA, Jang H-B, et al. Phage-specific metabolic reprogramming of virocells. *ISME J*. 2020;14:881–95.
130. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, et al. Minimum information about an uncultivated virus genome (MIUVIG). *Nat Biotechnol*. 2019;37:29–37.
131. Roux S, Emerson JB, Eloë-Fadrosch EA, Sullivan MB. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*. 2017;5:e3817.
132. Roux S, Solonenko NE, Dang VT, Poulos BT, Schwenck SM, Goldsmith DB, et al. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ*. 2016;4:e2777.
133. Gregory AC, Zayed AA, Sunagawa S, Wincker P, Sullivan MB, Ferland J, et al. Marine DNA viral macro-and microdiversity from pole to pole. *Cell*. 2019;177:1109–23.
134. Ibarbalz FM, Henry N, Brandão MC, Martini S, Busseni G, Byrne H, et al. Global trends in marine plankton diversity across kingdoms of life. *Cell*. 2019;179:1084–97.
135. Trubl G, Jang HBin, Roux S, Emerson JB, Solonenko N, Vik DR, et al. Soil viruses are underexplored players in ecosystem carbon processing. *mSystems*. 2018;3:e00076–18.
136. Emerson JB, Roux S, Brum JR, Bolduc B, Woodcroft B, Jang HB, et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat Microbiol*. 2018;3:870–80.
137. Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, et al. Scientists' warning to humanity: microorganisms and climate change. *Nat Rev Microbiol*. 2019;17:569–86.