

# Virus-Host Interactions Reveal Potential Roles Towards Phosphorus Cycling in South Atlantic Ocean Euphotic Zones

**Percy Mutseka Lunga**

University of Pretoria

**Oliver K.I Bezuidt**

University of Pretoria

**Miho Hirai**

JAMSTEC: Kaiyo Kenkyu Kaihatsu Kiko

**Yoshihiro Takaki**

JAMSTEC: Kaiyo Kenkyu Kaihatsu Kiko

**Taichi Yokokawa**

JAMSTEC: Kaiyo Kenkyu Kaihatsu Kiko

**Takuro Nunoura**

JAMSTEC: Kaiyo Kenkyu Kaihatsu Kiko

**J. Colin Murrell**

University of East Anglia

**Thulani Makhalanyane** (✉ [thulani.makhalanyane@up.ac.za](mailto:thulani.makhalanyane@up.ac.za))

University of Pretoria <https://orcid.org/0000-0002-8173-1678>

---

## Research

**Keywords:** Virus-host, phosphorus cycling, Atlantic Ocean

**Posted Date:** November 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1021375/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Due to their role as obligate parasites of marine microorganisms, viruses are primary mediators of marine biogeochemical cycles. Recent studies have provided irrevocable evidence showing that viruses augment the metabolisms of bacteria and archaea through expression of auxiliary metabolic genes (AMGs). Several studies have shown that AMGs affect the biogeochemical recycling of sulphur and nitrogen but comparatively less is known regarding their influence on phosphorus recycling.

## Results

Here, we provide the first insights regarding the potential effects of phosphorus limitation and AMGs on putative prokaryotic hosts in the euphotic zone of the South Atlantic Ocean (SAO). We identified 7,176 viral contigs that were clustered into 5,999 viral operational taxonomic units (vOTUs, >5kb). These SAO viral communities appear to be unique, as over 89% had no taxonomic assignment, possibly due to the genetic endemism in this ocean. Three phosphatases, *phoN*, *gmhB* and *rnhA-cobC*, were identified as P-cycle AMGs in both prokaryotic double-stranded DNA viruses and eukaryotic Nucleocytoplasmic Large DNA viruses. These genes are associated with the acquisition of inorganic phosphate from phosphate esters, the largest reservoir of P-containing compounds in the marine environment. AMGs were identified in both uncultured and unclassified prokaryotic double-stranded DNA viruses predicted to infect *Bacteroidetes*, *Proteobacteria*, *Chloroflexota* and *Poseidonales* lineages.

## Conclusion

Together, these results suggest that viruses modulate P-cycling in euphotic zones of the ocean and that the acquisition of these phosphatase genes may be cues of P-ester stress.

## Background

Viruses are the most numerically abundant entities on Earth and substantially regulate the structure and function of microbial communities [1–3]. As obligate parasites of marine microorganisms, viruses play essential roles in marine environments [1, 4–6]. Previous studies have shown that viruses influence nutrient cycling, functional diversity of microbes and particle sinking rates across a diverse range of marine ecosystems [7–10]. The complexity of microbial interactions such as widespread horizontal gene transfer complicates our efforts to understand the precise impacts of viruses on ecosystem services [4, 11, 12].

There has been some evidence showing the genetic potential of viruses to augment microbial metabolism through auxiliary metabolic genes (AMGs) [13–17]. AMGs encoded by marine bacteriophages are metabolically diverse and include various processes such as photosynthesis and the sequestration of nitrogen, sulphur, and carbon [18–21]. Metabolic genes that are equally suggested to

augment host cellular processes including central nitrogen metabolism, iron and phosphorous uptake have also been reported for eukaryotic Nucleocytoplasmic Large DNA viruses (NCLDV) [22, 23]. The discovery of diverse AMGs and elucidating their metabolic control on microbiomes has become the subject of extensive research over the past few years [24–28]. In oligotrophic environments, viruses may maximize host fitness through the expression of AMGs, which have been shown to increase metabolic flexibility [29–31]. For instance, cyanophages in oligotrophic systems with particularly low phosphorus levels have been shown to encode alkaline phosphatase (*PhoA*) and high-affinity phosphate-binding (*pstS*) genes. These findings suggest that cyanophages play important regulatory roles linked to host responses to phosphate stress [32–34]. Moreover, genes for inorganic phosphate transporters (*Pho4*) have also reported to be encoded by NCLDVs that infect phytoplankton, further suggesting the overall importance and contributions of viral families towards host responses to nutrient uptake under limiting conditions [35]. Dissolved inorganic phosphate (DIP) is one of the key macronutrients, which limits microbial growth and primary production [36, 37]. In oligotrophic oceanic regions, P is assimilated as fractions of the dissolved organic phosphorus (DOP) that constitute P-esters, P-anhydrides, and phosphonates [38, 39]. This inventory of DOP exceeds the pool of inorganic phosphate which is a key chemical component of DNA, RNA and phospholipids [40, 41]. As a result microbial communities and viruses contribute substantially to P-acquisition and turnover [41, 42]. Oligotrophic plankton communities from the North Atlantic have been reported to replace membrane phospholipids with those lacking phosphate in response to P-starvation [43, 44]. In another study, conducted in the Mediterranean Sea, C-P-lyase complex (*phnGHIJKLMOP*), high-affinity inorganic phosphate (*pstSCAB*) and organophosphate transport operons (*phnCDE-phoU*) were prevalent amongst *Gammaproteobacteria*, *Alphaproteobacteria* and *Actinobacteria* lineages [32]. Understanding of host-cell reprogramming through AMGs has been focused on coastal and marine environments in the Pacific [45], Indian [46] and North Atlantic Ocean [33, 36]. However, we lack a mechanistic understanding of virus-host dynamics in other marine environments.

The South Atlantic Ocean (SAO) extends from the equator to 40° south and is delimited by the subtropical convergence [47]. The SAO is limited in phosphorus, with concentrations ranging from 70 - 150nM although, these levels are comparatively higher than those in the North Atlantic Ocean (NAO) [48]. This is likely due to extensive nitrogen fixation in the NAO, which results in phosphorous depletion, subsequently affecting microbial communities in the NAO [49–51]. However, we lack a clear understanding regarding the importance of viruses in biogeochemical cycling and in particular P acquisition. Previous studies suggest that the SAO is dominated by *Alphaproteobacteria*, *Gammaproteobacteria*, *Thaumarcheota* and *Actinobacteria* microbial lineages [52, 53]. Viruses linked to these microbial communities were mostly from members of the *Caudovirales* order. These viruses were shown to harbour AMGs involved in amino acid and photosynthetic metabolism [52]. An understanding of AMGs involved in P acquisition may help in identifying the effects of viruses on marine ecosystems.

We primarily characterized the host-virus linkages in euphotic zones of the SAO and used shotgun metagenomic analysis to predict host-virus interactions in the SAO and reveal their role in P-cycling through AMGs. In addition to providing a compendium of metagenome assembled genomes (MAGs) from these environments, we demonstrate that the SAO metagenomes harboured 27 P-related AMGs.

Several of these AMG were directly linked to assimilation of organic phosphorus. These genes included phosphatases, *gmbB* and *rhnA-cobC*, which have not previously been described as AMGs. Our results indicate the potential of viruses to augment P-acquisition from phosphate-esters and phosphonates in the SAO and in other marine environments.

## Results And Discussion

# A compendium of phylogenetically diverse bacterial and archaeal communities

To estimate the diversity of the epipelagic microbial communities, we generated 620 metagenome assembled genomes (MAGs). These were dereplicated to 104 high and medium quality MAGs (>50% completeness and <10% contamination) following the MIMAG reporting standards [54]. In this study, we were able to retrieve a disproportionate number of high (32) and medium (72) draft quality MAGs compared to a recent study conducted in the SAO [52] (Supplementary Table 2). These MAGs represent the largest compendium of SAO genomes and comprised of taxonomically diverse archaeal (2) and bacterial (8) phyla (Fig. 1b and 1c). MAGs from our study ranged from 0.6Mbp to 6.92Mbp and encoded between 854 to 6,648 proteins consistent with genome completeness estimates. Within bacterial MAGs, members of *Bacteriodia* (17), *Gammaproteobacteria* (26) and *Alphaproteobacteria* (14) classes were highly represented in each sampling station. For archaea, members of *Thaumarchaeota* (11) and *Thermoplasmatota* (5) phyla were only recovered from a few sampling sites (sampling stations 1, 2 and 3).

The five *Thermoplasmatota* archaeal MAGs obtained were affiliated with marine group IIb (MG-IIb) classified as *Poseidoniales* ord. nov. (Supplementary Table 2) [55]. An important aim of our study was to elucidate the spatial distributions of this group in the SAO. For instance, recent studies have shown that MG-IIb distributions increased with inorganic nutrient availability [56, 57]. Marine euphotic zones are characterised by sunlight, photosynthetic driven metabolism, relatively low nutrient availability and high temperatures [58–61]. Similar to previous studies, the inferred metabolic capacity showed that functions characteristic of chemoheterotrophs including TCA cycle, glycolysis, amino acid synthesis and aerobic respiratory chain were common among bacterial and archaeal MAGs [52, 55, 62] (Supplementary Table 3). The compendium of South Atlantic Ocean metagenomes suggests substantial metabolic versatility in bacteria and archaea from this environment. In this study, MAGs were binned to predict viral-host assignments and assess the potential of viral mediated augmentation of phosphorus cycling amongst different microbial lineages.

## SAO harbours novel viral communities

To explore viral communities associated with the SAO euphotic zones, microbial communities were processed using VirSorter2. The viral population obtained in this study were expected to be derived from either proviruses, viruses in infection and replication stages in microbial cells or giant viruses because

microbial cell fractions larger than 0.2  $\mu\text{m}$  were used in this study. The results suggest that the 24 metagenomes harboured 7,363 viral contigs. These contigs were assessed for quality using CheckV and the resultant contigs, with qualities ranked as complete (0.2%), high quality (27.8%), medium quality (19.8%) and low quality (43.4%) (Supplementary Table 4 and Supplementary Figure 1) were further used for downstream analyses. Of these, 7,176 contigs (approximately 3% reduction) with sizes  $>5\text{kb}$  were clustered at  $>95\%$  ANI over 80% of the shortest sequence. These resulted in 5,999 viral operational taxonomic units (vOTUs) of which 1,663 were  $>10\text{kb}$ . A large proportion of the total vOTUs comprised 3,581 putative NCLDVs as well as 2,348 dsDNA viruses (Supplementary Figure 2). Only 423 of 3,581 NCLDVs had  $\geq 1$  hallmark genes, and of these only 2 had one *poB* and MCP proteins, respectively. Based on the underrepresentation of these marker genes within the putative NCLDV contigs we did not conduct taxonomic classifications. Comparisons of prokaryotic-specific dsDNA vOTUs against known prokaryotic and archaeal viruses from NCBI RefSeq (version 85) using vconTACT2 indicated that only 8.5% of these were classified within the *Caudovirales* order (Figure 2, Supplementary Figure 2). These comprised *Myoviridae* (383), *Siphoviridae* (100), *Podoviridae* (78) as well as two previously unidentified genome-based viral families in the SAO, namely *Autographiviridae* ( $n=36$ ), *Demereciviridae* ( $n=16$ ) (Supplementary Figure 2) [63, 64]. Assessments of viral-host interactions suggest that a large proportion of uncharacterised viruses are putative host of reconstructed MAGs (Supplementary Table 5), suggesting that these lineages may be sources of organic matter which may shunt nutrients to the aphotic zone. The remaining 28.9% (excluding putative NCLDV contigs) of prokaryotic dsDNA vOTUs, that could not be taxonomically assigned, suggests that these could represent a wide range of uncharacterised viral communities in the SAO (Supplementary Figure 3). Consistent with previous studies, our results suggest that the SAO viral communities possess protein coding genes which have no homologous hits against the NCBI nr database. However, a much lower proportion of these genes (only 2-3%) were viral [52].

## Euphotic zone viruses encode AMGs that potentially reprogram phosphorus cycling

To further understand how viruses influence nutrient cycling and biogeochemical processes in the SAO, we explored vOTUs for the presence of auxiliary metabolic genes (AMGs). Our analysis revealed 1,279 AMGs involved in a diverse range of metabolic processes among 804 vOTUs of which 356 and 448 were identified in both putative NCLDVs as well as dsDNA viruses. Overall, AMGs involved in carbohydrate, amino acids and cofactors/vitamins metabolic pathways were the most overrepresented (Supplementary Figure 4). This suggests that the composition and abundance of AMGs may be influenced by host diversity and intrinsic environmental conditions [65]. Therefore, the diversity and abundance of AMGs involved in nutrient cycling reflects specific genes that potentially contribute to host adaptation and nutrient acquisition. Among these, were several AMGs, identified in seven uncharacterized vOTUs, which were linked to P assimilation and acquisition. These included acid phosphatases (*phoN*) and other previously undescribed phosphatases such as D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase (*gmhB*) and ribonuclease H/adenosylcobalamin/ $\alpha$ -ribazole phosphatase (*rhnA-cobC*) (Figure 3, Supplementary Table 6). The presence of phosphorus cycling AMGs in the oligotrophic SAO indicate that viruses harbour niche-specific AMGs for particular environments in which they coexist with

their hosts [18, 20, 26, 66, 67]. As these were of major interest, we further explored the genetic content and structure of these AMGs following established methods [26, 67]. The screening involved: (i) using Vibrant v1.2.0 to search for AMGs as previously described [11], (ii) structural modelling using SWISS-MODEL and Phyre2 for quaternary and tertiary structures respectively and (iii) assessment of promoters using BPPROM. All three sequences of P-AMGs genes were annotated as those linked to phosphate acquisition, based on protein structure models (Figure 4, Supplementary Table 6).

**Acid phosphatase** The *phoN* acid phosphatase genes encode for enzymes that catalyse the hydrolysis of phosphomonoesters (P-esters) to inorganic phosphate at optimal pH under phosphorus stress [68]. P-esters form part of the most widespread pool of dissolved organic phosphorus (DOP) in phosphorus-limited marine environments [34, 69]. The presence of *phoN*, as an AMG, suggests the potential of viruses in the SAO to enhance and facilitate the assimilation of organic phosphorus in bacterial (*Gammaproteobacteria*, *Dehalococcoidia* and *Bacteroidia*) and archaeal (*Poseidonii*) lineages following viral-host association analysis using VirHostmatcher (Supplementary Table 6). As the *phoN* gene was identified in viral contigs, we conducted phylogenomic analyses to assess its evolution and check for possible horizontal gene transfer events. The analysis resulted in incongruent ML gene trees, indicating that these *phoN* AMGs were likely acquired from bacterial and archaeal microorganisms (Figure 5). This result suggests that these lineages may be highly abundant in the SAO [52, 53] or that they likely respond rapidly to P-stress.

**D-glycero-D-manno-heptose 1,7-bisphosphate** The *gmhB* gene encodes a phosphatase, which hydrolyses D-glycero-D-manno-heptose 1,7-bisphosphate, producing phosphate as a by-product [70]. This gene is a member of the haloalkanoic acid dehalogenase (HAD) enzyme super-family and has not been previously detected in viral genomes. The putative giant viral contig (S11\_k141\_598071), which harboured this new AMG, had 45 other genes (Figure 4). Of these genes, three had functional assignments to affiliated sequences previously recovered from unclassified *Phycodnaviridae* which are known to infect algae in the *Pyramimonas* order [71]. *gmhB* catalyses two pathways that result in the production of precursors for the S-layer glycoprotein and Lipid A biosynthesis [70, 72]. Previous studies have shown that microbial communities in low phosphorus environments use alternative P-recycling mechanisms [43, 73]. These mechanisms may include the replacement of membrane phospholipids with those that do not contain phosphorus [43, 74, 75]. Our results suggest the potential augmentation of the P-cycle by giant viruses in the SAO utilising *gmhB* to replace membrane phospholipids with lipopolysaccharides during P limitation. In addition, our analysis suggests that carbohydrate phosphates such D-glycero-D-manno-heptose 1,7-bisphosphate may be abundant organic phosphate compounds in marine environments playing essential roles as sources of phosphorus during P starvation.

**Ribonuclease H/adenosylcobalamin/alpha-ribazole** The *rhnA-cobC* encodes an enzyme which catalyses the hydrolysis of adenosylcobalamin-5'-phosphate to phosphate and adenosylcobalamin, the active form of vitamin B [76]. Many marine bacteria lack the ability to synthesise adenosylcobalamin, and as a result are dependent on archaea and phytoplankton for this function [77]. In this study, the *rhnA-cobC* gene was associated with putative giant viral contigs. All MAGs in this study did not have the *rhnA-cobC* gene

(Figure 6). This may be due to the assembly approach used or other sequencing related issues as *rnhA-cobC* genes are ubiquitous in prokaryotic genomes [78]. In addition, the ML tree suggests that the gene was likely acquired from bacteria further highlighting mosaicism of giant viruses (Figure 6). To the best of our knowledge, this is the first description of the *rnhA-cobC* as an AMG. Functional characterisation of the *rnhA-cobC* has mostly been linked to its cleavage of RNA/DNA hybrids formed following replication and repair [79]. Our results suggest that adenosylcobalamin-5'-phosphate could be a potential alternative source of phosphorus when its hydrolysis liberates phosphate, thus facilitating microbial use of phosphate. Alternatively, *rnhA-cobC* could be used to cleave host RNA/DNA, molecules which are rich in phosphate, during infection and lysis. Other compounds such as methylphosphonic acid, various phosphonates and monophosphate esters have been shown to be alternative sources of phosphorus [75, 80, 81]. Similar to phosphate esters, adenosylcobalamin-5'-phosphate has a C-O-P bond which suggests its potential use as a source of P in euphotic zones. Viruses may be shifting their ecological strategies to acquire non-canonical P-genes, thus enabling the assimilation of alternative sources of P by their prokaryotic hosts.

## Conclusion

The South Atlantic Ocean is characterized by distinct genetic endemism and generally low levels of nitrogen and phosphorus [53, 82]. The paucity of SAO metagenomics studies has resulted in a clear knowledge deficit regarding the roles played by viruses in shaping microbial diversity, ecology and evolution. This study revealed a large proportion of uncharacterised viruses, which putatively infect diverse consortia of archaea and bacteria. Virus-encoded AMGs, comprised of genes related to carbohydrate, amino acids and cofactors/vitamins metabolic pathways. Signifying that viruses have the potential to augment host metabolism during infection. Our analysis also revealed several AMGs, potentially related to P acquisition and assimilation by bacteria. These AMGs include *phoN*, *ghmB* and *rnhA-cobC*, which are outside the canonical repertoire of phosphorus genes. These results suggest that viruses potentially augment the acquisition of alternative sources of P in the form of monoesters and organic phosphorus compounds such as adenosylcobalamin. In addition, these P-AMG signatures suggest that their specific pathways in *Bacteroidetes*, *Proteobacteria*, *Chloroflexota* and *Poseidonales* are those that potentially respond rapidly to P-starvation. As regions of high productivity, nutrient depletion and high prokaryotic diversity, euphotic zones represent model ecosystems for studying the contributions of viruses to the physiology of their hosts under nutrient starvation. Since only a small proportion of viral contigs could be classified in this study, there remains large gaps in the understanding of the microbiology of SAO. Together, the data from our study provide fundamental insights regarding potential viral-host dynamics in euphotic zones of the SAO.

## Materials And Methods

### Cruise details and sampling

Sampling was carried out during the 63rd Gough Cruise (10th September – 16th October 2018) aboard the *RV SA Agulhas II*. Water samples were retrieved from eight oceanographic stations in the South Atlantic Ocean. These sites spanned from 37° and 47° South East (Supplementary Table 1). We collected 45 litres of seawater from epipelagic waters at each station (5 m). These samples were stored in acid-washed 5 litre high-density polyethylene bottles at 4°C. From this water, 15 litres were filtered through 0.2 µm Polyethersulfone (PES) filter membranes (Merck, RSA) respectively, resulting in three replicates per station ( $n=24$ ). These membrane filters were stored at -80°C until downstream molecular analysis was carried out.

## **DNA extraction, library Preparation and sequencing**

We performed DNA extraction and library preparation as previously described [83, 84] at JAMSTEC laboratories. Briefly, membrane filter pieces were incubated in 400 µl of DNA extraction buffer (400 mM Tris-HCl pH8, 60 mM EDTA pH8, 150 mM NaCl and 1 % w/v SDS) prepared at Nippon Gene (Toyama, Japan) for 10 min at 60°C. 120 µl of 3 M potassium acetate buffer (Nippon Gene) was used to precipitate organic matter followed by incubation on ice for five min and centrifugation at  $33,516 \times g$  at 4°C for one min. DNA retrieved from the supernatant was subsequently purified through a two-step process by initially adding 800 µl of solution C4 and 500 µl of solution C5 from the DNAeasy Power soil kit (Qiagen, Hilden, Germany). Nucleic acid concentrations were determined using the Qubit 4 Fluorimeter (ThermoFisher, Massachusetts, USA) following manufacturers specifications. High-quality DNA was used to construct libraries using the KAPA Hyper Prep Kit (KAPA biosystems, Massachusetts, USA) as detailed in the manufacturer's protocol. Samples were sequenced using an Illumina HiSeq2000 instrument at MacroGen next-generation sequencing service (<http://www.macrogen.com/en/main/index.php>).

## **Bioinformatics analysis**

### **Metagenome preparation and assembly**

Trimmomatic v0.36 was used to remove adaptors and overrepresented sequences from raw reads [85]. From these sequences, low quality reads were removed using Prinseq lite v0.20.4 (-lc\_method dust) [86]. The resultant reads were *de novo* assembled into contiguous segments using MEGAHIT v1.2.3 (default parameters) [87]. All contigs below 500bp were removed from each sample and the remaining sequences were used to predict open reading frames (ORFs) using Prodigal v2.6.3 (-p meta) [88]. To determine the percentage of assembled reads, the resultant quality-filtered sequences were mapped back to the contigs using BMap Aligner (BMap) (minid=0.90 maxindel=3) [89]. The alignments were 'sorted' and 'indexed' using SAMtools v1.9 as previously described [90].

### **MAGs construction, classification and annotation**

Metagenome-assembled genomes (MAGs) were constructed using MetaBat v2.12.2 which computes tetra-nucleotide frequencies and contig abundances to produce high quality bins [91]. Contigs were binned using the following parameters: -minContig 2500 -minCVSum 2. Assembled MAGs were de-replicated using DRep v2.6.2 [92]. The completeness and contamination levels of the resultant archaeal



and bacterial MAGs were evaluated using CheckM v1.0.18 [93]. In accordance with genome reporting standards, all MAGs reported in this study were characterized as high or medium draft genomes and had completeness > 50% with contamination levels < 10% [54]. The taxonomy of all MAGs were inferred using the genome taxonomy database toolkit (GTDB-Tk) v1.0.2 GTDB release 89 [94]. Functional annotation was determined using the KofamScan tool, which is based on KEGG Orthology and hidden markov models [95]. KEGG-Decoder was used to analyse the results as described previously [95, 96]. rRNAs and tRNAs were determined using barrnap v0.7 and the tRNAScan-SE v2.0 tool, respectively [97, 98]. The coverage of each MAG was determined using CoverM (genome -p bwa-mem -m relative abundance) [99]. We determined phylogenies for archaeal and bacterial MAGs independently using GToTree (-G 0.4) [100]. All phylogenomic trees were visualised and annotated using iTOL [101].

## **Viral identification and taxonomic assignments**

Viral signatures in assembled contigs (> 5 kb) were identified using VirSorter v2.1 as previously described [102]. Briefly, contigs were analysed using VirSorter v2.1 (-include-groups all -min-length 5000 -min-score 0.5) and the quality of the resultant contigs was checked using CheckV [103]. The output results final-viral-score.tsv and contamination.tsv from VirSorter and CheckV were used to screen viral contigs for downstream analysis respectively. Quality checked contigs were screened based on the following criteria: VirSorter\_max\_score (>=0.95), viral gene (>0) and viral gene =0 (and host gene =0 or score >=0.95 or hallmark >2) as previously described [104]. The resultant outputs were further manually checked for proportions of contigs putatively assigned as double-stranded DNA viruses (dsDNA), nucleocytoplasmic large DNA viruses (NCLDV) and single stranded DNA (ssDNA). Viral populations were determined by clustering viral contigs > 5kb at 95% average nucleotide identity over 80% of the shortest sequence using CD-HIT-EST from the CD-HIT package [105]. vconTACT2 was used to classify viral contigs [106]. Briefly, viral open reading frames were firstly predicted using Prodigal v2.6.3 as described previously [88, 107]. Predicted open reading frames were then queried against the NCBI Bacterial and Archaeal Viral RefSeq V85 with ICTV and NCBI taxonomy [108] used in vconTACT2 (<http://www.cyverse.org>). The results of viral classification were then visualised on Cytoscape v3.8.1 and colour coded based on respective order taxonomic rank [109, 110].

## **AMG identification, annotation and validation**

Auxiliary metabolic genes were determined using VIBRANT v1.2.0 (in “virome” mode), which was accessed via the CyVerse platform [111]. Contigs containing phosphorus acquisition and assimilation genes were queried and subjected to a series of validation steps to ensure that they could be attributed to viral sequences and to confirm the functional annotations [112]. Genetic architecture plots to visualise the organisation of genes on contigs were generated using the R package gggenes (<https://wilcox.org/gggenes/>). Conserved domains and active sites of phosphorus AMGs were identified using the NCBI conserved domain search (e-value 0.001) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Protein domains and structural homology of all AMGs were identified using Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2>) [113]. Predicted secondary structures, with a 100% confidence score and alignment coverage above 70%, were further analysed with

the SWISS-MODEL to predict quaternary structure (GMQE > 0.5) [114]. Promoter regions in the protein domains were predicted using BPPROM [115].

## Viral relative abundance, distribution and host-viral matches

To determine the relative abundances of viral populations containing P-cycle AMGs, all contigs were concatenated and then used as a database to recruit the quality trimmed reads using BMap [89]. The relative abundance of each population per sample was estimated from the resulting bam files and converted into table using a custom wrapper script from 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 as previously described [26, 116]. The resultant data were used to assess the distribution of viral populations harbouring P-cycle AMGs among sampling stations and visualised using ggplot2 in R v4.0.3 [117, 118]. To infer putative links between viruses and hosts, VirHostMatcher was used to compute oligonucleotide frequency between viral contigs and MAGs [119]. As the goal of the study was to primarily study the interactions and influence of prokaryotic viruses towards bacterial hosts, all putative giant viral contigs were removed prior the analysis. Putative virus-host interactions were filtered using a d2\* dissimilarity cut-off < 0.3 and visualised using Cytoscape v3.8.1.

## Phylogenetic tree reconstruction

Functional analyses of *rnhA-cobC* (ribonuclease H/adenosylcobalamin/alpha-ribazole phosphatase) and *phoN* (acid phosphatase (class A)) genes were performed to investigate the evolutionary origin of these AMGs. Protein sequences were queried against the NCBI nr database (blastp 1000 bit score cut-off, and e-value 0.001) [120]. Closely related sequences were retrieved from the NCBI to estimate the non-viral context of sequences in the phylogenetic trees. The closest relatives were selected and reduced to representative sequences using CD-HIT (-c 0.9 -n 5). Protein sequences were aligned using MAFFT (-auto) and trimmed using trimAL (default parameters) [121, 122]. Functional gene trees reconstruction were performed using PhyML with the Chi<sup>2</sup>-based branch support and visualised using iTOL [101, 123].

## List Of Abbreviations

AMGs - Auxiliary metabolic genes

SAO - South Atlantic Ocean

vOTUs - Viral operational taxonomic units

*phoN* - Acid phosphatase

*gmhB* - D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase

*rnhA-cobC* - ribonuclease H/adenosylcobalamin/alpha-ribazole phosphatase

NCLDVs - Nucleocytoplasmic large DNA viruses

*PhoA* - Alkaline phosphatase

*pstS* - High affinity phosphatase

*Pho4* - Phosphate transporter

DIP - Dissolved inorganic phosphate

DOP - Dissolved inorganic phosphate

NAO - North Atlantic Ocean

MAGs - Metagenome assembled genomes

MIMAG - Minimum information about a metagenome-assembled genome

TCA - Tricarboxylic acid cycle

*poB* - DNA polymerase B

MCP- Major capsid protein

PES - Polyethersulfone

JAMSTEC - Japan Agency for Marine-Earth Science and Technology

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Availability of data and material

Nucleotide sequence data have been submitted to NCBI SRA under the study BioProject ID PRJNA748242.

### Competing interests

All authors declare that they have no competing interests.

## **Funding**

We are grateful to the National Research Foundation (NRF) (Grant ID 118981, the South African National Antarctic Programme (SANAP 110717), the South Africa Japan Bilateral Program (114681) and the University of Pretoria for funding. TPM and OKIB acknowledge funding received from the European Union's Horizon 2020 research and innovation programme as part of the AtlantECO project under grant agreement No 862923.

## **Authors' contributions**

T.P.M conceived the project with input from J.C.M; P.M.L, T.P.M planned and conducted sampling; P.M.L, M.H prepared metagenomic libraries and performed sequencing; P.M.L, O.K.I.B, Y.T assembled, curated and analysed sequence data; T.Y, T.N provided computational resources; P.M.L wrote the manuscript with contributions from O.K.I.B. and T.P.M.; all authors discussed and revised the manuscript.

## **Acknowledgements**

We thank the Centre for High Performance Computing (Cape Town, South Africa) and the Centre for Bioinformatics and Computational Biology, University of Pretoria, for providing computational resources.

## **Authors' information**

### **Affiliations**

**Centre for Microbial Ecology and Genomics, Department of Biochemistry, Genetics and Microbiology, Natural Sciences 2 Building, Hatfield, University of Pretoria, Pretoria, South Africa, 0028**

Percy Mutseka Lunga, Thulani P. Makhalanyane & Oliver K.I Bezuidt

**Super-cutting-edge Grand and Advanced Research (SUGAR) Program, Institute for Extra-cutting-edge Science and Technology Avant-garde Research (X-star), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Natushima 2-15, Yokosuka, Kanagawa 237-0061, Japan**

Miho Hirai, Yoshihiro Takaki & Taichi Yokokawa

**Research Centre for Bioscience and Nanoscience (CeBN), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Natushima 2-15, Yokosuka, Kanagawa 237-0061, Japan**

Takuro Nunoura

**School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom**

J. Colin Murrell

# References

1. Suttle CA (2005) Viruses in the sea. *Nature* 437: 356–361.
2. Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *science* 320: 1034–1039.
3. Aevarsson A, Kaczorowska A-K, Adalsteinsson BT, Ahlqvist J, Al-Karadaghi S, Altenbuchner J, Arsin H, Átlason ÚÁ, Brandt D, Cichowicz-Cieślak M (2021) Going to extremes—a metagenomic journey into the dark matter of life. *FEMS microbiology letters*.
4. Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* 399: 541–548.
5. Bachy C, Yung CC, Needham DM, Gazitúa MC, Roux S, Limardo AJ, Choi CJ, Jorgens DM, Sullivan MB, Worden AZ (2021) Viruses infecting a warm water picoeukaryote shed light on spatial co-occurrence dynamics of marine viruses and their hosts. *The ISME Journal*: 1–19.
6. Mojica KD, Huisman J, Wilhelm SW, Brussaard CP (2016) Latitudinal variation in virus-induced mortality of phytoplankton across the North Atlantic Ocean. *The ISME journal* 10: 500–513.
7. Endo H, Blanc-Mathieu R, Li Y, Salazar G, Henry N, Labadie K, de Vargas C, Sullivan MB, Bowler C, Wincker P, Karp-Boss L, Sunagawa S, Ogata H (2020) Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nature Ecology & Evolution* 4: 1639–1649. doi: 10.1038/s41559-020-01288-w
8. Sheyn U, Rosenwasser S, Lehahn Y, Barak-Gavish N, Rotkopf R, Bidle KD, Koren I, Schatz D, Vardi A (2018) Expression profiling of host and virus during a coccolithophore bloom provides insights into the role of viral infection in promoting carbon export. *The ISME journal* 12: 704–713.
9. McAllister SM, Vandzura R, Keffer JL, Polson SW, Chan CS (2021) Aerobic and anaerobic iron oxidizers together drive denitrification and carbon cycling at marine iron-rich hydrothermal vents. *The ISME journal* 15: 1271–1286.
10. Jian H, Yi Y, Wang J, Hao Y, Zhang M, Wang S, Meng C, Zhang Y, Jing H, Wang Y (2021) Diversity and distribution of viruses inhabiting the deepest ocean on Earth. *The ISME Journal*: 1–17.
11. Breitbart M, Thompson LR, Suttle CA, Sullivan MB (2007) Exploring the Vast Diversity of Marine Viruses. *Oceanography* 20: 135–139. doi: DOI 10.5670/oceanog.2007.58
12. Brito IL (2021) Examining horizontal gene transfer in microbial communities. *Nature Reviews Microbiology* 19: 442–453.
13. Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003) Marine ecosystems: Bacterial photosynthesis genes in a virus. *Nature* 424: 741–741. doi: 10.1038/424741a
14. Kieft K, Zhou Z, Anderson RE, Buchan A, Campbell BJ, Hallam SJ, Hess M, Sullivan MB, Walsh DA, Roux S (2021) Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nature communications* 12: 1–16.
15. Zhong Z-P, Rapp JZ, Wainaina JM, Solonenko NE, Maughan H, Carpenter SD, Cooper ZS, Jang HB, Bolduc B, Deming JW (2020) Viral ecogenomics of arctic cryopeg brine and sea ice. *Msystems* 5:

e00246-00220.

16. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB (2020) The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell host & microbe* 28: 724–740. e728.
17. Fuchsman CA, Carlson MC, Garcia Prieto D, Hays MD, Rocap G (2020) Cyanophage host-derived genes reflect contrasting selective pressures with depth in the oxic and anoxic water column of the Eastern Tropical North Pacific. *Environmental Microbiology*.
18. Hurwitz BL, Hallam SJ, Sullivan MB (2013) Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biology* 14: R123. doi: ARTN R1231186/gb-2013-14-11-r123
19. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA (2019) Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. *The ISME journal* 13: 618–631.
20. Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ (2014) Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. *Science* 344: 757–760. doi: 10.1126/science.1252229
21. Kieft K, Zhou Z, Anderson RE, Buchan A, Campbell BJ, Hallam SJ, Hess M, Sullivan MB, Walsh DA, Roux S, Anantharaman K (2020) Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *bioRxiv*: 2020.2008.2024.253096. doi: 10.1101/2020.08.24.253096
22. Needham DM, Yoshizawa S, Hosaka T, Poirier C, Choi CJ, Hehenberger E, Irwin NAT, Wilken S, Yung CM, Bachy C, Kurihara R, Nakajima Y, Kojima K, Kimura-Someya T, Leonard G, Malmstrom RR, Mende DR, Olson DK, Sudo Y, Sudek S, Richards TA, DeLong EF, Keeling PJ, Santoro AE, Shirouzu M, Iwasaki W, Worden AZ (2019) A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc Natl Acad Sci U S A* 116: 20574–20583. doi: 10.1073/pnas.1907517116
23. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO (2020) Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat Commun* 11: 1710. doi: 10.1038/s41467-020-15507-2
24. Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, Waldbauer JR, Coleman ML (2020) Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews Microbiology* 18: 21–34. doi: 10.1038/s41579-019-0270-x
25. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaque D, Bork P, Acinas SG, Wincker P, Sullivan MB, Coordinators TO (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537: 689+. doi: 10.1038/nature19366
26. Gazitúa MC, Vik DR, Roux S, Gregory AC, Bolduc B, Widner B, Mulholland MR, Hallam SJ, Ulloa O, Sullivan MB (2020) Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *The ISME Journal*. doi: 10.1038/s41396-020-00825-6

27. Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit M-A (2017) Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *The ISME journal* 11: 237–247.
28. Roux S, Krupovic M, Debroas D, Forterre P, Enault F (2013) Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. *Open biology* 3: 130160.
29. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B (2019) Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology Journal* 16: 15. doi: 10.1186/s12985-019-1120-1
30. McDaniel EA, Peterson BD, Stevens SL, Tran PQ, Anantharaman K, McMahon KD (2020) Expanded phylogenetic diversity and metabolic flexibility of mercury-methylating microorganisms. *Msystems* 5: e00299-00220.
31. Patwardhan S, Smedile F, Giovannelli D, Vetriani C (2021) Metaproteogenomic profiling of chemosynthetic microbial biofilms reveals metabolic flexibility during colonization of a shallow-water gas vent. *Frontiers in microbiology* 12: 627.
32. Sosa OA, Repeta DJ, DeLong EF, Ashkezari MD, Karl DM (2019) Phosphate-limited ocean regions select for bacterial populations enriched in the carbon–phosphorus lyase pathway for phosphonate degradation. *Environmental microbiology* 21: 2402–2414.
33. Martiny AC, Coleman ML, Chisholm SW (2006) Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proceedings of the National Academy of Sciences* 103: 12552-12557.
34. Sebastian M, Ammerman JW (2009) The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *The ISME journal* 3: 563–572.
35. Monier A, Welsh RM, Gentemann C, Weinstock G, Sodergren E, Armbrust EV, Eisen JA, Worden AZ (2012) Phosphate transporters in marine phytoplankton and their viruses: cross-domain commonalities in viral-host gene exchanges. *Environ Microbiol* 14: 162–176. doi: 10.1111/j.1462-2920.2011.02576.x
36. Diaz JM, Holland A, Sanders JG, Bulski K, Mollett D, Chou C-W, Phillips D, Tang Y, Duhamel S (2018) Dissolved organic phosphorus utilization by phytoplankton reveals preferential degradation of polyphosphates over phosphomonoesters. *Frontiers in Marine Science* 5: 380.
37. Moore C, Mills M, Arrigo K, Berman-Frank I, Bopp L, Boyd P, Galbraith E, Geider R, Guieu C, Jaccard S (2013) Processes and patterns of oceanic nutrient limitation. *Nature geoscience* 6: 701–710.
38. Young CL, Ingall ED (2010) Marine dissolved organic phosphorus composition: insights from samples recovered using combined electro dialysis/reverse osmosis. *Aquatic Geochemistry* 16: 563–574.
39. Sosa OA, Burrell TJ, Wilson ST, Foreman RK, Karl DM, Repeta DJ (2020) Phosphonate cycling supports methane and ethylene supersaturation in the phosphate-depleted western North Atlantic Ocean. *Limnology and Oceanography* 65: 2443–2459.

40. Bertilsson S, Berglund O, Karl DM, Chisholm SW (2003) Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. *Limnology and oceanography* 48: 1721–1731.
41. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS (2014) The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nature Reviews Microbiology* 12: 519–528.
42. Lomas MW, Burke A, Lomas D, Bell D, Shen C, Dyhrman ST, Ammerman JW (2010) Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP). *Biogeosciences* 7: 695–710.
43. Carini P, Van Mooy BAS, Thrash JC, White A, Zhao Y, Campbell EO, Fredricks HF, Giovannoni SJ (2015) SAR11 lipid renovation in response to phosphate starvation. *Proceedings of the National Academy of Sciences* 112: 7767. doi: 10.1073/pnas.1505034112
44. Geiger O, Röhrs V, Weissenmayer B, Finan TM, Thomas-Oates JE (1999) The regulator gene *phoB* mediates phosphate stress-controlled synthesis of the membrane lipid diacylglycerol-N,N,N-trimethylhomoserine in *Rhizobium* (*Sinorhizobium*) *meliloti*. *Molecular Microbiology* 32: 63–73. doi: <https://doi.org/10.1046/j.1365-2958.1999.01325.x>
45. Huang X, Jiao N, Zhang R (2021) The genomic content and context of auxiliary metabolic genes in roseophages. *Environmental Microbiology*.
46. Williamson SJ, Allen LZ, Lorenzi HA, Fadrosch DW, Bami D, Thiagarajan M, McCrow JP, Tovchigrechko A, Yooseph S, Venter JC (2012) Metagenomic exploration of viruses throughout the Indian Ocean.
47. Tyler PA (2003) *Ecosystems of the deep oceans*. Elsevier
48. Mather RL, Reynolds SE, Wolff GA, Williams RG, Torres-Valdes S, Woodward EMS, Landolfi A, Pan X, Sanders R, Achterberg EP (2008) Phosphorus cycling in the North and South Atlantic Ocean subtropical gyres. *Nature Geoscience* 1: 439–443.
49. Mills MM, Moore C, Langlois R, Milne A, Achterberg E, Nachtigall K, Lochte K, Geider R, J. La R (2008) Nitrogen and phosphorus co-limitation of bacterial productivity and growth in the oligotrophic subtropical North Atlantic. *Limnology and Oceanography* 53: 824–834.
50. Zubkov MV, Mary I, Woodward EMS, Warwick PE, Fuchs BM, Scanlan DJ, Burkill PH (2007) Microbial control of phosphate in the nutrient-depleted North Atlantic subtropical gyre. *Environmental Microbiology* 9: 2079–2089.
51. Reynolds SE, Mather RL, Wolff GA, Williams RG, Landolfi A, Sanders R, Woodward EMS (2007) How widespread and important is N<sub>2</sub> fixation in the North Atlantic Ocean? *Global Biogeochemical Cycles* 21.
52. Coutinho F, von Meijenfheldt F, Walter J, Haro-Moreno J, López-Pérez M, van Verk M, Thompson C, Cosenza C, Appolinario L, Paranhos R (2021) Ecogenomics and metabolic potential of the South Atlantic Ocean microbiome. *Science of The Total Environment* 765: 142758.
53. Alves Junior N, Meirelles PM, de Oliveira Santos E, Dutilh B, Silva GGZ, Paranhos R, Cabral AS, Rezende C, Iida T, de Moura RL, Kruger RH, Pereira RC, Valle R, Sawabe T, Thompson C, Thompson F



- (2014) Microbial community diversity and physical–chemical features of the Southwestern Atlantic Ocean. *Archives of Microbiology* 197: 165–179. doi: 10.1007/s00203-014-1035-6
54. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloë-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Kyrpides NC, Schriml L, Garrity GM, Hugenholtz P, Sutton G, Yilmaz P, Meyer F, Glöckner FO, Gilbert JA, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat Eren A, Schriml L, Banfield JF, Hugenholtz P, Woyke T, The Genome Standards C (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* 35: 725–731. doi: 10.1038/nbt.3893
55. Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvochina M, Brown M, Jeffries T, Tyson GW, Seymour JR, Hugenholtz P (2019) A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *The ISME Journal* 13: 663–675. doi: 10.1038/s41396-018-0282-y
56. Hugoni M, Taib N, Debroas D, Domaizon I, Dufournel IJ, Bronner G, Salter I, Agogué H, Mary I, Galand PE (2013) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proceedings of the National Academy of Sciences* 110: 6004–6009.
57. Mincer TJ, Church MJ, Taylor LT, Preston C, Karl DM, DeLong EF (2007) Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environmental microbiology* 9: 1162–1175.
58. Phoma S, Vikram S, Jansson JK, Ansoorge IJ, Cowan DA, Van de Peer Y, Makhalanyane TP (2018) Agulhas Current properties shape microbial community diversity and potential functionality. *Scientific Reports* 8. doi: 10.1038/s41598-018-28939-0
59. Chen S, He Y-B, Xie Z-X, Kong L-F, Yan K-Q, Li D-X, Wu P-F, Zheng R-W, Peng L, Chen J-W (2021) Metaproteomics reveals nutrient availability shaping distinct microbial community and metabolic niche in the nutrient-depleted and replete layers of an oligotrophic euphotic zone. *Science of The Total Environment* 774: 145123.
60. Giner CR, Pernice MC, Balagué V, Duarte CM, Gasol JM, Logares R, Massana R (2020) Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. *The ISME journal* 14: 437–449.
61. Raes EJ, Karsh K, Sow SL, Ostrowski M, Brown MV, van de Kamp J, Franco-Santos RM, Bodrossy L, Waite AM (2021) Metabolic pathways inferred from a bacterial marker gene illuminate ecological changes across South Pacific frontal boundaries. *Nature communications* 12: 1–12.
62. Zhang CL, Xie W, Martin-Cuadrado A-B, Rodriguez-Valera F (2015) Marine Group II Archaea, potentially important players in the global ocean carbon cycle. *Frontiers in Microbiology* 6: 1108.

63. Turner D, Kropinski AM, Adriaenssens EM (2021) A Roadmap for Genome-Based Phage Taxonomy. *Viruses* 13. doi: 10.3390/v13030506
64. Adriaenssens EM, Sullivan MB, Knezevic P, van Zyl LJ, Sarkar B, Dutilh BE, Alfenas-Zerbini P, Łobocka M, Tong Y, Brister JR (2020) Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Archives of virology*: 1-8.
65. Hurwitz BL, U'Ren JM (2016) Viral metabolic reprogramming in marine ecosystems. *Current Opinion in Microbiology* 31: 161–168. doi: <https://doi.org/10.1016/j.mib.2016.04.002>
66. Hurwitz BL, Brum JR, Sullivan MB (2015) Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME Journal* 9: 472–484. doi: 10.1038/ismej.2014.143
67. Mara P, Vik D, Pachiadaki MG, Suter EA, Poulos B, Taylor GT, Sullivan MB, Edgcomb VP (2020) Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *The ISME Journal* 14: 3079–3092. doi: 10.1038/s41396-020-00739-3
68. Kathuria S, Martiny AC (2011) Prevalence of a calcium-based alkaline phosphatase associated with the marine cyanobacterium *Prochlorococcus* and other ocean bacteria. *Environmental microbiology* 13: 74–83.
69. White AE (2009) New insights into bacterial acquisition of phosphorus in the surface ocean. *Proceedings of the National Academy of Sciences* 106: 21013-21014.
70. Wang L, Huang H, Nguyen HH, Allen KN, Mariano PS, Dunaway-Mariano D (2010) Divergence of biochemical function in the HAD superfamily: D-glycero-D-manno-heptose-1, 7-bisphosphate phosphatase (GmhB). *Biochemistry* 49: 1072–1081.
71. Sandaa R-A, Heldal M, Castberg T, Thyrrhaug R, Bratbak G (2001) Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* 290: 272–280.
72. Huddleston JP, Raushel FM (2019) Biosynthesis of GDP-d-glycero- $\alpha$ -d-manno-heptose for the Capsular Polysaccharide of *Campylobacter jejuni*. *Biochemistry* 58: 3893–3902. doi: 10.1021/acs.biochem.9b00548
73. Martin P, Dyhrman ST, Lomas MW, Poulton NJ, Van Mooy BAS (2014) Accumulation and enhanced cycling of polyphosphate by Sargasso Sea plankton in response to low phosphorus. *Proceedings of the National Academy of Sciences* 111: 8089. doi: 10.1073/pnas.1321719111
74. Carini P, White AE, Campbell EO, Giovannoni SJ (2014) Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. *Nature Communications* 5: 4346. doi: 10.1038/ncomms5346
75. Dyhrman S, Chappell P, Haley S, Moffett J, Orchard E, Waterbury J, Webb E (2006) Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. *Nature* 439: 68–71.
76. Zayas CL, Escalante-Semerena JC (2007) Reassessment of the late steps of coenzyme B12 synthesis in *Salmonella enterica*: evidence that dephosphorylation of adenosylcobalamin-5'

- phosphate by the CobC phosphatase is the last step of the pathway. *J Bacteriol* 189: 2210–2218. doi: 10.1128/JB.01665-06
77. Bonnet S, Tovar-Sánchez A, Panzeca C, Duarte CM, Ortega-Retuerta E, Sañudo-Wilhelmy SA (2013) Geographical gradients of dissolved Vitamin B12 in the Mediterranean Sea. *Frontiers in microbiology* 4: 126–126. doi: 10.3389/fmicb.2013.00126
  78. Tadokoro T, Kanaya S (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J* 276: 1482–1493.
  79. Cerritelli SM, Crouch RJ (2009) Ribonuclease H: the enzymes in eukaryotes. *FEBS J* 276: 1494–1505.
  80. Sañudo-Wilhelmy SA (2006) A phosphate alternative. *Nature* 439: 25–26. doi: 10.1038/439025a
  81. Kamat SS, Williams HJ, Raushel FM (2011) Intermediates in the transformation of phosphonates to phosphate by bacteria. *Nature* 480: 570–573. doi: 10.1038/nature10622
  82. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Coordinators TO (2015) Structure and function of the global ocean microbiome. *Science* 348: 1261359. doi: ARTN 1261359 1126/science.1261359
  83. Hirai M, Nishi S, Tsuda M, Sunamura M, Takaki Y, Nunoura T (2017) Library construction from subnanogram DNA for pelagic sea water and deep-sea sediments. *Microbes and environments*: ME17132.
  84. Phoma BS, Makhalanyane TP (2020) Depth-Dependent Variables Shape Community Structure and Functionality in the Prince Edward Islands. *Microbial Ecology*: 1-14.
  85. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. doi: 10.1093/bioinformatics/btu170
  86. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864.
  87. Li DH, Liu CM, Luo RB, Sadakane K, Lam TW (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31: 1674–1676. doi: 10.1093/bioinformatics/btv033
  88. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics* 11: 119. doi: ArtN 119 1186/1471-2105-11-119
  89. Bushnell B (2014) BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).
  90. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079. doi:

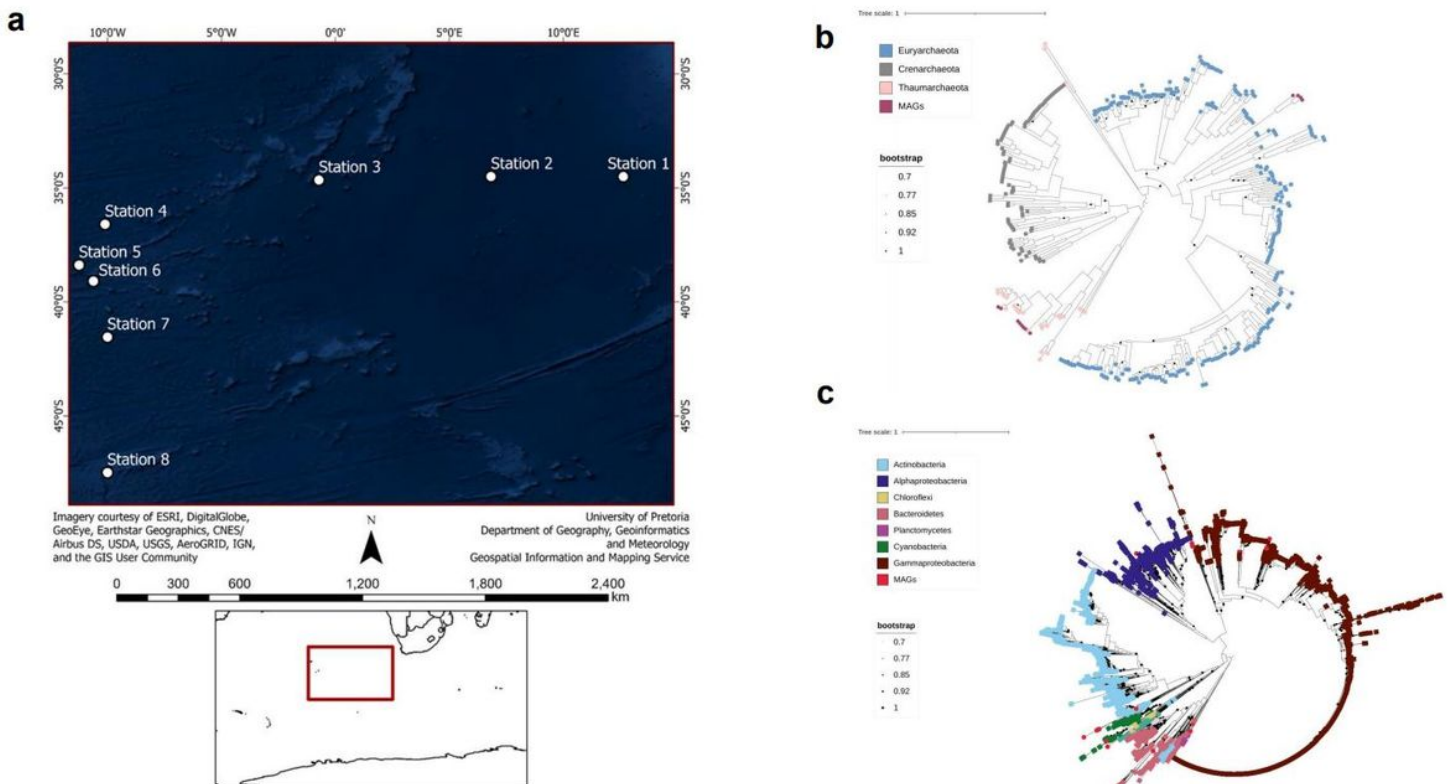
- 10.1093/bioinformatics/btp352
91. Kang DD, Froula J, Egan R, Wang Z (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3: e1165.
  92. Olm MR, Brown CT, Brooks B, Banfield JF (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal* 11: 2864–2868. doi: 10.1038/ismej.2017.126
  93. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* 25: 1043–1055.
  94. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36: 1925–1927. doi: 10.1093/bioinformatics/btz848
  95. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H (2020) KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36: 2251–2252. doi: 10.1093/bioinformatics/btz859
  96. Graham ED, Heidelberg JF, Tully BJ (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. *The ISME Journal* 12: 1861–1866. doi: 10.1038/s41396-018-0091-3
  97. Loman T (2017) A Novel Method for Predicting Ribosomal RNA Genes in Prokaryotic Genomes. doi: <http://lup.lub.lu.se/student-papers/record/8914064>
  98. Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research* 44: W54-W57. doi: 10.1093/nar/gkw413
  99. Robbins SJ, Chan CX, Messer LF, Singleton CM, Geers AU, Ying H, Baker A, Bell SC, Morrow KM, Ragan MA (2017) wwood/CoverM: Read coverage calculator for metagenomics.
  100. Lee MD (2019) GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 35: 4162–4164. doi: 10.1093/bioinformatics/btz188
  101. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23: 127–128. doi: 10.1093/bioinformatics/btl529
  102. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB (2021) VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9: 1–13.
  103. Nayfach S, Camargo AP, Eloë-Fadrosh E, Roux S, Kyrpides N (2020) CheckV: assessing the quality of metagenome-assembled viral genomes. *bioRxiv*: 2020.2005.2006.081778. doi: 10.1101/2020.05.06.081778
  104. Jiarong Guo DV, Akbar Adjie Pratama, Simon Roux, Matthew Sullivan (2021) Viral sequence identification SOP with VirSorter2. *protocols.io*. doi: <https://dx.doi.org/10.17504/protocols.io.btv8nn9w>

105. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659. doi: 10.1093/bioinformatics/btl158
106. Zablocki O, Jang HB, Bolduc B, Sullivan MB (2019) vConTACT 2: A Tool to Automate Genome-Based Prokaryotic Viral Taxonomy. *Plant and Animal Genome XXVII Conference (January 12-16, 2019)*. PAG.
107. Bolduc B, Jang HB, Doucier G, You ZQ, Roux S, Sullivan MB (2017) vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* 5: e3243. doi: ARTN e3243 7717/peerj.3243
108. Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Turner D (2019) Gene sharing networks to automate genome-based prokaryotic viral taxonomy. *Nature Biotechnology* 37.
109. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27: 431–432. doi: 10.1093/bioinformatics/btq675
110. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13: 2498–2504. doi: 10.1101/gr.1239303
111. Kieft K, Zhou Z, Anantharaman K (2020) VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8: 90. doi: 10.1186/s40168-020-00867-0
112. Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* 428: 726–731. doi: 10.1016/j.jmb.2015.11.006
113. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10: 845–858. doi: 10.1038/nprot.2015.053
114. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TA P, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* 46: W296-W303. doi: 10.1093/nar/gky427
115. Umarov RK, Solovyev VV (2017) Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. *PloS one* 12: e0171410.
116. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, Ardyna M, Arkhipova K, Carmichael M, Cruaud C, Dimier C, Domínguez-Huerta G, Ferland J, Kandels S, Liu Y, Marec C, Pesant S, Picheral M, Pisarev S, Poulain J, Tremblay J-É, Vik D, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G, de Vargas C, Follows M, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D, Jaillon O, Kandels-Lewis S, Karp-Boss L, Karsenti E, Not F, Ogata H, Pesant S, Poulton N, Raes J, Sardet C, Speich S, Stemmann L, Sullivan MB, Sunagawa S, Wincker P, Babin M, Bowler C, Culley AI, de Vargas C, Dutilh BE, Iudicone D, Karp-Boss L, Roux S, Sunagawa S, Wincker P, Sullivan MB (2019) Marine

DNA Viral Macro- and Microdiversity from Pole to Pole. Cell 177: 1109-1123.e1114. doi: <https://doi.org/10.1016/j.cell.2019.03.040>

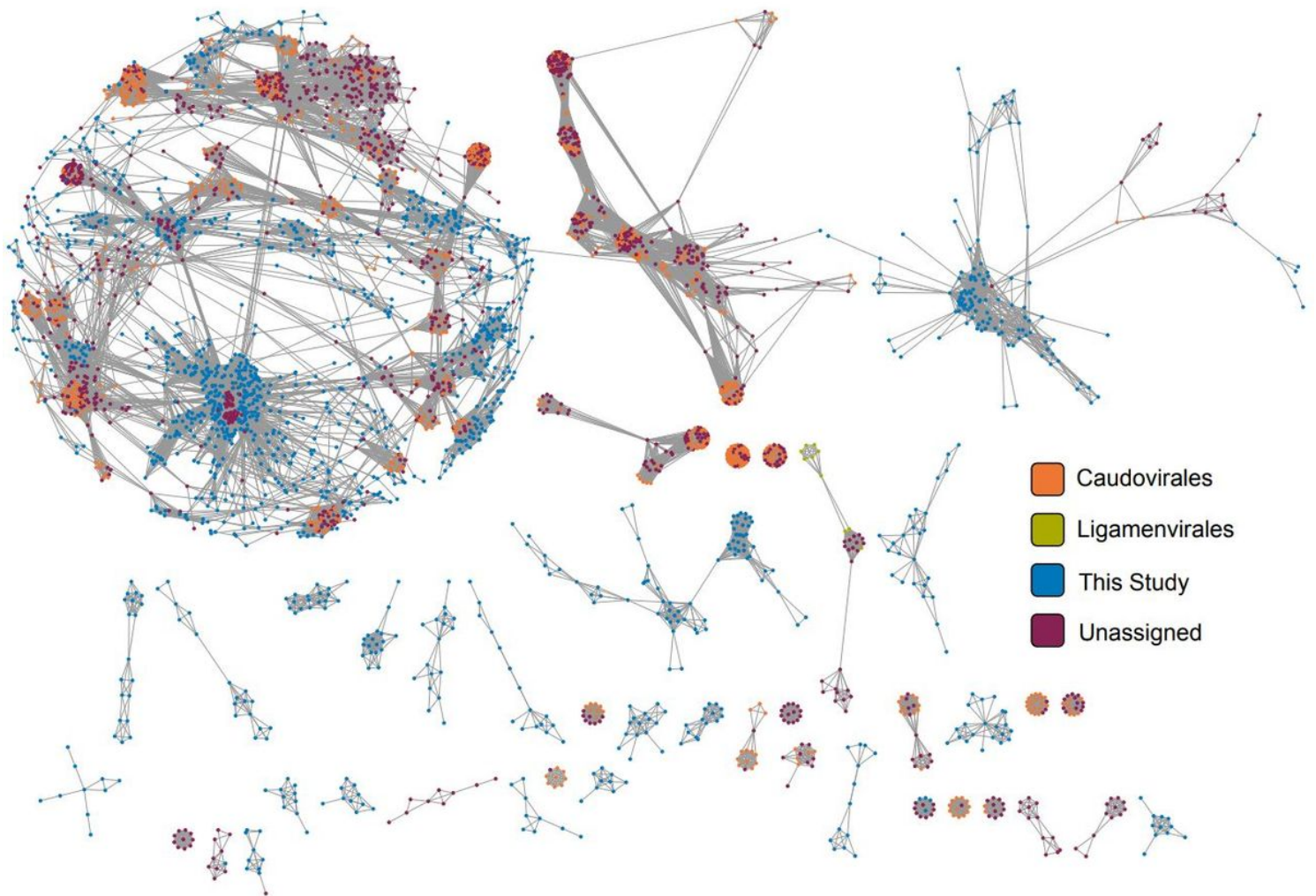
117. Team RC (2013) R: A language and environment for statistical computing.
118. Wickham H, Wickham MH (2007) The ggplot package. Google Scholar.
119. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F (2017) Alignment-free  $\$d_2^*\$$  oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. Nucleic Acids Research 45: 39–53. doi: 10.1093/nar/gkw1002
120. Coordinators NR (2018) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 46: D8-D13. doi: 10.1093/nar/gkx1095
121. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30: 3059–3066. doi: DOI 10.1093/nar/gkf436
122. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973. doi: 10.1093/bioinformatics/btp348
123. Poon AFY, Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE 5: e9490. doi: 10.1371/journal.pone.0009490

## Figures



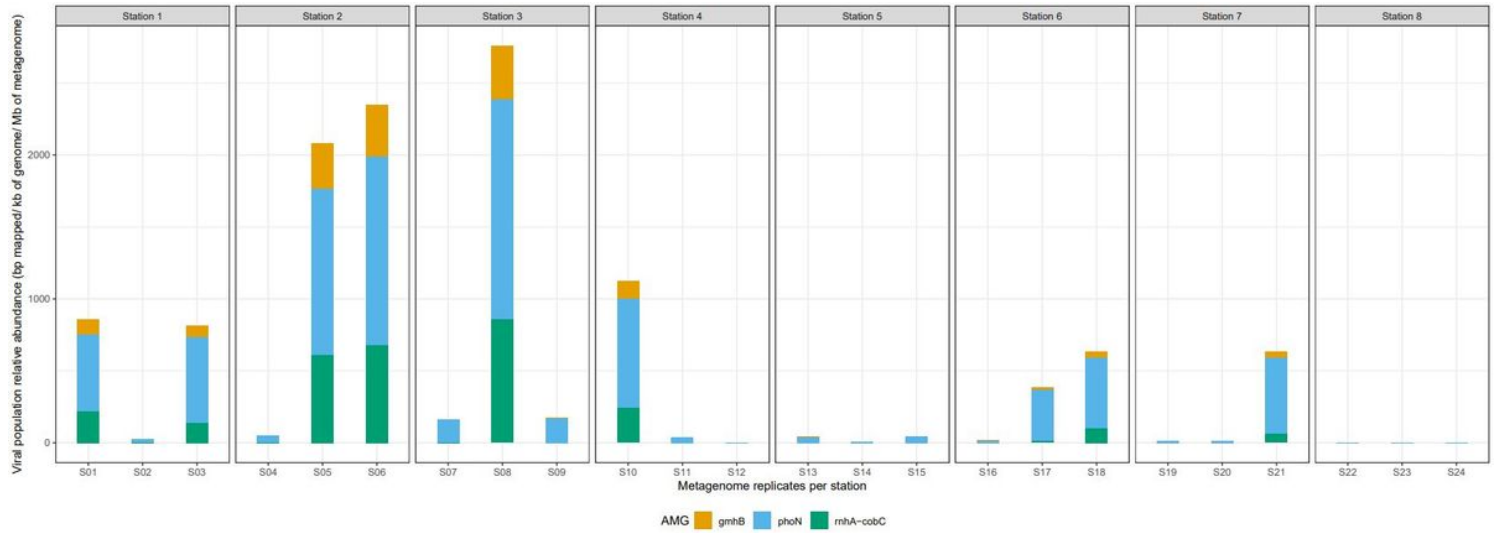
**Figure 1**

(a) Map showing sampling stations in the South Atlantic (courtesy of Lauren Pijper, University of Pretoria, Department of Geography, Geoinformatics and Meteorology). Sampling stations are denoted as white bullets and the outcrop shows the entire sampling site relative to South Africa and the South Atlantic Ocean. (b) Phylogenetic placement of 14 reconstructed metagenome-assembled Archaeal genomes. A maximum-likelihood phylogenomic tree was built based on concatenated amino acid sequences of 76 conserved single-copy genes using FastTree, two MAGs (S18.bin.27 and S20.bin.19) were removed from the tree because of insufficient single copy genes. (c) Phylogenetic placement of 88 reconstructed metagenome-assembled genomes. A maximum-likelihood phylogenomic tree was built based on concatenated amino acid sequences of 74 conserved single-copy genes using Fast Tree, 13 MAGs (S10.bin.8, S10.bin.9, S13.bin.1, S16.bin.2, S16.bin.25, S18.bin.14, S18.bin.16, S22.bin.40, S23.bin.29, S23.bin.36, S24.bin.26, S24.bin.3, S24.bin.39 and S8.bin.7) were removed from the tree because of insufficient single copy genes. The scale represents one amino acid substitution per sequence position.



**Figure 2**

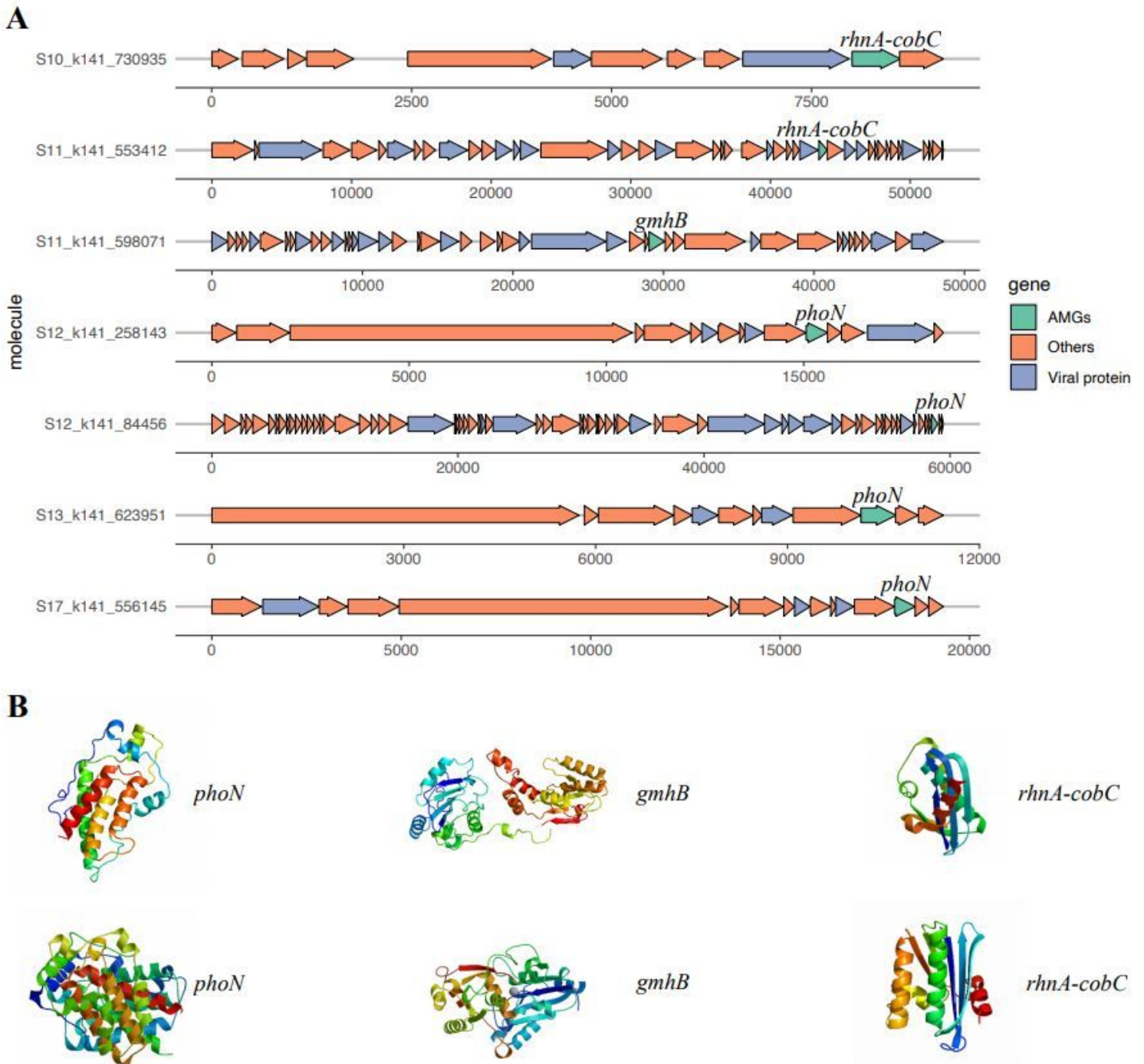
Taxonomic assignment of viral populations protein network clustering with reference phages. In the protein network, each shape represents a single viral population or reference phage, and shapes are connected by lines respective to shared protein content. Viral population taxonomy was coloured according to family level. The full untrimmed network is supplied as Supplementary Figure 3.



**Figure 3**

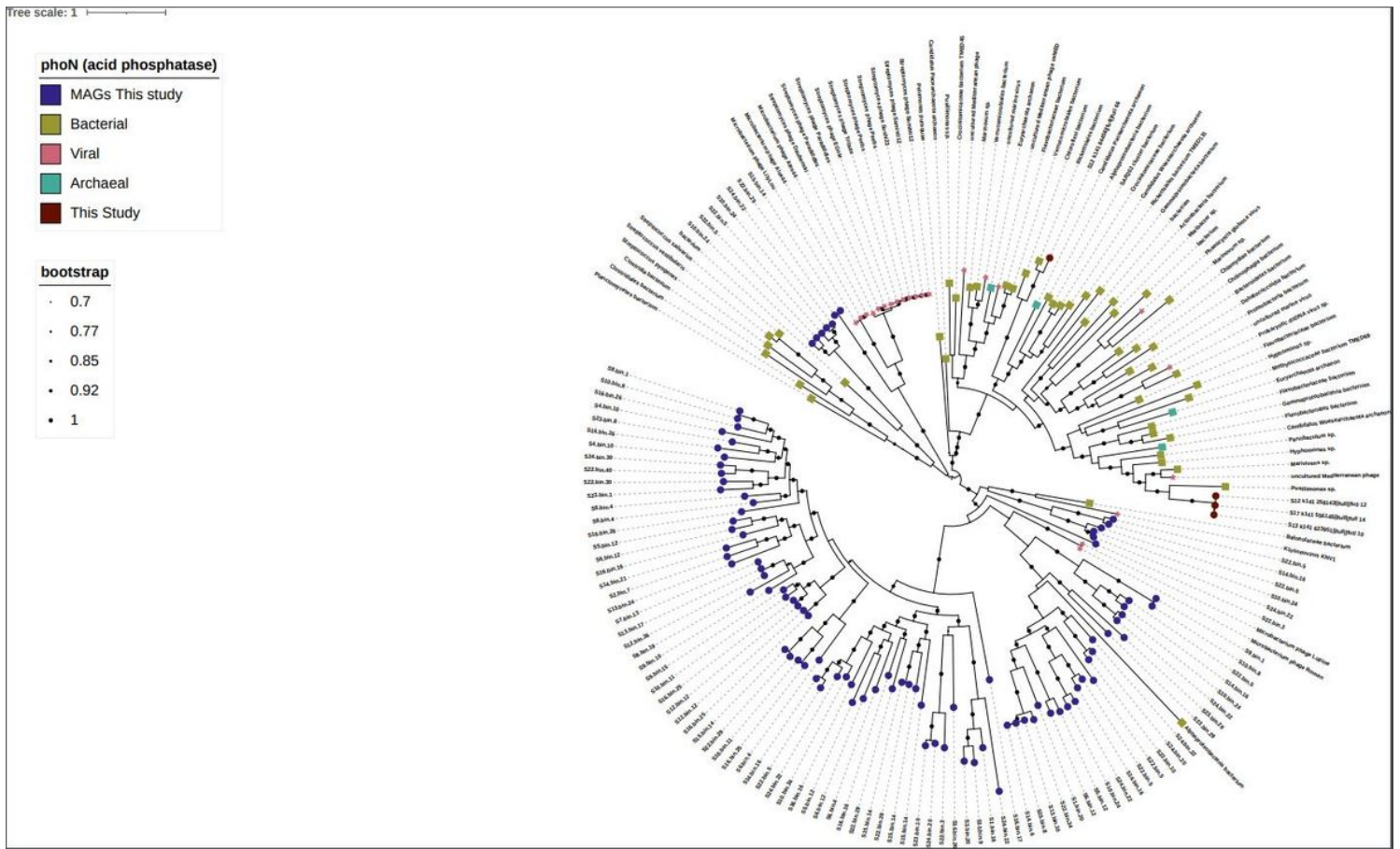
Viral population relative abundance of phosphorus AMGs containing contigs along the SAO. Relative abundance of AMG encoding viral contigs (coverage values normalized by metagenome size and contig length) detected in metagenomes from different SAO sampling stations.





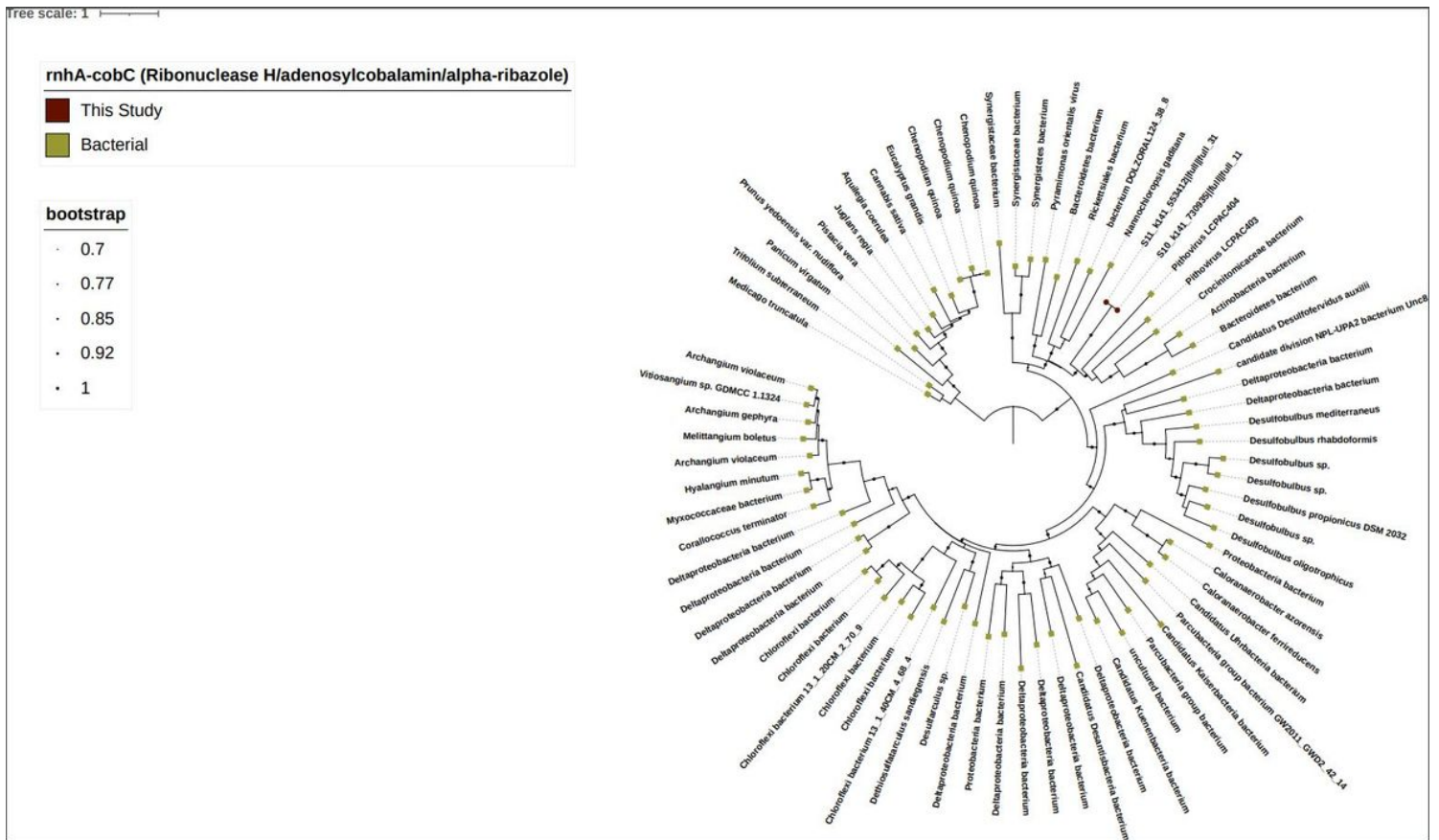
**Figure 4**

Genomic context and protein structure of P-cycle AMGs. a Genome architecture of P-cycle AMGs on viral contigs *rhoN* (acidic phosphatase), *gmhB* (D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase) and *rhoN-cobC* (ribonuclease H/adenosylcobalamin/alpha-ribazole phosphatase). Detailed annotations can be found in Supplementary Table 7. b tertiary structures of AMGs based on modelling using Phyre2 and Quaternary structures of AMGs based on modelling using SWISS\_MODEL (colour annotations online).



**Figure 5**

A maximum likelihood tree from an amino acid alignment of the bacterial like viral phoN found in SAO and reference microbial sequences. The viral AMG found in this study are circled in maroon. Internal nodes and SH-like supports are represented by proportional circles (all nodes with support <0.70 were collapsed)



**Figure 6**

A maximum likelihood tree from an amino acid alignment of the bacterial like viral rnhA-cobC found in SAO and reference microbial sequences. The viral AMG's found in this study are circled in marron. Internal nodes and SH-like supports are represented by proportional circles (all nodes with support <0.70 were collapsed).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.pdf](#)
- [SupplementaryFigure2.pdf](#)
- [SupplementaryFigure3.pdf](#)
- [SupplementaryFigure4.pdf](#)
- [SupplementaryTables.xlsx](#)