



# Auxiliary Metabolic Gene Functions in Pelagic and Benthic Viruses of the Baltic Sea

Benedikt Heyerhoff, Bert Engelen\* and Carina Bunse

Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany

Marine microbial communities are facing various ecosystem fluctuations (e.g., temperature, organic matter concentration, salinity, or redox regimes) and thus have to be highly adaptive. This might be supported by the acquisition of auxiliary metabolic genes (AMGs) originating from virus infections. Marine bacteriophages frequently contain AMGs, which allow them to augment their host's metabolism or enhance virus fitness. These genes encode proteins for the same metabolic functions as their highly similar host homologs. In the present study, we analyzed the diversity, distribution, and composition of marine viruses, focusing on AMGs to identify their putative ecologic role. We analyzed viruses and assemblies of 212 publicly available metagenomes obtained from sediment and water samples across the Baltic Sea. In general, the virus composition in both compartments differed compositionally. While the predominant viral lifestyle was found to be lytic, lysogeny was more prevalent in sediments than in the pelagic samples. The highest proportion of AMGs was identified in the genomes of *Myoviridae*. Overall, the most abundantly occurring AMGs are encoded for functions that protect viruses from degradation by their hosts, such as methylases. Additionally, some detected AMGs are known to be involved in photosynthesis, 7-cyano-7-deazaguanine synthesis, and cobalamin biosynthesis among other functions. Several AMGs that were identified in this study were previously detected in a large-scale analysis including metagenomes from various origins, i.e., different marine sites, wastewater, and the human gut. This supports the theory of globally conserved core AMGs that are spread over virus genomes, regardless of host or environment.

**Keywords:** bacteriophage, AMGs, salinity, marine, sediment

## OPEN ACCESS

### Edited by:

Minh-Thu Nguyen,  
University Hospital Münster, Germany

### Reviewed by:

Yosuke Nishimura,  
The University of Tokyo, Japan  
Xupeng Cao,  
Dalian Institute of Chemical Physics  
(CAS), China

### \*Correspondence:

Bert Engelen  
engelen@icbm.de

### Specialty section:

This article was submitted to  
Virology,  
a section of the journal  
Frontiers in Microbiology

Received: 27 January 2022

Accepted: 02 June 2022

Published: 07 July 2022

### Citation:

Heyerhoff B, Engelen B and  
Bunse C (2022) Auxiliary Metabolic  
Gene Functions in Pelagic  
and Benthic Viruses of the Baltic Sea.  
*Front. Microbiol.* 13:863620.  
doi: 10.3389/fmicb.2022.863620

## INTRODUCTION

Viruses are the most abundant biotic entities on Earth and are ubiquitous in the marine environment. Bacteriophages, viruses that infect bacteria, occur in concentrations of up to  $10^7$  viruses per ml marine surface waters, often outnumbering their hosts by 10-fold (Wommack and Colwell, 2000). Abundances of viruses in marine sediments are even higher with  $10^7$ - $10^{10}$  viral particles per g of dry sediment (Danovaro et al., 2008a). With an estimated number of  $10^{30}$  viruses in the world's oceans (Breitbart, 2012), viruses play an important role in controlling marine bacterial populations through virus-induced mortality and represent a substantial reservoir of genetic diversity (Suttle, 2007). The exact numbers of virus-induced mortality are environment-dependent but increase with water depth and are as high as 90% at depths below 1,000 m

(Danovaro et al., 2008b; Breitbart et al., 2018). Virus-induced mortality has major implications on global carbon and nutrient cycling, as it leads to a conversion of biomass to dissolved organic matter (DOM), enabling a reuptake by prokaryotes as well as preventing the transfer of DOM into higher trophic levels (Fuhrman, 1999; Wilhelm and Suttle, 1999; Suttle, 2005).

The Baltic Sea is one of the largest brackish water bodies on Earth, characterized by high rates of sedimentation (Ilus et al., 2001), high nutrient and DOC concentrations, and seasonal temperature variations of  $> 15^{\circ}\text{C}$  (Bunse et al., 2019), as well as substantial riverine influx of freshwater that establishes the north-southerly salinity gradient. These mechanisms result in a stratification of the Baltic Sea and a constant halocline at a water depth of 40–80 m (Vali et al., 2013). However, the connection to the North Sea allows inflow events of saline and oxygenated water to occur irregularly (Meier et al., 2006; Reissmann et al., 2009). Prolonged stagnation further divides, e.g., deep basins of the Baltic Sea into an oxygenated layer and underlying anoxic waters, separated by the pelagic redoxcline (Labrenz et al., 2007). This distinct zonation also divides bacterial mortality factors, such as grazing and viral lysis (Pernthaler, 2005). The majority of grazing occurs in oxygenated waters, while viral lysis becomes the predominant mortality factor in anoxic layers (Weinbauer et al., 2003; Kostner et al., 2017). The functional importance of viruses in the Baltic Sea becomes apparent at the phosphorous (P)-limited Ore Estuary in the northern Baltic Sea. Here, viral lysis is supplying the dissolved DNA pool with up to 25% of its total volume. The uptake of dissolved DNA covers up to 70% of the bacterioplankton's P-demand and thus supports their growth (Riemann et al., 2009). Stratification continues through Baltic Sea sediments, which are, like other marine sediments, vertically stratified and follow a redox gradient exhibiting decreasing free energy yield (Sørensen et al., 1979). High sedimentation rates and high organic matter concentrations hence harbor highly active bacterial communities and associated phages, even in deep subsurface sediments (Jørgensen et al., 2020). The most studied viruses of the Baltic Sea today are the bacteriophages of the Bacteroidetes phylum (Šulčius and Holmfeldt, 2016). Studies investigating other phyla such as Proteobacteria and Cyanobacteria remain scarce (Zeigler Allen et al., 2017; Nilsson et al., 2022, 2019).

A typical trait of marine bacteriophages is the ability to augment their host's metabolism through the promotion of auxiliary metabolic genes (AMGs) (Breitbart et al., 2007; Williamson et al., 2008). Viral AMGs are genes of high similarity to host homologs. They are introduced during viral infection and encode for the same metabolic functions as those proteins of the hosts they originate from (Thompson et al., 2011). AMGs were first discovered in marine heterotrophs and Cyanobacteria in the early 2000s (Rohwer et al., 2000; Mann et al., 2003; Lindell et al., 2004b). AMGs of cyanophages are associated with a variety of functions, such as energy conservation as part of Photosystem II (Mann et al., 2003; Sharon et al., 2011). Some cyanophage genomes contain over 20 AMGs that can alter the electron transport chain or enhance the carbon metabolism of their hosts (Hellweger, 2009; Sullivan et al., 2010; Thompson et al., 2011; Crummett et al., 2016). Other functions of AMGs include, e.g., the acceleration of nucleotide biosynthesis in roseophages or

sulfur oxidation genes in deep-sea viruses (Anantharaman et al., 2014; Zheng et al., 2021). Through contextual distribution and maintenance of particular AMGs in the environment, viruses increase their own fitness. While most AMGs seem to affect functions of global biogeochemical cycles, genes that increase host virulence occur as well. Here, the most famous example is the filamentous CTX bacteriophage, which carries the toxin that causes the virulence of *Vibrio cholerae* (Waldor and Mekalanos, 1996). In the past, AMG identification was performed through manual inspection and functional annotation. The advance toward scalable approaches in AMG identification through new bioinformatic tools has recently allowed for large-scale assessments across whole ecosystems and has further emphasized the ecological importance of viruses (Kieft et al., 2020, 2021). The aim of the present study was to examine the diversity and composition of benthic and pelagic viral assemblies across the Baltic Sea. We focused on how salinity as an environmental driver influenced their latitudinal distribution. Therefore, we downloaded and analyzed 212 publicly available Baltic Sea metagenomes from the National Center for Biotechnology Information (NCBI) sequence read archive (SRA). We separately analyzed the viral composition and distribution in Baltic Sea sediments and the water column. In the current study, we hypothesize that virus diversity differs in both compartments due to characteristic environmental factors. We further identified AMGs within the metagenomes and analyzed their composition and distribution along the north-southerly salinity gradient of the Baltic Sea. We hypothesize that the identified AMGs enhance the fitness of the viruses and putatively support their host.

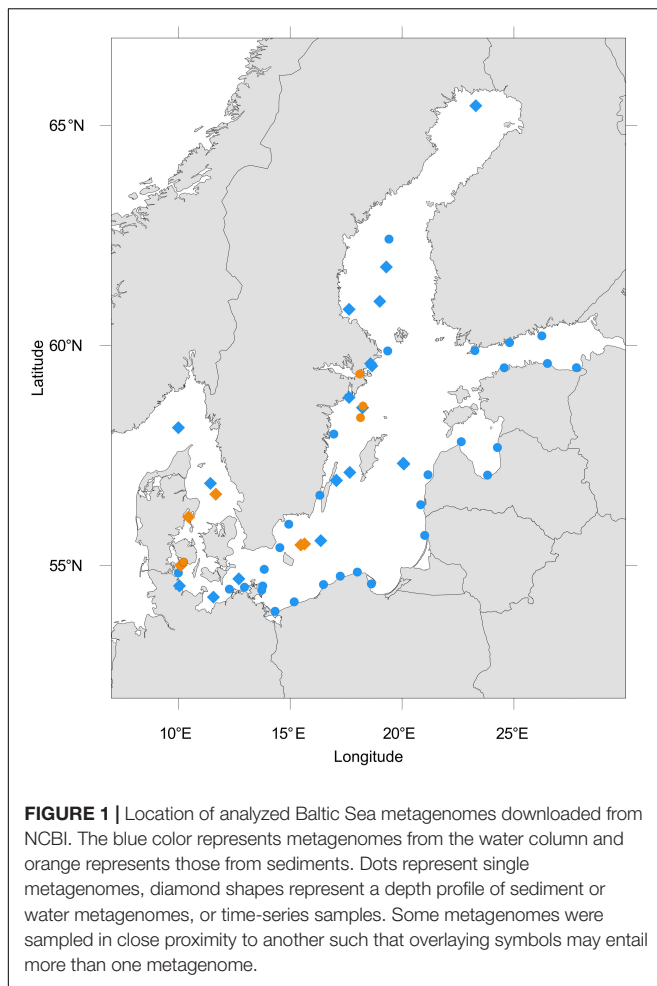
## MATERIALS AND METHODS

### Metagenomic Data From the Baltic Sea

We retrieved a total of 212 publicly available metagenomes from sequence-based metagenomic studies of Baltic Sea sediments and water samples (Figure 1). The included data result from the project IDs PRJEB22997 (Alneberg et al., 2018), PRJEB34883 (Alneberg et al., 2020), PRJEB6616 (Thureborn et al., 2016), PRJEB8682 (Kopf et al., 2015), PRJNA308531 (Andren et al., 2015), PRJNA273799 (Hugert et al., 2015), PRJNA297401, PRJNA322246 (Asplund-Samuelsson et al., 2016), PRJNA367442, PRJNA337783, PRJNA433242 (Zinke et al., 2017), and PRJNA337783 (Espínola et al., 2018), which were obtained from the NCBI SRA (accessed May–June 2020). The data originate from different sample sets that comprise different filter fractions, community members, and environments and likely also differ in sampling method as well as DNA extraction. The location of all metagenomic samples analyzed within this study were plotted using the R package oceanmaps (Bauer, 2020). All metadata and published environmental data available from these projects are summarized in **Supplementary Table 1**.

### Sequence Quality Analysis and Assembly

Sequence quality control analysis was performed using FastQC (Andrews, 2010). Metagenomic read files were then trimmed of adapters using BBDuk (Bushnell, 2018). Quality trimming was also performed using BBDuk with the quality threshold



set to Q30. High-quality metagenomes were assembled using MEGAHIT (Li et al., 2015) with the meta-large flag, as suggested for complex metagenomes. Statistics about the quality of assembled metagenomes were analyzed using MetaQUAST v5.0.2 (Mikheenko et al., 2016), **Supplementary Table 2**.

## Identification of Viral Contigs

The *in silico* prediction of phage scaffolds and viral AMGs was done on all 212 metagenomes using VIBRANT v1.2.1 (Kieft et al., 2020) with default settings. Vibrant can accurately recover viruses and AMGs by applying machine learning and a protein similarity approach. The quality of contigs identified by VIBRANT was further assessed with CheckV (Nayfach et al., 2021) retaining contigs > 3 kb and filtering any non-viral contigs. Abundance profiles of AMGs identified by VIBRANT were generated by mapping quality-controlled metagenome reads to the AMGs using Bowtie2 (v1.2.2) (Langmead and Salzberg, 2012). The sequence mapping files were handled and converted using SAMtools (v1.9-58) (Li et al., 2009). Fragments Per Kilobase per Million (FPKM) mapped reads were calculated as the number of mapped reads times  $10^9$  divided by the total number of mapped reads per sample multiplied by the gene length. Viral taxonomy of AMGs located on filtered contigs was assigned using DIAMOND

BLASTp v0.9.30 ( $E$ -value of  $< 0.0001$ , bit score  $\geq 50$ ) and the “—very-sensitive” preset (Buchfink et al., 2015). Viral hosts of identified contigs were assigned with VirHostMatcher-Net using default settings (Wang et al., 2020).

To gain broader context over the identified viral contigs, they were compared to viral contigs in the following public databases: (1) Global oceans virome (GOV) 2.0 Seawater (Gregory et al., 2019) and (2) Stordalen thawing permafrost (Emerson et al., 2018). Open reading frames for each viral contig were called using Prodigal V2.6.3 (Hyatt et al., 2010). Predicted protein sequences were used as input from vConTACT2 (Bin Jang et al., 2019). Viral Refseq (211) was used as a reference database (O’Leary et al., 2016). Diamond BLASTp was used for the protein-protein similarity method. All other parameters were set as default. The gene network was visualized using Cytoscape v3.9.1 (Shannon et al., 2003).

Inferring viral taxonomy through clusters identified by vConTACT2 resulted in small numbers of contigs that could be taxonomically assigned. Thus, to gain an overview of viral families present in Baltic Sea metagenomes, we used Kraken 2 to assign viral taxonomy from filtered high-quality unassembled reads applying default settings and using viral sequences from the NCBI non-redundant nucleotide database (release 211) as reference (O’Leary et al., 2016; Wood et al., 2019). Kraken 2 infers taxonomic classification by using exact k-mer matching and assigning query sequences to the lowest common ancestor. Accurate species abundance re-estimation was calculated using Bayesian Reestimation of Abundance with Kraken (Bracken) with default settings for all metagenomes (Lu et al., 2017).

## Statistical Analysis

Data processing and visualization were carried out with R version 4.0.5 (R Core Team, 2021) and the tidyverse package (Wickham et al., 2019). Bray–Curtis distances of relative viral abundances at each station were visualized by non-metric multidimensional scaling (NMDS) ( $k = 2$ ; 999 permutations) using the vegan (v2.5-7) package (Oksanen et al., 2013). The top nine most abundant virus families were fitted to the ordination using the vegan envfit-function with 999 permutations and removal of unavailable data enabled. Salinity isobars were added using the ordiplot function. Alpha diversity was calculated with the Shannon diversity index and centered log-ratio normalized counts using the phyloseq package in R (McMurdie and Holmes, 2013; Gloor et al., 2017). A Wilcoxon test was conducted to test the significant difference in alpha diversity between the viral composition of water and sediment metagenomes using the Wilcox test R function. Beta diversity was analyzed by using the Aitchinson distance by applying principal component analysis (PCA) to the centered log-ratio transformed counts. Zero counts were avoided by adding a pseudo count to avoid errors during clr transformation. Differential abundance testing was done using DESeq2 normalized counts (Love et al., 2014). A permutational multivariate analysis of variance (PERMANOVA; function Adonis, method = “Euclidean,” Permutations = 999) was done to test if beta diversity was significantly different in water or sediment metagenomes. The 20 most differentially abundant taxa with the smallest  $p$ .adj values ( $p$ .adj  $< 0.001$ ) were plotted

in a heatmap using the ComplexHeatmap package, using the dendextend R package for hierarchical clustering analysis (Galili, 2015; Gu et al., 2016).

## RESULTS

### The Viral Composition Differs Between Baltic Sea Sediments and Water Column Samples

In this study, we assembled and analyzed 212 publicly available metagenomes from Baltic Sea sediments and water column samples. The metagenomes were constructed from samples collected between 53°N and 65°N latitudes in the years 2008–2015. In total, 37 of the analyzed metagenomes originated from sediments and 175 from the water column samples. We identified 102,892 viral contigs > 3 kb after quality filtering with CheckV of which 7,540 contained at least one AMG (Supplementary Table 3).

We investigated the relationship of Baltic Sea viral contigs, with other publicly available viral sequences from different ecosystems (Figure 2A). Baltic Sea sediment and water column viral contigs as well as viral contigs from permafrost, GOV seawater, and RefSeq were grouped into 2,638 clusters (Supplementary Table 4). Viral contigs originating from the Baltic Sea water column overlapped relatively well with GOV seawater viruses; however, some small outliers occurred with the largest one displayed on the right side of the network graph (Figure 2A). The separate cluster of water column viruses was exclusively lytic but appeared throughout the Baltic Sea from 53° N to 65° N from surface water to 241.7 m water depth in the Skagerrak and seemed not to be impacted by salinity or temperature. None of the clusters contained vOTUs of all analyzed ecosystems. Rather, the Baltic Sea water column shared 35 clusters with GOV seawater and 11 clusters with Baltic Sea sediments but only 5 clusters contained vOTUs from Baltic Sea sediments, water column, and GOV seawater. Baltic Sea sediment shared 4 clusters with Stordalen permafrost and only very few vOTUs (0.4%) clustered with taxonomically known genomes from viral RefSeq. The limited number of clusters between viral genomes of analyzed ecosystems may reflect the high habitat specificity of viruses. The limited number of taxonomically identifiable viral genomes led us to use another method of taxonomic identification of viruses via Kraken 2 and allowed for a more detailed look at present viral families.

The Baltic Sea viral composition was versatile and locally differentiated. The most evenly distributed viruses in analyzed water and sediment samples were *Myoviridae*, showing abundances of around 20–40% (Figure 2B). One outlier was observed at 58°N, where they comprised 83.98% of the total viral composition. In sediments, the lowest abundance of *Myoviridae* occurred at around 55°N. *Siphoviridae* represented the second most abundant viral family and occurred in a less evenly distributed manner than *Myoviridae*. They dominated the sediment viral communities between 55°N and 56°N and

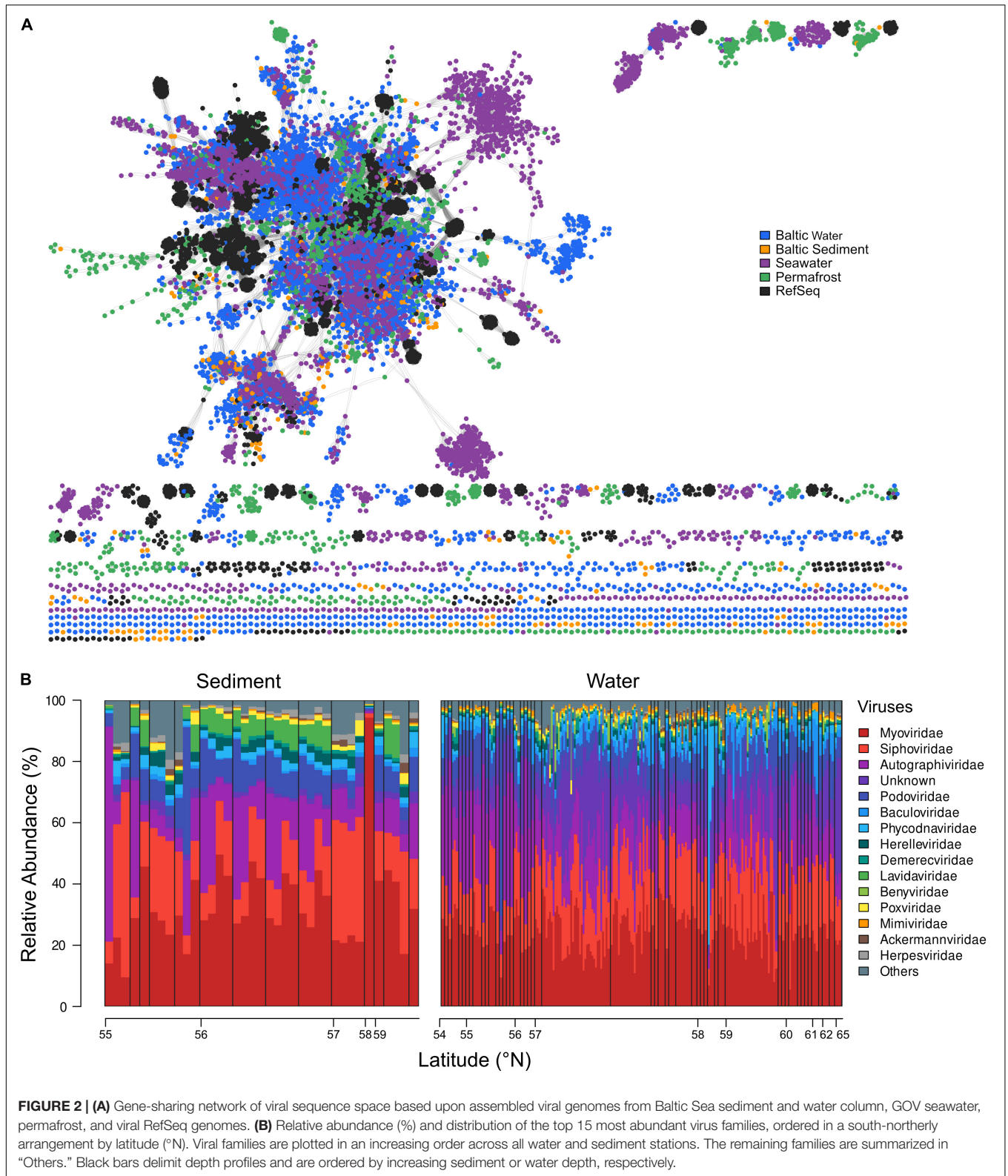
between 57° and 58°N, where they made up to 55% of the viral assemblies. In the water column, a larger fraction of unknown and *Phycodnaviridae* viral families were found compared to sediments, of which the latter accounted for more than 80% of some communities in the water column. Overall, we did not observe a major trend along the latitudinal gradient.

Members of the *Phycodnaviridae* family were the most differentially abundant viruses and distinguished pelagic from benthic viral assemblies. *Phycodnaviridae* infect *Bathycoccus*, *Micromonas*, and *Ostreococcus* genera, which belong to the green algae (Figure 3). Sediment stations from the Bornholm Basin and the Bay of Aarhus (SRR3081534, SRR3085416, SRR3085435, SRR3085585, SRR3089827, SRR3091743, SRR3095933, SRR3095939, and SRR7067081) sampled at depths of 0.75–3 m below sea floor (mbsf) displayed higher counts of the differentially abundant Mycobacterium phage Sparkdehlily. Deep subsurface stations (SRR12059190, SRR12059191, and SRR12059199) sampled at 24.1, 24.1, and 67.5 mbsf, close to the island of Anholt and the Little Belt, were defined by the differentially abundant Ralstonia Phage RSS30.

### Viruses in Sediments and Water Column are Similarly Diverse

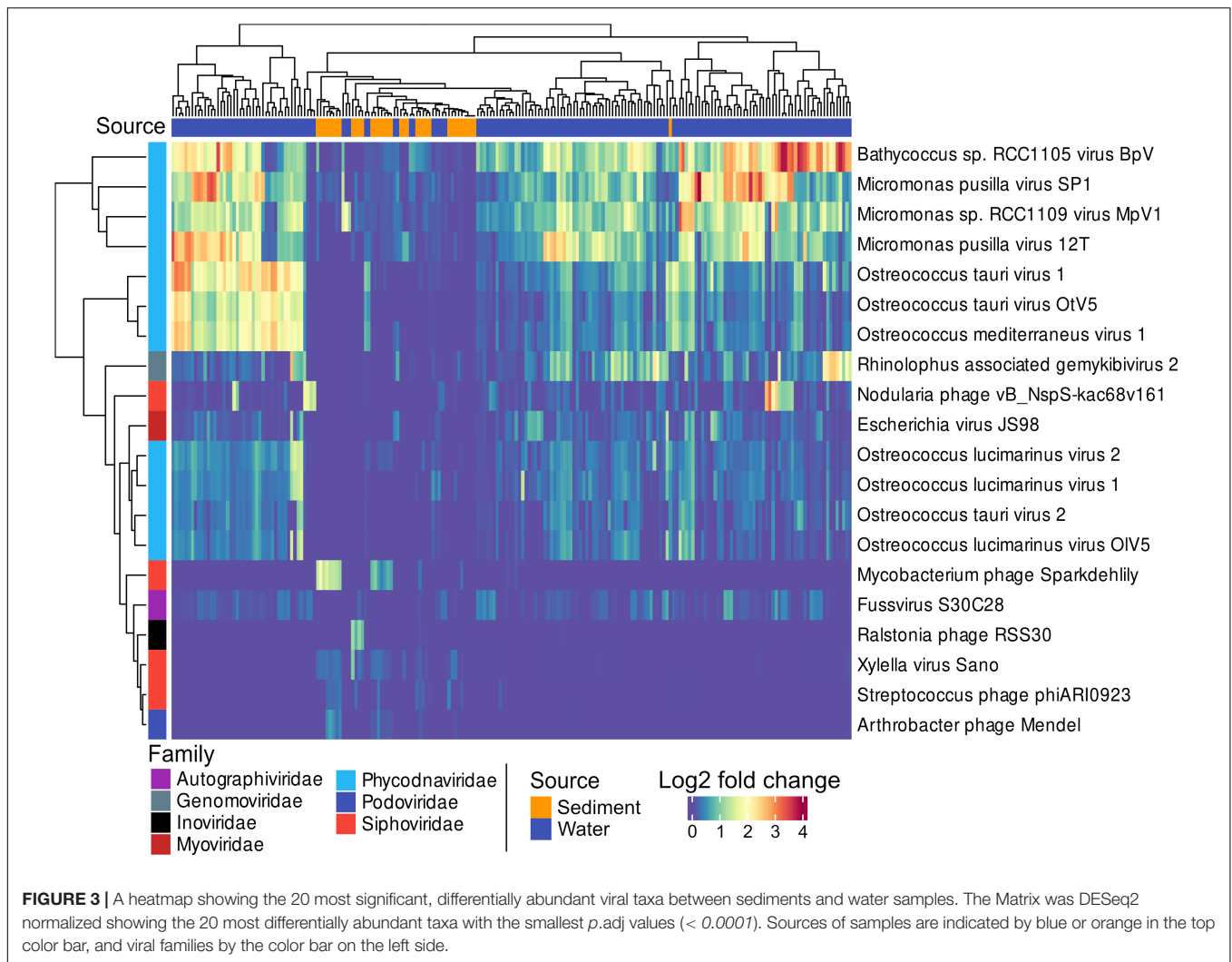
Beta diversity of viral communities revealed a distinct pattern in virus composition between sediment and water viral assemblies, explaining 22.6% of total variation (Adonis,  $p = 0.001$ ). The two plotted components separated viral taxa from sediments and water samples distinctively, with a small remaining overlap (Figure 4A). The median viral alpha diversities (Shannon index) of the sediment stations were 4.25, and 4.2 for water stations, indicating no statistically significant difference calculated by the Wilcoxon test ( $p$ -value = 0.7861) (Figure 4B). Due to the compositional nature of the metagenomic data used in this study, we assessed a possible batch effect via Bray–Curtis distance and visualized the results in an NMDS ordination (Supplementary Figure 3). We additionally conducted a PERMANOVA to test whether the different sequencing projects caused a batch effect within our NMDS ordinations. While some project-specific clustering could be observed within the ordination, the PERMANOVA showed these effects to be less important ( $R^2 = 0.30723$ ,  $P < 0.001$ ) (Supplementary Figure 3). The batch effect among just water and sediment stations was also not relevant ( $R^2 = 0.23696$ ,  $p < 0.001$  and  $R^2 = 0.28178$ ,  $p < 0.001$ ).

Viral community dissimilarity was investigated using non-metric multidimensional scaling (NMDS) analysis (Figure 4C). Samples close to the center of the ordination represented samples with similar viral compositions. Sediment samples spread more throughout the ordination and appeared more toward higher salinity, while water stations clustered closer together around the center of the ordination and around the 10 PSU isobar. The top nine most abundant viral families were plotted into the ordination. Among these, *Autographiviridae* and *Myoviridae* aligned more with the  $y$ -axis, whereas the other viral families aligned more with the cluster of samples that aligned with the  $x$ -axis in the lower left quadrant. The *Siphoviridae* and



*Phycodnaviridae* vectors were located somewhat separately from other viral families. While the lytic lifestyle aligned more with water stations, the lysogenic lifestyle appeared more in the

sediments. However, the lytic lifestyle was found to be the overall dominant viral lifestyle in the Baltic Sea as shown in **Supplementary Figure 1**.



### Viruses From Water and Sediment Carry Auxiliary Metabolic Genes Specific to the Environment

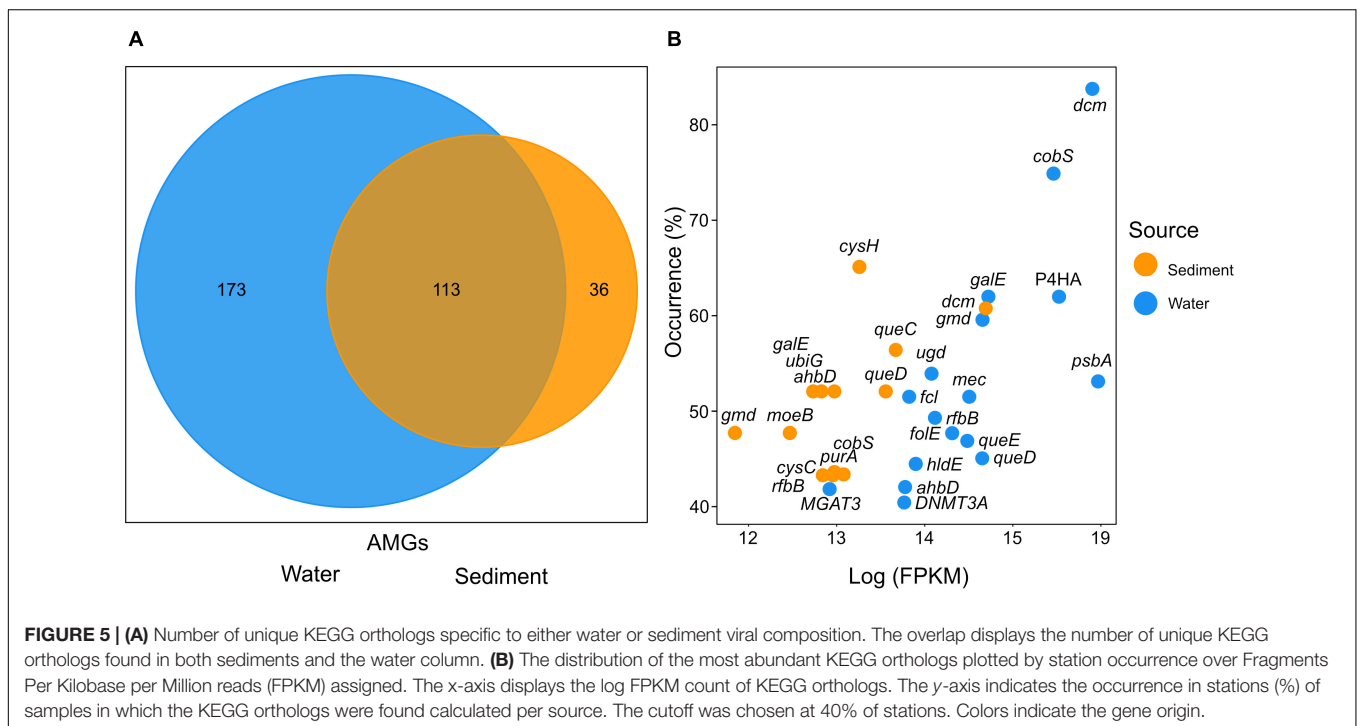
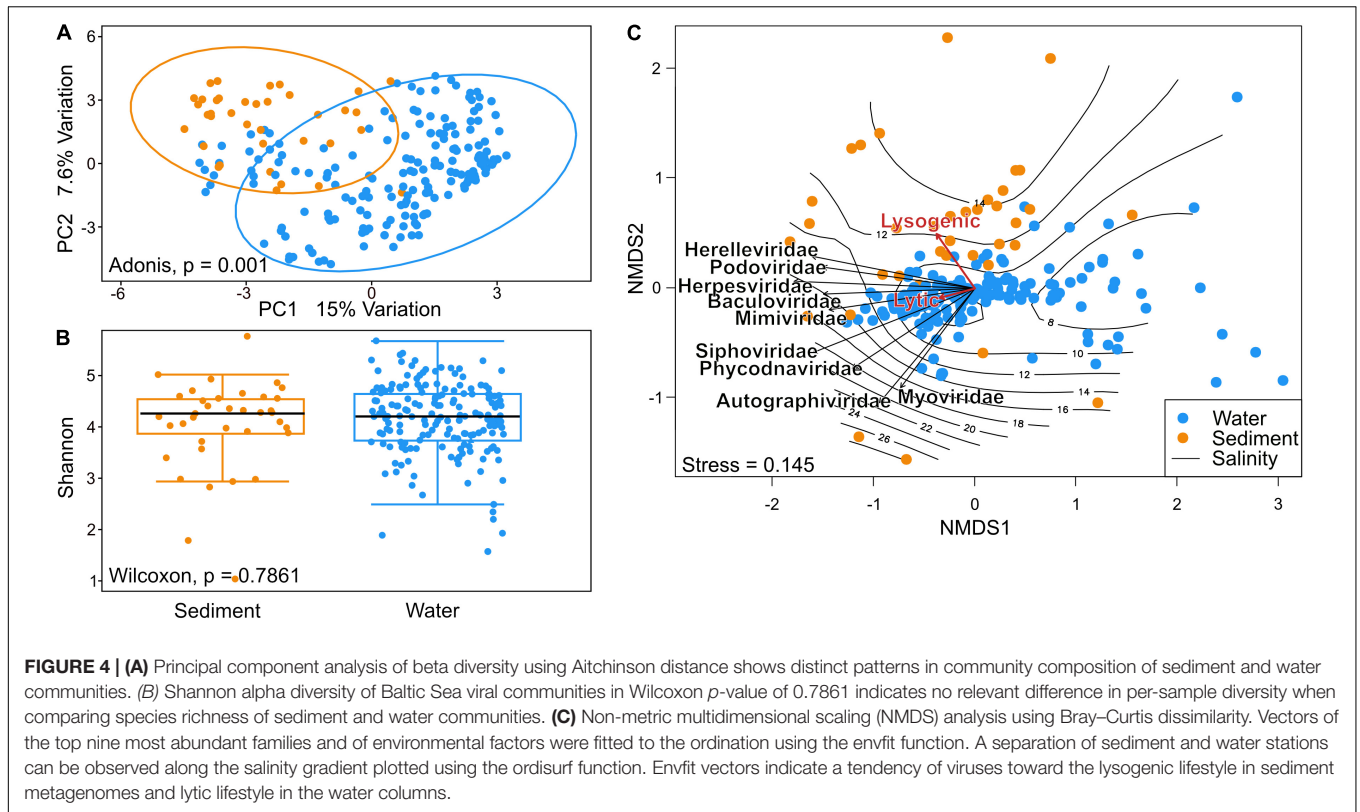
In both, the sediment and water metagenome AMGs could be assigned to 322 unique KEGG orthologs (Supplementary Table 5). The water column was the more diverse habitat with 173 unique KEGG orthologs assigned in total. While 36 unique KEGG orthologs could be identified in sediments, 113 identified AMGs were found in both habitats (Figure 5A). In the water column, AMGs of unknown function accounted for 13.1% of all mappable FPKM and 15.2% in sediments.

The most abundant AMGs are visualized by the percent occurrence of stations over their log sum of FPKM (Figure 5B and Table 1). Six outliers stand out among the AMGs identified in the water column: *dcm*, *cobS*, *gale*, *P4HA*, *gmd*, and *psbA*. The two most abundantly occurring AMGs in this study were encoding for *dcm/DNMT1* and *cobS* occurring in 84 and 75% of stations, while *gale*, *P4HA*, *gmd*, and *psbA* occurred in 62, 62, 60, and 54% of water respectively. In contrast, the most abundant

outliers of sediment AMGs, *cysH*, *dcm*, and *queC* occurred in 65, 61, and 57% of stations, respectively. These genes encode for the phosphoadenosine phosphosulfate reductase, DNA (cytosine-5)-methyltransferase 1, and the 7-cyano-7-deazaguanine synthase.

### Auxiliary Metabolic Gene Distribution Along the Salinity Gradient of the Baltic Sea

While the most predominant salinity gradient stretches from the south to the north, differences in salinity also occur by depth due to the higher density of saline water. “Amino acid” and “carbohydrate metabolism” AMGs as well as “Metabolism of cofactors and vitamins” appeared most consistently along the salinity gradient. The highest number of fragments were assigned to the “Energy metabolism” at salinities of 6.7, 6.71, and 12.3. However, while “Metabolism of cofactors and vitamins” did not occur in high numbers at certain salinities, 23.2% of all assigned FPKM were assigned to this pathway as it appears most consistently along the salinity gradient. It is closely followed by “Amino acid metabolism” with 18.1% and



“Energy metabolism” and unknown pathways with 14.9 and 14.1%, respectively (Figure 6A and Supplementary Table 6). PCA of Hellinger transformed AMG counts clustered most of the analyzed metagenomic samples evenly throughout the ordination

(Figure 6B). While most pathway vectors cluster relatively closely together, “Amino acid metabolism”, “Metabolism of cofactors and vitamins”, and “Carbohydrate metabolism” are separated from the other vectors.

**TABLE 1** | Most abundant KEGG orthologs with assigned metabolic pathway, the sum of assigned FPKM, and occurrence (% samples) in which the KEGG orthologs were found.

Gene name	KEGG ortholog	Log (FPKM)	Station occurrence (%)
<b>Sediment</b>			
<i>cysH</i>	K00390	12.85	65.2
<i>DNMT1, dcm</i>	K00558	14.29	60.9
<i>queC</i>	K06920	13.27	56.5
<i>ubiG</i>	K00568	12.42	52.2
<i>queD</i>	K01737	13.15	52.2
<i>galE</i>	K01784	12.33	52.2
<i>ahbD</i>	K22227	12.57	52.2
<b>Water</b>			
<i>DNMT1, dcm</i>	K00558	15.50	83.9
<i>cobS</i>	K09882	15.05	75.0
<i>P4HA</i>	K00472	15.12	62.1
<i>galE</i>	K01784	14.32	62.1
<i>gmd</i>	K01711	14.25	59.7
<i>ugd</i>	K00012	13.67	54.0
<i>psbA</i>	K02703	15.56	53.2
<i>fcl</i>	K02377	13.42	51.6
<i>mec</i>	K21140	14.10	51.6

## Cyanobacteria and Proteobacteria Are the Most Abundant Prokaryotic Hosts

Prokaryotic hosts of AMGs were assigned with the VirHostMatcher-Net tool, allowing the identification of hosts affected most by viral metabolic interference. In the water column, *Cyanobacteria* were found to be the hosts with the highest number of assigned FPKM with 43.9% of all FPKM assigned to them. *Proteobacteria* and *Bacteroidetes* followed closely with 34 and 20.2% of total assigned FPKM respectively. In sediments, most FPKM were assigned to *Proteobacteria* (53.5%), *Cyanobacteria* (24.1%), and *Bacteroidetes* (21.9%) (Figures 7A,B and Supplementary Table 7).

## Most FPKM Assigned to Auxiliary Metabolic Genes of Myo- and Siphoviridae

In the water column, 48.8% of FPKM were assigned to AMGs identified in the *Myoviridae* family and 21.7% in the *Siphoviridae* family. While some FPKM were also assigned to AMGs of other viral families (i.e., *Herelleviridae* and *Ackermannviridae*), they only accounted for 1.9 and 0.6%, respectively (Figure 7A). In the sediment, 40.5% of FPKM could be assigned to *Siphoviridae* and 23% to *Myoviridae*, while *Podoviridae* and *Herelleviridae* only accounted for 1.88 and 0.31%, respectively (Supplementary Table 7).

*Myoviridae* procured the most diverse pathways. AMGs assigned to this family were found in nine out of twelve KEGG pathways that are detected in all metagenomes. The second most versatile family of viruses was *Siphoviridae*, which carried AMGs belonging to 8 out of 12 of all KEGG pathways, while *Siphoviridae* carried AMGs from 7 out of 12 KEGG hierarchies.

Notably, more AMGs of the “Metabolism of cofactors and vitamins” pathways were present in *Myoviridae* in sediments, whereas AMGs of “Biosynthesis of other secondary metabolites” and “Energy metabolism” in the water columns were more pronounced compared to sediments. AMGs assigned to the “Amino acid metabolism” and “Metabolism of cofactors and vitamins” were the most abundantly occurring type of AMGs that were present in the top four most abundant viral families, which made up 98% of all assigned AMGs.

## DISCUSSION

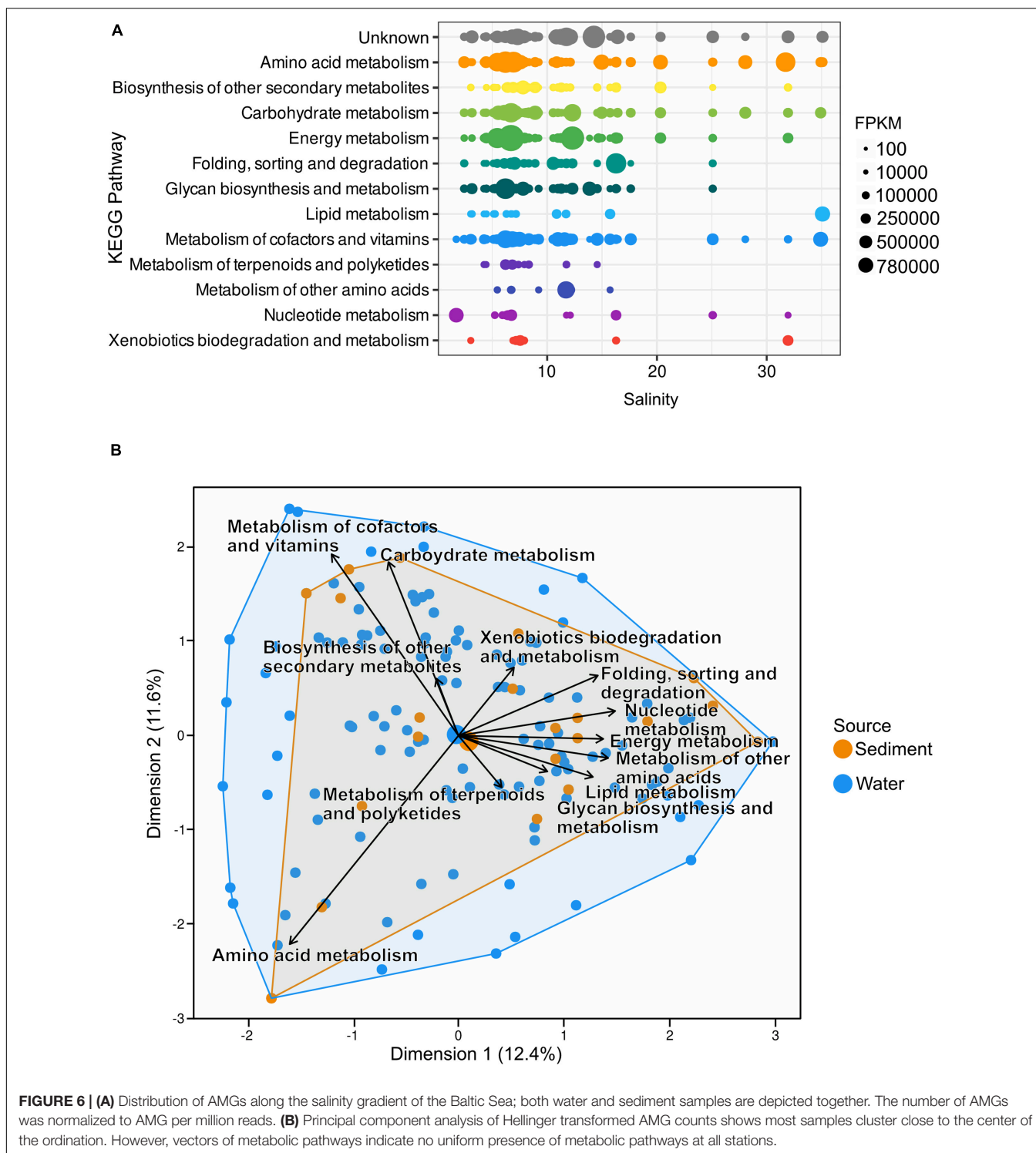
### Diverse Viral Composition in Sediments and the Water Column

In general, *Myoviridae* and *Siphoviridae* were detected as the most abundant viral families within the whole metagenomic data set. However, while some overlaps occur, the viral composition of sediment and water stations clustered separately from each other. Even though the viral assembly in sediments and the water columns were similar in species richness, they differed in beta diversity. Minor differences in viral diversity of the samples can be expected due to varying sampling methods and filter fractions used in the individual projects, that are summarized in this meta-analysis. Specifically, differences in the viral composition of sediments and the water column were defined through the differentially abundant *Phycodnaviridae*, which contributed up to 80% of total relative abundance at the surface. While in compositional datasets, relative abundances cannot be used to infer absolute viral absolute viral particles, cell counts, or gene abundances, we noted decreasing relative abundances of *Phycodnaviridae* with decreasing water depth. The high relative abundances of *Phycodnaviridae* in the subsurface sediments at 10.1 mbsf are likely the result of high sedimentation rates in the Baltic Sea and a lack of benthic animals due to limited oxygen supply, allowing the undisturbed formation of sediment layers. While most sediment stations in our study were similar to each other, small increases of *Ralstonia* phage RSS30 were observed in subsurface stations sampled from Aarhus Bay sediments, Denmark. *Cyanobacteria* in our study were abundant hosts, yet the identification of *Prochloraceae* as hosts likely resulted from a database bias, as previous studies have not detected them in the Baltic Sea (Bertos-Fortis et al., 2016; Celepli et al., 2017). This interpretation is supported by Celepli et al. (2017), who have generated hits of *Prochloraceae* in their metagenomic data set, while their 16S rRNA analysis also revealed the absence of *Prochloraceae*. Other unicellular *Cyanobacteria* (*Synechococcus* and *Cyanobium*) are frequently found in datasets of the Baltic Sea and likely contributed as cyanophage hosts also in these metagenomes (Haverkamp et al., 2009; Larsson et al., 2014; Broman et al., 2021).

### The Lytic Lifestyle Prevails in the Baltic Sea

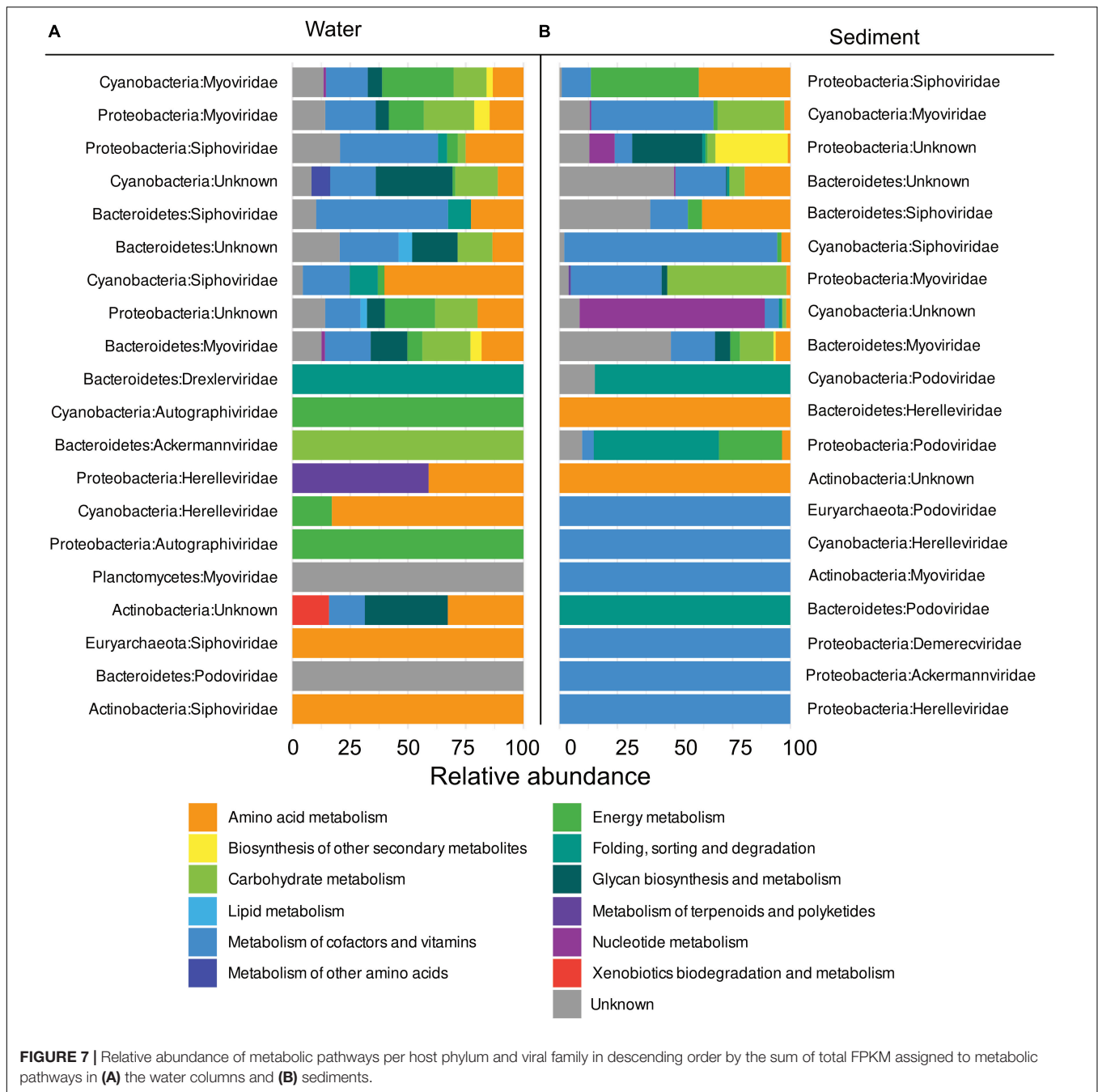
In the past, viruses have mostly been studied under laboratory conditions with a focus on three life cycles: chronic, lytic, and lysogenic. During the chronic lifestyle, phages enter





a productive replication cycle, releasing virions without lysing their host. Exhibiting the lytic lifestyle, viruses lyse their hosts upon infection, releasing viral progeny into the environment (Correa et al., 2021). In coastal waters, Wilcox and Fuhrman (1994) reported the lytic lifestyle to be the most abundant. During the lysogenic cycle, a non-productive

infection occurs by integrating the virus into the host's genome and replicating it along with the host. A virus may exit the lysogenic cycle and become lytic through specific factors or by spontaneous switching of lifestyles. Previous laboratory studies have looked at abiotic factors, where life lifestyle switching was influenced by phosphate availability or salinity



(Wilson et al., 1996; McDaniel et al., 2002; Bettarel et al., 2011).

However, while lytic, lysogenic, and chronic lifestyles are reflective of viral behavior in the laboratory, they are not entirely representative of natural behavior. Instead, viral lifestyles are controlled by complex interactions and represent a continuum rather than infection categories (Weitz et al., 2019; Correa et al., 2021). External mechanisms such as diel and seasonal changes may influence viral lifestyles (Ballaud et al., 2016; Brum et al., 2016; Puxty et al., 2018). However, recent studies postulate that

switching between lysogeny and lysis is especially influenced by host density (Erez et al., 2017). In the environment, lysogeny has been found to be a low-density refugium occurring at low host abundance. The refugium theory assumes exponentially growing communities to be rich in intracellular energy which favors lysis, whereas communities of low abundances are depleted of intracellular energy sources, which favors lysogeny (Erez et al., 2017). Yet, lysogeny has been found to be a survival strategy in low and high host-density conditions (Mizuno et al., 2016; Kim and Bae, 2018; Coutinho et al., 2019; Luo et al., 2020).

Lysogeny, as a result of high host density, has been described as the “piggyback-the-winner” model (Knowles et al., 2016). Additionally, the “killing the winner” theory predicts that hosts of the highest density are lysed (Thingstad, 2000). Low density and energy availability favor lysogeny, but increasing host density facilitates induction and lysis as denser communities administer more internal energy. Lysogeny continues to decrease until host densities of  $\sim 10^6$  cells  $\text{ml}^{-1}$  are reached, which are typically observed in the open oceans (Luque and Silveira, 2020). Higher densities increase the chances of coinfections with other viruses. Thus, lysogeny becomes a favorable lifestyle (Luque and Silveira, 2020). This switching might be communicated among phages as observed by Erez et al. (2017). Here, the *Bacillus* infecting phages of the SPbeta group released small peptides into the medium, signaling switching to the lysogenic lifestyle at higher concentrations of the respective compounds. This mechanism has been identified in different phages, each of which utilized different versions of the communication peptide (Erez et al., 2017).

Considering the abovementioned complexity of viral lifestyles, the observed dominance of identified lytic viral contigs in the Baltic Sea (**Supplementary Figure 1**) provides just a snapshot derived from genomic sequences of complex and dynamic systems. For instance, the metagenomic samples which are the basis of this study were taken at different time points and locations, thus allowing only remarks about the moment of sampling. Furthermore, the methodological limitation of the VIBRANT tool used to classify contigs as lytic or lysogenic has to be considered. VIBRANT assigns the lysogenic lifestyle by using surrounding host genome elements or integrases as evidence, limiting the identification of lysogenic contigs. The detected viral contigs might be lysogenic but the absence of the aforementioned properties in partial genomes could lead VIBRANT to falsely categorize them as lytic rather than lysogenic.

## Auxiliary Metabolic Genes Catalyze Virus-Host Interactions

Viruses utilize AMGs to alter their host's rate-limited cellular processes during infection (Sullivan et al., 2006). The roles of such AMGs are not random but critical for the successful proliferation of the viruses. Here, the most abundantly distributed AMG was the *dcm* gene, encoding for a methyltransferase. These enzymes are ubiquitously found in prokaryotes and are often associated with cognate restriction endonucleases, forming a restriction-modification system that protects bacterial cells from foreign DNA invasion. In bacteriophages, the so-called orphan methyltransferases appear without these endonucleases and are involved in regulatory activities to protect the phage DNA from being digested (Schlagman et al., 1986; Boye and Løbner-Olesen, 1990; Palmer and Marinus, 1994; Kosykh et al., 1995). While methylation is a well-known way of escaping host restriction in viruses, they also procured other means of nucleotide modifications. The genes *folE*, *queD*, *queE*, and *queC*, are necessary for 7-cyano-7-deazaguanine (preQ<sub>0</sub>) synthesis. These genes were among the most abundantly occurring AMGs in our dataset with the genes *queD* and *queC* both occurring

in 52% of sediment and *queE* in 46% of water samples. This indicates that preQ<sub>0</sub> is important in viral replication. Queuosine is a hypermodified guanosine found in tRNAs specific for four amino acids (Asp, Asn, His, Tyr) and increases translation efficiency (Sabri et al., 2011; El Yacoubi et al., 2012). The presence of preQ<sub>0</sub> synthesis genes in viruses has been reported previously and protects the virus from host restriction enzymes (Hutinet et al., 2019).

Photosynthesis genes, such as the *cobS* gene among the *psbA* gene, are considered core AMGs in cyanophages (Ignacio-Espinoza and Sullivan, 2012). In our study, these occurred especially abundantly in water samples. The *cobS* gene encodes for a protein that catalyzes the final step in bacterial cobalamin (vitamin B12) biosynthesis (Magnusdottir et al., 2015). Speculations about the involvement of viruses in the cobalamin biosynthesis in the pelagic ecosystem are tempting, yet more targeted analyses and experimental evidence would be needed for a conclusive answer. The *psbA* gene is among the best-studied AMGs. It encodes for the photosystem II protein D1. Together with photosystem II protein D2, it forms a heterodimer and binds P680, which is a specific chlorophyll a and the primary electron donor of photosystem II. Marine picocyanobacteria, such as those of the genus *Synechococcus*, are among the most abundant photosynthetic organisms on Earth and are responsible for the fixation of approximately 25% of the carbon in the marine environment (Scanlan et al., 2009; Flombaum et al., 2013). Viral production in *Cyanobacteria* is limited by the availability of energy for protein synthesis during late infection. Cyanophage production correlates with irradiation intensity and is inhibited by darkness (Puxty et al., 2016; Thompson et al., 2016). To circumvent energy limitations, cyanophages augment their hosts' metabolism by introducing genes for the photosynthetic light reactions. In the early stage of infection, CO<sub>2</sub> fixation can be actively inhibited by the phages, diverting the hosts' metabolism toward the pentose phosphate pathway, thus increasing NADPH and ribose-5-phosphate production, facilitating viral protein and DNA synthesis rather than increasing photosynthetic activity (Thompson et al., 2011). The regulatory ability of cyanophages on global carbon cycling and primary production through lysis and active augmentation of carbon fixation rates implies the importance of these phages. The AdoMet-dependent heme synthase *ahbD* is involved in protoheme biosynthesis by catalyzing the conversion of Fe-coproporphyrin III into heme. This has been studied in sulfate-reducing bacteria of the *Desulfivibrio* genus and in methanogenic Archaea (Buchenau et al., 2006). Heme is an essential prosthetic group and, among other biological processes such as respiration, is very important in photosynthesis (Layer et al., 2010). Procuring such genes could increase the energy metabolism and speed up virus production by reducing the latent period (Mann et al., 2003; Lindell et al., 2004a; Millard et al., 2004). However, while we found 261 instances of the *ahbD* AMG, only three contigs contained both, the radical SAM domain PF13186 and radical SAM domain PF04055, and were not classified as Archaea nor of *Desulfovibrio*. Hence, inferences about the function of this AMG are rather speculative. While the *ahbD* gene requires further investigation, the genes *psbA* and *cobS* highlight the importance and distribution of

AMGs involved in photosynthesis in pelagic phages, emphasized also by high cyanobacteria abundances in the Baltic Sea.

The *ubiG*, *galE*, and *P4HA* genes cannot easily be assigned to greater metabolic functions such as photosynthesis but appear to be of similar importance. The *ubiG* gene encoding for the last step in the pathway of ubiquinone biosynthesis likely provides the phages with the ability to affect the electron transport chain. The *galE* gene encoding the UDP-glucose 4-epimerase placed third among the most abundant AMGs in the water column. It mediates the conversion of UDP-galactose and UDP-glucose in galactose metabolism (Thoden and Holden, 1998). Thus, the introduction of *galE* likely allows the virus to participate in the carbohydrate metabolism to generate energy. The similarly abundant *P4HA* gene, encoding for the prolyl 4-hydroxylase, catalyzes the hydroxylation of proline residues in peptide linkages in collagens, forming 4-hydroxyproline (Myllyharju, 2003). In viruses, collagen can be part of the tail fibers and was first detected in *Paramecium bursaria* *Chlorella* virus-1 (Eriksson et al., 1999; Rasmussen et al., 2003). In which way viruses utilize prolyl 4-hydroxylase remains cryptic. However, biological consequences of prolyl hydroxylation include altering protein conformation and protein-protein interactions but also contributing to collagen-helix stability in general (Rao and Adams, 1979).

Analogous to water samples in our study, nucleotide modification through the *dcm* and *queC* gene are the most important functions provided by viruses. As photosynthesis is less relevant in sediments, especially those of greater depths, other functions prevail. Here, the most abundantly occurring AMG not involved in DNA modification is the *cysH* gene, encoding for the phosphoadenosine 5'-phosphosulfate reductase. It was found in 65% of all sediment stations but also in 39.5% of the water stations. The presence of the *cysH* gene suggests the viral involvement in sulfur cycling, especially in viruses found in Baltic Sea sediments. The enzyme is involved in the synthesis of sulfite from phosphoadenosine 5'-phosphosulfate (PAPS) and thus part of the sulfate reduction pathway (Bick et al., 2000). The *cysH* gene has been identified in phages infecting members of the SAR11 clade, which lack the phosphoadenosine 5'-phosphosulfate reductase and other genes required in assimilatory sulfate reduction but has recently also been found to be generally widespread in marine phages (Du et al., 2021; Kieft et al., 2021).

## Conserved Core Auxiliary Metabolic Genes

Recently, Kieft et al. (2020) identified a set of AMGs found in highly different viral assemblies from various origins, i.e., human gut, marine sediment, deep subsurface, and others. Their set of globally conserved AMGs consisted of the *dcm*, *cysH*, *folE*, *phnP*, *ubiG*, *ubiE*, *waaF*, *moeB*, *ahbD*, *cobS*, *mec*, *queE*, *queD*, *queC* genes and occurred in at least 10 out of 12 of their studied samples. These genes were also identified in the Baltic Sea, though not all of them are among the most abundant. However, the genes *dcm*, *cysH*, *folE*, *cobS*, *mec*, *queE*, *queD*, *queC* are concurrent with the findings of Kieft et al.

(2020) and suggest the existence of a globally distributed set of conserved core AMGs, which are present regardless of host and environment. Locally, the core AMGs might then be extended by genes specific to an environment of interest such as genes for photosynthesis. A definition of core AMGs is difficult, as setting a threshold for their occurrence at a given station or environment would be arbitrary. However, the similarity in composition of most abundantly occurring AMGs in the Baltic Sea and other environments is striking.

## CONCLUSION

The metagenomic analysis revealed a predominantly lytic viral life mode in the Baltic Sea, possibly aided by high nutrient availabilities and increasing lysogeny traits in sediments. We did not find major virus community differences along the north-southerly salinity gradient of the Baltic Sea. Yet, the composition of pelagic and benthic virus assemblies differed, especially in the relative abundances of *Phycodnaviridae*. Also, the functional virus AMGs differed between the pelagic and benthic samples. While viruses from the water columns procured AMGs specific for photosynthesis, viruses in sediments acquired AMGs that are part of the nutrient cycling pathways such as sulfur cycling. Other AMGs that exclusively occurred in sediments or the water column were found in low abundances and are likely linked to functions that specifically increase virus fitness in the respective ecosystem. Viruses use AMGs to evade host restriction mechanisms, i.e., by modifying their DNA through methylation or utilization of preQ<sub>0</sub>. These DNA modification AMGs were highly abundant in the Baltic Sea and have also been observed to be globally conserved. Our findings, therefore, strengthen the hypothesis on the existence of global core AMGs that are central to viral replication, regardless of environment and host.

## DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

BH performed the conceptualization, datamining, bioinformatics, and analyses. BE secured the funding. BH and BE wrote the manuscript. CB helped conceptualize the study and helped with data evaluation. All authors read and approved the manuscript.

## FUNDING

This research was supported by a grant from the German Research Foundation (DFG) as part of the Collaborative Research Center TRR51 Roseobacter. Funding to CB was generously

provided by the Ministry for Science and Culture of Lower Saxony Vorab grant “Ecology of Molecules, EcoMol.”

## ACKNOWLEDGMENTS

We thank the personnel and shipboard scientists for sampling and data provision of the public metagenomes included in our study: OSD-CONSORTIUM, Kungliga Tekniska Hogskolan Science for Life Laboratory, UNIMIB, Aarhus University, Stockholm University, JGI, University of Southern California, IODP Leg 347. We would like to express our gratitude to the scientist at the mentioned institutes who sampled and

sequenced the metagenomic data that is the basis of this study. We thank the reviewers for their constructive feedback, which helped improve our manuscript. Furthermore, we also thank Leon Dlugosch for his help with scripting and discussing general ideas.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.863620/full#supplementary-material>

## REFERENCES

- Alneberg, J., Bennke, C., Beier, S., Bunse, C., Quince, C., Ininbergs, K., et al. (2020). Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun. Biol.* 3:119. doi: 10.1038/s42003-020-0856-x
- Alneberg, J., Sundh, J., Bennke, C., Beier, S., Lundin, D., Hugerth, L. W., et al. (2018). Barm and balticmicrobedb, a reference metagenome and interface to meta-omic data for the baltic sea. *Sci. Data* 5:180146. doi: 10.1038/sdata.2018.146
- Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M., and Dick, G. J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344, 757–760. doi: 10.1126/science.1252229
- Andren, T., Barker Jørgensen, B., Cotterill, C., Green, S., and The IODP expedition 347 scientific party (2015). IODP expedition 347: Baltic sea basin paleoenvironment and biosphere. *Sci. Dril.* 20, 1–12.
- Andrews, S. (2010). *Babraham Bioinformatics-fastqc a Quality Control Tool for High Throughput Sequence Data*. Available online at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed April, 2020).
- Asplund-Samuelsson, J., Sundh, J., Dupont, C. L., Allen, A. E., McCrow, J. P., Celepli, N. A., et al. (2016). Diversity and expression of bacterial metacaspases in an aquatic ecosystem. *Front. Microbiol.* 7:1043. doi: 10.3389/fmicb.2016.01043
- Ballaud, F., Dufresne, A., Francez, A.-J., Colombet, J., Sime-Ngando, T., and Quaiser, A. (2016). Dynamics of viral abundance and diversity in a sphagnum-dominated peatland: temporal fluctuations prevail over habitat. *Front. Microbiol.* 6:1494. doi: 10.3389/fmicb.2015.01494
- Bauer, R. K. (2020). “Oceanmap: a Plotting Toolbox for 2D Oceanographic Data.” *R Package Version*.
- Bertos-Fortis, M., Farnelid, H. M., Lindh, M. V., Casini, M., Andersson, A., Pinhassi, J., et al. (2016). Unscrambling cyanobacteria community dynamics related to environmental factors. *Front. Microbiol.* 7:625. doi: 10.3389/fmicb.2016.00625
- Bettarel, Y., Bouvier, T., Bouvier, C., Carre, C., Desnues, A., Domaizon, I., et al. (2011). Ecological traits of planktonic viruses and prokaryotes along a full-salinity gradient. *FEMS Microbiol. Ecol.* 76, 360–372.
- Bick, J. A., Dennis, J. J., Zylstra, G. J., Nowack, J., and Leustek, T. (2000). Identification of a new class of 5'-adenylylsulfate (aps) reductases from sulfate-assimilating bacteria. *J. Bacteriol.* 182, 135–142. doi: 10.1128/JB.182.1.135-142.2000
- Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., et al. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* 37, 632–639. doi: 10.1038/s41587-019-0100-8
- Boye, E., and Løbner-Olesen, A. (1990). The role of dam methyltransferase in the control of dna replication in e. coli. *Cell* 62, 981–989. doi: 10.1016/0092-8674(90)90272-g
- Breitbart, M. (2012). Marine viruses: truth or dare. *Annu. Rev. Mar. Sci.* 4, 425–448. doi: 10.1146/annurev-marine-120709-142805
- Breitbart, M., Bonnain, C., Malki, K., and Sawaya, N. A. (2018). Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* 3, 754–766. doi: 10.1038/s41564-018-0166-y
- Breitbart, M., Thompson, L. R., Suttle, C. A., and Sullivan, M. B. (2007). Exploring the vast diversity of marine viruses. *Oceanography* 20, 135–139.
- Broman, E., Holmfeldt, K., Bonaglia, S., Hall, P. O., and Nascimento, F. J. (2021). Cyanophage diversity and community structure in dead zone sediments. *mSphere* 6:e00208-21. doi: 10.1128/mSphere.00208-21
- Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W., and Sullivan, M. B. (2016). Seasonal time bombs: dominant temperate viruses affect southern ocean microbial dynamics. *ISME J.* 10, 437–449.
- Buchenaus, B., Kahnt, J., Heinemann, I. U., Jahn, D., and Thauer, R. K. (2006). Heme biosynthesis in *Methanosarcina barkeri* via a pathway involving two methylation reactions. *J. Bacteriol.* 188, 8666–8668. doi: 10.1128/JB.01349-06
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using diamond. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Bunse, C., Israelsson, S., Baltar, F., Bertos-Fortis, M., Fridolfsson, E., Legrand, C., et al. (2019). High frequency multi-year variability in baltic sea microbial plankton stocks and activities. *Front. Microbiol.* 9:3296. doi: 10.3389/fmicb.2018.03296
- Bushnell, B. (2018). *Bbtools: A Suite of Fast, Multithreaded Bioinformatics Tools Designed for Analysis of DNA and RNA Sequence Data*. Available online at: <https://sourceforge.net/projects/bbmap/> (accessed April, 2020).
- Celepli, N., Sundh, J., Ekman, M., Dupont, C. L., Yooshep, S., Bergman, B., et al. (2017). Meta-omic analyses of baltic sea cyanobacteria: diversity, community structure and salt acclimation. *Environ. Microbiol.* 19, 673–686. doi: 10.1111/1462-2920.13592
- Correa, A. M., Howard-Varona, C., Coy, S. R., Buchan, A., Sullivan, M. B., and Weitz, J. S. (2021). Revisiting the rules of life for viruses of microorganisms. *Nat. Rev. Microbiol.* 19, 501–513. doi: 10.1038/s41579-021-00530-x
- Coutinho, F. H., Rosselli, R., and Rodríguez-Valera, F. (2019). Trends of microdiversity reveal depth-dependent evolutionary strategies of viruses in the Mediterranean. *mSystems* 4:e00554-19. doi: 10.1128/mSystems.00554-19
- Crummett, L. T., Puxty, R. J., Weihe, C., Marston, M. F., and Martiny, J. B. (2016). The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* 499, 219–229. doi: 10.1016/j.virol.2016.09.016
- Danovaro, R., Corinaldesi, C., Filippini, M., Fischer, U. R., Gessner, M. O., Jacquet, S., et al. (2008a). Viriobenthos in freshwater and marine sediments: a review. *Freshw. Biol.* 53, 1186–1213.
- Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., et al. (2008b). Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature* 454, 1084–1087. doi: 10.1038/nature07268
- Du, S., Qin, F., Zhang, Z., Tian, Z., Yang, M., Liu, X., et al. (2021). Genomic diversity, life strategies and ecology of marine htvc010p-type pelagiphages. *Microb. Genom.* 7:000596. doi: 10.1099/mgen.0.000596
- El Yacoubi, B., Bailly, M., and de Crecy-Lagard, V. (2012). Biosynthesis and function of posttranscriptional modifications of transfer rnas. *Annu. Rev. Genet.* 46, 69–95.
- Emerson, J. B., Roux, S., Brum, J. R., Bolduc, B., Woodcroft, B. J., Jang, H. B., et al. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* 3, 870–880. doi: 10.1038/s41564-018-0190-y
- Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., et al. (2017). Communication between viruses guides lysis-lysogeny decisions. *Nature* 541, 488–493.

- Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999). Evidence for 4-hydroxyproline in viral proteins: characterization of a viral prolyl 4-hydroxylase and its peptide substrates. *J. Biol. Chem.* 274, 22131–22134. doi: 10.1074/jbc.274.32.22131
- Espínola, F., Dionisi, H. M., Borglin, S., Brislawn, C. J., Jansson, J. K., Mac Cormack, W. P., et al. (2018). Metagenomic analysis of subtidal sediments from polar and subpolar coastal environments highlights the relevance of anaerobic hydrocarbon degradation processes. *Microb. Ecol.* 75, 123–139. doi: 10.1007/s00248-017-1028-5
- Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9824–9829. doi: 10.1073/pnas.1307701110
- Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548.
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31, 3718–3720. doi: 10.1093/bioinformatics/btv428
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* 8:2224. doi: 10.3389/fmicb.2017.02224
- Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., et al. (2019). Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177, 1109–1123.e14. doi: 10.1016/j.cell.2019.03.040
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi: 10.1093/bioinformatics/btw313
- Haverkamp, T. H., Schouten, D., Doeleman, M., Wollenzien, J., Ute, denm Huisman, et al. (2009). Colorful microdiversity of synechococcus strains (picocyanobacteria) isolated from the baltic sea. *ISME J.* 3, 397–408. doi: 10.1038/ismej.2008.118
- Hellweger, F. L. (2009). Carrying photosynthesis genes increases ecological fitness of cyanophage in silico. *Environ. Microbiol.* 11, 1386–1394. doi: 10.1111/j.1462-2920.2009.01866.x
- Hugerth, L. W., Larsson, J., Alneberg, J., Lindh, M. V., Legrand, C., Pinhassi, J., et al. (2015). Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* 16:279. doi: 10.1186/s13059-015-0834-7
- Hutin, G., Kot, W., Cui, L., Hillebrand, R., Balamkundu, S., Gnanakalai, S., et al. (2019). 7-Deazaguanine modifications protect phage DNA from host restriction systems. *Nat. Commun.* 10:5442. doi: 10.1038/s41467-019-13384-y
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Ignacio-Espinoza, J. C., and Sullivan, M. B. (2012). Phylogenomics of t4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ. Microbiol.* 14, 2113–2126. doi: 10.1111/j.1462-2920.2012.02704.x
- Ilus, E., Mattila, J., Klemola, S., Ikaheimonen, T., and Niemisto, L. (2001). *Sedimentation Rate in the Baltic Sea. Tech. Rep. NKS-8*. Available online at: <https://www.osti.gov/etdweb/servlets/purl/20226330> (accessed October, 2021).
- Jørgensen, B. B., Andren, T., and Marshall, I. P. (2020). Sub-seafloor biogeochemical processes and microbial life in the Baltic Sea. *Environ. Microbiol.* 22, 1688–1706.
- Kieft, K., Zhou, Z., and Anantharaman, K. (2020). Vibrant: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8:90. doi: 10.1186/s40168-020-00867-0
- Kieft, K., Zhou, Z., Anderson, R. E., Buchan, A., Campbell, B. J., Hallam, S. J., et al. (2021). Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat. Commun.* 12:3503. doi: 10.1038/s41467-021-23698-5
- Kim, M. S., and Bae, J. W. (2018). Lysogeny is prevalent and widely distributed in the murine gut microbiota. *ISME J.* 12, 1127–1141. doi: 10.1038/s41396-018-0061-9
- Knowles, B., Silveira, C. B., Bailey, B. A., Barott, K., Cantu, V. A., Cobián-Güemes, A. G., et al. (2016). Lytic to temperate switching of viral communities. *Nature* 531, 466–470.
- Kopf, A., Bicak, M., Kottmann, R., Schnetzer, J., Kostadinov, I., Lehmann, K., et al. (2015). The ocean sampling day consortium. *Gigascience* 4:27. doi: 10.1186/s13742-015-0066-5
- Kosykh, V. G., Schlagman, S. L., and Hattman, S. (1995). Phage t4 dna [n]-adenine6methyltransferase. overexpression, purification, and characterization. *J. Biol. Chem.* 270, 14389–14393.
- Kostner, N., Scharnreiter, L., Jürgens, K., Labrenz, M., Herndl, G. J., and Winter, C. (2017). High viral abundance as a consequence of low viral decay in the Baltic sea Redoxcline. *PLoS One* 12:e0178467. doi: 10.1371/journal.pone.0178467
- Labrenz, M., Jost, G., and Jürgens, K. (2007). Distribution of abundant prokaryotic organisms in the water column of the central Baltic Sea with an oxic-anoxic interface. *Aquat. Microb. Ecol.* 46, 177–190.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Larsson, J., Celepli, N., Ininbergs, K., Dupont, C. L., Yooseph, S., Bergman, B., et al. (2014). Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic sea. *ISME J.* 8, 1892–1903. doi: 10.1038/ismej.2014.35
- Layer, G., Reichelt, J., Jahn, D., and Heinz, D. W. (2010). Structure and function of enzymes in heme biosynthesis. *Protein Sci.* 19, 1137–1161. doi: 10.1002/pro.405
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (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
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004a). Photosynthesis genes in *Prochlorococcus* cyanophage. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11013–11018.
- Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004b). Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11013–11018. doi: 10.1073/pnas.0401526101
- Love, M., Anders, S., and Huber, W. (2014). Differential analysis of count data—the deseq2 package. *Genome Biol.* 15, 10–1186.
- Lu, J., Breitwieser, F. P., Thielen, P., and Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 3:e104.
- Luo, E., Eppley, J. M., Romano, A. E., Mende, D. R., and DeLong, E. F. (2020). Double-stranded DNA viroplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J.* 14, 1304–1315. doi: 10.1038/s41396-020-0604-8
- Luque, A., and Silveira, C. B. (2020). Quantification of lysogeny caused by phage coinfections in microbial communities from biophysical principles. *mSystems* 5:e00353-20. doi: 10.1128/mSystems.00353-20
- Magnusdottir, S., Ravcheev, D., de Crecy-Lagard, V., and Thiele, I. (2015). Systematic genome assessment of b-vitamin biosynthesis suggests co-operation among gut microbes. *Front. Genet.* 6:148. doi: 10.3389/fgene.2015.00148
- Mann, N. H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003). Marine ecosystems: Bacterial photosynthesis genes in a virus. *Nature* 424:741. doi: 10.1038/424741a
- McDaniel, L., Houchin, L., Williamson, S., and Paul, J. (2002). Lysogeny in marine synechococcus. *Nature* 415, 496–496. doi: 10.1038/415496a
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Meier, H., Feistel, R., Piechura, J., Arneborg, L., Burchard, H., Fiekas, V., et al. (2006). Ventilation of the baltic sea deep water: a brief review of present knowledge from observations and models. *Oceanologia* 48, 133–164.
- Mikheenko, A., Saveliev, V., and Gurevich, A. (2016). Metaquast: evaluation of metagenome assemblies. *Bioinformatics* 32, 1088–1090. doi: 10.1093/bioinformatics/btv697
- Millard, A., Clokie, M. R., Shub, D. A., and Mann, N. H. (2004). Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11007–11012. doi: 10.1073/pnas.0401478101

- Mizuno, C. M., Ghai, R., Saghai, A., López-García, P., and Rodríguez-Valera, F. (2016). Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7:e00805-16. doi: 10.1128/mBio.00805-16
- Myllyharju, J. (2003). Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* 22, 15–24. doi: 10.1016/s0945-053x(03)00006-4
- Nayfach, S., Camargo, A. P., Schulz, F., Eloë-Fadrosch, E., Roux, S., and Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* 39, 578–585. doi: 10.1038/s41587-020-00774-7
- Nilsson, E., Li, K., Fridlund, J., Sulčius, S., Bunse, C., Karlsson, C. M., et al. (2019). Genomic and seasonal variations among aquatic phages infecting the baltic sea *Gammmaproteobacterium Rheinheimera* sp. strain BAL341. *Appl. Environ. Microbiol.* 85:e01003-19. doi: 10.1128/AEM.01003-19
- Nilsson, E., Li, K., Hoetzinger, M., and Holmfeldt, K. (2022). Nutrient driven transcriptional changes during phage infection in an aquatic gammaproteobacterium. *Environ. Microbiol.* 24, 2270–2281. doi: 10.1111/1462-2920.15904
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., et al. (2013). *Package 'vegan'. Community Ecology Package, version 2.*
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciuflo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189
- Palmer, B. R., and Marinus, M. G. (1994). The dam and DCM strains of *Escherichia coli*—a review. *Gene* 143, 1–12. doi: 10.1016/0378-1119(94)90597-5
- Pernthaler, J. (2005). Predation on prokaryotes in the water column and its ecological implications. *Nat. Rev. Microbiol.* 3, 537–546. doi: 10.1038/nrmicro1180
- Puxty, R. J., Evans, D. J., Millard, A. D., and Scanlan, D. J. (2018). Energy limitation of cyanophage development: implications for marine carbon cycling. *ISME J.* 12, 1273–1286. doi: 10.1038/s41396-017-0043-3
- Puxty, R. J., Millard, A. D., Evans, D. J., and Scanlan, D. J. (2016). Viruses inhibit CO<sub>2</sub> fixation in the most abundant phototrophs on Earth. *Curr. Biol.* 26, 1585–1589. doi: 10.1016/j.cub.2016.04.036
- R Core Team (2021). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.
- Rao, N. V., and Adams, E. (1979). Collagen helix stabilization by hydroxyproline in (ala-hyp-gly)<sub>n</sub>. *Biochem. Biophys. Res. Commun.* 86, 654–660. doi: 10.1016/0006-291x(79)91763-7
- Rasmussen, M., Jacobsson, M., and Björck, L. (2003). Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* 278, 32313–32316.
- Reissmann, J. H., Burchard, H., Feistel, R., Hagen, E., Lass, H. U., Mohrholz, V., et al. (2009). Vertical mixing in the Baltic Sea and consequences for eutrophication - A review. *Prog. Oceanogr.* 82, 47–80.
- Riemann, L., Holmfeldt, K., and Titelman, J. (2009). Importance of viral lysis and dissolved DNA for bacterioplankton activity in a P-limited estuary, Northern Baltic sea. *Microb. Ecol.* 57, 286–294. doi: 10.1007/s00248-008-9429-0
- Rohwer, F., Segall, A., Steward, G., Seguritan, V., Breitbart, M., Wolven, F., et al. (2000). The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol. Oceanogr.* 45, 408–418.
- Sabri, M., Hauser, R., Ouellette, M., Liu, J., Dehbi, M., Moeck, G., et al. (2011). Genome annotation and intraviral interactions for the streptococcus pneumoniae virulent phage dp-1. *J. Bacteriol.* 193, 551–562.
- Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* 73, 249–299. doi: 10.1128/MMBR.00035-08
- Schlagman, S. L., Hattman, S., and Marinus, M. G. (1986). Direct role of the *Escherichia coli* dam DNA methyltransferase in methylation-directed mismatch repair. *J. Bacteriol.* 165, 896–900.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sharon, I., Battchikova, N., Aro, E.-M., Giglione, C., Meinel, T., Glaser, F., et al. (2011). Comparative metagenomics of microbial traits within oceanic viral communities. *ISME J.* 5, 1178–1190. doi: 10.1038/ismej.2011.2
- Sørensen, J., Jørgensen, B. B., and Revsbech, N. P. (1979). A comparison of oxygen, nitrate, and sulfate respiration in coastal marine sediments. *Microb. Ecol.* 5, 105–115. doi: 10.1007/BF02010501
- Sulčius, S., and Holmfeldt, K. (2016). Viruses of microorganisms in the baltic sea: current state of research and perspectives. *Mar. Biol. Res.* 12, 115–124.
- Sullivan, M. B., Huang, K. H., Ignacio-Espinoza, J. C., Berlin, A. M., Kelly, L., Weigele, P. R., et al. (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* 12, 3035–3056. doi: 10.1111/j.1462-2920.2010.02280.x
- Sullivan, M. B., Lindell, D., Lee, J. A., Thompson, L. R., Bielawski, J. P., and Chisholm, S. W. (2006). Prevalence and evolution of core photosystem ii genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* 4:e234. doi: 10.1371/journal.pbio.0040234
- Suttle, C. (2005). Crystal ball. The virosphere: the greatest biological diversity on Earth and driver of global processes. *Environ. Microbiol.* 7, 481–482. doi: 10.1111/j.1462-2920.2005.803\_11.x
- Suttle, C. A. (2007). Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* 5, 801–812. doi: 10.1038/nrmicro1750
- Thingstad, T. F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* 45, 1320–1328.
- Thoden, J. B., and Holden, H. M. (1998). Dramatic differences in the binding of udp-galactose and udp-glucose to udp-galactose 4-epimerase from *Escherichia coli*. *Biochemistry* 37, 11469–11477. doi: 10.1021/bi9808969
- Thompson, L. R., Zeng, Q., and Chisholm, S. W. (2016). Gene expression patterns during light and dark infection of *Prochlorococcus* by cyanophage. *PLoS One* 11:e0165375. doi: 10.1371/journal.pone.0165375
- Thompson, L. R., Zeng, Q., Kelly, L., Huang, K. H., Singer, A. U., Stubbe, J. A., et al. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 108, E757–E764. doi: 10.1073/pnas.1102164108
- Thureborn, P., Franzetti, A., Lundin, D., and Sjöling, S. (2016). Reconstructing ecosystem functions of the active microbial community of the baltic sea oxygen depleted sediments. *PeerJ* 4:e1593.
- Vali, G., Meier, M., and Elken, J. (2013). Simulated halocline variability in the baltic sea and its impact on hypoxia during 1961-2007. *J. Geophys. Res. Oceans* 118, 6982–7000.
- Waldor, M. K., and Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 272, 1910–1913. doi: 10.1126/science.272.5270.1910
- Wang, W., Ren, J., Tang, K., Dart, E., Ignacio-Espinoza, J. C., Fuhrman, J. A., et al. (2020). A network-based integrated framework for predicting virus-prokaryote interactions. *NAR Genom. Bioinform.* 2:lqaa044. doi: 10.1093/nargab/lqaa044
- Weinbauer, M. G., Brettar, I., and Höfle, M. G. (2003). Lysogeny and virus-induced mortality of bacterioplankton in surface, deep, and anoxic marine waters. *Limnol. Oceanogr.* 48, 1457–1465.
- Weitz, J. S., Li, G., Gulbudak, H., Cortez, M. H., and Whitaker, R. J. (2019). Viral invasion fitness across a continuum from lysis to latency. *Virus Evol.* 5:vez006. doi: 10.1093/ve/vez006
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686.
- Wilcox, R. M., and Fuhrman, J. A. (1994). Bacterial viruses in coastal seawater: lytic rather than lysogenic production. *Mar. Ecol. Prog. Ser.* 114, 35–45.
- Wilhelm, S. W., and Suttle, C. A. (1999). Viruses and nutrient cycles in the sea. *Bioscience* 49, 781–788.
- Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., et al. (2008). The sorcerer II global ocean sampling expedition: Metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* 3:1456. doi: 10.1371/journal.pone.0001456
- Wilson, W. H., Carr, N. G., and Mann, N. H. (1996). The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *synechococcus* sp. wh7803 1. *J. Phycol.* 32, 506–516.

- Wommack, K. E., and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69–114. doi: 10.1128/MMBR.64.1.69-114.2000
- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0
- Zeigler Allen, L., McCrow, J. P., Ininbergs, K., Dupont, C. L., Badger, J. H., Hoffman, J. M., et al. (2017). The baltic sea virome: diversity and transcriptional activity of DNA and RNA viruses. *mSystems* 2:e00125-16. doi: 10.1128/mSystems.00125-16
- Zheng, X., Liu, W., Dai, X., Zhu, Y., Wang, J., Zhu, Y., et al. (2021). Extraordinary diversity of viruses in deep-sea sediments as revealed by metagenomics without prior virion separation. *Environ. Microbiol.* 23, 728–743. doi: 10.1111/1462-2920.15154
- Zinke, L. A., Mullis, M. M., Bird, J. T., Marshall, I. P., Jørgensen, B. B., Lloyd, K. G., et al. (2017). Thriving or surviving? evaluating active microbial guilds in baltic sea sediment. *Environ. Microbiol. Rep.* 9, 528–536. doi: 10.1111/1758-2229.12578

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Heyerhoff, Engelen and Bunse. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.