

# Arctic Ocean virus communities and their seasonality, bipolarity, and prokaryotic associations

Received: 3 September 2024

Accepted: 25 June 2025

Published online: 11 July 2025

Alyzza M. Calayag<sup>1,2</sup>, Taylor Priest<sup>1,3</sup>, Ellen Oldenburg<sup>1,4,5,6</sup>, Jan Muschiel<sup>1</sup>, Ovidiu Popa<sup>5</sup>, Matthias Wietz<sup>1,4,7,8</sup> & David M. Needham<sup>1,2</sup> ✉

Viruses of microbes play important roles in ocean environments as agents of mortality and genetic transfer, influencing ecology, evolution and biogeochemistry. However, we know little about the diversity, seasonality, and host interactions of viruses in polar waters. Here, we study dsDNA viruses in the Arctic Fram Strait across four years via 47 long-read metagenomes of the cellular size-fraction. Among 5662 vOTUs, 98% and 2% are *Caudoviricetes* and *Megaviricetes*, respectively. Viral coverage is, on average, 5-fold higher than cellular coverage, and 8-fold higher in summer. Viral community composition shows annual peaks in similarity and strongly correlates with prokaryotic community composition. Using network analysis, we identify putative virus-host interactions and six ecological modules associated with distinct environmental conditions. The network reveals putative novel cyanophages with time-lagged correlations to their hosts (in late summer) as well as diverse viruses correlated with *Flavobacteriaceae*, *Pelagibacteraceae*, and *Nitrosopumilaceae*. Via global metagenomes, we find that 42% of Fram Strait vOTUs peak in abundance in high latitude regions of both hemispheres, and encode proteins with biochemical signatures of cold adaptation. Our study reveals a rich diversity of polar viruses with pronounced seasonality, providing a foundation for understanding viral regulation and ecosystem impacts in changing polar oceans.

Polar regions are subject to the strongest seasonal cycles on Earth, and experience intense pressure from climate change<sup>1,2</sup>. The functioning of polar ecosystems is critical to biogeochemical cycles<sup>3–5</sup>, and is under marked ecological and evolutionary constraints<sup>6</sup>. Such ecosystem processes are strongly driven by microorganisms and their interactions, including viral dynamics. In addition to cell death, viruses also drive evolution via gene exchange, frequency-dependent selection<sup>7,8</sup>, and transmission and expression of auxiliary metabolic genes<sup>9–11</sup>.

Therefore, characterizing the diversity and dynamics of viruses is paramount for understanding the function and stability of polar ocean ecosystems.

Relative to more accessible temperate, subtropical, and tropical oceans, few studies have examined virus diversity and ecology in polar waters<sup>12,13</sup>. It is known, however, that polar viromes are distinct from their warm-water counterparts<sup>13–22</sup> and are typically dominated by bacteriophages with a high level of diversity<sup>14</sup>. In addition to spatial

<sup>1</sup>GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany. <sup>2</sup>Faculty of Mathematics and Natural Sciences, Kiel University, Kiel, Germany. <sup>3</sup>Institute of Microbiology, ETH Zurich, Zurich, Switzerland. <sup>4</sup>Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany. <sup>5</sup>Institute of Quantitative and Theoretical Biology, Heinrich Heine University, Düsseldorf, Germany. <sup>6</sup>Cluster of Excellence on Plant Sciences, Heinrich Heine University, Düsseldorf, Germany. <sup>7</sup>Max Planck Institute for Marine Microbiology, Bremen, Germany. <sup>8</sup>Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Oldenburg, Germany. ✉e-mail: [dneedham@geomar.de](mailto:dneedham@geomar.de)

structuring, polar viral communities also shift over time, for example in the Antarctic, following bloom dynamics and assembling into distinct communities across seasons<sup>23</sup>. One year-round study showed strong seasonal variation among virus communities, with sharp decreases in virus-to-prokaryote ratios at the onset of the spring bloom and highest ratios in winter<sup>22</sup>. Others have reported increases during spring-summer, and highest abundances of viruses in winter<sup>24,25</sup>. Some evidence suggests that different lifestyles of viruses (i.e., lytic vs. lysogenic) may exhibit distinct dynamics across seasons, with lytic infection more prevalent in the spring bloom<sup>26</sup>.

However, due to the challenges of continuous sampling in the polar regions across multiple years, the degree of seasonality among polar viruses—here, meaning annually repeating patterns of populations and communities across the same seasons of different years<sup>27</sup>—and the potential ecological implications remain to be discovered. Such seasonality has been previously observed in marine temperate environments<sup>28–30</sup>. In the Arctic, time-series have been critical to advancing an understanding of the biological carbon pump<sup>31</sup>, elucidating benthopelagic coupling and biotic interactions<sup>32</sup>, both in the water column<sup>33</sup> and on sinking particles<sup>34</sup>. Furthermore, continuous sampling, over the long-term, can discern the impact of ‘Arctic Atlantification’<sup>35–37</sup>: the northward expansion of subarctic habitats through the Fram Strait—the major connection between Atlantic and Arctic Oceans<sup>38</sup>. Here, polar water outflowing the central Arctic Ocean via the East Greenland Current (EGC) meets the West Spitsbergen Current (WSC), transporting warmer Atlantic water into the Arctic Ocean<sup>39–41</sup>. In the WSC, prokaryotic communities exhibit pronounced seasonality, underpinned by changes in photosynthetically active radiation (PAR) and mixed layer depth<sup>42</sup>.

Given this, and the tight coupling of viruses and their hosts<sup>9–11</sup>, we aim to examine the degree that viral populations are seasonally structured in the Arctic. The strong seasonal gradients in the Arctic provide an opportunity to observe host-virus dynamics and how they relate to prevailing conceptual models such as the Piggyback-the-winner<sup>43–45</sup>, Constant-Diversity<sup>46,47</sup>, and Red-Queen<sup>29,48</sup> hypotheses. We examine the diversity and seasonality of dsDNA virus communities in the Arctic Fram Strait over four complete annual cycles at roughly monthly resolution. This time-series of virus communities in the Arctic demonstrates considerable seasonality in viral diversity, their association with environmental conditions and potential microbial hosts, and their distribution across the global ocean.

## Results

We examined 47 long-read metagenomes from samples collected at near-monthly resolution over a four-year period (Sep 2016–Jul 2020)

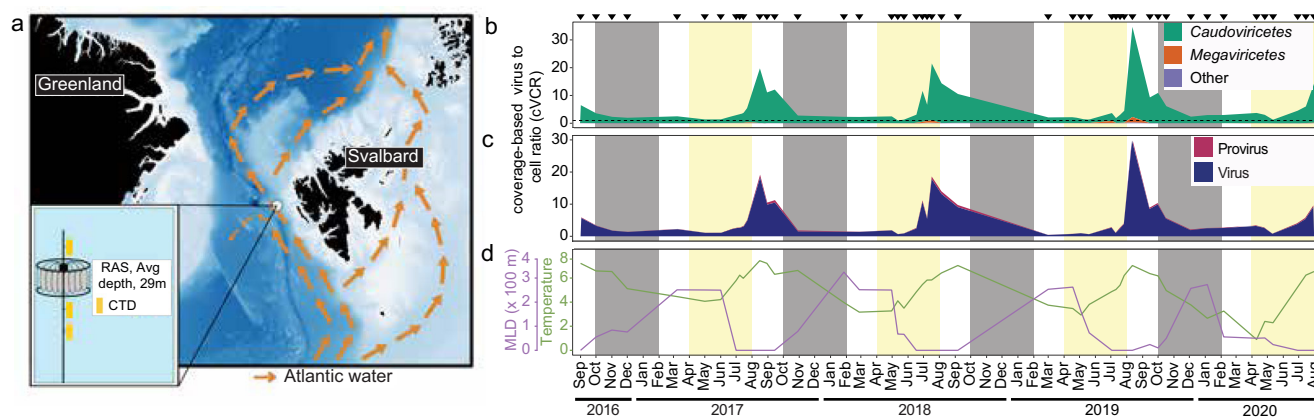
in the WSC (Fig. 1a, and Supplementary Data 1), from an average depth of 29 m. As samples originate from the cellular size-fraction ( $>0.2\ \mu\text{m}$ ), our study characterizes the diversity of actively infecting (i.e., intracellular) viruses, free-living viruses with a size of  $>0.2\ \mu\text{m}$ , viruses attached to particles, and/or integrated viruses.

The total number of reads per sample were  $196,489 \pm 18,358$ , with an average length of  $5435 \pm 405$  bp. Viral sequences were predicted based on both mapping to assembled contigs and raw reads (see Virus Prediction, Methods). The overall number of predicted viral sequences was concordant between the two approaches (Supplementary Data 1). On average, each sample harbored  $9815 \pm 965$  viral reads based on raw read counts ( $5.3 \pm 0.5\%$ ), and  $8825 \pm 1020$  reads mapping to vOTUs (viral Operational Taxonomic Units) ( $6.8\% \pm 1.8\%$ ). Notably, despite not pre-filtering to remove eukaryotes, the large majority of cellular reads in the dataset were prokaryotic ( $90.2\% \pm 1.4\%$ ), with eukaryotic reads making up a smaller proportion ( $7.2\% \pm 1.0\%$ ). Hereafter, we focus on the contig-based predictions because of their improved genomic context.

### Diversity of Fram Strait viruses over time

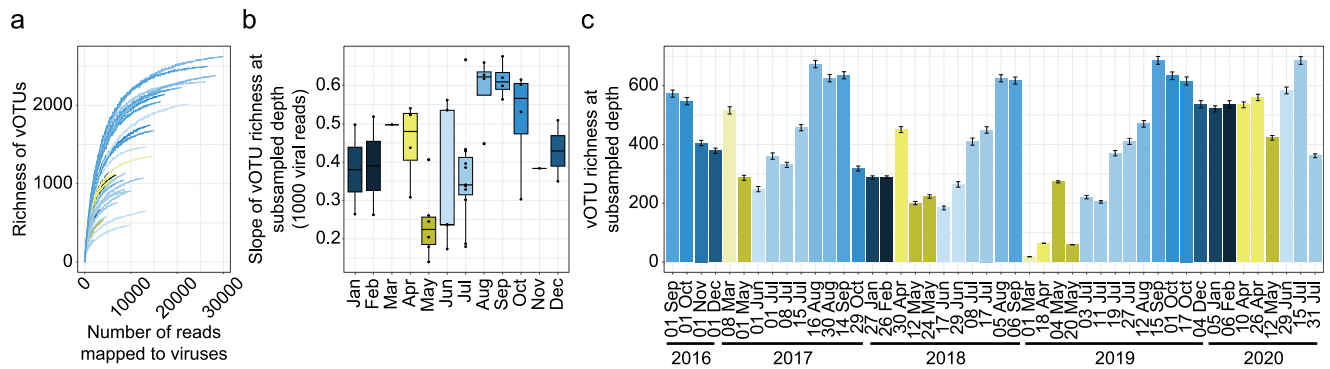
We first investigated how the community of Fram Strait viruses is structured over time. Through contig-based predictions, we identified 5662 vOTUs (95% nucleotide identity clusters, see Methods, Supplementary Data 2) which were derived from 7775 viral contigs that were greater than 10 Kb in length. These vOTUs predominantly represented the classes *Caudoviricetes* and *Megaviricetes*, with *Caudoviricetes* being the largest group, comprising 5531 vOTUs.

To assess the temporal dynamics of vOTUs, we normalized their abundance based on the estimated number of cells sequenced in each metagenome—a metric we term coverage-based Virus to Cell Ratio (cVCR). The cVCR approach not only accounts for differences in sequencing depth but also captures shifts in virus to host ratios across samples. Using this metric, we observed clear seasonal structuring in the cVCR of viral communities and taxa. Community cVCR values reached annual maxima of 20–35 and were four-fold higher during July–September (average cVCR 8.6) than during October–June (Fig. 1b). These patterns were consistent also when we sub-sampled the total reads to 100,000 reads per sample, as a test of the robustness of this metric to sequencing depth, as well as when we calculated coverage per Gigabase (Supplementary Fig. 1), thus we focus the remaining analyses on cVCR from the full dataset. This seasonal variation in cVCR was generally consistent across *Caudoviricetes*. Notably, the most abundant vOTU, a member of the *Caudoviricetes*, was also the most persistent, occurring in 45 of the 47 sampled time points. Similarly, the most persistent vOTUs overall, present in half of the samples,



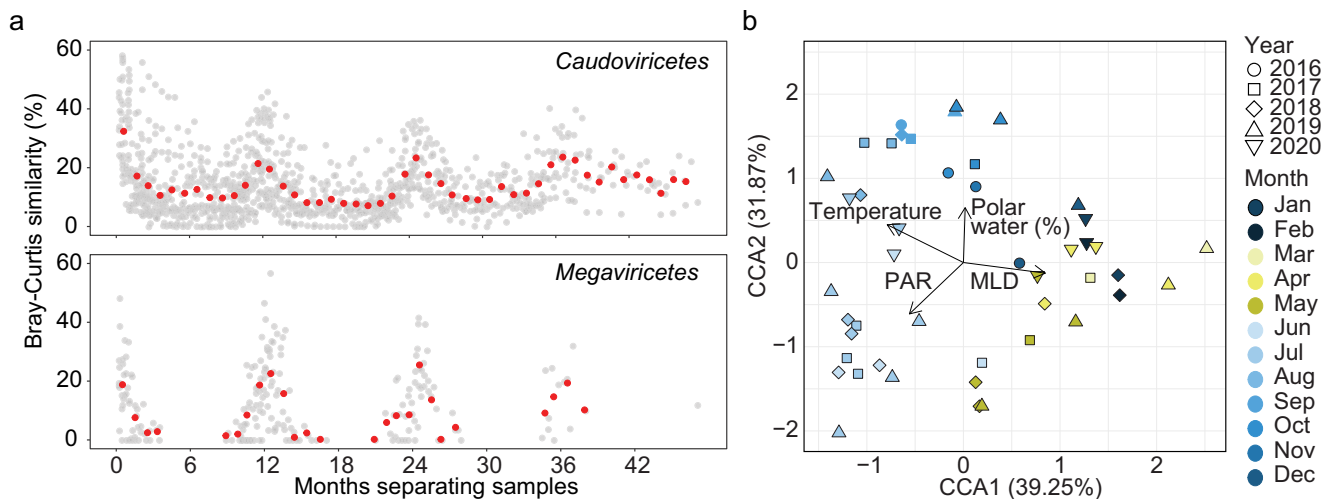
**Fig. 1 | Overview of study site, environmental conditions and viral distributions. a** The mooring site in the West Spitsbergen Current of Fram Strait. **b** cVCR of viral communities across the two major classes detected and **(c)** lifestyles during

the light and dark cycles from September 2016 to July 2020. **d** Dynamics of mixed layer depth and temperature which are major community structuring factors. Triangles (top) indicate metagenome sampling points.



**Fig. 2 | Intra- and inter-annual viral richness variability.** **a** To assess differences in sequencing depth, we iteratively subsampled viral read counts from 50 up to 30,000 at 50 count intervals and at each interval determined the mean richness from 100 iterations. The mean richness across subsampled intervals was visualized in a rarefaction-style curve. **b** Boxplots illustrating the slope of the vOTU richness rarefaction curves up until the chosen subsampling depth (1000 viral reads) where each box encompasses the 25th, median, and 75th percentile, and the whiskers capture the minimum and maximum values of richness estimated from samples

collected within the same sampling month. The slopes represent the trajectory of vOTU discovery, and thus indicate how richness may appear if a higher sequencing depth was achieved. The boxplots illustrate the 25%, median and 75% percentile of richness estimations for samples within each sampling month. **c** vOTU richness at subsampled depth across sampling months. Illustrated values represent the mean (bars) and standard error (error bars) of richness determined from 100 iterations of subsampling (1000 viral read counts per sample).



**Fig. 3 | Seasonality of viral communities and their association with environmental conditions.** **a** Bray-Curtis similarity of *Caudoviricetes* and *Megaviricetes* over time. Each point represents the time between two individual sampling points (x-axis) and their similarity (y-axis) for all pairs of samples. The red points indicate

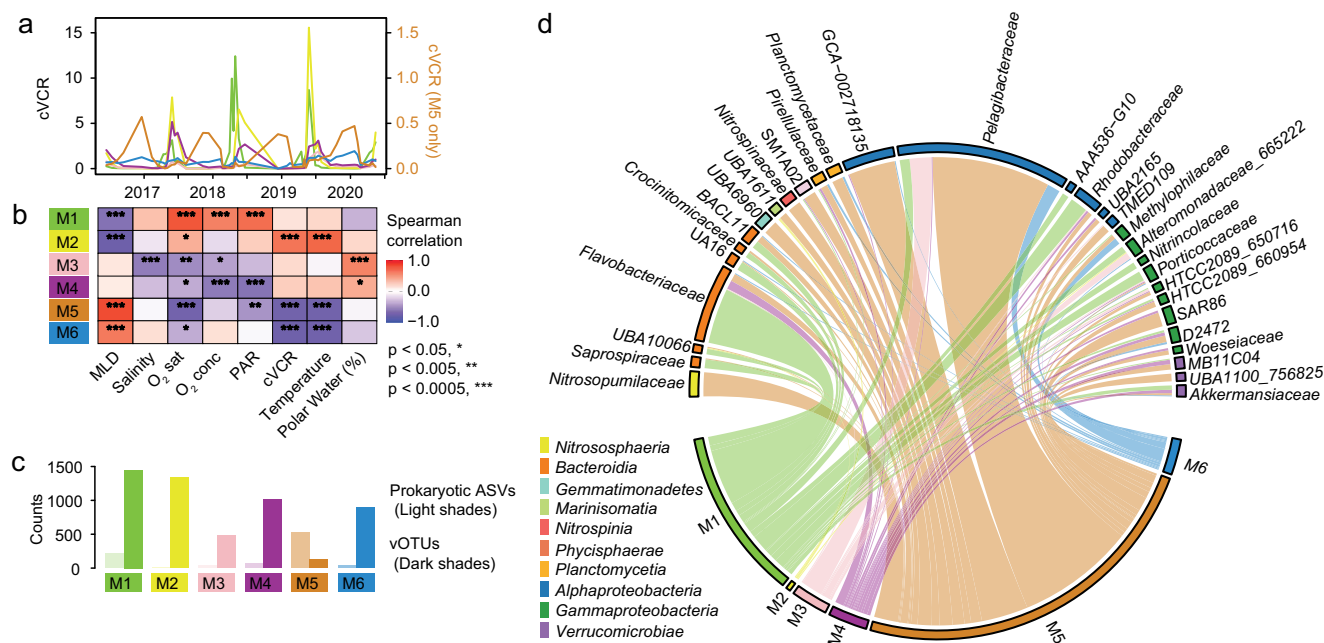
the average similarity for 30.5-day (-monthly) intervals. **b** CCA analysis of viral community composition colored by month, with vectors representing environmental conditions.

were predominantly members of the *Caudoviricetes*. In contrast, *Megaviricetes* were found less frequently, but were prevalent in late summer (July–September) (Fig. 1b, and Supplementary Fig. 1a), when eukaryotic phytoplankton, their presumed hosts, are most abundant<sup>49</sup>. Overall, lytic lifestyle was predicted to be predominant, with an average of  $94.0 \pm 0.4\%$  of viruses, a pattern that was invariable across seasons (Fig. 1c). The Southern Ocean featured higher rates of lysogeny during periods of low production<sup>26</sup>. As for the latter point, it is important to point out that some lysogens may not be recognized due to missing integrases or other factors, and, vOTUs may not represent the full prophage, which could then be misinterpreted as lytic rather than integrated. Additionally, lytic viruses may outnumber lysogens in terms of copies within the cellular metagenome, making them easier to detect and assemble, which could bias the recovery or obscure lysogens.

We next investigated how the overall diversity of viruses changes across seasons. As viral diversity was not saturated at any sequencing depth (Fig. 2a), we compared the diversity of vOTUs across months

after subsampling to a normalized depth. The diversity of viruses reached a maximum between late summer–autumn (Aug–Oct) (Fig. 2b, Supplementary Fig. 2), with lowest richness values typically in May. These patterns were consistent both by estimates from the slopes of rarefaction curves to the sub-sampled depth (Fig. 2b) as well as the extrapolated richness (Supplementary Fig. 2). Notably, prokaryotic community richness was also lowest in May, though the timing of highest prokaryotic richness was different than for viruses, with the highest prokaryotic richness observed in winter (vs. late summer–autumn for viruses)<sup>42</sup> (Supplementary Data 1). Inter-annual variability in richness was also observed, for example, with May 2019 showing elevated richness relative to other years (Fig. 2c). The dynamics in richness calls for further sequencing efforts, with increased depth of sequencing to further unravel the viral diversity that persists, especially, through late polar night.

In addition to the variations in overall relative abundance (cVCR) and diversity, we also observed strong seasonality in the composition of the major bacteriophage groups (Fig. 3a). The composition of



**Fig. 4 | Viral modules and their association with environmental conditions.**

**a** Sum of cVCR of all vOTUs per module over time; note separate y-axis for M5. **b** Two-sided Spearman's rank correlation coefficients between abundances (cVCR) of viral modules and environmental parameters, with significance levels of

correlations indicated by asterisks. **c** Count of prokaryotic and viral members of each major module. **d** Chord diagram showing the taxonomy of bacterial ASVs in each module. The taxa along the arc are colored based on their corresponding phylum.

vOTUs within both *Caudoviricetes* and *Megaviricetes* showed a sinusoidal-like pattern over time, with peaks in Bray-Curtis similarity at 12, 24, and 36 month intervals (Fig. 3a). Furthermore, in some cases, similarity between samples from opposing seasons was zero, which is likely a combination of strong seasonality, low relative abundance (and hence detection) of vOTUs in winter, as well as variable sequencing depth. Notably, the seasonality observed in virus community composition was consistent regardless of minimum coverage value utilized for detected viral presence, though similarity was less at higher thresholds (Supplementary Fig. 3).

To further contextualize the dynamics at the vOTU level, we examined the underlying sequence diversity of vOTUs in two ways. First, we explicitly compared all non-clustered viral contigs >10,000 bp across years. Out of 7360 such contigs, we found 21 clusters of contigs, encompassing 55 contigs total that were 100% matches between at least two years. This suggests that detection of viruses with no genetic changes across years is rare, though there is a limitation of sequencing depth. Second, based on analysis of single nucleotide variations in reads to vOTUs (via InStrain, see Methods), we find that there tends to be microdiversity underlying the vOTU populations, though the amount varies between vOTUs ranging from likely insignificant and potentially related to sequencing error, to high differentiation (e.g., 97% similarity) (Supplementary Fig. 4). These analyses highlight that, in general, vOTUs are made up of populations of highly similar viruses, rather than being clonal. This has also been observed elsewhere, including the Arctic<sup>14</sup>.

### Virus-host and environmental relationships

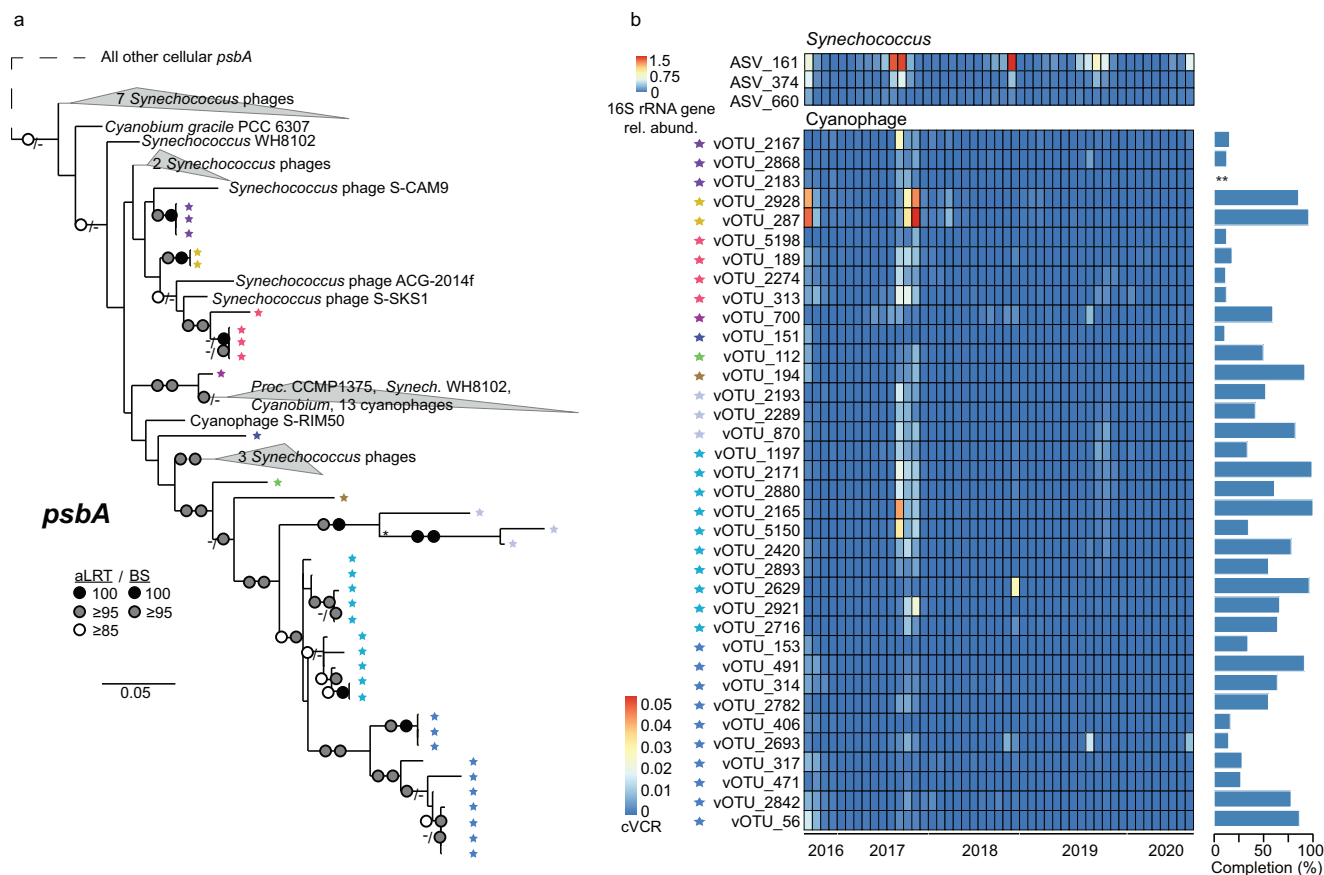
To explore the temporal structuring of viruses and their association with prokaryotic hosts, we employed community- and taxon-level analyses in the context of eight physicochemical parameters and prokaryotic community composition data, comprising 3748 prokaryotic amplicon sequence variants (ASV).

At the whole community level, Mantel tests demonstrated the strongest correlation to prokaryotic community composition (Mantel  $\rho = 0.632$ ,  $p = 0.001$ ,  $n = 46$ ) followed by oxygen (Mantel  $\rho = 0.371$ ,

$p = 0.001$ ,  $n = 46$ ) and mixed layer depth (MLD) (Mantel  $\rho = 0.259$ ,  $p = 0.001$ ,  $n = 46$ ) (Supplementary Data 3, Supplementary Fig. 5). To further elucidate physicochemical drivers, we used canonical correspondence analysis (CCA) to reveal a seasonal clustering of samples, with temperature (summer), MLD (winter), photosynthetically active radiation (PAR; late spring-early summer), and polar water fraction (late summer) accounting for 15.4% of the total variance (Fig. 3b). These results suggest that viral communities are primarily correlated with prokaryotic community composition which is in turn driven by environmental conditions plus biological interactions—leading to a complex network of interdependencies and physicochemical linkages, similar to dynamics in temperate environments<sup>7,50</sup>.

Given the coupling between viral and prokaryotic communities, we next explored potential virus-host associations over time at the individual vOTU and 16S rRNA gene ASV level. We constructed a Convergent Cross Mapping (CCM) network based on co-occurrences, similar to the approach described in Oldenburg et al.<sup>49</sup>, which includes information about the direction of associations (i.e., predicting causal relations) between vOTUs and ASVs. The CCM network comprised 5136 vOTUs and 850 prokaryotic ASVs (Supplementary Fig. 6a), with directional associations (correlations > 0.7) where vOTUs dynamics are 'following', or are 'caused' by, prokaryotic ASV dynamics (akin to Lotka-Volterra dynamics<sup>51</sup>) totaling 15,930. Louvain clustering resulted in six distinct modules, comprising vOTU-ASV associations occurring during the same temporal period. We named the six modules based on their seasonal abundance patterns (Fig. 4a); for example, M1 peaks in late spring, followed by M2. The temporal distinction between modules was confirmed by their correlations with specific physicochemical conditions (Fig. 4b). M1 showed the strongest positive correlation with PAR, while M2 had the strongest positive correlation with temperature (Fig. 4b). Interestingly, these two modules also have the most diverse virus communities, with M1 containing 1448 and M2 containing 1334 vOTUs, respectively (Fig. 4c). M3 and M4 peaked in late summer, with M3 having the strongest positive correlation with polar water fraction, and M4 having the strongest negative correlation with PAR. M5 and M6 both negatively correlated with temperature, cVCR, and positively with





**Fig. 5 | Cyanophages. a** *psbA* gene phylogeny of cyanophage vOTUs and reference genomes. Support values are reported as aLRT (approximate likelihood ratio test) / BS (maximum likelihood bootstrap). For visualization purposes, “\*\*” indicates branches from this node are scaled 33% of their actual length. **b** Dynamics of

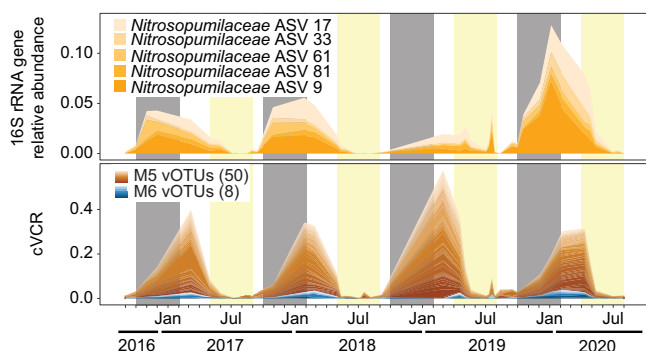
correlated *Synechococcus* ASVs (top) and cyanophage vOTUs (bottom), along with completion estimates based on CheckV<sup>108</sup>. \*\* indicates no completion prediction as the vOTU was classified as a provirus.

MLD. However, among these two, only M5 negatively correlated with PAR, being particularly abundant during winter (December–April), while M6 was persistent throughout the year. Generally, the number of vOTUs in a given module outnumbered the number of prokaryotic ASVs, except M5 where ASVs outnumbered vOTUs. Notably, M5 also contained the fewest vOTUs, 132 (Fig. 4c). The overall trend of more vOTUs than ASVs in modules corresponds to the overall larger number of vOTUs vs. prokaryotic ASVs examined.

The modules were consistently dominated by *Caudoviricetes* (Supplementary Fig. 6b), and the proportion of lytic to lysogenic viruses was rather invariable (Supplementary Fig. 6c). *Megaviricetes* were primarily present in the spring and early summer modules M1 and M2 (Supplementary Fig. 6b) where they constituted 6.3% and 1.4% of the vOTUs, respectively. Together, the major differences between modules, especially relating to bacteriophages, are hence at the vOTU level.

In terms of prokaryotic membership, the modules varied considerably. M1 was dominated by taxa typically associated with copiotrophic conditions or phytoplankton blooms, including *Flavobacteriaceae*, *Rhodobacteraceae*, and *Porticoccaceae* (Fig. 4d). In contrast, M5, which contained the largest number of ASVs, comprised diverse prokaryotic taxa, including *Pelagibacteraceae*, *Nitrosopumilaceae*, GCA-002718135 (aka HIMB59), *Nitrospinaceae*, SAR86, *Pir-ellulaceae*, and *Planctomycetaceae* (Fig. 4d). The smaller modules M2, M3, M4, and M6 comprised a variety of taxa, with the most prevalent taxon within each being UBA1611 (Marinimicrobia), and *Pelagibacteraceae*, respectively (Fig. 4d).

In order to evaluate host association patterns, we utilized both the CCM network correlations and host predictions via iPhoP<sup>52</sup>. In total, 42.5% of vOTUs received a host prediction by iPhoP. Among bacterial families with a high number of predicted vOTUs, *Flavobacteriaceae* and *Pelagibacteraceae* were the most frequently predicted. In M1, *Flavobacteriaceae* and *Akkermansiaceae* (Verrucomicrobia) dominated, while *Pelagibacteraceae* was less common. Conversely, in M5, *Pelagibacteraceae* was the most predominant, with fewer *Flavobacteriaceae* and *Akkermansiaceae*. Meanwhile, vOTUs with *Cyanobiaceae* as their predicted host were more common in M2–M4 (Supplementary Fig. 6d). Among the vOTUs correlated with an ASV in the network, the majority of host-virus pairwise correlations were to diverse prokaryotes present in M5, with *Pelagibacteraceae* (20.3%) and *Nitrosopumilaceae* (6.0%) being the most common (Supplementary Fig. 6e). M1 had the second most bacteria-to-virus connections with 37.5% of these being to *Flavobacteriaceae*. Among the families with most pairwise correlations (i.e., over 250), the overall pattern of host prediction of the viruses and module membership was similar, with vOTUs with predicted hosts of *Flavobacteriaceae* and *Pelagibacteraceae* being the dominant host predictions for M1 and M5, respectively (Supplementary Fig. 6f). However for individual vOTUs and ASVs, only 3.7% of the predictions overlapped at the family level between CCM network correlations and iPhoP (i.e., 122 out of 3,280 correlations with associated host predictions). Among these, again, *Flavobacteriaceae* and *Pelagibacteraceae* made up 46.7% and 6.6%. For *Flavobacteriaceae*, these vOTUs were diverse, coming from 14 different novel orders (via VConTACT3<sup>53</sup>, see Methods), while for *Pelagibacteraceae* they all came



**Fig. 6 | Dynamics of *Nitrosopumilaceae* and associated viruses.** The shown *Nitrosopumilaceae* are the most abundant ASVs, which are all members of M5. Different ASVs and vOTUs are shown by different shades of orange and blue, corresponding to M5 and M6 colors, respectively. The shown vOTUs are those that are correlated with *Nitrosopumilaceae* ( $p < 0.05$ ) in the Convergent Cross Mapping network (see Methods, Supplementary Fig. 5a, 6).

from a single novel order (Supplementary Data 4). However, overall, the low number of matching CCM correlations and host predictions at the vOTU-ASV level demonstrates the challenge to discern host-virus relationships in diverse and complex ecosystems where a variety of technical and biological challenges may complicate such analyses (see Discussion). Nonetheless, the results provide valuable indications for particular lineages at an overall module level.

### Diversity and dynamics of cyanophages and *Nitrosopumilaceae*-associated viruses

From the predicted interactions between vOTUs and ASVs, we further explored the diversity and dynamics of putative cyanophages—as cyanobacteria are of particular interest with respect to changing conditions in Arctic ecosystems<sup>5,54,55</sup>. We identified putative cyanophage vOTUs based on phylogenetic analysis of *psbA* genes, classification based on VPF-Class<sup>56</sup> and host prediction via iPhoP<sup>52</sup>. The phylogenetic reconstruction of *psbA* revealed that putative cyanophages are distinct from cultivated relatives from more temperate locations (Fig. 5a). The putative cyanophage vOTUs represented some of the largest viral contigs recovered, with half of the assembled cyanophages being medium- to high-quality, and a quarter > 80% complete (Fig. 5b, and Supplementary Data 2). The cyanophage vOTU abundance peaked during August–September 2017 (Fig. 5b), complementing previous observations of *Synechococcus* in the WSC<sup>42,54</sup>. Although the seasonal dynamics were consistent across the cyanophage vOTUs, their maximal cVCR varied from 0.003 to 0.06 (Fig. 5b). In addition, all of the cyanophage vOTUs exhibited lower abundances in 2018–2020 compared to 2017, indicating interannual variation of these viruses and their presumed hosts (Fig. 5b). Overall, the abundances of the cyanophage vOTUs with *Synechococcus* ASVs were correlated when considering no time-lag (Spearman's  $\rho = 0.48$ ,  $p = 0.0008$ ,  $df = 44$ ), but a stronger correlation (Spearman's  $\rho = 0.51$ ,  $p = 0.0008$ ,  $df = 44$ ) occurred with a vOTU time-lag of one time-point (average interval of 32 days). Thus, putative cyanophages increased in abundance ~1 month after *Synechococcus* peaks. For individual ASVs and vOTUs, the strongest correlations were without time delay ( $n = 53$ ), followed by one ( $n = 44$ ) and two ( $n = 32$ ) time-points, indicating some temporal variability between individual cyanophage vOTUs and their putative hosts compared to the groups at-large (Supplementary Data 5). Our results reveal that both cyanobacteria and their viruses are present in the Arctic, highlighting a need to further understand their interactions for the future Arctic Ocean, as they are expected to increase due to Atlantification.

Nevertheless, given the scarcity of data on viruses during the polar night, we also examined the microbial dynamics of viruses related to

prokaryotes dominant in winter. To do so, we focused on viruses associated with *Nitrosopumilaceae* due to their importance for wintertime nitrogen and carbon cycling<sup>57,58</sup>, their prevalence in the dataset with the top five *Nitrosopumilaceae* ASVs peaking from October–May (average 4.2%) vs. lower abundances from June–September (average 0.5%), and being among the taxa with the most correlations to viruses (Supplementary Fig. 7). We identified associations between 58 vOTUs and five *Nitrosopumilaceae* ASVs (Fig. 6), which were primarily associated with the winter module M5. In total, 53 of the 58 vOTUs were classified as *Caudoviricetes* (the other five were unclassified), and came from 14 different novel orders, based on vConTACT3 (Supplementary Data 6). Notably, among vOTUs with correlations to hosts in the correlation network, only one had a predicted host (based on iPhoP) of *Nitrosopumilaceae*. This vOTU was correlated with an unclassified Gammaproteobacterium (no family-level prediction) within the correlation network, highlighting again the difficulty comparing directly these two complex and independent approaches. The persistence of these vOTUs despite low virus cVCR demonstrates the ability to track pronounced seasonality even amongst low abundance viruses. However, more focused investigations on the temporal linkages of these vOTUs and their potential hosts are needed to better understand the environmental impacts of such associations.

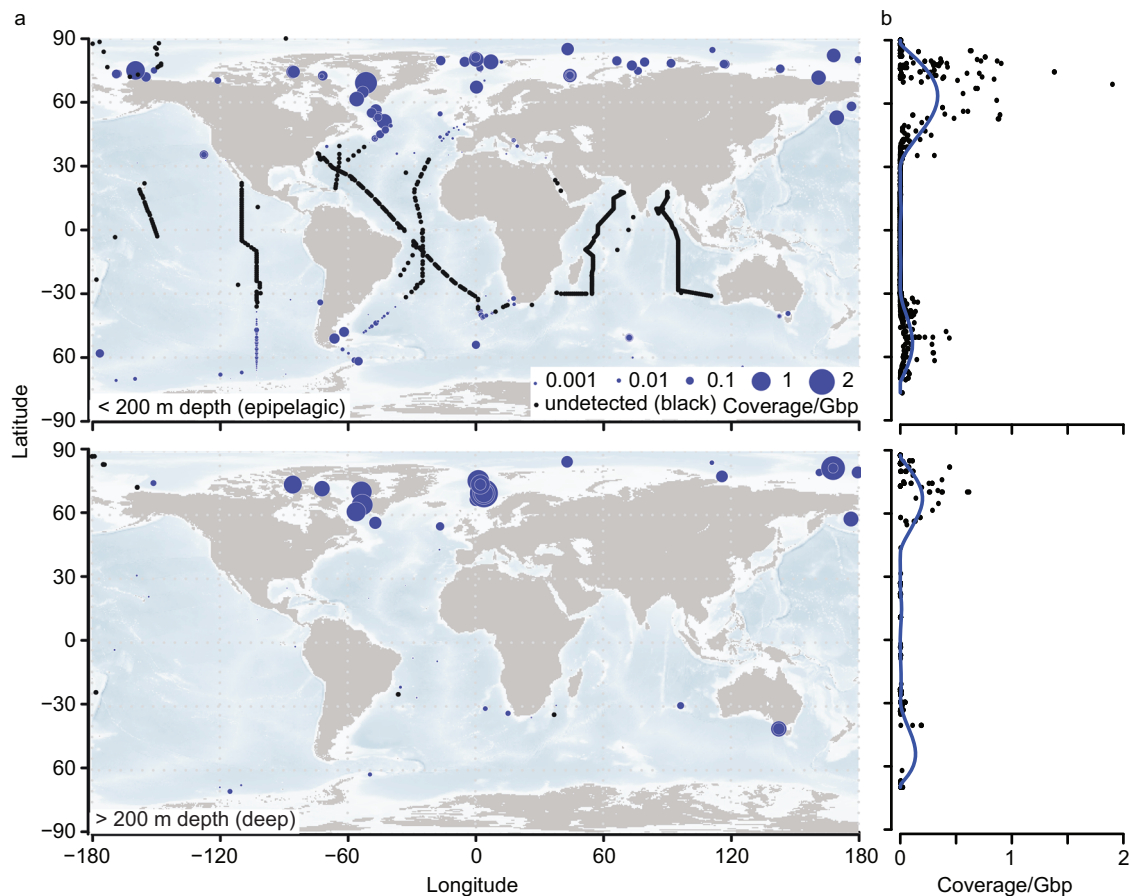
### Bipolarity of Fram Strait viruses

Considering the long-standing discussion on latitudinal microbial diversity gradients<sup>59–62</sup> and the endemism of Arctic and Antarctic microbiomes<sup>63,64</sup>, we assessed the distribution of Fram Strait viruses across the global oceans through metagenomic datasets from various large-scale sampling campaigns, such as Malaspina, Tara Oceans and Bio-GO-SHIP<sup>14,65–69</sup> (Supplementary Data 7).

Overall, the abundance of Fram Strait viruses peaked in the epipelagic (<200 m) around 60–70°N, with decreasing abundances towards 50°N, and typically no detection in subtropical and tropical epipelagic waters (Fig. 7a). A similar pattern occurred in southern hemisphere epipelagic waters, with peaks in abundance around 45–55°S (Fig. 7a). At individual vOTU level, 42% of vOTUs displayed bimodal peaks in abundance with separate maxima in each hemisphere (i.e., the two highest peaks in abundance occurred in separate hemispheres). The average peak latitude of vOTUs was 61°N and 51°S, respectively (Fig. 7b). Fram Strait viruses were also commonly detected in the deep ocean (here, operationally defined as sampling depth of > 200 m). In particular, they were detected in 87% (338 of 390) of deep global samples. Like the epipelagic waters, deep-water viruses were more prevalent in northern and southern higher latitudes (Fig. 7b), though in deep samples the viruses were more commonly detected. Overall, however, when detected, their relative abundances were lower in deep samples than in shallow samples. Notably, there was significant variation in community structure of the detected Fram Strait viruses between epipelagic and deep samples (PERMANOVA:  $F(1, 767) = 99.16$ ,  $p = 0.001$ ;  $R^2 = 0.1145$ , 95 % CI [0.102 – 0.131])

which aligns with established differences in microbial communities in the deep vs. epipelagic ocean, and past results of virus biogeography<sup>14</sup>. Future work could consider the differences in the total virus communities between epipelagic and deep, as the current analysis utilizes only mapping to the Fram Strait viruses.

Expanding on these observations, we investigated the distributional patterns of the six Fram Strait modules on the global scale. All modules were more prevalent in the northern than in the southern hemisphere, but their relative distributions varied, with M1 and M2 being the most abundant in northern epipelagic waters, while M6 being more prominent in higher southern latitudes (Supplementary Fig. 8a), indicating a relatively restricted distribution of M1 and M2. In deeper waters, all modules were less prevalent than in epipelagic waters, with higher prevalence in the northern hemisphere. In deep waters, M6 was the most prevalent (Supplementary Fig. 8b).



**Fig. 7 | Global distribution of Fram Strait vOTUs by comparison with short-read (0.2–3  $\mu$ m size-fraction) metagenomes (Supplementary Data 7). **a** Upper and lower map correspond to samples collected within the upper 200 m and deeper than 200 m, respectively. Abundance is calculated as coverage per gigabase pair of sequence (Coverage/Gbp). In the deep map, latitude and longitude were jittered to**

allow visualization of multiple depths at the same location. **b** Viral Coverage/Gbp plotted by latitude across all samples along with a Generalized Additive Model prediction, illustrating the trend of total viral Coverage/Gbp by latitude. For more information on samples, see Supplementary Data 1.

### Distinctive amino acid signatures of polar viruses

To expand on the global perspective, we assessed potential adaptations of viruses to polar waters by examining properties that have been attributed to cold adaptation in prokaryotes, in particular amino acid signatures that increase protein flexibility in cold environments. To do this, we examined the amino acid features of proteins from the Fram Strait viruses and from the GOV2.0 dataset, which spans both polar and non-polar sampling locations<sup>14</sup>. To ensure fair comparisons between the datasets, we focused only on GOV samples originating from  $\leq 35$  m water depth (i.e., the average depth of Fram Strait sampling) and first analysed the environmental context, which shows strong clustering of GOV2.0 polar samples to the Fram Strait, based on their similarity in distance to equator, oxygen, and temperature (Fig. 8a). Linking those environmental data to the examined protein features for all viruses, revealed that the aliphatic index and the nitrogen usage score were positively correlated with temperature (Fig. 8b), whereas other traits of potential cold adaptation, in particular polar charged and uncharged amino acids, showed significant correlation with one or more of the other environmental parameters, but not directly with temperature. This was also the case for *Caudoviricetes*, while *Megaviricetes* amino acid traits were more generally positively correlated with oxygen, temperature, and chlorophyll, and other vOTUs (non-*Caudoviricetes*, non-*Megaviricetes*) were generally positively correlated with salinity and oxygen (Supplementary Fig. 9). In the latter cases, these groups are more poorly sampled due to their low abundance resulting in weaker correlations overall, necessitating

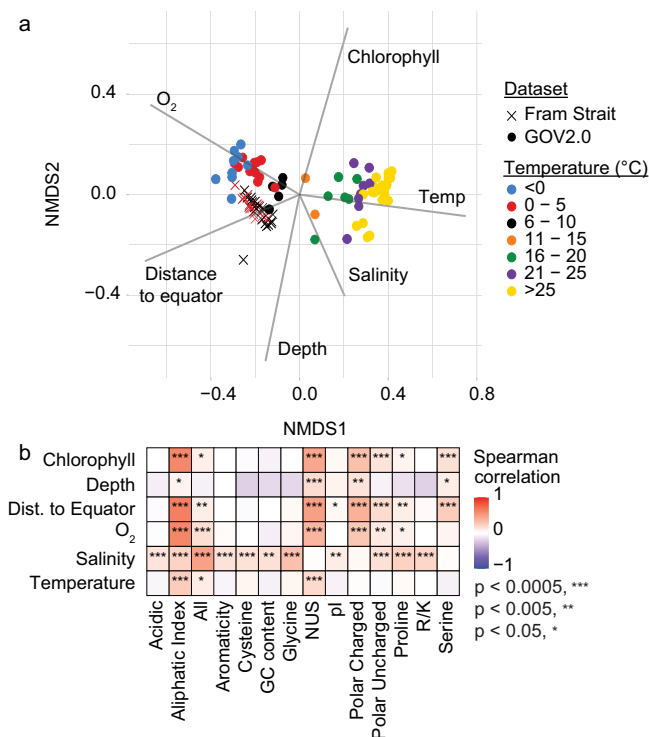
further study. Examining the average values for these protein parameters according to sample and temperature, across all viruses, reflected the patterns observed via other statistical analyses (Supplementary Fig. 10), reinforcing our observations.

Beyond the amino acid level, we found that 17.0% of viral protein annotations were significantly enriched in high latitudes (253 of 1490), while another 7.1% were enriched in lower latitudes (106 of 1490) (Supplementary Data 8). Notably, among a variety of annotations enriched at high latitudes, most significant were chaperone proteins and stress response such as Cold (CSD) and Heat Shock Proteins (HSP70), DnaJ (Supplementary Fig. 11), as well as genes common in cyanophages in the lower latitudes including a photosystem gene (Photo\_RC) and Transaldolase/Fructose-6-phosphate aldolase (TAL\_FC)<sup>70</sup>. Together, our evidence suggests that biochemical and biological traits distinguish polar viruses from their tropical and subtropical counterparts, linked to amino acid sequences.

### Discussion

Harnessing long-read metagenomic data in high temporal resolution, we demonstrate a pronounced, annually repeating seasonality of viruses across multiple years in the Arctic Ocean. The seasonal viral dynamics correlated strongly with both the predominant prokaryotic host communities, as well as environmental parameters. We found that Fram Strait viruses and seasonal modules occur across cold waters of both hemispheres, while being mostly absent from subtropical and tropical waters.





**Fig. 8 | Viral amino acid traits across environmental gradients. a** Nonmetric multidimensional scaling (NMDS) plot showing the environmental diversity of the sampled Fram Strait and GOV2.0 ecosystems. The Bray-Curtis distance similarity matrix was calculated based on the available environmental parameters of 38 Fram Strait and 69 GOV2.0 samples and used to generate NMDS coordinates of each sample. Point shapes represent dataset origin, and colors distinguish different temperature ranges. Vectors show correlations with environmental variables. This ordination provides context for the correlations to environmental parameters of amino acid traits shown in part b. **b** Heatmap plot illustrating the Spearman correlation coefficients of environmental parameters to amino acid traits, with *P* values represented by asterisks as indicated. The Spearman coefficients were calculated using a two-sided Mantel test using pairwise distances of each environmental parameter (Euclidean distance) vs. each amino acid trait (Bray-Curtis distance).

In terms of total virus prevalence, we found strong increases in the ratio of viruses to prokaryotes in the late summer months, mirroring findings from the Antarctic where viruses peaked in mid to late summer<sup>23</sup>. This increase was consistent regardless of the normalization used, namely either coverage-based virus to cell ratio (our main metric), or coverage per gigabase pair (which would also account for eukaryotic DNA). Our observation of strong virus seasonality in the cellular size-fraction contrasts with some of those of free viruses via fluorescence microscopy. This seeming contradiction may indicate a difference in the dynamics of infection and host-association, or loss factors of free-living viruses<sup>71,72</sup>, for example photo-degradation of free viral particles<sup>73,74</sup>. Our results suggest that the number of viruses in the cellular size-fraction is in the same range (order of magnitude) as that of free viruses in seawater, and complement similar metagenomic marker-gene derived virus-to-prokaryote ratios which have focused so far on size-fractions that include free viruses<sup>75</sup>. This observation is consistent with estimated infection rates<sup>10,76</sup>, especially considering that each infected cell can harbor a large number of viral genomes<sup>77–80</sup>. A further consideration is that our study focused on dsDNA viruses, thus not investigating the RNA virome<sup>81–85</sup>. Investigating RNA viruses, their seasonality and impacts on the prokaryotic community is an important future direction.

At the virus community level, the long-read and cellular size-fraction metagenomic approach likely aided assembly of the dominant viruses<sup>86–88</sup>, which often suffer from assembly problems when short-

reads are utilized. However, we observed low similarity in opposing seasons compared to studies from more temperate areas<sup>28–30</sup>. The low similarity, reflective of low viral persistence in the ecosystem, may be due to the focus on host-associated viruses, thus excluding free-viruses with no prevalent host that could increase persistence (viral seed-bank hypothesis)<sup>89,90</sup>. Furthermore, the low persistence may also be in part related to the low frequency of predicted lysogenic viruses that we detected, which would otherwise be a potential persistence mechanism for viruses with rare hosts<sup>91</sup>. The very low similarity could also correspond to the relatively low coverage overall due to the long-read technology as well as dilution of reads from prokaryotes. Nevertheless, the strong seasonality reflects the strongly seasonal host communities<sup>33,92</sup>.

The use of CCM in studying virus-host interactions presents challenges due to their rapid temporal dynamics. Viral latent periods range from hours to days, complicating interpretations based on monthly sampling intervals. The low overlap (3.7%) between CCM associations and iPhoP predictions may reflect limitations of our sampling resolution. Integrating multiple methodologies is critical for understanding the complexities of microbial interactions. Combining CCM findings with naive network analysis and iPhoP results provides a more comprehensive view of microbial community dynamics, with visual comparisons highlighting the biological significance of identified interactions. It should be considered that factors such as diversity and abundance can impact the total number of viruses detected in modules, and co-occurrence patterns overall. M5, for example, which has the lowest vOTU to prokaryotic ASV ratio, corresponds to the period during which vOTUs were the least relatively abundant. Additionally, during this period, vOTU richness was relatively high. The combination of these two factors challenges vOTU detection, and could result in vOTUs being excluded during pre-processing or missed altogether.

We observe a pronounced bi-modality in the latitudinal distribution of Fram Strait viruses, with peaks at high latitudes in both the southern and northern hemispheres. In particular, the peaks occur around the northern and southern polar fronts (70°N / 60°S)<sup>93–96</sup>, suggesting that these boundaries of oceanic realms represent a hot-spot of polar-adapted viruses, possibly due to the pronounced biological and physicochemical gradients that are present, together with overall greater productivity. It might be, however, somewhat biased by a larger number of samples collected along these fronts. The high-latitude preference is also reflected within viral amino acid sequences, with signatures of cold adaptation that may contribute to their success in colder waters, thus expanding previous observations from the Southern Ocean<sup>18</sup> and polar eukaryotic viruses<sup>97</sup>. Further investigations to unravel the diversity and dynamics of viruses in Arctic and Antarctic are crucial, given that climate shifts are causing major biological and physicochemical perturbations in these regions. Additionally, in general, the detection of Fram Strait viruses in the deep ocean is notable, and potentially related to export from the surface and deep ocean currents, some of which originate at the Arctic. Further inspection of the specific vOTUs and other datasets at the global level could help further evaluate these ideas. Part of the explanation may be that vOTUs primarily observed in deep waters are from M6, the persistently present Fram Strait module; thus, they may be associated also with generally persistent/ubiquitous host lineages and/or have broad host ranges. Furthermore, the relative lack of seasonality in the deep sea may help with their consistent detection.

The peaks of Arctic viruses during summer and their specialization to high-latitude regions call for further process and targeted studies to elucidate the host communities that they interact with and influence. Previously, viruses of phytoplankton (cyanobacteria and eukaryotic phytoplankton) have been experimentally observed to exert less top-down pressure than eukaryotic grazers at high latitudes<sup>98,99</sup>, but how this relates to the total microbial community



remains unknown. In any case, our study demonstrates that the impacts of Arctic viruses are likely very different depending on season and ecosystem state. Our results indicated a low percentage of predicted prophages without a seasonal enrichment, even in the cellular size-fraction studied. This suggests that the Piggyback-the-winner scenario is neither very prevalent, nor seasonally variable, though again this may be currently limited by computational predictions, and other biases like lytic viruses occurring within cells in more than one copy, and thus we may underestimate prophage prevalence. We found that most viruses are part of seasonally variable modules, and that often there is no similarity between opposing seasons (e.g., six months apart), at least at the depth sequenced at in the investigated cellular size-fraction. Hence rather than having a Constant-Diversity<sup>46,47</sup>, Arctic virus communities are undergoing substantial seasonal change. Likewise, the prokaryotic community is also highly seasonal, thus the viruses are generally following and strongly correlated with host abundances. Thus, unlike how viral ecological models are sometimes idealized in a chemostat-like setting<sup>100</sup>, viral communities' strong seasonality impacts how such models can be of use in the Arctic Ocean (such as in the Red-Queen hypothesis<sup>29,48</sup>). Arctic virus communities, with their strong seasonality where many taxa are not present year-round (or at least often not detected), may be impacted by how these dynamics play out elsewhere; likely, such dynamics occur primarily within seasons. Our analysis of microdiversity of Arctic virus communities demonstrated both instances of identical viruses across years, as well as a general tendency for vOTU to be overlaid by sequences with very high similarity but not exact matches to the vOTUs, akin to that seen elsewhere in the ocean<sup>29,30,87</sup>, including the Arctic Ocean<sup>14</sup>, though also some exact matches were observed across years. Thus, based on these results we cannot rule out either the Red-Queen hypothesis nor Constant-diversity. In each case, these dynamics could be further evaluated via more spatially-resolved temporal sampling.

In conclusion, our study advances the basis for understanding how viruses regulate and impact the dynamic and changing polar ecosystem, setting the stage for more detailed population dynamics studies, as well as process- and host-specific studies. Extended time-series observations will allow for improved understanding of the Arctic ecosystem impacts, in this ecosystem undergoing rapid change.

## Methods

### Seawater collection and eDNA sequencing

Research and sampling complied with relevant ethical and international regulations. Moored Remote Access Samplers (RAS; McLane) autonomously collected and fixed seawater from an average depth of 29 m in the eastern and western Fram Strait (Fig. 1) at weekly to fortnightly intervals between 2016–2020<sup>42,58</sup>. Sampling occurred in the framework of the FRAM / HAUSGARTEN Observatory. The resulting eDNA was used to sequence 16S rRNA gene fragments using primers 515F–926R<sup>101</sup>, processed into ASVs using DADA2 as described under ([https://github.com/matthiaswietz/FRAM\\_eDNA](https://github.com/matthiaswietz/FRAM_eDNA)). Subsequently, we only considered ASVs with  $\geq 3$  reads in  $\geq 3$  samples, corresponding to a total of 3748 prokaryotic ASVs. To complement iPHoP host assignment predictions (below), we assigned taxonomy of prokaryotic ASVs via Greengenes<sup>2102</sup> using the classify-consensus-vsearch method<sup>103</sup> in the q2-feature-classifier<sup>104</sup> of QIIME 2<sup>105</sup>.

DNA extracts from selected timepoints were additionally used to generate PacBio HiFi metagenomes at the Max Planck Genome Center, Cologne, Germany. Further details about the molecular analyses are described in previous reports<sup>42,58</sup>. In total, we herein analyze 94 amplicon samples and 47 metagenomes from the WSC, and 9 metagenomes from the EGC (Supplementary Data 1).

### Environmental parameters

Attached to the RAS were Seabird SBE37-ODO CTD sensors that measured temperature, depth, salinity, and oxygen concentration.

Sensor measurements were averaged over 4 h around each seawater sampling event. Physical sensors were manufacturer-calibrated and processed in accordance with <https://epic.awi.de/id/eprint/43137>. Employing multiple CTD sensors along the mooring depths enabled the determination of the minimum MLD at each sampling time point. Chlorophyll concentrations were measured via Wetlab Ecotriplet sensors. Surface water PAR data, with a 4 km grid resolution, was obtained from AQUA-MODIS (Level-3 mapped; SeaWiFS, NASA) and extracted in QGIS v3.14.16 (<http://www.qgis.org>). Polar water fraction, which is the proportion of polar water in the mixing of Atlantic and polar water masses, was calculated based on salinity and temperature<sup>33</sup>. Figure 1a map was created in QGIS v3.14.16 (<http://www.qgis.org>) using publicly available bathymetry data obtained from GEBCO. RAS illustration in Fig. 1a was generated via Inkscape (<https://inkscape.org>).

### Virus assembly, prediction, classification, and host prediction

Long-reads from each sample were assembled individually using hifiasm-meta v. 0.13-r308<sup>106</sup> with default settings. Viral sequences were predicted from both assembled contigs and the long-reads themselves using a combination of tools, based on a VirSorter2-based Standard Operating Procedure (SOP, <https://www.protocols.io/view/viral-sequence-identification-sop-with-virsorter2-5qpvoqebg4o/v3>). First, in VirSorter2 v.2.2.3<sup>107</sup> we included all possible viral groups: dsDNA-phages, RNA viruses, ssDNA viruses, nucleocytoplasmic large DNA viruses (NCLDV), and Lavidaviridae. CheckV v.1.0.1<sup>108</sup> was then used to quality check the long-reads and contigs identified as viruses by VirSorter2<sup>107</sup> as well as to trim potential host regions from identified proviruses. Contigs identified as viruses by VirSorter2<sup>107</sup> ( $\sim$ min-score 0.5) with at least one viral gene (predicted by CheckV<sup>108</sup>) were further screened using DeepMicroClass v.1.0.3<sup>109</sup> (not included in the original SOP, but intended to further reduce false positives). Contigs that were classified as either eukaryotic or prokaryotic by DeepMicroClass<sup>109</sup> were discarded. Contigs that were identified as RNA viruses by VirSorter2<sup>107</sup> were not analyzed. We clustered the remaining viral contigs into vOTUs, as defined previously<sup>110</sup>, using CD-HIT v.4.8.1<sup>111,112</sup> with the following parameters: cd-hit-est -M 100000 -c 0.95 -d 100 -g 1 -s 0.85.

For classification, we used the *genomad annotate* function (default settings) of *genomad* v.1.8.1<sup>113</sup> which uses taxon-specific marker proteins to assign taxonomy. To further resolve taxonomy to the subfamily level (Supplementary Data 2), we used vConTACT3 v.3.0.0b65 (<https://bitbucket.org/MAVERICLab/vcontact3/src/master/>). To predict hosts, we employed iPHoP v.1.3.3<sup>52,114–117</sup> with default settings, along with a custom database by adding previously assembled bacterial and archaeal MAGs from the Fram Strait (PRJEB67368) and removed MAGs from one oceanic study that was not manually curated, due to their high degree of potential virus contamination (PRJNA385857), which could result in false positives. When multiple host predictions were available, we selected the prediction with the highest confidence score.

### Mapping and normalization of viral abundance

Metagenomic long-reads from all samples were mapped to vOTUs ( $> 10$  Kb, from assembly- and read-based approaches) using minimap2 v.2.28-r1209 (parameter: -x asm5)<sup>118</sup>, and only the primary alignments were considered in subsequent analyses.

To quantify the abundances of vOTUs across metagenome samples, we employed a two-step normalization process to account for differences in read lengths and sequencing depth. First, we determined the mean coverage across the viral sequence length from mapped read counts (with a threshold that  $> 25\%$  of the viral sequence must be covered). Second, we divided the mean coverage by the estimated number of cellular genomes in each metagenome, as predicted based on the mean coverage across 16 universal single-copy ribosomal proteins<sup>119,120</sup>. We call this metric coverage-based virus to cell ratio (cVCR). This approach is similar to that used to derive virus to

microbial ratios on virus to prokaryotic marker genes elsewhere<sup>75</sup>, but considers the full viral coverage. In addition to this, we also calculated the coverage of each virus per gigabase pair (cVGB) of metagenome for a given sample and demonstrated that the two metrics yield comparable values (Supplementary Fig. 1). To gain an insight into the composition of the metagenome samples in terms of prokaryotic and eukaryotic reads, we employed Tiara v.1.0.3, a machine-learning tool, to assign domain-level classifications to the raw reads<sup>121</sup>. This revealed that the large majority of reads across the metagenomes, about 90%, were of prokaryotic origin.

### Estimation of viral diversity

Given that the vOTUs were derived from metagenomes with different sequencing depths, we employed an iterative subsampling approach to quantify and compare the alpha diversity of viruses across samples. For this, we applied 100 iterations of subsampling the vOTU count table at a range of different depths, spanning from 50 up to 30,000 counts at 50 count intervals. For this step, all vOTUs that exceeded the 25% breadth of coverage cutoff when using all the data for a given sample were utilized. During each iteration, richness, evenness and Shannon diversity were calculated, with mean values being determined for each 50-count interval. The functions *rrarefy*, *specnumber* and *diversity* from the *vegan* package<sup>122</sup> were used to perform subsampling and alpha diversity calculations. The mean values were used to generate a rarefaction curve of alpha diversity using the *ggplot2* package<sup>123</sup>. To assess shifts in diversity over time, we compared the mean richness values obtained from subsampling at the 1000 count interval. In addition, we computed and compared the slopes of the sample rarefaction curves up until the subsampled depths, which represents the rate of vOTU discovery with increasing viral read counts. Comparing the slopes across samples enabled an assessment of whether the sample richness patterns observed at the subsampled depth would change if the viral read count was increased. To further explore this, we also employed the iNEXT v.3.0.1 package<sup>124,125</sup> to estimate the vOTU richness based on extrapolations to 30,000 viral read counts.

### Co-occurrence network

To identify temporal co-occurrence patterns, cVCR values per vOTU and relative abundance of prokaryotic ASVs were converted into temporal profiles by Fourier transformation. Temporal profiles were constructed based on 16 Fourier coefficients, which capture the majority of observed vOTU and ASV peaks within the four-year period. Pairwise correlations between individual temporal profiles were then computed between all vOTUs and ASVs. Higher Pearson correlation values indicated similar temporal profiles. For network construction, we first calculated Pearson correlations for all pairs resulting in an undirected graph, from which we only considered correlations > 0.7 after multiple testing corrections using the Benjamini-Hochberg procedure. This threshold was determined based on previous findings on similar data<sup>49,58</sup> as well as preliminary analyses conducted on our dataset, which indicated that interactions exhibiting a correlation below this value often lacked biological relevance. To delineate strongly connected components representing co-occurring taxa, the Louvain community detection algorithm was applied to the entire graph<sup>126</sup>. Next, the putative associations were further evaluated based on Convergent Cross Mapping (CCM) in order to discern causal relationships between taxa in time-series data, as outlined by ref. 127. CCM enables the prediction of a species' time-series based on the knowledge of another species' time-series. Given the nature of lytic virus-host systems, where interactions can be short-lived, we acknowledge that this method must be carefully interpreted within the context of our sampling strategy (i.e., biweekly to monthly on average). Initially, we constructed a CCM network encompassing all pairwise combinations. Subsequently, we extracted the in- and outgoing edges between

nodes that were also connected in the co-occurrence network, utilizing resources from (<https://gitlab.com/qtb-hhu/marine/publications/framphages2024>). To quantify the strength of relationships, we employed Normalized Mutual Information (NMI) to account for non-linear relations<sup>49</sup>. A permutation approach was employed to compute significance values for edge weights, with the objective of determining whether the NMI values exceeded those expected for random edges. Further details on the construction and validation of the CCM network can be found in ref. 49. The CCM network was visualized in Cytoscape v.3.10.1<sup>128</sup>. Additionally, to identify potential time-lagged correlations, we employed extended local similarity analysis<sup>129</sup> with a maximum delay of two sampling points on the original non-Fourier transformed data.

### Statistical analyses

Bray-Curtis similarity, Mantel tests, CCA, and Spearman correlation were performed in R v. 4.3.1<sup>130</sup> using the *vegan*<sup>122</sup> and *Hmisc*<sup>131</sup> packages. We included only vOTUs with a breadth of coverage > 0.25.

### Cyanophage phylogenetics and comparative genomics

Putative cyanophages were predicted via VPF-class v.0.1.2<sup>56</sup> and iPhoP<sup>52</sup> host predictions. Subsequently, we focused on the most confident predictions by selecting cyanophages harboring the *psbA* core photosystem, identified by hmmscan<sup>132</sup> of viral predicted proteins (prodigal<sup>133</sup>) PFAM for Photo\_RC (PF00124). Viral and prokaryotic reference *psbA* were extracted from the vConTACT2 viral reference database (ViralRefSeq-prokaryotes-v211<sup>53</sup>) and GTDB prokaryotic reference database (version 214<sup>134</sup>) by hmmscan<sup>132</sup> with PFAM Photo\_RC. Eukaryotic reference *psbA* genes were extracted via Uniprot (gene\_exact:psbA) AND (taxonomy\_id:2759, reviewed Swiss-Prot, *n* = 157)<sup>135</sup>. A preliminary tree was constructed from these sequences to differentiate between D1 (*psbA*) and D2 (*psbD*) photosystem II (PSII) reaction center proteins, by alignment using MAFFT<sup>136</sup> followed by trimming with trimAl (-gt 0.2)<sup>137</sup>, and FastTree<sup>138</sup> for phylogenetic reconstruction. *psbD* were identified manually and excluded. The remaining sequences were considered, and phylogenetic reconstruction performed a second time with the same alignment and trimming steps (but with -gt 0.8); at this stage, tree building was performed with IQ-TREE (settings: -B 1000, -alrt 1000, -m MFP)<sup>139</sup>.

### Mapping to global metagenomes

Metagenomes from ten oceanic datasets<sup>14,65–69,140–145</sup>, including Tara Oceans, Malaspina, and Bio-GO-SHIP (Supplementary Data 7), were downloaded from NCBI, and quality-filtered with Trimmomatic (parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:50:30 MINLEN:50)<sup>146</sup>. All data utilized is previously published. Only DNA samples from discrete depths (e.g., not composites of the mixed layers or similar) and the prokaryotic size-fraction (0.2–3 µm) were used to allow comparison with our dataset. Multiple other size fractions are available, but the 0.2–3 µm was utilized because of our focus on prokaryotic dsDNA viruses, its comparability to the present Fram Strait virus dataset, and because it is among the most common range sampled (i.e., many samples available). Reads were mapped to the clustered contigs using bbmap.sh from the BBTools package (version 39.01, parameters: minid=95 idfilt=95) (<https://sourceforge.net/projects/bbmap/>). Trimmed mean coverage was calculated via CoverM<sup>147</sup> on contigs with minimum covered fraction of > 25% for each sample. Trimmed mean coverage values were then divided by total Gbp per quality filtered paired reads. Maps were generated via publicly available NOAA bathymetry data via getNOAA.bathy and the *maps* package in R. For determining abundance trends in the total dataset and at the module level, we computed a Generalized Additive Module (GAM) between cVCR and latitude using the *gam* function from the *mgcv* package<sup>148</sup>, with ten degrees of freedom (*k* = 10). For plotting the predicted GAM, negative values were converted to 0. For

calculating the maximum latitude that viruses occurred in each hemisphere, we used the GAM model, computing where peaks occur and then determining which two peaks had the highest abundance. For evaluating statistical differences in community similarity between epipelagic (< 200 m) and deep (> 200 m) samples, we first removed samples that had less than 0.0001 cVGB ( $n = 769$  samples remaining, including 435 and 334, shallow and deep samples, respectively) and vOTUs that had less than 0.0001 summed cVGB across all samples ( $n = 4,693$  vOTUs remaining). Then, we ran *adonis2*<sup>122,149</sup> with 999 permutations based on the epipelagic and deep grouping, with 95% confidence intervals for the  $R^2$  statistic estimated using 1000 bootstrapped replicates.

### Microdiversity of viral contigs

To examine microdiversity among vOTUs, we used two separate approaches. First, to examine the extent that exact, long virus sequences re-occur across years, we used CD-HIT v.4.8.1<sup>111</sup> to cluster viral contigs > 10 Kb at a 100% sequence identity threshold (cd-hitest -M 100000 -c 0.95 -d 100 -g 1 -aS 0.85). We then identified clusters that appeared across more than one year throughout the four-year sampling period. Second, we used inStrain v.1.9.0<sup>150</sup>, which compares reads mapped to vOTUs to characterize virus populations based on nucleotide diversity, total counts of single nucleotide substitution (SNS) and single nucleotide variation (SNV), and other related metrics. We applied a 98% minimum ANI, the most stringent threshold recommended by the developers, and used a minimum coverage of five to call a variant. To confirm a SNV, we required a minimum SNV frequency of 0.05 and a false discovery rate of  $1e-06$ . Additionally, we filtered for contigs where more than 25% of their bases were covered by at least one read, i.e., breadth > 0.25. Only vOTUs that exceeded the coverage minima more than three times throughout the sampling period were included in the analysis.

### Calculation of amino acid traits

Proteins were predicted from GOV2.0 and Fram Strait viruses via Prodigal v.2.6.3 (parameter: -p meta)<sup>133</sup>. Then, amino acid composition and various biochemical properties of the predicted protein sequences were assessed using custom R scripts available at [https://github.com/alyzzabc/fram\\_strait\\_viruses\\_2016-2020](https://github.com/alyzzabc/fram_strait_viruses_2016-2020). Briefly, protein sequences containing the ambiguous amino acid “X” were removed, followed by counting the occurrences of each proteinogenic amino acid. The frequencies of single amino acids (G, S, P, C) and of amino acid groups (acidic, polar uncharged, polar charged, aromatic) were calculated relative to the protein length. For the amino acids arginine (R) and lysine (K), the ratio R/K was calculated instead of occurrence. The molecular weight was calculated according to<sup>151</sup> and for calculation of the aliphatic index we followed<sup>152</sup>. The pI function from the *Peptides* package<sup>153</sup> was used to calculate the pI of the predicted proteins using the EMBOSS scale. The nitrogen usage score (NUS) was calculated using the following formula:

$$NUS = \frac{4F(R) + 3F(H) + 2F(KNQW) + F(DESTGPCAVILMFY)}{n(\text{protein})} \quad (1)$$

where  $F(X)$  corresponds to the absolute count of each amino acid and  $n(\text{protein})$  to length of the protein.

To analyze the variation of the viral protein features in relation to the environmental conditions, we omitted samples collected from a water depth > 35 m followed by calculation of the Bray-Curtis distance similarity matrix and generation of NMDS coordinates using the *metaMDS* function from the *vegan* package<sup>122</sup>.

To calculate pairwise correlations, the *vegan* package<sup>122</sup> was used: the distances for each protein feature (Bray-Curtis) and environmental condition (Euclidean) were calculated using the *vegdist* and *dist*

function, respectively. Pairwise Mantel tests of each environmental parameter vs. each amino acid trait were done using the Spearman method with 9999 permutations via the *mantel* function. The NMDS plot and the correlation heatmap were prepared using the *ggplot2* package<sup>123</sup>.

### Calculation of protein family enrichments by latitude

For determination of proteins enriched in polar regions, we used the prodigal-predicted proteins for GOV2.0 and the Fram Strait virus dataset. These predicted proteins were annotated with pfam<sup>154</sup> via hmmer<sup>132</sup> using the protein-specific cutoffs for significant (cut\_ga). Then, individual protein families were summed within a given sample. Proteins that had prevalences of less than 10% were removed ( $n = 1,490$  remaining). Then, ANCOM-BC<sup>155</sup> was used to determine which annotations were enriched at latitudes greater than 60°N or S, or less than 60°N or S.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Raw metagenomic reads are available under ENA BioProjects PRJEB67368 (WSC) and PRJEB52171 (EGC). 16S rRNA amplicon reads are available under PRJEB43890 (2016–2017), PRJEB43889 (2017–2018), PRJEB67813 (2018–2019), and PRJEB66202 (2019–2020). Physicochemical parameters are available at PANGAEA under (<https://doi.org/10.1594/PANGAEA.904565>) (2016–2017), (<https://doi.org/10.1594/PANGAEA.904534>) (2017–2018), (<https://doi.org/10.1594/PANGAEA.941126>) (2018–2019), and <https://doi.org/10.1594/PANGAEA.946508> (2019–2020). vOTU contigs and network files are available via figshare DOIs (<https://doi.org/10.6084/m9.figshare.28045856>) and (<https://doi.org/10.6084/m9.figshare.28045826>), respectively. The *psbA* alignment, trimmed alignment, treefile, and labels for cyanophage references are available via figshare (<https://doi.org/10.6084/m9.figshare.28398173>).

### Code availability

Code and inputs for reproducing workflows and figures are available via GitHub [https://github.com/alyzzabc/fram\\_strait\\_viruses\\_2016-2020/](https://github.com/alyzzabc/fram_strait_viruses_2016-2020/) and <https://doi.org/10.5281/zenodo.15490229>. All scripts to reproduce the CON and CCM networks are available via gitlab <https://gitlab.com/qtb-hhu/marine/publications/framphages2024>. CON and CCM calculations can also be performed using the tool OTTER, available as a GUI (<https://doi.org/10.5281/zenodo.13840702>).

### References

- Overland, J. et al. The urgency of Arctic change. *Polar Sci.* **21**, 6–13 (2019).
- Post, E. et al. The polar regions in a 2 °C warmer world. *Sci. Adv.* **5**, eaaw9883 (2019).
- Teng, Z.-J. et al. Biogeographic traits of dimethyl sulfide and dimethylsulfoniopropionate cycling in polar oceans. *Microbiome* **9**, 207 (2021).
- Garneau, M.-È., Roy, S., Lovejoy, C., Gratton, Y. & Vincent, W. F. Seasonal dynamics of bacterial biomass and production in a coastal arctic ecosystem: Franklin Bay, western Canadian Arctic. *J. Geophys. Res. C: Oceans* **113**, C07S91 (2008).
- Harding, K. et al. Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean. *Proc. Natl Acad. Sci. USA.* **115**, 13371–13375 (2018).
- Clark, M. S. et al. Multi-omics for studying and understanding polar life. *Nat. Commun.* **14**, 7451 (2023).
- Tominaga, K. et al. Prevalence of viral frequency-dependent infection in coastal marine prokaryotes revealed using monthly time series virome analysis. *mSystems* **8**, e0093122 (2023).



8. Cordero, O. X. & Polz, M. F. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* **12**, 263–273 (2014).
9. Zimmerman, A. E. et al. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* **18**, 21–34 (2020).
10. Suttle, C. A. et al. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
11. Breitbart, M., et al. Marine viruses: truth or dare. *Ann. Rev. Mar. Sci.* **4**, 425–448 (2012).
12. Yau, S. & Seth-Pasricha, M. Viruses of polar aquatic environments. *Viruses* **11**, 189 (2019).
13. Heinrichs, M. E. et al. Breaking the ice: a review of phages in polar ecosystems. *Methods Mol. Biol.* **2738**, 31–71 (2024).
14. Gregory, A. C. et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**, 1109–1123 (2019).
15. Winter, C., Matthews, B. & Suttle, C. A. Effects of environmental variation and spatial distance on Bacteria, Archaea and viruses in sub-polar and arctic waters. *ISME J.* **7**, 1507–1518 (2013).
16. Lopez-Simon, J. et al. Viruses under the Antarctic Ice Shelf are active and potentially involved in global nutrient cycles. *Nat. Commun.* **14**, 8295 (2023).
17. Meng, L. et al. Genomic adaptation of giant viruses in polar oceans. *Nat. Commun.* **14**, 6233 (2023).
18. Alarcón-Schumacher, T., Guajardo-Leiva, S., Martinez-Garcia, M. & Díez, B. Ecogenomics and adaptation strategies of Southern Ocean viral communities. *mSystems* **6**, e0039621 (2021).
19. Buchholz, H. H. et al. Novel pelagiphage isolate *Polarivirus skadi* is a polar specialist that dominates SAR11-associated bacteriophage communities at high latitudes. *ISME J.* **17**, 1660–1670 (2023).
20. Kim, K. E. et al. Ecological interaction between bacteriophages and bacteria in sub-arctic Kongsfjorden bay, Svalbard, Norway. *Microorganisms* **12**, 276 (2024).
21. Rocchi, A. et al. Abundance and activity of sympagic viruses near the Western Antarctic Peninsula. *Polar Biol.* **45**, 1363–1378 (2022).
22. Sandaa, R.-A. et al. Seasonality drives microbial community structure, shaping both eukaryotic and prokaryotic host–viral relationships in an arctic marine ecosystem. *Viruses* **10**, 715 (2018).
23. Piedade, G. J. et al. Seasonal dynamics and diversity of Antarctic marine viruses reveal a novel viral seascape. *Nat. Commun.* **15**, 9192 (2024).
24. De Corte, D., Sintes, E., Yokokawa, T. & Herndl, G. J. Changes in viral and bacterial communities during the ice-melting season in the coastal Arctic (Kongsfjorden, Ny-Ålesund). *Environ. Microbiol.* **13**, 1827–1841 (2011).
25. Collins, R. E. & Deming, J. W. Abundant dissolved genetic material in Arctic sea ice Part II: Viral dynamics during autumn freeze-up. *Polar Biol.* **34**, 1831–1841 (2011).
26. Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W. & Sullivan, M. B. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J.* **11**, 588 (2017).
27. Fuhrman, J. A., Cram, J. A. & Needham, D. M. Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**, 133–146 (2015).
28. Dart, E., Fuhrman, J. A. & Ahlgren, N. A. Diverse marine T4-like cyanophage communities are primarily comprised of low-abundance species including species with distinct seasonal, persistent, occasional, or sporadic dynamics. *Viruses* **15**, 581 (2023).
29. Ignacio-Espinoza, J. C., Ahlgren, N. A. & Fuhrman, J. A. Long-term stability and red queen-like strain dynamics in marine viruses. *Nat. Microbiol.* **5**, 265–271 (2020).
30. Bolaños, L. M., Michelsen, M. & Temperton, B. Metagenomic time series reveals a Western English Channel viral community dominated by members with strong seasonal signals. *ISME J.* **18**, wræ216 (2024).
31. von Appen, W.-J. et al. Sea-ice derived meltwater stratification slows the biological carbon pump: results from continuous observations. *Nat. Commun.* **12**, 7309 (2021).
32. Soltwedel, T. et al. Natural variability or anthropogenically-induced variation? Insights from 15 years of multidisciplinary observations at the arctic marine LTER site HAUSGARTEN. *Ecol. Indic.* **65**, 89–102 (2016).
33. Wietz, M. et al. The polar night shift: seasonal dynamics and drivers of Arctic Ocean microbiomes revealed by autonomous sampling. *ISME Commun.* **1**, 76 (2021).
34. Cardozo-Mino, M. G. et al. A decade of microbial community dynamics on sinking particles during high carbon export events in the eastern Fram Strait. *Front. Mar. Sci.* **10**, 1173384 (2023).
35. Mioduchowska, M., Pawłowska, J., Mazanowski, K. & Weydmann-Zwolicka, A. Contrasting marine microbial communities of the Fram Strait with the first confirmed record of cyanobacteria *Prochlorococcus marinus* in the Arctic region. *Biology* **12**, 1246 (2023).
36. Asbjørnsen, H., Årthun, M., Skagseth, Ø. & Eldevik, T. Mechanisms underlying recent Arctic Atlantification. *Geophys. Res. Lett.* **47**, e2020GL088036 (2020).
37. Ingvaldsen, R. B. et al. Physical manifestations and ecological implications of Arctic Atlantification. *Nat. Rev. Earth Environ.* **2**, 874–889 (2021).
38. Muilwijk, M. et al. Divergence in climate model projections of future Arctic Atlantification. *J. Clim.* **36**, 1727–1748 (2023).
39. Mosby, H. Water, salt and heat balance of the North Polar Sea and of the Norwegian Sea. *Geophysica Nor.* **24**, 289–313 (1962).
40. Beszczynska-Möller, A., Fahrbach, E., Schauer, U. & Hansen, E. Variability in Atlantic water temperature and transport at the entrance to the Arctic Ocean, 1997–2010. *ICES J. Mar. Sci.* **69**, 852–863 (2012).
41. Aagaard, K., Coachman, L. K. & University of Washington. Department of Oceanography. *The East Greenland Current North of Denmark Strait*. (University of Washington Department of Oceanography, 1969).
42. Priest, T. et al. Seasonal recurrence and modular assembly of an Arctic pelagic marine microbiome. *Nat. Commun.* **16**, 1326 (2025).
43. Knowles, B. et al. Lytic to temperate switching of viral communities. *Nature* **531**, 466–470 (2016).
44. Coutinho, F. H. et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* **8**, 15955 (2017).
45. Alrasheed, H., Jin, R. & Weitz, J. S. Caution in inferring viral strategies from abundance correlations in marine metagenomes. *Nat. Commun.* **10**, 501 (2019).
46. Rodriguez-Valera, F. et al. Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* **7**, 828–836 (2009).
47. Emerson, J. B., Thomas, B. C., Andrade, K., Heidelberg, K. B. & Banfield, J. F. New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. *Appl. Environ. Microbiol.* **79**, 6755–6764 (2013).
48. Brockhurst, M. A. et al. Running with the red queen: the role of biotic conflicts in evolution. *Proc. Biol. Sci.* **281**, 20141382 (2014).
49. Oldenburg, E. et al. Beyond blooms: the winter ecosystem reset determines microeukaryotic community dynamics in the Fram Strait. *Commun. Earth Environ.* **5**, 1–16 (2024).
50. Needham, D. M., Sachdeva, R. & Fuhrman, J. A. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* **11**, 1614–1629 (2017).
51. Thingstad, T. F. 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 (2000).

52. Roux, S. et al. iPhoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* **21**, e3002083 (2023).
53. Bin Jang, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
54. Paulsen, M. L. et al. *Synechococcus* in the Atlantic gateway to the Arctic Ocean. *Front. Mar. Sci.* **3**, 191 (2016).
55. von Friesen, L. W. & Riemann, L. Nitrogen fixation in a changing Arctic Ocean: an overlooked source of nitrogen? *Front. Microbiol.* **11**, 596426 (2020).
56. Pons, J. C. et al. VPF-Class: taxonomic assignment and host prediction of uncultivated viruses based on viral protein families. *Bioinformatics* **37**, 1805–1813 (2021).
57. Tolar, B. B. et al. Contribution of ammonia oxidation to chemolithotrophy in Antarctic coastal waters. *ISME J.* **10**, 2605–2619 (2016).
58. Priest, T. et al. Atlantic water influx and sea-ice cover drive taxonomic and functional shifts in Arctic marine bacterial communities. *ISME J.* **17**, 1612–1625 (2023).
59. Milici, M. et al. Low diversity of planktonic bacteria in the tropical ocean. *Sci. Rep.* **6**, 19054 (2016).
60. Fuhrman, J. A. et al. A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl. Acad. Sci. USA* **105**, 7774–7778 (2008).
61. Ladau, J. et al. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**, 1669–1677 (2013).
62. Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–463 (2017).
63. Ghiglione, J.-F. et al. Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc. Natl. Acad. Sci. USA* **109**, 17633–17638 (2012).
64. Sul, W. J., Oliver, T. A., Ducklow, H. W., Amaral-Zettler, L. A. & Sogin, M. L. Marine bacteria exhibit a bipolar distribution. *Proc. Natl. Acad. Sci. USA* **110**, 2342–2347 (2013).
65. Clayton, S. et al. Bio-GO-SHIP: The time is right to establish global repeat sections of ocean biology. *Front. Mar. Sci.* **8**, 767443 (2022).
66. Larkin, A. A. et al. High spatial resolution global ocean metagenomes from Bio-GO-SHIP repeat hydrography transects. *Sci. Data* **8**, 107 (2021).
67. Roux, S. et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).
68. Sánchez, P. et al. Marine picoplankton metagenomes and MAGs from eleven vertical profiles obtained by the Malaspina Expedition. *Sci. Data* **11**, 154 (2024).
69. Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
70. Ignacio-Espinoza, J. C. & Sullivan, M. B. Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes: Molecular evolution of T4 myoviruses. *Environ. Microbiol.* **14**, 2113–2126 (2012).
71. Bongiorno, L., Magagnoli, M., Armeni, M., Noble, R. & Danovaro, R. Viral production, decay rates, and life strategies along a trophic gradient in the North Adriatic Sea. *Appl. Environ. Microbiol.* **71**, 6644–6650 (2005).
72. Chen, P. W.-Y., Olivia, M., Chou, W.-C., Mukhanov, V. & Tsai, A.-Y. Differences in viral decay and production following exposure to sunlight and dark. *Terr. Atmos. Ocean. Sci.* **34**, 8 (2023).
73. Jacquet, S. & Bratbak, G. Effects of ultraviolet radiation on marine virus-phytoplankton interactions. *FEMS Microbiol. Ecol.* **44**, 279–289 (2003).
74. Mojica, K. D. A. & Brussaard, C. P. D. Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *FEMS Microbiol. Ecol.* **89**, 495–515 (2014).
75. López-García, P. et al. Metagenome-derived virus-microbe ratios across ecosystems. *ISME J.* **17**, 1552–1563 (2023).
76. Brüwer, J. D. et al. Globally occurring pelagiphage infections create ribosome-deprived cells. *Nat. Commun.* **15**, 3715 (2024).
77. Jang, H. B. et al. Viral tag and grow: a scalable approach to capture and characterize infectious virus-host pairs. *ISME Commun.* **2**, 12 (2022).
78. Kirzner, S., Barak, E. & Lindell, D. Variability in progeny production and virulence of cyanophages determined at the single-cell level. *Environ. Microbiol. Rep.* **8**, 605–613 (2016).
79. Duhaime, M. B. et al. Comparative omics and trait analyses of marine *Pseudoalteromonas* phages advance the phage OTU concept. *Front. Microbiol.* **8**, 1241 (2017).
80. Parada, V., Herndl, G. J. & Weinbauer, M. G. Viral burst size of heterotrophic prokaryotes in aquatic systems. *J. Mar. Biol. Assoc. U. K.* **86**, 613–621 (2006).
81. Zayed, A. A. et al. Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science* **376**, 156–162 (2022).
82. Dominguez-Huerta, G. et al. Diversity and ecological footprint of Global Ocean RNA viruses. *Science* **376**, 1202–1208 (2022).
83. Steward, G. F. et al. Are we missing half of the viruses in the ocean? *ISME J.* **7**, 672–679 (2013).
84. Miranda, J. A., Culley, A. I., Schvarcz, C. R. & Steward, G. F. RNA viruses as major contributors to Antarctic viroplankton. *Environ. Microbiol.* **18**, 3714–3727 (2016).
85. Kaneko, H. et al. Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience* **24**, 102002 (2021).
86. Warwick-Dugdale, J. et al. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. *PeerJ* **7**, e6800 (2019).
87. Warwick-Dugdale, J. et al. Long-read powered viral metagenomics in the oligotrophic Sargasso Sea. *Nat. Commun.* **15**, 4089 (2024).
88. Zaragoza-Solas, A., Haro-Moreno, J. M., Rodríguez-Valera, F. & López-Pérez, M. Long-read metagenomics improves the recovery of viral diversity from complex natural marine samples. *mSystems* **7**, e0019222 (2022).
89. Brum, J. R. et al. Patterns and ecological drivers of ocean viral communities. *Science* **348**, (2015).
90. Short, C. M., Rusanova, O. & Short, S. M. Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. *ISME J.* **5**, 810–821 (2011).
91. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.* **11**, 1511–1520 (2017).
92. Oldenburg, E. et al. Sea-ice melt determines seasonal phytoplankton dynamics and delimits the habitat of temperate Atlantic taxa as the Arctic Ocean atlantifies. *ISME Commun.* **4**, ycae027 (2024).
93. Freeman, N. M., Lovenduski, N. S. & Gent, P. R. Temporal variability in the Antarctic Polar Front (2002–2014). *J. Geophys. Res. C: Oceans* **121**, 7263–7276 (2016).
94. Kolås, E. H., Fer, I. & Baumann, T. M. The Polar Front in the north-western Barents Sea: structure, variability and mixing. *Ocean Sci.* **20**, 895–916 (2024).
95. Landry, M. R. et al. Seasonal dynamics of phytoplankton in the Antarctic Polar Front region at 170°W. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **49**, 1843–1865 (2002).
96. Smetacek, V., De Baar, H. J. W., Bathmann, U. V., Lochte, K. & Rutgers Van Der Loeff, M. M. Ecology and biogeochemistry of the antarctic circumpolar current during austral spring: a summary of southern ocean JGOFS cruise ANT X/6 of R.V. Polarstern. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **44**, 1–21 (1997).

97. Buscaglia, M., Iriarte, J. L., Schulz, F. & Diez, B. Adaptation strategies of giant viruses to low-temperature marine ecosystems. *ISME J.* **18**, wrae162 (2024).
98. Evans, C. & Brussaard, C. P. D. Viral lysis and microzooplankton grazing of phytoplankton throughout the Southern Ocean. *Limnol. Oceanogr.* **57**, 1826–1837 (2012).
99. Mojica, K. D. A., Huisman, J., Wilhelm, S. W. & Brussaard, C. P. D. Latitudinal variation in virus-induced mortality of phytoplankton across the North Atlantic Ocean. *ISME J.* **10**, 500–513 (2016).
100. Thingstad, T. F., Våge, S., Storesund, J. E., Sandaa, R.-A. & Giske, J. A theoretical analysis of how strain-specific viruses can control microbial species diversity. *Proc. Natl. Acad. Sci. USA* **111**, 7813–7818 (2014).
101. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18**, 1403–1414 (2016).
102. McDonald, D. et al. Greengenes2 unifies microbial data in a single reference tree. *Nat. Biotechnol.* **42**, 715–718 (2024).
103. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* **4**, e2584 (2016).
104. Bokulich, N. A. et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, (2018).
105. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
106. Feng, X., Cheng, H., Portik, D. & Li, H. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nat. Methods* **19**, 671–674 (2022).
107. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
108. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
109. Hou, S. et al. DeepMicroClass sorts metagenomes into prokaryotes, eukaryotes and viruses. *NAR Genom. Bioinform.* **6**, lqae044 (2024).
110. Roux, S. et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2019).
111. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
112. Li, W., Fu, L., Niu, B., Wu, S. & Wooley, J. Ultrafast clustering algorithms for metagenomic sequence analysis. *Brief. Bioinform.* **13**, 656–668 (2012).
113. Camargo, A. P. et al. Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* **42**, 1303–1312 (2024).
114. Coutinho, F. H. et al. RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. *Patterns* **2**, 100274 (2021).
115. Galiez, C., Siebert, M., Enault, F., Vincent, J. & Söding, J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**, 3113–3114 (2017).
116. Ahlgren, N. A., Ren, J., Lu, Y. Y., Fuhrman, J. A. & Sun, F. Alignment-free d<sub>2</sub><sup>2</sup>\* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res* **45**, 39–53 (2017).
117. Lu, C. et al. Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. *BMC Biol.* **19**, 5 (2021).
118. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).
119. Priest, T., Vidal-Melgosa, S., Hehemann, J.-H., Amann, R. & Fuchs, B. M. Carbohydrates and carbohydrate degradation gene abundance and transcription in Atlantic waters of the Arctic. *ISME Commun.* **3**, 130 (2023).
120. Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
121. Karlicki, M., Antonowicz, S. & Karnkowska, A. Tiara: deep learning-based classification system for eukaryotic sequences. *Bioinformatics* **38**, 344–350 (2022).
122. Oksanen, J. et al. vegan: Community Ecology Package. <https://CRAN.R-project.org/package=vegan> (2024).
123. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*, (Springer-Verlag New York, 2016).
124. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: Interpolation and Extrapolation for Species Diversity. [http://chao.stat.nthu.edu.tw/wordpress/software\\_download/](http://chao.stat.nthu.edu.tw/wordpress/software_download/) (2024).
125. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
126. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
127. Sugihara, G. et al. Detecting causality in complex ecosystems. *Science* **338**, 496–500 (2012).
128. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).
129. Xia, L. C. et al. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Syst. Biol.* **5**, S15 (2011).
130. R. Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022).
131. Harrell, F. E. Jr. Hmisc: Harrell Miscellaneous. <https://CRAN.R-project.org/package=Hmisc> (2024).
132. Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
133. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
134. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* **50**, D785–D794 (2022).
135. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
136. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
137. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
138. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
139. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
140. Zhang, R., Debeljak, P., Blain, S. & Obernosterer, I. Seasonal shifts in Fe-acquisition strategies in Southern Ocean microbial communities revealed by metagenomics and autonomous sampling. *Environ. Microbiol.* **25**, 1816–1829 (2023).
141. Dlugosch, L. et al. Significance of gene variants for the functional biogeography of the near-surface Atlantic Ocean microbiome. *Nat. Commun.* **13**, 456 (2022).
142. Cao, S. et al. Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. *Microbiome* **8**, 47 (2020).



143. Collins, E. Arctic Ocean microbial metagenomes sampled aboard CGC Healy during the 2015 GEOTRACES Arctic research cruise. *SCAR - Microb. Antarct. Resour. Syst.* <https://doi.org/10.15468/ILJMUN> (2019).
  144. Priest, T., Orellana, L. H., Huettel, B., Fuchs, B. M. & Amann, R. Microbial metagenome-assembled genomes of the Fram Strait from short and long read sequencing platforms. *PeerJ* **9**, e11721 (2021).
  145. Akpudo, Y. M., Bezuidt, O. K. & Makhanyane, T. P. Metagenome-assembled genomes of four Southern Ocean Archaea harbor multiple genes linked to polyethylene terephthalate and polyhydroxybutyrate plastic degradation. *Microbiol. Resour. Announc.* **12**, e0109822 (2023).
  146. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
  147. Aroney, S. T. N. et al. *CoverM: Read Coverage Calculator for Metagenomics*. (2024). <https://doi.org/10.5281/zenodo.10531253>.
  148. Wood, S. N. Fast Stable Restricted Maximum Likelihood and Marginal Likelihood estimation of semiparametric Generalized Linear Models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 3–36 (2010).
  149. Warton, D. I., Wright, S. T. & Wang, Y. Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* **3**, 89–101 (2012).
  150. Olm, M. R. et al. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. *Nat. Biotechnol.* **39**, 727–736 (2021).
  151. Gasteiger, E. et al. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana Press, Totowa, NJ, 2005).
  152. Ikai, A. Thermostability and aliphatic index of globular proteins. *J. Biochem.* **88**, 1895–1898 (1980).
  153. Osorio, D., Rondón-Villarreal, P. & Torres, R. Peptides: a package for data mining of antimicrobial peptides. *The R Journal* **7**, 4–14 (2015).
  154. Finn, R. D. et al. Pfam: the protein families database. *Nucleic Acids Res* **42**, D222–D230 (2014).
  155. Lin, H. & Peddada, S. D. Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* **11**, 3514 (2020).
- Polarstern grants AWI\_PS99\_00, AWI\_PS107\_05, AWI\_PS114\_01, AWI\_PS121\_07, AWI\_PS126\_05, and AWI\_PS126\_07. Further support came from the Helmholtz Association, the Max Planck Society, and a Helmholtz Young Investigator Grant to D.M.N. M.W. was supported by the German Research Foundation grant 522416631 within the SPP 1158.

## Author contributions

A.M.C. identified, annotated, mapped, and assembled viruses. A.M.C. and D.M.N. analyzed viral and prokaryotic community data. T.P. contributed normalization techniques and alpha diversity metrics. E.O. and O.P. performed network analysis and module identification. J.M. performed amino acid analyses. M.W. and D.M.N. instigated the study. D.M.N. designed and coordinated the analyses. A.M.C. and D.M.N. wrote the paper with input from all co-authors, especially T.P. and M.W.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-61568-6>.

**Correspondence** and requests for materials should be addressed to David M. Needham.

**Peer review information** *Nature Communications* thanks Holger Buchholz, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2025

## Acknowledgements

Christina Bienhold, Katja Metfies, Ian Salter and Antje Boetius co-designed the mooring strategy and coordinated sample collection and processing. Katja Metfies coordinated amplicon sequencing. Wilken-Jon von Appen, Sinhué Torres-Valdés, and Daniel Scholz carried out physicochemical and oceanographic measurements. We thank Jana Bäger, Theresa Hargesheimer, Jakob Barz, Anja Batzke, Rafael Stiens, and Lili Hufnagel for RAS operations; Normen Lochthofen, Janine Ludszuweit, Lennard Frommhold and Jonas Hagemann for mooring operations; Jakob Barz, Swantje Ziemann and Anja Batzke for DNA extraction, amplicon library preparation and sequencing; and Bruno Huettel, Christian Woehle and the technicians at the Max Planck Genome Center in Cologne for metagenome sequencing. Captains, crew and scientists of RV Polarstern cruises PS99.2, PS107, PS114, PS121 and PS126 are gratefully acknowledged. This project has received funding from