ISME

ARTICLE   OPEN

Novel pelagiphage isolate *Polarivirus skadi* is a polar specialist that dominates SAR11-associated bacteriophage communities at high latitudes

Holger H. Buchholz[1], Luis M. Bolaños [1], Ashley G. Bell [1], Michelle L. Michelsen[1], Michael J. Allen[1] and Ben Temperton [1 ✉]

The SAR11 clade are the most abundant members of surface marine bacterioplankton and a critical component of global biogeochemical cycles. Similarly, pelagiphages that infect SAR11 are ubiquitous and highly abundant in the oceans. Pelagiphages are predicted to shape SAR11 community structures and increase carbon turnover throughout the oceans. Yet, ecological drivers of host and niche specificity of pelagiphage populations are poorly understood. Here we report the global distribution of a novel pelagiphage called "*Polarivirus skadi*", which is the sole representative of a novel genus. *P. skadi* was isolated from the Western English Channel using a cold-water ecotype of SAR11 as bait. *P. skadi* is closely related to the globally dominant pelagiphage HTVC010P. Along with other HTVC010P-type viruses, *P. skadi* belongs to a distinct viral family within the order Caudovirales, for which we propose the name Ubiqueviridae. Metagenomic read recruitment identified *P. skadi* as one of the most abundant pelagiphages on Earth. *P. skadi* is a polar specialist, replacing HTVC010P at high latitudes. Experimental evaluation of *P. skadi* host range against cold- and warm-water SAR11 ecotypes supported cold-water specialism. Relative abundance of *P. skadi* in marine metagenomes correlated negatively with temperature, and positively with nutrients, available oxygen, and chlorophyll concentrations. In contrast, relative abundance of HTVC010P correlated negatively with oxygen and positively with salinity, with no significant correlation to temperature. The majority of other pelagiphages were scarce in most marine provinces, with a few representatives constrained to discrete ecological niches. Our results suggest that pelagiphage populations persist within a global viral seed bank, with environmental parameters and host availability selecting for a few ecotypes that dominate ocean viromes.

INTRODUCTION

Bacteriophages are the most abundant organisms in the oceans and play a critical role in driving microbial global biogeochemical cycles, particularly carbon cycling [1–4]. During an infection, phages hijack their hosts, alter cellular metabolism, and shape the structure of microbial populations [5–7]. Ultimately, phages kill bacterial host cells, releasing the carbon and other cell-bound nutrients back into the environment, thereby recycling organic compounds [3, 8]. These compounds are remineralised by other microbes within the "viral shunt", which is thought to be an essential part of the marine food web [3, 9]. Phages drive microbial evolution and are agents of frequent genetic exchange between hosts, as well as other phages, creating highly diverse genomic populations [10, 11]. In marine systems, discrete viral populations persist within spatial and temporal niches, in which the majority of sequence variation occurs in metabolic genes associated with niche specificity [12, 13]. Genomic clusters for T4-like phages infecting Cyanobacteria, were found to exhibit spatial and temporal abundance patterns suggesting that they represent distinct viral ecotypes [13]. Viral communities at the Bermuda Atlantic Time-Series (BATS) station also revealed recurrent seasonal patterns linking viral communities with stratification and mixing events in the water column [14]. Temporal analysis of

T4-like phages and their hosts at the San Pedro Ocean Time series station further suggested that the relative abundance of virus populations is tightly linked to those of their associated hosts [15]. Therefore, spatiotemporal distribution and ecotypes of viral populations are interlinked and influenced by both the physical and biological parameters that shape the ecotypic specificity of their hosts [16].

The ubiquitous alphaproteobacterial order *Pelagibacterales* (clade SAR11) comprises up to one-third of total marine microbial communities, making them likely the most abundant bacteria on Earth [17, 18]. SAR11 are an ancient and diverse clade, characterised by small, streamlined genomes with low G + C% content, high gene synteny, and are phylogenetically organised into specialised temperature-associated ecotypes [19–21]. Phages infecting members of the SAR11 (pelagiphages) are the most abundant viruses in global ocean viromes [22, 23] and have been isolated across most major oceans, resulting in 40 different cultivated species, including 37 of podophage morphotypes [22–27], two myophages [22, 25], and one siphophage [26]. Owing to the challenges of culturing SAR11 [28, 29] and isolating phages for them [26], experimental evaluation of predicted pelagiphage ecology is in its infancy. Read recruitment from global viral metagenomes to genomes of pelagiphage isolates has

[1]School of Biosciences, University of Exeter, Exeter, UK. ✉email: b.temperton@exeter.ac.uk

Table 1. Overview of general features and accession numbers for *P. skadi* -like viral isolate genome.

Phage	Collection number	Genome size (bp)	G + C (%)	Terminal repeat (bp)	ANI (%) to Skadi-1	ORFs	Accession number	Collection date
Skadi-1	EXVC101P	35,392	31.5	125	100.0	65	OP131302	Sept 24, 2018
Skadi-2	EXVC102P	35,392	31.5	125	99.8	65	OP131299	Sept 24, 2018
Skadi-3	EXVC103P	35,083	31.5	n.a.	99.5	65	OP131297	Sept 24, 2018
Skadi-4	EXVC104P	34,897	31.5	n.a.	99.3	65	OP131301	Sept 24, 2018
Skadi-5	EXVC105P	35,374	31.5	125	99.6	65	OP131293	Sept 24, 2018
Skadi-6	EXVC106P	35,392	31.6	125	99.8	65	OP131294	Sept 24, 2018
Skadi-7	EXVC107P	35,247	31.6	n.a.	99.6	65	OP131300	Nov 05, 2018
Skadi-8	EXVC108P	35,276	31.6	n.a.	99.6	65	OP131296	Nov 05, 2018
Skadi-10	EXVC110P	35,267	31.5	n.a.	99.8	65	OP131295	Nov 05, 2018
Skadi-11	EXVC111P	9865	32.2	n.a.	n.a.	n.a.	OP131298	Sept 24 2018

identified abundant and cosmopolitan distributions of several pelagiphages irrespective of temperature (such as HTVC010P [22], HTVC023P [23], and vSAG-37-F6 [30, 31]). However, when conservative read recruitment standards are applied, including a minimum genome coverage cutoff (read identity 90% and minimum covered fraction of 40%) to avoid false discovery rate of 20–25% [32], the vast majority of pelagiphages (22 out of 32 with mean RPKM < 100, 31 out of 32 with median RPKM of zero [26]) have low abundance distributed across few geographic locations, in support of the viral seed-bank hypothesis [33], which states that ocean currents distribute phages across the globe, but only a fraction of phages in the ocean are actively replicating, depending on local host availability. HTVC010P is thought to be the sole representative of cultured pelagiphages with high abundance in epipelagic samples from both temperate, tropical, and polar regions.

Two possibilities may explain the extraordinary abundance of HTVC010P across the global oceans. One hypothesis is that HTVC010P has a broad host range and is capable of infecting multiple host ecotypes and are thus propagated in both warm and cold-water environments. Experimentally confirmed host ranges for HTVC010P are not available and the isolate has since been lost from culture. But for the majority of other pelagiphages [34], evidence suggests that almost all of them have narrow host ranges restricted to specific host ecotypes [23, 24, 26]), at least under laboratory conditions. To date, only a single, low abundance pelagiphage (Bylgja) has been found to experimentally infect both warm and cold-water ecotypes of SAR11 [26]. An alternative hypothesis is that abundant variants of HTVC010P exist which are ecotype-specific, but which have not yet been captured in pelagiphage cultures, causing reads to recruit to the most closely related strain in the databases [23]. Viral single-amplified genomes (vSAGs) and metagenomically assembled viral genomes (MAVGs) of predicted pelagiphages (without cultured representatives) suggest that current viral isolates only cover a fraction of the available pelagiphage diversity [30, 35, 36].

In this study, we isolated and sequenced ten strains of a novel pelagiphage species *Polarivirus skadi*, which is the sole representative of a novel genus. *P. skadi* was named after the Norse goddess of winter. We demonstrate that *P. skadi* is closely related to the globally abundant HTVC010P, both in gene sharing networks and single-gene phylogeny of multiple shared genes. Together with our previously isolated pelagiphages *Iscavirus greipi* (formerly Greip) and *Diaboliformavirus lederbergi* (formerly Lederberg), as well as seven other HTVC010P-type pelagiphages, they form a distinct novel viral family, for which we propose the name *Ubiqueviridae*. Our analysis suggests that each phage within the *Ubiqueviridae* family is a unique species and sole representatives of their own genera. Recruitment of reads from global marine viromes revealed that *P. skadi* is a polar pelagiphage ecotype that

dominates Arctic and Antarctic marine viral communities, with read recruitment rapidly decreasing in viromes from lower latitudes. At tropical and subtropical latitudes, the representative Skadi-1 genome is near or below limits of detection, apart from regions with direct influx of polar water, such as the Humboldt Current System in the Southern Pacific Ocean. The global distribution pattern, in addition to experimentally confirmed host ranges, suggest that abundant populations of *P. skadi* replicate in polar waters on cold-water ecotypes of SAR11 and are potentially transported around the globe in currents of sinking polar water which are part of the ocean conveyor belt. *P. skadi* therefore serves as an exemplar of niche specialisation in pelagiphages and a model organism for determining how ocean currents might shape regional viral communities through the viral seed bank.

RESULTS AND DISCUSSION
Polarivirus skadi is closely related to Pelagibacter phage HTVC010P

A total of 10 *Pelagibacter* phages were isolated on the SAR11 cold-water ecotype *Pelagibacter ubique* HTCC1062 from three coastal surface water samples taken from the Western English Channel (WEC) in September, October, and November of 2018 [26]. Nine out of 10 genomes were assembled into single circular contigs, with one assembly returning a fragmented, linear genome. Four out of the nine circular contigs contained a direct terminal repeat of 125 bp at the end of the linearised genome indicative of *pac* headful packaging similar to the T7-like P-SSP7 cyanophage [37]. General features of the genomes are shown in Table 1. Genome sizes ranged from 34,897 base pairs (bp) to 35,392 bp with a G + C content between 31.5% and 32.5%. All full-length genomes contained 65 putative open reading frames (ORFs) (Fig. 1). Pairwise average nucleotide identity (ANI) values between strains ranged from 99.3% to 99.8%, all strains were therefore considered to be the same viral species [38]. For analysis purposes, we used Skadi-1 as reference for the entire species unless specified otherwise. Skadi-1 was related to HTVC010P (ANI 32.9%) and other HTVC010P-type pelagiphages, based on whole-genome phylogenetic analysis for all *Pelagibacter* phages (Fig. 2). A smaller capsid (39 ± 0.7 nm (Fig. 1B), compared to 50 ± 3 nm for HTVC010P [22]) was observed in Skadi-1 virions, but the difference in capsid size reported for HTVC010P [22, 23] and here for Skadi-1 may be partly due to different operators and protocols for performing transmission electron microscopy (TEM). Like HTVC010P, the short stubby tail observed in Skadi-1 suggested a podophage morphology.

Polarivirus skadi and other HTVC010P-type pelagiphages constitute a novel viral family within the Caudoviricetes class

Analysis of average nucleotide identity between Skadi-1 and all other known phages infecting *Pelagibacter* showed that the closest
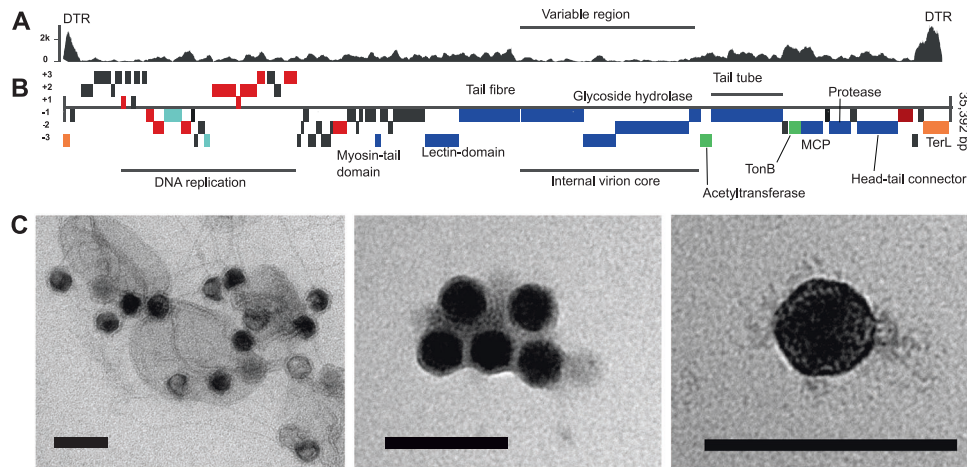
Fig. 1 General features of *Polarivirus skadi*. A Per nucleotide coverage of the Skadi-1 genome. High coverage at contig-ends suggest direct terminal repeats; the region encoding for virion proteins has low average coverage, suggestive of a hypervariable region. **B** Genome map of Skadi-1, translation frames of open reading frames (ORFs) are indicated in the figure, functional annotations of ORFs are colour coded with red: DNA replication and metabolism; teal: transcription; blue: Virus structural genes; green: virulence related genes; Orange: packaging. **C** Transmission electron microscopy (TEM) images, black scale bar represents 100 nm; left: *P. skadi* viral particles and cellular debris of lysed HTCC1062 host cells; middle: icosahedral viral capsids of *P. skadi* with a size of 39 ± 0.7 nm; right: *P. skadi* single viral particle.

related phages were Greip (ANI 37%) [26], Lederberg (ANI 33%) [25], and the highly abundant HTVC010P-type pelagiphages (ANI 30–49%) [22, 27]. The average nucleotide identity between all these phages was <60%, which, based on established ICTV thresholds, suggests that each HTVC010P-like phage isolate is a unique species and sole representative of a separate genus respectively [38]. We propose these phages as representatives of 11 novel genera, for which we suggest latinised binomial names based on the names of the reference genomes and sampling locations (Table 2).

Other known pelagiphages with podovirus morphology had ANI values <5%. A phylogenetic tree based on complete whole-genome protein sequences placed Skadi-1 and HTVC010P-like phages on a monophyletic branch distinct from other phage groups, providing support for a family-level taxonomic assignment (Fig. 2). This comparative phylogenomic analysis against all dsDNA prokaryotic and archaeal viruses within the DB-B:Baltimore Group also suggested that there are no other cultivated phages in the database which are related to *P. skadi* and HTVC010P-type phages, and that the closest related families were *Schitoviridae* and *Autographviridae*. A full-length genome based phylogenetic analysis using 17 arbitrarily chosen representatives of the *Schitoviridae* and *Autographviridae* families, which are two main ICTV recognised families with podovirus morphology, together with all 40 *Pelagibacter* phages also placed Skadi-1 and HTVC010P-type phages on a well-supported branch distinct to all other branches (Fig. 2B). This suggests that these phages are members of the same family-level taxonomic group, for which we suggest the name *Ubiqueviridae* – derived from the Latin word *ubique* ("everywhere") to reflect both the high abundance (HTVC010P-type phages are widely accepted to be one of the most abundant marine viruses [22, 27]) and the ubiquitous distribution of these phages, as well as in recognition of their common host, *Pelagibacter ubique*.

Further evidence that *Ubiqueviridae* represents a distinct taxonomic cluster from other pelagiphages was provided by analysing the hypergeometric probability of shared protein clusters (Fig. 2C, D), showing high probability of shared genes within groups corresponding to pelagiphage types, with *Ubiqueviridae* forming a single viral cluster. The taxonomic assignment tool vConTACT2 against database INPHARED [39] (accessed March 1, 2023) and RefSeq V88 placed the *Ubiqueviridae* into a viral cluster separate from other phages. A shared gene network

analysis with all RefSeq phages showed that phages with podovirus morphology infecting SAR11, as well as the taxonomically undefined pelagisiphophage Kolga, share more genes than compared to other pelagiphages and viruses of other hosts. PIRATE analysis [40] identified ten shared core genes within all eleven members of the proposed *Ubiqueviridae* family (~16% of ORFs in Skadi-1) (Supplementary Table 1). *Ubiqueviridae* did not share core genes with pelagiphages of the HTVC019P-types (Unconfirmed part of the *Autographviridae* family) or taxonomically unclassified HTVC111P-types, HTVC103P-types, or HTVC023P-types. One shared core gene (*TerL*) was found between *Ubiqueviridae* and HTVC106P. Phylogenetic analysis of *TerL* shows that the *Ubiqueviridae* genes form a monophyletic branch, with HTVC106P as a closely related, but separate branch (Supplementary Fig. 1). Additional phylogenetic analysis on individual structural genes (tail tube and major capsid proteins) that are commonly found in dsDNA phages also supports the proposed *Ubiqueviridae* as a monophyletic group, and places other pelagiphage types onto distinct and well-supported (bootstrap value > 0.94) monophyletic branches (Supplementary Figs. 2 and 3). Other pelagiphage types had 15–33 shared core genes in their respective groups (Supplementary Table 2). No core genes were shared between *Ubiqueviridae* and phages of other morphotypes that have been experimentally confirmed to infect *Pelagibacter* [22–24, 27, 41], the unassigned siphophage Kolga, and myophages Melnitz and HTVC008M.

Polarivirus skadi uses host RNA and DNA polymerases for viral replication

Whole-genome alignment showed that the Skadi-1 genome is arranged into early and late genomic regions, similar to other HTVC010P-type genomes [27]. A TonB-dependent receptor protein located between the two tail tube proteins A and B and the major capsid protein suggests TonB as a putative receptor for adsorption and initiating infection [42, 43] (Fig. 1). TonB-dependent transporters are one of the dominant proteins found in surface ocean metaproteomes, with increasing abundance in nutrient-rich, coastal waters [44]. We did not find Skadi-like TonB-receptors in other cultivated pelagiphages, suggesting that these species may use different receptor proteins. We identified three internal virion proteins similar to gp14, gp15, and gp16 found in T7, which likely share a similar role in injecting phage DNA through the host periplasm during infection [45, 46]. A proximate
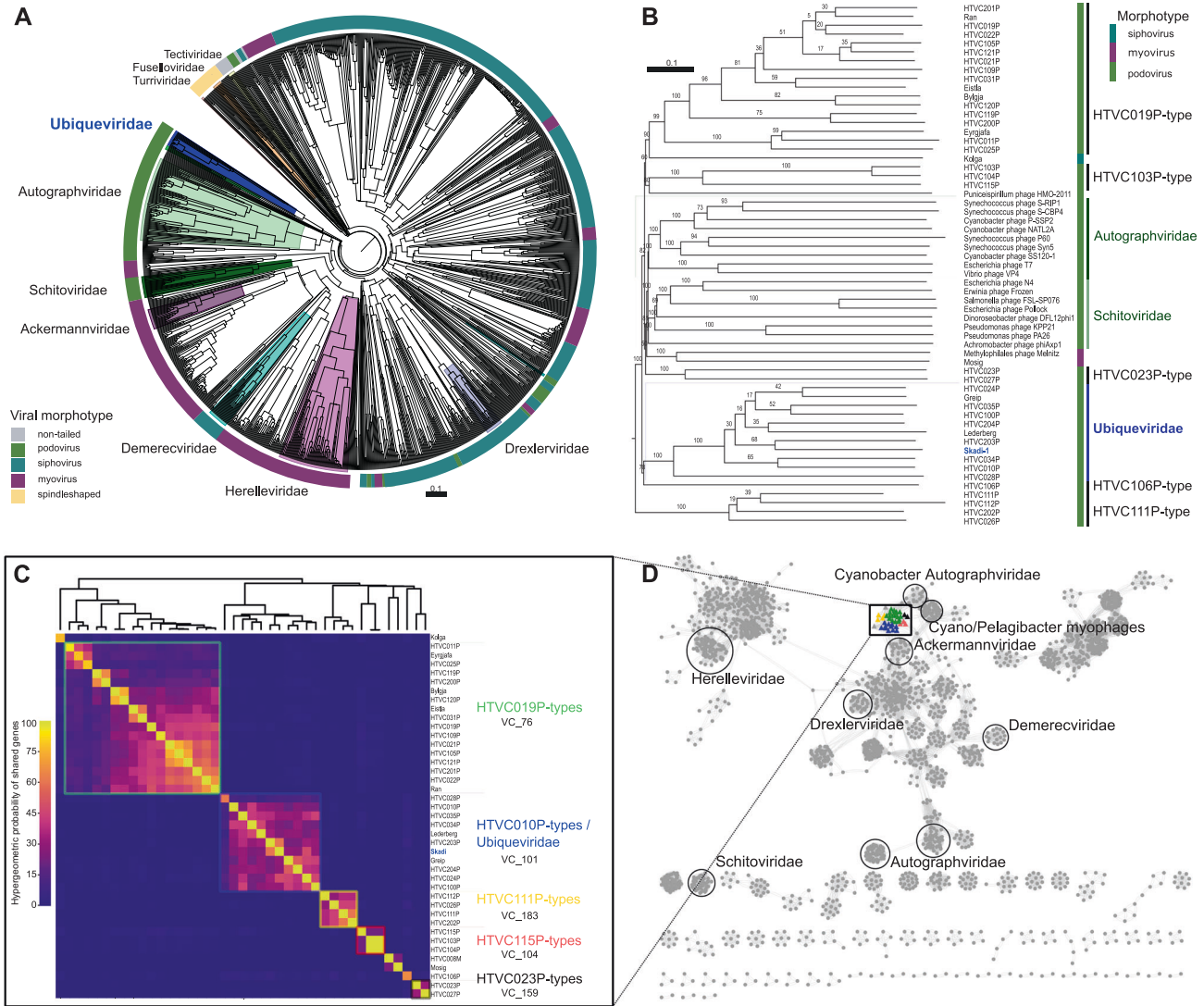
Fig. 2 *Pelagibacter* **phages can be grouped into distinct phylogenetic clusters. A** All full-length genomes were screened against DB-B:Baltimore Group Ib prokaryotic and archaeal, calculated using GRAViTy 1.1.0 (http://gravity.cvr.gla.ac.uk/). Outer ring represents the main virus morphotypes, branch highlights indicate ICTV recognised viral families. Blue highlighted branch represents the proposed *Ubiqueviridae* family. **B** Full-genome phylogenetic tree created in VICTOR [86] using known *Pelagibacter* phage genomes, leaf-label for Skadi-1 is highlighted in red. Randomly chosen representative genomes of the *Autographviridae* (light green) and *Schitoviridae* (dark green) families were included and highlighted, the branch with all members of the proposed *Ubiqueviridae* is highlighted in blue. All bootstrap values are shown. **C** Heatmap of hypergeometric probability of shared genes between phages isolated on SAR11 hosts based on protein clusters identified by vConTACT2, highlights indicate corresponding vConTACT2 viral cluster assignments, no highlight indicates singleton viral clusters in vConTACT2 gene sharing analysis. Hypergeometric probability of shared genes between Skadi-like and HTVC010P-like phages suggests both are relatively closely related. **D** Visualised vConTACT2 analysis of shared gene network. The location of genomes of ICTV recognised bacteriophage families and genomes of phages infecting SAR11 with podovirus morphology are indicated, coloured nodes represent viral cluster identified by vConTACT2 and correspond to viral clusters colour coded in (**A**).

glycoside hydrolase gene found in Skadi-1 is likely a structural component of the internal virion complex, facilitating the injection of phage DNA by breaking glycosidic bonds in the cell wall, as was found to be the role of the glycoside hydrolase (Gp255) in *Bacillus* phage vB_BpuM_BpSp [47].

Skadi-1 encodes for an acetyltransferase gene downstream of the internal core complex that may play a role in metabolic hijacking and host programming via host quorum sensing networks [48]. In a phiKMV-like virus infecting *Pseudomonas aeruginosa*, acetyltransferases were responsible for the shutdown of host transcription by cleaving the bacterial RNA polymerase (RNAP) during the early stages of infection [49]. However, unlike pelagiphages in the *Autographviridae* family, Skadi-1does not encode a viral RNA polymerase (RNAP), making cleavage of host

RNAP unlikely. Temporal regulation in the absence of endogenous RNAP is common in T4- and λ-like phages, and involves sophisticated mechanisms that rely on early to late promoter sequences for recognition by the host RNAP [50]. We identified four σ^{70} promoters and two Pribnow boxes (TATAAT-promoter sequences) in proximity of DNA replication and manipulation genes, but were unable to identify any terminator sequences, tRNA, tmRNA or ribonucleotide switches. This could suggest that Skadi-1 uses host tRNAs, reinforcing similar codon usage between the host and phage, and potentially constraining host range [51]. Skadi-1 also lacks a phage encoded DNA polymerase (DNAP), but encodes for two transcriptional regulator proteins, similar to HTVC010P-type phages [27]. Likely these transcriptional regulators are used to hijack the cellular transcriptional machinery, indicating

that Skadi-1, as predicted in other HTVC010P-type phages, relies on host RNAP and DNAP proteins for translation and transcription.

An additional methylase gene likely provides Skadi-1 with protection from host restriction endonucleases found in *Pelagibacter ubique* HTCC1062 (SAR11_RS00560), or might help to protect Skadi-1 virocells against superinfecting DNA from phage competitors as endonucleases are a common feature in pelagiphages [52]. Skadi-1 further encodes peptide hydrolase and nuclease genes that are likely involved in breaking down host proteins and nucleic acids during the early infection stage to recycle material for phage synthesis [53]. Skadi-1 encodes a helicase loader protein similar to the λ-like DNA replication machinery, without virally encoded genes for helicases or primases. In addition to the aforementioned lack of a terminator sequence, this suggests that Skadi-1 utilises "rolling-circle"-like DNA replication. Skadi-1 encodes several common phage structure proteins, namely: three internal virion proteins, major capsid protein, tail tube proteins A and B, tail fibres, head-tail connector, and prohead-protease. Located within the direct terminal repeat region is a conserved TerL gene for packaging. For cell lysis, Skadi-1 encodes a peptidase M15, similar to other HTVC010P-type pelagiphages [27].

Skadi-1 contains an arms-race associated hypervariable region that is conserved across HTVC010P-like phages

Alignment of HTVC010P-type genomes showed that regions related to DNA replication and metabolism, as well as the terminase and phage head related proteins are conserved, using recommended amino acid identity (AAI) thresholds of >30% [38] (Supplementary Fig. 4). In contrast, the region of the genome encoding virion core proteins and tail fibre related genes had an AAI identity of <30% between HTVC010P-types and Skadi-1. Previous analyses of long-read sequencing based viromes from the Western English Channel resolved the microdiversity of HTVC010P across niche-defining genomic islands, and revealed a hypervariable region of ~5000 bp in length that contained a ribonuclease and an internal virion protein [54]. Such regions of low coverage occur when reads from a diverse population fail to recruit to the genome of a specific strain and are indicative of genetic population variance [55]. Mapping randomly subsampled (5 million reads per virome) from all environmental GOV2 viromes against HTVC010P-type genomes revealed an average per nucleotide coverage of 4745 X for Skadi-1 per sample. The region between 13,646 and 15,713 bp returned per nucleotide coverage of <100, marking it as a putative hypervariable region (HVR) using previously defined cut-offs (<20% of the median contig coverage and at least 500 kb in length) [56]. Five out of 11 other HTVC010P-like phages recruited at least the minimum number of mapped reads and coverage to putatively identify HVRs, and these genomes showed consistently low recruitment in the same genomic region as Skadi-1, identifying it as a conserved feature within this group. Skadi-1's putative HVR spanned two ORFs encoding an uncharacterised phage protein and a structural lectin-domain containing protein. Lectin-folds in association with phage tail fibre proteins are often involved in host receptor binding and cell attachment [57], supporting the idea that the hypervariable region is a hotspot for arms-race co-evolution and selection within global *P. skadi* populations.
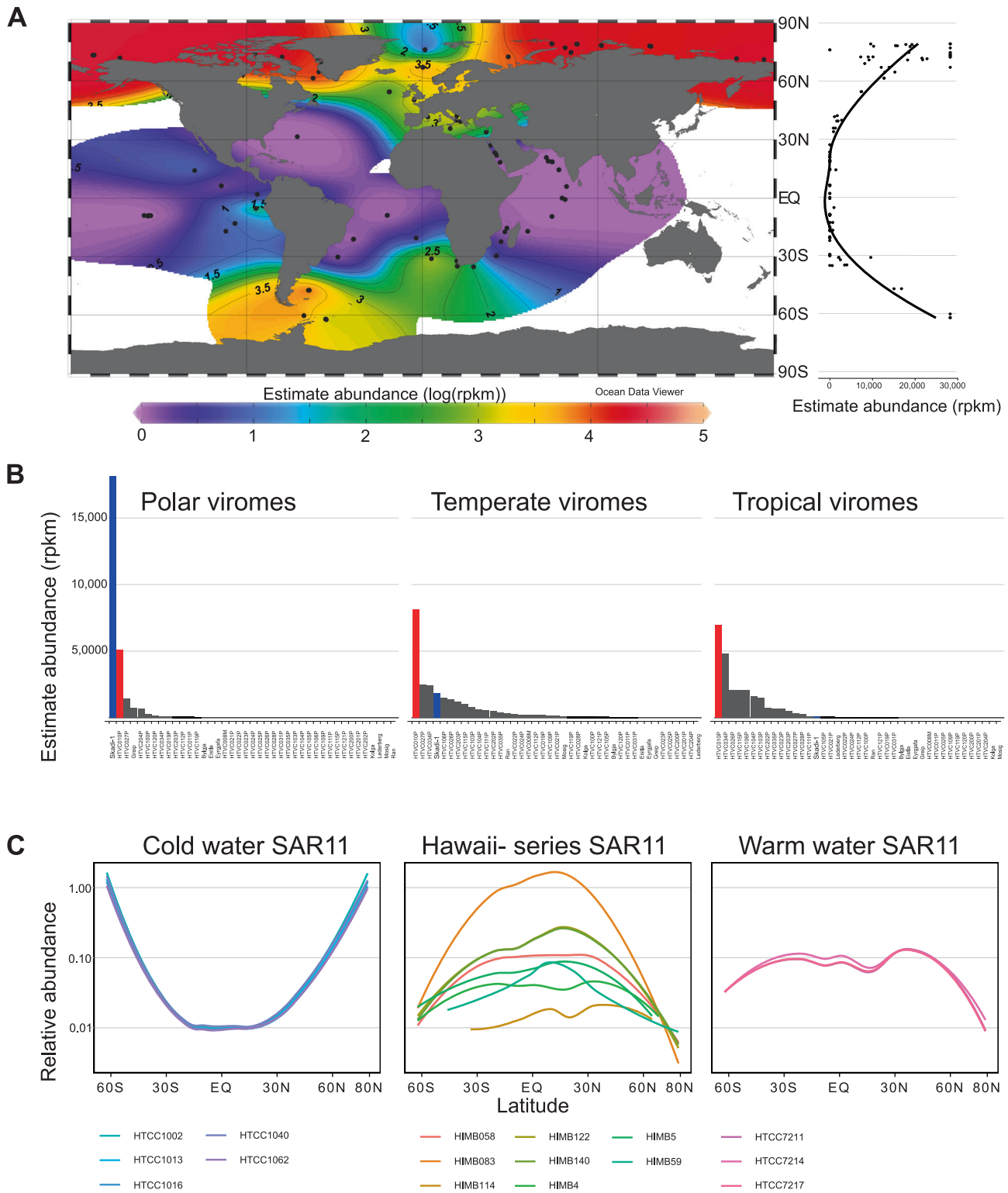
Global distribution patterns reveal *Polarivirus skadi* as a polar surface pelagiphage ecotype

Skadi-1 was observed almost exclusively in high latitude surface waters in both the Arctic and Southern Ocean, despite both bodies of water being separated by the Atlantic and Pacific Ocean, respectively (Fig. 3). Skadi-1 was near or below limits of detection in viromes from lower latitudes (below 66° absolute latitude). Based on the absence of Skadi-1-like populations at lower latitudes, we speculate that this species is unable to replicate

sufficiently to maintain sizable populations in these waters. Therefore, the Atlantic and Pacific Oceans form an evolutionary barrier separating southern and northern populations. As ^{14}C carbon isotope analysis has shown that a single "parcel" of seawater can take up to 1000 years to be transported around the world by global ocean circulations [58], the high abundance of genetically similar *P. skadi* populations in both polar regions (based on read recruitment) suggests that they form conserved populations despite the implied relatively low rates of direct genetic exchange.

To further evaluate differences within *P. skadi*-like genomes, we applied single nucleotide variant (SNV) profiles [21] using 22 GOV2 metagenomes in which Skadi-1 was sufficiently abundant to meet a minimum coverage cutoff of 70%. In total, 10,142 (461 SNV positions across 22 metagenomes) entries were evaluated (91,410 were removed, following recommended cut-offs for coverage and minimum departure constraints [59]). These 461 positions spanned 13 out of 60 genes encoded by Skadi-1. Population structures were evaluated using clustering of SNV profiles identified in each metagenome and revealed two groups (Supplementary Fig. 5): The first group comprised 12 samples obtained from latitudes greater than 60° N (considered here as polar influenced biomes), as well as two samples (sample 82, surface and DCM) obtained from 47° Southern Atlantic off the Argentinian coast ([FKLD] Southwest Atlantic Shelves Province). The second group comprised 10 samples retrieved from tropical and subtropical samples from the South Atlantic, South Pacific, and the Mediterranean, spanning between latitudes 34.95° S and 42.17° N. Metagenomes from different depths but same locations clustered tightly together except one (25_SRF and 25_DCM), suggesting depth and associated intrinsic changes in physio-chemical conditions is not a factor that differentiate *P. skadi* populations, compared to geographical location. Skadi-1 was virtually absent from metagenomes associated with the North Pacific Subtropical Gyre, with only three out of 382 metagenomes (SRR9178452, SRR9178352, SRR9178296 [60]; all from 200 m and 250 m depth) from Station ALOHA (22°45′ N, 158° W) meeting minimum coverage cutoff requirements. This suggests that cold- (deep-) water at lower latitudes does not support large *P. skadi* populations. SNV profiles of all *P. skadi* genomes clustered against Skadi-1 were evaluated along with the metagenomic SNV profiles. This distance-based clustering pattern showed that all genomes of the 10 isolated *P. skadi* strains clustered together when comparing *P. skadi* populations (Supplementary Fig. 6). This may suggest that despite not seeing a clear genetic barrier between Arctic and Southern Ocean, localised clusters of *P. skadi* can occur. Reported mutation rates in dsDNA viruses range from 10^{-4} to 10^{-6} substitutions per site per cell infection [61–63]. Assuming the latent period of *P. skadi* is comparable to the latent periods of 22–24 h reported for HTVC010P [22], and considering Skadi-1's 35,392 bp genome, a rough estimate of 0.04–3.5 substitutions per year would be expected over the full genome, or 35.4 to 3539 substitutions in the 1000 years that it can take to transport water around the world by ocean circulations [58]. Therefore, we propose that the barrier formed by the Atlantic and Pacific Oceans does not create genetically distinct *P. skadi* populations at either polar region, suggesting that ocean mixing occurs faster than mutation rates.

Based on the study of T4-like marine cyanophages, it has been suggested that the abundance of virus populations reflects that of their hosts [15]. HTVC010P was initially isolated from the subtropical Sargasso Sea on cold-water ecotype *Pelagibacter ubique* HTCC1062 [22], which decreases in a north-south gradient to be near-absent at low latitudes, where *Pelagibacter bermudensis* HTCC7211 dominates [64, 65]. We used competitive recruitment of short reads from marine viromes (GOV2) to evaluate relative abundances of all 41 isolate genomes of pelagiphages across global oceans (≥95% nucleotide identity over 90% of read length,

Fig. 3 Global distribution pattern of *Polarivirus skadi*. A Estimated abundance in reads per genome kilobase mapped per million reads (RPKM) of Skadi-1's full-length genome from the GOV2 dataset; **B** Estimated abundance (RPKM) of isolated phages infecting SAR11 (blue: Skadi; green HTVC010P) grouped by absolute latitude into "Polar" (90° to 66°), "Temperate" (66° to 33°) and "Tropical" (33° to 0°) regions; **C** Relative abundance of SAR11 bacterial strains based on single amino acid variants from Delmont et al. [21].

with 40% minimum genome coverage threshold to reduce false positives [32]). HTVC010P did not have any discernible spatial distribution patterns and recruited a broadly uniform number of reads across all latitudes (Fig. 3B). Therefore, we speculate that HTVC010P can infect both warm- and cold-water ecotypes,

possibly with lower efficiencies in cold-water ecotypes, explaining its abundance and uniform distribution across the global oceans, and its replacement by *P. skadi* at higher latitudes. As with HTVC010P, all Skadi strains and Greip were isolated using the SAR11 cold-water ecotype HTCC1062, but when these phages

Table 2. Summary of proposed taxonomy for phages related to up to a family-level with *P. skadi*-type and HTVC010P-type viruses.

Reference phage name	Class	Order	Family	Genus	Species
HTVC010P	Caudoviricetes	Undefined	Ubiqueviridae	Decempevirus	decempe
HTVC024P	Caudoviricetes	Undefined	Ubiqueviridae	Vigintiquatropevirus	vigintiquatrope
HTVC028P	Caudoviricetes	Undefined	Ubiqueviridae	Vigintioctopevirus	vigintioctope
HTVC034P	Caudoviricetes	Undefined	Ubiqueviridae	Trigintaquatropevirus	trigintaquatrope
HTVC035P	Caudoviricetes	Undefined	Ubiqueviridae	Trigintaquinquepevirus	trigintaquinquepe
HTVC100P	Caudoviricetes	Undefined	Ubiqueviridae	Centumpevirus	centumpe
HTVC203P	Caudoviricetes	Undefined	Ubiqueviridae	Ducentitripevirus	ducentitripe
HTVC204P	Caudoviricetes	Undefined	Ubiqueviridae	Ducentiquatropevirus	ducentiquatrope
Lederberg	Caudoviricetes	Undefined	Ubiqueviridae	Diaboliforamvirus	lederbergi
Greip	Caudoviricetes	Undefined	Ubiqueviridae	Iscavirus	greipi
Skadi-1	Caudoviricetes	Undefined	Ubiqueviridae	Polarivirus	skadi

were used to challenge two warm-water (*Pelagibacter spp.* HTCC7211 and H2P3α) host ecotypes they failed to lyse host cells, indicating preferential infection of the cold-water host ecotype. Low abundance of *P. skadi* at temperate and tropical latitudes is most likely explained by a lack of suitable hosts for efficient replication (Fig. 3C), highlighting how phage ecotypes are constrained by environmental selection of their preferred hosts, similar to observations made in cyanobacteria and roseobacteria virus-host systems [66, 67].

Across all GOV2 samples, global median relative abundance of HTVC010P (3100–7713 RPKM) was significantly higher than that of Skadi-1 (0–1673 RPKM) by an average effect size of 5141 RPKM (2699–7530 RPKM). Repeating this analysis with mean RPKM (sensitive to outliers) instead of median RPKM removed this significant difference, as expected given the highly skewed abundance of Skadi-1 within polar regions. The mean difference between Skadi-1 and HTVC034P (the next highest recruiting pelagiphage) was 3902 RPKM (2088–5776 RPKM), indicating that Skadi-1 and HTVC010P are the two most abundant pelagiphages in the oceans by ~2-20-fold. Mean relative abundance of the 39 remaining pelagiphage ranged from 0–2562 RPKM, suggesting that global oceans are dominated by few highly successful pelagiphages and a long tail of less abundant species. This supports the viral seed bank hypothesis for SAR11 virus-host systems, which suggests that marine viral genotypes are ubiquitously distributed albeit relatively rare, but can become abundant when environmental conditions enable efficient replication in available suitable hosts [33]. Most marine phages decay within hours in situ [68, 69], which could explain why Skadi-like phages could not be detected in low latitude metagenomes. Further host range experiments would be required to confirm if Skadi-like phages maintain a seed population at low latitude by infecting local sub-optimal hosts, or if a fraction of the inactive population survives the decay during transport between poles.

Relative abundance of Skadi-1 was significantly negatively correlated to temperature (RPKM~temperature linear regression, $p < 0.001$, $R^2 = 0.69$) (Supplementary Fig. 7). In contrast, relative abundance of HTVC010P and the majority of other pelagiphages (21 out of 28 pelagiphages with non-zero RPKM in >3 viromes) was not significantly correlated to temperature ($p = 0.06$, $R^2 = 0.08$ for HTVC010P, $R^2 < 0.5$, $p \geq 0.05$ for other pelagiphages). HTVC027P was the only phage other than Skadi-1 to have a significant, but weak, negative correlation with temperature ($p < 0.001$, $R^2 = 0.27$). HTVC023P significantly and positively correlated with high temperatures ($p = 0.003$, $R^2 = 0.59$), suggesting an association with warm-water SAR11 ecotypes, though the original isolating host was the cold-water HTCC1062, indicating potential generalism in HTVC023P, similar to that predicted for HTVC010P. This is in line with two long-term 16S rRNA time series conducted at BATS station in the Sargasso Sea and L4 station in the WEC, which found

that SAR11 Ia.3 have a broad temperature range in which it can thrive (from 7–30 °C), whilst the maximum temperature for Ia.1 was ~20 °C [70]. This suggests that phages specialising for SAR11 Ia.3 hosts are less likely to display significant trends with temperature, whereas abundance of phages specialising on cold-water SAR11 Ia.1 would be expected to inversely correlate with temperature. A similar temperature dependent relationship between viruses and environmental parameters has recently been demonstrated for HMO-2011-type phages infecting Roseobacteria [67]. Our results indicate a similar ecotypical distribution is present for pelagiphages, suggesting that this is a common occurrence for viruses of marine heterotrophs.

Given the importance of sinking polar water in driving large-scale global water circulation [71, 72], we evaluated whether *P. skadi* may also be found in lower latitude viromes known to be under the influence of deep-water upwelling. An intriguing exception to the otherwise exclusively polar distribution pattern for *P. skadi* was found in three GOV2 samples (100_DCM, 102_DCM, 102_MES), from which Skadi-1 recruited between 517 and 912 RPKM. These samples were taken in the eastern parts of the Pacific equatorial divergence province (PEOD) and the South Pacific subtropical gyre province (SPSG) in autumn 2011 [73]. An important feature of this region is the Humboldt current system (HCS), which is a cold-water current flowing from Antarctica in parallel to the South American west coast. This causes upwelling of nutrients along the coastline, creating one of the most productive regions in the ocean, but also feeds cold water from the Southern Ocean into the Pacific Ocean [74, 75]. We hypothesise that dominant *P. skadi* populations in the Southern Ocean were transported via the HCS into the Southern Pacific Ocean, where over time *P. skadi* populations are replaced by other actively replicating phages like HTVC010P as the most dominant pelagiphages (4829 to 5191 RPKM). As Skadi-1 was only the 8th to 14th most abundant pelagiphage in the Southern Pacific (compared to the Southern Ocean where Skadi-1 was the most abundant pelagiphage), it is likely that *P. skadi* populations are either outcompeted by phages that are better adapted to local conditions, or that regional SAR11 populations do not provide enough suitable host cells for *P. skadi* to maintain a dominant population. In this scenario, if the population dynamics conform to the Royal-Family hypothesis [16], the Southern Pacific *P. skadi* population is likely in a state of decay and being succeeded by the closely related HTVC010P.

Abundance of *I. greipi*, previously identified as a potential polar ecotypic phage [26], did not significantly correlate to temperature ($p = 0.38$, $R^2 = 0.08$), but was detected in five viromes taken in the Arctic with temperatures of <5 °C, and only two viromes from below 66° north and south. Greip was isolated through enrichment on HTCC1062 in the WEC when ambient surface temperature was 14.8 °C. However, *I. greipi* was below the limit of detection
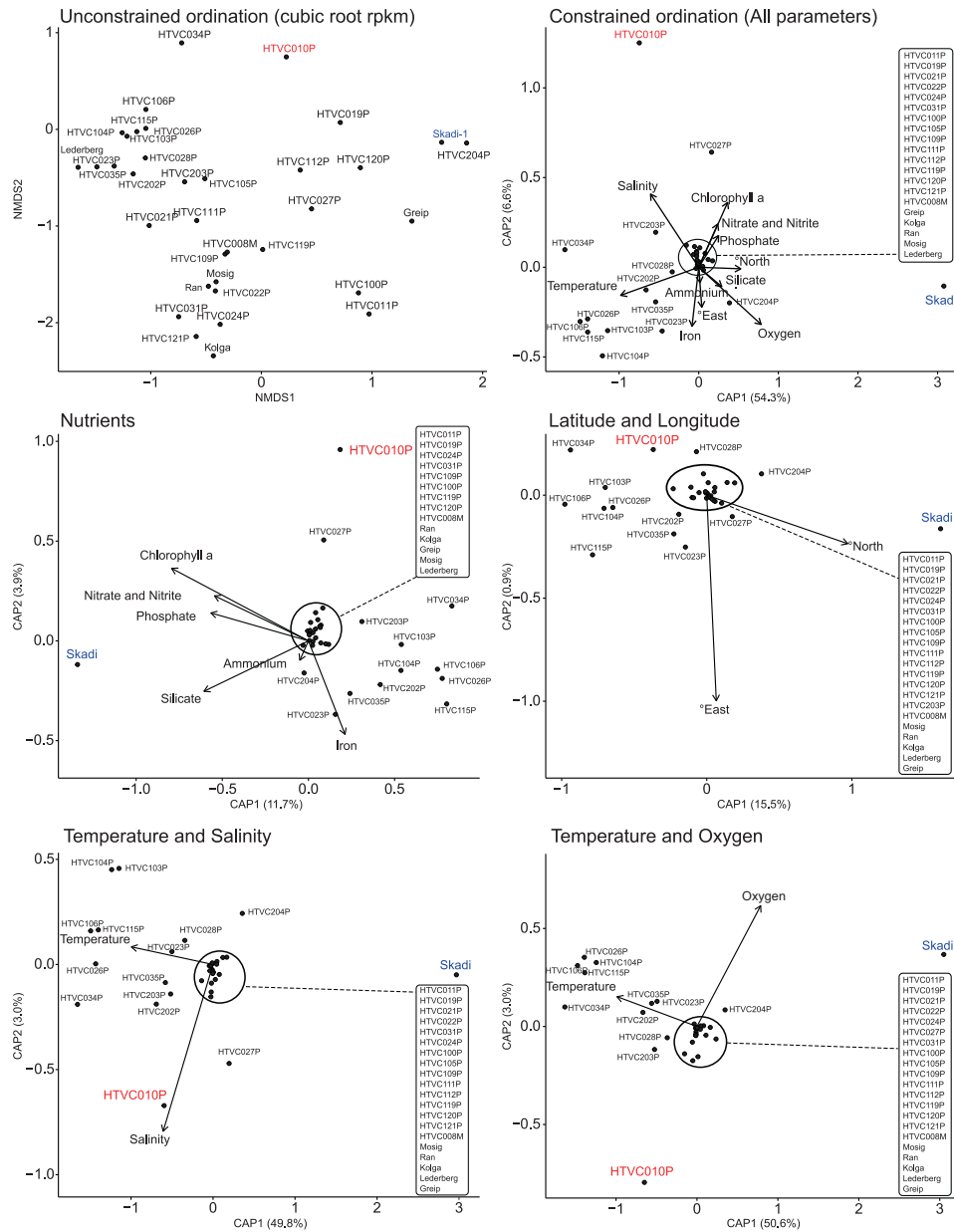
Fig. 4 *Polarivirus skadi* **abundance is driven by typical cold-water parameters.** Plot shows ordinations of RPKM values obtained from GOV2 viromes constrained with different environmental parameters provided by the GOV2 metadata. *P. skadi* abundance is driven by low temperature, high latitude, and high nutrient concentrations, HTVC010P abundance is driven by low oxygen and high salinity, but not temperature.

in a WEC virome [51]. All *P. skadi* strains were also isolated from the WEC, with water temperatures ranging from 14.1 to 15.5 °C, but were present in the WEC virome (5300 RPKM). This could suggest that unlike *I. greipi*, *P. skadi* is able to replicate on local SAR11 populations in this environment, which indicates different optimal hosts for both species, despite both of them being able to replicate on HTCC1062. Successful propagation of *I. greipi* and all *P. skadi* strains in cultures kept at 15 °C further indicates that host availability, rather than temperature itself, limits viral replication in these putative viral polar ecotypes.

Ecotypic pelagiphages are an important aspect of global viral communities
Cold-water environments are often defined as areas with average annual temperatures below 15 °C [76], with polar oceans temperatures between 5 °C and −2 °C, which allows for up to double the

available oxygen concentrations compared to water at 20 °C [77]. The Arctic Ocean also experiences high riverine influx of terrestrial nutrients [78], and subsequently high primary production [79] (for which chlorophyll *a* serves as a proxy). Constrained ordinations provide further evidence that Skadi-1 abundance is correlated with high latitude and/or associated low temperature (Fig. 4), but also suggest that high oxygen, and increased concentration of nutrients and chlorophyll *a* correlate positively with Skadi-1 abundance. Absolute longitude did not impact ordinations for any pelagiphage, indicating that distribution patterns are a function of shared environmental conditions associated with geographical provinces, but not the marine geographical provinces themselves. This is in line with our observations of Skadi-1-like populations at each pole that are putatively connected via the global conveyor belt, as ocean currents would continuously mix global viral communities, maintaining the viral seed-bank. These results therefore support *P. skadi*

as a polar specialist. High temperatures correlated positively with relative abundance of five other (lower abundance) pelagiphages in addition to HTVC023P (HTVC103P, HTVC104P, HTVC115P, HTVC106P, HTVC026P). However, 18 out of 41 pelagiphage genomes did not recruit reads from enough viral metagenome samples (non- zero RPKM in five or more metagenomes) to establish any robust patterns, indicating that these phages are part of the respective seed bank at sampling locations. Considering that the seed bank hypothesis addresses the turnover of dominant phages, the dominance of both HTVC010P and Skadi-1 across space and time, suggests that the aforementioned ecological constraints maintain the host population and subsequently the dominance of a small group of phages – supporting the Royal-Family hypothesis [16]. Deep-sequencing of viral *phoH* genes in the Sargasso Sea also suggested that the majority of operational taxonomic units (OTUs) remained rare throughout the sampling period, with a small number of OTUs dominating throughout the seasons, depths, and years [80]. Similarly, in SAR11 virus-host systems our results could mean that abiotic factors shaping SAR11 ecotypic niche partitioning also indirectly maintain the dominance of a small number of phages out of a large diverse pelagiphage community akin to pelagiphage ecotypes.

CONCLUSION
In this study we isolated and characterised the highly abundant pelagiphage *P. skadi*, defining it as one of the most abundant phages infecting the ubiquitous SAR11 clade. Using strain Skadi-1 as reference, phylogenetic analysis and comparison to other pelagiphages shows that *P. skadi* and HTVC010P-type phages are a monophyletic and distinct taxonomic clade for which we propose the creation of a novel viral family named *Ubiquiviridae*. Skadi-1 and HTVC010P are examples showing that both putative generalist and specialist pelagiphages can be successful and dominant. Our metagenomic analysis of the GOV2 viromes shows that pelagiphage distribution is ubiquitous as thought previously, but the majority of pelagiphage species remain in the low abundance fraction, conforming to the seed-bank model. Skadi-1 abundance is strongly correlated to polar environmental conditions, marking it as a pelagiphage cold-water specialist. The high abundance of Skadi-1 at both poles further suggests that SAR11 ecotypic niche partitioning indirectly shape the pelagiphage community akin to pelagiphage ecotypes, but also that the dominance of these viral species is being maintained across space and time as suggested by the Royal-Family model. Furthermore, considering the virome sampling locations across global oceans, observations of increased abundance for Skadi-1 along the Humboldt Current Systems demonstrates how ocean currents could potentially maintain the viral seed-bank globally. If polar oceans keep increasing in temperature due to escalating climate concerns, a shift in the polar microbial community is likely [81]. By uncovering polar ecotypic distribution of an abundant pelagiphage, our results suggest that polar pelagiphage communities could be altered under predicted warming patterns, with polar specialists lost due to the inability of further poleward niche expansion. SAR11 ecotypes are known to possess specific preferences for different carbon moieties, with different pathways that produce climatically active gasses [19]. Coupled with the importance of viral predation on community productivity and carbon export from the surface ocean [82], it is possible that this loss of viral diversity and shifting of virus-host dynamics may influence biogeochemical cycles.

METHODS SUMMARY
Methods deployed in this study are described in detail in the supplementary materials. Briefly, all phages were isolated on host *Pelagibacter ubique* HTCC1061 using Dilution-to-Extinction methods with surface water samples from the Western English Channel

(50° 15.00 N; 4° 13.00 W) as described previously [26]. Viral particles were precipitated using a modified PEG8000/NaCl DNA isolation method [83] and purified with Wizard DNA Clean-up kits (Promega), following the manufacturer's instructions. Sequencing of phage DNA was performed using 2×250 PE sequencing on a HiSeq 2500 (Illumina). Genomes were annotated and manually curated using an approach developed for the SEA-PHAGES programme [84], where the output from multiple annotation programmes is imported into DNA Master (v5.23.3) and evaluated by using a scoring system to minimise human bias. Phylogenetic analysis was performed using ICTV recommended tools and cut-offs for viral phylogeny [38]. To estimate global relevance of Skadi-like phages, sequencing reads were mapped against the GOV2 database [85] and viromes from the ALOHA station time series (22° 45′ N, 158° W) in a subtropical Pacific ocean gyre [60].

REFERENCES
1. Puxty RJ, Evans DJ, Millard AD, Scanlan DJ. Energy limitation of cyanophage development: implications for marine carbon cycling. ISME J. 2018;12:1273–86.
2. Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Viruses inhibit CO_2 fixation in the most abundant phototrophs on Earth. Curr Biol. 2016;26:1585–9.
3. Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. Science. 2015;347:1257594.
4. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles: implications for marine biogeochemical cycles. Nat Rev Microbiol. 2014;12:519–28.
5. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: how phages command the microbial high seas. Virol J. 2019;16:15.
6. Motegi C, Nagata T, Miki T, Weinbauer MG, Legendre L, Rassoulzadegand F. Viral control of bacterial growth efficiency in marine pelagic environments. Limnol Oceanogr. 2009;54:1901–10.
7. Hurwitz BL, Hallam SJ, Sullivan MB. Metabolic reprogramming by viruses in the sunlit and dark ocean. Genome Biol. 2013;14:R123.
8. Danovaro R, Dell'Anno A, Corinaldesi C, Magagnini M, Noble R, Tamburini C, et al. Major viral impact on the functioning of benthic deep-sea ecosystems. Nature. 2008;454:1084–7.
9. Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, et al. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. Nat Rev Microbiol. 2019;18:21–34.
10. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Reticulate representation of evolutionary and functional relationships between phage genomes. Mol Biol Evol. 2008;25:762–77.
11. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. Nat Rev Microbiol. 2020;18:125–38.
12. Deng L, Ignacio-Espinoza JC, Gregory AC, Poulos BT, Weitz JS, Hugenholtz P, et al. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. Nature. 2014;513:242–5.
13. Marston MF, Martiny JBH. Genomic diversification of marine cyanophages into stable ecotypes. Environ Microbiol. 2016;18:4240–53.
14. Parsons RJ, Breitbart M, Lomas MW, Carlson CA. Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. ISME J. 2012;6:273–84.
15. Chow C-ET, Kim DY, Sachdeva R, Caron DA, Fuhrman JA. Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. ISME J. 2014;8:816–29.
16. Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. Nat Microbiol. 2018;3:754–66.
17. Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, et al. SAR11 clade dominates ocean surface bacterioplankton communities. Nature. 2002;420:806–10.
18. Stingl U, Tripp HJ, Giovannoni SJ. Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. ISME J. 2007;1:361–71.

19. Giovannoni SJ. SAR11 bacteria: the most abundant plankton in the oceans. Ann Rev Mar Sci. 2017;9:231–55.

20. Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, et al. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. mBio. 2012;3:e00252–12.

21. Delmont TO, Kiefl E, Kilinc O, Esen OC, Uysal I, Rappé MS, et al. Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. Elife. 2019;8:e46497.

22. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11 viruses in the ocean. Nature. 2013;494:357–60.

23. Zhang Z, Qin F, Chen F, Chu X, Luo H, Zhang R, et al. Culturing novel and abundant pelagiphages in the ocean. Environ Microbiol. 2020;23:1145–61.

24. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae family integrate into host genomes. Environ Microbiol. 2018;21:1989–2001.

25. Buchholz HH, Michelsen M, Parsons RJ, Bates NR, Temperton B. Draft genome sequences of pelagimyophage Mosig EXVC030M and pelagipodophage Lederberg EXVC029P, isolated from Devil's Hole, Bermuda. Microbiol Resour Announc. 2021;10:e01325–20.

26. Buchholz HH, Michelsen ML, Bolaños LM, Browne E, Allen MJ, Temperton B. Efficient dilution-to-extinction isolation of novel virus–host model systems for fastidious heterotrophic bacteria. ISME J. 2021;15:1585–98.

27. Du S, Qin F, Zhang Z, Tian Z, Yang M, Liu X, et al. Genomic diversity, life strategies and ecology of marine HTVC010P-type pelagiphages. Micro Genom. 2021;7:000596.

28. Carini P, Steindler L, Beszteri S, Giovannoni SJ. Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. ISME J. 2013;7:592–602.

29. Carini P, Campbell EO, Morré J, Sañudo-Wilhelmy SA, Thrash JC, Bennett SE, et al. Discovery of a SAR11 growth requirement for thiamin's pyrimidine precursor and its distribution in the Sargasso Sea. ISME J. 2014;8:1727–38.

30. Martinez-Hernandez F, Fornas O, Lluesma Gomez M, Bolduc B, de la Cruz Peña MJ, Martínez JM, et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. Nat Commun. 2017;8:15892.

31. Martinez-Hernandez F, Fornas Ò, Lluesma Gomez M, Garcia-Heredia I, Maestre-Carballa L, López-Pérez M, et al. Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. ISME J. 2019;13:232–6.

32. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ. 2017;5:e3817.

33. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 2005;13:278–84.

34. Koskella B, Meaden S. Understanding bacteriophage specificity in natural microbial communities. Viruses. 2013;5:806–23.

35. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the marine virosphere using metagenomics. PLoS Genet. 2013;9:e1003987.

36. Zaragoza-Solas A, Rodriguez-Valera F, López-Pérez M. Metagenome mining reveals hidden genomic diversity of pelagimyophages in aquatic environments. mSystems. 2020;5:e00905–19.

37. Sabehi G, Lindell D. The P-SSP7 cyanophage has a linear genome with direct terminal repeats. PLoS ONE. 2012;7:e36710.

38. Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. Viruses. 2021;13:506.

39. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, et al. INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. Phage. 2021;2:214–23.

40. Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. PIRATE: a fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. Gigascience. 2019;8:giz119.

41. Buchholz HH, Bolaños LM, Bell AG, Michelsen ML, Allen MJ, Temperton B, et al. A novel and ubiquitous marine methylophage provides insights into viral-host coevolution and possible host-range expansion in streamlined marine heterotrophic bacteria. Appl Environ Microbiol. 2022;88:e00255–22.

42. Rabsch W, Ma L, Wiley G, Najar FZ, Kaserer W, Schuerch DW, et al. FepA- and TonB-dependent bacteriophage H8: receptor binding and genomic sequence. J Bacteriol. 2007;189:5658–74.

43. Bertozzi Silva J, Storms Z, Sauvageau D. Host receptors for bacteriophage adsorption. FEMS Microbiol Lett. 2016;363:fnw002.

44. Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J. 2010;4:673–85.

45. Chang C-Y, Kemp P, Molineux IJ. Gp15 and gp16 cooperate in translocating bacteriophage T7 DNA into the infected cell. Virology. 2010;398:176–86.

46. Lupo D, Leptihn S, Nagler G, Haase M, Molineux JI, Kuhn A. The T7 ejection nanomachine components gp15-gp16 form a spiral ring complex that binds DNA and a lipid membrane. Virology. 2015;486:263–71.

47. Yuan Y, Gao M. Proteomic analysis of a novel bacillus jumbo phage revealing glycoside hydrolase as structural component. Front Microbiol. 2016;7:745.

48. Hendrix H, Zimmermann-Kogadeeva M, Zimmermann M, Sauer U, De Smet J, Muchez L, et al. Metabolic reprogramming of Pseudomonas aeruginosa by phage-based quorum sensing modulation. Cell Rep. 2022;38:110372.

49. Ceyssens P-J, De Smet J, Wagemans J, Akulenko N, Klimuk E, Hedge S, et al. The phage-encoded N-acetyltransferase Rac mediates inactivation of Pseudomonas aeruginosa transcription by cleavage of the RNA polymerase alpha subunit. Viruses. 2020;12:976.

50. Yang H, Ma Y, Wang Y, Yang H, Shen W, Chen X. Transcription regulation mechanisms of bacteriophages: recent advances and future prospects. Bioengineered. 2014;5:300–4.

51. Delesalle VA, Tanke NT, Vill AC, Krukonis GP. Testing hypotheses for the presence of tRNA genes in mycobacteriophage genomes. Bacteriophage. 2016;6:e1219441.

52. Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D. Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. Appl Environ Microbiol. 2013;79:7547–55.

53. Li H-F, Wang X-F, Tang H. Predicting bacteriophage enzymes and hydrolases by using combined features. Front Bioeng Biotechnol. 2020;8:183.

54. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, et al. Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. PeerJ. 2019;7:e6800.

55. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF, et al. Genomic islands and the ecology and evolution of Prochlorococcus. Science. 2006;311:1768–70.

56. Mizuno CM, Ghai R, Rodriguez-Valera F. Evidence for metaviromic islands in marine phages. Front Microbiol. 2014;5:27.

57. Salazar AJ, Sherekar M, Tsai J, Sacchettini JC. R pyocin tail fiber structure reveals a receptor-binding domain with a lectin fold. PLoS ONE. 2019;14:e0211432.

58. Gebbie G, Huybers P. The mean age of Ocean waters inferred from radiocarbon observations: sensitivity to surface sources and accounting for mixing histories. J Phys Oceanogr. 2012;42:291–305.

59. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.

60. Luo E, Eppley JM, Romano AE, Mende DR, DeLong EF. Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. ISME J. 2020;14:1304–15.

61. Kupczok A, Dagan T. Rates of molecular evolution in a marine Synechococcus phage lineage. Viruses. 2019;11:720.

62. Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010;84:9733–48.

63. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol Biol Evol. 2010;27:2038–51.

64. Brown MV, Lauro FM, DeMaere MZ, Muir L, Wilkins D, Thomas T, et al. Global biogeography of SAR11 marine bacteria. Mol Syst Biol. 2012;8:595.

65. Bolaños LM, Choi CJ, Worden AZ, Baetge N, Carlson CA, Giovannoni S. Seasonality of the microbial community composition in the North Atlantic. Front Mar Sci. 2021;8:624164.

66. Marston MF, Taylor S, Sme N, Parsons RJ, Noyes TJE, Martiny JBH. Marine cyanophages exhibit local and regional biogeography. Environ Microbiol. 2013;15:1452–63.

67. Qin F, Du S, Zhang Z, Ying H, Wu Y, Zhao G, et al. Newly identified HMO-2011-type phages reveal genomic diversity and biogeographic distributions of this marine viral group. ISME J. 2022;16:1363–75.

68. Heldal M, Bratbak G. Production and decay of viruses in aquatic environments. Mar Ecol Prog Ser. 1991;72:205–12.

69. Noble RT, Fuhrman JA. Virus decay and its causes in coastal waters. Appl Environ Microbiol. 1997;63:77–83.

70. Bolaños LM, Tait K, Somerfield PJ, Parsons RJ, Giovannoni SJ, Smyth T, et al. Influence of short and long term processes on SAR11 communities in open ocean and coastal systems. ISME Commun. 2022;2:116.

71. Jacobs SS. Bottom water production and its links with the thermohaline circulation. Antarct Sci. 2004;16:427–37.

72. Cai W. Zonal extent of oceans, high-latitude fresh water supplies and the thermohaline circulation. Quart J R Meteor Soc. 1998;124:811–28.

73. Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, et al. Open science resources for the discovery and analysis of Tara Oceans data. Sci Data. 2015;2:150023.

74. Igarza M, Dittmar T, Graco M, Niggemann J. Dissolved organic matter cycling in the coastal upwelling system off central peru during an 'El Niño' year. Front Mar Sci. 2019;6:198.

75. Montecino V, Lange CB. The Humboldt Current System: ecosystem components and processes, fisheries, and sediment studies. Prog Oceanogr. 2009;83:65–79.

76. Miller RV, Whyte LG, editors. Polar microbiology - life in a deep freeze. Washington, DC: ASM Press; 2012.

77. Still PC, Johnson TA, Theodore CM, Loveridge ST, Crews P. Scrutinizing the scaffolds of marine biosynthetics from different source organisms: Gram-negative cultured bacterial products enter center stage. J Nat Prod. 2014;77:690–702.

78. Terhaar J, Lauerwald R, Regnier P, Gruber N, Bopp L. Around one third of current Arctic Ocean primary production sustained by rivers and coastal erosion. Nat Commun. 2021;12:169.

79. Arrigo KR, van Dijken GL. Continued increases in Arctic Ocean primary production. Prog Oceanogr. 2015;136:60–70.

80. Goldsmith DB, Parsons RJ, Beyene D, Salamon P, Breitbart M. Deep sequencing of the viral *phoH* gene reveals temporal variation, depth-specific composition, and persistent dominance of the same viral *phoH* genes in the Sargasso Sea. PeerJ. 2015;3:e997.

81. Comeau AM, Li WKW, Tremblay J-É, Carmack EC, Lovejoy C. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLoS ONE. 2011;6:e27492.

82. Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, et al. Plankton networks driving carbon export in the oligotrophic ocean. Nature. 2016;532:465–70.

83. Solonenko N. Isolation of DNA from phage lysate. 2016. https://doi.org/10.17504/protocols.io.c36yrd.

84. Salisbury A, Tsourkas PK. A method for improving the accuracy and efficiency of bacteriophage genome annotation. Int J Mol Sci. 2019;20:3391.

85. Gregory AC, Zayed AA, Sunagawa S, Wincker P, Sullivan MB, Temperton B, et al. Marine DNA viral macro- and microdiversity from pole to pole. Cell. 2019;177:1109–23.

86. Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics. 2017;33:3396–404.

AUTHOR CONTRIBUTIONS
HHB led cultivation experiments, nucleic acid extractions, and analyses. HHB, LMB, AGB, and BT provided bioinformatics analyses; MLM led sample collections and assisted in cultivation maintenance. BT designed the study. HHB and BT led manuscript writing; and all authors contributed to and reviewed the text.

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/s41396-023-01466-1.

Correspondence and requests for materials should be addressed to Ben Temperton.

Reprints and permission 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.