Highly diverse flavobacterial phages as mortality factor during North Sea spring blooms

Nina Bartlau[1], Antje Wichels[2], Georg Krohne[3], Evelien M. Adriaenssens[4], Anneke Heins[1], Bernhard M. Fuchs[1], Rudolf Amann[1*], Cristina Moraru[5*]

*corresponding authors

Affiliations

(1) Max Planck Institute for Marine Microbiology, Bremen, Germany
(2) Alfred-Wegner Institute Helmholtz Centre for Polar and Marine Research, Biologische Anstalt Helgoland, Germany
(3) Imaging Core Facility, Biocenter, University of Würzburg, Germany
(4) Quadram Institute Bioscience, Norwich Research Park, Norwich, UK
(5) Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Germany

Addresses of corresponding authors

Cristina Moraru (liliana.cristina.moraru@uni-oldenburg.de)
Institute for Chemistry and Biology of the Marine Environment,
Postbox 2503, Carl-von-Ossietzky –Str. 9 -11, 26111 Oldenburg, Germany
TEL: 0049 441 798 3218

Rudolf Amann (ramann@mpi-bremen.de)
Max Planck Institute for Marine Microbiology,
Celsiusstr. 1, 28359 Bremen, Germany
TEL: 0049 421 2028 930

Key words

Marine flavophages, Helgoland, virus counts, phage isolation, phage genomes

Running Title

North Sea flavophages

Abstract

It is generally recognized that phages have a modulating role in the marine environment. Therefore, we hypothesized that phages can be a mortality factor for the dense heterotrophic bacterial population succeeding in phytoplankton blooms. For the marine carbon cycle, spring phytoplankton blooms are important recurring events. In this study, we focused on *Flavobacteriia*, because they are main responders during these blooms and have an important role in the degradation of polysaccharides. A cultivation-based approach was used, obtaining 44 lytic flavobacterial phages (flavophages), representing twelve new species from two viral realms – *Duplodnaviria* and *Monodnaviria*. Taxonomic analysis allowed us to delineate ten new phage genera and seven new families, from which nine and four, respectively, had no previously cultivated representatives. Genomic analysis predicted various life styles and genomic replication strategies. A likely eukaryote-associated host habitat was reflected in the gene content of some of the flavophages. Detection in cellular metagenomes and by direct-plating indicated that part of these phages were actively replicating in the environment during the 2018 spring bloom. Furthermore, CRISPR/Cas spacers and re-isolation during two consecutive years indicated that, at least part of the new flavophages are stable components of the microbial community in the North Sea. Together, our results indicate that these diverse flavophages have the potential to modulate their respective host populations.

Introduction

Marine bacteriophages outnumber their hosts by one order of magnitude in surface seawater and infect 10-45% of the bacterial cells at any given time (1-3). They have a major impact on bacterioplankton dynamics. This impact can be density dependent (4) and take many forms. By lysing infected cells, viruses decrease the abundance of their host population, shifting the dominant bacterial population, and recycling the intracellular nutrients inside the same trophic level (5). By expressing auxiliary metabolic genes, phages can enhance the metabolic capabilities of the virocells (6, 7). By transferring pieces of host DNA, they can drive bacterial evolution (8). By blocking superinfections with other phages (9), they can protect from immediate lysis. Potentially phages even influence carbon export to the deep ocean due to aggregation of cell debris resulted from cell lysis, called the viral shuttle (10, 11). Marine phages modulate not only their hosts, but also the diversity and function of whole ecosystems. This global impact is reflected in a high phage abundance (12) and diversity (13, 14).

Phages are also well known for modulating bacterial communities in temperate coastal oceans. Here, the increase in temperature and solar radiation in spring induces the formation of phytoplankton blooms (15), which are globally important components of the marine carbon cycle. These ephemeral events release high amounts of organic matter, which fuels subsequent blooms of heterotrophic bacteria. *Flavobacteriia* and *Gammaproteobacteria* are the main responders. Recurrent genera like *Polaribacter*, *Maribacter*, and *Pseudoalteromonas* succeed each other in a highly dynamic fashion. This bacterial succession is likely triggered by the availability and quality of substrates

such as polysaccharides (16, 17), yet it cannot be fully understood without considering mortality factors such as grazing by protists, and viral lysis. Grazing by protists is mostly size dependent (18), whereas viral lysis is highly host-specific (19).

Based on the availability of suitable host bacteria, marine phages can be obtained with standard techniques. Over the years notable numbers of phages infecting marine *Alphapoteobacteria* (e.g. (20, 21)), *Gammproteobacteria* (e.g. (22)), and *Cyanobacteria* (e.g. (23, 24)) have been isolated. Despite the importance of *Flavobacteriia* as primary degraders of high molecular weight algal derived matter only few marine flavobacterial phages, to which we refer in the following as flavophages, have been characterized. This includes several *Cellulophaga* phage isolates from the Baltic Sea, covering all morphological types in *Duplodnaviria* and also two different phage groups in *Monodnaviria* (19, 25). In addition, eight tailed phages, one having a myoviral and the rest a siphoviral morphology, were isolated for members of the genera *Polaribacter* (26), *Flavobacterium (27, 28)*, *Croceibacter* (29) or *Nonlabens* (30). However, still the coverage of the class *Flavobacteriia* and the diversity of marine flavophages remains low. With the exception of the *Cellulophaga* phages most of the other flavophages have only been briefly characterized in genome announcements.

In the context of a large project investigating bacterioplankton successions during North Sea spring bloom season, we isolated and characterized new flavophages, with the purpose of assessing their ecological impact and diversity. In total, more than 100 phage isolates were obtained, sequenced, annotated, and classified. This diverse collection is here presented in the context of virus and bacterioplankton abundances. All newly isolated flavophage genomes were mapped to metagenomes obtained for Helgoland

waters of different size fractions, testing the environmental relevance of the flavophage isolates. This study indicates that flavophages are indeed a mortality factor during spring blooms in temperate coastal seas. Furthermore it provides twelve novel phage-host systems of six genera of *Flavobacteriia*, doubling the number of known hosts.

Material and methods

Sampling campaigns

Surface water samples were taken off the island Helgoland at the long term ecological research station Kabeltonne (54° 11.3’ N, 7° 54.0’ E). The water depth was fluctuating from 7 to 10 m over the tidal cycle. In 2017, a weekly sampling was conducted over five weeks starting on March 14 (Julian days 73-106) and covered the beginning of a spring phytoplankton bloom. In 2018, sampling was conducted over eight weeks starting on March 29 and ending on May 24 (Julian days 88-145). It covered the full phytoplankton bloom. Additional measurements were performed to determine the chlorophyll concentration, total bacterial cell counts and total virus abundances (SI file 1 text). Viruses were counted both by epifluorescence microscopy of Sybr Gold stained samples and by transmission electron microscopy (TEM) of uranyl acetate stained samples (SI file 1 text).

Phage isolation

Phage isolates were obtained either after an intermediate liquid enrichment step or by direct plating on host lawns (SI file 1 Table1). In both cases, seawater serially filtered through 10 µm, 3 µm, 0.2 µm polycarbonate membranes served as phage source. To ensure purity, single phage plaques were transferred three times and then a phage stock was prepared. For more details, see SI file 1 text. These phage stocks were then used for assessing phage morphology, host range and genome size, and for DNA extraction (SI file 1 text).

Determination of phage genomes

Phage DNA was extracted using the Wizard resin kit (Promega, Madison, USA) and eluted in TE buffer (after (31)). The DNA was sequenced on a Illumina HiSeq3000 applying the paired-end 2 x 150 bp read mode. For most *Cellulophaga* phages (except Ingeline) and *Maribacter* phages a ChIPSeq library was prepared. For the other phages a DNA FS library was prepared. The raw reads were quality trimmed and checked, then assembled using SPAdes (v3.13.0, (32)) and Tadpole (v35.14, sourceforge.net/projects/bbmap/). Assembly quality was checked with Bandage (32). The genome ends were determined using PhageTerm (33). For more details about all these procedures, see SI file 1 text.

Retrieval of related phage genomes

Several publicly available datasets of environmental phage contigs (13, 33-38) were queried for sequences related with the flavophages isolated in this study, in a multistep procedure (SI file 1 text). All environmental phage contigs and the reference genomes that formed a cluster in vConTact2 (39) with the flavophage isolates were selected for further analysis. Only environmental contigs with a length >80% compared to their related flavophage isolate were kept.

Phage genome annotation

All phage genomes analysed in this study were annotated using a custom bioinformatics pipeline described elsewhere (20), with modifications (SI file 1 text). The final protein annotations were evaluated manually.

Phage genome taxonomic assignment

The dataset used for phylogenetic analysis comprised of i) all flavophage isolates from this study; ii) related phage genomes (see above); and iii) all *Caudovirales* phage genomes from the Master Species List release 35 (March 2020) (40) from the International Committee of Taxonomy of Viruses (ICTV).

To determine the family level classification of the flavophages we used ViPTree (41) and VICTOR (42) to calculate whole proteome phage trees. First, the command line ViPTree was applied on the full phage genome dataset and different clades were assigned to families, taking as reference the phylogenetic distances for the *Herelleviridae* family (43). Then, a subset of phages comprising all family level clades with the new flavophages was selected for analysis with VICTOR. The parameters for VICTOR were "amino acid" data type and the "d6" intergenomic distance formula. In addition to phylogenetic trees, VICTOR used the following predetermined distance thresholds to suggest taxon boundaries at subfamily (0.888940) and family (0.985225) level (42).

To calculate the sequence relatedness of the ssDNA virus Omtje, the web service of GRAViTy (http://gravity.cvr.gla.ac.uk, (44)) was used with the "Obscuriviridae" and the Baltimore Group II - ssDNA viruses + *Papillomaviridae* and *Polyomaviridae* (VMRv34).

To determine the intra-familial structure and relationships, smaller phage genome datasets corresponding to each family were analysed using i) nucleic acid based intergenomic similarities calculated with VIRIDIC (viridic.icbm.de) and ii) core protein phylogeny. The thresholds used for species and genus definition were 95% and 70% intergenomic similarity, respectively. The core genome analysis was conducted as

follows: i) core genes were calculated with the CoGePh web tool (cogeph.icbm.de); ii) a multiple alignment of all concatenated core proteins was constructed with MUSCLE (v3.8.425,(45)) and iiia) used for the calculation of an approximately maximum-likelihood phylogenetic tree using FastTree v 2.1.11 (46), integrated as plugin in Geneious v 11.1.5 (http://www.geneious.com, (47). In addition, iiib) IQ-Trees with SH-aLRT (48) and ultrafast boostrap values (49) using ModelFinder (50) were calculated. All phylogenetic trees were visualized using FigTree v1.4.4. (51), available at http://tree.bio.ed.ac.uk/software/figtree/).

Host assignment for environmental phage genomes

To determine potential hosts for the environmental phage genomes, a BLASTN (52) search (standard parameters) was performed against the nucleotide collection (nr/nt, taxid:2, bacteria), for all phages and environmental contigs belonging to the newly defined viral families. The hit with the highest bitscore and annotated genes was chosen to indicate the host.

Detection of flavophages and their hosts in Helgoland metagenomes by read mapping

The presence of flavophages, flavobacterial hosts and environmental phages in unassembled metagenomes from the North Sea and their relative abundances were determined by read mapping, using a custom bioinformatics pipeline (20). Two datasets were analyzed: i) cellular metagenomes (0.2 – 3 µm fraction) from the spring 2016 algal bloom (SI file 1 Table 2) and ii) cellular metagenomes (0.2 - 3 µm, 3 – 10 µm and >10 µm fractions) from the spring 2018 algal bloom (SI file 1 Table 2).

A bacterium was considered present when > 60% of the genome was covered by reads with at least 95% identity. A phage was present when > 75% of its genome was covered by reads with at least 90% identity. Relatives of a phage were present when > 60% of its genome was covered by reads with at least 70% identity. The relative abundance of a phage/host genome in a sample was calculated by the following formula: "number of bases at ≥x% identity aligning to the genome / genome size in bases / library size in gigabases (Gb)".

Host 16S rRNA analysis

The genomes of those bacteria which gave phages were sequenced with Sequel I technology (Pacific Biosciences, Menlo Park, USA) (SI file 1 text). After genome assembly the quality was checked and 16S rRNA operons were retrieved using the MiGA online platform (53). Additionally, the 16S rRNA gene from all the other bacterial strains was amplified and sequenced using the Sanger technology (see SI file 1).

For phylogenetic analysis, a neighbour joining tree with Jukes-Cantor correction and a RaxML tree (version 8, (54)) were calculated using ARB (55). The reference data set Ref132 was used, with the termini filter and Capnocytophaga as outgroup (56). Afterwards, a consensus tree was calculated.

CRISPR spacer search

CRISPR spacers and cas systems were identified in the host genomes by CRISPRCasFinder (57). Extracted spacers were mapped with Geneious to the flavophage genomes.

The IMG/VR (36) web service was used to search for spacers targeting the flavophage isolates and the related environmental genomes. A BlastN against the viral spacer database and the metagenome spacer database were run with standard parameters.

Data availability

Sequencing data and phage genomes are available at the National Center for Biotechnology Information (NCBI) with the accession number PRJNA639310. Phages (DSM111231-111236, DSM111238-111241, DSM111252, and DSM111256-111257) and bacterial hosts (DSM111037-111041, DSM111044, DSM111047-111048, DSM111061, and DSM111152) were deposited at the German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany.

Results

Spring phytoplankton blooms were monitored by chlorophyll *a* measurements (Figure 1, SI file 1 Figure 1). In 2018, the bloom had two chlorophyll *a* peaks, and it was more prominent than in 2017. Diatoms and green algae dominated the 2018 blooms (SI file 1 Figure 2). During both blooms, bacterial cell numbers almost tripled, from ~ 6.5×10^5 cells ml^{-1} to ~ 2×10^6 cells ml^{-1}. The *Bacteroidetes* population showed a similar trend, as revealed by 16S rRNA data (Figure 1).

Viral counts

Viral particles were counted at three time points during the 2018 bloom, both by SYBR Gold staining and TEM. Numbers determined by TEM were almost two orders of magnitude lower than those determined by SYBR Gold (Figure 1), a typical phenomenon when comparing these two methods (58). However, both methods showed a strong increase of viral particle numbers over the course of the bloom. All viruses counted by TEM were tailed, a strong indication that they were infecting bacteria or archaea, but not algae (59) (SI file 1 Figure 3). The capsid size ranged between 54 and 61 nm, without any significant differences between the three time points (SI file 1 Figure 4). The virus to bacteria ratio increased throughout the bloom, almost doubling (Figure 1, SI file 1 Table 3).

Flavophage isolation and classification

For phage enrichment, 23 bacterial strains previously isolated from algal blooms in the North Sea were used as bait (SI file 1 Table 1). A total of 108 phage isolates were obtained for 10 of the bacterial strains. These were affiliated with the bacterial genera

Polaribacter, *Cellulophaga*, *Olleya*, *Tenacibaculum*, *Winogradskyella*, and *Maribacter* (Figure 2, SI file 1 Table 4).

Intergenomic similarities at the nucleic acid level allowed the grouping of the 108 flavophages into 44 strains (100% similarity threshold) and 12 species (95% similarity threshold) (SI file Figure 5). The species were tentatively named "Olleya virus Harreka" (Harreka), "Tenacibaculum virus Gundel" (Gundel), "Maribacter virus Molly" (Molly), "Maribacter virus Colly" (Colly), "Winogradskyella virus Peternella" (Peternella), "Polaribacter virus Danklef" (Danklef), "Polaribacter virus Freya" (Freya), "Polaribacter virus Leef" (Leef), "Cellulophaga virus Nekkels" (Nekkels), "Cellulophaga virus Ingeline" (Ingeline), "Cellulophga virus Calle" (Calle), and "Cellulophaga virus Omtje" (Omtje). The names have a Frisian origin, to reflect the flavophage place of isolation. Molly and Colly were isolated only in 2017, Ingeline and Omtje both in 2017 and 2018, and the remaining in 2018 (SI file Figure 1).

Virion morphology as determined by TEM (Figure 3) showed that 11 of the phages were tailed. According to the new megataxonomy of viruses, tailed phages belong to the realm *Duplodnaviria*, kingdom *Heunggongvirae*, phylum *Uroviricota*, class *Caudoviricetes*, order *Caudovirales (60).* The only non-tailed phage was Omtje. Further digestion experiments with different nucleases showed that Omtje has a ssDNA genome (SI file 1, Figure 6). Therefore, this phage belongs to the realm *Monodnaviria*. The non-tailed, icosahedral capsid morphology (Figure 3) places Omtje in the *Sangevirae* kingdom.

Phylogenetic comparison using ViPTree of the 11 tailed flavophages with all *Caudoviricetes* phages in the ICTV35 database showed that they fall into 7 monophyletic

clades of similar rank with the newly defined *Herelleviridae* family (61) (SI file Figure 7). In a further analysis of selected phage genomes using VICTOR, the same monophyletic clades were obtained (Figure 3). All were maximally supported, with pseudo boostrap values of 100, with the exception of the clade containing the Harreka_1 phage, which had a pseudo bootstrap value of 67. To most of these clades VICTOR assigned the rank of subfamily (SI file 1 Figure 7). In the light of the current changes in the ICTV phage classification, this corresponds to the family rank (61). Therefore, taking both the ViPTree and the VICTOR results into consideration, we are proposing that the 7 clades correspond to 7 new families, which we tentatively named as follows: “Flosverisviridae”, “Winoviridae”, “Molycolviridae”, “Pompuviridae”, “Assiduviridae”, “Forsetiviridae” and “Aggregaviridae” (Figure 3).

Four of the new families (“Aggregaviridae”, “Forsetiviridae”, “Molycolviridae” and “Winoviridae”) were formed only from new flavophages and from environmental phage genomes. The remaining three also contained previously cultivated phages, infecting bacteria from the genus *Cellulophaga* (“Pompuviridae” and “Assiduviridae”) and *Flavobacterium* (“Flosverisviridae”).

Using a 70% threshold for the intergenomic similarities at nucleotide level indicated that Harreka, Nekkels, Gundel, Peternella, Leef and Ingeline phages form genera on their own, tentatively named here “Harrekavirus”, “Nekkelsvirus”, “Gundelvirus”, “Peternellavirus”, “Leefvirus” and “Ingelinevirus”. The other new flavophages formed genera together with isolates from this study or with previously isolated flavophages, as follows: the genus “Freyavirus” formed by Danklef and Freya, the genus “Callevirus” formed by Calle, Cellulophaga phage phi38:1, Cellulophaga

phage phi40:1, and the genus “Mollyvirus” formed by Molly and Colly (SI file Figure 9). For each family containing more than 3 phages, a phylogenetic analysis of the core proteins, based on the determined core genes, was performed (SI file 1 Figure 10 – 20). The trees obtained supported the intra-familial structure obtained with VICTOR, and allowed us to define two subfamilies in “Pompuviridae” (“Pachyvirinae” and “Pervagovirinae”) and two in “Flosverisviridae” (“Helgolandvirinae” and “Duenevirinae”), see Figure 3 and SI file 1 Figure 10, 11, 14, 15.

Phylogenetic analysis using both ViPTree and VICTOR of a dataset including the genome of Omtje and all ssDNA phages from the ICTV35 release, and of Omtje using GRAViTy, showed that Omtje and previously isolated *Cellulophaga* phages form a clade separated from the known *Microviridae* family (Figure 4, SI file 1 Figure 21 - 23). We propose here that this clade represents a new family, tentatively called here “Obscuriviridae”. The placement of this family into higher taxonomic ranks remains to be determined in the future. Intergenomic similarity calculations indicate that Omtje forms a genus by itself (SI file 1 Figure 24).

During the phylogenetic analysis we have worked closely with ICTV members, to ensure a good quality of the phage taxonomic affiliations. Two taxonomic proposals for the new defined taxons are being prepared, one for flavophages in *Duplodnaviria* and one for *Monodnaviria*.

Features of the new flavophage isolates

Phages of *Polaribacter*

Three *Polaribacter* phages were obtained: i) Danklef and Freya, part of the family “Forsetiviridae”, infected *Polaribacter* sp. R2A056_3_33 and HaHaR_3_91,

respectively; and ii) Leef, part of the family "Flosverisviridae", infected *Polaribacter* sp. AHE13PA (SI file Figure 1). These phages infected only their isolation host (SI file 1 Figure 18). They all had a siphoviral morphology (Figure 3 and Table 1).

The genome size ranged between ~ 38 and ~ 49 kbp. The GC content was very low (28.9-29.7%). For Leef, PhageTerm predicted genome ends of type Cos 3' (Table 1). Three types of structural proteins were identified in all phages, a major capsid protein, a tail tape measure protein and a portal protein. Several genes for DNA replication, modification and nucleotide metabolism genes were found (Figure 5 and SI file 2). An N-acetylmuramidase was detected in Danklef and Leef, surrounded by transmembrane domains (TMDs) containing proteins. All three phages encoded an integrase. Leef had also a LuxR protein and a pectin lyase.

Phages of *Cellulophaga*

Four Cellulophaga phages were obtained: i) Calle, part of the family "Pompuviridae", infected *Cellulophaga* sp. HaHa_2_95; ii) Nekkels, part of the family "Assiduviridae", infected *Cellulophaga* sp. HaHa_2_1; iii) Ingeline, part of the family "Flosverisviridae", and Omtje, part of the family "Obscuriviridae", infected *Cellulophaga* sp. HaHaR_3_176 (SI file 1 Figure 1). Nekkels and Calle also infected other bacterial strains than their isolation host, although at a low efficiency: *Polaribacter* sp. AHE13PA (Nekkels) and Cellulophaga sp. HaHa_2_1 (Calle) (SI file 1 Figure 25). The virion morphology varied from icosahedral, non-tailed, microvirus-like for Omtje, to tailed, podovirus-like for Calle and siphovirus-like for Nekkels and Ingeline (Figure 3 and Table 1).

The genome size ranged from ~ 6.5 kb (Omtje) to ~ 73 kbp (Calle). The G+C content varied between 31.2 and 38.1% (Table 1). The few structural genes recognized were in agreement with the virion morphology (Figure 5). For example, Ingeline and Nekkels had a tape measure protein, and a portal protein was found in all three tailed phages. From the replication genes, we detected in Omtje a replication initiation protein specific for ssDNA phages, and in Calle a DNA polymerase A (Figure 5).

Potential endolysins found were the mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase in Omtje and the N-aceylmuramoyl-L-alanine-amidase in Calle. The latter had in its vicinity two proteins with a TMD. In Ingeline we annotated a unimolecular spanin. Nekkels encoded a glucoside-hydrolase of the GH19 family, with a peptidoglycan-binding domain. In its vicinity there were three proteins with three to six TMDs and one potential unimolecular spanin (Table 1, SI file 2).

Pectin and chondroitin/alginate lyase domains were found in Ingeline and Nekkels, as part of bigger proteins, also carrying signaling peptides. Nekkels also encoded a Yersinia outer protein X (YopX), a potential virulence factor against eukaryotes. Ingeline encoded an integrase and a LuxR gene. Calle had 20 tRNA and one tmRNA genes (Figure 5, SI file 2).

Phages isolated from other *Flavobacteriaceae*

Five phages were obtained from four other flavobacterial hosts: i) Harreka, part of "Aggregaviridae", infected *Olleya* sp. HaHaR_3_96; ii) Peternella, part of "Winoviridae", infected *Winogradskyella* sp. HaHa_3_26; iii) Gundel, part of "Pompuviridae", infected *Tenacibaculum* sp. AHE14PA and AHE15PA; and iv) Molly

and Colly, part of “Molycolviridae”, infected *Maribacter forsetii* DSM18668 (Figure 3). Harreka infected also *Tenacibaculum* sp. AHE14PA and AHE15PA with a significantly lower infection efficiency (SI file 1 Figure 25). All virions were tailed, with a podoviral morphology for Gundel, and a myoviral morphology for Molly, Peternella and Harreka (Table 1). The genome size ranged from ~ 40 kbp to ~ 125 kbp and the G+C content between 30.4 and 36.2%. Gundel had short direct terminal repeats (DTRs) as genome ends (Table 1).

In these phages we recognized several structural, DNA replication and modification, and nucleotide metabolism genes (Figure 5). With respect to replication, Molly and Colly had a DNA polymerase I gene, plus a helicase and a primase. In Gundel and Harreka we only found a DNA replication protein. In Peternella we found a MuA transposase, several structural genes similar to the Mu phage (62) and hybrid phage/host genome reads, indicating that Peternella is likely a transducing phage.

Potential lysins were a N-acetylmuramoyl-L-alanine-amidase in Molly and Peternella and a L-Alanine-D-glutamine-peptidase in Gundel. Harreka had a glycoside-hydrolase of the GH19 family. In the genomic vicinity of the potential lysins, several proteins having 1-4 TMDs were found, including a holin in Peternella (Table 1, SI file annots).

Additional features of these phages were: i) 10 tRNA genes in Gundel; ii) a relatively short (199 aa) zinc-dependent metallopeptidase, formed from a lipoprotein domain and the peptidase domain in Molly, and iii) a YopX protein in Harreka (Figure 5).

Environmental phage genomes

Five of the 7 proposed new families, five ("Aggregaviridae", "Forsetiviridae", "Pompuviridae", "Winoviridae" and "Flosverisviridae") had members for which the genomes were assembled from environmental metagenomes (Figure 3, SI file 1 Table 5). We have briefly investigated which bacterial groups are potential hosts for these phages. Most of them gave BLASTN hits with a length between 190 and 5845 bases with bacterial genomes from the *Bacteroidetes* phylum (Figure 3SI file 1 Tab 4). Furthermore, two of the marine environmental genomes from "Aggregaviridae" and "Flosverisviridae" had hits with viral spacers from CRISPR arrays in *Bacteroidetes* genomes (SI file 1 Table 5). Additionally, several marine contigs from the "Flosverisviridae" contained Bacteroidetes Associated Carbohydrate-binding Often N-terminal (BACON) domains (SI file Table 6). Together, these results suggest that at least part of the environmental phage contigs in the new families are also infecting members of the phylum *Bacteroidetes.*

Integrase encoding genes were found on all environmental phages from the "Forsetiviridae", and part of the genomes from "Winoviridae" and "Flosverisviridae". LuxR encoding genes were found in many genomes from "Flosverisviridae", and one had also a transporter for the auto-inducer 2 (Figure 3). In the "Winoviridae", all phages encoded Mu-like structural proteins, including the MuD terminase, and two environmental phages also encoded a MuA transposase.

Flavophages in the environment

CRISPR/Cas spacers indicate flavophage presence in the environment

CRISPR/Cas systems were identified in *Polaribacter* sp. HaHaR_3_91 and *Polaribacter* sp. R2A056_3_33 genomes. Spacers from the first strain matched Freya genomes. From the second strains, several spacers matched Danklef genomes, one matched Freya and another Leef (SI file 1 Table 7). This shows that Freya, Danklef and Leef, or their relatives, have infected *Polaribacter* strains in the Helgoland sampling site before 2016, when the host *Polaribacter* strains were isolated. Spacers matching Nekkels, were found in a metagenome of a *Rhodophyta* associated bacterial community (SI file 1 Table 8), showing the presence of Nekkels or its relatives in this habitat.

Read mapping for phages and hosts show presence in North Sea waters

To assess the presence of flavophages in the North Sea, we mined by read mapping cellular metagenomes from the 2016 and 2018 spring blooms. We found five of the new flavophages in the cellular metagenomes from the 2018 spring phytoplankton bloom, at three different time points (Table 2). The complete genomes of Freya, Harreka and Ingeline were covered by reads with 100% identity, signifying that exactly these phage isolates were present in the environment. About 85% from Danklef's genome was covered with reads having 100% identity, indicating that close relatives of this phage (e.g. same species) were present. The genome of Leef was covered only 62% with reads of >70% identity, suggesting that more distant relatives (e.g. genus level) were detected. All phages and their relatives were exclusively found in the >3 µm and >10 µm metagenomes. The most abundant flavophages were Freya and Danklef, reaching 53.8 and 10.3 normalized genome coverage.

Further, we searched for the presence of the five flavophage hosts in the 2018 spring bloom (Tab 2). *Polaribacter* sp. was found in the >10 µm and >0.2 µm fractions,

at different time points during the bloom, including in the two same samples where its phages (Freya, Danklef and Leef) and their relatives were detected. At the beginning of the bloom, in April, there were significantly more phages than hosts. The phage/host genome ratio was 384 for Freya, and 74 for Danklef. One and a half months later, only relatives of Freya, Danklef, and Leef were found, the phage/host ratio being lower than 0.1. Discrimination between the different *Polaribacter* strains was not possible, due to their high similarity (>98% ANI). *Olleya* sp. was found in the > 0.2 µm and > 10 µm fractions, at low abundance and dates preceding the detection of its phage, Harreka. *Cellulophaga* sp., the host for Ingeline, was not found (SI file 1 Table 9-13).

Discussion

We have performed a cultivation-based assessment of the diversity of flavophages potentially modulating *Flavobacteriia*, a key group of heterotrophic bacteria. With our results, we are able to connect phages with their host species and even strain, in contrast to high-throughput viromics-based studies. In addition, these phages were isolated in the ecological context of spring bloom events in two consecutive years. In 2017, we implemented a method of enriching flavophages on six host bacteria. A much larger and more diverse collection of 21 mostly recently isolated *Flavobacteriia* was used in 2018.

Highly diverse flavophages were isolated. Their taxonomic diversity is evident immediately by having representatives in two out of four of the newly delineated viral realms that is in *Duplodnaviria* and *Monodnaviria*. As a point of reference, a viral realm is the equivalent of a domain in the cellular world (60). Furthermore, inside *Duplodnaviria*, the new flavophages fall into 7 new families, four of which have no

previously cultivated representatives. At the genus level, out of ten new genera, only "Callevirus" includes previously cultivated phages. The newly delineated families are relevant not only for the marine environment. Besides cultivated flavophages, five of the families also include environmental phages from marine, freshwater, wastewater and soil samples.

Genomic analysis indicates that the new flavophages have various life styles. The presence of integrases in Danklef, Freya, Peternella, Leef and Ingeline, as well as in many of the other phages in their families, points toward the potential to perform a temperate life cycle. However, the absence of recognizable gene markers for lysogeny in the genomes of Harreka, Nekkels, Gundel, Calle, Molly, Colly suggests that most likely these phages are strictly lytic. With respect to genome replication, several strategies can be recognized. First, the genome ending in DTRs of Gundel and the Cos-3' ends of Leef are indicators for replication through long concatemers (63). Second, the MuA-transposase of Peternella, as well as of other environmental phages in this family, indicate a replicative transposition strategy (64). And third, Omtje, a ssDNA phage, most likely replicates through a ssDNA rolling circle mechanism (65). At the translation level, Calle and Gundel likely increase the protein synthesis efficiency by encoding a high number of tRNAs and even a tmRNA (Table 1), similar to other cultivated phages of the "Pompuviridae" family.

With regard to lysis, flavophages encode various peptidoglycan degrading proteins potentially functioning as endolysins. These genes are usually surrounded by genes coding for proteins with several TMDs, likely holins and anti-holins, with the exception of the endo-beta-N-acetylglucosaminidase of Omtje. Harreka and Nekkels do

not encode easily recognizable lysis enzymes. Instead, they encode each a GH19. Usually, this hydrolase family is known for chitin degradation yet also peptidoglycan may be degraded (66). A phage GH19 expressed in *Escherichia coli* caused cellular lysis (67). Furthermore, in Harreka and Nekkels, the vicinity with potential holins, antiholins and spanin, and the peptidoglycan-binding domain in Nekkels, suggest that the GH19 proteins of these two phages likely function as endolysins and degrade bacterial peptidoglycan. It cannot be excluded though that these enzymes have a dual function. Once released extracellularly due to lysis, they could degrade chitin, an abundant polysacharide in the marine environment, produced for example by green-algae or copepods (68). The lysis mechanism in the new dsDNA flavophages likely follows the canonical holin/endolysins model, as suggested by the lack of membrane binding domains in the potential endolysins.

The presence of lyases in the genomes of Leef (pectin lyase), Ingeline (pectin lyase) and Nekkels (pectin and alginate lyases) suggests that their bacterial hosts, *Polaribacter* sp. and *Cellulophaga* sp., are surrounded by polysaccharides. In the marine environment, both alginate and pectin, are produced in large quantities by micro and macro algae (69-71), which serve as substrate for marine *Flavobacteriia*, especially *Polaribacter* (72-74). Bacteria are not only able to degrade these polysaccharides, but also to produce them and form capsules or an extracellular matrix of biofilms. In phages, such polysaccharide degrading enzymes are usually located on the tails, as part of proteins with multiple domains to reach the bacterial cell. Another enzyme potentially degrading proteins in the extracellular matrix is the zinc-dependent metallopeptidase of Molly, which infects a *Maribacter* strain. By depolymerizing the extracellular matrix

surrounding the cells, lyases and peptidases help the phage to reach the bacterial membranes for infection or allow the new progeny to escape the cell debris and the extracellular matrix (75, 76). It remains to be proven if phages carrying these enzymes contribute significantly to the degradation of algal excreted polysaccharides, as a byproduct of their quest to infect new bacterial cells.

Previous studies indicate that flavobacteriia can have a surface-associated life style (77). Our results paint a similar picture. For example, we detected phages for *Polaribacter*, *Cellulophaga* and *Olleya*, as well as the *Polaribacter* and *Olleya* genomes themselves in the particulate fraction of the cellular metagenomes. Therefore, it is likely that these bacteria are associated with particles, protists or zooplankton. Viral spacers matching Nekkels, a *Cellulophaga* phage, with English Channel metagenomes suggest an association with red macro-algae (SI file 1 Table 8). An association with eukaryotes is also supported by the presence of YopX proteins in Harreka, infecting *Olleya*, and in all three "Assiduviridae" phages infecting *Cellulophaga*. Scanning electron microscopy images of *Cellulophaga* sp. HaHaR_3_176 cells show high amounts of extracellular material (SI file 1 Figure 26), indicating the ability for an attached life style.

In the "Flosverisviridae" family, Leef and Ingeline, as well as all other cultivated flavophages and most of the environmental genomes, encode an integrase gene, which is a marker for a temperate life style, and a LuxR gene, which is a quorum sensing dependent transcriptional activator (Figure 3 and SI file 1 Table 6). This suggests that "Flosverisviridae" phages potentially respond to changes in the hosts cell densities, for example by entering the lytic cycle, as recently observed in *Vibrio cholera* (pro)-phages (78, 79). This type of quorum-sensing response would make sense in a habitat in which

the host cells can achieve high densities, as for example in association with eukaryotes or other particles.

All of our flavophages, with the exception of Molly and Colly, were actively produced at different times throughout the 2018 phytoplankton bloom, as shown by their isolation by enrichment or direct-plating. Several of these flavophages were also found with 100% read identity covering the whole genome in the cellular metagenomes from the 2018 bloom, indicating not only an intracellular presence, but also higher abundances. Additionally, the genus "Callevirus" is related to the 5th most abundant cluster in the Global Ocean Virome (14). Calle and Nekkels, infecting *Cellulophaga*, although not detected in the cellular metagenomes, were in high abundance (at least 100 PFU ml^{-1}) in the viral fraction of our samples, as revealed by their isolation by direct-plating without previous enrichment. This apparent discrepancy between the presence in either the viral or cellular fraction, could be explained either by the phage lysis preceding the metagenome collection, or by their host habitat not being sampled for metagenomics, because *Cellulophaga* are known to grow on macro-algae (80). The presence of Omtje, Peternella and Gundel in enrichments, but not in direct-plating or in the cellular metagenomes, indicate their presence in the environment in low abundances. Further investigations of phage and host habitat are necessary to confirm these findings.

From the five flavophages detected in the 2018 spring bloom metagenomes, Freya, Danklef, Leef and Ingeline have a temperate potential, and Harreka is most likely strictly lytic. The ratio between phage and host normalized read abundance can be used to determine their lytic or temperate state in environmental samples. *Cellulophaga*, the host of Ingeline, was not detected throughout the spring bloom. However, because Ingeline

was detected presumably in the intracellular fraction, we can hypothesize that either its host was in a low abundance, or its genome was degraded under the phage influence. Either way, it points toward Ingeline being in a lytic cycle at the time of detection. For Freya, Danklef and their relatives, the high phage to host genome ratios (as high as 454 x) from the 12th of April suggest that these phages were actively replicating and lysing their cells. In contrast, when they reappeared on the 22nd of May, these phages were not only three orders of magnitude less abundant than in April, but also ~ 10 times less abundant then their host. This indicates that in May, only a small proportion of the *Polaribacter* cells were infected with relatives of Freya and Danklef, either in a lytic or a temperate state.

A temporal dynamics in the phage populations is also visible from the isolation results for Freya and Gundel and it is reminiscent of the boom and bust behavior previously observed with T4-like marine phages (81) . Furthermore, the read mapping results indicate a change in the dominant phage genotypes, from the same species and even the same strain with Danklef and Freya in April, towards more distant relatives at the level of genus or even family in May. This suggests a selection of resistant *Polaribacter* strains after the April infection and differs from the genotypic changes previously observed for marine phages, which were at the species level (82). Persistence over longer times in the environment was be shown for *Polaribacter* phages by the CRISPR/Cas spacers, and for Omtje and Ingeline, by their isolation both in 2017 and 2018.

The North Sea flavophages isolated in this study have a high taxonomic, genomic and life style diversity. They are a stable and active part of the microbial community in

Helgoland waters. With their influence on the dynamics of their host populations by lysis or on other bacterial population by substrate facilitation, they are modulating the carbon cycling in coastal shelf seas. The increase in bacterial numbers, mirrored by the phage numbers and the phage to bacterial cell ratio, indicate that phages actively replicate through the 2018 phytoplankton bloom, matching previous observations of the North Sea microbial community (83). Our read mapping data indicate a complex dynamics, which can now be further investigated with the novel phage-host systems obtained in this study.

Acknowledgements

The authors would like to thank Carlota Alejandre-Colomo, and Jens Harder for providing several bacterial isolates. For sampling and 16S rRNA analysis we would like to thank Lilly Franzmeier and Mirja Meiners. Many thanks to the Biologische Anstalt Helgoland (BAH), the Aade Crew for providing the infrastructure for the sampling campaign, and Karen Wiltshire for providing the chlorophyll *a* and algae fluorescence data. The authors acknowledge the help in the lab of Jan Brüwer, Jörg Wulf and Sabine Kühn, the funding of the German Research Foundation (DFG) project FOR2406 – "Proteogenomic of Marine Polysaccaride Utilisation (POMPU)" project and the Max Planck Society. Part of the bioinformatics analyses were performed on the High Performance Computing Cluster CARL, located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Programme (INST 184/157-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State. EMA was funded by the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1.

Conflict of Interest

E.M.A. is the current Chair of the Bacterial and Archaeal Viruses Subcommittee of the International Committee on Taxonomy of Viruses, and member of its Executive Committee. The other authors declare no conflict of interest.

Supplementary information is available at ISME Journal's website.

References

1. Proctor LM, Fuhrman JA. Viral mortality of marine bacteria and cyanobacteria. Nature. 1990;343:60.
2. Steward G, Wikner J, Cochlan W, Smith D, Azam F. Estimation of Virus Production in the Sea: II. Field Results. Marine Microbial Food Webs. 1992;6:79-90.
3. Suttle CA. The significance of viruses to mortality in aquatic microbial communities. Microbial Ecology. 1994;28(2):237-43.
4. Thingstad TF. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnology and Oceanography. 2000;45(6):1320-8.
5. Wilhelm SW, Suttle CA. Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. BioScience. 1999;49(10):781-8.
6. Breitbart MYA, Thompson LR, Suttle CA, Sullivan MB. Exploring the Vast Diversity of Marine Viruses. Oceanography. 2007;20(2):135-9.
7. Coutinho FH, Silveira CB, Gregoracci GB, Thompson CC, Edwards RA, Brussaard CPD, et al. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. Nature Communications. 2017;8(1):15955.

8. Touchon M, Moura de Sousa JA, Rocha EPC. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Current Opinion in Microbiology. 2017;38:66-73.

9. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, et al. Lytic to temperate switching of viral communities. Nature. 2016;531(7595):466-70.

10. Yamada Y, Tomaru Y, Fukuda H, Nagata T. Aggregate Formation During the Viral Lysis of a Marine Diatom. Frontiers in Marine Science. 2018;5(167).

11. Nissimov JI, Vandzura R, Johns CT, Natale F, Haramaty L, Bidle KD. Dynamics of transparent exopolymer particle production and aggregation during viral infection of the coccolithophore, Emiliania huxleyi. Environmental Microbiology. 2018;20(8):2880-97.

12. Bergh Ø, BØrsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. Nature. 1989;340(6233):467-8.

13. Gregory AC, Zayed AA, Conceição-Neto N, Temperton B, Bolduc B, Alberti A, et al. Marine DNA Viral Macro- and Microdiversity from Pole to Pole. Cell. 2019;177(5):1109-23.e14.

14. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature. 2016;537(7622):689-93.

15. Gerdts G, Wichels A, Döpke H, Klings K-W, Gunkel W, Schütt C. 40-year long-term study of microbial parameters near Helgoland (German Bight, North Sea): historical view and future perspectives. Helgoland Marine Research. 2004;58(4):230-42.

16. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, et al. Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. Science. 2012;336(6081):608-11.

17. Teeling H, Fuchs BM, Bennke CM, Kruger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. eLife. 2016;5.

18. Gonzalez JM, Sherr EB, Sherr BF. Size-selective grazing on bacteria by natural assemblages of estuarine flagellates and ciliates. Applied and Environmental Microbiology. 1990;56(3):583-9.

19. Holmfeldt K, Middelboe M, Nybroe O, Riemann L. Large Variabilities in Host Strain Susceptibility and Phage Host Range Govern Interactions between Lytic Marine Phages and Their Flavobacterium Hosts. Applied and Environmental Microbiology. 2007;73(21):6730-9.

20. Bischoff V, Bunk B, Meier-Kolthoff JP, Spröer C, Poehlein A, Dogs M, et al. Cobaviruses – a new globally distributed phage group infecting Rhodobacteraceae in marine ecosystems. The ISME Journal. 2019;13(6):1404-21.

21. Chan JZM, Millard AD, Mann NH, Schäfer H. Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like Roseophage specific genes. Frontiers in Microbiology. 2014;5:506-.

22. Wichels A, Biel SS, Gelderblom HR, Brinkhoff T, Muyzer G, Schütt C. Bacteriophage Diversity in the North Sea. Applied and Environmental Microbiology. 1998;64(11):4128-33.

23. Sabehi G, Shaulov L, Silver DH, Yanai I, Harel A, Lindell D. A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. Proceedings of the National Academy of Sciences. 2012;109(6):2037-42.

24. Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, et al. Genomic and structural analysis of Syn9, a cyanophage infecting marine Prochlorococcus and Synechococcus. Environmental Microbiology. 2007;9(7):1675-95.

25. Holmfeldt K, Odić D, Sullivan MB, Middelboe M, Riemann L. Cultivated Single-Stranded DNA Phages That Infect Marine Bacteroidetes Prove Difficult To Detect with DNA-Binding Stains. Applied and Environmental Microbiology. 2012;78(3):892-4.

26. Kang I, Jang H, Cho J-C. Complete genome sequences of bacteriophages P12002L and P12002S, two lytic phages that infect a marine Polaribacter strain. Standards in Genomic Sciences. 2015;10(1):82.

27. Borriss M, Helmke E, Hanschke R, Schweder T. Isolation and characterization of marine psychrophilic phage-host systems from Arctic sea ice. Extremophiles. 2003;7(5):377-84.

28. Jiang SC, Kellogg CA, Paul JH. Characterization of Marine Temperate Phage-Host Systems Isolated from Mamala Bay, Oahu, Hawaii. Applied and Environmental Microbiology. 1998;64(2):535-42.

29. Kang I, Kang D, Cho J-C. Complete Genome Sequence of *Croceibacter* Bacteriophage P2559S. Journal of Virology. 2012;86(16):8912-3.

30. Kang I, Jang H, Cho J-C. Complete Genome Sequences of Two *Persicivirga* Bacteriophages, P12024S and P12024L. Journal of Virology. 2012;86(16):8907-8.

31. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environmental Microbiology. 2010;12(11):3035-56.

32. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. Epub 2012/04/16.

33. Mizuno CM, Ghai R, Saghaï A, López-García P, Rodriguez-Valera F. Genomes of Abundant and Widespread Viruses from the Deep Ocean. mBio. 2016;7(4):e00805-16.

34. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the Marine Virosphere Using Metagenomics. PLOS Genetics. 2013;9(12):e1003987.

35. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, et al. Environmental Viral Genomes Shed New Light on Virus-Host Interactions in the Ocean. mSphere. 2017;2(2):e00359-16.

36. Paez-Espino D, Roux S, Chen IMA, Palaniappan K, Ratner A, Chu K, et al. IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. Nucleic Acids Research. 2019;47(D1):D678-D86.

37. Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, et al. Single-cell genomics-based analysis of virus–host interactions in marine surface bacterioplankton. The ISME Journal. 2015;9(11):2386-99.

38. 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. Nature Communications. 2017;8(1):15892.

39. Jang HB, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nature Biotechnology. 2019;37(6):632-9.

40. Adriaenssens EM, Sullivan MB, Knezevic P, van Zyl LJ, Sarkar BL, Dutilh BE, et al. Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Archives of Virology. 2020;165(5):1253-60.

41. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017;33(15):2379-80.

42. Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics. 2017;33(21):3396-404.

43. Barylski J, Kropinski AM, Alikhan N-F, Adriaenssens EM, Consortium IR. ICTV Virus Taxonomy Profile: Herelleviridae. Journal of General Virology. 2020;101(4):362-3.

44. Aiewsakun P, Adriaenssens EM, Lavigne R, Kropinski AM, Simmonds P. Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. Journal of General Virology. 2018;99(9):1331-43.

45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004;32(5):1792-7.

46. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010;5(3):e9490.

47. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the

organization and analysis of sequence data. Bioinformatics (Oxford, England). 2012;28(12):1647-9. Epub 2012/04/27.

48. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution. 2014;32(1):268-74.

49. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution. 2017;35(2):518-22.

50. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods. 2017;14(6):587-9.

51. Rambaut A. 1.4.4—a graphical viewer of phylogenetic trees and a program for producing publication-ready figures. 2018.

52. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.

53. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, et al. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Research. 2018;46(W1):W282-W8.

54. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.

55. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. Nucleic Acids Research. 2004;32(4):1363-71.

56. Peplies J, Kottmann R, Ludwig W, Glöckner FO. A standard operating procedure for phylogenetic inference (SOPPI) using (rRNA) marker genes. Systematic and Applied Microbiology. 2008;31(4):251-7.

57. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Research. 2018;46(W1):W246-W51.

58. Forterre P, Soler N, Krupovic M, Marguet E, Ackermann H-W. Fake virus particles generated by fluorescence microscopy. Trends in Microbiology. 2013;21(1):1-5.

59. Nagasaki K. Dinoflagellates, diatoms, and their viruses. The Journal of Microbiology. 2008;46(3):235-43.

60. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, et al. Global Organization and Proposed Megataxonomy of the Virus World. Microbiology and Molecular Biology Reviews. 2020;84(2):e00061-19.

61. Barylski J, Enault F, Dutilh BE, Schuller MB, Edwards RA, Gillis A, et al. Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of Tailed Phages. Systematic Biology. 2019;69(1):110-23.

62. Taylor AL. Bacteriophage-induced mutation in *Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America. 1963;50(6):1043-51.

63. Casjens SR, Gilcrease EB. Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions. In: Clokie MRJ, Kropinski AM, editors. Bacteriophages: Methods and Protocols, Volume 2 Molecular and Applied Aspects. Totowa, NJ: Humana Press; 2009. p. 91-111.

64. Montaño SP, Pigli YZ, Rice PA. The Mu transpososome structure sheds light on DDE recombinase evolution. Nature. 2012;491(7424):413-7.

65. Krupovic M. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Current Opinion in Virology. 2013;3(5):578-86.

66. Wohlkönig A, Huet J, Looze Y, Wintjens R. Structural Relationships in the Lysozyme Superfamily: Significant Evidence for Glycoside Hydrolase Signature Motifs. PLoS ONE. 2010;5(11):e15388.

67. Dziewit L, Oscik K, Bartosik D, Radlinska M. Molecular Characterization of a Novel Temperate Sinorhizobium Bacteriophage, ΦLM21, Encoding DNA Methyltransferase with CcrM-Like Specificity. Journal of Virology. 2014;88(22):13111-24.

68. Souza CP, Almeida BC, Colwell RR, Rivera ING. The Importance of Chitin in the Marine Environment. Marine Biotechnology. 2011;13(5):823.

69. Desikachary TV, Dweltz NE. The chemical composition of the diatom frustule. Proceedings of the Indian Academy of Sciences - Section B. 1961;53(4):157-65.

70. Khotimchenko Y, Khozhaenko E, Kovalev V, Khotimchenko M. Cerium binding activity of pectins isolated from the seagrasses Zostera marina and Phyllospadix iwatensis. Mar Drugs. 2012;10(4):834-48. Epub 2012/04/05.

71. Hay ID, Ur Rehman Z, Moradali MF, Wang Y, Rehm BHA. Microbial alginate production, modification and its applications. Microb Biotechnol. 2013;6(6):637-50. Epub 2013/08/19.

72. Krüger K, Chafee M, Ben Francis T, Glavina del Rio T, Becher D, Schweder T, et al. In marine Bacteroidetes the bulk of glycan degradation during algae blooms is

mediated by few clades using a restricted set of genes. The ISME Journal. 2019;13(11):2800-16.

73. Avcı B, Krüger K, Fuchs BM, Teeling H, Amann RI. Polysaccharide niche partitioning of distinct Polaribacter clades during North Sea spring algal blooms. The ISME Journal. 2020.

74. Kappelmann L, Krüger K, Hehemann J-H, Harder J, Markert S, Unfried F, et al. Polysaccharide utilization loci of North Sea Flavobacteriia as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. The ISME Journal. 2019;13(1):76-91.

75. Latka A, Maciejewska B, Majkowska-Skrobek G, Briers Y, Drulis-Kawa Z. Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. Applied Microbiology and Biotechnology. 2017;101(8):3103-19.

76. Pires DP, Oliveira H, Melo LDR, Sillankorva S, Azeredo J. Bacteriophage-encoded depolymerases: their diversity and biotechnological applications. Applied Microbiology and Biotechnology. 2016;100(5):2141-51.

77. Bižić-Ionescu M, Zeder M, Ionescu D, Orlić S, Fuchs BM, Grossart H-P, et al. Comparison of bacterial communities on limnic versus coastal marine particles reveals profound differences in colonization. Environmental Microbiology. 2015;17(10):3500-14.

78. Silpe JE, Bassler BL. Phage-Encoded LuxR-Type Receptors Responsive to Host-Produced Bacterial Quorum-Sensing Autoinducers. mBio. 2019;10(2):e00638-19.

79. Silpe JE, Bassler BL. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. Cell. 2019;176(1):268-80.e13.

80. Bowman JP. The Marine Clade of the Family Flavobacteriaceae: The Genera Aequorivita, Arenibacter, Cellulophaga, Croceibacter, Formosa, Gelidibacter, Gillisia, Maribacter, Mesonia, Muricauda, Polaribacter, Psychroflexus, Psychroserpens, Robiginitalea, Salegentibacter, Tenacibaculum, Ulvibacter, Vitellibacter and Zobellia. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes: Volume 7: Proteobacteria: Delta, Epsilon Subclass. New York, NY: Springer New York; 2006. p. 677-94.

81. Needham DM, Chow C-ET, Cram JA, Sachdeva R, Parada A, Fuhrman JA. Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. The ISME Journal. 2013;7(7):1274-85. Epub 2013/02/28.

82. Ignacio-Espinoza JC, Ahlgren NA, Fuhrman JA. Long-term stability and Red Queen-like strain dynamics in marine viruses. Nature Microbiology. 2020;5(2):265-71.

83. Wiltshire KH, Kraberg A, Bartsch I, Boersma M, Franke H-D, Freund J, et al. Helgoland Roads, North Sea: 45 Years of Change. Estuaries and Coasts. 2010;33(2):295-310.

Figures and Tables

Figure 1: Flavophage detection during the 2018 spring phytoplankton bloom, as inferred from phage isolation and metagenome read mapping.

Upper panel: Results are presented in the context of chlorophyll a concentration (green), total cell counts (black line), Bacteroidetes counts (orange line), phage counts by transmission electron microscopy (TEM, black bar), and phage counts by epifluorescence light microscopy (LM, grey bar).

Lower panel: Phage isolation is shown for different time points (x axis, Julian days), with the phage identity verified by sequencing (black dots) or not determined (gray dots). Most of the isolations were done by enrichment, and just a few by direct plating (asterisks). Phage detection in metagenomes was performed by read mapping, with a 90% read identity threshold for phages in the same species (full red circles) and 70% read identity threshold for related phages (dashed red circles). Alternating gray shading of isolates indicates phages belonging to the same family.

Figure 2: Phylogenetic tree (consensus between raxml and the NJ) of the 16S rRNA gene from all bacterial strains used to enrich for phages in 2017 and 2018 (red), plus reference genomes. Red squares indicate successful phage isolation.

Figure 3: Whole genome phylogeny determined with VICTOR (amino-acid based) for the dsDNA flavophages, including environmental sequences and reference phage genomes. Our new phage isolates are depicted in red. Pseudo-bootstrap values are

indicated at branches. Column 6 shows the TEM image of the negative stained new flavophage (scale bar in each TEM image 50 nm) and column 7 gives the new family name.

Figure 4: Whole genome phylogeny determined with VICTOR (amino-acid based) for the ssDNA phages. Our new phage isolate is depicted in red. Pseudo-bootstrap values are indicated at branches. Column 6 shows the TEM image of the negative stained new flavophage (scale bar in each TEM image 50 nm) and column 7 gives the new family name.

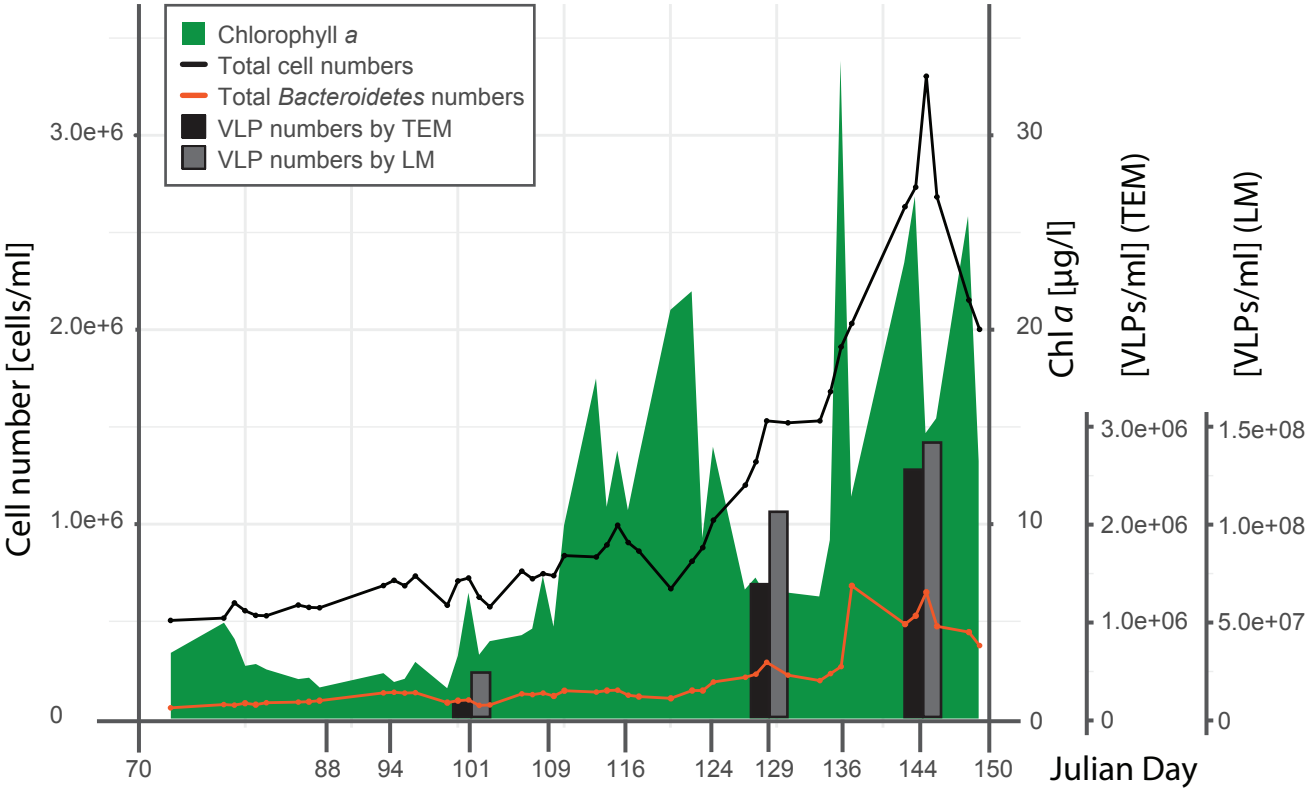
Figure 5: Genome maps of phage isolates with color-coded gene annotations.

Table 1: Phage characteristics of each phage group. As Colly has the same characteristics as Molly, it is not especially mentionied in the table. Lysis genes are stated as N-acetylmuramidase (M), N-aceylmuramoyl-L-alanine-amidase (A), Glycosid Hydrolase 19(GH), L-Alanine-D-glutamine-peptidase (P), spanin (S), holin(H).

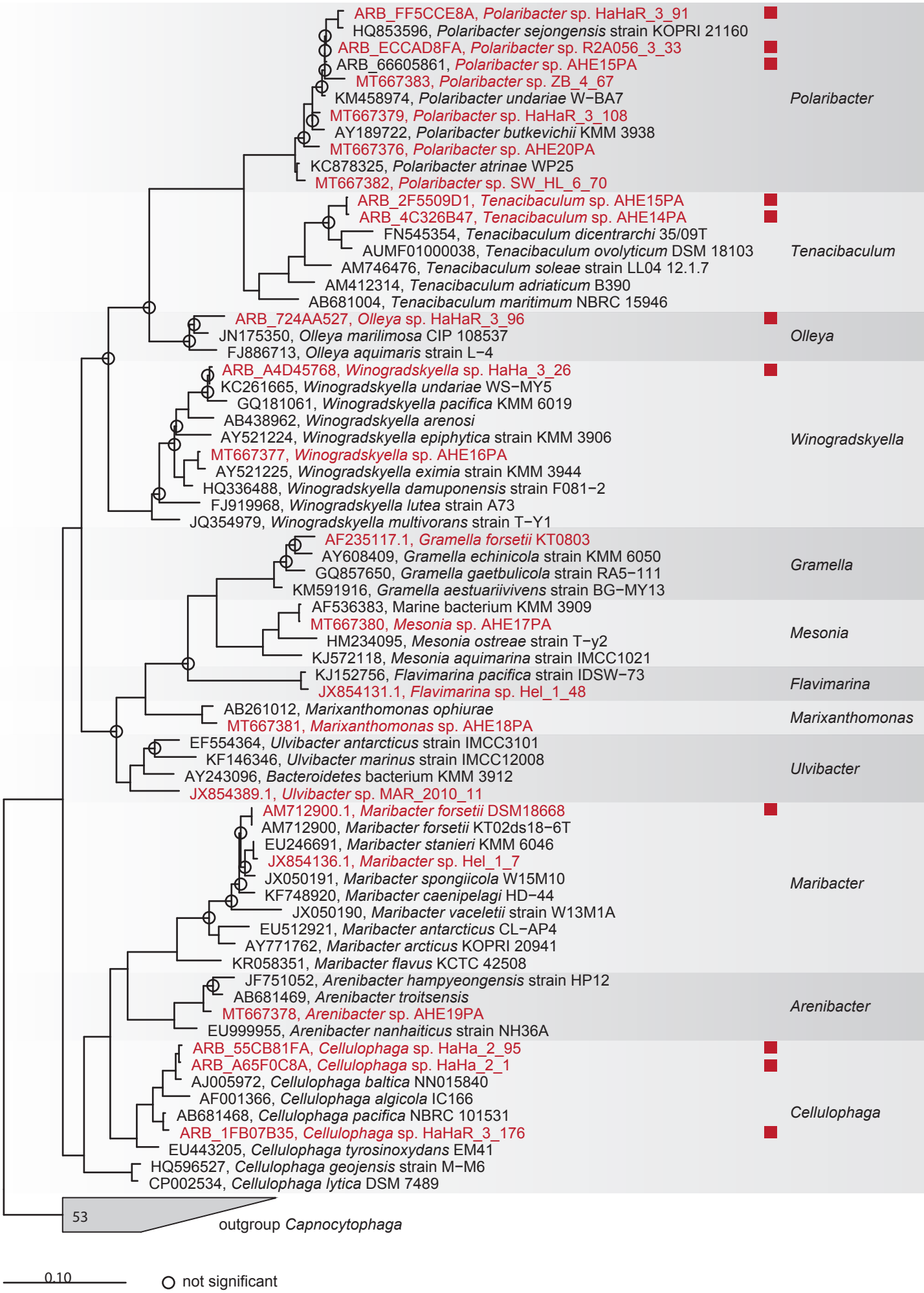
Phage genus	host genus	genome size [kbp]	%GC	ORFs	# tRNAs	# tmRNAs	# integrases	lysis genes	genome ends	circularly closed	capsid diameter [nm]	tail length [nm]	tail width [nm]	morphology	# strains
Molly	*Maribacter*	124.2	36.2	193-201	0	0	0	A	nd	yes	74.9±3.6	101.5±6.3	18.1±1.6	myoviridal	7
Gundel	*Tenacibaculum*	78.5	30.4	118	10	0	0	P	short DTRs	yes	60.5±5.2	22.7±3.2	-	podoviridal	1
Calle	*Cellulophaga*	73.0	38.1	85-86	20	1	0	A	nd	yes	60.3±3.0	23.0±5.5	13.5±2.4	podoviridal	3
Nekkels	*Cellulophaga*	53.3-54.3	31.5	93-94	0	0	0	GH, S	nd	yes	57.8±5.0	141.0±7.9	13.0±1.5	siphoviridal	2
Omtje	*Cellulophaga*	6.6	31.2	13	0	0	0	A	nd	yes	52.3±4.6	-	-	non tailed	5
Ingeline	*Cellulophaga*	42.6	32.2	49-50	0	0	2	S	nd	yes	59.0± 5.3	132.9±19.3	11.2±1.7	siphoviridal	8
Leef	*Polaribacter*	37.5	29.7	48	0	0	2	M	Cos 3'	yes	49.2±3.6	138.7±9.6	11.1±2.0	siphoviridal	1
Danklef	*Polaribacter*	47.2-48.2	28.9	76-78	1	0	1	M	nd	yes	46.1±2.2	157.4±4.6	12.1±1.8	siphoviridal	5
Freya	*Polaribacter*	44.0-48.9	28.9	66-78	0	0	1	0	nd	yes	53.9±4.7	151.0±8.2	13.1±2.0	siphoviridal	10
Peternella	*Winogradskyella*	39.6	35.3	62	0	0	1	A, H x	Mu-like	no	52.3±4.2	105.8±7.4	16.4±2.4	myoviridal	1
Harreka	*Olleya*	43.2	32.0	80	0	0	0	GH	nd	yes	44.3±3.6	123.8±8.0	14.3±2.0	myoviridal	1

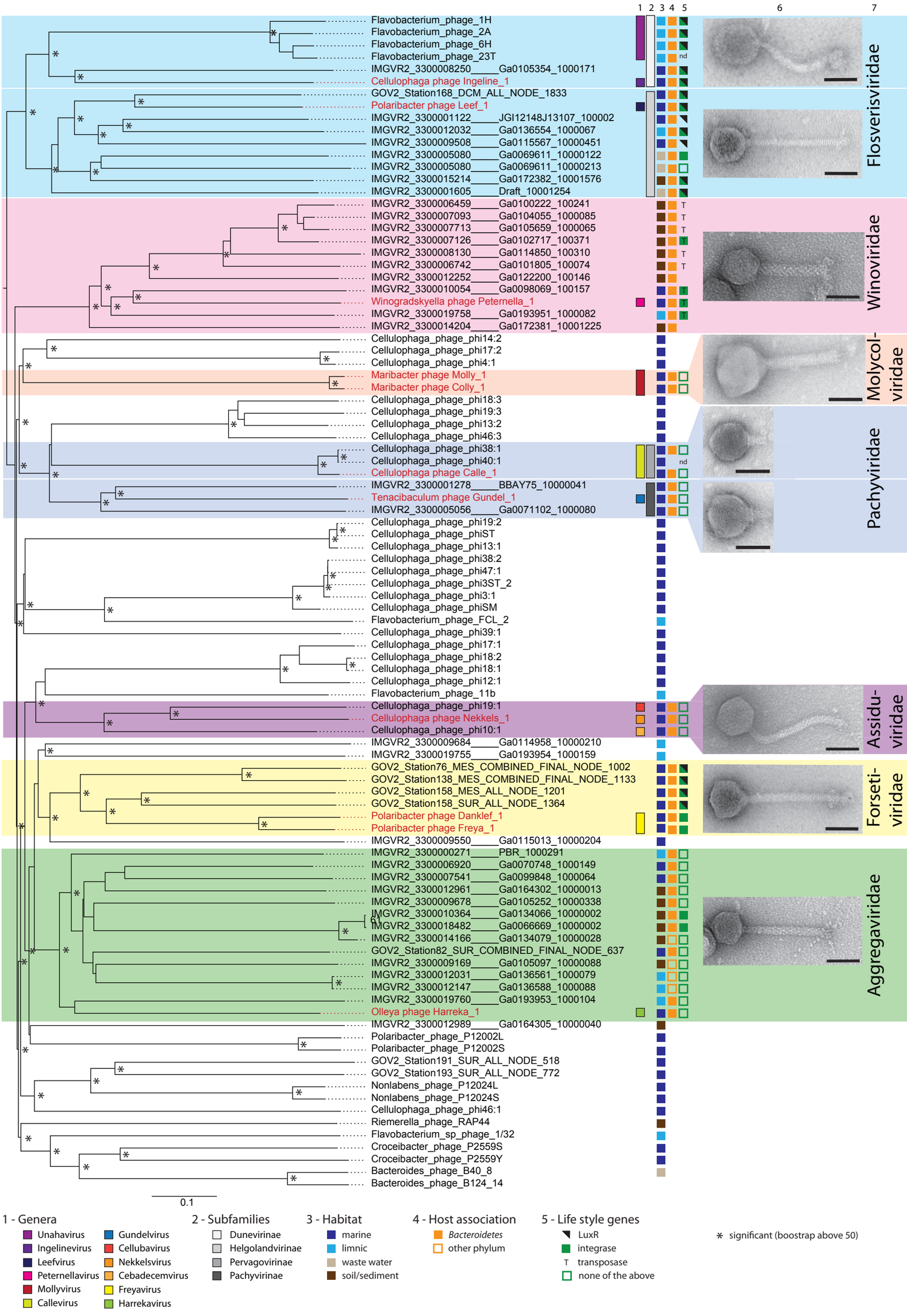
Table 2: Read mapping results from 2018 metagenomes for isolated flavophages and their hosts.

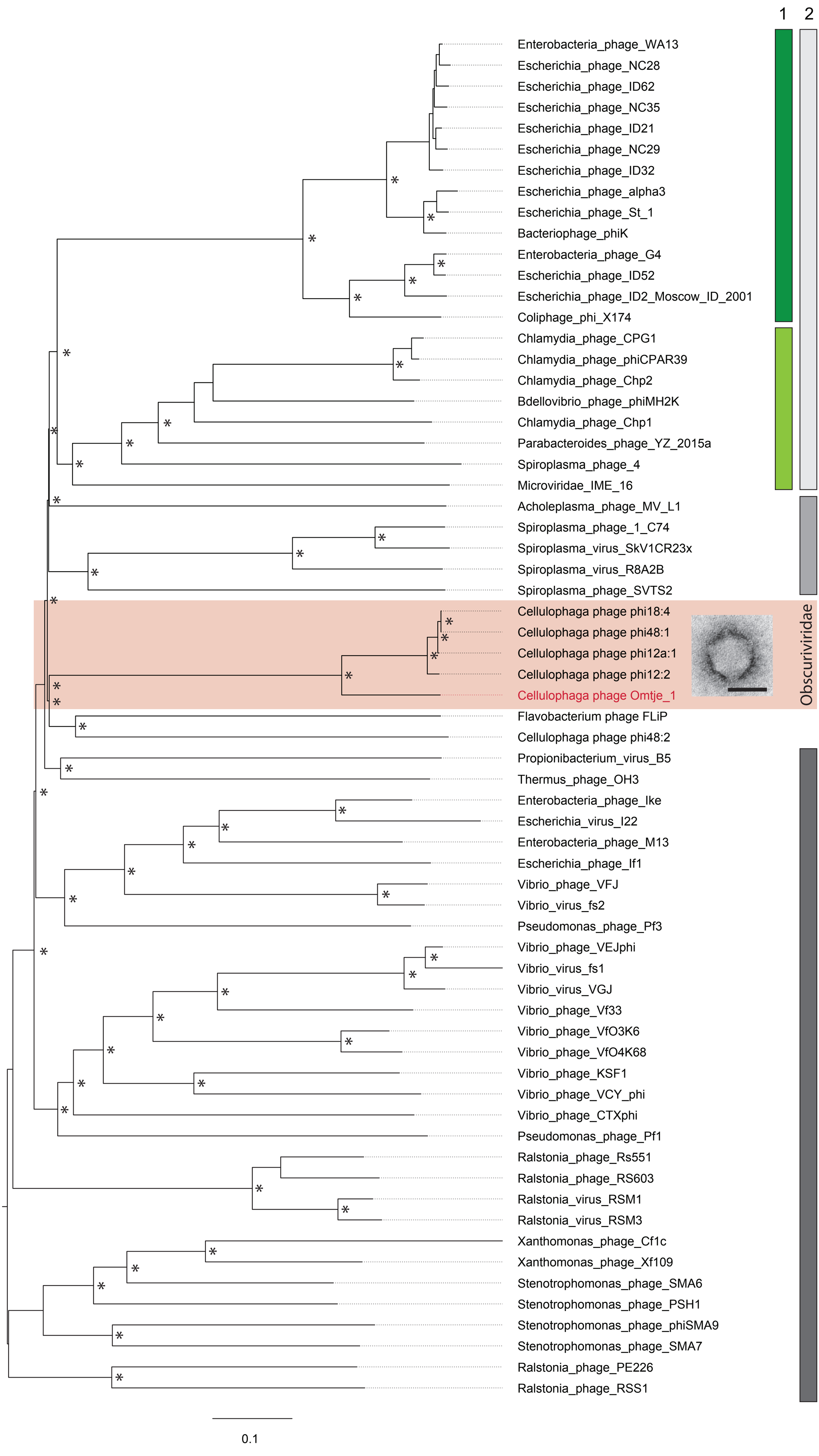
Julian day /Greorian date	**Fraction**	**Phage**		**Host**		**Phage/Host ****
		Name	Norm. genome coverage	Name	Norm. genome coverage	
102 / 12.04.2018	10 µm	Freya*	53.8	*Polaribacter* sp. (HaHaR_3_91 and R2A056_3_33)	0.14	384
		Freya relatives	63.3			454
		Danklef	10.4			74
		Danklef relatives	45.8			327
		Ingeline*	0.7			
		Ingeline relatives	0.8			
				Olleya sp. HaHaR_3_96	0.05	
116 / 26.04.2018	10 µm			*Polaribacter* sp. HaHaR_3_91 and R2A056_3_33	0.06	
				Olleya sp. HaHaR_3_96	0.03	
122 / 02.05.2018	0.2 µm			*Polaribacter* sp. HaHaR_3_91 and	0.07	
				Olleya sp. HaHaR_3_96	0.05	
128 / 08.05.2018	3 µm	Harreka*	1.8			
		Harreka relatives	5.7			
142 / 22.05.2018	10 µm	Freya relatives	0.04	*Polaribacter* sp. (HaHaR_3_91, R2A056_3_33 and AHE53.11)	0.46	0.09
		Danklef relatives	0.03			0.07
		Leef relatives	0.02			0.04

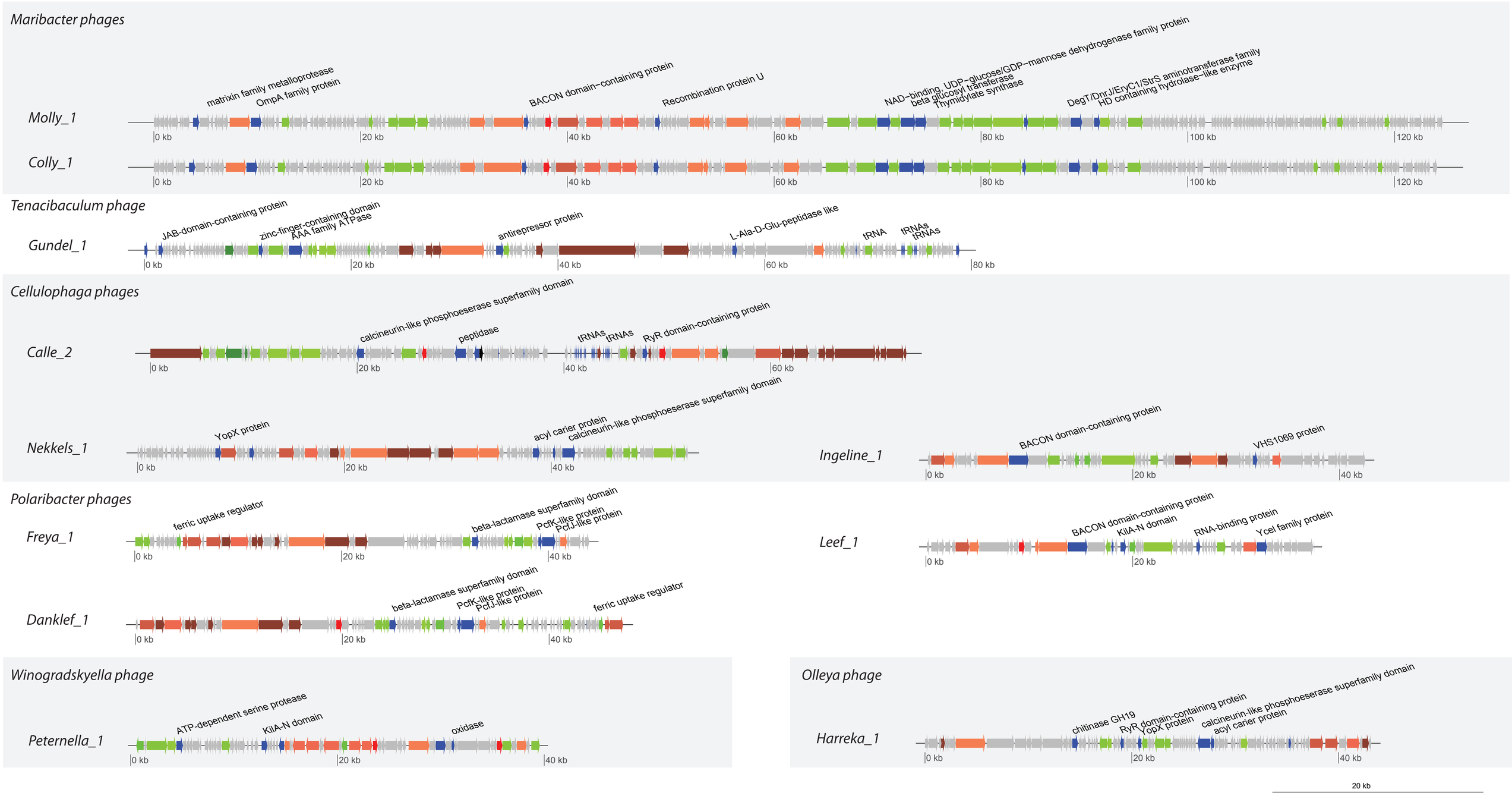


Gundel		○	○	●	○	○	○	●	○	*Tenacibaculum*
Calle		●	●	◐	●	●*				*Cellulophaga*
Nekkels		●	○	◐	◐	●*				*Cellulophaga*
Omtje	○					●	●	○	○	*Cellulophaga*
Ingeline	○		○			●	●	○	○	*Cellulophaga*
Leef								○	◉	*Polaribacter*
Danklef			○					●	◉	*Polaribacter*
Freya		●	●	○	○	●	○	○	◌	*Polaribacter*
Peternella		○	○	○	○	○	○	●	○	*Winogradskyella*
Harreka							○	●	●	*Olleya*


ARB_FF5CCE8A, Polaribacter sp. HaHaR_3_91
HQ853596, Polaribacter sejongensis strain KOPRI 21160
ARB_ECCAD8FA, Polaribacter sp. R2A056_3_33
ARB_66605861, Polaribacter sp. AHE15PA
MT667383, Polaribacter sp. ZB_4_67
KM458974, Polaribacter undariae W−BA7
MT667379, Polaribacter sp. HaHaR_3_108
AY189722, Polaribacter butkevichii KMM 3938
MT667376, Polaribacter sp. AHE20PA
KC878325, Polaribacter atrinae WP25
MT667382, Polaribacter sp. SW_HL_6_70
Polaribacter
ARB_2F5509D1, Tenacibaculum sp. AHE15PA
ARB_4C326B47, Tenacibaculum sp. AHE14PA
FN545354, Tenacibaculum dicentrarchi 35/09T
AUMF01000038, Tenacibaculum ovolyticum DSM 18103
AM746476, Tenacibaculum soleae strain LL04 12.1.7
AM412314, Tenacibaculum adriaticum B390
AB681004, Tenacibaculum maritimum NBRC 15946
Tenacibaculum
ARB_724AA527, Olleya sp. HaHaR_3_96
JN175350, Olleya marilimosa CIP 108537
FJ886713, Olleya aquimaris strain L−4
Olleya
ARB_A4D45768, Winogradskyella sp. HaHa_3_26
KC261665, Winogradskyella undariae WS−MY5
GQ181061, Winogradskyella pacifica KMM 6019
AB438962, Winogradskyella arenosi
AY521224, Winogradskyella epiphytica strain KMM 3906
MT667377, Winogradskyella sp. AHE16PA
AY521225, Winogradskyella eximia strain KMM 3944
HQ336488, Winogradskyella damuponensis strain F081−2
FJ919968, Winogradskyella lutea strain A73
JQ354979, Winogradskyella multivorans strain T−Y1
Winogradskyella
AF235117.1, Gramella forsetii KT0803
AY608409, Gramella echinicola strain KMM 6050
GQ857650, Gramella gaetbulicola strain RA5−111
KM591916, Gramella aestuariivivens strain BG−MY13
Gramella
AF536383, Marine bacterium KMM 3909
MT667380, Mesonia sp. AHE17PA
HM234095, Mesonia ostreae strain T−y2
KJ572118, Mesonia aquimarina strain IMCC1021
Mesonia
KJ152756, Flavimarina pacifica strain IDSW−73
JX854131.1, Flavimarina sp. Hel_1_48
Flavimarina
AB261012, Marixanthomonas ophiurae
MT667381, Marixanthomonas sp. AHE18PA
Marixanthomonas
EF554364, Ulvibacter antarcticus strain IMCC3101
KF146346, Ulvibacter marinus strain IMCC12008
AY243096, Bacteroidetes bacterium KMM 3912
JX854389.1, Ulvibacter sp. MAR_2010_11
Ulvibacter
AM712900.1, Maribacter forsetii DSM18668
AM712900, Maribacter forsetii KT02ds18−6T
EU246691, Maribacter stanieri KMM 6046
JX854136.1, Maribacter sp. Hel_1_7
JX050191, Maribacter spongiicola W15M10
KF748920, Maribacter caenipelagi HD−44
JX050190, Maribacter vaceletii strain W13M1A
EU512921, Maribacter antarcticus CL−AP4
AY771762, Maribacter arcticus KOPRI 20941
KR058351, Maribacter flavus KCTC 42508
Maribacter
JF751052, Arenibacter hampyeongensis strain HP12
AB681469, Arenibacter troitsensis
MT667378, Arenibacter sp. AHE19PA
EU999955, Arenibacter nanhaiticus strain NH36A
Arenibacter
ARB_55CB81FA, Cellulophaga sp. HaHa_2_95
ARB_A65F0C8A, Cellulophaga sp. HaHa_2_1
AJ005972, Cellulophaga baltica NN015840
AF001366, Cellulophaga algicola IC166
AB681468, Cellulophaga pacifica NBRC 101531
ARB_1FB07B35, Cellulophaga sp. HaHaR_3_176
EU443205, Cellulophaga tyrosinoxydans EM41
HQ596527, Cellulophaga geojensis strain M−M6
CP002534, Cellulophaga lytica DSM 7489
Cellulophaga
53
outgroup Capnocytophaga
0.10
not significant


1 2 3 4 5 6 7
Flavobacterium_phage_1H
Flavobacterium_phage_2A
Flavobacterium_phage_6H
Flavobacterium_phage_23T
IMGVR2_3300008250_____Ga0105354_1000171
Cellulophaga phage Ingeline_1
GOV2_Station168_DCM_ALL_NODE_1833
Polaribacter phage Leef_1
IMGVR2_3300001122_____JGI12148J13107_100002
IMGVR2_3300012032_____Ga0136554_1000067
IMGVR2_3300009508_____Ga0115567_10000451
IMGVR2_3300005080_____Ga0069611_10000122
IMGVR2_3300005080_____Ga0069611_10000213
IMGVR2_3300015214_____Ga0172382_10001576
IMGVR2_3300001605_____Draft_10001254
IMGVR2_3300006459_____Ga0100222_100241
IMGVR2_3300007093_____Ga0104055_1000085
IMGVR2_3300007713_____Ga0105659_1000065
IMGVR2_3300007126_____Ga0102717_100371
IMGVR2_3300008130_____Ga0114850_100310
IMGVR2_3300006742_____Ga0101805_100074
IMGVR2_3300012252_____Ga0122200_100146
IMGVR2_3300010054_____Ga0098069_100157
Winogradskyella phage Peternella_1
IMGVR2_3300019758_____Ga0193951_1000082
IMGVR2_3300014204_____Ga0172381_10001225
Cellulophaga_phage_phi14:2
Cellulophaga_phage_phi17:2
Cellulophaga_phage_phi4:1
Maribacter phage Molly_1
Maribacter phage Colly_1
Cellulophaga_phage_phi18:3
Cellulophaga_phage_phi19:3
Cellulophaga_phage_phi13:2
Cellulophaga_phage_phi46:3
Cellulophaga_phage_phi38:1
Cellulophaga_phage_phi40:1
Cellulophaga phage Calle_1
IMGVR2_3300001278_____BBAY75_10000041
Tenacibaculum phage Gundel_1
IMGVR2_3300005056_____Ga0071102_1000080
Cellulophaga_phage_phi19:2
Cellulophaga_phage_phiST
Cellulophaga_phage_phi13:1
Cellulophaga_phage_phi38:2
Cellulophaga_phage_phi47:1
Cellulophaga_phage_phi3ST_2
Cellulophaga_phage_phi3:1
Cellulophaga_phage_phiSM
Flavobacterium_phage_FCL_2
Cellulophaga_phage_phi39:1
Cellulophaga_phage_phi17:1
Cellulophaga_phage_phi18:2
Cellulophaga_phage_phi18:1
Cellulophaga_phage_phi12:1
Flavobacterium_phage_11b
Cellulophaga_phage_phi19:1
Cellulophaga phage Nekkels_1
Cellulophaga_phage_phi10:1
IMGVR2_3300009684_____Ga0114958_10000210
IMGVR2_3300019755_____Ga0193954_1000159
GOV2_Station76_MES_COMBINED_FINAL_NODE_1002
GOV2_Station138_MES_COMBINED_FINAL_NODE_1133
GOV2_Station158_MES_ALL_NODE_1201
GOV2_Station158_SUR_ALL_NODE_1364
Polaribacter phage Danklef_1
Polaribacter phage Freya_1
IMGVR2_3300009550_____Ga0115013_10000204
IMGVR2_3300000271_____PBR_1000291
IMGVR2_3300006920_____Ga0070748_1000149
IMGVR2_3300007541_____Ga0099848_1000064
IMGVR2_3300012961_____Ga0164302_10000013
IMGVR2_3300009678_____Ga0105252_10000338
IMGVR2_3300010364_____Ga0134066_10000002
IMGVR2_3300018482_____Ga0066669_10000002
IMGVR2_3300014166_____Ga0134079_10000028
GOV2_Station82_SUR_COMBINED_FINAL_NODE_637
IMGVR2_3300009169_____Ga0105097_10000088
IMGVR2_3300012031_____Ga0136561_1000079
IMGVR2_3300012147_____Ga0136588_1000088
IMGVR2_3300019760_____Ga0193953_1000104
Olleya phage Harreka_1
IMGVR2_3300012989_____Ga0164305_10000040
Polaribacter_phage_P12002L
Polaribacter_phage_P12002S
GOV2_Station191_SUR_ALL_NODE_518
GOV2_Station193_SUR_ALL_NODE_772
Nonlabens_phage_P12024L
Nonlabens_phage_P12024S
Cellulophaga_phage_phi46:1
Riemerella_phage_RAP44
Flavobacterium_sp_phage_1/32
Croceibacter_phage_P2559S
Croceibacter_phage_P2559Y
Bacteroides_phage_B40_8
Bacteroides_phage_B124_14
Flosverisviridae
Winoviridae
Molycol-viridae
Pachyviridae
Assidu-viridae
Forseti-viridae
Aggregaviridae
0.1
1 - Genera
Unahavirus
Ingelinevirus
Leefvirus
Peternellavirus
Mollyvirus
Callevirus
Gundelvirus
Cellubavirus
Nekkelsvirus
Cebadecemvirus
Freyavirus
Harrekavirus
2 - Subfamilies
Dunevirinae
Helgolandvirinae
Pervagovirinae
Pachyvirinae
3 - Habitat
marine
limnic
waste water
soil/sediment
4 - Host association
Bacteroidetes
other phylum
5 - Life style genes
LuxR
integrase
T transposase
none of the above
* significant (boostrap above 50)


1
2
Enterobacteria_phage_WA13
Escherichia_phage_NC28
Escherichia_phage_ID62
Escherichia_phage_NC35
Escherichia_phage_ID21
Escherichia_phage_NC29
Escherichia_phage_ID32
Escherichia_phage_alpha3
Escherichia_phage_St_1
Bacteriophage_phiK
Enterobacteria_phage_G4
Escherichia_phage_ID52
Escherichia_phage_ID2_Moscow_ID_2001
Coliphage_phi_X174
Chlamydia_phage_CPG1
Chlamydia_phage_phiCPAR39
Chlamydia_phage_Chp2
Bdellovibrio_phage_phiMH2K
Chlamydia_phage_Chp1
Parabacteroides_phage_YZ_2015a
Spiroplasma_phage_4
Microviridae_IME_16
Acholeplasma_phage_MV_L1
Spiroplasma_phage_1_C74
Spiroplasma_virus_SkV1CR23x
Spiroplasma_virus_R8A2B
Spiroplasma_phage_SVTS2
Cellulophaga phage phi18:4
Cellulophaga phage phi48:1
Cellulophaga phage phi12a:1
Cellulophaga phage phi12:2
Cellulophaga phage Omtje_1
Obscuriviridae
Flavobacterium phage FLiP
Cellulophaga phage phi48:2
Propionibacterium_virus_B5
Thermus_phage_OH3
Enterobacteria_phage_Ike
Escherichia_virus_I22
Enterobacteria_phage_M13
Escherichia_phage_If1
Vibrio_phage_VFJ
Vibrio_virus_fs2
Pseudomonas_phage_Pf3
Vibrio_phage_VEJphi
Vibrio_virus_fs1
Vibrio_virus_VGJ
Vibrio_phage_Vf33
Vibrio_phage_VfO3K6
Vibrio_phage_VfO4K68
Vibrio_phage_KSF1
Vibrio_phage_VCY_phi
Vibrio_phage_CTXphi
Pseudomonas_phage_Pf1
Ralstonia_phage_Rs551
Ralstonia_phage_RS603
Ralstonia_virus_RSM1
Ralstonia_virus_RSM3
Xanthomonas_phage_Cf1c
Xanthomonas_phage_Xf109
Stenotrophomonas_phage_SMA6
Stenotrophomonas_phage_PSH1
Stenotrophomonas_phage_phiSMA9
Stenotrophomonas_phage_SMA7
Ralstonia_phage_PE226
Ralstonia_phage_RSS1
0.1
1 - Subfamily
Microviridae
Plectroviridae
Inoviridae
2 - Family
Bullavirinae
Gokushovirinae
* significant


Maribacter phages
Molly_1
matrixin family metalloprotease
OmpA family protein
BACON domain-containing protein
Recombination protein U
NAD-binding, UDP-glucose/GDP-mannose dehydrogenase family protein
beta glucosyl transferase
Thymidylate synthase
DegT/DnrJ/EryC1/StrS aminotransferase family
HD containing hydrolase-like enzyme
Colly_1
0 kb
20 kb
40 kb
60 kb
80 kb
100 kb
120 kb
Tenacibaculum phage
Gundel_1
JAB-domain-containing protein
zinc-finger-containing domain
AAA family ATPase
antirepressor protein
L-Ala-D-Glu-peptidase like
tRNA
tRNAs
tRNAs
0 kb
20 kb
40 kb
60 kb
80 kb
Cellulophaga phages
Calle_2
calcineurin-like phosphoeserase superfamily domain
peptidase
tRNAs
tRNAs
RyR domain-containing protein
0 kb
20 kb
40 kb
60 kb
Nekkels_1
YopX protein
acyl carier protein
calcineurin-like phosphoeserase superfamily domain
0 kb
20 kb
40 kb
Ingeline_1
BACON domain-containing protein
VHS1069 protein
0 kb
20 kb
40 kb
Polaribacter phages
Freya_1
ferric uptake regulator
beta-lactamase superfamily domain
PcfK-like protein
PcfJ-like protein
0 kb
20 kb
40 kb
Danklef_1
beta-lactamase superfamily domain
PcfK-like protein
PcfJ-like protein
ferric uptake regulator
0 kb
20 kb
40 kb
Leef_1
BACON domain-containing protein
KilA-N domain
RNA-binding protein
Ycel family protein
0 kb
20 kb
Winogradskyella phage
Peternella_1
ATP-dependent serine protease
KilA-N domain
oxidase
0 kb
20 kb
40 kb
Olleya phage
Harreka_1
chitinase GH19
RyR domain-containing protein
YopX protein
calcineurin-like phosphoeserase superfamily domain
acyl carier protein
0 kb
20 kb
40 kb
20 kb
Cellulophaga phage (ssDNA)
Omtje_1
0 kb
1 kb
2 kb
3 kb
4 kb
5 kb
6 kb
1 kb
hypothetical protein/ DUF
functional annotated protein
structural protein
genome ends
tRNA
tmRNA
tail associated
fibers
capsid
packaging
lysis associated
DNA replication
nucleotide/DNA manipulation
regulators and transcription factors
chaperon