Highly diverse flavobacterial phages isolated from 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 are a mortality factor for their bacterial hosts. This could be particularly true in spring phytoplankton blooms, which are known to be closely followed by a highly specialized bacterial community. We hypothesized that phages modulate these dense heterotrophic bacteria successions following phytoplankton blooms. 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. Taxonomic analysis allowed us to delineate ten new phage genera and ten 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 showed 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 suggested 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 likely 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, which are often dominated by diatoms (15), and 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* belong to the main responders (16, 17) and their increase is linked to the release of phytoplankton derived polysaccharides (18, 19). These

polysaccharides are produced by microalgae as storage compounds, cell wall building blocks, and exudates (20-22). This highly complex organic matter is likely converted by the *Flavobacteriia* to low molecular weight compounds and thus they are important for the carbon turnover during phytoplankton blooms (18, 23-25). Recurrent genera like *Polaribacter*, *Maribacter*, and *Tenacibaculum* succeed each other in a highly dynamic fashion (19). This bacterial succession is likely triggered by the availability and quality of substrates such as polysaccharides (18, 19), 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 (26), whereas viral lysis is highly host-specific (27).

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 *Alphaproteobacteria* (e.g. (28, 29)), *Gammproteobacteria* (e.g. (30)), and *Cyanobacteria* (e.g. (31-36)) 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 phage morphotypes in the realm *Duplodnaviria* and also two different phage groups in the realm *Monodnaviria* (27, 37). Phages were also isolated for members of the genera *Polaribacter* (38), *Flavobacterium* (39, 40), *Croceibacter* (41) or *Nonlabens* (42) and included eight tailed phages, one having a myoviral and the rest having a siphoviral morphology. However, 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. Metagenomes obtained for Helgoland waters of different size fractions were mapped to all newly isolated flavophage genomes, 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, a weekly 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, absolute *Bacteroidetes* cell numbers 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). The *Bacteroidetes* cell numbers were determined by 16S rRNA targeted fluorescence *in situ* hybridization (FISH) with specific probes (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 pore size filters served as phage source. To ensure purity, three subsequent isolation rounds were performed by picking single phage plaques and using them as inoculum for new plaque assays. 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 (43)). The DNA was sequenced on an Illumina HiSeq3000 applying the paired-end 2 x 150 bp read mode. For most *Cellulophaga* phages (except Ingeline) and *Maribacter* phages a ChIP-seq (chromatin immunoprecipitation DNA-sequencing) library was prepared. This was done to generate a NGS library for ssDNA phages and due to library preparation issues with the *Maribacter* phages for Illumina and PacBio. 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, (44)) and Tadpole (v35.14, sourceforge.net/projects/bbmap/). Assembly quality was checked with Bandage (32). The genome ends were predicted using PhageTerm (33), but not experimentally verified. For more details about all these procedures, see SI file 1 text.

Retrieval of related phage genomes and taxonomic assignment

Several publicly available datasets of cultivated and environmental phage genomes were queried for sequences related with the flavophages isolated in this study, in a multistep procedure (SI file 1 text). The datatsets included GenBank Viral (ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/viral/, downloaded on 17.03.2021), the GOV2 dataset, IMG/VR2, as well as further environmental datasets (13, 45-50). To determine the relationship of the new flavophages and their relatives with taxa recognized by the the International Committee of Taxonomy of Viruses (ICTV), we have added

these phages to larger datasets, including ICTV recognized phages (for details, see SI file 1 text).

To determine the family level classification of the flavophages, we used VirClust (ref, www.virclust.icbm.de), ViPTree (51) and VICTOR (52) to calculate protein based hierarchical clustering trees. For dsDNA flavophages, a VirClust hierarchical tree was first calculated for the isolates, their relatives and the ICTV dataset. Based on this, a reduced dataset was compiled, from family level clades containing our flavophages. The reduced dsDNA flavophages dataset and the complete ssDNA flavophage dataset were further analysed with VirClust, VipTree and VICTOR. The parameters for VirClust were: i) protein clustering based on "evalue", after reciprocal BLASTP hits were removed if e-value > 0.0001 and bitscore < 50; ii) hierarchical clustering based on protein clusters, agglomeration method "complete", 1000 bootstraps, tree cut at a distance of 0.9. 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 (52). Furthermore, the web service of GRAViTy (http://gravity.cvr.gla.ac.uk, (53)) was used to determine the similarity of ssDNA phages and their relatives with other ssDNA viruses in the Baltimore Group II, *Papillomaviridae* and *Polyomaviridae* (VMRv34).

To determine the intra-familial relationships, smaller phage genome datasets corresponding to each family were analysed using i) nucleic acid based intergenomic similarities calculated with VIRIDIC (54) and ii) core protein phylogeny. The thresholds used for species and genus definition were 95% and 70% intergenomic similarity,

respectively. The core protein analysis was conducted as follows: i) core genes were calculated with the VirClust web tool (virclust.icbm.de), based on protein clusters calculated with the above parameters; ii) duplicated proteins were removed; iii) a multiple alignment of all concatenated core proteins was constructed with MUSCLE (v3.8.425,(55)) and iv) used for the calculation of IQ-Trees with SH-aLRT (56) and ultrafast bootstrap values (57) using ModelFinder (58). All phylogenetic trees were visualized using FigTree v1.4.4. (59), available at http://tree.bio.ed.ac.uk/software/figtree/).

Phage genome annotation

All phage genomes analysed in this study were annotated using a custom bioinformatics pipeline described elsewhere (28), with modifications (SI file 1 text). The final protein annotations were evaluated manually.

Host assignment for environmental phage genomes

To determine potential hosts for the environmental phage genomes, several methods were used. First we did, a BLASTN (60) search (standard parameters) 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. Second, with the same genomes a BLASTN against the CRISPR/cas bacterial spacers from the metagenomic and isolate spacer database was run with standard settings using the IMG/VR website (https://img.jgi.doe.gov/cgi-bin/vr/main.cgi). Third, WIsH (61) was used to predict hosts (standard parameters) with the GEM metagenomic contig database (62) as host database.

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 (28). Two datasets were analysed: 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 bacteria from which phages were isolated 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 (63). Additionally, the 16S rRNA gene from all the other bacterial strains was amplified and sequenced using the Sanger technology (see SI file 1 text).

For phylogenetic analysis, a neighbour joining tree with Jukes-Cantor correction and a RaxML tree (version 8, (64)) were calculated using ARB (65). The reference data set Ref132 was used, with the termini filter and *Capnocytophaga* as outgroup (66). Afterwards, a consensus tree was calculated.

CRISPR spacer search

CRISPR spacers and cas systems were identified in the host genomes by CRISPRCasFinder (67). Extracted spacers were mapped with the Geneious Assembler to the flavophage genomes in highest sensitivity mode without trimming. Gaps were allowed up to 20% of the spacer and with a maximum size of 5, word length was 10, and a maximum of 50% mismatches per spacer were allowed. Gaps were counted as mismatches and only results up to 1 mismatch were considered for the phage assignment to the hosts used in this study.

The IMG/VR (48) 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 (e-value of 1e-5). Only hits with less than two mismatches were taken into account.

Data availability

Sequencing data and phage genomes are available at the National Center for Biotechnology Information (NCBI) with the accession number PRJNA639310 (for details see SI file Table 1 and Table 3). 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 FISH 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 (68). 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 (69) (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 4).

Flavophage isolation and classification

For phage enrichment, 23 bacterial strains previously isolated from algal blooms in the North Sea were used as potential hosts (SI file 1 Table 1). In 2017, we implemented a method for enriching flavophages on six host bacteria. A much larger and

more diverse collection of 21 mostly recently isolated *Flavobacteriia* was used in 2018. A total of 108 phage isolates were obtained for 10 of the bacterial strains, either by direct plating or by enrichment (see Table 1) These were affiliated with the bacterial genera *Polaribacter*, *Cellulophaga*, *Olleya*, *Tenacibaculum*, *Winogradskyella*, and *Maribacter* (Figure 2, SI file 1 Table 5).

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 3). A summary of the new phage species and their exemplar isolate phage, including their binomial name and isolation data, is found in Table 1. For brevity, we are mentioning here only the short exemplar isolate phage names of each new species: Harreka, Gundel, Molly, Colly, Peternella, Danklef, Freya, Leef, Nekkels, Ingeline, Calle and Omtje.

Virion morphology as determined by TEM showed that 11 of the exemplar phages were tailed (Figure 3). According to the new megataxonomy of viruses, tailed phages belong to the realm *Duplodnaviria*, kingdom *Heunggongvirae*, phylum *Uroviricota*, class *Caudoviricetes*, order *Caudovirales (70).* The only non-tailed, icosahedral phage was Omtje (Figure 4). Further digestion experiments with different nucleases showed that Omtje has a ssDNA genome (SI file 1 Figure 5).

Hierarchical clustering with VirClust placed the new dsDNA, tailed flavophages into 9 clades of similar rank with the current eleven *Caudovirales* families (Figure 3, SI file 4). Similar clades were obtained with VICTOR and ViPTree (SI file 1 Figure 8 and Figure 9). These clades formed individual clusters when the VirClust tree was cut at a 0.9 distance threshold (Figure 3), which was shown to delineate the majority of *Caudovirales*

families (https://rhea.icbm.uni-oldenburg.de/VIRCLUST/). In agreement, these clades were assigned to different subfamilies by the VICTOR analysis (SI file 1 Figure 8), which, in the light of the current changes in the ICTV phage classification, represent different families (71). Therefore, we are proposing that the 9 clades correspond to 9 new families, which we tentatively named “Helgolandviridae”, “Duneviridae”, “Winoviridae”, “Molycolviridae”, “Pachyviridae”, “Pervagoviridae” “Assiduviridae”, “Forsetiviridae” and “Aggregaviridae” (Figure 3). The core proteins for each family consisted mainly of structural proteins (Table 2). With the exception of few proteins from “Pachyviridae” and “Pervagoviridae”, the core proteins were not shared between the families (they belonged to separate protein clusters, see Figure 3 and Table 2). Four of the new families (“Aggregaviridae”, “Forsetiviridae”, “Molycolviridae” and “Helgolandviridae”) were formed only from new flavophages and from environmental phage genomes (Figure 3, SI file 1 Figure 8). The remaining five families also contained previously cultivated phages, infecting bacteria from the genus *Cellulophaga* (“Pervagoviridae”, “Pachyviridae” and “Assiduviridae”) and *Flavobacterium* (“Duneviridae”, “Winoviridae”). Part of the cultivated flavobacterial phages in the “Duneviridae” and “Winoviridae” are currently classified by ICTV in three genera in the families *Siphoviridae* and *Myoviridae*. However, because the *Siphoviridae* and *Myoviridae* families are based on phage morphologies, they are being slowly dissolved and split into new families, based on sequence data (72).

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 5). The assignment to new genera was supported by the core proteins phylogenetic analysis (SI file 1 Figure 10 – 15).

Hierarchical clustering using VirClust (Figure 4), ViPTree (SI file 1 Figure 18) and VICTOR (SI file 1 Figure 19) of a dataset including Omtje and all related and reference ssDNA phages showed that Omtje is clustering with previously isolated ssDNA phages infecting *Cellulophaga*, separately from other ssDNA phage families, the *Microviridae, Inoviridae* and *Plectroviridae.* This was supported also by GRAViTy (SI file 6). Only one protein cluster was shared outside this cluster, with Flavobacterium phage FliP (Figure 4), even when forming protein-superclusters based on HMM similarities (SI file 1 Figure 20). We propose here that this cluster represents a new family, tentatively called here “Obscuriviridae”. The placement of this family into higher taxonomic ranks, including the realm, remains to be determined in the future. Intergenomic similarity calculations (SI file 7), supported by core gene phylogenetic analysis (SI file 1 Figure 21), indicate that Omtje forms a genus by itself, tentatively named here “Omtjevirus”.

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 "Helgolandviridae", infected *Polaribacter* sp. AHE13PA (SI file Figure 1). These phages infected only their isolation host (SI file 1 Figure 22). They all had a siphoviral morphology (Figure 3 and Table 3).

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 3). 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, potentially functioning as an endolysin, was detected in Danklef and Leef, surrounded by transmembrane domain (TMD) containing proteins. All three phages encoded an integrase, and thus have the potential to undergo a temperate lifestyle. Leef had also a LuxR protein, which is a quorum sensing dependent transcriptional activator, and a pectin lyase.

Phages of *Cellulophaga*

Four Cellulophaga phages were obtained: i) Calle, part of the family "Pervagoviridae", 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 "Duneviridae", 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 22). 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 3).

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 3). 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-acetylmuramoyl-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 3, 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, pointing toward a potential temperate life style. Calle had 20 tRNAs and one tmRNA gene (Figure 5, SI file 2).

Phages infecting 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 "Pachyviridae", 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 22). All virions were tailed, with a podoviral morphology for Gundel, and a myoviral morphology for Molly, Peternella and Harreka (Table 3). The genome size ranged from ~ 40 kbp to ~ 125 kbp and the G+C content between 30.4 and 36.2%. For Gundel short direct terminal repeats (DTRs) were predicted as genome ends (Table 3).

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 (73) and hybrid phage/host genome reads, indicating that Peternella is likely a transducing phage.

Potential endolysins 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 3, SI file 2).

Additional features of these phages were: i) ten 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

Six of the nine proposed new families ("Forsetiviridae", "Pachyviridae", "Pervagoviridae", "Winoviridae", "Helgolandviridae" and "Duneviridae") include members for whose genomes were assembled from environmental metagenomes (Figure 3). We have briefly investigated which bacterial groups are potential hosts for these phages. Most of them gave BLASTN hits with a length between 74 and 5844 bases with bacterial genomes from the *Bacteroidetes* phylum (Figure 3, SI file 1 Table 6 and Table 7), likely due to the presence of prophages and horizontal gene transfer events. Some of the environmental viral genomes gave *Bacteroidetes* associated hits against the metagenome CRISPR spacer database (SI file 1 Table 8). The host prediction using WIsH supported the results by BLASTN and CRISPR spacer matching, and predicted members of *Bacteroidetes* as hosts for most of the environmental viral genomes. Some hosts were identified down to the family level, as *Flavobacteriaceae* (SI file 1 Table 9). Only phages in "Winoviridae", "Pervagoviridae" and "Helgolandviridae" had other families than *Flavobacteriaceae* as hosts, but all in the *Bacteroidetes* phylum (Figure 3). Additionally, several marine contigs from the "Helgolandviridae" contained *Bacteroidetes* Associated Carbohydrate-binding Often N-terminal (BACON) domains (SI file 1 Table 11). Together, these results suggest that most of the environmental viral genomes in the new families are infecting members of the phylum *Bacteroidetes*.

Integrase encoding genes were found on all environmental phages from the "Forsetiviridae", and some of the genomes from "Winoviridae", "Pachyviridae", "Helgolandviridae" and "Duneviridae". LuxR encoding genes were found in many genomes from "Helgolandviridae", "Duneviridae" and "Forsetiviridae", and one had also a transporter for the auto-inducer 2 (Figure 3, SI file 1 Table 11). In the "Winoviridae", all phages encoded Mu-like structural proteins, including the MuD terminase, and two environmental viral genomes 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 strain, several spacers matched Danklef genomes, one matched Freya and another Leef (SI file 1 Table 10). 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 7), 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 and dynamics of flavophages in the North Sea, we mined by read mapping cellular metagenomes (> 10 µm, 3 – 10 µm, 0.2 – 3 µm) 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 4). The complete genomes of Freya, Harreka and Ingeline were covered by reads with 100% identity, signifying that these exact 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 10.4 normalized genome coverage, respectively.

Further, we searched for the presence of the five flavophage hosts in the 2018 spring bloom (Table 4). *Polaribacter* sp. was found in the >10 µm and 0.2 - 3 µm fractions, at different time points during the bloom, including the same two samples in which its phages (Freya, Danklef and Leef) and their relatives were detected. At the beginning of the bloom, in April, significantly more phages than hosts were present in the > 10 µm fraction. 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 - 3 µm and > 10 µm fractions, at low abundances and dates preceding the detection of its phage, Harreka. *Cellulophaga* sp., the host for Ingeline, was not found (SI file 1 Table 12-16).

Discussion

We have performed a cultivation-based assessment of the diversity of flavophages potentially modulating *Flavobacteriia*, a key group of heterotrophic bacteria, in the ecological context of spring bloom events in two consecutive years. In contrast to high-throughput viromics-based studies, our approach enabled us to link phages to their host species and even to strains.

Highly diverse flavophages, belonging to two distinct viral realms were isolated. As a point of reference, a viral realm is the equivalent of a domain in the cellular world (70). Furthermore, four of the ten new families, and nine of the ten new genera had no previously cultivated representatives. This study not only uncovers a novel flavophage diversity, but also, structures a substantial part of the known marine flavophage diversity into families. These novel flavophage families are relevant not only for the marine environment. Besides cultivated flavophages, six of the families also include environmental phages from marine, freshwater, wastewater and soil samples, which most likely infect *Bacteroidetes.*

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 taxa are being submitted, one for flavophages in *Duplodnaviria* and one for “Obscuriviridae”.

Genomic analysis indicates that the new flavophages have various life styles and diverse replication strategy characteristics. Some families are dominated by potentially temperate phages, and others by potentially strictly lytic phages, as indicated by the presence/absence of integrases. Genome replication can take place (i) through long

concatemers (74) (Gundel and Leef), (ii) replicative transposition (75) (Peternella), and (iii) the rolling circle mechanism (76) (Omtje).

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. 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 peptidoglycan may also be degraded (77). A phage GH19 expressed in *Escherichia coli* caused cellular lysis (78). 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, the endolysins could degrade chitin, an abundant polysaccharide in the marine environment, produced for example by green-algae or copepods (79).

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, which are produced in large quantities by micro- and macro-algae (80-82), serve as substrate for marine *Flavobacteriia*, especially *Polaribacter* (83-85). Bacteria are not only able to degrade these polysaccharides, but also to produce them and form capsules or an extracellular matrix of biofilms (86-88). In phages, such polysaccharide degrading enzymes are usually located on the tails, as part of proteins with multiple domains to reach the bacterial cell membrane. Another enzyme

potentially degrading proteins in the extracellular matrix was the zinc-dependent metallopeptidase of Molly, a “Mollyvirus”, which infected 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 (89, 90). 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 exhibit a surface-associated life style (91). 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 phytoplankton or zooplankton. Spacers in metagenomes matching Nekkels, a *Cellulophaga* phage, suggest an association with red macro-algae (SI file 1 Table 7). 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*. The ability for an attached life style is indicated by scanning electron microscopy images of *Cellulophaga* sp. HaHaR_3_176 cells showing high amounts of extracellular material (SI file 1 Figure 23).

The presence of integrases and LuxR genes in phages from the “Helgolandviridae” and Duneviridae” (Figure 3 and SI file 1 Table 11) suggests that they likely respond to changes in the host cell densities, for example by switching from the temperate to the lytic cycle, as recently observed in *Vibrio cholerae* (pro)-phages (92, 93). 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 planktonic organisms or particles.

All of our flavophages, except Molly and Colly, replicated successfully at different times throughout the 2018 phytoplankton bloom. The isolation by enrichment or direct-plating indicated their presence in the viral fraction, and the retrieval by read-mapping indicated a likely presence inside the host or association with surfaces. Calle and Nekkels, infecting *Cellulophaga*, although not detected in the cellular metagenomes, were present in high abundance (at least 100 plaque forming units ml^{-1}) in the viral fraction of our samples, as revealed by successful isolation by direct-plating without previous enrichment. This apparent gap between the presences in either the viral or cellular fraction, could be either explained by phage lysis prior the metagenome collection, or by the fact that their host habitat was not sampled for metagenomics. A potential explanation is that members of the genus *Cellulophaga* are known to grow predominantly on macro-algae (94). The detection of Omtje, Peternella and Gundel in enrichment cultures, but not by direct-plating or in the cellular metagenomes, indicate that their presence in the environment is low. Further investigations of the specific habitat of both phage and host are necessary to confirm these findings.

For flavophages with temperate potential, the ratio between phage and host normalized read abundance can be used to predict their lytic or temperate state in the environmental samples. *Cellulophaga*, the host of Ingeline, was not detected throughout the spring bloom. However, because Ingeline was detected presumably in the particle fraction and thus might be inside its host, 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 April suggest that these phages were actively replicating and lysing their cells. In contrast, when they reappeared in May, these phages were not only three orders of magnitude less abundant than in April, but also ~ 10 times less abundant than 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 phage populations is also reflected in the appearance of Freya and Gundel. Moreover, it is reminiscent of the boom and bust behavior previously observed with T4-like marine phages (95). In addition, the read mapping results indicated a change in dominant phage genotypes, from the same species and even strain with Danklef and Freya in April, towards more distant relatives at the genus or even family level 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 (96). Persistence over longer times in the environment was shown for *Polaribacter* phages by the CRISPR/Cas spacers, and for Omtje and Ingeline, by their isolation in two consecutive years.

The North Sea flavophages isolated in this study display a very 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 either lysis or with regard to specific bacterial population by substrate facilitation, they are modulating the carbon cycling in coastal shelf seas. The increase in bacterial numbers, reflected in the phage numbers and the ratio of phage to bacterial cells,

indicate that phages actively replicate through the 2018 phytoplankton bloom, matching previous observations of the North Sea microbial community (97). Our read mapping data indicate complex dynamics, which can now be further investigated with a large number of valuable 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 Franzmeyer 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 Polysaccharide Utilisation (POMPU)" project by a grant of RA (AM73/9-1) 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 Program (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 Program Gut Microbes and Health BB/R012490/1 and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356.

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.

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 M, 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. Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol Ò, Malits A, et al. Changes in bacterioplankton composition under different phytoplankton regimens. Applied and Environmental Microbiology. 2004;70(11):6753-66.

17. Simon M, Glöckner F, Amann R. Different community structure and temperature optima of heterotrophic picoplankton in various regions of the Southern Ocean. Aquatic Microbial Ecology - AQUAT MICROB ECOL. 1999;18:275-84.
18. 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.
19. 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.
20. Gügi B, Le Costaouec T, Burel C, Lerouge P, Helbert W, Bardor M. Diatom-Specific Oligosaccharide and Polysaccharide Structures Help to Unravel Biosynthetic Capabilities in Diatoms. Mar Drugs. 2015;13(9):5993-6018.
21. Haug A, Myklestad S. Polysaccharides of marine diatoms with special reference to Chaetoceros species. Marine Biology. 1976;34(3):217-22.
22. Painter TJ. Algal polysaccharides. The polysaccharides: Elsevier; 1983. p. 195-285.
23. Unfried F, Becker S, Robb CS, Hehemann J-H, Markert S, Heiden SE, et al. Adaptive mechanisms that provide competitive advantages to marine bacteroidetes during microalgal blooms. The ISME Journal. 2018;12(12):2894-906.
24. Reintjes G, Arnosti C, Fuchs BM, Amann R. An alternative polysaccharide uptake mechanism of marine bacteria. The ISME Journal. 2017;11(7):1640-50.
25. Thomas F, Le Duff N, Wu T-D, Cébron A, Uroz S, Riera P, et al. Isotopic tracing reveals single-cell assimilation of a macroalgal polysaccharide by a few marine Flavobacteria and Gammaproteobacteria. The ISME Journal. 2021.
26. 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.
27. 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.
28. 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.
29. 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-.
30. 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.
31. 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.
32. Suttle CA, Chan AM. Marine cyanophages infecting oceanic and coastal strains of Synechococcus: abundance, morphology, cross-infectivity and growth characteristics. Marine Ecology-Progress Series. 1993;92:99-.

33. Wilson WH, Carr NG, Mann NH. The effect of phosphate status on the kinetics of cyanophage inefction in the oceanic cyanobacterium *Synechococcus* sp. WH78031 Journal of Phycology. 1996;32(4):506-16.
34. Fuller NJ, Wilson WH, Joint IR, Mann NH. Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. Applied and Environmental Microbiology. 1998;64(6):2051-60.
35. Proctor LM, Fuhrman JA. Viral mortality of marine bacteria and cyanobacteria. Nature. 1990;343(6253):60-2.
36. Suttle CA, Chan AM, Cottrell MT. Infection of phytoplankton by viruses and reduction of primary productivity. Nature. 1990;347(6292):467-9.
37. 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.
38. 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.
39. 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.
40. 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.
41. Kang I, Kang D, Cho J-C. Complete Genome Sequence of *Croceibacter* Bacteriophage P2559S. Journal of Virology. 2012;86(16):8912-3.
42. 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.
43. 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.
44. 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.
45. 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.
46. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R. Expanding the Marine Virosphere Using Metagenomics. PLOS Genetics. 2013;9(12):e1003987.
47. 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.
48. 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.

49. 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.
50. 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.
51. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017;33(15):2379-80.
52. Meier-Kolthoff JP, Göker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics. 2017;33(21):3396-404.
53. 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.
54. Moraru C, Varsani A, Kropinski AM. VIRIDIC – a novel tool to calculate the intergenomic similarities of prokaryote-infecting viruses. bioRxiv. 2020:2020.07.05.188268.
55. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004;32(5):1792-7.
56. 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.
57. 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.
58. 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.
59. Rambaut A. 1.4.4—a graphical viewer of phylogenetic trees and a program for producing publication-ready figures. 2018.
60. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.
61. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics. 2017;33(19):3113-4.
62. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nature Biotechnology. 2020.
63. 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.
64. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.
65. 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.

66. 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.
67. 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.
68. 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.
69. Nagasaki K. Dinoflagellates, diatoms, and their viruses. The Journal of Microbiology. 2008;46(3):235-43.
70. 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.
71. 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.
72. Turner D, Kropinski AM, Adriaenssens EM. A Roadmap for Genome-Based Phage Taxonomy. Viruses. 2021;13(3):506.
73. 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.
74. 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.
75. Montaño SP, Pigli YZ, Rice PA. The Mu transpososome structure sheds light on DDE recombinase evolution. Nature. 2012;491(7424):413-7.
76. Krupovic M. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Current Opinion in Virology. 2013;3(5):578-86.
77. 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.
78. 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.
79. Souza CP, Almeida BC, Colwell RR, Rivera ING. The Importance of Chitin in the Marine Environment. Marine Biotechnology. 2011;13(5):823.
80. 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.
81. 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.
82. 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.

83. 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.
84. 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.
85. 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.
86. Maleki S, Almaas E, Zotchev S, Valla S, Ertesvåg H. Alginate Biosynthesis Factories in Pseudomonas fluorescens: Localization and Correlation with Alginate Production Level. Applied and Environmental Microbiology. 2016;82(4):1227-36.
87. Singh JK, Adams FG, Brown MH. Diversity and Function of Capsular Polysaccharide in Acinetobacter baumannii. Frontiers in Microbiology. 2019;9(3301).
88. Limoli DH, Jones CJ, Wozniak DJ. Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. Microbiol Spectr. 2015;3(3):10.1128/microbiolspec.MB-0011-2014.
89. 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.
90. 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.
91. 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.
92. Silpe JE, Bassler BL. Phage-Encoded LuxR-Type Receptors Responsive to Host-Produced Bacterial Quorum-Sensing Autoinducers. mBio. 2019;10(2):e00638-19.
93. Silpe JE, Bassler BL. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. Cell. 2019;176(1):268-80.e13.
94. 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.
95. 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.

96. 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.
97. 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.
98. Shimodaira H, Terada Y. Selective Inference for Testing Trees and Edges in Phylogenetics. Frontiers in Ecology and Evolution. 2019;7(174).
99. Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22(12):1540-2.

Table 1: Phylogenetic characterization and isolation details of each phage group. The names have a Frisian origin, to reflect the flavophage place of isolation.

Exemplar phage abbreviation	Phage species	Phage genus	Family	Year of isolation	Isolation sources	Number of strains
Molly	"Mollyvirus Molly"	"Mollyvirus"	"Molycolviridae"	2017	Enrichment	6
Colly	"Mollyvirus Colly"	"Mollyvirus"	"Molycolviridae"	2017	Enrichment	1
Gundel	"Gundelvirus Gundel"	"Gundelvirus"	"Pachyviridae"	2018	Enrichment	1
Calle	"Callevirus Calle"	"Callevirus"	"Pervagoviridae"	2018	Enrichment, direct plating	3
Nekkels	"Nekkelsvirus Nekkels"	"Nekkelsvirus"	"Assiduviridae"	2018	Enrichment, direct plating	2
Omtje	"Omtjevirus Omtje"	"Omtjevirus"	"Obscuriviridae"	2017 & 2018	Enrichment	5
Ingeline	"Ingelinevirus Ingeline"	"Ingelinevirus"	"Duneviridae"	2017 & 2018	Enrichment	8
Leef	"Leefvirus Leef"	"Leefvirus"	"Helgolandviridae"	2018	Enrichment	1
Danklef	"Freyavirus Danklef"	"Freyavirus"	"Forsetiviridae"	2018	Enrichment	5
Freya	"Freyavirus Freya"	"Freyavirus"	"Forsetiviridae"	2018	Enrichment	10
Peternella	"Peternellavirus Peternella"	"Peternellavirus"	"Winoviridae"	2018	Enrichment	1
Harreka	"Harrekavirus Harreka"	"Harrekavirus"	"Aggregaviridae"	2018	Enrichment	1

Table 2: Core genes of the newly defined families.

Family	Total core genes	Core gene annotation	Number of core genes used for phylogeny	Core genes used for phylogeny
"Forsetiviridae"	8	tape measure protein, major capsid protein, portal protein, structural proteins (5)	7	tape measure protein, major capsid protein, portal protein, structural proteins (4)
"Winoviridae"	10	Clp protease, neck protein, major capsid protein, sheath, phage protein D, head morphogenesis protein/portal protein, hp (4)	9	Clp protease, neck protein, major capsid protein, sheath, phage protein D, hp (4)
"Pachyviridae"	11	portal protein*, sheath, structural protein** (5), hp*** (4)	9	portal protein, structural protein (4), hp (4)
"Pervagoviridae"	8	portal protein*, chaperonin cpn10, structural protein** (4), hp*** (2)	3	chaperonin cpn10, structural protein (2)
"Duneviridae"	9	major capsid protein, portal protein, adaptor protein, hp (5)	7	major capsid protein, adaptor protein, hp (4)
"Helgolandviridae"	6	hp (6)	5	hp (5)
"Assiduviridae"	33	phage tail tape measure protein, DNA primase/helicase, NinG recombination protein, pectate lyase (2), structural protein, hp (27)	n.d.	
"Obscuriviridae"	11	replication initiation factor, mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase, structural protein (8), hp	10	replication initiation factor, mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase, structural protein (8)

*the portal proteins are shared (belongs to the same protein cluster) between *Pachyviridae* and *Pevagoviridae*
**2 structural proteins each are shared (belong to the same protein cluster) between *Pachyviridae* and *Pevagoviridae*
*** one hypothetical protein is shared (belongs to the same protein cluster) between *Pachyviridae* and *Pevagoviridae*

Table 3: Phage characteristics of each phage group. As Colly has the same characteristics as Molly, it is not especially mentioned in the table. Lysis genes are stated as N-acetylmuramidase (M), N-acetylmuramoyl-L-alanine-amidase (A), Glycosid Hydrolase 19 (GH), L-Alanine-D-glutamine-peptidase (P), spanin (S), holing (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
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
Gundel	*Tenacibaculum*	78.5	30.4	118	10	0	0	P	short DTRs	yes	60.5±5.2	22.7±3.2	-	podoviridal
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
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
Omtje	*Cellulophaga*	6.6	31.2	13	0	0	0	A	nd	yes	52.3±4.6	-	-	non tailed
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
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
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
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
Peternella	*Winogradskyella*	39.6	35.3	62	0	0	1	A, Hx	Mu-like	no	52.3±4.2	105.8±7.4	16.4±2.4	myoviridal
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

Table 4: Read mapping results from 2018 metagenomes for isolated flavophages and their hosts. An * indicates those phages covered by reads with 100% identity over the whole genome length.

Julian day /Gregorian date	Fraction	Phage		Host		Phage/Host ratio
		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)		384
		Freya relatives	63.3			454
		Danklef	10.4			74
		Danklef relatives	45.8			327
		Ingeline*	0.7	not detected		
		Ingeline relatives	0.8			
		not detected		*Olleya* sp. HaHaR_3_96		
116 / 26.04.2018	10 µm	not detected		*Polaribacter* sp. HaHaR_3_91 and R2A056_3_33		
		not detected		*Olleya* sp. HaHaR_3_96		
122 / 02.05.2018	0.2 µm	not detected		*Polaribacter* sp. HaHaR_3_91 and		
		not detected		*Olleya* sp. HaHaR_3_96		
128 / 08.05.2018	3 µm	Harreka*	1.8	not detected		
		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)		0.09
		Danklef relatives	0.03			0.07
		Leef relatives	0.02			0.04

Figure Legends

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 bacterial cell numbers (black line), *Bacteroidetes* numbers (orange line), phage numbers by transmission electron microscopy (TEM, black bar), and phage numbers 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 some 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.

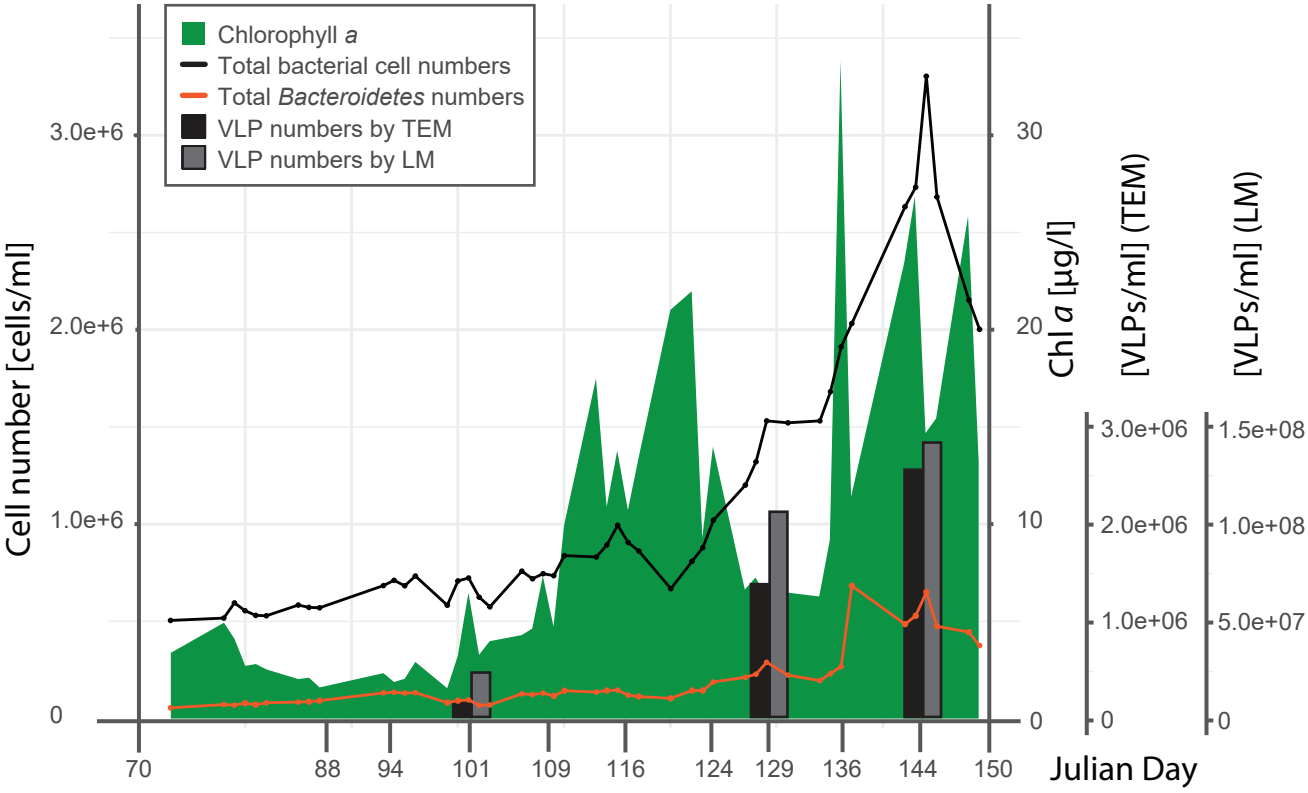
Figure 2: Phylogenetic tree (consensus between RaxML and neighbor-joining) of the 16S rRNA gene from all bacterial strains used to enrich for phages in 2017 and 2018 (red), plus reference genomes. Red squares indicate successful phage isolation.

Figure 3: VirClust hierarchical clustering of the new dsDNA flavophages and their relatives (reduced dataset), based on intergenomic distances calculated using the protein cluster content. For an extended tree, including the larger ICTV dataset, see SI file 4. **1.** Hierarchical clustering tree. Two support values, selective inference (si, (98)) and approximately unbiased (au, (99)), are indicated at branching points (si/au) only for the major clades (see SI file 1 Figure 6 and Figure 7 for all support values). The tree was cut

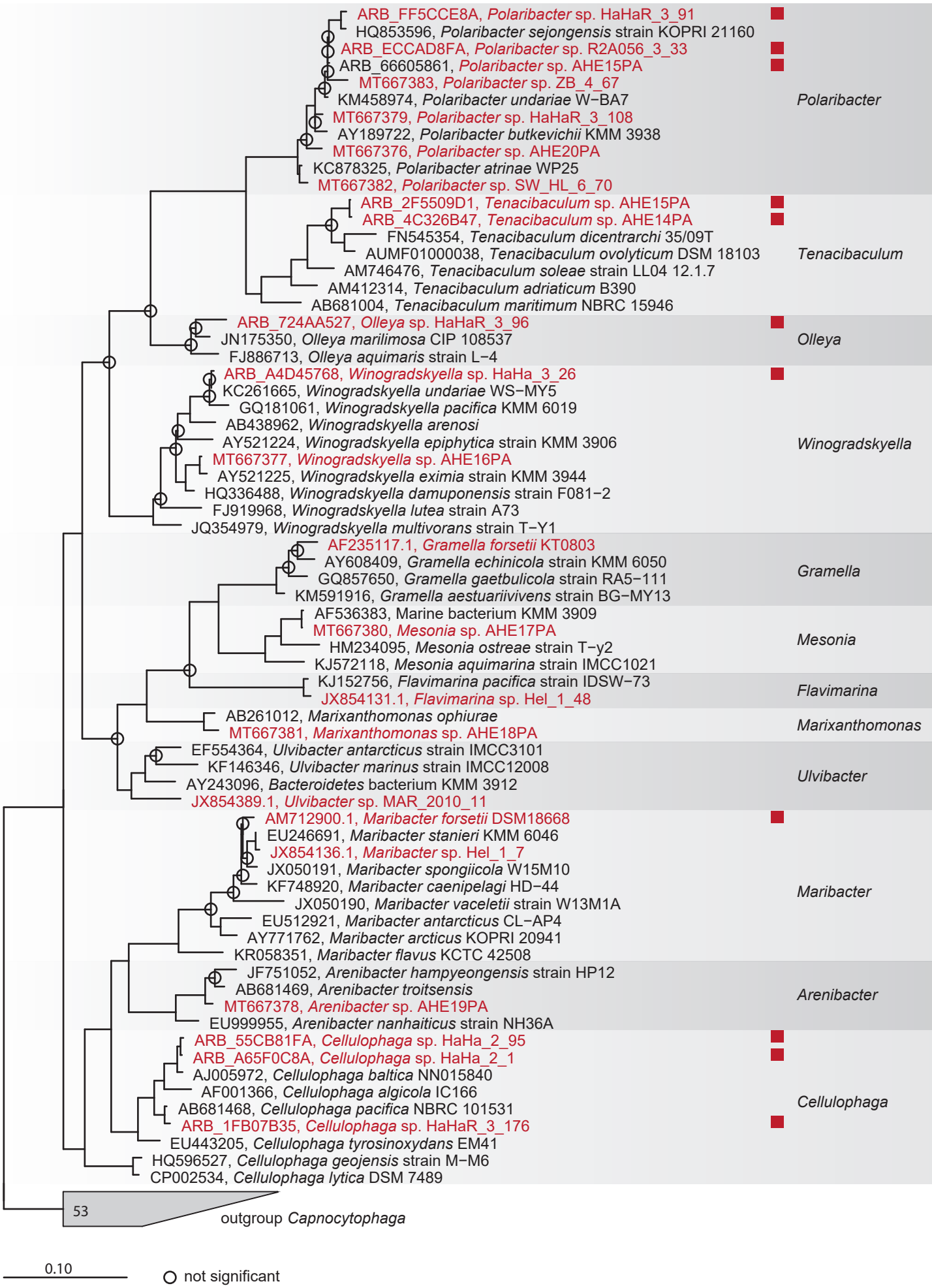
into smaller viral genome clusters (VGCs) using a 0.9 distance threshold. Each VGC containing our flavophages was proposed here as a new family. Exceptions were made for the "Aggregaviridae" and "Forsetiviridae", for which only part of the VGC were included in the families, to exlude some genomes with lower support values. Each VGC is framed in a rectangle in 2 and 3. **2.** Silhouette width, measures how related is a virus with other viruses in the same VGCs. Similarity to other VGCs is indicated by values closer to -1 (red). Similarity to viruses in the same VGC is indicated by values closer to 1 (green). **3.** Distribution of the protein clusters (PCs) in the viral genomes. **4.** Genome length (bps). **5.** Fraction of proteins shared with other viruses (dark grey), based on protein assignment to PCs. **6.** Virus names, with flavophages isolated in this study marked in red. **7.** Genus assignment. Gray bars indicate already defined genera by the ICTV with the following names from top to bottom: *Silviavirus*, *Labanvirus*, *Unahavirus*, *Pippivirus*, *Lillamyvirus*, *Muminvirus*, *Lillamyvirus*, *Helsingorvirus*, and *Incheonvirus*. **8.** Habitat. **9.** Host association, as determined by: 9a) BlastN; 9b) CRISPR spacers; 9c) WIsH with host database GEM. **10.** Life style genes. **11.** TEM images of the new flavophages, uranyl acetate negative staining. Scale bar in each TEM image has 50 nm **12.** Newly proposed families. Full name of environmental phages in "Winoviridae" and unclassified: AP013511.1_Uncultured_Mediterranean_phage_uvMED,_group_G21,_isolate__uvMED-CGR-C117A-MedDCM-OCT-S32-C49 and , AP013358.1_Uncultured_Mediterranean_phage_uvMED,_group_G1,_isolate__uvMED-CGR-U-MedDCM-OCT-S27-C45, respectively.

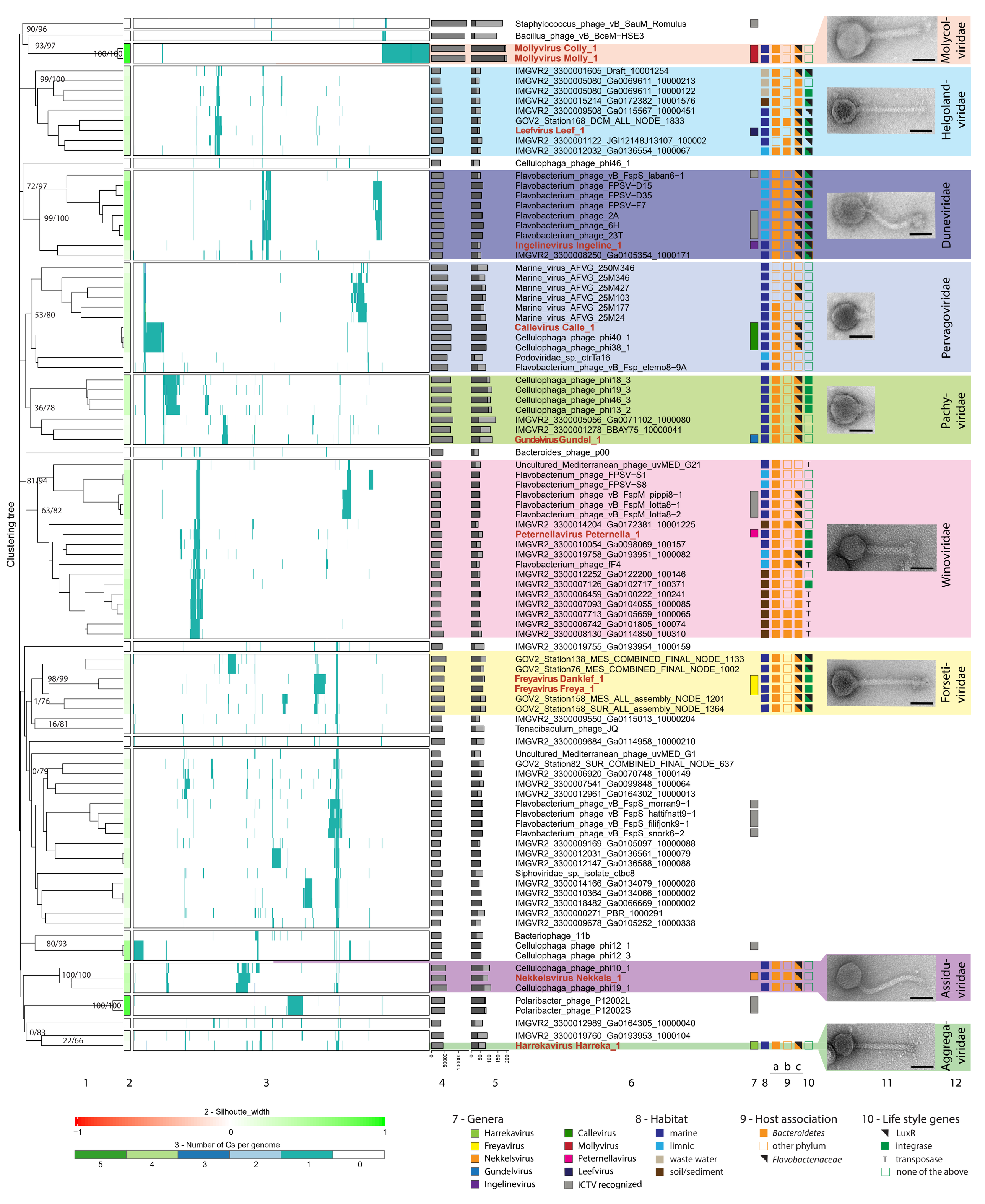
Figure 4: VirClust hierarchical clustering of the new ssDNA flavophages and their relatives (reduced dataset), based on intergenomic distances calculated using the protein cluster content. **1.** Hierarchical clustering tree. Two support values, selective inference (si (98)) and approximately unbiased (au (99)) are indicated at branching points (si/au) only for the major clades (see SI file 1 Figure 16 and Figure 17 for all support values). The tree was cut into smaller viral genome clusters (VGCs) using a 0.9 distance threshold. Each VGC is framed in a rectangle in 2 and 3. **2.** Silhouette width, measures how related is a virus with other viruses in the same VGCs. Similarity to other VGCs is indicated by values closer to -1 (red). Similarity to viruses in the same VGC is indicated by values closer to 1 (green). **3.** Distribution of the protein clusters (PCs) in the viral genomes. **4.** Genome length (bps). **5.** Fraction of proteins shared with other viruses (dark grey), based on protein assignment to PCs. **6.** Virus names, with flavophages isolated in this study marked in red. **7.** TEM images of the new flavophages, uranyl acetate negative staining. Scale bar in each TEM image has 50 nm. **8.** Family (ICTV). **9.** Kingdom (ICTV). **10.** Realm (ICTV). Lighter colours in column 8- 10 represent phages not recognized by the ICTV, but by publications.

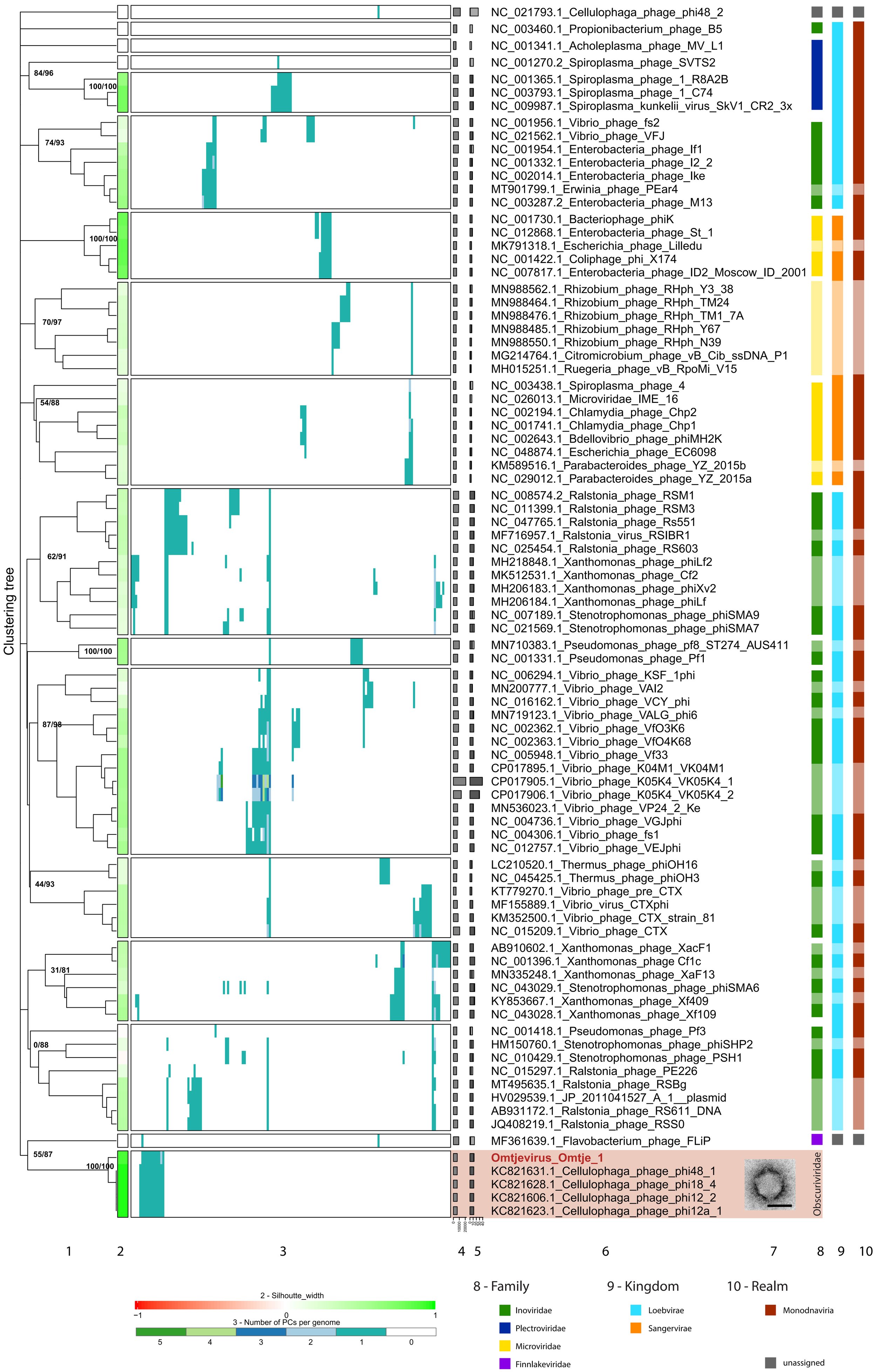
Figure 5: Genome maps of phage isolates with color-coded gene annotations

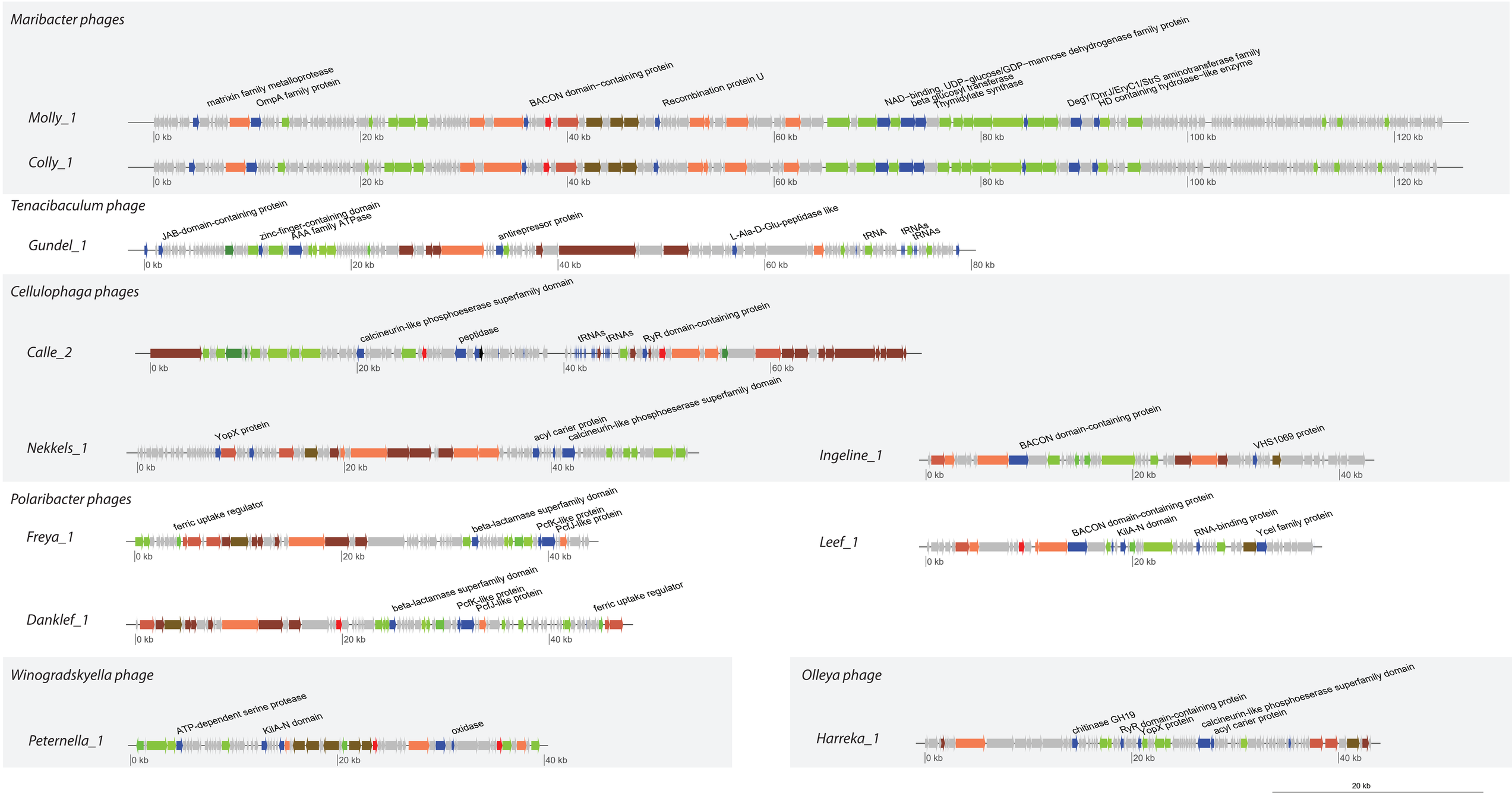


Gundel		○	○	●	○	○	○	●	○	*Tenacibaculum*
Calle		●	●	◐	●	●*				*Cellulophaga*
Nekkels		●	○	◐	◐	●*				*Cellulophaga*
Omtje	○					●	●	○	○	*Cellulophaga*
Ingeline	○		○			●	●	○	○	*Cellulophaga*
Leef								○	◉	*Polaribacter*
Danklef			○					●	◉	*Polaribacter*
Freya		●	●	○	○	●	○	○	◌	*Polaribacter*
Peternella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


Clustering tree
Staphylococcus_phage_vB_SauM_Romulus
Bacillus_phage_vB_BceM-HSE3
Mollyvirus Colly_1
Mollyvirus Molly_1
IMGVR2_3300001605_Draft_10001254
IMGVR2_3300005080_Ga0069611_10000213
IMGVR2_3300005080_Ga0069611_10000122
IMGVR2_3300015214_Ga0172382_10001576
IMGVR2_3300009508_Ga0115567_10000451
GOV2_Station168_DCM_ALL_NODE_1833
Leefvirus Leef_1
IMGVR2_3300001122_JGI12148J13107_100002
IMGVR2_3300012032_Ga0136554_1000067
Cellulophaga_phage_phi46_1
Flavobacterium_phage_vB_FspS_laban6-1
Flavobacterium_phage_FPSV-D15
Flavobacterium_phage_FPSV-D35
Flavobacterium_phage_FPSV-F7
Flavobacterium_phage_2A
Flavobacterium_phage_6H
Flavobacterium_phage_23T
Ingelinevirus Ingeline_1
IMGVR2_3300008250_Ga0105354_1000171
Marine_virus_AFVG_250M346
Marine_virus_AFVG_25M346
Marine_virus_AFVG_25M427
Marine_virus_AFVG_25M103
Marine_virus_AFVG_25M177
Marine_virus_AFVG_25M24
Callevirus Calle_1
Cellulophaga_phage_phi40_1
Cellulophaga_phage_phi38_1
Podoviridae_sp._ctrTa16
Flavobacterium_phage_vB_Fsp_elemo8-9A
Cellulophaga_phage_phi18_3
Cellulophaga_phage_phi19_3
Cellulophaga_phage_phi46_3
Cellulophaga_phage_phi13_2
IMGVR2_3300005056_Ga0071102_1000080
IMGVR2_3300001278_BBAY75_10000041
Gundelvirus Gundel_1
Bacteroides_phage_p00
Uncultured_Mediterranean_phage_uvMED_G21
Flavobacterium_phage_FPSV-S1
Flavobacterium_phage_FPSV-S8
Flavobacterium_phage_vB_FspM_pippi8-1
Flavobacterium_phage_vB_FspM_lotta8-1
Flavobacterium_phage_vB_FspM_lotta8-2
IMGVR2_3300014204_Ga0172381_10001225
Peternellavirus Peternella_1
IMGVR2_3300010054_Ga0098069_100157
IMGVR2_3300019758_Ga0193951_1000082
Flavobacterium_phage_fF4
IMGVR2_3300012252_Ga0122200_100146
IMGVR2_3300007126_Ga0102717_100371
IMGVR2_3300006459_Ga0100222_100241
IMGVR2_3300007093_Ga0104055_1000085
IMGVR2_3300007713_Ga0105659_1000065
IMGVR2_3300006742_Ga0101805_100074
IMGVR2_3300008130_Ga0114850_100310
IMGVR2_3300019755_Ga0193954_1000159
GOV2_Station138_MES_COMBINED_FINAL_NODE_1133
GOV2_Station76_MES_COMBINED_FINAL_NODE_1002
Freyavirus Danklef_1
Freyavirus Freya_1
GOV2_Station158_MES_ALL_assembly_NODE_1201
GOV2_Station158_SUR_ALL_assembly_NODE_1364
IMGVR2_3300009550_Ga0115013_10000204
Tenacibaculum_phage_JQ
IMGVR2_3300009684_Ga0114958_10000210
Uncultured_Mediterranean_phage_uvMED_G1
GOV2_Station82_SUR_COMBINED_FINAL_NODE_637
IMGVR2_3300006920_Ga0070748_1000149
IMGVR2_3300007541_Ga0099848_1000064
IMGVR2_3300012961_Ga0164302_10000013
Flavobacterium_phage_vB_FspS_morran9-1
Flavobacterium_phage_vB_FspS_hattifnatt9-1
Flavobacterium_phage_vB_FspS_filifjonk9-1
Flavobacterium_phage_vB_FspS_snork6-2
IMGVR2_3300009169_Ga0105097_10000088
IMGVR2_3300012031_Ga0136561_1000079
IMGVR2_3300012147_Ga0136588_1000088
Siphoviridae_sp._isolate_ctbc8
IMGVR2_3300014166_Ga0134079_10000028
IMGVR2_3300010364_Ga0134066_10000002
IMGVR2_3300018482_Ga0066669_10000002
IMGVR2_3300000271_PBR_1000291
IMGVR2_3300009678_Ga0105252_10000338
Bacteriophage_11b
Cellulophaga_phage_phi12_1
Cellulophaga_phage_phi12_3
Cellulophaga_phage_phi10_1
Nekkelsvirus Nekkels_1
Cellulophaga_phage_phi19_1
Polaribacter_phage_P12002L
Polaribacter_phage_P12002S
IMGVR2_3300012989_Ga0164305_10000040
IMGVR2_3300019760_Ga0193953_1000104
Harrekavirus Harreka_1
90/96
93/97
100/100
99/100
72/97
99/100
53/80
36/78
81/94
63/82
98/99
1/76
16/81
0/79
80/93
100/100
100/100
0/83
22/66
Molycol-viridae
Helgoland-viridae
Dunevirdae
Pervagoviridae
Pachy-viridae
Winoviridae
Forseti-viridae
Assidu-viridae
Aggrega-viridae
a b c
1 2 3 4 5 6 7 8 9 10 11 12
2 - Silhoutte_width
−1 0 1
3 - Number of Cs per genome
5 4 3 2 1 0
7 - Genera
Harrekavirus
Freyavirus
Nekkelsvirus
Gundelvirus
Ingelinevirus
Callevirus
Mollyvirus
Peternellavirus
Leefvirus
ICTV recognized
8 - Habitat
marine
limnic
waste water
soil/sediment
9 - Host association
Bacteroidetes
other phylum
Flavobacteriaceae
10 - Life style genes
LuxR
integrase
T transposase
none of the above


Clustering tree
84/96
100/100
74/93
100/100
70/97
54/88
62/91
100/100
87/98
44/93
31/81
0/88
55/87
100/100
NC_021793.1_Cellulophaga_phage_phi48_2
NC_003460.1_Propionibacterium_phage_B5
NC_001341.1_Acholeplasma_phage_MV_L1
NC_001270.2_Spiroplasma_phage_SVTS2
NC_001365.1_Spiroplasma_phage_1_R8A2B
NC_003793.1_Spiroplasma_phage_1_C74
NC_009987.1_Spiroplasma_kunkelii_virus_SkV1_CR2_3x
NC_001956.1_Vibrio_phage_fs2
NC_021562.1_Vibrio_phage_VFJ
NC_001954.1_Enterobacteria_phage_If1
NC_001332.1_Enterobacteria_phage_I2_2
NC_002014.1_Enterobacteria_phage_Ike
MT901799.1_Erwinia_phage_PEar4
NC_003287.2_Enterobacteria_phage_M13
NC_001730.1_Bacteriophage_phiK
NC_012868.1_Enterobacteria_phage_St_1
MK791318.1_Escherichia_phage_Lilledu
NC_001422.1_Coliphage_phi_X174
NC_007817.1_Enterobacteria_phage_ID2_Moscow_ID_2001
MN988562.1_Rhizobium_phage_RHph_Y3_38
MN988464.1_Rhizobium_phage_RHph_TM24
MN988476.1_Rhizobium_phage_RHph_TM1_7A
MN988485.1_Rhizobium_phage_RHph_Y67
MN988550.1_Rhizobium_phage_RHph_N39
MG214764.1_Citromicrobium_phage_vB_Cib_ssDNA_P1
MH015251.1_Ruegeria_phage_vB_RpoMi_V15
NC_003438.1_Spiroplasma_phage_4
NC_026013.1_Microviridae_IME_16
NC_002194.1_Chlamydia_phage_Chp2
NC_001741.1_Chlamydia_phage_Chp1
NC_002643.1_Bdellovibrio_phage_phiMH2K
NC_048874.1_Escherichia_phage_EC6098
KM589516.1_Parabacteroides_phage_YZ_2015b
NC_029012.1_Parabacteroides_phage_YZ_2015a
NC_008574.2_Ralstonia_phage_RSM1
NC_011399.1_Ralstonia_phage_RSM3
NC_047765.1_Ralstonia_phage_Rs551
MF716957.1_Ralstonia_virus_RSIBR1
NC_025454.1_Ralstonia_phage_RS603
MH218848.1_Xanthomonas_phage_phiLf2
MK512531.1_Xanthomonas_phage_Cf2
MH206183.1_Xanthomonas_phage_phiXv2
MH206184.1_Xanthomonas_phage_phiLf
NC_007189.1_Stenotrophomonas_phage_phiSMA9
NC_021569.1_Stenotrophomonas_phage_phiSMA7
MN710383.1_Pseudomonas_phage_pf8_ST274_AUS411
NC_001331.1_Pseudomonas_phage_Pf1
NC_006294.1_Vibrio_phage_KSF_1phi
MN200777.1_Vibrio_phage_VAI2
NC_016162.1_Vibrio_phage_VCY_phi
MN719123.1_Vibrio_phage_VALG_phi6
NC_002362.1_Vibrio_phage_VfO3K6
NC_002363.1_Vibrio_phage_VfO4K68
NC_005948.1_Vibrio_phage_Vf33
CP017895.1_Vibrio_phage_K04M1_VK04M1
CP017905.1_Vibrio_phage_K05K4_VK05K4_1
CP017906.1_Vibrio_phage_K05K4_VK05K4_2
MN536023.1_Vibrio_phage_VP24_2_Ke
NC_004736.1_Vibrio_phage_VGJphi
NC_004306.1_Vibrio_phage_fs1
NC_012757.1_Vibrio_phage_VEJphi
LC210520.1_Thermus_phage_phiOH16
NC_045425.1_Thermus_phage_phiOH3
KT779270.1_Vibrio_phage_pre_CTX
MF155889.1_Vibrio_virus_CTXphi
KM352500.1_Vibrio_phage_CTX_strain_81
NC_015209.1_Vibrio_phage_CTX
AB910602.1_Xanthomonas_phage_XacF1
NC_001396.1_Xanthomonas_phage Cf1c
MN335248.1_Xanthomonas_phage_XaF13
NC_043029.1_Stenotrophomonas_phage_phiSMA6
KY853667.1_Xanthomonas_phage_Xf409
NC_043028.1_Xanthomonas_phage_Xf109
NC_001418.1_Pseudomonas_phage_Pf3
HM150760.1_Stenotrophomonas_phage_phiSHP2
NC_010429.1_Stenotrophomonas_phage_PSH1
NC_015297.1_Ralstonia_phage_PE226
MT495635.1_Ralstonia_phage_RSBg
HV029539.1_JP_2011041527_A_1__plasmid
AB931172.1_Ralstonia_phage_RS611_DNA
JQ408219.1_Ralstonia_phage_RSS0
MF361639.1_Flavobacterium_phage_FLiP
Omtjevirus_Omtje_1
KC821631.1_Cellulophaga_phage_phi48_1
KC821628.1_Cellulophaga_phage_phi18_4
KC821606.1_Cellulophaga_phage_phi12_2
KC821623.1_Cellulophaga_phage_phi12a_1
Obscuriviridae
1
2
3
4
5
6
7
8
9
10
2 - Silhoutte_width
−1
0
1
3 - Number of PCs per genome
5
4
3
2
1
0
8 - Family
Inoviridae
Plectroviridae
Microviridae
Finnlakeviridae
9 - Kingdom
Loebvirae
Sangervirae
10 - Realm
Monodnaviria
unassigned


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