

**Part 1:** **TITLE, AUTHORS, APPROVALS, etc**

|  |  |  |
| --- | --- | --- |
| **Code assigned:** | **2021.029B** |  |
| **Short title:**  Create nine new families (*Pachyviridae, Pervagoviridae, Assiduviridae, Helgolandviridae, Duneviridae, Molycolviridae, Winoviridae, Forsetiviridae*, and *Aggregaviridae*) including 13 new genera and 18 new species (*Caudoviricetes*) | | |
|  | | |

**Author(s) and email address(es)**

|  |  |
| --- | --- |
| Bartlau N, Moraru C, Wichels A, Holmfeldt K, Adriaenssens EM, Amann R | nbartlau@mpi-bremen.de; liliana.cristina.moraru@uni-oldenburg.de; Antje.Wichels@awi.de; karin.holmfeldt@lnu.se; Evelien.Adriaenssens@quadram.ac.uk; ramann@mpi-bremen.de |

**Author(s) institutional address(es) (optional)**

|  |
| --- |
| Max Planck Institute for Marine Microbiology, Bremen, Germany [BAN, AMR]  Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Germany [MOC]  Alfred-Wegner Institute Helmholtz Centre for Polar and Marine Research, Biologische Anstalt Helgoland, Germany [WIA]  Centre for Ecology and Evolution in Microbial Model Systems (EEMiS), Department of Biology and Environmental Science, Faculty of Health and Life Sciences, Linnaeus University, Kalmar, Sweden [HOK]  Quadram Institute Bioscience, Norwich Research Park, Norwich, UK [ADE] |

**Corresponding author**

|  |
| --- |
| Nina Bartlau |

**List the ICTV Study Group(s) that have seen this proposal**

|  |
| --- |
| Bacterial Viruses Subcommittee |

**ICTV study group comments and response of proposer**

|  |
| --- |
|  |

**Authority to use the name of a living person**

|  |  |
| --- | --- |
| **Is any taxon name used here derived from that of a living person (Y/N)** | N |

|  |  |  |
| --- | --- | --- |
| **Taxon name** | **Person from whom the name is derived** | **Permission attached (Y/N)** |
|  |  |  |
|  |  |  |
|  |  |  |

**Submission dates**

|  |  |
| --- | --- |
| Date first submitted to SC Chair | May 2021 |
| Date of this revision (if different to above) |  |

**ICTV-EC comments and response of the proposer**

|  |
| --- |
| Acceptance of proposal 2021.001B.abolish\_Caudovirales by EC53 results in removal of the order *Caudovirales* and families *Myoviridae*, *Podoviridae* and *Siphoviridae*. All underlying taxa are to be assigned directly to the class *Caudoviricetes*. The Excel module of this proposal has been altered to reflect the future changes; however, the Word module has been unaltered while awaiting the ratification vote. |

**Part 3:** **TAXONOMIC PROPOSAL**

**Name of accompanying Excel module**

|  |
| --- |
| 2021.029B.R.Flavophages\_9\_families |

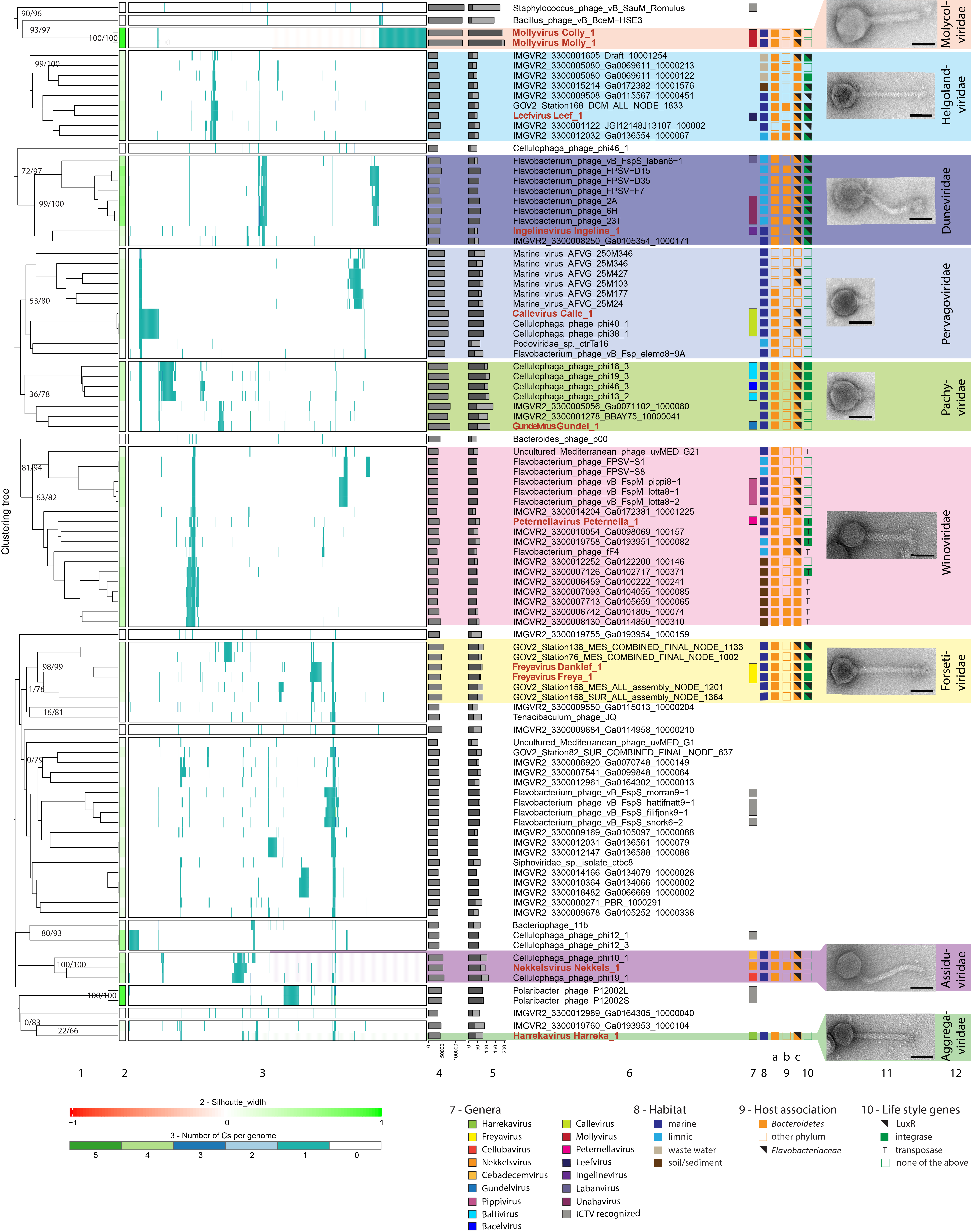
**Abstract**

|  |
| --- |
| We propose here nine new families in the order *Caudovirales*. The first family is called *Pachyviridae* and consists of the genus *Gundelvirus* with the species *Gundelvirus Gundel*, the genus *Baltivirus* with the species *Baltivirus phi19tres, Baltivirus phi13duo,* and *Baltivirus phi18tres*, and the genus *Bacelvirus* with the species *Bacelvirus phi46tres*. The second family *Pervagoviridae* consists of the genus *Callevirus* with *Callevirus Calle* and *Callevirus phi38una* as species. The third family is called *Assiduviridae*, which contains the three genera *Nekkelsvirus*, *Cebadecemvirus*, and *Cellubavirus* with the species *Nekkelsvirus Nekkels*, *Cebadecemvirus phi10una*, and *Cellubavirus phi19una*, respectively. The fourth family is called *Helgolandviridae* and has a single genus the *Leefvirus* with *Leefvirus Leef* as species. The fifth family *Duneviridae* has three genera the *Ingelinevirus* with the species *Ingelinevirus Ingline* the *Labanvirus* with the species *Labanvirus Laban*, and the *Unahavirus* with the four species *Unahavirus 1H*, *Unahavirus 6H,* *Unahavirus 2A,* and *Unahavirus 23T*. The sixth family is named *Molycolviridae* with the single genus *Mollyvirus*, which contains the species *Mollyvirus Molly* and *Mollyvirus Colly*. The seventh family is called *Winoviridae* with *Peternellavirus* as genus and the species *Peternellavirus Peternella*. The genus *Pippivirus* with its species *Pippivirus Lotta* and *Pippivirus Pippi* is moved to this family. The eighth family is called *Forsetivirida*e with *Freyavirus* as genus and *Freyavirus Freya* and *Freyavirus Danklef* as species. The ninth family is called *Aggregaviridae* with *Harrekavirus* as genus and *Harrekavirus Harreka* as species. |

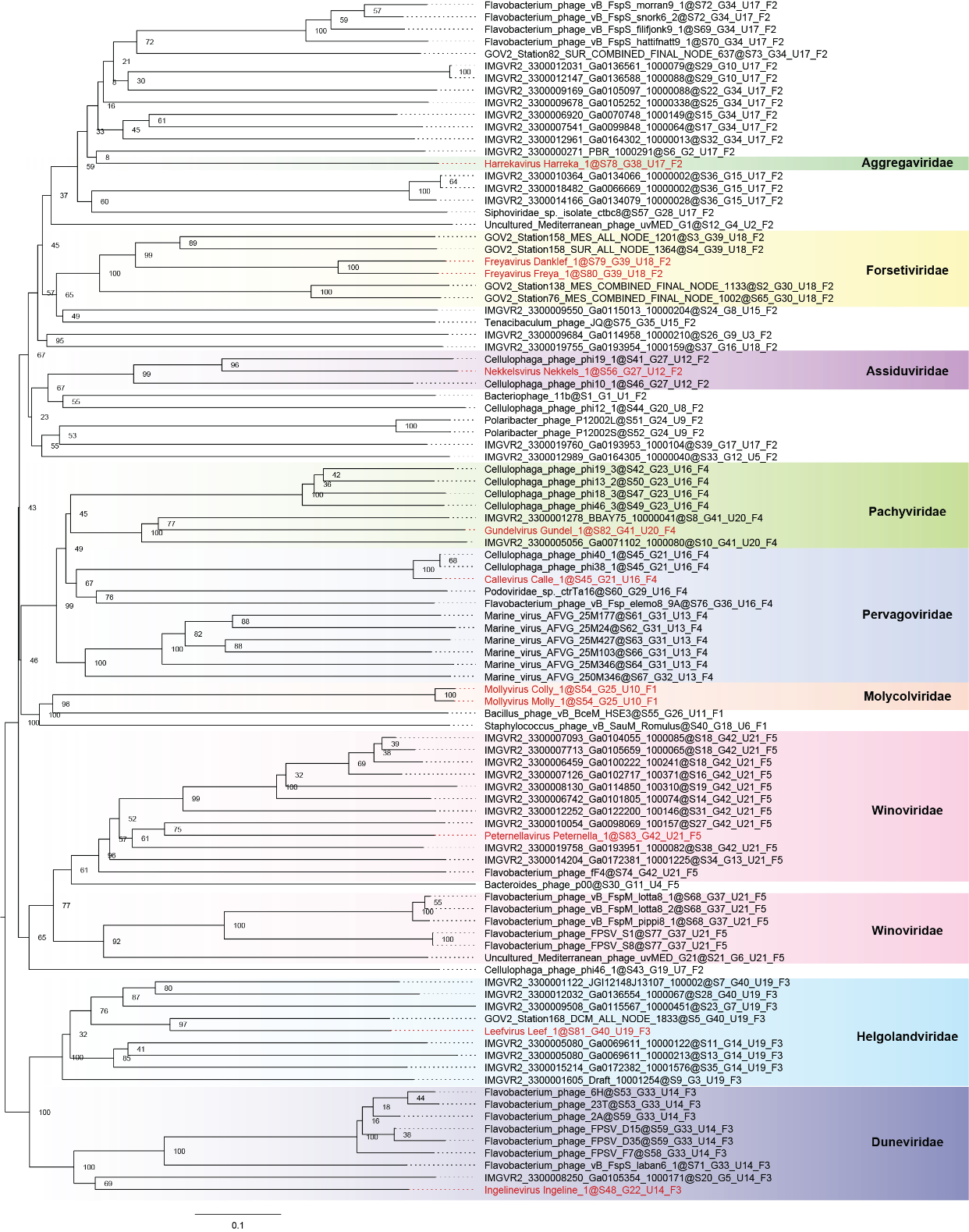
**Text of proposal**

|  |  |
| --- | --- |
| |  | | --- | | This taxonomic proposal aims to classify new phages infecting members of the class *Flavobacteriia* (shortly - flavophages)*,* as well as related previously cultured phages. The isolation and characterization of these new flavophages, as well as the retrieval of related phages from sequence databases is described in [2]. Most of the phages were isolated from surface seawater sampled during spring phytoplankton blooms at the long-term ecological research station Helgoland Roads.  The here proposed taxonomic assignments are based on five different methods: i) nucleotide-based intergenomic similarities calculated with the VIRIDIC web-service (viridic.icbm.de, [11]); ii) protein-based hierarchical clustering trees calculated with VirClust ([www.virclust.icbm.de](http://www.virclust.icbm.de)) , iii) amino acid-based whole genome trees calculated with the the Viral Proteomic Tree (ViPTree) ; iv) amino acid-based whole genome trees calculated with the Virus Classification and Tree Building Online Resource (VICTOR) [10]; and v) core-protein based trees using IQ-Trees [12] were calculated.  To calculate nucleotide-based intergenomic similarities, VIRIDIC was used with the default parameters.  To calculate protein-based hierarchical clustering trees VirClust was used with the following parameters: 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, aglomeration method “complete”, 1000 bootstraps, tree cut at a distance of 0.9. As reference for the ICTV dataset, the Master Species List release 35 (March 2020, <https://talk.ictvonline.org/files/master-species-lists/m/msl/9601>) [1] updated with the newly approved proposals for 2020 (<https://talk.ictvonline.org/files/master-species-lists/m/msl/12314>) was used.  To calculate phylogenetic trees with ViPTree, the command line tool was used, with default parameters.  To calculate the intergenomic distances and the phylogenetic trees, VICTOR was used with the d6 formula [9]. The resulted trees were visualized with FigTree (tree.bio.ed.ac.uk/software/figtree). In addition to phylogenetic trees, VICTOR uses the following predetermined distance thresholds to suggest taxon boundaries at the species, genus, subfamily and family level: 0.118980, 0.749680, 0.888940 and 0.985225, respectively [10]. These thresholds yield the highest agreement with the ICTV 2014 classification regarding the investigated taxonomic ranks [10], but not necessarily with the current ICTV recommendations.  The nucleotide-based thresholds for the rank of species and genus used in this proposal are 95% and 70%, as recommended by ICTV. Based on previous observations [3], the VICTOR threshold for subfamily matches the new ICTV recommendations for family, and were considered as such in this proposal.  The intra-familial structure and relationships was determined by concatenated core proteins, calculated with VirClust, aligned with MUSCLE (v3.8.425,[4]), manually trimmed and trees calculated IQ-Tree with SH-aLRT [12] and ultrafast boostrap values [5] using ModelFinder [8].  The following figures and descriptions are taken from the supplementary information of [2] and demarcation criteria are added. *Pachyviridae* family The *Pachyviridae* are named after the morphology of these phages being chubby and having no long attachments. It is coming from ancient Greek *παχύς* (pakhús, “thick”).  **Family demarcation criteria**: monophyletic group in VirClust and VICTOR, sharing 11 core genes (portal protein, sheath, structural protein (5), hypothetical protein (4), see protein clusters 1, 3 - 12 (Table 2, [2])). *Gundelvirus* genus The genus is named after its species *Gundelvirus Gundel*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Gundelvirus Gundel* species The *Gundelvirus Gundel* was isolated twice, before and after the phytoplankton peak in 2018, with *Tenacibaculum* sp. AHE14PA (DSM111040) and AHE15PA (DSM111039). It had a podoviral morphology with a capsid diameter of 60.5 ± 5.2 nm and a tail length of 22.7 ± 3.2 nm. Ten isolates were obtained, all having a genome size of 78,511 bp and a GC content of 30.4%. The circular genome had short direct terminal repeats (DTRs) at the ends. Gundel can only infect its two hosts of isolation, AHE14PA and AHE15PA, which had 99.87% similar 16S rRNA and an ANI of 99.99%. Gundel had genes coding for tail fiber, portal, L-alanine-D-glutamine-peptidase, and a phage antirepressor protein. In addition, Gundel had 10 tRNAs.  Gundel is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Baltivirus* genus The genus is named after the isolation source of the phages of this genus: the Baltic Sea.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Baltivirus phi19tres* species The *Baltivirus phi19tres* was isolated with *Cellulophaga* sp. #19 in 2005 from the Öresund strait [6]. It had a podoviral morphology and a genome size of 75,991 bp [7].  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Baltilvirus phi13duo* species The *Baltivirus phi13duo* was isolated with *Cellulophaga* sp. #13 in 2005 from the Öresund strait [6]. It had a genome size of 72,369 bp and a podoviral morphology [7].  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Baltivirus phi18tres* species The *Baltivirus phi18tres* was isolated in 2005 from the Öresund strait with the host *Cellulophaga* sp. #18 [6]. It had a podoviral morphology and a genome size of 71,442 bp [7].  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Bacelvirus* genus The genus is named after the Baltic Sea, which is the source of isolation and the host genus *Cellulophaga* of its species.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Bacelvirus phi46tres* species The *Bacelvirus phi46tres* was isolated with the host *Cellulophaga* sp. NN016046 in 2005 from the Öresund strait [6]. It had a genome size of 72,960 bp and a podoviral morphology [7].  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Pervagoviridae* family The name of the *Pervagoviridae* is derived from the Latin word *pervagus*, which is translated as widely roaming. Because the species *Callevirus phi38una* in the genus *Callevirus* is found in various ocean regions [13], this name was chosen.  **Family demarcation criteria**: monophyletic group in VirClust and VICTOR, sharing 8 core genes (portal protein, chaperonin cpn10, structural protein (4), hypothetical protein (2), see protein clusters 1, 2, 4, 5, 7-9, 13 (Table 2, [2])). *Callevirus* genus The genus is named after its species *Callevirus Calle*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Callevirus Calle* species With *Cellulophaga* sp. HaHa\_2\_95 (DSM111037) the podoviral *Callevirus Calle* was isolated. Three strains Calle\_1-3, which were 99.95-99.98% similar, belong to this species. This species was present throughout the 2018 phytoplankton bloom. It had a podoviral morphology with the capsid having a diameter of 60.3 ± 3.0 nm. The tail is 23.0 ± 5.5 nm long, and 13.5 ± 2.4 nm wide. The circular genome length of the strains ranged from 72,979 to 72,980 bp and the GC content was 38.1%. The genome encoded a capsid, tail, and DNA polymerase protein. In addition, it contained 20 tRNAs and a tRNA-splicing ligase RtcB. This phage also encoded a tmRNA. Both types of RNA are suggesting a more efficient phage replication and might increase the host range. Indeed, Calle is able to infect another *Cellulophaga* strain (HaHa\_2\_1). Two chaperonin proteins, which are associated with the GroEL system were also encoded in the genome. Its closest relative was the *Cellulophaga phage phi38:1* (KC821614.1) with 92.2% similarity, which was isolated in the Baltic Sea in 2005 [6] with a closely related host (99.5% 16S rRNA sequence identity and 94.2% ANI). Phi38:1 and Calle belong to an abundant cluster of marine phages [13].  Calle is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Callevirus phi38una* species The *Callevirus phi38una* species consists of two different strains *Cellulophaga phage phi38:1* and *Cellulophaga phage phi40:1*. Both were isolated in 2005 from the Öresund strait using *Cellulophaga* sp. NN016038 and Cellulophaga sp. NN015840 as hosts, respectively [6]. Both strains had a podoviral morphology and a genome size of 72,534 and 72,529 bp, respectively [7]. The two different strains share > 99.9% nucleotide similarity.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Assiduviridae* family The *Assiduviridae* are named after the Latin word *assiduous* with the meaning of constant and regular. The name is due to the regular isolation from enrichments and even from seawater.  **Family demarcation criteria:** monophyletic group in VICTOR and ViPTree, sharing 32 core genes (phage tail tape measure protein, DNA primase/helicase, NinG recombination protein, pectate lyase (2), structural protein, hypothetical protein (27), see protein clusters 6, 13, 14, 17, 21, 22, 23. 28 – 30, 41, 47, 48 – 55, 57 – 59, 61 – 68, 85 (Table 2, [2])). *Nekkelsvirus* genus The genus is named after its species *Nekkelsvirus Nekkels*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Nekkelsvirus Nekkels* species The *Nekkelsvirus Nekkels* was isolated by direct plating of sea surface water with the host *Cellulophaga* sp. HaHa\_2\_1 (DSM111038). Spot tests indicated that this phage group is present during most of the sampling period. The second strain, Nekkels\_2, had a nucleotide similarity of 97.1%. Nekkels had a capsid diameter of 54.8 ± 5.0 nm, its tail was 141.0 ± 7.9 nm long, and 13.0 ± 1.5 nm wide. The two genomes varied in length. Nekkels\_2 encoded in its 54,332 bp genome two ORFs more than Nekkels\_1 (53,385 bp). Both circular genomes have a 31.5% GC content. The genomes encoded a major capsid, tail tape measure, neck, terminase, pectate lyase, a lysozyme (GH19), and a potential spanin protein. Additional genes were the acyl carrier protein and a Yersinia outer protein X (YopX). Although Nekkels had a 40.5% nucleotide similarity with the *Cellulophaga phage phi19:1* (KC821607.1). Nekkels had a long tail and was morphological identified as siphoviral. Nekkels was able to infect another flavobacterial host: AHE13PA, a *Polaribacter* sp., but with a lower efficiency.  Nekkels is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Cebadecemvirus* genus This genus is named after the host of its species *Cellulophaga baltica* and the species itself phi10una, where *decem* is the Latin word for ten.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Cebadecemvirus phi10una* species The *Cebadecemvirus phi10una* was isolated with *Cellulophaga* sp. #10 in 2000 from the Öresund strait [6]. It had a siphoviral morphology and a genome size of 53,664 bp [7].  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Cellubavirus* genus This genus is named after the host of its species *Cellulophaga baltica.*  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Cellubavirus phi19una* species The *Cellubavirus phi19una* was isolated with *Cellulophaga* sp. #19 in 2000 [6]. It had a siphoviral morphology and a genome size of 57,447 bp [7].  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Helgolandviridae* family The family *Helgolandviridae* is named after the island Helgoland, where all samples of the study were taken.  **Family demarcation criteria**: monophyletic group in VirClust and VICTOR, sharing 6 core genes (hypothetical protein (6), see protein cluster 4, 5, 9-11, 13 (Table 2, [2])). *Leefvirus* genus The genus is named after its species *Leefvirus Leef*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Leefvirus Leef* species *Leefvirus Leef* infected *Polaribacter* sp. AHE13PA (DSM111061). Leef was isolated at the peak of the bacterial bloom. It had a capsid size of 49.2 ± 3.6 nm, a tail length of 138.7 ± 9.6 nm, and tail width of 11.1 ± 2.0 nm. The appearance was siphoviral. Its circular genome had Cos 3’ ends, a size of 37,547 bp and a GC content of 29.7%. Leef encoded proteins related to the HK97 Phage, a LuxR, a BACON (*Bacteroidetes*-Associated Carbohydrate-binding Often N-terminal) domain, a pectin lyase, and two integrases. Genes encoding capsid, tail tape measure, and neck proteins were annotated along with a terminase and a portal protein. An N-acetylmuramidase was detected in Leef, surrounded by transmembrane domains (TMDs) containing proteins. The closest relative to Leef was node 1833 from the GOV2 dataset, which was sampled in the Barents Sea. They were 30.4% similar.  Leef is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Duneviridae* family The family *Duneviridae* is named after the island close to Helgoland that is called Dune.  **Family demarcation criteria**: monophyletic group in VirClust and VICTOR, sharing 9 core genes (major capsid protein, portal protein, adaptor protein, hypothetical protein (6), see protein clusters 2-5, 7-9, 14, 15 (Table 2, [2])). *Ingelinevirus* genus The genus is named after its species *Ingelinevirus Ingeline*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Ingelinevirus Ingeline* species The *Ingelinevirus Ingeline* infected *Cellulophaga* sp. HaHaR\_3\_176 (DSM111152). This phage group is very diverse, as indicated by eight isolates, Ingeline 1-8, of a high nucleotide similarity above 99.9%. Ingeline\_7 and Ingeline\_8 were isolated in 2017 and 2018. Ingeline had a capsid diameter of 59.0 ± 5.3 nm and a tail, which is 132.9 ± 19.3 nm long, and 11.2 ± 1.7 nm wide. It had a circular genome ranging between 42,624 and 42,797 bp and a GC content of 32.2%. We annotated genes encoding a capsid, tail, tail tape measure, adaptor, portal, and a potential spanin protein. The morphology observed by TEM was that of a siphovirus. Interestingly, the genome also encodes a LuxR gene and a BACON domain-containing protein. Its closest environmental relative is Ga0105354\_1000171 from the Norwegian Sea with 12.4% similarity. Although Ingeline is lytic to its original *Cellulophaga* host, it contains two integrases, indicating the potential for lysogeny.  Ingeline is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Unahavirus* genus The genus is moved to the *Duneviridae*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Unahavirus 1H* species The species *Unahavirus 1H* was renamed according to the new binomial nomenclature. *Unahavirus 6H* species The species *Unahavirus 6H* was renamed according to the new binomial nomenclature. *Unahavirus 2A* species The species *Unahavirus 2A* was renamed according to the new binomial nomenclature. *Unahavirus 23T* species The species *Unahavirus 23T* was renamed according to the new binomial nomenclature. *Labanvirus* genus The genus is moved to the *Duneviridae.*  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Labanvirus Laban* species The species *Labanvirus Laban* was renamed according to the new binomial nomenclature. *Molycolviridae* family The name of the family *Molycolviridae* originates from the two species in the genus *Mollyvirus*: *Mollyvirus Molly* and *Mollyvirus Colly*.  **Family demarcation criteria:** monophyletic group in VirClust and VICTOR. *Mollyvirus* genus The genus is named after its species *Mollyvirus Molly*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Mollyvirus Molly* species Using the type strain *Maribacter forsetii* T (DSM18668) the *Mollyvirus Molly* was isolated from two time points in 2017. Six highly similar 99.56 to 99.99% strains (Molly1-7) were obtained. The phage has a capsid diameter of 74.9 ± 3.6 nm, a tail length of 101.5 ± 6.3 nm, and a tail width of 18.1 ± 1.6 nm. The morphology was myoviral-like. The circular genome had 124,169 to 124,898 bp with a GC% of 36.2. This phage group was very difficult to sequence, which might be due to a high degree of DNA modifications indicated by the respective genes. Furthermore, genes encoding a baseplate, tail fiber, tail, tail sheath, tape measure, neck, portal, major capsid protein, and a DNA polymerase I were identified. Additionally, it had a ribonucleotide reductase with two subunits A and B and a relatively short (199 aa) zinc-dependent metallopeptidase, formed from a lipoprotein domain and a peptidase domain.  Molly is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Mollyvirus Colly* species *Mollyvirus Colly* is 94% similar to *Mollyvirus Molly*. Following the ICTV guidelines it is a different species in the genus *Mollyvirus*. Colly had the same functional genes as Molly. The difference is due to genes encoding hypothetical proteins.  Colly is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Winoviridae* family The origin of the family *Winoviridae* is the host genus *Winogradyskyella* of the *Peternellavirus*.  **Family demarcation criteria:** monophyletic group in VirClust and VICTOR, sharing 10 core genes (Clp protease, neck protein, major capsid protein, sheath, phage protein D, head morphogenesis protein/portal protein, hypothetical protein (4), see protein clusters 1, 3, 4, 6, 8, 10, 11, 13, 15, 18 (Table 2, [2])). *Peternellavirus* genus The genus is named after its species *Peternellavirus Peternella*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Peternellavirus Peternella* species From *Winogradskyella* sp. HaHa\_3\_26 (DSM111041) *Peternellavirus Peternella* was isolated once in 2018 during the bacterial bloom. It had a capsid diameter of 52.3 ± 4.2 nm, a tail length of 105.8 ± 7.4 nm, and a tail width of 16.4 ± 2.4 nm. The morphology was myoviral. From four phage stocks, the same 39,649 bp long linear genome with 35.3% GC content was retrieved. This phage belongs to the Mu-like phages, due to its overlapping reads with the host genome and the typical genes like the MuA transposase and several structural proteins like the capsid, tail, tail fiber, baseplate, neck. This phage also encodes a portal, a holin, and a L-Alanine-D-glutamine-peptidase protein.  Peternella is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Pippivirus* genus The genus is moved to the *Winoviridae*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Pippivirus Lotta* species The species *Pippivirus Lotta* was renamed according to the new binomial nomenclature. *Pippivirus Pippi* species The species *Pippivirus Pippi* was renamed according to the new binomial nomenclature. *Flapsvirus* genus The genus is named after the host *Flavobacterium psychrophilum* of the species *Flapsvirus FPSV-S1*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Forsetiviridae* family The family is named after the god Forseti, which is associated to the island Helgoland in the Nordic mythology.  **Family demarcation criteria:** monophyletic group in VirClust and VICTOR, sharing 8 core genes (tape measure protein, major capsid protein, portal protein, structural proteins (5), see protein clusters 1-5, 7, 11, 17 (Table 2, [2])). *Freyavirus* genus The genus is named after its species *Freyavirus Freya*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Freyavirus Freya* species The *Freyavirus Freya* was isolated very early in the bloom with its host *Polaribacter* sp. HaHaR\_3\_91 (DSM111048). The ten strains obtained share a similarity of 94.68-99.05%. All genomes are circular, ranging from 43,978 up to 48,920 bp and had a GC content of 28.9%. The diameter of the capsid was 53.9 ± 4.7 nm, the tail was 151.0 ± 8.2 nm long, and 13.1 ± 2.0 nm wide. The morphology was siphoviral-like. Genomes varied in size by 11 genes, which were all annotated as hypothetical proteins. Genes encoding capsid, tail, tape measure, terminase, and portal proteins were annotated. Closely related sequences are from the Norwegian Sea and are 16.2% and 13.7% similar.  Freya is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Freyavirus Danklef* species *Freyavirus Danklef* with its five strains Danklef\_1-5 (similarity > 95.6%) were isolated late in the 2018 bloom, infecting *Polaribacter* sp. R2A056\_3\_33 (DSM111047). The capsid of Danklef was 46.1 ± 2.2 nm in diameter, the tail was 157.4 ± 4.6 nm long, and 12.1 ± 1.8 nm wide. Danklef had a siphoviral morphology. Its circular genome had a size of 47,177-47,426 bp with a GC content of 28.9%. Danklef’s major capsid protein was HK97-like. It had genes coding for a portal, tape measure, terminase, integrase, and N-acetylmuramidase protein. Interestingly, Danklef had a ferric uptake regulator family related gene. Danklef’s closest environmental relatives are coming from the Norwegian Sea and are 16.8 and 15.14% similar.  Danklef is a first name of Frisian origin, the language spoken on Helgoland.  Freya and Danklef were closely related (71.38%) and belong to the same genus. Comparing the two phages, Freya had a peptidase and a DNA replication protein, whereas Danklef had two endonucleases, a methylase, and an endolysin, which is not found in Freya.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. *Aggregaviridae* family The host of the *Harrekavirus Harreka* is mainly found on aggregates. Following this, this family is named *Aggregaviridae*.  **Family demarcation criteria:** monophyletic group in VirClust and VICTOR. *Harrekavirus* genus The genus is named after its species *Harrekavirus Harreka*.  **Genus demarcation criteria:** 70% intergenomic nucleotide similarity. *Harrekavirus Harreka* species The *Harrekavirus Harreka* was isolated during the bacterial bloom in 2018 with the host *Olleya* sp. HaHaR\_3\_96 (DSM111044). The capsid diameter was 44.3 ± 3.6 nm, the tail was 123.8 ± 8.0 nm long and 14.3 ± 2.0 nm wide. Harreka had a myoviral morphology. In total seven isolates were obtained and all of them revealed the 100% identical 43,175 bp circular genome with a GC content of 32%. Genes for capsid, tail, portal, replication proteins, GH19 and YopX were identified. Harreka was able to infect the two closely related *Tenacibaculum* strains AHE14PA and AHE15PA.  Harreka is a first name of Frisian origin, the language spoken on Helgoland.  **Species demarcation criteria:** 95% intergenomic nucleotide similarity. | |

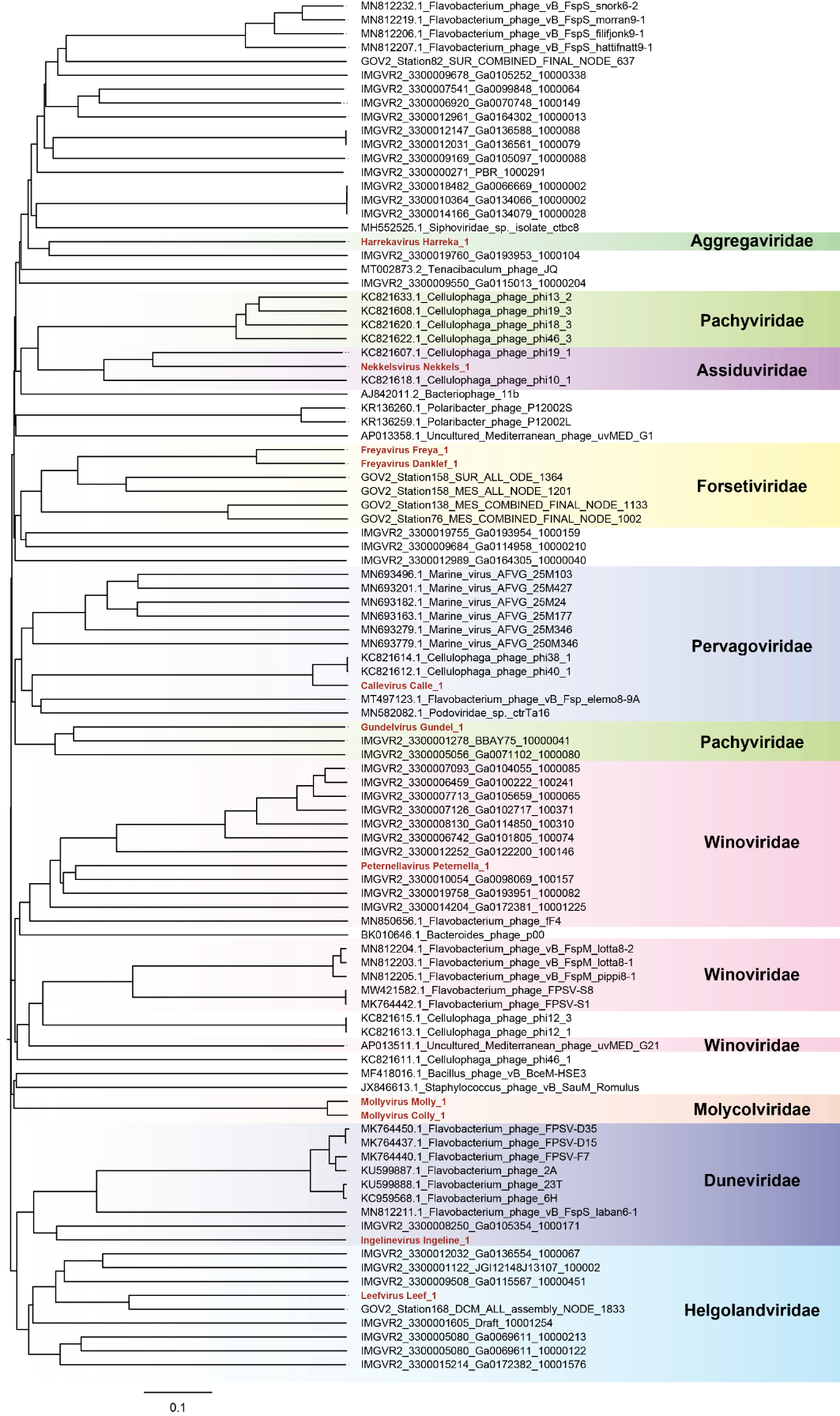
**Supporting evidence**



**Figure 1:** 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 Figure 13. 1. Hierarchical clustering tree. Two support values, selective inference (si, [14]) and approximately unbiased (au, [15]), are indicated at branching points (si/au) only for the major clades. 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. Genus assignment. Gray bars indicate already defined genera by the ICTV with the following names from top to bottom: *Silviavirus, Labanvius, Unahavirus, Pippivirus, Lillamyvirus, Muminvirus, Lillamyvirus, Helsingorviru*s, 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.

**

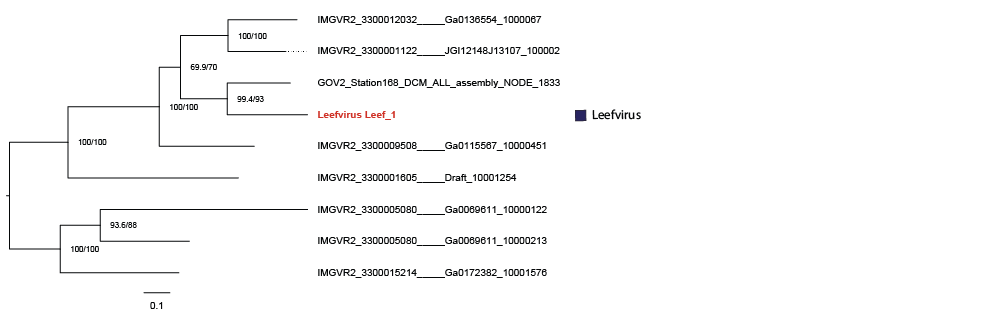
**Figure 2:** Whole-genome phylogeny determined with VICTOR (amino-acid based) for the dsDNA flavophages, including isolates and their relative phage genomes. Our new phage isolates are depicted in red. Pseudo-bootstrap values are indicated at branches. Family (F) and subfamily (U) clustering is indicated at the end of the genome names.

****

**Figure 3:** Whole-genome clustering determined with ViPTree (amino-acid based) for the dsDNA flavophages, including isolates and their relative phage genomes.



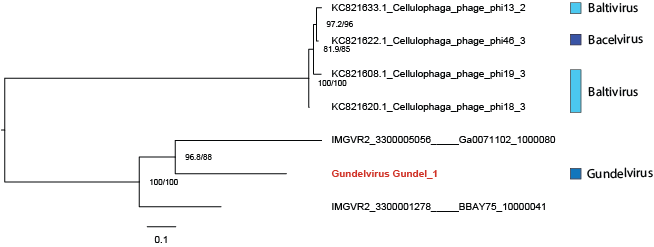
**Figure 4:** Core gene phylogeny of *Duneviridae* using MUSCLE aligned core proteins and IQ-Tree. This phylogeny is based on seven core genes which can be found in the annotation file of Ingeline in the following protein clusters: 3 (hp), 4 (hp), 5 (major capsid protein), 7 (hp), 8 (adaptor protein), 14 (hp), 15 (hp) (SI file 2). The first branch support value is the SH-aLRT support in %, the second value is the ultrafast bootstrap support. Using the model finder the VT+F+I substitution model was determined as best fitting substitution model and used for the tree calculation.



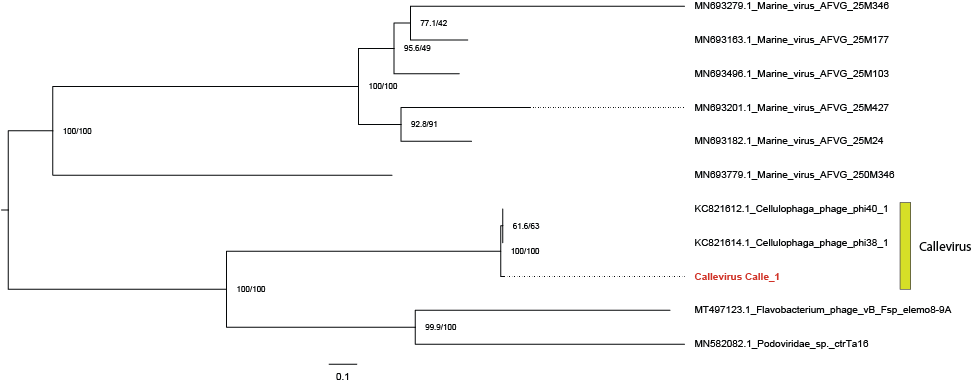
**Figure 5:** Core gene phylogeny of *Helgolandviridae* using MUSCLE aligned core proteins and IQ-Tree. This phylogeny is based on five core genes which can be found in the annotation file of Leef in the following protein clusters: 5 (hp), 9 (hp), 10 (hp), 11 (YceI family protein), 13 (hp) (SI file 2). The first branch support value is the SH-aLRT support in %, the second value is the ultrafast bootstrap support. Using the model finder the VT+F+I+G4 substitution model was determined as best fitting substitution model and used for the tree calculation.



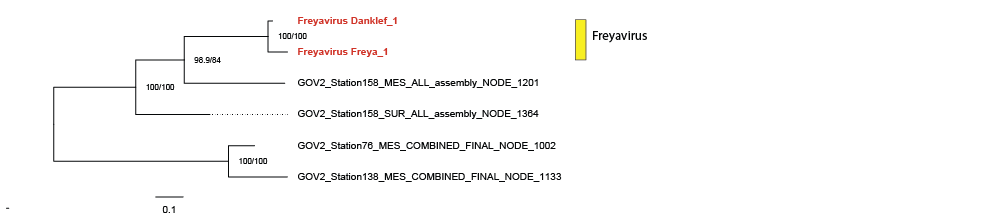
**Figure 6;** Core gene phylogeny of *Winoviridae* using MUSCLE aligned core proteins and IQ-Tree. This phylogeny is based on nine core genes which can be found in the annotation file of Peternella in the following protein clusters: 3 (major capsid protein), 4 (Clp protease), 6 (DUF2586/sheath), 8 (hp), 10 (Mu-like prophage protein gpG/neck), 11 (hp), 13 (phage protein D), 15 (oxidase), 18 (nucleotidyltransferase) (SI file 2). The first branch support value is the SH-aLRT support in %, the second value is the ultrafast bootstrap support. Using the model finder the VT+F+I+G4 substitution model was determined as best fitting substitution model and used for the tree calculation.



**Figure 7:** Core gene phylogeny of *Pachyviridae* using MUSCLE aligned core proteins and IQ-Tree. This phylogeny is based on nine core genes which can be found in the annotation file of Gundel in the following protein clusters: 4 (hp), 5 (structural protein), 6 (structural protein), 7 (hp), 8 (portal protein), 9 (structural protein), 10 (structural protein), 11 (hp), 12 (hp) (SI file 2). The first branch support value is the SH-aLRT support in %, the second value is the ultrafast bootstrap support. Using the model finder the VT+F+G4 substitution model was determined as best fitting substitution model and used for the tree calculation.

**

**Figure 8:** Core gene phylogeny of *Pervagoviridae* using MUSCLE aligned core proteins and IQ-Tree. This phylogeny is based on three core genes which can be found in the annotation file of Calle in the following protein clusters: 4 (chaperonin cpn10), 8 (structural protein), 13 (structural protein) (SI file 2). The first branch support value is the SH-aLRT support in %, the second value is the ultrafast bootstrap support. Using the model finder the LG+F+G4 substitution model was determined as best fitting substitution model and used for the tree calculation.

**

**Figure 9:** Core gene phylogeny of *Pervagoviridae* using MUSCLE aligned core proteins and IQ-Tree. This phylogeny is based on seven core genes which can be found in the annotation file of Freya in the following protein clusters: 1 (tape measure protein), 3 (structural protein), 4 (portal protein), 5 (structural protein), 7(structural protein), 11 (structural protein), 17 (major capsid protein) (SI file 2). The first branch support value is the SH-aLRT support in %, the second value is the ultrafast bootstrap support. Using the model finder the VT+F+I substitution model was determined as best fitting substitution model and used for the tree calculation.

**Table 1:** Additional strains belonging to the same species, classified in this proposal.

|  |  |  |
| --- | --- | --- |
| **Species** | **Exemplar isolate** | **Additional genomes belonging to this species** |
| *Callevirus phi38una* | Cellulophaga phage phi38:1 | KC821612 |
| *Inglinevirus Ingeline* | Cellulophaga phage Ingeline\_1 | MT732436, MT732437, MT732438, MT732439, MT732440, MT732441, MT732442 |
| *Mollyvirus Molly* | Maribacter phage Molly\_1 | MT732452, MT732453, MT732454, MT732455, MT732456 |
| *Callevirus Calle* | Cellulophaga phage Calle\_1 | MT732433, MT732434 |
| *Nekkelsvirus Nekkels* | Cellulophaga phage Nekkels\_1 | MT732444 |
| *Freyavirus Freya* | Polaribacter phage Freya\_1 | MT732464, MT732465, MT732466, MT732467, MT732468, MT732469, MT732470, MT732471, MT732472 |
| *Freyavirus Danklef* | Polaribacter phage Danklef\_1 | MT732459, MT732460, MT732461, MT732462 |

**Table 2:** Core genes used for family demarcation. \* indicate duplicated core genes, which were not included in the core protein trees. Genes of new isolates with their position in the genome including their annotation are depicted. In addition the protein clusters according [2] are indicated.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Family** | **Phage\_gene** | **Protein cluster** | **Start** | **End** | **Annotation** |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_43 | 15 | 34360 | 36519 | Hypothetical protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_44 | 4 | 36585 | 37406 | Hypothetical protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_47 | 14 | 38902 | 39807 | Hypothetical protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_42 | 5 | 33506 | 34336 | Major capsid protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_2 | 9\* | 501 | 1844 | Portal protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_5 | 2\* | 3083 | 3988 | Hypothetical protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_3 | 8 | 1844 | 2770 | Adaptor protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_6 | 7 | 3981 | 4439 | Hypothetical protein |
| *Duneviridae* | Cellulophaga\_phage\_Ingeline\_1\_gene\_7 | 3 | 4602 | 4982 | Hypothetical protein |
| *Helgolandviridae* | Polaribacter\_phage\_Leef\_1\_gene\_47 | 9 | 35385 | 35966 | Hypothetical protein |
| *Helgolandviridae* | Polaribacter\_phage\_Leef\_1\_gene\_46 | 10 | 34630 | 35388 | Hypothetical protein |
| *Helgolandviridae* | Polaribacter\_phage\_Leef\_1\_gene\_42 | 11 | 32000 | 32998 | YceI family protein |
| *Helgolandviridae* | Polaribacter\_phage\_Leef\_1\_gene\_39 | 13 | 29520 | 29993 | Hypothetical protein |
| *Helgolandviridae* | Polaribacter\_phage\_Leef\_1\_gene\_48 | 5 | 35950 | 37464 | Hypothetical protein |
| *Helgolandviridae* | Polaribacter\_phage\_Leef\_1\_gene\_44 | 4\* | 33666 | 34229 | Hypothetical protein |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_33 | 1\* | 15678 | 16841 | head morphogenesis protein, SPP1 gp7 family /portal protein |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_40 | 3 | 22360 | 23367 | major capsid protein |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_39 | 4 | 21135 | 22277 | Clp protease/major capsid protein |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_44 | 6 | 24627 | 25859 | DUF2586/sheath |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_45 | 8 | 25862 | 26305 | Hypothetical protein |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_30 | 10 | 14866 | 15378 | Mu-like prophage protein gpG /Neck |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_49 | 13 | 29506 | 30477 | phage protein D |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_51 | 15 | 31036 | 31317 | oxidase |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_58 | 18 | 35955 | 36866 | nucleotidyltransferase |
| *Winoviridae* | Winogradskyella\_phage\_Peternella\_1\_gene\_48 | 11 | 28895 | 29509 | Hypothetical protein |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_9 | 7\* | 11467 | 12807 | structural protein |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_11 | 2\* | 14114 | 16600 | portal protein |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_12 | 5\* | 16605 | 19244 | Hypothetical protein |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_5 | 1\* | 4970 | 8944 | structural protein |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_6 | 8 | 8950 | 9801 | structural protein |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_7 | 13 | 9813 | 10511 | structural protein |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_13 | 4 | 19241 | 19825 | chaperonin cpn10 |
| *Pervagoviridae* | Cellulophaga\_phage\_Calle\_2\_gene\_8 | 9\* | 10597 | 11397 | Hypothetical protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_101 | 12 | 68927 | 69292 | DUF3134 |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_57 | 4 | 26322 | 27089 | Hypothetical protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_55 | 9 | 24653 | 26053 | structural protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_77 | 8 | 50209 | 52674 | portal protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_76 | 7 | 47889 | 50159 | Hypothetical protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_60 | 1\* | 28729 | 32889 | sheath |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_74 | 3\* | 40108 | 47556 | structural protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_59 | 6 | 27920 | 28729 | structural protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_58 | 5 | 27226 | 27918 | structural protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_71 | 11 | 37380 | 37790 | Hypothetical protein |
| *Pachyviridae* | Tenacibaculum\_phage\_Gundel\_1\_gene\_72 | 10 | 37889 | 38566 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_25 | 1 | 14887 | 18426 | tape measure protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_26 | 2\* | 18416 | 20779 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_28 | 3 | 21336 | 22517 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_13 | 4 | 6934 | 8337 | portal protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_15 | 5 | 8458 | 9300 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_18 | 7 | 11318 | 11902 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_19 | 11 | 11899 | 12432 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Freya\_1\_gene\_16 | 17 | 9305 | 11005 | major capsid protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_14 | 1 | 8376 | 11915 | tape measure protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_15 | 2\* | 11905 | 14268 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_17 | 3 | 14825 | 16006 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_2 | 4 | 432 | 1835 | portal protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_4 | 5 | 1953 | 2783 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_7 | 7 | 4803 | 5387 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_8 | 11 | 5384 | 5917 | structural protein |
| *Forsetiviridae* | Polaribacter\_phage\_Danklef\_1\_gene\_5 | 17 | 2787 | 4487 | major capsid protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_86 | 6 | 46984 | 47691 | NinG recombination protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_71 | 13 | 41077 | 42336 | calcineurin-like phosphoesterase superfamily domain protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_85 | 14 | 46542 | 46919 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_55 | 17 | 33026 | 35002 | pectate lyase |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_32 | 21 | 11784 | 12125 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_90 | 22 | 49934 | 51823 | DNA primase/helicase, TOPRIM |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_53 | 23 | 29102 | 30568 | structural protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_11 | 28 | 3689 | 4102 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_18 | 29 | 6232 | 6438 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_54 | 30 | 30576 | 33023 | pectate lyase |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_89 | 41 | 49095 | 49931 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_69 | 47 | 40422 | 40601 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_70 | 48 | 40598 | 41071 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_73 | 49 | 43082 | 43405 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_74 | 50 | 43405 | 43551 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_77 | 51 | 44010 | 44288 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_79 | 52 | 44451 | 44693 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_91 | 53 | 51834 | 52064 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_93 | 54 | 53009 | 53233 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_2 | 55 | 225 | 533 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_5 | 57 | 1382 | 1519 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_7 | 58 | 2036 | 2530 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_9 | 59 | 3207 | 3365 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_14 | 61 | 5207 | 5377 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_16 | 62 | 5720 | 5929 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_30 | 63 | 11287 | 11589 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_33 | 64 | 12263 | 12451 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_56 | 65 | 35014 | 35427 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_57 | 66 | 35746 | 36078 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_58 | 67 | 36143 | 36772 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_62 | 68 | 37915 | 38163 | Hypothetical protein |
| *Assiduviridae* | Cellulophaga\_phage\_Nekkels\_1\_gene\_80 | 85 | 44693 | 44812 | Hypothetical protein |

**Appendix**

**Figure 10:**VirClust hierarhical clustering of the new dsDNA flavophages and their relatives, extended dataset, based on intergenomic distances calculated using the protein cluster content. **1.** Hierarchical clustering tree. 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.** Silhoutte 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**. VGCs - numbering. **4.** Family (ICTV). **5.** Subfamily (ICTV). **6.** Genus (ICTV). **7.** Species (ICTV). **8.** Genome length (bps). **9.** Fraction of proteins shared with other viruses (dark grey), based on protein assignment to PCs. **10.** Virus genome names.

**To note: the new flavophages are not yet assigned to any ICTV recognized family.**

**Figure 11:** Intergenomic nucleotide similarity matrix of the obtained isolates and their relatives. Phage species are indicated on the right side.

**References**

1. Adriaenssens EM, Sullivan MB, Knezevic P, van Zyl LJ, Sarkar BL, Dutilh BE, Alfenas-Zerbini P, Łobocka M, Tong Y, Brister JR, Moreno Switt AI, Klumpp J, Aziz RK, Barylski J, Uchiyama J, Edwards RA, Kropinski AM, Petty NK, Clokie MRJ, Kushkina AI, Morozova VV, Duffy S, Gillis A, Rumnieks J, Kurtböke İ, Chanishvili N, Goodridge L, Wittmann J, Lavigne R, Jang HB, Prangishvili D, Enault F, Turner D, Poranen MM, Oksanen HM, Krupovic M (2020) Taxonomy of prokaryotic viruses: 2018-2019 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. Archives of Virology 165:1253-1260, PMCID:32162068, Doi:10.1007/s00705-020-04577-8

2. Bartlau N, Wichels A, Krohne G, Adriaenssens EM, Heins A, Fuchs BM, Amann R, Moraru C (2021) Highly diverse flavobacterial phages isolated from North Sea spring blooms. bioRxiv:2021.2005.2020.444936, Doi:10.1101/2021.05.20.444936

3. Barylski J, Enault F, Dutilh BE, Schuller MB, Edwards RA, Gillis A, Klumpp J, Knezevic P, Krupovic M, Kuhn JH, Lavigne R, Oksanen HM, Sullivan MB, Jang HB, Simmonds P, Aiewsakun P, Wittmann J, Tolstoy I, Brister JR, Kropinski AM, Adriaenssens EM (2019) Analysis of Spounaviruses as a Case Study for the Overdue Reclassification of Tailed Phages. Systematic Biology 69:110-123, PMCID:PMC7409376, Doi:10.1093/sysbio/syz036

4. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797, PMCID:PMC390337, Doi:10.1093/nar/gkh340

5. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2017) UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35:518-522, PMCID:PMC5850222, Doi:10.1093/molbev/msx281

6. Holmfeldt K, Middelboe M, Nybroe O, Riemann L (2007) Large Variabilities in Host Strain Susceptibility and Phage Host Range Govern Interactions between Lytic Marine Phages and Their Flavobacterium Hosts. Applied and Environmental Microbiology 73:6730-6739, PMCID:PMC2074958, Doi:10.1128/aem.01399-07

7. Holmfeldt K, Solonenko N, Shah M, Corrier K, Riemann L, VerBerkmoes NC, Sullivan MB (2013) Twelve previously unknown phage genera are ubiquitous in global oceans. Proceedings of the National Academy of Sciences 110:12798-12803, PMCID:PMC3732932, Doi:10.1073/pnas.1305956110

8. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587-589, PMCID:PMC5453245, Doi:10.1038/nmeth.4285

9. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60, PMCID:PMC3665452, Doi:10.1186/1471-2105-14-60

10. Meier-Kolthoff JP, Göker M (2017) VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33:3396-3404, PMCID:PMC5860169, Doi:10.1093/bioinformatics/btx440

11. Moraru C, Varsani A, Kropinski AM (2020) VIRIDIC-A novel tool to calculate the intergenomic similarities of prokaryote-infecting Viruses. Viruses 12:1268, PMCID:PMC7694805, Doi:10.3390/v12111268

12. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2014) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32:268-274, PMCID:PMC4271533, Doi:10.1093/molbev/msu300

13. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan MB, Tara Oceans C (2016) Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature 537:689-693, PMCID:27654921, Doi:10.1038/nature19366

14. Shimodaira H, Terada Y (2019) Selective inference for testing trees and edges in phylogenetics. Frontiers in Ecology and Evolution 7Doi:10.3389/fevo.2019.00174

15. Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540-1542, Doi:10.1093/bioinformatics/btl117