

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

|  |  |  |
| --- | --- | --- |
| **Code assigned:** | ***2023.009F*** |  |
| **Short title:** Establishing formal demarcation criteria for the *Maveriviricetes* class (“virophages”), and creating or re-assigning 4 orders, 7 families, 7 genera, and 8 species within this class | | |
|  | | |

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

|  |  |
| --- | --- |
| Roux S, Yutin N, Fischer M, Schulz F | siroux1@gmail.com; yutin@ncbi.nlm.nih.gov; fschulz@lbl.gov; mfischer@mr.mpg.de |

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

|  |
| --- |
| DOE Joint Genome Institute, Lawrence Berkeley National Laboratory [SR, FS]  National Center for Biotechnology Information, National Library of Medicine [NY]  Max Planck Institute for Medical Research, Department of Biomolecular Mechanisms [MF] |

**Corresponding author**

|  |
| --- |
| Simon Roux |

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

|  |
| --- |
| Virophage SG |

**ICTV Study Group comments and response of proposer**

|  |
| --- |
|  |

**ICTV Study Group votes on proposal**

|  |  |  |  |
| --- | --- | --- | --- |
| **Study Group** | **Number of members** | | |
| **Votes support** | **Votes against** | **No vote** |
| Virophage SG | 4 | 0 |  |
|  |  |  |  |

**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 | 21st June 2023 |
| Date of this revision (if different to above) |  |

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

|  |
| --- |
|  |

**Part 2:** **NON-TAXONOMIC PROPOSAL**

**Text of proposal**

|  |
| --- |
|  |

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

**Name of accompanying Excel module**

|  |
| --- |
| 2023.009F.v1.Virophages\_reorg.xlsx |

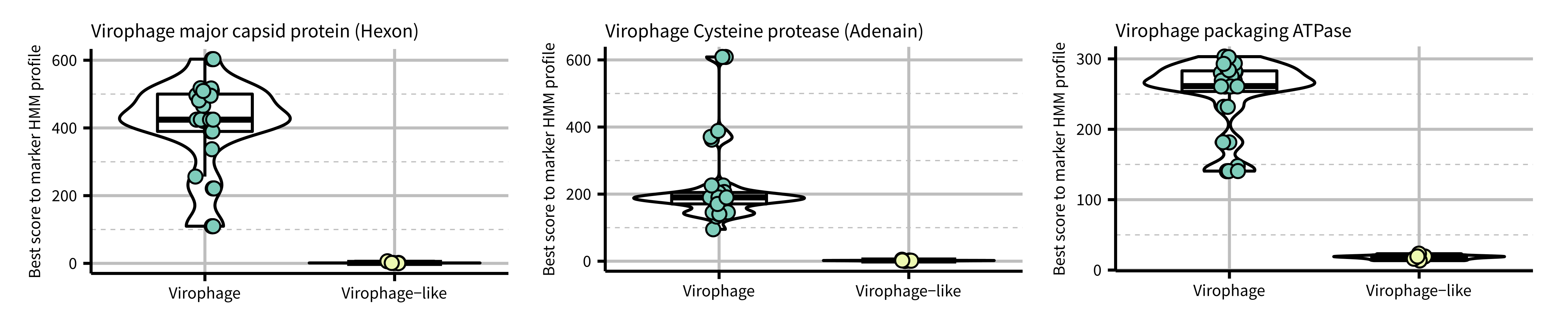
**Abstract**

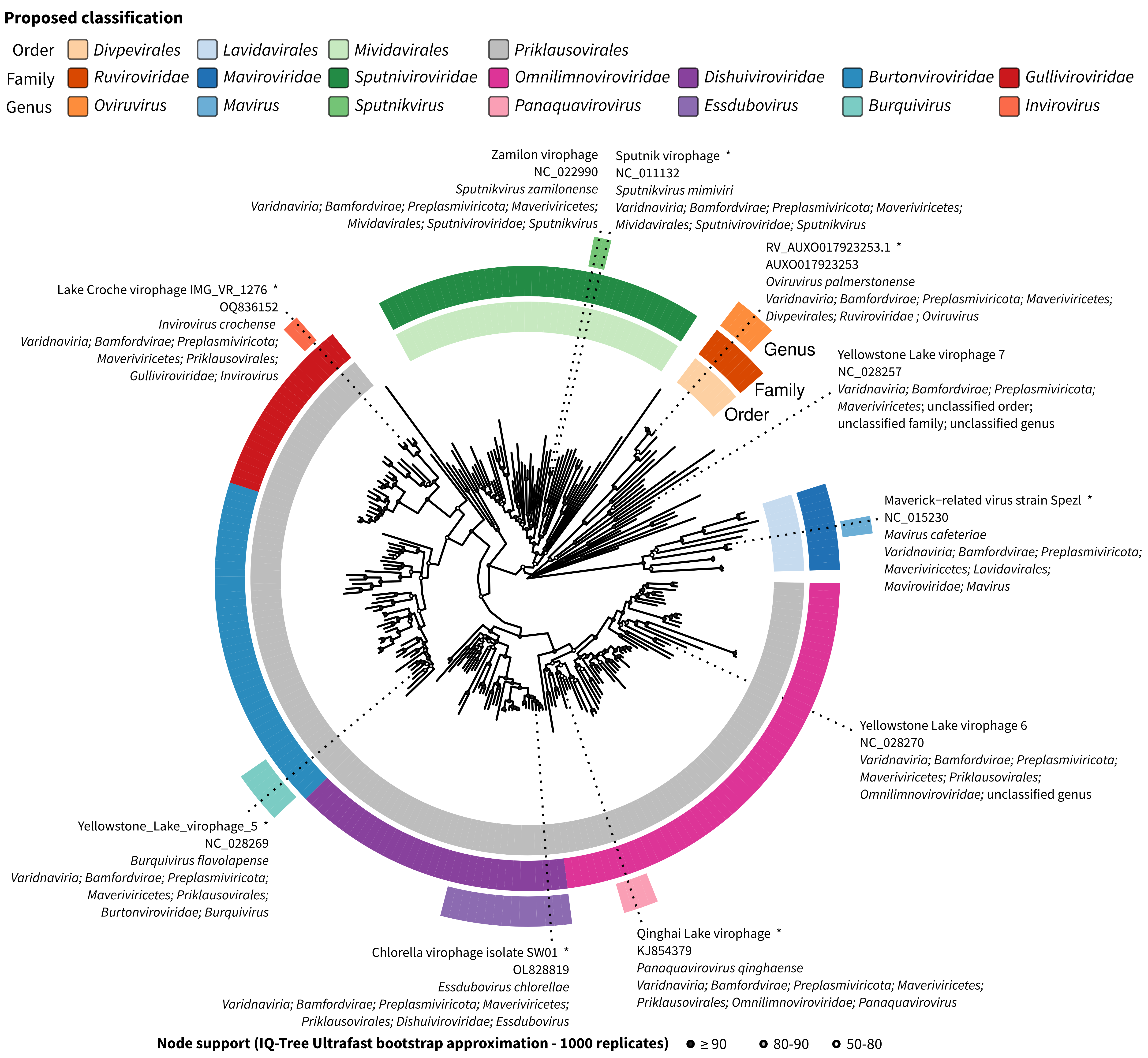
|  |
| --- |
| Virophages, dsDNA viruses dependent on giant viruses for their replication, were recently reclassified in the *Maveriviricetes* class, part of the *Preplasmiviricota* phylum which also includes the related *Polintoviricetes* and *Tectiliviricetes* classes. There are currently no clear demarcation criteria to determine whether a new virus belongs to the *Maveriviricetes*, and classification within this taxon has not been updated since 2015, when the *Lavidaviridae* family and the two virophage genera (*Mavirus* and *Sputnikvirus*)were originally created. Here we leverage a recent analysis of isolate- and metagenome-derived virophage sequences to propose (i) formal delineation criteria for the *Maveriviricetes* based on the detection of 3 core genes, (ii) updated definition and/or creation for 4 orders, 7 families, 5 genera, and 5 species based on a phylogenomic analysis of all available virophage genomes, and (iii) updated species name for currently recognized virophage species to conform to the Latin binomial format. We also provide bioinformatic resources to help users assign new virophage sequences in this updated framework (https://github.com/simroux/ICTV\_VirophageSG). |

**Text of proposal**

|  |  |
| --- | --- |
| |  | | --- | | Virophages are dsDNA viruses that depend for their replication on the co-infection of a eukaryotic host by giant virus(es) from the *Megaviricetes* class. Two distinct groups of virophages were originally isolated, and led to the identification of virophages as unique viral entities, are currently classified in two different genera. Sputnik, classified in the *Sputnikvirus* genus, replicates in *Acanthamoeba polyphaga* cells that are co-infected with Acanthamoeba polyphaga mimivirus (APMV) [1, 2]. Mavirus, classified in the *Mavirus* genus, replicates in the marine heterotrophic nanoflagellate *Cafeteria roenbergensis* in the presence of Cafeteria roenbergensis virus (CroV) [3, 4]. In a 2015 proposal (2015.001a-kF), these two genera were grouped together in the *Lavidaviridae* family.  Following the description of these isolates, thousands of virophage-like genomes, partial or complete, were identified in various metagenomes (see e.g. [5, 6, 7, 8]). These metagenome-derived sequences vastly expanded our knowledge about the diversity of genes encoded by virophages, including several instances of virophage-like genomes closely related to polintons. In addition, virophage-like sequences were also found to be integrated into the genome of the marine flagellate *Cafeteria burkhardae* [9]. These endogenous virophages were shown to reactivate upon superinfection with the giant virus CroV, and thus likely provide adaptive defense against giant viruses.  Most “virophage-like” sequences reported so far, either from viral isolates, metagenomes, or from endogenous forms integrated into protist genomes, have circular or linear double-stranded DNA genomes of similar length (~ 15 - 40kbp). The virions of isolated virophages are icosahedral with a diameter of 60-75 nm, and are composed of at least two different proteins with jelly-roll folds, the major capsid protein (“hexon” protein) and the penton protein. In a recent “mega-proposal” (2019.003G), based on their dsDNA genomes, genome size, and “jelly-roll” major capsid proteins, members of the *Lavidaviridae* family were assigned to the *Varidnaviria* realm, *Bamfordvirae* kingdom, *Preplasmiviricota* phylum, *Maveriviricetes* class, and *Priklausovirales* order. However, there are no formal demarcation criteria for many of these taxa (especially for members of the new class and order). Among the large diversity of virophage-like sequences assembled from metagenomes, it is thus unclear which ones belong to currently recognized taxa, and which should be the basis for new ones. In addition, the virophage taxonomy at lower ranks (i.e. below the *Priklausovirales* order), originally defined in 2015, should now be updated in light of new metagenome-derived genomes and to account for the recent description of a third virophage isolate, Chlorella virophage isolate SW01 [10].  Here, we leverage a recent comparative genomic analysis of isolate and metagenome-derived virophage genomes [11] to propose the following changes to the virophage taxonomy:   * Establish demarcation criteria for the *Maveriviricetes* class, in particular to provide a better distinction from the related *Polintoviricetes* * Create demarcation criteria for 3 additional orders (*Lavidavirales*, *Mividavirales*, *Divpevirales*) on the basis of a phylogenomic analysis of 257 complete and near-complete virophage genomes, in addition to the current *Priklausovirales* order * Replace the current *Lavidaviridae* family with 7 new families (*Maviroviridae*, *Sputniviroviridae*, *Dishuiviroviridae*, *Omnilimnoviroviridae*, *Burtonviroviridae*, *Gulliviroviridae*, *Ruviroviridae*) classified in these 4 orders, and provide associated demarcation criteria * Establish 5 new genera and species (*Essdubovirus chlorellae*, *Panaquavirovirus qinghaense*, *Burquivirus flavolapense*, *Invirovirus crochense*, *Oviruvirus palmerstonense*), to represent 5 new families, based on criteria updated from TaxoProp 2015.001a-kF * Rename the 3 existing virophage species (*Cafeteriavirus-dependent mavirus* to *Mavirus cafeteriae*, *Mimivirus-dependent virus Sputnik* to *Sputnikvirus mimiviri*, *Mimivirus-dependent virus Zamilon* to *Sputnikvirus zamilonense*)and re-assign them to families *Maviroviridae* and *Sputniviroviridae.*   ## Maveriviricetes delineation criteria  Based on a comparative analysis of current virophage and virophage-like sequences, we propose to update the delineation criteria for the *Maveriviricetes* classas follows:   * required features: the complete genome should encode a virophage-like hexon protein, a virophage-like ATPase, and a virophage-like cysteine protease, all of which can be detected based on established HMM profiles of protein sequences encoded by each of these marker genes (<https://github.com/simroux/ICTV_VirophageSG>) * other expected (but not required) features: the genome should consist of dsDNA with a length between 15 kbp and 45 kbp and encoding a virophage-like penton protein detected based on established HMM profile(s) (<https://github.com/simroux/ICTV_VirophageSG>)   The specificity of the marker genes selected is illustrated in Fig. 1, which shows the score obtained for the three required markers (Hexon, Adenain, and ATPase) in hmmsearch comparisons for virophage (i.e. members of the *Maveriviricetes* class)and virophage-like (i.e. polinton-like viruses, closest neighbors of the *Maveriviricetes* class) sequences.  ## New orders and families  Based on a phylogenomic analysis of 257 (near)complete genomes assigned to the *Maveriviricetes* (using the demarcation criteria described above), we further propose to reorganize the class into 4 orders (currently 1) and 7 families (currently 1), by proposing the following actions (Fig. 2):  Create three new orders with three new associated families:   * To create a new *Divpevirales* order, named after virophages with Divergent penton proteins, including the new *Ruviroviridae* family, to host therumenvirophages. * A new *Lavidavirales* order, a name adapted from the current *Lavidaviridae* family named for “Large virus dependent or associated”, to host the new *Maviroviridae* family, named after the first isolated member of the taxon (Mavirus) * A new *Mividavirales* order, for “Mimivirus dependent or associated”, hosting the new *Sputniviroviridae* family, named after the first isolated member of the taxon (Sputnikvirus)   Create four new families in an existing order *Priklausovirales*:   * A new *Dishuiviroviridae* family, named after the lake from which the first member of the taxon (SW01) was isolated. * A new *Omnilimnoviroviridae* family, named by combining the prefixes “omni” (“all”, “everywhere”) and “limno” denoting a link to freshwater environments, since members of this clade were detected across a broad geographic range of freshwater lakes. * A new *Gulliviroviridae* family, named from Lemuel Gulliver, the main character of “Gulliver’s travel”, to classify virophages can be considered as “giant” compared to others, but are still relatively small compared to their associated giant viruses. * A new *Burtonviroviridae* family named from Mary Burton, the wife of Lemuel Gulliver in “Gulliver’s travel”, as this clade is the most closely related to the large virophages.   These taxa are monophyletic in an MCP (hexon) tree (Fig. 2), remain broadly monophyletic on ATPase and penton trees [11], and correspond to individual clusters when comparing genomes based on genome-wide Amino Acid Identity (AAI) [1].  ## Updated genera demarcation criteria  Current genera in the *Maveriviricetes* class (*Sputnikvirus* and *Mavirus*) were delineated based on manually curated lists of shared genes between members. Currently valid criteria for classification of virophages in the genus Sputnikvirus, are: “In addition to the two capsid protein genes, members of the Sputnikvirus species encode a FtsK-HerA family DNA-packaging ATPase, a cysteine protease, a primase-superfamily 3 helicase, a lambda-type integrase, a transposase, a Zinc-ribbon domain protein, a collagen-like protein, and six proteins of unknown function”. The common features for members of the genus *Mavirus* include: “7 homologous proteins involved in virion morphogenesis (minor and major capsid proteins, FtsK-HerA-type genome packaging ATPase and cysteine protease homologous to adenoviral maturation proteases), genome replication (protein-primed family B DNA polymerase and superfamily 3 helicase) and integration (retrovirus-like integrase which belongs to a broad superfamily of DDE transposases)”.  Therefore, we applied similar criteria to identify additional genera, specifically:   * a minimum of 7 genes shared, with shared genes being defined as a minimum gene AAI of 40%. * a minimum genome-wide AAI (i.e. average across all genes) of 15%.   Here, to apply these criteria uniformly to the 257 new genomes of putative *Maveriviricetes*, we leveraged the gwAAI tool [12] (options --min\_num\_shared 7, --min\_aai 40), and removed poorly connected genomes using mcl (inflation 2.0). This clustering recovers the same genera as the ones previously defined (including especially a single genus gathering Ace Lake Mavirus and Mavirus strain Spezl), and allowed the definition of new genera for all exemplar genomes of the new, above-listed taxa.  ## New exemplars and associated species and genera  Exemplar genomes are already available for the species classified in the genus *Mavirus* (family *Maviroviridae*, order *Lavidavirales*) and in the genus *Sputnikvirus* (family *Sputniviroviridae*, order *Mividavirales*). We propose to establish the following new species and genera represented by the corresponding exemplar genomes:   * a new genus *Essdubovirus* and a new species *Essdubovirus chlorellae*, in the *Dishuiviroviridae* family. The names are based on the name of the exemplar (“Essdub” for Chlorella virophage isolate SW01), and the host (currently classified as a *Chlorella* sp.). * a new genus *Panaquavirovirus* and a new species *Panaquavirovirus* *qinghaense* in the *Omnilimnoviridae* family. The prefix “Panaqua” was chosen to highlight the fact that these viruses were found across different aquatic environments, while “qinghaense” is based on the location from which this virus was identified (“Qinghai Lake”, a Tibetan mountain lake [13]) * a new genus *Burquivirus* and a new species *Burquivirus flavolapense* in the *Burtonviroviridae*. The prefix “Burqui” was built as a contraction of the family name (*Burtonviroviridae*) and “quinques”, as a reference to the number associated with the representative genome for this taxon (“Yellowstone Lake Virophage 5”), “*flavolapense*” is based on the location from which this virus was identified (“Yellowstone lake”, using the Latin “flavo” and “lacus”) * a new genus *Invirovirus* and a new species *Invirovirus crochense*. *Invirovirus* refers to the terminal inverted repeats observed in the representative of this taxon, while *crochense* is based on the location from which the exemplar genome was obtained (“Lake Croche”) * a new genus *Oviruvirus* and a new species *Oviruvirus* *palmerstonense*. *Oviruvirus* reflects the type of environment from which the exemplar sequence was obtained (sheep rumen, in latin *ovis rumen*), while *palmerstonense* is based on the location from which the specific genome was obtained (Palmerston North, N-Z).   ## Renaming existing species names to fit binomial format  Finally, following latest ICTV guidelines, names of the three currently recognized species in the *Maveriviricetes* class must be changed to comply with the binomial format. The proposed new names are as follows:   * *Cafeteriavirus*-*dependent* *mavirus* into *Mavirus* *cafeteriae*, based on the genus name and the original name of the giant virus and eukaryotic host associated with this virophage (Cafeteria roenbergensis virus (CroV) and *Cafeteria roenbergensis*) * *Mimivirus-dependent virus Sputnik* into *Sputnikvirus* *mimiviri*, based on the genus name and the name of the giant virus associated with this virophage (Mimivirus) * *Mimivirus-dependent virus Zamilon* into *Sputnikvirus zamilonense*, based on the genus name and the original name of this virophage (“Zamilon”) | |

**Supporting evidence**

**Fig. 1.** Distribution of hmmsearch score for each marker across virophage and virophage-like elements. The best hit for each genome (virophage or virophage-like) was considered for each marker.

****

**Fig. 2. MCP (Hexon) protein sequence phylogeny for (near-)complete genomes in the *Maveriviricetes* class.** The multiple alignment was built by using an iterative clustering-alignment-phylogeny procedure specifically adapted for aligning highly diverging sequences [15], automatically trimmed using clipkit v1.3.0 using the kpi-smart-gap mode to remove uninformative positions, and used as input for tree building with IQ-Tree v2.2.0.3 with automatic detection of the most appropriate substitution matrix (here Q.pfam+F+R8), and ultrafast bootstrap with 1000 replicates. Proposed taxa are indicated with colored rings (order to genus from inside to outside), and node supports are indicated with a colored circle. All nodes with a support < 50 were collapsed. Exemplar genomes of newly proposed species as well as other *Maveriviricetes* currently included in the NCBI RefSeq database (r216) are indicated on the tree. Exemplars are highlighted with a star.

**References**

1. La Scola B, Desnues C, Pagnier I, et al. (2008) The virophage as a unique parasite of

the giant mimivirus. Nature 455:100–4. doi: 10.1038/nature07218

2. Mougari S, Sahmi-Bounsiar D, Levasseur A, Colson P, and La Scola B (2019). Virophages of Giant Viruses: An Update at Eleven. Viruses 11, no. 8: 733. doi: 10.3390/v11080733

3. Fischer MG, Suttle CA (2011) A Virophage at the Origin of Large DNA Transposons.

Science 332:231–234. doi: 10.1126/science.1199412

4. Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010) Giant virus with a remarkable

complement of genes infects marine zooplankton. Proc Natl Acad Sci U S A

107:19508–13. doi: 10.1073/pnas.1007615107

5. Paez-Espino, D., Zhou, J., Roux, S. et al. (2019) Diversity, evolution, and classification of virophages uncovered through global metagenomics. Microbiome 7, 157. doi: 10.1186/s40168-019-0768-5

6. Yutin, N., Kapitonov, V.V. & Koonin, E.V. (2015) A new family of hybrid virophages from an animal gut metagenome. Biol Direct 10, 19. doi: 10.1186/s13062-015-0054-9

7. Stough J, Yutin N, Chaban Y, Moniruzzaman M, et al. (2019) Genome and Environmental Activity of a Chrysochromulina parva Virus and Its Virophages. Frontiers in Microbiology 10. doi: 10.3389/fmicb.2019.00703

8. Bellas, C.M., Sommaruga, R. (2021) Polinton-like viruses are abundant in aquatic ecosystems. Microbiome 9, 13 doi: 10.1186/s40168-020-00956-0

9. Hackl T, Duponchel S, Barenhoff K, Weinmann A, Fischer M. (2021) Virophages and retrotransposons colonize the genomes of a heterotrophic flagellate eLife 10:e72674.

doi: 10.7554/eLife.72674

10. Sheng, Y., Wu, Z., Xu, S. & Wang, Y. (2022) Isolation and Identification of a Large Green Alga Virus (Chlorella Virus XW01) of Mimiviridae and Its Virophage (Chlorella Virus Virophage SW01) by Using Unicellular Green Algal Cultures. Journal of Virology 96, e02114-21. doi: 10.1128/jvi.02114-21

11. Roux, S., Fischer, M.G., Hackl, T., Katz, L.A., Schulz, F., Yutin, N. (2023) Updated Virophage Taxonomy and Distinction from Polinton-like Viruses. Biomolecules 13, 204. doi: 10.3390/biom13020204

12. Nayfach, S., Páez-Espino, D., Call, L., Low, S.J., Sberro, H., Ivanova, N.N., Proal, A.D., Fischbach, M.A., Bhatt, A.S., Hugenholtz, P., Kyrpides, N.C., (2021) Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. Nature Microbiology 6, 960–970. doi: 10.1038/s41564-021-00928-6

13. Oh, S., Yoo, D., Liu, W.-T.T. (2016) Metagenomics reveals a novel virophage population in a Tibetan mountain lake. Microbes and Environments 31, 173–177. doi: 10.1264/jsme2.ME16003

14. Tran, P., Ramachandran, A., Khawasik, O., Beisner, B.E., Rautio, M., Huot, Y., Walsh, D.A., 2018. Microbial life under ice: Metagenome diversity and in situ activity of Verrucomicrobia in seasonally ice‐covered Lakes. Environmental Microbiology 20, 2568–2584. doi: 10.1111/1462-2920.14283

15. Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V., Koonin, E.V. (2018) Origins and Evolution of the Global RNA Virome. mBio 9, e02329-18. doi: 10.1128/mBio.02329-18