

The International Committee on Taxonomy of Viruses

Taxonomy Proposal Form, 2024

**Part 1a: Details of taxonomy proposals**

|  |  |  |
| --- | --- | --- |
| **Title:** | Create one new order *Grandevirales* (*Duplodnaviria*) | |
| **Code assigned:** | 2024.014B |

|  |  |  |  |
| --- | --- | --- | --- |
| **Author(s), affiliation and email address(es):** | | | |
| **Name** | **Affiliation** | **Email address** | **Corresponding author(s)** X |
| Cook R | Quadram Institute Bioscience, Norwich, United Kingdom | [Ryan.cook@quadram.ac.uk](mailto:Ryan.cook@quadram.ac.uk) |  |
| Pye HV | Quadram Institute Bioscience, Norwich, United Kingdom | [Hannah.pye@quadram.ac.uk](mailto:Hannah.pye@quadram.ac.uk) |  |
| Crisci MA | University College London, London, United Kingdom | [mcrisci@partners4access.com](mailto:mcrisci@partners4access.com) |  |
| Telatin A | Quadram Institute Bioscience, Norwich, United Kingdom | [Andrea.telatin@quadram.ac.uk](mailto:Andrea.telatin@quadram.ac.uk) |  |
| Santini JM | University College London, London, United Kingdom | [j.santini@ucl.ac.uk](mailto:j.santini@ucl.ac.uk) | **X** |
| Adriaenssens EM | Quadram Institute Bioscience, Norwich, United Kingdom | [Evelien.adriaenssens@quadram.ac.uk](mailto:Evelien.adriaenssens@quadram.ac.uk) | **X** |

**Part 1b: Taxonomy Proposal Submission**

|  |  |  |  |
| --- | --- | --- | --- |
| **ICTV Subcommittee:** | | | |
| Animal DNA Viruses and Retroviruses |  | Bacterial viruses | **X** |
| Animal minus-strand and dsRNA viruses |  | Fungal and protist viruses |  |
| Animal positive-strand RNA viruses |  | Plant viruses |  |
| Archaeal viruses |  | General - |  |

|  |
| --- |
| **List the ICTV Study Group(s) that have seen or have been involved in creating this proposal:** |
| No Study Group available |

|  |  |  |  |
| --- | --- | --- | --- |
| **Optional – complete only if formally voted on by an ICTV Study Group:** | | | |
| **Study Group** | **Number of members** | | |
| **Votes in support** | **Votes against** | **No vote** |
|  |  |  |  |
|  |  |  |  |

|  |  |
| --- | --- |
| **Submission date:** | 04/06/2023 |

**Part 1c: Feedback from ICTV Executive Committee (EC) meeting**

|  |  |
| --- | --- |
| **Executive Committee Meeting Decision code:** | **X** |
| A – Accept |  |
| Ac – Accept subject to revision by relevant subcommittee chair. No further vote required | **X** |
| U – Accept without revision but with re-evaluation and email vote by the EC |  |
| Uc – Accept subject to revision and re-evaluation and email vote by the EC |  |
| Ud – Deferred to the next EC meeting, with an invitation to revise based on EC comments |  |
| J - Reject |  |
| W - Withdrawn |  |

|  |
| --- |
| **Comments from the Executive Committee:** |
| Proposal originally withdrawn by authors because accession numbers were available in time.  One species lack a correct exemplar genome accession number. |

**Part 1d: Revised Taxonomy Proposal Submission**

|  |
| --- |
| **Response of proposer:** |
| Resubmission of proposal with updated accession numbers and transferred and edited to fit the new proposal template.  Resubmission October 2024 to remove one species for which the exemplar genome accession number was not correct. |

|  |  |
| --- | --- |
| **Revision date:** | 07/10/2024 |

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

|  |
| --- |
| **Name of accompanying Excel module:** |
| 2024.014B.A.v2.Grandevirales\_no.xlsx |

|  |  |  |  |
| --- | --- | --- | --- |
| **Taxonomic changes proposed:** | | | |
| Establish new taxon | **X** | Split taxon |  |
| Abolish taxon |  | Merge taxon |  |
| Move taxon |  | Promote taxon |  |
| Rename taxon |  | Demote taxon |  |
| Move and rename |  |

|  |  |  |
| --- | --- | --- |
| **Is any taxon name used here derived from that of a living person:** | | **N** |
| **Taxon name** | **Person from whom the name is derived** | **Attached X** |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
| **Abstract of Taxonomy Proposal:** |
| *Taxonomic rank(s) affected*: Realm *Duplodnaviria*, kingdom *Heunggongvirae*, phylum *Uroviricota*, class *Caudoviricetes*  *Description of current taxonomy*: The viruses classified in this proposal do not have a current taxonomic assignment.  *Proposed* *taxonomic change(s):* We propose the creation of one new order, *Grandevirales*, comprising two new families, *Lakviridae* and *Epsomviridae*, three sub-families (*Quingentivirinae*, *Quadringentisvirinae* and *Sescentorumvirinae*), and four genera (*Vetruanivirus*, *Hatfieldvirus*, *Amboselivirus* and *Wendovervirus*). We also propose the creation of eight novel species within these genera.  *Justification*: We investigated the evolutionary relationships of 23 megaphage genomes with sizes greater than 400 kb and propose a taxonomy for their classification. Analysis of their putative proteins revealed that Lak phages formed a deeply branching monophyletic clade within the class *Caudoviricetes* that contained no other genomes, and hence justifies the creation of a new order *Grandevirales*. One of the interesting features of this clade is that all current members are characterised by an alternative genetic code, where the TAG stop codon is repurposed to an amino acid. |

|  |
| --- |
| **Text of Taxonomy proposal:** |
| *Taxonomic rank(s) affected*:  Realm *Duplodnaviria*, kingdom *Heunggongvirae*, phylum *Uroviricota*, class *Caudoviricetes*  *Description of current taxonomy*:  The viruses classified in this proposal do not have a current taxonomic assignment.  *Proposed* *taxonomic change(s)*:  We propose the creation of one new order, *Grandevirales*, comprising two new families, *Lakviridae* and *Epsomviridae*, three sub-families (*Quingentivirinae*, *Quadringentisvirinae* and *Sescentorumvirinae*), and four genera (*Vetruanivirus*, *Hatfieldvirus*, *Amboselivirus* and *Wendovervirus*). We also propose the creation of eight novel species within these genera.    *Demarcation criteria:*  **Order demarcation criteria:** Members of the order *Grandevirales* formed a deeply branching monophyletic clade in VipTree analysis (Figure 1). All members are predicted to use an alternative genetic code (Figure 1). Core genes/proteins are only shared across the whole order when amino acid identity for clustering is reduced to 25%.  **Family demarcation criteria:** The two families proposed are *Lakviridae*, which encompasses 22 of the Lak megaphages which share 72 core genes (most of which have unknown functions) and *Epsomviridae*, which only has one current member and only one shared core gene with the other members of the order *Grandevirales*. We propose a shared core gene content of minimum 10% at 70% amino acid identity and robust clustering in the core genome phylogenetic tree as family demarcation criteria.  **Sub-family demarcation criteria:** Robust clustering in the core genome phylogenetic tree with a suggested shared core gene content of minimum 25%. Members of the same subfamily typically share > 25% nucleotide identity across the genome length.  **Genus demarcation criteria:** An intergenomic similarity cut-off of 70%, a combination of average nucleotide identity and alignment fraction, is used alongside the pangenome analysis to determine genera demarcation. Members of the same genus have >70% intergenomic similarity and cluster tightly in marker gene phylogenies.  **Species demarcation criteria:** A demarcation value of 95% intergenomic similarity was used to define different species according to intergenomic similarity.Members of the same genus have >95% intergenomic similarity.  *Justification*:  The 23 Lak megaphage genomes used for classification are all publicly available and were regarded as complete genomes from the original publications (strain names indicated in Table 1) [1-5]. All phages were comprised of dsDNA and have been assigned to the class *Caudoviricetes* due to the presence of tail proteins, portal proteins, terminase large subunits and HK97-like major capsid proteins.  Predicted protein analysis of 56 unclassified megaphages (phages with genomes >400kb) using VipTree to produce a hierarchical clustered tree based on tBLASTx scores, alongside 3,539 currently classified *Caudoviricetes*, revealed that the 23 previously described “Lak megaphages” formed a deeply branching monophyletic clade (Figure 1). The other 33 phage genomes were interspersed into known orders. Additionally, all the 23 Lak megaphages which grouped together on the proteomic tree used an alternative genetic code (Figure 1). Therefore, the monophyletic grouping of these phage genomes and their alternative genetic code warrants the creation of a new order, termed *Grandevirales*.  To infer family level taxonomy, the 23 Lak megaphage genomes were subjected to core protein analysis (Figure 2). At an amino acid threshold of 25%, only one core gene was shared amongst all 23 Lak mega phages, and 22 of the phage genomes shared 72 core genes at 70% amino acid identity, hence the justification of the creation of two families based on core genes (Figure 2). The two families proposed are *Lakviridae*, which encompasses 22 of the Lak megaphages which share 72 core genes (most of which have unknown functions) and *Epsomviridae*, which only has one current member and only one shared core gene with the other members of the order *Grandevirales*.  The concatenated protein phylogeny of the 72 core genes identified within the members of the family *Lakviridae* formed two distinct clusters, and hence justifies the creation of two sub-families (*Quingentivirinae* and *Quadringentisvirinae*) (Figure 3).The pangenome analysis showing the presence and absence of proteins clustered at 70% amino acid identity provides further evidence for the creation of sub-families within the *Lakviridae* and *Epsomviridae* families, due to the distinct clustering of proteins present in each sub-family (Figure 2). Although there is only one member of the *Epsomviridae* family, the creation of a sub-family (*Sescentorumvirinae)* is still informative, and demarcation can be determined when more genomes are incorporated into the *Epsomviridae* family.  Intergenomic similarity was used alongside the pangenome analysis to determine genera demarcation. A demarcation cut-off of 70% similarity (whole genome length) was used to determine genus (Figure 4). This demarcation criteria divided the 23 Lak megaphages into four genera, which have been proposed to be named *Wendovervirus*, *Hatfieldvirus*, *Amboselivirus* and *Vetruanivirus.* The latter three of which are all members of the *Lakviridae* family, with only one genome incorporated into the *Wendovervirus* genus.  A demarcation value of 95% intergenomic similarity was used to define different species according to intergenomic similarity (Figure 4). Using this demarcation value, we propose the creation of nine novel species.  *Derivation of names*:  **Origin of Names Proposed:** All taxonomic names proposed follow the guidance suggested and use a latinised binomial name 1.  **Order “*Grandevirales*”:** The name “*Grandevirales*” is proposed for the order, as Grande means large (or great) in Latin and in other European languages. Based on the present analyses, some related phages in this order have genomes under 500 kb, therefore it would be unsuitable to denote the order as ‘Mega’, given that ‘megaphage’ has historically been used to describe phages with genomes over 500 kb 2. Furthermore, the order *Megavirales* has been used to describe large eukaryotic viruses (>100 kb) 3. Given that all phages within this order are ‘huge’, i.e., > 200 kb, and that the order encompasses the largest known complete phage genomes, the order name “*Grandevirales*” was chosen.  **Family “*Lakviridae*”:** The family “*Lakviridae*” was selected as the founding members of this family (A1 and A2) were all identified from the gut microbiomes of arsenic-impacted individuals in Laksam Upazila, Bangladesh 2.  **Family “*Epsomviridae*”:** The second family proposed in this study is “*Epsomviridae*”, named after Epsom, Surrey (UK). Epsom is a town famous for horse racing and is the location where the racehorse “Sonny” was stabled and trained. In the current proposal, only one megaphage belongs to this family, which was identified from the horse gut microbiome 4.  **Sub-families “*Quingenti-*”, “*Quadringenti-*” and “*Sescentorum-*” -*virinae:*** Subfamily names were chosen based on genome sizes of the founding members of each subfamily. ‘Quingenti’, ‘Quadringentis’, and ‘Sescentorum’ are Latin for 400, 500 and 600, respectively. For “*Quingentivirinae*”, founding members have genome lengths within the 500 kb range. For “*Quadringentisvirinae*”, current members have ~476 kb genomes, and one phage has currently been classified into the subfamily “*Sescentorumvirinae*”, with a ~660 kb genome. Although the genome lengths of founding members have been used as the basis of subfamily names, this is not a criterion for taxonomic classification and members of each subfamily form distinct clusters within *Lakviridae* and *Epsomviridae*.  **Genus “*Vetruanivirus*”:** The genus name “*Vetruaniviru*s” is an amalgamation of the words Veterinary and Eruani, which encompasses the isolation source of current phages in this genus. Some phages were identified from samples of pigs at the Royal Veterinary College (RVC), hence “*Vet-* “, whilst other strains were identified from samples of individuals living in a village called Eruani in Laksam, Bangladesh, hence “*ruani-*“ which forms “*Vetruanivrus*”.  **Genus “*Amboselivirus*”:** Phage genomes belonging to the proposed “*Amboselivirus*” genus (B1-B9) were resolved from faecal samples collected from Kenyan yellow baboons living in the Amboseli national park. The current sole species has been given the name *Amboselivirus simi* (‘simia’ is Latin for primate/monkey, with simi in the genitive form). Strains designated as “*Amboselivirus*” (B1-B9) are maintained as described in the original paper 2.  **Genus “*Hatfieldvirus*”:** Two of the phage genomes identified from pig samples form the “*Hatfieldvirus*” genus, which has been named according to their isolation location. These phages were identified from faecal samples from pigs reared and cared for at the RVC in Hatfield, Hertfordshire (UK). The sole species name has been given as, *Hatfieldvirus porci*, as porci is Latin for pig. Strains of this proposed genus (and those predicted to be placed in it), have so far only been found in pig gastrointestinal tracts.  **Genus “*Wendovervirus*”:** The sole genome belonging to this genus, Sonny, was assembled from sequencing data of a microbiome sample from a horse stabled at the Wendover stables in Epsom.  *Please note that the species “Vetruanivirus cani” will not be officially proposed in this Taxonomy Proposal. The accession number of the exemplar virus genome is currently not available in the correct format. We plan to submit a new taxonomy proposal to formalize this species when the accession number becomes available.* |

|  |
| --- |
| **References:** |
| 1 Postler, T. S. *et al.* Guidance for creating individual and batch latinized binomial virus species names. *Journal of General Virology* **103** (2022). <https://doi.org/https://doi.org/10.1099/jgv.0.001800>  2 Devoto, A. E. *et al.* Megaphages infect Prevotella and variants are widespread in gut microbiomes. *Nature Microbiology* **4**, 693-700 (2019). <https://doi.org/10.1038/s41564-018-0338-9>  3 Colson, P., de Lamballerie, X., Fournous, G. & Raoult, D. Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. *Intervirology* **55**, 321-332 (2012). <https://doi.org/10.1159/000336562>  4 Crisci, M. A. *et al.* Closely related Lak megaphages replicate in the microbiomes of diverse animals. *iScience* **24**, 102875 (2021). <https://doi.org/10.1016/j.isci.2021.102875>  5 Cook, R. *et al.* Decoding huge phage diversity: a taxonomic classification of Lak megaphages. *J Gen Virol* **105** (2024). <https://doi.org/10.1099/jgv.0.001997> |

|  |
| --- |
| **Tables, Figures:** |

The tables and figures presented here have been published as: Cook, Crisci, Pye, Adriaenssens, Santini, Decoding Huge Phage Diversity: A Taxonomic Classification of Lak Megaphages, bioRxiv preprint: doi: <https://doi.org/10.1101/2024.02.01.578382>, Journal of General Virology vol 105 doi: <https://doi.org/10.1099/jgv.0.001997> 5.

**Table 1**. Proposed taxonomic classification of Lak megaphages. The species “Vetruanivirus cani” will not be officially proposed in this Taxonomy Proposal due to a missing accession number.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Order** | *Grandevirales* | | | | | | | | |
| **Family** | *Lakviridae* | | | | | | | | *Epsomviridae* |
| **Subfamily** | *Quingentivirinae* | | | | | | | *Quadringentisvirinae* | *Sescentorumvirinae* |
| **Genus** | *Vetruanivirus* | | | | | | *Amboselivirus* | *Hatfieldvirus* | *Epsombirus* |
| **Species** | *Vetruanivirus primi* | *Vetruanivirus secundi* | *Vetruanivirus dhakaense* | *Vetruanivirus porcinprimi* | *Vetruanivirus porcinsecundi* | *Vetruanivirus*  *cani* | *Amboselivirus simi* | *Hatfieldvirus porci* | *Wendovervirus sonnii* |
| **Genomes/Strains** | As-22-1  As-22-2  As-22-3  As-22-4 | As-20 | Dhaka | RVC-AP1-GC26 | RVC-AP4-GC26  RVC-AP3-GC26 | Wal-1  Wal-2 | B1-B9 | RVC-AP3-GC31  RVC-JS4-GC31 | Sonny |

A picture containing circle, colorfulness, screenshot, hula hoop

Description automatically generated**Figure 1**. *Grandevirales* are a distinct clade of phages with likely alternative genetic code. (**A**) VipTree proteomic tree of “Megaphages” and *Duplodnaviria* with viral family shown in the coloured ring. Blue bar chart (inner) represents genome length and red bar chart (outer) shows difference in coding capacity when using translation table 15 rather than 11 (i.e., coding capacity of 90% using 15 and 70% using 11 would lead to a difference of 20). The *Grandevirales* are shown with red branches and other Megaphages are shown with green. (**B**) A pruned tree showing the *Grandevirales* with nearest sister clades only.

A screen shot of a computer screen

Description automatically generated with low confidence

**Figure 2**. Shared protein clusters for phages of *Grandevirales.* Heatmap showing presence/absence of proteins clustered at 70% identity with the dendrogram showing hierarchical clustering. Colour strips on the y-axis show prosed taxonomy, and colour strips on the x-axis show predicted function of the protein cluster derived from PHROGs. X-axis labels show strain name with proposed species name in brackets.

A picture containing screenshot, design

Description automatically generated**Figure 3**. Core genome phylogeny of *Lakviridae* phages. A concatenated protein phylogeny of 72 “core” genes present on the 22 members of proposed *Lakviridae.* Alignments were performed using MAFFT, and the tree was produced using IQ-Tree with 1000 rapid bootstraps and -m TEST to optimize model fits for each alignment. Tree is rooted at the midpoint and bootstraps ≥ 95% are shown. The coloured strips indicate proposed genera and sub-families. Node labels are based on strain names with proposed species names shown in brackets.

A screenshot of a computer

Description automatically generated with medium confidence**Figure 4**. Intergenomic similarity of *Grandevirales* phages. Heatmap showing intergenomic distance (ANI multiplied by aligned fraction) of *Grandevirales* phages with proposed taxonomy shown in coloured strips. Those shown as having a similarity of 100.0\* are highly similar but not identical. Axis labels are based on strain names with proposed species names shown in brackets.