

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

|  |  |  |
| --- | --- | --- |
| **Code assigned:** |  |  |
| **Short title:** Create 418 new genera and 1,706 new species in the class *Leviviricetes*, move and rename 375 species, and abolish 114 genera and 48 species. | | |
|  | | |

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

|  |  |
| --- | --- |
| Muscatt G, Kuhn JH, Millard A, Bending GD, Jameson E | george.s.muscatt@warwick.ac.uk; kuhnjens@mail.nih.gov; adm39@le.ac.uk; gary.bending@warwick.ac.uk; e.jameson@bangor.ac.uk |

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

|  |
| --- |
| University of Warwick, UK [GM, GDB]  NIH/NIAID/DCR/Integrated Research Facility at Fort Detrick, USA [JHK]  University of Leicester, UK [AM]  Bangor University, UK [EJ] |

**Corresponding author**

|  |
| --- |
| George Muscatt; george.s.muscatt@warwick.ac.uk |

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

|  |
| --- |
| ICTV *Leviviricetes* Study Group |

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

|  |
| --- |
|  |

**ICTV Study Group votes on proposal**

Study Group Chair(s) please report voting by Study Group members on the proposal prior to final submission to the Subcommittee Chair.

|  |  |  |  |
| --- | --- | --- | --- |
| **Study Group** | **Number of members** | | |
| **Votes support** | **Votes against** | **No vote** |
| ICTV *Leviviricetes* Study Group |  |  |  |

**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)** |
| N/A | N/A | N/A |

**Submission dates**

|  |  |
| --- | --- |
| Date first submitted to SC Chair | June 23, 2023 |
| Date of this revision (if different to above) | November 2, 2023 |

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

|  |
| --- |
| **EC comments:** Please provide the evidence in a format that we can read and assess the proposed taxonomic changes. This evidence could be the raw date like alignment files, tree files, heatmap and annotations and/or pdf versions of the current figures in which we can zoom in on the individual phage labels. We need this for the nucleotide comparison heatmap and the phylogenetic trees.  The additional evidence can be provided as supplementary files that can be added to the proposal and referenced in the word document.  **Proposer response:** The additional evidence requested has been provided in the supplementary materials. Enlarged, higher resolution versions of the heatmap and phylogenetic trees have been provided as supplementary figures and referenced in the word document, where relevant. An interactive phylogenetic tree, displaying the proposed reorganization of leviviricete taxonomy, has also been made available at the following link: <https://itol.embl.de/tree/861725212587391698939501>. |

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

**Text of proposal**

|  |
| --- |
| N/A |

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

**Name of accompanying Excel module**

|  |
| --- |
| 2023.###B.S.v1.Leviviricetes\_reorg.xlsx |

**Abstract**

|  |
| --- |
| Recent environmental metatranscriptomics studies have revealed many bacterial positive-sense RNA (+ssRNA) virus genomes, most notably in soils. Here, we propose reorganizing the established taxonomy of lenarviricot class *Leviviricetes* via the addition of 6,685 coding-complete virus genome sequences (previously defined as sequences encoding a maturation protein with a minimum length of 350 amino acid residues and an RNA-directed RNA polymerase (RdRp) longer than 500 amino acid residues) to the previously available dataset.  Application of the current leviviricete species demarcation criterion of 80% pairwise amino-acid identity of the RdRp protein resulted in an increase in the number of leviviricete species from 882 to 2,540. Similarly, application of the current genus demarcation criterion of 50% pairwise amino acid identity of the RdRp resulted in an increase of leviviricete genera from 428 to 732.  Phylogenetic assessment confirms that the established taxonomic structure of two leviviricete orders and six families remains conserved. |

**Text of proposal**

|  |  |
| --- | --- |
| |  | | --- | | **Addition of +ssRNA viral genomes**  A query database of bacterial +ssRNA virus sequences was generated from the 882 previously classified leviviricetes, 4,006 additional sequences available from GenBank, and 16,541 new genome sequences determined using soil sample metatranscriptomes [1]. These sequences were queried using profile hidden Markov models (HMMs) designed to detect +ssRNA viral proteins (available at <https://figshare.com/articles/dataset/Bacterial_ssRNA_virus_Hidden_Markov_Models/12745394>).  Subsequently, 6,685 coding complete +ssRNA virus genome sequences were identified using the previously established definition for “coding-complete” [2], i.e., sequences that contain three core proteins in order maturation protein—coat protein—RdRp, a minimum-length maturation protein (350 amino acid residues), and a minimum-length RdRp (500 amino acid residues).  **Species and genus demarcation criterion**  The previously established species and genus demarcation criteria, 50% and 80% pairwise amino-acid identity of the RdRp, respectively, was maintained (Figure 1; enlarged version in Figure S1). This resulted in a total of 2,540 leviviricete species and 732 leviviricete genera.  **Genus and species name generation**  Genus names:  The generation of genus names for the 418 proposed genera was performed automatically, through the manipulation of strings from a seed database containing names of cities, towns, and villages in Warwickshire, UK. String manipulation was achieved using a method presented previously [3]. Briefly, letters of each seed word were mutated randomly to modify the string while maintaining pronounceable syllables. Long strings were shortened to less than seven letters and forced to end on a vowel to prevent a hard consonant before the genus level suffix “*-virus*”. For example, the input seed of “Welford-on-Avon” was used to generate the genus name *Wehlfuvirus*. Resulting genus names were checked against the ICTV Species Master List 2022.v1 to ensure the uniqueness of taxon name word stems (Realm*→*Species).  Species names:  Latinized binomial virus species names were generated by combining the relevant genus name with a Latin species epithet based on the source of the representative (exemplar) virus, following guidance from Postler *et al.* [3]. For example, *Bardovirus agrivicinum* belongs to the genus *Bardovirus* and its exemplar virus was recovered from soil (*agri*- from ager, land or field; -*vicinum*, a neighbor). A full list of root words used for generating species epithets, along with their English translations/meanings, are provided in Table 1. One of the following suffixes for indicating “inhabiting” were then added to the root words: -*adaptatum*, -*cola*, -*enecus*, -*habitans*, -*vicinum*, and -*vivens*. For genera represented by many species, the following additional prefixes were used to create unique species epithets: *allo*-, *neo*-, and *pseudo-*.  **Changes to established species and genera**  Given the addition of >5,000 novel coding-complete +ssRNA viral genome sequences, a greater extent of within-taxon diversity was captured in the analysis presented here, than previously. Therefore, upon clustering of the RdRp protein for demarcating species and genera across the entire leviviricete sequence dataset, the following changes to the established taxonomy were required.  Moving and renaming of established species:  Upon the clustering of RdRp protein sequences to produce leviviricete genera, there were 375 previously established species (previously belonging to 84 genera) which were moved into 45 established genera and 23 new leviviricete genera. The respective binomial species names were updated to reflect the new genus assignments. For example, the previously taxonomized species *Rainacovirus pelenecus* was moved from the genus *Rainacovirus* to the newly formed genus Alkesdovirus, and hence renamed *Alkesdovirus pelenecus*.  Abolition of established species and genera:  Previously established leviviricete species were abolished when the RdRp protein sequences from genomes belonging to multiple established species clustered into a single new species. Similarly, previously established leviviricete genera were abolished when the RdRp protein sequences from genomes belonging to multiple established genera clustered into a single new genus. To prevent confusion with the merging of multiple species/genera, the previously established species and genera were abolished and replaced with newly named species and genera, respectively.  **Conservation of established order and familial taxonomic ranks**  Conservation of two orders:  Phylogenetic assessment of the bacterial +ssRNA viral RdRp protein confirmed the conservation of the two leviviricete orders, assigned using RdRp protein cluster detection (Figure 2; enlarged version in Figure S2).  Conservation of six families:  Phylogenetic assessment of core protein concatenations for each +ssRNA virus species confirmed the conservation of the six leviviricete families, assigned using coat-protein cluster detection (Figure 3; enlarged version in Figure S3). There were 10/2,540 leviviricete species (belonging to three existing and two novel genera) for which the best predicted coat-protein cluster did not confidently match its corresponding best predicted RdRp cluster. This implied assignment to a familial rank within the wrong order. Therefore, these five genera (*Grandbuvirus*, *Mahrahovirus*, *Nicedsevirus*, *Nordovirus*, and *Skrubnovirus*) remain unassigned to families and orders.  An interactive phylogenetic tree visualizing the proposed reorganization of Leviviricetes can be accessed here: <https://itol.embl.de/tree/861725212587391698939501>. | |

**Supporting evidence**

A picture containing screenshot, colorfulness, line, diagram

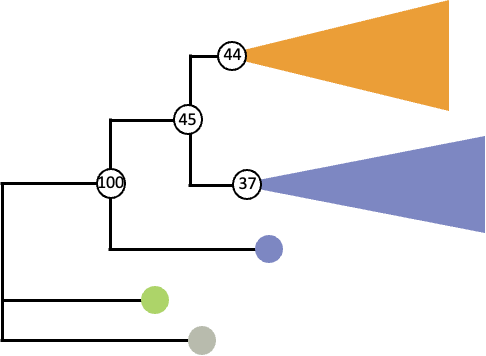
Description automatically generated

**Figure 1 - Example of species and genus demarcation criteria based on the previously established 80% and 50% pairwise amino acid identity of the RdRp.**

Pairwise amino acid identities of the RdRp protein sequences for the 738 members of the leviviricete family *Atkinsviridae*. Dotted boxes indicate the formation of a genus and a species cluster, respectively. Pairwise identities in shades of blue represent similarities below the genus demarcation threshold. Shades of green represent genus-level similarities below the species demarcation threshold. Shades of red represent species-level similarities.

**Table 1 - List of root words based on the source of the representative virus used to generate species epithets.**

|  |  |
| --- | --- |
| **Root** | **Translation/meaning** |
| Acridae- | Acrida-associated virus |
| Agri- | Land, field |
| Arvi- | Ploughed land |
| Asio- | Mud |
| Caeni- | Mud |
| chilonis | Chilo-associated virus |
| Choro- | Fields, land |
| Chthono- | Soil, world |
| Conocephali- | Conocephalus-associated virus |
| Cyperi- | Cyperus-associated virus |
| Dinebrae- | Dinebra-associated virus |
| Edapho- | Soil |
| escherichiae | Escherichia-infecting virus |
| Faeci- | Feces, sediment |
| Frumenti- | Grain |
| Fundi- | Farm |
| Geo- | Earth |
| Graminis- | Soil |
| Humi- | Land, soil |
| Infecundi- | Infertile |
| Ischnurae- | Ischnura-associated virus |
| Leersiae- | Leersia-associated virus |
| Limi- | Mud |
| Luti- | Mud |
| Messi- | Crop |
| Montis- | Mountain |
| Nemoris- | Forest |
| Paspali- | Paspalum-associated virus |
| Pedo- | Soil |
| Pelo- | Mud |
| Plasmoparae- | Plasmopara-associated virus |
| Pontederiae- | Pontederia-associated virus |
| Prati- | Meadow |
| Radici- | Root |
| Segeti- | Crop, grain |
| Soli- | Land, soil |
| Telluri- | Land |
| Terri- | Land, soil |
| Titanodulae- | Titanodula-associated virus |
| Vitis- | Vitis-associated virus |



**Figure 2 - Phylogenetic assessment of bacterial positive-sense RNA (+ssRNA) virus RNA-directed RNA polymerase (RdRp) sequences.**

Phylogeny of +ssRNA virus RdRp sequences generated using IQ-TREE, the VT+F+R10 model, and 1,000 ultrafast bootstrap replicates. The phylogenetic tree shows RdRps deduced from 2,816 coding-complete genome sequences, made non-redundant at 95% BLASTP identity across 95% of coverage length. Representative sequences of viruses assigned to non-leviviricete lenarviricots (howeltoviricete *Mitoviridae* and amabiliviricete *Narnaviridae*) were used to root the tree. Branch tip colors indicate viral orders. Inset image (top left) represents a simplified version of the phylogenetic tree with bootstrap support values for the major branches. The RdRp phylogeny is used to confirm leviviricete *Norzivirales* and *Timlovirales*.



**Figure 3 -** **Phylogenetic assessment of bacterial positive-sense (+ssRNA) virus core-protein concatenations.**

Phylogeny of +ssRNA virus concatenated core-protein sequences (maturation protein, coat protein, and RdRp) generated using IQ-TREE, the VT+F+R10 model, and 1,000 bootstrap replicates. The phylogenetic tree shows 2,540 +ssRNA virus species and was midpoint rooted. The branch tip shapes indicate RdRp clusters, the demarcation criterion proposed for establishing leviviricete *Norzivirales* and *Timlovirales*. Branch tip colors indicate coat-protein clusters, the demarcation criterion proposed for establishing leviviricete families. Sequences without specific corresponding RdRp and coat protein sequences, and which were not assigned order and familial taxonomic ranks (see text), are highlighted by shapes with black outlines.

**References**

1. Muscatt G, Hilton S, Raguideau S, et al (2022) Crop management shapes the diversity and activity of DNA and RNA viruses in the rhizosphere. Microbiome 10:181. https://doi.org/10.1186/s40168-022-01371-3

2. Callanan J, Stockdale SR, Shkoporov A, et al (2020) Expansion of known ssRNA phage genomes: from tens to over a thousand. Sci Adv 6:eaay5981. https://doi.org/10.1126/sciadv.aay5981

3. Postler TS, Rubino L, Adriaenssens EM, et al (2022) Guidance for creating individual and batch latinized binomial virus species names. J Gen Virol 103:. https://doi.org/10.1099/jgv.0.001800