

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

|  |  |  |
| --- | --- | --- |
| **Code assigned:** | **2020.095B** |  |
| **Short title:** Rename one class (*Leviviricetes* - formerly *Allassoviricetes*), rename one order (*Norzivirales* - formerly *Levivirales*), create one new order (*Timlovirales*), and expand the class to a total of six families, 420 genera and 883 species | | |
|  | | |

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

|  |  |
| --- | --- |
| Callanan J, Stockdale SR, Adriaenssens EM, Kuhn JH, Pallen M, Rumnieks J, Shkoporov A, Draper LA, Ross RP, Hill C | [julie.callanan@umail.ucc.ie](mailto:julie.callanan@umail.ucc.ie); [stephen.stockdale@ucc.ie](mailto:stephen.stockdale@ucc.ie); [evelien.adriaenssens@quadram.ac.uk](mailto:evelien.adriaenssens@quadram.ac.uk); [kuhnjens@niaid.nih.gov](mailto:kuhnjens@niaid.nih.gov); mark.pallen@quadram.ac.uk; [j.rumnieks@biomed.lu.lv](mailto:j.rumnieks@biomed.lu.lv); [andrey.shkoporov@ucc.ie](mailto:andrey.shkoporov@ucc.ie); [l.draper@ucc.ie](mailto:l.draper@ucc.ie); [p.ross@ucc.ie](mailto:p.ross@ucc.ie); [c.hill@ucc.ie](mailto:c.hill@ucc.ie) |

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

|  |
| --- |
| APC Microbiome Ireland [JC, SS, AS, LAD, RPR, CH]  Quadram Institute Bioscience, UK [EMA, MP]  NIH/NIAID/DCR/Integrated Research Facility at Fort Detrick, USA [JHK]  Latvian Biomedical Research and Study Centre, Latvia [JR] |

**Corresponding author**

|  |
| --- |
| Colin Hill; [c.hill@ucc.ie](mailto:c.hill@ucc.ie) |

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

|  |
| --- |
| ICTV Bacterial and Archaeal Viruses Subcommittee; ICTV *Leviviridae* Study Group |

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

|  |
| --- |
|  |

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

|  |  |  |
| --- | --- | --- |
| **Taxon name** | **Person from whom the name is derived** | **Permission attached (Y/N)** |
| *Atkinsviridae* | John F. Atkins | Y |
| *Blumeviridae* | Thomas Blumenthal | Y |
| *Steitzviridae* | Joan A. Steitz | Y |

**Submission dates**

|  |  |
| --- | --- |
| Date first submitted to SC Chair | 31/07/2020 |
| Date of this revision (if different to above) | 26/11/2020 |

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

|  |
| --- |
| **EC comment:** Make species names legal according to the virus code.  **Proposer response:** We have created legal species names and have chosen to create Latin binomials for all species names.  **EC comment:** Check monophyly of the families in Fig 3 and Fig 4.  **Proposer response:** We have made the classification based on the figures more clear. The RdRP tree (Fig 3) has been remade to be more clear is used for the delineation of the orders. The phylogenetic tree based on the concatenated maturation, coat protein and RdRP (Fig 4) is used for the delineation of families. |

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

**Name of accompanying Excel module**

|  |
| --- |
| 2020.095B.R.Leviviricetes.xlsx |

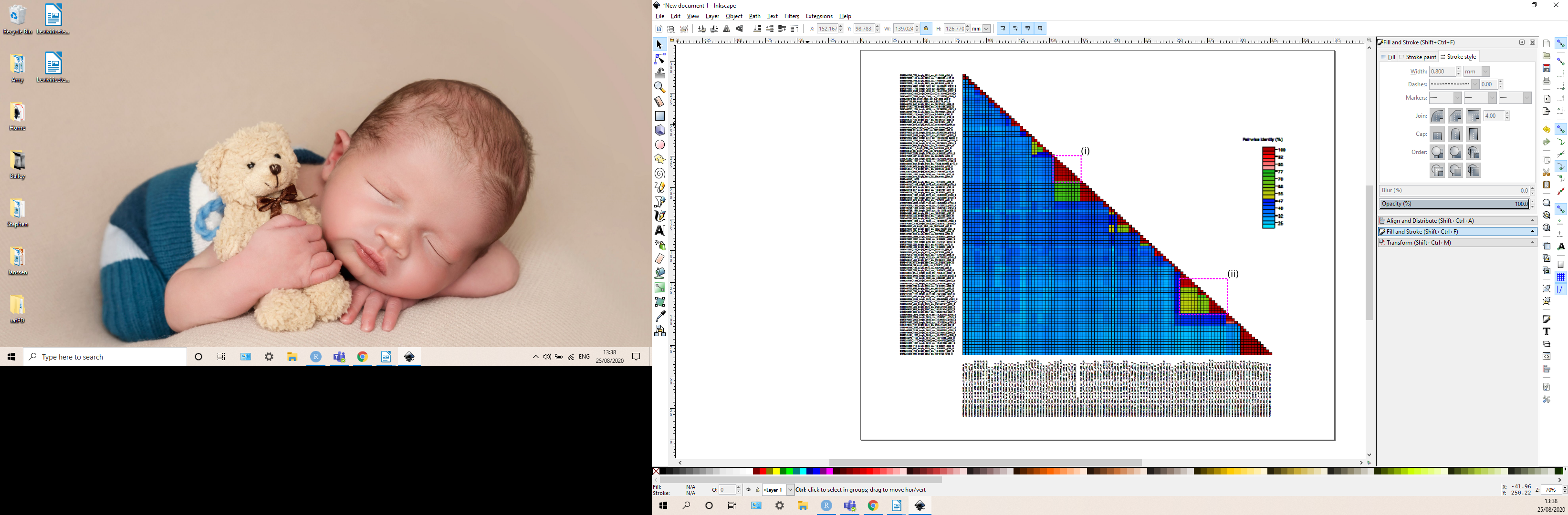
**Abstract**

|  |
| --- |
| The relatively simple genome architecture of all bacterial positive-sense single-stranded (+ssRNA) viruses identified to date contain three core genes; a maturation protein, a coat protein (CP), and an RNA-directed RNA polymerase (RdRP), in that order. We present the characterization of 1,868 near-complete bacterial +ssRNA virus genomes, defined as sequences encoding a maturation protein with a minimum length of 350 amino acids and an RdRP greater than 500 amino acids.  As nucleotide sequences are poorly conserved between bacterial +ssRNA, following the existing demarcation criteria for viruses classified in family *Leviviridae* (*Allassoviricetes*: *Levivirales*): pairwise amino-acid comparisons of the RdRP for species and genus demarcation were determined as 80% and 50% identity, respectively [1]. Profile hidden Markov models (HMMs) were used to detect more distant relationships between core bacterial +ssRNA virus proteins.  Phylogenetic relationships between bacterial +ssRNA virus RdRPs are in agreement with protein clustering, resulting in a proposed taxonomic structure of two orders, six families, 420 genera, and 883 species, encompassed within a single class. |

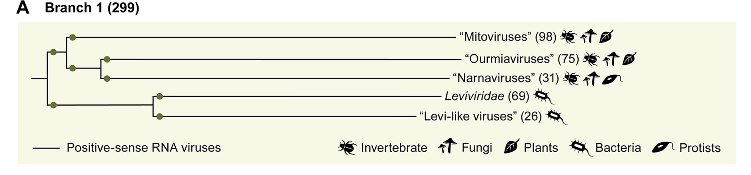
**Text of proposal**

|  |  |
| --- | --- |
| |  | | --- | | **Species demarcation criteria**  We have chosen 80% pairwise amino-acid identity of the viral core-encoded RdRP protein as the criterion for establishing species (Figure 1). This cutoff was applied in a bottom-up approach to 1,868 +ssRNA viruses that met specific criteria, including a minimum-length maturation protein (350 amino acids) and RdRP (500 amino acids). The 1,868 sequences originated from NCBI available sequences, and the studies of Callanan *et al.* [1], Starr *et al.* [2], Shi *et al.* [3], and Krishnamurthy *et al.* [4]. This yielded 883 species, with all sequences assigned a species membership contained within a distinct genus.  **Genus demarcation criteria**  We determined that the currently classified bacterial +ssRNA viruses, which are presently classified in genera *Levivirus* and *Allolevivirus*, share 50% amino-acid identity in their RdRPs. Applying this criterion, the 1,868 bacterial +ssRNA viruses clustered into 420 genera. All sequences classified with the 420 genera are contained within a distinct family taxon.  **New higher taxa and naming origins**  Order and family names, which are derived from scientists that studied bacterial +ssRNA viruses, were arbitrarily assigned to groups. No scientist’s name was deliberately associated with a particular group of viruses, in order to prevent author bias towards interpreting the merits or achievements of any individual.  **Class**  ***Leviviricetes* (formerly named *Allassoviricetes*)**: The class is based on the current highest taxonomic rank encompassing all bacteria-infecting +ssRNA viruses that share the same genome architecture of their three core genes. Previous analysis of +ssRNA viruses suggested that bacterial‑specific +ssRNA viruses form two distinct groups (Figure 2) [5].  **Orders**  ***Norzivirales* (formerly named *Levivirales*):** This order is based on the phylogeny and clustering of bacterial +ssRNA virus RdRP protein sequences. It is named after Norton Zinder (1928-2012), who isolated the first bacterial virus that contained RNA as its genetic material and continued to make crucial findings about these entities.  ***Timlovirales:*** This order is based on the phylogeny and clustering of bacterial +ssRNA virus RdRP protein sequences. It is named after Timothy Loeb (1935-2016) who, with Norton Zinder, isolated the first bacterial +ssRNA virus.  **Families**  Familial taxonomic groups were based on distinct phylogeny of bacterial +ssRNA virus RdRP protein sequences, which is supported by coat protein (CP) clustering using OrthoMCL. There are nine instances (out of 883 bacterial +ssRNA viruses) for which the predicted CP cluster did not confidently match its predicted corresponding RdRP cluster (difference in RdRP E-values < 1E-10). Therefore, no order or familial taxonomic rank is designated for these bacterial +ssRNA viruses until they are further investigated (see example AVE006, Figure 4).  ***Atkinsviridae***: named after John Atkins (1944-present) for his discovery of the lysin protein from Escherichia virus MS2.  ***Blumeviridae***: named after Thomas Blumenthal (1943-present) for his findings on the replication of bacterial ssRNA viruses, in particular the structure and function of the replicase.  ***Duinviridae***: named after Jan van Duin (1937-2017) for his discoveries related to novel bacterial ssRNA viruses and RNA folding within bacterial ssRNA viruses.  ***Fiersviridae* (formerly named *Leviviridae*)**: named after Walter Fiers (1931-2019), who sequenced the first gene and genome of any organism, *Escherichia* virus MS2.  ***Solspiviridae***: named after Sol Spiegelman (1914-1983) who discovered an RNA chain of only 218 nucleotides that could be reproduced by an RdRP.  ***Steitzviridae***: named after Joan Argetsinger Steitz (1941-present) for her determination of an initiation sequence that is central to modern-day ribosome profiling.  **Genus and species name generation**  **Genera**  Establishing a nomenclature for the 420 proposed genera was conducted as follows: A bacterial +ssRNA virus was chosen to represent the genus if (1) it was previously described and available in the ICTV archives, (2) its sequence had been deposited in GenBank, (3) or it was the longest contig sequence of all remaining available.  Some genera names for isolated phages were manually designed based on their current type species exemplar isolate names, including their phonetics: *Emesvirus* for Escherichia virus MS2, *Qubevirus* for Escherichia virus Qbeta, *Pepevirus* for Pseudomonas virus PP7, *Cunavirus* for virus C-1, *Empivirus* for the M-pili dependent virus, *Hagavirus* for enterobacteria virus Hgal1, *Perrunavirus* for Pseudomonas virus PRR1, *Apeevirus* for Acinetobacter virus AP205, and *Cebevirus* for Caulobacter virus Cb5.  While several additional unique genera names were generated manually, others were manipulated using three different scripts, written to subtly alter the spelling of chosen terms. For bacterial +ssRNA virus sequences from the Starr *et al*. study [2], random grass names were chosen from a list of world grasses, as this was the plant-soil interaction study. All metagenomically assembled bacterial +ssRNA virus sequences identified in the Callanan *et al*. study [1] include within their strain name the accession code for the raw sequence reads (i.e., SRR1234567). This code also enables the tracking of each sequence to its original study location. Unique names were therefore derived from the sequence’s original study location by modifying the anglicized names of cities, towns, or villages.  The name-modifying scripts altered terms as follows. Each letter of the term to be mutated was randomly assigned number 1, 2, or 3. Characters were then passed through a three-step script to create the following changes:   * *a1→ah, b1→p, c1→k, d1→t, e1→eh, g1→j, i1→ih, k1→c, o1→oh, t1→d, u1→uh* * *a2→e, e2→i, i2→o, o2→u, u2→a* * *a3→i, e3→o, i3→u, o3→a, u3→e, j3→g, k3→c, p3→b, t3→**d*   To maximize the likelihood that the mutated term was pronounceable, only the first occurrence of a repeating letter was kept. All long terms were truncated after the seventh character and shortened further to the last occurring vowel, if needed (to prevent a hard consonant before the genus level suffix “*-virus*”). Each mutated name needed to be a minimum of five characters in length and contain two consonants and two vowels. All names were checked against the ICTV Species Master List 2019.v1 to ensure the uniqueness of taxon name word stems (Realm*→*Species).  As an example, a unique genus name was derived from a representative phage that was isolated from a metagenomic study of Japanese environmental samples. A randomly chosen Japanese city, Kakunodate, was modified to ultimately generate the proposed genus name *Kecuhnavirus*.  **Species**  Binomial species names were generated by combining the genus name with a Latin species epithet based on a characteristic of the exemplar isolate or characteristic of the sample it was found in. The etymology of the species epithets is indicated in the comments section of the Excel module. The species naming was inspired by the preprint on Latin binomials for bacteria and archaea [6].  \* The authors would like to acknowledge and thank Aharon Oren for corrections of the Latin grammar of the proposed species epithets.  Taxonomy assigning profile Hidden Markov Models  Profile hidden Markov models (HMMs designed to detect bacterial +ssRNA viral proteins were first presented in the Callanan *et al.* study [1]. Updated HMMs are now available at *https://figshare.com/articles/dataset/Bacterial\_ssRNA\_virus\_Hidden\_Markov\_Models/12745394*. These HMMs are designed to aid researchers in finding bacterial +ssRNA viruses and inferring higher taxonomic assignments. Concatenated HMM profiles for bacterial +ssRNA virus proteins detect two RdRP protein clusters, three maturation-protein clusters, and nine CP clusters.  HMM profile searches return order and family taxonomic information for RdRP and CP hits, respectively. By curating the HMM search output to determine the best hit, a scheme for rapidly advancing bacterial +ssRNA phage taxonomy is available: RdRP\_A→*Norzivirales*; RdRP\_B→*Timlovirales*; CP\_A, CP\_B, and CP\_H→*Fiersviridae*; CP\_C→*Atkinsviridae*; CP\_D and CP\_F→*Steitzviridae;* CP\_E→*Blumeviridae;* CP\_G→*Solspiviridae*; and CP\_AP205-like→*Duinviridae*.  The phylogeny of bacterial +ssRNA viral RdRP proteins agree with the current RdRP and CP clusters used to generate the taxonomy assigning HMM profiles (Figure 3). | |

**Supporting evidence**

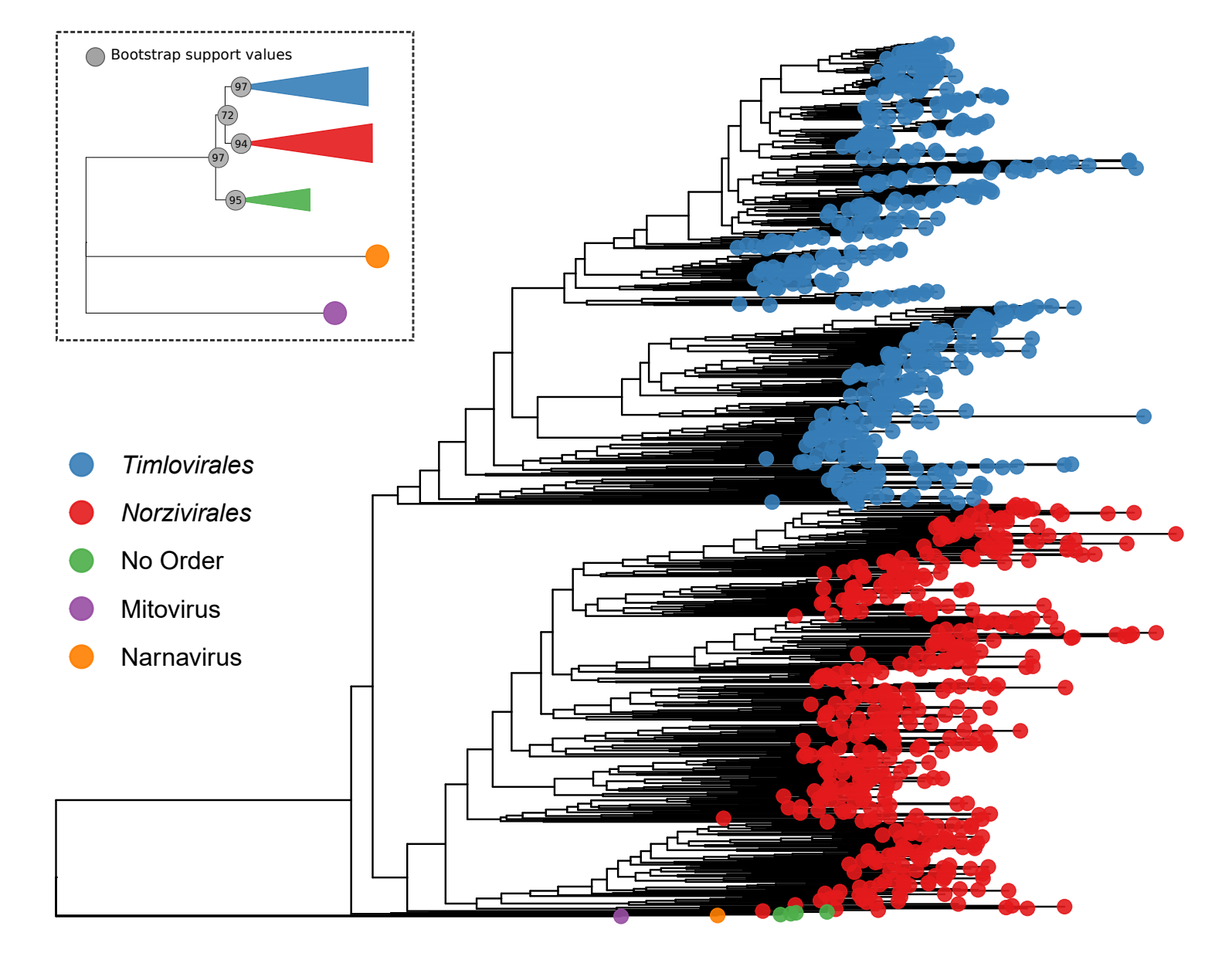
**Figure 1. Example of species and genus demarcation cutoffs of 80% and 50%, respectively, applied to pairwise RNA-directed RNA polymerase (RdRP) amino-acid comparisons.**

The pairwise amino-acid comparisons of the RdRP protein sequences for the members of the proposed *Atkinsviridae*. The image inset dotted-box **(i)** shows a distinct species clustering (red-colored boxes), whereas the dotted-box **(ii)** shows three species represented by multiple sequences and a species representing a single sequence, clustered into a genera (yellow-green-colored boxes). Pairwise comparisons in shades of blue do not meet the species or genera clustering criteria.

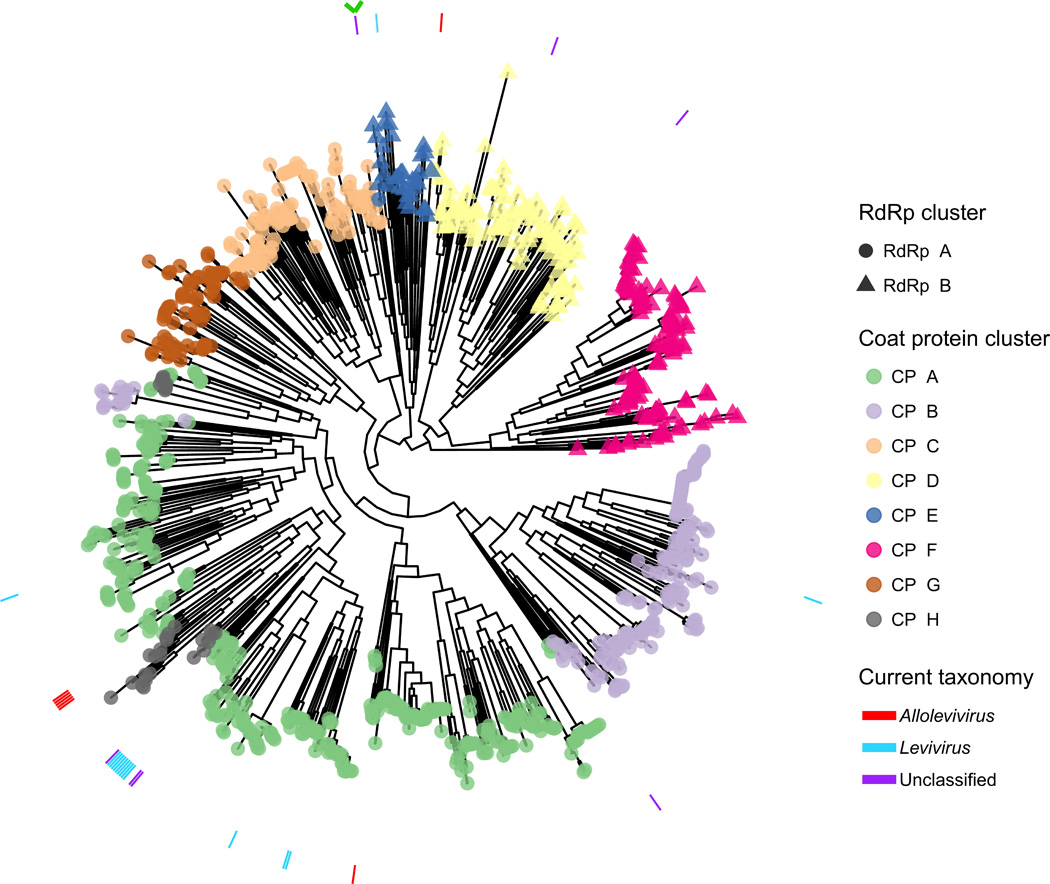


**Figure 2. Phylogenetic tree of positive-sense single-stranded (+ssRNA) viral RNA-directed RNA polymerases (RdRPs) of *Leviviricetes*.**

This information was sourced from Figure 2A of Wolf *et al.* (2018) [5]. This phylogenetic tree indicates the predicted separation of bacterial +ssRNA viruses into two clades, termed “*Leviviridae*” and “Levi-like viruses”. The numbers in parentheses indicate approximately how many distinct virus RdRPs are present in each respective branch. Symbols to the right indicate presumed virus host(s). Olive-green dots indicate that these branches are well-supported (≥0.7).

**Figure 3. Phylogenetic analysis of bacterial positive-sense single-stranded (+ssRNA) virus RNA-directed RNA polymerase (RdRP) protein sequences.**

Mitovirids and narnavirids were used to root the bacterial +ssRNA viral RdRP tree, generated using maximum-likelihood-based phylogenetic reconstruction in IQ-TREE with the VT+F+R10 model and 1,000 bootstrap replicates. RdRP sequences used to generate the tree were made non-redundant at 95% BLASTp identity across 95% of coverage length. The image inset, top left, shows a simplified version of the phylogenetic tree with bootstrap support values for the major branches. Sequences without specific corresponding coat protein and RdRP sequences, and which were not assigned order and familial taxonomic ranks (see text), are highlighted as “No Order”. RdRP phylogeny is the demarcation criteria proposed for establishing the *Norzivirales* and *Timlovirales* +ssRNA viral orders.



**Figure 4.** **Phylogenetic assessment of bacterial positive-sense single-stranded (+ssRNA) virus core proteins.**

This information was sourced from Figure 3 of Callanan *et al.* (2020) [1]. Phylogeny of concatenated bacterial +ssRNA viral maturation protein, coat proteins (CPs), and RNA-directed RNA polymerase (RdRP) sequences, which closely agree with the phylogeny of RdRP alone. Twenty-nine previously characterized and 1,015 newly identified viruses were included in this core protein phylogenetic analysis. Branch tip shapes indicate the specific RdRP protein cluster: circular = *Norzivirales,* triangular = *Timlovirales*, while branch tip colors indicate CP clusters. The +ssRNA viral coat protein clusters are used as the family demarcation criteria for *Leviviricetes* viruses. The family *Fiersviridae* is represented by coat protein clusters CP\_A, CP\_B, and CP\_H, the family *Steitzviridae* by clusters CP\_D and CP\_F, while all other families are represented by singular coat protein clusters. The encircling annotation ring depicts *Leviviridae* ICTV taxonomy (ICTV Master Species List 2018.v2). A green arrowhead points to virus AVE006, which encodes a unique RdRP and CP association and is therefore not assigned to an order or family within the *Leviviricetes* proposal.

**References**

1. Callanan J, Stockdale SR, Shkoporov A, et al (2020) Expansion of known ssRNA phage genomes: From tens to over a thousand. Science Advances 6:eaay5981. PMID: 32083183 DOI: 10.1126/sciadv.aay5981

2. Starr EP, Nuccio EE, Pett-Ridge J, et al (2019) Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. PNAS 116:25900–25908. PMID: 31772013 DOI: 10.1073/pnas.1908291116

3. Shi M, Lin X-D, Tian J-H, et al (2016) Redefining the invertebrate RNA virosphere. Nature 540:539–543. PMID: 27880757 DOI: 10.1038/nature20167

4. Krishnamurthy SR, Janowski AB, Zhao G, et al (2016) Hyperexpansion of RNA Bacteriophage Diversity. PLOS Biology 14:e1002409. PMID: 27010970 DOI: 10.1371/journal.pbio.1002409

5. Wolf YI, Kazlauskas D, Iranzo J, et al (2018) Origins and Evolution of the Global RNA Virome. mBio 9:. PMID: 30482837 DOI: 10.1128/mBio.02329-18

6. Pallen MJ, Telatin A, Oren A (2020) The Next Million Names for Archaea and Bacteria. Preprint DOI: 10.20944/preprints202010.0160.v1

Figure permissions:

Callanan, J., et al. "Expansion of known ssRNA phage genomes: From tens to over a thousand." *Science advances* 6.6 (2020): eaay5981. © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC) <http://creativecommons.org/licenses/by-nc/4.0/>