

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

|  |  |  |
| --- | --- | --- |
| **Code assigned:** | **2020.114B** |  |
| **Short title:** Create one new family (*Paulinoviridae*)including two genera moved from the family *Inoviridae* (*Tubulavirales*) | | |
|  | | |

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

|  |  |
| --- | --- |
| Roux S, Krupovic M | siroux1@gmail.com; mart.krupovic@pasteur.fr |

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

|  |
| --- |
| DOE Joint Genome Institute, Lawrence Berkeley National Institute [SR]  Institut Pasteur [MK] |

**Corresponding author**

|  |
| --- |
| Simon Roux (siroux1@gmail.com) |

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

|  |
| --- |
| *Tubulavirales* Study Group, Bacterial and Archaeal Viruses Subcommittee |

**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)** |
|  |  |  |
|  |  |  |
|  |  |  |

**Submission dates**

|  |  |
| --- | --- |
| Date first submitted to SC Chair | April 2020 |
| Date of this revision (if different to above) |  |

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

|  |
| --- |
|  |

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

**Name of accompanying Excel module**

|  |
| --- |
| 2020.114B.R.Paulinoviridae.xlsx |

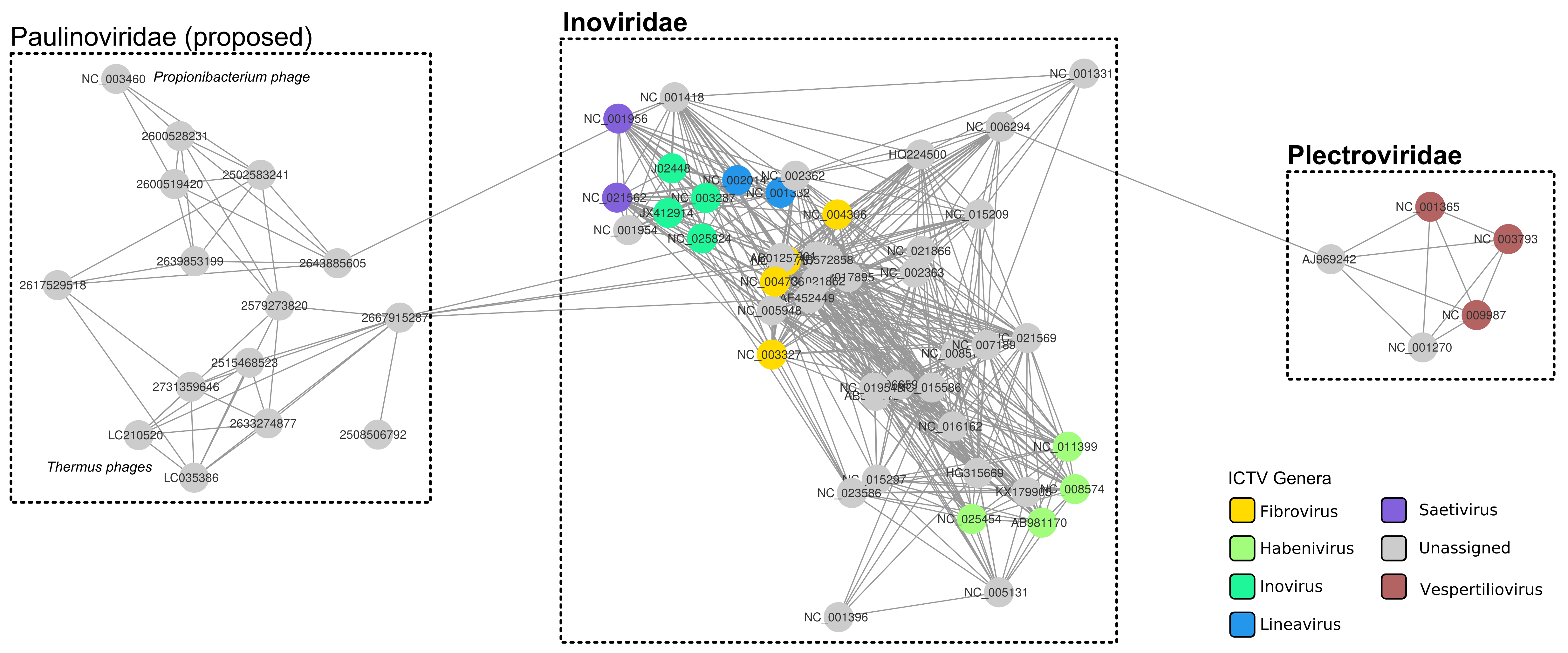
**Abstract**

|  |
| --- |
| Order *Tubulavirales* currently includes two families, *Inoviridae* and *Plectroviridae*. Phylogenetic and whole-genome-alignment criteria have been outlined to define species and genera within this order, whereas higher-order ranks (e.g., family) remain challenging to delineate based on the same features. Here we propose to use gene content as a criterion to delineate family-level groups in the *Tubulavirales* order. We verified that this criterion was consistent with currently established families, genera, and species. Then, we use this criterion to delineate a group within the *Tubulavirales*, which is distantly related to the currently existing families. To more accurately represent the evolutionary relationship of these divergent species with other members of the *Tubulavirales*, we propose to create a third family, the *Paulinoviridae*, and move genera *Bifilivirus* and *Thomixvirus* into this new family. Furthermore, the new *Paulinoviridae* family would include the only known member of the *Tubulavirales* infecting hosts within the phyla Actinobacteria, Deinococcus-Thermus, and Firmicutes. |

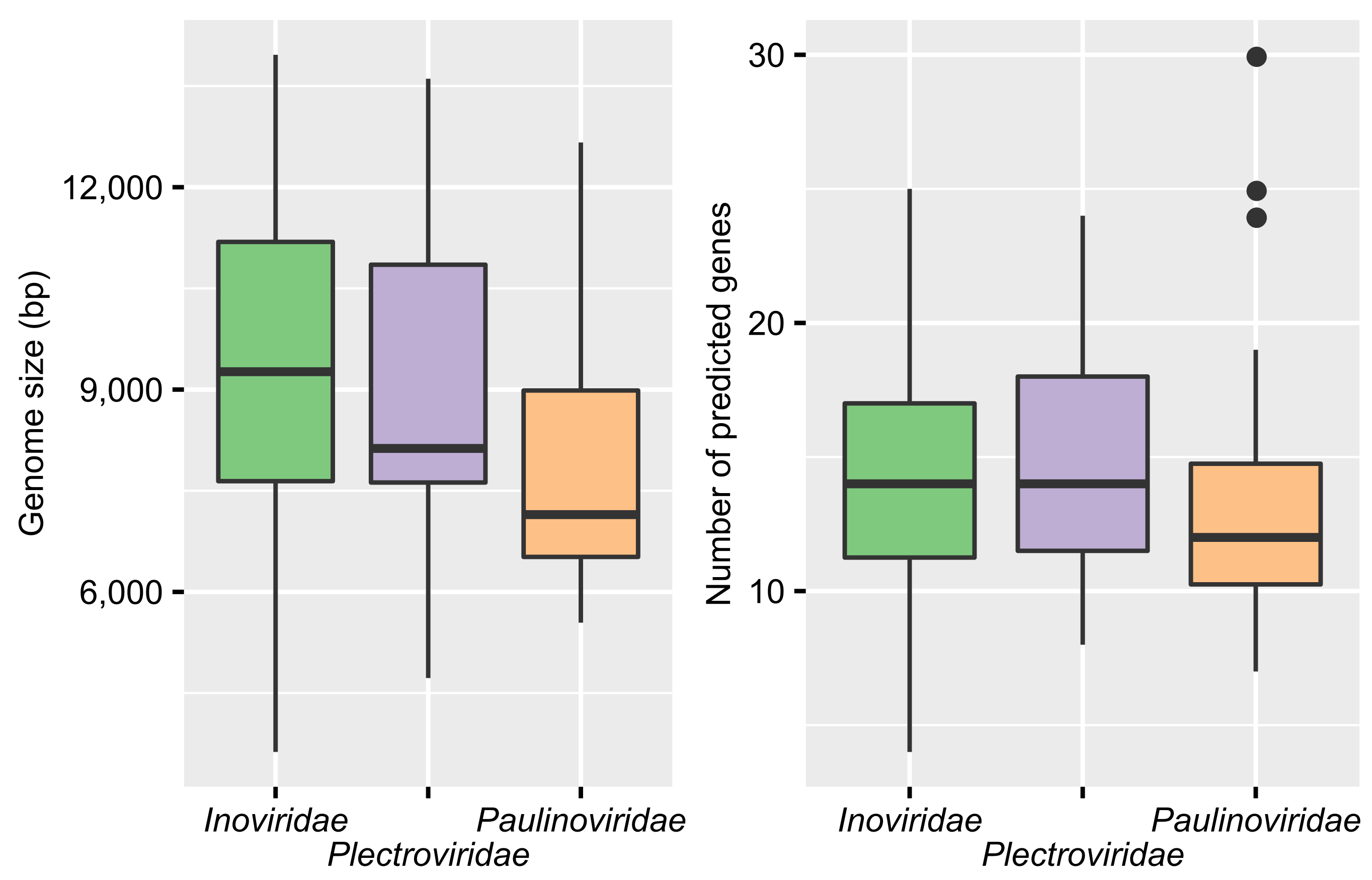
**Text of proposal**

|  |  |
| --- | --- |
| |  | | --- | | Order *Tubulavirales* includes two families, *Inoviridae* and *Plectroviridae*, of filamentous and rod-shaped bacteriophages, respectively. Criteria have been outlined to define species and genera within the *Tubulavirales* order. Namely, monophyletic clades including taxa with ≥95% global DNA sequence identity is used as a species demarcation criterion, whereas genera are defined using ≥90% amino acid identity in the core proteins. However, establishing higher-order ranks (e.g., families) is more challenging. Only a handful of genes are recognizably similar between members of different *Tubulavirales* genera and the sequence similarity is typically very low. Even for the single protein shared across all members of the *Tubulavirales*, namely, the morphogenesis protein, the amino acid similarity when comparing different genera typically ranges between 15 and 25% (as calculated with SDT; Muhire et al., 2014). Hence, while this gene can (and is) used to infer phylogenies and delineate monophyletic clades corresponding to genera and species, the deep-branching clades in these trees, which would correspond to ~ family-level groups, have low support values and confidence.  Instead of monophyletic clades and high sequence similarity, we propose using gene content as a criterion to delineate family-level groups in the *Tubulavirales* order. First, we verified that this approach was relevant by including known *Tubulavirales* isolates and representative prophages (Roux et al., 2019) in a gene-content-based network (Fig. 1). The network mostly recapitulated the established taxonomy within the *Tubulavirales* order, with the two families, *Inoviridae* and *Plectroviridae*, forming two well-separated modules. However, a group of sequences including 12 prophages and three isolates, Propionibacterium phage B5 (genus *Bifilivirus*) and Thermus phages OH3 and OH16, which are classified into a single species (genus *Thomixvirus*), all currently included into the family *Inoviridae*, clustered separately, suggesting a greater divergence of this group. We propose creating a third family in the order *Tubulavirales*, the *Paulinoviridae* (see below), and moving genera *Bifilivirus* and *Thomixvirus* into this new family. This grouping is also consistent with a gene-content network analysis performed on a larger set of inovirus genomes, including >10,000 inovirus-like sequences (Roux et al., 2019). Notably, the three families would each be associated with a different type of host membrane, as also noted in the larger analysis: hosts of viruses in the *Inoviridae* family would be associated with Gram-negative bacteria with a lipopolysaccharide (LPS)-containing outer membrane, members of the *Plectroviridae* family would be *associated* with wall-less hosts (i.e. bacteria within the Mollicutes order), and members of the new *Paulinoviridae* family would be associated with hosts without an LPS-containing outer membrane. In particular, the *Paulinoviridae* would include the three isolates, which have been cultivated on bacteria within the *Thermus* and *Propionibacterium* genera, and several prophages identified in the genomes of bacteria belonging to the phyla *Actinobacteria*, *Deinococcus-Thermus*, and *Firmicutes*, composed almost exclusively of monoderm or “simple diderm” bacteria lacking LPS.  While gene-content networks are useful for exploration of genome diversity within the *Tubulavirales* order, they are challenging to use directly for classification of a newly sequenced genome. Instead, we propose using a comparison to a curated set of reference genomes, composed of isolates and high-quality prophages (available at https://genome.jgi.doe.gov/portal/Inovirus/; Roux et al., 2019). If a new genome encodes 2 predicted proteins with significant similarity (blastp, E value < 0.001) to 2 distinct proteins of the same reference genome, it would be classified within the same family as this reference. We verified that, based on the current set of isolates and after removing hits to the same species, this approach would be entirely consistent with the gene-content network displayed in Fig. 1. We also provide a companion tool that performs a gene prediction, comparison to references, and automatic affiliation of a new inovirus sequence (available at https://github.com/simroux/Inovirus/tree/master/Inovirus\_classifier).  Finally, we propose to name the new family as “*Paulinoviridae*” from the latin “paulus, pauli”, meaning “small”, as we noted that members of this new family tend to have shorter genomes and encode less predicted proteins than members of the two other families within the *Tubulavirales* (Fig. 2). This pattern was also observed on the larger dataset, and seems to be a consistent feature of the *Paulinoviridae* (Roux et al., 2019). | |

**Supporting evidence**

****

**Figure 1: Gene content network of *Tubulavirales* isolates and high-quality prophages.** Protein sequences from all inoviruses were compared with an all-vs-all blastp (v2.9.0, E value cutoff of 0.001). Genomes are represented as nodes, colored according to currently defined genera with > 2 genomes, and connected by edges if significant hits were observed on ≥ 2 genes for each genome. The network was displayed in Cytoscape v3.5.0 using an edge-weighted Embedded Layout, with edge weight defined as the cumulative AAI across all reciprocal best hits for each genome pair.

****

**Figure 2: Distribution of genome size (left) and number of predicted genes (right).** Both plots include data for genomes and high-quality prophages in the two existing families (*Inoviridae* and *Plectroviridae*) and the proposed new family (*Paulinoviridae*).

**References**

1. Muhire BM, Varsani A, Martin DP. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One. 2014; 9(9):e108277. doi: 10.1371/journal.pone.0108277. PMID: 25259891
2. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, Sharrar A, Matheus Carnevali PB, Cheng JF, Ivanova NN, Bondy-Denomy J, Wrighton KC, Woyke T, Visel A, Kyrpides NC, Eloe-Fadrosh EA. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol. 2019; 4(11):1895-1906. doi: 10.1038/s41564-019-0510-x. PMID: 31332386