

The International Committee on Taxonomy of Viruses

Taxonomy Proposal Form, 2024

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

|  |  |
| --- | --- |
| **Title:** | Create a new family, *Sarkviridae* for the Jersey-likesiphophages (Class: *Caudoviricetes*) |
| **Code assigned:** | 2024.031B | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Author(s), affiliation and email address(es):** | | | |
| **Name** | **Affiliation** | **Email address** | **Corresponding author(s)** X |
| Moraru C | Carl von Ossietzky Universität Oldenburg, Germany | liliana.cristina.moraru@uol.de |  |
| Tolstoy I | National Center for Biotechnology Information, MD, USA | tolstoy@ncbi.nlm.nih.gov |  |
| Kropinski AM | University of Guelph, Ontario, Canada [AMK] | Phage.Canada@gmail.com | 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:** |
| Caudoviricetes Study Group |

|  |  |  |  |
| --- | --- | --- | --- |
| **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:** | 04/06/2024 |

**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| W - Withdrawn |  |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:** |
| Update figure 3 to include leaf labels. |

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

|  |
| --- |
| **Response of proposer:** |
| Corrected. |

|  |  |
| --- | --- |
| **Revision date:** | 30/09/2024 |

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

|  |
| --- |
| **Name of accompanying Excel module:** |
| 2024.031B.A.v2.Sarkviridae\_nf.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*:  *Jerseyvirus, Cornellvirus* and *Kagunavirus* in the subfamily *Guernseyvirinae;* and, amongst the *Serratia* phages *Seretavirus* and *Otakuvirus*  *Proposed* *taxonomic change(s):*  **A. To create a new family *Sarkviridae***  *Justification*:  Genomic, proteomic and phylogenetic data indicate that this group of phages is a family. In addition, Taxonomy Proposal 2023.068B.Caudoviricetes\_Serratia\_3ng suggested a higher level relationship between the *Otakuvirus* and the *Guernseyvirinae*. |

|  |
| --- |
| * **Text of Taxonomy proposal:** |
| *Taxonomic rank(s) affected*: Species, genus and Family  *Description of current taxonomy*:  At present the following taxa exist: *Jerseyvirus, Cornellvirus* and *Kagunavirus* in the subfamily *Guernseyvirinae;* and, amongst the *Serratia* phages *Seretavirus* and *Otakuvirus*  *Proposed* *taxonomic change(s)*:  **A. To create a new family *Sarkviridae***  *Demarcation criteria:*  **Species demarcation criteria:** Two phages are assigned to the same species if their genomes are more than 95% identical over their genome length for isolates.  These values can be calculated by a number of tools, such as BLASTn [1,2] – usually calculated using intergenomic distance calculator VIRIDIC [3].  **Genus demarcation criteria:** In search for criteria that create cohesive and distinct genera that are reproducible and monophyletic, the Bacterial Viruses Subcommittee has established 70% nucleotide identity of the genome length as the cut-off for genera. Genus-level groupings should always be monophyletic in the signature genes, as tested with a phylogenetic tree. [10]  **Subfamily demarcation criteria:** Subfamilies are to be created when two or more genera are related below the family level. In practical terms, this usually means that they share a low degree of sequence similarity (usually about 40-50%) and that the genera form a clade in a marker tree phylogeny. [10]  **Family demarcation criteria:** The family is represented by a cohesive and monophyletic group in the main predicted proteome-based clustering tools (VirClust, ViPTree, GRAViTy dendrogram, vConTACT2 network). Members of the family share a significant number of orthologous genes (the number will depend on the genome sizes and number of coding sequences of members of the family). [10]  *Justification*: Genomic, proteomic and phylogenetic data indicate that this group of phages is a family. In addition, Taxonomy Proposal 2023.068B.Caudoviricetes\_Serratia\_3ng suggested a higher level relationship between the *Otakuvirus* and the *Guernseyvirinae*. This assignment is in keeping with the criteria laid out in [10] |

|  |
| --- |
| **References:** |
| 1. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, Marchler-Bauer A, Landrum M, Lathrop S, Lu Z, Madden TL, O'Leary N, Phan L, Rangwala SH, Schneider VA, Skripchenko Y, Wang J, Ye J, Trawick BW, Pruitt KD, Sherry ST. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021 Jan 8;49(D1):D10-D17. doi: 10.1093/nar/gkaa892. PMID: 33095870  2. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733-45. doi: 10.1093/nar/gkv1189. PMID: 26553804.  3. Moraru C, Varsani A, Kropinski AM. VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. Viruses. 2020 Nov 6;12(11):1268. doi: 10.3390/v12111268. PMID: 33172115; PMCID: PMC7694805. http://kronos.icbm.uni-oldenburg.de/viridic/  4. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. ViPTree: the viral proteomic tree server. Bioinformatics. 2017; 33(15):2379-2380. doi:10.1093/bioinformatics/btx157. PubMed PMID: 28379287. https://www.genome.jp/viptree/  5. Rohwer F, Edwards R. The Phage Proteomic Tree: a genome-based taxonomy for phage. J Bacteriol. 2002 Aug;184(16):4529-35. PubMed PMID: 12142423  6. Turner D, Reynolds D, Seto D, Mahadevan P. CoreGenes3.5: a webserver for the determination of core genes from sets of viral and small bacterial genomes. BMC Res Notes. 2013;6:140. doi: 10.1186/1756-0500-6-140. PMID: 23566564.  7. Davis P, Seto D, Mahadevan P. CoreGenes5.0: An Updated User-Friendly Webserver for the Determination of Core Genes from Sets of Viral and Bacterial Genomes. Viruses. 2022 Nov 16;14(11):2534. doi: 10.3390/v14112534. PMID: 36423143; PMCID: PMC9693508.  8. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 2008;36(Web Server issue):W465-9. doi: 10.1093/nar/gkn180. Epub 2008 Apr 19. PMID: 18424797.  9. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst Biol. 2006;55(4):539-52. PMID: 16785212. DOI: 10.1080/10635150600755453.  10. Turner D, Kropinski AM, Adriaenssens EM. A Roadmap for Genome-Based Phage Taxonomy. Viruses. 2021 Mar 18;13(3):506. doi: 10.3390/v13030506. PMID: 33803862; PMCID: PMC8003253.  11. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. Nat Biotechnol. 2019 Jun;37(6):632-639. doi: 10.1038/s41587-019-0100-8. Epub 2019 May 6. PMID: 31061483.  12. Bolduc B, Jang HB, Doulcier G, You ZQ, Roux S, Sullivan MB. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. PeerJ. 2017 May 3;5:e3243. doi: 10.7717/peerj.3243. PMID: 28480138; PMCID: PMC5419219.  13. Moraru C. VirClust-A Tool for Hierarchical Clustering, Core Protein Detection and Annotation of (Prokaryotic) Viruses. Viruses. 2023 Apr 19;15(4):1007. doi: 10.3390/v15041007. PMID: 37112988; PMCID: PMC10143988.  14. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007 Jan 1;23(1):127-8. doi: 10.1093/bioinformatics/btl529. Epub 2006 Oct 18. PMID: 17050570.  15. Zhou T, Xu K, Zhao F, Liu W, Li L, Hua Z, Zhou X. itol.toolkit accelerates working with iTOL (Interactive Tree of Life) by an automated generation of annotation files. Bioinformatics. 2023 Jun 1;39(6):btad339. doi: 10.1093/bioinformatics/btad339. PMID: 37225402; PMCID: PMC10243930.  16. Nguyen LT, Schmidt HA, von Haeseler A, and Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Molecular Biology and Evolution, 32:268-274. https://doi.org/10.1093/molbev/msu300  17. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution, 35:518–522. <https://doi.org/10.1093/molbev/msx281>  18. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, and Jermiin JS (2017) ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates, Nature Methods, 14:587–589. https://doi.org/10.1038/nmeth.4285 |

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

<Start here>



Figure 1. VIRIDIC heat map of a small portion of the members of this family: VIRIDIC (Virus Intergenomic Distance Calculator; VIRIDIC (Virus Intergenomic Distance Calculator; [3]; http://rhea.icbm.uni-oldenburg.de/VIRIDIC/) computes pairwise intergenomic distances/similarities amongst phage genomes. Abbreviations: phg = phage; vir = virus; Esch = *Escherichia;* Salm = *Salmonella,* Serr = *Serratia*. The full heatmap is provided as supplementary material.

A circular object with different colored lines

Description automatically generated  
Figure 2. ViPTree [4] analysis Proteomic tree of 4,408 bacterial viruses with proposed viral families labeled by the coloured ring. The *Sarkviridae* are marked with a star symbol. The hierarchical tree was created using ViPTreeGen (version 1.1.2) [4] and annotated using iToL [15-16]. The tree is based on a dissimilarity matrix generated by pairwise tBLASTx scores between each of the genomes.

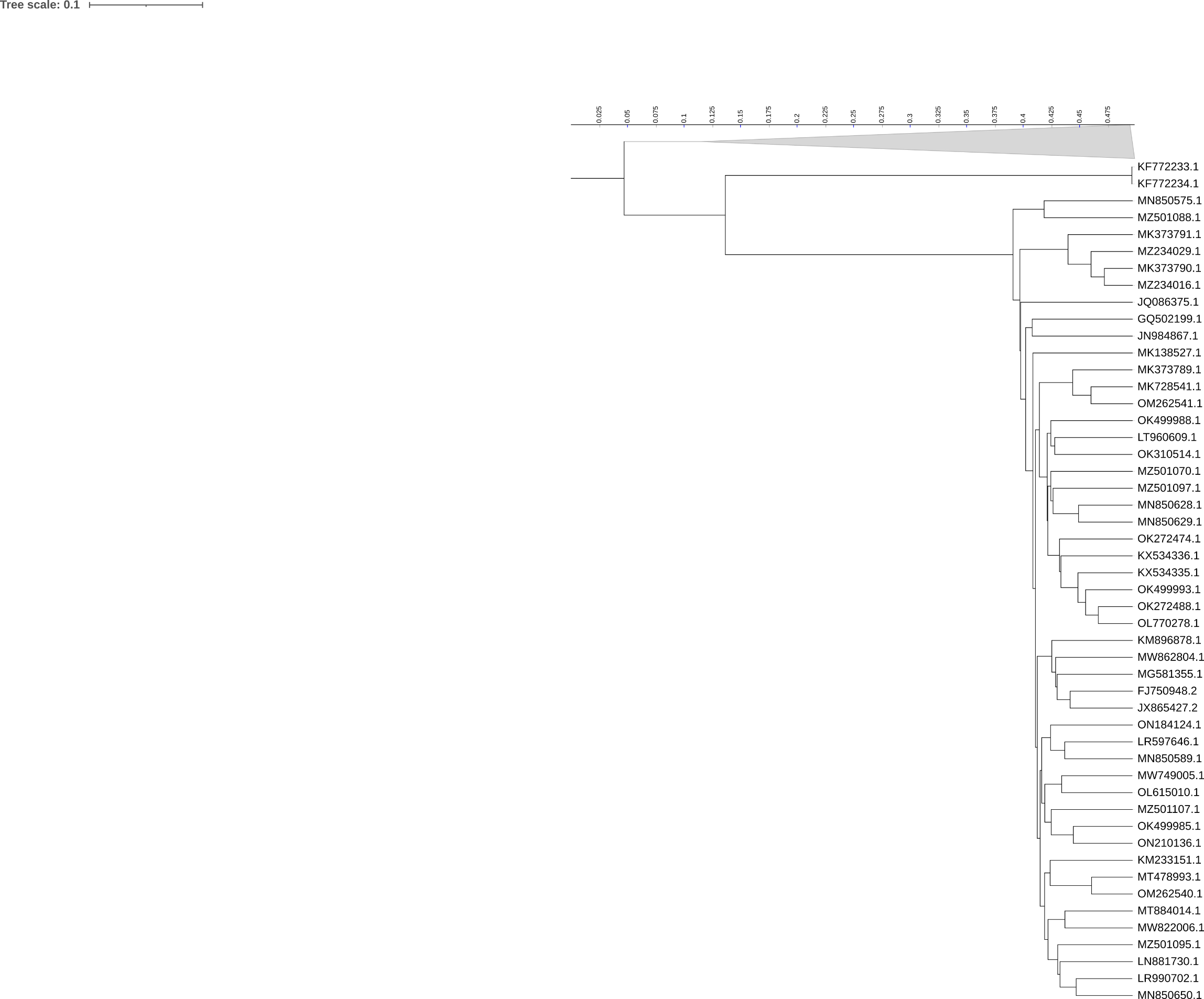
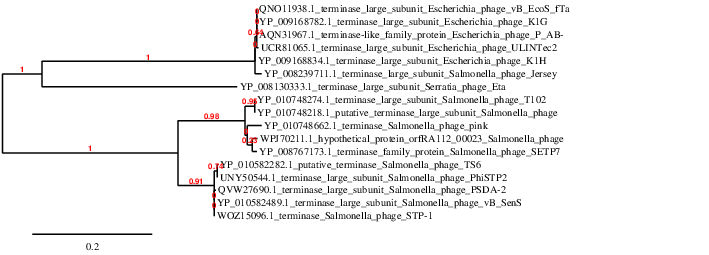


Figure 3. ViPTree [4] hierarchical tree pruned to show the proposed *Sarkviridae,* delineated as a light grey coloured collapsed clade, alongside neighbouring clades.



**Figure 4. Phylogeny:** The phylogenetic tree was constructed using the large subunit terminase (TerL) of a subset of the *Sarkviridae* and related phages with phylogeny.fr in “one click” mode [6]. "The "One Click mode" targets users that do not wish to deal with program and parameter selection. By default, the pipeline is already set up to run and connect programs recognized for their accuracy and speed (MUSCLE for multiple alignment and PhyML for phylogeny) to reconstruct a robust phylogenetic tree from a set of sequences. The usual bootstrapping procedure is replaced by a new confidence index that is much faster to compute. See: Anisimova M., Gascuel O. Approximate likelihood ratio test for branches: A fast, accurate and powerful alternative [7] for details."

A screenshot of a computer

Description automatically generated

Figure 5. Core genome phylogeny of the proposed *Sarkviridae* family of bacterial viruses. A partitioned protein ML phylogeny was created from 3 genes present in all species of the proposed family. Alignments were performed using MAFFT in e-insi mode and trimmed using trimAl with a gap threshold of 0.5. The tree was calculated using IQ-Tree2 with 1000 ultrafast (UF) bootstrap replicates and SH-Alrt tests with -m TEST to optimise models for each alignment [16-18]. The tree is rooted at the midpoint and UF bootstrap support ≥ 95% are shown. The coloured strips indicate proposed genera and subfamilies.



Figure 6. VirClust protein heatmap of representative species of each genus. At the first level, proteins are grouped based on their reciprocal BLASTP similarities into protein clusters, or PCs. At the second level, PCs are grouped based on their Hidden Markov Model (HMM) similarities into protein superclusters, or PSCs. AT the third, still experimental level, PSCs are grouped based on their HMM similarities into protein super-superclusters, or PSSCs [13].

Table 1. Signature genes in the proposed *Sarkviridae* family of bacterial viruses. Genes were identified by clustering with MMSeqs2, with thresholds of 35% sequence similarity and 50% coverage.

|  |  |  |  |
| --- | --- | --- | --- |
| **protein cluster** | **No. of genomes (88 total)** | **Percentage of genomes present in protein cluster** | **Predicted gene function** |
| 1 | 88 | 100% | head-tail joining protein |
| 2 | 88 | 100% | head morphogenesis protein |
| 3 | 88 | 100% | portal vertex protein |
| 4 | 87 | 98.86% | tail component protein |
| 5 | 87 | 98.86% | tail protein |
| 6 | 87 | 98.86% | hypothetical protein |
| 7 | 87 | 98.86% | hypothetical protein |
| 8 | 87 | 98.86% | structural protein |
| 9 | 87 | 98.86% | terminase, large subunit |
| 10 | 87 | 98.86% | hypothetical protein |
| 11 | 87 | 98.86% | DNA helicase |
| 12 | 87 | 98.86% | hypothetical protein |
| 13 | 87 | 98.86% | hypothetical protein |
| 14 | 87 | 98.86% | hypothetical protein |

**Proposals Data:**

**A. To create a new family *Sarkviridae***

1. **To create a new family *Sarkviridae* for these taxa.**

**Origin of the name of this taxon:** This taxon *is named to be consistent with the use of the Channel Island theme – Jerseyvirus*, *Guernseyvirinae*, and now *Sarkviridae*, named after the island of Sark.

**Rationale:** Genomic, proteomic and phylogenetic data indicate that this group of phages is a family. In addition, Taxonomy Proposal 2023.068B.Caudoviricetes\_Serratia\_3ng suggested a higher level relationship between the *Otakuvirus* and the *Guernseyvirinae*.