This Word module should be used for all taxonomic proposals.

Please complete **Part 1** and:

either **Part 3** for proposals to create new taxa or change existing taxa

or **Part 2** for proposals of a general nature.

Submit the completed Word module, together with the accompanying Excel module named in Part 3, to the appropriate ICTV Subcommittee Chair.

The Word module explains and justifies your proposal. The Excel module is a critical document that will be used to implement the proposed taxonomic changes once they are approved and ratified. If proposals presented in the Word module are not presented accurately in the Excel module, the taxonomic changes cannot proceed.

For guidance, see the notes written in blue, below, and the Help Notes in file Taxonomic\_Proposals\_Help\_2019.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Code assigned:** | ***2019.011D*** | |  |
| **Short title:** Create one new family (*Redondoviridae*) for circular, Rep-encoding DNA viruses | | | |
|  | | | |
| **Author(s) and email address(es):** | | | |
| List authors in a single line *Archives of Virology* citation format (e.g. Smith AB, Huang C-L, Santos, F) | | Provide email address for each author in a single line separated by semi-colons | |
| Abbas AA, Taylor LJ, Collman RG, Bushman FD | | [arwaa@pennmedicine.upenn.edu](mailto:arwaa@pennmedicine.upenn.edu); [louist@pennmedicine.upenn.edu](mailto:louist@pennmedicine.upenn.edu); [collmanr@pennmedicine.upenn.edu](mailto:collmanr@pennmedicine.upenn.edu); [bushman@pennmedicine.upenn.edu](mailto:bushman@pennmedicine.upenn.edu) | |
| **Author(s) institutional address(es) (optional):**   |  | | --- | | Provide institutional addresses, each on a single line followed by author(s) initials (e.g. University of Woolloomooloo [SAB, HCL]) | | Department of Microbiology, University of Pennsylvania Perelman School of Medicine [LJT, AAA, FDB]  Department of Medicine, Pulmonary, Allergy and Critical Care Division, University of Pennsylvania Perelman School of Medicine [RGC] | | | | |
| **Corresponding author** | | | |
| Frederic D. Bushman | | | |
| **List the ICTV study group(s) that have seen this proposal:** | | | |
| A list of study groups and contacts is provided at <http://www.ictvonline.org/subcommittees.asp> . If in doubt, contact the appropriate subcommittee chair (there are six virus subcommittees: animal DNA and retroviruses, animal ssRNA-, animal ssRNA+, fungal and protist, plant, bacterial and archaeal) | | Circoviridae Study Group chair (Arvind Varsani)  Animal DNA Viruses and Retroviruses Subcommittee chair (Balázs Harrach) | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | |
|  | | | |
|  | | | |
| Date first submitted to ICTV: | | | 7 May 2019 |
| Date of this revision (if different to above): | | | 14 October 2019 |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
| Family is not a problem, it is how many species - it should be two species not two genera or two families.  **Response**: Accepted; TP rewritten for one genus including two species. |

**Part 3:** **PROPOSED TAXONOMY**

|  |
| --- |
| **Name of accompanying Excel module:** 2019.011D.A.v1.Redondoviridae.xlsx |

The taxonomic changes you are proposing should be presented on an accompanying Excel module, 2019\_TP\_Template\_Excel\_module. Please enter the file name of the completed module in this box.

**Supporting material:**

| additional material in support of this proposal |
| --- |
| Please explain the reasons for the taxonomic changes you are proposing and provide evidence to support them. The following information should be provided, where relevant:   * **Species demarcation criteria**: Explain how new species differ from others in the genus and demonstrate that these differences meet the criteria previously established for demarcating between species. If no criteriahave previously been established, and if there will now be more than one species in the genus, please state the demarcation criteria you are proposing. * **Higher taxa**:   + There is no formal requirement to state demarcation criteria when proposing new genera or other higher taxa. However, a similar concept should apply in pursuit of a rational and consistent virus taxonomy.   + Please indicate the **origin of names** assigned to new taxa at genus level and above.   + For each new genus a **type species** must be designated to represent it. Please explain your choice. * **Supporting evidence**: The use of Figures and Tables is strongly recommended (note that copying from publications will require permission from the copyright holder). For phylogenetic analysis, please provide a tree where branch length is **proportional to genetic** distance, generated using an appropriate algorithm (Neighbour-Joining, Maximum Likelihood, or Bayesian) and provide evidence of the reliability of the branching (e.g., by bootstrapping).   Please refer to the Help Notes file (Taxonomic\_Proposals\_Help\_2019) for more information. |

**Introduction**

Metagenomic analysis, have uncovered numerous small, circular DNA viruses in diverse environments1–3. New families of circular, Rep-encoding single-strand (CRESS) DNA viruses, including *Genomoviridae*4 and *Smacoviridae*5, were recently discovered through metagenomic approaches. Here, we propose the establishment of a new family of CRESS DNA viruses called *Redondoviridae*, which have been identified mostly in human oral and respiratory tract samples through metagenomic sequencing, and validated by targeted genome amplification from human samples6.

**Discovery of 19 new genomes and prevalence across organisms and sample types**

In two studies focusing on the viral microbiome of lung transplant donors and recipients7,8, we identified metagenomic sequence reads from two samples that aligned with limited coverage (<20%) of an unclassified CRESS DNA virus, porcine stool-associated circular virus 5 (PoSCV-5)9, NCBI nucleotide accession number KJ433989. After building contigs with reads from those samples, we recovered complete or near-complete circular genomes, suggesting that the limited coverage of the reference sequence resulted from low homology between the viruses. We confirmed both genome sequences by PCR and Sanger sequencing. We used them to query additional metagenomic sequences from bronchoalveolar lavage (BAL) samples from human subjects previously analysed by our group, resulting in the identification and cloning of five additional genomes. Thus, a total of seven novel genomes were identified and used in subsequent searches.

A danger is that small circular viruses may be derived from environmental contaminants in laboratory reagents10,11. We thus used qPCR targeting these novel genomes to test 24 negative controls that reflect the entire sample procurement and specimen processing pipeline, sterile saline solution that had passed through bronchoscopes prior to use for bronchoscopy—and failed to detect these sequences. Additionally, 132 contamination controls generated in metagenomics sequencing studies by ourselves and others were interrogated for these sequences by alignment of reads; no positives were found.

To identify more of these divergent viruses and assess their abundance in different niches, we aligned reads from more than 6000 publicly available metagenomic samples from different human body sites, animals, environments, and reagent controls to the novel genomes. We identified more than 60 positive samples (using a threshold of >25% genome coverage) including 12 samples with sufficient coverage to construct additional complete genomes. This group appears to be specific to the human oro-respiratory tract. Positive hits were primarily found in human oral, lung, and nasopharyngeal samples, with fewer hits in human gut samples. We identified no positives in other human body sites, other animals, environmental metagenomes, or reagent controls. Concurrently, another group independently published a single similar genome sequence identified in respiratory samples from a febrile human patient12: human respiratory-associated brisavirus (HRAPLV) isolate LC (proposed name), NCBI nucleotide accession number KY052047, consistent with our detections in the human respiratory tract. Two additional similar genomes have been deposited in the NCBI database by another independent group (NCBI nucleotide accession numbers KY244146.1 and KY349925.1). Thus, a total of 22 novel genomes fall within the family proposed in this report.

**Evidence for a new family by comparison to existing families**

To determine whether these genomes belong in an existing viral family, we compared genomic features including genome size, open reading frames and orientation, nucleotide and protein sequence, and predicted replication origin sequence and location, between this group and other CRESS DNA virus families including *Circoviridae*, *Smacoviridae*, *Geminiviridae*, *Genomoviridae*, and *Nanoviridae*. A detailed comparison is shown in Table 1 and a genome map showing nucleotide identity across all 22 genomes is shown in Figure 1. While some families share one or two features with our novel group, no existing family matches more than two of the above attributes. Additionally, no currently characterized family includes the presence of a conserved ORF, overlapping the capsid gene in this new group. This ORF displays no homology to any known protein.

Assessment of both replication-associated protein (Rep) and capsid (Cp) phylogenies (Figures 2 and 3, protein alignment by MUSCLE, phylogeny built using PhyML and visualized using FigTree13–15) reveals that this group forms a clade distinct from other CRESS DNA virus families. Based on the above, we propose the creation of a new family named *Redondoviridae*, which includes all of our 19 virus strains and the three strain discovered by others. This name refers to the circular nature of the genomes—*redondo* is the Spanish word for “round”.

**Species demarcation criteria; nomenclature**

Based on feedback from the ICTV on the initial submission of this proposal, we propose that the members of the *Redondoviridae* family proposed here should consist of two species within a single genus. In Figure 4, we show the distributions of pairwise identities across Rep and Cp protein and genome sequences. Pairwise identities were calculated by the global Needleman–Wunsch algorithm in R using the Biostrings package16. The Rep pairwise identities show a distribution with a valley at 50% identity, which is similar to the 40% threshold used to divide *Smacoviridae* genera5. Based on feedback from the ICTV EC, we use this cutoff as our species demarcation criterion. A 50% identity cut-off results in two robust species, which we propose to be named *Vientovirus* and *Brisavirus*. *Viento* means “wind” and *brisa* means “breeze” in Spanish, alluding to the prevalence of these viruses in the respiratory tract. We propose the name *Torbevirus* for the genus containing these species, from the Spanish *torbellino*, meaning “whirlwind”, as this incorporates both the circular nature of redondovirus genomes, plus the prevalence of redondoviruses in the respiratory tract.

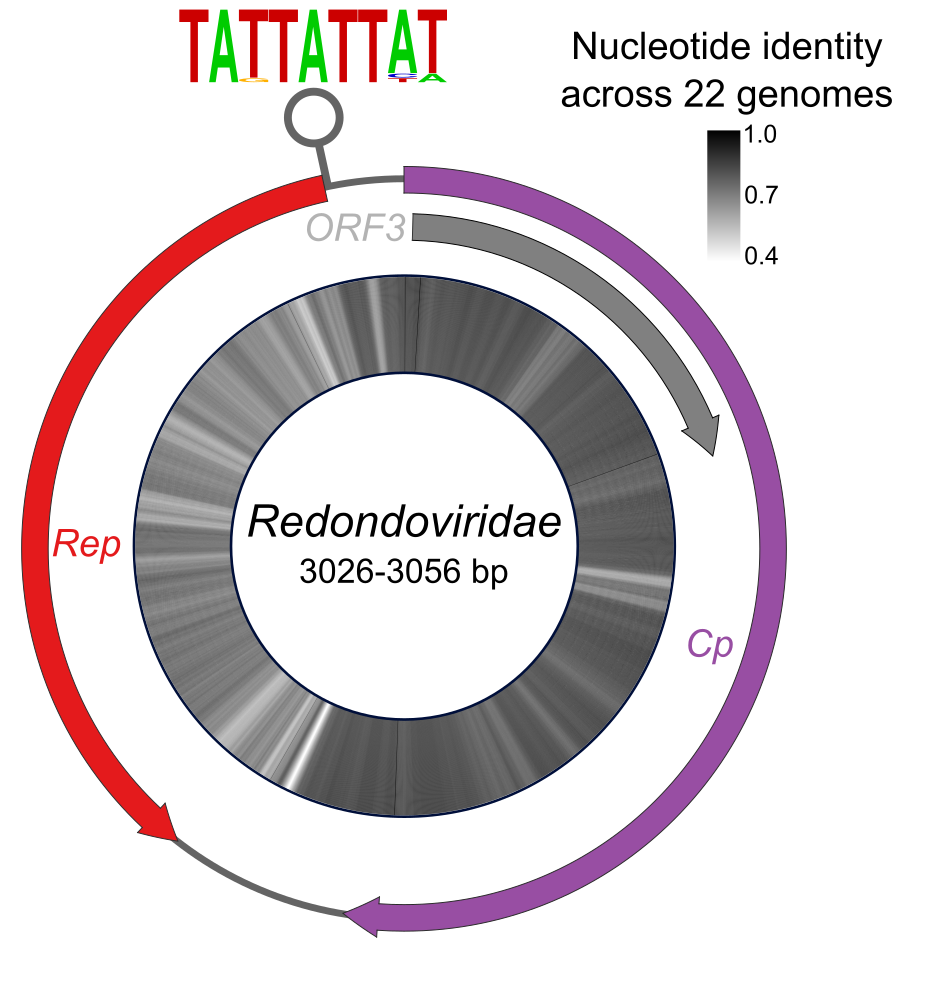
Genome identity is often used for species and type demarcation4,5. In the case of *Redondoviridae*, the process of choosing a genome identity cutoff is complicated by the fact that redondovirus Rep and Cp phylogenies do not show the same relationship between viral genomes (Figures 2, 3, and 5), potentially due to recombination found common in other CRESS viruses17. Additionally, most of the 3 kb genome is coding—thus, the genomic nucleotide sequence phylogeny is likely an intermediate between the Rep and Cp trees. Given these difficulties and the current lack of published characterization supporting biologically-relevant type-level differences, we do not currently propose a type cutoff, but show potential options for such in Figure 5.

Using these criteria, we propose the creation of 2 redondovirus species from the 22 genomes identified. Table 2 enumerates all members of this proposed new family with associated references and metadata. We selected the human respiratory-associated brisavirus LC (GenBank acc. KY052047; formerly HRAPLV) as the exemplar member of a new species *Brisavirus,* and selected *Brisavirus* as the type species of the proposed genus *Torbevirus* as this was the first redondovirus genome described. We designate human lung-associated vientovirus FB (GenBank acc. MK059763) as the exemplar member of the other presently proposed species *Vientovirus*, as it was the first vientovirus we discovered.

In the case of PoSCV-59, we do not recommend any classification presently. While the Rep protein identity and genome size fall within the bounds of other proposed *Redondoviridae* family members, the Cp protein is much less similar (~40-50% identity with members of *Redondoviridae*).(This value is >70% identity among the Cp proteins of the proposed members of *Redondoviridae*). Additionally, PoSCV-5 lacks an open reading frame overlapping Cp, whereas all members of *Redondoviridae* possess this ORF. Based on these observations, we do not think there is a strong case for proposing an official taxonomy for PoSCV-5 currently—future genomes discovered may provide evidence for or against including PoSCV-5 in the family *Redondoviridae*.

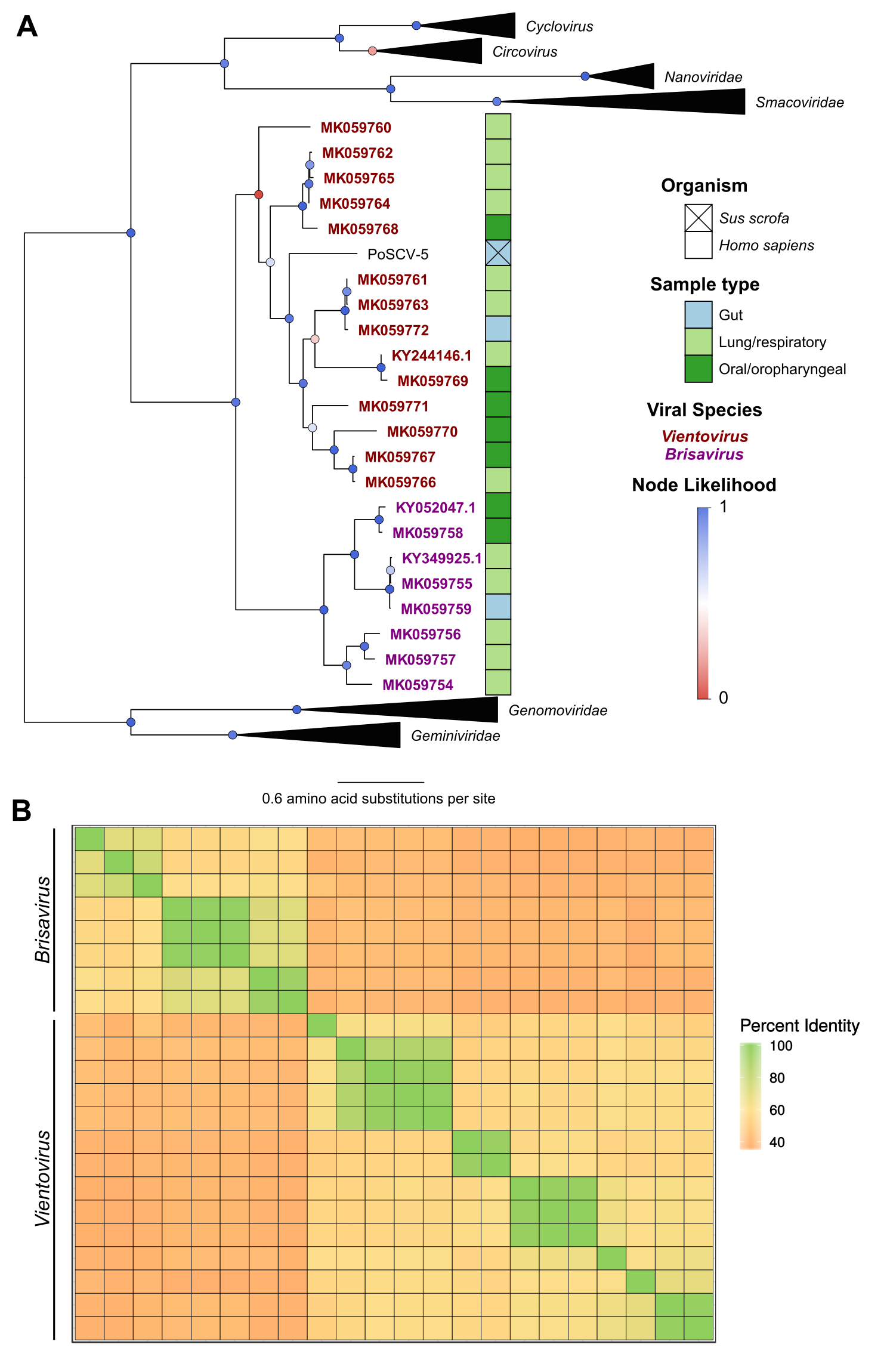
**Figures and Tables**

**Figure 1: Genome map of members of *Redondoviridae***

****

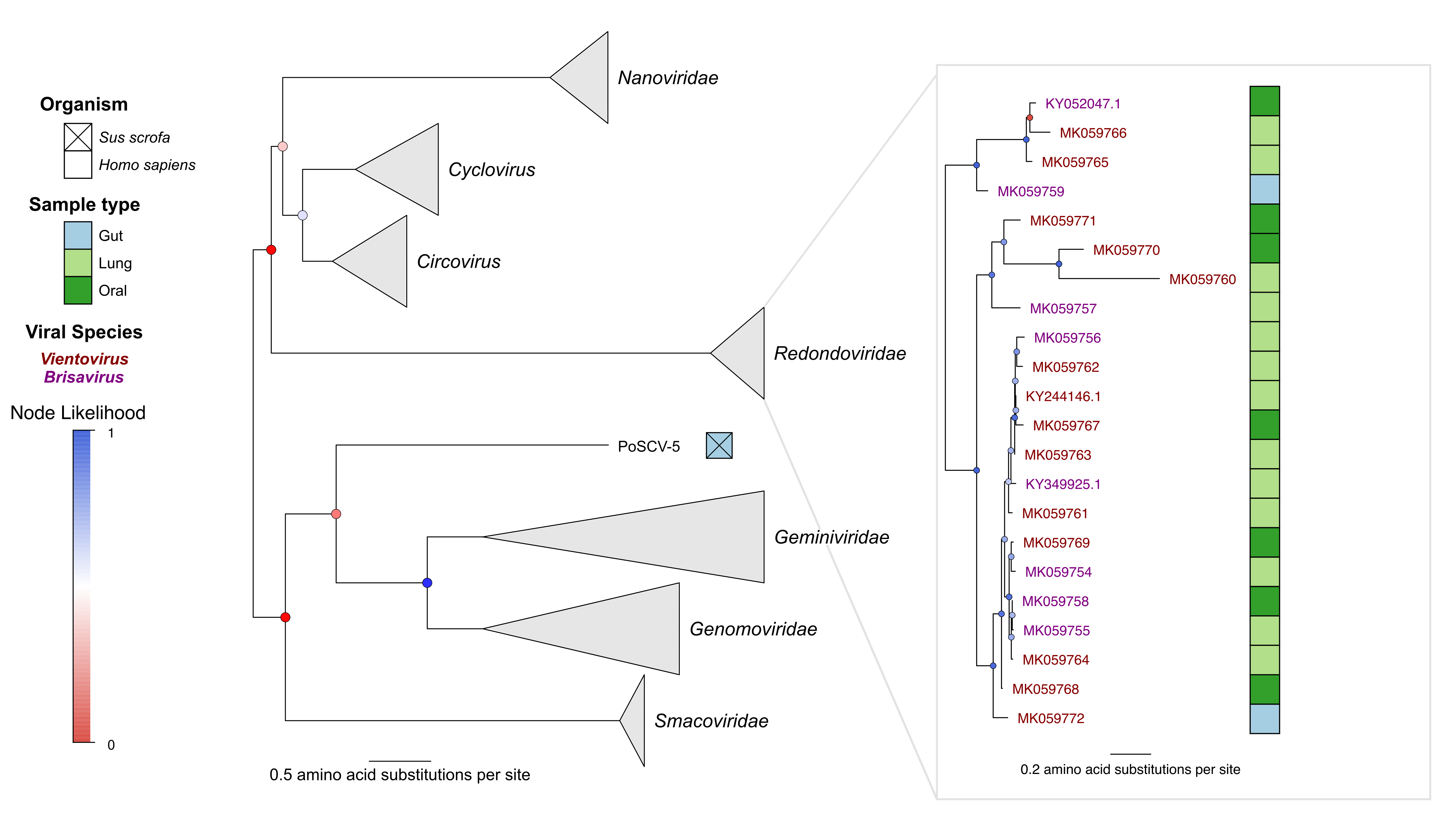
Genome organization of members of *Redondoviridae* (outer ring) with percent nucleotide identity after multiple sequence alignment by MUSCLE13 (inner ring). The conserved nonanucleotide motif is shown as a sequence logo above the stem loop. Redondoviruses encode three proteins: Rep, Cp, and an ORF3 of unknown function.

**Figure 2: *Redondoviridae* Rep protein amino acid sequence phylogeny and identity matrix**



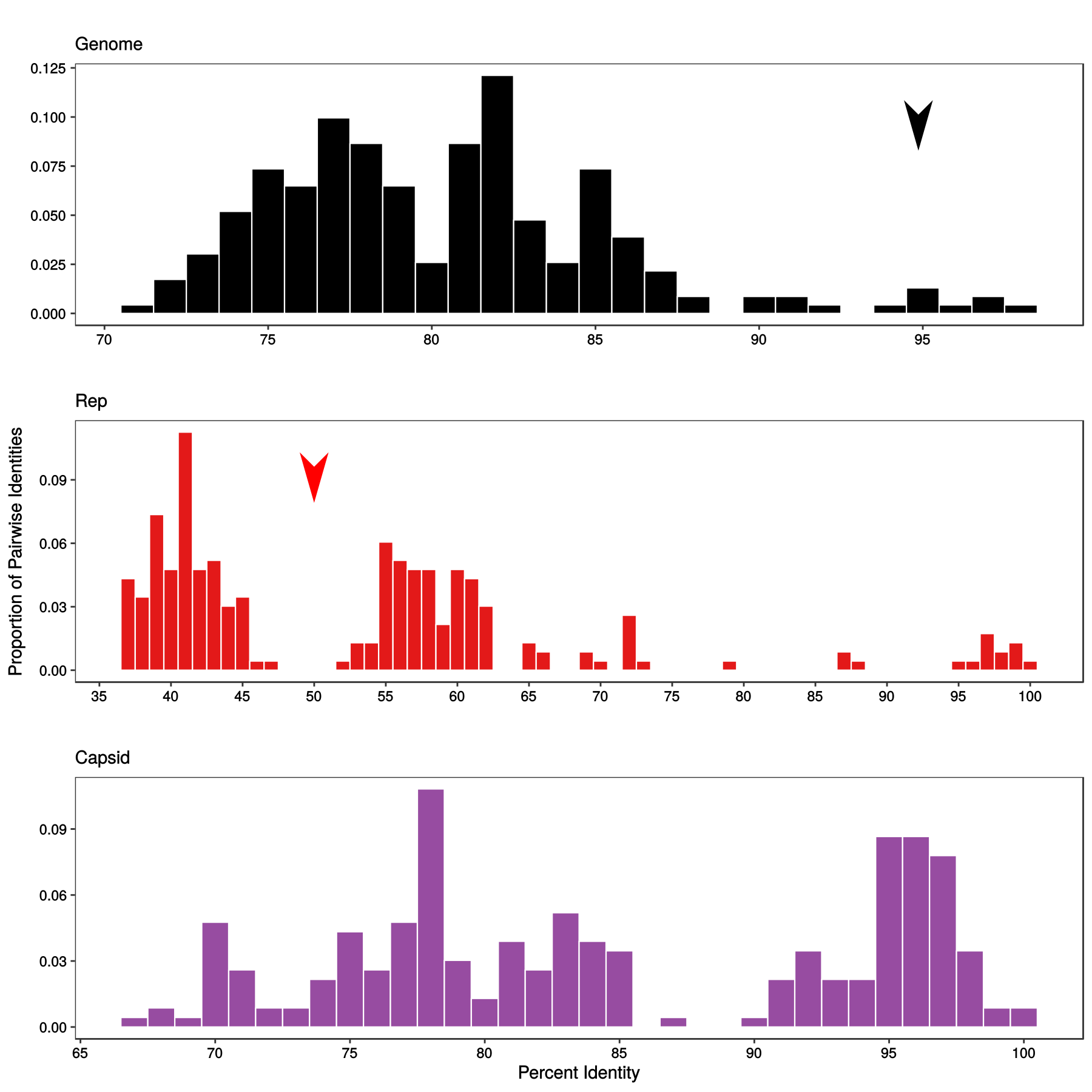
**A.** Phylogenetic tree of Rep proteins from CRESS DNA viruses (2-3 representatives per family), compared with PoSCV-5 and all 22 redondoviruses. Amino acid sequences were aligned with MUSCLE13, trees built using PHYML with branch support determined by approximate likelihood ratio test14 and visualized using FigTree15. **B.** Pairwise identity matrix of redondovirus Rep proteins showing segregation into two distinct groups. Pairwise alignment performed by global Needleman–Wunsch algorithm as implemented in the Biostrings R package16.

**Figure 3: *Redondoviridae* Cp protein amino acid sequence phylogeny**

****

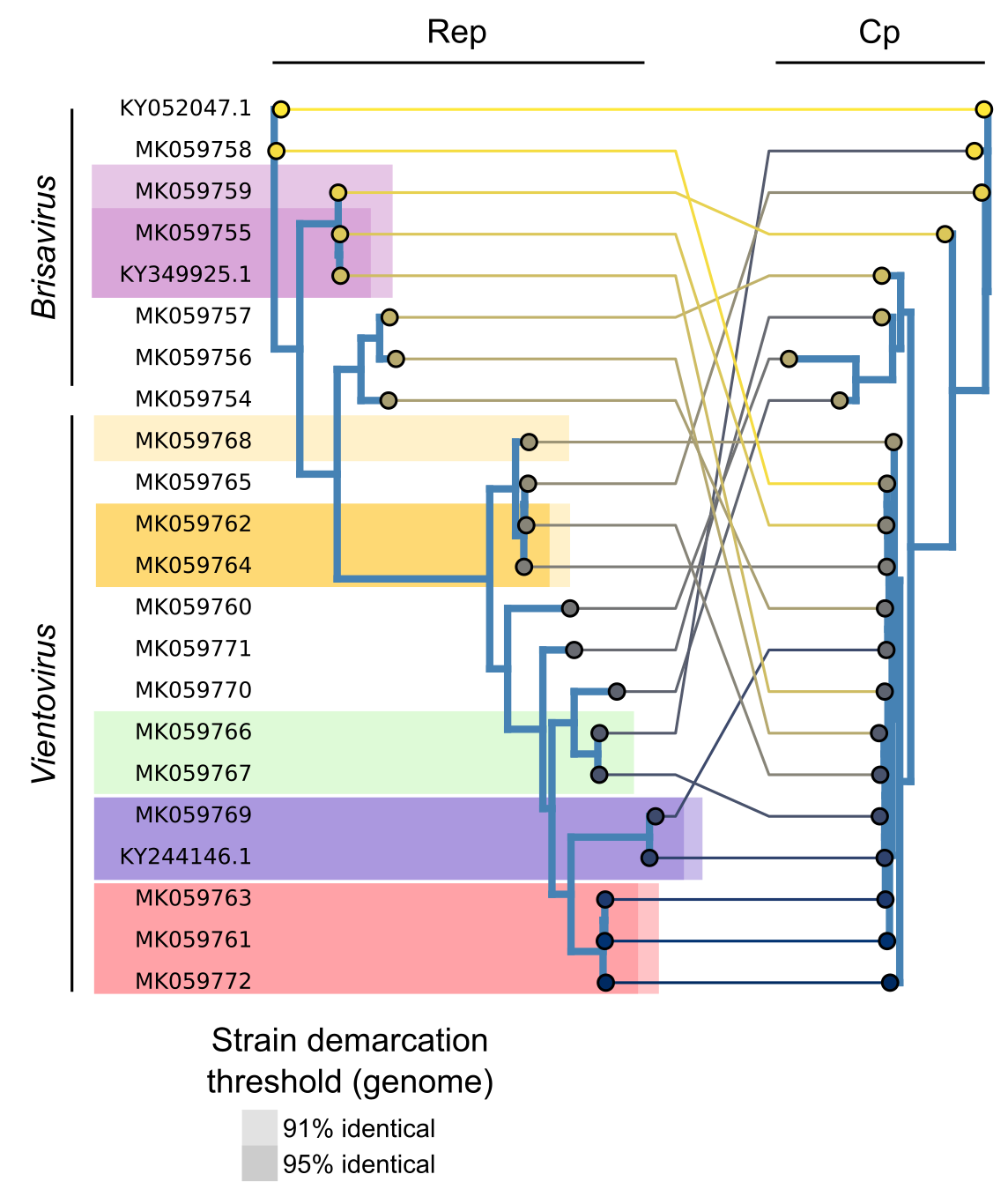
Phylogenetic tree of Cp proteins from CRESS DNA viruses (2-3 representatives per family) and redondoviruses. Amino acid sequences were aligned with MUSCLE13, trees built using PHYML with branch support determined by approximate likelihood ratio test14, and visualized using FigTree15. Redondovirus Cp phylogeny shown in inset at larger scale.

**Figure 4: Distribution pairwise identities of genomic nucleotide and Cp/Rep protein sequence identities**



Pairwise identities of 22 *Redondoviridae* Rep and capsid amino acid sequences, and genomic nucleotide sequences. Pairwise alignment performed by global Needleman–Wunsch algorithm as implemented in the Biostrings R package16. The red arrow above the Rep plot denotes the proposed species demarcation cutoff. The black arrow above the genome plot indicates a potential type cutoff.

**Figure 5: Comparative phylogeny between redondovirus Rep and Cp proteins**



Comparative phylogeny of redondovirus Rep (left) and Cp (right) proteins exemplifying potential type demarcation cutoffs. Phylogenetic trees of redondovirus proteins generated as in Figures 2 and 3 were compared using baltic (<https://github.com/evogytis/baltic>)18. Colored boxes denote isolates that fall within example type demarcation cutoffs (91% or 95%); each color designates a possible different type. Isolates that would be members of the same type at a 91% genome identity cutoff are boxed with a light background; isolates of the same species at a 95% genome identity cutoff are boxed in a darker background.

**Table 1: Comparison of genomic features between *Redondoviridae* and other CRESS-DNA viruses**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Feature** | *Redondoviridae* | *Circoviridae* | *Nanoviridae* | *Geminiviridae* | *Genomoviridae* | *Smacoviridae* |
| **Size (kb)** | 3.0-3.1 | 1.7-2.0 | 1.0 \* 6 segments | 2.5-3.0 | 2.1-2.2 | 2.6-2.9 |
| **ORFs** | Cp, Rep, ORF3 | Cp, Rep, ORF3/4 | Cp, Rep, others | Cp, Rep, others | Cp, Rep | Cp, Rep |
| **ORF orientation** | Ambisense | Ambisense | Segmented | Ambisense (or segmented) | Ambisense | Ambisense |
| **Origin sequence** | TATTATTAT | TAGTATTAC | TATTATTAC | TAATATTAC | TAATATTAT | NAGTATTAC |
| **Origin location** | Noncoding (upstream) / in Rep | Noncoding (upstream) / in Rep | Noncoding (upstream) | Noncoding (upstream) | Noncoding (upstream) | Noncoding (downstream) |

**Table 2: List of redondovirus genomes and metadata**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Genus** | **Species** | **Representative virus** | **Strain/ isolate** | **GenBank Accession** | **Country** |
| Torbevirus | **Brisavirus** | Human respiratory-associated brisavirus, isolate LC | LC | KY052047 | China |
| Torbevirus | Brisavirus | Human respiratory-associated brisavirus, isolate II | II | MK059755 | USA |
| Torbevirus | Brisavirus | Human lung-associated brisavirus, isolate RC | RC | MK059757 | USA |
| Torbevirus | Brisavirus | Human lung-associated brisavirus, isolate AA | AA | MK059754 | USA |
| Torbevirus | Brisavirus | Human lung-associated brisavirus, isolate MD | MD | MK059756 | USA |
| Torbevirus | Brisavirus | Human gut-associated brisavirus, isolate VW | VW | MK059759 | USA |
| Torbevirus | Brisavirus | Human oral-associated brisavirus, isolate YH | YH | MK059758 | USA |
| Torbevirus | Vientovirus | Human lung-associated vientovirus, isolate FB | FB | MK059763 | USA |
| Torbevirus | Vientovirus | Human lung-associated vientovirus, isolate DC | DC | MK059761 | USA |
| Torbevirus | Vientovirus | Human gut-associated vientovirus, isolate MW | MW | MK059772 | UK |
| Torbevirus | Vientovirus | Human lung-associated vientovirus, isolate JY | JY | MK059765 | USA |
| Torbevirus | Vientorvirus | Human lung-associated vientovirus, isolate ES | ES | MK059762 | USA |
| Torbevirus | Vientorvirus | Human lung-associated vientovirus, isolate JB | JB | MK059764 | USA |
| Torbevirus | Vientorvirus | Human lung-associated vientovirus, isolate AL | AL | MK059760 | USA |
| Torbevirus | Vientorvirus | Human lung-associated vientovirus, isolate LT | LT | MK059766 | USA |
| Torbevirus | Vientovirus | Human oral-associated vientovirus, isolate EC | EC | MK059768 | USA |
| Torbevirus | Vientovirus | Human oral-associated vientovirus, isolate XM | XM | MK059771 | China |
| Torbevirus | Vientovirus | Human oral-associated vientovirus, isolate AV | AV | MK059767 | Spain |
| Torbevirus | Vientovirus | Human oral-associated vientovirus, isolate MC | MC | MK059770 | USA |
| Torbevirus | Vientovirus | Human oral-associated vientovirus, isolate LZ | LZ | MK059769 | USA |
| Torbevirus | Vientovirus | Human oral-associated vientovirus, isolate 15040 | 15040 | KY244146.1 | USA |
| Torbevirus | Vientovirus | Human respiratory-associated vientovirus, isolate 15278 | 15278 | KY349925.1 | USA |

The type species for the genus is in bold; the exemplar viruses for both proposed species are underlined.

**Acknowledgements**

This work was supported by NIH grants R61-HL137063 (Lung Virome in Health and Disease) and R01-HL113252 (Lung Transplant Microbiome and Chronic Allograft Dysfunction). A.A.A was supported by NSF grant DGE-1321851 and L.J.T by T32-AI-007324. We thank J. Christie and members of the lung transplant program for access to transplant specimens and ICU staff for assistance with specimen collection. We are grateful to members of the Bushman and Collman laboratories for suggestions, Arvind Varsani and Balázs Harrach for help with virus taxonomy, and for the laboratories of Xiang-Jin Meng and Matthew Weitzman for assistance and support.

| **References** |
| --- |
| 1. Rosario, K., Duffy, S. & Breitbart, M. A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Arch. Virol.* **157**, 1851–1871 (2012).  2. Delwart, E. & Li, L. Rapidly expanding genetic diversity and host range of the Circoviridae viral family and other Rep encoding small circular ssDNA genomes. *Virus Res.* **164**, 114–121 (2012).  3. Simmonds, P. *et al.* Consensus statement: Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* **15**, 161–168 (2017).  4. Krupovic, M., Ghabrial, S. A., Jiang, D. & Varsani, A. Genomoviridae: a new family of widespread single-stranded DNA viruses. *Arch. Virol.* **161**, 2633–2643 (2016).  5. Varsani, A. & Krupovic, M. Smacoviridae: a new family of animal-associated single-stranded DNA viruses. *Arch. Virol.* **163**, 2005–2015 (2018).  6. Abbas, A. A. *et al.* Redondoviridae, a family of small, circular DNA viruses of the human oro-respiratory tract associated with periodontitis and critical illness. *Cell Host Microbe* **25**, 719–729 (2019).  7. Abbas, A. A. *et al.* The perioperative lung transplant virome: torque teno viruses are elevated in donor lungs and show divergent dynamics in primary graft dysfunction. *Am. J. Transplant.* **17**, 1313–1324 (2017).  8. Abbas, A. A. *et al.* Bidirectional transfer of Anelloviridae lineages between graft and host during lung transplantation. *Am. J. Transplant.* **19**, 1086–1097 (2018).  9. Cheung, A. K. *et al.* Identification of a novel single-stranded circular DNA virus in pig feces. *Genome Announc.* **2**, e00347-14 (2014).  10. Phan, T. *et al.* The perils of pathogen discovery: Origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin c olumns. *J. Virol.* **87**, 11966–11977 (2013).  11. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 1–12 (2014).  12. Cui, L. *et al.* Identification and genetic characterization of a novel circular single-stranded DNA virus in a human upper respiratory tract sample. *Arch. Virol.* **162**, 3305–3312 (2017).  13. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).  14. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).  15. Rambaut, A. FigTree, a graphical viewer of phylogenetic trees. Available at: http://tree.bio.ed.ac.uk/software/figtree/. (Accessed: 10th July 2018)  16. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: Efficient manipulation of biological strings. (2017).  17. Martin, D. P. *et al.* Recombination in eukaryotic single stranded DNA viruses. *Viruses* **3**, 1699–1738 (2011).  18. Dudas, G., Bedford, T., Lycett, S. & Rambaut, A. Reassortment between influenza B lineages and the emergence of a coadapted PB1-PB2-HA gene complex. *Mol. Biol. Evol.* **32**, 162–172 (2015). |