This form should be used for all taxonomic proposals. Please complete all those modules that are applicable.



For guidance, see the notes written in blue and the separate document “Help with completing a taxonomic proposal”

Please try to keep related proposals within a single document.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Code assigned:** | ***2017.001D*** | | | | (to be completed by ICTV officers) |
| **Short title: Create 3 species within genus *Semotivirus* and move genus from family *Metaviridae* to a new family *Belpaoviridae*** | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| Mart Krupovic,  Carlos Llorens,  Eugene V. Koonin,  Jens H. Kuhn | | | | | |
| **Corresponding author with e-mail address:** | | | | | |
| Mart Krupovic; E-mail: [krupovic@pasteur.fr](mailto:krupovic@pasteur.fr) | | | | | |
| **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) | | | **Chair of the Animal DNA Viruses and Retroviruses Subcommittee (N.B.: currently there is no Study Group responsible for the family *Metaviridae*)** | | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | | | |
|  | | | | | |
|  | | | | | |
| Date first submitted to ICTV: | | | | June 8, 2017 | |
| Date of this revision (if different to above): | | | |  | |

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

**Part 2**: **PROPOSED TAXONOMY**

|  |
| --- |
| Present the proposed new taxonomy on accompanying spreadsheet |
| **Name of accompanying spreadsheet: 2017.001D.N.v1.Belpaoviridae.xlsx** |

Please display the taxonomic changes you are proposing on the accompanying spreadsheet module 2017\_TP\_Template\_Excel\_module. Submit both this and the spreadsheet to the appropriate ICTV Subcommittee Chair.

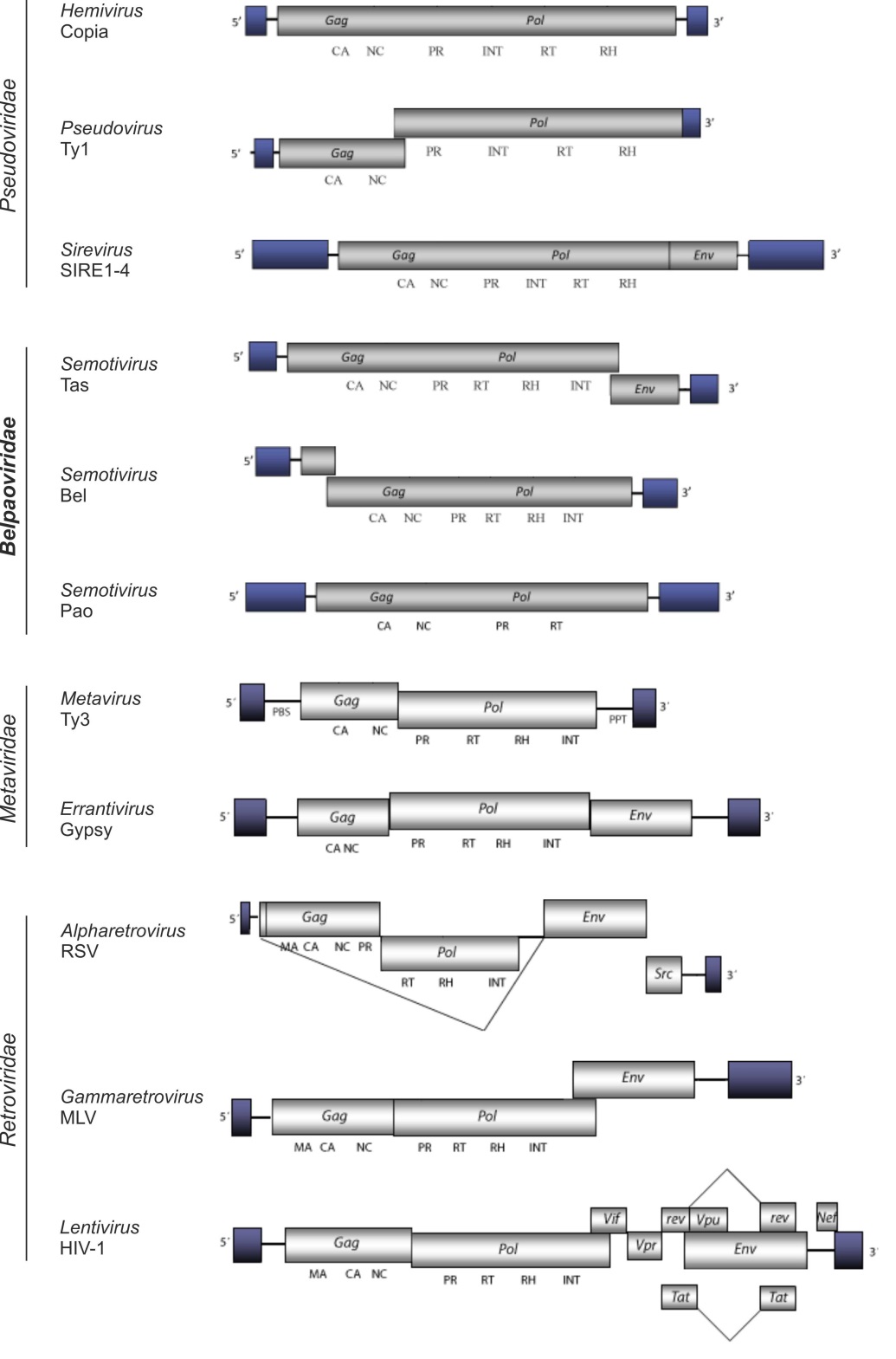
**Part 4:** **APPENDIX**: supporting material

The family *Metaviridae* has been established in 1998 for reverse-transcribing ssRNA viruses that have been found in all major eukaryotic taxa. The family originally included two genera, *Metavirus* and *Errantivirus*, typified by Saccharomyces cerevisiae Ty3 virus and Drosophila melanogaster gypsy virus, respectively (<https://talk.ictvonline.org/ICTV/proposals/Ratification_1998.pdf>). In 2004, a third genus, *Semotivirus* (typified by Ascaris lumbricoides Tas virus) was added to the family (<https://talk.ictvonline.org/ICTV/proposals/2003.F187-190.Semotivirus.pdf>). Viruses of the three genera have similar genome organization, with the *gag* gene encoding a structural polyprotein being followed by the *pol* gene encoding a polyprotein with protease, reverse transcriptase (RT), ribonuclease H, and the integrase (INT) domains (Eickbush and Jamburuthugoda, 2008). Some family members also possess *env* genes encoding surface proteins. A similar genome organization is also typical of members of the families *Pseudoviridae* and *Retroviridae*, even though in the case of pseudoviruses the INT domain is found upstream of the RT domain, whereas in metaviruses and retroviruses the INT domain is located downstream of the RT domain (Figure 1). Similar to retroviruses and pseudoviruses, metavirus genomes have long terminal repeats (LTR).

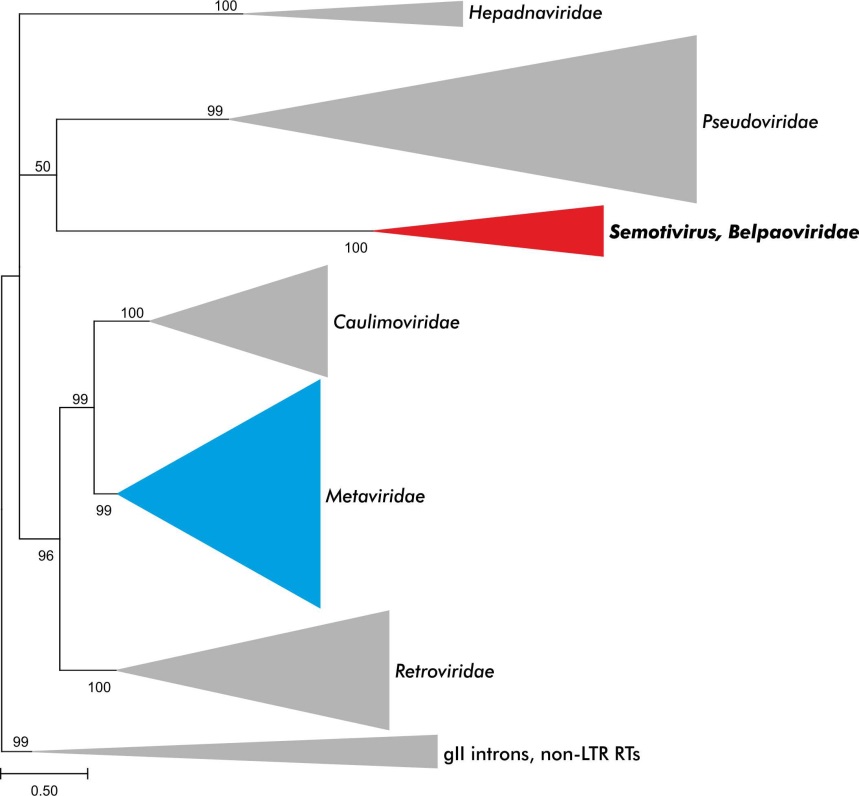
In phylogenies based on RT, the signature gene of reverse-transcribing viruses, semotiviruses are not monophyletic with errantiviruses and metaviruses. This has been recognized in the original proposal establishing the genus *Semotivirus* (<https://talk.ictvonline.org/ICTV/proposals/2003.F187-190.Semotivirus.pdf>). Due to similarities in the gene content and order of the functional domains of the Pol polyprotein, it was considered to be more appropriate at that time to include the genus *Semotivirus* in the family *Metaviridae*. However, such a placement does not pay justice to the distinct evolutionary history of semotiviruses and is at odds with the efforts of the ICTV to enforce phylogeny-guided classification practices.

Thus, we propose moving the genus *Semotivirus* into a separate new family “*Belpaoviridae*” (after viruses Bel and Pao, two representatives of the *Semotivirus* genus). Such a classification would be consistent with the one adopted for retrotransposons, in which semotiviruses are referred to as Bel/Pao elements and are typically considered as belonging to a family different from the Ty3/gypsy (*Metaviridae*) elements (Llorens et al., 2009).

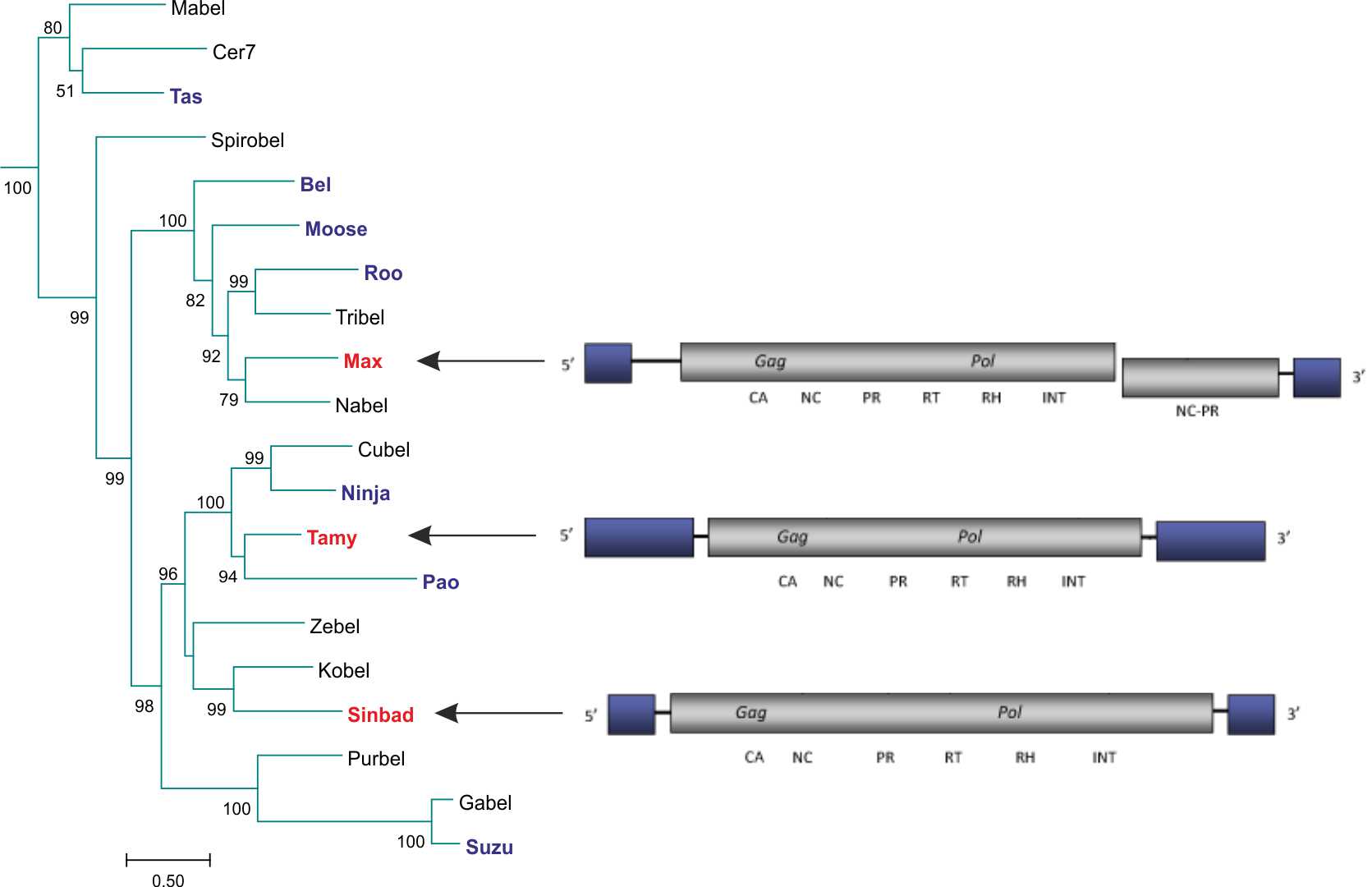
In addition, we propose to create 3 new species, namely *Drosophila semotivirus Max*, *Antheraea semotivirus Tamy* and *Schistosoma semotivirus Sinbad*, within the genus *Semotivirus* to accommodate Drosophila melanogaster Max virus, Antheraea mylitta Tamy virus and Schistosoma mansoni Sinbad virus, respectively. Figure 3 shows that in the RT-based phylogeny, the 3 viruses are nested among other semotiviruses and have similar genome organizations. In accordance with the established demarcation criteria for the genus *Semotivirus* (<https://talk.ictvonline.org/ICTV/proposals/2003.F187-190.Semotivirus.pdf>), viruses in the 3 new species have less than 50% identity in their Gag protein sequences compared to all other species (Figure 4). Furthermore, we propose to rename the species *Fugu rubripes Suzu virus* to *Takifugu rubripes Suzu virus,* to maintain consistency with the nomenclature of the host organism. We note that whereas family *Metaviridae* includes viruses infecting hosts from different branches of the eukaryotic tree, semotiviruses are exclusively associated with animal hosts (10 species are associated with invertebrates and one, *Takifugu rubripes Suzu virus*, with a vertebrate host).



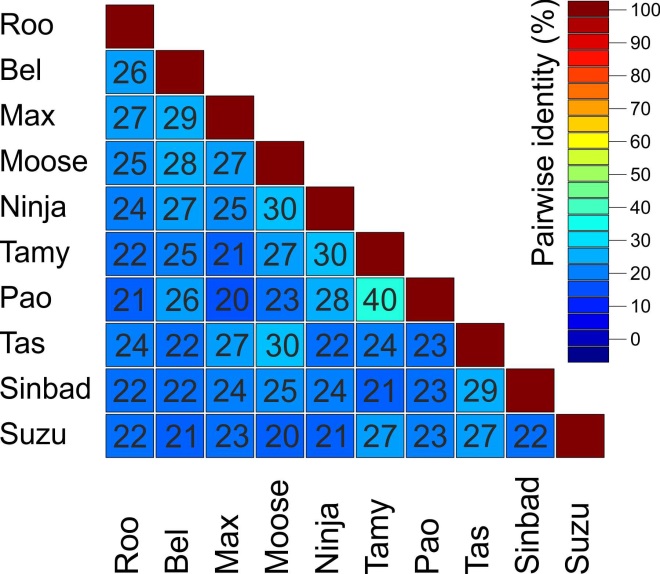
**Figure 1**. Schematic genome maps of several groups of reverse-transcribing viruses. Long terminal repeats are depicted with blue rectangles. Abbreviations: CA, capsid; NC, nucleocapsid; MA, matrix protein; PR, protease; RT, reverse transcriptase; RH, ribonuclease H; INT, integrase; Env, envelope protein; Gag, group-specific antigen; Pol, polymerase; RSV, Rous sarcoma virus; MLV, murine leukemia virus. Note the distinct positions of the INT domain in pseudoviruses on the one hand and metaviruses, retroviruses, and semotiviruses, on the other hand. The genome maps are not drawn to scale and were downloaded from the Gypsy Database (Llorens *et al*., 2011).



**Figure 2.** Maximum likelihoodphylogenetic tree of reverse transcriptases representing different groups of reverse-transcribing viruses. The clade including members of the genus *Semotivirus* (currently within the family *Metaviridae*; proposed to be moved into the family *Belpaoviridae*) is shown in red, whereas the clade encompassing members of the genera *Metavirus* and *Errantivirus* (family *Metaviridae*) is shown in blue. RT sequences were aligned using MAFFT (Katoh and Standley, 2013). Non-informative positions were removed with trimAL using the *gappyout* option (Capella-Gutiérrez et al., 2009). Maximum likelihood analysis was performed using PhyML v3 (Guindon et al., 2010) with the best substitution model (LG+G+F) determined by the program (Lefort et al., 2017). The tree is rooted with RT sequences of bacterial group II introns and non-LTR retroelements.

****

**Figure 3.** Maximum likelihoodphylogenetic tree of reverse transcriptases of semotiviruses. The clade is extracted from the phylogeny shown in Figure 2. The previously classified semotiviruses are colored blue, whereas the 3 new viruses are highlighted in red and their corresponding genome maps are shown on the right. N.B. Other putative semotiviruses (Mabel, Cer7, Spirobel, Tribel, Nabel, Cubel, Zebel, Kobel, and Purbel) could represent new species; however, the complete genome sequences of these viruses were not deposited to GenBank, precluding their official classification. By contrast, Gag polyprotein of Gabel shows 64% sequence identity to the Gag of Suzu and, according to the established demarcation criteria, would be included in the species *Takifugu rubripes Suzu virus*.



**Figure 4.** Matrix of sequence similarities between the Gag proteins of semotiviruses. The figure was prepared using SDT v1.2 (Muhire *et al*., 2014).

**References**

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25(15):1972-3.

Eickbush TH, Jamburuthugoda VK. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res. 2008;134(1-2):221-34.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80.

Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 2017. doi: 10.1093/molbev/msx149.

Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 2011; 39(Database issue):D70-4.

Llorens C, Muñoz-Pomer A, Bernad L, Botella H, Moya A. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biol Direct. 2009;4:41.

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.