

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

|  |  |  |
| --- | --- | --- |
| **Code assigned:** | **2020.016M** |  |
| **Short title:** Create three new subfamilies in the family *Rhabdoviridae* (*Mononegavirales*) | | |
|  | | |

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

|  |  |
| --- | --- |
| Walker PJ, Blasdell KR, Breyta R, Dietzgen RG, Fooks AR, Freitas-Astúa J, Kondo H, Kurath G, Kuzmin IV, Stone DM, Tesh RB, Tordo N, Vasilakis N, Whitfield AE | peter.walker@uq.edu.au; kim.blasdell@csiro.au; rbjmax@uw.edu; r.dietzgen@uq.edu.au; tony.fooks@apha.gsi.gov.uk; juliana.astua@embrapa.br; hkondo@rib.okayama-u.ac.jp; gkurath@usgs.gov; ivkuzmin@yandex.ru; david.stone@cefas.co.uk; rtesh@utmb.edu; ntordo@pasteur.fr; nivasila@utmb.edu; awhitfi@ncsu.edu |

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

|  |
| --- |
|  |

**Corresponding author**

|  |
| --- |
| Peter J Walker |

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

|  |
| --- |
| *Rhabdoviridae* SG |

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

|  |
| --- |
| Approved by all responding SG members (12 of 14) with minor revisions. |

**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 | 2 August 2020 |
| Date of this revision (if different to above) | 2 December 2020 |

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

|  |
| --- |
| Concerns with this proposal have been due to phylogenetic analyses of the L protein RdRp domain which suggest that the family *Rhabdoviridae* may not currently be monophyletic.  As discussed below, issues relating to the classification of novirhabdoviruses are not straightforward and will require further consideration. The approach described here of dealing with the issue of monophyly of the *Rhabdoviridae* separately and subsequently to the establishment of three subfamilies has been discussed with and is supported by the (-) ssRNA Virus Subcommittee Chair and the ICTV Vice-President.  To facilitate future removal of the novirhabdoviruses the family (if that is ultimately decided as appropriate), the proposal has been modified to rearrange the subfamily nomenclature. The genus *Novirhabdovirus* will now be assigned to the subfamily *Gammarhabdovirinae* rather than *Betarhabdovirinae* which will now comprise the clade of plant rhabdoviruses. This rearrangement means that, if decided in future, removal of the *Gammarhabdovirinae* from the family will not be disruptive to the new subfamily structure.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **The issue of novirhabdovirus classification**  The genus *Novirhabdovirus* comprises four species for viruses that include important pathogens of farmed and wild fish. The genus has been assigned to the *Rhabdoviridae* since it was first created in 1998.  It has recently been recognized that Bayesian trees for all mononegaviruses place viruses assigned to the genus *Novirhabdovirus* (proposed subfamily *Gammarhabdovirinae*) in a deeply rooted clade that includes viruses representing the *Paramyxoviridae*, *Sunviridae* and *Filoviridae*, rather than with other members of the *Rhabdoviridae*.  However, in contrast to the RdRP-based phylogeny, several other genetic, structural and morphological characteristics support the assignment of the novirhabdoviruses to the *Rhabdoviridae.*  **Glycoproteins**  Rhabdovirus G proteins are unique class I transmembrane glycoproteins (single membrane-spanning proteins with a luminal N-terminal domain and a cytoplasmic C-terminal domain). Animal rhabdoviruses share a conserved set of 12 cysteine residues that have been shown by mass spectrometry and chystallography to form 6 disulphide bridges, stabilizing the folded structures of the proteins. Genus-specific variations may abolish certain pairs of these cysteine residues and the associated disulphide bridge. The novirhabdovirus G proteins are homologous with those of other animal rhabdoviruses, sharing 10 of the 12 conserved cysteine residues with the G proteins of vesiculoviruses (for which the folded structures have been resolved crystallographically) [1, 5, 6].  **Nucleoproteins**  We have considered that perhaps trees based on the nucleoproteins (N proteins) may be informative. However, these sequences are also quite diverse and, although closely related taxa form well supported clades, the deeper nodes are not well supported, even for the various sets (proposed subfamilies) of rhabdoviruses.  **Matrix proteins**  Other common mononegavirus structural proteins (M and P) are extremely diverse in sequence, even for viruses in different genera of the same family. Nevertheless, novirhabdovirus M proteins display obvious homology with the VSIV M protein. Importantly, the sequence motif characteristic of the late budding domain (PPPH/Y) is preserved (see **Figure 3**).  **Transcription regulatory sequences**  Novirhabdoviruses share with all other animal rhabdoviruses the nucleotide sequence motif NNNC[U]7 which serves as a signal for the RdRP to terminate transcription and commence polyadenylation. In all other mononegaviruses, this sequence motif differs, typically with the string of U residues interrupted by multiple other different nucleotides (see for example Norton and Fearns [4]; Hume & Muhlberger [3]). This indicates that, despite the inferred phylogeny, the novirhabdovirs L protein has functional aspects that more resemble those of rhabdoviruses.  **Morphology and morphogenesis**  The novirhabdoviruses share with other animal rhabdoviruses the iconic bullet-shaped morphology that we associate with rabies, VSIV, etc. (see Granzow et al [2] Fig 2). The assembly processes associated with the morphogenesis of these bullet shapes are complex and driven by interactions between the N-RNA complex, the matrix protein and the C-terminal cytoplasmic domain of the G protein. Indeed, in morphology and several other of these considerations, novirhabdoviruses are much more closely related to animal rhabdoviruses than are the plant rhabdoviruses (which do cluster phylogenetically with rhabdoviruses based on the RdRP core domain).  **Genetic recombination in rhabdoviruses**  Evidence of recombination in rhabdoviruses (and we believe all mononegaviruses) is exceedingly rare and the few reports suggesting it have been viewed with some scepticism. That is assumed to be due to the tight and instantaneous N-RNA interaction that happens during RNA synthesis to form the new template for replication and transcription. Genomic or antigenomic RNA are never in free form and the binding of L and P to the N-RNA complex is highly specific. DI particle formation and gene duplication are not uncommon - but not with template switching as would be required for recombination. That is not to say that exceedingly rare events cannot happen over exceeding long timeframes and rhabdoviruses certainly appear to be very ancient. However, it would not be reasonable to discuss rhabdovirus evolution in terms of modularity in the same way it has been characterised for other viruses such as nidoviruses.  **Interpretation of these data**  In considering how to interpret these data, we favour the argument that the reductionist approach using only phylogenetic relationships based only on the polymerase core module is not necessarily the most suitable. A more holistic approach at family level, considering a range of factors including genome organisation, replication and expression strategy, protein structural data, morphogenesis and virion morphology may be more useful and informative. Arguably, all of these factors are relevant to the structure and function of extant viruses rather than deep phylogeny of the polymerase core domain which may simply reflect very rare events in distant geological history. It is also arguable that the evolutionary history inferred from trees is entirely dependent on the accuracy of sequence alignments and we note that these can be quite variable when different algorithms (e.g., Clustal, Muscle, MAFFT) are applied to the same sequence sets. We also observe that none of these alignment programs correctly aligns the sets of cysteine residues that we know should align in rhabdovirus G proteins (or the G proteins of other mononegaviruses for that matter). Unless we determine the crystallographic structures of the RdRP domains of all viruses in a set (or at least model them structurally), how do we know which alignment is correct? There is also the issue of sequence reversions which could occur randomly over deep evolutionary history, confounding phylogenetic inferences.  It can also be argued that, if we are using structural relationships to classify viruses at the highest ranks of taxonomy (when sequence-based phylogeny breaks down), why is it not equally legitimate to base classification on structural properties at the lower taxonomic ranks (when sequence-based phylogeny does not reflect relationships of extant viruses)? However, it is equally arguable that we insist on monophyly based on L protein sequence alignments for assignment of genera within the *Rhabdoviridae* so why should we not apply the same principle when defining the family itself? The reasons outlined above may counter this but, nevertheless, we recognise that the issue needs to be addressed.  **The way forward**  We propose to deal separately with the two issues here.  The first is to improve the taxonomy of the *Rhabdoviridae* as it is currently assigned, i.e., to introduce sub-families to the current taxonomic structure of the family. That will help reverse the informal taxonomy that is being promulgated describing the large clade of animal rhabdovirus as "dimarhabdoviruses". This would formally become the sub-family *Alpharhabdovirinae*.  The second more complex issue is the apparent absence of monophyly of the *Rhabdoviridae*based on trees inferred from alignments of the RdRP domain, and the possible removal of the genus *Novirhabdovirus* from the *Rhabdoviridae*. Consideration of that issue will be done separately and subsequently, whether as a genus (as it now stands) or as a subfamily (as it would be if this proposal is approved). To facilitate that discussion, we have rearranged the proposed subfamily nomenclature so that the novirhabdoviruses would be assigned to the subfamily *Gammarhabdovirinae* rather than *Betarhabdovirinae* which would then be proposed for the clade of plant rhabdoviruses. Removal of the *Gammarhabdovirinae* would then not be disruptive to the new subfamily structure. |

**Part 2:** **NON-TAXONOMIC PROPOSAL**

**Text of proposal**

|  |
| --- |
|  |

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

**Name of accompanying Excel module**

|  |
| --- |
| 2020.001M\_014M\_015M\_016M.R.Rhabdoviridae.xlxs |

**Abstract**

|  |
| --- |
| The family *Rhabdoviridae* currently comprises 30 genera for 191 species for viruses infecting vertebrates, invertebrates and plants. Here we propose the establishment of three subfamilies to accommodate all currently assigned genera and species. Each subfamily comprises viruses that form a distinct and well-supported clade in Maximum Likelihood or Bayesian phylogenetic trees inferred from the complete L protein (RdRp) sequences. The *Alpharhabdovirinae* will comprise 24 genera of viruses infecting vertebrates and/or invertebrates. The *Betarhabdovirinae* will comprise six genera of viruses infecting plants and invertebrates. The *Gammarhabdovirinae* will comprise one genus of viruses infecting fish. |

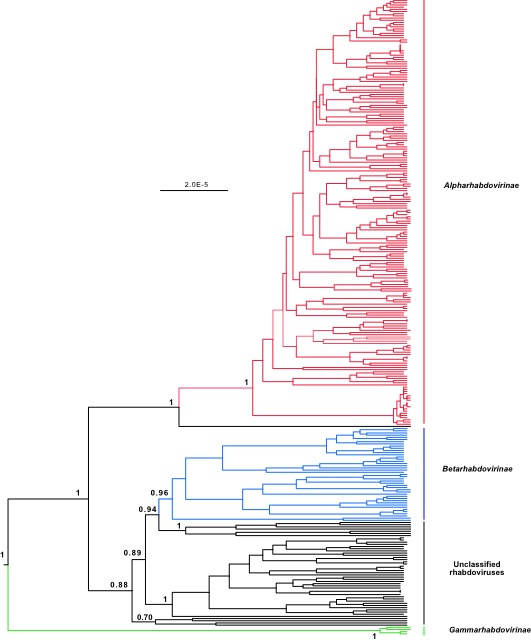
**Text of proposal**

|  |  |
| --- | --- |
| |  | | --- | | The family *Rhabdoviridae* currently comprises 30 genera and 191 species for viruses infecting vertebrates, invertebrates and plants. As homologous genetic recombination occurs very rarely, if at all, amongst rhabdoviruses, taxonomic assignment of viruses within the family has been conducted primarily from evolutionary histories inferred using the L protein. As in other mononegaviruses, L is a large (> 200 kDa) protein with multiple functions involved in replication and transcription, including the RdRp domain. More limited analyses of other structural proteins (e.g., G or N) using sets of closely related rhabdoviruses support the evolutionary histories inferred from L protein sequences. However, due to lower levels of sequence conservation in these proteins, are not suitable for family-wide analyses.  Phylogenetic trees inferring the evolutionary history from complete L protein sequences of 215 rhabdoviruses representing 30 currently assigned and three newly proposed genera consistently identifies three well-supported clades in Maximum Likelihood trees (**Figure 1**). Clade 1 comprises 24 existing genera and three proposed new genera of viruses infecting vertebrates (mammals, fish, birds, reptiles, amphibians) and/or invertebrates (arthropods, nematodes); we propose to assign all current and proposed new genera and species in this clade to the new subfamily *Alpharhabdovirinae*. Clade II comprises six genera of viruses infecting plants and arthropods; we propose to assign all genera and species in this clade to the new sub-family *Betarhabdovirinae*. Clade III comprises a single genus (*Novirhabdovirus*) of viruses infecting only teleost fish; we propose to assign this genus and its assigned species to the new subfamily *Gammarhabdovirinae*.  Bayesian analysis was conducted on a larger set of 291 rhabdoviruses including all available complete L protein sequences of currently classified rhabdoviruses and all other unclassified rhabdoviruses (as of January 2020)(**Figure 2**). The analysis also indicated strong support for the three clades of viruses representing the three proposed subfamilies. Other rhabdoviruses represented in this tree will be the subject of future taxonomic proposals.  Subfamily demarcation is based on the association of viruses within three major clades in phylogenetic analyses using complete L protein sequences and the specific host associations of each clade.  The origin of names for the three subfamilies is derived from simple alphabetical designations. Other designations (e.g., dimarhabdoviruses, phytorhabdoviruses and piscirhabdoviruses) have been considered by the Study Group, noting particularly that the term dimarhabdovirus has been used commonly in the scientific literature. However, this term dimarhabdovirus is derived from dipteran and ­mammalian rhabdoviruses and, as this clade is now known also to include many viruses isolated from reptiles, birds, fish and nematodes, the term is considered inappropriate and potentially misleading. We also consider the term piscirhabdovirus to be misleading as rhabdoviruses isolated from fish occur in both clade I (two genera) and clade III (one genus). | |

**Supporting evidence**

****

**Figure 1.** The evolutionary history was inferred from a Clustal W alignment of 215 complete L protein sequences of rhabdoviruses currently assigned to 30 genera and those rhabdoviruses recently proposed for assignment to three new genera *Alphapaprhavirus\**, *Merhavirus\** and *Alpharicinrhavirus\**. Phylogenetically informative sites were selected from the alignment using Gblocks resulting in 428 positions in the final dataset. The tree was inferred in MEGA7 by using the Maximum Likelihood method based on the Whelan and Goldman + Freq. model. The tree with the highest log likelihood (-70173.33) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bootstrap values (100 iterations) are shown for each node. Bootstrap values for the three key nodes defining clades of viruses assigned to each of the three subfamilies are shown in blue font. The number of viruses included for each genus are indicated in parenthesis.

****

**Figure 2.** The complete L protein sequences of 291 taxonomically assigned and currently unclassified rhabdoviruses were aligned using the MUSCLE algorithm and gaps in the alignment were pruned according to permissive parameters in the GBlocks algorithm. The resulting alignment of 409 characters was phylogenetically analysed under a Bayesian framework using the Birth-Death incomplete sampling demographic model, the LG amino acid substitution model, and the strict molecular clock model. Branch lengths are in units of substitutions/site and node support of posterior probability are labelled for key nodes. The assignments of viruses to subfamilies *Alpharhabdovirinae*, *Betarhabdovirinae* and *Gammarhabdovirinae* corresponds to those shown in Figure 1.

VSIV\_M MSSLKKILGLKGKGKKSKKLGIAPPPYEEDTSMEYAPSAPIDKSYFGVDEMDTHDPNQLR

IHNV\_M MSIFKR----------AKKTVLIPPPHLLSGDEERVTILSAE------GEIKVTGRRPTT

\*\* :\*: :\*\* : \*\*\*: . . \* .. . : .\*:.. . .

VSIV\_M YEKSFFTVKMTVRSNRPFRTYSDVAAAVSHWDHMYIGMAGKRPFYKILAFLGSSNLKATP

IHNV\_M LEEKIY--------------YSMNLAAAIVGGDLHPSFKSMTYLFQKEMEFGSTQEKVNF

\*:.:: \*\* \*\*. ..:: .: . ::: :\*\*:: \*..

VSIV\_M AVLADQGQPEYHAHCEGRAYLPHR-MGKTPPMLNVPEHFRRPFNIGLYKGTIELT-MTIY

IHNV\_M GSRKPAPQTTYQVTKAREVYLQTQPLEKKIPM--------QTYSVSTEGATITFTGRFLF

. \*. \*:. ..\*\* : : \*. \*\* :.:.:. .\*\* :\* ::

VSIV\_M DDESLEAAPMIWDHFNSSKFSDFREKALMFGLIVEEEASGAWVLDSVRHSKWASLASSF

IHNV\_M SSSHVGCDDNRTKLAGLDGFTTSNSYQRVKDYYAQETALALTFAAPEKRGKEK------

... : . . . . \*: .. : . .:\* \* . . . ::.\*

**Figure 3.** Clustal alignment of the M proteins of vesicular stomatitis Indiana virus (VSIV; *Rhabdoviridae*: *Vesiculovirus*) and infectious hematopoeitic necrosis virus (IHNV; *Rhabdoviridae*: *Novirhabdovirus*). The sequence motif characteristic of the late budding domain (PPPH/Y) is highlighted.

**References**

1. Einer-Jensen K, Krogh TN, Roepstorff P, Lorenzen N (1998) Characterization of intramolecular disulfide bonds and secondary modifications of the glycoprotein from viral hemorrhagic septicemia virus, a fish rhabdovirus. J Virol 72:10189-10196. PMID: 9811760 DOI: 10.1128/JVI.72.12.10189-10196.1998

2. Graznow H, Weiland F, Fichtner D, Enzmann PJ (1997) Studies of the ultrastructure and morphogenesis of fish pathogenic viruses grown in cell culture. Journal of Fish Diseases 20:1-10. DOI: 10.1046/j.1365-2761.1997.00267.x

3. Hume AJ, Muhlberger E (2019) Distinct genome replication and transcription strategies within the growing filovirus family. Journal of Molecular Biology 431:4290-4320. PMID: 31260690 DOI: 10.1016/j.jmb.2019.06.029

4. Noton SL, Fearns R (2015) Initiation and regulation of paramyxovirus transcription and replication. Virology 479-480:545-554. PMID: 25683441 DOI: 10.1016/j.virol.2015.01.014

5. Roche S, Bressanelli S, Rey FA, Gaudin Y (2006) Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science 313:187-191. PMID: 16840692 DOI: 10.1126/science.1127683

6. Walker PJ, Kongsuwan K (1999) Deduced structural model for animal rhabdovirus glycoproteins. Journal of General Virology 80:1211-1220. PMID: 10355768 DOI: 10.1099/0022-1317-80-5-1211