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:** | ***2017.014S*** | | | | (to be completed by ICTV officers) |
| **Short title:** Establishing taxa at the ranks of subfamily, genus, sub-genus and species in six families of invertebrate nidoviruses | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| A.E. Gorbalenya, Chair of the *Nidovirales* & *Mesoniviridae* Study Groups  M.A. Brinton, Member of the *Nidovirales* SG; Chair of the *Arteriviridae* Study Group  J. Cowley, Member of the *Nidovirales* SG; Chair of the *Roniviridae* Study Group  R. de Groot, Member of the *Nidovirales*, *Coronaviridae* & *Roniviridae* Study Groups  A. Gulyaeva, Non-member  C. Lauber, Member of the *Nidovirales* & *Polyomaviridae* SGs  B. Neuman, Member of the *Nidovirales* & *Coronaviridae* Study Groups  J. Ziebuhr, Member of the *Nidovirales* & *Mesoniviridae* SGs; Chair of the *Coronaviridae* SG | | | | | |
| **Corresponding author with e-mail address:** | | | | | |
| Alexander E. Gorbalenya (A.E.Gorbalenya@lumc.nl) | | | | | |
| **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) | | | This proposal is filed on behalf of the ***Nidovirales*, *Mesoniviridae* & *Roniviridae*** **Study Groups** in consultation with:  *Arteriviridae* Study Group  *Coronaviridae* Study Group | | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | | | |
|  | | | | | |
|  | | | | | |
| Date first submitted to ICTV: | | | | 23.06.2017 | |
| Date of this revision (if different to above):  Date of this revision (if different to above):  Date of this revision (if different to above): | | | | 12.07.2017  17.11.2017  08.08.2018 | |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
| **The proposer (17.11.2017)**: Numerous changes were introduced during two revisions, prior and after the EC meeting in Singapore. They concerned different aspects of the presentation and several corrections of inaccuracies. They did not revise the major results and conclusions of this proposal. They included: text refining; improved labelling of trees; summary of demarcation criteria in Table 1; extension of Figure 2 with genome and domain organization of viruses of three extra major groups. We would like to stress that we expect the proposed taxonomy structure to provide a framework for the rationalization of the molecular and biological properties of viruses in these two families, which, in many cases, remain to be determined and, therefore, cannot be used to evaluate the validity of the proposed structure.  **The proposer (08.08.2018)**: Fig. 1 has been updated to reflect genus reassignment of an arterivirus species. The name of the accompanying spreadsheet has been updated. |

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

|  |
| --- |
| Present the proposed new taxonomy on accompanying spreadsheet |
| **Name of accompanying spreadsheet:** 2017.012-015S.A.v4.Nidovirales |

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.

This and three accompanying proposals are based on analyses of the genomic diversity of viruses in the *Nidovirales* order and related unclassified viruses undertaken by Gorbalenya’s group and presented at the International Symposium Nido2017 (1). Implications of these analyses for the taxonomy of nidoviruses were discussed at a joint meeting of chairs and members of five SGs concerned with either all nidoviruses or its different members. All attendees were in favor of advancing nidovirus taxonomy according to recommendations made in this study, and a summary of this meeting was presented at the Symposium.

This proposal describes extensive changes at the ranks of subfamily, genus, subgenus and species in the current families *Mesoniviridae* and *Roniviridae*, and the putative families *Medioniviridae,* *Euroniviridae*, *Abyssoviridae*, and *Mononiviridae*, as recommended in an accompanying proposal to revise and extend the family structure of the order *Nidovirales*. Two other accompanying proposals describe novelties in the family *Arteriviridae* and the current family *Coronaviridae*. An overview of the proposed revisions to the taxonomy of the order *Nidovirales,* based on these proposals, together with datasets supporting the revised ranks are depicted in **Fig. 1 (see Appendix**).

The analyses included >3500 (near) complete genome sequences of nidoviruses from diverse vertebrate and invertebrate hosts; many of which are currently the only or major source of information about respective viruses, particularly for viruses reported at the International Symposium Nido2017 (2, 3). Sequences were analyzed in the computational comparative genomics framework DEmARC (DivErsity pArtitioning by hieRarchical Clustering) using profiles of multiple sequence alignments (MSA), Bayesian and Maximum-likelihood phylogenetic trees, and profiles of clustering cost (CC) function that were produced for weighted hierarchical clustering of pairwise patristic distances (PPD). In the CC profiles, all local minima (smallest CC values in a range of PPD values) were considered as candidate thresholds for ranks because they satisfied two requirements, (i) the clusters formed under these thresholds were monophyletic in the ML tree of the respective nidovirus subset, and (ii) all intra- and inter-cluster PPDs were (predominantly) smaller and (predominantly) larger, respectively, than the respective threshold. If *all* intra- and inter-cluster PPDs were smaller and larger than the respective threshold, respectively, such clustering has clustering cost of zero (CC=0). We have also measured persistence of a clustering as a range of PPD values over which this clustering was favored with the support of CC=0. The respective “threshold PPD ranges” were considered best candidates for demarcation. Those clustering and threshold (ranges) supported independently by several datasets were used to set demarcation criteria of a rank, as these assignments were less likely to be fortuitous due to biased virus sampling and/or domain selection.

Genome sequences were assigned to nidovirus taxa using either the Haygens tool (<http://veb.lumc.nl/HAYGENS/>) or by authors who described the viruses. Assignments were verified by alignments and phylogenetic analyses of five replicative protein domains characteristic of nidoviruses (synteny of molecular markers), namely the 3CLpro, NiRAN, RdRp, ZBD and HEL1. As shown in **Fig. 1**, 9 groups of nidovirus lineages, ranging from separate subfamilies to the entire order, were analyzed. For each group, from 3 to 4 MSAs of concatenated replicative domains including from 1 to 18 domains conserved within a group (**Fig. 2**), in total 29 MSAs, were generated and used in phylogenetic and DEmARC analyses. Data from these analyses provided support for monophyletic clusters, levels and clusters of classification, agreement between phylogeny and classification for each virus group, and inter-group agreement regarding classification levels.

This proposal is based on data generated with 14 datasets for 4 groups, namely 1) Invertebrate nidoviruses (current families *Mesoniviridae* and *Roniviridae*, and four putative families *Medioniviridae*, *Euroniviridae*, *Abyssoviridae*, and *Mononiviridae*, Inv group); 2) ExoN-encoding viruses of the order *Nidovirales* (family *Coronaviridae* + Invertebrate nidoviruses, CTI group); 3) all viruses of the order *Nidovirales* (the family *Arteriviridae* + family *Coronaviridae* + Invertebrate nidoviruses, ACTI group); 4) all viruses of the order *Nidovirales* plus 3 unclassified invertebrate viruses that uniquely share 3 key replicative enzymes, 3CLpro, RdRp, and HEL1, with nidoviruses and formed an outgroup in the phylogeny, PACTI group. The data strongly supported delineation of 8 clusters (taxa) at the subfamily/genus ranks, respectively (**Figs. 3, 4**). These cluster assignments were most consistent for MSAs generated from analyses of the more conserved 3CLpro, NiRAN, RdRp, ZBD and HEL1 replicative domains (5d proteins). Assignments for the 2 ranks below genus (subgenus, species) were made using datasets for the Inv group, as illustrated in **Figs. 5, 6,** or for the *Mesoniviridae* family, as shown in **Table 1**, which produced similar results. Based on a demarcation criterion of 0.0568 ppd for the 5d proteins (Inv dataset), 10 new species, including 2 in the existing family *Mesoniviridae*, were established, while this threshold with compatible with the existence of 7 “old” species established earlier in this family **(Fig. 6)**. An overview of intra-cluster genetic divergence in the 13-level hierarchical clustering of the Inv 5d dataset delineated by DEmARC is shown in **Fig. 7**.

**Table 1** Demarcation thresholds for species and subgenus ranks of the *Mesoniviridae* family

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **rank** | **Taxa #** | **PPD range1** | **PUD (%) range2** | **Dataset used3** |
| subgenus | 8 | 0.101-0.170 | 0.067-0.112 | Mesoni\_5d |
| species | 9 | 0.055-0.095 | 0.035-0.063 | Mesoni\_5d |

**1**Demarcation threshold depicted as a range of PPD values for which number of clusters (taxa) remained constant and CC=0. PPD values account for repeated replacements of amino acid residues.

**2**Demarcation threshold depicted as a range of PUD values for which number of clusters (taxa) remained constant and CC=0. PUD values are calculated as % of different residues in compared proteins.

**3**See Figure 2.

| **References:** |
| --- |
| 1. Gulyaeva, A.A., Lauber, C., Samborskiy, D.V., Leontovich, A.M., Sidorov, I.A. and A.E. Gorbalenya (2017) Evolutionary based classification of genomic diversity of nidoviruses connects metagenomics and experimental research. Proceedings for the XIVth International Nidovirus Symposium, S4. P-05, Kansas City, MO, USA, June 4-9, 2017. 2. Neuman, B. W., Bukhari, K., Mutlk, S. T., Alrashedi, H. S. H., Abdulsattar, B. O., Shu, G., Zhao, L., Jianping, J., Moroz, L. L., di Palma, F., Ayoub, N., Garb, J., and W. Sun (2017) Novel nido-like virus genomes associated with eukaryotic intracellular RNA pools. Proceedings for the XIVth International Nidovirus Symposium, S4. O-05, Kansas City, MO, USA, June 4-9, 2017. 3. Saberi A., Gulyaeva A., Brubacher J.L., Newmark P.A. and A.E. Gorbalenya (2017) Planarian virus with giant RNA genome redefines nidoviruses. Proceedings for the XIVth International Nidovirus Symposium, S4. P-04, Kansas City, MO, USA, June 4-9, 2017. |

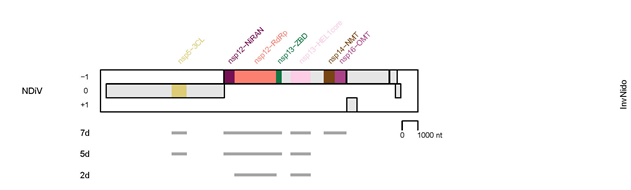
**Part 4:** **APPENDIX**: supporting material

| additional material in support of this proposal |
| --- |

|  |
| --- |
| **Annex:** |



**Fig. 1**. Proposed taxonomy of the order *Nidovirales* and sequences datasets analysed to produce it. The color panel depicts the proposed seven-rank taxonomy of the order *Nidovirales* along with a monophyletic sister group of unclassified invertebrate viruses, and with each suborder colored differently. Each taxon at every rank is depicted with a black rhomb and acronyms are given for species. Genome sequences of nine groups of nidoviruses, depicted with acronyms, were used to generate DEmARC classifications that were merged to produce this taxonomy. PACTI, all viruses of order *Nidovirales* plus three unclassified invertebrate viruses; ACTI, all viruses of the order *Nidovirales;* CTI, ExoN-encoding viruses of the order *Nidovirales* (family *Coronaviridae* + Invertebrate nidoviruses); Inv, Invertebrate nidoviruses; Me, family *Mesoniviridae*; CoTo, family *Coronaviridae;* Co, subfamily *Coronavirinae*; To, subfamily *Torovirinae*; Ar, family *Arteriviridae.* The bottom panel shows the taxa coverage of each group of sequences. The left panel specifies ranks that are largely supported by DEmARC classifications of the respective group of sequences.



**Fig. 2**. Domain combinations used for phylogenetic and DEmARC analyses to revise and/or establish the taxonomic structure below the family rank for invertebrate nidoviruses. Shown are the domain compositions of three combinations of conserved replicative domains used in this analysis for four virus groups. They are depicted relative to the genome and open reading frames of the representative virus specified at the left and the virus group (see **Fig. 1**) specified at the right. EAV, Equine arteritis virus; NDiV, Nam Dinh virus; SARS-CoV, SARS coronavirus; BRV, Breda torovirus. 13d, 12d, 9d, 7d, 5d, and 2d, respectively, indicate the respective numbers of concatenated domains whose locations are indicated by gray lines. Results shown in **Figures 3-7** were obtained for 5d combinations of two virus datasets (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 3**. Cluster partitioning of the phylogenetic tree of invertebrate nidoviruses by DEmARC. Shown is the ML tree of 18 invertebrate nidoviruses (one for each invertebrate nidovirus species delineated by DEmARC) inferred by IQ-Tree 1.5.5 using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to the 8 clusters (subfamily taxa) at level 4 of the DEmARC classification of the ACTI 5d dataset (see **Fig. 1**). The current and proposed subfamily structures of invertebrate nidoviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



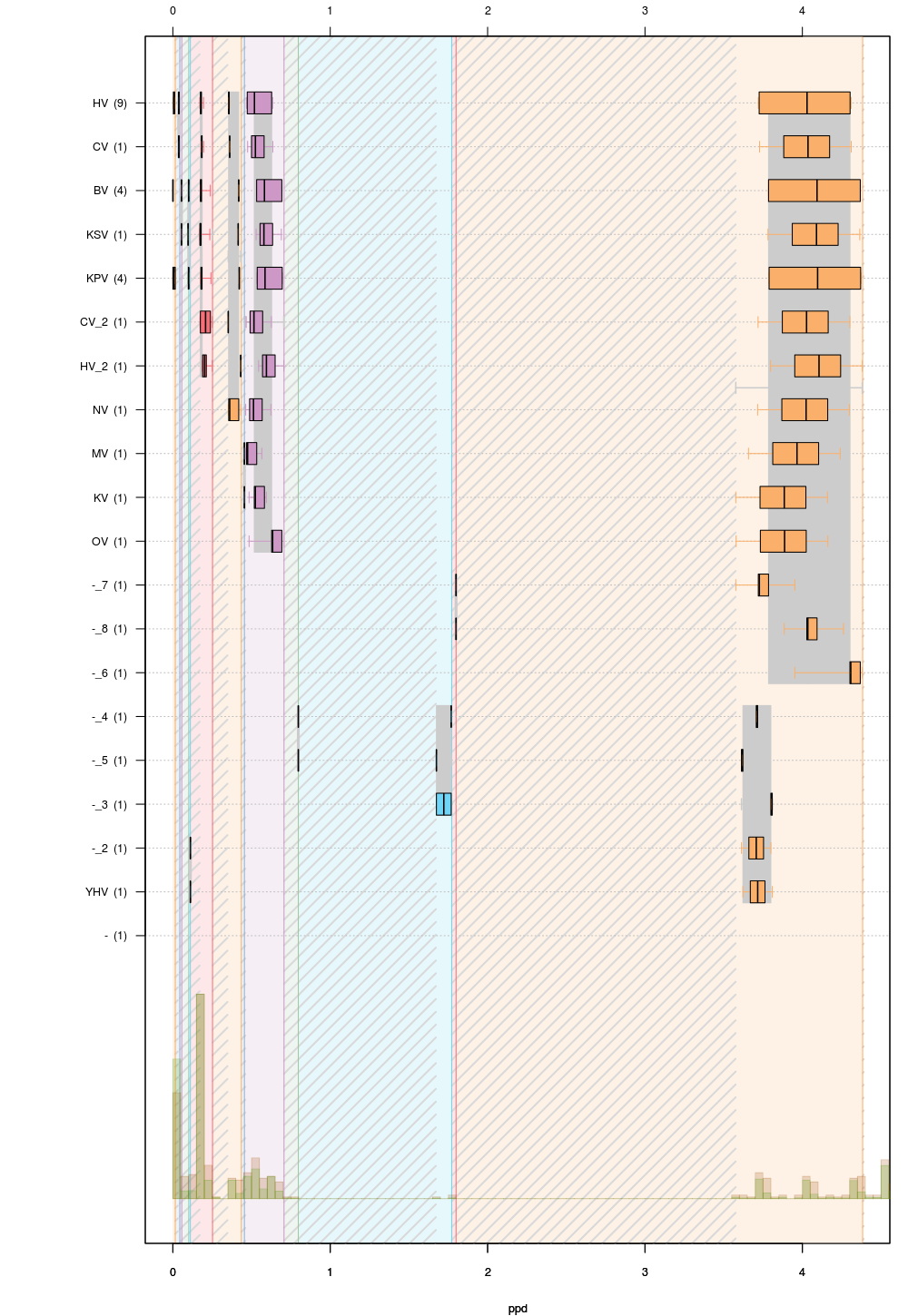
**Fig. 4**. Cluster partitioning of the phylogenetic tree of invertebrate nidoviruses by DEmARC. Shown is the ML tree of 18 invertebrate nidoviruses (one for each invertebrate nidovirus species delineated by DEmARC) inferred by IQ-Tree 1.5.5 using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to the 8 clusters (genera taxa) at level 3 of the DEmARC classification of the ACTI 5d dataset (see **Fig. 1**). The current and proposed genus structures of invertebrate nidoviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 5**. Cluster partitioning of the phylogenetic tree of invertebrate nidoviruses by DEmARC. Shown is the ML tree of 18 invertebrate nidoviruses (one for each invertebrate nidovirus species delineated by DEmARC) inferred by IQ-Tree 1.5.5 using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to the 16 clusters (subgenera taxa) at level 5 of the DEmARC classification of the Inv 5d dataset (see **Fig. 1**). The current and proposed subgenus structures of invertebrate nidoviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 6**. Cluster partitioning of the phylogenetic tree of invertebrate nidoviruses by DEmARC. Shown is the ML tree of 18 invertebrate nidoviruses (one for each invertebrate nidovirus species delineated by DEmARC) inferred by IQ-Tree 1.5.5 using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to the 18 clusters (species taxa) at level 3 of the DEmARC classification of the Inv 5d dataset (see **Fig. 1**). The current and proposed species structures of invertebrate nidoviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 7.** Intra-group genetic divergence in the hierarchical clustering of all invertebrate nidoviruses (Inv 5d dataset, **Fig. 1**) by DEmARC. Levels are defined by the thirteen strongest PPD thresholds. For simplicity, identities of clusters at the lowest level are indicated via arbitrary acronyms (left axis); the number of viruses in the identified clusters are shown in brackets. All identified clusters correspond to monophyletic groups on the phylogenetic tree. Box-and-whisker graphs were used to plot distributions of distances between viruses belonging to the same taxon on the candidate classification level under consideration, but different taxa on the previous level. The boxes span from the first to the third quartile and include the median (bold line), and the whiskers (dashed lines) extend to the extreme values. The corresponding part of the PPD distribution is depicted at the bottom. Thresholds for species and subgenus ranks correspond to the third (left-most violet) and fifth (left-most green) levels, respectively, of this classification (Gulyaeva et al. & Gorbalenya, unpublished).