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:** | ***2017.015S*** | | | | (to be completed by ICTV officers) |
| **Short title:** Reorganization and expansion of the order *Nidovirales* at the family and sub-order ranks | | | | | |
| **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* SG  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 by the ***Nidovirales* Study Group** in consultation with: *Arteriviridae* Study Group  *Coronaviridae* Study Group  *Mesoniviridae* Study Group  *Roniviridae* 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. Name of one suborder was slightly modified from "*Menidovirineae*" to "*Me****s****nidovirineae*”. 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 a study of genomic diversity of viruses of the *Nidovirales* order and related unclassified viruses that was conducted by Gorbalenya’s group and whose results were presented at the International Symposium Nido2017 (1). Implications of this study to 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 subsets. All attendees were in favor of advancing the nidovirus taxonomy according to recommendations of this study, and a summary of this meeting was presented at the Symposium.

This proposal deals with extensive changes at the family rank and introduction of taxa at the new rank sub-order, which was proposed to establish in a concurrently submitted proposal. Three other proposals deal with novelties concerning all the ranks below the family rank in (i) the family *Arteriviridae*, (ii) the current family *Coronaviridae*, and (iii) several families of invertebrate nidoviruses, respectively. A summary of the proposed taxonomy of the order *Nidovirales* along with support of its ranks and datasets involved in the analysis are depicted in **Fig. 1 (see Appendix**).

The underlying analysis involved >3500 (near) complete genome sequences of nidoviruses of diverse vertebrate and invertebrate hosts; many of these sequences were the only or major source of information about the respective viruses, particularly for viruses reported at the International Symposium Nido2017 (2, 3). Genome 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 if they satisfied two requirements: (i) the clusters formed under these thresholds were monophyletic in the ML tree of the respective nidovirus subsets, and (ii) all intra- and inter-cluster PPDs, respectively, were (predominantly) smaller or (predominantly) larger than the respective threshold. If *all* intra- and inter-cluster PPDs, respectively, were smaller and larger than the respective threshold, such clustering has a cost of zero, CC=0, according to DEmARC. We have also calculated 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) that were independently supported by several datasets were selected for setting demarcation criteria of a rank because these assignments were considered less likely to be fortuitous due to biased virus sampling and/or domain selection.

The initial assignment of genome sequences to nidoviruses or their subset was produced using either the Haygens tool (<http://veb.lumc.nl/HAYGENS/> ) or by authors who described these viruses. These assignments were subsequently verified in the conducted study through the delineation of an array of five replicative protein domains characteristic of nidoviruses (synteny of molecular markers), including 3CLpro, NiRAN, RdRp, ZBD and HEL1, and phylogenetic analysis. Nine (sub)groups of nidoviruses of different virus divergence, from separate subfamily to the entire order were analyzed (bottom panel of **Fig. 1**). For each group, from three to four MSAs of concatenated replicative domains including from 1 to 18 domains conserved in a group (**Fig. 2**), in total 29 MSAs, were generated and submitted to phylogenetic and DEmARC analyses. These analyses assessed the 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.

For this proposal, we used the results obtained with 12 datasets for five highly diverse groups, including: 1) Invertebrate nidoviruses (families *Mesoniviridae* and *Roniviridae*, and unclassified invertebrate nidoviruses, Inv group); 2) the current family *Coronaviridae* (subfamily *Coronavirinae* and subfamily *Torovirinae,* CoTo group); 3) ExoN-encoding viruses of the order *Nidovirales* (family *Coronaviridae* + Invertebrate nidoviruses, CTI group); 4) all viruses of the order *Nidovirales* (the family *Arteriviridae* + family *Coronaviridae* + Invertebrate nidoviruses, ACTI group); 5) all viruses of the order *Nidovirales* plus three unclassified invertebrate viruses that uniquely share three key replicative enzymes, 3CLpro, RdRp, and HEL1, with nidoviruses and formed an outgroup in the phylogeny, PACTI group. They strongly (CC=0) supported delineation of seven and nine clusters (taxa) at two levels corresponding to the sub-order and family ranks, respectively. Most consistently, this assignment was observed for MSAs encompassing the five most conserved replicative domains (5d), 3CLpro, NiRAN, RdRp, ZBD and HEL1. At the sub-order rank, seven taxa, including three of vertebrate viruses and four of invertebrate viruses, were recognized (**Table 1**). At the family rank, nine taxa, including three of vertebrate viruses and six of invertebrate viruses, were recognized. The respective results obtained for the ACTI group including virus representatives of all species of the order *Nidovirales* are shown in **Figs. 3-4**. An overview of the intra-cluster genetic divergence in the nine-level hierarchical clustering of the PACTI 2d dataset by DEmARC is shown in **Fig. 5**.

**Table 1** Demarcation thresholds for family and suborder ranks of the *Nidovirales* order

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **rank** | **Taxa #** | **PPD range1** | **PUD (%) range2** | **Dataset used3** |
| suborder | 7 | 3.240-3.476 | 0.734-0.747 | ACTI\_5d |
| family | 9 | 2.620-2.649 | 0.691-0.693 | ACTI\_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.

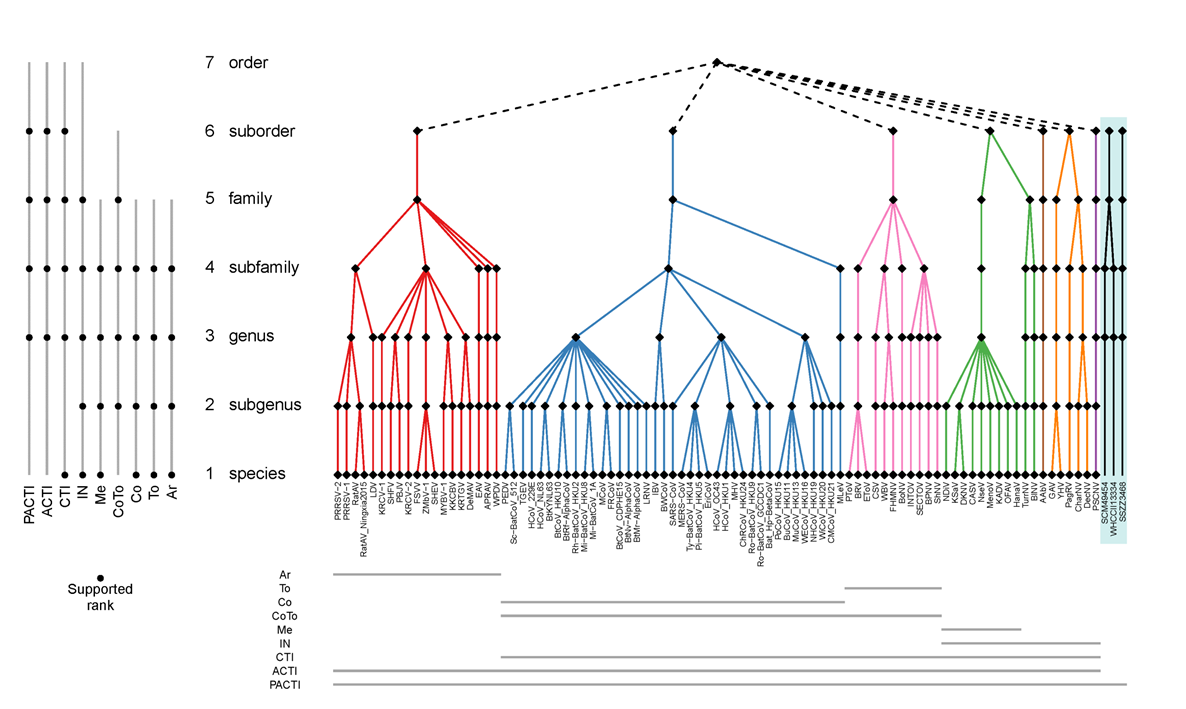
**2**Demarcation threshold depicted as a range of PUD values for which number of clusters (taxa) remained constant and CC=0. PUD values are calculated as % of different residues in compared proteins.

**3**See Figure 2.

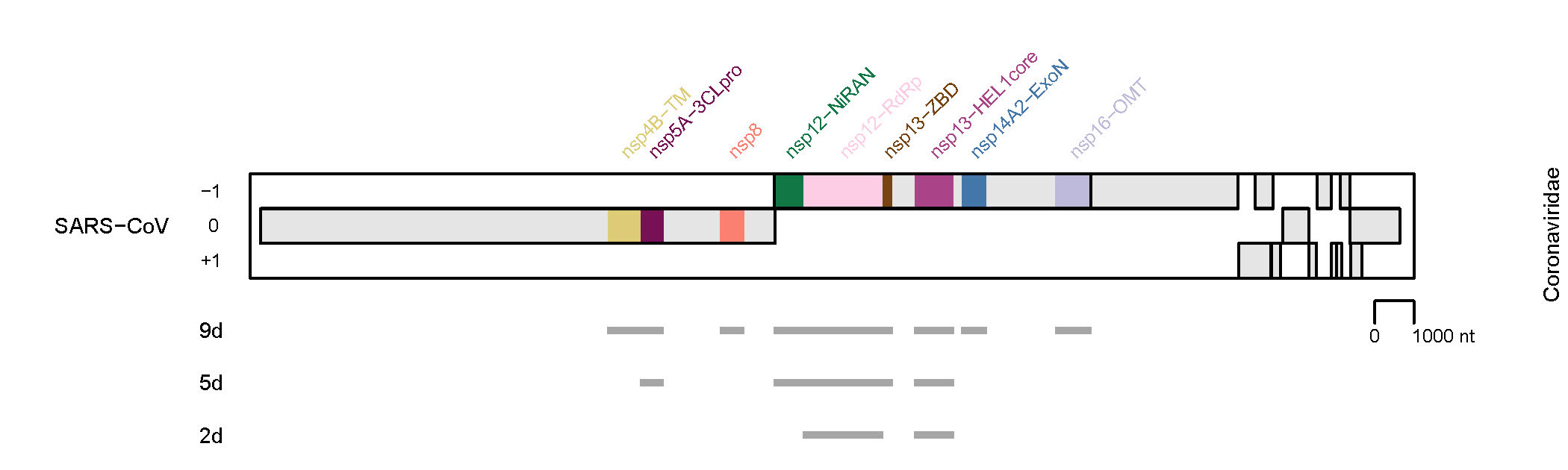
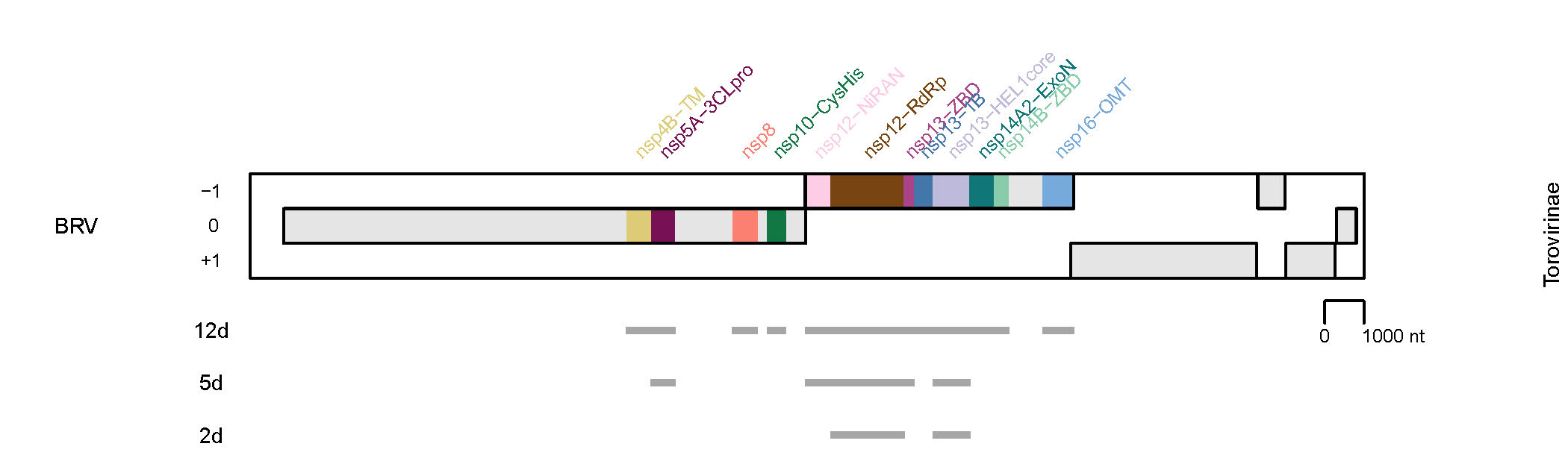
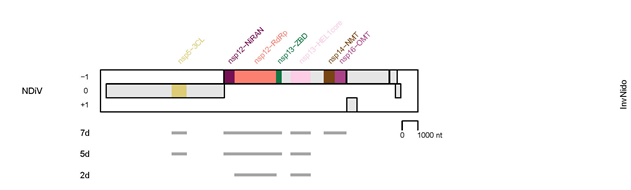
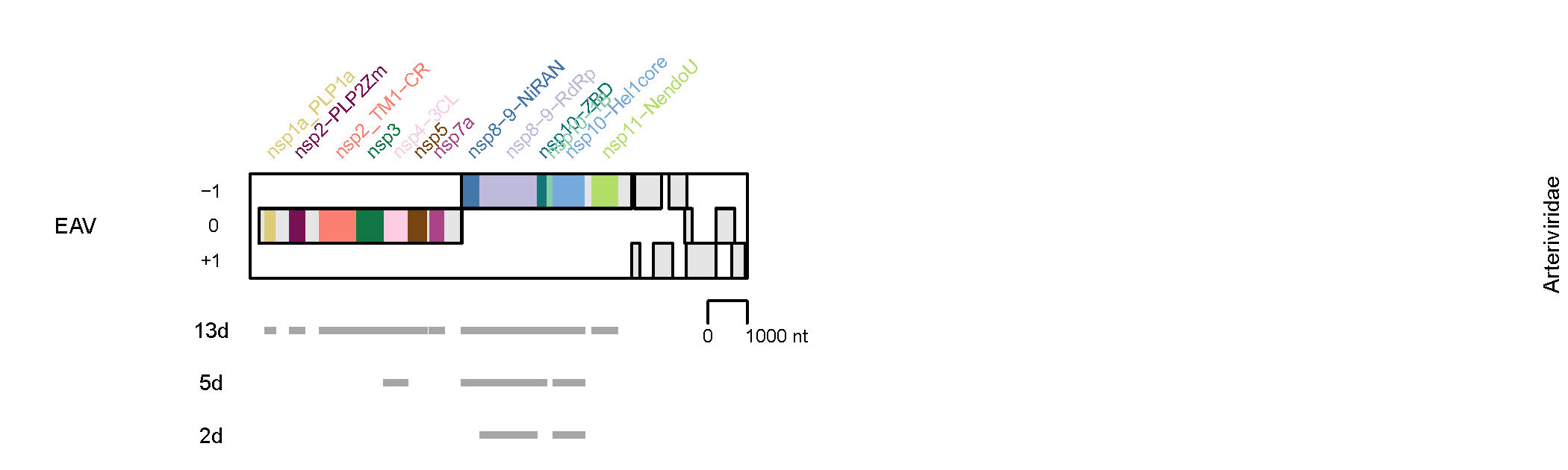
| **References:** |
| --- |
| 1. Gulyaeva, A.A., Lauber, C., Samborskiy, D.V., Leontovich, A.M., Sidorov, I.A. and A.E. Gorbalenya (2017) Evolutionary based classification of genomic diversity of nidoviruses connects metagenomics and experimental research. Proceedings for the XIVth International Nidovirus 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. 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 the respective 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 the 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, *Mesoniviridae* family; 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.



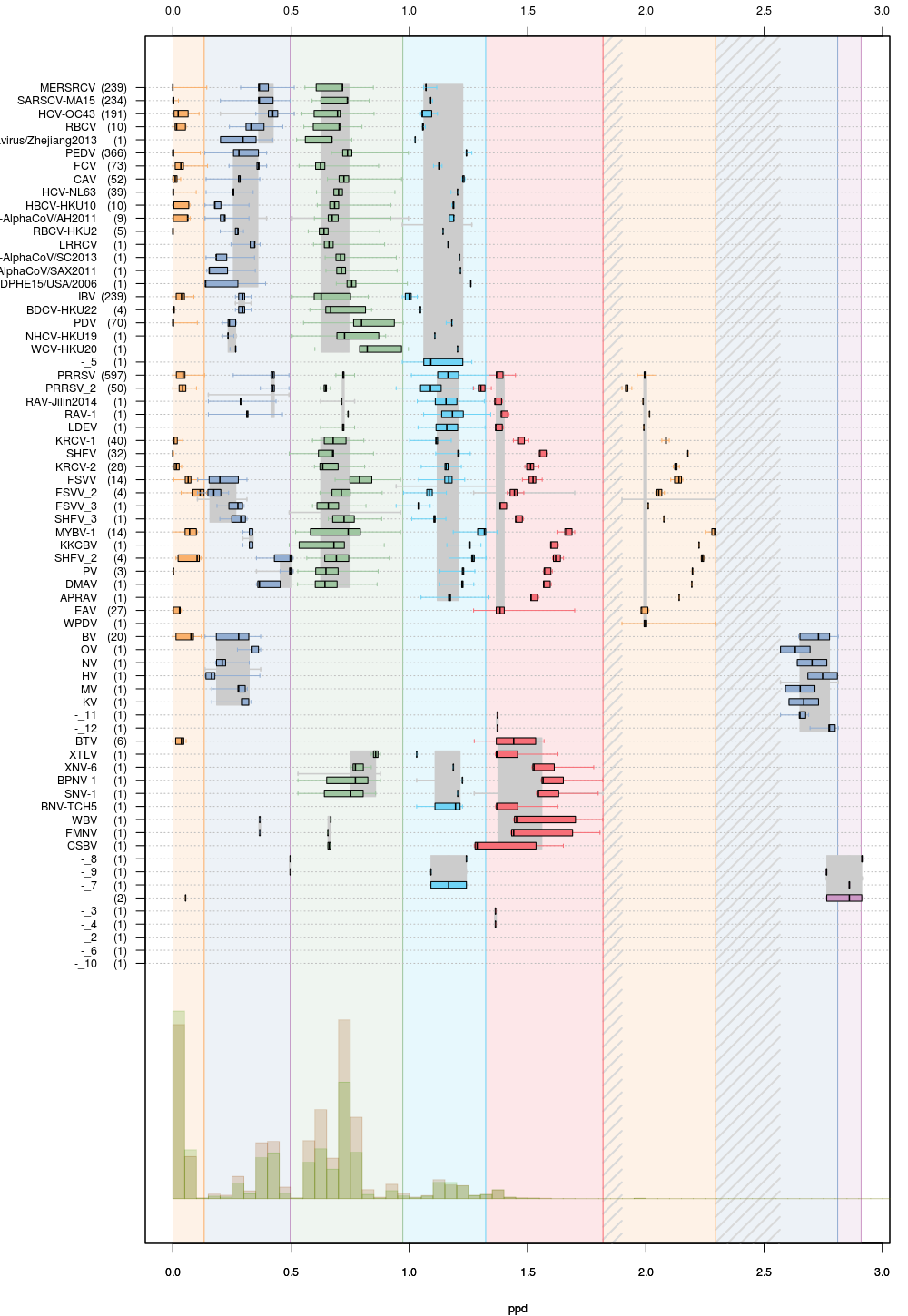
**Fig. 2**. Domain combinations used for phylogenetic and DEmARC analyses to revises family ranks structure and establish suborder rank structure of the *Nidovirales* order. Shown are the domain compositions of three combinations of conserved replicative domains used in this analysis for four virus groups. 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 nidoviruses by DEmARC. Shown is the ML tree of 87 nidoviruses (one for each 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 7 clusters (**suborder taxa**) at level 10 of the DEmARC classification of the ACTI 5d dataset (see **Fig. 1**). The current and proposed suborder structures of the order *Nidovirales* are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 4**. Cluster partitioning of the phylogenetic tree of nidoviruses by DEmARC. Shown is the ML tree of 87 nidoviruses (one for each 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 9 clusters (**family taxa**) at level 8 of the DEmARC classification of the ACTI 5d dataset (see **Fig. 1**). The current and proposed family structures of the order *Nidovirales* are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 5.** Intra-group genetic divergence in the nine-level hierarchical clustering of all nidoviruses and a monophyletic sister group of unclassified invertebrate viruses (PACTI 2d dataset, **Fig. 1**) by DEmARC. Levels are defined by the nine 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 family and sub-order ranks correspond to the seventh (right-most orange) and ninth (right-most violet) levels, respectively, of this classification (Gulyaeva et al. & Gorbalenya, unpublished).