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:** | ***2017.013S*** | | | | (to be completed by ICTV officers) |
| **Short title:** Reorganization of the family *Coronaviridae* into two families, *Coronaviridae* (including the current subfamily *Coronavirinae* and the new subfamily *Letovirinae*) and the new family *Tobaniviridae* (accommodating the currentsubfamily *Torovirinae* and threeother subfamilies), revision of the genus rank structure and introduction of a new subgenus rank | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| Ziebuhr J., Chair  Baric, R.S., Member  Baker, S., Member  de Groot, R.J., Member  Drosten, C., Member  Gulyaeva, A., Non-member  Haagmans, B.L., Member  Neuman, B.W., Member  Perlman, S., Member  Poon, L.L.M., Member  Sola, I., Member  Gorbalenya, A.E., Member | | | | | |
| **Corresponding author with e-mail address:** | | | | | |
| John Ziebuhr, e-mail: [john.ziebuhr@viro.med.uni-giessen.de](mailto:john.ziebuhr@viro.med.uni-giessen.de)  Alexander E. Gorbalenya, e-mail: [A.E.Gorbalenya@lumc.nl](mailto: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 ***Coronaviridae* Study Group** in consultation with:  *Nidovirales* Study Group  *Arteriviridae* 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. The changes included: text refining; improved labelling of trees; summary of demarcation criteria in Tables 1 and 2; extension of Figure 2 with genome and domain organization of viruses of three extra major groups. The names of three coronavirus species were corrected during 2nd revision to avoid the use of forward slashes. 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. The study was undertaken by Alexander Gorbalenya’s group. The results of these analyses were presented at the International Nidovirus Symposium 2017 (1) and implications of this study for the taxonomy of nidoviruses were discussed at a joint meeting of the chairs and members of the five SGs concerned with the order *Nidovirales* as a whole and the 4 currently established families in this order. All participants of this joint meeting were in favor of advancing the nidovirus taxonomy according to recommendations made in this study, and a summary of this meeting was presented at the Symposium.

A summary of the proposed taxonomy of the order *Nidovirales* along with support of its ranks and information on datasets used in these analyses is provided in **Fig. 1 (see Appendix**) and detailed in proposal 2017.015S.N. This proposal deals with an extensive reorganization of the current family *Coronaviridae* into two families. The proposed taxonomy structure will provide a framework for the rationalization of the molecular and biological properties of the involved viruses that, in the majority of cases, remain to be characterized in future studies.

The analyses of the order *Nidovirales* included >3500 (near) complete genome sequences of nidoviruses from diverse vertebrate and invertebrate hosts; many of which are currently the only source of information on the respective viruses. Genome sequences were analyzed in the computational framework DEmARC (DivErsity pArtitioning by hieRarchical Clustering) for comparative genomics. The analysis involved profiles of multiple sequence alignments (MSA), Bayesian and Maximum-likelihood phylogenetic trees, and profiles of the clustering cost (CC) function that were produced for weighted hierarchical clusterings of pairwise patristic distances (PPD). In profiles of the CC function, 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 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, respectively, than the respective threshold, such clustering has a clustering cost of zero, CC=0. 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 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 these viruses. Assignments were subsequently verified by alignments and phylogenetic analyses of an array 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 nidovirus (sub)groups of different divergence, ranging from separate subfamilies to the entire order, were analyzed. For each group, from 3 to 4 MSA of concatenated replicative domains including 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 18 datasets for 6 (out of 9) groups, namely 1) the current subfamily *Coronavirinae* (Co group); 2) the current subfamily *Torovirinae* (To group); 3) the current family *Coronaviridae* (subfamily *Coronavirinae* and subfamily *Torovirinae,* CoTo group); 4) viruses of the order *Nidovirales* encoding an ExoN domain (family *Coronaviridae* + Invertebrate nidoviruses, CTI group); 5) all viruses of the order *Nidovirales* (the family *Arteriviridae* + the current family *Coronaviridae* + Invertebrate nidoviruses, ACTI group); 6) 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 form a phylogeny outgroup (PACTI group) **(Fig. 1)**. The data obtained for groups PACTI, ACTI and CTI strongly supported (CC=0) an elevation of the current subfamilies *Coronavirinae* and *Torovirinae* to the family rank (**Fig. 3**), including two and four subfamilies, respectively (**Fig. 4)**. The assignments were most consistent for MSAs generated from analyses of the five most conserved replicative domains, namely 3CLpro, NiRAN, RdRp, ZBD and HEL1. Further analyses of less divergent datasets of groups CoTo, To and Co delineated monophyletic clusters at three levels, which we assigned to taxa at genus (**Fig. 5)**, sub-genus (**Fig. 6),** and species rank, respectively. These assignments were either fully compatible with the existing taxonomy or led to taxonomic assignments of floating species and several other viruses discovered very recently. An overview of the intra-cluster genetic divergence in a ten-level hierarchical clustering of the CoTo 5d dataset by DEmARC is shown in **Fig. 7**.

**Demarcation criteria.** The following values of patristic pairwise distances (PPD) calculated for ML phylogeny of MSA of five concatenated domains (3CLpro, NiRAN, RdRp, ZBD and HEL1) were used as demarcation criteria for taxa at each of the four ranks of the two families (**Table 1** and **Table 2**):

**Table 1** Demarcation thresholds for four ranks of the revised *Coronaviridae* family

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **rank** | **Taxa #** | **PPD range1** | **PUD (%) range2** | **Dataset used3** |
| subfamily | 2 | 1.583-1.733 | 0.542-0.561 | Co\_5d |
| genus | 5 | 0.789-0.853 | 0.363-0.385 | Co\_5d |
| subgenus | 24 | 0.186 | 0.075 | Co\_5d |
| species | 39 | 0.075-0.079 | 0.021-0.023 | Co\_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.

**Table 2** Demarcation thresholds for four ranks of the new *Tobaniviridae* family**1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **rank** | **Taxa #** | **PPD range** | **PUD (%) range** | **Dataset used** |
| subfamily | 4 | 1.382-1.646 | 0.555-0.597 | To\_5d |
| genus | 8 | 0.532-0.879 | 0.321-0.443 | To\_5d |
| subgenus | 9 | 0.194-0.532 | 0.140-0.321 | To\_5d |
| species | 11 | 0.049-0.119 | 0.036-0.088 | To\_5d |

**1**See footnote for Table 1

**Naming.** Names of taxa of the current *Coronaviridae* family were revised only in cases of reassignments to other ranks. New names were designed to facilitate communication and provide connections to other taxa and viruses in the family.

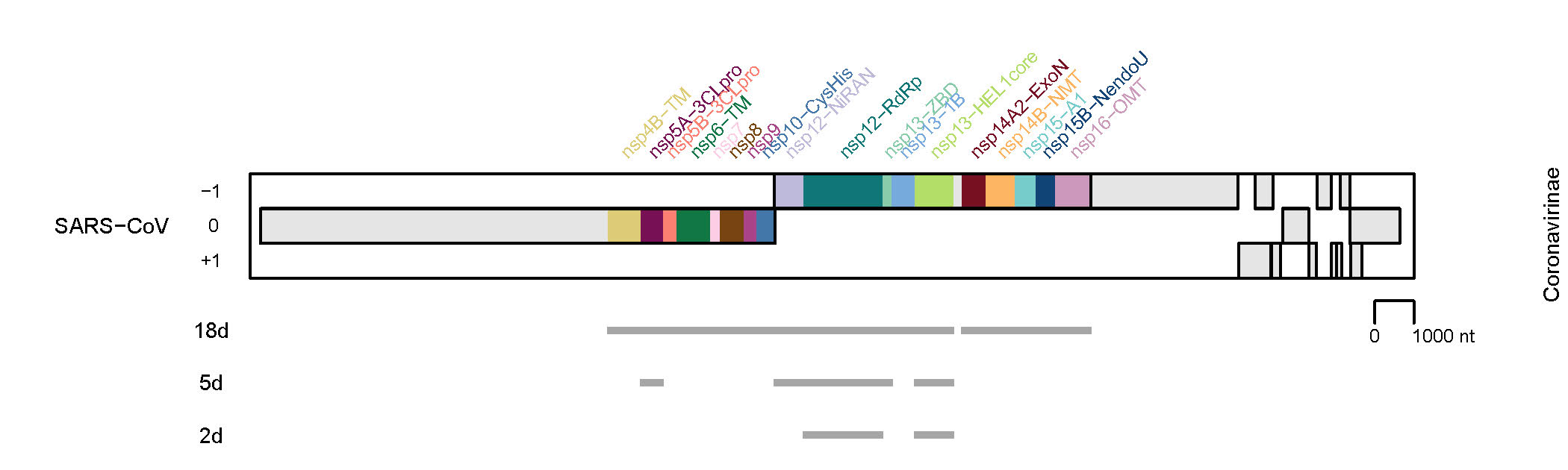
| **References:** |
| --- |
| 1. Gulyaeva, Anastasia A., Lauber, Chris, Samborskiy, Dmitry V., Leontovich, Andrey M., Sidorov, Igor A. and Alexander E. Gorbalenya (2017) Evolutionary based classification of genomic diversity of nidoviruses connects metagenomics and experimental research. Proceedings for the XIVth International Nidovirus Symposium, S4. P-05, Kansas City, MO, USA, June 4-9, 2017. |

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

| additional material in support of this proposal |
| --- |
| **Annex:** |

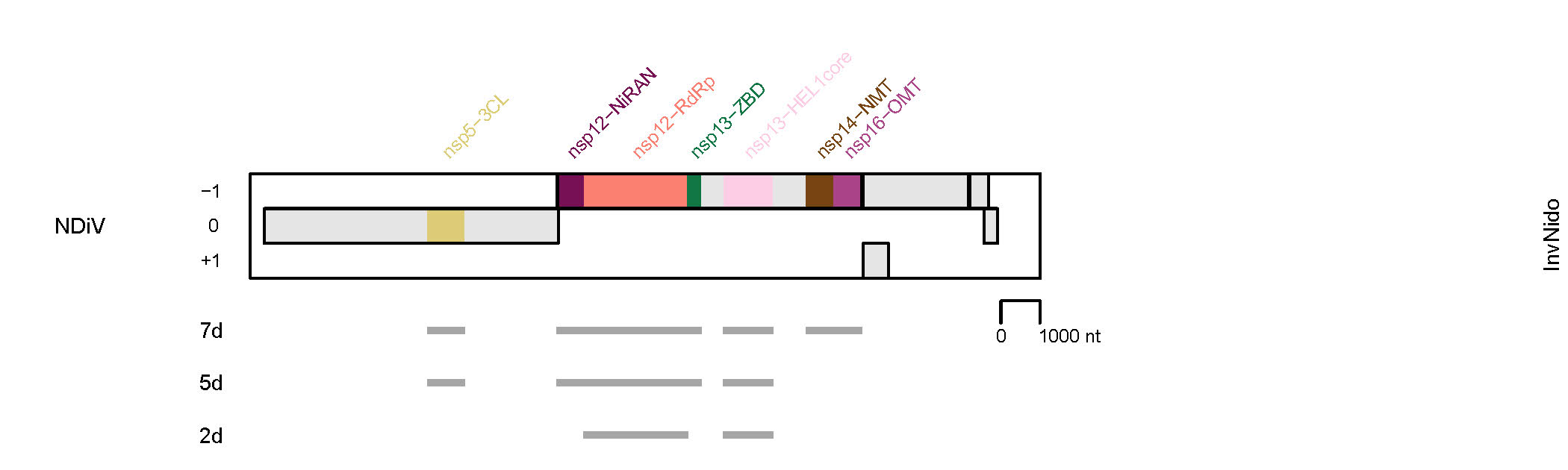


**Fig. 1**. Proposed taxonomy of the order *Nidovirales* and sequences datasets analysed to produce it. The color panel depicts the proposed seven-rank taxonomy of the order *Nidovirales* along with a monophyletic sister group of unclassified invertebrate viruses, and with each suborder colored differently. Each taxon at every rank is depicted with a black rhomb and acronyms are given for 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, 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 groups of sequences.









**Fig. 2**. Domain combinations used for phylogenetic and DEmARC analyses to revise the *Coronaviridae* family. 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. NDiV, Nam Dinh virus; SARS-CoV, SARS coronavirus; BRV, Breda torovirus. 18d, 12d, 9d, 7d, 5d, and 2d, respectively, indicate the respective numbers of concatenated domains which are also indicated by gray lines. The results shown in **Figs 3-6** were obtained for two datasets with 5d composition. (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 the 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).



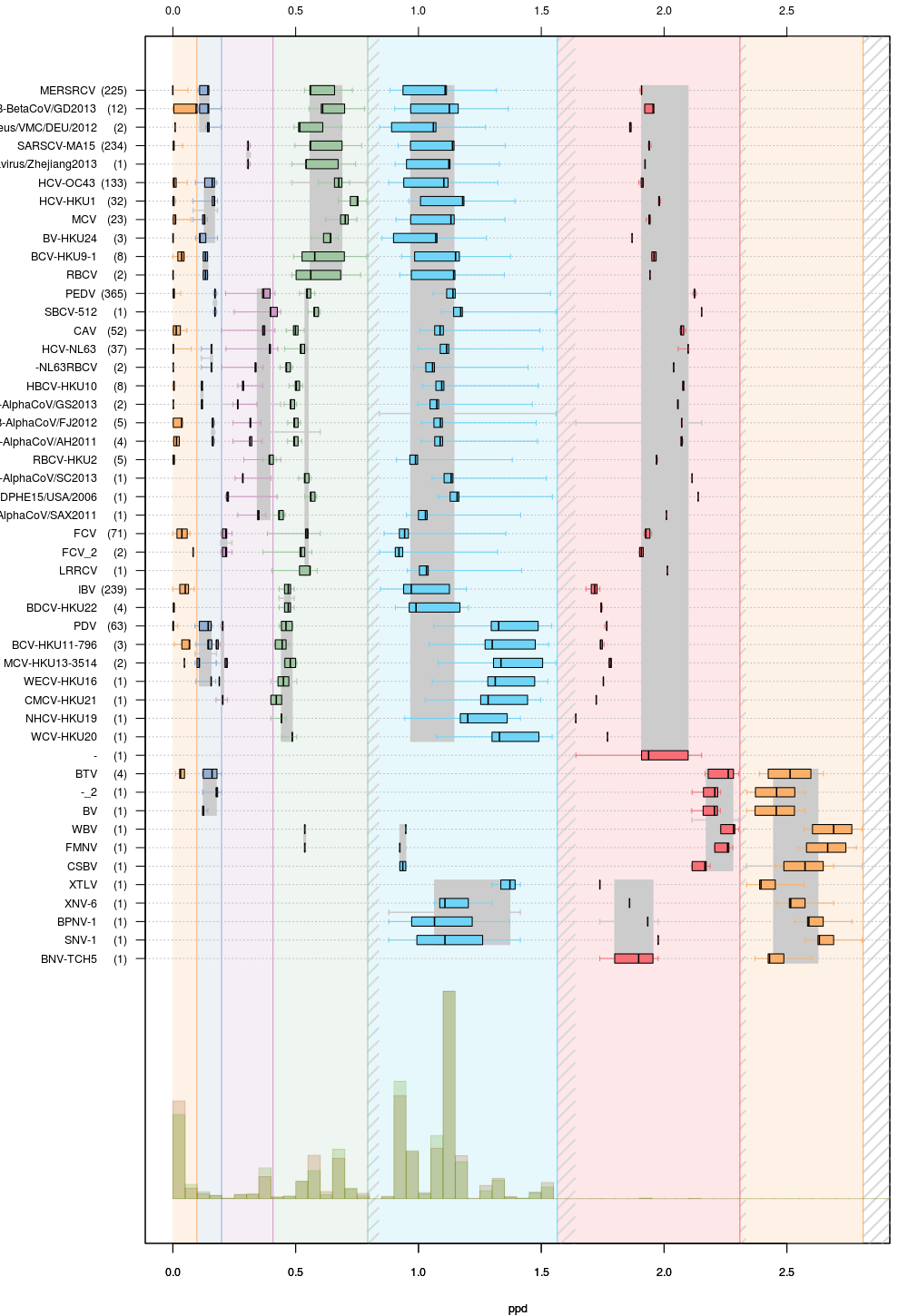
**Fig. 4**. Cluster partitioning of the phylogenetic tree of the current *Coronaviridae* family by DEmARC. Shown is the ML tree of 50 coronaviruses and toroviruses (one for each coronavirus/torovirus species delineated by DEmARC) inferred by IQ-Tree 1.5.5 using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to 6 clusters (subfamily taxa) at level 7 of the DEmARC classification of the ACTI 5d dataset (see **Fig. 1** for designations). The current and proposed subfamily structures of the current family *Coronaviridae* are detailed at the right (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 5**. Cluster partitioning of the phylogenetic tree of the current *Coronaviridae* family by DEmARC. Shown is the ML tree of 50 coronaviruses and toroviruses (one for each coronavirus/torovirus 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. Colors are used to indicate the 13 clusters (genus taxa) at level 6 of the DEmARC classification of the ACTI 5d dataset. The current and proposed genus structures of the current family *Coronaviridae* are detailed at the right. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 6**. Cluster partitioning of the phylogenetic tree of the current *Coronaviridae* family by DEmARC. Shown is the ML tree of 50 coronaviruses and toroviruses (one for each coronavirus/torovirus 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. Colors were used to indicate the 33 clusters (subgenus taxa) at level 2 of the DEmARC classification of the CoTo 5d dataset (see **Fig. 1** for designations). Names of the current and proposed subgenera of the current family *Coronaviridae* are detailed at the right. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 7.** Intra-cluster genetic divergence in the seven-level hierarchical clustering of the current family *Coronaviridae* (CoTo 5d dataset, **Fig. 1**) by DEmARC. Levels are defined by the seven 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), while the whiskers (dashed lines) extend to the extreme values. The corresponding part of the PPD distribution is depicted at the bottom. Threshold for subfamily rank corresponds to the second (dark blue) level of this classification (Gulyaeva et al. & Gorbalenya, unpublished).