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:** | ***2017.012S*** | | | | (to be completed by ICTV officers) |
| **Short title:** Expansion of the rank structure of the family *Arteriviridae* and renaming its taxa | | | | | |
| **Modules attached**  (Modules 1, 4 and either 2 or 3 are required. | | **1**  **2  3  4** | | | |
| **Author(s):** | | | | | |
| M.A. Brinton, Chair  A. Gulyaeva, Non-member  U.B.R. Balasuriya, Member  M. Dunowska, Member  K.S. Faaberg, Member  T. Goldberg, Member  F.C.-C. Leung, Member  H.J. Nauwynck, Member  E.J. Snijder, Member  T. Stadejek, Member  A.E. Gorbalenya, Member | | | | | |
| **Corresponding author with e-mail address:** | | | | | |
| Margo A. Brinton, mbrinton@gsu.edu  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 ***Arteriviridae* Study Group** in consultation with:  *Nidovirales* Study Group  *Coronaviridae* Study Group  *Mesoniviridae* Study Group  *Roniviridae* Study Group | | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | | | |
| **ICTV-EC (30.07.2018):** The statement, on page 4, that “Names of all subgenera taxa include ‘artevirus’ which is preceded by a letter that matches the first letter of the respective genus name” is not true. Please correct or remove. | | | | | |
|  | | | | | |
| 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  17-23.11.2017  08.08.2018 | |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
| **The proposer (17-23.11.2017)**: Numerous changes were introduced during two revisions, prior and post 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 Table 1 and names origins. The names of two subfamilies, nine subgenera and five species were modified in the second revision to address concerns of the EC, including that about the name stability and ease to use.  **The proposer (08.08.2018)**: The statement on page 4 has been improved to address the confusion. Fig. 1 has been updated to reflect genus reassignment of an arterivirus species. Typos and inaccuracies have been also corrected. 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.

A summary of the proposed taxonomy of the entire order *Nidovirales* along with support of its ranks and datasets involved in the analysis are summarized in **Fig. 1 (see Appendix**) and detailed in the proposal 2017.015S.N. This proposal describes extensive reorganization of the current family *Arteriviridae* due to elevation of the current five genera to subfamily rank, establishing eleven new genera and use of new subgenus rank.

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 about respective viruses. 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 profiles of CC function, 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 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, respectively, were smaller and larger than the respective threshold, such clustering has a cost of zero, CC=0, according to DEmARC. 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 thresholds 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 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, 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.

Our prior and current DEmARC analyses of the genetic divergence in the family of the *Arteriviridae* (Ar group in **Fig. 1**) identified from seven to nine candidate PPD thresholds with utmost support (CC=0). Most of these thresholds were not sensitive to choice of the number of domains used to base classification, either 13, five or two (**Fig. 2**). We then used six datasets of two other nidovirus groups including arteriviruses to verify support of the observed levels and clustering: 1) all viruses of the order *Nidovirales* (the family *Arteriviridae* + family *Coronaviridae* + Invertebrate nidoviruses, ACTI group); 2) 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 outgroup in phylogeny, PACTI group. Results of these analyses supported strongly (CC=0) a level of arterivirus classification with five clusters corresponding to the current arterivirus genera. When arteriviruses were analyzed in combination with other nidoviruses, this five-taxa clustering was observed at the level of subfamily rank for other nidovirus lineages. For the sake of consistency, the ASG decided to reassign these five taxa to subfamily rank as well (**Fig. 3**). From the four levels of the DEmARC classification below the subfamily rank (DEmARC level 5 in Ar 5d dataset) three levels were consistently supported by datasets that included five and thirteen domains. Accordingly they were used to designate ranks genera, subgenera and species, respectively (**Fig. 4-6**). Taxa designated at the species rank included 2 new species and 17 that matched existing species, further corroborating the validity of this rank designation.

Due to good agreement between classifications obtained using thirteen and five domains, the ASG decided to adopt analysis of five replicative domains for future taxonomic development of arterivirus taxonomy.

**Demarcation criteria**. We used either a range or a particular value of patristic pairwise distances (PPD) calculated for FastTree 2.1.4 SSE3 ML phylogeny of MSA of five concatenated domains 3CLpro, NiRAN, RdRp, ZBD and HEL1 as demarcation criterion to taxa at each of four ranks: subfamily, genus, subgenus, and species (**Table 1**). They were selected as local minima in the CC distribution, commonly corresponding to the CC=0 (see above).

**Table 1** Demarcation thresholds for four ranks of the *Arteriviridae* family

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **rank** | **Taxa #** | **PPD range1** | **PUD (%) range2** | **Dataset used3** |
| subfamily | 5 | 1.172-1.315 | 0.457-0.486 | Arteri\_5d |
| genus | 11 | 0.659-0.680 | 0.325-0.332 | Arteri\_5d |
| subgenus | 16 | 0.304-0.380 | 0.192-0.225 | Arteri\_5d |
| species | 19 | 0.206 | 0.144 | Arteri\_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.

**Naming**. Names of all taxa of the *Arteriviridae* family were revised in this proposal as a result of adopting a binominal species nomenclature.

Origins of names of taxa:

**Subfamilies**

*Equarterivirinae* - (equine viruses)

*Simarterivirinae* - (simian viruses)

*Variarterivirinae* - (LDV, 2 x PRRSV, 2 x rat) from heterogeneous

*Zealarterivirinae* - (WPDV) from New Zealand - location of first virus isolated in this subfamily

*Heroarterivirinae* - (Gambian pouched rat virus) HeroRATS – trained as buried bomb sniffers

**Genera**:

Names of all genera taxa include ‘arterivirus’ which is preceded by a prefix corresponding to a Greek letter

**Subgenera**:

Names of all subgenera are formed from a unique part followed by the common ending with ‘artevirus’. The unique part ends with the first letter of the corresponding genus name.

**Species**:

*Alphaarterivirus equid* (EAV) equine virus

*Betaarterivirus suid 1* (PRRSV-1) virus of suid, hoofed mammal

*Betaarterivirus suid 2* (PRRSV-2) virus of suid, hoofed mammal

*Betaarterivirus chinrav 1* (RatAV) KP280006.1 Chinese RatAV

*Betaarterivirus ninrav* (RatAV\_Ningxia2015) KU302440.1 Ningxia RatAV

*Gammaarterivirus lacdeh* (LDV) lactate dehydrogenase elevating virus

*Deltaarterivirus hemfev* (SHFV) simian hemorrhagic fever virus

*Deltaarterivirus pejah* (PBJV) Pebjah virus

*Epsilonarterivirus hemcep* (SHEV) simian hemorrhagic encephalitis virus

*Epsilonarterivirus safriver* (FSVV) South African free state vervet virus KR862307.1

*Epsilonarterivirus zamalb* (ZMbV-1) Zambian malbrouck virus 1KT166441.1

*Zetaarterivirus ugarco 1* (KRCV-1) Ugandan red colobus virus 1

*Etaarterivirus ugarco 1* (KRCV-2) Ugandan red colobus virus 2

*Thetaarterivirus mikelba 1* (MYBV) Mikumi yellow baboon virus 1

*Thetaarterivirus kafuba* (KKCBV) Kafue kinda chacma baboon virus

*Iotaarterivirus kibreg 1* (KRTGV) Kibale red-tailed guenon virus 1

*Iotaarterivirus debrazmo* (DeMAV) DeBrazza monkey arterivirus

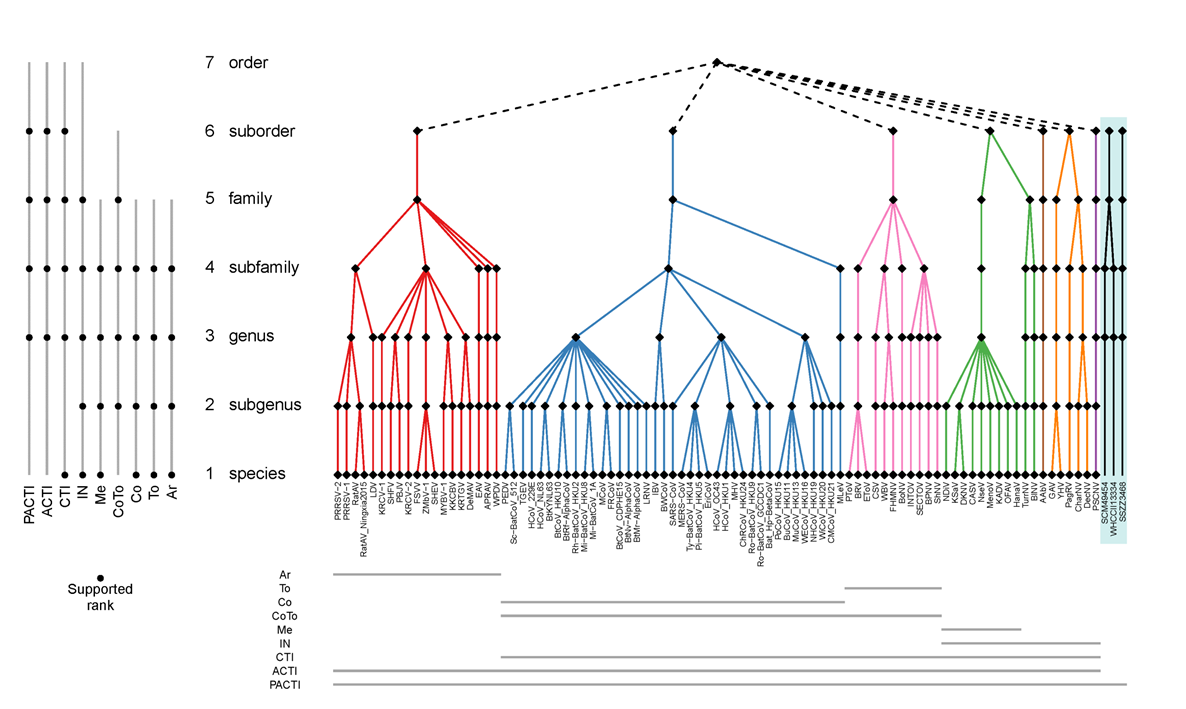
*Kappaarterivirus wobum* (WPDV) wobbly possum disease virus

*Lambdaarterivirus afriporav* (APRAV) African pouched rat arterivirus

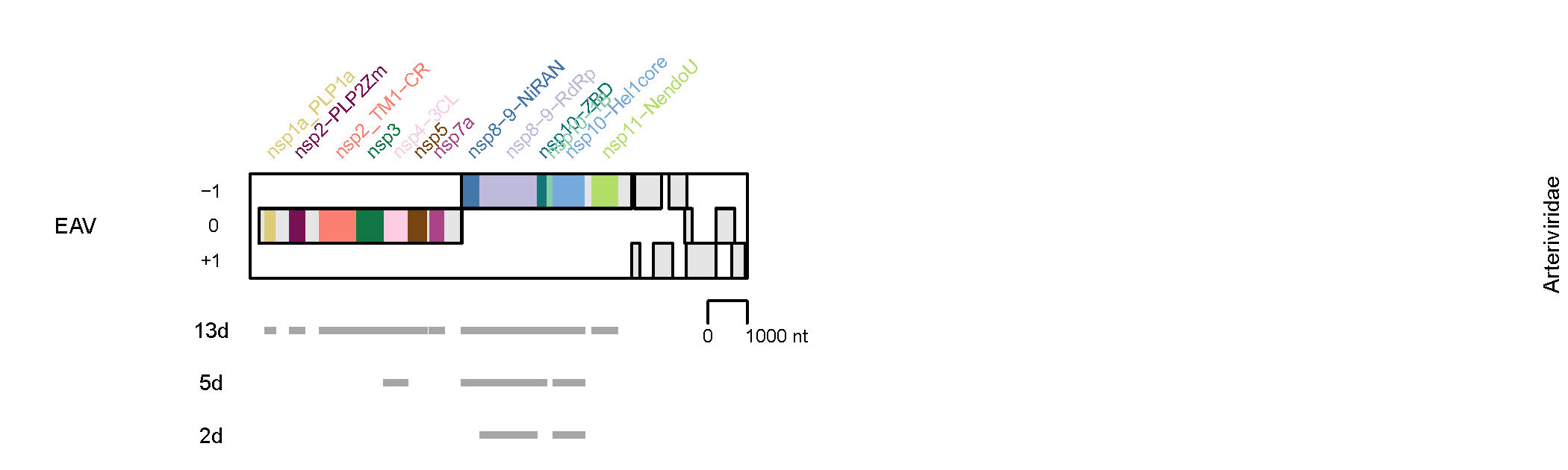
| **References:** |
| --- |
| 1. Gulyaeva, Anastasia A., Lauber, Chris, Samborskiy, Dmitry V., Leontovich, Andrey M., Sidorov, Igor A. and Alexander E. Gorbalenya (2017) Evolutionary based classification of genomic diversity of nidoviruses connects metagenomics and experimental research. Proceedings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

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

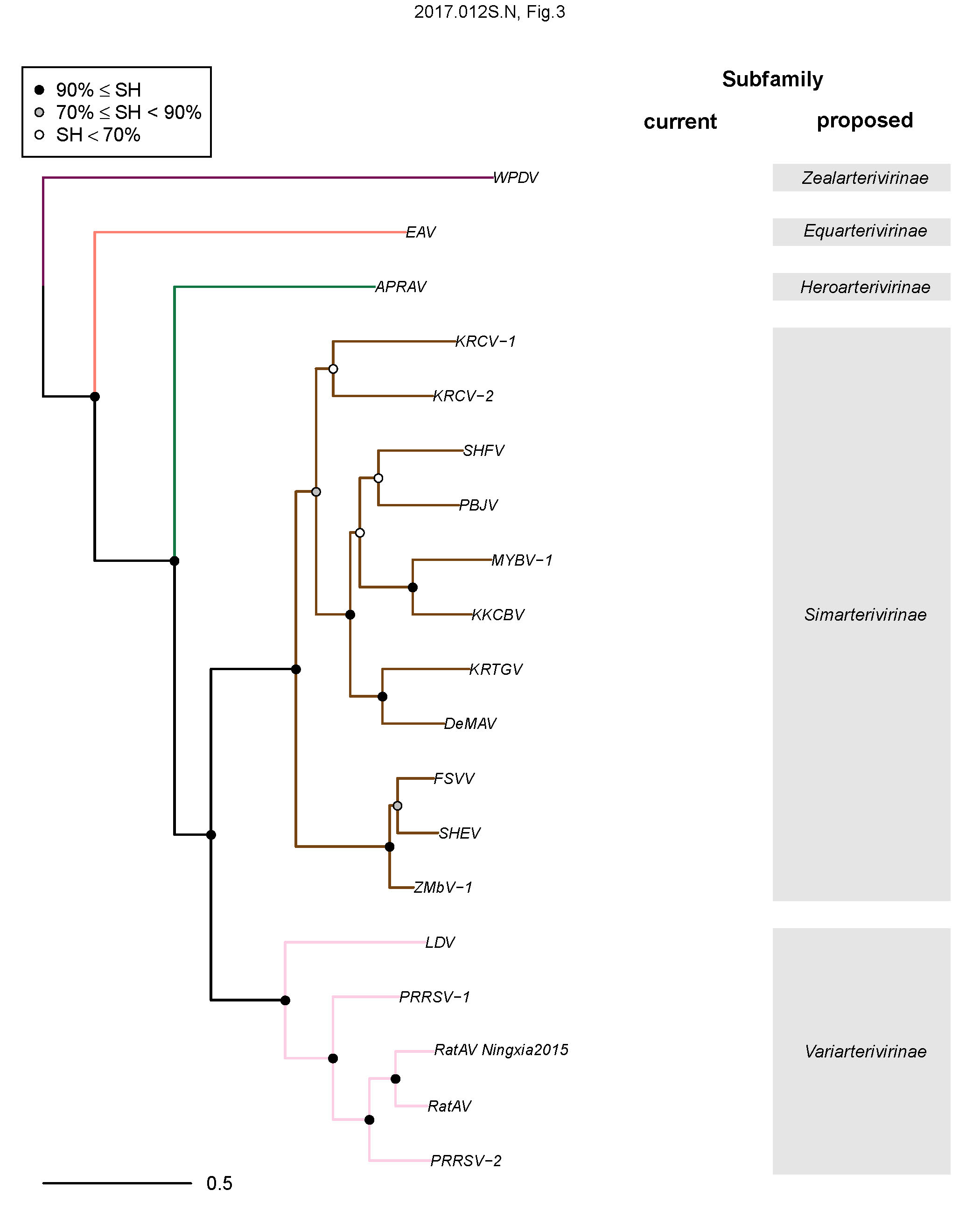
| additional material in support of this proposal |
| --- |
| **Annex:** |



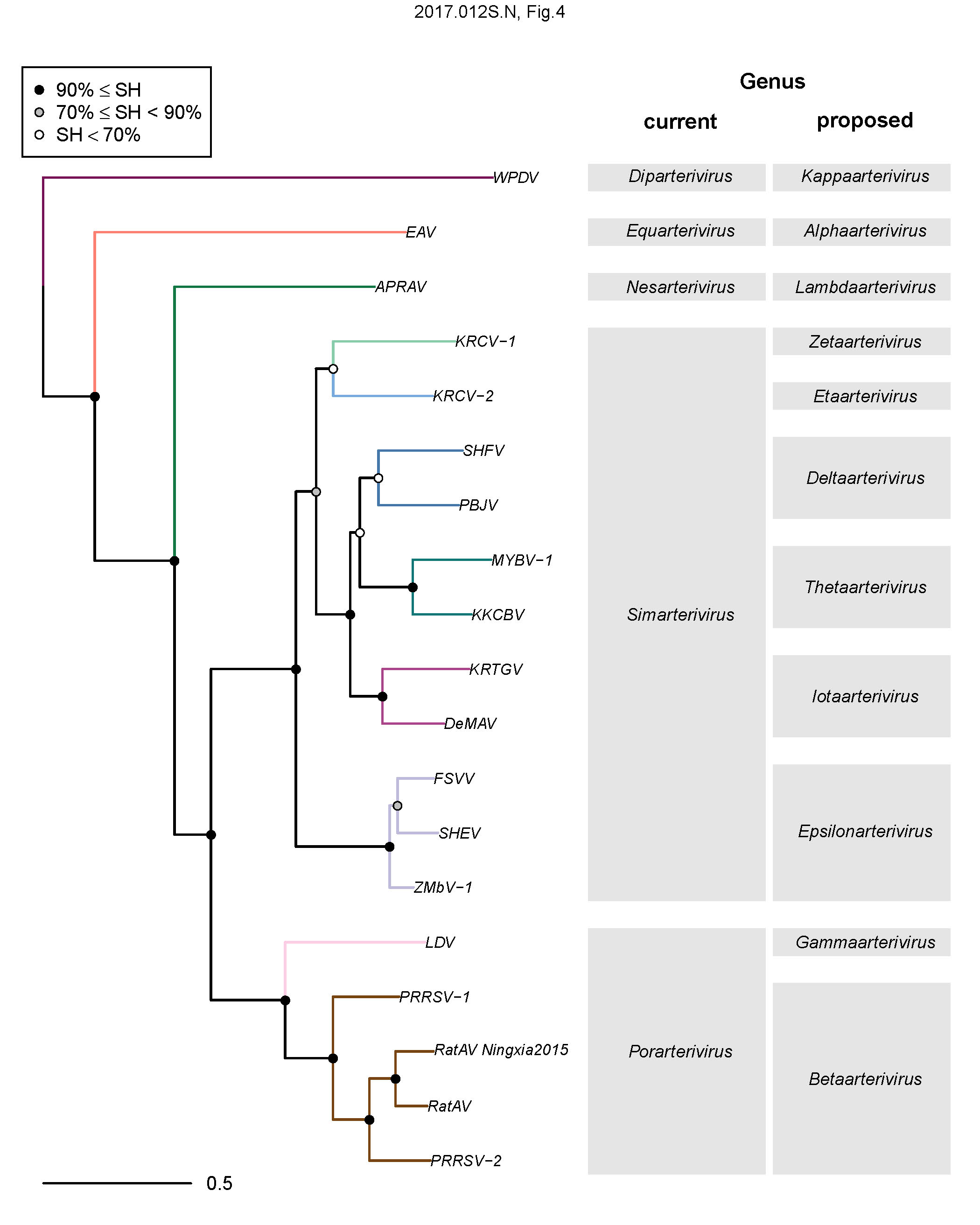
**Fig. 1**. Proposed taxonomy of the order *Nidovirales* and sequences datasets analysed to produce it. Color panel depicts the proposed seven-rank taxonomy of the order *Nidovirales* along with a monophyletic sister group of unclassified invertebrate viruses, and with each suborder colored differently. 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.



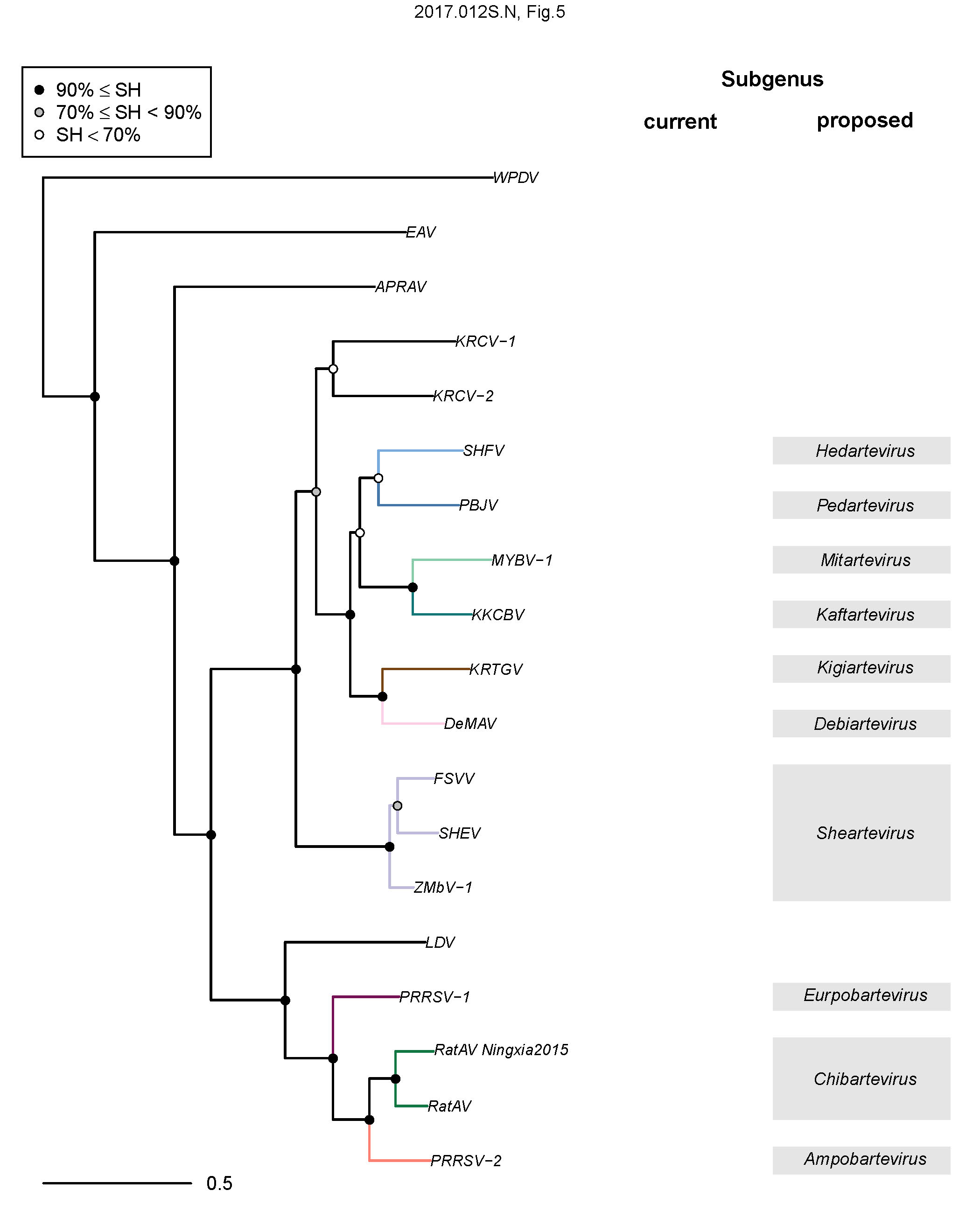
**Fig. 2**. Domain combinations used for phylogenetic and DEmARC analyses of the *Arteriviridae* family. Shown are composition of three combinations of conserved replicative domains (13d, 5d, and 2d) used in this analysis of the Ar group (see Fig. 1) and depicted relative to the genome and open reading frames of Equine arteritis virus (EAV). 13d, 13 domains; 5d, 5 domains; 2d, two domains. The results shown in **Figs 3-6** were obtained using the 5d combination. (Gulyaeva et al & Gorbalenya, unpublished).



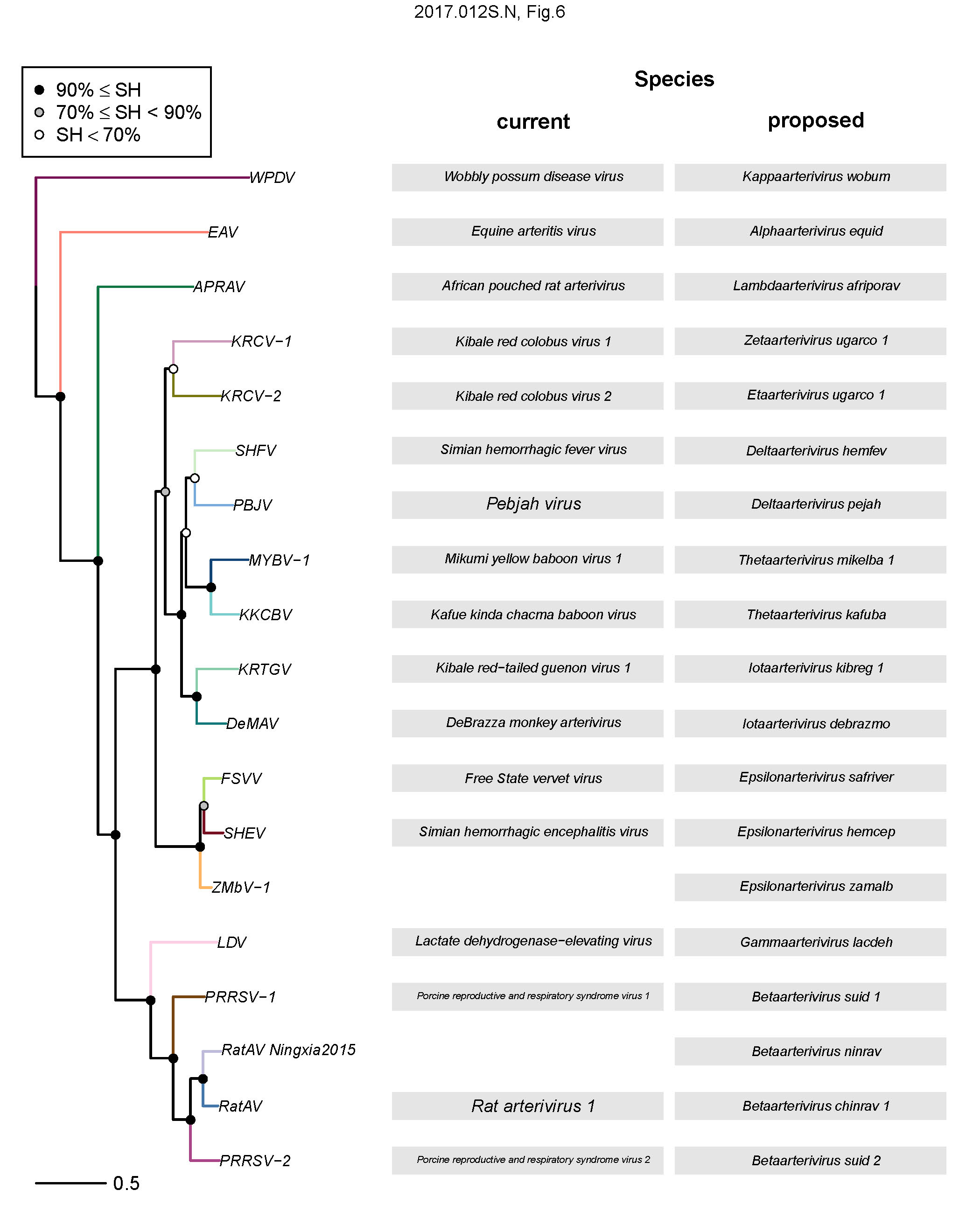
**Fig. 3**. Cluster partitioning of the phylogenetic tree of the *Arteriviridae* to subfamilies by DEmARC. Shown is the ML tree of 19 arteriviruses, which was derived from a tree, reconstructed by FastTree for a dataset of 822 arteriviruses using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains, by pruning off all tips except a single representative of each species delineated by DEmARC. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to 5 clusters at level 5 of DEmARC classification (**subfamily taxa**) of the Ar 5d dataset. The current and proposed subfamily structure of arteriviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



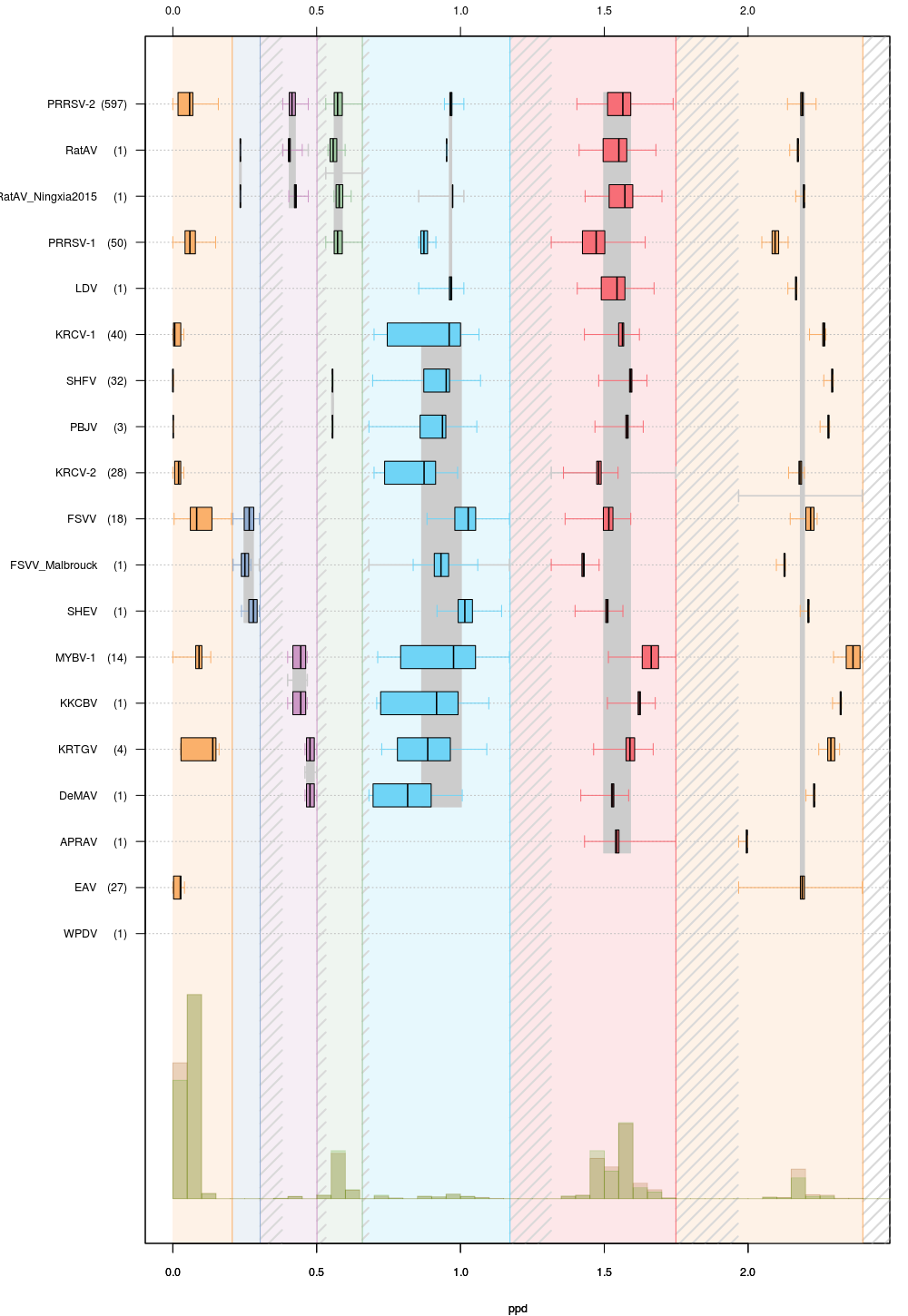
**Fig. 4**. Cluster partitioning of the phylogenetic tree of the *Arteriviridae* to genera by DEmARC. Shown is the ML tree of 19 arteriviruses, which was derived from a tree, reconstructed by FastTree for a dataset of 822 arteriviruses using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains, by pruning off all tips except a single representative of each species delineated by DEmARC. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to 11 clusters at level 4 of DEmARC classification (**genus taxa**) of the Ar 5d dataset. The current and proposed genus structure of arteriviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 5**. Cluster partitioning of the phylogenetic tree of the *Arteriviridae* to subgenera by DEmARC. Shown is the ML tree of 19 arteriviruses, which was derived from a tree, reconstructed by FastTree for a dataset of 822 arteriviruses using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains, by pruning off all tips except a single representative of each species delineated by DEmARC. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to 10 of 16 clusters at level 2 of DEmARC classification (**subgenus taxa**) of the Ar 5d dataset. The remaining six clusters, each comprising an entire cluster at the level 4 (genus rank), were excluded from the taxonomy assignment at the subgenus rank (black colored branches in the tree).The current and proposed subgenus structure of arteriviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 6.** Cluster partitioning of the phylogenetic tree of the *Arteriviridae* to species by DEmARC. Shown is the ML tree of 19 arteriviruses, which was derived from a tree, reconstructed by FastTree for a dataset of 822 arteriviruses using concatenated MSAs of 3CLpro, NiRAN, RdRp, ZBD, HEL1 domains, by pruning off all tips except a single representative of each species delineated by DEmARC. For the virus abbreviations, see the accompanying spreadsheet 2017.012-015S.N.v3.Nidovirales\_rev. The tree is colored according to 19 clusters at level 1 of DEmARC classification (**species taxa**) of the Ar 5d dataset. The current and proposed species structure of arteriviruses are detailed at the right of the virus names. (Gulyaeva et al & Gorbalenya, unpublished).



**Fig. 7.** Intra-cluster genetic divergence in seven-level hierarchical clustering of the current family *Arteriviridae* (Ar 5d dataset, **Fig. 1**) by DEmARC. Levels are defined by seven strongest PPD thresholds. For simplicity, identities of clusters at the lowest level are indicated via species acronyms (left axis; FSVV\_Malbrouck stands for ZMbV-1 in other figures); the number of viruses in the identified clusters are shown in brackets. All identified clusters correspond to monophyletic groups on a FastTree 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 subfamily, genus, subgenus and species ranks corresponds to fifth (light blue), fourth (green), second (dark blue) and first (left-most orange) levels, respectively, of this classification. (Gulyaeva et al. & Gorbalenya, unpublished)