This Word module should be used for all taxonomic proposals.

Please complete **Part 1** and:

either **Part 3** for proposals to create new taxa or change existing taxa

or **Part 2** for proposals of a general nature.

Submit the completed Word module, together with the accompanying Excel module named in Part 3, to the appropriate ICTV Subcommittee Chair.

The Word module explains and justifies your proposal. The Excel module is a critical document that will be used to implement the proposed taxonomic changes once they are approved and ratified. If proposals presented in the Word module are not presented accurately in the Excel module, the taxonomic changes cannot proceed.

For guidance, see the notes written in blue, below, and the Help Notes in file Taxonomic\_Proposals\_Help\_2019.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Code assigned:** | ***2019.012D*** | |  |
| **Short title:** Create 1 new phylum (*Cressdnaviricota*) including 2 classes and 6 orders for classification of CRESS-DNA viruses | | | |
|  | | | |
| **Author(s) and email address(es):** | | | |
| List authors in a single line *Archives of Virology* citation format (e.g. Smith AB, Huang C-L, Santos, F) | | Provide email address for each author in a single line separated by semi-colons | |
| Krupovic M, Varsani A, Kuhn J, Kazlauskas D, Breitbart M, Delwart E, Rosario K, Yutin N, Wolf YI, Harrach B, Zerbini FM, Dolja VV, Koonin EV | | [mart.krupovic@pasteur.fr](mailto:mart.krupovic@pasteur.fr); [Arvind.varsani@asu.edu](mailto:Arvind.varsani@asu.edu); [kuhnjens@niaid.nih.gov](mailto:kuhnjens@niaid.nih.gov); [dariausk@gmail.com](mailto:dariausk@gmail.com); [mya@usf.edu](mailto:mya@usf.edu); [Eric.Delwart@ucsf.edu](mailto:Eric.Delwart@ucsf.edu); [krosari2@mail.usf.edu](mailto:krosari2@mail.usf.edu); [yutin@ncbi.nlm.nih.gov](mailto:yutin@ncbi.nlm.nih.gov); [wolf@ncbi.nlm.nih.gov](mailto:wolf@ncbi.nlm.nih.gov); [balazs.harrach@gmail.com](mailto:balazs.harrach@gmail.com); [zerbini@ufv.br](mailto:zerbini@ufv.br); [doljav@science.oregonstate.edu](mailto:doljav@science.oregonstate.edu); [koonin@ncbi.nlm.nih.gov](mailto:koonin@ncbi.nlm.nih.gov) | |
| **Author(s) institutional address(es) (optional):**   |  | | --- | | Provide institutional addresses, each on a single line followed by author(s) initials (e.g. University of Woolloomooloo [SAB, HCL]) | | Institut Pasteur, France [KM]  Arizona State University, USA [VA]  NIH/NIAID/DCR Integrated Research Facility at Fort Detrick, USA [KJ]  Vilnius University, Lithuania [KD]  University of South Florida, USA [BM, RK]  University of California San Francisco, USA [DE]  NIH, USA [YN, WY, KEV]  Centre for Agricultural Research, Hungary [HB]  Universidade Federal de Viçosa, Viçosa [ZFM]  Oregon State University, USA [DVV] | | | | |
| **Corresponding author** | | | |
| Mart Krupovic; [mart.krupovic@pasteur.fr](mailto:mart.krupovic@pasteur.fr) | | | |
| **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) | | **Direct submission to ICTV Executive Committee** | |
| **ICTV Study Group comments (if any) and response of the proposer:** | | | |
|  | | | |
|  | | | |
| Date first submitted to ICTV: | | | June 19, 2019 |
| Date of this revision (if different to above): | | | October 8, 2019 |

|  |
| --- |
| **ICTV-EC comments and response of the proposer:** |
| More explanation about the taxonomic levels that are proposed for all those new clades.  Response: Inserted |

**Part 3:** **PROPOSED TAXONOMY**

|  |
| --- |
| **Name of accompanying Excel module:** 2019.012D.A.v1.Cressdnaviricota |

**Supporting material:**

Eukaryotic ssDNA viruses are classified into 10 families, although many uncultured ssDNA viruses remain unclassified (Simmonds et al., 2017). Viruses from seven of these families, namely, *Bacilladnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Nanoviridae* and *Smacoviridae* as well as the proposed family “*Redondoviridae*” (Abbas et al., 2019), have small circular genomes, many encoding only 2 proteins, one for the genome replication initiation protein and the other for the capsid protein. The genomes are known or predicted to be replicated by the rolling-circle mechanism, which is initiated by the virus-encoded Rep protein of the HUH endonuclease superfamily, characterized by the signature HUH (or HUQ) motif, in which two histidine (or histidine and glutamine) residues are separated by a bulky hydrophobic residue (Figure 1) (Chandler et al., 2013; Ilyina and Koonin, 1992; Koonin and Ilyina, 1993; Krupovic, 2013; Rosario et al., 2012). In eukaryotic ssDNA viruses, the HUH endonuclease domain is followed by a superfamily 3 helicase domain, which is not found in bacterial or archaeal viruses, suggesting a closer evolutionary relationship among eukaryotic ssDNA viruses. Informally, these viruses are often collectively referred to as circular, Rep-encoding ssDNA (CRESS-DNA) viruses (Rosario et al., 2012; Zhao et al., 2019). Viruses from the 7 families infect hosts across the eukaryotic domain, including plants (*Nanoviridae* and *Geminiviridae*), fungi (*Genomoviridae*), animals (*Circoviridae*), and algae (*Bacilladnaviridae*), whereas members of the families *Smacoviridae* and “*Redondoviridae*” have been discovered by metagenomics and are suspected to infect animals, although association of smacovirids with methanogenic archaea has also been suggested (Díez-Villaseñor and Rodriguez-Valera, 2019).

Unlike the Rep proteins, the capsid proteins of CRESS DNA viruses are not orthologous, but were suggested to be acquired from RNA viruses on multiple independent occasions (Diemer and Stedman, 2012; Kazlauskas et al., 2017; Krupovic and Koonin, 2017; Krupovic et al., 2009; Roux et al., 2013). Thus, the characteristic two-domain Rep protein of CRESS DNA viruses is the only universally conserved gene in this virus super group and can serve as a convenient scaffold for mapping evolutionary events, much like RNA-directed RNA polymerase is for RNA viruses (Koonin et al., 2015; Koonin et al., 2008; Wolf et al., 2018), or rRNA genes are for cellular organisms across the tree of life. Thus, we propose to use Rep proteins as a basis for megataxonomic classification of CRESS-DNA viruses.

Phylogenetic analysis of classified and unclassified CRESS-DNA viruses revealed 13 well supported clades (Figure 2). Six of these clades correspond to unclassified CRESS-DNA viruses, denoted CRESSV1 through CRESSV6 (Kazlauskas et al., 2018). The phylogeny splits into 2 large clades: (i) the first clade includes members of the families *Geminiviridae* and *Genomoviridae*, and CRESSV6 viruses; (ii) the second clade includes all other CRESS-DNA viruses.

Reps in the clade 1 share the GRS motif located between motifs II and III (not shown in Figure 1), which is not found in Reps of CRESS-DNA viruses from the second clade and is thought to enable the appropriate spatial arrangement of motifs II and III (Nash et al., 2011; Varsani and Krupovic, 2017). Furthermore, geminivirids and genomovirids, but not CRESSV6, lack the arginine finger in the helicase motif (Figure 1). Thus, we propose to unify families *Geminiviridae* and *Genomoviridae* into the order *Geplafuvirales* (*ge*- for gemini/genomo; *pla*- for plants; *fu*- for fungi). In the future, CRESSV6 are expected to be grouped with “geplafuviruses” at the level of class, which we propose to name *Repensiviricetes* (a portmanteaux of Rep-encoding single strand; and the suffix -*vircetes* for class taxa).

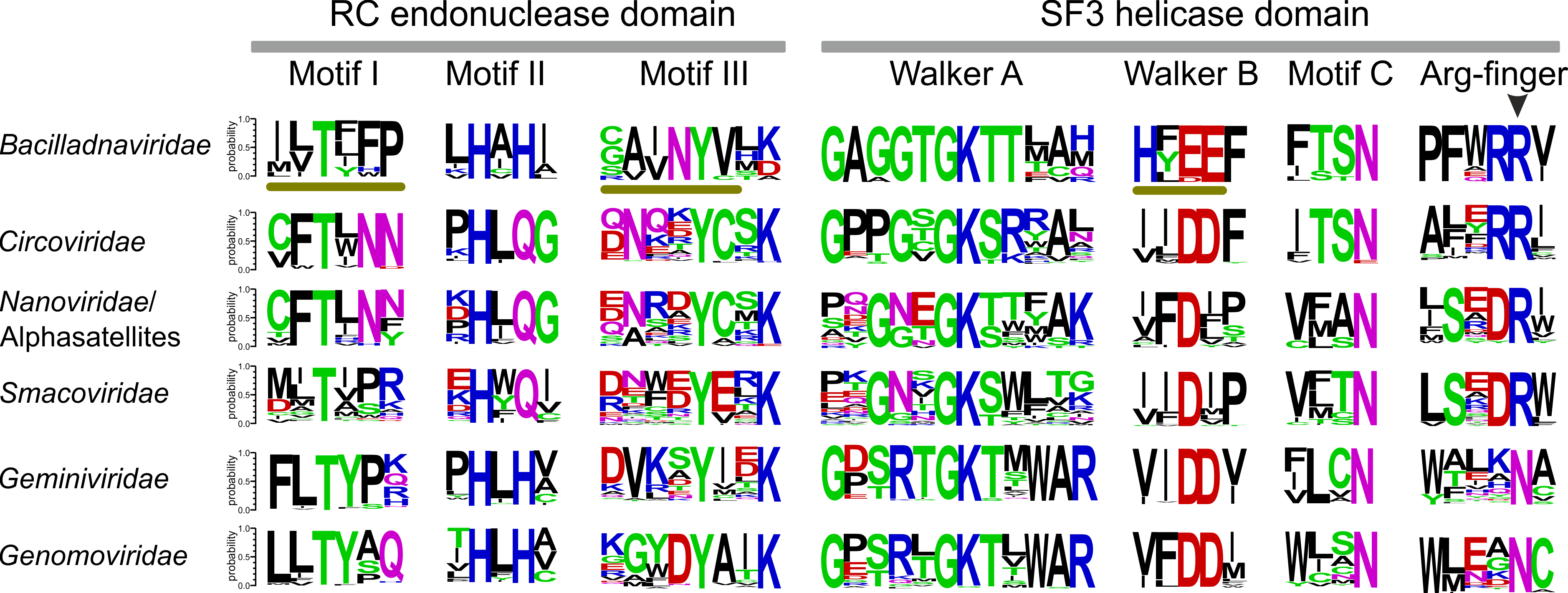
Clade 2 includes families *Bacilladnaviridae*, *Circoviridae*, *Smacoviridae*, *Nanoviridae*, and the proposed family “*Redondoviridae*” as well as unclassified virus groups CRESSV1-5 (Figure 2). Unlike members of the proposed class *Repensiviricetes*, viruses in this assemblage encode Reps lacking the GRS motif. Furthermore, Reps of all members of clade 2 contain a conserved arginine finger in the helicase domain, absent in the geminivirids and genomovirids. Collectively, these characters support the monophyly of the Reps in clade 2. Nevertheless, the established/proposed virus families include viruses which are sufficiently distinct from each other (and from viruses in the proposed order “*Geplafuvirales*”), precluding their unification into a single order. The differences extend beyond the phylogenetic relationships of their Reps and include non-homologous capsid proteins (e.g., bacilladnaviruses are predicted to build T=3 icosahedral capsids, whereas capsids of other viruses, where known, are built on the T=1 icosahedral lattice), considerably different genome lengths (bacilladnavirid genomes are ~5.5–6 kb, whereas those of circovirids are ~1.2 kb) and genome architectures (e.g., nanovirids have multipartite genomes, whereas other viruses in clade 2 have monopartite genomes). Importantly, circovirids cluster with CRESSV1-3, whereas nanovirids cluster with CRESSV4 and CRESSV5, suggesting that order-rank taxa will be needed for these assemblages. Thus, we propose to create five orders, which will include the five families:

* *Baphyvirales* (to include *Bacilladnaviridae*): after bacillariophytes (diatoms), hosts of viruses in this taxon; and the suffix -virales for order taxa;
* *Cirlivirales* (to include *Circoviridae* and, in due course, CRESSV1-3): *cir*- for circoviruses; and *li*- for circo-like; and the suffix -virales for order taxa;
* *Cremevirales* (to include *Smacoviridae*): *Cre*- for CRESS; and *me*- for metagenomics; and the suffix -virales for order taxa;
* *Mulpavirales* (to include *Nanoviridae* and, in due course, CRESSV4 and CRESSV5): after multipartite genomes of viruses in this taxon; and the suffix -virales for order taxa;
* *Recrevirales* (to include “*Redondoviridae*”): *Re*- for redondoviruses; and *cre*- for CRESS; and the suffix -virales for order taxa.

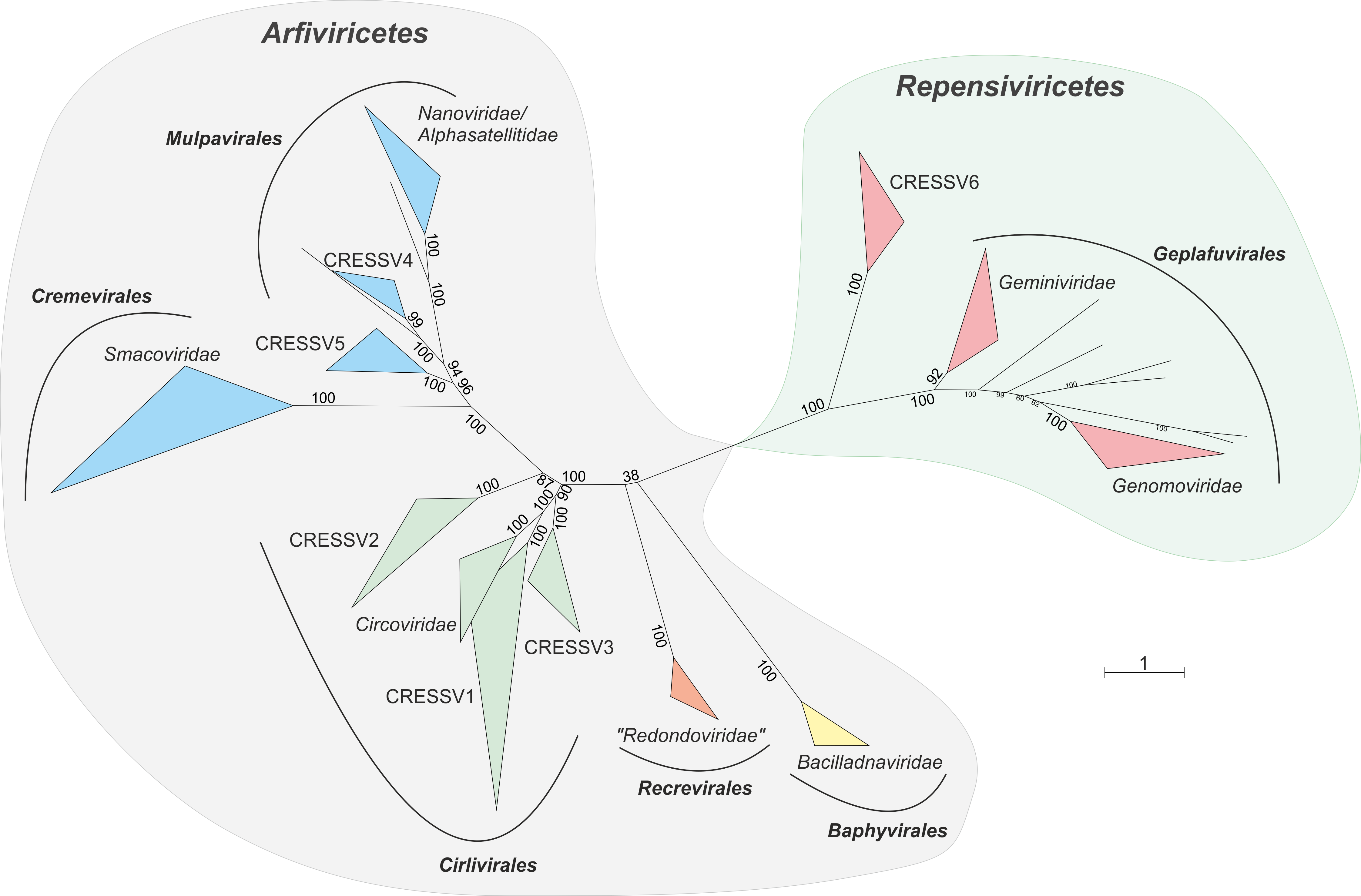
Viruses from established families within the 5 proposed orders encode order-specific capsid proteins, which are not recognizably similar at the sequence level across the orders. This feature as well as Rep phylogeny might help ascribing the newly identified viruses to a particular taxon. We propose to unify the 5 orders into a class *Arfiviricetes* (*Ar*- for arginine; and *fi*- for finger; and the suffix -*vircetes* for class taxa.

Overall, taxonomic levels are suggested as to be comparable to the relatedness level of the higher taxa proposed in the parallel taxonomic proposals for *Divdnaviria*, *Duplodnaviria*, *Monodnaviria* and *Riboviria*.

Finally, we propose to unify the two classes into a phylum *Cressdnaviricota*: a contraction of CRESS DNA; and the suffix -*viricota* for phylum taxa.



**Figure 1.** Sequence motifs of Rep proteins from CRESS-DNA viruses. Bacilladnavirus-specific motifs are underlined. Residues are coloured according to their chemical properties (polar, green; basic, blue; acidic, red; hydrophobic, black; neutral, purple). The figure is adapted from (Kazlauskas et al., 2017).



**Figure 2.** Unrooted maximum likelihood phylogenetic tree of Rep proteins from CRESS-DNA viruses. Closely related sequence groups are collapsed into triangles, whose side lengths are proportional to the distances between closest and farthest leaf nodes. The Rep alignment used for the tree reconstruction was taken from (Kazlauskas et al., 2017) and supplemented with sequences of “redondovirids” (Abbas et al., 2019). The maximum likelihood phylogenetic tree was constructed using the PhyML program (Guindon et al., 2010) with automatic selection of the best-fit substitution model for a given alignment. The best model identified by PhyML was RtREV. The branch support was assessed using aBayes implemented in PhyML. The scale bar represents the number of substitutions per site.

| **References:** |
| --- |
| Abbas, A.A., Taylor, L.J., Dothard, M.I., Leiby, J.S., Fitzgerald, A.S., Khatib, L.A., Collman, R.G., Bushman, F.D., 2019. Redondoviridae, a family of small, circular DNA viruses of the human oro-respiratory tract associated with periodontitis and critical illness. Cell Host Microbe 25, 719-729.e4.  Chandler, M., de la Cruz, F., Dyda, F., Hickman, A.B., Moncalian, G., Ton-Hoang, B., 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol 11, 525-538.  Diemer, G.S., Stedman, K.M., 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. Biol Direct 7, 13.  Díez-Villaseñor, C., Rodriguez-Valera, F., 2019. CRISPR analysis suggests that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. Nat Commun 10, 294.  Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307-321.  Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res 20, 3279-3285.  Kazlauskas, D., Dayaram, A., Kraberger, S., Goldstien, S., Varsani, A., Krupovic, M., 2017. Evolutionary history of ssDNA bacilladnaviruses features horizontal acquisition of the capsid gene from ssRNA nodaviruses. Virology 504, 114-121.  Kazlauskas, D., Varsani, A., Krupovic, M., 2018. Pervasive chimerism in the replication-associated proteins of uncultured single-stranded DNA viruses. Viruses 10, E187.  Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479-480, 2-25.  Koonin, E.V., Ilyina, T.V., 1993. Computer-assisted dissection of rolling circle DNA replication. Biosystems 30, 241-268.  Koonin, E.V., Wolf, Y.I., Nagasaki, K., Dolja, V.V., 2008. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. Nat Rev Microbiol 6, 925-939.  Krupovic, M., 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Curr Opin Virol 3, 578-586.  Krupovic, M., Koonin, E.V., 2017. Multiple origins of viral capsid proteins from cellular ancestors. Proc Natl Acad Sci U S A 114, E2401-E2410.  Krupovic, M., Ravantti, J.J., Bamford, D.H., 2009. Geminiviruses: a tale of a plasmid becoming a virus. BMC Evol Biol 9, 112.  Nash, T.E., Dallas, M.B., Reyes, M.I., Buhrman, G.K., Ascencio-Ibanez, J.T., Hanley-Bowdoin, L., 2011. Functional analysis of a novel motif conserved across geminivirus Rep proteins. J Virol 85, 1182-1192.  Rosario, K., Duffy, S., Breitbart, M., 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch Virol 157, 1851-1871.  Roux, S., Enault, F., Bronner, G., Vaulot, D., Forterre, P., Krupovic, M., 2013. Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. Nat Commun 4, 2700.  Simmonds, P., Adams, M.J., Benko, M., Breitbart, M., Brister, J.R., Carstens, E.B., Davison, A.J., Delwart, E., Gorbalenya, A.E., Harrach, B., Hull, R., King, A.M., Koonin, E.V., Krupovic, M., Kuhn, J.H., Lefkowitz, E.J., Nibert, M.L., Orton, R., Roossinck, M.J., Sabanadzovic, S., Sullivan, M.B., Suttle, C.A., Tesh, R.B., van der Vlugt, R.A., Varsani, A., Zerbini, F.M., 2017. Consensus statement: Virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15, 161-168.  Varsani, A., Krupovic, M., 2017. Sequence-based taxonomic framework for the classification of uncultured single-stranded DNA viruses of the family *Genomoviridae*. Virus Evol 3, vew037.  Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucia-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V., Koonin, E.V., 2018. Origins and evolution of the global RNA virome. MBio 9, e02329-02318.  Zhao, L., Rosario, K., Breitbart, M., Duffy, S., 2019. Eukaryotic circular Rep-encoding single-stranded DNA (CRESS DNA) viruses: Ubiquitous viruses with small genomes and a diverse host range. Adv Virus Res 103, 71-133. |