AMERICAN SOCIETY FOR MICROBIOLOGY Journal of Virology®

LETTER TO THE EDITOR

Cressdnaviricota: a Virus Phylum Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes

Mart Krupovic,[a] Arvind Varsani,[b,c] Darius Kazlauskas,[d] Mya Breitbart,[e] Eric Delwart,[f,g] Karyna Rosario,[e] Natalya Yutin,[h] Yuri I. Wolf,[h] Balázs Harrach,[i] F. Murilo Zerbini,[j] Valerian V. Dolja,[k] Jens H. Kuhn,[l] Eugene V. Koonin[h]

[a]Archaeal Virology Unit, Institut Pasteur, Paris, France
[b]The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[c]Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
[d]Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania
[e]University of South Florida, College of Marine Science, St. Petersburg, Florida, USA
[f]Vitalant Research Institute, San Francisco, California, USA
[g]UCSF Department of Laboratory Medicine, University of California—San Francisco, San Francisco, California, USA
[h]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA
[i]Institute for Veterinary Medical Research, Centre for Agricultural Research, Budapest, Hungary
[j]Departamento de Fitopatologia/Bioagro, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil
[k]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, USA
[l]Integrated Research Facility at Fort Detrick, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Frederick, Maryland, USA

KEYWORDS *Bacilladnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Nanoviridae*, *Redondoviridae*, *Smacoviridae*, International Committee on Taxonomy of Viruses (ICTV), CRESS DNA viruses, virus classification, virus evolution

Eukaryotic single-stranded DNA (ssDNA) viruses are classified into 10 families (Table 1), but many remain unclassified (1, 2). Viruses from seven families (*Bacilladnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Nanoviridae*, *Redondoviridae*, and *Smacoviridae*) have small circular genomes, with some encompassing only two genes, encoding the replication initiation protein (Rep) and the capsid protein (3–5). The genomes are replicated by the rolling-circle mechanism initiated by the virus-encoded Rep proteins of the HUH endonuclease superfamily (6–10). In eukaryotic ssDNA viruses, the HUH endonuclease domain is followed by a superfamily 3 helicase (S3H) domain, which is not found in prokaryotic ssDNA virus Reps, although homologous two-domain Reps are encoded by several groups of bacterial plasmids (5). Informally, this assemblage of ssDNA viruses has been referred to as "circular, Rep-encoding single-stranded (CRESS) DNA viruses" (4, 8). Viruses from the seven families infect hosts across the eukaryotic domain, including animals (*Circoviridae*), diatoms (*Bacilladnaviridae*), fungi (*Genomoviridae*), and plants (*Nanoviridae* and *Geminiviridae*), whereas members of the *Smacoviridae* and the newly created *Redondoviridae* have been discovered by metagenomics and are suspected to infect animals (3, 11), although an association of smacovirids with methanogenic archaea has also been suggested (12).

Phylogenetic analysis of classified and unclassified CRESS DNA viruses revealed 13 well-supported groups (Fig. 1). Six groups correspond to unclassified viruses, denoted CRESSV1 through CRESSV6 (13). The phylogeny splits into two large clades: (i) geminivirids, genomovirids, and CRESSV6 viruses and (ii) all other CRESS DNA

Citation Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, Rosario K, Yutin N, Wolf YI, Harrach B, Zerbini FM, Dolja VV, Kuhn JH, Koonin EV. 2020. *Cressdnaviricota*: a virus phylum unifying seven families of Rep-encoding viruses with single-stranded, circular DNA genomes. J Virol 94:e00582-20. https://doi.org/10.1128/JVI.00582-20.

Editor Rozanne M. Sandri-Goldin, University of California, Irvine

Address correspondence to Mart Krupovic, mart.krupovic@pasteur.fr.

Published 1 June 2020

TABLE 1 Families of ssDNA viruses associated with eukaryotic hosts[a]

Family	Hosts	Virion diam (nm)	Genome topology	Genome size (kb)	No. of ORFs	Replication protein[b]
Anelloviridae	Animals	30–32	Circular	2–4	2–4	Rep
Bidnaviridae	Animals, insects	20–24	Linear	6–6.5 per segment (2 segments)	6	pPolB
Parvoviridae	Animals, insects	23–28	Linear	4–6.3	3–4	HUH-Rep
Bacilladnaviridae	Algae (diatoms)	33–38	Circular	5.5–6	4	HUH-Rep
Circoviridae	Animals	15–25	Circular	1.7–2.1	2	HUH-Rep
Geminiviridae	Plants	22 × 38	Circular	2.5–3 per segment (1–2 segments)	4–8	HUH-Rep
Genomoviridae	Fungi, insects	20–22	Circular	2–2.4	2	HUH-Rep
Nanoviridae	Plants	17–20	Circular	0.98–1.1 per segment (6–9 segments)	5–8	HUH-Rep
Redondoviridae	?	?	Circular	3.0–3.1	3	HUH-Rep
Smacoviridae	?	?	Circular	2.3–2.9	2	HUH-Rep

[a] Families unified into the phylum *Cressdnaviricota* are in boldface type. ORFs, open reading frames.

[b] HUH-Rep, Rep of the HUH endonuclease superfamily; pPolB, protein-primed family B DNA polymerase.

viruses (Fig. 1). These clades are supported by detailed analyses of the conserved catalytic motifs of Rep proteins (5, 14), with Reps of viruses in clade 1 sharing the GRS motif (15, 16), which is not found in Reps of viruses from clade 2. In contrast, Reps of all members of clade 2 contain a conserved arginine finger in the helicase

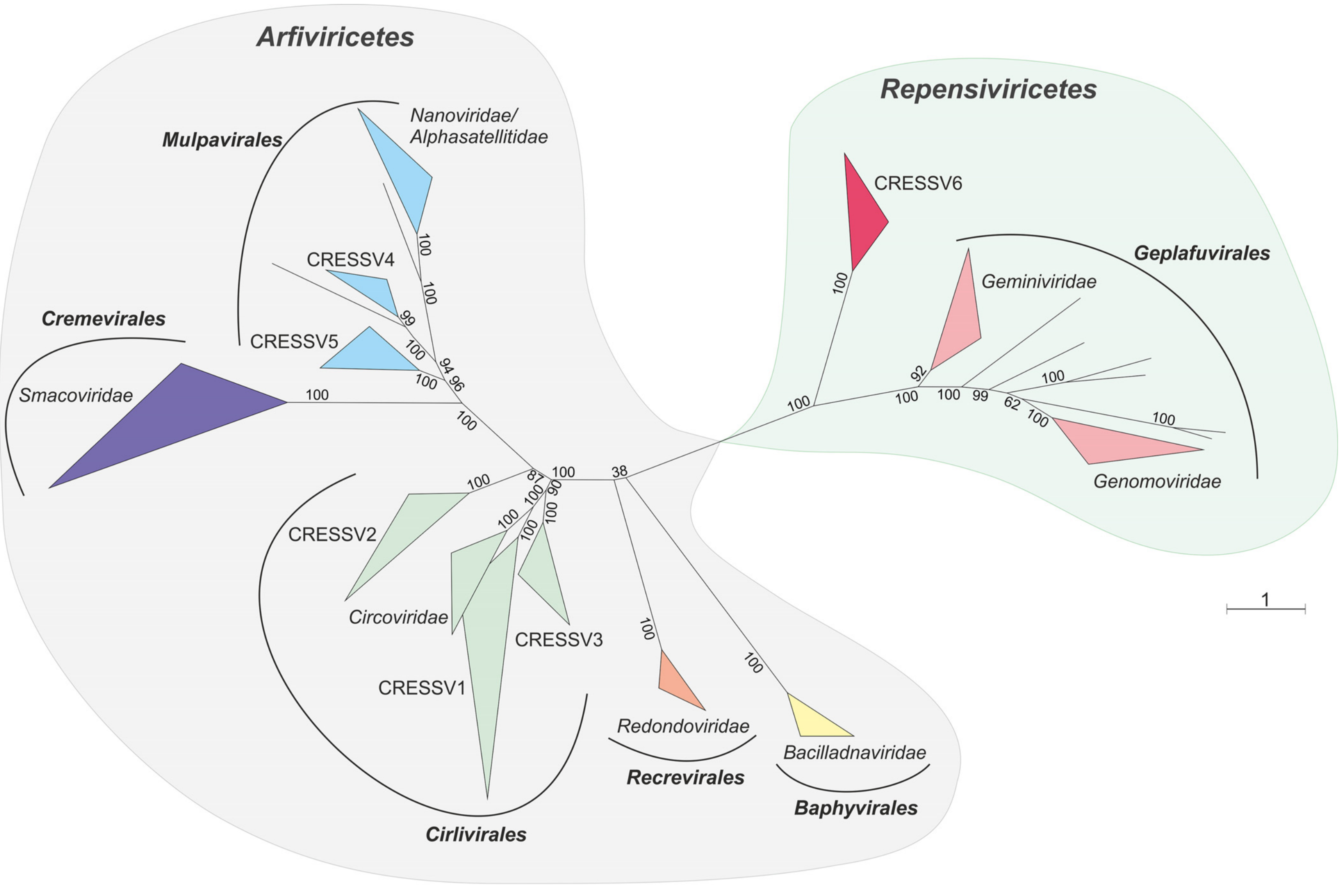

FIG 1 Unrooted maximum likelihood phylogenetic tree of Rep proteins from CRESS DNA viruses. Closely related sequence groups are collapsed into triangles, the side lengths of which are proportional to the distances between the closest and farthest leaf nodes. The Rep alignment used for the tree reconstruction was taken from the work of Kazlauskas et al. (5, 13) and supplemented with sequences of redondovirids (3). The maximum likelihood phylogenetic tree was constructed using PhyML (24) 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.

domain, absent in geminivirids and genomovirids, supporting the monophyly of the Reps in clade 2 (5).

Unlike the Reps, the capsid proteins of CRESS DNA viruses are not orthologous and apparently were acquired from RNA viruses on multiple, independent occasions (5, 14, 17–19). Thus, the gene encoding the characteristic two-domain Rep is the only universally conserved gene in this virus supergroup and thus can serve as a scaffold for mapping evolutionary events. Accordingly, it has been proposed that Reps be used as a basis for megataxonomic classification of CRESS DNA viruses (20).

Recently, the International Committee on Taxonomy of Viruses (ICTV) extended its taxonomic rank structure to facilitate the classification of the entire spectrum of virus diversity (21, 22). This new taxonomic framework was adopted to formalize the evolutionary relationships among CRESS DNA viruses. Below, we outline these taxonomic changes, which were ratified by the ICTV in March 2020 and are now part of the official virus taxonomy:

- Clade 1. The families *Geminiviridae* and *Genomoviridae* were included in the new order *Geplafuvirales* (gemini/genomo, plants, fungi, and the suffix -*virales* for order taxa). In anticipation of the future classification of the CRESSV6 virus group, the order *Geplafuvirales* has been assigned into a new class, *Repensiviricetes* (a portmanteau of Rep-encoding single strand and the suffix -*viricetes* for class taxa).
- Clade 2. Viruses of the families *Bacilladnaviridae*, *Circoviridae*, *Nanoviridae*, *Redondoviridae*, and *Smacoviridae* and unclassified groups CRESSV1 to -5 are sufficiently distinct from each other, precluding their unification into a single order. The differences extend beyond the separation of the Reps in the phylogenetic tree and include nonhomologous capsid proteins and different genome lengths (Table 1) and genome architectures (e.g., nanovirids have multipartite genomes, whereas other viruses have monopartite genomes). Notably, circovirids cluster with CRESSV1 to -3 viruses, whereas nanovirids cluster with CRESSV4 and -5 viruses, justifying order-rank taxa for these assemblages. Thus, five orders have been created for the classification of clade 2 viruses:
    - *Baphyvirales* (to include *Bacilladnaviridae*): after bacillariophytes (diatoms, hosts of viruses in this taxon) and the suffix -*virales*
    - *Cirlivirales* (to include *Circoviridae* and, in due course, CRESSV1 to -3 viruses): circo-like and the suffix -*virales*
    - *Cremevirales* (to include *Smacoviridae*): CRESS, metagenomics, and the suffix -*virales*
    - *Mulpavirales* (to include *Nanoviridae* and, in due course, CRESSV4 and CRESSV5 viruses): after multipartite genomes of viruses in this taxon and the suffix -*virales*
    - *Recrevirales* (to include the newly approved *Redondoviridae*): redondoviruses, CRESS, and the suffix -*virales*

Viruses from the established families within the five orders encode order-specific capsid proteins, which are not recognizably similar at the sequence level across the orders. This feature and the Rep phylogeny should help in assigning newly identified viruses to a particular taxon. Notably, Reps of satellite nucleic acids from the family *Alphasatellitidae* (23) and nanovirids form a clade (Fig. 1); thus, a taxon unifying *Alphasatellitidae* and *Nanoviridae* might be considered in the future. The five orders were further assigned to the class *Arfiviricetes* (arginine finger and the suffix -*viricetes* for class taxa). Finally, the two classes, *Repensiviricetes* and *Arfiviricetes*, were grouped into the phylum *Cressdnaviricota* (a contraction of CRESS DNA and the suffix -*viricota* for phylum taxa), thereby unifying all officially recognized families of CRESS DNA viruses.

ACKNOWLEDGMENTS

M.K. was supported by l'Agence Nationale de la Recherche (project ENVIRA, no. ANR-17-CE15-0005-01). E.V.K., Y.I.W., and N.Y. are supported by the Intramural Research Program of the U.S. National Institutes of Health (National Library of Medicine). B.H. was supported by the Hungarian National Research, Development, and Innovation Office (NN128309). This work was supported in part through the Laulima Government Solutions, LLC, prime contract with the U.S. National Institute of Allergy and Infectious Diseases (NIAID) under contract no. HHSN272201800013C. J.H.K. performed this work as an employee of Tunnell Government Services (TGS), a subcontractor of Laulima Government Solutions, LLC, under contract no. HHSN272201800013C.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Health and Human Services or of the institutions and companies affiliated with the authors.

REFERENCES

1. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AMQ, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz EJ, Nibert ML, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, van der Vlugt RA, Varsani A, Zerbini FM. 2017. Consensus statement: virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15:161–168. https://doi.org/10.1038/nrmicro.2016.177.
2. Tisza MJ, Pastrana DV, Welch NL, Stewart B, Peretti A, Starrett GJ, Pang Y-YS, Krishnamurthy SR, Pesavento PA, McDermott DH, Murphy PM, Whited JL, Miller B, Brenchley J, Rosshart SP, Rehermann B, Doorbar J, Ta'ala BA, Pletnikova O, Troncoso JC, Resnick SM, Bolduc B, Sullivan MB, Varsani A, Segall AM, Buck CB. 2020. Discovery of several thousand highly diverse circular DNA viruses. Elife 9:e51971. https://doi.org/10.7554/eLife.51971.
3. Abbas AA, Taylor LJ, Dothard MI, Leiby JS, Fitzgerald AS, Khatib LA, Collman RG, Bushman FD. 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. https://doi.org/10.1016/j.chom.2019.04.001.
4. 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. https://doi.org/10.1016/bs.aivir.2018.10.001.
5. Kazlauskas D, Varsani A, Koonin EV, Krupovic M. 2019. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nat Commun 10:3425. https://doi.org/10.1038/s41467-019-11433-0.
6. Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B. 2013. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol 11:525–538. https://doi.org/10.1038/nrmicro3067.
7. Krupovic M. 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Curr Opin Virol 3:578–586. https://doi.org/10.1016/j.coviro.2013.06.010.
8. 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. https://doi.org/10.1007/s00705-012-1391-y.
9. Ilyina TV, Koonin EV. 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. https://doi.org/10.1093/nar/20.13.3279.
10. Koonin EV, Ilyina TV. 1993. Computer-assisted dissection of rolling circle DNA replication. Biosystems 30:241–268. https://doi.org/10.1016/0303-2647(93)90074-M.
11. Varsani A, Krupovic M. 2018. *Smacoviridae*: a new family of animal-associated single-stranded DNA viruses. Arch Virol 163:2005–2015. https://doi.org/10.1007/s00705-018-3820-z.
12. 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. https://doi.org/10.1038/s41467-018-08167-w.
13. Kazlauskas D, Varsani A, Krupovic M. 2018. Pervasive chimerism in the replication-associated proteins of uncultured single-stranded DNA viruses. Viruses 10:187. https://doi.org/10.3390/v10040187.
14. 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. https://doi.org/10.1016/j.virol.2017.02.001.
15. Nash TE, Dallas MB, Reyes MI, Buhrman GK, Ascencio-Ibañez JT, Hanley-Bowdoin L. 2011. Functional analysis of a novel motif conserved across geminivirus Rep proteins. J Virol 85:1182–1192. https://doi.org/10.1128/JVI.02143-10.
16. 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. https://doi.org/10.1093/ve/vew037.
17. Diemer GS, Stedman KM. 2012. A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. Biol Direct 7:13. https://doi.org/10.1186/1745-6150-7-13.
18. Krupovic M, Ravantti JJ, Bamford DH. 2009. Geminiviruses: a tale of a plasmid becoming a virus. BMC Evol Biol 9:112. https://doi.org/10.1186/1471-2148-9-112.
19. 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. https://doi.org/10.1038/ncomms3700.
20. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH. 2020. Global organization and proposed megataxonomy of the virus world. Microbiol Mol Biol Rev 84:e00061-19. https://doi.org/10.1128/MMBR.00061-19.
21. Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, Gorbalenya AE, Harrach B, Harrison RL, Junglen S, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Nibert M, Rubino L, Sabanadzovic S, Sanfacon H, Simmonds P, Varsani A, Zerbini FM, Davison AJ. 2019. Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). Arch Virol 164:943–946. https://doi.org/10.1007/s00705-018-04136-2.
22. Gorbalenya AE, Krupovic M, Mushegian A, Kropinski AM, Siddell S, Varsani A, Adams MJ, Davison AJ, Dutilh BA, Harrach B, Harrison RL, Junglen S, King AMQ, Knowles NJ, Lefkowitz EJ, Nibert ML, Rubino L, Sabanadzovic S, Sanfaçon H, Simmonds P, Walker PJ, Zerbini FM, Kuhn JH. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat Microbiol, in press.
23. Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olive E, Moriones E, Lett JM, Zerbini FM, Varsani A. 2018. *Alphasatellitidae*: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. Arch Virol 163:2587–2600. https://doi.org/10.1007/s00705-018-3854-2.
24. Guindon S, Dufayard JF, 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. https://doi.org/10.1093/sysbio/syq010.