



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

Mart Krupovic, Arvind Varsani, Darius Kazlauskas, Mya Breitbart, Eric A Delwart, Karyna Rosario, Natalya Yutin, Yuri I Wolf, Balázs Harrach, F. Murilo Zerbini, et al.

## **► To cite this version:**

Mart Krupovic, Arvind Varsani, Darius Kazlauskas, Mya Breitbart, Eric A Delwart, et al.. Cressdnaviricota : a Virus Phylum Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes. *Journal of Virology*, 2020, 94 (12), pp.e00582-20. 10.1128/JVI.00582-20 . pasteur-02861253

**HAL Id: pasteur-02861253**

**<https://pasteur.hal.science/pasteur-02861253>**

Submitted on 8 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

***Cressdnaviricota*: a virus phylum unifying seven families of Rep-encoding viruses with single-stranded, circular DNA genomes**

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

a – Archaeal Virology Unit, Institut Pasteur, 75015 Paris, France

b – The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, 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, Saulėtekio av. 7, Vilnius, 10257, Lithuania

e – University of South Florida, College of Marine Science, 140 7th Avenue South, St. Petersburg, FL, 33701, USA

f – Vitalant Research Institute, San Francisco, CA, USA

g – UCSF Dept. of Laboratory Medicine, University of California-San Francisco, San Francisco, CA, USA

h – National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, 20894, USA

i – Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungária krt. 21, Budapest, 1143, 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

# – Correspondence

E-mail: mart.krupovic@pasteur.fr

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

## Text

Eukaryotic single-stranded (ss) DNA viruses are classified into ten 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, 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 association of smacovirids with methanogenic archaea has also been suggested (12).

Phylogenetic analysis of classified and unclassified CRESS-DNA viruses revealed 13 well supported clades (Figure 1). Six clades correspond to unclassified viruses, denoted CRESSV1 through CRESSV6 (13). The phylogeny splits into two large clades: (1) geminivirids, genomovirids, and CRESSV6 viruses; (2) all other CRESS-DNA viruses (Figure 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. By contrast, Reps of all members of clade 2 contain a conserved arginine finger in the helicase domain, absent in geminivirids and genomovirids, supporting the monophyly of the Reps in clade 2.

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) has 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 have been ratified by the ICTV in March of 2020 and are now part of the official virus taxonomy:

- Clade 1: Families *Geminiviridae* and *Genomoviridae* were included into 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 portmanteaux 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 CRESSV1–5 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 non-homologous capsid proteins, different genome lengths (Table 1) and genome architectures (e.g., nanovirids have multipartite genomes, whereas other viruses have monopartite genomes). Notably, circovirids cluster with CRESSV1–3 viruses, whereas nanovirids

cluster with CRESSV4–5 viruses, justifying order-rank taxa for these assemblages. Thus, five orders have been created for classification of clade 2 viruses:

- *Baphyvirales* (to include *Bacilladnaviridae*): after bacillariophytes (diatoms, hosts of viruses in this taxon), and the suffix -virales;
- *Circlivirales* (to include *Circoviridae* and, in due course, CRESSV1–3 viruses): circo-like, and the suffix -virales;
- *Cretevirales* (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 (Figure 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.

## Funding

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 US 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.

## Figure legend

**Figure 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 closest and farthest leaf nodes. The Rep alignment used for the tree reconstruction was taken from 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.

## References

1. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AM, 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.
2. Tisza MJ, Pastrana DV, Welch NL, Stewart B, Peretti A, Starrett GJ, Pang YS, Krishnamurthy SR, Pesavento PA, McDermott DH, Murphy PM, Whited JL, Miller B, Brenchley J, Rosshart SP, Rehmann 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.
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.
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.
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.
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* 11:525-38.
7. Krupovic M. 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Curr Opin Virol* 3:578-86.
8. Rosario K, Duffy S, Breitbart M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* 157:1851-71.
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 archaeobacteria. *Nucleic Acids Res* 20:3279-85.
10. Koonin EV, Ilyina TV. 1993. Computer-assisted dissection of rolling circle DNA replication. *Biosystems* 30:241-68.
11. Varsani A, Krupovic M. 2018. *Smacoviridae*: a new family of animal-associated single-stranded DNA viruses. *Arch Virol* 163:2005-2015.
12. Díez-Villaseñor C, Rodríguez-Valera F. 2019. CRISPR analysis suggests that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. *Nat Commun* 10:294.
13. Kazlauskas D, Varsani A, Krupovic M. 2018. Pervasive chimerism in the replication-associated proteins of uncultured single-stranded DNA viruses. *Viruses* 10:E187.
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.
15. Nash TE, Dallas MB, Reyes MI, Buhrman GK, Ascencio-Ibanez JT, Hanley-Bowdoin L. 2011. Functional analysis of a novel motif conserved across geminivirus Rep proteins. *J Virol* 85:1182-92.
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.
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.

- 180 18. Krupovic M, Ravantti JJ, Bamford DH. 2009. Geminiviruses: a tale of a plasmid becoming a virus.  
181 BMC Evol Biol 9:112.
- 182 19. Roux S, Enault F, Bronner G, Vaulot D, Forterre P, Krupovic M. 2013. Chimeric viruses blur the  
183 borders between the major groups of eukaryotic single-stranded DNA viruses. Nat Commun  
184 4:2700.
- 185 20. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH. 2020. Global  
186 organization and proposed megataxonomy of the virus world. Microbiol Mol Biol Rev 84:e00061-  
187 19.
- 188 21. Siddell SG, Walker PJ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, Gorbalenya AE, Harrach  
189 B, Harrison RL, Junglen S, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Nibert M, Rubino L,  
190 Sabanadzovic S, Sanfacon H, Simmonds P, Varsani A, Zerbini FM, Davison AJ. 2019. Additional  
191 changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of  
192 Viruses (October 2018). Arch Virol 164:943-946.
- 193 22. Gorbalenya AE, Krupovic M, Mushegian A, Kropinski AM, Siddell S, Varsani A, Adams MJ, Davison  
194 AJ, Dutilh BA, Harrach B, Harrison RL, Junglen S, King AMQ, Knowles NJ, Lefkowitz EJ, Nibert ML,  
195 Rubino L, Sabanadzovic S, Sanfacon H, Simmonds P, Walker PJ, Zerbini FM, Kuhn JH. 2020. The  
196 new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat  
197 Microbiol 5:668-674.
- 198 23. Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olive E, Moriones E, Lett JM,  
199 Zerbini FM, Varsani A. 2018. *Alphasatellitidae*: a new family with two subfamilies for the  
200 classification of geminivirus- and nanovirus-associated alphasatellites. Arch Virol 163:2587-2600.
- 201 24. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and  
202 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.  
203 Syst Biol 59:307-21.

204

205 Table 1. Families of ssDNA viruses associated with eukaryotic hosts. Families unified into the phylum  
 206 *Cressdnaviricota* are in bold.

Family	Host	Virion $\phi$ , nm	Genome topology	Genome size, kb	ORFs	Replication protein <sup>#</sup>
<i>Anelloviridae</i>	animals	30–32	circular	2–4	2–4	Rep
<i>Bidnaviridae</i>	animals (insects)	20–24	linear	6–6.5 per segment (2 segments)	6	pPolB
<i>Parvoviridae</i>	animals	23–28	linear	4–6.3	3–4	HUH-Rep
<b><i>Bacilladnaviridae</i></b>	algae (diatoms)	33–38	circular	5.5–6	4	HUH-Rep
<b><i>Circoviridae</i></b>	animals	15–25	circular	1.7–2.1	2	HUH-Rep
<b><i>Geminiviridae</i></b>	plants	22x38	circular	2.5–3 per segment (1–2 segments)	4–8	HUH-Rep
<b><i>Genomoviridae</i></b>	fungi, insects	20–22	circular	2–2.4	2	HUH-Rep
<b><i>Nanoviridae</i></b>	plants	17–20	circular	0.98–1.1 per segment (6–9 segments)	5–8	HUH-Rep
<b><i>Redondoviridae</i></b>	?	?	circular	3.0–3.1	3	HUH-Rep
<b><i>Smacoviridae</i></b>	?	?	circular	2.3–2.9	2	HUH-Rep

207 <sup>#</sup>HUH-Rep, Rep of the HUH endonuclease superfamily; pPolB, protein-primed family B DNA  
 208 polymerase.