



HAL
open science

Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses

Mart Krupovic, Jonas Blomberg, John Coffin, Indranil Dasgupta, Hung Fan, Andrew Geering, Robert Gifford, Balázs Harrach, Roger Hull, Welkin Johnson, et al.

► **To cite this version:**

Mart Krupovic, Jonas Blomberg, John Coffin, Indranil Dasgupta, Hung Fan, et al.. Ortervirales: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. *Journal of Virology*, 2018, 92 (12), pp.e00515-18. 10.1128/JVI.00515-18 . pasteur-01977336

HAL Id: pasteur-01977336

<https://pasteur.hal.science/pasteur-01977336v1>

Submitted on 10 Jan 2019

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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

1 ***Ortervirales*: A new viral order unifying five families of reverse-transcribing viruses**
2
3 **Mart Krupovic^{1#}, Jonas Blomberg², John M. Coffin³, Indranil Dasgupta⁴, Hung Fan⁵, Andrew**
4 **D. Geering⁶, Robert Gifford⁷, Balázs Harrach⁸, Roger Hull⁹, Welkin Johnson¹⁰, Jan F. Kreuze¹¹,**
5 **Dirk Lindemann¹², Carlos Llorens¹³, Ben Lockhart¹⁴, Jens Mayer¹⁵, Emmanuelle Muller^{16,17}, Neil**
6 **Olszewski¹⁸, Hanu R. Pappu¹⁹, Mikhail Pooggin²⁰, Katja R. Richert-Pöggeler²¹, Sead**
7 **Sabanadzovic²², Hélène Sanfaçon²³, James E. Scholze²⁴, Susan Seal²⁵, Livia Stavolone^{26,27},**
8 **Jonathan P. Stoye²⁸, Pierre-Yves Teycheney^{29,30}, Michael Tristem³¹, Eugene V. Koonin³², Jens H.**
9 **Kuhn³³**

10

11 1 – Department of Microbiology, Institut Pasteur, Paris, France;

12 2 – Department of Medical Sciences, Uppsala University, Uppsala, Sweden;

13 3 – Department of Molecular Biology and Microbiology, Tufts University School of Medicine,

14 Boston, MA, USA;

15 4 – Department of Plant Molecular Biology, University of Delhi, New Delhi, India;

16 5 – Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, USA;

17 6 – Queensland Alliance for Agriculture and Food Innovation, The University of Queensland,

18 Brisbane, Queensland, Australia;

19 7 – MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom;

20 8 – Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy

21 of Sciences, Budapest, Hungary;

22 9 – 3 Portman Drive, Child Okeford, Blandford Forum, Dorset DT11 8HU, United Kingdom;

23 10 – Biology Department, Boston College, Chestnut Hill, MA, USA;

24 11 – Global Program of Integrated Crop and Systems Research, International Potato Center (CIP),

25 Lima, Peru;

26 12 – Institute of Virology, Technische Universität Dresden, Dresden, Germany;

27 13 – Biotechvana, Parc Científic, Universitat de Valencia, Valencia, Spain;

28 14 – Department of Plant Pathology, University of Minnesota, St. Paul, MN, USA;

- 29 15 – Institute of Human Genetics, University of Saarland, Homburg, Germany;
- 30 16 – CIRAD, UMR BGPI, 34398 Montpellier, France;
- 31 17 – BGPI, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France;
- 32 18 – Department of Plant Biology, University of Minnesota, Minneapolis, MN, USA;
- 33 19 – Department of Plant Pathology, Washington State University, Pullman, WA, USA;
- 34 20 – INRA, UMR BGPI, 34398 Montpellier, France;
- 35 21 – Julius Kühn-Institut, Institute for Epidemiology and Pathogen Diagnostics, Braunschweig,
- 36 Germany;
- 37 22 – Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi
- 38 State University, MS, USA;
- 39 23 – Agriculture and Agri-Food Canada, Summerland Research and Development Centre,
- 40 Summerland, BC, Canada;
- 41 24 – Division of Plant Sciences, University of Missouri, Columbia, MO, USA;
- 42 25 – Natural Resources Institute, University of Greenwich, Chatham, Kent, United Kingdom;
- 43 26 – Consiglio Nazionale delle Ricerche, Istituto per la Protezione Sostenibile delle Piante, Bari, Italy;
- 44 27 – International Institute of Tropical Agriculture, Ibadan, Nigeria;
- 45 28 – Department of Medicine, Faculty of Medicine, Imperial College London, London, United
- 46 Kingdom;
- 47 29 – CIRAD, UMR AGAP, 97130 Capesterre Belle eau, Guadeloupe, France;
- 48 30 – AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France;
- 49 31 – Imperial College London, Silwood Park Campus, Ascot, Berkshire, United Kingdom;
- 50 32 – National Center for Biotechnology Information, National Library of Medicine, National Institutes
- 51 of Health, Bethesda, MD, USA;
- 52 33 – Integrated Research Facility at Fort Detrick; National Institute of Allergy and Infectious Diseases,
- 53 National Institutes of Health, Fort Detrick, Frederick, MD, USA.
- 54
- 55 #Corresponding author: Mart Krupovic, krupovic@pasteur.fr
- 56

57 **Text**

58 Reverse-transcribing viruses, which synthesize a copy of genomic DNA from an RNA template, are
59 widespread in animals, plants, algae and fungi (1, 2). This broad distribution suggests ancient origin(s)
60 of these viruses, possibly concomitant with the emergence of eukaryotes (3). Reverse-transcribing
61 viruses include prominent human pathogens, such as human immunodeficiency viruses 1 and 2 (HIV-
62 1/2) and hepatitis B virus, as well as plant pathogens that cause considerable economic losses (4).

63 The International Committee on Taxonomy of Viruses (ICTV) traditionally classified reverse-
64 transcribing viruses into five families: *Caulimoviridae*, *Hepadnaviridae*, *Metaviridae*, *Pseudoviridae*,
65 and *Retroviridae* (5). In 2018, the ICTV recognized an additional family, *Belpaoviridae*, which
66 contains the genus *Semotivirus* (previously included in *Metaviridae* (6)). The infection cycles, nucleic
67 acid types, genome organizations, and virion morphologies of these viruses are very diverse. Indeed,
68 reverse-transcribing viruses are distributed between two Baltimore Classes of viruses. Belpaoviruses,
69 metaviruses, pseudoviruses — better known as Bel/Pao, Ty3/Gypsy, and Ty1/Copia retrotransposons,
70 respectively (1, 7) — and retroviruses typically have single-stranded RNA genomes (Table 1) and
71 frequently integrate into the host genomes as part of their replication cycles (Baltimore Class VI). In
72 contrast, members of the families *Caulimoviridae* and *Hepadnaviridae*, often referred to as
73 “pararetroviruses” (8), encapsidate circular double-stranded DNA genomes and do not actively
74 integrate into host chromosomes (Baltimore Class VII). However, capture of pararetroviral DNA in
75 host genomes, presumably by illegitimate recombination, is commonplace, particularly in plants,
76 giving rise to the corresponding endogenous elements (9, 10).

77 Mechanistic studies on the replication cycles of reverse-transcribing viruses of different
78 families have revealed many similarities that have been reinforced by comparative genomics of the
79 viral reverse transcriptases (RTs), the hallmark enzymes encoded by all reverse-transcribing viruses.
80 Indeed, phylogenetic analyses support the monophyly of all viral RTs, to the exclusion of those
81 encoded by non-viral retroelements from both eukaryotes and prokaryotes (11, 12). In addition to the
82 evidence from the RT phylogeny, belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and
83 retroviruses share several conserved features that hepadnaviruses lack (Table 1). In particular, the
84 polymerase (Pol) polyproteins of belpaoviruses, metaviruses, pseudoviruses, and retroviruses possess

85 similar domain architectures. These Pol polyproteins contain an aspartate protease, which is
86 responsible for the processing of viral polyproteins, and an integrase of the DDE recombinase
87 superfamily. The genomes of these viruses also share long terminal repeats (LTRs) (13). Within
88 certain clades, Pol polyproteins of retroviruses and metaviruses share additional features, such as a
89 dUTPase domain (14-16) and the GPY/F subdomain of the integrase (17, 18). Caulimoviruses also
90 possess a homologous aspartate protease domain in their Pol polyprotein (19), but lack an integrase
91 and LTR. However, RT-based phylogenies consistently place these plant-infecting viruses as a sister
92 clade to the metaviruses (Figure 1), suggesting that among “pararetroviruses”, encapsidation of a DNA
93 genome is a homoplasious character and therefore not a reliable criterion for classification. The basal
94 branches of the RT tree are not resolved and are presented as a multifurcation in Figure 1. This
95 topology is at least compatible with placing the *Hepadnaviridae* clade outside the viral group that
96 includes belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses.

97 Belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses share not only
98 homologous proteins involved in genome replication and polyprotein processing, but also the two
99 principal protein components of the virions, namely, the capsid and nucleocapsid proteins/domains
100 (20-22), although the nucleocapsid domain appears to be absent in spumaretroviruses (family
101 *Retroviridae*; Table 1). By contrast, hepadnaviruses encode an unrelated capsid protein (23). These
102 findings suggest that belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses have
103 evolved from a common viral ancestor, rather than from distinct capsid-less retrotransposons (20).

104 Finally, similarities between belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and
105 retroviruses extend to the mechanism of replication priming. All these viruses utilize host tRNA
106 molecules as primers for genome replication by reverse transcription (24), whereas hepadnaviruses use
107 a specific protein priming mechanism mediated by the polymerase terminal protein domain (25).

108 Taken together, the common complement of proteins required for genome replication,
109 polyprotein processing, and virion formation, the topology of the RT phylogenetic tree, and
110 mechanistic similarities in genome replication present strong evidence that belpaoviruses,
111 caulimoviruses, metaviruses, pseudoviruses, and retroviruses share a common evolutionary origin. The
112 hepadnaviruses, which typically branch out at the base of the viral RT clade (Figure 1), possess a

113 unique capsid protein and employ a distinct replication mechanism, appear to be more distantly related
114 to all these virus families. In recognition of these relationships, the ICTV has recently regrouped the
115 families *Belpaoviridae*, *Caulimoviridae*, *Metaviridae*, *Pseudoviridae* and *Retroviridae* into an order
116 *Ortervirales* (*orter*: an inversion of *retro*, which was derived from reverse transcription; *virales*: suffix
117 for an order). This change in taxonomy acknowledges and formalizes the long-proposed evolutionary
118 relationship among most groups of reverse-transcribing viruses (26). We note that although
119 hepadnaviruses are not included in the order, they might be unified with other reverse-transcribing
120 viruses at a higher taxonomic level in the future.

121

122

123 **Acknowledgments**

124 This work was supported in part through Battelle Memorial Institute's prime contract with the US
125 National Institute of Allergy and Infectious Diseases (NIAID) under Contract No.
126 HHSN272200700016I (JHK). EVK is supported by intramural funds of the US Department of Health
127 and Human Services (to the National Library of Medicine). SS acknowledges support from SRI Funds
128 from Mississippi Agriculture and Forestry Experiment Station of Mississippi State University.

129 M. Krupovic, B. Harrach, S. Sabanadzovic, H. Sanfaçon, and J. H. Kuhn are members of the
130 2014–2017 International Committee on Taxonomy of Viruses (ICTV) Executive Committee. J.
131 Blomberg, J. M. Coffin, H. Fan, R. Gifford, W. Johnson, D. Lindemann, J. Mayer, J. P. Stoye, and M.
132 Tristem are members of the 2014–2017 ICTV *Retroviridae* Study Group. I. Dasgupta, A. D. Geering,
133 R. Hull, J. F. Kreuze, B. Lockhart, E. Muller, N. Olszewski, H. R. Pappu, M. Pooggin, K. R. Richert-
134 Pöggeler, J. E. Schoelz, S. Seal, L. Stavalone, and P.-Y. Teycheney are members of the 2014–2017
135 ICTV *Caulimoviridae* Study Group. R. Hull is retired from the John Innes Centre, Norwich, Norfolk,
136 United Kingdom.

137

138 References

- 139 1. **Llorens, C., R. Futami, L. Covelli, L. Dominguez-Escriba, J. M. Viu, D. Tamarit, J. Aguilar-Rodriguez, M. Vicente-**
140 **Ripolles, G. Fuster, G. P. Bernet, F. Maumus, A. Munoz-Pomer, J. M. Sempere, A. Latorre, and A. Moya.** 2011. The
141 Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res* **39**:D70-4.
- 142 2. **Ahlquist, P.** 2006. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA
143 viruses. *Nat Rev Microbiol* **4**:371-82.
- 144 3. **Koonin, E. V., V. V. Dolja, and M. Krupovic.** 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity.
145 *Virology* **479-480**:2-25.
- 146 4. **Geering, A. D. W.** 2014. *Caulimoviridae* (plant pararetroviruses), eLS. John Wiley & Sons, Ltd.
- 147 5. **King, A. M. Q., M. J. Adams, E. B. Carstens, and E. J. Lefkowitz.** 2011. Virus taxonomy: Ninth report of the International
148 Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego.
- 149 6. **Eickbush, T., J. D. Boeke, S. B. Sandmeyer, and D. F. Voytas.** 2011. *Metaviridae*, p. 457-466. In A. M. Q. King, M. J.
150 Adams, E. B. Carstens, and E. J. Lefkowitz (ed.), *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report*
151 *of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego.
- 152 7. **Arkhipova, I. R.** 2017. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand
153 their complex evolutionary histories. *Mob DNA* **8**:19.
- 154 8. **Hull, R., and H. Will.** 1989. Molecular biology of viral and nonviral retroelements. *Trends Genet* **5**:357-9.
- 155 9. **Feschotte, C., and C. Gilbert.** 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev*
156 *Genet* **13**:283-96.
- 157 10. **Diop, S. I., A. D. W. Geering, F. Alfama-Depauw, M. Loaec, P. Y. Teycheney, and F. Maumus.** 2018. Tracheophyte
158 genomes keep track of the deep evolution of the *Caulimoviridae*. *Sci Rep* **8**:572.
- 159 11. **Gladyshev, E. A., and I. R. Arkhipova.** 2011. A widespread class of reverse transcriptase-related cellular genes. *Proc Natl*
160 *Acad Sci U S A* **108**:20311-6.
- 161 12. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences.
162 *Embo J* **9**:3353-62.
- 163 13. **Benachenhou, F., G. O. Sperber, E. Bongcam-Rudloff, G. Andersson, J. D. Boeke, and J. Blomberg.** 2013. Conserved
164 structure and inferred evolutionary history of long terminal repeats (LTRs). *Mob DNA* **4**:5.
- 165 14. **Mayer, J., and E. U. Meese.** 2003. Presence of dUTPase in the various human endogenous retrovirus K (HERV-K) families. *J*
166 *Mol Evol* **57**:642-9.
- 167 15. **Rodriguez, F., A. W. Kenefick, and I. R. Arkhipova.** 2017. LTR-retrotransposons from bdelloid rotifers capture additional
168 ORFs shared between highly diverse retroelement types. *Viruses* **9**:E78.
- 169 16. **Novikova, O. S., and A. G. Blinov.** 2008. dUTPase-containing metaviridae LTR retrotransposons from the genome of
170 *Phanerochaete chrysosporium* (Fungi: Basidiomycota). *Dokl Biochem Biophys* **420**:146-9.

- 171 17. **Jern, P., G. O. Sperber, and J. Blomberg.** 2005. Use of endogenous retroviral sequences (ERVs) and structural markers for
172 retroviral phylogenetic inference and taxonomy. *Retrovirology* **2**:50.
- 173 18. **Malik, H. S., and T. H. Eickbush.** 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR
174 retrotransposons. *J Virol* **73**:5186-90.
- 175 19. **Marmey, P., A. Rojas-Mendoza, A. de Kochko, R. N. Beachy, and C. M. Fauquet.** 2005. Characterization of the protease
176 domain of Rice tungro bacilliform virus responsible for the processing of the capsid protein from the polyprotein. *Viol J* **2**:33.
- 177 20. **Krupovic, M., and E. V. Koonin.** 2017. Homologous capsid proteins testify to the common ancestry of retroviruses,
178 caulimoviruses, pseudoviruses, and metaviruses. *J Virol* **91**:e00210-17.
- 179 21. **Vo, J. N., P. R. Campbell, N. N. Mahfuz, R. Ramli, D. Pagendam, R. Barnard, and A. D. Geering.** 2016. Characterization
180 of the banana streak virus capsid protein and mapping of the immunodominant continuous B-cell epitopes to the surface-
181 exposed N terminus. *J Gen Virol* **97**:3446-3457.
- 182 22. **Sandmeyer, S., K. Patterson, and V. Bilanchone.** 2015. Ty3, a position-specific retrotransposon in budding yeast. *Microbiol*
183 *Spectr* **3**:MDNA3-0057-2014.
- 184 23. **Steven, A. C., J. F. Conway, N. Cheng, N. R. Watts, D. M. Belnap, A. Harris, S. J. Stahl, and P. T. Wingfield.** 2005.
185 Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv Virus Res* **64**:125-64.
- 186 24. **Menendez-Arias, L., A. Sebastian-Martin, and M. Alvarez.** 2017. Viral reverse transcriptases. *Virus Res* **234**:153-176.
- 187 25. **Nassal, M.** 2008. Hepatitis B viruses: reverse transcription a different way. *Virus Res* **134**:235-49.
- 188 26. **Hull, R.** 2001. Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. International Committee on
189 Taxonomy of Viruses. *Arch Virol* **146**:2255-61.
- 190 27. **Richert-Poggeler, K. R., F. Noreen, T. Schwarzacher, G. Harper, and T. Hohn.** 2003. Induction of infectious petunia vein
191 clearing (pararetro) virus from endogenous provirus in petunia. *Embo J* **22**:4836-45.
- 192 28. **Richert-Poggeler, K. R., and R. J. Shepherd.** 1997. Petunia vein-clearing virus: a plant pararetrovirus with the core
193 sequences for an integrase function. *Virology* **236**:137-46.
- 194 29. **Miller, R. H., and W. S. Robinson.** 1986. Common evolutionary origin of hepatitis B virus and retroviruses. *Proc Natl Acad*
195 *Sci U S A* **83**:2531-5.
- 196 30. **Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel.** 2010. New algorithms and methods to
197 estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**:307-21.
- 198
199
200

201 **Figure legend**

202 **Figure 1.** Maximum likelihood phylogeny of viral reverse transcriptases. The tree includes sequences of 290 viruses belonging to all ICTV-
203 recognized genera of reverse-transcribing viruses. The phylogeny was inferred using PhyML (30) with the LG+G+F substitution model and is
204 rooted with sequences from non-viral retroelements (bacterial group II introns and eukaryotic LINE retroelements). Genomic organizations of
205 selected representatives of reverse-transcribing viruses are shown next to the corresponding branches. Long terminal repeats (LTR) are shown as
206 black triangles. Note that members of the virus families display considerable variation in gene/domain content (5), which is not captured in this
207 figure. Abbreviations: 6, 6-kDa protein; ATF, aphid transmission factor; CA/CP, capsid protein; CHR, chromodomain (only present in the INT of
208 particular clades of metaviruses of plants, fungi and several vertebrates); *gag*, group-specific antigen; *env*, envelope genes; SU, surface
209 glycoprotein; TM, transmembrane glycoprotein; INT, integrase; MA, matrix protein; NC, MP, movement protein; nucleocapsid; *nef*, *tat*, *rev*, *vif*,
210 *vpr*, and *vpu*, genes that express regulatory proteins via spliced mRNAs; TP, terminal protein domain; TT/SR, translation trans-activator/suppressor
211 of RNA interference; P, polymerase; *pol*, polymerase gene; PR, protease; PreS, pre-surface protein (envelope); PX/TA, protein X/transcription
212 activator; RH, RNase H; RT, reverse transcriptase; VAP, virion-associated protein.

Table 1. Features shared by reverse-transcribing viruses.

| Family | <i>Retroviridae</i> | | <i>Metaviridae</i> | <i>Pseudoviridae</i> | <i>Belpaoviridae</i> | <i>Caulimoviridae</i> | <i>Hepadnaviridae</i> |
|-------------|--------------------------|--------------------------|--------------------|----------------------|----------------------|-----------------------|-----------------------|
| Subfamily | <i>Orthoretrovirinae</i> | <i>Spumaretrovirinae</i> | | | | | |
| Pol | RT-RH | + | + | + | + | + | + |
| | Protease | + | + | + | + | + | - |
| | Integrase | + | + | + | + | - | - |
| Gag | CA/CP | + | + | + | + | + | - |
| | NC | + | - | + | + | + | - |
| LTR | + | + | + | + | + | -\$ | -# |
| Priming | tRNA | tRNA | tRNA | tRNA | tRNA | tRNA | TP |
| Genome type | ssRNA | ssRNA/dsDNA* | ssRNA | ssRNA | ssRNA | dsDNA | dsDNA |

* – Members of the subfamily *Spumaretrovirinae* contain both ssRNA and dsDNA in extracellular particles and reverse transcription occurs during virus assembly and disassembly.

\$ – In the genus *Petuvirus* (*Caulimoviridae*) an inactivated integrase-like domain and quasi (long) terminal repeats have been identified (27, 28), suggesting that certain ancestral elements have been lost during the evolution of caulimoviruses.

– Upstream of the capsid protein gene, hepadnavirus genomes contain a sequence showing similarity to the U5 region of the retroviral LTR (29).

Abbreviations: CA/CP, capsid protein; Gag, group-specific antigen; LTR, long terminal repeats; NC, nucleocapsid protein; RH, RNase H; RT, reverse transcriptase; Pol, polymerase polyprotein; TP, terminal protein.

