JVI Accepted Manuscript Posted Online 4 April 2018
J. Virol. doi:10.1128/JVI.00515-18

***Ortervirales*: A new viral order unifying five families of reverse-transcribing viruses**

Mart Krupovic[1]#, Jonas Blomberg[2], John M. Coffin[3], Indranil Dasgupta[4], Hung Fan[5], Andrew D. Geering[6], Robert Gifford[7], Balázs Harrach[8], Roger Hull[9], Welkin Johnson[10], Jan F. Kreuze[11], Dirk Lindemann[12], Carlos Llorens[13], Ben Lockhart[14], Jens Mayer[15], Emmanuelle Muller[16,17], Neil Olszewski[18], Hanu R. Pappu[19], Mikhail Pooggin[20], Katja R. Richert-Pöggeler[21], Sead Sabanadzovic[22], Hélène Sanfaçon[23], James E. Schoelz[24], Susan Seal[25], Livia Stavolone[26,27], Jonathan P. Stoye[28], Pierre-Yves Teycheney[29,30], Michael Tristem[31], Eugene V. Koonin[32], Jens H. Kuhn[33]

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

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

3 – Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA, USA;

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

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

6 – Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Queensland, Australia;

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

8 – Institute for Veterinary Medical Research, Centre for Agricultural Research, Hungarian Academy of Sciences, Budapest, Hungary;

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

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

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

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

13 – Biotechvana, Parc Cientific, Universitat de Valencia, Valencia, Spain;

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

15 – Institute of Human Genetics, University of Saarland, Homburg, Germany;

16 – CIRAD, UMR BGPI, 34398 Montpellier, France;

17 – BGPI, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France;

18 – Department of Plant Biology, University of Minnesota, Minneapolis, MN, USA;

19 – Department of Plant Pathology, Washington State University, Pullman, WA, USA;

20 – INRA, UMR BGPI, 34398 Montpellier, France;

21 – Julius Kühn-Institut, Institute for Epidemiology and Pathogen Diagnostics, Braunschweig, Germany;

22 – Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, MS, USA;

23 – Agriculture and Agri-Food Canada, Summerland Research and Development Centre, Summerland, BC, Canada;

24 – Division of Plant Sciences, University of Missouri, Columbia, MO, USA;

25 – Natural Resources Institute, University of Greenwich, Chatham, Kent, United Kingdom;

26 – Consiglio Nazionale delle Ricerche, Istituto per la Protezione Sostenibile delle Piante, Bari, Italy;

27 – International Institute of Tropical Agriculture, Ibadan, Nigeria;

28 – Department of Medicine, Faculty of Medicine, Imperial College London, London, United Kingdom;

29 – CIRAD, UMR AGAP, 97130 Capesterre Belle eau, Guadeloupe, France;

30 – AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France;

31 – Imperial College London, Silwood Park Campus, Ascot, Berkshire, United Kingdom;

32 – National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA;

33 – Integrated Research Facility at Fort Detrick; National Institute of Allergy and Infectious Diseases, National Institutes of Health, Fort Detrick, Frederick, MD, USA.

#Corresponding author: Mart Krupovic, krupovic@pasteur.fr

Text

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

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

Mechanistic studies on the replication cycles of reverse-transcribing viruses of different families have revealed many similarities that have been reinforced by comparative genomics of the viral reverse transcriptases (RTs), the hallmark enzymes encoded by all reverse-transcribing viruses. Indeed, phylogenetic analyses support the monophyly of all viral RTs, to the exclusion of those encoded by non-viral retroelements from both eukaryotes and prokaryotes (11, 12). In addition to the evidence from the RT phylogeny, belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses share several conserved features that hepadnaviruses lack (Table 1). In particular, the polymerase (Pol) polyproteins of belpaoviruses, metaviruses, pseudoviruses, and retroviruses possess similar domain architectures. These Pol polyproteins contain an aspartate protease, which is responsible for the processing of viral polyproteins, and an integrase of the DDE recombinase superfamily. The genomes of these viruses also share long terminal repeats (LTRs) (13). Within certain clades, Pol polyproteins of retroviruses and metaviruses share additional features, such as a dUTPase domain (14-16) and the GPY/F subdomain of the integrase (17, 18). Caulimoviruses also possess a homologous aspartate protease domain in their Pol polyprotein (19), but lack an integrase and LTR. However, RT-based phylogenies consistently place these plant-infecting viruses as a sister clade to the metaviruses (Figure 1), suggesting that among "pararetroviruses", encapsidation of a DNA genome is a homoplasious character and therefore not a reliable criterion for classification. The basal branches of the RT tree are not resolved and are presented as a multifurcation in Figure 1. This topology is at least compatible with placing the *Hepadnaviridae* clade outside the viral group that includes belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses.

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

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

Taken together, the common complement of proteins required for genome replication, polyprotein processing, and virion formation, the topology of the RT phylogenetic tree, and mechanistic similarities in genome replication present strong evidence that belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses share a common evolutionary origin. The hepadnaviruses, which typically branch out at the base of the viral RT clade (Figure 1), possess a unique capsid protein and employ a distinct replication mechanism, appear to be more distantly related to all these virus families. In recognition of these relationships, the ICTV has recently regrouped the families *Belpaoviridae*, *Caulimoviridae*, *Metaviridae*, *Pseudoviridae* and *Retroviridae* into an order *Ortervirales* (*orter*: an inversion of *retro*, which was derived from reverse transcription; *virales*: suffix for an order). This change in taxonomy acknowledges and formalizes the long-proposed evolutionary relationship among most groups of reverse-transcribing viruses (26). We note that although hepadnaviruses are not included in the order, they might be unified with other reverse-transcribing viruses at a higher taxonomic level in the future.

Acknowledgments

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

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

References

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

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

Figure legend

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

Table 1. Features shared by reverse-transcribing viruses.

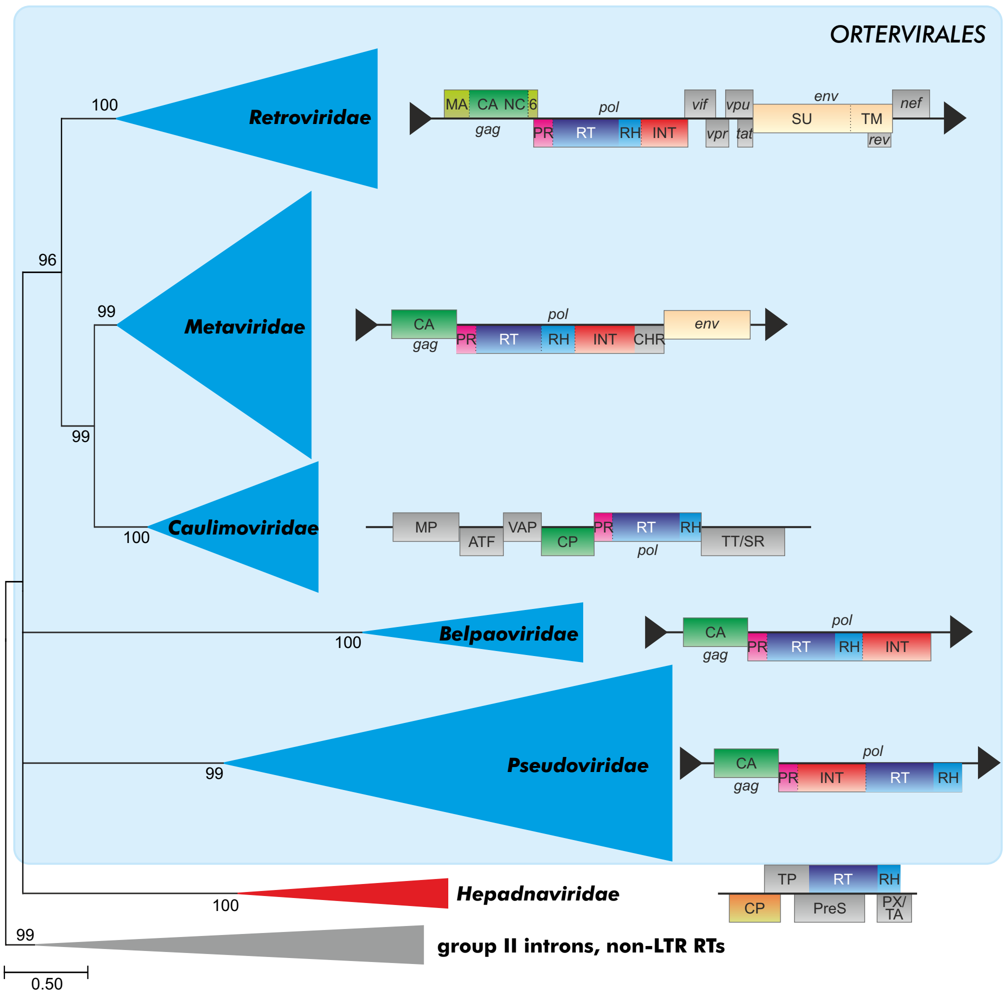
Family		*Retroviridae*		*Metaviridae*	*Pseudoviridae*	*Belpaoviridae*	*Caulimoviridae*	*Hepadnaviridae*
Subfamily		*Orthoretrovirinae*	*Spumaretrovirinae*					
Pol	RT-RH	+	+	+	+	+	+	+
	Protease	+	+	+	+	+	+	−
	Integrase	+	+	+	+	+	−	−
Gag	CA/CP	+	+	+	+	+	+	−
	NC	+	−	+	+	+	+	−
LTR		+	+	+	+	+	−$	−#
Priming		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.


ORTERVIRALES
100
Retroviridae
MA
CA
NC
6
gag
PR
RT
RH
INT
pol
vif
vpr
vpu
tat
env
SU
TM
rev
nef
96
99
Metaviridae
CA
gag
PR
RT
RH
INT
CHR
pol
env
99
100
Caulimoviridae
MP
ATF
VAP
CP
PR
RT
RH
pol
TT/SR
100
Belpaoviridae
CA
gag
PR
RT
RH
INT
pol
99
Pseudoviridae
CA
gag
PR
INT
RT
RH
pol
100
Hepadnaviridae
TP
RT
RH
CP
PreS
PX/
TA
99
group II introns, non-LTR RTs
0.50