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

Mart Krupovic1,5,#, Jonas Blomberg2,^, John M. Coffin3,^, Indranil Dasgupta4,^, Hung Fan5,^,
Andrew D. Geering6,^, Robert Gifford7,^, Balázs Harrach8,^, Roger Hull9,^, Welkin Johnson10,^,
Jan F. Kreuze11,^, Dirk Lindemann12,^, Carlos Llorens13,^, Ben Lockhart14,^, Jens Mayer15,^,
Emmanuelle Muller16,17,^, Neil Olszewski18,^, Hanu R. Pappu19,^, Mikhail Pooggin20,^, Katja R.
Richert-Pöggeler21,^, Sead Sabanadzovic22,5, Hélène Sanfaçon23,^, James E. Schoelz24,^, Susan
Seal25,^, Livia Stavolone26,27,^, Jonathan P. Stoye28,^, Pierre-Yves Teycheney29,30,^, Michael
Tristem31,^, Eugene V. Koonin31, Jens H. Kuhn33,5

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;
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, Germany;
36 22 – Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, MS, USA;
37 23 – Agriculture and Agri-Food Canada, Summerland Research and Development Centre, Summerland, BC, Canada;
38 24 – Division of Plant Sciences, University of Missouri, Columbia, MO, USA;
39 25 – Natural Resources Institute, University of Greenwich, Chatham, Kent, United Kingdom;
40 26 – Consiglio Nazionale delle Ricerche, Istituto per la Protezione Sostenibile delle Piante, Bari, Italy;
41 27 – International Institute of Tropical Agriculture, Ibadan, Nigeria;
42 28 – Department of Medicine, Faculty of Medicine, Imperial College London, London, United Kingdom;
43 29 – CIRAD, UMR AGAP, 97130 Capesterre Belle eau, Guadeloupe, France;
44 30 – AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France;
45 31 – Imperial College London, Silwood Park Campus, Ascot, Berkshire, United Kingdom;
46 32 – National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA;
47 33 – Integrated Research Facility at Fort Detrick; National Institute of Allergy and Infectious Diseases, National Institutes of Health, Fort Detrick, Frederick, MD, USA.
48 34 –
49 35 – Correspondence
50 E-mail: 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.

Funding

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.
Table 1. Features shared by reverse-transcribing viruses.

<table>
<thead>
<tr>
<th>Family</th>
<th>Retroviridae</th>
<th>Metaviridae</th>
<th>Pseudoviridae</th>
<th>Belpaoviridae</th>
<th>Caulimoviridae</th>
<th>Hepadnaviridae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subfamily</td>
<td>Orthoretrovirinae</td>
<td>Spumaretrovirinae</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pol</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Protease</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Integrase</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Ga</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>g</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>NC</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>LTR</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Priming</td>
<td>ssRNA</td>
<td>ssRNA/dsDNA*</td>
<td>ssRNA</td>
<td>ssRNA</td>
<td>ssRNA</td>
<td>dsDNA</td>
</tr>
<tr>
<td>Genome type</td>
<td>ssRNA</td>
<td>ssRNA/dsDNA*</td>
<td>ssRNA</td>
<td>ssRNA</td>
<td>ssRNA</td>
<td>dsDNA</td>
</tr>
</tbody>
</table>

$ –$ 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.
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-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.
References


Malik, H. S., and T. H. Eickbush. 1999. Modular evolution of the integrase domain in

2005. Characterization of the protease domain of Rice tungro bacilliform virus

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.

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

Sandmeyer, S., K. Patterson, and V. Bilanchone. 2015. Ty3, a position-specific

Steven, A. C., J. F. Conway, N. Cheng, N. R. Watts, D. M. Belnap, A. Harris, S. J.


134:235-49.

Hull, R. 2001. Classifying reverse transcribing elements: a proposal and a challenge to

Induction of infectious petunia vein clearing (pararetro) virus from endogenous provirus

pararetrovirus with the core sequences for an integrase function. Virology 236:137-46.


2010. New algorithms and methods to estimate maximum-likelihood phylogenies:
group II introns, non-LTR RTs