



This form should be used for all taxonomic proposals. Please complete all those modules that are applicable.

For guidance, see the notes written in blue and the separate document "Help with completing a taxonomic proposal"

Please try to keep related proposals within a single document.

Part 1: **TITLE, AUTHORS, etc**

| | | |
|---|---------------------------------------|--|
| Code assigned: | 2017.013D | (to be completed by ICTV officers) |
| Short title: creation of new order, <i>Ortervirales</i> , for 5 families of reverse-transcribing viruses | | |
| Modules attached (Modules 1, 4 and either 2 or 3 are required.) | 1 <input checked="" type="checkbox"/> | 2 <input checked="" type="checkbox"/> 3 <input type="checkbox"/> 4 <input checked="" type="checkbox"/> |

Author(s):

Mart Krupovic,
Balázs Harrach,
Sead Sabanadzovic,
Hélène Sanfaçon,
Eugene V. Koonin,
Jens H. Kuhn

Retroviridae SG

Welkin Johnson,
Jonas Blomberg,
John Coffin,
Hung Fan,
Robert Gifford,
Dirk Lindemann,
Jens Mayer,
Jonathan Stoye,
Michael Tristem

Caulimoviridae SG

Pierre-Yves Teycheney,
Idranil Dasgupta,
Andrew Geering,
Roger Hull,
Jan F. Kreuze,
Ben Lockhart,
Emmanuelle Muller,
Neil Olszewski,
Hanu Pappu,
Mikhail Pooggin,
Katja Richert-Pöggeler,
James E. Schoelz,
Susan Seal,
Livia Stabolone

Corresponding author with e-mail address:

Mart Krupovic; E-mail: krupovic@pasteur.fr

List the ICTV study group(s) that have seen this proposal:

A list of study groups and contacts is provided at <http://www.ictvonline.org/subcommittees.asp> . If in doubt, contact the appropriate subcommittee chair (there are six virus subcommittees: animal DNA and retroviruses, animal ssRNA-, animal ssRNA+, fungal and protist, plant, bacterial and archaeal)

ICTV Retroviridae and Caulimoviridae Study Groups as well as Chairs of the Plant Viruses Subcommittee, Animal DNA Viruses and Retroviruses Subcommittee, as well as Fungal and Protist Viruses Subcommittee

ICTV Study Group comments (if any) and response of the proposer:

Date first submitted to ICTV: June 8, 2017
Date of this revision (if different to above): June 22, 2017

ICTV-EC comments and response of the proposer:**Part 2: PROPOSED TAXONOMY**

Present the proposed new taxonomy on accompanying spreadsheet

Name of accompanying spreadsheet: 2017.013D.N.v1.Ortervirales

Please display the taxonomic changes you are proposing on the accompanying spreadsheet module 2017_TP_Template_Excel_module. Submit both this and the spreadsheet to the appropriate ICTV Subcommittee Chair.

Part 4: APPENDIX: supporting material

Reverse-transcribing viruses are currently classified into five families: *Caulimoviridae*, *Hepadnaviridae*, *Metaviridae*, *Pseudoviridae*, and *Retroviridae* (<https://talk.ictvonline.org/taxonomy/>). In addition, a taxonomic relocation of the genus *Semotivirus* from the family *Metaviridae* to a new family “Belpaoviridae” has been proposed in 2017 (submitted TaxoProp 2017.001D.N.v1.Belpaoviridae). The only protein shared by viruses from all these families is the reverse-transcriptase (RT) including an RNase H (RH) domain (Table 1). Phylogenetic analyses support the monophyly of these viral RTs (Xiong and Eickbush, 1990; Gladyshev and Arkhipova, 2011), to the exclusion of those encoded by non-viral retroelements from both eukaryotes and prokaryotes (Figure 1). However, besides RTs, members of the families “Belpaoviridae”, *Caulimoviridae*, *Metaviridae*, *Pseudoviridae*, and *Retroviridae* share several conserved features that hepadnaviruses lack (Table 1).

Table 1. Features shared by reverse-transcribing viruses.

| | Retroviruses | Metaviruses | Pseudoviruses | Belpaoviruses | Caulimoviruses | Hepadnaviruses |
|-----------|--------------|-------------|---------------|---------------|----------------|----------------|
| RT-RH | + | + | + | + | + | + |
| Pol | + | + | + | + | + | - |
| Protease | + | + | + | + | + | - |
| Integrase | + | + | + | + | - | - |
| Gag | + | + | + | + | + | - |
| CA/CP | + | + | + | + | + | - |
| NC | +* | + | + | + | + | - |
| LTR | + | + | + | + | - | - |
| Priming | tRNA | tRNA | tRNA | tRNA | tRNA | TP |

* – members of the subfamily *Spumaretrovirinae* do not contain the canonical NC domain within their Gag polyproteins. Abbreviations: RT, reverse transcriptase; RH, RNase H; CA/CP, capsid protein; NC, nucleocapsid protein; LTR, long terminal repeats; Gag, group-specific antigen; Pol, polymerase polyprotein; TP, terminal protein.

In particular, the Pol polyproteins of belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses possess similar domain architectures. These Pol polyproteins encode an aspartate protease, which is responsible for the processing of viral polyproteins. Furthermore, belpaoviruses, metaviruses, pseudoviruses, and retroviruses integrate into host-cell chromosomes as part of their life cycles, share long terminal repeats (LTR), and encode homologous integrases of the DDE recombinase superfamily, which are expressed as part of the corresponding Pol polyproteins. Although members of the *Caulimoviridae* lack an integrase, in RT-based phylogenies, they consistently form a sister clade to metaviruses, suggesting that the integrase has been lost in the caulimovirus branch (Figures 1 and 2). In agreement with this hypothesis, petunia vein clearing virus (genus *Petuvirus*, family *Caulimoviridae*), which in RT-based phylogenies occupies a basal position within the *Caulimoviridae* clade, contains sequence motifs resembling those of retroviral integrases (Richert-Pöggeler and Shepherd, 1997), although no evidence for integrase activity of the corresponding protein domain has been presented. We note that the basal branches of the RT tree are not resolved and are presented as a multifurcation in Figure 2. This topology is at least compatible with the *Hepadnaviridae* clade being outside of the viral group including 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 viral particles, namely the capsid and nucleocapsid proteins/domains (Figure 3, Table 1) (Vo et al., 2016; Krupovic and Koonin, 2017). By contrast, hepadnaviruses encode an unrelated capsid protein (Steven et al., 2005). These findings indicate that belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses have evolved from a common viral ancestor, rather than from distinct capsid-less retrotransposons (Krupovic and Koonin, 2017).

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 (Menéndez-Arias et al., 2017), whereas hepadnaviruses use a specific protein priming mechanism mediated by the terminal protein (TP) domain of the viral RT (Nassal, 2008).

The common complement of proteins required for genome replication, polyprotein processing, and virion formation, the topology of the phylogenetic tree of the RTs, as well as mechanistic similarities in genome replication, strongly suggest that belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses share a common evolutionary origin. By contrast, hepadnaviruses that typically branch out at the base of the viral RT clade (Figures 1 and 2) appear to be more distantly related. Thus, we propose to include the families *Caulimoviridae*, *Metaviridae*, *Pseudoviridae*, *Retroviridae* and the putative family “Belpaoviridae”, into an order to be named *Ortervirales* (*orter*: an inversion of *retro*, which stands for reverse transcription; *virales*: suffix for an order).

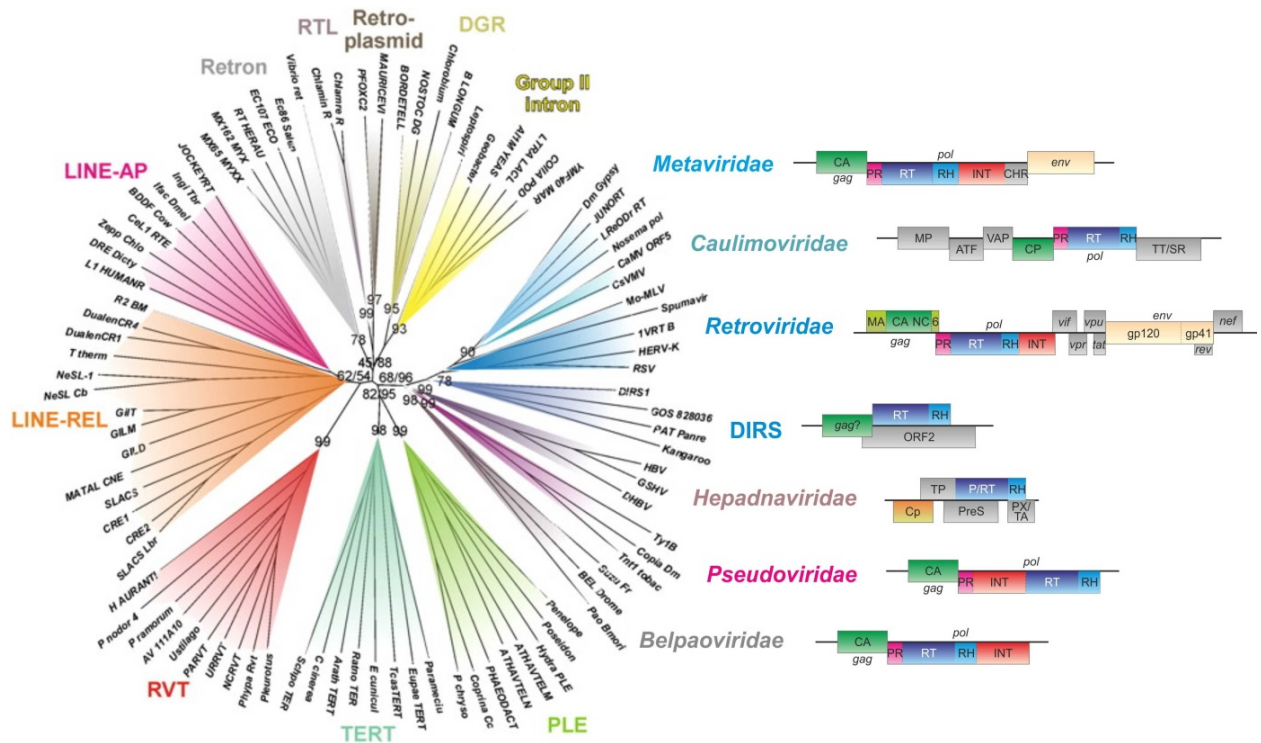


Figure 1. Genomic organizations of selected representatives of reverse-transcribing viruses overlay the phylogenetic tree of reverse transcriptases (RTs). A phylogram indicating both minimum evolution (ME) and maximum-likelihood (ML) support values for the most basal branches and ME support for each colored clade in cases where it exceeds 70%. The tree is reproduced from (Gladyshev and Arkhipova, 2011). Abbreviations: DGR, diversity-generating retroelements; LINE, long interspersed nucleotide elements; *gag*, group-specific antigen gene; *env*, envelope gene; *pol*, polymerase gene; PR, aspartate protease; RT, reverse transcriptase; RH, RNase H; INT, integrase; CHR, chromodomain. The sites of Pol processing by PR are shown as vertical dashed lines. MA, matrix protein; CA/Cp, capsid protein; NC, nucleocapsid; 6, 6-kDa protein; *vif*, *vpr*, *vpu*, *tat*, *rev*, and *nef*, genes that express regulatory proteins via spliced mRNAs; gp120 and gp41, 120- (surface) and 41-kDa (transmembrane) glycoproteins; ATF, aphid transmission factor; VAP, virion-associated protein; TT/SR, translation trans-activator/suppressor of RNA interference; TP, terminal protein; P, polymerase; PreS, pre-surface protein (envelope); PX/TA, protein X/transcription activator; RVT, RT-related cellular genes; TERT, telomerase reverse transcriptase; PLE, Penelope-like retroelements.

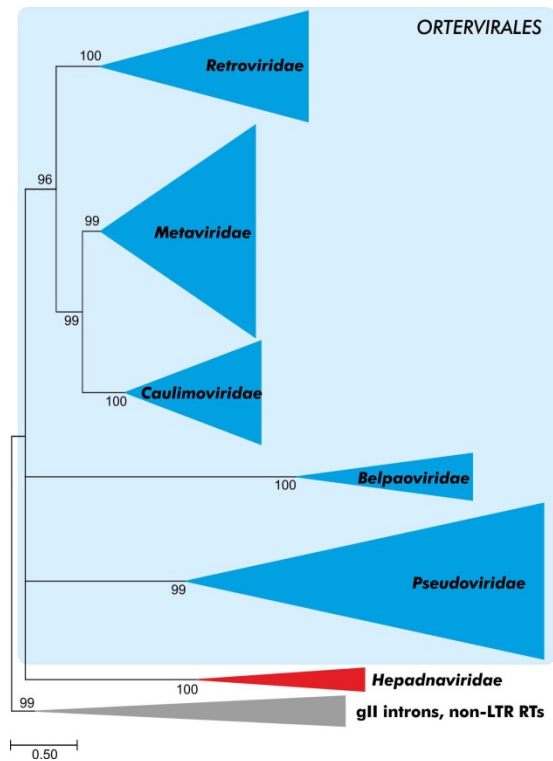


Figure 2. Maximum likelihood phylogeny of viral RTs. The tree includes sequences of 290 taxa representing all ICTV-recognized genera of RT viruses. The phylogeny was inferred using PhyML (Guindon et al., 2010) with the LG+G+F substitution model and is rooted with sequences from non-viral retroelements (bacterial group II introns and eukaryotic LINE retroelements).

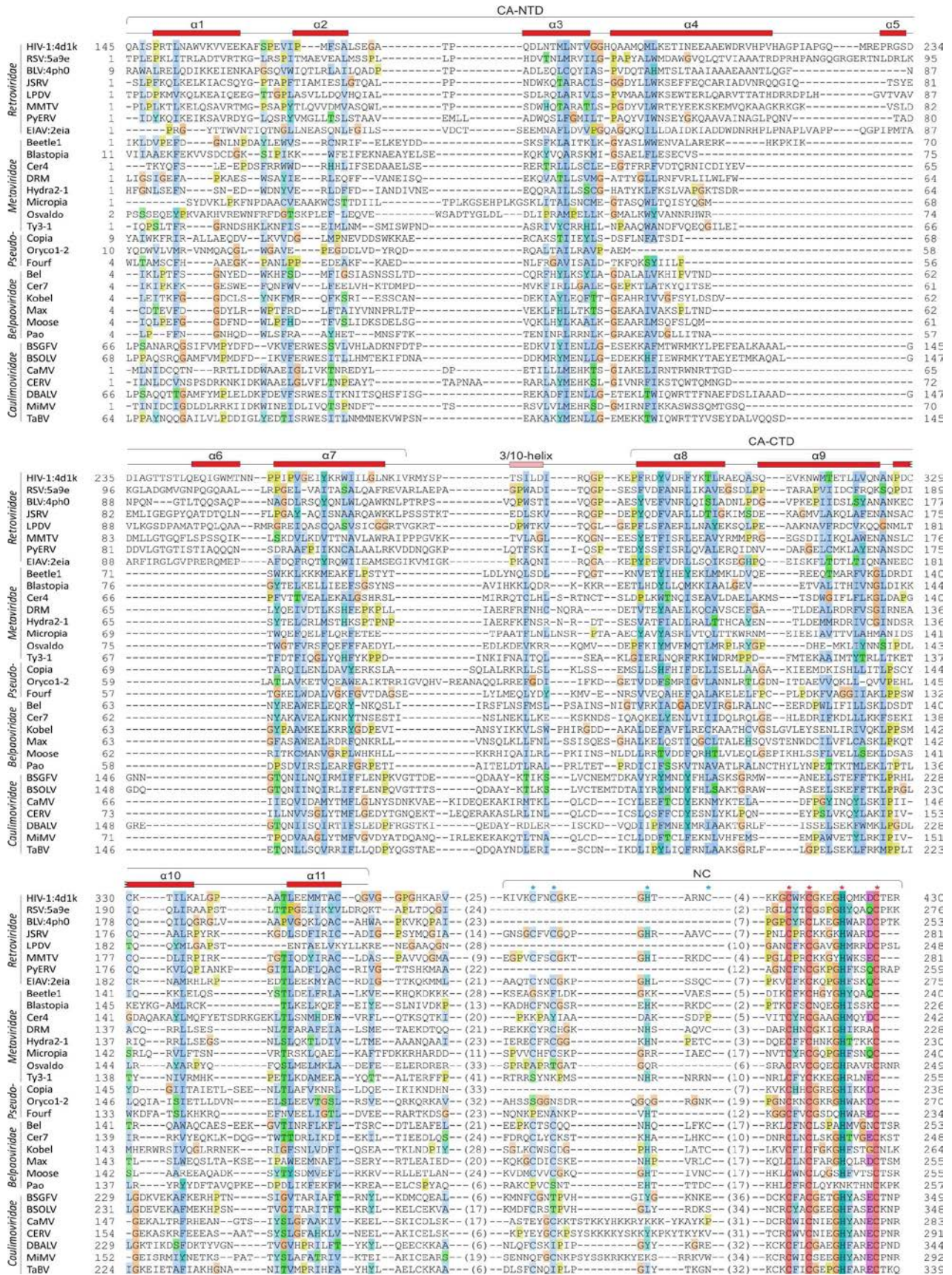


Figure 3. Multiple-sequence alignment of CAs/CPs of reverse-transcribing viruses belonging to the families “Belpaoviridae”, *Caulimoviridae*, *Metaviridae*, *Pseudoviridae*, and *Retroviridae*. The figure is modified from Krupovic and Koonin (2017). Secondary-structure elements above the alignment are indicated for Rous sarcoma virus CA (PDB accession number 5A9E). Red and blue asterisks indicate the conserved residues in the single or tandem Zn-knuckle motifs of the nucleocapsid (NC) domain. CP

sequences are conserved throughout the *Caulimoviridae* family; however, for convenient representation, only CP sequences from viruses classified into genera *Caulimovirus* and *Badnavirus* are shown in Figure 3. Abbreviations: HIV-1, human immunodeficiency virus 1; BLV, bovine leukemia virus; JSRV, Jaagsiekte sheep retrovirus; LPDV, lymphoproliferative disease virus; MMTV, mouse mammary tumor virus; MPMV, Mason-Pfizer monkey virus; PyERV, Python molurus endogenous retrovirus; EIAV, equine infectious anemia virus; DRM, Danio rerio Mag element; BSGFV, banana streak Goldfinger virus; BSOLV, banana streak OL (badna)virus; CaMV, cauliflower mosaic virus; CERV, carnation etched ring virus; CSSV, cacao swollen shoot virus; DBALV, Dioscorea bacilliform AL virus; MiMV, Mirabilis mosaic virus; TaBV, taro bacilliform virus.

References

- Gladyshev EA, Arkhipova IR. A widespread class of reverse transcriptase-related cellular genes. *Proc Natl Acad Sci U S A*. 2011; 108(51):20311-6.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010; 59(3):307-21.
- Krupovic M, Koonin EV. Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses, pseudoviruses, and metaviruses. *J Virol*. 2017; 91(12):e00210-17.
- Menéndez-Arias L, Sebastián-Martín A, Álvarez M. Viral reverse transcriptases. *Virus Res*. 2017; 234:153-76.
- Nassal M. Hepatitis B viruses: reverse transcription a different way. *Virus Res*. 2008; 134(1-2):235-49.
- Richert-Pöggeler KR, Shepherd RJ. Petunia vein-clearing virus: a plant pararetrovirus with the core sequences for an integrase function. *Virology*. 1997; 236(1):137-46.
- Steven AC, Conway JF, Cheng N, Watts NR, Belnap DM, Harris A, Stahl SJ, Wingfield PT. Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv Virus Res*. 2005; 64:125-64.
- Vo JN, Campbell PR, Mahfuz NN, Ramli R, Pagendam D, Barnard R, Geering AD. 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*. 2016; 97(12):3446-3457.
- Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J*. 1990; 9(10):3353-62.
-

