



HAL
open science

Homologous Capsid Proteins Testify to the Common Ancestry of Retroviruses, Caulimoviruses, Pseudoviruses, and Metaviruses

Mart Krupovic, Eugene Koonin

► **To cite this version:**

Mart Krupovic, Eugene Koonin. Homologous Capsid Proteins Testify to the Common Ancestry of Retroviruses, Caulimoviruses, Pseudoviruses, and Metaviruses. *Journal of Virology*, 2017, 91 (12), pp.e00210-17. 10.1128/JVI.00210-17. pasteur-01977363

HAL Id: pasteur-01977363

<https://pasteur.hal.science/pasteur-01977363v1>

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 **Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses,**
2 **pseudoviruses and metaviruses**
3

4 **Mart Krupovic^{a*} and Eugene V. Koonin^b**

5
6 a – Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Institut Pasteur, 25 rue du Docteur Roux,
7 75015

8 b – National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894,
9 USA

10
11 * – Correspondence
12 E-mail: krupovic@pasteur.fr
13

14

15 **Text**

16 Reverse-transcribing viruses are classified into 5 different families, *Retroviridae*, *Metaviridae*,
17 *Pseudoviridae*, *Caulimoviridae* and *Hepadnaviridae* (1). Retroviruses, hepadnaviruses and caulimoviruses
18 are full-fledged viruses, whereas metaviruses and pseudoviruses are more often referred to as
19 retrotransposons. Nevertheless, similar to other reverse-transcribing viruses, certain members of the
20 families *Metaviridae* and *Pseudoviridae* form icosahedral or irregularly-shaped particles that play
21 important roles in their infection cycles (2, 3). Phylogenetic analyses reveal monophyly of the reverse
22 transcriptases (RT) of all these viruses, to the exclusion of the RTs of other retroelements (1, 4). By
23 contrast, the evolutionary relationships between the capsid proteins (CP) of these viruses remain
24 obscure.

25

26 Evolutionary connections between viruses that have diverged in a distant past are often difficult to trace
27 due to high mutation and recombination rates in the viral genomes, which abrogate their taxonomic
28 classification, especially at the level of higher taxa. However, it has been argued that comparison of
29 sequences and structures of the major virion proteins provides a deeper reach into the evolutionary

30 history of the virosphere (5-8). High-resolution CP structures are available for several retroviruses and
31 hepadnaviruses. Members of the *Retroviridae* encode their CP as part of a larger polyprotein precursor
32 known as Gag, which is proteolytically processed by the virus-encoded protease to release several major
33 proteins, three of which, namely matrix, capsid (CA), nucleocapsid (NC), are conserved in all retroviruses
34 (9). The CA protein is composed of two α -helical domains, the N-terminal domain (NTD) and the C-
35 terminal domain (CTD) (10, 11). The CP of hepadnaviruses also adopts an α -helical fold, but does not
36 bear recognizable similarity to retroviral CA (12). By contrast, members of the family *Metaviridae*,
37 including Ty3/Gypsy retrotransposons, encode a Gag polyprotein, which contains the CA and NC domains
38 related to those of retroviruses (13, 14). Additionally, the Zn-binding module, known as 'Zn-knuckle', that
39 is found in many retrovirus and metavirus NC domains is also present in the CPs of caulimoviruses and
40 members of the family *Pseudoviridae*, such as Ty1/Copia retrotransposons (15). However, the CPs from
41 the latter two groups of viruses are generally considered to be unrelated to the CA of retroviruses or
42 metaviruses (1, 16), suggesting that different groups of reverse-transcribing viruses could have evolved
43 from non-viral retroelements via acquisition of capsid-encoding genes from different sources. As a case
44 in point, envelope proteins responsible for host recognition and membrane fusion have been recruited
45 by reverse-transcribing viruses on several independent occasions (1, 17).

46

47 To investigate the provenance of the pseudovirus and caulimovirus CPs, we performed sensitive profile-
48 profile searches using HHpred (18) seeded with sequences of CPs of pseudoviruses, caulimoviruses and
49 metaviruses. In all cases, we obtained highly significant matches (>98% HHpred probability) to the Gag of
50 retroviruses. The region of similarity encompasses the end of the CA-NTD and extends throughout the
51 CA-CTD and NC (data not shown). To validate these matches, we aligned the retroviral CA-NC sequences
52 to the corresponding protein sequences from the three other families of reverse-transcribing viruses.
53 Inspection of the alignment showed that the sequence conservation among the aligned proteins

54 encompasses not only CA-CTD and NC domains but also CA-NTD, except for two short α -helices (α 5 and
55 α 6) that have no equivalents in CPs of viruses other than retroviruses (Figures 1 and 2). Importantly, non-
56 polar residues, which stabilize the hydrophobic core of the α -helical bundles of both CA-NTD and CA-CTD
57 (10, 19), are conserved in all CP/CA proteins. Thus, patterns of sequence and secondary structure
58 conservation strongly support the homology of the CA/CP-NC module of all reverse-transcribing viruses,
59 except for hepadnaviruses. This result, combined with the monophyly of viral RTs (4), strongly suggests
60 that retroviruses, caulimoviruses, pseudoviruses and metaviruses evolved from a common viral ancestor,
61 rather than from distinct capsid-less retrotransposons. Such an ancestral virus encoding the CA/CP-NC,
62 protease (14) and the RT, including the RNase H domain (20), is likely to have existed prior to the
63 divergence of plants and opisthokonts (fungi and animals) ~1.6 billion years ago (21). Furthermore, given
64 that, in the RT phylogeny, hepadnaviruses cluster with pseudoviruses (1, 4), it appears most likely that
65 the conserved CP/CA was acquired by the common ancestor of all reverse-transcribing viruses, followed
66 by replacement with an unrelated protein in hepadnaviruses. Finally, our results seem to justify the
67 creation of a single, high-order viral taxon unifying all families of reverse-transcribing viruses (or
68 alternatively, four families, excluding hepadnaviruses), consistent with previous suggestions based on
69 the shared mechanisms of genome replication employed by these viruses (22).

70 References

- 71 1. **Koonin, E. V., V. V. Dolja, and M. Krupovic.** 2015. Origins and evolution of viruses of eukaryotes:
72 The ultimate modularity. *Virology* **479-480**:2-25.
- 73 2. **King, A. M. Q., M. J. Adams, E. B. Carstens, and E. J. Lefkowitz.** 2011. *Virus Taxonomy*. Ninth
74 Report of the International Committee on Taxonomy of Viruses. Elsevier Academic, London.
- 75 3. **Palmer, K. J., W. Tichelaar, N. Myers, N. R. Burns, S. J. Butcher, A. J. Kingsman, S. D. Fuller, and**
76 **H. R. Saibil.** 1997. Cryo-electron microscopy structure of yeast Ty retrotransposon virus-like
77 particles. *J Virol* **71**:6863-8.
- 78 4. **Gladyshev, E. A., and I. R. Arkhipova.** 2011. A widespread class of reverse transcriptase-related
79 cellular genes. *Proc Natl Acad Sci U S A* **108**:20311-6.
- 80 5. **Krupovic, M., and D. H. Bamford.** 2010. Order to the viral universe. *J Virol* **84**:12476-9.
- 81 6. **Sinclair, R. M., J. J. Ravantti, and D. H. Bamford.** 2017. Nucleic and amino acid sequences
82 support structure-based viral classification. *J Virol*.
- 83 7. **Krupovic, M., and E. V. Koonin.** 2017. Multiple origins of viral capsid proteins from cellular
84 ancestors. *Proc Natl Acad Sci U S A* **114**:E2401-E2410.
- 85 8. **Abrescia, N. G., D. H. Bamford, J. M. Grimes, and D. I. Stuart.** 2012. Structure unifies the viral
86 universe. *Annu Rev Biochem* **81**:795-822.
- 87 9. **Mattei, S., F. K. Schur, and J. A. Briggs.** 2016. Retrovirus maturation-an extraordinary structural
88 transformation. *Curr Opin Virol* **18**:27-35.
- 89 10. **Ganser-Pornillos, B. K., A. Cheng, and M. Yeager.** 2007. Structure of full-length HIV-1 CA: a
90 model for the mature capsid lattice. *Cell* **131**:70-9.
- 91 11. **Ganser-Pornillos, B. K., M. Yeager, and W. I. Sundquist.** 2008. The structural biology of HIV
92 assembly. *Curr Opin Struct Biol* **18**:203-17.
- 93 12. **Steven, A. C., J. F. Conway, N. Cheng, N. R. Watts, D. M. Belnap, A. Harris, S. J. Stahl, and P. T.**
94 **Wingfield.** 2005. Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv*
95 *Virus Res* **64**:125-64.
- 96 13. **Larsen, L. S., M. Zhang, N. Beliakova-Bethell, V. Bilanchone, A. Lamsa, K. Nagashima, R. Najdi,**
97 **K. Kosaka, V. Kovacevic, J. Cheng, P. Baldi, G. W. Hatfield, and S. Sandmeyer.** 2007. Ty3 capsid
98 mutations reveal early and late functions of the amino-terminal domain. *J Virol* **81**:6957-72.
- 99 14. **Llorens, C., M. A. Fares, and A. Moya.** 2008. Relationships of gag-pol diversity between
100 Ty3/Gypsy and *Retroviridae* LTR retroelements and the three kings hypothesis. *BMC Evol Biol*
101 **8**:276.
- 102 15. **Covey, S. N.** 1986. Amino acid sequence homology in gag region of reverse transcribing elements
103 and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res* **14**:623-33.
- 104 16. **Pachulska-Wieczorek, K., S. F. Le Grice, and K. J. Purzycka.** 2016. Determinants of Genomic RNA
105 Encapsulation in the *Saccharomyces cerevisiae* Long Terminal Repeat Retrotransposons Ty1 and
106 Ty3. *Viruses* **8**:E193.
- 107 17. **Malik, H. S., S. Henikoff, and T. H. Eickbush.** 2000. Poised for contagion: evolutionary origins of
108 the infectious abilities of invertebrate retroviruses. *Genome Res* **10**:1307-18.
- 109 18. **Söding, J.** 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**:951-
110 60.
- 111 19. **Ball, N. J., G. Nicastro, M. Dutta, D. J. Pollard, D. C. Goldstone, M. Sanz-Ramos, A. Ramos, E.**
112 **Müllers, K. Stirrnagel, N. Stanke, D. Lindemann, J. P. Stoye, W. R. Taylor, P. B. Rosenthal, and I.**
113 **A. Taylor.** 2016. Structure of a spumaretrovirus Gag central domain reveals an ancient retroviral
114 capsid. *PLoS Pathog* **12**:e1005981.
- 115 20. **Malik, H. S., and T. H. Eickbush.** 2001. Phylogenetic analysis of ribonuclease H domains suggests
116 a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Res* **11**:1187-
117 97.

- 118 21. **Hedges, S. B., J. E. Blair, M. L. Venturi, and J. L. Shoe.** 2004. A molecular timescale of eukaryote
119 evolution and the rise of complex multicellular life. *BMC Evol Biol* **4**:2.
- 120 22. **Hull, R.** 2001. Classifying reverse transcribing elements: a proposal and a challenge to the ICTV.
121 International Committee on Taxonomy of Viruses. *Arch Virol* **146**:2255-61.
- 122 23. **Llorens, C., R. Futami, D. Bezemer, and A. Moya.** 2008. The Gypsy Database (GyDB) of mobile
123 genetic elements. *Nucleic Acids Res* **36**:D38-46.
- 124 24. **Pei, J., and N. V. Grishin.** 2014. PROMALS3D: multiple protein sequence alignment enhanced
125 with evolutionary and three-dimensional structural information. *Methods Mol Biol* **1079**:263-71.
126
127

128 **Figure legend**

129 **Figure 1.** Multiple sequence alignment of CA/CP proteins of reverse-transcribing viruses belonging to
130 families *Retroviridae*, *Metaviridae*, *Pseudoviridae* and *Caulimoviridae*. Representative sequences from all
131 four virus groups were downloaded from the Gypsy Database (23), supplemented with sequences of
132 retroviral CA proteins for which X-ray structures are available and were aligned using PROMALS3D (24).
133 Secondary structure elements above the alignment are indicated for the Rous sarcoma virus CA (PDB:
134 5A9E), which was retrieved as the best hit to CP proteins of metaviruses, pseudoviruses and
135 caulimoviruses in HHpred search (data not shown). Red and blue asterisks indicate the conserved
136 residues in the single or tandem Zn-knuckle motifs of the nucleocapsid (NC) domain. Abbreviations: HIV-
137 1, human immunodeficiency virus 1; BLV, bovine leukemia virus; JSRV, Jaagsiekte sheep retrovirus; LPDV,
138 lymphoproliferative disease virus; MMTV, mouse mammary tumor virus; PyERV, *Python molurus*
139 endogenous retrovirus; EIAV, equine infectious anemia virus; BSGFV, banana streak goldfinger virus;
140 BSOLV, banana streak OL (badna)virus; CaMV, cauliflower mosaic virus; CERV, carnation etched ring
141 virus; DBV, dasheen bacilliform virus; DrMV, *Dracaena* mottle virus; MIMV, *Mirabilis* mosaic virus; TaBV,
142 taro bacilliform virus.

143

144 **Figure 2.** Conservation of the amino acid residues contributing to the stabilization of the hydrophobic
145 core of the α -helical bundles of CA-NTD and CA-CTD. Sequence conservation derived from the alignment
146 shown in Figure 1 was mapped onto the CA structure of Rous sarcoma virus (PDB: 5A9E). The least
147 conserved regions are colored red, whereas those most conserved are in blue.



