Virus Genes (2026) 62:846–850
https://doi.org/10.1007/s11262-026-02277-0

BRIEF REPORT

Genome-wide identification of rhabdoviral sequences in alfalfa (*Medicago sativa* L.)

Sam Grinstead[1] · **Lev G. Nemchinov**[1]

Received: 4 June 2026 / Accepted: 18 September 2026 / Published online: 26 September 2026
This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2026

Abstract

We recently reported the identification of endogenous viral elements (EVEs) originating from the *Caulimoviridae* family within the alfalfa (*Medicago sativa* L.) genome. Our subsequent identification of ubiquitous rhabdoviral elements in infected and healthy alfalfa tissues by high-throughput sequencing suggested that the alfalfa genome might be populated with integrated rhabdoviruses as well. Bioinformatics analysis using 26 publicly available alfalfa genomes proved the suggestion accurate. We found multiple nonretroviral segments of the *Rhabdoviridae* family belonging to the genera *Betanucleorhabdovirus* and *Betacytorhabdovirus* that appeared to be stable constituents of the host genome. In that capacity, they could potentially acquire functional roles in alfalfa development and response to environmental stresses. We believe this study reveals the first documented case of rhabdoviruses integrated into the alfalfa genome.

Keywords Rhabdoviruses · Endogenous viral elements · Alfalfa · (*Medicago sativa* L) · Genome

Endogenous viral elements (EVEs) are viral sequences integrated into host chromosomes and inherited by the host progeny as genetic alleles [1–3]. Plant genomes contain two main types of EVEs: endogenous pararetroviruses (EPRVs), which originate from reverse-transcribing dsDNA viruses of the family *Caulimoviridae*, and endogenous nonretroviral elements (ENREs) derived from ssDNA, ssRNA, and dsRNA viruses [4–7]. While integration of EPRVs into plant genomes is characterized relatively well [8], insertion mechanisms of the ENREs remain largely obscure [5]. Their integration may be facilitated by reverse transcriptase (RT) provided by pararetroviruses or retrotransposons [5]. In this scenario, the host or pararetroviral RT could inadvertently bind to abundant viral mRNAs or genomic RNA fragments rather than its own templates, converting rhabdoviral RNA into cDNA.

As of today, no ENREs were documented in alfalfa. We recently reported that alfalfa genome contains EVEs of the EPRV type and proposed that they are stable constituents of the alfalfa genome [3]. Our subsequent discoveries of rhabdoviral sequences in both healthy and infected alfalfa via HTS indicated that the plant genome may contain integrated rhabdoviruses as well [9–13].

In this work, using publicly available alfalfa genomes, we performed bioinformatics analysis to identify the presence of endogenous viral elements, specifically those of the rhabdoviral origin. A profile Hidden Markov Model (HMM) [14, 15] for the *Rhabdoviridae* family was constructed using HMMER v3.4 (http://hmmer.org/). Briefly, 85 complete plant *Rhabdoviridae* sequences were obtained from NCBI's Refseq database; the nucleotide regions making up each conserved ORF (N, P, M, G, and L) (Online Resource 1) were aligned in MAFFT v7.055b [16] and this alignment was used to generate a probabilistic model reflecting conserved family domains. The compiled HMM profile was then used to search for rhabdoviral-related EVEs in 26 alfalfa genomes obtained from diverse alfalfa cultivars originating from different countries [17] (BioProjec PRJNA1220045); and [18] (BioProject PRJNA1193221) (Online Resource 1). To the best of our knowledge, these included most, if not all, publicly available complete alfalfa genomes.

Only hits with an E-value of 1×10^{-10} and lower were used in order to minimize false positives. Individual identified sequences were then searched via BLASTx against

Edited by Seung-Kook Choi.

✉ Lev G. Nemchinov
lev.nemchinov@usda.gov; lnemchinov.usda@gmail.com

[1] Molecular Plant Pathology Laboratory, USDA-ARS, Beltsville Agricultural Research Center, Beltsville, MD, USA

the ClusteredNR database to identify virus species. In a separate analysis, we searched the latest alfalfa genome annotation, ZM4_V2.0 [19], (BioProject PRJNA1179654), and extracted uncategorized regions marked as viral that lacked taxonomic classification. These sequences were then aligned against the NCBI non-redundant (nr) database using BLASTx to identify the closest relative virus species. Phylogenetic analysis was performed with MEGA12.1 software [20] using Maximum Likelihood method and Tamura-Nei model of nucleotide substitutions [21].

Screening 26 alfalfa genomes with Hidden Markov Model profiles revealed multiple rhabdoviral nucleocapsid protein (N) sequences across three chromosomes (Table 1, Online Resource 2). No other sequences were identified using the strict E-value threshold. However, sequences matching the P (phosphoprotein) and L (RNA-dependent RNA polymerase) were also identified when evaluating marginal or subthreshold E-values (Online Resource 2). While the integration of these sequences remains a possibility, we excluded them from our analysis as not sufficiently robust.

The identified N sequences spanned 1,065–1,069 nt on chromosome 3 across five genomes, 1,016–1,022 nt on chromosome 4 across nine genomes, and 1,078–1,388 nt on chromosome 5 across all 26 genomes (Table 1, Online Resource 2).

Individual analysis of all detected EVEs via BLASTx searches against the ClusteredNR database identified sequences that shared 78–100% identity (E-value=0; query coverage 89–100%) with the nucleocapsid proteins of two previously described viruses: alfalfa nucleorhabdovirus 1 (ANRV1, UGO47147.1) and alfalfa cytorhabdovirus 1 (ACRV1, UGO47148.1) (Online Resource 3) [12]. Phylogenetic analysis performed with MEGA12.1 software [20] clustered ANRV1 and ACRV1 EVEs found in 26 genomes with representative members of the genera *Betanucleorhabdovirus* and *Betacytorhabdovirus,* respectively (Fig. 1). When unclassified viral sequences from alfalfa genome ZM4_V2 were extracted [19], (BioProject PRJNA1179654) and subjected to BLASTx search, several rhabdoviral EVEs were found in all four haplotypes of Chr 5 that mapped to ANRV1 (~88-96% identity; E-value=0; Online Resource 3), partially confirming results of our 26-genome analysis [17,18].

In summary, bioinformatics analysis revealed numerous endogenous viral elements of the nucleoprotein (N) gene integrated into alfalfa chromosomes 3, 4, and 5. Although marginal or subthreshold E-values revealed potential integration of P and L sequences within alfalfa genomes, they were excluded from our analysis as not sufficiently robust. However, given that 185 endogenous rhabdoviral elements encoding not only nucleoproteins, but also glycoproteins, and RNA-dependent RNA polymerases have been documented in the genomes of the mosquito (*Aedes aegypti*) and the blacklegged tick (*Ixodes scapularis*) [22], similar integration events may occur in plants – in this case, *M. sativa*.

The N gene sequences are the most abundant and highly transcribed mRNA during rhabdovirus infection and thus likely primarily available for accidental endogenization. Although these elements share phylogenetic relationship with the *Betanucleorhabdovirus* and *Betacytorhabdovirus* genera of the family *Rhabdoviridae*, ANRV1 and ACRV1 are known to be relatively divergent from other representatives of the family [12], which may suggest that their EVEs are uniquely adapted to *Medicago sativa.* Consequently, they may have evolved and potentially acquired functional roles in alfalfa development and response to environmental stresses, as was reported for integrated pararetroviruses linked to heritable RNA-mediated immunity in plants [23]. Similarly, the biological significance of integrated rhabdoviral N sequences, which, during natural infection, are responsible for encapsidating the negative-sense RNA genome, may primarily relate to a targeted immune response. This defense mechanism could potentially halt both viral replication and cell-to-cell movement.

We believe this study reveals the first documented case of rhabdoviruses integrated into the genomes of *M. sativa*. These findings also imply that some previously reported alfalfa rhabdoviruses may potentially represent EVEs rather than active infections, though verification is required.

Table 1 Rhabdoviral sequences identified in 26 alfalfa genomes (PRJNA1220045 and PRJNA1193221)

№	Accession	Query	Align from	Align to	E-value	Length	Target description
	CM108033.1	N*	100591199	100589812	3.00E−118	1388	M. sativa isolate S58 Chr 5
	CM107961.1	N	86335110	86333845	8.00E−117	1266	M. sativa isolate S227 Chr 5
	CM107993.1	N	89087491	89086226	1.00E−116	1266	M. sativa isolate S147 Chr 5
	CM121174.1	N	77730637	77729374	2.00E−116	1264	M. sativa cv. Bolivia Chr 5
	CM107969.1	N	90220257	90218994	2.00E−116	1264	M. sativa isolate S206 Chr 5
	CM108009.1	N	93133461	93132196	2.00E−116	1266	M. sativa isolate S89 Chr 5
	CM108113.1	N	92244351	92243086	2.00E−116	1266	M. sativa isolate B62 Chr 5
	CM107953.1	N	100444835	100443573	3.00E−116	1263	M. sativa isolate S387 Chr 5
	CM108049.1	N	90584661	90583399	3.00E−116	1263	M. sativa isolate S44 Chr 5
	CM108041.1	N	86457080	86455815	3.00E−116	1266	M. sativa isolate S55 Chr 5
	CM108081.1	N	88733969	88732707	3.00E−116	1263	M. sativa isolate B97 Chr 5
	CM107977.1	N	88724662	88723400	3.00E−116	1263	M. sativa isolate S199 Chr 5
	CM107969.1	N	90898112	90896847	7.00E−116	1266	M. sativa isolate S206 Chr 5
	CM107937.1	N	84162729	84163995	2.00E−115	1267	M. sativa isolate ZM4 LG05
	CM108105.1	N	84414687	84413422	3.00E−115	1266	M. sativa isolate B88 Chr 5
	CM108073.1	N	82032750	82031487	5.00E−115	1264	M. sativa isolate B108 Chr 5
	CM107945.1	N	84122908	84121536	6.00E−115	1373	M. sativa isolate ZM3 Chr 5
	CM108017.1	N	89226219	89224956	1.00E−113	1264	M. sativa isolate S115 Chr 5
	CM108065.1	N	91424109	91422846	1.00E−113	1264	M. sativa isolate B160 Chr 5
	CM108097.1	N	89006326	89005063	1.00E−113	1264	M. sativa isolate B92 Chr 5
	CM108129.1	N	90152801	90151538	1.00E−113	1264	M. sativa isolate B32 Chr 5
	CM108001.1	N	86771033	86769757	5.00E−112	1277	M. sativa isolate S122 Chr 5
	CM107985.1	N	90950912	90949651	9.00E−112	1262	M. sativa isolate S150 Chr 5
	CM108121.1	N	86108228	86,106,988	2.00E−111	1241	M. sativa isolate B52 Chr 5
	CM108089.1	N	86792919	86791655	6.00E−111	1265	M. sativa isolate B106 Chr 5
	CM108025.1	N	89328635	89327419	2.00E−107	1217	M. sativa isolate S65 Chr 5
	CM108033.1	N	100592668	100591591	5.00E−105	1078	M. sativa isolate S58 Chr 5
	CM108057.1	N	88346797	88345643	9.00E−101	1155	M. sativa isolate S9 Chr 5
	CM107975.1	N	70112560	70113628	1.00E−56	1069	M. sativa isolate S199 Chr 3
	CM108039.1	N	75075996	75077060	1.10E−56	1065	M. sativa isolate S55 Chr 3
	CM108063.1	N	74001673	74002741	1.10E−56	1069	M. sativa isolate B160 Chr 3
	CM107991.1	N	77960484	77961548	1.2E−56	1065	M. sativa isolate S147 Chr 3
	CM108055.1	N	74364313	74365377	1.2E−56	1065	M. sativa isolate S9 Chr 3
	CM108080.1	N	49500693	49501713	1.1E−54	1021	M. sativa isolate B97 Chr 4
	CM107944.1	N	53841332	53842353	1.2E−54	1022	M. sativa isolate ZM3 Chr 4
	CM107960.1	N	50846482	50847503	1.2E−54	1022	M. sativa isolate S227 Chr 4
	CM107976.1	N	53913708	53914729	1.2E−54	1022	M. sativa isolate S199 Chr 4
	CM108000.1	N	51479660	51480681	1.2E−54	1022	M. sativa isolate S122 Chr 4
	CM108016.1	N	55348410	55349430	1.2E−54	1021	M. sativa isolate S115 Chr 4
	CM108048.1	N	53332936	53333957	1.3E−54	1022	M. sativa isolate S44 Chr 4
	CM107952.1	N	50836138	50837159	1.9E−54	1022	M. sativa isolate S387 Chr 4
	CM108008.1	N	55864427	55865442	3.40E−54	1016	M. sativa isolate S89 Chr 4

N*—rhabdoviral nucleoprotein

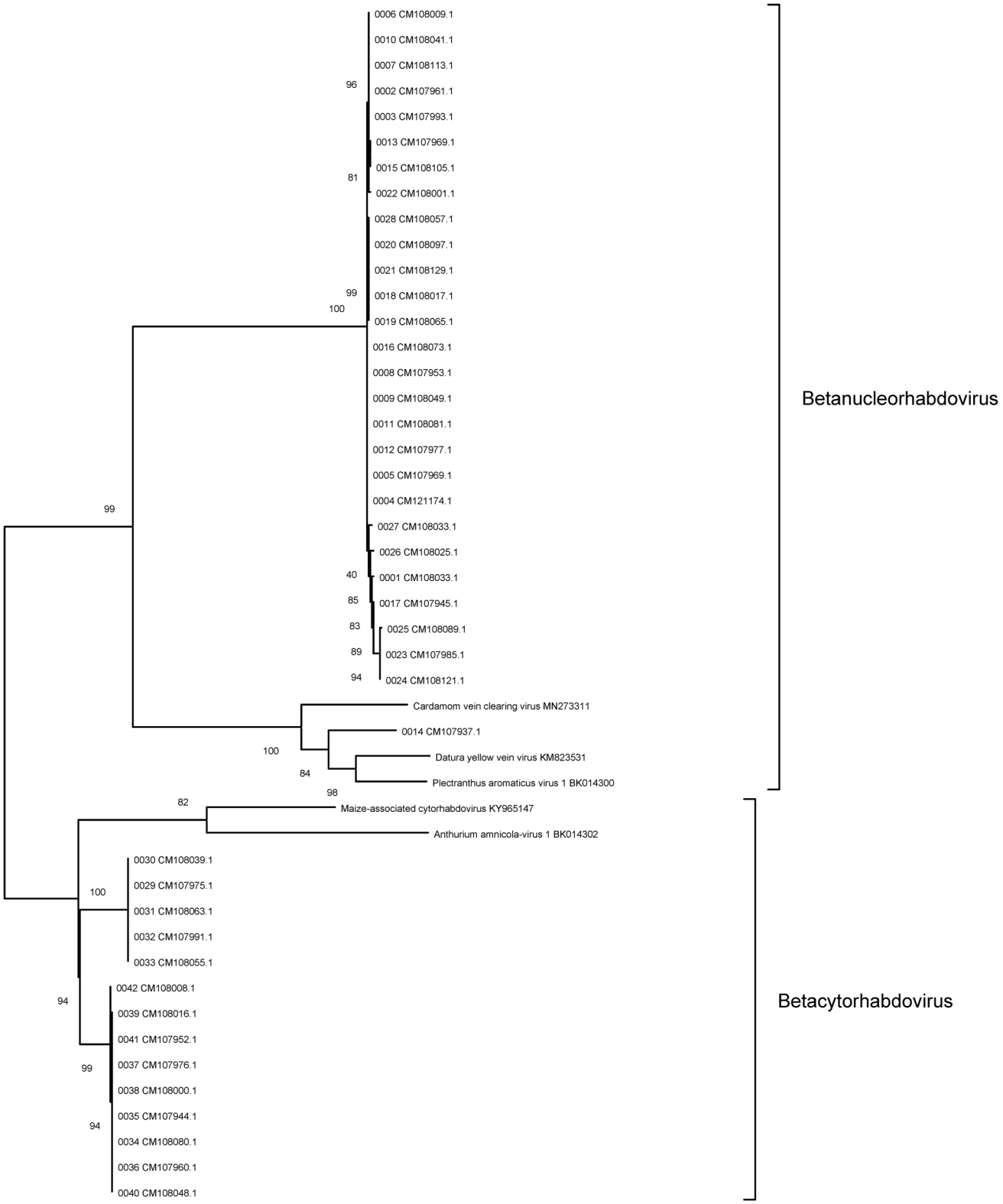

Fig. 1 Phylogenetic tree depicting rhabdoviral sequences identified in 26 alfalfa genomes (PRJNA1220045 and PRJNA1193221). The tree was constructed by MEGA 12 software (Kumar et al. 2024) using Maximum Likelihood method. The percentage of replicate trees in which the associated taxa clustered together (1,000 replicates) is shown next to the branches. The numbers 0001–0042 represent serial numbers, which are immediately followed by the corresponding accession numbers (e.g., CM108033.1)

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11262-026-02277-0.

Authors contribution SG: concept, HMM-profiling, bioinformatics analysis. LGN: concept, phylogenetic analysis, first draft and final version of the manuscript. Both authors edited the final version of the manuscript.

Funding This study was supported by the United States Department of Agriculture, the Agricultural Research Service, CRIS number 8042–21500-003-000D.

Data availability The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Holmes EC (2011) The evolution of endogenous viral elements. Cell Host Microbe 10:368–377
2. Bejarano ER, Khashoggi A, Witty M, Lichtenstein C (1996) Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proc Natl Acad Sci U S A 93:7555–7560
3. Boutanaev AM, Nemchinov LG (2021) Genome wide identification of endogenous viral sequences in alfalfa (*Medicago sativa* L.). Virol J 18:185
4. Diop SI, Geering ADW, Alfama-Depauw F, Loaec M, Teycheney PY, Maumus F (2018) Tracheophyte genomes keep track of the deep evolution of the Caulimoviridae. Sci Rep 8:572
5. Chiba S, Kondo H, Tani A, Saisho D, Sakamoto W (2011) Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. PLoS Pathog 7:e1002146
6. Takahashi H, Fukuhara T, Kitazawa H, Kormelink R (2019) Virus latency and the impact on plants. Front Microbiol 10:2764
7. Chu H, Jo Y, Cho WK (2014) Evolution of endogenous non-retroviral genes integrated into plant genomes. Curr Plant Biol 1:55–59
8. Staginnus C, Richert-Pöggeler KR (2006) Endogenous pararetroviruses: two faced travelers in the plant genome. Trends Plant Sci 11:485–491
9. Nemchinov LG, Grinstead S, Postnikova OA, Irish BM (2025) Characterization of the alfalfa pollen virome. Viruses 18:408
10. Nemchinov LG, Irish BM, Grinstead S, Postnikova OA (2023) Characterization of the seed virome of alfalfa (*Medicago sativa* L). Virol J 20:96
11. Nemchinov LG, Irish BM, Grinstead S, Postnikova OA (2025) Alfalfa transcriptomic responses to the field pathobiome. Plant Biol (Stuttg) 27:492–503
12. Nemchinov LG, Irish BM, Grinstead S, Shao J, Vieira P (2022) Diversity of the virome associated with alfalfa (*Medicago sativa* L.) in the U.S. Pacific Northwest. Sci Rep 12:8726
13. Nemchinov LG, Irish BM, Ushapovsky IV, Grinstead S, Shao J, Postnikova OA (2023) Composition of the alfalfa pathobiome in commercial fields. Front Microbiol 14:1225781
14. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763
15. Eddy SR (2011) Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195
16. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066
17. He F, Chen S, Zhang Y, Chai K, Zhang Q, Kong W, Wang Z, Chang Z, Zhao Z, Jiang F, Wang X, Zhang T (2025) Pan-genomic analysis highlights genes associated with agronomic traits and enhances genomics-assisted breeding in alfalfa. Nat Genet 57:1262–1273
18. Zhang H, Zhou L, Zhao H, Liang J, Liu Y, Wang C (2025) Chromosome-scale haplotype-resolved genome assembly of the autotetraploid alfalfa cultivar Bolivia. Plant Biotechnol J 23:4773–4775
19. Zhang F, Wei C, Shi X, Cao S, Xu X, Ma Z, Chen L, Yang X, Wang Y, Zhang J (2025) Medicago super-pangenome reveals adaptive advantages and evolutionary constraints in autotetraploid alfalfa. Nat Commun 17:583
20. Kumar S, Stecher G, Suleski M, Sanderford M, Sharma S, Tamura K (2024) MEGA12: molecular evolutionary genetics analysis version 12 for adaptive and green computing. Mol Biol Evol 41:msae263
21. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526
22. Fort P, Albertini A, Van-Hua A, Berthomieu A, Roche S, Delsuc F, Pasteur N, Capy P, Gaudin Y, Weill M (2012) Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. Mol Biol Evol 29(1):381–390
23. Valli AA, Gonzalo-Magro I, Sanchez DH (2023) Rearranged endogenized plant pararetroviruses as evidence of heritable RNA-based immunity. Mol Biol Evol 40:msad011

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.