Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of a putative novel cholivirus in the transcriptome of Gymnema sylvestre R. Br

V. Kavi Sidharthan (kavi.icfre@gmail.com)
Institute of Forest Biodiversity (ICFRE)
Gayatri Vanamala
Institute of Forest Biodiversity (ICFRE)
V Rajeswari
Malla Reddy University
V. K. Baranwal
ICAR-Indian Agricultural Research Institute

Short Report

Keywords: Cholivirus, novel, public domain, database

Posted Date: February 13th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2329992/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Archives of Microbiology on April 12th, 2023. See the published version at https://doi.org/10.1007/s00203-023-03517-9.

Abstract

Gymnema sylvestre is a tropical climber species that is widely used in traditional medicine since ages. In the present study, the transcriptome datasets of *G. sylvestre* available in public domain were screened for the presence of novel plant viral sequences and a putative novel virus tentatively named as Gymnema sylvestre virus 1 (GysV1) was identified. Coding-complete genome segments of GysV1 that are 6.35 kb (RNA1) and 3.98 kb (RNA2) long possessed a single large open reading frame coding for a polyprotein. BLASTp, sequence identity and phylogenetic analyses revealed the relatedness of GysV1 to the members of the subgenus *Cholivirus* (genus *Sadwavirus*; family *Secoviridae;* order *Picornavirales*). Based on the species demarcation criteria of the family *Secoviridae*, GysV1 can be regarded as a new cholivirus member.

Introduction

Gymnema sylvestre R. Br. (family Asclipiadaceae) is a medicinally important, perennial, woody climber species that grows in the tropical regions [1, 2]. Commonly known as madhunashini in India, *G. sylvestre* is used to treat diabetes since ages [2, 3]. Besides maintaining blood glucose homeostasis, the leaves of the plant enhance uptake and activities of glucose-utilizing enzymes by insulin-dependent pathways [3]. The plant is also used for the treatment of obesity and other ailments of animals and humans [1]. The medicinal properties of *G. sylvestre* could mainly be attributed to the presence of dammarane and oleanane classes of triterpene saponins and polyphenols [1, 3].

In recent years, several studies have generated the transcriptome data of *G. sylvestre* to gain insights into the genes regulating various biosynthetic pathways [1, 3, 4]. On the other hand, viral taxonomy is witnessing a shift from the traditional taxonomy to the sequence-based taxonomy, which advocates the inclusion of viral sequences known only from metagenomic data in the International Committee on Taxonomy of Viruses (ICTV) taxonomy scheme for comprehensively characterizing global virome [5]. The acceptance of sequence-based viral taxonomy paved way for data-driven virus discovery studies that identified several putative novel plant viruses in various plant transcriptomes [6, 7, 8, 9, 10, 11, 12]. Thus, in the present study, we explored the transcriptome datasets of *G. sylvestre* available in public domain for novel plant viral sequences and identified a putative novel cholivirus.

A total of ten transcriptome datasets of *G. sylvestre* (belonging to four Bioprojects: PRJNA325591, PRJNA399247, PRJNA491644, PRJNA543558) available in Sequence Read Archive (SRA) database of National Centre for Biotechnology Information (NCBI) were searched individually for the presence of reads of novel plant viruses using the RNA-dependent RNA Polymerase (RdRp) search tool available in Serratus [13] (alignment identity: ≥45%; score: ≥10). Raw reads of all the putative virus-positive libraries were imported directly into the Galaxy Australia server (https://usegalaxy.org.au/) [14] and trimmed using Trimmomatic v 0.36 [15]. Trimmed reads were assembled using rnaviralspades v 3.15.4 [16] and subjected to BLASTn analysis against the genome sequences of related viruses using blastn tool of ncbi blast+ v 2.10.1 [17] in the Galaxy server. Obtained viral contigs were examined for the presence of intact open reading frame (ORF) through NCBI ORF Finder (https://www.ncbi.nlm.nih.gov/orfnder/). Molecular weight estimation and motif prediction were performed using the tools mentioned in [18] while BLASTp analysis of encoded proteins were performed against 'non-redundant' database of NCBI for obtaining the maximum percent identity values at maximum query coverage (qcov) and minimum expectation value (evalue). Putative cleavage sites were predicted in viral polyproteins by comparing the known polyprotein sequences of related viruses. Trimmed reads of virus-positive libraries were mapped onto the recovered viral genomes and mean coverage values were obtained as described in [12]. Neighbourhood-joining (NJ) trees were generated using Poisson model with 1000 bootstrap replicates after MUSCLE alignment of protein sequences in MEGA7 v 7.0.26 [19] and the sequence identity matrices were generated using the Sequence Demarcation Tool v 1.2 [20].

Reads of a putative novel virus tentatively named as Gymnema sylvestre virus 1 (GysV1) that shared sequence similarities with known choliviruses were identified in all the libraries of the Bioproject PRJNA399247 through Serratus RdRp search and the details of each cholivirus-positive library are provided in Table 1. After assembly and BLASTn analysis, the two longest non-redundant viral contigs of lengths 6.35 kb and 3.98 kb that shared sequence identities with RNA1 and 2 of known choliviruses were regarded as RNA1 and 2 of GysV1, respectively.

GysV1 RNA 1 (BK062888) encoded for a 1929 aa (219.42 kDa) long polyprotein with RNA helicase (PF00910) and viral RdRp (PF00680) motifs that shared maximum sequence identity of 44.82% (at maximum qcov of 94%) with polyprotein 1 of Dioscorea mosaic associated virus (DMaV), a member of the subgenus *Cholivirus*. Putative cleavage sites predicted in GysV1 polyprotein 1 are Q(427)/T, Q(957)/G and Q(1186)/G. GysV1 RNA2 (BK062889) encoded for a 1132 aa (125.57 kDa) long polyprotein with no predicted motif that shared 34.22% sequence identity (at maximum qcov of 93%) with polyprotein 2 of DMaV. Putative cleavage site predicted in GysV1 polyprotein 2 is K(329)/G (Fig. 1).

Phylogenetic analysis based on the conserved protease-polymerase (Pro-Pol) region placed GysV1 in a distinct sub-clade that was related to Ananas comosus secovirus (AcSV) and DMaV (Fig 2a) while GySV1 was placed in a sister clade to DMaV in the polyprotein 2-based phylogenetic tree (Fig. 2b). Sequence identity matrix revealed the maximum identities of GysV1 Pro-Pol (60.80%) and polyprotein 2 (36.10%) sequences with AcSV Pro-Pol and DMaV polyprotein 2 sequences, respectively (Fig. 2c, d).

Recently, the genus *Sadwavirus* under the family *Secoviridae* was reorganized into three sub genera- *Satsumavirus*, *Stramovirus* and *Cholivirus* [21]. Currently, the sub genus *Cholivirus* contains three recognized members- chocolate lily virus A, DMaV, pineapple secovirus A and two putative members- AcSV and pineapple secovirus B [22, 11]. Based on the <80% and <75% amino acid sequence identity criteria for the conserved Pro-Pol and polyprotein 2 sequences, respectively to regard a member of the family *Secoviridae* as a new species [23], GysV1 can be designated as a new member of the subgenus *Cholivirus* (genus *Sadwavirus*). Further studies are needed to understand the biological properties and prevalence of GysV1.

Declarations

Acknowledgements

We thank Galaxy Australia, a service provided by the Australian Biocommons and its partners. The authors are grateful to the Head, Division of Genetics and Tree Improvement, the GCR and the Director, IFB (ICFRE), Hyderabad for their support. The authors are also thankful to the original submitters of datasets used in the study.

Author contributions

V. Kavi Sidharthan: conceptualization, methodology, formal analysis and investigation, writing- original draft preparation; Gayatri Vanamala, V. Rajeswari: formal analysis and investigation, writing- review and editing; V. K. Baranwal: conceptualization, resources, supervision, writing- review and editing.

Data availability

The viral genome sequences described in the study are deposited at NCBI Third Party Annotation database with accession numbers BK062888, BK062889 and are available as supplementary material (ESM_1).

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Research involving human participants and/or animals

This work does not contain any animal or human participants.

Informed consent

This work does not contain any animal or human participants.

References

1. Kalariya KA, Gajbhiye N, Minipara D, Meena RP, Kumar S, Saha A, Trivedi A, Manivel P (2019) Deep sequencing-based *de novo* transcriptome analysis reveals biosynthesis of gymnemic acid in *Gymnema sylvestre* (Retz.) Schult. Ecol Genet Genom 1(13):100047
2. Kalariya KA, Meena RP, Poojara L, Shahi D, Patel S (2021) Characterization of squalene synthase gene from *Gymnema sylvestre* R. Br. Beni-Suef University J Basic Appl 10(1):1–1
3. Ayachit G, Shaikh I, Sharma P, Jani B, Shukla L, Sharma P, Bhairappanavar SB, Joshi C, Das J (2019) *De novo* transcriptome of *Gymnema sylvestre* identified putative lncRNA and genes regulating terpenoid biosynthesis pathway. Sci Rep 9(1):1–3
4. Kalariya KA, Minipara DB, Manivel P (2018) *De novo* transcriptome analysis deciphered polyoxypregnane glycoside biosynthesis pathway in *Gymnema sylvestre*. 3 Biotech 8(9):1–1.
5. Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R (2017) Consensus statement: virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15:161–168
6. Bejerman N, Dietzgen RG, Debat H (2021) Illuminating the plant rhabdovirus landscape through metatranscriptomics data. Viruses 13:1304
7. Bejerman N, Debat H (2022) Exploring the tymovirales landscape through metatranscriptomics data. Arch Virol 167(9):1785–803
8. Bejerman N, Dietzgen RG, Debat H (2022) Unlocking the hidden genetic diversity of varicosaviruses, the neglected plant rhabdoviruses. Pathogens 11(10):1127
9. Sidharthan VK, Chaturvedi KK, Baranwal VK (2022). Diverse RNA viruses in a parasitic flowering plant (spruce dwarf mistletoe) revealed through RNA-seq data mining. J Gen Plant Pathol 88(2):138–44
10. Sidharthan VK, Nagendran K, Baranwal VK (2022) Exploration of plant transcriptomes reveals five putative novel poleroviruses and an enamovirus. Virus Genes 58:244–253
11. Sidharthan VK, Rajeswari V, Baranwal VK (2022) Analysis of public domain plant transcriptomes expands the phylogenetic diversity of the family *Secoviridae*. Virus Genes 58:598–604
12. Sidharthan VK, Rajeswari V, Vanamala G, Baranwal VK (2022) Revisiting the amalgaviral landscapes in plant transcriptomes expands the host range of plant amalgaviruses. Virology 577:65–73
13. Edgar RC, Taylor J, Lin V, Altman T, Barbera P, Meleshko D, Lohr D, Novakovsky G, Buchfink B, Al-Shayeb B, Banfield JF (2022) Petabase-scale sequence alignment catalyses viral discovery. Nature 602(7895):142–7
14. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M et al (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3-W10
15. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
16. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes *de novo* assembler. Curr Protoc Bioinform 70:e102
17. Cock PJ, Chilton JM, Grüning B, Johnson JE, Soranzo N (2015) NCBI BLAST + integrated into Galaxy. Gigascience 4:s13742-015
18. Sidharthan VK, Baranwal VK (2021) Mining of the water hyssop (*Bacopa monnieri*) transcriptome reveals genome sequences of two putative novel rhabdoviruses and a solendovirus.<background-color:#CCFF99;bvertical-align:super;></background-color:#CCFF99;bvertical-align:super;> Arch Virol 166:

1985–1990
19. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874
20. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PloS one 9(9):e108277
21. Sanfaçon H, Dasgupta I, Fuchs M, Karasev AV, Petrzik K, Thompson JR, … Yoshikawa N(2020) Proposed revision of the family *Secoviridae* taxonomy to create three subgenera,"*Satsumavirus*","*Stramovirus*" and "*Cholivirus*", in the genus *Sadwavirus*. Arch Virol 165:527–533
22. Larrea-Sarmiento A, Geering AD, Olmedo-Velarde A, Wang X, Borth W, Matsumoto TK, Suzuki JY, Wall MM, Melzer M, Moyle R, Sharman M (2022) Genome sequence of pineapple secovirus B, a second sadwavirus reported infecting *Ananas comosus*. Arch Virol 21:1–4
23. Thompson JR, Dasgupta I, Fuchs M, Iwanami T, Karasev AV, Petrzik K, … Consortium IR(2017) ICTV virus taxonomy profile: *Secoviridae*. J Gen Virol 98(4):529

Tables

Table 1. Summary of Gymnema sylvestre virus 1 identification in transcriptomes of *Gymnema sylvestre*

Library	Biosample	Total Paired-end Reads (in million)	Serratus RdRp search			Mean coverage (x)		Plant Cultivar	Geographic Location	Tissue	Reference
			Score	Identity (%)[a]	Reads	RNA1	RNA2				
SRR5965320	SAMN07528739	28.50	100	65	259299	9,782.5757	8,579.41	DGS 22	India: Andra Pradesh, Visakhapatnam	Flower	[1]
SRR5965321	SAMN07528740	22.91	100	65	59230	3,469.3372	2,529.52			Fruit	
SRR5965322	SAMN07528737	14.25	100	64	910	466.7836	459.75	DGS 03	India: Gujarat, Dang	Leaf	[4]
SRR5965323	SAMN07528738	30.48	53	64	110	58.0606	57.79	DGS 22	India: Andra Pradesh, Visakhapatnam	Leaf	[1]

[a]Alignment identity with other choliviruses

Figures

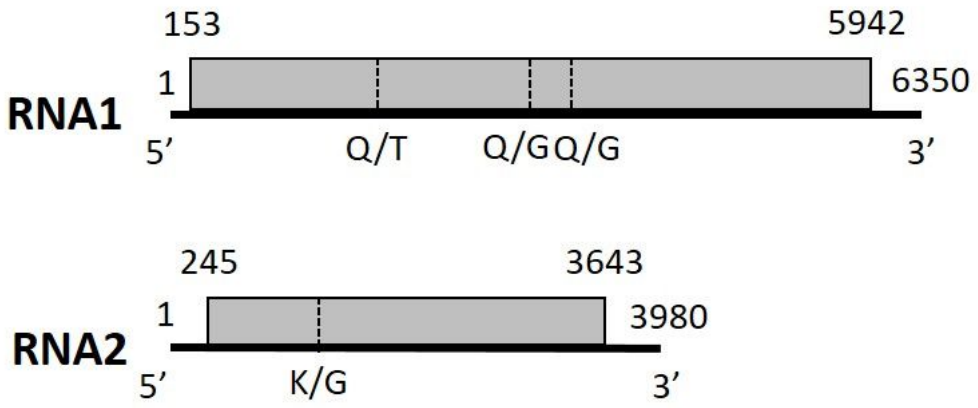


Figure 1

Genome organization of Gymnema sylvestre virus 1 (GysV1) identified in the present study. Cleavage sites predicted in polyproteins encoded by GysV1 are indicated.

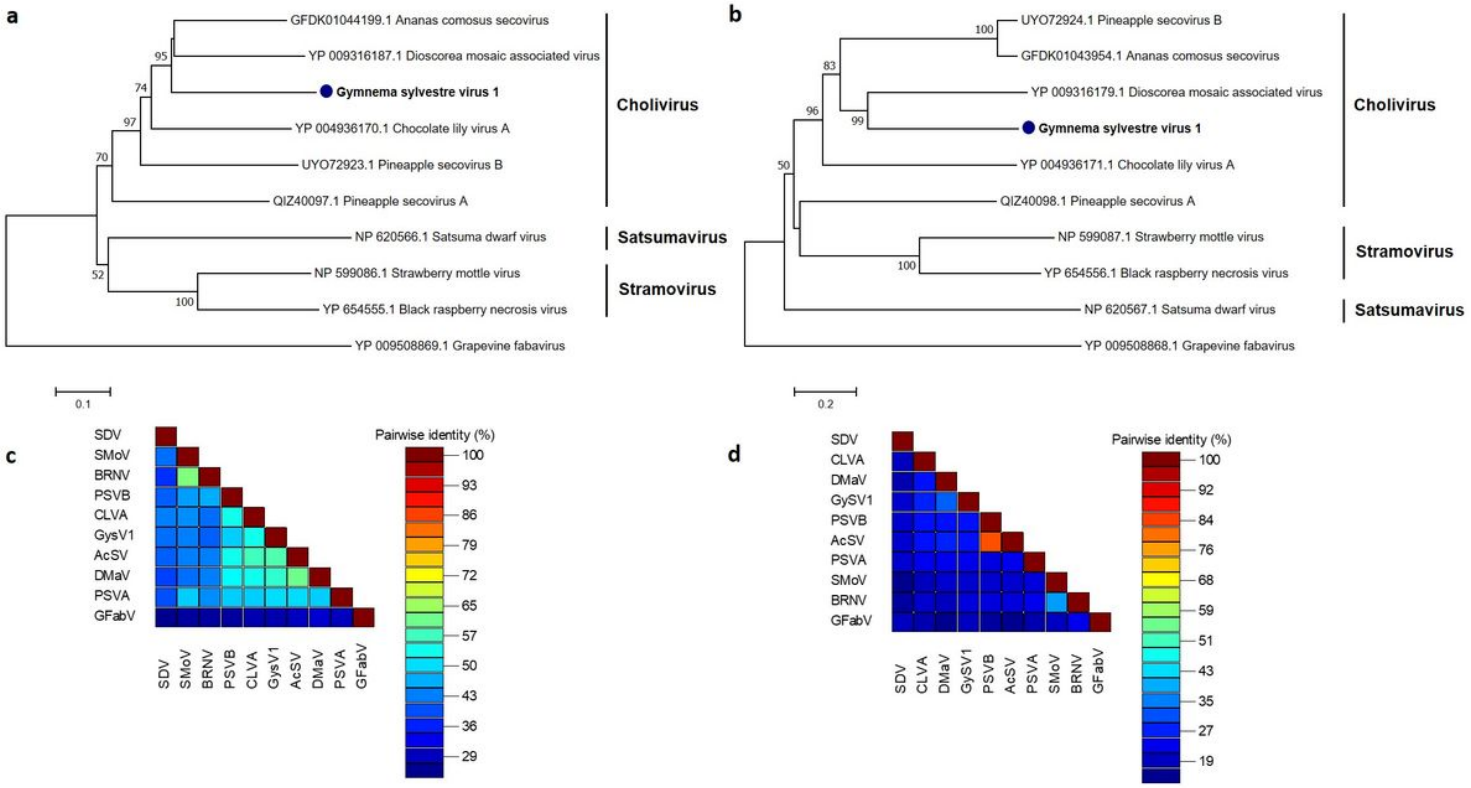


Figure 2

Phylogenetic trees **(a, b)** and sequence identity matrices **(c, d)** showing the relationship of Gymnema sylvestre virus 1 identified in the present study with other sadwaviruses based on the conserved Pro-Pol **(a, c)** and polyprotein **(b, d)** regions. Phylogenetic trees were constructed using Neighbour-Joining (NJ) method and Poisson model with 1000 bootstrap replicates. Bootstrap values >50% are only indicated. Grapevine fabaviral sequences were used as outgroups during phylogenetic tree construction. AcSV: Ananas comosus secovirus, BRNV: black raspberry necrosis virus, CLVA: chocolate lily virus A, DMaV: Dioscorea mosaic associated virus, GFabV: grapevine fabavirus, GysV1: Gymnema sylvestre virus 1, PSVA: pineapple secovirus A, PSVB: pineapple secovirus B, SDV: satsuma dwarf virus, SMoV: strawberry mottle virus.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.txt