Acta virologica **66**: 376 – 379, 2022 doi: 10.4149/av_2022_409

LETTER TO THE EDITOR

Genome sequence of a Papaya ringspot virus from khejri (*Prosopis cineraria*) transcriptome from India

Malyaj Rajeev Prajapati[1], Jitender Singh[1,2*], Pankaj Kumar[1], Virendra Kumar Baranwal[3]

[1]College of Biotechnology, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut, Uttar Pradesh, India-250110; [2]Department of Microbiology, Chaudhary Charan Singh University, Meerut, Uttar Pradesh, India-250004; [3]Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi, India-110012

Received May 19, 2022; revised August 15, 2022; accepted September 6, 2022

Summary. – *Prosopis cineraria* Druce (khejri), a member of the *Fabaceae* family, is an essential leguminous plant and agri-horticultural crop, which helps to protect the ecology of desert and semi-arid areas. Using the public domain RNA-seq dataset, we aimed to reveal the viral spectrum of khejri, which is still unexplored. Following quality control, 111,497 contigs (> 200 nt) were created by *de novo* assembly of 24,081,663 reads. The presence of one contig of 10316 nt related to Papaya ringspot virus (PRSV) with sequence identity 83.20%–93.21% was identified through BLASTx analysis, and 3,650 (0.02%) reads were mapped using full-length PRSV sequence as reference. Pairwise sequencing analysis revealed that the nucleotide and amino acid sequences of the complete genome sequence (BK061298) shared 84.30%–95.54% and 86.38%–96.48% similarity with other PRSV genomes, respectively. Phylogenetic analysis revealed that this virus isolate is closely related to PRSV isolate WB (host: *Carica papaya*; Acc. No. LC482263). Our findings show that *Prosopis cineraria* (khejri) is a novel host plant of PRSV in India by utilizing khejri transcriptome data and bioinformatics analysis.

Keywords: Prosopis cineraria; Potyvirus; Papaya ringspot virus; RNA-seq; phylogenetics; India

Cineraria Prosopis (L.) Druce, a member of the *Fabaceae* family, is a significant leguminous tree and agri-horticultural crop that helps to protect the ecology of arid and semi-arid regions. It is a key species native to the Indian subcontinent and recognized as a backbone of the rural economy of the people of these regions due to its great utility in ecological balance, agroforestry, economic and medicinal value. It is commonly known as 'Khejri' or 'Jandi' in India, 'Jand' in Pakistan and 'Ghaf' in Arab. It is being regarded as 'Golden tree of deserts', 'Love Tree', 'King of the desert', 'Pride of the Desert' because every plant part of this multi-purpose tree is used. It is the state tree of Rajasthan, India and the national tree of the United Arab Emirates in the Arabian Gulf *(1)*.

Papaya ringspot virus (PRSV) is a member of the genus *Potyvirus*. Potyviruses in the family *Potyviridae* are a large group of positive-strand RNA viruses that infect and cause diseases in diverse plant species. Some members have a narrow host range, whilst others can infect a wide range of hosts *(2)*. Several of the members cause disease symptoms in commercially significant crops, rendering this genus one of the largest thoroughly researched virus groups by plant virologists *(3)*. Potyviruses are primarily spread by aphids in a non-persistent manner as well as by mechanical means *(2)*. Vertical transmission of poty-

*Corresponding author. E-mail: jeets80@gmail.com; phone: +91-7017785669.

Abbreviations: HTS = high-throughput sequencing; NCBI = National Center for Biotechnology Information; ORF= open reading frame; PIPO = pretty interesting *Potyviridae* ORF; PRSV = Papaya ringspot virus

viruses through seeds has also been documented *(3, 4)*. Considering the economic importance of khejri, identifying and characterizing the disease-causing viruses of khejri is a prerequisite for devising suitable management strategies.

The rapid development of high-throughput sequencing (HTS) technology in recent years has not only enabled the investigation of diseases with unknown etiology but also facilitated the identification and characterization of known/novel viruses *(5)*. Through bioinformatic analysis of HTS reads, viruses associated with a particular sample can be unveiled and their genomes can be recovered *(6)*. Besides, HTS facilitates the detection of multiple infections of a plant species, thereby, helping us to comprehensively capture the entire virome and viral variability *(7, 8, 9)*. In the present study, khejri RNA-seq data downloaded from NCBI were analyzed to identify viruses infecting khejri.

An increasing number of viral isolates have been unraveled utilizing HTS around the world, allowing for major advancements in diagnostic tools. Furthermore, HTS technology may provide detailed insights into the etiological and epidemiological correlations of viruses associated with diseased plant samples *(6, 7, 8, 9)*. Thus, khejri RNA-seq data was utilized to identify the viral genome sequence in the present study. Only one khejri RNA-seq dataset (SRR5234756) is available in the public domain (submitted from India) and it was mined for the presence of viral sequences. Dataset SRR5234756 (BioProject: PRJNA371524) was derived from leaf tissues and the dataset was sequenced using Illumina HiSeq 2500 *(11)*. Raw reads obtained were imported into CLC

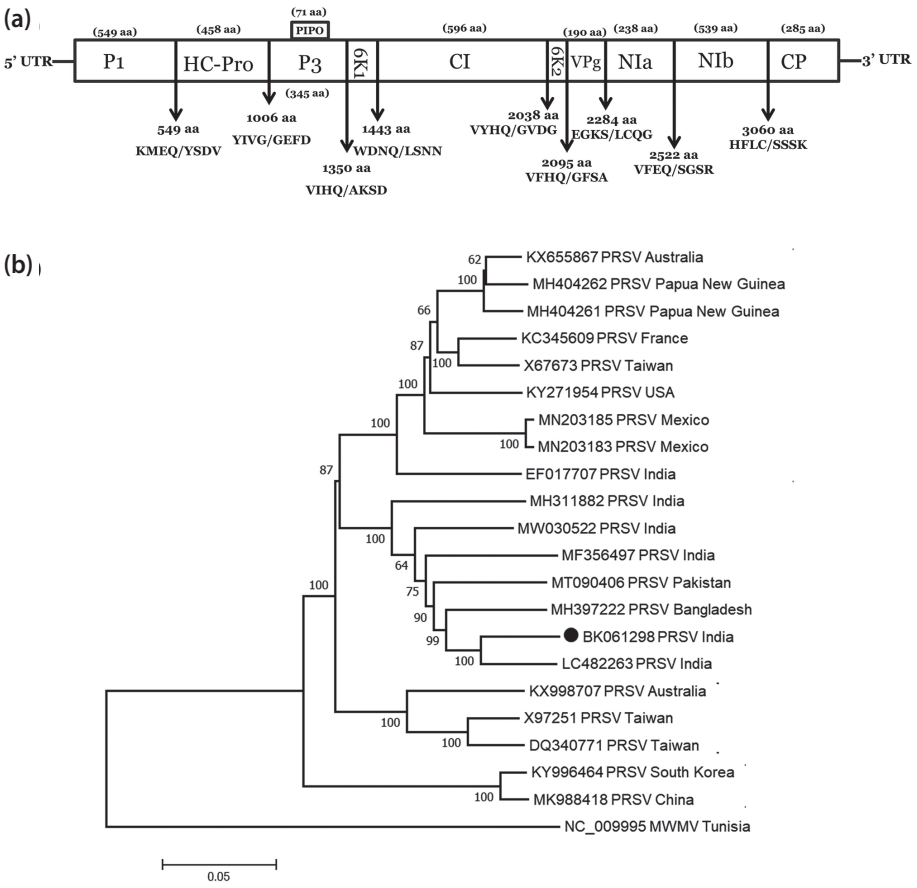

Fig. 1

Genome organization and Phylogenetic tree of PRSV (BK061298)

(a) Schematic presentation of PRSV genome (5'-UTR; P1; HC-Pro; P3; PIPO; 6K1; CI; 6K2; VPg; NIa; NIb; CP; 3'-UTR). **(b)** Phylogenetic analysis of nucleotide sequence of the PRSV isolate (Acc. No. BK061298) with other PRSV isolates: cucurbit crops (KX655867; Australia, KC345609; France, KX998707; Australia, KY996464; South Korea, MK988418; China) and papaya (MH404261; Papua New Guinea, MH404262; Papua New Guinea, X67673; Taiwan, KY271954; USA, MN203183; Mexico, MN203185; Mexico, EF017707; India, MH311882; India, MW030522; India, MF356497; India, MT090406; Pakistan, MH397222; Bangladesh, LC482263; India, X97251; Taiwan, DQ340771; Taiwan). Phylogenetic analyses were conducted in MEGA-X and the evolutionary history was inferred using the Neighbor-Joining method with 1000 bootstrap values and the dendrogram was rooted with distant member- Moroccan watermelon mosaic virus (NC_009995; Tunisia) of *Potyvirus*.

Genomic workbench (20.0.4) and quality trimmed using the trimming tool to remove ambiguous and adaptor sequences. After quality control, *de novo* assembly of 24,081,663 reads generated 111,497 contigs (> 200 nt) using CLC Genomic workbench. Obtained contigs were subjected to BLASTx analysis against the non-redundant (NR) database in the workbench. The alignment against reference viral genomes in the database was made using OmicsBox 1.3 (*https://www.biobam.com/omicsbox*). BLASTx analysis revealed the presence of one contig of 10316 nt related to PRSV, excluding the poly(A) tail, with sequence identity 83.20%–93.21% and 3,650 (0.02%) reads were mapped using full-length PRSV sequences as references. The complete virus genomic sequence was submitted to NCBI GenBank third party annotation (TPA) with Acc. No. BK061298. PRSV genomic RNA containing a large open reading frame (ORF), encoding for a polyprotein of 3344 aa, which began with an ATG codon at 76 nt and ended with a UAG termination codon at 10110 nt was obtained using NCBI ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder). The characteristic domains, conserved cleavage sites, and motifs specific to the genus *Potyvirus* were predicted in the polyprotein sequence encoded by the genomes *(12)*.

Furthermore, a PIPO ORF resulting from viral RNA polymerase slippage in the PRSV P3 cistrons was predicted downstream at positions 3552–3765 nt. The conserved $G_{1\text{-}2}A_{5\text{-}7}$ motif of PIPO was identified at position 3547 and this motif resembled the highly conserved motif found in other members of the family *Potyviridae (13)*. Like other potyviruses polyproteins, the PRSV polyprotein is proteolytically processed into 10 mature proteins - P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb and CP, at aa positions 549, 1006, 1350, 1443, 2038, 2095, 2284, 2522, and 3060 *(4, 7, 12)* (Fig 1a). Potyviruses are non-persistently transmitted by aphids in nature and this mode of transmission accounts for serious disease epidemics caused in various cultivated plants. Furthermore, potyvirus-specific conserved motifs were discovered in virus. The predicted motifs in PRSV cleavage products was ^{600}K-I-T-C^{603} (aphid transmission), ^{745}C-D-N-Q-L-D^{750} (symptomatology) and ^{858}P-T-K^{860} (aphid transmission) in HC-Pro; ^{1757}G-R-V-G-R^{1761} in CI; ^{1237}G-$(X)_2$-G-X-G-K-S^{1244}, ^{2710}F-T-A-A-P^{2714}, ^{2724}C-V-D-D^{2727}, ^{2828}G-N-N-S-G-Q-P-S-T-V-V-D-N-S-L-M-V^{2844} (RNA-dependent polymerase activity), ^{2872}G-D-D^{2874} (RNA-dependent polymerase activity) and ^{2914}W-F-M-S^{2917} in NIb; ^{3067}D-A-G^{3069} (aphid transmission), and ^{3297}Q-M-K-A-A-A^{3302} in CP *(14)*. Few of these motifs are suggested to play critical roles in the transmission of potyviruses by aphids. Despite the presence of all of these conserved motifs, aphid transmission necessitates species-specific interactions *(15)*, Further research is needed to determine the aphid species transmitting PRSV.

A phylogenetic tree was constructed using the neighbor-joining method and 1000 bootstrap replicates in MEGA-X *(16)* and pairwise nucleotide and amino acid sequence compositions and percent sequence identities were calculated using BioEdit 7.2 *(17)*. Pairwise sequence analysis indicated that the nucleotide and amino acid sequence of the complete genome (BK061298) shared 84.30%–95.54% and 86.38%–96.48% similarity with other PRSV isolates sequences, respectively. Phylogenetic analysis showed a close relationship of this virus isolate with PRSV isolate WB (host: *Carica papaya*; Acc. No. LC482263) (Fig. 1b). Our findings show that the *Prosopis cineraria* (khejri) as a novel host plant of PRSV in India by utilizing khejri transcriptome data and bioinformatics analysis. Furthermore, a global survey and molecular detection to investigate the distribution, severity of symptoms, and genetic diversity of PRSV should be conducted in the future to estimate the risk of PRSV to khejri production in India.

Acknowledgments. The authors acknowledge the Vice Chancellor, Sardar Vallabhbhai Patel University of Agriculture and Technology, Meerut-250110, Uttar Pradesh, CoE in Agri Biotech., Council of Science & Technology, Uttar Pradesh and Bioinformatics facility, Department of Biotechnology, India for providing the facilities to carry out this research work.

References

1. Singh P, Bangarwa KS, Dhillon RS, J. Pharmacogn. Phytochem. 8(5), 817–820, 2019.
2. Wylie SJ, Adams M, Chalam C, Kreuze J, López-Moya JJ et al., J. Gen. Virol. 98(3), 352, 2017. https://doi.org/10.1099/jgv.0.000740
3. Revers F, García JA, Adv. Virus Res. 92, 101–199, 2015. https://doi.org/10.1016/bs.aivir.2014.11.006
4. Simmons HE, Dunham JP, Zinn KE, Munkvold GP, Holmes EC, Stephenson AG, Virus Res. 176(1–2), 259–264, 2013. https://doi.org/10.1016/j.virusres.2013.06.016
5. Sidharthan VK, Kalaivanan NS, Baranwal VK, Gene 786, 145626, 2021. https://doi.org/10.1016/j.gene.2021.145626
6. Prajapati MR, Manav A, Singh J, Kumar P, Kumar A, Kumar R, Prakash S, Baranwal VK, Plants 11(2), 224, 2022. https://doi.org/10.3390/plants11020224
7. Villamor DEV, Ho T, Al Rwahnih M, Martin RR, Tzanetakis IE, Phytopathology 109(5), 716–725, 2019. https://doi.org/10.1094/PHYTO-07-18-0257-RVW
8. Singh J, Truong TN, An D, Prajapati MR, Manav A, Quoc NB, Baranwal VK, Acta Virol. 64(4), 427–432, 2020. https://doi.org/10.4149/av_2020_405
9. Prajapati MR, Manav A, Singh J, Singh MK, Ranjan K, Kumar A, Kumar P, Kumar R, Baranwal VK, J. Hortic. Sci. Biotechnol. 97(1), 96–105, 2022. https://doi.org/10.1080/14620316.2021.1963848

10. Adams IP, Miano DW, Kinyua ZM, Wangai A, Kimani E, Phiri N, Boonham N, Plant Pathol. 62(4), 741–749, 2013. https://doi.org/10.1111/j.1365-3059.2012.02690.x

11. Rai MK, Shekhawat JK, Kataria V, Shekhawat NS, Plant Gene. 12, 88–97, 2017. https://doi.org/10.1016/j.plgene.2017.09.002

12. Gorane A, Verma R, Naik A, Nikam T, Ade A, Mahapatro G, Tripathi S, J. Plant Pathol. 101(4), 1203–1209, 2019. https://doi.org/10.1007/s42161-019-00340-4

13. Rodamilans B, Valli A, Mingot A, San León D, Baulcombe D, López-Moya JJ, García JA, J. Virol. 89(13), 6965–6967, 2015. https://doi.org/10.1128/JVI.00337-15

14. Worrall EA, Hayward AC, Fletcher SJ, Mitter N, Arch. Virol. 164(1), 181–194, 2019. https://doi.org/10.1007/s00705-018-4065-6

15. Gibbs AJ, Mackenzie AM, Wei KJ, Gibbs MJ, Arch. Virol. 153(8), 1411–1420. https://doi.org/10.1007/s00705-008-0134-6

16. Kumar S, Stecher G, Li M, Knyaz C, Tamura K, Mol. Biol. Evol. 35(6), 1547, 2018. https://doi.org/10.1093/molbev/msy096

17. Hall TA. BioEdit Sequence Alignment Editor 7.0. 1. Carlsbad, CA, USA: Isis Pharmaceuticals, 2004.