Acta virologica **66**: 182 – 185, 2022 doi: 10.4149/av_2022_207

Dwarf polish wheat hosts a novel closterovirus: Revelation by transcriptome data-mining

Venkidusamy Kavi Sidharthan[1], Virendra Kumar Baranwal[2*]

[1]Division of Genetics and Tree Improvement, Institute of Forest Biodiversity (ICFRE), Hyderabad, India; [2]Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi, India

Received December 1, 2021; revised February 14, 2022; accepted April 4, 2022

Summary. – Closteroviruses are positive sense single-stranded RNA genome-containing plant viruses with narrow natural host range and wide distribution. In the present study, a putative novel closterovirus, Triticum polonicum closterovirus (TriPCV) was identified in the transcriptome assembled contigs of dwarf polish wheat available in public domain. The genome of TriPCV (15.36 kb; TPA Acc. No.: BK059767) contained nine open reading frames (ORFs) that encode for proteins involved in viral replication, cell-to-cell movement, encapsidation and suppression of host RNA silencing. Phylogenetic analysis revealed that TriPCV was distantly related to other members of the genus *Closterovirus*. Based on genome organization, sequence similarities in BLAST analysis, predicted motifs and phylogeny, TriPCV can be regarded as a putative novel member of the genus *Closterovirus*.

Keywords: Closterovirus; Triticum polonicum; transcriptome; public domain

Introduction

The family *Closteroviridae* comprises four filamentous plant infecting viral genera – *Ampelovirus*, *Closterovirus*, *Crinivirus* and *Velarivirus*. Members of *Closteroviridae* possess positive sense single-stranded RNA genomes of lengths13 kb to 19 kb that are largest among the plant-infecting viruses. Generally, closteroviruses are phloem-limited and have a limited natural host range (Candresse and Fuchs, 2020). They are regarded as important pathogens of various herbaceous and woody crops, mainly dicots (Fuchs *et al.,* 2020). Members of the genera *Ampelovirus*, *Closterovirus* and *Velarivirus* possess monopartite genomes while those of the genus *Crinivirus* contain bi-/tripartitite genomes with a 5'-cap and lack poly(A) tail or tRNA-like structure at the 3' end (Fuchs *et al.,* 2020). Except for velariviruses whose vectors are unknown, each closterovirus genus is transmitted by a unique group of hemipteran insects in a semi-persistent manner (Candresse and Fuchs, 2020).

Despite differences in the number and position of open reading frames (ORF), closteroviral genome organization largely remains conserved. At the 5' terminal region, closteroviral genomes possess a dual-gene module (ORF1a, ORF1b) encoding for replication-associated proteins with papain-like leader protease (L-Pro), methyltransferase (MTR), RNA helicase (HEL) and RNA-dependent RNA polymerase (RdRP) domains (Rubio *et al.,* 2013; Fuchs *et al.,* 2020). Downstream of the dual-gene module, closteroviral genomes contain a quintuple gene module coding for proteins like hydrophobic protein (p6), homologs of heat shock protein 70 (HSP70h) and heat shock protein 90 (HSP90h), minor coat protein (CPm) and coat protein (CP) (Rubio *et al.,* 2013; Orílio *et al.,* 2018). HSP70h is involved in viral cell-to-cell movement while CP and CPm are the

*Corresponding author. E-mail: vbaranwal2001@yahoo.com; phone: +011-25843474.

Abbreviations: CP = coat protein; CPm = minor coat protein; HEL = RNA helicase; HSP70h = heat shock protein 70 homolog; HSP90h = heat shock protein 90 homolog; ICTV = International Committee on Taxonomy of Viruses; L-Pro = papin-like protease; NCBI = National Center for Biotechnology Information; ORF(s) = open reading frame(s); RdRP = RNA-dependent RNA polymerase; TMH = Transmembrane helixes; TriPCV = Triticum polonicum closterovirus; TSA = Transcriptome Shotgun Assembly

coat proteins that encapsidate the viral genomic RNA (Rubio *et al.,* 2013). Few of the proteins encoded by ORFs near the 3' terminal region downstream of the quintuple gene module are probably involved in suppression of host RNA silencing (Kwon *et al.,* 2018).

The advent of Next Generation Sequencing (NGS) technologies has led to the increased deposition of raw reads and assembled contigs of plant transcriptomes in Sequence Read Archive (SRA) and Transcriptome Shotgun Assembly (TSA) databases of the National Center for Biotechnology Information (NCBI), respectively (Bejerman *et al.,* 2021; Sidharthan *et al.,* 2021). Since the total RNA isolated from plants for transcriptome studies may also contain viral RNA if they are infected at the time of sample collection, the assembled contigs of plant transcriptomes available in TSA database can be mined for novel viral sequences (Bejerman *et al.,* 2021). Putative novel viral sequences discovered from plant transcriptomes could be considered as *bona fide* viruses based on the consensus statement report by Simmonds *et al.* (2017) which suggests the inclusion of viruses discovered solely from metagenomic data in the International Committee on Taxonomy of Viruses (ICTV) taxonomy for comprehensive characterization of the global virome. In the present study, we explored the TSA sequences derived from plants for the discovery of novel closteroviral sequences and identified a putative novel closterovirus in *Triticum polonicum*.

Materials and Methods

Sequence analysis. Replicase sequence of beet yellows virus (NP733949.1) was used as query in tBLASTn analysis (e-value: 0.05; word size: 6; matrix: BLOSUM62) against TSA database and the search was restricted to *Viridiplantae* (taxid: 33090). The resulting hits that corresponded to the genome length of closteroviruses were manually examined.

ORF analysis. Presence of intact ORFs in the putative viral contigs was predicted using NCBI ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/). Molecular weight of proteins encoded by each ORF of putative coding-complete viral genome was estimated using the tool available at https://web.expasy.org/compute_pi/ while the viral motifs in encoded proteins were predicted using MOTIF search tool available at https://www.genome.jp/tools/motif/. Transmembrane helixes (TMH) in viral proteins were predicted using TMHMM-2.0 server (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0).

Protein sequence analysis. BLASTp analysis of viral protein sequences was performed against NCBI non-redundant database and percent sequence identity of the hit with maximum query coverage was indicated. Reference HSP70h amino acid sequences of different closteroviruses were retrieved from NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) and aligned along with the respective sequence of the recovered novel virus using MUSCLE option in MEGA7 v.7.0.26 (Kumar *et al.,* 2016). Aligned sequences were subjected to Neighbourhood-Joining (NJ) tree construction using Poisson model with 1000 bootstrap replicates in MEGA7. The family level taxonomic assignment of the recovered viral genome was confirmed using the GRAViTy pipeline (http://gravity.cvr.gla.ac.uk) (Aiewsakun and Simmonds, 2018).

Results and Discussion

Complete genome-coding sequence (TSA Acc. No.: GEDL01038631.1; 15, 366 nt long) of a putative novel closterovirus tentatively named as Triticum polonicum closterovirus (TriPCV) was identified in the transcriptome-assembled contigs of *T. polonicum* in tBLASTn searches (e-value: 7e-156; percent identity: 59.65; query coverage: 51%). The viral contig was derived from four transcriptome datasets (SRR2973592–SRR2973595) of *T. polonicum* leaves collected during heading stage (Biosample: SAMN04316857; Bioproject: PRJNA305056) (Wang *et al.,* 2016a,b). Like other closteroviruses (Rubio *et al.,* 2013; Fuchs *et al.,* 2020), a total of nine ORFs were predicted in TriPCV genome (Fig. 1). ORF1a (nt 97–7353) encodes a 266.0 kDa protein (2148 aa) with L-Pro (aa 439–519), MET (aa 577–930), closteroviral polyprotein central region (aa 1222–1319) and HEL (aa 2045–2304) domains. ORF1b (nt 97–8746), probably expressed by a +1 ribosomal frame

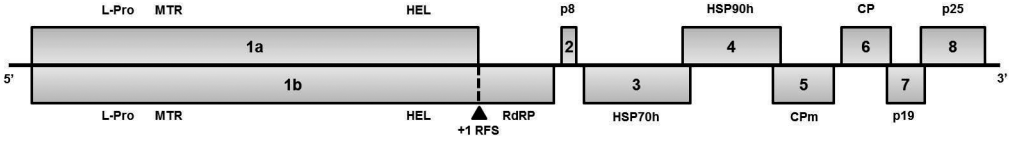

Fig. 1

Genome organization of Triticum polonicum closterovirus (TriPCV)

L-Pro - papin-like protease motif; MTR - methyltransferase motif; HEL - RNA helicase motif; RdRP - RNA-dependent RNA polymerase motif; HSP70h - heat shock protein 70 homolog; HSP90h - heat shock protein 90 homolog; CPm - minor coat protein; CP - coat protein, +1 RFS indicates the +1 ribosomal frameshift site.

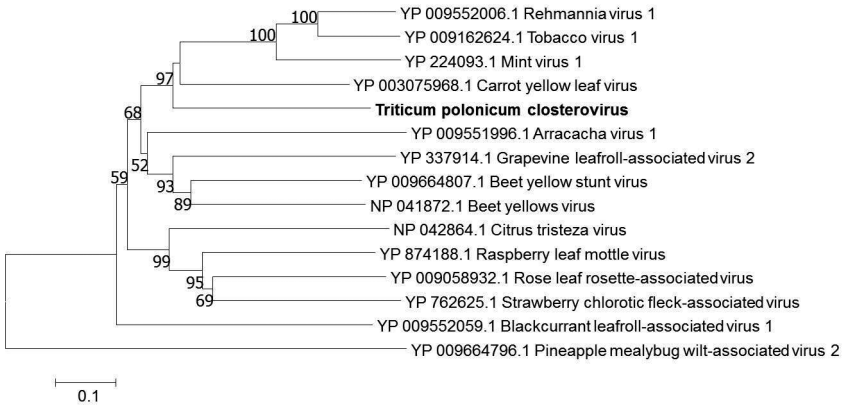

Fig. 2

Phylogenetic relationship of TriPCV to other closteroviruses based on HSP70h protein sequences
Pineapple mealybug wilt-associated virus 2 HSP70h sequence is used as an out-group. Only bootstrap values more than 50% are indicated. TriPCV is shown in bold.

shift at the stop codon of ORF1a (Kwon *et al.*, 2018), codes for a 320.0 kDa protein (2,882 aa) with an additional RdRP motif (aa 2389–2834) besides containing the motifs present in ORF1a encoded proteins. ORF2 (nt 8984–9199) encodes a small 7.9 kDa protein (71 aa) with a TMH (aa 48–70) for mediating cell-to-cell movement (Peremyslov *et al.*, 2004). ORF3 (nt 9393–11210) and ORF4 (nt 11122–12774) encodes for 65.4 kDa (605 aa) protein with HSP70 motif (aa 3–447) and 62.2 kDa (550 aa) protein with HSP90h motif (aa 8–485), respectively. 23.1 kDa CPm (210 aa) and 21.1 kDa CP (197 aa) encoded by ORF5 (nt 12704–13336) and ORF6 (nt 13440–14033), respectively contained a closteroviral CP domain (aa 22–210 in CPm; aa 12–195 in CP). ORF7 (nt 14030–14518 nt) codes for 18.6 kDa protein (162 aa) while the 24.7 kDa protein (223 aa) encoded by ORF8 (nt 14484–15155) contained RNA silencing suppressor N- (aa 1–87) and C-terminal (aa 110–176) domains (Supplementary Table S1).

BLASTp analysis showed that TriPCV encoded proteins, except for proteins encoded by ORF2, 7 and 8 that did not show significant similarity with any of the sequences at GenBank, shared 31.78–52.68% aa sequence similarities with the respective sequences of carrot closterovirus, grapevine leafroll-associated virus 2, raspberry leaf mottle virus, rehmannia virus 1 and tobacco virus 1 (Supplementary Table S1). Phylogenetic analysis showed that TriPCV fell in a distinct subclade to closteroviruses isolated from carrot, mint, rehmannia and tobacco (Fig. 2). Analysis of TriPCV genome in the GRAViTy pipeline showed that TriPCV could be a member of the family *Closteroviridae.*

Closteroviruses are generally reported from dicot plants (Fuchs *et al.*, 2020). To our knowledge, this is the first report of a putative member of the genus *Closterovirus* in a monocot host. Based on the <75% aa sequence identities of RdRP, HSP70h and CP proteins to other closteroviral sequences (the closterovirus species demarcation criteria (Kwon *et al.*, 2018), TriPCV can be regarded as a novel closterovirus. The recovered genome sequence will help in developing detection assays for TriPCV. Further studies are needed for understanding the biological properties, economic importance and distribution of TriPCV.

Acknowledgments. The authors are grateful to the Head, Division of Genetics and Tree Improvement, the GCR, the Director, IFB (ICFRE), Hyderabad, the Head, Division of Plant Pathology and the Director, ICAR-IARI, New Delhi for their support. The authors are also thankful to the original submitters of datasets/contigs used in the study.

Supplementary information is available in the online version of paper.

References

Aiewsakun P, Simmonds P (2018): The genomic underpinnings of eukaryotic virus taxonomy: creating a sequence-based framework for family-level virus classification. Microbiome 6, 38. https://doi.org/10.1186/s40168-018-0422-7

Bejerman N, Dietzgen RG, Debat H (2021): Illuminating the plant rhabdovirus landscape through metatranscriptomics data. Viruses 13, 1304. https://doi.org/10.3390/v13071304

Candresse T, Fuchs M (2020): Closteroviridae. eLS, Wiley, Chichester, UK, pp. 1–10. https://doi.org/10.1002/9780470015902.a0000747.pub4

Fuchs M, Bar-Joseph M, Candresse T, Maree HJ, Martelli GP, Melzer MJ, Menzel W, Minafra A, Sabanadzovic S, ICTV

Report Consortium (2020): ICTV virus taxonomy profile: Closteroviridae. J. Gen. Virol. 101, 364–365. https://doi.org/10.1099/jgv.0.001397

Kumar S, Stecher G, Tamura K (2016): MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054

Kwon SJ, Jin M, Cho IS, Yoon JY, Choi GS (2018): Identification of rehmannia virus 1, a novel putative member of the genus Closterovirus, from Rehmannia glutinosa. Arch. Virol. 163, 3383–3388. https://doi.org/10.1007/s00705-018-4014-4

Orílio AF, Blawid R, Costa GA, Gomes SS, Nagata T, Madeira NR, Inoue-Nagata AK, Resende RO (2018): High-throughput sequencing reveals a novel closterovirus in arracacha (Arracacia xanthorrhiza). Arch. Virol. 163, 2547–2550. https://doi.org/10.1007/s00705-018-3863-1

Peremyslov VV, Pan YW, Dolja VV (2004): Movement protein of a closterovirus is a type III integral transmembrane protein localized to the endoplasmic reticulum. J. Virol. 78, 3704–3709. https://doi.org/10.1128/JVI.78.7.3704-3709.2004

Rubio L, Guerri J, Moreno P (2013): Genetic variability and evolutionary dynamics of viruses of the family Closteroviridae. Front. Microbiol. 4, 151. https://doi.org/10.3389/fmicb.2013.00151

Sidharthan VK, Sharma SK, Baranwal VK (2021): The first near-complete genome sequence of large cardamom chirke virus mined from the transcriptome dataset of large cardamom. Plant Gene 28, 100324. https://doi.org/10.1016/j.plgene.2021.100324

Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, Carstens EB, Davison AJ, Delwart E, Gorbalenya AE, Harrach B, Hull R, King AMQ, Koonin EV, Krupovic M, Kuhn JH, Lefkowitz EJ, Nibert M, Orton R, Roossinck MJ, Sabanadzovic S, Sullivan MB, Suttle CA, Tesh RB, van der Vlugt R, Varsani A, Zerbini FM (2017): Virus taxonomy in the age of metagenomics. Nat. Rev. Microbiol. 15, 161–168. https://doi.org/10.1038/nrmicro.2016.177

Wang Y, Wang C, Wang X, Peng F, Wang R, Jiang Y, Zeng J, Fan X, Kang H, Sha L, Zhang H, Xiao X, Zhou Y (2016a): De Novo sequencing and characterization of the transcriptome of dwarf polish wheat (Triticum polonicum L.). Int. J. Genomics 2016, 5781412. https://doi.org/10.1155/2016/5781412

Wang Y, Xiao X, Wang X, Zeng J, Kang H, Fan X, Sha L, Zhang H, Zhou Y (2016b): RNA-Seq and iTRAQ reveal the dwarfing mechanism of dwarf polish wheat (Triticum polonicum L.). Int. J. Biol Sci. 12, 653. https://doi.org/10.7150/ijbs.14577

Acta virologica **66**: 165 – 168, 2022 doi: 10.4149/av_2022_207

SUPPLEMENTARY INFORMATION

Dwarf polish wheat hosts a novel closterovirus: Revelation by transcriptome data-mining

Venkidusamy Kavi Sidharthan[1], Virendra Kumar Baranwal[2*]

[1]Division of Genetics and Tree Improvement, Institute of Forest Biodiversity (ICFRE), Hyderabad, India; [2]Division of Plant Pathology, ICAR-Indian Agricultural Research Institute, New Delhi, India

Received December 1, 2021; revised February 14, 2022; accepted April 4, 2022

[*]Corresponding author. E-mail: vbaranwal2001@yahoo.com; phone: +011-25843474.

Table S1. Annotation of Triticum polonicum closterovirus genome recovered in the study

	Start (nt)	Stop (nt)	nt length	aa length	Molecular weight (kDa)	Predicted viral motifs against Pfam database	Note	BLASTp analysis				
								Greatest % identity	e value	Maximum query coverage (%)	Subject	Acc. No.
5'UTR	1	96	96	–	–	–	–	–	–	–	–	–
ORF1a	97	7353	7257	2418	266.08641	PF05533, Beet yellows virus-type papain-like endopeptidase C42; PF01660, Viral methyltransferase; PF17646, Closterovirus 1a polyprotein central region; PF01443, Viral (Superfamily 1) RNA helicase	–	31.78	1.00E-168	76	1a [Tobacco virus 1]	YP_009162621.1
ORF1b	97	8746	8649	2882	320.06705	PF05533, Beet yellows virus-type papain-like endopeptidase C42; PF01660, Viral methyltransferase; PF17646, Closterovirus 1a polyprotein central region; PF01443, Viral (Superfamily 1) RNA helicase; PF00978, RNA dependent RNA polymerase; PF00998, Viral RNA dependent RNA polymerase	+1 ribosome frame shift (join 97-7350-352-8746)	52.68	0	76	330 kDa protein [Rehmannia virus 1]	YP_009552003.1
ORF2	8984	9199	216	71	7.91486	Nil	–	NS	NS	NS	NS	NS
ORF3	9393	11210	1818	605	65.42858	PF00012, Hsp70 protein	–	51.64	0	100	ORF4 [Carrot closterovirus]	AHA85412.1
ORF4	11122	12774	1653	550	62.24022	PF03225, Viral heat shock protein Hsp90 homologue	–	36.31	6.00E-115	96	heat shock protein 90 [Grapevine leafroll-associated virus 2]	ABK60004.1
ORF5	12704	13336	633	210	23.09025	PF01785, Closterovirus coat protein	–	32.71	1.00E-23	99	coat protein minor [Raspberry leaf mottle virus]	CAY54348.1, YP_874190.1
ORF6	13440	14033	594	197	21.14507	PF01785, Closterovirus coat protein	–	32.98	3.00E-16	94	major capsid protein [Grapevine leafroll-associated virus 2]	AHW79716.1
ORF7	14030	14518	489	162	18.58796	Nil	–	NS	NS	NS	NS	NS
ORF8	14484	15155	672	223	24.65958	PF11757, Suppressor of RNA silencing P21-like N-terminal domain, PF11479, RNA silencing suppressor P21 C-terminal domain	–	NS	NS	NS	NS	NS
3'UTR	15156	15366	211	–	–	–	–	–	–	–	–	–

NS = not significant.