jmb

Genome Sequence of Spinach Cryptic Virus 1, a New Member of the Genus *Alphapartitivirus* (Family *Partitiviridae*), Identified in Spinach ^S

Dongbin Park and Yoonsoo Hahn*

Department of Life Science, Research Center for Biomolecules and Biosystems, Chung-Ang University, Seoul 06974, Republic of Korea

A distinct double-stranded RNA (dsRNA) cryptic virus, named spinach cryptic virus 1 (SpCV1), was identified from spinach transcriptome datasets. The SpCV1 genome has two dsRNA genome segments. The larger dsRNA1 has an open reading frame for a conserved RNA-dependent RNA polymerase (RdRp). The smaller dsRNA2 encodes a putative coat protein (CP). The sequence identity of SpCV1 RdRp and CP to the closest cryptic virus is 81% and 60%, respectively. Phylogenetic analysis indicates that SpCV1 is a novel member of the genus *Alphapartitivirus* (family *Partitiviridae*).

Keywords: Spinach cryptic virus 1, *Alphapartitivirus*, *Partitiviridae*, spinach, transcriptome

Partitiviridae is a taxonomic family of double-stranded RNA (dsRNA) viruses that infect plants (mostly angiosperms), fungi (ascomycetes and basidiomycetes), and an apicomplexan protist of the genus *Cryptosporidium* [1]. Partitiviruses are classified into five genera on the basis of phylogenetic analysis: *Alphapartitivirus*, *Betapartitivirus*, *Gammapartitivirus*, *Deltapartitivirus*, and *Cryspovirus*. Most partitiviruses have segmented genomes of two linear 1–3-kb-long dsRNAs, known as dsRNA1 and dsRNA2, respectively. dsRNA1 encodes the virion-associated RNA-dependent RNA polymerase (RdRp); dsRNA2 encodes a viral coat protein (CP). These RNA molecules are encapsidated in separate virion particles, which must simultaneously infect the host for propagation [2, 3]. The RNA molecules of these viruses can be isolated along with the host RNA, and their genome sequences can be identified in the host transcriptome or RNA-Seq data collected by next-generation sequencing [4]. In this study, the spinach transcriptome data were analyzed to identify the spinach cryptic virus 1 (SpCV1), a novel member of the genus *Alphapartitivirus* (family *Partitiviridae*).

Spinach transcriptome data collected from the seedlings of two wild spinach species (*Spinacia tetrandra* and *S. turkestanica*) and one cultivated spinach species (*S. oleracea*) were downloaded from the Sequence Read Archive of the National Center for Biotechnology Information (NCBI) [5]. Their accession numbers were as follows: SRR1766329, SRR1766330, and SRR1766331 (*S. tetrandra*); SRR1766332, SRR1766333, and SRR1766334 (*S. turkestanica*); and SRR1766310, SRR1766311, and SRR1766312 (*S. oleracea*). Each of these transcriptome data was separately assembled into contigs using the SPAdes Genome Assembler (ver. 3.7.1) [6].

The complete genome sequences of RNA viruses were downloaded from the NCBI Nucleotide database and converted to a BLAST-searchable database. BLASTN search was performed to identify potential viral RNA sequences in the assembled spinach transcriptome contigs [7]. The E value cut-off for the match was set at 1e-5. Some of the contigs were identified to be virus-related sequences. Contigs assembled from the two spinach transcriptome datasets, namely, SRR1766311 from *S. oleracea* and SRR1766329 from *S. tetrandra*, showed sequence similarity to full-length cryptic virus genome sequences, and were chosen for further in-depth analysis.

From each transcriptome dataset (SRR1766311 and SRR1766329), two sequence contigs were deduced: one showed sequence similarity to the cryptic virus dsRNA1 sequences, and the other to dsRNA2 sequences. Therefore, these contigs were considered to be genomes of putative cryptic viruses, which were designated as SpCV1. To examine potential heterogeneity in the assembled genome sequences, SpCV1-associated RNA-Seq reads were mapped back to the corresponding contigs using BWA, and sequence variations were called using SAMtools and BCFtools [8, 9]. There was no sequence variation in the four sequence contigs, suggesting that each assembled genomic segment

was derived from clones propagated from one or a few viruses with the same sequence. The SpCV1 genome sequences were deposited in the NCBI Nucleotide database under the accession numbers KX784754 (dsRNA1 from SRR1766311), KX784755 (dsRNA2 from SRR1766311), KX784756 (dsRNA1 from SRR1766329), and KX784757 (dsRNA2 from SRR1766329).

The SpCV1 dsRNA1 sequences were 1,966 and 1,961 bp long and encoded a 616-amino-acid-long RdRp (Fig. S1). The dsRNA2 sequences were 1,762 and 1,753 bp long and encoded a 488-amino-acid-long CP. The two dsRNA1 sequences showed only six nucleotide sequence mismatches,

Table 1. Characteristics of the *Alphapartitivirus* species used for comparison of spinach cryptic virus 1 (SpCV1).

Virus	Abbr.	dsRNA1	dsRNA2	RdRp	Identity[a]	CP	Identity
Carrot cryptic virus	CaCV	FJ550604.1	FJ550605.1	ACL93278.1	505/616 (81%)	ACL93279.1	260/446 (58%)
Beet cryptic virus 1	BCV1	NC_011556.1	NC_011557.1	YP_002308574.1	504/616 (81%)	YP_002308575.1	294/488 (60%)
Dill cryptic virus 1	DCV1	NC_022614.1	NC_022615.1	YP_008719880.1	504/616 (81%)	YP_008719881.1	278/488 (56%)
Red clover cryptic virus 1	RcCV1	NC_022616.1	NC_022617.1	YP_008719882.1	496/616 (80%)	YP_008719883.1	285/487 (58%)
White clover cryptic virus 1	WcCV1	NC_006275.1	NC_006276.1	YP_086754.1	495/616 (80%)	YP_086755.1	277/488 (56%)
Vicia cryptic virus	VCV	NC_007241.1	NC_007242.1	YP_272124.1	480/616 (77%)	YP_272125.1	267/488 (54%)
Heterobasidion partitivirus 4	HetPV4	HQ541325.1	NA[b]	ADV15443.1	388/622 (62%)	NA	
Rhizoctonia solani dsRNA virus 2	RsRV2	NC_023684.1	NC_023685.1	YP_009011230.1	364/618 (58%)	YP_009011231.1	126/466 (27%)
Heterobasidion partitivirus 5	HetPV5	HQ541326.1	NA	ADV15444.1	370/619 (59%)	NA	
Diuris pedunculata cryptic virus	DpCV	JX156424.1	JX891460.1	AFQ95555.1	376/620 (60%)	AFY23215.1	108/372 (29%)
Rhizoctonia fumigata partitivirus	RfPV	KM668042.1	KM668043.1	AJE25830.1	380/618 (61%)	AJE25831.1	136/517 (26%)
Amasya cherry disease-associated mycovirus	ACDMV	NC_006441.1	NC_006440.1	YP_138537.1	351/619 (56%)	YP_138536.1	132/488 (27%)
Cherry chlorotic rusty spot-associated partitivirus	CCRSPV	AJ781401.1	AJ781402.1	CAH03668.1	351/619 (56%)	CAH03669.1	127/492 (25%)
Heterobasidion partitivirus 9	HetPV9	JN606085.2	NA	AEX87909.2	341/622 (54%)	NA	
Heterobasidion partitivirus 1	HetPV1	HQ541329.1	HQ541324.1	ADV15447.1	330/622 (53%)	ADV15442.1	121/492 (24%)
Heterobasidion partitivirus 17	HetPV17	KJ873060.1	NA	AIF33767.1	324/623 (52%)	NA	

The NCBI accession numbers of dsRNA1, dsRNA2, RNA-dependent RNA polymerase (RdRp), and viral coat protein (CP) are presented in the table.

[a]The amino acid sequence identity of SpCV1 RdRp and CP with the corresponding proteins of other viruses calculated using BLASTP are shown in a format of identical residues/aligned length (identity %).

[b]NA, not available.

but encoded identical RdRp proteins. The two dsRNA2 sequences showed 13 nucleotide mismatches and encoded CPs with only one difference in the amino acid sequence. Therefore, these two viruses seemed to have very recently descended from a common ancestor. Because the two viral genome sequences were almost identical, subsequent analysis was performed using the SpCV1 genome sequence derived from SRR1766329.

The SpCV1 RdRp encoded by dsRNA1 showed up to 81% amino acid sequence identity to several plant-infecting cryptic viruses, including carrot cryptic virus (CaCV), beet cryptic virus 1 (BCV1), dill cryptic virus 1 (DCV1), red clover cryptic virus 1 (RcCV1), white clover cryptic virus 1 (WcCV1), Vicia cryptic virus (VCV), and Diuris pedunculata cryptic virus (DCV) (Table 1). It also showed 52–62% identity with fungus-infecting partitiviruses, including Heterobasidion partitivirus 5 (HetPV5), Rhizoctonia fumigata partitivirus (RfPV), Amasya cherry disease-associated mycovirus (ACDMV), cherry chlorotic rusty spot-associated partitivirus (CCRSPV), Heterobasidion partitivirus 9 (HetPV9), Heterobasidion partitivirus 1 (HetPV1), and Heterobasidion partitivirus 17 (HetPV17). These viruses belong to the genus *Alphapartitivirus* of the family *Partitiviridae*, indicating that SpCV1 is a new member of *Alphapartitivirus* [1].

Multiple sequence alignment of RdRp sequences of SpCV1 and related *Alphapartitivirus* species generated using the MUSCLE program (ver. 3.8.31) revealed highly conserved positions (Fig. S2) [10]. Phylogenetic analysis was performed using the neighbor-joining method implemented in MEGA7 [11]. The phylogenetic tree thus generated confirmed our idea that SpCV1 is a new member of the genus *Alphapartitivirus* (Fig. 1A). SpCV1 formed a strong clade with six other cryptic viruses, including CaCV, DCV, BCV1, VCV, WcCV1, and RcCV1. These seven cryptic viruses were isolated from flowering plants: SpCV1 from spinach; CaCV from carrot; DCV from dill; BCV1 from beet; VCV from a *Vicia* species; WcCV1 from white clover; and RcCV1 from red clover. The plants associated with these seven cryptic viruses can be classified into three distantly related plant orders of eudicots: Caryophyllales (spinach and beet), Apiales (carrot and dill), and Fabales (*Vicia*, white clover, and red clover). Interestingly, BCV1 that infects beet is not a sister taxon of SpCV1, which is found in spinach. Instead, BCV1 forms a subclade with CaCV1 and DCV1, which were isolated from carrot and dill, respectively. The discordant phylogenetic relationship between viruses and their hosts, along with a wide diversity of host plants, suggests that these cryptic viruses might have been laterally transmitted by pathogenic fungal
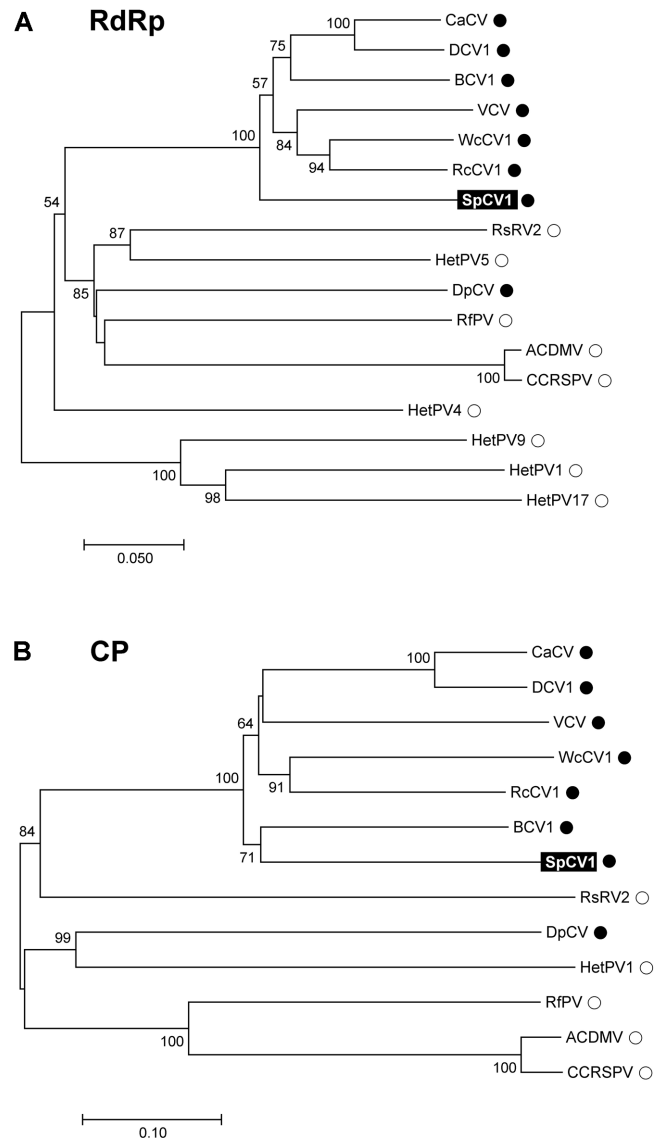
Fig. 1. Phylogenetic trees of RNA-dependent RNA polymerase (RdRp) (**A**) and viral coat protein (CP) (**B**) of *Alphapartitivirus* species related to spinach cryptic virus 1 (SpCV1).

The phylogenetic relationship among viruses was generated using the neighbor-joining method implemented in MEGA7. Bootstrap percentages are shown at the nodes supported in a minimum of 50% of 1,000 replicates. Viruses that were isolated from plants and fungi are marked by a closed circle (●) and an open circle (○), respectively. See Table 1 for the full names of viruses and accession numbers.

vectors, instead of supporting the notion that viruses and hosts have co-evolved [12].

The SpCV1 CP encoded by dsRNA2 also showed higher sequence identity to that of plant-infecting cryptic viruses than to that of mycoviruses (Table 1). The sequence identity of SpCV1 CP to that of CaCV, BCV1, DCV1, RcCV1,

WcCV1, and VCV ranged from 54% to 60%, whereas mycovirus CPs showed 24–27% sequence identity. Multiple sequence alignment of CPs revealed conserved residues among related *Alphapartitivirus* species (Fig. S3). However, the overall conservation level of CP was lower than that of RdRp.

Phylogenetic analysis of CPs also supported a strong clade of the seven cryptic viruses that were isolated from plants (Fig. 1B). Interestingly, CPs revealed a different topology to that of RdRp proteins. For example, the tree generated using CPs placed BCV1 as a sister taxon of SpCV1, instead of CaCV1 and DCV1. This discrepancy between RdRp and CP can be attributed to their different evolutionary rates across viruses. Another possible explanation is the mechanism of reassortment by which two different viruses exchange their genomic segments during co-infection [13].

The 5′-untranslated regions of SpCV1 dsRNA1 and dsRNA2 exhibit a strong sequence similarity to each other, although they encoded completely different proteins (Fig. S4). The sequences are also conserved in closely related cryptic viruses, including CaCV, BCV1, DCV1, WcCV1, and VCV. These terminal sequence domains might be important for recognition by the viral RdRp during viral RNA replication and/or may act as a packaging signal during assembly [14, 15].

In conclusion, the genomic sequence of SpCV1 has been identified from spinach transcriptome datasets. The sequence identity of SpCV1 RdRp and CP to the closest cryptic virus is 81% and 60%, respectively, which indicates that SpCV1 is a distinct cryptic virus. Phylogenetic analysis confirmed that SpCV1 is a new member of the genus *Alphapartitivirus* (family *Partitiviridae*).

Acknowledgments

References

1. Nibert ML, Ghabrial SA, Maiss E, Lesker T, Vainio EJ, Jiang D, Suzuki N. 2014. Taxonomic reorganization of family *Partitiviridae* and other recent progress in partitivirus research. *Virus Res.* **188:** 128-141.

2. Willenborg J, Menzel W, Vetten HJ, Maiss E. 2009. Molecular characterization of two alphacryptovirus dsRNAs isolated from *Daucus carota. Arch. Virol.* **154:** 541-543.

3. Szego A, Enunlu N, Deshmukh SD, Veliceasa D, Hunyadi-Gulyas E, Kuhne T, *et al.* 2010. The genome of Beet cryptic virus 1 shows high homology to certain cryptoviruses present in phylogenetically distant hosts. *Virus Genes* **40:** 267-276.

4. Kim DS, Jung JY, Wang Y, Oh HJ, Choi D, Jeon CO, Hahn Y. 2014. Plant RNA virus sequences identified in kimchi by microbial metatranscriptome analysis. *J. Microbiol. Biotechnol.* **24:** 979-986.

5. Xu C, Jiao C, Zheng Y, Sun H, Liu W, Cai X, *et al.* 2015. De novo and comparative transcriptome analysis of cultivated and wild spinach. *Sci. Rep.* **5:** 17706.

6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19:** 455-477.

7. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389-3402.

8. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754-1760.

9. Li H. 2011. Improving SNP discovery by base alignment quality. *Bioinformatics* **27:** 1157-1158.

10. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32:** 1792-1797.

11. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33:** 1870-1874.

12. Lubicz JV, Rush CM, Payton M, Colberg T. 2007. Beet necrotic yellow vein virus accumulates inside resting spores and zoosporangia of its vector *Polymyxa betae* BNYVV infects *P. betae*. *Virol. J.* **4:** 37.

13. McDonald SM, Nelson MI, Turner PE, Patton JT. 2016. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nat. Rev. Microbiol.* **14:** 448-460.

14. Li L, Tian Q, Du Z, Duns GJ, Chen J. 2009. A novel double-stranded RNA virus detected in *Primula malacoides* is a plant-isolated partitivirus closely related to partitivirus infecting fungal species. *Arch. Virol.* **154:** 565-572.

15. Lesker T, Rabenstein F, Maiss E. 2013. Molecular characterization of five betacryptoviruses infecting four clover species and dill. *Arch. Virol.* **158:** 1943-1952.