Acta virologica **65**: 373 – 380, 2021 doi: 10.4149/av_2021_404

Identification of the genome sequence of Zostera associated varicosavirus 1, a novel negative-sense RNA virus, in the common eelgrass (*Zostera marina*) transcriptome

Chaerim Shin, Dongjin Choi, Yoonsoo Hahn*

Department of Life Science, Chung-Ang University, Seoul, South Korea

Received May 12, 2021; accepted June 3, 2021

Summary. – Varicosaviruses (the genus *Varicosavirus*) are bipartite, negative-sense, single-stranded RNA viruses that infect plants. We analyzed a transcriptome dataset isolated from the common eelgrass (*Zostera marina*) and identified a novel varicosavirus named Zostera associated varicosavirus 1 (ZaVV1). The ZaVV1 genome consists of two genomic segments: RNA1 (6,632-nt) has an open reading frame (ORF) encoding a large multi-functional polymerase protein (L), while RNA2 (4,304-nt) has four ORFs: one for a nucleocapsid protein and three for proteins with unknown functions (P2, P3, and P4). Sequence comparison and phylogenetic analysis using L proteins showed that ZaVV1 is a novel member of the genus *Varicosavirus* of the family *Rhabdoviridae*. The conserved regulatory elements involved in transcription termination/polyadenylation and transcription initiation were identified in the ZaVV1 gene-junction regions with the consensus sequence 3'-UAUUAUUCUUUUUGCUCU-5' (in the negative-sense genome). The ZaVV1 genome sequence may be useful for studying the phylogenetic relationships of varicosaviruses and genome evolution of rhabdoviruses.

Keywords: Zostera associated varicosavirus 1; Varicosavirus; Rhabdoviridae; common eelgrass; Zostera marina

Introduction

Varicosaviruses (the genus *Varicosavirus*) are negative-sense RNA viruses of the family *Rhabdoviridae*, order *Mononegavirales* (Kormelink *et al*., 2011; Dietzgen *et al*., 2017; Walker *et al*., 2018). Virions of varicosaviruses are filamentous and lack an envelope, although other rhabdoviruses (the family *Rhabdoviridae*) usually have a bullet-shaped virion particle enveloped by a lipid membrane. The family *Rhabdoviridae*, the members of which infect various animals and plants, comprises 30 genera officially approved by the International Committee on Taxonomy of Viruses (ICTV) (https://talk.ictvonline.org, last accessed on February 13, 2021). Plant-infecting rhabdoviruses are classified into six genera: *Alphanucleorhabdovirus*, *Betanucleorhabdovirus*, *Cytorhabdovirus*, *Dichorhavirus*, *Gammanucleorhabdovirus*, and *Varicosavirus*. Their genomes are either unsegmented (*Alphanucleorhabdovirus*, *Betanucleorhabdovirus*, *Cytorhabdovirus*, and *Gammanucleorhabdovirus*) or bi-segmented (*Dichorhavirus* and *Varicosavirus*) (Dietzgen *et al*., 2014, 2017; Walker *et al*., 2018).

The genus *Varicosavirus* currently has one recognized species, the lettuce big-vein associated varicosavirus, which is represented by lettuce big-vein associated virus (LBVaV), and two tentative species: red clover associated varicosavirus (RCaVV) and Alopecurus myosuroides varicosavirus 1 (AMVV1) (also known as black grass varicosavirus-like virus) (Sasaya *et al*., 2002, 2004; Sab-

*Corresponding author. E-mail: hahny@cau.ac.kr; phone: +82-2-820-5812.

Abbreviations: AMVV1 = Alopecurus myosuroides varicosavirus 1; LBVaV = lettuce big-vein associated virus; ORF = open reading frame; RCaVV = red clover associated varicosavirus; RdRp = RNA-dependent RNA polymerase; TTP = transcription termination/polyadenylation; TI = transcription initiation; ZaVV1 = Zostera associated varicosavirus 1

badin *et al.*, 2017; Koloniuk *et al.*, 2018b). Varicosaviruses are reported to be transmitted by soil-borne chytrid fungi (Whitfield *et al.*, 2018).

Varicosaviruses have linear, bipartite, negative-sense, single-stranded RNA genomes, which possess two segments called RNA1 and RNA2. The RNA1 segments of three known varicosaviruses commonly encode a large multifunctional protein (L) (Sasaya *et al.*, 2002; Sabbadin *et al.*, 2017; Koloniuk *et al.*, 2018b). The L protein, which contains an RNA-dependent RNA polymerase (RdRp) domain and mRNA capping domain, is a polymerase involved in the genome replication and mRNA transcription of rhabdoviruses (Jackson *et al.*, 2005; Walker *et al.*, 2018). The LBVaV RNA1 segment was predicted to have an additional small ORF with unknown function in front of the L protein ORF, whereas RCaVV and AMVV1 have no additional ORFs in their RNA1 segments (Sasaya *et al.*, 2002).

The RNA2 segments of known varicosaviruses have three to five ORFs (Sasaya *et al.*, 2004; Sabbadin *et al.*, 2017; Koloniuk *et al.*, 2018a). The first ORF encodes a nucleocapsid protein (N), also known as coat protein, which encapsidates the viral genomic RNAs and is commonly shared with other rhabdoviruses (Jackson *et al.*, 2005; Walker *et al.*, 2018). Two or more additional ORFs with unknown functions were predicted in the RNA2 segment. These additional ORFs were named protein 2 (P2), protein 3 (P3), protein 4 (P4), and protein 5 (P5) in LBVaV; 47-kDa protein (47K) and 20-kDa protein (20K) in AMVV1; and P2 and P3 in RCaVV. There were no meaningful amino acid (aa) sequence similarities among these ORFs, except between the RCaVV P2 and AMVV1 47K proteins. Although LBVaV and RCaVV have proteins with the same name, P2 and P3 proteins have no notable sequence similarities.

The negative-sense RNA genomes of rhabdoviruses are used as templates for the transcription of viral gene mRNAs that are translated into proteins (Jackson *et al.*, 2005; Walker *et al.*, 2011). There are conserved regulatory elements in the gene-junction regions of rhabdoviruses, which are responsible for transcription termination/polyadenylation (TTP) of the preceding gene and transcription initiation (TI) of the following gene (Jackson *et al.*, 2005; Goh *et al.*, 2020; Zhou *et al.*, 2020). The TTP and TI elements are separated by an untranscribed spacer that is one or several nucleotides (nt) in length (Goh *et al.*, 2020; Orfanidou *et al.*, 2020; Zhou *et al.*, 2020). The gene-junction region consensus sequence of LBVaV, RCaVV, and AMVV1 was reported as 3'-NAUNNNNNUUUUU–G–CUCU-5', where the dashes separate three elements: TTP (3'-NAUNNNNNUUUUU-5'), untranscribed spacer (3'-G-5'), and TI (3'-CUCU-5') (Koloniuk *et al.*, 2018b).

RNA-seq data obtained from plant tissues often contain viral sequences derived from latently infected RNA viruses. Contigs assembled from these viral sequences can be detected by bioinformatics analysis (Nibert *et al.*, 2016; Lee *et al.*, 2019; Bejerman *et al.*, 2020). Previously, we reported numerous novel RNA virus genome sequences identified in plant transcriptome datasets (Goh and Hahn, 2019; Park *et al.*, 2019, 2020, 2021; Goh *et al.*, 2021). The common eelgrass (*Zostera marina*) is an aquatic plant that predominantly grows in temperate coastal waters and plays key roles in coastal ecosystems (Dahl *et al.*, 2016; Reynolds *et al.*, 2016). In the present study, we analyzed a common eelgrass transcriptome dataset and identified the genome sequence of a novel virus belonging to the genus *Varicosavirus* (Tan *et al.*, 2020).

Materials and Methods

The common eelgrass transcriptome dataset (six sequencing runs totaling 41.9 gigabase pairs) was downloaded from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) (Tan *et al.*, 2020). The SRA accession numbers were SRR11192591, SRR11192592, SRR11192593, SRR11192594, SRR11192595, and SRR11192596. RNA-seq reads were filtered to collect high-quality sequences using the sickle program (version 1.33; https://github.com/najoshi/sickle) with the parameter "-q 30 -l 55." High-quality reads from all six runs were pooled into a single dataset and subjected to *de novo* contig assembly using the SPAdes program (version 3.14.1; http://cab.spbu.ru/software/spades) (Bushmanova *et al.*, 2019).

Tentative virus-derived genome contigs were detected by the sequence comparison of the common eelgrass transcriptome contigs and known viral RdRp sequences (a total of 2,565 sequences), which were obtained from the Pfam database (release 33.1; https://pfam.xfam.org). The Pfam accession numbers for the viral RdRp domain sequences were PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03035, PF03431, PF04196, PF04197, PF05788, PF05919, PF06317, PF07925, PF08467, PF08716, PF08717, PF12426, and PF17501. The DIAMOND program (version 2.0.4.142; https://github.com/bbuchfink/diamond) in blastx mode was used for sequence similarity searches (Buchfink *et al.*, 2015).

Protein-coding ORFs were predicted using the ORFfinder web server (https://www.ncbi.nlm.nih.gov/orffinder). The pepstats program of the EMBOSS package (version 6.6.0.0; http://emboss.open-bio.org) was used to calculate the molecular weights of the predicted proteins. The InterPro web server (version 5.50-84.0; https://www.ebi.ac.uk/interpro) was used to predict the protein domains.

The NCBI BLAST web server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to collect protein sequences. Pairwise identities of the protein sequences were calculated using the EMBOSS needle program. Multiple sequence alignment was performed using the MAFFT program (version 7.475; https://mafft.cbrc.jp/alignment/software), with the parameter "--auto" (Nakamura *et al.*, 2018). The aligned sequences were filtered using the trimAl

Table 1. Sequence comparison of the L proteins of ZaVV1 and representative viruses from the family *Rhabdoviridae*

No.	Genus	Virus	Acronym	NCBI[a]	Identity[b]
1	*Varicosavirus*	Red clover associated varicosavirus	RCaVV	AUD57853.1	821/2085 (39.4%)
2		Alopecurus myosuroides varicosavirus 1	AMVV1	YP_009130620.1	800/2124 (37.7%)
3		Lettuce big-vein associated virus	LBVaV	YP_002308576.1	795/2121 (37.5%)
4	*Cytorhabdovirus*	Northern cereal mosaic virus	NCMV	NP_597914.1	516/2291 (22.5%)
5		Barley yellow striate mosaic virus	BYSMV	YP_009177231.1	545/2284 (23.9%)
6		Rice stripe mosaic virus	RSMV	YP_009553369.1	543/2296 (23.6%)
7		Tomato yellow mottle-associated virus	TYMaV	YP_009352236.1	532/2324 (22.9%)
8		Lettuce necrotic yellows virus	LNYV	YP_425092.1	571/2273 (25.1%)
9		Trichosanthes associated rhabdovirus 1	TrARV1	DAC81998.1	531/2351 (22.6%)
10	*Alphanucleorhabdovirus*	Eggplant mottled dwarf virus	EMDV	YP_009094358.1	510/2308 (22.1%)
11		Physostegia chlorotic mottle virus	PhCMoV	AOT55662.1	504/2256 (22.3%)
12		Potato yellow dwarf virus	PYDV	YP_004927971.1	499/2255 (22.1%)
13		Peach virus 1	PeV1	QIQ60850.1	522/2252 (23.2%)
14		Wheat yellow striate virus	WYSV	AVV48080.1	506/2241 (22.6%)
15		Rice yellow stunt virus	RYSV	NP_620502.1	514/2305 (22.3%)
16		Morogoro maize-associated virus	MMaV	AZP55481.1	514/2268 (22.7%)
17		Taro vein chlorosis virus	TaVCV	YP_224083.1	500/2249 (22.2%)
18		Maize Iranian mosaic virus	MIMV	YP_009444713.1	531/2283 (23.3%)
19		Maize mosaic virus	MMV	YP_052855.1	528/2232 (23.7%)
20	*Gammanucleorhabdovirus*	Maize fine streak virus	MFSV	YP_052849.1	552/2258 (24.4%)
21	*Betanucleorhabdovirus*	Datura yellow vein virus	DYVV	YP_009176977.1	556/2435 (22.8%)
22		Zhuye pepper nucleorhabdovirus	ZPNRV	AZN18347.1	539/2436 (22.1%)
23		Sowthistle yellow vein virus	SYVV	QJQ80127.1	550/2373 (23.2%)
24		Cardamom vein clearing virus	CdVCV	QJZ27984.1	519/2370 (21.9%)
25		Black currant-associated rhabdovirus	BCaRV	AUW36419.1	530/2437 (21.7%)
26		Sonchus yellow net virus	SYNV	NP_042286.1	555/2338 (23.7%)
27		Alfalfa-associated nucleorhabdovirus	AaNV	QAB45076.1	523/2321 (22.5%)
28	*Dichorhavirus*	Orchid fleck virus	OFV	YP_001294929.1	509/2230 (22.8%)
29		Clerodendrum chlorotic spot virus	ClCSV	YP_009666993.1	495/2210 (22.4%)
30		Coffee ringspot virus	CoRSV	YP_009507905.1	505/2241 (22.5%)
31	*Vesiculovirus*	Vesicular stomatitis Indiana virus	VSIV	NP_041716.1	510/2413 (21.1%)

[a]NCBI Acc. Nos. for the L proteins. [b]Protein sequence identity to the ZaVV1 L protein in the form of "number of identical residues/aligned length (percent identity)" calculated using the EMBOSS needle program.

software (version 1.4.rev22; http://trimal.cgenomics.org) with the parameter "-gappyout" (Capella-Gutierrez *et al.*, 2009). A maximum-likelihood phylogenetic tree was inferred using the IQ-TREE program (version 2.1.2; http://www.iqtree.org), with the parameter "-B 1000" (Minh *et al.*, 2020).

High-quality RNA-seq reads were mapped to viral genome sequences using the URMAP program (version 1.0.1480; https://drive5.com/urmap), and sequencing depth was determined using the samtools program (version 1.11; http://www.htslib.org).

Putative regulatory elements conserved in gene-junction regions were detected using the MEME web server (version 5.3.2; http://meme-suite.org/tools/meme) (Bailey and Elkan, 1994). The sequence logo representation of the gene-junction region sequences was created using the WebLogo server (version 3; http://weblogo.threeplusone.com).

Results and Discussion

RNA-seq reads obtained from the leaf tissues of the common eelgrass were assembled to generate contig sequences (Tan *et al.*, 2020). Putative virus-derived contigs

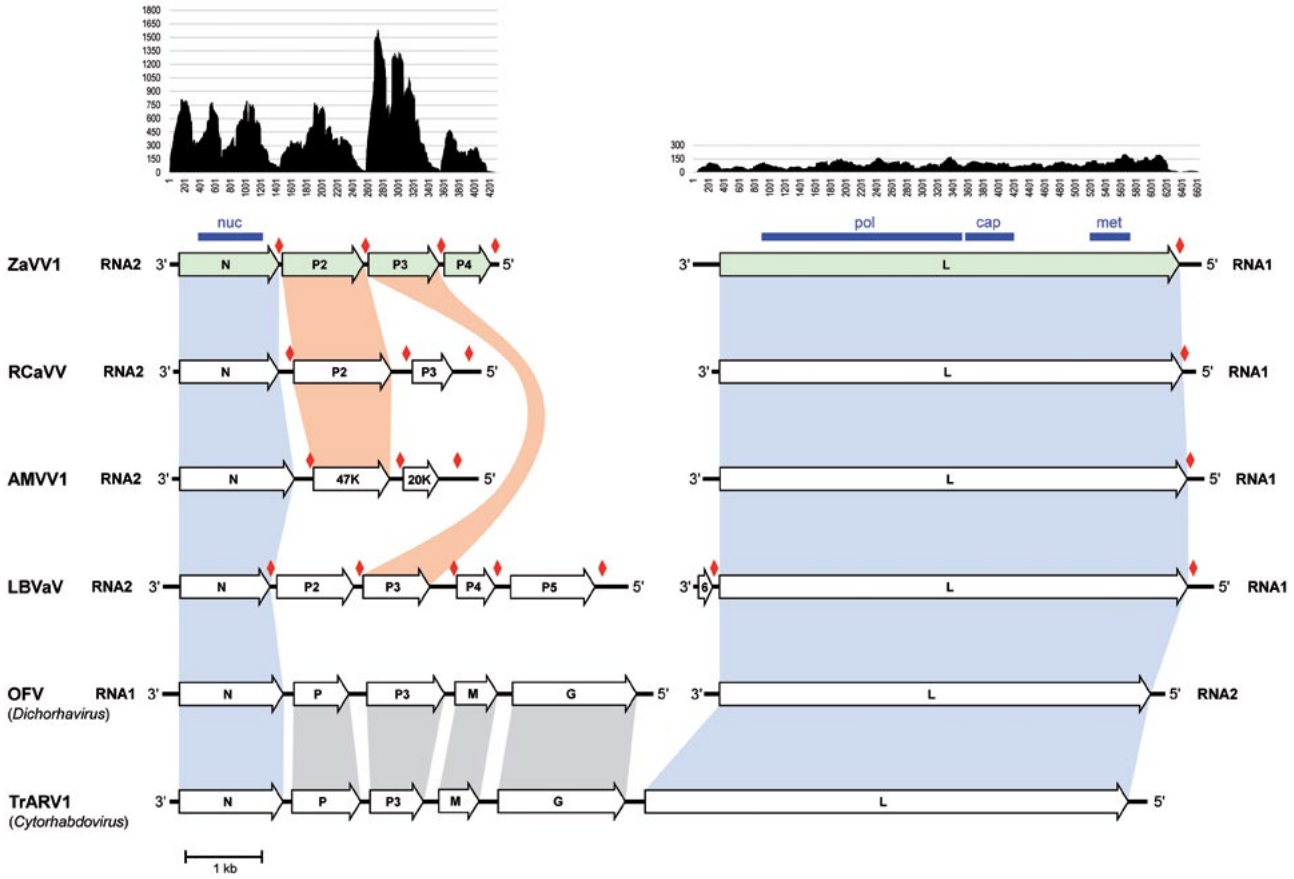

Fig. 1

Comparison of the genome organizations of ZaVV1 and related viruses

Genome organizations of ZaVV1, three known members of the genus *Varicosavirus* (RCaVV, AMVV1, and LBVaV), orchid fleck virus (OFV, the genus *Dichorhavirus*), and Trichosanthes associated rhabdovirus 1 (TrARV1, the genus *Cytorhabdovirus*) are depicted. Known or predicted ORFs are represented as light green (ZaVV1) or white (other viruses) arrows on the negative-sense strand (3' to 5'). Tentative orthologous proteins are connected by shaded areas: light blue, shared by all rhabdoviruses; light orange, shared among some varicosaviruses; and gray, shared by other rhabdoviruses. Predicted InterPro domains of ZaVV1 proteins are marked by blue lines above the ORF: nuc, "Rhabdovirus nucleoprotein"; pol, "Mononegavirales RNA-directed RNA polymerase catalytic domain"; cap, "Mononegavirales mRNA-capping region V"; and met, "Mononegavirus L protein 2-O-ribose methyltransferase." The RNA-seq read depth of the ZaVV1 genome is shown at the top. Predicted regulatory elements in varicosavirus genomes are indicated by red diamonds.

were identified by a similarity search against known viral RdRp domain sequences. One contig that was 6,632 nt in length showed a strong sequence similarity to the RdRp domain of the RCaVV L protein. A subsequent sequence similarity search of the NCBI protein database using the contig as a query confirmed that it encodes a protein that is markedly similar to the L proteins of RCaVV, AMVV1, and LBVaV (Sasaya *et al.*, 2002; Sabbadin *et al.*, 2017; Koloniuk *et al.*, 2018b). These three viruses are members of the genus *Varicosavirus* of the family *Rhabdoviridae*, the order *Mononegavirales*, suggesting that the contig may be a genomic segment of a novel varicosavirus.

Varicosavirus genomes are composed of two genomic segments: RNA1, which encodes the L protein, and RNA2, which encodes the N and additional proteins. The 6,632-nt contig was identified as the RNA1 segment of a novel varicosavirus. To identify the RNA2 segment, all known varicosavirus protein sequences were downloaded and compared against all the common eelgrass transcriptome contigs. A contig that was 4,304 nt in length was identified to have an ORF specific for a protein similar to the N proteins of known varicosaviruses. Therefore, these two contigs were considered the RNA1 (6,632-nt contig) and RNA2 (4,304-nt contig) segments of a novel varicosavirus, which was tentatively named Zostera associated varicosavirus 1 (ZaVV1).

The ZaVV1 RNA1 genome segment was predicted to have a single ORF encoding a 2,001-aa L protein, while

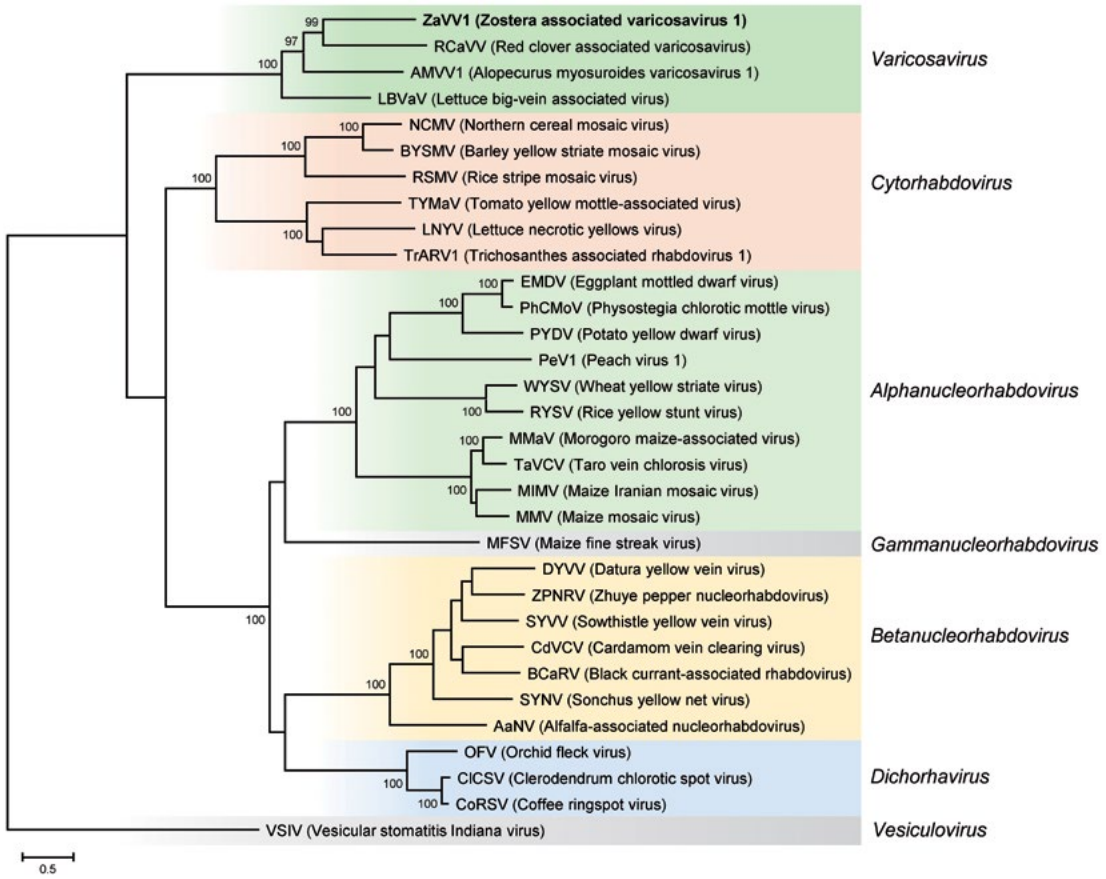

Fig. 2

Phylogenetic position of ZaVV1 among the family *Rhabdoviridae*

A maximum-likelihood phylogenetic tree was constructed based on the L protein sequences of ZaVV1 and related rhabdoviruses. ZaVV1 was placed within the genus *Varicosavirus*. RCaVV was identified to be the closest member of ZaVV1. The tree was rooted using vesicular stomatitis Indiana virus (VSIV) (the genus *Vesiculovirus*) as the outgroup. Bootstrap support percentages, 95% or greater, calculated from 1,000 replicates, are shown at the nodes.

the RNA2 segment had four ORFs encoding a 438-aa N protein, 353-aa protein 2 (P2), 306-aa protein 3 (P3), and 202-aa protein 4 (P4) (Fig. 1). The ZaVV1 genome sequence and annotation information are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK014484 and BK014485.

The ZaVV1 L protein had three predicted InterPro protein domains: a "Mononegavirales RNA-directed RNA polymerase catalytic domain" (InterPro accession number IPR014023) at aa positions 184–1054, a "Mononegavirales mRNA-capping region V" domain (IPR026890) at positions 1067–1283, and a "Mononegavirus L protein 2-O-ribose methyltransferase" domain (IPR025786) at positions 1610–1789. These domains are commonly found in the L proteins of viruses from the order *Mononegavirales* and function in viral genomic segment replication, viral gene transcription, and viral mRNA capping (Paesen *et al.*, 2015; Liang, 2020). The ZaVV1 L protein showed the highest sequence similarities with L proteins of known varicosaviruses, followed by those of other rhabdoviruses (Table 1).

The ZaVV1 N protein encoded in the RNA2 segment contained a "Rhabdovirus nucleoprotein" domain (IPR004902) at positions 88–369. This domain is commonly found in rhabdovirus N proteins, which encapsidate viral genomes (Kormelink *et al.*, 2011; Dietzgen *et al.*, 2017; Goh *et al.*, 2020).

The ZaVV1 RNA2 segment was predicted to have three additional ORFs encoding hypothetical proteins designated P2, P3, and P4, in the order of 3'-N-P2-P3-P4-5'. Their predicted molecular weights were 40, 34, and 23 kDa, respectively. The ZaVV1 P2 protein has a marginal sequence similarity with the RCaVV 47K protein (20.2% identity in the 377 aa overlap) and the AMVV1 P2 protein

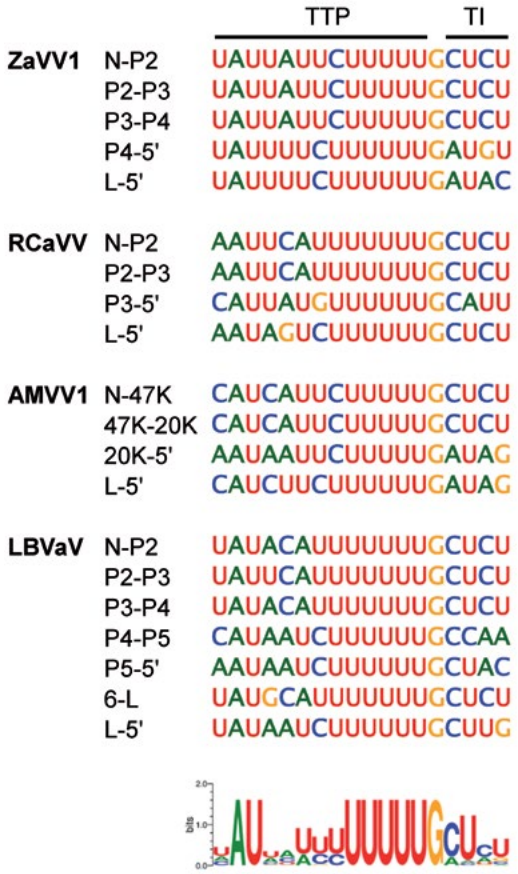

Fig. 3

Conserved genomic sequences in varicosavirus genomes
Putative regulatory elements possibly involved in transcription termination/polyadenylation (TTP) and transcription initiation (TI) were identified in the gene-junction regions of ZaVV1, RCaVV, AMVV1, and LBVaV. The RNA sequences are presented in the 3' to 5' direction. A sequence logo representation created from all predicted sequences is shown at the bottom. See Fig. 1 for the locations of the predicted regulatory elements in the viral genomes.

(19.4% identity in the 248 aa overlap). The RCaVV 47K and AMVV1 P2 genes are located after the N gene in the RNA2 segments of the respective viruses. The sequence similarity and shared genomic locations suggest that these proteins may have similar functions. However, no molecular function has been reported for the RCaVV 47K or AMVV1 P2 proteins (Sabbadin *et al.*, 2017; Koloniuk *et al.*, 2018b). No other sequences in the NCBI protein database yielded a significant hit with the ZaVV1 P2 protein.

The ZaVV1 P3 protein showed sequence similarity only with the LBVaV P3 protein (28.6% identity in the 241 aa overlap) and did not have any meaningful similarity to other proteins in the NCBI protein database. The LBVaV P3 protein was proposed to be involved in viral cell-to-cell movement; however, no direct evidence for this phenomenon has been observed (Dietzgen *et al.*, 2017). The ZaVV1 P4 protein showed no significant similarity to known varicosavirus proteins or any other protein in the NCBI database.

To investigate the phylogenetic position of ZaVV1 among rhabdoviruses, L protein sequences were collected from three known viruses of the genus *Varicosavirus* and representative viruses of the genera *Betanucleorhabdovirus*, *Cytorhabdovirus*, *Dichorhavirus*, *Gammanucleorhabdovirus*, and *Vesiculovirus*. Pairwise sequence identities between the ZaVV1 L protein and those of the 30 rhabdoviruses were calculated (see Table 1). The ZaVV1 L protein showed 37.5%–39.4% aa sequence identity with three known varicosavirus L proteins and 21.1%–25.1% identity with the other rhabdovirus L proteins.

Multiple sequence alignment of the collected L protein sequences was performed using the MAFFT program, followed by sequence filtering using the trimAl program (Capella-Gutierrez *et al.*, 2009; Nakamura *et al.*, 2018). A maximum-likelihood phylogenetic inference obtained using the IQ-TREE program (Minh *et al.*, 2020) placed ZaVV1 within the genus *Varicosavirus*, a clade supported by 100% bootstrap value (Fig. 2). The virus closest to ZaVV1 was identified as RCaVV (bootstrap value, 99%).

Because viral genome contigs may be assembled from viral mRNAs as well as genomic RNAs and varicosavirus genes were suggested to be individually transcribed, the expression levels of ZaVV1 viral genes may affect the sequencing depths of ORFs in the genomic segments (Koloniuk *et al.*, 2018b). To validate this possibility, the sequencing depths of the ZaVV1 genomic contigs were examined by mapping high-quality RNA-seq reads to the genome sequence (see Fig. 1, top). A total of 4,237 and 23,938 reads were mapped to the RNA1 and RNA2 segments, respectively. ORFs in the RNA2 segment exhibited higher read depths than the L protein ORF in the RNA1 segment. The P3 gene showed the highest average sequencing depth. The lower depth of the L gene was in agreement with previous reports that examined other rhabdoviruses (Koloniuk *et al.*, 2018b; Goh *et al.*, 2020).

Interestingly, the sequencing depths sharply decreased at the gene-junction regions (N-P2, P2-P3, and P3-P4 in the RNA2 segment) and end regions of the last genes (P4 and L). The sharp drop in read depth may indicate the presence of a transcription termination site after each gene. A similar phenomenon of low read-depth in gene-junction regions has also been observed in other rhabdoviruses (Franova *et al.*, 2019).

Rhabdoviruses have conserved regulatory elements in gene-junction regions of their genomes, which are involved in the TTP and TI of individual viral gene mRNAs (Jackson *et al.*, 2005; Ogino and Green, 2019; Goh *et al.*, 2020; Zhou *et al.*, 2020). To identify putative conserved

elements in varicosavirus genomes, the gene-junction regions and non-coding regions (NTRs) of ZaVV1, RCaVV, AMVV1, and LBVaV were investigated. From each of the four viral genome sequences, 3'-NTRs, gene-junction regions, and 5'-NTRs were extracted. Analysis of these sequences using the MEME web server yielded an 18-nt conserved sequence motif that was present in 20 genomic regions of four varicosaviruses: five regions in ZaVV1, four regions in RCaVV, four regions in AMVV1, and seven regions in LBVaV (Fig. 3, see Fig. 1 for locations).

The consensus sequence deduced from the 20 sequences by the MEME program was 3'-HAUWMUYUUUUUUG-CUCU-5', where H is A, C, or U, W is A or U, and M is A or C. The consensus sequence can be divided into three elements: a 13-nt TTP motif, 3'-HAUWMUYUUUUUU-5', 1-nt untranscribed spacer (3'-G-5'), and 4-nt TI motif (3'-CUCU-5'). The predicted TTP/TI motifs were virtually identical to previously deduced motifs in LBVaV, AMVV1, and RCaVV (Sasaya *et al.*, 2004; Koloniuk *et al.*, 2018b).

The three gene-junction regions of the ZaVV1 RNA2 segment (N-P2, P2-P3, and P3-P4) had exactly the same sequence, 3'-UAUUAUUCUUUUU–G–CUCU-5', in which the dashes separate the TTP motif, untranscribed spacer, and TI motif. The 5'-NTRs of RNA1 and RNA2 segments had the same TTP motif sequence, 3'-UAUUUUCUUUUUU-5', which was similar to the TTP motif of the gene-junction regions. However, two ZaVV1 5'-NTRs did not have a TI motif sequence; the 5'-NTRs of RNA1 and RNA2 had 3'-AUGU-5' and 3'-AUAC-5', respectively, instead of 3'-CUCU-5'. This difference is probably due to the absence of genes after the P4 and L genes and no requirement of transcription initiation.

In conclusion, the genome sequence of a novel, bipartite, negative-sense, single-stranded RNA virus, ZaVV1, was identified from the common eelgrass transcriptome dataset. The ZaVV1 genome was predicted to contain five proteins, including the L and N proteins that are commonly shared by rhabdoviruses. Sequence comparison and phylogenetic analysis revealed that ZaVV1 is a novel virus belonging to the genus *Varicosavirus* of the family *Rhabdoviridae*. The genome sequence of ZaVV1 may be useful for studying the evolution and phylogenetic relationships of varicosaviruses.

Acknowledgment. This work was supported by the Chung-Ang University Graduate Research Scholarship Grants in 2021 and the National Research Foundation of Korea grant (2020R1A2C1013403) funded by the Government of Korea.

References

Bailey TL, Elkan C (1994): Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36.

Bejerman N, Debat H, Dietzgen RG (2020): The plant negative-sense RNA virosphere: Virus discovery through new eyes. Front. Microbiol. 11, 588427. https://doi.org/10.3389/fmicb.2020.588427

Buchfink B, Xie C, Huson DH (2015): Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019): rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience 8, giz100. https://doi.org/10.1093/gigascience/giz100

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009): trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Dahl M, Deyanova D, Gutschow S, Asplund ME, Lyimo LD, Karamfilov V, Santos R, Bjork M, Gullstrom M (2016): Sediment properties as important predictors of carbon storage in Zostera marina meadows: A comparison of four European areas. PLoS One 11, e0167493. https://doi.org/10.1371/journal.pone.0167493

Dietzgen RG, Kondo H, Goodin MM, Kurath G, Vasilakis N (2017): The family Rhabdoviridae: Mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. Virus Res. 227, 158–170. https://doi.org/10.1016/j.virusres.2016.10.010

Dietzgen RG, Kuhn JH, Clawson AN, Freitas-Astua J, Goodin MM, Kitajima EW, Kondo H, Wetzel T, Whitfield AE (2014): Dichorhavirus: a proposed new genus for Brevipalpus mite-transmitted, nuclear, bacilliform, bipartite, negative-strand RNA plant viruses. Arch. Virol. 159, 607–619. https://doi.org/10.1007/s00705-013-1834-0

Franova J, Sarkisova T, Jakesova H, Koloniuk I (2019): Molecular and biological properties of two putative new cytorhabdoviruses infecting Trifolium pratense. Plant Pathol. 68, 1276–1286. https://doi.org/10.1111/ppa.13065

Goh CJ, Hahn Y (2019): Identification of a novel member of the family Betaflexiviridae from the hallucinogenic plant Salvia divinorum. Acta Virol. 63, 373–379. https://doi.org/10.4149/av_2019_401

Goh CJ, Park D, Hahn Y (2020): Identification of Trichosanthes associated rhabdovirus 1, a novel member of the genus Cytorhabdovirus of the family Rhabdoviridae, in the Trichosanthes kirilowii transcriptome. Acta Virol. 64, 36–43. https://doi.org/10.4149/av_2020_105

Goh CJ, Park D, Hahn Y (2021): A novel tepovirus, Agave virus T, identified by the analysis of the transcriptome data of blue agave (Agave tequilana). Acta Virol. 65, 68–71. https://doi.org/10.4149/av_2021_107

Jackson AO, Dietzgen RG, Goodin MM, Bragg JN, Deng M (2005): Biology of plant rhabdoviruses. Annu. Rev. Phytopathol. 43, 623–660. https://doi.org/10.1146/annurev.phyto.43.011205.141136

Koloniuk I, Franova J, Sarkisova T, Pribylova J (2018a): Complete genome sequences of two divergent isolates of strawberry crinkle virus coinfecting a single strawberry plant. Arch. Virol. 163, 2539–2542. https://doi.org/10.1007/s00705-018-3860-4

Koloniuk I, Franova J, Sarkisova T, Pribylova J, Lenz O, Petrzik K, Spak J (2018b): Identification and molecular characterization of a novel varicosa-like virus from red clover. Arch. Virol. 163, 2213–2218. https://doi.org/10.1007/s00705-018-3838-2

Kormelink R, Garcia ML, Goodin M, Sasaya T, Haenni AL (2011): Negative-strand RNA viruses: the plant-infecting counterparts. Virus Res. 162, 184–202. https://doi.org/10.1016/j.virusres.2011.09.028

Lee JS, Goh CJ, Park D, Hahn Y (2019): Identification of a novel plant RNA virus species of the genus Amalgavirus in the family Amalgaviridae from chia (Salvia hispanica). Genes Genomics 41, 507–514. https://doi.org/10.1007/s13258-019-00782-1

Liang B (2020): Structures of the Mononegavirales polymerases. J. Virol. 94, e00175–20. https://doi.org/10.1128/JVI.00175-20

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020): IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Nakamura T, Yamada KD, Tomii K, Katoh K (2018): Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34, 2490–2492. https://doi.org/10.1093/bioinformatics/bty121

Nibert ML, Pyle JD, Firth AE (2016): A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. Virology 498, 201–208. https://doi.org/10.1016/j.virol.2016.07.002

Ogino T, Green TJ (2019): RNA synthesis and capping by non-segmented negative strand RNA viral polymerases: Lessons from a prototypic virus. Front. Microbiol. 10, 1490. https://doi.org/10.3389/fmicb.2019.01490

Orfanidou CG, Beta C, Reynard JS, Tsiolakis G, Katis NI, Maliogka VI (2020): Identification, molecular characterization and prevalence of a novel cytorhabdovirus infecting zucchini crops in Greece. Virus Res. 287, 198095. https://doi.org/10.1016/j.virusres.2020.198095

Paesen GC, Collet A, Sallamand C, Debart F, Vasseur JJ, Canard B, Decroly E, Grimes JM (2015): X-ray structure and activities of an essential Mononegavirales L-protein domain. Nat. Commun. 6, 8749. https://doi.org/10.1038/ncomms9749

Park D, Goh CJ, Hahn Y (2021): Two novel closteroviruses, fig virus A and fig virus B, identified by the analysis of the high-throughput RNA-sequencing data of fig (Ficus carica) latex. Acta Virol. 65, 42–48. https://doi.org/10.4149/av_2021_104

Park D, Goh CJ, Lee JS, Sebastiani F, Hahn Y (2020): Identification of Pistacia-associated flexivirus 1, a putative mycovirus of the family Gammaflexiviridae, in the mastic tree (Pistacia lentiscus) transcriptome. Acta Virol. 64, 28–35. https://doi.org/10.4149/av_2020_104

Park D, Zhang M, Hahn Y (2019): Novel Foveavirus (the family Betaflexiviridae) species identified in ginseng (Panax ginseng). Acta Virol. 63, 155–161. https://doi.org/10.4149/av_2019_204

Reynolds LK, DuBois K, Abbott JM, Williams SL, Stachowicz JJ (2016): Response of a habitat-forming marine plant to a simulated warming event is delayed, genotype specific, and varies with phenology. PLoS One 11, e0154532. https://doi.org/10.1371/journal.pone.0154532

Sabbadin F, Glover R, Stafford R, Rozado-Aguirre Z, Boonham N, Adams I, Mumford R, Edwards R (2017): Transcriptome sequencing identifies novel persistent viruses in herbicide resistant wild-grasses. Sci. Rep. 7, 41987. https://doi.org/10.1038/srep41987

Sasaya T, Ishikawa K, Koganezawa H (2002): The nucleotide sequence of RNA1 of lettuce big-vein virus, genus Varicosavirus, reveals its relation to nonsegmented negative-strand RNA viruses. Virology 297, 289–297. https://doi.org/10.1006/viro.2002.1420

Sasaya T, Kusaba S, Ishikawa K, Koganezawa H (2004): Nucleotide sequence of RNA2 of lettuce big-vein virus and evidence for a possible transcription termination/initiation strategy similar to that of rhabdoviruses. J. Gen. Virol. 85, 2709–2717. https://doi.org/10.1099/vir.0.80061-0

Tan Y, Zhang QS, Zhao W, Liu Z, Ma MY, Zhong MY, Wang MX, Xu B (2020): Chlororespiration serves as photoprotection for the photo-inactivated oxygen-evolving complex in Zostera marina, a marine angiosperm. Plant Cell Physiol. 61, 1517–1529. https://doi.org/10.1093/pcp/pcaa075

Walker PJ, Blasdell KR, Calisher CH, Dietzgen RG, Kondo H, Kurath G, Longdon B, Stone DM, Tesh RB, Tordo N, Vasilakis N, Whitfield AE, ICTV Report Consortium (2018): ICTV virus taxonomy profile: Rhabdoviridae. J. Gen. Virol. 99, 447–448. https://doi.org/10.1099/jgv.0.001020

Walker PJ, Dietzgen RG, Joubert DA, Blasdell KR (2011): Rhabdovirus accessory genes. Virus Res. 162, 110–125. https://doi.org/10.1016/j.virusres.2011.09.004

Whitfield AE, Huot OB, Martin KM, Kondo H, Dietzgen RG (2018): Plant rhabdoviruses-their origins and vector interactions. Curr. Opin. Virol. 33, 198–207. https://doi.org/10.1016/j.coviro.2018.11.002

Zhou J, Cao K, Zhang Z, Wang L, Li S (2020): Identification and characterization of a novel rhabdovirus infecting peach in China. Virus Res. 280, 197905. https://doi.org/10.1016/j.virusres.2020.197905