Two novel poty-like viruses identified from the transcriptome data of purple witchweed (*Striga hermonthica*)

Dongjin Choi[1], Chaerim Shin[1], Ken Shirasu[2,3], Yoonsoo Hahn[1*]

[1]Department of Life Science, Chung-Ang University, Seoul 06974, South Korea; [2]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan; [3]Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan

Received May 12, 2021; accepted July 20, 2021

Summary. – Potyvirids (the family *Potyviridae*) are the largest family of plant RNA viruses. Two novel potyvirid viruses, Striga-associated poty-like virus 1 (SaPlV1) and Striga-associated poty-like virus 2 (SaPlV2), were identified from the transcriptome data of purple witchweed (*Striga hermonthica*). SaPlV1 was most closely related to bellflower vein mottle virus (BVMoV), the only member of the genus *Bevemovirus*, and then to macluraviruses (the genus *Macluravirus*). The SaPlV1 genome encodes a 2462-amino acid (aa) polyprotein that may be cleaved into nine mature peptides. The cleavage sites of SaPlV1, BVMoV, and macluravirus polyproteins shared strong sequence similarities. SaPlV2 was most closely related to celery latent virus, the sole species of the genus *Celavirus*, which is the most

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

Abbreviations: BVMoV = bellflower veinal mottle virus; CeLV = celery latent virus; ORF = open reading frame; PIPO = pretty interesting *Potyviridae* ORF; RdRp = RNA-dependent RNA polymerase; SaPlV1 = Striga-associated poty-like virus 1; SaPlV2 = Striga-associated poty-like virus 2

divergent potyvirid genus. The SaPlV2 polyprotein contained 3329 aa; it may be cleaved into at least seven or eight mature peptides. Phylogenetic analysis suggested that SaPlV1 and SaPlV2 may be novel species of the genera *Bevemovirus* and *Celavirus*, respectively. The genome sequences of SaPlV1 and SaPlV2 are useful resources for studying the genome evolution of potyvirids.

Keywords: Striga-associated poty-like virus 1; Striga-associated poty-like virus 2; Potyviridae; Bevemovirus; Celavirus; purple witchweed, Striga hermonthica

Introduction

Potyviridae is the largest family of plant-infecting RNA viruses; it contains 235 species approved by the International Committee on Taxonomy of Viruses (ICTV) (https://talk.ictvonline.org; ICTV Master Species List #36, March 2021) (Gibbs *et al.*, 2020; Palani *et al.*, 2021). Potyvirids (members of the family *Potyviridae*) are classified into 12 genera: *Arepavirus*, *Bevemovirus*, *Brambyvirus*, *Bymovirus*, *Celavirus*, *Ipomovirus*, *Macluravirus*, *Poacevirus*, *Potyvirus*, *Roymovirus*, *Rymovirus*, and *Tritimovirus* (Palani *et al.*, 2021). Members of the genus *Bymovirus* have a bipartite genome, whereas those from the remaining 11 genera have a monopartite positive-sense single-stranded RNA genome (Revers and Garcia, 2015).

The genomic RNA of potyvirids encodes an open reading frame (ORF) for a large polyprotein, which undergoes proteolytic cleavage by virus-encoded proteases (Adams *et al.*, 2005; Goh and Hahn, 2021; Palani *et al.*, 2021). The polyprotein of potyviruses (genus *Potyvirus*) is processed into ten mature peptides (from N-terminus to C-terminus): protein 1 protease (P1), helper component-protease (HC-Pro), protein 3 (P3), 6-kilodalton (kDa) peptide 1 (6K1), cylindrical inclusion protein (CI), 6-kDa peptide 2 (6K2), viral protein

genome-linked (VPg), nuclear inclusion a protease (NIa-Pro), nuclear inclusion b protein (NIb), and coat protein (CP) (Adams *et al.*, 2005; Goh and Hahn, 2021). P1 and HC-Pro proteases cleave themselves from the polyprotein, and NIa-Pro cleaves the remaining seven junctions. The NIb protein is an RNA-dependent RNA polymerase (RdRp) that is responsible for viral genome replication (Shen *et al.*, 2020). Potyvirids from other genera have the same or similar genomic organization.

Eight mature peptides, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP, are commonly produced from potyvirid polyproteins (Revers and Garcia, 2015). However, the P1 and HC-Pro segments among potyvirids are variable, probably resulting from diversifying evolution involving recombination and gene duplication (Valli *et al.*, 2007). For example, macluraviruses (genus *Macluravirus*) lack the P1 coding region and HC-Pro is absent in cassava brown streak virus (Mbanzibwa *et al.*, 2009; Revers and Garcia, 2015; Elangovan *et al.*, 2019).

An additional short ORF termed "pretty interesting *Potyviridae* ORF" (PIPO) is universally present within the P3 coding region of potyvirid genomes (Chung *et al.*, 2008). A polymerase slippage inserts an additional A nucleotide within the highly conserved GAAAAAA (GA_6) motif. This insertion, termed "+1A insertion event," results in the formation of a "trans-frame" fusion protein with the N-terminal half of P3 and PIPO; this protein is named the P3N-PIPO protein (Olspert *et al.*, 2015; Rodamilans *et al.*, 2015; White, 2015).

The majority of ICTV-approved *Potyviridae* species (190 out of 235) belong to the genus *Potyvirus*. The other 11 genera each contain one to ten recognized species. The genus *Bevemovirus* has a single approved species, the *Bellflower veinal mottle virus* (BVMoV), which was isolated from a bellflower (*Campanula takesimana*) (Seo *et al.*, 2017). The BVMoV polyprotein is predicted to be cleaved to produce nine mature proteins, HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP; this is the same as that of macluravirus polyproteins. Phylogenetic analysis using BVMoV and selected potyvirid polyprotein

sequences suggested that the genus *Bevemovirus* is the sister taxon of the genus *Macluravirus* (Seo *et al.*, 2017).

Celavirus is another genus represented by only one species, the *Celery latent virus* (CeLV), which was first identified in celery (*Apium graveolens* var. *dulce*) and celeriac (*A. g*. var. *rapaceum*) without visible symptoms (Bos *et al.*, 1978; Rose *et al.*, 2019). The CeLV polyprotein has a low sequence similarity to other potyvirid polyproteins, suggesting that *Celavirus* is the most divergent genus of the family *Potyviridae* (Rose *et al.*, 2019).

RNA-Seq data obtained from plant tissue samples often contain genomic RNA or mRNA fragments derived from latently infected RNA viruses (Bejerman *et al.*, 2020; Park and Hahn, 2021). Comprehensive analyses of diverse plant transcriptome data have yielded many novel RNA virus genome sequences (Park *et al.*, 2018; Goh *et al.*, 2021; Park *et al.*, 2021; Park and Hahn, 2021). In this study, we identified the genome sequences of two novel viruses belonging to the genera *Bevemovirus* and *Celavirus* (family *Potyviridae*) from the transcriptome data of purple witchweed (*Striga hermonthica*) (Yoshida *et al.*, 2019). Purple witchweed is a hemiparasitic plant that infests grain crops, such as sorghum, maize, and sugar cane, and causes major grain losses (Spallek *et al.*, 2013).

Materials and Methods

The purple witchweed transcriptome data analyzed in this study are available in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) (Yoshida *et al.*, 2019). The SRA Acc. Nos. are DRR183243, DRR183244, DRR183245, DRR183246, DRR183247, DRR183248, DRR183249, and DRR183250. RNA-Seq reads were trimmed using the sickle program (version 1.33; https://github.com/najoshi/sickle) with the parameter "-q 30 -l 55." Filtered high-quality reads from all nine sequencing runs were pooled into a single dataset and assembled into contigs using the rnaviralSPAdes pipeline of the SPAdes

assembler (version 3.15.1; http://cab.spbu.ru/software/spades) (Bushmanova *et al.*, 2019). Known viral RdRp domain sequences were downloaded from the Pfam database (release 33.1; https://pfam.xfam.org). Pfam Acc. Nos. for the viral RdRp families are 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; http://www.diamondsearch.org/index.php) was used to compare the purple witchweed contigs and known viral RdRp sequences.

Sequencing depth was examined by mapping high-quality RNA-Seq reads to the contig sequence using the bwa-mem2 program (version 2.0pre2; https://github.com/bwa-mem2/bwa-mem2). Conserved domains in the polyprotein sequences were predicted using the InterPro web server (version 84.0; https://www.ebi.ac.uk/interpro). The SignalP web server (version 5.0; https://services.healthtech.dtu.dk/service.php?SignalP) was used to predict the signal peptide.

Pairwise identities of the protein sequences were calculated using the needle program of the EMBOSS package (version 6.6.0.0; http://emboss.open-bio.org). Multiple sequence alignments were performed using the MAFFT program (version 7.475; https://mafft.cbrc.jp/alignment/software) (Nakamura *et al.*, 2018). Gap-rich segments in the aligned sequences were removed using the trimAl program (version 1.4.rev22; http://trimal.cgenomics.org) with the "-gappyout" option (Capella-Gutierrez *et al.*, 2009). A maximum-likelihood phylogenetic tree was constructed using the IQ-TREE program (version 2.1.2; http://www.iqtree.org) (Minh *et al.*, 2020). Bootstrap supports were calculated from 1,000 replicates using the UFBoot2 method implemented in the IQ-TREE program. Sequence logos were generated using the WebLogo 3 server (http://weblogo.threeplusone.com) (Crooks *et al.*, 2004).

Results and Discussion

We re-analyzed the transcriptome data obtained for the investigation of gene expression dynamics during haustorium development in purple witchweed (Yoshida *et al.*, 2019). Transcriptome contigs assembled from high-quality RNA-Seq reads were compared with known viral RdRp sequences. Several contigs were found to contain a viral RdRp motif. Two different contigs showing significant sequence similarity to potyvirid RdRp motifs were chosen for further studies. To obtain a high-quality contig sequence, the sequencing depth of each contig was examined by mapping RNA-Seq reads to a contig, and bases supported by only one read were trimmed off at both ends. Two putative viral contigs that were 7853 and 10767 nucleotides (nt) in length were obtained.

BLASTX searches of the NCBI protein database using two contigs as query sequences confirmed that they encoded a large protein showing significant sequence similarity to known potyvirid polyproteins. Therefore, the 7853-nt and 10767-nt contigs were considered genome sequences of novel potyvirids and tentatively named Striga-associated poty-like virus 1 (SaPlV1) and Striga-associated poty-like virus 2 (SaPlV2), respectively. The genome sequences were deposited in the NCBI database (Acc. Nos. MW699352 and MW699353, respectively).

Genome sequence of Striga-associated poty-like virus 1 (SaPlV1)

The SaPlV1 genome was 7853-nt long and predicted to encode a 2462-amino acid (aa) polyprotein (Fig. 1a and Table 1). The InterPro domain analysis indicated that the SaPlV1 polyprotein has conserved domains that are typically found in other potyvirid polyproteins, including the HC-Pro cysteine protease, helicase, NIa protease, and RdRp domains.

A sequence similarity search of all known viral proteins revealed that the SaPlV1 polyprotein

was the most similar to the BVMoV polyprotein (NCBI Acc. No. YP_009508455.1), with 52.6% identity in the 2496-aa overlap. BVMoV is a species within the genus *Bevemovirus* (Seo *et al.*, 2017). The next similar sequences were polyproteins of various macluraviruses (genus *Macluravirus*) such as the narcissus latent virus (QDC21212.1), cardamom mosaic virus (YP_009508901.1), and Chinese yam necrotic mosaic virus (YP_006590058.1), with 32%–34% identity in an approximately 2700-aa overlap (Kondo and Fujita, 2012; Elangovan *et al.*, 2019; Wylie *et al.*, 2019). Thus, SaPlV1 is the most closely related to BVMoV and possibly, a novel member of the genus *Bevemovirus*.

The polyprotein of potyvirids is post-translationally cleaved at conserved cleavage sites to produce mature peptides (Adams *et al.*, 2005; Revers and Garcia, 2015; Goh and Hahn, 2021; Palani *et al.*, 2021). BVMoV and macluravirus polyproteins are processed into nine mature peptides: HC-Pro, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, and CP (Revers and Garcia, 2015; Seo *et al.*, 2017). The HC-Pro/P3 cleavage site is self-cleaved by HC-Pro, and the remaining seven sites are processed by NIa-Pro. To predict the putative cleavage sites and mature peptides of the SaPlV1 polyprotein, the polyprotein sequences of SaPlV1, BVMoV, and nine macluraviruses were multiply aligned (Supplementary Fig. S1 and Fig. 1b).

Based on the annotated cleavage sites of three macluraviruses (narcissus latent virus, cardamom mosaic virus, and Chinese yam necrotic mosaic virus), eight cleavage sites were predicted in the SaPlV1 polyprotein (see Fig. 1a for their positions). The predicted HC-Pro/P3 cleavage site was YIIG/G, which is highly similar to the potyvirus HC-Pro recognition sequence YXVG/G, where 'X' indicates any aa and the slash (/) indicates the cleaved bond (Adams *et al.*, 2005; Goh and Hahn, 2021). Seven other predicted cleavage sites, which were processed by NIa-Pro, were TKLQ/A (P3/6K1), TRLQ/F (6K1/CI), MELQ/A (CI/6K2), YTLE/G (6K2/VPg), TRLE/V (VPg/NIa-Pro), DRLQ/S (NIa-Pro/NIb), and THLQ/V (NIb/CP). When the SaPlV1 NIa-Pro recognition sites were compared with those of BVMoV and macluraviruses, two consensus sequences were deduced: XXLQ/X for

five sites (P3/6K1, 6K1/CI, CI/6K2, NIa-Pro/NIb, and NIb/CP) and XXLE/X for two sites (6K2/VPg, VPg/NIa-Pro) (see Fig. 1b). The same pattern was also observed in bymoviruses (genus *Bymovirus*), which are more closely related to BVMoV and macluraviruses than to other potyvirids (Palani *et al.*, 2021).

Potyviruses have an additional overlapping ORF termed PIPO within the P3 segment (Revers and Garcia, 2015; Gibbs *et al.*, 2020). The trans-frame fusion protein P3N-PIPO is produced by a +1A insertion in the A-tract of the GA_6 motif by polymerase slippage (Olspert *et al.*, 2015; White, 2015). There was only one GA_6 motif in the SaPlV1 genome sequence within the P3 segment (nt 1406–1412) (see Fig. 1a). To examine if the +1A insertion product was present in the purple witchweed transcriptome data, RNA-Seq reads were mapped to the SaPlV1 genome sequence. There were 18 reads spanning the GA_6 motif. Among them, 17 (94.44%) were the wild-type and 1 (5.56%) was the +1A insertion product, suggesting that this position may be a polymerase slippage site.

The SaPlV1 small polyprotein sequence containing the 54-aa PIPO segment was deduced by inserting an A nucleotide within the A-tract of the GA_6 motif. When the predicted SaPlV1 and BVMoV small polyprotein sequences were compared, the P3N and PIPO boundaries of both viruses exactly matched (Supplementary Fig. S2). The PIPO segments of the two viruses exhibited high sequence similarity, further indicating that the predicted GA_6 motif was a genuine polymerase slippage site for the production of the trans-frame fusion protein P3N-PIPO.

Genome sequence of Striga-associated poty-like virus 2 (SaPlV2)

The SaPlV2 genome was 10767-nt long and predicted to encode a 3329-aa polyprotein (Fig. 2 and Table 1). The SaPlV2 polyprotein had InterPro domains that are shared by potyvirid polyproteins, including helicase, NIa protease, and RdRp domains. No HC-Pro cysteine

protease domains were detected. Sequence similarity search revealed that the CeLV polyprotein (3640-aa, NCBI Acc. No. AZJ53460.1) was the most similar to the SaPlV2 polyprotein, with 44.9% identity in a 3917-aa overlap, whereas the other potyvirid polyproteins showed lower (15%–20%) sequence identities. Thus, SaPlV2 is the most closely related to CeLV among the currently known viruses.

CeLV is the sole species in the genus *Celavirus* (Rose *et al.*, 2019). Because CeLV is distantly related to other potyvirids, annotation of proteolytic cleavage sites and mature peptides of its polyprotein is highly limited. The CeLV polyprotein was assumed to be cleaved into at least seven or eight mature peptides: one or two P1-like proteases, a P3-like peptide, CI, VPg, NIa-Pro, NIb, and CP (Rose *et al.*, 2019). However, only two cleavage sites (NIa-Pro/NIb and NIb/CP) have been previously predicted. Therefore, the annotation of the SaPlV2 polyprotein is difficult.

The SaPlV2 and CeLV polyprotein sequences shared a strong sequence similarity with their N-terminal regions (Supplementary Fig. S3). The SaPlV2 polyprotein was approximately 300-aa shorter than the CeLV polyprotein at the N-terminus. There was an in-frame stop codon upstream of the SaPlV2 polyprotein ORF, indicating that the SaPlV2 polyprotein was full-length. The CeLV polyprotein was predicted to contain a signal peptide at its N-terminus (Rose *et al.*, 2019). However, no signal peptide was predicted for the SaPlV2 polyprotein. Despite the difference in their N-terminal regions, it was assumed that the SaPlV2 polyprotein may be processed into seven or eight mature peptides similar to those of CeLV. The SaPlV2 and CeLV polyproteins lacked the HC-Pro segment. Among the six or seven potential cleavage sites, two were predicted in the SaPlV2 polyprotein: LNRQ/D (NIa-Pro/NIb) and QQIQ/N (NIb/CP) (see Fig. 2 and Supplementary Fig. S3).

A putative GA_6 polymerase slippage motif for the P3N-PIPO fusion was found within the SaPlV2 P3-like segment (nt 2041–2047). A total of 191 purple witchweed RNA-Seq reads mapped to the putative GA_6 motif: 175 (91.62%) were the wild-type and 16 (8.38%) were the

+1A insertion products, implying that the position was a polymerase slippage site. The predicted SaPlV2 PIPO contained 130 aa whereas the CeLV PIPO contained 189 aa. Sequence alignment of the SaPlV2 and CeLV small polyproteins with the P3N-PIPO fusion revealed that the PIPO segments of the two viruses shared many conserved residues, although their start positions were different (Supplementary Fig. S4).

Phylogenetic positions of SaPlV1 and SaPlV2 among potyvirids

To investigate the phylogenetic positions of SaPlV1 and SaPlV2 among potyvirids, polyprotein sequences of 76 selected potyvirids were collected from 12 genera of the family *Potyviridae*. In the case of the genus *Bymovirus*, which has a bipartite genome, two polyprotein sequences were concatenated. Multiple sequence alignment was generated using the MAFFT program, followed by filtering out the gap-rich segments using the trimAl program. A maximum-likelihood tree was constructed using the IQ-TREE program (Fig. 3). SaPlV1 and BVMoV, the founding and only member of the genus *Bevemovirus*, clearly formed a strong clade with 100% bootstrap support. SaPlV2 and CeLV, the type species of the genus *Celavirus*, also formed a distinct clade with 100% bootstrap support. Phylogenetic relationships and similarities of genome organizations strongly suggested that SaPlV1 and SaPlV2 may be novel members of the genera *Bevemovirus* and *Celavirus*, respectively.

Conclusion

In conclusion, the genome sequences of SaPlV1 and SaPlV2, novel members of the family *Potyviridae*, were identified from the transcriptome data of purple witchweed. Sequence comparisons and phylogenetic analyses suggested that SaPlV1 and SaPlV2 may be novel species from the genera *Bevemovirus* and *Celavirus*, respectively. The genome sequences of

SaPlV1 and SaPlV2 may be useful for studying the evolution of potyvirid genome organizations.

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

Supplementary information is available in the online version of the paper.

References

Adams MJ, Antoniw JF, Beaudoin F (2005): Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. Mol. Plant Pathol. 6, 471-487. https://doi.org/10.1111/j.1364-3703.2005.00296.x

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

Bos L, Diaz-Ruiz JR, Maat DZ (1978): Further characterization of celery latent virus. Neth. J. Plant Pathol. 84, 61-79. https://doi.org/10.1007/BF01976409

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

Chung BY, Miller WA, Atkins JF, Firth AE (2008): An overlapping essential gene in the *Potyviridae*. Proc. Natl. Acad. Sci. USA 105, 5897-5902. https://doi.org/10.1073/pnas.0800468105

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004): WebLogo: a sequence logo generator. Genome Res. 14, 1188-1190. https://doi.org/10.1101/gr.849004

Elangovan S, Srikakulam N, Pandi G, Jacob T, Ramakrishnan U, Madhugiri Narayanarao V, Tennyson J (2019): The first complete genomic sequence of cardamom mosaic virus, a member of the genus *Macluravirus* (family *Potyviridae*). Arch. Virol. 164, 1723-1726. https://doi.org/10.1007/s00705-019-04203-2

Gibbs AJ, Hajizadeh M, Ohshima K, Jones RAC (2020): The potyviruses: an evolutionary

synthesis is emerging. Viruses 12, 132. https://doi.org/10.3390/v12020132

Goh CJ, Hahn Y (2021): Analysis of proteolytic processing sites in potyvirus polyproteins revealed differential amino acid preferences of NIa-Pro protease in each of seven cleavage sites. PLoS One 16, e0245853. https://doi.org/10.1371/journal.pone.0245853

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

Kondo T, Fujita T (2012): Complete nucleotide sequence and construction of an infectious clone of Chinese yam necrotic mosaic virus suggest that macluraviruses have the smallest genome among members of the family *Potyviridae*. Arch. Virol. 157, 2299-2307. https://doi.org/10.1007/s00705-012-1429-1

Mbanzibwa DR, Tian Y, Mukasa SB, Valkonen JP (2009): Cassava brown streak virus (*Potyviridae*) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. J. Virol. 83, 6934-6940. https://doi.org/10.1128/JVI.00537-09

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

Olspert A, Chung BY, Atkins JF, Carr JP, Firth AE (2015): Transcriptional slippage in the positive-sense RNA virus family *Potyviridae*. EMBO Rep. 16, 995-1004. https://doi.org/10.15252/embr.201540509

Palani SN, Sankaranarayanan R, Tennyson J (2021): Comparative study of potyvirid NIa proteases and their cleavage sites. Arch. Virol. 166, 1141-1149.

https://doi.org/10.1007/s00705-021-04997-0

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, Kim H, Hahn Y (2018): Identification of two novel amalgaviruses in the common eelgrass (*Zostera marina*) and *in silico* analysis of the amalgavirus +1 programmed ribosomal frameshifting sites. Plant Pathol. J. 34, 150-156. https://doi.org/10.5423/PPJ.NT.11.2017.0243

Park D, Hahn Y (2021): Identification of genome sequences of novel partitiviruses in the quinoa (*Chenopodium quinoa*) transcriptome datasets. J. Gen. Plant Path. 87, 236-241. https://doi.org/10.1007/s10327-021-01002-z

Revers F, Garcia JA (2015): Molecular biology of potyviruses. Adv. Virus Res. 92, 101-199. https://doi.org/10.1016/bs.aivir.2014.11.006

Rodamilans B, Valli A, Mingot A, San Leon D, Baulcombe D, Lopez-Moya JJ, Garcia JA (2015): RNA polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the *Potyviridae* family. J. Virol. 89, 6965-6967. https://doi.org/10.1128/JVI.00337-15

Rose H, Doring I, Vetten HJ, Menzel W, Richert-Poggeler KR, Maiss E (2019): Complete genome sequence and construction of an infectious full-length cDNA clone of celery latent virus - an unusual member of a putative new genus within the *Potyviridae*. J. Gen. Virol. 100, 308-320. https://doi.org/10.1099/jgv.0.001207

Seo JK, Kwak HR, Kim MK, Kim JS, Choi HS (2017): The complete genome sequence of a novel virus, bellflower veinal mottle virus, suggests the existence of a new genus within the family *Potyviridae*. Arch. Virol. 162, 2457-2461. https://doi.org/10.1007/s00705-017-3374-5

Shen W, Shi Y, Dai Z, Wang A (2020): The RNA-dependent RNA polymerase NIb of potyviruses plays multifunctional, contrasting roles during viral infection. Viruses 12, 77.

https://doi.org/10.3390/v12010077

Spallek T, Mutuku M, Shirasu K (2013): The genus *Striga*: a witch profile. Mol. Plant Pathol. 14, 861-869. https://doi.org/10.1111/mpp.12058

Valli A, Lopez-Moya JJ, Garcia JA (2007): Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. J. Gen. Virol. 88, 1016-1028. https://doi.org/10.1099/vir.0.82402-0

White KA (2015): The polymerase slips and PIPO exists. EMBO Rep. 16, 885-886. https://doi.org/10.15252/embr.201540871

Wylie SJ, Tran TT, Nguyen DQ, Koh SH, Chakraborty A, Xu W, Jones MGK, Li H (2019): A virome from ornamental flowers in an Australian rural town. Arch. Virol. 164, 2255-2263. https://doi.org/10.1007/s00705-019-04317-7

Yoshida S, Kim S, Wafula EK, Tanskanen J, Kim YM, Honaas L, Yang Z, Spallek T, Conn CE, Ichihashi Y, Cheong K, Cui S, Der JP, Gundlach H, Jiao Y, Hori C, Ishida JK, Kasahara H, Kiba T, Kim MS, Koo N, Laohavisit A, Lee YH, Lumba S, McCourt P, Mortimer JC, Mutuku JM, Nomura T, Sasaki-Sekimoto Y, Seto Y, Wang Y, Wakatake T, Sakakibara H, Demura T, Yamaguchi S, Yoneyama K, Manabe RI, Nelson DC, Schulman AH, Timko MP, dePamphilis CW, Choi D, Shirasu K (2019): Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. Curr. Biol. 29, 3041-3052.e3044. https://doi.org/10.1016/j.cub.2019.07.086

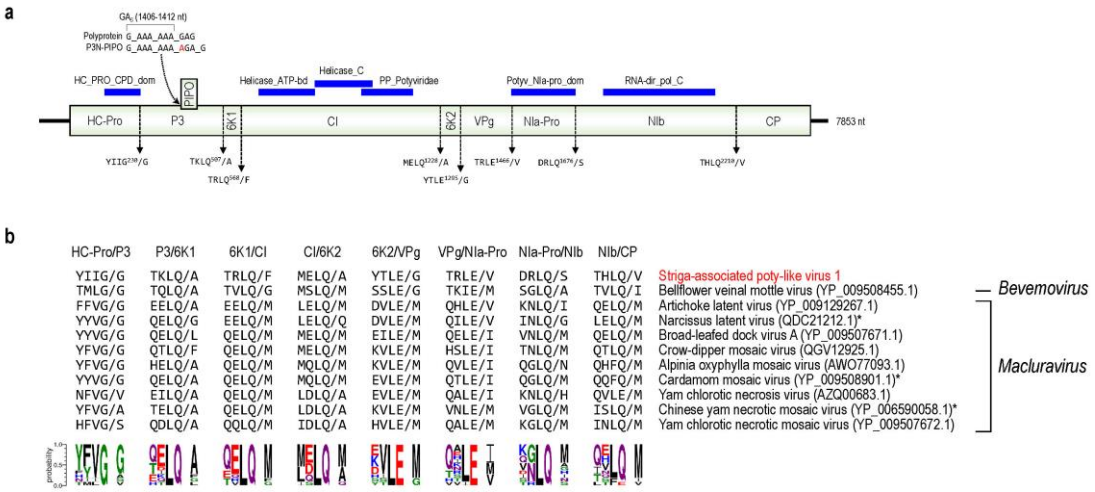

Fig. 1

Genome organization and cleavage sites of SaPlV1

(a) Schematic figure of the SaPlV1 genome is presented. ORFs for the large polyprotein and PIPO are depicted as boxes. Predicted InterPro domains are marked above the polyprotein ORF with the corresponding InterPro short name. The GA_6 motif causing a polymerase slippage is shown above the genome (the +1A insertion is indicated in red and the underscores ('_') indicate the codon boundaries). Predicted cleavage sites are presented below the polyprotein ORF. **(b)** Five aa resides at the eight cleavage sites of the SaPlV1, BVMoV, and nine macluravirus polyprotein sequences are presented. The slash ('/') indicates the cleaved bond. Three macluraviruses with annotated cleavage site information are indicated by asterisks. Sequence logos showing the frequencies of residues at each position are displayed at the bottom.

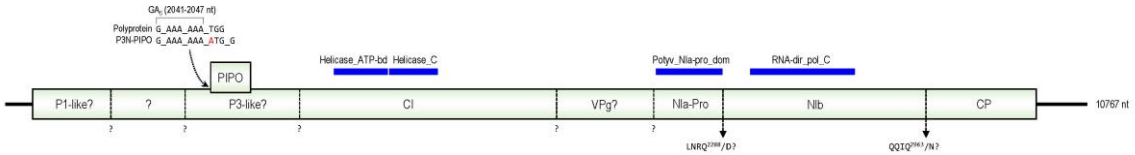

Fig. 2

Schematic figure of the putative genome organization of SaPlV2

Predicted genomic features of SaPlV2 are presented. See Fig. 1a for details.

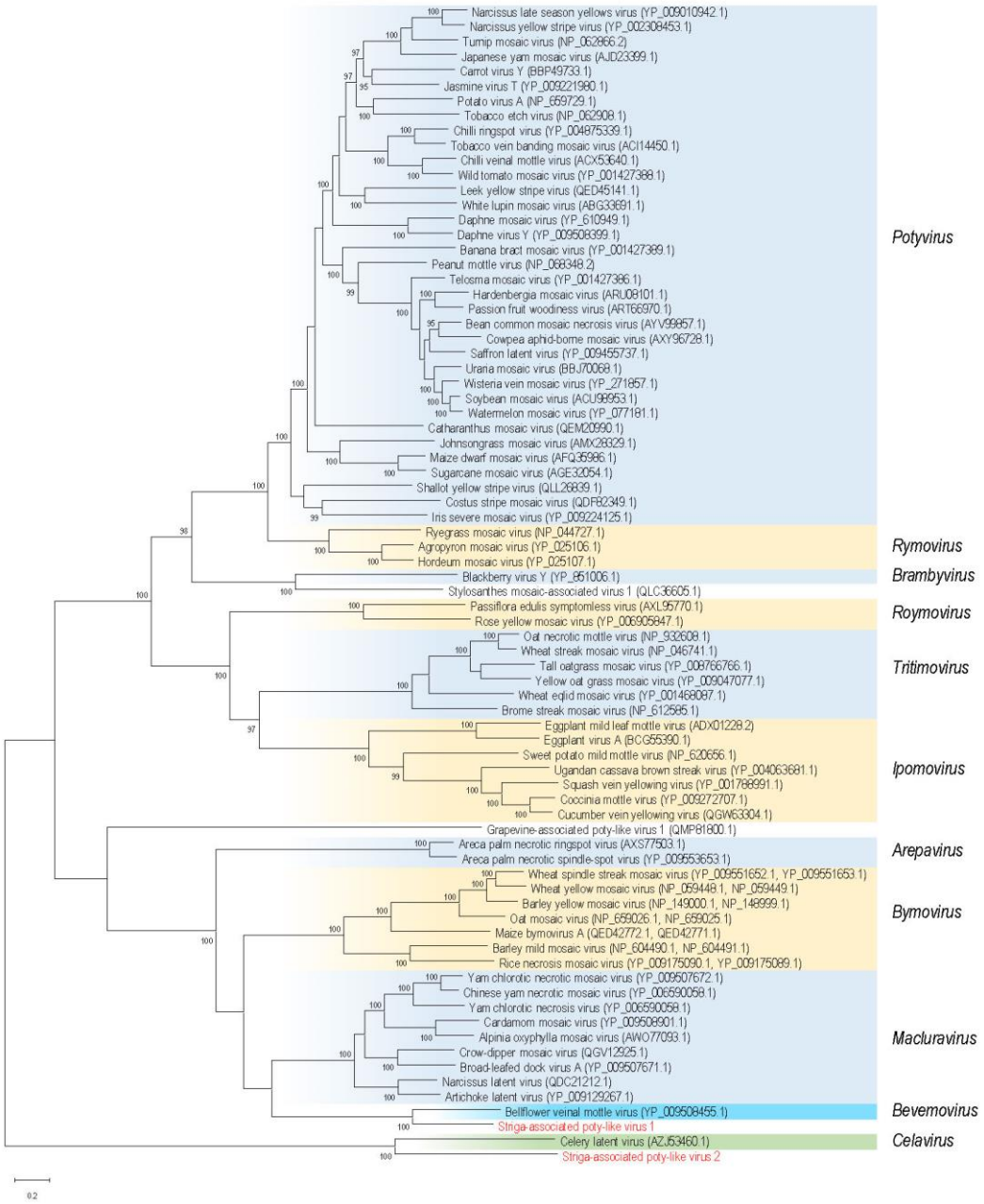

Fig. 3

Phylogenetic relationships of SaPlV1 and SaPlV2 among *Potyviridae* viruses

A maximum-likelihood phylogenetic tree was constructed from the polyprotein sequences of SaPlV1, SaPlV2, and selected members of the family *Potyviridae*. The NCBI protein sequence accession numbers are shown in parentheses. Bootstrap supporting values of 95% or higher are shown at the nodes.

Table 1. ORFs and domains of SaPlV1 and SaPlV2 genome sequences

Virus	ORF	ORF position (nt)	Protein length (aa)	Domain position (aa)	InterPro domain[a]
SaPlV1	Polyprotein	309–7679	2462	113–230	HC_PRO_CPD_dom (IPR031159): Helper-component proteinase (HC-Pro) cysteine protease (CPD) domain-like methyltransferase (MT) domain
				625–810	Helicase_ATP-bd (IPR014001): Helicase superfamily 1/2, ATP-binding domain
				810–1000	Helicase_C (IPR001650): Helicase, C-terminal
				967–1136	PP_Potyviridae (IPR013648): Polyprotein, Potyviridae
				1463–1674	Potyv_NIa-pro_dom (IPR001730): Potyvirus NIa protease (NIa-pro) domain
				1769–2139	RNA-dir_pol_C (IPR001205): RNA-directed RNA polymerase, C-terminal domain
	PIPO	1412–1573	53		
SaPlV2	Polyprotein	278–10267	3329	1180–1342	Helicase_C (IPR001650): Helicase, C-terminal
				2067–2285	Potyv_NIa-pro_dom (IPR001730): Potyvirus NIa protease (NIa-pro) domain
				2381–2726	RNA-dir_pol_C (IPR001205) RNA-directed RNA polymerase, C-terminal domain
	PIPO	2047–2439	130		

[a]InterPro domain in the format of "short name (accession number): full name."