

Novel *Foveavirus* (the family *Betaflexiviridae*) species identified in ginseng (*Panax ginseng*)

D. PARK¹, M. ZHANG², Y. HAHN^{1*}

¹Department of Life Science, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, South Korea;

²College of Life Science, Jilin Agricultural University, Changchun 130118, Jilin, P. R. China

Received July 17, 2018; accepted August 7, 2018

Summary. – Ginseng (*Panax ginseng*) is a valuable herb that is widely cultivated in Korea, China, and Japan because it contains a variety of pharmacologically active substances with a wide range of positive effects on human health. Identification and prevention of disease-causing viral pathogens of ginseng is important for improving the yield and quality of ginseng-derived bioactive molecules. In this study, the genome sequence of the virus *Panax ginseng flexivirus 1* (PgFV1) was identified from a ginseng root transcriptome data set. Sequence comparison and phylogenetic analysis showed that PgFV1 is a novel plant RNA virus species of the genus *Foveavirus* (the family *Betaflexiviridae*). Foveaviruses have flexuous and filamentous virions with a single-stranded positive-sense mono-segmented RNA genome. Its infection causes diseases with mosaic and ringspot symptoms in the stems and leaves. The PgFV1 genome encodes for 5 open reading frames: a replicase polyprotein for viral genome replication, 3 triple gene block proteins for viral cell-to-cell movement, and coat protein. Phylogenetic trees inferred from replicase polyprotein or coat protein sequences showed that PgFV1 is most closely related to grapevine virus T. PgFV1 is the first foveavirus identified to be associated with ginseng. Given the potential pathogenic features of previously known foveaviruses and importance of ginseng in the health industry, the PgFV1 genome sequence may be highly useful for studying ginseng foveaviruses.

Keywords: ginseng; *Panax ginseng flexivirus 1*; Foveavirus; *Betaflexiviridae*

Introduction

Ginseng (*Panax ginseng*) is a slowly growing perennial plant belonging to the genus *Panax* of the family *Araliaceae*, which is widely cultivated in Korea, China, and Japan. Ginseng has long been used as a valuable herb in oriental medicine because it contains various pharmacologically active substances such as ginsenosides (dammarane-type tri-

terpenoid saponins) and gintonin (Briskin, 2000; Yun, 2001; Baeg and So, 2013). Ginseng is known to have a wide range of positive effects on human health including preventing some types of cancers, improving diabetes and vascular diseases, enhancing host immunity, and protecting against some virus infections (Vuksan *et al.*, 2010; Kim and Park, 2011; Im *et al.*, 2016). Several comprehensive transcriptomic analyses using next-generation sequencing technology have been performed to understand the underlying molecular genetic mechanisms of the beneficial traits of ginseng (Jayakodi *et al.*, 2015; Wang *et al.*, 2015; Zhen *et al.*, 2015; Jo *et al.*, 2017a; Xu *et al.*, 2017; Zhang *et al.*, 2017).

Identification of pathogens that cause disease in ginseng is important for improving the quality and yield of ginseng-derived substances. For example, fungal pathogens, such as *Cylindrocarpon destructans* and *Fusarium solani*, cause root rot of ginseng and lead to yield losses (Ohh, 1981; Jang *et al.*, 2010). Several viruses, such as a *Closterovirus* species (the

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

Abbreviations: ASPV = apple stem pitting virus; CP = coat protein; GRSPaV = grapevine rupestris stem pitting-associated virus; GVT = grapevine virus T; ORF = open reading frame; PgFV1 = *Panax ginseng flexivirus 1*; PVM = potato virus M; REP = replicase polyprotein; RdRp = RNA-dependent RNA polymerase; sgRNA = subgenomic RNA; SRA = Sequence Read Archive; TGB = triple gene block

family *Closteroviridae*) and watermelon mosaic virus (the genus *Potyvirus*, the family *Potyviridae*), have been identified to infect or be associated with ginseng (Mishchenko *et al.*, 2009; Jung *et al.*, 2013; Park *et al.*, 2017).

Application of next-generation sequencing methods has been demonstrated to be a rapid, efficient, and inexpensive method for the detection and characterization of viral sequences (Barba *et al.*, 2014). When transcriptomic analyses of plant samples are performed, genome sequences of plant RNA viruses present in the sample may be isolated together with host RNA molecules. Thus, RNA virus genome sequences can be found in plant transcriptome data deposited in publicly available databases such as the Sequence Read Archive (SRA) or Transcriptome Shotgun Assembly databases of the National Center for Biotechnology Information (NCBI) (Liu *et al.*, 2012; Jo *et al.*, 2017b). In this study, a ginseng transcriptome dataset was analyzed to identify RNA viruses that potentially infect ginseng (Wang *et al.*, 2015). As a result, a genome sequence of a novel *Foveavirus* species was found and annotated.

Materials and Methods

RNA-seq assembly. A previously published ginseng transcriptome data set (SRA accession number SRP066368) was analyzed (Wang *et al.*, 2015). The data set contained approximately 26.5 gigabases of paired-end reads obtained from 18 RNA samples, which were prepared from various tissues of ginseng at different ages. High-quality sequences were collected by filtering raw RNA-seq reads using the sickle program (version 1.33; <https://github.com/najoshi/sickle>) with the option “-q 30 -l 55.” The SPAdes Genome Assembler (version 3.10.1; <http://spades.bioinf.spbau.ru>) with the “-rna” option was used for *de novo* assembly of RNA-seq reads (Bankevich *et al.*, 2012). Eighteen transcriptome data sets were separately assembled into contigs.

Detection of virus-derived contigs. All RNA sequence contigs were analyzed to detect potentially virus-derived contigs containing a viral RNA-dependent RNA polymerase (RdRp) motif. The reference sequences of viral RdRps were prepared from the Pfam database (release 31.0; <http://pfam.xfam.org>). Pfam accession numbers for viral RdRp motifs were PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03431, PF04196, PF04197, PF05788, PF05919, PF07925, PF08467, PF08716, PF08717, and PF12426. A total of 394 non-redundant viral RdRp motif sequences were collected and converted into a custom-built BLAST-searchable database. BLASTX searches were performed against RdRp motif sequences using assembled contigs as queries with the parameter “-evalue e-5.” Mapping of RNA-seq reads to a viral genome contig was carried out using BWA program (version 0.7.16a-r1181; <https://github.com/lh3/bwa>) with the “mem” method (Li and Durbin, 2009). Sequence variants were

analysed using SAMtools/BCFtools programs (version 1.6; <http://www.htslib.org>) (Li *et al.*, 2009).

Multiple sequence alignments. The NCBI BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) searches were performed to identify and collect closely related viruses. Open reading frames (ORFs) were predicted based on BLASTX searches and ORF finder analysis (<https://www.ncbi.nlm.nih.gov/orffinder>). Functional domains of viral proteins were identified by using Pfam and TMHMM (version 2.0; <http://www.cbs.dtu.dk/services/TMHMM>) (Sonnhammer *et al.*, 1998). Pair-wise identities among protein sequences were calculated using the FASTA program (version 36.3.6; https://fasta.bioch.virginia.edu/fasta_www2/fasta_down.shtml).

Multiple alignments of protein sequences were generated using MUSCLE software (version 3.8.425; <https://www.drive5.com/muscle>) (Edgar, 2004). A phylogenetic relationship among viruses was inferred using the neighbor-joining method implemented in the ClustalW2 program (version 2.1; <http://www.clustal.org>) (Larkin *et al.*, 2007).

Results and Discussion

A 9002 nucleotide (nt) long contig isolated from a 12-year old ginseng root sample (SRA accession number SRR2952882) showed amino acid (aa) sequence similarity with the RdRp motif sequence of apple stem pitting virus (ASPV) (UniProt Acc. No., Q64962; Pfam Acc. No., PF00978). ASPV is the type species of the genus *Foveavirus* of the family *Betaflexiviridae* (Jelkmann, 1994), suggesting that the putative viral contig was derived from a foveavirus or related virus.

A BLASTX search against the non-redundant protein database using the NCBI BLAST confirmed that the putative viral contig was related to foveaviruses, including grapevine virus T (GVT) (Jo *et al.*, 2017b) and grapevine rupestris stem pitting-associated virus (GRSPaV) (Meng *et al.*, 1998). The contig was considered a novel ginseng RNA virus of the genus *Foveavirus* of the family *Betaflexiviridae* and named as *Panax ginseng flexivirus 1* (PgFV1). The PgFV1 genome sequence with annotation information is available in the NCBI nucleotide database under Acc. No. MH036372.

The ginseng root RNA-seq reads (SRA Acc. No., SRR2952882) were mapped to the PgFV1 genome contig. A total of 22,513 reads were mapped to the PgFV1 genome. There were only two single-nucleotide polymorphism sites at nt positions 52 and 55, indicating that the PgFV1 genome contig was assembled from a highly homogeneous viral population. The two single-nucleotide polymorphism sites were in the 5'-untranslated region of the virus genome.

Viruses in the genus *Foveavirus* (the family *Betaflexiviridae*) have flexuous and filamentous virions with a single-stranded positive-sense mono-segmented RNA genome

Table 1. ORFs and functional domains of PgFV1

Protein	nt position	nt length	aa length	Domain	Pfam	aa position
Replicase polyprotein (REP)	57-6476	6420	2139	Viral methyltransferase	PF01660	43-354
				Carlavirus endopeptidase	PF05379	1163-1250
				Viral (superfamily 1) RNA helicase	PF01443	1334-1599
				RNA-dependent RNA polymerase	PF00978	1711-2130
Triple gene block protein 1 (TGB1)	6559-7224	666	221	Viral (superfamily 1) RNA helicase	PF01443	24-221
Triple gene block protein 2 (TGB2)	7228-7575	348	115	Plant viral movement protein	PF01307	4-107
				Transmembrane		12-29
				Transmembrane		71-93
Triple gene block protein 3 (TGB3)	7502-7738	237	78	7 kDa viral coat protein	PF02495	2-63
				Transmembrane		7-26
Coat protein (CP)	7755-8633	879	292	Viral coat protein	PF00286	110-247

of approximately 8.4–9.3 kb (Martelli and Jelkmann, 1998; Martelli *et al.*, 2007). The natural hosts of foveaviruses include dicotyledonous woody plants such as grapevines, apple trees, peach trees, or apricot trees (Jelkmann, 1994; Meng *et al.*, 1998; James *et al.*, 2007; Youssef *et al.*, 2011). In many cases, foveavirus infection causes diseases manifested as mottle, mosaic, and ringspot symptoms in the stems and leaves. Foveaviruses are transmitted via grafting and no biological vectors have been reported.

The PgFV1 genome sequence was predicted to contain 5 ORFs (Table 1 and Fig. 1, top), which are commonly shared with other foveaviruses (Jelkmann, 1994; Martelli and Jelkmann, 1998; Martelli *et al.*, 2007). ORF1, which is the longest ORF, encodes a 2139 aa long replicase polyprotein (REP). REP was predicted to contain 4 known domains: a viral methyltransferase, carlavirus endopeptidase, viral RNA helicase, and RdRp.

The next 3 ORFs (ORF2, ORF3, and ORF4) encode for triple gene block protein 1 (TGB1), triple gene block protein 2 (TGB2), and triple gene block protein 3 (TGB3), respectively, which constitute tripartite movement proteins involved in the cell-to-cell movement of viruses (Kalinina *et al.*, 2002; Rebelo *et al.*, 2008). The sizes of these proteins are 221, 115, and 78 aa, respectively. TGB1 contains a viral RNA helicase domain, which shows approximately 17.5% aa sequence identity with the viral RNA helicase domain of REP. The two RNA helicase domains belong to the same Pfam domain family PF01443. TGB2 has a plant viral movement protein domain and TGB3 has a 7 kDa viral coat protein domain. Transmembrane domain prediction using the TMHMM showed that the TGB2 and TGB3 proteins contain 2 and 1 transmembrane domains, respectively. The presence of transmembrane domains in these proteins agrees with a previous report showing that these two proteins are membrane-associated and are essential for viral mobilization (Rebelo *et al.*, 2008).

ORF5 encodes a 292 aa long coat protein (CP). CP contains a viral coat protein domain that is shared with coat proteins of other *Betaflexiviridae* viruses such as members of the genera *Potexvirus* and *Carlavirus* (Rupasov *et al.*, 1989; Querci *et al.*, 1993).

ORF1 (REP) of the family *Betaflexiviridae* is directly translated from genomic RNA. In contrast, the other ORFs are translated from subgenomic RNAs (sgRNAs), which were transcribed from genomic RNA (Martelli *et al.*, 2007). It is thought that foveaviruses generate two sgRNAs, one for the 3 TGB proteins and another for the CP, although their presence in the plant cells has not been confirmed (Fig. 1, arrows). When RNA-seq reads isolated from the ginseng root sample (SRR2952882) were analyzed, a total of 22,513 reads were found to be derived from PgFV1. Interestingly, the sequencing depth plot of the PgFV1 genome showed an elevated sequencing depth at the 3' region, which approximately coincides with putative sgRNA segments (Fig. 1, graph). The observed read depth elevation suggests the presence of sgRNAs in ginseng cells. However, it is also possible that the pattern resulted from biased cDNA synthesis because of a local RNA structure or the experimental procedure employed.

For phylogenetic analysis of PgFV1, a BLAST search against the NCBI protein database was performed using the PgFV1 genome sequence as a query. A total of 11 related viral genomes were collected, including GVT, GRSPaV, peach chlorotic mottle virus (PCMV), apricot latent virus (ApLV), ASPV, apple green crinkle associated virus (AGCaV), rubus canadensis virus 1 (RuCV1), Asian prunus virus 1 (APV1), Asian prunus virus 2 (APV2), Asian prunus virus 3 (APV3), and potato virus M (PVM). The first 10 viruses are members of the genus *Foveavirus*. PVM, the type species of the most closely related genus *Carlavirus*, was included as an outgroup (Rupasov *et al.*, 1989).

Protein sequence identities between PgFV1 proteins and their respective orthologs of related viruses were calculated

Table 2. Protein sequence comparison of the REP, TGB1, TGB2, TGB3, and CP of PgFV1 and related viruses

No.	Acronym	Full name	RefSeq	REP	TGB1	TGB2	TGB3	CP
1	GVT	grapevine virus T	NC_035203.1	1282/2147 (59.7%) ^a	144/221 (65.2%)	69/115 (60.0%)	33/78 (42.3%)	157/219 (71.7%)
2	GRSPaV	grapevine rupestris stem pitting-associated virus	NC_001948.1	1146/2170 (52.8%)	137/221 (62.0%)	53/115 (46.1%)	24/74 (32.4%)	125/245 (51.0%)
3	PCMV	peach chlorotic mottle virus	NC_009892.1	980/2171 (45.1%)	94/221 (42.5%)	55/115 (47.8%)	19/62 (30.6%)	78/208 (37.5%)
4	AplV	apricot latent virus	NC_014821.1	979/2208 (44.3%)	96/222 (43.2%)	60/114 (52.6%)	23/64 (35.9%)	74/191 (38.7%)
5	ASPV	apple stem pitting virus	NC_003462.2	965/2216 (43.5%)	93/222 (41.9%)	55/114 (48.2%)	28/64 (43.8%)	71/193 (36.8%)
6	AGCaV	apple green crinkle associated virus	NC_018714.1	956/2216 (43.1%)	91/222 (41.0%)	55/114 (48.2%)	23/68 (33.8%)	68/202 (33.7%)
7	RuCV1	rubus canadensis virus 1	NC_019025.1	887/2208 (40.2%)	82/231 (35.5%)	42/108 (38.9%)	24/67 (35.8%)	91/286 (31.8%)
8	APV2	Asian prunus virus 2	NC_028868.1	887/2201 (40.3%)	78/203 (38.4%)	47/106 (44.3%)	26/67 (38.8%)	65/193 (33.7%)
9	APV1	Asian prunus virus 1	NC_025388.1	873/2202 (39.6%)	77/204 (37.7%)	45/106 (42.5%)	25/66 (37.9%)	87/289 (30.1%)
10	APV3	Asian prunus virus 3	NC_028975.1	872/2201 (39.6%)	76/203 (37.4%)	46/106 (43.4%)	23/66 (34.8%)	67/193 (34.7%)
11	PVM	potato virus M	NC_001361.2	816/2158 (37.8%)	89/222 (40.1%)	36/106 (34.0%)	15/52 (28.8%)	84/288 (29.2%)

^aAmino acid sequence identities in a format of "identical residues/aligned length (% identity)".

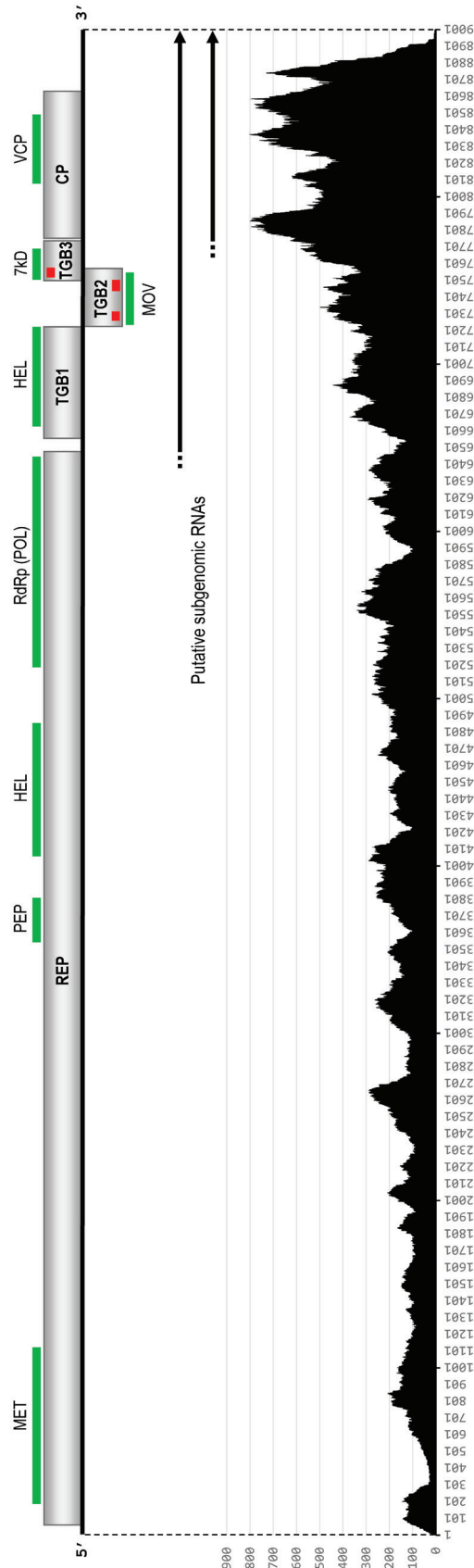


Fig. 1

Schematic representation of PgFV1 genome organization

At the top, ORFs are depicted as gray boxes: replicase polyprotein (REP), triple gene block 1 (TGB1), triple gene block 2 (TGB2), triple gene block 3 (TGB3), and coat protein (CP). Domains identified by Pfam are marked by green lines: viral methyltransferase (MET), carlaviral endopeptidase (PEP), viral RNA helicase (HEL), RNA-dependent RNA polymerase (RdRp or POL), plant viral movement protein (MOV), 7 kDa viral coat protein (7kD), and viral coat protein (VCP). Predicted transmembrane domains of TGB2 and TGB3 are indicated by red lines. Coordinates and lengths of ORFs and domains are presented in Table 1. Arrows indicate putative subgenomic RNAs, which were presumed based on other foveavirus genomes. Graph represents sequencing depth of the PgFV1 genome contig. The X-axis represents genomic coordinates and Y-axis represents sequencing depth. Note the elevated sequencing depth towards the 3'-end of the PgFV1 genome.

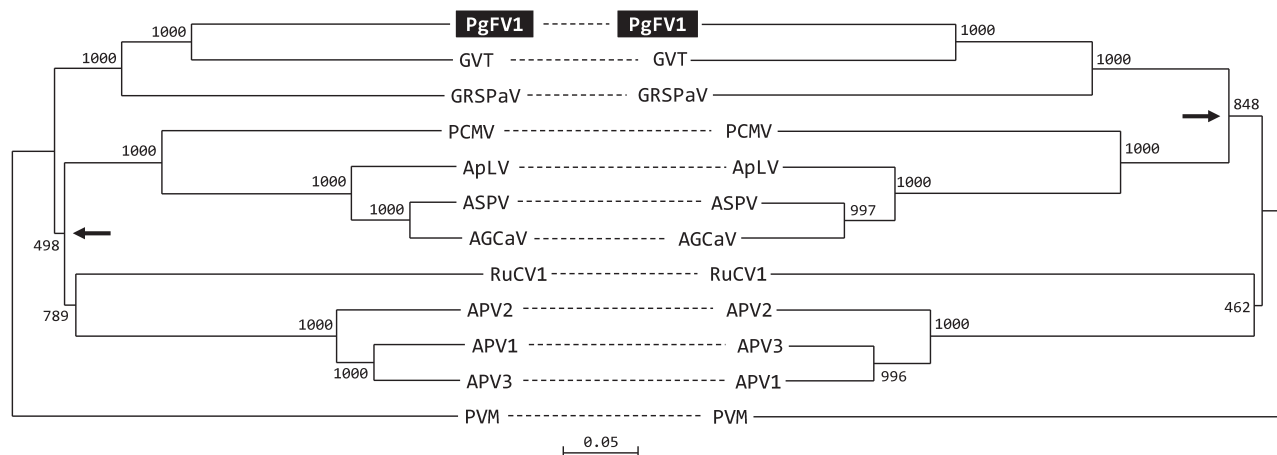


Fig. 2

Phylogenetic tree of PgFV1 and related viruses

Phylogenetic trees were inferred from multiple alignments of REP (left) and CP (right) sequences. Trees were rooted using sequences of PVM, the type species of the genus *Carlavirus*. The node showing a topology discrepancy between two trees is marked by an arrow in each tree. The bootstrap values obtained from 1000 replicates are shown at the nodes. Full names and accession numbers of viruses are presented in Table 2.

(Table 2). The highest aa sequence identity of the PgFV1 REP was found with the GVT REP (59.7% identity over 2147 aa overlap) and the second highest with the GRSPaV REP (52.8% identity over 2170 aa overlap). REP proteins of the remaining viruses showed lower identities (37–45%) with the PgFV1 REP.

The other PgFV1 proteins (TGB1, TGB2, TGB3, and CP) also showed the highest identities with respective proteins of GVT, indicating that GVT is the most closely related known virus to PgFV1. The sequence identity between the PgFV1 CP and GVT CP was 71.7% over 219 aa overlap. The CP or REP protein sequence identity threshold for assigning foveaviruses to different species is approximately 80% (Adams *et al.*, 2004), confirming that PgFV1 is a novel species.

Multiple sequence alignment of PgFV1 proteins with the respective orthologs of 10 other foveaviruses and 1 carlavirus revealed conserved regions mainly corresponding to known functional domains (Supplementary Figs. 1–5). Phylogenetic trees inferred from REP and CP sequences confirmed that PgFV1 is a member of the genus *Foveavirus* and formed a subclade with GVT and GRSPaV (Fig. 2). Phylogenetic tree topologies from two proteins were nearly the same except for a difference of the sister clade from the PgFV1/GVT/GRSPaV subclade. The tree inferred from REP sequences showed that PgFV1, GVT, and GRSPaV formed a sister clade of all other foveaviruses (Fig. 2 left). However, the tree from CP showed that the PgFV1/GVT/GRSPaV subclade was a sister clade of the PCMV/ApLV/ASPV/AGCaV subclade (Fig. 2 right). This discrepancy may be related to ancestral recombination events which have been reported for several foveaviruses (Komorowska *et al.*, 2011; Glasa *et al.*, 2017).

In conclusion, the genome sequence of PgFV1, a novel member of the genus *Foveavirus* of the family *Betaflexiviridae*, was identified from a ginseng root transcriptome data set. Sequence comparison and phylogenetic analysis indicated that PgFV1 is most closely related to GVT. PgFV1 is thought to be the first foveavirus found to be associated with ginseng although no disease symptoms have been confirmed. Given the important status of ginseng in the health industry and tendency of foveaviruses to cause diseases, further studies of PgFV1 and other foveaviruses potentially associated with ginseng are needed. The genome sequence of PgFV1 reported in this report may be useful for investigating ginseng foveaviruses.

Acknowledgments. This research was supported by the Chung-Ang University Graduate Research Scholarship in 2017.

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

References

- Adams MJ, Antoniw JE, Bar-Joseph M, Brunt AA, Candresse T, Foster GD, Martelli GP, Milne RG, Zavriev SK, Fauquet CM (2004): The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation. *Arch. Virol.* 149, 1045–1060. <https://doi.org/10.1007/s00705-004-0304-0>
- Baeg IH, So SH (2013): The world ginseng market and the ginseng (Korea). *J. Ginseng Res.* 37, 1–7. <https://doi.org/10.5142/jgr.2013.37.1>

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012): SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Barba M, Czosnek H, Hadidi A (2014): Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 6, 106–136. <https://doi.org/10.3390/v6010106>
- Briskin DP (2000): Medicinal plants and phytomedicines. Linking plant biochemistry and physiology to human health. *Plant Physiol.* 124, 507–514. <https://doi.org/10.1104/pp.124.2.507>
- Edgar RC (2004): MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Glasa M, Predajna L, Soltys K, Sihelska N, Nagyova A, Wetzels T, Sabanadzovic S (2017): Analysis of Grapevine rupestris stem pitting-associated virus in Slovakia reveals differences in intra-host population diversity and naturally occurring recombination events. *Plant Pathol. J.* 33, 34–42. <https://doi.org/10.5423/PPJ.OA.07.2016.0158>
- Im K, Kim J, Min H (2016): Ginseng, the natural effectual antiviral: Protective effects of Korean red ginseng against viral infection. *J. Ginseng Res.* 40, 309–314. <https://doi.org/10.1016/j.jgr.2015.09.002>
- James D, Varga A, Croft H (2007): Analysis of the complete genome of peach chlorotic mottle virus: identification of non-AUG start codons, in vitro coat protein expression, and elucidation of serological cross-reactions. *Arch. Virol.* 152, 2207–2215. <https://doi.org/10.1007/s00705-007-1050-x>
- Jang CS, Lim JH, Seo MW, Song JY, Kim HG (2010): Direct detection of *Cylindrocarpon destructans*, root rot pathogen of ginseng by nested PCR from soil samples. *Mycobiology* 38, 33–38. <https://doi.org/10.4489/MYCO.2010.38.1.033>
- Jayakodi M, Lee SC, Lee YS, Park HS, Kim NH, Jang W, Lee HO, Joh HJ, Yang TJ (2015): Comprehensive analysis of *Panax ginseng* root transcriptomes. *BMC Plant Biol.* 15, 138. <https://doi.org/10.1186/s12870-015-0527-0>
- Jelkmann W (1994): Nucleotide sequences of apple stem pitting virus and of the coat protein gene of a similar virus from pear associated with vein yellows disease and their relationship with potex- and carlaviruses. *J. Gen. Virol.* 75, 1535–1542. <https://doi.org/10.1099/0022-1317-75-7-1535>
- Jo IH, Lee J, Hong CE, Lee DJ, Bae W, Park SG, Ahn YJ, Kim YC, Kim JU, Lee JW, Hyun DY, Rhee SK, Hong CP, Bang KH, Ryu H (2017a): Isoform sequencing provides a more comprehensive view of the *Panax ginseng* transcriptome. *Genes (Basel)* 8. <https://doi.org/10.3390/genes8090228>
- Jo Y, Song MK, Choi H, Park JS, Lee JW, Lian S, Lee BC, Cho WK (2017b): Genome sequence of Grapevine virus T, a novel foveavirus infecting grapevine. *Genome Announc.* 5. <https://doi.org/10.1128/genomeA.00995-17>
- Jung W-K, Nam M, Lee JH, Park CY, Kim BH, Park EH, Lee M-A, Kim M-K, Choi H-S, Lee JS, Kim J-S, Choi JK, Kwon TR, Lee K-W, Lee S-H (2013): Novel pathogenic strain of Watermelon mosaic virus occurred on insam (*Panax ginseng*). *Res. Plant Dis.* 19, 331–337. <https://doi.org/10.5423/RPD.2013.19.4.331>
- Kalinina NO, Rakitina DV, Solovyev AG, Schiemann J, Morozov SY (2002): RNA helicase activity of the plant virus movement proteins encoded by the first gene of the triple gene block. *Virology* 296, 321–329. <https://doi.org/10.1006/viro.2001.1328>
- Kim SK, Park JH (2011): Trends in ginseng research in 2010. *J. Ginseng Res.* 35, 389–398. <https://doi.org/10.5142/jgr.2011.35.4.389>
- Komorowska B, Siedlecki P, Kaczanowski S, Hasiow-Jaroszewska B, Malinowski T (2011): Sequence diversity and potential recombination events in the coat protein gene of Apple stem pitting virus. *Virus Res.* 158, 263–267. <https://doi.org/10.1016/j.virusres.2011.03.003>
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007): Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Li H, Durbin R (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009): The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, Yi X, Jiang D (2012): Discovery of novel dsRNA viral sequences by in silico cloning and implications for viral diversity, host range and evolution. *PLoS One* 7, e42147. <https://doi.org/10.1371/journal.pone.0042147>
- Martelli GP, Adams MJ, Kreuze JF, Dolja VV (2007): Family Flexiviridae: a case study in virion and genome plasticity. *Annu. Rev. Phytopathol.* 45, 73–100. <https://doi.org/10.1146/annurev.phyto.45.062806.094401>
- Martelli GP, Jelkmann W (1998): Foveavirus, a new plant virus genus. *Arch. Virol.* 143, 1245–1249. <https://doi.org/10.1007/s007050050372>
- Meng B, Pang SZ, Forsline PL, McFerson JR, Gonsalves D (1998): Nucleotide sequence and genome structure of grapevine rupestris stem pitting associated virus-1 reveal similarities to apple stem pitting virus. *J. Gen. Virol.* 79 (Pt 8), 2059–2069. <https://doi.org/10.1099/0022-1317-79-8-2059>
- Mishchenko LT, Korenieva AA, Molchanets OV, Boiko AL (2009): Detection of viral infection pathogens in medicinal plants grown in Ukraine. *Mikrobiol. Z* 71, 55–61.
- Ohh SH (1981): Diseases of ginseng: Environmental and host effect on disease outbreak and growth of pathogens. *J. Ginseng Res.* 5, 73–84.
- Park D, Kim H, Hahn Y (2017): Genome sequence of a distinct watermelon mosaic virus identified from ginseng (*Panax*

- ginseng) transcriptome. *Acta Virol.* 61, 479–482. <https://doi.org/10.4149/av.2017.410>
- Querci M, van der Vlugt R, Goldbach R, Salazar LF (1993): RNA sequence of potato virus X strain HB. *J. Gen. Virol.* 74 (Pt 10), 2251–2255. <https://doi.org/10.1099/0022-1317-74-10-2251>
- Rebelo AR, Niewiadomski S, Prosser SW, Krell P, Meng B (2008): Subcellular localization of the triple gene block proteins encoded by a foveavirus infecting grapevines. *Virus Res.* 138, 57–69. <https://doi.org/10.1016/j.virusres.2008.08.011>
- Rupasov VV, Morozov S, Kanyuka KV, Zavriev SK (1989): Partial nucleotide sequence of potato virus M RNA shows similarities to protexviruses in gene arrangement and the encoded amino acid sequences. *J. Gen. Virol.* 70 (Pt 7), 1861–1869. <https://doi.org/10.1099/0022-1317-70-7-1861>
- Sonnhammer EL, von Heijne G, Krogh A (1998): A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 175–182.
- Vuksan V, Sievenpiper J, Jovanovski E, Jenkins AL (2010): Current clinical evidence for Korean red ginseng in management of diabetes and vascular disease: A Toronto's ginseng clinical testing program. *J. Ginseng Res.* 34, 264–273. <https://doi.org/10.5142/jgr.2010.34.4.264>
- Wang K, Jiang S, Sun C, Lin Y, Yin R, Wang Y, Zhang M (2015): The spatial and temporal transcriptomic landscapes of ginseng, *Panax ginseng* C. A. Meyer. *Sci. Rep.* 5, 18283. <https://doi.org/10.1038/srep18283>
- Xu J, Chu Y, Liao B, Xiao S, Yin Q, Bai R, Su H, Dong L, Li X, Qian J, Zhang J, Zhang Y, Zhang X, Wu M, Zhang J, Li G, Zhang L, Chang Z, Zhang Y, Jia Z, Liu Z, Afreh D, Nahurira R, Zhang L, Cheng R, Zhu Y, Zhu G, Rao W, Zhou C, Qiao L, Huang Z, Cheng YC, Chen S (2017): *Panax ginseng* genome examination for ginsenoside biosynthesis. *Gigascience* 6, 1–15. <https://doi.org/10.1093/gigascience/gix093>
- Youssef F, Marais A, Faure C, Barone M, Gentil P, Candresse T (2011): Characterization of *Prunus*-infecting apricot latent virus-like foveaviruses: evolutionary and taxonomic implications. *Virus Res.* 155, 440–445. <https://doi.org/10.1016/j.virusres.2010.11.013>
- Yun TK (2001): Brief introduction of *Panax ginseng* C. A. Meyer. *J. Korean Med. Sci.* 16 (Suppl.), S3–5. <https://doi.org/10.3346/jkms.2001.16.S.S3>
- Zhang JJ, Su H, Zhang L, Liao BS, Xiao SM, Dong LL, Hu ZG, Wang P, Li XW, Huang ZH, Gao ZM, Zhang LJ, Shen L, Cheng RY, Xu J, Chen SL (2017): Comprehensive characterization for ginsenosides biosynthesis in ginseng root by integration analysis of chemical and transcriptome. *Molecules* 22. <https://doi.org/10.3390/molecules22060889>
- Zhen G, Zhang L, Du Y, Yu R, Liu X, Cao F, Chang Q, Deng X, Xia M, He H (2015): De novo assembly and comparative analysis of root transcriptomes from different varieties of *Panax ginseng* C. A. Meyer grown in different environments. *Sci. China Life Sci.* 58, 1099–1110. <https://doi.org/10.1007/s11427-015-4961-x>

Supplementary information

Novel *Foveavirus* (family *Betaflexiviridae*) species identified in ginseng (*Panax ginseng*)

D. PARK¹, M. ZHANG², Y. HAHN^{1*}

¹Department of Life Science, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, South Korea;

²College of Life Science, Jilin Agricultural University, Changchun 130118, Jilin, P. R. China

Received July 17, 2018; accepted August 7, 2018

Fig. S1. Continued

PgFV1 386 KNFLNEM SLFFSCVGRLESIEEEKCLADFIYDLKPKFKVTKVA K--RDFEDBIHFSLT-----DPNPDYD IAKL
 GVT 386 KDCKNEL GLFFGFI SRMFOIKDKCLSDFIYDLEPFNFYVDIVD K--GDESDFEFDWIDH-----DLLDSFDVSEGI
 GRSPaV 387 KLVISFFCR MPNARLSSPRECSLDSFYSLPEFNFVSNLVD T--PDFFHFLFSL-----NELIEEDVEVM
 PCMV 398 KLSNFFI GLFPDSVARNKVEQQMSLDNFISEMEFNFVOTST S--LSWLDDIRFDIN-----FDIKPVDIDLMF
 ApLV 388 RIVQGFY KLFENP SRNEKVVQQLHLDNFIEIETLEEFNFNTES S--LNKDDIEFVNT-----FGTDFNVEDSF
 ASPV 388 RIVQGFY KLFENP SRNEKVVQQLHLDNFIEIETLEEFNFNTES S--LNKDDIEFVNT-----FGTDFNVEDSF
 AGCaV 388 RIVQGFY KLFENP SRNEKVVQQLHLDNFIEIETLEEFNFNTES S--LNKDDIEFVNT-----FGDIDFNVEESF
 RuCV1 388 KIFLHSL SLFFSSWVRNNSFOEITLDNFISLALLRVVELK--K--LDESDDL FSDPGLSHEEEFPSIPFDPQML
 APV2 386 KLGALFASFLPFKARFEDCKSSSLDKFIHDLEPFSSFSVPEVIN--CNTMPAFEAISIG-----ESMSPEVVIQIL
 APV1 387 KLGSLFSSLPFKARFEDCKSSSLDKFIHDLEPFSSFSVPEVIN--SRSNPMEAVTIG-----ETMSSEVILKIL
 APV3 386 KLGALFASFLPFKARFEDCKSSSLDKFIHDLEPFSSFSVPEVIN--CRFSQTEAVTIG-----ETMGPEVMOIIL
 PVM 387 KEFMGNW GKMPSVLARRSSVRAVCNKGIRGLKQVSEFIRLNE TWWNLWENSYAWFF-----DTDAEVDVPEKL
 consensus 401 * * *

PgFV1 458 DDDVFGVDKSA--DRIAQRVC NSQFWRISRLRFYSKFEC-----SAVYSICDHFIDCPGLNFEIVINIRSV
 GVT 458 DDMKVTIKNWN--TRSPQSMF GGREWSSSSSKFYSKFGA---FERVGD--LYRYLIQDL-----SHDMVNEEVVRRIL
 GRSPaV 459 DNSFGLGLDQF--NRQRAPF GSSYWLNSKFSVEHKFSQ-----TNSQMQVILSLIPFSDPTFRPSSTEVVALS
 PCMV 470 QKGGLVKEVTIL--ERSREPMFYSKEDSYDAQTSVPSTN---FEGFAAISYIRSR-----KPLLPCSWYFENVRDM
 ApLV 460 DSTGVKKDHVNI TVVHHPMFVSKFEAYEYQESLNDS--TAAATR--FAKIVVSLY-----SVSLVESYDCKSCSL
 ASPV 460 AEA GTKKDVVNI TVVHSPMLVSKFESYDHOESLSKS--ISALTR--TAKIVLSLY-----DPCVVEAFSESRVTNL
 AGCaV 460 ASAGTKRDVVDIT TVHSPMLVSKFESYDHOESLSLSTNS--TTALTR--TAKIVLSLY-----DPCVVGAFSECRVSNL
 RuCV1 463 ES--SIGRSKKAIPDRVSPMNNLIE-----EGAEVVDTRDKSVLIQMYHCVFNA-----SNSDGFVSSRTFFSV
 APV2 458 NRMEMG-----AFAGSCNLV GKSKYSNEAVCHHLDGCKPEVFKR--LIRRAGSE-----CNEFGYRKSYSVVASV
 APV1 459 AKLDMG-----APTSLEGLTKKYSNEAVCHHLDGCGPHVFR--LVRKADSE-----CNEFGYRKSYSVVASV
 APV3 458 RRMELS-----APVRLNLTGKYSNEAVCHHLDGCGPHVFR--LVRKADSE-----CNEFGYRKSYSVVASV
 PVM 459 DSLMGEV--AGLVAHITSRPYGVTVPLADRENAALCDSQKLLHAR MFMGAWGAHMCVISREFLLYVEAR KSS
 consensus 481

PgFV1 529 ---KARKL-----DPQLCVVIND---FFDVSFKLSR---KLRDRLNKVDARVLDLWF--LKLNRVNSKFTSNTGK-
 GVT 526 RDFRRRQ-----TESLAMISCNLDI--TALKITVEN---WTLEARKENLRVVRPWF--LLVGRINTKFTCESSSS-
 GRSPaV 532 EVKAALAT-----GQSKLRFLVDDCAMREV--RSSYKVLGFKHIKALTHCNSCGLQWF--LLQORSNLKFLKDRASS-
 PCMV 542 AIGSLVKG-----GLGKSVSISCTADLRLAFDSSVKP-----KLHQPPPLIEWVSVLWF--FKNRINCHFTREYSE--
 ApLV 532 ASEVIVSA-----NLRACEVVDLWKTFRALIKREGNR---SKGRMRKRYFFELGKWF--LFTNAANVKELTPGRD--
 ASPV 532 AVNVIIAA-----NLRACEVVDLWRIFFGILLRECKR---AQGKMRKRYFFELGIRWF--LFVDVSNQWFLPPCRD--
 AGCaV 532 AVNVIIAA-----NLRACEVVDLWKTFFGILLRECKK---AQGKMRKRYFFELGIRWF--LFVDVSNQWFLPPCRD--
 RuCV1 530 ADGYARNSIFSNAKALRLAFTMSCSGLRLVDVYLNIRRKIFFVRSPTSQAQIRTANLQWF--FLKRGNOKFLRDSAFSD
 APV2 525 ANVVFQRRS-----NLNLDLILGSDLNQKVVCIENFES---SKSSQLIWFSTNSMAWRFS--LRANOKFLRSPIFP-
 APV1 526 AITVLRKKT-----NSNLSIFGTDLNPMLIGFVENYFA---SRSSLLIWFHRPESWQFS--LRANOKFLRSPIWP-
 APV3 525 AITVLRKKT-----NSNLSIFGADLNPMLIGFVENYFS---SKKASLLIWFHRPSSWQFS--LRANOKFLRSPIFP-
 PVM 537 CIIAKARRR-----GQHKE--KLEAWEVLGLKSSDALRAMTYLCNARLEPMES--SGRRE--LTRGNLNYGLTNYTEG-
 consensus 561 * *

PgFV1 593 ENGDRTSQSRYNAL I-----SDLTSRKDFNLKVVWGF-----KLAISKDVR
 GVT 593 ISNPSVESVNLRSIV-----KEVEKFKLESLDVFW-----QTSSA
 GRSPaV 603 FADFDCEVIKVYQLT-----SQAILPEALLSLTKVFRD-----SDSGVSTP
 PCMV 609 --D-YCR-----ADFRSFKIFKSRFLN-----EVSNGSKIL
 ApLV 599 --GFVRSLNEDRFK-----NCQCNLSFHRGRMALH-----QTIKGERIE
 ASPV 599 --GFIARSVSFDQFK-----GCQRDNSLHNCRMSLR-----QVLGPKIQ
 AGCaV 599 --GFIARSVLFQDFK-----GCQRHNSFHNGRMALR-----QVFKFKIQ
 RuCV1 609 ESNPETKVFTSWMKRVDEVLINSSSSSLGMRSTLSDSLNLWEYNCSTTSYPAPSATESVEKV-----EEIEEESA
 APV2 595 --ASNCKDLSTRKN I-----LDEFVGFCAGLKFGSSGVSSVVPKPVGGFEFVP-----PGEEAEDDIE
 APV1 596 --NDGLVEFRMRKNA-----LEISNDFRNSNLKFGSFQAPTNKSNSTSSSEGASNHQREQ--RDQCDEATIQ
 APV3 595 --DQGLSDFRIQKST-----AEVKEFDRKLGKFGSFLNSRPYQVQDQLPPLKSVNEANCVSGDIEQGSNPE
 PVM 608 --KRAVTGVQNLWSNV-----HEVSTKRHKCMIRLE-----KARVTEQPR
 consensus 641

PgFV1 637 VSINGDELND--SPHSLKSD--LVSSNFQNEEVLTEQNGSVK-----GDCSIREDSGEDK-----SFEV
 GVT 631 MEFGGPKFEISEEPVVEVEEQC--APEEVIDVGSKQTSVAH--HAES--QATAPSYVDCT-----SVSI
 GRSPaV 647 RLVSRNELELAHPANSALEP--QSVDCNAGVQAS--SSSQ--QLADTHSLGSKSSSIETANK-----AFNL
 PCMV 638 NHCTSPIFFGLESSLTSTITD---EVPPICTPKS--TPDALSDS-----NFPTFIHSC--GNLR-----CAKL
 ApLV 638 SLLDEYELNSHKKFKVEKQDVRT--VEEIEKQDPHTFENHENSSP-----ATVGC SLVPCACSTN-----CFVQ
 ASPV 638 ALFDVSELSIIHNVENAPBA--GSTLDAGIKPTSSPLEVPIE-----NARCNLAPCKCDLN-----CFIQ
 AGCaV 638 ALLSELSVVMHVEIRTVPEV--EPTVEDEQHLTAPSEVPIIT-----TVGNVVPCKCNLN-----CFIQ
 RuCV1 682 THENVPILKRDVSSFLKIAFNV--PTVQEEVVEKDVET--GTEE-----EFEAPALEKFGDAGI-----VFGR
 APV2 658 VKSNTMAKMFDPSPQAVLGI--G--PAAGVHEIDPQDTLSPKAADLEERNGLIFGCVPEPSLMNFSFSSANQDRLFANIV
 APV1 663 VNFQRIEFCMGLPEANSFKSFTPCV--EVAEAETSVEVQPEPEP--EANGLVLCI--PDASTPAFTFNSANQDRLFSSV
 APV3 663 VCFAMEMEFGRRS--PMSMGTSE--ILAPN--AAELAVKEV--SDSTAPER--KHEGLVFGC--IPDSS--TGAF--FCSANQHE--TFAAVV
 PVM 647 SEFASCVLPEVWRDVEAAL--IELGEVACACNARFVQGV--LSNQ-----
 consensus 721

Fig. S1. Continued

PgFV1 696 EGFNIKVAEPP-DLAQFASLLSEPDQ-----GRKAGFFAVPEVSHYQVNCGFHENRWGPPF-SNLIKICGG-FGHYNS
 GVT 689 DEIKIDIGNSS-ELKFHYGVLKSPDVI-----KGRKAGFAKSGIKSYAVNCGSEHDSMGWEKALDEIILRLVDG-GLVYNS
 GRSPaV 711 EELRIMRVLPEDFNWVAKN-GERDRI-----RGRGASFFSKPGISCHSYNCGSEHDSMGWEKALDEIILRLVDG-GLVYNS
 PCMV 696 RSDCTAIP-----LNLDPEPKI-----NNTRVACFYSR-RGDGYSYTCFSEKSMGWEDFLDKELLNQIALENYDH
 ApLV 700 NAKVDAPH-----NNLTPEDFI-----GGARGGATFFSR-DLKGYSYTCFSESRGWFRFLDDLDLONIPKHBYDQ
 ASPV 699 PADVNSH-----GNLVLDFFI-----GGSKRGASFYSR-DLKGYSYTCFSEVSRGWFAFLDKELSNKIPLNBYNQ
 AGCaV 699 QADVSSH-----GNLVLDFFI-----GGARGGASFYSR-NLKGYSYTCFSEVSRGWFTFLDKELDNKIPOKBYNQ
 RuCV1 744 VKDITLGLCLD-MV----FNENLHDVI-----HLKGRSAVETRRCPCCLRYGHNLQYLPQKWEIFEDWELG-NL---GKYNS
 APV2 737 NENVDGILP-P----TELKIPLDAKVTOINGRDCYFFTRCCSIDYGHNKIKYTANNWFSGLDQELSNEG---TAYNA
 APV1 741 EENTDGILS-P----MELVLPSSAEATPLNGROCYFFTRCCGIDYGHNKIRYKPNRWFSGLDQELSNEG---TAYNA
 APV3 741 NENTDGVLV-P----TELVLPDAEATIVRGROCYFFTRCCGIDYGHNKIRYKPNRWFSGLDQELSNEG---TAYNA
 PVM 691 -----AGLVNRQV-----AGASVGLVTK-DRSNLKGNSELLSNVGRSLSVMEINSV-SQKEDV
 consensus 801 *

PgFV1 769 CIAQLYDDGSDIKHRDNEKCYNENHKVLTVC-RGSCITNFCSTCKSKSRGDWVSEVGPFEWEMPRSFQAKMLHGVYN
 GVT 762 CIAQRVDDGASIGMHSNDRCYAVGHKVLTVNIGSCQFYTCRKQDRNLKERHINFRLREGDYFEMPRNFQENLHGVR
 GRSPaV 785 CIAQIYEENSKLAIHKDDPSCYEIGHKVLTVNIIGSATFTI-SKSRNLVGGNHCSTIGPNEFEMPRGMOCNPHGVSN
 PCMV 761 CLVQKYSQCAALCWHSDNEDCYDLDHQVLTVNIIGSATFTIIGSATFTI-SKSRNLVGGNHCSTIGPNEFEMPRGMOCNPHGVSN
 ApLV 767 CLVQEYSTCHGLAMHRDDPEIYDPNEQVLTVNIIGSATFTIIGSATFTI-SKSRNLVGGNHCSTIGPNEFEMPRGMOCNPHGVSN
 ASPV 766 CLVQEYSTCHGLS-NHKDDPEIYDPNEQVLTVNIIGSATFTIIGSATFTI-SKSRNLVGGNHCSTIGPNEFEMPRGMOCNPHGVSN
 AGCaV 766 CLVQEYSTCHGLS-NHKDDPEIYDPNEQVLTVNIIGSATFTIIGSATFTI-SKSRNLVGGNHCSTIGPNEFEMPRGMOCNPHGVSN
 RuCV1 813 CLVQRBEVCAKICFHSDDPKHYSSDNDIYTVNIMGNAQLSTRPKCDKRK-ANEITRALVSQDS-IMPSGFDKREHSRS
 APV2 808 CLVQVYRAGSGICFHSDDPKHYSSDNDIYTVNIMGNAQLSTRPKCDKRK-ANEITRALVSQDS-IMPSGFDKREHSRS
 APV1 812 CLVQVYRAGSGICFHSDDPKHYSSDNDIYTVNIMGNAQLSTRPKCDKRK-ANEITRALVSQDS-IMPSGFDKREHSRS
 APV3 812 CLVQVYRAGSGICFHSDDPKHYSSDNDIYTVNIMGNAQLSTRPKCDKRK-ANEITRALVSQDS-IMPSGFDKREHSRS
 PVM 746 AVRLSYSKETQANLLPSLDGIERGAGATVVNIRKCAPIVRCARGWR-----DALAWMDHICLVMANVAAGHECYM
 consensus 881 *

PgFV1 849 CKGERISLTFR-----RHIVQDNETLQIP---NLNKNLAAFERVKALDVKVW
 GVT 842 TSDGRISLTFR-----RQLVPDSDGDEEE---CSPFAFGSIDRIAGR--KFW
 GRSPaV 864 CTPGRVSLTFR-----RQKLEDDDLIFINPQVPIELNHEKLDL---SMW
 PCMV 835 TSEGRVSLTFR-----KSKACLNGIS-TLVQGAQGGPIS-----TPSDASHSLAHLQ
 ApLV 841 PSKGRISLTFRLRSASNQVPVQEVVVKV-DQDGAVKGASLEKLPSSQQTQDESMLGTGVSSVFDEGFSADSSGSSPVQEFM
 ASPV 840 PSKGRISLTFRLRKEGDSQVPIQEVVVICDHGSDDDRAALKALE-RSHQSGGRPAVELEGHEREKVNSSDSDSAPVQEFM
 AGCaV 840 PSKGRISLTFRLAKEDDHQVPIQEVVAI-EHGDSDDDRAALKQME-KGFFDGGKPPPTDSGEQTRKVTGDCPGSAPVQEFM
 RuCV1 892 MTEGRVSLTFR-----RQKLEDDDLIFINPQVPIELNHEKLDL---SMW
 APV2 886 CSEGRVSLTFR-----RQKLEDDDLIFINPQVPIELNHEKLDL---SMW
 APV1 890 CSEGRVSLTFR-----RQKLEDDDLIFINPQVPIELNHEKLDL---SMW
 APV3 890 CSEGRVSLTFR-----RQKLEDDDLIFINPQVPIELNHEKLDL---SMW
 PVM 819 RSWGTMDVVF-----LKRATVSEQVTFESA-----
 consensus 961

PgFV1 893 -----NLGKMMADLKTKRSCFISDLCAICAFSCSNFDFKFSADLSKYSKALGAEVVAEDRI-ILFDQTAKEPIFRDALSR
 GVT 884 -----PECSNYAAMIDGSGKGRELDLDCDFSCPNHFRAKNIDGRKVFSAICLAAGDRV-LVLNSKSSKATLEKFERL
 GRSPaV 905 -----QMGLHGKKSISMNGTSFTSDICSCGFSCNHFHKFDLNNRLALGAQGLGQCDRVVFATGPGLSKVLMPRSK
 PCMV 883 VDVDGSVVLEIRKVGKFGKGYQSDICCCNMSWATDEDEPILETIRSLSFACFSNVDRV-LISDVNSITLSSLLLE
 ApLV 920 IQIDSSLEAYAKSISLCKNVDLINDICLGNPWLKNEELKFSEARDLAFASCFNPTDRFSLAGVSGVRGNRIISEL
 ASPV 920 IQIDSSLEAYAKSISLCKNVDLINDICLGNPWLKNEELKFSEARDLAFASCFNPTDRFSLAGVSGVRGNRIISEL
 AGCaV 919 IQIDSSILEYAKSISLCKNVDLINDICLGNPWLKNEELKFSEARDLAFASCFNPTDRFSLAGVSGVRGNRIISEL
 RuCV1 903 -----KVVDHSLGQ-----
 APV2 897 -----HHVNNVAGL-----
 APV1 901 -----YHVNNAAGL-----
 APV3 901 -----YHVNNAAGL-----
 PVM 845 -----QEVGPIECK-----SDSGAPCV-----
 consensus 1041

PgFV1 967 PVKIFVLECEQRVSDSEDLILKSMQCGVV---RNFSISKATCQLGFAF-REKISSRGSIRVHIECEKRSVGSSEVLLGC
 GVT 958 RVTIAVFSCKIEFLDGLVPGIVSDGLREV---KNFSVONLKVLAIVIF-RGSTVSGSKVRVHGECKEFGASEVLLGC
 GRSPaV 980 KQSLVLECALSIETDYGPKVLGSEVFK---GDFHKMEEGSIFVITYKAPIRSTGRVHSSSECSFVGSKEVLLGC
 PCMV 962 MGTIWCISCFILVKSEGEVVKIGEMMASKF---MKMSIIGWSKDFLSFFYKPRLGKGMQIRTHNECELSDFTEQLGFC
 ApLV 1000 PSHVPLRCMSIILDIDDKTVKGNVKEGFSGFRRWK-VSCSTDLVVAFKPKMTQGGERTHEDECELSDLTEKILHG
 ASPV 999 PTHVPLRCMHIIVDLDDKSIKGDVKEGFSGFRRWK-VSCSTDLMVAFKPKMTLGGERTHEDECELSDLTEKILHG
 AGCaV 999 PTHVPLRCMHIIVDLDDKSIKGDVKEGFSGFRRWK-VSCSTDLMVAFKPKMTLGGERTHEDECELSDLTEKILHG
 RuCV1 912 -----EIVHDSICLLDCFENLFDV-----
 APV2 906 -----PIKHTCGEECDTEILFDV-----
 APV1 910 -----PIKHTCGEECDTEILFDV-----
 APV3 910 -----PIKHTCGEECDTEILFDV-----
 PVM 862 -----GVNLDLGGV-----
 consensus 1121 *

Fig. S1. Continued

PgFV1 1424 LFPPGYDLICLLICDSSCTVAAGDPCQSTYDYSKEDRNHLCTMDPDIFFELRNKEMSYNLESRRREKNPFEKRLPCEFLR
GVT 1415 LFPPGYDLICLLVSDKSCIFSVAGDPCQSGYDSKSDRNILGALDSDIFFELRNKEMSYNLESRRREKNPFEKRLPCEFLR
GRSPaV 1442 LFPPGYDLICLLIRSDAFISLAGDPCQSTYDYSKDRALGAEQSDILRLLEGKRYRYNLESRRRFVNPFEKRLPCHFKK
PCMV 1418 LFPPGYDLILVCTSSDINVLVAGDPCQSDYDSKDRHLEANSYSDIHLNNGKYRYNLSORERFNPFEKRLPCEFLR
ApLV 1461 LFPPGYDLILGLKQNVIIILAGDPCQSDYDSSDRHLEFAGS SDIMRILSGRKYKFNLSORERFNPFEKRLPCEFLR
ASPV 1460 LFPPGYDLILGLKPNVIIILAGDPCQSDYDSSDRHLEFAGS SDIMRILSGRKYKFNLSORERFNPFEKRLPCEFLR
AGCaV 1460 LFPPGYDLILGLKPNVIIILAGDPCQSDYDSSDRHLEFAGS SDIMRILSGRKYKFNLSORERFNPFEKRLPCEFLR
RuCV1 1322 LPPGYDILLCVLTSSDCKVFLTGDPQSDYDSDKDRMLEYGMPEDIMHLLNEKSYNFNLESRRRFVSGFEKRLPCEFLR
APV2 1288 LPPGYDLVMLCISGNCIYLSGDPQSDYDSAKDRALDGLKGDIFELEKGYKFNASSRRRFQSQMFVGRLPCEFLR
APV1 1305 LPPGYDLVMLCISLNCIYLAGDPCQSDYDSAKDRALDGLKGDIFELEKGYKFNASSRRRFQSQMFVGRLPCEFLR
APV3 1305 LPPGYDLVMLCISLNCIYLAGDPCQSDYDSAKDRALDGLKGDIFELEKGYKFNASSRRRFQSQMFVGRLPCEFLR
PVM 1249 LPPGYDILVSMKVDVRFVLDGPAQSDYDSEKDRILGAMENMSVILGAEYNYKRSRHLNLCNIGLRLPCEFLR
consensus 1601 *****

PgFV1 1504 GSHLKIEYAIYNSMADFIYAKKEKP-DVYLVSSFEKKIVSSHAGSSKCLTFGESTGLNFQGVILLIYDSIHTDDR
GVT 1495 GSHEKHQDFVWTKSMEDLVFGEKFKI-DAFLVSSFEKKIVSSHFQSAATCLTFGESTGLNFQGVILLIYDSIQDDN
GRSPaV 1522 GSTAAFAFYAIYHNHDFLLARSGPLDAVYLVSSFEKKIVQSYFGMKQLTLTFGESTGLNFQGVILLIYDSIHTDDR
PCMV 1498 KRIMDEEYTLWESITQFELAGGKNF-PVYLVSSFEKKIVAAHLGLK KCLTFGESTGLNFQGVILLIYDSIHTDDR
ApLV 1541 SRITLDEEYTLWESIQEFMSYGRKDC-PVYLVSSFEKKIVAAHLGLK KCLTFGESTGLNFQGVILLIYDSIHTDDR
ASPV 1540 TRITLDEEYTLWESIQEFMSYGRKDC-PVYLVSSFEKKIVAAHLGLK KCLTFGESTGLNFQGVILLIYDSIHTDDR
AGCaV 1540 TRITLDEEYTLWESIQEFMSYGRKDC-PVYLVSSFEKKIVAAHLGLK KCLTFGESTGLNFQGVILLIYDSIHTDDR
RuCV1 1402 GTACFNPKNMEYMGHIKSIDKNLIKSV-DVYLVSSFEKKIVCWSHLGAGNKCLTFGESTGLTFESCIILLIYDSIHTDDR
APV2 1368 KDIVGDEDEHWLSEIEAAAEVSNTEY-DVYLVSSFEKKIVAAHLGRDEVLTFGESTGLTFENGIILLIYDSIHTDDR
APV1 1385 KAVTENENDEHWLSEIEAAAEVSNTEY-DVYLVSSFEKKIVAAHLGRDEVLTFGESTGLTFENGIILLIYDSIHTDDR
APV3 1385 KAVTENENDEHWLSEIEAAAEVSNTEY-DVYLVSSFEKKIVAAHLGRDEVLTFGESTGLTFENGIILLIYDSIHTDDR
PVM 1329 DDCHIDEPHIRMHENLLDVAEYK-SVYLVSSFEKKIVAAHL-PEAKVLTFGESTGLTFENGIILLIYDSIHTDDR
consensus 1681 *****

PgFV1 1583 RWTALSRFRMNVSEINLIGMPLSGAHEEAGKPLYHFTLESSGQSVIIDMLPGSEFFSFFKISVGRDEGVKRNKAVG
GVT 1574 RWTALSRFRMNVSEINLIGMPLSGAHEEAGKPLYHFTLESSGQSVIIDMLPGCEKFKNGFSISVGRDEGVKRNKAVG
GRSPaV 1602 RWTALSRFRSHNDLVNITGLRVSFLESAGKPLYHFTLAKSGENVIRDMLPGSEFFSFFKISVGRDEGVKRNKAVG
PCMV 1577 RWTALSRFRSHETHEINGSTIENWVSEIHEGKALNFEFKRASHDDVVDMLPGSEFFSFFKISVGRDEGVKRNKAVG
ApLV 1620 RWTALSRFRSHDHEINGGVTVNALTIEVVKPLHKKFTKACNDIIDMLPGSEFFSFFKISVGRDEGVKRNKAVG
ASPV 1619 RWTALSRFRSHDHEINGGVTVNALTIEVVKPLHKKFTKACNDIIDMLPGSEFFSFFKISVGRDEGVKRNKAVG
AGCaV 1619 RWTALSRFRSHDHEINGGVTVNALTIEVVKPLHKKFTKACNDIIDMLPGSEFFSFFKISVGRDEGVKRNKAVG
RuCV1 1481 RWTALSRFRKLVESNLTNLSYDIAQAFCEPFLYKFTKASTDDILEILPGSEFFSFFKISVGRDEGVKRNKAVG
APV2 1446 RWTALSRFRNLIEFNVLGNNDVQCVEHDTLEKFNLRATATIGDIIKOLPGPELNTDFGDRVGRSEGVMEAKLSD
APV1 1463 RWTALSRFRNLIEFNVLGNCLDACQVEHDTLEKFNLRATATIANIDOLPGPELNTDFGDRVGRSEGVMEAKLSD
APV3 1463 RWTALSRFRNLIEFNVLGNCAVQCVEHDTLEKFNLRATATIANIDOLPGPELNTDFGDRVGRSEGVMEAKLSD
PVM 1406 RWTALSRFRFNLCVNCSEMDYQQLAGYKGRVRSKFLCKTAPDDINSILPGQALFKSEYPRILGKDEGVKRNKAVG
consensus 1761 *****

PgFV1 1663 PWLKTMIIFLGQEDMEVEHEKEMSSSEWFKTHIPISSVENCRRARWVKIALKEAREFRIGMEVEEQFQKDDHDFRG-ER
GVT 1654 PWLKTMIIFLGQEDIEHEEEMEEFKNEWFKTHIPISSVENCRRARWVKIALRESREVRCCMNTTEQFQKDDHDLKGV-L
GRSPaV 1682 PWLKTMIIFLGQEDCEVEEEMEESECSNEWFKTHIPISSVENCSTRARWVKIALKEAREFRIGMEVEEQFQKDDHDFRG-EQ
PCMV 1657 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAVAVRQVVKIAREAREFRIGMEVEEQFQKDDHDFRG-KQ
ApLV 1700 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-LE
ASPV 1699 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-LE
AGCaV 1699 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-LE
RuCV1 1560 PWLKEKIFLGQSEDIQEEIYQYBEIKNEWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-SA
APV2 1526 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-VT
APV1 1543 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-IT
APV3 1543 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-IT
PVM 1486 PWLKTMIIFLGQEDIEHEEEMEEIIEQWFKTHIPISSVAVRAQVRSRLAKEDREFRIGDITTEQFQKDDHDFRG-IT
consensus 1841 *****

PgFV1 1742 MTNACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
GVT 1733 MTNACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
GRSPaV 1761 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
PCMV 1736 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
ApLV 1779 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
ASPV 1778 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
AGCaV 1778 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
RuCV1 1639 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
APV2 1605 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
APV1 1622 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
APV3 1622 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
PVM 1566 LANAACERYESIYPRHKGNDSVTFMLAVKRLRFSSEIYETELLRAMPFGHFLKTFLLKRPPLRIHDPSSMMEKSVAEFE
consensus 1921 *****

Fig. S2. Multiple sequence alignment of TGB1 protein sequences of PgFV1 and related viruses

PgFV1	1	MNNLLAALDDLYNFECVSDKLSFPVIVHCVPGCGKTSLLIRDLIKIDSNFVFAFTAGEPDIPNLECKYIKRYSK-D--CAVKG
GVT	1	MNNLVSALELEFGVRISEEARYPVIVHSVPGSGKTSLLIRSLIKIDGDFEAFAGVDPDPNLEGGYIRSHFE-G--CASNK
GRSPaV	1	MNNLVKALSABEFGVGFVSLKFPVIVHSVPGSGKSSLIRELISEDENFIIFTAGVDPSPNLTGRYIKPYSP-G--CAVFG
PCMV	1	MDTVVSLLEFGYGFERTSVPLEDKLIVHAVPGSGKTTLLIRALNRLNGIEAFSFGEPDLNLTWCRYIKKAIS-G--QKGTG
ApLV	1	MEIVLSLLEFGYGFERTVEPLSDPVVHAVPGSGKTTLLIQALIRNHNIEAVTEGVPEKANLHCTYIKKARQ-G--QRGRG
ASPV	1	MEIVLSLLEFGYGFERTVEPLSDPVVHAVPGSGKTTLLIQALIRNHNIEAVTEGVPEKANLHCTYIKKARQ-G--QRGRG
AGCaV	1	MEIVLSLLEFGYGFERTVEPLSDPVVHAVPGSGKTTLLIQALLRNHNIEAVTEGVPEKANLHCTYIKKARQ-G--QRGRG
RuCV1	1	MEDLLHLLEISSFERTSVPLSSVVVHCVACAGKTSLLIRKWLGRNPNGEVRTCGVPDKENLTGRRIKAWGG-NLESKSD
APV2	1	MDFVYDKLLDAGLIRTRRLSFPPIVHCVACAGKSTLIRELIDEVDHKFEAFYGVDPDINLSCVRIKNSNDIP--SARAD
APV1	1	MEFVYDKLLDAGLIRTRRLSFPPIVHCVACAGKSTLIRELIDESRFEAFYGVDPDINLSCGIRIRGAELR--NARPE
APV3	1	MDFVYDKLLDAGLIRTRRLSFPPIVHCVACAGKSTLIRELIDEADNRFYGVDPDINLSCVRIKGAADIG--RARAD
PVM	1	MDVLDLIVKIKFERLSNKLVCPIVVHCVPCAGKSSLLIRELLEDSDRECAITAGVEDQPRLSGNIIRKWS--G--QOPEG
consensus	1	* . . . * * * * * * * * *
PgFV1	78	KLNLLDEYLIAE---DWIGFVSVLFSDDPQNVK-APLEAHEVGRSRFRFKETCKYLRRNGHEVNSTS-----EDIVVIGS
GVT	78	KLSILDEYLVVE---NWEFGDALFSDPYQNDK-SPLVASVSKKPKRFKSTOAYLDADYGEVESEI-----EDLVVIRGS
GRSPaV	78	KVNNLLDLSVQ---DFGEDVLFSDPYQNIS-IPKEAHEKSKCRFGVNTCKVLSSEFKVSSDG-----IDRVVIRGS
PCMV	78	SFCILDEYLSGE---FGIGDCFFSDPHQNSG-DCAPAHEVGRSSQRFGRNTAGLLQSLGNSVNSAK-----DDELIFEN
ApLV	78	NFSILDEYLSGE---YSIGENCLFSDPYQNHG-DCLRAHETGRCSHRFGNCTVQVIRNLGNNTASSK-----EDIVVEKKN
ASPV	78	NYSILDEYLSGE---YSIGENCLFSDPYQNHG-DCLRAHETGRCSHRFGNCTVQVIRDLGNNTASSK-----EDIVVEKKN
AGCaV	78	NFSILDEYLSGE---YSIGENCLFSDPYQNHG-DCLRAHETGRCSHRFGNCTVQVIRNLGNNTASSK-----QDIVERKN
RuCV1	80	QVWVLDVEYCELKGFHDYREGAIFCDNLOFPEDSCLEAHYICLNSHRLSRNTISFDNKEGCHNHSLSKEVSKEEPEVWFGG
APV2	79	SLKILDEYLGQD---LPECTTFCADENCFPY-TCPDAHETCYQSKRFEDOTCALLGKLDCAAFSYK-----QDQIFDK
APV1	79	SFKILDEYLGQH---RPECTAVCFADPNCFPY-SSPNAHETCYQSKRFEDOTCAFLGKLDCAAFSYK-----SDQLIFET
APV3	79	SIRKILDEYLGQV---LPDCTAFCFADENCFPY-TCPDAHETSYQSKRFEDOTCSFLGKLDCAAFSYK-----SDQLIFEK
PVM	77	KFVWLDVEYTLLT---EVPVFALFGDPIQSNTSAVQRADDEVCSVSRRFGSATCGLLRELGNVRSK-----ADIVQVSD
consensus	81	. . . * * * * * * *
PgFV1	149	PEEV-KTEGQLICFGKGAIDLAVSHSASKLPCEVIRGSTEFEVVTVLK-SSEPNSENHLEFYIALTRHRKLLILE-----
GVT	149	PEEL-KVEGQLICFGKAAVELALSHCAEIKLPCEVIRGSTEFDVVTLK-SSEPPSSNRHLEFYVCCTRHRKLLILE-----
GRSPaV	149	PFTL-DVEGVLICFGKEAVDLAVAHNSEIKLPCEVIRGSTEFEVVTVLK-SRDPTPEDRHWFYIAATRHRKLLIMQ-----
PCMV	149	VEEA-EIEGAVICVEKNVEDFLRWNHCEYKLPQVRGSTEFEVVTFTH-ELPLDQVGPDLVIALTRHRSKLLQILN----
ApLV	149	IERLVEPEGAVIICLEKEVEDFLWVHVEYKLPQVRGATFDIVTFTH-EKPLEEAVGPDVWVALTRHRKLVLVSN----
ASPV	149	IEQLIEPEGVLIICLEKEVEDFLKWHSVEYKFPQVRGATFDIVTFTH-EKPLEEAVGPDLEVALTRHRKLVLVSN----
AGCaV	149	IEQLVDEPEGVLIICLEESVEDFLKWHSVEYKLPQVRGATFDIVTFTH-EKPLEEAVSPDLEVALTRHRKLVLVSN----
RuCV1	160	IYSE-EIEGQLIISLDREAELIKNHSVAKRTFEVRGLFERYVCLVS-SRELVELEPHVYLAFTTRHRSVMFLDPDASH
APV2	150	IEAG-DIEGQIVCYEKEVEDLLDRHGADYKRYCQVRGSTEFDIVTFITASDTFEPEDRYKVYLCVLRHRSVDRILPEGKF
APV1	150	VEEG-SIEGQIVCYEKEVEDLLDRHGADYKRDQVRGSTEFDIVTFITSSSEFEPEDRYKVYLCVLRHRSVDRILPEGMF
APV3	150	IEEG-SIEGQIVCYEKEVEDLLDRHGADYKRDQVRGSTEFDIVTFITASSEFEPEDRYKVYLCVLRHRSVDRILPEGMF
PVM	149	IYTK-DPLCKVVFSEEEVGCILRSEHGVEALSQETGCTFEVVVTFVT-SENSPVNRAAAYQCCTRHRATDHIICPDATY
consensus	161 * * * * * * *
PgFV1	-----	
GVT	-----	
GRSPaV	-----	
PCMV	-----	
ApLV	-----	
ASPV	-----	
AGCaV	-----	
RuCV1	238	TS-----
APV2	229	LREDAKFDITTT
APV1	229	LRDNAKFDATS
APV3	229	LRDNAKFDATS
PVM	227	TAA-----
consensus	241	

Fig. S3. Multiple sequence alignment of TGB2 protein sequences of PgFV1 and related viruses
Two predicted transmembrane domains of PgFV1 TGB2 protein are marked by number signs (#) at the top.

```

#####
PgFV1      1 MS-FQOPADWSKNRPELLTGGGVALLLFFFRONNLEPHTGDNHSLPHGGYODGTRKINCYGPKKNFPGPGILSIGSSST
GVT        1 MS-FQOPANWSKSRPEVLTGAGVALVTHFTRESNLEHVGDNLHHLPHGGSYODGTRKITCYGPKKEPPTGTFQLGNSQT
GRSPaV     1 MP-FQOPANAKRTPTPLTCLGIGIVLHFRKSNLEPYSGDNIHQEPHGGYRVDGTRKSITCYGPKOSFPSSGIFGQSENFV
PCMV       1 MP-LAOPPDFSKSVPLAICLAVGIIVIFALTRSTLEHAGDNIHHLPHGGSYVDGTRKINCYGPKESFPTGGL--GLKFY
ApLV       1 MP-FSOPPDYSKSVPEIATCAAVAVLFLTRSTLEOVGDNIHNLPHGGYODGTRKISYCGPKNSFPSSSLIS-SGTPM
ASPV       1 MP-FAOPPDYSKSVPEIAVCIAVAVLFLLTRSTLEOVGDNIHNLPHGGYODGTRKISYCGPDSFPSSSLIS-SGTPM
AGCaV     1 MP-FSOPPDYSKSVPEVAVCVAVAVLFLTRSTLEOVGDNIHNLPHGGYODGTRKISYCGPKDSFPSSSLIS-SGTPM
RuCV1     1 MP-LTPPKDYTGAAISVVICLCIAFAFHSLTRSNLEHAGDNIHHLPHGGYFKDGTQVAVYCGPQSRFPSSNLF--SSSFS
APV2      1 MPSLTPPPDNTRVLPPIALCLGVGVVITWCLTRSTLEOVGDNIHSLPHGGYLDGTRKISYCGPKNSFPSSNLFK-GGAFS
APV1      1 MPSLTPPPDNTRVLPPIAVCLGVGVVITWCLTRSTLESVGDNVHSLPHGGYLDGTRKISYCGPDSFPSSNLFK-GGTFSS
APV3      1 MPSLTPPPDNTRVLPPIAVCLGVGVVITWCLTRSTLESVGDNVHSLPHGGYMDGTRKINCYGPRDSFPSSNLFK-GGTFSS
PVM       1 MP-LTPPDEFTRKVVLSAALCVSLAVVWLLTRSTLEVVDGDRDHLPHGGYMRDGTQVAVYVFNYSPEGR----LNSIEARKAPL
consensus 1 *. . * . . . . . * . . . . . * * . * . * * . . . * . . . . .
    
```

```

#---#####
PgFV1      80 A---FVILIVLVALTYASERFARIVRRC-PCVPGTCASR-----
GVT        80 A---VVIVLGLSLLTYAVEKFGIGTYRGC-ACNPNPCVCRR----
GRSPaV     80 P---IMTVLGLTAFTHVLSVNSGLGRNC-NCHPNPCSCRQQ---
PCMV       78 A---FCLACCLLAYHAFEGNNSSVR-CPTCINNPOHCVRS---
ApLV       79 I---ITVILITAAITYSERWFGSGHRRCSCLPGAPACTATNHE
ASPV       79 I---IGLITFLIFATYVSEKWSRSGRRCSCCVPGAPACTATVHE
AGCaV     79 I---VGIITFLIFATYVSEKWSRSGNRRCCICIPGSPACTATSHE
RuCV1     78 SLSVLCVILLLSGLTYASNKEGGGAGQC-----VCSRAHNRR
APV2      80 A---ICVVVLLVFATHVSELENRPNRRTC-----CGGSASHA
APV1      80 A---ICTVVLLVFATHVSELENRPNRRTC-----CGGSAAHA
APV3      80 A---ICTVVLLVFATHVSELENRPNRRTC-----CGGSAAHS
PVM       76 LGQPWATVVLVLLIWFASHQLGRPNCRAC-----AGSHHT
consensus 81 . . . . . * . . . . .
    
```

Fig. S4. Multiple sequence alignment of TGB3 protein sequences of PgFV1 and related viruses
A predicted transmembrane domain of PgFV1 TGB3 protein is marked by number signs (#) at the top.

```

#####
PgFV1      1 -MLQKDLLGLGAAVYVAIV---LVDNFKES--CSIRITGESVSVINCQDLEAVAKLFRNSKPL--LSLGGRGDFE
GVT        1 -MQWRSTALSLICAVLATLILA---YVGGERGG--SCFIRITGESVIVSNCGNLQVPEVVKAVKGV--WSSLGAGVNF
GRSPaV     1 MYCLFGLTIVLVGI VAIQQLA---HVDSSSGNHQCFIRATGESLITENCPSSEALASTVKEVLGG--LALGVSRAVE
PCMV       1 -MQPLEVITAVFALALALITL---NIVFGANSCDEPQIITGESVVKVFNCLSPFLEVLKGLKPY--HPTL
ApLV       1 MYPRGGDAQGIVAAVVVYLVLLLAQQLMSTRAN--SCTIITGESVSVIVGCTYSDAFIELVKGLKPY--YHPLG
ASPV       1 MFPRSGIGLAVAAAVVAYLVLLLAQQLYMNSNS--QCTIITGESVSVVGVVYSEAFIELVKGLKPY--YHPLG
AGCaV     1 MFLKSGGLATAASAAAYLVLLLAQQLMNSQK--QCTIITGESVSVIVGCVYSDAFIELVKGLKPY--YHPLG
RuCV1     1 ---MLQINLVVVLVSVFVVLVLTIIDKFEREN--PCFIQITGESVVIKGCLEDKAFIELVKGLKPY--HPELG
APV2      1 --MFQSYLIGLITVALVAVVIT---HMLNSGQE--GCLIIITGESVVIKNCVYTSDFIELVKGLKPEHNSIT
APV1      1 --MYQSYLIGLITVALVAVVIT---HMLNSNHD--GCLIIITGESVVIKNCVYTSDFIELVKGLKPEHNSIT
APV3      1 --MYQSYLIGLITVALAVAGIT---HITSSAQE--GCLIIITGESVVIKNCVYTSDFIELVKGLKPEHNSIT
PVM       1 MIVYVVLGLSAFCIVL-----YLSIQGQS--DCVVIITGESVVRVQGRIDGEGSIVLSKLPKPE----CGGSFRS
consensus 1 . . . . . * . . . . .
    
```

```

PgFV1      73 IIEERC-
GVT        73 -LEGNCS
GRSPaV     75 EIDYHC-
PCMV       -----
ApLV       -----
ASPV       -----
AGCaV     -----
RuCV1     -----
APV2      -----
APV1      -----
APV3      -----
PVM       -----
consensus 81
    
```

Fig. S5. Multiple sequence alignment of coat proteins of PgFV1 and related viruses

PgFV1 1 -----
 GVT 1 -----
 GRSPaV 1 -----
 PCMV 1 MLYSVYELLSNLLIFDRNSAF-----VPFMATTAAATASATVQPTTQOE
 ApLV 1 MATSQDPTTSANPAVTRNEETPVVTEVQATNAVATTPEVVPVATAPLPASTOPTATSSFEVSTFPAISISATPMTFPEP
 ASPV 1 MTSNGSQPQASTPMVSAEEPAAAA-----VPNSTPMVSAEGPAAAVSAPNSSVVSAPASAPTASEP
 AGCaV 1 MASNGVSSQSSTPMVSIENPLVTS-----SPNPSVVAP-----LPVSMPVVASAPEVTFPAISEA
 RuCV1 1 -----
 APV2 1 MTTSQSTATADVETSKVETSPPPVV-----TSEPVLQQ-----VVFSDTSNAATPPPPVTTKTST
 APV1 1 MSVSSAAVVSTVSSSTSEAVPTSS-----ENEPVISQ-----LDLPPISSGVSGVSTTPTSTATSS
 APV3 1 MTTSVSAATDVMSTSVETAATIAP-----ANEPVIOQ-----VIYPKSTAVASVAPVMTHTAQR
 PVM 1 -----
 consensus 1 .

PgFV1 1 -----MEKPEQLQSEKESLRLARLSEERAAEKKAANDKKLELQH-----NRNAHRSQSSRSRAP
 GVT 1 -----MSNEELRKLKLEKQRDEAKTL-----
 GRSPaV 1 -----MSQIGKLPGESENEAFEARLKSLELAR-----
 PCMV 45 VTAGLQSTTS--STF-PTVTPVEIPARVATTRSDPVTSGYSLGNVTAALTRPTGSLPITVGLG--ALNLGDGCAVAVGGV
 ApLV 81 VVSQLVFPPLVATGQSSVQTTAVPDTSLRLOQMAANRGFSEG-----FRVHPLPITPA-----SSNPFTTGNITSSV
 ASPV 64 VISQVQSLAPIVSGFDPNLHGRLNEQMRQAQNEAMQGYEEGSRRNPRPSSSTAHNDYASM--NSNPFETGTAYGGAP
 AGCaV 58 VISQVQSVAPMVNGFDPNLHGRLNEQHRAAQEEARLGFQSGIQORROPVTSAYAITSS-----NPFETGNAYSVAP
 RuCV1 1 -----MTSEKLRDLEKKLSELATDSNERAGTEKDLKAEKAA-----
 APV2 57 AATKGVERSWFGFLPSKTTGLGTLATATSTSPLLSASGRASQLDENLKKPKAKDESFFVL-----NSEVEDLEAKASGKV
 APV1 57 VENPFKPRTSFSLPSLTLAGLTLATATSSAPAAATSEARTSSLESWKEKPKNKDEEPRRV-----LSEPESLGCFEAGAG
 APV3 57 VEDPFKTRTSFSLPSLTLNLGTLSTATAAASSTSDVKSSTFENWREKPKTKDDLNFPRITGVEHLGSEFAGEGSRGGS
 PVM 16 TSQERREARPLPTAADFEGKDTSENTDGRAADACEMSLERRLDSLREPRRERRGAIRVT-----
 consensus 81 .

PgFV1 56 -----CRTNEKREDGREDL-----ERPALPDSSVIRKRRKKV
 GVT 23 -----CKEIEGAGGSRPAL-----PDLCTVIKRRKKQ
 GRSPaV 28 -----AQKQPEGSNAP-----PTTSGILAKRRRI
 PCMV 120 TAIT--TPAAATVPMGRTESMEWFNRTYNNPASSAPFAFGVQPGTNLG-VPTQLGSSTGPVRRRQKFTKPINRESSV
 ApLV 150 FSGRGSNAATSSEAIICEPTP-----QRVFQSSQGSNPPAQGHYSQOHTSGSVGNVITP---FTLENRAPRRAAS
 ASPV 142 RVSF--GSYPTFPFGSGSASE-----PNSQRIFPQHGVMNPPAHASDLVPHQATSGGNTGTP---FTLENRAPRNATA
 AGCaV 131 QNMN--GYPYTLQDQGTSE-----PNSQRIFPIQGGVSPSAHDGNLRPQTESSQSASITP---FTLENRAPRTASS
 RuCV1 37 -----IAAGSVSS-----DKKKAARIKQKF
 APV2 132 QGPR--GVVPSMMSYERPEK-----GKVVEESKGLSLGSRQRLVFEKARQRAQVNIEDRDSLAPPF
 APV1 132 SGPR--GVSPSTMIHSKLEV-----KKGPDLEATGVSLGARQRMVFEAARKRAQANSEDRDSLAPPF
 APV3 137 GGPVHGVSPSTMAYERQNE-----AEDSNVKTGVSILGSRQRMVFEVARKRAQSNIEDRDNMAPPF
 PVM 76 -----NPGLETGRPRL-----QLLENMRPDPT--
 consensus 161 .

PgFV1 88 VESPLSKKVDMAITMQRNVOQSNVMEDIDVLAISSEIFSEAGLPEDKQQAVALIETANAADVGVSSKHSFETGKSAVC--
 GVT 50 IENALSAKVNKIDMEREGLQSNVMNNDILALSSITEAKLPEDLQOHTAIEVARMAVDVGVSSKHSFEMGKSVVC--
 GRSPaV 52 IENALSKTDMREVLKHEVTVVISPNVMDEGATDEIRAFGESGIAESVQFDVAIDIARHCSDVGVSSQSRSLTGKSPFC--
 PCMV 197 VSGSMKRSIRDAADYKPOAGSVAAEEKISQGLEFTKIGLSSDQLTEVGVYIARHCADVGVSSNQSRLTGTFFGS--
 ApLV 216 SIGTRRRSDSVGKSIMYEPQAGVVAIDAKIRATGRALTEMGIREDDQLTEVGVYIARHCADVGVSSDKSLTGTFFGS--
 ASPV 209 NTGGMRRRSDSVGKSNIRYEPQAGVVAIDQKIRATGMALGMGIEGHQLETEVGVYIARHCADVGVSSDKSLTGTFFGS--
 AGCaV 198 ISGGMRRRSDSVGKSIIEYEPQAGVVAIDQKIRATGMALGMGIEGHQLETEVGVYIARHCADVGVSSDKSLTGTFFGS--
 RuCV1 57 SGSSITSIPTNKMIREIKISTEVRNVCLSNAELIAAEFVELGIPEDKLAEAAWDIALHCADVGVSSSELTEACTCTFA-P
 APV2 192 ASDPFSRPKVQVQRFSEYESSPDVIAENIEYIRADLRAGVETKDTIFAMWDIARYCADAGSSESTEFVGTSSYG-G
 APV1 193 ASGDPFSPKVVQVQRFSEYESSPDVIAENIEYIRADLRAGVETKDTIFAMWDIARYCADAGSSESTEFVGTSSYG-G
 APV3 200 ASSDPFSRPKVQVQRFSEYESSPDVIAENIEYIRADLRAGVETKDTIFAMWDIARYCADAGSSESTEFVGTSSYG-G
 PVM 98 ---NPNYRPSIEAHSRIKPIAISNNMAISEDMMRIYVNEGLGVTEHQVVVQAVLFCQDASSVFLPRCSFEWPRG
 consensus 241 .

PgFV1 166 DIDLPTCVGLIKREV--TILRRFCMAYAKVIWNIRVKSRYPPANWARKGFKDETKEFAAFDFEVGVFDESALNPEGGLVRSKSP
 GVT 128 GHDLSAIVCYIKREV--TILRRFCMAYSKVIWNLLIKEKTPPANWARKGFKDETKEFAAFDFEVGVFDESALNPEGGLVRSKSP
 GRSPaV 130 DINRSEIAGIIEV--TILRRFCMAYAKVIWNIRVLETGTIPANWARKGFKENENEKFAAFDFEVGVFDESALNPEGGLVRSKSP
 PCMV 275 DVELEELATIKSTAGCTLRQFCAYAKVIWNIMLETQIPPAIWSKKGVDENKFAAFDFEVGVFDESALNPEGGLVRSKSP
 ApLV 294 DITLEEVCITRIKQTEGCTLRQFCAYAKVIWNIMLOTQIPPAIWSKKGVDENKFAAFDFEVGVFDESALNPEGGLVRSKSP
 ASPV 287 DITLEEVCITRIKQTEGCTLRQFCAYAKVIWNIMLOTQIPPAIWSKKGVDENKFAAFDFEVGVFDESALNPEGGLVRSKSP
 AGCaV 276 DITLEEVCITRIKQTEGCTLRQFCAYAKVIWNIMLOTQIPPAIWSKKGVDENKFAAFDFEVGVFDESALNPEGGLVRSKSP
 RuCV1 136 NVTRSDMGAVVKSII--CTLRQFCSLYAKVIWNIMTCDRPPANWARKGFKDETKEFAAFDFEVGVFDESALNPEGGLVRSKSP
 APV2 271 RVTIRMEIAAVIKKH--TILRRFCGYAKVIWNIMLSTNIPPSGVMKKGKENTKFAAFDFEVGVFDESALNPEGGLVRSKSP
 APV1 272 RVTIRMEIAAVIKKH--TILRRFCGYAKVIWNIMLSTNIPPSGVMKKGKENTKFAAFDFEVGVFDESALNPEGGLVRSKSP
 APV3 279 RVTIRMEIAAVIKKH--TILRRFCGYAKVIWNIMLSTNIPPSGVMKKGKENTKFAAFDFEVGVFDESALNPEGGLVRSKSP
 PVM 175 AITADAVIAVIRKDA--ETILRRVCRLYAPVIWNIMLSTNIPPAIWAAMGFOYEDRFAAFDFEVGVFDESALNPEGGLVRSKSP
 consensus 321 .

