Identification of Pistacia-associated flexivirus 1, a putative mycovirus of the family *Gammaflexiviridae*, in the mastic tree (*Pistacia lentiscus*) transcriptome

D. PARK[1], C. J. GOH[1], J. S. LEE[1], F. SEBASTIANI[2], Y. HAHN[1*]

[1]Department of Life Science, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea;
[2]Institute for Sustainable Plant Protection, Department of Biology, Agriculture and Food Sciences, The National Research Council of Italy, Sesto Fiorentino, Italy

Summary. – In this study, we identified the genome sequence of the novel virus Pistacia-associated flexivirus 1 (PAFV1), a putative member of the mycovirus family *Gammaflexiviridae* (the order *Tymovirales*), via analysis of a transcriptome dataset for the mastic tree (*Pistacia lentiscus*, the family *Anacardiaceae*). PAFV1 was predicted to have three open reading frames (ORFs): ORF1, encoding a replicase (REP) with RNA-dependent RNA polymerase activity; ORF2, a movement protein (MP); and ORF3, a hypothetical protein. The PAFV1 REP sequence showed high similarity to those of three known members of the family *Gammaflexiviridae* i.e., Entoleuca gammaflexivirus 1 (EnFV1), Entoleuca gammaflexivirus 2 (EnFV2), and Botrytis virus F (BVF). A genome contig of the fungus *Monosporascus cannonballus* also contained a sequence of an endogenous virus similar to that of PAFV1. Sequence comparison and phylogenetic analysis indicated that PAFV1, EnFV1, and the endogenous virus of *M. cannonballus* formed a distinct subgroup (apart from EnFV2 and BVF), and may be the founding members of a novel genus in the family *Gammaflexiviridae*. Notably, MP sequences of PAFV1/EnFV1 showed similarity to the MP sequences of the mycovirus group called tobamo-like mycoviruses (an unassigned taxon), implying that genomic recombination occurred between members of the family *Gammaflexiviridae* and tobamo-like mycoviruses. Since PAFV1 is phylogenetically related to mycoviruses, PAFV1 may also be a mycovirus that infected a fungus associated with the mastic tree sample, which is evidenced by the presence of fungal ribosomal RNA sequences in the mastic tree transcriptome. Thus, the PAFV1 genome sequence may be useful in elucidating the genome evolution of *Gammaflexiviridae* and tobamo-like mycoviruses.

Keywords: Pistacia-associated flexivirus 1; Gammaflexiviridae; mycovirus, mastic tree

Introduction

Gammaflexiviridae is a family of positive-strand single-stranded RNA viruses and is one of the five approved

families in the order *Tymovirales* (Svanella-Dumas *et al.*, 2018). The other four families are *Tymoviridae*, *Alphaflexiviridae*, *Betaflexiviridae*, and *Deltaflexiviridae*. Currently, the family *Gammaflexiviridae* contains one genus, *Mycoflexivirus*, of which a prototype species called Botrytis virus F (BVF) (Howitt *et al.*, 2001; Donaire and Ayllon, 2017; Svanella-Dumas *et al.*, 2018) is the only approved species in the family *Gammaflexiviridae*.

BVF is a flexuous mycovirus (fungus-infecting virus) that was first isolated from the plant-pathogenic necrotrophic fungus, *Botrytis cinerea* (Howitt *et al.*, 2001). It has a single-stranded RNA genome of around 6.8 kb with two open reading frames (ORFs) encoding a replicase (REP) and a coat

protein (CP). BVF REP contains three conserved domains i.e., a methyltransferase, helicase, and RNA-dependent RNA polymerase (RdRp). Interestingly, there is an in-frame stop codon between the helicase and RdRp domains, which is expressed by a ribosomal readthrough translation strategy (Howitt *et al.*, 2001; Svanella-Dumas *et al.*, 2018).

The genome sequences of two proposed members of the family *Gammaflexiviridae* i.e., Entoleuca gammaflexivirus 1 (EnFV1) and Entoleuca gammaflexivirus 2 (EnFV2), are present in the National Center for Biotechnology Information (NCBI) nucleotide database (Acc. Nos. MF375883 and MF375884, respectively). These sequences were identified as belonging to two of many mycoviruses infecting the fungus *Entoleuca* sp. associated with avocados (Velasco *et al.*, 2018). EnFV1 and EnFV2 were predicted to have three ORFs each: EnFV1 ORFs were predicted to encode a REP, putative movement protein (MP), and hypothetical protein (HP), whereas EnFV2 ORFs were predicted to encode a REP, CP, and HP, respectively. The REP sequences of BVF, EnFV1, and EnFV2 are significantly similar to one another, which supports their close phylogenetic relationship.

However, there are some differences among BVF, EnFV1, and EnFV2. Unlike BVF, EnFV1 and EnFV2 have no readthrough stop codon in the REP ORF. Moreover, BVF has two ORFs, whereas EnFV1 and EnFV2 have three ORFs each. Furthermore, the BVF and EnFV2 CPs exhibit no sequence similarity even though they have the same name. Likewise, the EnFV1 and EnFV2 HP sequences also exhibit no similarity to each other. The difference in genome organization implies that genome rearrangement is common in viruses of the family *Gammaflexiviridae*.

Transcriptome data obtained from plant, animal, and fungal samples often contain sequences derived from infecting RNA viruses, which can be identified by a comprehensive bioinformatic-based method (Goh *et al.*, 2018; Kim *et al.*, 2018; Park *et al.*, 2018). Therefore, publicly available transcriptome datasets are invaluable resources for identifying novel RNA virus genome sequences. In this study, we analyzed a transcriptome dataset obtained from two leaf samples of the mastic tree (*Pistacia lentiscus*, the family *Anacardiaceae*) and identified the genome sequence of a novel virus that may belong to the family *Gammaflexiviridae*. The mastic tree is a common hard-leaved shrub widely distributed throughout the Mediterranean region whose resin possesses numerous and diverse biomedical and pharmacological properties (Albaladejo *et al.*, 2008; Dimas *et al.*, 2012).

Materials and Methods

Transcriptome dataset. A transcriptome dataset (7.4 gigabases; two paired-end runs; F. Sebastiani, unpublished data)

obtained from two leaf samples of the mastic tree was downloaded from the Sequence Read Archive (SRA) of the NCBI. SRA Acc. Nos. of the two runs are SRR5043661 and SRR5043662, respectively. The Sickle program (version 1.33; https://github.com/najoshi/sickle) was used to filter out low-quality reads using the parameters, "-q 30 -l 50." High-quality RNA-seq data were assembled *de novo* into contigs using the SPAdes Genome Assembler (version 3.12.0; http://cab.spbu.ru/software/spades) using the parameter, "--rna" (Bankevich *et al.*, 2012). Each of the two sequencing runs was assembled separately.

Known viral RdRp motif sequences. The known virus-derived RdRp motif sequences were collected from the Pfam database release 32.0 (https://pfam.xfam.org). First, Pfam families containing the keyword "RNA-dependent RNA polymerase" were retrieved. The records were manually reviewed to select virus-derived RdRp motif sequences. Initially, 2539 RdRp motif sequences were collected from 22 Pfam families (PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03035, PF03431, PF04196, PF04197, PF05788, PF05919, PF06317, PF07925, PF08467, PF08716, PF08717, PF12426, and PF17501). Subsequently, sequences were clustered using the USEARCH program (version 11.0.667, https://drive5.com/usearch) (Edgar, 2010) with the following parameters: "-cluster_fast *input* -id 0.9 -centroids *output* -sort length". As the result, 2195 non-redundant viral RdRp motif sequences were obtained.

Identification and annotation of virus genome contig. The assembled mastic tree transcriptome contigs were compared with viral RdRp motif sequences using the BLASTX program to identify potential RNA virus-derived contig sequences. Nucleotide (nt) and protein sequence similarity searches of putative viral genome contigs equal to or longer than 1000 nt were performed using the BLAST server from the NCBI website (https://blast.ncbi.nlm.nih.gov/Blast.cgi). ORFs and their protein sequences were predicted using the BLAST results and ORF finder analysis (https://www.ncbi.nlm.nih.gov/orffinder) with default settings. Conserved protein domain prediction was conducted using Pfam and SMART (http://smart.embl-heidelberg.de) (Letunic and Bork, 2018).

Sequence comparison and phylogenetic analysis. Pairwise identities of the protein sequences were calculated using the GGSEARCH program in the FASTA package (version 36.3.8g; https://github.com/wrpearson/fasta36). Multiple sequence alignment was generated using the MAFFT program (version 7.419; https://mafft.cbrc.jp/alignment/software) (Nakamura *et al.*, 2018) with the parameter, "--auto". Phylogenetic analysis was conducted using the RAxML program (version 8.2.12; https://cme.h-its.org/exelixis/web/software/raxml/index.html) (Stamatakis, 2014). First, the maximum likelihood tree was inferred using the parameter, "-m PROTGAMMAWAG -p 12345 -s *input* -# 20 -n T1." Then, 100 bootstrap replicate trees were computed using the parameter, "-m PROTGAMMAWAG -p 12345 -b 12345 -# 100 -s *input* -n T2." Lastly, branch support values

Table 1. Open reading frames (ORFs) of the PAFV1 genome sequence

ORF	Position (nt)	ORF length (nt)	Protein length (aa)	Domain name	Acc. No.	Position (aa)
Replicase	279–6026	5748	1915	Viral methyltransferase	PF01660[a]	42–320
				Viral RNA helicase	PF01443[a]	1056–1292
				RNA-dependent RNA polymerase	PF00978[a]	1585–1818
Movement protein	5995–8412	2418	805	DEAD-like helicases superfamily	SM000487[b]	176–372
Hypothetical protein	8445–9425	981	326			

[a]Pfam database Acc. No. [b]SMART database Acc. No.

were calculated using the parameter, "-m PROTGAMMAWAG -p 12345 -f b -t RAxML_bestTree.T1 -z RAxML_bootstrap.T2 -n T3."

Identification of ribosomal RNA contigs. FASTA-formatted eukaryotic ribosomal RNA (rRNA) sequences, including those of fungal rRNAs, were downloaded from the UNITE database (version 18.11.2018; https://unite.ut.ee) (Nilsson *et al.*, 2019) and converted to a BLAST-searchable database. A MEGABLAST search against the UNITE database was used for the initial collection of potential rRNA contigs in the mastic tree transcriptome contigs. Contigs showing 80% or higher identity over 1000 nt or longer with a known rRNA sequence were selected for further analysis. These contigs were then used as queries for MEGABLAST searches against the NCBI nucleotide database.

Results and Discussion

A total of 7.4 Gb of transcriptome data obtained from the leaves of mastic trees were assembled into contigs. A BLASTX search was performed using the mastic tree contigs as queries against the Pfam viral RdRp motif sequences. A 9525 nt-long contig showed high sequence similarity to the RdRp region of the BVF REP sequence. BVF is the prototype species of the genus *Mycoflexivirus* in the family *Gammaflexiviridae* (Howitt *et al.*, 2001; Svanella-Dumas *et al.*, 2018).

A BLASTX search against the NCBI protein database using the 9525-nt contig as a query revealed that the REP
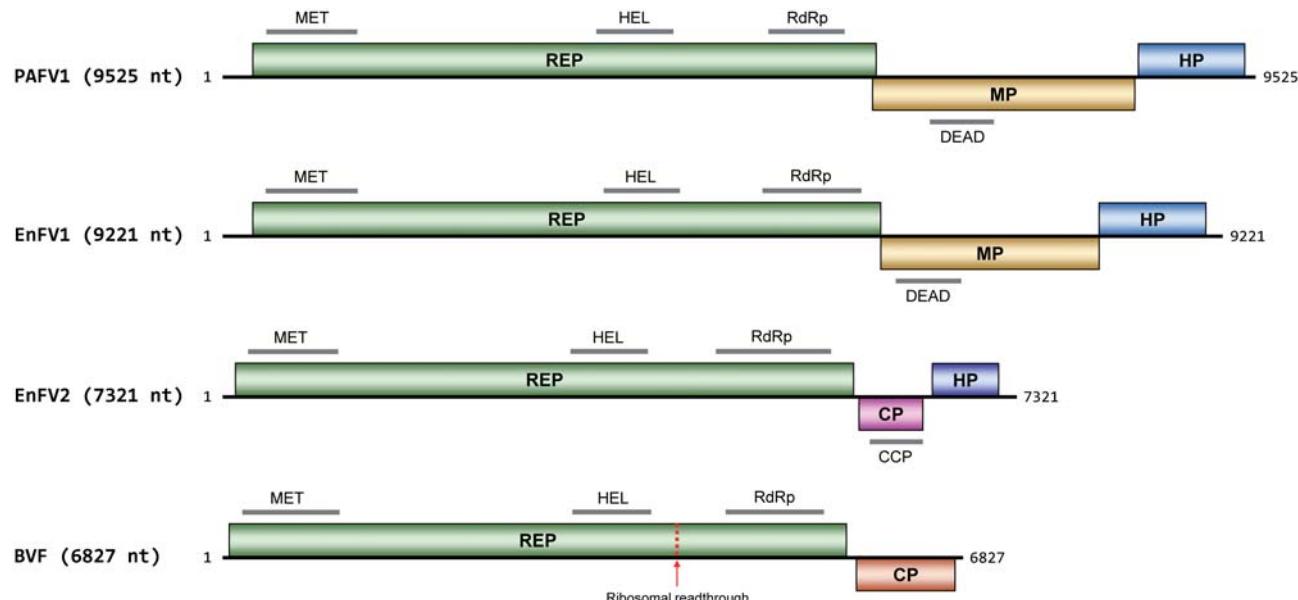
Fig. 1

Comparison of the genome organizations of PAFV1 and other *Gammaflexiviridae* viruses
Genome organizations of PAFV1, EnFV1, EnFV2, and BVF are schematically depicted. Protein-coding ORFs are represented by boxes: REP, replicase; MP, movement protein; CP, coat protein; and HP, hypothetical protein. Homologous proteins are shown in the same color. Predicted protein domains are marked by lines above or below the ORF: MET, viral methyltransferase; HEL, viral RNA helicase; RdRp, RNA-dependent RNA polymerase; DEAD, DEAD-like helicases superfamily; and CCP, closterovirus coat family. The readthrough stop codon in the BVF REP ORF is indicated by a red dotted line.

sequence of an unassigned *Gammaflexiviridae* virus named EnFV1 showed the highest sequence similarity. REP sequences of EnFV2, another unassigned *Gammaflexiviridae* virus, and BVF, the only approved species of the family *Gammaflexiviridae*, also showed significant sequence similarities. Therefore, the mastic tree contig was assumed to be a viral genome sequence derived from a novel member of the family *Gammaflexiviridae*, and was named Pistacia-associated flexivirus 1 (PAFV1). Its genome sequence is available in the NCBI nucleotide database under the Acc. No. MK605686.

The PAFV1 genome sequence was predicted to have three ORFs (Table 1), each of which has a corresponding homolog in the EnFV1 genome. The first ORF encodes a 1915-amino acid (aa) REP that is involved in viral reproduction. The PAFV1 REP has three predicted protein domains: viral methyltransferase (Pfam PF01660), viral RNA helicase (Pfam PF01443), and RdRp (Pfam PF00978). The second ORF encodes an 805-aa protein annotated as

movement protein (MP) in the EnFV1 genome. Notably, PAFV1 MP has a DEAD-like helicases superfamily domain (SMART SM000487). The third ORF encodes a 326-aa hypothetical protein (HP). However, PAFV1 HP has no known protein domain.

Notably, comparison of the PAFV1 genome sequences and those of known members of the family *Gammaflexiviridae* (EnFV1, EnFV2, and BVF) revealed similarities and differences in genome organization among these viruses (Fig. 1). EnFV1, the closest virus of PAFV1, has three ORFs that are homologous to PAFV1 ORFs. EnFV2 also has three ORFs (REP, CP, and HP). However, BVF only has two ORFs (REP and CP). The EnFV2 and BVF REP sequences have significant sequence similarity to the PAFV1 and EnFV1 REP sequences. However, EnFV2 HP has no sequence similarity to the PAFV1 and EnFV1 HP sequences. On the other hand, EnFV2 CP, which has a closterovirus coat protein domain (Pfam PF01785), has no sequence similarity to the BVF CP. Moreover, the ribosomal readthrough stop codon,

Table 2. Sequence identities of the PAFV1 proteins and their homologs

PAFV1 protein	Virus or fungus	Acronym	Taxonomy[a]	NCBI Acc. No.	Identity[b]
REP	Entoleuca gammaflexivirus 1	EnFV1	*Gammaflexiviridae*	AVD68667.2	771/1977 (39.0%)
–	*Monosporascus cannonballus*	–	Fungi; *Ascomycota*	RYO87574.1	668/2060 (32.4%)
–	Entoleuca gammaflexivirus 2	EnFV2	*Gammaflexiviridae*	AVD68668.2	502/2006 (25.0%)
–	Botrytis virus F	BVF	*Gammaflexiviridae*; *Mycoflexivirus*	NP_068549.1	506/2027 (25.0%)
–	Watercress white vein virus	WWVV	*Tymoviridae*; *Tymovirus*	AFC95826.1	491/2042 (24.0%)
–	Ullucus tymovirus 1	UTyV1	*Tymoviridae*	AZF99027.1	495/2019 (24.5%)
–	Diascia yellow mottle virus	DiaYMV	*Tymoviridae*; *Tymovirus*	YP_002048673.1	499/2017 (24.7%)
–	Nemesia ring necrosis virus	NeRNV	*Tymoviridae*; *Tymovirus*	YP_002308442.1	477/2011 (23.7%)
–	Grapevine virus A	GVA	*Betaflexiviridae*; *Vitivirus*	AGT17853.1	419/1984 (21.1%)
–	Grapevine virus E	GVE	*Betaflexiviridae*; *Vitivirus*	YP_002117775.1	429/1995 (21.5%)
–	Cherry virus A	CVA	*Betaflexiviridae*; *Capillovirus*	ARQ83952.1	395/2394 (16.5%)
–	Garlic yellow mosaic-associated virus	GYMaV	*Betaflexiviridae*	AZM69107.1	440/2165 (20.3%)
–	Cherry rusty mottle associated virus	CRMaV	*Betaflexiviridae*; *Robigovirus*	YP_007761581.1	419/2118 (19.8%)
–	Cherry twisted leaf associated virus	CTLaV	*Betaflexiviridae*; *Robigovirus*	YP_009046478.1	441/2151 (20.5%)
–	Cherry necrotic rusty mottle virus	CNRMV	*Betaflexiviridae*	ATB18070.1	422/2134 (19.8%)
MP	Entoleuca gammaflexivirus 1	EnFV1	*Gammaflexiviridae*	AZG06254.1	328/807 (40.6%)
–	*Monosporascus cannonballus*	–	Fungi; *Ascomycota*	QJNS01000099.1:260663-262990[c]	295/820 (36.0%)
–	Acidomyces richmondensis tobamo-like virus 1	ArTlV1	Tobamo-like mycovirus	AZT88675.1	211/891 (23.7%)
–	Macrophomina phaseolina tobamo-like virus 1a	MpTLV1a	Tobamo-like mycovirus	ANE37542.1	210/888 (23.6%)
–	Podosphaera prunicola tobamo-like virus	PpTlV1	Tobamo-like mycovirus	ATS94408.1	237/1046 (22.7%)
–	*Golovinomyces cichoracearum*	–	Fungi; *Ascomycota*	RKF78024.1	187/830 (22.5%)
–	*Eutypa lata*	–	Fungi; *Ascomycota*	EMR61651.1	214/985 (21.7%)
HP	Entoleuca gammaflexivirus 1	EnFV1	*Gammaflexiviridae*	AZG06255.1	123/334 (36.8%)
–	*Monosporascus cannonballus*	–	Fungi; *Ascomycota*	QJNS01000099.1:262997-263980[c]	98/339 (28.9%)

[a]Taxonomic family and genus (if available) for viruses; kingdom and division for fungi. [b]Amino acid sequence identities in the format: "identical residues/aligned length (% identity)". [c]Nucleotide positions for predicted protein sequences.

(a) REP
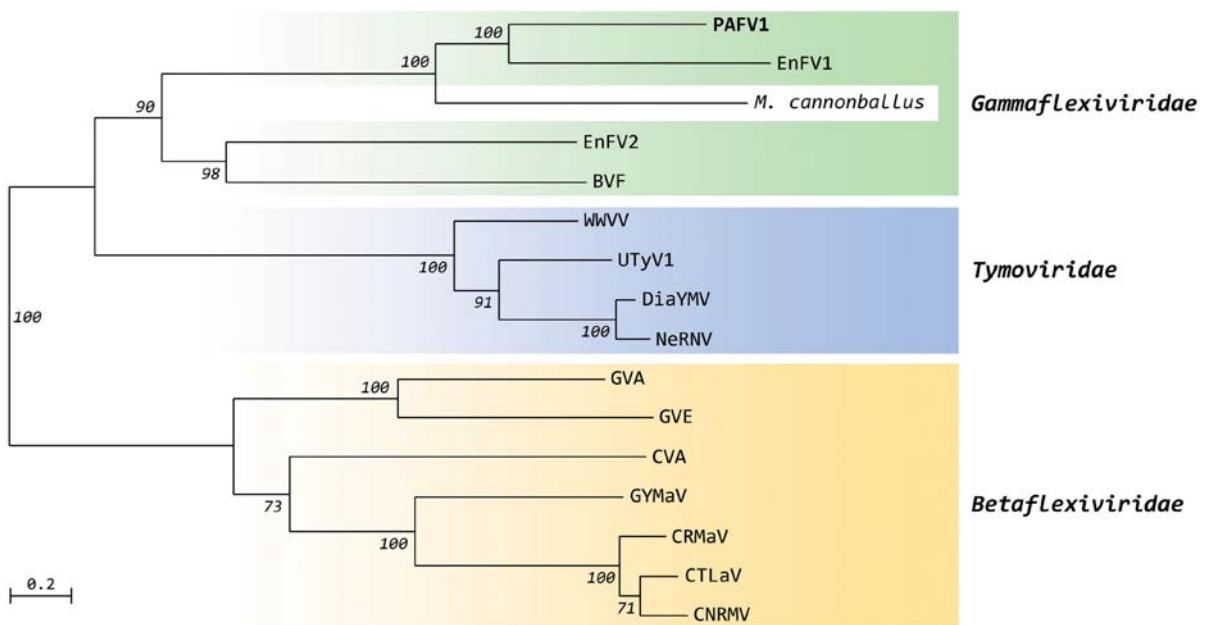
(b) MP
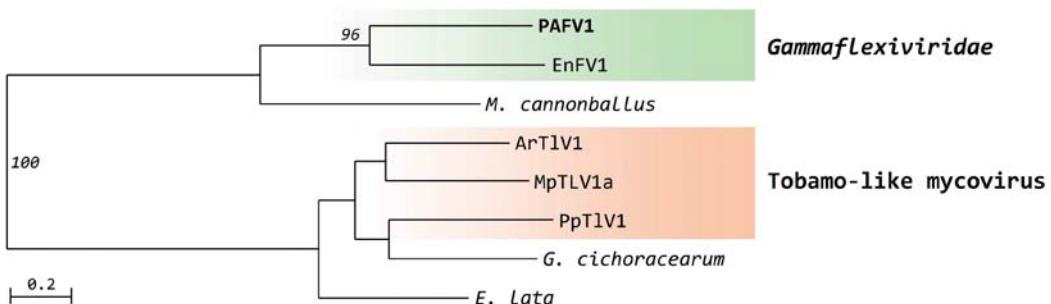
Fig. 2

Phylogenetic position of PAFV1

(a) The phylogenetic tree inferred from REP sequences reveals that PAFV1 is a member of the family *Gammaflexiviridae* in the order *Tymovirales*. The virus closest to PAFV1 is EnFV1. Note that PAFV1, EnFV1, and the endogenous virus of *M. cannonballus* form a strong subclade. **(b)** The phylogenetic tree inferred from MP sequences also shows that PAFV1, EnFV1, and the endogenous virus of *M. cannonballus* form a subclade. No homologous sequences were found in other viruses from the order *Tymovirales*. Similar sequences were found in tobamo-like mycoviruses and the endogenous viruses of two fungi (*G. cichoracearum* and *E. lata*). Bootstrap branch support values of 70% or higher, calculated from 100 replicates, are shown at the tree nodes. Viruses of the same family are shaded in the same color. Fungal species are not shaded. Full names of the viruses are provided in Table 2.

which is present in the BVF REP ORF, is not found in the PAFV1, EnFV1, and EnFV2 REP ORFs. Therefore, based on genome organization and sequence similarity, these four viruses can be divided into three groups: (1) PAFV1/EnFV1, (2) EnFV2, and (3) BVF.

Next, homologous PAFV1 REP protein sequences were collected via BLASTP searches against the NCBI protein database, which showed significant sequence similarity to the REP sequences from members of the families in the order *Tymovirales*, including *Gammaflexiviridae*, *Tymoviridae*, and *Betaflexiviridae*. REP sequences of representa-

tive viruses in these three families were retrieved and pairwise sequence identities with respect to the PAFV1 REP sequence were calculated (Table 2 and Supplementary Data S1). As noted above, PAFV1 REP showed the most significant sequence similarity to the REP sequences from *Gammaflexiviridae* viruses i.e., 39.0% identity over 1977 aa in EnFV1 REP; 25.0% identity over 2006 aa in EnFV2 REP; and 25.0% identity over 2027 aa in BVF REP. *Tymoviridae* viruses, such as watercress white vein virus (WWVV) and Nemesia ring necrosis virus (NeRNV), showed 23.7–24.0% sequence identities over 2011–2042 aa. *Betaflexiviridae*

viruses, such as grapevine virus A (GVA) and cherry necrotic rusty mottle virus (CNRMV), showed 16.5–21.5% sequence identities over 1984–2394 aa. Multiple alignment of the PAFV1 REP sequence and its homologous sequences showed that the methyltransferase and RdRp domains were more conserved than other regions.

Notably, a hypothetical protein (NCBI Acc. No. RYO87 574.1) in the *Ascomycota* fungus, *Monosporascus cannonballus*, showed the second highest sequence identity (32.4% identity over 2060 aa) with the PAFV1 REP. The *M. cannonballus* protein is encoded by a predicted gene in the genomic DNA contig (Acc. No. QJNS01000099.1). The presence of a highly homologous virus REP protein in the fungal genome suggests that a *Gammaflexiviridae* virus might have recently integrated into the nuclear genome of *M. cannonballus*. Integration of non-retroviral RNA virus sequences into fungal and plant nuclear and mitochondrial genomes has been rather frequently observed in the past (Taylor and Bruenn, 2009; Chiba *et al.*, 2011; Bruenn *et al.*, 2015). Further analysis of the *M. cannonballus* genomic contig QJNS01000099.1 sequence revealed two more viral proteins that were homologous to the PAFV1 MP and HP, which indicated that the endogenous virus had the same genomic organization with PAFV1 and EnFV1.

Next, a maximum-likelihood phylogenetic tree inferred from the multiple alignment of 16 REP sequences confirmed PAFV1 to be a member of the family *Gammaflexiviridae* (Fig. 2a). Notably, PAFV1, EnFV1, and the endogenous virus of *M. cannonballus* formed a strong subclade with a bootstrap value of 100%. Additionally, EnFV2 and BVF formed a sister clade to the PAFV1/EnFV1 subclade. The phylogenetic tree inferred in this study placed *Tymoviridae* as a sister family to *Gammaflexiviridae*.

Homologous protein sequences of PAFV1 MP were also collected and their pairwise identities with PAFV1 MP were calculated (Table 2 and Supplementary Data S2). EnFV1 MP showed the highest sequence identity (40.6% over 807 aa). The second closest sequence was found in the endogenous virus of *M. cannonballus* (36.0% identity over 820 aa). EnFV2 and BVF exhibited no sequences homologous to that of the PAFV1 MP.

Notably, sequences homologous to PAFV1 MP were also found in viruses from the group tentatively called tobamo-like mycovirus, a sister group of the family *Virgaviridae*. It includes viruses such as Acidomyces richmondensis tobamo-like virus 1 (ArTlV1), Macrophomina phaseolina tobamo-like virus 1a (MpTLV1a), and Podosphaera prunicola tobamo-like virus (PpTlV1) (Marzano *et al.*, 2016; Pandey *et al.*, 2018; Gilbert *et al.*, 2019). Two more MP-like sequences were found in the genomic sequences of two *Ascomycota* fungi, *Golovinomyces cichoracearum* and *Eutypa lata*. These sequences could be parts of endogenous viruses that originated from tobamo-like mycoviruses.

Notably, MP sequences of these viruses contain the DEAD-like helicases domain as the PAFV1 MP, thus implying that these proteins may be involved in a similar molecular function in the host cell.

A maximum-likelihood phylogenetic tree inferred from the multiple alignment of 8 MP sequences confirmed the close relationship between PAFV1, EnFV1, and the endogenous virus of *M. cannonballus* (Fig. 2b). Furthermore, MP sequences from tobamo-like mycoviruses and two fungal endogenous viruses formed a sister taxon. The presence of homologous MP sequences in these viruses suggests that a genome arrangement occurred between *Gammaflexiviridae* and tobamo-like mycoviruses.

However, homologous sequences of PAFV1 HP were found only in EnFV1 and the endogenous virus of *M. cannonballus* (Table 2 and Supplementary Data S3). No homologous sequences were found in other *Gammaflexiviridae* or tobamo-like mycoviruses. It is possible that the HP ORF was acquired recently by the ancestor of PAFV1, EnFV1, and the endogenous virus of *M. cannonballus* from a yet unknown source.

Notably, all the homologous sequences of PAFV1 proteins are either from fungus-infecting viruses or endogenous viruses of fungi. This implies that PAFV1 is also a mycovirus infecting a fungus associated with the mastic tree. A plant transcriptome often contains sequences derived from a mycovirus and its host fungus. For example, sequences from BVF and its host *B. cinerea* were previously detected in a grapevine transcriptome dataset (Svanella-Dumas *et al.*, 2018). If PAFV1 is indeed a mycovirus, sequences derived from its fungal host must be present in the mastic tree transcriptome. To address this possibility and identify a potential fungal host, rRNA contigs were identified in the mastic tree transcriptome contigs by comparing them with eukaryotic rRNA sequences in the UNITE database (Nilsson *et al.*, 2019) and then with the NCBI nucleotide database.

Consequently, four contigs (6473, 1855, 1371, and 1344 nt in size) were identified to be rRNA sequences (Supplementary Data S4). The longest contig (6473 nt) showed 94.85% sequence identity and 98% coverage with the nuclear rRNA of the plant species *Spondias tuberosa* of the family *Anacardiaceae*, which also includes the mastic tree. The second longest contig (1855 nt) was identified to be the chloroplast rRNA sequence of the genus *Pistacia* (100% identity and 100% coverage), to which the mastic tree belongs. Therefore, the two longest rRNA contigs appear to belong to the nuclear and chloroplast genomes of the mastic tree, respectively.

The other two contigs (1371 and 1344 nt, respectively), which were derived from different segments of the rRNA, showed about 96% sequence identity and 100% coverage with rRNA sequences reported from various fungal spe-

cies. The top five fungal species that showed similarity to the 1371-nt contig belonged to the same family i.e., *Teratosphaeriaceae* of the order *Capnodiales*. In the case of the 1344-nt contig, all top five species belonged to the same order (*Capnodiales*), but different families (an unassigned family, *Capnodiaceae*, *Mycosphaerellaceae*, and *Teratosphaeriaceae*). The presence of fungal rRNA fragments in the mastic tree transcriptome strongly suggests that the mastic tree sample was infected or associated with a fungus of the order *Capnodiales*, which may be the host for PAFV1.

In conclusion, PAFV1, a putative member of the mycovirus family *Gammaflexiviridae*, was identified in the mastic tree *P. lentiscus* transcriptome data. The presence of fungal rRNA fragments in the mastic tree transcriptome strongly implied that PAFV1 was derived from a fungus associated with the mastic tree sample. Comparison of genome organizations and phylogenetic analysis indicated that PAFV1, EnFV1, and an endogenous virus of the fungus *M. cannonballus* were members of a distinct subgroup, possibly a novel genus, in the family *Gammaflexiviridae*. Thus, the PAFV1 genome sequence may provide a novel insight into the genomic evolution of mycoviruses.

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

References

Albaladejo RG, Sebastiani F, Aparicio A, Buonamici A, Gonzalez-Martinez SC, Vendramin GG (2008): Development and characterization of eight polymorphic microsatellite loci from Pistacia lentiscus L. (Anacardiaceae). Mol. Ecol. Resour. 8, 904–906. https://doi.org/10.1111/j.1755-0998.2008.02110.x

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

Bruenn JA, Warner BE, Yerramsetty P (2015): Widespread mitovirus sequences in plant genomes. PeerJ 3, e876. https://doi.org/10.7717/peerj.876

Chiba S, Kondo H, Tani A, Saisho D, Sakamoto W, Kanematsu S, Suzuki N (2011): Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. PLoS Pathog. 7, e1002146. https://doi.org/10.1371/journal.ppat.1002146

Dimas KS, Pantazis P, Ramanujam R (2012): Chios mastic gum: a plant-produced resin exhibiting numerous diverse pharmaceutical and biomedical properties. In Vivo 26, 777–785.

Donaire L, Ayllon MA (2017): Deep sequencing of mycovirus-derived small RNAs from Botrytis species. Mol. Plant Pathol. 18, 1127–1137. https://doi.org/10.1111/mpp.12466

Edgar RC (2010): Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Gilbert K, Holcomb EE, Allscheid RL, Carrington J (2019): Discovery of new mycoviral genomes within publicly available fungal transcriptomic datasets. bioRxiv, 510404. https://doi.org/10.1101/510404

Goh CJ, Park D, Kim H, Sebastiani F, Hahn Y (2018): Novel Divavirus (the family Betaflexiviridae) and Mitovirus (the family Narnaviridae) species identified in basil (Ocimum basilicum). Acta Virol. 62, 304–309. https://doi.org/10.4149/av_2018_224

Howitt RL, Beever RE, Pearson MN, Forster RL (2001): Genome characterization of Botrytis virus F, a flexuous rod-shaped mycovirus resembling plant ,potex-like' viruses. J. Gen. Virol. 82, 67–78. https://doi.org/10.1099/0022-1317-82-1-67

Kim H, Park D, Hahn Y (2018): Identification of novel RNA viruses in alfalfa (Medicago sativa): an Alphapartitivirus, a Deltapartitivirus, and a Marafivirus. Gene 638, 7–12. https://doi.org/10.1016/j.gene.2017.09.069

Letunic I, Bork P (2018): 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 46, D493–D496. https://doi.org/10.1093/nar/gkx922

Marzano SL, Nelson BD, Ajayi-Oyetunde O, Bradley CA, Hughes TJ, Hartman GL, Eastburn DM, Domier LL (2016): Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens. J. Virol. 90, 6846–6863. https://doi.org/10.1128/JVI.00357-16

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

Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glockner FO, Tedersoo L, Saar I, Koljalg U, Abarenkov K (2019): The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. 47, D259–D264. https://doi.org/10.1093/nar/gky1022

Pandey B, Naidu RA, Grove GG (2018): Next generation sequencing analysis of double-stranded RNAs from sweet cherry powdery mildew fungus Podosphaera prunicola. J. Plant Pathol. 100, 435–446. https://doi.org/10.1007/s42161-018-0092-0

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.

Stamatakis A (2014): RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Svanella-Dumas L, Marais A, Faure C, Theil S, Lefebvre M, Candresse T (2018): Genome characterization of a divergent isolate of the mycovirus Botrytis virus F from a grapevine metagenome. Arch. Virol. 163, 3181–3183. https://doi.org/10.1007/s00705-018-3975-7

Taylor DJ, Bruenn J (2009): The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biol. 7, 88. https://doi.org/10.1186/1741-7007-7-88

Velasco L, Arjona-Girona I, Ariza-Fernandez MT, Cretazzo E, Lopez-Herrera C (2018): A novel hypovirus species from Xylariaceae fungi infecting avocado. Front. Microbiol. 9, 778. https://doi.org/10.3389/fmicb.2018.00778

SUPPLEMENTARY INFORMATION

Identification of Pistacia-associated flexivirus 1, a putative mycovirus of the family *Gammaflexiviridae*, in the mastic tree (*Pistacia lentiscus*) transcriptome

D. PARK[1], C. J. GOH[1], J. S. LEE[1], F. SEBASTIANI[2], Y. HAHN[1*]

[1]Department of Life Science, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea;
[2]Institute for Sustainable Plant Protection, Department of Biology, Agriculture and Food Sciences, The National Research Council of Italy, Sesto Fiorentino, Italy

Supplementary Data S1. Multiple sequence alignment of the PAFV1 replicase (REP) and related proteins

```
PAFV1          MS-ATNPLVLLGSSLTV---TEHLFNLQTDYLVQYRHLRQQVGSYSQFQLPKPAAAYCAQLGLLHDPHAPGFHSHPLSKL   76
EnFV1          M--AANPLTVLSGNLTA---SEHVFNVQTDYLAQYRALRQVVSGYSQHQLPKCARTHLQSIGICYDPDAPGTHSHPLSKT   75
M.cannonballus MSGGPETLNAVQAQITS---AEHRFNIELDHVVRVRLMEHEIARYGGFRLPPIGHSYFRNLGISTNSEAAGSHSHPVSKT   77
EnFV2          M----ENTLYAHGIVSR---TNHADALVAPTLNELLAKQKITKLIAPFAIPPNLSKLLDSNGVPVDPEALATHRHPAHKT   73
BVF            M---AVPLERAFASVTQ---SFHKDATQGPTTNEYLVARAETQRYAPYAVRKGPAALLAQVGINTIHDSPLSHSHPACYA   74
WWVV           M-----AFQLALDALAP---TTHRDPSLHPILESTVDSIRSSITTYPWAVPKLLLPLLNSYGIPTSGLGTSHHPHAAHKT   72
UTyV1          M-----SFQLALEALNS---TSHRDASTNPILNSIVKPLTSSLEDYPWLLPKEAIPFLLSSGIPNSGLGTTPHPHPVHKT   72
DiaYMV         M-----AFQLALEALST---TSHRDSSLNPVLYHSVSPLQRSLDLFPWIITRDALPFLNSCGIPVSGFGTTPHPHAVHKA   72
NeRNV          M-----AFQLALDALST---TSHRDSSLNPVLNHSVTPLQRSLDLFPWIITKDSLPFLNSCGIPVSGFGTTPHPHAVHKA   72
GVA            MSISVSSQRVAVSNLYTNGSEESVKAIKELKSKRLLETETRLDGLFDYYIPDTLREILTGYGMEFSVHSFQGHAHPVSKM   80
GVE            MSLGASSQRVAYANLYANIGSDKLSEVRDRKASTVNSIEAYASGLFDYYVSDDVYDFLASKGLPLSINCFRTHSHPISKM   80
CVA            M---AFVAKFAEENYFNSLPSNVTDAFLRDGFNAEHNRFEILSRHFAFALKPSQRTYLNDCGIQLAPIASKTHPHPVSKI   77
GYMaV          M---AFSYKTPVESLLSKFTSDEQSKICSTAAEKIASLEKAEHNFFSYYLSDEAKKILTERGVRLSVVAHQVHSHPVCKT   77
CRMaV          M---ALHTITPAEGVLAQFSSEEASRIGASAISNFSKLESEYHSLFHFHLPAYAKSKLSNRGFYLSPFSYETHSHPVSKT   77
CTLaV          M---ALHTITPAENVLAQFSSEEASRIGASAISNFSKLEADYHNLFHFHLPAYAKKKLSERGFYLSPFSYETHSHPVSKT   77
CNRMV          M---ALHTITPAESVLAQFSSEEASRIGSSAIANFSRLEADNHNLFHYHLPAYAKKKLSARGFYLSPFSYETHSHPVSKT   77
               *                          :             *.          * *.
```

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

```
PAFV1          LEDHLLRIVLPPLLT--GKAYDV-MFSKPRKFVRLLQ--DAPPLA-GYTPRLSNYVITPSDFRRYG-RAPA---------  140
EnFV1          IENYVLNVVVPPILR--NMEYDV-AFMKPAKFARLML--AAPPSQ-NHDSALINRCVTDRDRLRYGLDCPT---------  140
M.cannonballus LEDYVLRVVLPAWFTVADKPFAS-WFLKEGKHNRLLNDLGVPDNQ-RHEYPLLNFRITEPDGGRYR-ILRN---------  145
EnFV2          IENYFLNGPVRSAIR---EDCTA-YFVKENSNKLALA--GIPNA------PLANTSLTAIDINRYP-EAPT---------  131
BVF            LNAFHFLKVIPRYLH--GKAEV-WGTKDNWFRKLQ---AQIDSQ-TTTLTHRNYAITARDHVRYP-GTVV---------  136
WWVV           IETFLLHIHWSFQAT---TPSSV-MFMKPSKFHKLAS--VNPNFK-----ELKNYRLHPNDSTRYP-TTSP---------  131
UTyV1          IETFMLYNHWSSLAR---TPSTV-MFMKPSKFQKLQA--FNPNFL-----HLVNYRLTAADTARYP-ETST---------  131
DiaYMV         IETFLLFSHWSSLAS---TLSSV-MFMKPSKFRKLAA--INPNFD-----ELTNFRLTPADTVRYP-TTSS---------  131
NeRNV          IETFLLFSHWSSLAS---TLSSV-MFMKPSKFRKLAA--INPNFD-----ELTNFRLTPADTVRYP-TTSS---------  131
GVA            IENHLLYRIAPNYFS---SNTLI-VSCKESKIKRLRL--KNANSR-NLNFTQYNRLVHAKDHHRYE-NAFR---------  143
GVE            IENHFIFNLIGNNLA---KDSTF-ISFKEDKLVLLE----NKKSRIDGNTCIINRLVHAKDALRYK-DPLR---------  142
CVA            IENHLLYCVVSNMIS--NFKFLVFLSIKEVKAEYIW----NKNTA-DTVREISNRILDIKDAFRYG-PTNTVNGGI----  145
GYMaV          VENHILFNVIKSYID---NSFYV-VSIKSKKLEYLR-----ARNK-MPTLSYINRYIAAKDISRYG-MDKYFNEIAESEV  146
CRMaV          IESHLINVKLPNYIT---EDFLI-VGIKENKLSVLR-----KDKK-MRFLEALNRCVTSHDIQRYG-PSFHFEKAR----  142
CTLaV          IESHLINIKLPNYIN---EDFLI-VGIKDNKLSVLR-----KEKK-LRFLEAVNRCVTSHDIQRYG-PSFHFEKAK----  142
CNRMV          IESHLINIKLPNYIS---EDFLI-VGIKDNKLSVLR-----KEKK-LRFLEAVNRCVSSHDVQRYG-PSFHFEQAK----  142
                ::  .  :                *                                 *  :   *  **
```

```
PAFV1          --------------------------------------PPIIGVGAETLFVHDALHFATPDDVLHLFRSSPQLHTLYAS  181
EnFV1          --------------------------------------DPFPEVTSRVLFVHDAFHFFTPSDVLTLFRSSPSLDLVVAS  181
M.cannonballus --------------------------------------LPPPYVEQEVLFLHDALHHLTPADIVKIFGMSPELNCIIGT  186
EnFV2          --------------------------------------GAPPPCTTPLAFMHDALHYMTAADIAEFFRRSPHCHTLIAT  172
BVF            --------------------------------------DHPGSCTAGTLFMHDALQYMTPLDVYTLFATSPEMHSLVAT  177
WWVV           ---------------------------------DLP---------SHPTVFMHDALMYYHPSQILHLFQAKPNLEKLYAS  169
UTyV1          ---------------------------------HLP---------TTPSVFVHDALMYFTPAQILDLFLSCPNLETLYCS  169
DiaYMV         ---------------------------------SLP---------QHEIVFMHDALMYFTPHQIVDLFERCPLLNRLHCS  169
NeRNV          ---------------------------------FLP---------QHEIVFMHDALMYFTPHQIVDLFERCPLLNRLHCS  169
GVA            -----------------------ELDVGNLTN-------LINKETQSECIFIHDEVQYWSLDEMQRFLGSLSKVDRVVYS  193
GVE            ----------------------NLWFDEDQK-------TMKGLAEPDRVIIHDEVHYWNLKDFQRFLSYINC--PLIYT  190
CVA            --------------NNFSFFCANLNRR----------FNNRAIKPDCFFIHDEVHFWSPAELCEFLFTVEP-KNVLAT  198
GYMaV          TGILNRKKKRFFNLFESAKKEFNSANYDPSNSETLKDLVPQALTTKAKRFFLHDELHYWSLEELDLFLCTVKP-EIIVAT  225
CRMaV          ---------------SNWRSDFSGVNLSAGVQSLLPRILFDKGKMFDSQIFLYDELHYWSMKDIVDFLEISKA-KTIIGS  206
CTLaV          ---------------SNWKTDFSEVNLSAGVQSLLPRVLFDKGRIQDAQIFLYDELHYWSMRDIVDFLEITRA-KTLIGS  206
CNRMV          ---------------SNWKNDFSDLNLAPGVQSLLPRILFDKGKTRDAQIFLYDELHYWSMKDIVDFLEISRA-KTVIGS  206
                                                     :::*  .  .      :.  ::        :  :
```

```
PAFV1          IVIPPGLLLNSA-RGHTPLYDYEVKG--ALIHYAPDGRWDEHYEQPLSCLWLLKTGFLRSLDLT----ISITTAATYDSV  254
EnFV1          IIIPPGILFNRS-RGSTPMYDYRTEG--NVLHYAPDGEWSQSYSQPLDCSWLLLTKFVEDLELH----ISITTTFQADAS  254
M.cannonballus VEYPVPLLLGAR-AFECAAYAYRLTGD-GNFEHAPDRVWSESYTQPLSGRWLLDTASMHANGID----LGVHIQWVY-PC  259
EnFV2          MVYPPEARAYAP-SAYPELYELRYSD--NNVTYIPEGHTAGSYTQPRLN-EFIDCHTIATENHR----LTVTFHETLFAH  244
BVF            AVIPPESVDRLP-AFWPELYQLAYYE--DHLCYAPDGNFADAYNQPLAAHQWMTMKSLHGPDFT----LSVDVPASRYSH  250
WWVV           LVVPPEALLSDH-SFYPQLYRYSTTR--NTLHYVPEGHEAGSYNQPADALSWLRINELSLDDLH----LSVTILESWGPV  242
UTyV1          LIVPPESNFTDL-SLFPNLYTYQINQ--QTLHYIPEGHHAGSYNQPLQALSWLKIGSILGSQLS----LSVTKLESWGPV  242
DiaYMV         LVVPPESSFTDL-SLFPTIYTYQLNS--NNLHYTPEGHHAGSYDQPRSAIRWLKLRRITSPFFN----LSVTILESWGPV  242
NeRNV          LVVPPESSFTDL-SLHPTIYTYTLNS--NNLHYTPEGHHAGSYDQPRSAIRWLKLRRISSPFFN----LSVTILESWGPV  242
GVA            IIYPSEVEAGYSQSLFPEAYTFDLKY--GRLIWYPDGKAEGAYTQPINP-WLLRCSKTEDSKGR---PWTITKLQTIGAH  267
GVE            VIYPAELHAGYPFSLYPELYDFRVSEDGKTFTWMPDGKCCGSYKQPVNP-WLLSTNKTMDLKGR---TWTLTKLESVGSH  266
CVA            VVIPPELLEGLDYSFNSVAYDFKKVD--GNLYYFPDKSKGKPYQQPMDP-WLLKCNKISMIKNGEIFSYSIGLLESVGAN  275
GYMaV          HVFPKEILKGVHHSYNKWMYDFKVEG--EKLFYYPDGCLEEGYEQPMHGGLMYYLNSFTTTTGE---IYSISLVYNLFSH  300
CRMaV          FVFPSEILAGARTSLNPWAYEFKIKG--DKLIYAPDGVWSESYEQPLSAGQLLKFNKIMTRNGS----YSVQVRDSIYSH  280
CTLaV          FVFPTEILAGSDRSLNPWAYDFKIQG--DKLIYAPDGVWAESYEQPLAAGQILKYNKIITQQGV----YSVQVRDSIYSH  280
CNRMV          FVFPTEVLAGTYKSLNPWAYDFKIKG--DKLIYAPDGVWAESYEQPLAAGQLLKYNKILTRLGP----YSVQVRDSIFSH  280
                *                *         . *:      * **                          :        .
```

```
PAFV1          HLLQLTRGITHAG-----------------------------------------SPTYVTMPNTVVLPLDGAPFSDYSRR  293
EnFV1          FVITLQRGHSLQH-----------------------------------------STTFCVVPDKTILPS--SPLLPLQEP  291
M.cannonballus HTVVITRGLSLVN-----------------------------------------SPRTLQVPGYKMVQTSRGP------R  292
EnFV2          HLVIISTKDRLPQ-----------------------------------------RHRQFRNPAYVTLPAFLHPFTPVQDR  283
BVF            HIFVISKRPGLPD-----------------------------------------THRDFMCPDLVELPKDFFPGANKADK  289
WWVV           HSLLIQRGVPHPDPALLSPLPLTGHDLFSSYHQPRID----------------LVSFRTPDAVALPEATFLDQPLRHR  304
UTyV1          HSLLIQRGLPPSHPSRSPFLSPHSTPSISRKRSPQRDLLQLTSAALFPPPSLQDQFSTFKVPDCLELPQATFLHQPLRHR  322
DiaYMV         HSILIQRGLPISD-----------------------------------SLAP------PHASFKVPQARVLPEATFLNQPLRHR  285
NeRNV          HSILIQRGSPVSD-----------------------------------SLAP------PHASFNVPQARVLPEATFLNQPVRHR  285
GVA            HLFSAIKGSYLTE-----------------------------------------ESYKYDNFTIINPNDILR-GRRGGKP  305
GVE            HLFLCTPGDLVTE-----------------------------------------DEAVYNDFTLVDPKLFSSLTQRMPK-  304
CVA            HLFSFQRNKVVE------------------------------------------STRFFNDFDCLDMRKLLPINVENGKI  313
GYMaV          SVFQISKGNLKTE-----------------------------------------TQRSFSGFQASPTNELTAFGGTILKT  339
CRMaV          CLVIINRDELLCE-----------------------------------------EFRVFSDFDAISIRKIGYLGGNADDI  319
CTLaV          CLVIINRDNLLNE-----------------------------------------EFRVYSEFDAVSIRRLNYLGGNSDDI  319
CNRMV          CLVIINRDDLITE-----------------------------------------NYRVYSDFDAISIKRLNYLGGNADDV  319
                .
```

```
PAFV1          L---VNKAALTFLVG-YAKSVTLKSPQQLAARLRLWLGDHRENAISSSENDILAQYVYAAVTTY----NHCSDRHLFPCR  365
```

```
EnFV1          VPH-VNSAAAHFLISQYSNCVSAPTKVQLRARLRMWASDHPEHHLGPDEYETLASAARDLIKHG----DAIGTGGLPVLS    366
M.cannonballus L---LRQEPIDFAIN-YSYPVGHATDASVQGRVNMWCNDHRGVLVPTADRRLMADVAY-------RERLRLKQTGLPSGF    361
EnFV2          M---FPKQVAVANLA-YARSVNRLTPLQATGRIRAVVNDTK----TPLSLEQIACLEHHMVAVSHLHSPILSSPEGPRGL    355
BVF            L---LPRTLVNQMVE-YAASVKRATIKDATSRTRAYVKDQKYAAVTPVQQMYLSWFGSGL--------SQMQFPDRPVTP    357
WWVV           L---VPRTVYNALFT-YTRAVRTLRTSDPAAFVRMHSSKPEHDWVTPTAWDNLQTF-----------ALLNVPLRPNVV    368
UTyV1          L---VPTTVYNALFT-YTRAVRTLRTSDPAGFVRTQSNKPEYAWVTPNAWDNLQTF-----------ALLNACHRPAVV    386
DiaYMV         M---VPSEVYDALFT-YTRAVRTLRTSDPAGFVRTHSNKPKHAWVTPQAWDNLQTY-----------ALLNAPVRPRVV    349
NeRNV          L---VPSEVYDALFT-YTRAVRTLRTSDPAGFVRTHSNKPKHAWVTPQAWDNLQTY-----------ALLNAPVRPRVV    349
GVA            LY--LRARMIKPTLL-YLLVLKKSDSNSAVAKLRMLSSREE----NMDEALFVAQLAKQIKDTG-LYDKMGN-PNLRSIL    376
GVE            ----LRAGYMHKVIH-YLMALRKPDAASAVSKLRQLSKGDE----TTDEMIFSGTIATQIAELK-YFTEVGGLIDLTRAL    374
CVA            KGYNIRTWVFKKILS-YIVCLKKGDSESSLAKLRQLSDSSP----SSDELLLIGDFFDLMTRVK---------------    372
GYMaV          FG--VKSSLIKKIYL-YLRTLKKPDIESAMAKLRQLEPDP-----SGQTIVFVENFSEFLLSHK-TGRNLIE-EGLASKF    409
CRMaV          IP--VRHEVVLSIFK-YIRTLKKPDLQSGMAKHRQLVDNP-----TGFEIRFIEDFVQFILEHH-EKFNLIE-QKFSNFF    389
CTLaV          IP--VRYEVILSVFK-YIRTLKKPDLQSGMAKHRQLVDEP-----TGFEVRFIEDFVQFILENH-EKFNLIG-QKFSNFL    389
CNRMV          IP--VRYEVILSIFK-YIRTLKKPDLQSGMAKHRQLVDDP-----TGFEIRFIEDFVQFILENH-ERFNLIG-QSFSNFL    389
                       .              *    :        .   .   .
```

```
PAFV1          PHHFITRRG----P----------IRWARLFAHSLDGRILDLARDALDPPAVRYRCH----TEDLHLSSDTRPSMLTTGA    427
EnFV1          FHHLLTRRG----P----------LGFLRGVIQAAPRLLLSANTGEHGLLPATTMTH----HIPVYLNGRHTNSVLFGGP    428
M.cannonballus IGSVYTGNG--------------ALVDLALTTTLGRFTKLIPSHPDPGLGTELAIWHSPRKALYADGTWQLSQLFVNY    425
EnFV2          LELVWA--F-ATFPVHLLLRQP--LAFAEFAINSRYTQLETVKQTVTPGIT-------------YLHAGHRLSAAYTGK    416
BVF            LDNFATALWYRLLP-SRLYRLP--ATIRNFQLKAFVDQLRTPQYLLHFKLD----------EYHLSSNDKYSDFYRNE    422
WWVV           YHVLQS-------P----------LAALKLYLKQHWRRLAATAAPILSSLTLLQHFL------PLSL-------------    412
UTyV1          YSFFLN-------P----------FKRLLLRLQQHWRTYALCAAPVLALFPLIPLLS------NISLP-----------    431
DiaYMV         YNFFLS-------P----------IQKAHLYLAQHLNPFLVRALPFLGLVLPALKVL------TTGL-------------    393
NeRNV          YNFFLS-------P----------LQKAQLYLSQHLNPFLVRALPFLGLVLPTLKVL------TRGL-------------    393
GVA            SESFYD-----IAG-SLFTRLFNRPEYDARCLEKFIRSCETTEIHVE---------------RRYMEG-----------    423
GVE            KVGFAR-----LFG-QSIEYFSNKKWYHLDCFTNTVQFMVKAEVRIK--------------------------------    415
CVA            -----------IFN----------------------------------------------------------------    375
GYMaV          EKFIIG-----LLP-NLVRKNF--ESHYSSNLCEFISNLEDDKVVVNCQC-----------------GDSKFSDW---S    460
CRMaV          SSACIN-----LLP-RYMQRFF--NSFKGYSLGKFIEEIEPFSFTLRCS--------------------TYSRF---G    436
CTLaV          SSACIE-----VLP-RYMQRFF--KSFKGYSLGKFIEEIEPFSFTVKCR--------------------TYSRF---S    436
CNRMV          SAACIE-----RLP-GYMQRFF--SSFKGYSLGRFIEEIESFNFTIRCK--------------------TYSRF---G    436
```

```
PAFV1          LPAPSIGRVRSAVHSLLYPPDRPPRALPF-----------DPLLGVGLRSTLTRRFVASGLRLG----------------    480
EnFV1          APARGIREYRAALISLILKPQATSDSTPH-----------IPLSGMGNAWSLSPRIARTLLRLV-------LIYASSEVP    490
M.cannonballus RKLDK-KDLTAAMASIIFRDEKIDWDMVN---------VLPPAMGQGSPRRLSDEAPRPLVLGV--AGLATLAFALRKGS    493
EnFV2          RPAPSARAIGQLIGKLI------TGHHPY---SSLGTNLGPNGRTTPSSLRRGGSPTRALICYAPMAILGGLAVAHNNP-    486
BVF            SRQASSKDMLSAFEAICF------GGAPELEKKILGNNNPAPFATVPNP------PSQTALGLA--TITAAIAFILPNR-    487
WWVV           -PLPEVKSICLFHKEVYR----KRELLPSFHPFQVC----PPLRT--------------------FFQETLRVKPNCRP    462
UTyV1          FSLPYLQSIL------------RRQIHPS-STLDLS----FPLLP--------------------FQKKRVSIPLHPR-    472
DiaYMV         -PVPRVSSLQVAHRQLVGP---HPEPSPT----------AHPILRQIVA---------------FSQQNPFFLELCQR-    442
NeRNV          -PAPSVSSLQIAHRVLVGP---QPEPSPL----------AHPLVAQLIL---------------HAQRNPLFLELCQK-    442
GVA            -----------------------TRRGAPFKVQNVID-----------------------------------------    437
GVE            ------------------------------------------------------------------------------    415
CVA            -----------------------KRSPW-------------------------------------------------    380
GYMaV          VVYPKIDEMRRKLINEV-----VKFKDPVMELFSNH-LKDGKLRRRPQKY--------------FISNAGTEMLGESRT    519
CRMaV          FKTSFIEEEEAA----------VADKDPLCLTLKLANNKPSCFESYPD-----------------LIFNAHTLVFASSHP    489
CTLaV          FKNSFIEEEEAR----------ERDGDPIYKQFSKSQSFNS--TAYPD-----------------CLFHASQSVFSNPHP    487
CNRMV          FKESWSELEEEE----------CSKLDPFFSQLSLRREMSS---DYPT-----------------FINQAVLSIFSEPHQ    486
```

```
PAFV1          -----LLCF---------------SYRAVLPAVSIA---------------RSLLPRFSWR-LVAKPSTLLFLALSAPAL    524
EnFV1          --LSRMVSLLVK------------SLRRAL--------------------KNLLTNPK-----RSIKTLLFIAQCSALF    530
M.cannonballus --ERKLLLFIVR------------TLRWIP--------------------KSALVR------LIRQRPLLLPLFNVLDRY    533
EnFV2          --NLRLGLLSTHK-----------------------------------PSMPPCPSLTTILRSAATTAAATGSRA--    524
BVF            --AARLTRLVARY-----------AIRAVKEIVN----------------PSFLTIPRFGLLSHSIASWLQTDGAVITY    537
WWVV           GSPQHF--LLQRV---------LNSLRPLAPLLA--------MTPLYSRKPKPLLPHAELSWTLKSFALPWQASLALL--    521
UTyV1          --LQHLNQLIRQH---------LPQLLPPLPIFP--------VFKLHQPMLRQIIPFPHLT-------LPLVSSLFF---    523
DiaYMV         --TPFLNRFFRS------------ACRPPSA--------------FRLELQRSTLPLPSWL------AVLLISMIPV---    485
NeRNV          --IPVFNKFFLA------------ACRPPSA--------------LRLELSRSMIHLPSWP------LTILISLIPV---    485
GVA            -------------WVEDVSANTLSEVNFLEVSWNDRTSTP------YSIGSTHMVDS-------------KVRVPLTRI    484
GVE            -------------------------------------------------MGYRPMVDGLG--------------------    426
CVA            -----------------------------------------------------------SFLSDAKNYVDSW    393
GYMaV          TMAKKAAQFFLYSFS-------CSGLKFAKKRILNEVMRNDFTFFMRNISFRRCVEAGN----IDHISFLIFLRNYFDER    588
CRMaV          SITLMLVKSFINMWVGKTNDRYYQSLVALRQALNQKGAK---LFMLHNESYNSLVILAN---LMDSHLFKNLLRNEIRKR    563
CTLaV          NLIRRLVTLFISTWVGKTEADYYTSLLSLKKSLSQKGLK---LFRLHDDRYFSLTRLAN---LMDSFSFKATLKNEVMKR    561
CNRMV          NLIKRLVTCFIAAWVGKTDEGYYTSLFSLKKSLSQKGHK---LFLLHDDKYSSLVILAN---LMNSFIFKLSLKNEVSRR    560
```

```
PAFV1          Q--------------------LALDASGAALYHLFRRGFDLTADCLTVPVRWSLYKLMWRVPALRSM------------    571
```

```
EnFV1          S--------------------LLAAALAATKELLIEAAVAKIAQGL----RAVWYPILASSPAAQSL-------------  573
M.cannonballus F--------------PPILPTLWNGIIKPALLLLDRKMEDLIASFL----RHLRTWLGDRFPIVAVI-------------  582
EnFV2          --LTGDITAICTIAGAYLAYRLTSGLLAGIADTVTDVLVTKIADLP----QSMAITILRFASAYAGPRVAEQLYKV----  594
BVF            ----GLAPKIASWSAPIW---YNAALLAPTDMRSCQVLVDIANDLR----FDYRYGTMRLAVTLGLTLLNPFLRRLVLLP  606
WWVV           --------------------AISELSILVHKMTSPPALQAQHDIYHRHMHPGSYVLQWERTPLRLS-------------  567
UTyV1          -----------------------LPNLTHYLLHPPTPQQMHDQYHLHLHPPQFHLTWMLQDLEVQ-------------  565
DiaYMV         -----------------------MASPTLTLVLPQSVQSLHDGYHTHLHPSEFNLSWPLESFHVP-------------  527
NeRNV          -----------------------VASPTLAFVLPQSIQSLHDGYHTHLHPPEFSLSWPLESFHVP-------------  527
GVA            L--------------------RAHELITGAQEKVEINLQRFVCSARAL-MYFRQYLIK------------LGRFKFMESR  531
GVE            --------------------PEHEVVEPA-------------------------------------GLEKRREYQ  444
CVA            I--------------------IQSPFLRRIFPVGSRAITELIRDWIA---NAESLKIQTTCSSLTFS-------------  437
GYMaV          FRKISDDPLLLTY-RSCSEEEEKKKAIKQAIEESKRKFSGVIKELF----NDNKFKIKGIISDLYTVE--SFRSQDPEPT  661
CRMaV          LRLRNSVRGLLRN------DLPPSCPDAKREVRFISCYKSLLADFKK--MSEECSPILTLIATGGPFEQLNAMKKDFVLI  635
CTLaV          LRLGGSLRGLLRY------DPISENPSDKKDQRARSNFSSVVSELLS--ISSECPPSSTLTKTSGLQQELWKMKKEVVTK  633
CNRMV          LRLRGSIRGLIRC------DLPMENPADKIQGRVINNYAKVVSELLA--ISNEGAITCILLETSWLSEQLLNLTNQFLSP  632


PAFV1          -DVALHQLLV------PNDEQAL----FEVDPELLFGI------------STSDQPPSSADSPPVASSPSPPPS------  622
EnFV1          -DFALHSFLA------PSHGSLA----FQTDPTLHANCSRNNPFGSRETTSEHDTTDDASSQAPDANPPSQPYR------  636
M.cannonballus -DLALHQLLS------PQYIMLP----GNHNPTGNNDD------------DSSDSGSNDDDSGDDGGDSDDDDSDGDDFH  639
EnFV2          -NLELEGTMAN-----DSHPFVP----RETCPTDEEEV---------PTTDPGNEPHTRSRRSNSPHPTAPSTSSTDEDD  655
BVF            FKVLWHSYLPSIKNKLDSMVLMS----ISTLPDNISVP---------LLTLRGNLNRQRCPDNKDPTPIHESDD------  667
WWVV           ---------------RSTAFLPFTPTISTAP----------LASSQAITAPLFSNMNARPLPPSTHPSRPPAPTTTEVL  621
UTyV1          ---------------TTFPFLP------------------------FVLP--------PTPTLIHPTTPPTPIPV-FL  595
DiaYMV         ---------------TRRNFLP--PGMASNP----------SPPPTPFVAPPFATNSPHPPP----PEPPTASL-----  570
NeRNV          ---------------ARRNFLP--PGMAANP----------PPAPLPFVSPPFPSTSASAAP-LITPVNPAQSV-----  573
GVA            AIKDIEDIQSG----LEEGMITEEEAELRLLP---------------SARWTEEENCVGDAKSEPSDERKADK------  585
GVE            GLRGVEAIADG----------------------------------------------------GVPERKRQP------  464
CVA            ---------------DSFEMIK----------------------------------------------------  444
GYMaV          HLNSVEQFQSE----FDSLVLPA--------------------VEGAVSSTEFDEKKGDCKSLIDISDEIVP------  709
CRMaV          NQPSVGKFEK----------VKKCISKSTDASPDGGKEHKLE--------PQHSEGVNAKTDEVIPKRVDQPEPKTEEGK  697
CTLaV          QAPVLNEEQKP-------YKIPSEREKCSTQAIEKNEEPKSEVVSTKGETPPETSKTTSKFGELLATPEATAVSEPTDNV  706
CNRMV          SAQKKNVQKQE-------EVIKKEQSEIKEMETQRKKRDREDSKDSKNAEEKE-EDTTFKDKSSLASSQMEIDSLIDGNH  704


PAFV1          ----SPPGPDT-----------PP-------PSLPLPEPA----------------------------------------  640
EnFV1          ----APTCEDAE------DDSTPPS-----RPSSPSPPPT----------------------------------------  661
M.cannonballus DAYPEPPGPDFDGDEPFDDDDWPP----------PPPRAPRRGRGGR------------------------GGGGRGG  683
EnFV2          D---TDPEEDNCPTC-----SGPPDYCHCTEPT-PTP----------------------------------------  683
BVF            ----NSTADDKCLHC------------------P--------------------IHCLK-----------------  684
WWVV           Q---TQTES-------------PPNHSLPDQPF-T----------------------------------------  639
UTyV1          Q---EPSIP-------------APLPSSRLEPS-P----------------------------------------  613
DiaYMV         ----GPSQ--------------PLAPHHPTPPTNPPAPA----------------------------------------  591
NeRNV          ----APPA--------------PPALASTLEPQLPP----------------------------------------  591
GVA            -------------------FGGARNLCLEEI-----------------------------------------YSEKLKG  604
GVE            -------------------YSMVFSDCERMC-----------------------------------------DAGDPVG  483
CVA            ----TDSISEFGEHILG---NISGGLSSAIEAC-RNWLTSKRFSGGSYSMVSRKGLLVDSMHQNTSFSYEVFVDLFPSTI  516
GYMaV          ----QDPFAAF---------FLSGSINSSVEMA-QPLEPTVYSSHYHYCGCGLK-IVVNSFNNEIAISGMIFCD--SMKG  772
CRMaV          T-----GSSDQ-------------FISSIIKVG-PFKEPSTIS----------------------FVEGLDFSKGHNHKG  736
CTLaV          L-----ASSDL-------------FISSIIKTG-PFGDHGVIE----------------------FIRSLSFNDGHSHNG  745
CNRMV          LK-QESSSCDT-------------FISSIIKVG-PFSDSNVIE----------------------FVRNLSFEQGFQHSG  747


PAFV1          ---AAAVDTPRAPRSAQNPSGLP-------------------LARLYPTIEADASFPERPSPPLTIGDTYSKVTLLVVN  697
EnFV1          ---AEAPTPTAIPRSARRVTPVR---------------------RNPTSIDVVIGGPNDAPDRFRWGVSFPPGCTLNLT  716
M.cannonballus R--GGRQDDDRGERRGRGDDQQPGRGGRGGRGRGRGGREAPPQPQRRFPGPAPPPPRPDNEIPALVHNVSTGVATGNWRN  761
EnFV2          ---SELFEPSPTPSRPTSPQPPP------------TEAEPETNAQTQQEAEPTPNAPPNNG------------------  729
BVF            ---SKLSEPAASPDTVPEDNPLQ------------EGDLTEEVYAAALAKKDTPPPYPTRND-CLLVALADGLLTKRILW  748
WWVV           ------TPPSLSPTAHPPTSPHP-----------------------------PPPHPSSDD------------------  665
UTyV1          ------VSPLFSPNAPTIPSDIP-----------------------------SEPTPTLPT------------------  639
DiaYMV         ---SSSAPPELSLSQNPRSETSS-----------------------------PPSQPSPS------------------  619
NeRNV          ----SQPPPEPNADPSTSHSRPP-----------------------------APSPCSPS------------------  618
GVA            REVAFYSRYSKEYKYNGGTHRSLGWDKALNELVQELGLD------------------DSYDHCLIQRYTAGGSIG----  661
GVE            E------------------------LNEFKRMLG--------------------HPNNCENQK------------  502
CVA            R-PAFYSDPN----FNKVESFEPEWDFLLGS------------------------------------------  542
GYMaV          RMGAFYSRDNSGYAYKGFSHSSQGWLNGLDKLISACGEE------------------PTDYNQCLAQRYEEGSGIG----  830
CRMaV          RKSLFFSEGGFSYGFGSIVYPSQGWPNAFKELYG-------------------DRFNSCLVQKYNADATLG----  788
CTLaV          RKALFFSRGGFAYGFNSVTYQSSGWPSAFEEIYG-------------------DRFNSCLVQKYEKSAKLG----  797
CNRMV          RKALFFSDGNFSYGFNSVKYESHGWPIVFKELYG-------------------DRFNSCLVQHYTKAATLG----  799


PAFV1          LPAPIPFAGLLSWLKYTDAPEVSFRIASGSWTGSDA--LEMVRSFPHFES-------SLNNPDLRVSWLHR-----VPTI  763
```

```
EnFV1          LLEPLHPYLIATWVCEGFVRQVNIPDEYTHLVPFDQ--LDEVVTSVSSSNGI-----TLTFPRILPYPADD-----VPCR   784
M.cannonballus FAQPGPQTQELHVLL-TDGTVEDFMKSQTAWIRGDMFVHEPDQGFMGFHL----------------------------   810
EnFV2          ------------------------------------HPPGRT-------------DLDPA-NVPLPESD---NEAED   753
BVF            FSCIKMFGNTACWV--------------------IHDTGKAFDWMHLKMLAEQLKLNVS-IEVAAEHDHHWTGLAKR   804
WWVV           ---------------------------------------------------------------------APLR   669
UTyV1          ---------------------------------------------------------------------ITLT   643
DiaYMV         ----------------------------------------------------------------------PSQ   622
NeRNV          ----------------------------------------------------------------------ASP   621
GVA            ----------------------------FHADDEPCYLPGGSVVTVNLLG-----DAT----------------FEVR   690
GVE            -----------WM--------------YF-----EATPTPGQNFVYQEVAK-----DIYIK-------------VYLI   532
CVA            -----------WI--------------FNHKQVRLCYEPSDS-------------------------------H   560
GYMaV          --------------------------FHSDDEAIYPKGNKILTVNASG-----SGQ----------------FGIR   859
CRMaV          --------------------------FHADDEQCYDQDHEVLTINLFG-----SAT----------------ICFT   817
CTLaV          --------------------------LHKDDEDCYDDEHEVMTVNLFG-----TAT----------------LIFT   826
CNRMV          --------------------------LHKDDEDCYEKDHEVMTINLFG-----EAT----------------LYFT   828


PAFV1          SYFLDAASAPGSV-------------------PCAQLPPPGESIDSPHVAVFCDELTSPDELVPWVERNPDTVFTLIG-   822
EnFV1          NYTTEFINCPAATY----NAISKDRIMIIDCEDPMAYLAHHSAILDANAIAYCAHPEFAALYRIPA--RCDDTLLEQLR-   857
M.cannonballus --PAEWEKVNEGVY----RDASNRRA-------PMACRARKLEDAYAKAVANPCRLWGIRHEDLPA-----------   863
EnFV2          PSPAATTNKSTKSH----TEAERDNAPP-----PAN------------------------RAVNRRNPKNTT-----   791
BVF            VGPIDGRKIALRWY----------------PYHWELSDRGR--------------EDYGPPPPPPPPST-----   843
WWVV           PDPLAMFPELIPAS----NPSSSRPAVP-----P----------------------GTH--PAPPTAPTTA------   707
UTyV1          SSDVSSDSTVLPSG----ALSHSEPIIA-----P----------------------ALDSQPSPTPPPIT-------   682
DiaYMV         SSPLPAGPPPHPSD---GSCRCADCQPE-----HQHLRSINE--------------FGALNQNPAPTPSPLL-------   672
NeRNV          DSPSPATHVENPTH--GSCQCSDCRPE-----PQHLRSINE--------------FGALNQNPTPSPSPIL-------   671
GVA            ENQSGRIEKK--------ELHDGDVYVM-----GPGMQQTHEHRVTS----------------HTDGRCSITLR-----   735
GVE            TNP--------------PLHE----LL-----LEGLITLEEYK----------------------------SMR-----   555
CVA            ADSEEDVNRTSETSSPEKTIADSSSVSP-------------------------------------------------   588
GYMaV          CNADE----------FYLNLNDGDFFIM-----PCGFQESHKHRVTALSV-----------------RVSLTFRSTVAN   906
CRMaV          KGDFSALNTSNPKLYLEVGLDHCDWLLM-----PRGFQRNYKHSIKG----------------TSEGRISLTFR-----   870
CTLaV          ADGAKGLERADPSKFLEITLSHGEYLLM-----PNGFQKKFKHGVSC----------------TSAGRISLTLR-----   879
CNRMV          SDSSEKIDRNDPKKFMEITLSHGEYILM-----PRGFQQSYKHGVKD----------------TQAGRISLTFR-----   881


PAFV1          --PELSPTLGSSLPLSVEYNHVGS-------------------------------------TCRLWS----------   850
EnFV1          --TQLNPLPPQDLASGSASSASFR-------------------------------------SC-------------   881
M.cannonballus --QDKRPPLPENVHSIDMGGNT-------------------------------------WFMFAP------   889
EnFV2          --NTPEPPRPNTRP-------------------QKAASTWNTAIAAWSRNDENDPN--------------------   826
BVF            ----TEPPPPPNPP-------------------------------------------------------------   853
WWVV           ----ARSIPPALLP------------------------PPLPSDHTAVGPVLPFHQL--------------------   736
UTyV1          ---------------------------------------LQNDPTAIGLSLPFSTL--------------------   699
DiaYMV         ----NRSEPKADLP----------------------GSDLLSDPSCVGPVVEFECL--------------------   702
NeRNV          ----NQSDPKSDLP----------------------GSDLLSDPSCAGPVVEFECL--------------------   701
GVA            --NKTIDYESR-------------------------------------------------------------------   744
GVE            --GQLPGTETSIV-----------------------------------------------------------------   566
CVA            -------------------------------------------------------------------------------   588
GYMaV          CCEKKNELVPENLRSDCIESSVQNLSL-PNSFYRVKLNSIHTLSQRFIISDIDKFNEFKVVGDGNC-FWHCVASLLGGSA   984
CRMaV          --KQRRTLEGSLIQSRAESGDSNSDDG-EGGFYFEEINKCSITSAPDSVK--CSLSVFPVKADGDC-FWHAVSSIFGLEA   944
CTLaV          --KQARSMSGATLHAGADGGNNNGSGNEEDDSYYEEMNKCSISSAPDSAK--CSLSVFPVKADGDC-FWHAVSSIFGLDA   954
CNRMV          --LQSRDLAGKKIPTLNNSNLEPDDGN-EGGGYYEEMNKCSISAAPSSMQ--CSLSTFPVKADGDC-FWHAVSSIFGLEA   955


PAFV1          -------------------------RPVSLASSDP---------------LLGRGSTFSIPEPPPPGNDT------   880
EnFV1          -----------------------------------SPSPV-------------VLSG------SSSASLPTQNT------   901
M.cannonballus -------------------------------------------------VLAGNGTDDLDPDLPAPPPRN------   910
EnFV2          -------------------------RPKWVPHQTG----------------------TVNYPVPVPTANT------   847
BVF            ---------------------------AASPYHSPTVE--------------------DEPLEEIPLANSNGEFSTF-   883
WWVV           --------------------------HPRNYPPNTAD-------------YHTRLRVLPPSPLPHPILN-------   766
UTyV1          --------------------------FPGDHFPDSAL-------------FLTRARHLPSSNLPMPSLN-------   729
DiaYMV         --------------------------FPGVYHNSNGT-------------FLTRQRAQASSSAPLPAKR-------   732
NeRNV          --------------------------FPGTYHHSNGT-------------FLTRQRVQPSSQAPLPQKN-------   731
GVA            ------------------------KGDEGSEYEEDKAE-----------------------------------LDEGI   763
GVE            ----------------------VRKDEVQETEADDS-----------------------------------FDNDM   585
CVA            ----------------------FVSSSHEEGTSSQQSE-----------------------EKPISEQEK------   613
GYMaV          ETCKEICKNHCEAHHLAPSEKL-----QFDGSNWAEDESIFIACTAFSIR--ITMFDLVSSVVHEFKPVSSNEE-LNVNI   1056
CRMaV          KELKQLVHDRAIAEGCIDKCHMKDFLHEMEPKVYASNASLAATCYLMNLKLIIKLTGLEDDSWVVVEPLALSNERASIGY   1024
CTLaV          LELKNLVKERAIEEGCVDQKHMKDFLHEMEAKVYASNASITATCFLMNIKLIIKLVESKHSGWVVVEPLNSSNEKISLGY   1034
CNRMV          LELKQIVKERAAAENCIDAVHAQDFKNEMEAKVYASNASVTATCFLMNIKLVIKLVDLDCDSWVVVEPLASSKEKTALGY   1035


PAFV1          ------------------CLLDAFSAAAAVPRDEVWRVLIGLIGLNR--AVVYSRKPF-ALGSPE---VEAFALAFSCHV   936
```

```
EnFV1         ------------------CLIDSLSRATGTSRHAVFSAITQGWRQEK--INEYLAKPF-PLDDVD---LAHYCKVTHHYC    957
M.cannonballus -------SGW--------CLIDAVQDTTGRPRQQIWRHLLQNCPDVEPPLPDVNMSPK-ELTHSH---LRRMARPMEICF    971
EnFV2         ------------------CLLDAFAPHL--TRAAAWLLLLRHCSPPP---TTEEVNST-GLHIGHA-AIIAYSHHMRLHV    902
BVF           ------------------AFDDMAADDL--NPAQPLSSLDLDEEPSA---KQEPLEPA-AIKTSEPEPLVPESEEV----    935
WWVV          ------------------CLLTAVSLQTNVSEEHLWHSLQTILPDSQ--LDNDEVRTF-GLSTDH---LTALAHLYNFQA    822
UTyV1         ------------------CLLVSFSSLTHIPVDTLWHTLQSKLPDSL--LTNPEIETL-GMSTDL---LSALCYFFHYSC    785
DiaYMV        ------------------CLLTSISPQLGVPEATLWNYLCEMLPDSL--LDNPEIRNY-GLSTDH---LTALASRLNFEC    788
NeRNV         ------------------CLLTSIAPQLGVSEAVLWSYLCEMLPDSL--LDNSEIRNF-GLSTDH---LTALSSRLNFEC    787
GVA           EYLQK--------NQGNMCSLKAFADHMQLSIPSVIAIVNGASPQ-----TLREIEDG-GISLAT---LVNLSKALDFPI    826
GVE           VKLKE--------IKKNLCLIQPIAEHFALKAPVLISKATVEIPN-----FARYLSDK-GLSMPG---LYMLCKNMGLTL    648
CVA           ---KE--------TRKNDCFFKAVGETIGIPANSLIERILCSDSEDLK-PVIEQLNLDHPISSKL---LEVCCKFLGYRV    678
GYMaV         VFDGS---HFNLLVPKEGCVIRAIAESLNRKEIEIVKVLALNENS----KILEELNTGLGLPMNL---LEDCFKIFGIRA   1126
CRMaV         LVLNQKCHHFDLAVPKEGCVVRAVSEFLKQNPTKILSVLSANCSK----DLLHELMSGLGIQEFH---LEEIFSIFDICA   1097
CTLaV         LVLNQRVQHFDLAVPKEGCVIRAISEFLKQNPTKVLSVLSANCSK----ELLHELMSGLGIQEFF---LEEIFKVFDICA   1107
CNRMV         LILNQKSHHFDLAVPKEGCVIRAVSEFLKQNPTKILSVLSANCSK----DLLHELMSGLGIQEFF---LEEIFSVFDINA   1108
                                   .                                        :            :
```

```
PAFV1         QVVTPQGKHS-------------FGP-RSGRPVILYHSAY---HW------SPIPTSRGRIPFSPTTP-----P------    982
EnFV1         TVNSSLGNFN-------------FGS-HSDTRISLCHRPG---HWFNADRPQRALINSGRRSRAPPHP-----------   1008
M.cannonballus RVHGRNNKFI-------------VGD-DRLPCFDIYHSTN---HW-SGTKPSPAPT---------PPPSSGGGPVSGYGN   1024
EnFV2         DNSTATGALVTHNRKTRAPSYVGYHRPTSTHTVRFELGDGT-AHW------SPRSTDTPNHKLKSIPL-----PT-----    965
BVF           -------------------F--FRPSPTPTPSDATPTPA----ARAPSSVSNENAQRPRHPLPP-----VPTGGASA    982
WWVV          RVHSDHGQLL-------------FGPHDSAQRINITHTVGP-HPT------SPPALDSSAVPRPNPPE--PSHP-----    874
UTyV1         LLHTPFGVHN-------------YGIFSSSKSVSIYYQPGPPAHF------SSSPRVVASYPRSNPSS----EP------    836
DiaYMV        VIHTGHTTLP-------------YSCVGASTRVQITFHPGPPKHF------SPNIRLSASAPGSNPSK----SP------    839
NeRNV         VIHTGQTTLP-------------YGCVGATNRIQITFHPGPPKHF------SPNIRLSASAPGSNPSK----SP------    838
GVA           AIHGERGYAGTP-----------GSYRRLHLKITSG---HV------EPFEGVTSKGGFREAML---LGDGVGLGR    879
GVE           SILSDEGYLHLQ-----------GSYKPLGLVIKDD---HA------TPGRYIQRQNRPAEALA---VNPGVGQME    701
CVA           HIYYGDSIIK-------------LNDDINMHAIHIGGKPG---HL------FCINQERSKIPKDSQIKVPEVGPQSFIGS    736
GYMaV         LVNFGTESIE-------------FNK-QGKLCRKFFLTDG---HI----EYIGLANFDSKLGQKKISL-----SNKNYDS   1180
CRMaV         EVSDGASSRV-------------LNK-KGSRSAKFIVDKD---HF----SFCPGTKASTNLGVFKAPSGCPMIAIEKYDE   1156
CTLaV         EVSDGEKSRI-------------LNA-NGSRSAKFTVEQD---HF----SFSPGVKASTNLGSFKAPSGGQTIPIEQYET   1166
CNRMV         EVSDGEKTRN-------------LNP-KGSRYAKFLAEQG---HF----SYCPGTKVSSNLGSFKAPSGGQMIPDDQYEN   1167
```

```
PAFV1         ----------GSHCQLPLADKLHFLCRQYSQQYRISVKRADATLKMFVSGSLGVLIRR----------------KPEVLE   1036
EnFV1         ---------------SFSKSLFELFWSYSRPFRPNISRARVSYAAMESRSLGALIPN----------------TPSHMA   1056
M.cannonballus G----QNKNKGARPRSELAERLWISCRSWGLTYRPSTKRAKVALRAIRDFQFGELLRK---------------DPQWIA   1084
EnFV2         -------------TAARFAKA----ITLPAQQLSIDKNRALSYLEAIRQ-NEGIANRM---------------KGFDYE   1011
BVF           PSNDKLNS-----AAKQYAR----LVAGKTRQMTVAQPRSKAYLRDLKRGNIGTI-------------------PTAEAE   1034
WWVV          -----LVR-----AMKSFKVSDHYLPFLEAHRHPTSISHAKNLVSNMKNGFDGVLSLIDVSSNPRPGHT-----PRERII    939
UTyV1         -----LVR-----AALRFKHSGFFLPFTQAHPHTSSLVHAKNLISNMKNGFDGILSTISTSS--RSGPT-----PRERIF    899
DiaYMV        -----LVR-----AALRFQLNGDFLPFMNSHKHKVSVPHAKNLVSNMKNGFDGITSQLSESS----GRS-----PKQKLL    900
NeRNV         -----LVR-----AALRFQLNGDFLPFLGSHKHKVSVPHAKNLISNMKNGFDGITSQLTESV----GRS-----PKQKLL    899
GVA           F----------------------------RVDKAKADRLAQSFYNGNTGVLLGKY-----NKGKMH----------    912
GVE           I----------------------------EVVSENARALQASFEKGFTGLILNDH--------------RGKWAA    734
CVA           I----FSKTYGSGSSAPIHLG------------QIDITKALVLVSAFESMNLGVRVDRKA---ILEGQLISNGFLAFLKR    797
GYMaV         I----FASM-ADGNVISY--------------VPSLDRAKLLEKSFKDGLTGKILATS-----FGG-------PEMIIK   1228
CRMaV         F----L--R-SSANVVPF--------------TPSLPLAKKLADSFLSGQTGVINSKI-----VAG-------QYDWLA   1202
CTLaV         F----L--R-GNANVIPF--------------TPSLIAAKKLANSFLSGQTGVINSKI-----ISG-------QYDWLA   1212
CNRMV         F----L--R-GNANVIPF--------------TPSLPAAKKLASSLLSGQTGVINSKI-----ISG-------QYDWLA   1213
                                                :         :        *
```

```
PAFV1         LWRSLCDAP-PTATRRSIKTCGILGEPGTGKTHGFSV-VVKDHWRAHP-SADDITIIVGTADLRSKLCSDLSPPQ-----   1108
EnFV1         NLRSLCKSA-PTRTTPTVEVMGRMGDPGTGKSYEVLM-LIKAWMHSPL-PNSKCTIVVNSNDLRHDIIKKLELTD-----   1128
M.cannonballus KAKSVVDCP-PRASVRPIRIAGILGEGGTGKSYDMLE-IIKSRYEHYP-EDHSWCIVVATTDLRSKIVKGLGLAD-----   1156
EnFV2         RWKATVKG----SHRPDCAVTGYLGLPGSAKSAKIKD-IARQILAENPDKVPTFKIVARNNSLANELAKDLALPA-----   1081
BVF           QLDAYIDSWAVNGFNRSVHVIYFCGLPGTGKSRRCMK-MTERLLKENKYLAQTVRIVTPTDNLRNSVARQLKPPT-----   1108
WWVV          ELDRHLDT----NPEKTVPLVHIAGFAGCGKTHPIQK-LLQSKLFKD------FRVSCPTTELRNEWKSSMSLPG-----   1003
UTyV1         TLDSLIDV----APAKTIPVIHIAGFAGCGKTHPIQK-LLQNRAFHD------YRISTPTNELRAEWKKDMKPTE-----    963
DiaYMV        ELDATIDV----AFPREVDVIHIAGFPGCGKSHPVQK-LLQTKAFRH------FRLSVPTNELRTEWKSDMSLPE-----    964
NeRNV         ELDATIDV----VFPRDVDVIHIAGFPGCGKSHPIQK-LLQTKAFKH------FRLSVPTNELRTEWKTDMALPE-----    963
GVA           ----------TGEIEEPKEVLTAFGFAGSGKSHWCQT-ILKHCSVEK------VLVISPRKVLKDDWVAKI---------    966
GVE           KIPQGARA--------QLTVSMCLGFAGSGKTSSITQ-MLKLGHQMA------VAVVSPRKNLCEEWKKEL---------    790
CVA           KNNEGHKV----IKIQSLPVYPFIGFAGSGKSFGLTEKLINGDCSQN------FMFTAPRKKIIGQIHERIDSRQYDDKL    867
GYMaV         KDVEAI----------KKKVWFMVGTFGSGKSFSVKN-KIKERTDIN------FLIISPRRKLADVFKEELGLKKEWKKA   1291
CRMaV         NTNKLCF--------EERRVGAIVGTFGSGKSHNVIE-LIRHNLGYQ------NLIISPRRNLKDQFINMLDLVNARSKG   1267
CTLaV         DTNKLCF--------DERKIGAIVGTFGSGKSHNVIE-LLRHNLGYQ------NLIISPRRSLKEQFINMLDLVQARSKG   1277
CNRMV         DSNKLCF--------DERKVGTIVGTFGSGKSHNVIE-LLRHNLGYQ------NLVISPRRSLKEQFISMLDLVKARSKG   1278
                      * *.*:          .                    .        :          :
```

```
PAFV1         -GRGYR-IKTWENALLQPL---APTILLDDAGLIPP-ILDLLLL----------------------------LNPQTT   1152
```

```
EnFV1           -DQKFM-VKTWERAIVEPC---ESIVLMDDIGDIPP-ITDLFLL------------------------------AHPNVR  1172
M.cannonballus  -EDTWR-VKTWEKAHIEPL---PSTVLFDDCGLIPT--FDLNMM------------------------------LQLGPT  1199
EnFV2           -KLNFV-ASTYETALMIHP---GDTLILDDVGLFPAGYTDLYLA-----------------------------LHPNIR  1126
BVF             -ESSYS-VQTYETPIRQPSL--GPVLIVDEYGKMPAGWLETVLF-----------------------------LNPRVQ  1154
WWVV            -NQSWR-FCTWESSLLKS----SKILVIDEIYKMPRGYLDLSIL-----------------------------SDPAVE  1047
UTyV1           -ENVWR-FSTWESSLLKH----STILVIDEIYKLPRGYLDLSIL-----------------------------ADPSLK  1007
DiaYMV          -SEIWR-LCTWETSLFKS----SSILVIDEIYKLPRGYLDLALL-----------------------------ADPSTS  1008
NeRNV           -SEVWR-LCTWETSLFKS----SNVLVIDEIYKLPRGYLDLALL-----------------------------ADPSTS  1007
GVA             -SKKHK-VVTFEVAFMDDYG--CKDIVIDEIGLLPPGYIDLVIA-----------------------------AH-QPR  1011
GVE             -LDTDVCVYTYESFLLNHKH-KLDLLILDEIPLFPPGYIDLVHF-----------------------------IK-SVE   837
CVA             KISRKKNFSTFENTLLSLVN--KPLVIMDECSLNPPGFIDLVLIKSLDSIIRKSNKDFDHFFSSSVLSEGIIANVASPIA   945
GYMaV           KKSNFE-VVTFETAIKHKSFKKSEIIILDELQLFPPGYLDLLLL-----------------------------IS-KSE  1339
CRMaV           KKTSTD-VVTFEVALKKNGLLKKARIFIDEAQLLPPGYLDLICL-----------------------------IAGSDS  1316
CTLaV           KKASTE-VATFEVALKKTGMLKKVRIFIDETQLLPPGYLDLVCL-----------------------------IAGPDA  1326
CNRMV           KKASTE-VVTFEVALKKSGLLKRVIFLDEVQLLPPGYLDLICL-----------------------------IAGPTA  1327
                      *:*              :..*:     *     :

PAFV1           KVFFTGDPCQGQYLLPADAPNQALSRSTLDLL--QPSA----------------------------------------  1188
EnFV1           TLCFTGDPAQVVRELSTDVPNHKKCKNGIDLL--STHA----------------------------------------  1208
M.cannonballus  DVFFTGEPGQSGMELPFGVPGWGDSRTMVECV--KGAA----------------------------------------  1235
EnFV2           LVVFTGDTAQGTHRPTRRPSPLDHLTSEITHL--APYA----------------------------------------  1162
BVF             MVIFTADPTQGIFRTRIPDAYCCRIPSSTEAI--ASYA----------------------------------------  1190
WWVV            LVIILGDPLQGEYHSLSKDSSNHRLPSETIRL--VDYI----------------------------------------  1083
UTyV1           LVIILGDPLQGEYHSTSSHSSNHHLPSELSRL--RAYI----------------------------------------  1043
DiaYMV          LVILLGDPLQGEYHSSHPSSSNNRLESETSRL--SKYI----------------------------------------  1044
NeRNV           LVILLGDPLQGEYHSSHPSSSNSRLESETSRL--SKYI----------------------------------------  1043
GVA             TLVLLGDPLQSTYHSKRDNVVLEASQEDVFNR--VQGR----------------------------------------  1047
GVE             HILLLGDPLQTSYHADGDALTLSGVEGDIFRR--LQAKKSGICPCGMHFKPSRYMGPNVEFDFDEADKLKGREAIFFSRG   915
CVA             CIAVTGDTLQSSFYSESCGKLMQ-YKNDIKTLCALSHTR---------------------------------------   983
GYMaV           NILVLGDPAQSSYDSEEDRAIFEGINNDLVNL--LSNQK----------------------------------------  1376
CRMaV           SILVMGDPAQSSYDSAEDRMMFAGDKGCLDRL--LEGKK----------------------------------------  1353
CTLaV           SILVMGDPAQSSYDSADDRMAFIGDRGCLDVL--LDNKR----------------------------------------  1363
CNRMV           SILVMGDPAQSSYDSADDRMAFIGDKGCLDTL--LEGKK----------------------------------------  1364
                  :  .  .:.  *

PAFV1           ------------------------------------------------------------------------------  1188
EnFV1           ------------------------------------------------------------------------------  1208
M.cannonballus  ------------------------------------------------------------------------------  1235
EnFV2           ------------------------------------------------------------------------------  1162
BVF             ------------------------------------------------------------------------------  1190
WWVV            ------------------------------------------------------------------------------  1083
UTyV1           ------------------------------------------------------------------------------  1043
DiaYMV          ------------------------------------------------------------------------------  1044
NeRNV           ------------------------------------------------------------------------------  1043
GVA             ------------------------------------------------------------------------------  1047
GVE             GEGYKYNGGDHKSSGWVGELDQIIDACNINADSFDHCLVQRYIPGGSIKPHADNEPIYPVNNPILTVQLTGSCTFTLSCR   995
CVA             ------------------------------------------------------------------------------   983
GYMaV           ------------------------------------------------------------------------------  1376
CRMaV           ------------------------------------------------------------------------------  1353
CTLaV           ------------------------------------------------------------------------------  1363
CNRMV           ------------------------------------------------------------------------------  1364


PAFV1           ------------------------------------------------------CEYKRLTHRLAADV-AGRLGAVTSSA  1213
EnFV1           ------------------------------------------------------GCYFREGHRLPRDV-ARAMGLTTRST  1233
M.cannonballus  ------------------------------------------------------CEFRRKNRRLGDGT-SSKLGVQSINP  1260
EnFV2           ------------------------------------------------------VKYIFHSTTLADQV-AAQLGLPREVG  1187
BVF             ------------------------------------------------------HEYRRVSDRPAPGV-ARALGLPTTSR  1215
WWVV            ------------------------------------------------------DAYCWWSYRIPQVI-ARLFSITSFNY  1108
UTyV1           ------------------------------------------------------DCYCWWTYRLPKRI-ADLFSINTFSS  1068
DiaYMV          ------------------------------------------------------DCYCWWTYRCPRAV-ADLFGVKTFNT  1069
NeRNV           ------------------------------------------------------DCYCWWTYRCPKAV-ADLFGVKTFNP  1068
GVA             ------------------------------------------------------LPYLCYSHRLPRNC--NLFEIECMGA  1071
GVE             KGDTSLHLEGAQFFLMPNGRQKGHKHSVVAHDHRVSLTFRSTRPLELMSGKIQAVPYLFLTNRLSSKQ--KIMEVPSYGV  1073
CVA             ------------------------------------------------------LPYLFGSKRFGYFT--GFLKLGYYNQ  1007
GYMaV           ------------------------------------------------------YNYLIQSKRFRNRFIDGRLPCKFDDL  1402
CRMaV           ------------------------------------------------------YVYYLSESKRFRNPMFVGRLPCTFDSS  1379
CTLaV           ------------------------------------------------------YVYYLSESKRFRNPMFLGRLPCTFDQS  1389
CNRMV           ------------------------------------------------------YVYYLSESKRFRNPMFLGRLPCTFDSS  1390
                                                                      :                    :

PAFV1           A----------------------AG--SIRFVGR--PFGPVVASTVGSVQSLTAL---GTQAFTPATCQGQTFRCAYT  1262
```

```
EnFV1          Q----------------------PG--SLKAAYT--RRGDHIMLSNGAASLAREL---GIQAHSAGSVQGLSFEGPCV  1282
M.cannonballus R-----KAFPLFSHTYRYLLALPIDPGDFYFNFNRCPPAHATVIASTNGQVKSYEAV---NRHAFTVAGSQGQTFPVT-E  1331
EnFV2          K----------YD-----------NRG--SIKLNQSFDPTIPTIAPDTETAKAIP-------GAHTYESCQGLRIHKPYN  1237
BVF            R----------------------PG--QIETSTELNVDWITVRPEGERAAWHNTW---GGKVYTYATCQGETFHRAYQ  1266
WWVV           D----------------------TG--IIGSIPTPVDCHPILTNSHAAALTFNNL---GYRACTISSSQGITLSDPAV  1159
UTyV1          H----------------------QG--FIRSISSHPPNSKNLVNAMNTAASLQQM---GHHAITISSSQGITFDEPHT  1119
DiaYMV         E----------------------QG--FIRGELSHPQGLPNLVNSIATATTMQNL---GHHALTISSSQGMTYSSPTT  1120
NeRNV          E----------------------RG--FIRGELSHPQGLPNLVNSIATATTMQNL---GHHALTISSSQGMTYSTPTT  1119
GVA            E----CEKRVVYRSNR---L-----------------KDEPTICATRAMKEE------KGSGWYTVSETQGLSFKSC-L  1119
GVE            S---D------------F-----------EVKEIKKLNKETLTICFSRATVEEER----DNLTICTVGQAQGLSRDVV-Q  1122
CVA            M---ESKAFTIDNMET------------LQKAIGTSMDKFGVLVTSRADKSDFELD---FPNVCTINEAQGSTFNSV-I  1067
GYMaV          S-NFPSEEYFVYEDMR--------------KESDLILTNDVILCSSFDEKKSISYILGRRREVLTFGESTGLTFRKV-C  1465
CRMaV          RLTLEKEEYAVFDSFK------------AFKADYLSPKIKTFLVSSFTEKTVVKANMGRNVSIFTFGESTGMNFDYV-C  1445
CTLaV          RMTLEKEEYAVFSSFK------------DFKNDYLSPKIKTFLVSSFTEKTVVKANMGRNVLVYTFGKSTGMNFDYV-C  1455
CNRMV          RVTLEKEEYAVFDNFK------------AFKGDYLAPKMKTFLVSSFTEKTVVKANMGKNVSVFTFGESTGMNFDYV-C  1456
                                                                       :                :      *

PAFV1          ILADKYLAHATDAMVYTALTRGTGDVLVMDQGLFGKKAMAP-RSSILR---ALVAASVTGSW----------------TD  1322
EnFV1          INLNKMAKHMNDRTAYTILTRSKGDCLISDDNGSLVSHLNN-RCPILT---GAVSYCTTGDN----------------QT  1342
M.cannonballus LFIDSYMMKADNKAIYTALYRSSGDMYISLPHNYQK--LQP-ASTILK---GLLHAAATGDW----------------DR  1389
EnFV2          LLISKKSLYFSDGDIFTALTRGTTHVHIVCPELTVADRLAA-TG-ILA---ALIRPHTSDDTHRGIRYTIVDHISTPTNE  1312
BVF            MVINGNTRFWDDRDLFAALTRGSGTLRLIYETQSSKPLPRS-TSKLYN---ALLDFRTDP----------------HK  1324
WWVV           IVLDNYTRYLSASNGLVALTRSRTGVQFMGPTTYIGGTNGS-SAMFSD---AVNRTPINLDQ-----------YFP---  1220
UTyV1          ILLDRHTNLLSPNNCLVALTRSRHGVSFLGNLHLASASFGT-NYMFSR---ALAGLEIDMIG-----------CFP---  1180
DiaYMV         ILLDRHSTLLSPQNCFVALTRSRKGIIFIGNMYQASGYFGT-SYMFTQ---ALTGSPVDLMC-----------AFP---  1181
NeRNV          ILLDRHSTLLSPQNCFVALTRSRKGVIFIGNMYQASGHFGT-SYMFTQ---ALTGSPVDLMC-----------AFP---  1180
GVA            IYLDEHWAKKEDEDVMVALTRSRGEIGIHVTPALKKNLINNAKSTLLK---KVLKGETYR-----------------RS  1178
GVE            IMFDHGSLKCADETIIIALTRARKAVHLFYKIG--KTDLKNCSSPILR---AFISNGKIP-----------------EK  1179
CVA            LIVTRDFFSNPIESIIVAITRHQKNLLIYFPAAIQGEM------DFLSRRFPIHSNVVLK-----------------NF  1123
GYMaV          IVLTQNFRMTDEKRILVALSRASYQTNFVNNTGLPFKD-------FIL---SMPNSVIFK-----------------YC  1517
CRMaV          VLLTQDSMLVDERRWVVALSRAKINISFINLSGLTLPE-------FCT---QMMGGVVHK-----------------FF  1497
CTLaV          VLLTQDSMLVDERRWVVALSRAKINMSFVNLSGLSLPE-------FCT---QMVGGVVHK-----------------FF  1507
CNRMV          VLLTQDSMLVDERRWVVALSRAKMNISFINLSGLSLSE-------FCT---QMVGGVVHK-----------------FF  1508
                :                  : *      .               :

PAFV1          FHSAVRDHKLRHTPPQLSD-----PIRP--------------------PGV----APPPIASLLNY--PLAGTG--FGD  1368
EnFV1          LLRAVRDHRIKYTPPQLAD-----PLRM--------------------PGK----APRSVEQLSPP--KLSGTG--FGD  1388
M.cannonballus LRAAVRDHKIRSCPPQLRD-----PLRN--------------------PGV----KPPPLKDLLFK--PLAGNGSVFND  1437
EnFV2          LTRAVINHISRHAPSHLLNLSP-PPTQPVHVVVDPNTPIIGIAAWTRLTAPATQLPHRPPPFPARPAA--P----------  1379
BVF            LPLAIEEHVNRYIPSHLRNVNQAVPRRN--------------------AAS----NPDALPSQRIL--P---GG-----  1369
WWVV           -------SLFHRLP-LIHS-----PIQS--------------RLLRL--TGA----TPSSSPTFRS--------------  1253
UTyV1          -------QIFPHLP-RMFQ-----PIST--------------RTTRL--VAG----FPAATSKLSTTFKP----------  1217
DiaYMV         --------LYHTLP-LIFD-----PIKS--------------RRQRL--VAG----DPLPISTSESA------------  1214
NeRNV          --------LYHTLP-LIFD-----PIRS--------------RRHRL--VAG----DHPTPSVDSTA------------  1213
GVA            EIVAM---VRKHIP----D--------------------------------------------------------------  1190
GVE            LLV---DKVRGKLG----D--------------------------------------------------------------  1191
CVA            SVLD--NLIKDKLN----------PF--------------------------------------------------------  1137
GYMaV          SATCNVSDLLPLLP---GE-----PN--------------------------------------------------------  1535
CRMaV          TSTATFNDLRELLP---GD-----PI--------------------------------------------------------  1515
CTLaV          TGTATFNDLRSLLP---GD-----PI--------------------------------------------------------  1525
CNRMV          TGTATFNDLRELLP---GE-----PI--------------------------------------------------------  1526


PAFV1          TIFAAIYHSAVYLI----DTVAPTSWYISKHDYNPDVMDSAL--ERVPFLDLHYNTL---PPANRPDDF--QFQGHQFID  1437
EnFV1          VISKAILGMSTFTP----AMATVESRYTPSSNFHPDALPNAI--SEVPSLALTQVPM---AATSFLEDMPRTRRAAASID  1459
M.cannonballus MVCRQLTHYMKPKPETMRQWVYDNGWEEGRSRRS--IVDRLLYFSKLPWLWLTHQLG-------YTDPKPDVEPGPVPKG  1508
EnFV2          ------------RPNRN-QINQAARAQFATRTPRPLPRGGRLHSEKLPHLHHLKADI--FAYPQIEDEDPAKEADQP-T  1442
BVF            -------QSTTAEPXQ----SQASALPTGGEAIDVPALDRQDP-QRFPQLKVLPHLCGYIGDYSEASYPEPEHPKTQMLD  1437
WWVV           --------TNFHLPPHV-PLSYSSDFVALNPSLDPKVLDSRLETHFLPPSRLPLHF--------DLESPSLPPPAPSK-P  1315
UTyV1          NIHMESFKSTVSLPPHF-QIDSAHDFVLTNPAVVSIDLEDRLPTNHLPPTRLPLHF--------DLPSSTPLEPNPSTVT  1288
DiaYMV         ------IRNYGRLPPHI-PTDYAKDCFVSSQVVFSEGEDRTLPTLHLPPSRLPLHL--------LTEPAAPSEVLLSE-T  1278
NeRNV          ------LRRYGHLPPHI-PVDYSKDCFVSHKVVYRDGEDRSLTTLHLPPARLPLHL--------QVEPSIPSEVETSS-T  1277
GVA            ------------------TTVLFEESRLA-ETVD---YEARL--AGDPYLKSLLAL----YDEIEMEEVE-IEEPVAL--  1239
GVE            -----------------CRLLTENVFIGADSAT---IGDHL--AGDPGLKAMLLI----LEAEEMEPEI-FEEETVP--  1241
CVA            -------------------QLIQEDPFGHD------FEVKL--EGDPFLKSELSL----VNEIKLQQ---IEENSIESK  1182
GYMaV          -------------------FISTRVRIGHDEVD---REARM--IGDPWLKTMLFL----GQRGSYEENVPFEEPVLE--  1584
CRMaV          -------------------FSKRFQRLGKDEVD---REARL--LGDPWLKAKVFL----GQREEKIESIHVNDEGLK-D  1565
CTLaV          -------------------FSKKFQRLGSDEVD---REARL--SGDPWLKTKVFL----GQREVRPVEEPISVENLK-D  1575
CNRMV          -------------------FSKKFQRLGNDEVD---REARL--SGDPWLKTKIFL----GQRSVPVLHEKVEDEGLK-E  1576
                                                          *


PAFV1          SSSDLPFSLIDDFQLRSPRELLLNPSFTSEIIESMFPSPSVY-GREVYFR--GQHTVQIDT-------------EDYATA  1501
```

```
EnFV1          HSNQLPFVLPGDLTAFSPREMFRDLSVCSQVVEMVWTNPAIY-KREFRFN--GLVTQQVAE-------------DDYATA 1523
M.cannonballus DEFSPYFTLTGPSLGYSTNDMVDNPEVRADVVETLWKG-EYF-NREVKVN--GEMTHQVDD-------------RDAATS 1571
EnFV2          TTTPSPTIRETRW-----------RDNATRLHGFLMPDSEVRAHRELWSDRVHSWSSQIDG------------LPYSIES 1499
BVF            ATFPTYLARNRGW-----------NDPALTDAVTANYKAPEH---RDIWVEAVGETTRQVHG-------------SGDDRD 1491
WWVV           EPTLPKATACYPG------------ENFESLAAFFLPAHDPA-QREILYR--DQSSNQFPW--FDRPFSLS---CQPSSL 1375
UTyV1          SLFPTPHSLGLFG------------ETFENLSAFFLPAHDPS-IKEVTIR--DQSSNQFPF--FDRPFHLS---CQPSSL 1348
DiaYMV         EPSKSPITLALLG------------ESFEELAAHFLPAHDPE-LKEIIFA--DQSSSQFPF--LDVPFELS---CQPSSL 1338
NeRNV          PPCRTPITMALHG------------ESFEELSAHFLPAHDPE-VKEIAFA--DQRSSQFPF--MDQPFELS---CQPSSL 1337
GVA            EPTKTHLALSTMT--------------NELAPFDLKAKEY---REQYTE--VGRTEQIDEEGYQG---EAGN-PMTHKA 1295
GVE            ETIRTHLGVTTFA--------------NEQFAFGLKAKEE---REHHIHG-TGFSTQIRDNIASE---FHPG-PSAPSS 1298
CVA            ENLKTHLPISYSG--------------LWNLEISEMRARED---REFKKFG-VGWSKQFKDEPNQKDQVEDNC-AMLPEA 1243
GYMaV          MRTLTHHPIVGDN------------VVRARISELFRSKEE---REFRID--DNVSEQFRDSYNVKDFFKSSNQCELFEA 1646
CRMaV          IKVKVHCPIGSIG------------STLADIQAGVRVKEA---REFRID--NLVTEQFSEVHKGKGKVLTAA-PDNFEA 1626
CTLaV          IKIKVHCPVGSMG------------ATFAEVQSKLKVKEA---REHRID--TIVTEQFAEVHKGRGKILTAA-PDNFEA 1636
CNRMV          IKVKVHCPVGSMG------------GLFSEMQSKLKAKEA---REHRID--SIVTEQFAEVHKGRGKIMTAA-PDNFES 1637
                                                      ::          : *.
```

```
PAFV1          VFLRHR-RQDAATEAWTFAERY-----VPPKRPSTTYLGGGLALFEAYVSTYS---PQVPAFNQAVYEQCEE-EDQ-AAL 1570
EnFV1          VFLRHR-RGDSALEKWTFKERH-----VKPSRPTTTYLGGGRALFNAWMHVYN---PKPIPFSEDMYERAVE-ESQ-AAL 1592
M.cannonballus IFLRHR-RNNSATEAWTFKERM-----VKPRPPKLTYLGGAEALKSHFDMVYD---PKYPPFNETIWEEYQD-EDT-DNF 1640
EnFV2          IFLKHR-TKDAATAQATLEKRVRRTSQRKKEMAFIASRHVGAALYNAFCEEYG---PQDEEWDPQLYEECRV-ENE-TVH 1573
BVF            VFLEHR-GDDKATANITYAKRLRFSQRAANERSIAATKVAGSELFEAFTRAIP---LPQETFNDSLLEECRT-ENDTVHL 1566
WWVV           ISAKHSPASDPTLLPASINKRLRFRPSEAPHAITSDDVILGLQLFHSLCRAYNRQPSQSIPFNPELFADCIS-LNEYAQL 1454
UTyV1          LAPVHSPSSDPTLLPFSIKKRLRFRLSDSPYQFTANDSLLGHHLFESLCRAYQRNPSHQVPFNPALFAECIN-INEYAQL 1427
DiaYMV         LAASHRPASDPTLLVSSIKKRLRFRPSDCPYFISSNDILLGQHLFNSLCRAFNRSPLEVIPFDPVLFAECIA-LNEYAQL 1417
NeRNV          LAASHRPASDPTLLVSSIKKRLRFASESPYFITPNDMILGHHLFDSLCRAYGRDPSVVIPFDPALFAECIA-LNEYAQL 1416
GVA            LYLRHT-SDDTATFMMSVKKRLRFRNYEANRRKYRACHGIGHQMFSVFKDTYQ---LKEIDSLPEL-ERCEM-EFM-KKR 1368
GVE            IYLHHT-AEDDVLFILSIKKRLRFADFEKNCASFERKKKLGESIFTEFLKRAD---FMNFTYPPQVDETSMELDFT-MKR 1373
CVA            VFPRHF-ANDDLTFWSAVKKRLVFKNPLSNAHDFEKAKTFGKELLNIFLRKV----PLMPNFDQRMYDESVS-EFE-EKK 1316
GYMaV          IYPRHK-GTDVVTFLMAVRKRLSFSDPAVNESKFNSAKTFGLLMFEHFVKYI----PLKSNRDEEMFETARS-DFE-RKK 1719
CRMaV          IYPRHK-AGDTATFVMAARKRLKFSFPARERQKYMAAIPYGVSMLQVFLKRI----KLQSNFDHRLFEEARA-DFE-EKK 1699
CTLaV          IYPRHK-AGDTATFVMAARKRLKFSLPAKEKQKFMSAIPYGDTMLKVFLNKV----RLKPNFDHRLFEEARN-DFE-EKK 1709
CNRMV          IYPRHK-SGDTATFLMAARKRLKFSFPANERQKYLAAIPYGDTMLKVFLEKV----KLKPAFDHLLFDEARN-DFE-EKK 1710
                :     *    :       : :*                 . :                          .          :
```

```
PAFV1          LEKGPKCLINISYRNDPSLDPHKAETFLKSQDVTKLGTQFRDAKKGQMITGFAAAVNSRFGPLSRYLYRATRTSLPPEIL 1650
EnFV1          TDKGIKSLKNISYRNDPSLDPAKAETFLKSQDITKLGTAFRQAKKGQMITGFITHVNTTFAPMARYLYSAFRTSLPPSIL 1672
M.cannonballus LDKGEKALKNIAYRSDPSMDPSKAEIFLKAQ------------------SGFCTLINARFGGLSRYIYRAMRTALPAHIF 1702
EnFV2          LSKPLPTLTANASRSDPSWPLHRIELFMKNQNVTKLGKLNADARAGQIIANFRAQVILTLGPLGRYLAKKILSKLPDHTY 1653
BVF            TSKPLATLINNAERSDPSWSLNMIKLFIKGQTVKKLEKMGSDATAGQSIASFRAEVLLAWGPYARYIDRRIRALLPPHVY 1646
WWVV           SSKTQATIVANASRSDPDWRHTTVKIFAKAQHKVNDGSIFGPWKACQTLALMHDFVILVLGPVKKYQRIFDNSDRPEHIY 1534
UTyV1          SSKTRATIVANASRSDPDWRHTSVRIFAKAQHKVNDGSIFSSWKACQTLALMHDFVILTLGPVKKYQRLFDSRDRPPHLY 1507
DiaYMV         SSKTKATIVANASRSDPDWRFTAVRIFAKAQHKVNDGSIFGSWKACQTLALMHDYVIMTLGPVKKYQRILDHQDRPSHIY 1497
NeRNV          SSKTKATIVASASRSDPDWRFTAVRIFAKAQHKVNDGSIFGSWKACQTLALMHDYVIMTLGPVKKYQRILDHNDRPRHIY 1496
GVA            IEKSTGLIEKHAGRSDPDWPSNYLKIFLKQQTCTKMEKRGVDAKAGQTIACFAHSVLCRFGPILRQTEKALRELLPENVM 1448
GVE            IQKSARILEAHSYRSDADWPSNYLKIFIKNQDCTKMEKRGSDAKAGQTIACFSHAVLCKFGPILRKTEAQLRKILPPHVM 1453
CVA            ISKNAAMIGAHHDRSTTDWPTNEIFLFIKSQLCTKKEKMFCDAKAGQTLACFSHLILCKFAPLNRYIEKKVTQSLPGNFY 1396
GYMaV          LEKNIATIENHSGRSSADWDIREAFVFMKSQLCTKFEKRFVDAKAGQTLACFSHIVLCRFAPWIRYIEKKVFEVLPPNFY 1799
CRMaV          LQKSMATLENHSGRSDPDWSVEKALIFMKSQLCTKFDNRFRNAKAGQTLACFHHDVLCRLAPYIRYIEKKVFKALPSNLY 1779
CTLaV          LQKSMATLENHSGRSDPDWEIEKALIFMKSQLCTKFDNRFRDAKAGQTLACFHHNVLCRLAPYIRYIEKKVFDALPRNLY 1789
CNRMV          LGKSMATLENHSGRSDPDWEVEKALIFMKSQLCTKFDNRYRDAKAGQTLACFHHNVLCRLAPYVRYIEKKVFKALPRNLY 1790
                *    :    *. ..      * * *                 ::  :   .  :              *
```

```
PAFV1          LLNGVTLEEQERWFADHWDFSSGCYEDDYTGFDGTQNEDFLAYQTLYMRAFGVPEPLIEDYLSWVTHLSCMLGPLGIMIA 1730
EnFV1          LLNGVTLDDQERWFNSHWDWAQPCYEDDYTGFDGTQNEDFLAFQVLLMGWYGIPTNIIETYVLWVTHLRDALGECGVWIA 1752
M.cannonballus LLNGVTLDDQEAWYQKFWNWAKKYYEDDFIAFDGTQNEEFLGFHILIMQALGIPQAVIDSYISWVTHLFGMLGDFGIFIA 1782
EnFV2          VHHGKTTAQLAEFVDRLWDFNQESMESDATAFDQSNGPDILHAETYFMRAKNVPQHLVDYYLDAKRLAYTFLGPLEFMRL 1733
BVF            IHSRRTNEDFEKFVAAHWDHTRESTDGDYTAYDASQDATFVNFETLLMRRLDFPLDIIEAYVEMKASITSHFGPLAIMRF 1726
WWVV           SHCGKTPIQLRDWCQQHLTSFTPKVANDYTAFDQSQHGESVVLEALKMKRLNIPAHLIELHVHLKTNVSTQFGPLTCMRL 1614
UTyV1          IHCGHTPQQLSAWCSTH-LNHSTYVANDYTSFDQSQHGEAVILEMLKMQRLSIPQHLINLHCHLKTRVSTQFGPLTCMRL 1586
DiaYMV         THCGKTPAQLSSWCQKF-SLDGPSLCNDYTSFDQSQHGEAVVLECLKMRRCSIPDNLIQLHLHLKTNISTQFGPLTCMRL 1576
NeRNV          THCGKTPSQLASWCQEF-ACEGLSLCNDYTSFDQSQHGEAVVLECLKMRRCSIPDSLIQLHLHLKTNISTQFGPLTCMRL 1575
GVA            IYSQKNYMDLDKWAKTW-VESMMGTDSDYEAFDRSQDEKVLDLEVEVLRFFLWPEELIKEYEELKLMMGCALGDLAVMRF 1527
GVE            IFSQKNYEDLDKWSKDY-FNDHSGTDSDYEAFDRSQDGAILAFEICLLRHFLWPEELIEEYKTLKLMMGCQLGDLAVMRF 1532
CVA            IHQKKNFDELERWVKSY-DFSGVCTESDYEAYDASQDSYTLAFEYELLRYLGVSNSLIEDYLYLKMHLNCKLGNLAIMRF 1475
GYMaV          IHSGKNFDELKEWVLRS-DFSGECTESDYEAFDASQDATILSFEVEIMKYLNIPHDVIEDYKFIKFNLFSKLGIFEIMRF 1878
CRMaV          IHSARNFDDLRDWVIKN-NFTGVCTESDYEAFDSSQDVNILAFEVSLMEYLRLPRDLIEDYKYLKFHTHSKLGQFAVMRF 1858
CTLaV          IHSGKNFDDLRDWVINS-NFSGMCTESDYEAFDSSQDANILAFEVSLMNYLNLPRDLIEDYKHLKFNTHSKLGQFAVMRF 1868
CNRMV          IHSGKNFDDLQDWVIQN-NFVGTCTESDYEAFDSSQDANILAFEVSLMNYLNLPRDLIEDYKHLKFNTHSKLGRFAVMRF 1869
                .  :  :             .* .:* ::     :   :     .  :::.:         :*
```

```
PAFV1          SGFKPTWFFNTIDSMAYQALKHSLTPHHGPHPRVARAFSGDDTLHNERVTVRPEFGKLPHQFLLVSTGAHTDMPHFCGTL 1810
```

```
EnFV1         SGFIPTWLFNTADNMAFQALKHALTTHTWSPSDVARSFSGDDSTHNELVTVRPLFARLEHQFKLVSTGTHSDIPHFCGTV   1832
M.cannonballus SGFKPTWVFNTLDNMAFEALKHELTPA--SVAPVARSFSGDDSNHNEEVKERLAFRQLKHKFRLKSKGVHTTMPTFCGTI   1860
EnFV2         TGEVFTLLFNTLAMMMYLAFKYIIPK------NTPRAHTGDDVIMNARPKLRPSSQRLLTQIRITTVDEYTKTPSFVGWR   1807
BVF           SGEVWTYLFNTLGNIAFTYAKYEVP-------SVAQVYGGDDKSINSPITVRTGWSQLVGKFNLVEKPVVGYEPTFCGWR   1799
WWVV          TGEPGTYDDNTDYNLAVIFSQYEVG-------SCPIMVSGDDSLIDRVLPTRSDWPDVLKRLHLKFKLEHTTNPLFCGYY   1687
UTyV1         TGEPGTYDDNSDYNLAVIYSQYNVQ-------GSPILISGDDSVIAGSPSLHPSWPHVKDLLHLKFKTEVCSHPLFCGYY   1659
DiaYMV        TGEPGTYDDNTDYNLAVIYSQYQMG-------ATPCMVSGDDSVIFANPPIHPTWPAVENLLHLRFKTESTTQPLFCGYY   1649
NeRNV         TGEPGTYDDNTDYNLAVIYSQYQMG-------STACMVSGDDSVIFSSPPTHPSWAAVKDLLHLRFKTESTFQPLFCGYY   1648
GVA           SGEFGTFFFNTVCNMVFSCMRYHMDR------NTPMCFAGDDMYSPGILRVKKDYEATLEQLTLKAKVHVSEEPLFCGWR   1601
GVE           SGEFGTFFFNTMCNMAFSYLRYQLGP------YQPIAFAGDDMVAPGRLVVNESMNSVLNQLELKAKVNYSDSPLFCGWR   1606
CVA           TGEFCTFLFNTLTNMLFTFMKYDVRK------THAICFAGDDMCANVRLPENHEYSSLLKKFSLKAKVDFTRSPTFCGWN   1549
GYMaV         TGEAGTFLFNTLANICFTLMRYKIRG------DECIAFAGDDMCANTCLRVSTEFENILDRLKLKAKVDYKSQASFCGWS   1952
CRMaV         TGEAGTFLFNTLANMVFTFMRYEING------REAICFAGDDMCANKLLRKKSEFEHILDRMTLKAKVQHTTEPTFCGWR   1932
CTLaV         TGEAGTFLFNTLANMVFTFMRYETNG------RESICFAGDDMCANKLLRKKKEYEHVLDRMTLKAKMQHTTEPTFCGWR   1942
CNRMV         TGEAGTFLFNTLANMVFTFMRYETNG------RESICFAGDDMCANKLLRKKGDFEHVLNRMTLKAKVQHTTEPTFCGWR   1943
              :*    *   *:   :    ::          ***                  ::        . * *

PAFV1         NTPAGSFANPELLLLRCLYKLRVGRIFQSSLGYAEHCYRLQKNLETAIPHLTSRQIACHSLTLRLL--------------   1876
EnFV1         NLPGASFCDPTLLLTRILFRLRRGNLASCALSYSEHCARLQSKYEECSWRLTERERYCHATSLRLL--------------   1898
M.cannonballus NTPGGSFAEPELLAQRVLYRAQLGRINEAALAYAEHCSRLSKNIEGCLPYLNGYQIEMHTMTRRFL--------------   1926
EnFV2         LTPGGLIKDPRLLLARAYHAESKNKELETCVSYYYDLTVGLRNWTNILPHLPTQEVAAVHELAWYYQSLAKRYP------   1881
BVF           IVPGGIVKDPQLLFWRTRYARIRYDAALWAPGYYDELVLSLKTSDRLMDHMSPNDLAYLQALVRFYTKLSRRLP------   1873
WWVV          VGPAGCLRNPLALFCKLMIAVDDDALPDRRLSYLTEFTTGHRLGEPLWSLLPSELVKYQSACFDFFCRHCPKHEKMLLSD   1767
UTyV1         VSNFGAVRNPFALFAKLMICTDDGTVGDKILSYLSEFSVGFLLGDSLLQTIPSHLHSYQSACHDFFCRNCTPSQKILLSI   1739
DiaYMV        VGPSGCCRNPLALFAKLMITTDKGNLDDTLSSYLYEYSIGHRLGDACLSLLPSHLHSYQSACFDFFCRKASPLQKTLLSF   1729
NeRNV         VGPHGCCRNPYALFAKLMISTDRGNISDTLPSYLYEYSIGHRLGDACLDLVPDHLHSYLSACFDFFCRKASPLQKTLLSF   1728
GVA           MSPFGIIKEPNLILDRWKIALRSGNLSLCLVNYAIEASFGYRLSEHLYDV--NIDVDAQQELVREIVIKKHLLPKKIADL   1679
GVE           MSPYGIVKDPNLLLDRLEMKRAEGKLDDCIANYALEASYGYRLSDHLYDL--NIDLDAFQELIRKIVMLKHKLPPAIASL   1684
CVA           LSRYGIVKKPELIAARLAVARQKGEVNLVLDSYFLEHLYAYNKGDHLFEILSEKELEHHYNLTRFFVKNSKFLKGESKKK   1629
GYMaV         LGPYGIYKKPQLVFERFMISKEKGTLHECIDNYAIEVSYGYRMGDRVFGYMTEEEIECQNLCIRTIVLNKQMMKETALSY   2032
CRMaV         LGNFGIVKRPQLVQERILIALEKGNFHECIDNYAIEVSYAYNLGERLISIMSEKELDAHYFCVRTFLQNKKLFSSNALEF   2012
CTLaV         LGPFGIIKRPQLVQERIPIALEKGNFECIDNYAIEVSHAYNLGDRLISIMSEKELDAHYFCVRTFLQHKSLFSSNALEF   2022
CNRMV         LGSFGIVKRPQLVQERILIALEKGNFIECIDNYAIEVSYAYKLGERLISIMSEKELDAHYFCVRTFLQNKKLFRSNALDY   2023
                .    *  : :                    *  .

PAFV1         --------------------------------------------------------------------------------   1876
EnFV1         --------------------------------------------------------------------------------   1898
M.cannonballus --------------------------------------------------------------------------------   1926
EnFV2         --------------------------------------------------------------------------------   1881
BVF           --------------------------------------------------------------------------------   1873
WWVV          EPPTASLLD-----------------------------------------------------------------------   1776
UTyV1         DPIPDSKLL-----------------------------------------------------------------------   1748
DiaYMV        EEPSPSLLK-----------------------------------------------------------------------   1738
NeRNV         DEPSPSVLK-----------------------------------------------------------------------   1737
GVA           FSEDECEA------------------------------------------------------------------------   1687
GVE           FKEEEDIV------------------------------------------------------------------------   1692
CVA           FMETKEIEGGLFGECDFGNDSIFKDYINRVKNKVEIDLLNERILRINTELNQFDPRIYMMNKIGFVTSTSMFEAGHIASN   1709
GYMaV         FNGSLSRLE-----------------------------------------------------------------------   2041
CRMaV         FSESEGCL------------------------------------------------------------------------   2020
CTLaV         FSEGESCK------------------------------------------------------------------------   2030
CNRMV         FSESGSCS------------------------------------------------------------------------   2031

PAFV1         --------------------------------------------------------------------------------   1876
EnFV1         --------------------------------------------------------------------------------   1898
M.cannonballus --------------------------------------------------------------------------------   1926
EnFV2         --------------------------------------------------------------------------------   1881
BVF           --------------------------------------------------------------------------------   1873
WWVV          --------------------------------------------------------------------------------   1776
UTyV1         --------------------------------------------------------------------------------   1748
DiaYMV        --------------------------------------------------------------------------------   1738
NeRNV         --------------------------------------------------------------------------------   1737
GVA           --------------------------------------------------------------------------------   1687
GVE           --------------------------------------------------------------------------------   1692
CVA           ESNQSKGLPPQTSRPYDEVKPYMPLSLRNSYESRTKGNRLLSILRNQKRLCDLGSGSILGLRVLEGLKMFKEECRSYQSR   1789
GYMaV         --------------------------------------------------------------------------------   2041
CRMaV         --------------------------------------------------------------------------------   2020
CTLaV         --------------------------------------------------------------------------------   2030
CNRMV         --------------------------------------------------------------------------------   2031

PAFV1         -----------------------------------------------------------------RLVLRVQGFPLVGE--   1890
```

```
EnFV1          ---------------------------------------------------------------------RLQLRISGIPLVGS--  1912
M.cannonballus ---------------------------------------------------------------------QETLRRQGRGLIGW--  1940
EnFV2          ---------------------------------------------------------------------PLAQALRNTGFSST--  1895
BVF            ---------------------------------------------------------------------SLADRRRCNPLPADS-  1888
WWVV           ---------------------------------------------------------------------RVTSSPR--------  1783
UTyV1          ---------------------------------------------------------------------QLILKVR--------  1755
DiaYMV         ---------------------------------------------------------------------KLASSSA--------  1745
NeRNV          ---------------------------------------------------------------------KLASSSA--------  1744
GVA            -------------------------------------------------------------------------------------  1687
GVE            -------------------------------------------------------------------------------------  1692
CVA            SFFKESQETNQEFSLMQFVRRTFTVMQMPSIPKCSQLLKDFSLQLLYLQAALAKAMLHSSIFLMKLSWRLSKKPLLSTQC  1869
GYMaV          -------------------------------------------------------------------------------------  2041
CRMaV          -------------------------------------------------------------------------------------  2020
CTLaV          -------------------------------------------------------------------------------------  2030
CNRMV          -------------------------------------------------------------------------------------  2031


PAFV1          --------------------------FLTKLLPHFYNLSGNHGP-----F-----------------------  1909
EnFV1          --------------------------FLTKLLSHNYNLLISFA----------------------------  1929
M.cannonballus --------------------------FWTKLRPGHYDLS---------F----------------------  1954
EnFV2          -------------------------------------------------------------------------------------  1895
BVF            -------------------------------------------------------------------------------------  1888
WWVV           ----------------------WLTKNAMYLLPAKLRLAISSLSQ------------------------------  1806
UTyV1          ----------------------WCTSAFFSLLPSKARELLIAKSS------------------------------  1778
DiaYMV         ----------------------WASGPLLAQLDNDSLQSLLERSN------------------------------  1768
NeRNV          ----------------------WASSPLLAELDDQSLQALLARSN------------------------------  1767
GVA            -------------------------------------------------------------------------------------  1687
GVE            -------------------------------------------------------------------------------------  1692
CVA            FILGQSSYVLHAFSSLKSRSMAELYTLIPGFWTKMMHAKQVLVSSCKLDQPITFIGQIIQCPHMIQTCIGLLESSLNSMQ  1949
GYMaV          -------------------------------------------------------------------------------------  2041
CRMaV          -------------------------------------------------------------------------------------  2020
CTLaV          -------------------------------------------------------------------------------------  2030
CNRMV          -------------------------------------------------------------------------------------  2031


PAFV1          --------------------------LEQAQG---------------------------------------  1915
EnFV1          -------------------------------------------------------------------------------------  1929
M.cannonballus -------------------------------------------------------------------------------------  1954
EnFV2          -------------------------------------------------------------------------------------  1895
BVF            --------------------------PSVV-----------------------------------------  1892
WWVV           --------------------------VQS-----------------------------------------  1809
UTyV1          --------------------------LPS-----------------------------------------  1781
DiaYMV         --------------------------LPS-----------------------------------------  1771
NeRNV          --------------------------LPS-----------------------------------------  1770
GVA            -------------------------------------------------------------------------------------  1687
GVE            -------------------------------------------------------------------------------------  1692
CVA            SMLLTIPTYSSLTLESCTSSVIRALQRRQLPQMLGHSFKHFLGLLDYQILNPFLRMKILSILQLWHSLTSVLTRVSGKVV  2029
GYMaV          -------------------------------------------------------------------------------------  2041
CRMaV          -------------------------------------------------------------------------------------  2020
CTLaV          -------------------------------------------------------------------------------------  2030
CNRMV          -------------------------------------------------------------------------------------  2031


PAFV1          -------------------------------------------------------------------------------------  1915
EnFV1          -------------------------------------------------------------------------------------  1929
M.cannonballus -------------------------------------------------------------------------------------  1954
EnFV2          -------------------------------------------------------------------------------------  1895
BVF            --------------LSRD-----------------------------------------  1896
WWVV           --------------FPESPEVSQVESEL-----------------------------------------  1823
UTyV1          --------------FHSDPKVSQLESEL-----------------------------------------  1795
DiaYMV         --------------SHLDARVQRLESEL-----------------------------------------  1785
NeRNV          --------------SHLDARVQRLESEL-----------------------------------------  1784
GVA            ---------------HSDGDDDFLSNDV-----------------------------------------  1700
GVE            ---------------SSDEEA-----------------------------------------  1698
CVA            SLKAHHVQQEPEGIMPEARDRVLSQSQKLLVKILNNKKEICLGQIHVGLKIFSSIQNRGFQLIKNSSTDLISAILKGGTQ  2109
GYMaV          -------------------------------------------------------------------------------------  2041
CRMaV          ---------------SPERNFG-----------------------------------------  2027
CTLaV          ---------------SPDRNFG-----------------------------------------  2037
CNRMV          ---------------SPDRNFS-----------------------------------------  2038


PAFV1          -------------------------------------------------------------------------------------  1915
```

```
EnFV1          -------------------------------------------------------------------  1929
M.cannonballus -------------------------------------------------------------------  1954
EnFV2          -------------------------------------------------------------------  1895
BVF            -------------------------------------------------------------------  1896
WWVV           -------------------------------------------------------------------  1823
UTyV1          -------------------------------------------------------------------  1795
DiaYMV         -------------------------------------------------------------------  1785
NeRNV          -------------------------------------------------------------------  1784
GVA            -------------------------------------------------------------------  1700
GVE            -------------------------------------------------------------------  1698
CVA            ISSMALSIVEQNYNEIRRGLGNYIWENMIDPRDLLHLTAKPAVEASEGVAATPAITLSENQRAVKNTIRNYYLRIMFGNL  2189
GYMaV          -------------------------------------------------------------------  2041
CRMaV          -------------------------------------------------------------------  2027
CTLaV          -------------------------------------------------------------------  2037
CNRMV          -------------------------------------------------------------------  2038


PAFV1          -------------------------------------------------------------------  1915
EnFV1          -------------------------------------------------------------------  1929
M.cannonballus -------------------------------------------------------------------  1954
EnFV2          -------------------------------------------------------------------  1895
BVF            -------------------------------------------------------------------  1896
WWVV           -------------------------------------------------------------------  1823
UTyV1          -------------------------------------------------------------------  1795
DiaYMV         -------------------------------------------------------------------  1785
NeRNV          -------------------------------------------------------------------  1784
GVA            -------------------------------------------------------------------  1700
GVE            -------------------------------------------------------------------  1698
CVA            AVMGTSEQTDYPGEHLAIPRPVIENQEALTAHLPAGMSLLTFATNVKAWGVVGAEGKFAGLTFRQLCEPFAEQAYNFFRE  2269
GYMaV          -------------------------------------------------------------------  2041
CRMaV          -------------------------------------------------------------------  2027
CTLaV          -------------------------------------------------------------------  2037
CNRMV          -------------------------------------------------------------------  2038


PAFV1          -------------------------------------------------------------------  1915
EnFV1          -------------------------------------------------------------------  1929
M.cannonballus -------------------------------------------------------------------  1954
EnFV2          ----------------------------------------------------------SLRF--  1899
BVF            -------------------------------------------------------------------  1896
WWVV           -----------------------------------------------------------LHYLQ  1828
UTyV1          -----------------------------------------------------------LPFLN  1800
DiaYMV         -----------------------------------------------------------LHSFQ  1790
NeRNV          -----------------------------------------------------------LHSFQ  1789
GVA            -----------------------------------------------------------AGMYRIE  1707
GVE            -------------------------------------------------------------------  1698
CVA            NHGAVSFIYLKNPGAYFNCPAVVFDFNKGLPLTIIKIGKNANAISACNQRLFNREGKKAVFAAQGEVNLSFDA  2342
GYMaV          -------------------------------------------------------------------  2041
CRMaV          -------------------------------------------------------------------  2027
CTLaV          -------------------------------------------------------------------  2037
CNRMV          -------------------------------------------------------------------  2038
```

Supplementary Data S2. Multiple sequence alignment of the PAFV1 movement protein (MP) and related proteins

```
PAFV1           MA-------------------------------------------------------PS--------LNRLKDKLRR----   14
EnFV1           M-------------------------------------------------------------------------------   1
M.cannonballus  M-------------------------------------------------------------------------------   1
ArTlV1          MALRFRHDPPSFKLTRYLLENVIEASSI-DTTEQIFGYAVLFTRDKKKKAKSLGLRPH---SFASFANSFKQWLRTH---   73
MpTLV1a         MAFKYRNNPKSFSLPREFVAHFRESQNV-DTDRTLFGFQVLFTKTKKKEWLEGGHAPV---TFLRFFREFGQWVKGF---   73
PpTlV1          MAFPYRHDPFNFYLDRSLYTYLHSTDPKEQTHFVIAGYA--FSISPHKKRKLFANKEHIIVSVKNFIDSFNKYMDRMIET   78
G.cichoracearum M-------------------------------------------------------------------------------   1
E.lata          MALRYRHDPPSFKLTKDLLRDVYTAKG--QFSRYIFGYTVLFTDDKINKLQQMRLRPH---SFRTFQKSFFEFLNRQ---   72
                *

PAFV1           SQASSSDSNAPPSSHPSPSPLPSSLPSHSSSGGDRPPAV-----------------------------------------   53
EnFV1           --------------------------------------------------------------------------------   1
M.cannonballus  ---------RPASRGHMPSRRPSSV-LHWLAG----PQY-----------------------------------------   26
ArTlV1          DAAQLSRAKSLLSGSS--AVGPYSRNGHYDSSGNFLGEY-----------------------------------------   110
MpTLV1a         SAKERGAAAIAATKHK--GVGLYPRRGVYRDNGDYLGTY-----------------------------------------   110
PpTlV1          KSVYLTHSRKLAQSGHYDQNSPFHYPGSYDRKGLLKESYVQGRPLLLFAERANTADDKNFKAELLHKSE-LLKHRETYEQ   157
G.cichoracearum --------------------------------------------------------------------------------   1
E.lata          NSTDRNAAQSDFKMKK--RKDKYAEPGEYDKDGNFLSKF-----------PSSVKENNYKNELKKEEDRYADNAETWKQ   138

PAFV1           -----------------VRSVPSDRKVAYGA-----------PLSR----------------------------------   71
EnFV1           --------------------------------------------------------------------------------   1
M.cannonballus  ----------------QRTSRS------GA-----------PYSR----------------------------------   38
ArTlV1          ------EVGDDEFEK------LSESEFG-DASTTHSK-----PFDKDMLHP----------------------------   143
MpTLV1a         ------------------QEVVSEGKI----------------DPTLLAM-----------------------------   126
PpTlV1          IPFEDLEEDSDAFEKGEKRRSTSQEKTR-GASISGVKASVELPLRRRTASGKDFDTNASIDNDVIEPLQTIKQPVSLKAR   236
G.cichoracearum --------------------------------------------------------------------------------   1
E.lata          Y-WESLSPEEKAFYA--KANIGSEDGT---------------NFDRKEMDPFKYQTDAQ--DETGESLRRAYQ-------   191

PAFV1           -------------------------------SYDTHSRNPSLPTRSAPTPPPFLPKASHVYSPGQSSGDDASDYHF   116
EnFV1           --------------------------------------------------------------------------------   1
M.cannonballus  --------------------------------QSSSRRDRNKHRSGAG-----YNSGYYYYPGSSQPSGSSTR--   74
ArTlV1          ------------------------GFTD-----SVRPNTGGKSQR--------------------PSEHSRENSFQS   171
MpTLV1a         ------------------------GMIDVLDSSVIESSTSQTSQATSSTTTSKKASKKKSFTANSAKTSSARVTHAQA   180
PpTlV1          RESAANADRLGSASSRKSFNIQEEPKKFSD-----TFQDFNKERSKTKEHNAGSRSADYSRTEFGNRASSYSQRQSDLGE   311
G.cichoracearum -------------------------------------SSKSKTEGSGEYSGQEKEQNNQRPENKSTGETKTHKEDQRKE   43
E.lata          -----------------------KGEQD---QSSGAASSSSTRQDHSSTGADASYRERKKENETQGQERQEKERERHE   243

PAFV1           A------------------PALDRLVP---------PVFSP---PTFHASLP-SAS--DHARSLHARHLDTTSPSSSGV   162
EnFV1           ------------------------------------PFFFS--DTARAKRH-WAR--RHASTFIKSLHDVPSSQP---   35
M.cannonballus  ---T-------SKQQRKPPRPEMDAFTAEDVPGSSGYPRLFS---GMAPSQMDIWAA--HHTRRARPRH------TSSGI   133
ArTlV1          V-AA-------RKNMQQMTMQQATTA------------------MTEGAVLN-KAT---VASTRHYKSFDLSS------   214
MpTLV1a         --KD-------TTQKAQVLDTS---------------------------------IQKRTSYSQFNLGD------   207
PpTlV1          SLAD-------IKNARHMSYSDFGSKAADDYKTYVPPPADFLKKLATETSQLEIPSS---IPSRYRIGNFNLSQ------   375
G.cichoracearum T-KD-------SKEQEQTFDEHFDNK----------------------QEEISAV---PDSSKKIRFIDLRS------   82
E.lata          D-KDKKRHDRNRKERQDQKDERKTKSSSEDKGKGKEKDKDTDKQDGKQTEESSPPASTKPAPKTFRLRNFNLSS------   316

PAFV1           RVGVLHLEANPPIKRLAVDRGDYHTVDAAFPGSARRLHTIVGPTGCGKSTFALLNAM--RGLTTLIISPTRANLANLLVE   240
EnFV1           ----FSLVPDPEMQRAPVTRGDYTNVHRLISKGRGRLNTITGPTGCGKSTFALLKLL--VGHSALIICPSSANQANLLTE   109
M.cannonballus  NP-HLDYVPFEKPQKLEVTRGQYQAITDFVGRGKRFGY-IVGPTACGKSSYALLDLLE-MGKSILLVQPSNTNVASCMAN   210
ArTlV1          ----LSRAP-LRETHVTVSRGRYSGVTREVVAGPRGGYVITGPTGCGKSTVALLPLFV-RKARVMVVEPTQANAANIFHE   288
MpTLV1a         ----RQRPLSFTEVHVPVARGNYRGVTQKIVAGPPGGYVITGPTGCGKSTVALLPLFS-GKSSVLIVEPTQANAANIFHE   282
PpTlV1          ----HMRSKNFSETHVPVSRGDYDKVRTKIVNSAKGGFVITGPTGCGKSTVALLPLFANVDSTALIVEPTQANASNILHE   451
G.cichoracearum ----LQRELSFKEVHVPVARGNYDEVRTKIVNGEPGGFVITGPTGCGKSTVALLPLFTRKGSTVLMIEPTQANAANILHE   158
E.lata          ----LQRPKDMVETHVPVKRGDYNKVRSSIIEGDDGGYVITGPTGCGKSTVALLKLFSSRKTRVLVVEPTQANAANIEHE   392
                             :   *  ** *    :    .     .        *.***.****: ***   :        :::  *:  :*  :.   :

PAFV1           FNTRIPNTVAKYKLSYHTPRAAHCHFLSYHDAPSPLSLCTAEEFCRYVGNHSTFPPVDLIVLDEYHLGTPHVIKSRFLLR   320
EnFV1           FNQRIPDTITRLGLDYISPTVSFCNFLTFQDPCTQLTLCTAQDFVNFVLRFKTFPPCDFLILDEYHLNTPSVIQSRFLLR   189
M.cannonballus  FKDRIPAAIETRNLPYSNPNVAHCGFLSCIDSPAQLNVCESTDLLDFFNQYGTFPPVDIIFFDEFHLPRQEHVICRLLLT   290
ArTlV1          FRNVLPNLFAAGIIDAPVPAVEFVAPTVSKPPFAPLAVTTTQKLLEFFDYFGSFPPVDYLVIDEFHLPIPSMVEVVEFVR   368
MpTLV1a         FSNVLPTLHEAGVIPWNVPPVEFTAPTTREGSYGPLSVTTTDKLLEYFEWKGSLPKVDYLIIDEFHLPIPSMVQTVELLR   362
PpTlV1          FTNILPNLVLNKIIKGVVPKASFIAPAIS-APVSRFSITTTEKLLEYFYFNGRLPKSDYIIIDEFHLPIPSMVEVVELIR   530
G.cichoracearum FQLILPNLVAKKIITGNIPPCSFVAPTVSRKPYTPLAVTTTDKLLEFFVAFDEFPKVDYMVIDEFHLPIPSMVKVVELLR   238
E.lata          FRHILPNLVTRIPSLGEVPDVAFRALTTAREPLPQLVVTTTEKVLEYFNYYGQILQFDYLVLDEFHLPIEDMVEMVELMR   472
                *   :*           *   .         .  : : :. .:.    :    * :::**:**       :     ::
```

```
PAFV1           HCVPDLAGKFRQ-WAIMVSATPPDEPPPVPRTSGITMIRIPIP------DPLTLPVPAIYRLSTHKKLGNHYLLIVVDGC   393
EnFV1           HCLPKGQTLKTQ-KIIFVSATPPDEPAPPVRTDGLTIVEMDIP------DMLATPTPPLYQLSKYPPYSNNYLLIVADSC   262
M.cannonballus  HL----VYKENPFTVLFASATPPDEAAPPPRMDGLTITKIKLR------DPTVEPLDDHYQLSKLPYFGNNMLLVIADSC   360
ArTlV1          ------TFTLVP-KYILVSATAQGFSVQPLPSAVTPIYGNLPMGTIPSEPVGSDLDP----RRWSNRGDTVAVVAPSI   437
MpTLV1a         ------TFALVP-KYIFVSATAVGYSVNPELPKAVTQTWGQLPIGQIPSKMEGSDLDP----RRWWKRGDGNVAVVAPSV   431
PpTlV1          ------TFDLCK-KYILVSATAIGFKINAQLPPAVTQIKGNIPIGKLPVNLTGTDLDP----RRWFHQGDGTIAIVAPSV   599
G.cichoracearum ------IFNLAP-KYIFVSATAKGYSVSPQLPSAVKVVKSRLPLGHLPKNIARTDLDP----RKWQRTGDGTVAVVAPSV   307
E.lata          ------TFNVAK-KYILVSATAEGFKVNAQLPPAVTRIQGVLPVGEIPAQLEGSDLDP----RIWADRADGTYGVVAPSV   541
                      ::.***.  .            .::     :        .              .:    ::. .

PAFV1           EAAHSLRDRLLDLGEPAFSLCKCPTAEQSSLFLTYNPSNATVIATPDTEAGITVPCSYMVNPGTALRVEFQESVIVPLVS   473
EnFV1           GTAHALTERLMAKSESVFCICECPTPESVAHDLTTNTYNHTVIATPATEAGITTPCCYMVNPGLASRTDFSNGVIRRDTF   342
M.cannonballus  LAAHHLQEKLSAMGQRAFLLCQHVSSERATELLITEKRNVTFIATPETEAGLDVPCSHFTNGGTALRTLFSRGVLFSSIF   440
ArTlV1          AVANKLTALYKSWHLRVYTITRETFVSDYMLAATNYVPGTVYVLEPGVEAGVTLSMSVLISMGCTTAVRYDGRVVVEDTQ   517
MpTLV1a         IVAKRLFNVYRDWQIRAFLITRETFVSEYMKAATNYRSMTTFVLEPGVEAGVTLCIDVLISMGASTAIRYDGKVVLEDTQ   511
PpTlV1          NVARKLFEIYKDWKLRVFLITRFTTVSVYAKACSNYLPYSVYVLEPGVEAGVTLSISVLISMGCSTAVRYDGKVVLEDTQ   679
G.cichoracearum EIARKLHALYLSWRLRSYLITRYTVVSRYLKATRDYRSETVYVLESGVEASLTLSMSVLVSMGASTAIRYDGKVVIEDTR   387
E.lata          TMAHTLYRQYKDWGVRTFLITRNTYASDYLKAVKNYSSRTAFVLEPGVEAGVTLSVAVLVSMGATMAVRYDGDVVIEDIQ   621
                    *. *             :: .  .              .: . .**.:    :. * :    :. *:

PAFV1           RLGPRQSNQRLGRAGRIGHTVVYLPPDGGEAP--SDAASFVLLAEAYLLI-LSLTGGHPLAPEATLVADRFSRLRKLTPA   550
EnFV1           PLGPRQSNQRLGRGGRLGHTVVFTQRSPGES---SDSPSAVDLGDAYLLV-LALTFGHPSSPEATLATTRFPRLKKVLGS   418
M.cannonballus  KLGSRQNVQRLGRAGRARHSLCWVDDSG-QADDRMDACSPVDAATGYLII-FRHTGKQPDSPDCRFAVKEFGRLAKVTSK   518
ArTlV1          PLDPIAAIQRGGRGGRVRPTLYIQPRAP-EA---VVAKSSADYYRAQAIIKAVALGAHTQHWQDDDIVETFPKLKTLTRS   593
MpTLV1a         PLDKIAAIQRGGRGGRVVPTLYITPRLP-EA---LTMVSSADYYRAQSIVRLIAAGADVSRINDRGLFQTFPRLRTVSRQ   587
PpTlV1          PLDEIAAIQRGGRGGRIKPTLYVSPVVP-KS---ETRSSTSDYYRAQAVIKMIALGADLRKINFKGLDNQFPKLKTVTKS   755
G.cichoracearum PLDKIVAIQRGGLGGRVIPTLYVTPEAP-EV---LDESPTADYYKVQAIICALAYGADISKIKCTELTKTFPKLLTITKE   463
E.lata          PLSAIAAIQRGGRGGRVVPTLYVEPEAP-KD---VAAGSSADYYRARAIIKMIALGARTEGMRDNGIFATFPRLKTVTRE   697
                 *.      ** * .**   ::      :          .         ::              * :*  .:

PAFV1           AARIALRSP--QPLTALYRSDNQGVLYAEYGGSASTFVADNAADFRLFKWPGGSAYAPYLDLCT-DHDLSVGMTLDLQRS   627
EnFV1           GASLFVETT--RPMIALYMHDNAGERYAEFGGSATGFIDDNASDFRLFKWPSGSAYAPFLDLCS-DHDLSSLHNPALQKS   495
M.cannonballus  AAELVLQAPDATPVLELYKRNTEGELFREFGGHEDGFIDECAADLRLFVYPGGAFFAPFVDLKQ-DYDPTQGQTLKSIRS   597
ArTlV1          LAISALSAP-GDPFVAVYSRNANGDIYRECGGSGTGFAELAARELLLYHYPGGFFIAPIADFSDIDSKPYEFVRRESQLD   672
MpTLV1a         LASAAVSVS-GDPFVALYQRADDGQIYSECGGTGSGFDRLAKQELFLYHYPGGFFVAPIVDFTDLHSNPDSFVLRQSQLS   666
PpTlV1          LAAQAVASR-SDPFIAVYKTDASGKVYRECGGDGTGFSDLAKNELFLYHYPGGFYIAPISDFTDPLTKPNTFVLRNSQFT   834
G.cichoracearum LATIAMLSG-GDPFVNIYKSSKEGILYKECGGTGEGFDELAKKELFVYYYDEGFYVAPITDFSNLNSKPDSFVLRKYQFS   542
E.lata          LALAALEAG-GDPFISIYRHNDAGQVVYECGGNGGGFSRLAEQELKLYYWPAGFYVAPITDFSRLNSEPDTFVLRSHQLR   776
                   *    :       *. :*    *   : ***     *        ::  ::  :    *     ** *:       .

PAFV1           IAGAIMDLHPEFQTM-LDLDSALSSALLNPAPFATAIWNALKPLDGDSNLR--------APSLDQVR-DRITP--AYMLG   695
EnFV1           IADAIVAQDPSLVEG-ITLESALEAAERDPETYANAIWGALQRLHGRSNLT--------GESPGHPGISNCTL--NYMLG   564
M.cannonballus  LAKAAINNLPHLRDKEVNVREALKYAERFPDMFTTPIWLAFKTLKGKSVLK--------TRSPEEDP-IFSDP--VALFG   666
ArTlV1          AAWAMASSVGGLIDR-YTLDDLVGMLVGKFDVYVGDLFTRLKRIFNKPEPEKFSLSGRDSEYPEMRDFLKDSPPMVKLFE   751
MpTLV1a         AAHQIVEAIPGLEDS-YSLDDLIGLLIAKFDVYVSDLFTRLTAVFSEPAPTQYGIGS-KTRSPDVTDFLTKAPEISKLFV   744
PpTlV1          AAKAIVDSIAGLSAK-YDLNNLVDILISKFDVYVADLFLRLKEIFSSDVATPFSIS--KSRPAEIEDFVKSSPPVLNLFA   911
G.cichoracearum AARLIVEAVPKLTEK-FEFDKLVEQLIRSFKNYVNDLFVQLKSVFPATQPSGYRVG---NKAAEVEDFFRPLPPLVKLFS   618
E.lata          AARTMVEAIPGLADR-YPLGELVNMLIAKIDIYVNDLFGLLKSIFIGDKPGRFSLRG-RDYAPTLADFLGSAPEVLKLFN   854
                   *        :           . . :           :. :: : :                            ::

PAFV1           QTGAKAWSVLVSLGAYTSFPVESSSS-----EPYIRRLLHFRGETVSYRSTKVLG-DDGKVSEAKIAALLSGDLLPVAAT   769
EnFV1           VNGTRAFKVIEQLGARFEVSSEPQGG-----KYVIHRELFFRDQRVRYNGKPLLG-PDNLVSPEKVAAILTPALSPVTAT   638
M.cannonballus  ELGARAWRI---LGAKASITISSSSSGDRERPYQIHRTLHYGGEQFEFTSRDIL--EDRTISEAKVTRLLITHLKPVAQI   741
ArTlV1          YMTSNP------SGVQYSRRE-FPGS-----PNRATHALNFKGTTLNFAFDSKYM-NQTTLNTKLLAKDVYELLQGLLAV   818
MpTLV1a         YMTSQP------CGVIYERGEFEKNG-----TKHSTHSFIYKDTSLHFAFSARFM-NDHVVNVDLLGKEIHSLLKGLLAI   812
PpTlV1          YLKTSP------SGVVYDR---VLSS-----PTQAKHSFSYKGEFLHFSFSSEYM-KGSEVDVNKLSRSIFVLLQGLLAV   976
G.cichoracearum FLKTQP------CDVKYQRHETKSGD-----WFLGYHSFSFNERSLHYAVPPKYIIDGTMVDTEKLARDTFEILKGILCI   687
E.lata          HMATEP------AGIKFVRKVEVEDG-----VHYSEHSFEYSGKQLNFSFPADYT-AKGTVDVEKLSREVYAKLQHILAV   922
                 : .         .         ..       : : :  .:  .             :. :       * :

PAFV1           QLLL-DHPQ-LSVDLASFLPYRSRSSNTWFKSLFPVTQ-----                                        805
EnFV1           QALL-GKPE-YTIDLTSLAHLAPTCENRWFRTLF---------                                        670
M.cannonballus  FMLQ-TDPEFKSCDLAEFAYAQNRCVNKWFCNLK--------F                                        775
ArTlV1          EMMIAGAPQ-KCVDLQDYCNRVSP-EHYWFKVNVTDRRSGVKF                                        859
MpTLV1a         EILIEGAPN-KCVNLFDYKHRVPR-EHLWFSQHVGRT------                                        847
PpTlV1          EILLDGASH-KCVDLVQYKGKISD-EHIWFREISQDWS-----                                       1012
G.cichoracearum EMLLSGAPE-KCVDLNQYKAMVPG-EHKWFKENVVDRRS---I                                        725
E.lata          EILVNGTPE-KCVDLMKYKDRVSR-EHVWFTRVREHSSKVNLK                                        963
                 :    ..    :*. .         :  **
```

Supplementary Data S3. Multiple sequence alignment of the PAFV1 hypothetical protein (HP) and related proteins

```
PAFV1         MAELSFPVVC-TQGHEARSLTSPEELATHHASAHPNVSINDYRNSLRSSGRGSDRINAPPPPPVFNNP---LHAWGLNVS   76
EnFV1         MA--SYPAVC-SLGHPPVTVNTDAELQSHITSEHPDSDPSSVNNALAGAGSSG---ASGPPVPSLSDP---LSKWGITIT   71
M.cannonballus MSMPKLPATCPVAACNNKELESAADLFTHHSSEHKAIDMDTYKGMFTDTAR---RVTNPSNTVDYVQPKLLISRAGMTVS   77
              *:  .  *..*   .         : :  :*  :*  :* *  . .  .. :  .:.        .      :*   :    *:.::

PAFV1         NFTAYSSDDVSSGSELSVRLLQEFFDAYGRSFSPPMPIDTAALLLDFFLYCLFNAATNDASPTGSFVLRESDGPRQQKIR   156
EnFV1         SLTSIRPETVCVPTAISITDFTQFLASYGKQVVPTVDPDVGAFLIEFFRYVLFTGATEDASSAGSFQLRDKSGGSAPEVS   151
M.cannonballus NFKNIRSTEVTSNSLITFQAAANFLLELGAKFG-CLDPSHERLQIDLFILILNHAGTDDVEGLGSTTLMSNDGTAKKTIT   156
              .:.    .  *   :  ::.    :*:   *  ..  :   .  : :::*    *    ..*:*..   ** *  ...*     :

PAFV1         WTAFFAAAQDHFDPA-GYAFTPRRMMRTCEPAFWELWQNSEVKALDRVRSEGTPISRQWRLPNGKHPPAYALVPQLFTSH   235
EnFV1         WSSFFNSCKDYFNSR-GSLFTPRRLARTCDEAFWEMWNNNKIEALNDVRTRGTPISRRFRMQGGRAPKAYVLVPELFDSR   230
M.cannonballus WVQFQEAAKEYFDGELGIKFTFRRCIPSFERLWWELWNMDDLEALDDVKQFGTRRSRRWQ-QSGEPVEPYVLVPDLFDDH   235
              * *  *  :.:::*:    *  ** **   :  :  :**:*:  .::**: *:  **  **::: .*.   .*.***:** .:

PAFV1         LTNDERACRKSHQATI---TKLSTAAARPEYEGLDPENTVVASDVERQALEHRSLDLQNKTAGLRNRGGLGSGNAPRADP   312
EnFV1         LTAEELEVRRLHQSTI---TKIATNTTRPTFEGLDQEGSHLAYEAAKA-----------RTTRIIQNAGASADEAQRAYQ   296
M.cannonballus LTARERKVRLLYNEIVELEKEAGATTDDITYVGRDSEHKIGKAKL-RDALE------DNARARAMLNGGKARSNASKEHD   308
              ** *  *   * ::  :     .: .:: :    : * * * .   . :          :    ..* . .:* :

PAFV1         VLEAMYD---------SRHGQGR--------   326
EnFV1         SLHHATNASVPPNPPQSRYGGSRFTFGESDF   327
M.cannonballus DWMDTLNKQAIANNARYRD-----------   327
                      :             *
```

Supplementary Data S4. Ribosomal RNA contigs and their top five hit species in the NCBI nucleotide database

A. NODE_192_length_6473_cov_262.897260_g154_i0: 6473 nt, plant nuclear genome

Acc. No.	Coverage	Identity	Genus and species	Family	Order
KX522674.1	98%	94.85%	*Spondias tuberosa*	*Anacardiaceae*	*Sapindales*
KX064011.1	90%	93.25%	*Laennecia sophiifolia*	*Asteraceae*	*Asterales*
KX063989.1	90%	93.14%	*Blakiella bartsiifolia*	*Asteraceae*	*Asterales*
KX063947.1	90%	93.15%	*Diplostephium azureum*	*Asteraceae*	*Asterales*
KX063974.1	90%	93.10%	*Hinterhubera ericoides*	*Asteraceae*	*Asterales*

B. NODE_11601_length_1855_cov_155.436877_g8258_i0: 1855 nt, plant chloroplast genome

Acc. No.	Coverage	Identity	Genus and species	Family	Order
NC_037471.1	100%	100.00%	*Pistacia weinmaniifolia*	*Anacardiaceae*	*Sapindales*
KY549635.1	100%	100.00%	*Pistacia vera*	*Anacardiaceae*	*Sapindales*
KX447140.1	100%	99.95%	*Rhus chinensis*	*Anacardiaceae*	*Sapindales*
KX871231.1	100%	99.84%	*Mangifera indica*	*Anacardiaceae*	*Sapindales*
KY635877.1	100%	99.36%	*Anacardium occidentale*	*Anacardiaceae*	*Sapindales*

C. NODE_17762_length_1371_cov_95.266263_g12380_i0: 1371 nt, fungal nuclear genome

Acc. No.	Coverage	Identity	Genus and species	Family	Order
GU250346.1	100%	95.71%	*Meristemomyces frigidus*	*Teratosphaeriaceae*	*Capnodiales*
GU250356.1	100%	95.64%	*Teratosphaeriaceae* sp.	*Teratosphaeriaceae*	*Capnodiales*
GU214583.1	100%	95.56%	*Teratosphaeria stellenboschiana*	*Teratosphaeriaceae*	*Capnodiales*
GU214520.1	100%	95.56%	*Neocatenulostroma microsporum*	*Teratosphaeriaceae*	*Capnodiales*
GU214518.1	100%	95.56%	*Neocatenulostroma germanicum*	*Teratosphaeriaceae*	*Capnodiales*

D. NODE_18168_length_1344_cov_5.720463_g12643_i0: 1344 nt, fungal nuclear genome

Acc. No.	Coverage	Identity	Genus and species	Family	Order
NG_057777.1	100%	96.06%	*Pseudotaeniolina globosa*	unassigned	*Capnodiales*
GU214696.1	99%	96.05%	*Scorias spongiosa*	*Capnodiaceae*	*Capnodiales*
GU214449.1	100%	95.84%	*Mycosphaerella* sp.	*Mycosphaerellaceae*	*Capnodiales*
EU019255.2	100%	95.77%	*Neocatenulostroma microsporum*	*Teratosphaeriaceae*	*Capnodiales*
NG_058782.1	100%	95.69%	*Capnobotryella renispora*	*Teratosphaeriaceae*	*Capnodiales*