Acta virologica **62**: 122 – 128, 2018 doi:10.4149/av_2018_201

Identification of a novel plant amalgavirus (*Amalgavirus*, *Amalgaviridae*) genome sequence in *Cistus incanus*

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

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

Received February 2, 2018; accepted March 3, 2018

Summary. – *Amalgaviridae* is a family of double-stranded, monosegmented RNA viruses that are associated with plants, fungi, microsporidians, and animals. A sequence contig derived from the transcriptome of a eudicot, *Cistus incanus* (the family *Cistaceae*; commonly known as hoary rockrose), was identified as the genome sequence of a novel plant RNA virus and named Cistus incanus RNA virus 1 (CiRV1). Sequence comparison and phylogenetic analysis indicated that CiRV1 is a novel species of the genus *Amalgavirus* in the family *Amalgaviridae*. The CiRV1 genome contig has two overlapping open reading frames (ORFs). ORF1 encodes a putative replication factory matrix-like protein, while ORF2 encodes a RNA-dependent RNA polymerase (RdRp) domain. An ORF1+2 fusion protein, which functions in viral RNA replication, is produced by a +1 programmed ribosomal frameshifting (PRF) mechanism. A +1 PRF motif UUU_CGU, which matches the conserved amalgavirus +1 PRF consensus sequence UUU_CGN, was found at the boundary of CiRV1 ORF1 and ORF2. Comparison of 25 amalgavirus ORF1+2 fusion proteins revealed that only three different positions within a 13-amino acid segment were recurrently used at the boundary, possibly being selected so as not to interfere with correct folding and function of the fusion protein. CiRV1 is the first virus found to be associated with the *Cistus* species and may be useful for studying amalgaviruses.

Keywords: Cistus incanus RNA virus 1; Amalgavirus; Cistus incanus; +1 programmed ribosomal frameshifting

Introduction

Amalgaviruses (the family *Amalgaviridae*) are double-stranded, monosegmented RNA viruses with one confirmed and two proposed genera, namely *Amalgavirus*, and "Zybavirus" and "Anlovirus", respectively (Depierreux *et al.*, 2016; Martin *et al.*, 2011; Pyle *et al.*, 2017; Sabanadzovic *et al.*, 2009). Species of the genus *Amalgavirus* have been isolated from various plants and are known as plant amalgaviruses (Liu and Chen, 2009; Martin *et al.*, 2011; Sabanadzovic *et al.*, 2010; 2009). Zygosaccharomyces bailii virus Z (ZbV-Z), a prototype species of the proposed "Zybavirus" genus, was isolated from the yeast *Zygosaccharomyces bailii* (Depierreux *et al.*, 2016). Antonospora locustae virus 1 (AnloV1) represents the other proposed genus, "Anlovirus," which infects *Antonospora locustae*, a microsporidian pathogen of grasshoppers (Pyle *et al.*, 2017). Two additional "Anlovirus" species are associated with giant springtails and two-pronged bristletails, respectively.

The amalgavirus genome contains two open reading frames (ORFs), of which ORF1 encodes a protein of unknown function. The ORF1 protein was initially thought to be a coat protein (Liu and Chen, 2009; Sabanadzovic *et al.*, 2009); however, its predicted tertiary structure has α-helical coiled coil, which is uncommon for viral capsid proteins (Nibert *et al.*, 2016; Pyle *et al.*, 2017). The ORF1 protein may thus have an alternate function, such as formation of the replication factory matrix (Isogai *et al.*, 2011; Nibert *et*

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

Abbreviations: AnloV1 = Antonospora locustae virus 1; CiRV1 = Cistus incanus RNA virus 1; ORF(s) = open reading frame(s); PRF = programmed ribosomal frameshifting; RdRp = RNA-dependent RNA polymerase; STV = Southern tomato virus

al., 2016). The second ORF (ORF2), which partially overlaps with ORF1, encodes a RNA-dependent RNA polymerase (RdRp). An ORF1+2 fusion protein, formed by fusing ORF1 and ORF2 using a +1 programmed ribosomal frameshift (PRF) mechanism, is involved in viral RNA genome replication (Depierreux *et al.*, 2016; Nibert *et al.*, 2016).

Amalgaviruses show a phylogenetic relationship to partitiviruses (the family *Partitiviridae*), which infect plants, fungi, and apicomplexans (Martin *et al.*, 2011; Nibert *et al.*, 2014). They also share similarities in genomic organization with totiviruses (the family *Totiviridae*), which infect fungi and single-celled eukaryotes (Kondo *et al.*, 2016). Due to the close relationship among these three viral families, amalgaviruses are suggested to represent a transitional intermediate between totiviruses and partitiviruses (Krupovic *et al.*, 2015; Martin *et al.*, 2011; Sabanadzovic *et al.*, 2009).

Transcriptome or metatranscriptome data generated from total RNA isolated from organism or environmental samples often contain sequence reads derived from viral genomic RNAs, which can be identified by comprehensive bioinformatics analysis (Kim *et al.*, 2014; Liu *et al.*, 2012; Nibert *et al.*, 2016). As a result, many plant RNA virus genome sequences were discovered by analyzing transcriptome datasets (Goh *et al.*, 2018; Kim *et al.*, 2018; Park *et al.*, 2018; Park and Hahn, 2017a,b). In this study, a novel plant amalgavirus genome sequence was identified in a transcriptome dataset obtained from leaves of *Cistus incanus* (the family *Cistaceae*; commonly known as hoary rockrose).

Materials and Methods

Transcriptome dataset. The transcriptome dataset analyzed in this study was downloaded from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). The *Cistus incanus* RNA-seq data obtained from leaves were deposited under Acc. No. SRP093603 (F. Sebastiani, manuscript in preparation) and contained 7.4 gigabases (Gb) of paired-end reads. The sickle program (version 1.33; https://github.com/najoshi/sickle; parameters, -q 30 -l 55) was used to screen raw RNA-seq reads and high-quality reads were collected. De novo sequence assembly was performed using the SPAdes Genome Assembler (version 3.10.1; parameter, --rna) (Bankevich *et al.*, 2012).

Collection of viral genome sequence contigs. To collect sequence contigs putatively derived from viral genomes, a BLASTx search was carried out against a local viral RNA-dependent RNA polymerase (RdRp) sequence database, using the following parameters: -outfmt 6 -evalue 1e-5 -max_target_seqs 1 -max_hsps 1. The local RdRp protein sequence database of reported RNA viruses was prepared using sequences obtained from the Pfam database (release 30.0; http://pfam.xfam.org). A total of 345 representative viral RdRp domain sequences, defined by the Pfam database, were obtained from 19 Pfam families with Acc. No. PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03431, PF04196, PF04197, PF05788, PF05919, PF07925, PF08467, PF08716, PF08717, and PF12426.

Mapping. Mapping of *C. incanus* RNA-seq reads to a virus genome contig sequence was performed using BWA software (version 0.7.16a-r1181; http://bio-bwa.sourceforge.net) (Li and Durbin, 2009). The variants were called using the SAMtools package (version 1.6; http://www.htslib.org) (Li, 2011).

Sequence comparison. Multiple sequence alignments were generated using MUSCLE software (https://www.drive5.com/muscle) (Edgar, 2004). A phylogenetic tree was inferred by the neighbor-joining method implemented in the ClustalW2 program (http://www.clustal.org) (Larkin *et al.*, 2007). Secondary structure was predicted using the PSIPRED webserver (version 3.3; http://bioinf.cs.ucl.ac.uk/psipred) (McGuffin *et al.*, 2000). Sequence logo representation was generated using the WebLogo webserver (version 3; http://weblogo.threeplusone.com) (Crooks *et al.*, 2004; Schneider and Stephens, 1990).

Results and Discussion

RNA-seq reads in total of 7.4 Gb were obtained from *C. incanus* leaves and assembled into 135,253 transcript contigs. One of the contigs showed a strong amino acid (aa) sequence similarity with a RdRp domain of Southern tomato virus (STV) (UniProt Acc. No. A8R3Y5; Pfam Acc. No. PF02123). STV is the reference virus strain for the genus *Amalgavirus* of the family *Amalgaviridae* (Sabanadzovic *et al.*, 2009), suggesting the *C. incanus* contig was derived from an amalgavirus or related virus genome.

A BLASTx search of the NCBI non-redundant protein database confirmed the contig was related to plant amalgaviruses, including Blueberry latent virus (BLV), Zostera marina amalgavirus 1 (ZmAV1), Zostera marina amalgavirus 2 (ZmAV2), Allium cepa amalgavirus 1 (AcAV1), Allium cepa amalgavirus 2 (AcAV2), Spinach amalgavirus 1 (SpAV1), STV, and Rhododendron virus A (RHV-A) (Martin *et al.*, 2011; Nibert *et al.*, 2016; Park *et al.*, 2018; Park and Hahn, 2017b; Sabanadzovic *et al.*, 2009, 2010). The contig was therefore considered to be derived from a novel plant RNA virus and named Cistus incanus RNA virus 1 (CiRV1). The CiRV1 genome sequence is available in the NCBI nucleotide database under Acc. No. MG833407.

To validate CiRV1 genome sequence homogeneity, raw *C. incanus* RNA-seq reads were mapped to the CiRV1 genome contig and possible variants were identified. There were 75 polymorphic sites (Supplementary Table S1), indicating the genome contig is a composite sequence derived from a CiRV1 population.

The CiRV1 genome contig was 3323 nucleotides (nt) long and contained two overlapping ORFs (Fig. 1a). ORF1 encodes a 385 aa protein, which showed sequence and structural similarities to ORF1 proteins from other amalgaviruses.

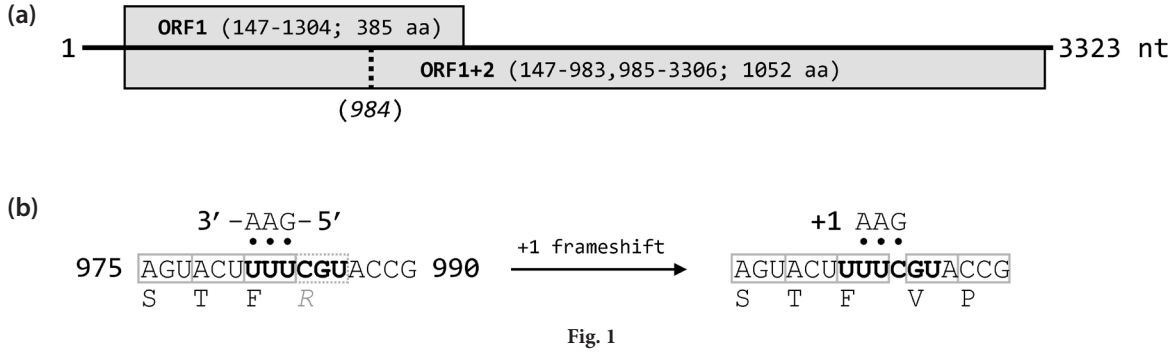

Fig. 1

Genomic structure of CiRV1 and proposed +1 PRF mechanism

(a) Genomic structure of CiRV1. CiRV1 has two overlapping ORFs. ORF1 encodes a 385 aa protein. The ORF1+2 fusion protein is produced by a +1 PRF mechanism and has 1052 aa. Nt position 984, which is skipped by a +1 PRF event, is marked by a dotted line. **(b)** Proposed +1 PRF mechanism of CiRV1. The CiRV1 +1 PRF region (positions 975–990) is shown. A $tRNA^{Phe}$ with an anticodon sequence 3'-AAG-5' initially binds to a UUU codon and may slip forward by one nt, thereby causing a C nt to be skipped at position 984. The consensus +1 PRF sequence is indicated in bold letters. Codon-anticodon base pairs are indicated by dots. Codons are marked by boxes. Single letter aa codes below the nt sequence are: S, serine; T, threonine; F, phenylalanine; R, arginine; V, valine; and P, proline.

CiRV1 ORF1 protein was predicted to be exclusively composed of α-helices, suggesting it may not function as a coat protein but a replication factory matrix-like protein (Isogai *et al.*, 2011; Krupovic *et al.*, 2015; Pyle *et al.*, 2017).

The second protein encoded by CiRV1 is an ORF1+2 fusion protein that requires a +1 PRF for proper translation. The consensus +1 PRF motif sequence UUU_CGN, where the underscore indicates the ORF1 codon boundary and N is any nt, is commonly found in amalgaviruses and influenza A virus (Depierreux *et al.*, 2016; Firth *et al.*, 2012; Nibert *et al.*, 2016; Park *et al.*, 2018). In the CiRV1 genome sequence, a +1 PRF motif (UUU_CGU) was identified at positions 981–986 (Fig. 1b). Initially, a phenylalanyl-tRNA ($tRNA^{Phe}$) with an anticodon sequence of 3'-AAG-5' would interact with the ORF1 UUU codon. The next codon CGU is a rare arginine codon in eukaryotic organisms, including plants (Li *et al.*, 2016). When the CGU codon is not bound by an arginyl-tRNA for an extended period, a $tRNA^{Phe}$ positioned on UUU may slip forward by one nt and bind to a UUC triplet, which is in the +1 frame relative to ORF1. As a result, when a +1 PRF occurs, the codon boundary changes from UUU_CGU_A (ORF1) to U_UUC_GUA (ORF2). This process skips a cytosine (C) residue at position 984, thereby causing a +1 frameshift for continued ORF2 translation. A UUU codon for ORF1 would subsequently be followed by a GUA codon for ORF2.

The CiRV1 +1 PRF motif UUU_CGU matches the consensus sequence UUU_CGN of other amalgavirus +1 PRF sites (Fig. 2). An uracil (U) residue is preferred both at the position before the motif and at the N position of the motif. Therefore, the most common 7 nt sequence of +1 PRF site is U_UUU_CGU, which was also identified in CiRV1.

The ORF2 component of the CiRV1 ORF1+2 fusion protein begins at nucleotide position 985, which is the first base after the +1 PRF site. The ORF2-encoded component has 773 aa and has a conserved viral RdRp motif (Pfam Acc. No. PF02123). The CiRV1 ORF1+2 fusion protein has 1052 aa, which may function in CiRV1 genome replication.

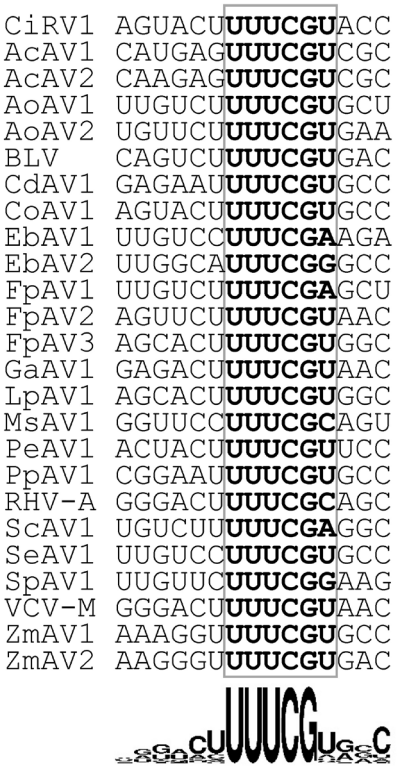

Fig. 2

Comparison of amalgavirus +1 PRF motif sequences

Sequences matching the +1 PRF consensus sequence UUU_CGN are marked in bold letters. Sequence logo representation is shown at the bottom. See Table 1 for full virus names.

Table 1. Sequence identities of ORF2 proteins of CiRV1 and related viruses

No.	Acronym	Full name	Accession No.[a]	Identity with CiRV1[b]
1	FpAV2	Festuca pratensis amalgavirus 2	GBXZ01002308.1	461/767 (60%)
2	FpAV3	Festuca pratensis amalgavirus 3	GBXZ01009138.1	440/767 (57%)
3	LpAV1	Lolium perenne amalgavirus 1	GAYX01076418.1	437/767 (57%)
4	BLV	Blueberry latent virus	NC_014593.1	399/763 (52%)
5	ZmAV1	Zostera marina amalgavirus 1	NC_034614.1	377/740 (51%)
6	ZmAV2	Zostera marina amalgavirus 2	NC_034615.1	378/745 (51%)
7	AcAV1	Allium cepa amalgavirus 1	NC_036580	356/721 (49%)
8	AcAV2	Allium cepa amalgavirus 2	NC_036581	355/722 (49%)
9	PeAV1	Phalaenopsis equestris amalgavirus 1	GDHJ01028335.1	371/737 (50%)
10	SpAV1	Spinach amalgavirus 1	NC_035070.1	341/739 (46%)
11	STV	Southern tomato virus	NC_011591.1	364/765 (48%)
12	EbAV1	Erigeron breviscapus amalgavirus 1	GDQF01098448.1	349/760 (46%)
13	EbAV2	Erigeron breviscapus amalgavirus 2	GDQF01120453.1	340/761 (45%)
14	CoAV1	Camellia oleifera amalgavirus 1	GEFY01004381.1	361/770 (47%)
15	GaAV1	Gevuina avellana amalgavirus 1	GEAC01063629.1	369/763 (48%)
16	CdAV1	Cleome droserifolia amalgavirus 1	GDRJ01026949.1	344/722 (48%)
17	MsAV1	Medicago sativa amalgavirus 1	GAFF01077243.1	336/718 (47%)
18	VCV-M	Vicia cryptic virus M	EU371896.1	321/716 (45%)
19	RHV-A	Rhododendron virus A	NC_014481.1	361/777 (46%)
20	AoAV1	Anthoxanthum odoratum amalgavirus 1	GBIE01024896.1	340/713 (48%)
21	FpAV1	Festuca pratensis amalgavirus 1	GBXZ01049574.1	345/726 (48%)
22	CaAV1	Capsicum annuum amalgavirus 1	JW101175.1	338/746 (45%)
23	ScAV1	Secale cereale amalgavirus 1	GCJW01039808	327/713 (46%)
24	PpAV1	Pinus patula amalgavirus 1	GECO01025317	340/728 (47%)
25	AnloV1	Antonospora locustae virus 1	NC_035189.1	125/536 (23%)

[a]Acc. Nos. of viral genome sequences; [b]Amino acid sequence identities have been described in the following format: identical residues/aligned length (% identity).

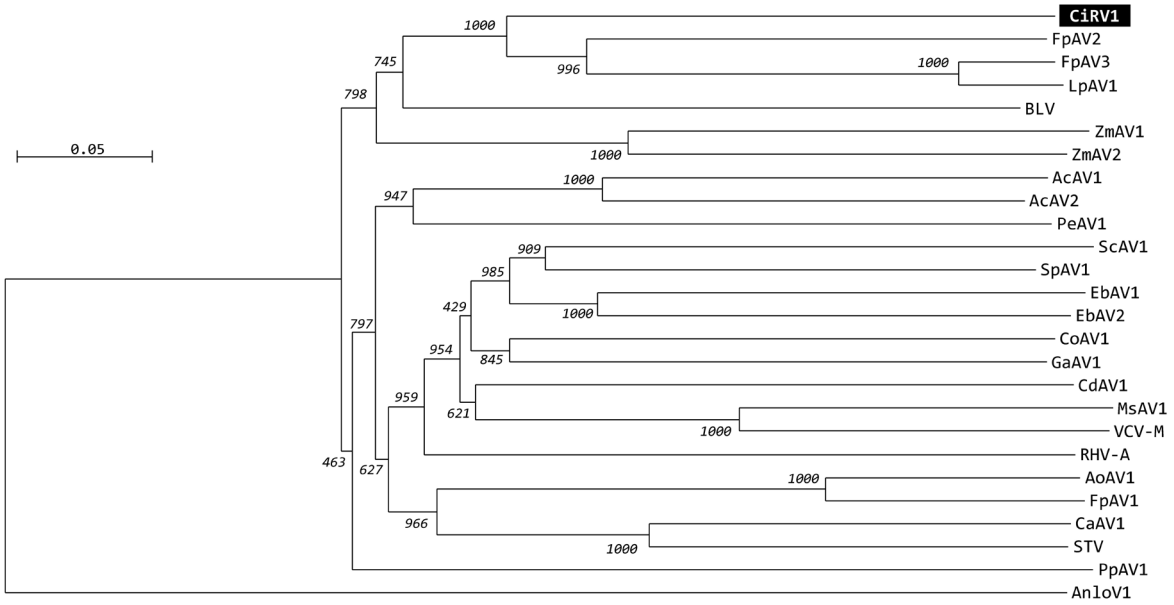

Fig. 3

Phylogenetic tree of CiRV1 and related amalgaviruses

A neighbor-joining phylogenetic tree was inferred based on ORF2 protein sequences. AnloV1 was used as an outgroup. Bootstrap values calculated from 1000 replicates are shown at the nodes. See Table 1 for full name of viruses and aa sequence identity with the CiRV1 ORF2 protein.

The ORF2 component of the CiRV1 fusion protein showed 45–60% aa sequence identity with ORF2 proteins previously reported for amalgaviruses (Table 1). The virus most closely related to CiRV1 was Festuca pratensis amalgavirus 2 (FpAV2), with 60% aa sequence identity. The RdRp protein sequence identity threshold for assigning amalgaviruses to different species is 65–70% (Nibert *et al.*, 2016), indicating that CiRV1 is a novel amalgavirus species. The CiRV1 ORF2 component showed approximately 23% aa sequence identity with the ORF2 protein of AnloV1, which is a species of the proposed "Anlovirus" genus, a sister genus to the *Amalgavirus* genus of the family *Amalgaviridae* (Pyle *et al.*, 2017).

The phylogenetic relationship between CiRV1 and other amalgaviruses was investigated based on a multiple sequence alignment of the RdRp-motif portion of the ORF2 sequence from CiRV1 and 24 other amalgaviruses. A neighbor-joining tree inferred from the ORF2 protein alignment confirmed that CiRV1 belongs to the genus *Amalgavirus*, of which members are plant amalgaviruses (Fig. 3). CiRV1 formed a strong clade together with FpAV2, Festuca pratensis amalgavirus 3 (FpAV3), and Lolium perenne amalgavirus 1 (LpAV1).

A previous study of the +1 PRF position (at the boundary between ORF1 and ORF2) of ORF1+2 fusion proteins revealed that only three positions are recurrently used in plant amalgaviruses (Park *et al.*, 2018). The multiple sequence alignment of CiRV1 and 24 other amalgavirus ORF1+2 fusion proteins revealed that the CiRV1 +1 PRF also occurred at one of the three positions, designated as positions #1, #2, and #3 (Fig. 4 and Supplementary Fig. S1). The +1 PRF occurs at positions #1, #2, and #3 in 9, 2, and 14 amalgaviruses, respectively. CiRV1 +1 PRF is at position #3, which is the most common.

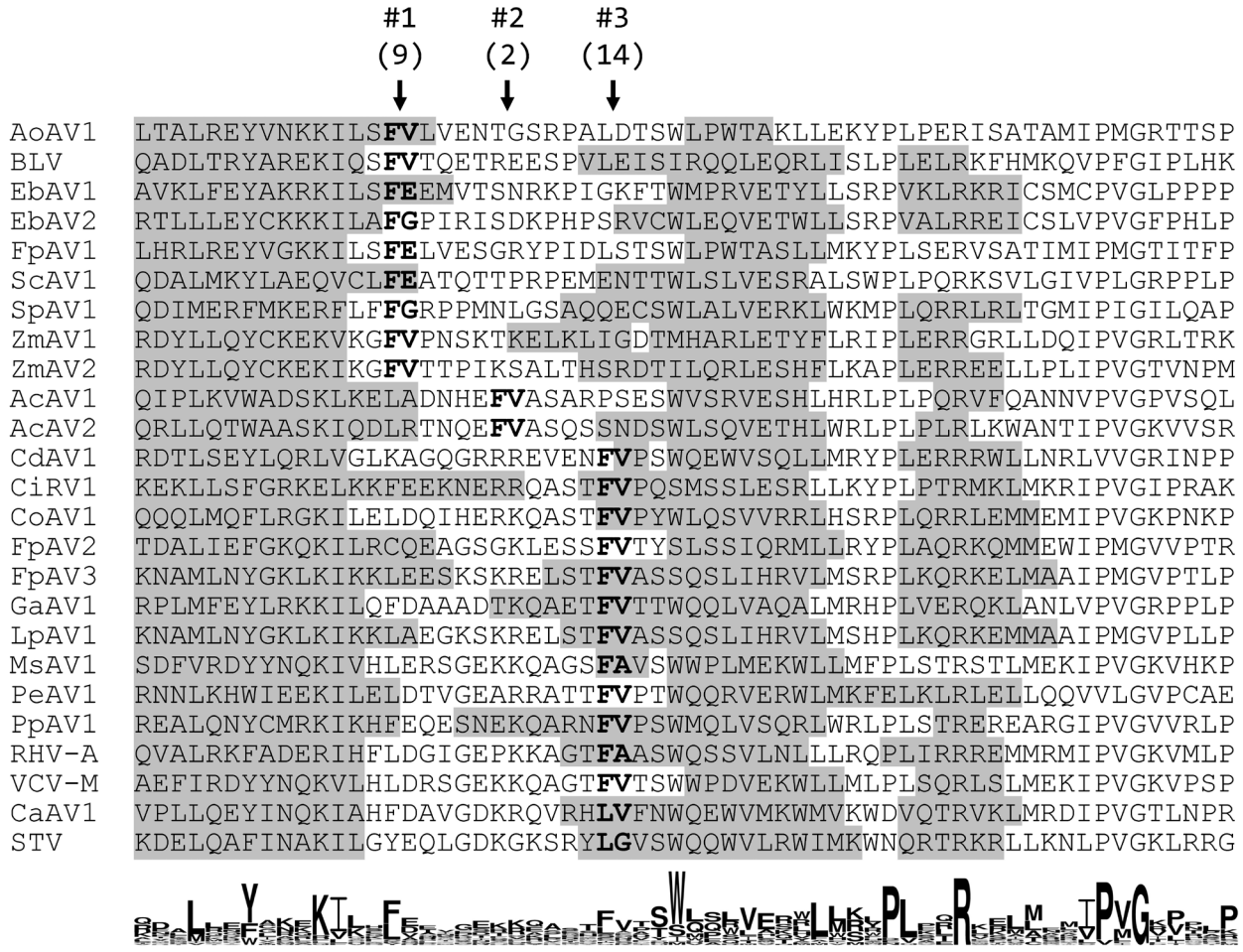

Fig. 4

Multiple sequence alignment of regions encompassing the +1 PRF sites of 25 amalgavirus ORF1+2 fusion protein sequences

Predicted α-helices are marked with a gray background. Three recurrent +1 PRF sites are labeled as #1, #2, and #3, with arrows indicating the ORF1 and ORF2 boundary. Numbers of viruses are shown in parenthesis. The last residue of ORF1 and the first residue of ORF2 are marked in bold letters. Sequence logo representation is shown at the bottom. See Supplementary Fig. S1 for the full-length alignment.

The three +1 PRF sites are closely located to each other within a 13 aa segment bounded by multiple conserved residues (Fig. 4). Distribution of +1 PRF sites among amalgaviruses indicates that these positions were switched repeatedly during virus evolution. However, only three positions within a short interval were recurrently used, implying that the ORF1/ORF2 boundary was highly conserved, which is likely the result of selection for proper folding of the fusion protein. Secondary structure prediction of 25 amalgavirus fusion proteins revealed that the +1 PRF positions were preferentially located within a random coil between two α-helices, one from ORF1 and the other from ORF2, or near the tip of an α-helix (Fig. 4). It is most likely that the ORF1+2 fusion protein position is under selection to ensure it does not interfere with proper folding and function of the fusion protein.

In conclusion, the full-length genome sequence of a novel amalgavirus CiRV1 was identified in the *C. incanus* transcriptome. CiRV1 is the first virus associated with any *Cistus* species (http://www.genome.jp/virushostdb; as of January 24, 2018) (Mihara *et al.*, 2016). CiRV1 genome sequence may be useful for studying evolution of amalgavirus genomic features, including a +1 PRF motif.

Acknowledgments. This research was supported by the National Research Foundation of Korea funded by the Korea Government (grant No. 2017R1A1B4005866).

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

References

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, Pevyer 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

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

Depierreux D, Vong M, Nibert ML (2016): Nucleotide sequence of Zygosaccharomyces bailii virus Z: Evidence for +1 programmed ribosomal frameshifting and for assignment to family Amalgaviridae. Virus Res. 217, 115–124. https://doi.org/10.1016/j.virusres.2016.02.008

Edgar RC (2004): MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Firth AE, Jagger BW, Wise HM, Nelson CC, Parsawar K, Wills NM, Napthine S, Taubenberger JK, Digard P, Atkins JF (2012): Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. Open Biol. 2, 120109. https://doi.org/10.1098/rsob.120109

Goh CJ, Park D, Kim H, Sebastiani F, Hahn Y (2018): Novel Divavirus (family Betaflexiviridae) and Mitovirus (family Narnaviridae) species identified in basil (Ocimum basilicum). Acta Virol. (in press).

Isogai M, Nakamura T, Ishii K, Watanabe M, Yamagishi N, Yoshikawa N (2011): Histochemical detection of Blueberry latent virus in highbush blueberry plant. J. Gen. Plant Pathol. 77, 304–306. https://doi.org/10.1007/s10327-011-0323-0

Kim DS, Jung JY, Wang Y, Oh HJ, Choi D, Jeon CO, Hahn, Y (2014): Plant RNA virus sequences identified in kimchi by microbial metatranscriptome analysis. J. Microbiol. Biotechnol. 24, 979–986. https://doi.org/10.4014/jmb.1404.04017

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

Kondo H, Hisano S, Chiba S, Maruyama K, Andika IB, Toyoda K, Fujimori F, Suzuki N (2016): Sequence and phylogenetic analyses of novel totivirus-like double-stranded RNAs from field-collected powdery mildew fungi. Virus Res. 213, 353–364. https://doi.org/10.1016/j.virusres.2015.11.015

Krupovic M, Dolja VV, Koonin EV (2015): Plant viruses of the Amalgaviridae family evolved via recombination between viruses with double-stranded and negative-strand RNA genomes. Biol. Direct 10, 12. https://doi.org/10.1186/s13062-015-0047-8

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins, DG (2007): Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Li, H (2011): A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li H, Durbin R (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li N, Li Y, Zheng C, Huang J, Zhang S (2016): Genome-wide comparative analysis of the codon usage patterns in plants. Genes Genom. 38, 723–731 https://doi.org/10.1007/s13258-016-0417-3

Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, Yi X, Jiang D (2012): Discovery of novel dsRNA viral sequences by in silico cloning and implications for viral diversity, host range and evolution. PLoS One 7, e42147. https://doi.org/10.1371/journal.pone.0042147

Liu W, Chen J (2009): A double-stranded RNA as the genome of a potential virus infecting Vicia faba. Virus Genes 39, 126–131. https://doi.org/10.1007/s11262-009-0362-1

Martin RR, Zhou J, Tzanetakis IE (2011): Blueberry latent virus: an amalgam of the Partitiviridae and Totiviridae.

Virus Res. 155, 175–180. https://doi.org/10.1016/j.virusres.2010.09.020

McGuffin LJ, Bryson K, Jones DT (2000): The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405. https://doi.org/10.1093/bioinformatics/16.4.404

Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H (2016): Linking virus genomes with host taxonomy. Viruses 8, 66. https://doi.org/10.3390/v8030066

Nibert ML, Ghabrial SA, Maiss E, Lesker T, Vainio EJ, Jiang D, Suzuki N (2014): Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. Virus Res. 188, 128–141. https://doi.org/10.1016/j.virusres.2014.04.007

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

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.

Park D, Hahn Y (2017a): Genome sequence of Spinach cryptic virus 1, a new member of the genus Alphapartitivirus (family Partitiviridae), identified in spinach. J. Microbiol. Biotechnol. 27, 834–837. https://doi.org/10.4014/jmb.1611.11026

Park D, Hahn Y (2017b): Genome sequences of Spinach deltapartitivirus 1, Spinach amalgavirus 1, and Spinach latent virus identified in spinach transcriptome. J. Microbiol. Biotechnol. 27, 1324–1330. https://doi.org/10.4014/jmb.1703.03043

Pyle JD, Keeling PJ, Nibert ML (2017): Amalga-like virus infecting Antonospora locustae, a microsporidian pathogen of grasshoppers, plus related viruses associated with other arthropods. Virus Res. 233, 95–104. https://doi.org/10.1016/j.virusres.2017.02.015

Sabanadzovic S, Abou Ghanem-Sabanadzovic N, Valverde RA (2010): A novel monopartite dsRNA virus from rhododendron. Arch. Virol. 155, 1859–1863. https://doi.org/10.1007/s00705-010-0770-5

Sabanadzovic S, Valverde RA, Brown JK, Martin RR, Tzanetakis IE (2009): Southern tomato virus: The link between the families Totiviridae and Partitiviridae. Virus Res. 140, 130–137. https://doi.org/10.1016/j.virusres.2008.11.018

Schneider TD, Stephens RM (1990): Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100. https://doi.org/10.1093/nar/18.20.6097

Acta virologica **62**: 122 – 128, 2018 doi:10.4149/av_2018_201

Supplementary information

Identification of a novel plant amalgavirus (*Amalgavirus*, *Amalgaviridae*) genome sequence in *Cistus incanus*

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

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

Received February 2, 2018; accepted March 3, 2018

Supplementary Table S1. Sequence variation of CiRV1

Position	Con[a]	Alt[b]	Quality[c]	Con #	Alt #	Con %	Alt %
254	T	C	103	8	7	53.33	46.67
257	G	A	87	10	6	62.5	37.5
272	C	T	49.474	10	4	71.43	28.57
296	T	C	69	19	6	76	24
347	G	C	168	21	10	67.74	32.26
350	A	G	176	20	10	66.67	33.33
368	G	A	203	22	13	62.86	37.14
398	C	T	222	22	14	61.11	38.89
455	A	G	182	21	12	63.64	36.36
458	G	A	187	19	12	61.29	38.71
542	A	G	142	14	13	51.85	48.15
602	C	T	222	30	25	54.55	45.45
620	T	C	222	36	27	57.14	42.86
656	T	C	222	34	23	59.65	40.35
662	A	G	222	36	24	60	40
681	C	T	222	28	20	58.33	41.67
683	G	A	222	27	20	57.45	42.55
713	T	C	183	21	16	56.76	43.24
736	C	A	177	20	13	60.61	39.39
752	C	T	101	19	10	65.52	34.48
795	T	C	218	25	18	58.14	41.86
815	C	T	222	29	18	61.7	38.3
861	C	A	222	28	17	62.22	37.78
863	A	G	222	26	18	59.09	40.91
910	A	G	222	35	19	64.81	35.19
920	C	T	222	35	22	61.4	38.6
1049	A	G	199	36	16	69.23	30.77
1103	G	A	222	37	22	62.71	37.29
1109	A	G	222	31	20	60.78	39.22
1141	T	C	218	38	17	69.09	30.91
1148	A	G	202	40	17	70.18	29.82

Position	Con[a]	Alt[b]	Quality[c]	Con #	Alt #	Con %	Alt %
1154	T	C	176	35	14	71.43	28.57
1221	A	G	130	25	9	73.53	26.47
1232	C	T	114	23	8	74.19	25.81
1238	A	G	134	25	9	73.53	26.47
1247	G	A	126	25	9	73.53	26.47
1248	T	A	127	24	9	72.73	27.27
1296	C	A	144	30	10	75	25
1548	T	C	93	44	9	83.02	16.98
1584	A	G	96	34	8	80.95	19.05
1599	G	A	103	42	10	80.77	19.23
1601	G	A	109	41	10	80.39	19.61
1662	A	C	147	59	17	77.63	22.37
1678	G	A	171	57	18	76	24
1839	G	T	124	41	11	78.85	21.15
1899	G	A	216	47	19	71.21	28.79
1911	A	T	198	48	18	72.73	27.27
1917	T	C	191	47	18	72.31	27.69
1959	T	C	169	41	13	75.93	24.07
1977	G	A	222	44	19	69.84	30.16
1980	A	G	222	46	19	70.77	29.23
1995	G	A	222	47	22	68.12	31.88
2011	C	T	222	40	20	66.67	33.33
2031	T	C	222	49	27	64.47	35.53
2056	A	G	210	71	25	73.96	26.04
2118	G	A	77	76	13	85.39	14.61
2139	G	A	71	69	12	85.19	14.81
2148	T	C	96	56	11	83.58	16.42
2187	A	G	42.5884	43	9	82.69	17.31
2229	C	G	184	34	13	72.34	27.66
2232	C	T	189	32	13	71.11	28.89
2244	A	G	202	36	15	70.59	29.41
2265	C	T	197	39	15	72.22	27.78
2307	G	A	221	55	21	72.37	27.63
2312	A	G	206	62	21	74.7	25.3
2382	T	C	204	51	18	73.91	26.09
2514	G	A	222	46	21	68.66	31.34
2562	T	C	222	42	21	66.67	33.33
2676	C	T	218	66	23	74.16	25.84
2763	C	T	222	61	32	65.59	34.41
2840	C	G	219	51	19	72.86	27.14
2892	A	G	222	24	20	54.55	45.45
2935	G	A	222	25	17	59.52	40.48
3030	T	C	149	19	21	47.5	52.5
3060	C	G	126	35	20	63.64	36.36

[a]Sequence of the assembled contig. [b]Alternative sequence observed in RNA-seq reads. [c]Phred-scaled quality score for the assertion made in Alt by BCFtools

Supplementary Fig. S1. Multiple sequence alignment of ORF1+2 fusion proteins. The first amino acid of ORF2 part is highlighted in red

```
CiRV1  GTRIKPPVARNLELMFDPLSLLATCKRVRLFGELGQVPKRAIATARSKWEAGVRHIIGGGELRTWRADNNMYRGGGNLHDAILLLGDADDTSDFIFLSSRLSLADARGILHLPVDLPVPDG   459
FpAV2  GARPGPPLPKAVEAISGNSLLSTPNRRIQIHRRSGQPGQAGIPTARSRWEAGVRHIIGGGELLEWRVDNNKYRGGGNLHDALLLIGSADDVTPYTALSSLLSVEEARRVLLLPDGLAVPDG   459
FpAV3  GGRPGPDLQQAIESCVTSFGLLRANKRLRVHRERSGLRQISIPHARSKWEAGIRHIIGGGEILNFRADNCKYRGGGNLFDALTLLARADDTTEYSTLSVHFSVEQARHVLRLPSGLPVPDG   459
LpAV1  GQRPGPDLRQALESCVSSFALLRKNKRLRVSRERSGLRQVSIPHARSKWEAGIRHVIGGGEILNFRADNCKYRGGGNLFDALTLLARADDTTEYSTLSVHFTVEQARHVLRLPSGLPVPDG   459
BLV    GHREKPPLGIRTELAQPPLKALEDCSRIGVLRSLAPQADKLIPRSRSKWEAGVRKIIGGGELKDWFKAKSMYRGGGNLFDALRLLSACDDYDGYTRLKGNYSVEAARREKLLLPSGLRVPDG   457
ZmAV1  GTKPTPPIQRNLEMTLLPMNLLKTNLRVQLLTEHVRGGVRQISIARSKWEAGVRRIIGGGEMRGWEKDSALYRGGGNLHDAIRLLATGRQDPPGSFLFEHFSLETAREILLLPCNLEVPDG   456
ZmAV2  GSKVKPPLPRKLEMLILPMNLLMANPRVNLLHAFVHASGRGLKIARSKWEAGVRRIVGGGEMRGWNEDSAKYRGGGNLHDAIRVLATGRHDLPGSFLFQHFNFRSARSILKLPSDLSVPDG   469
AcAV1  GMRKKPPLQKANESRSSKIDLLLRNPRILVLRKPERGPLGGIPTARSKFEAMVRKVIGGGEMLNWSIDSNMYRGGGNFTDALTLLADARYDAPEMFLSDYLSIEKARSLLCLPSDLKVPCH   465
AcAV2  GIKNKPPLARNAEVSASVTSLLDINARILVLRSIERSPTVGIPTARSRFEGAVRKVIGGGEMIDWRTVSNQYRGGGCFSDAILLLADARTDEPGKFLPDYFTLWKARDILRLPSDLKVPCN   464
PeAV1  GSKQPAPLSSEIERAVLIPPLLWQNQRLSILRPLRPDLGGRIPHSRSRYEAKVRKVIGGGEMRDWRAANAMYRGGGSFSDALKLLIDAREDAPGAILSEKWKVDSARRYLLLPCGLPVPRG   458
ScAV1  GVRPSQDLHPDLELKRPKPTLEE-NMRLRVVRSGPGEDRRRIPVSRSGYEGAVRKVIGGGALRSWKQDQAMYRGGGNNVDALLLMSQASEKRPGAFLRDRYSVLSARRALGLPSDLQVPDG   474
SpAV1  GVKLQQGISRKVELTSLGPKMLR-TSRIRVLRSLGPERPRAIPVARSRFEAGVRKVIGGGAMRSWEVDSQMYRGGGNSADALRLLGQARDDRPGAFLSGKFTQASARLALLLPNNLDVPDG   461
EbAV1  GTKKTNPLLRAFATVQLGAEVLR-STRVLVHHTFGPETQRRIPVARSRIEAGLRRIIGGGAMRGWDADSKMFRGGGNSSDALLLGQCDDNLPGGLLREHFSLLSAKRALSLPGGLRVPDG   456
EbAV2  GHRPIQVLARGIRAQELDSSIMR-NPRLKVYVNFQGEPGRVIPHSRSKYEAALRRIIGGGAMRSWGEDSKMYRGGGTSSDALLLLSQADFRLPGGLLKEHFSLRTAREALCLPEDLVVPDG   460
CoAV1  GIKVQRGLRRQIEVVEGVASEEFRLVRGKVYVEWAAEKARLLPVARSKFETGVRRVIGGGEMRNWHVASSMYRGGGNNADALRLLSNASFMNPEKYLHECFTIETARRELLLTTDQKVPDG   472
GaAV1  GLKGRPELGRGFELQGGLGSEALLTGRIKVLVHFEPERVRSMAVARSKFEAGVRKIIGGGEMRGWRSASSMYRGGGNSNDALRLLSQAKDDFPGRFLTDVFKVDMAREALCLESDLAVPDG   477
CdAV1  GIRHRPGLQRKIEEAGGGSRGAEMTGRVRVLVEWGRERHRTIPVARGKFEAGIRKIIGGGELLNWREDSSMYRGGGNSADAIRMLFSADPIPPGRVLTDFWREGSARQALFIEELFEVPDG   476
MsAV1  GYKSISRLERKHELQGCITSQGMPEGRIKIFKSWESEPGRMIPMSRSKYECAVRWVIGGGEVRNWKVDSSMYRGGGSSNDALKLLANASTVRPGKLLRDVYSFRVARARLQLPGDFSVPDG   466
VCV-M  GIKLTRGLQRNVEMKGTLPSQSLRGRRIRILRPWVTDHNRMIPTSMSRYEGAVRWVIGGGETRRWKEASTMYRGGGCSNDALKLLANVSTVLPGRLLRDLYSLRVARARLSLPADFCVPDG   466
RHV-A  GVKGNARLSRNVEVSAQVRLGELQTYRFRVHREMVEHRVRSIAVARSRFEAGIRRIIGGGEMRSWYTDSRMYRGGGNSNDAMKLLSAASVVTPGKFLDQCYRVSSARKALRLPSGLRVPDS   480
AoAV1  GTGPPPRLGAGVPAAVGTTPVLQGGARLHVIRAPVRPDRRGIPYARSKWEAGVRKIIGGGELSDWQAASSKVRCGGNASDALLLLADASDRLPGKTLRGLFTLRGARDALRLPSGLSVPDG   466
FpAV1  PSSAPSRLGASVPRAQGVAPLLQGGARLRVLRAPVAPDTRGIPFARSKWEAGVRKIIGGGEMNDWLQVSSKERCGGNASDALLMLADASDRLPGRILRGLFTLRGAREALRLPSGLSVPDG   466
CaAV1  SYRPAPQLSKMLECKAESLALLRANGRVEIIIMPGGCDLRAIPCARSRWEKGIRKIIGGGEMLNWKVAGNMYRGGGCYSDAIKLLSSASTIPPGRLLHQCFSLPGARAALGLPSNLSVPVG   468
STV    QYKLPPQLSRATEVEAEAVSLLYPNARIEVEVRPIGRDTRAIPFSRSRWEVGVRKVIGGGEVLNWDIDGNKYRGGGCFADAIKLLAGATQRPPQRLLEDCYTITTAREALGLPSDLAVPSG   468
PpAV1  GMRKVGKISPQMHNLVPSLSLFQGNGRLDVLRDVGSGKHRGIPTAMSMFEKGVRKVIGGGEMRDWNRASNFVRGGGDLGDALKFFSSCKTSPQQRFLCDVYSLEVAREILDLPTGLPVPDG   449
                :                  *  :          :  : .  *  :* ::***    :       * **   **: .:          *        *:  * :     **

CiRV1  PKACVMKYFNEDATGGPLLRSGGVRRKYGLKSVIEEFVWNIYDAVGYGELTVDQLPGMTSRVGYRSKLMTTDKALDKITRAKPLGRAVMMLDATEQSFSSPIFNVISGIVSQLNRDPSTGW   580
FpAV2  KACCVMKQFNDDATAGPLLRAFGVKGKYGLKAAIEQFVWGLYDRVGSEDLKPRQLPGLLARVGYRTKLLDMEKALKKIQAVEPLGRAVMMLDATEQCFSSPLFNAISEAVTELHSNPRSGW   580
FpAV3  AQCCFMKQFNDDASAGPLLRAFGVKNKYGLKSMVESFVWGMYDRVGSGDLTPDQLPCLLARLGFRTKLVDKDKAAKKIFDVEPVGRAVMMLDVTEQAFSSPLFNAVSEQVTLLHNDPRSGW   580
LpAV1  PQCCFMKQFNDDASAGPLLRAFGVRNKYGLKSIIEFFVWGMYDRVGAGTLNPEQLPCLLARLGFRTKLVDKDKAAKKIFDVEPVGRAVMMLDATEQAFSSPLFNAISEQVTFLHSDPRSGW   580
BLV    VGCVHMKNFNDEASAGPLCRAFGIRRKAGLKSRLEQFAWEIYDGIGNGDLDLGAIPPFLARVGYRTKLVTGEKAMRKISTGDSIGRAIMMLDAYEQAFSSPIFNVISDAVTHLHSDPTSGW   578
ZmAV1  PKAVRMKNFNEDATAGPVLRALRCKSKYGLKEGLERIAWDLYDRVGDGKLRRWQLPPLLARIGFRSKLVEQESAIKKIFSGQPIGRAVMMLDAMEQPFSSPLYNALCDVVSKLNRVPESGW   577
ZmAV2  RDSVTMRNFNEEATAGPVLRAIGCKTKYGLKRGLEDIVWNLYDRVGNGELRTWELPPLLARIGFRSKLIDQDKAEEKILTGQPLGRAVMMLDASEQAFSSPLYNVLSGVVSRLNCERRSGW   590
AcAV1  RNCVSVNNFNNEATAGPFFRAHGIRNKYGMRLQLEDFAWECYNAYVDSGGNPSMLPYISSRVGFRTKLVSTTEAFAKMKDNKPIGRCVMMLDAIEQMFSTPLYNVLSKTTAAQRFDIKSGF   586
AcAV2  RQALKVSNFNNDATAGPFFRAYGIKSKYGMRGLLEDFAWECYSSFVDNGGDVSFLPFVASRVGFRTKLVSQEEAFIRFSKNKAIGRCVMMLDAIEQMFSSPLYNVLSKLTADLRFDPASGF   585
PeAV1  PEATKMKNFNDDATAGPALRAFGILRKSGLKTSLEEFAWNCLDAFARGGAAEDCLPFVAARVGYRTKLLTLSDAWSKINSCKPLGRCVMMLDAHEQAFSSPLYNVLSNLTHLSFRQRNSGF   579
ScAV1  PAATKMKNFNNDATAGPFLKWCGVKSKRGLKCLLEEEMWGYYDAYAKGEIEDHQLPFLTARLGFRTKLLKKAEAMRRIGEGKAMGRAVMMMDALEQAASSPLYNAVSHYTFERRLEKDCGF   595
SpAV1  AKSTRMKNFNNDATAGPFLRSFGIKGKYGLKRKLEEEMWRYYDDYGAGRIDSSGLPFFTARVGFRTKLVSVEKAEEKFKTGQPFGRAVMMLDALEQAAASPLYNVLSHYTFERRLRRDCGF   582
EbAV1  PECLVMKNFNNDATAGPFLRAFGIKGKYGLKKLLEDTMWWFYDAYGRGEISDGEMPHFAARVGFRTKLVSETKAWEKLAAGAPVGRAVMMLDALEQAASSPLYNVMSNSTYQRRLERDCGF   577
EbAV2  KDCCRMKNFNNEATAGPFLRLFGVKGKYGLKKLLEDEMWRYYDDFAQGRIDERGLPYFAARLGFRTKLMSEKKAWEKMQKGDPYGRAVMMLDALEQAASSPLYNVLTGVTFERRLEKECGF   581
CoAV1  PESVKMKNFNNDATAGPFLRAFGIKGKYGLKTLLEQVMWQFYDYYANNEVSDATLPFLTARVGFRTKLMTERKALEKMMVNDTYGRAVMMLDALEQAASSPLYNILSSQCFHGRLRRESGF   593
GaAV1  FGCCSTKNFNNEATAGPFLRAFGVKVKHGLKTYLEQFMWGLYDRYGDGEINQKGLPHLTTRIGFRTKLVTREEALRKVQQGTTFGRAVMMLDALEQVASSPLYNVLSHKTFLMRNEPGSGF   598
CdAV1  PGCCVMKNFNNEATSGPYLRGQGLKGKYGLKRILEAEMWRYYDSFASGECGLKNLPWFAARVGFRTKLLTQEVAWAKMQKGEALGRAVMMMDALEQAASSPMYNVLSSYTFRRRLEPRCGF   597
MsAV1  PDACRVKNFNNFATSGPVLKAFGVRSKNGLRQLLQDEAWWYFNSFGNSDFGVEGLPWFGARLGFRSKLVTEEKARKKISEGDSVGRAVMMMDALEQCCSSPLYNVLSTYTFHKRLNRRSGF   587
VCV-M  PEACRVKNFNNFATAGPFLRRFGVKKKAGLRKLLEGEAWWYFNSFADGDLDVSGLPFFGARLGFRSKLVSEEKARKKISEGDSVGRAVMMMDALEQCCSSPLYNILSSYTYRRRLERECGF   587
RHV-A  LAECQCKTFNDQASAGPFLRAFGIKKKYGLRRMLEEFMWGIYDRYGDGRGDERSLPFITGRIGFRTKLLEEGEALRKIRNGEPLGRAVMMLDALEQISSSPLYNILSGYCGRNRNKGWCEF   601
AoAV1  RACCNIKHFNVEATAGPFLRAFGVKKKESLEQLLGDFVWECFDDFATNDADIRRLPFFGMRIGFRTKLLKKSEMLTKIRDFKPLGRCVMMLDAIEQFCSSPLYNVLSKLSADALKDPMSGF   587
FpAV1  RACCNVKHFNVEATAGPFLRAFGVKKKAALEQLLGDFVWDCFDDFASNDGDARRLPFFGMRIGFRTKLLPRSEMLTKIKDFKPLGRCVMMLDAIEQFCSSPLYNILSNLSADALKDPLSGF   587
CaAV1  KGSCRVKNYNDEATTGPFLYSFGIKKKYGLDNELQEIMENAYHHFAECKTSSSALPFFTARVGFRSKLLPMGEALRKFADNQPMGRCVMMMDALEQFASSPLYNVLSKYTSD-RSKGATSF   588
STV    KGSCHVRNYNDDATSGPFLWSFGIKKKYGLDKLLQSLMEDIYCHYSISEATDRALPYFAARVGFRSKLLTMGEAVKKFTECAPMGRCVMMLDALEQFASAPLYNVLSKYTAE-RSRGQTSF   588
PpAV1  PEACRIKNYNDEATAGPWLRAFGVRRKAGLKSSLESLMWSFYDAVGDGKLLPEDLPYLSARVGFRTKLLAREAAMEKLGKGEPMGRAVVMLDALEQAASSPLYNVMSGLAAQNHKKERGVF   570
              :*  *: **        * .:   :                    :* .  *:*:*:**:      :.    . **.::*:*. **  ::*::* :            :

CiRV1  RNYLVRASASWIDFWDELKGAGTIVELDWKKFDRERPREDILFFIDIIISCFTAKTKREERLLAAYKRMMQRCLVDRVLMMDNGAYFTIDGMVPSGSLWTGICDTALNIMYITAALRSLGF   701
FpAV2  RNYLVRASSAWADLWDELKSCGTIVELDWSKFDRDRPAEDIQFFVDVIISCFRPKNSRERRLLAGYRKMMENALVHRVMMLDDGSFFTLEGMVPSGSLWTGICDTALNIMYITSALRKLGF   701
FpAV3  RNYLVRASVAWVEFWHELRDAKVIVELDWAKFDRERPAEDIQFFIEVICSCFQPRTAREERLLAGYKKMMENALVHRLIVLDNGCFLKVDGMVPSGSLWTGICDTSLNILYITAALMSLGH   701
LpAV1  RNYLVRASVAWVEFWHELKDAKVIVELDWAKFDRERPAEDIQFFVDVICSCFQPKTAREENLLAGYKQMMENALVHRLIVLDNGCILKIDGMVPSGSLWTGICDTALNILYISAALISLGH   701
BLV    RNYLIRASSDWGKMYKEMRKAKVVVELDWKKFDRERPADHISFFIDVILSCFAPRTTREEKLLGAYRRAMENALLHRVLVTDDGGVLTVDGMVPSGSLWTGIIGTGLNILYIGYALSDIGI   699
ZmAV1  RNMVVRASSDWSRFWGDIRNSCAIVELDWSKFDRERPKEDIEFVINIFVSCFLPKNRRERRLLSAYKYMMRKALIDKVIMLDDGCAFTYDGMIPSGSLWTGLLGTALNILYINAAVRRLGI   698
ZmAV2  RNAVVRASSDWSRFWKDIRDAKCIVELDWSKFDRERPRADIQFCIDVLCSCFRPRGKRQRRLLSAYKYMLENALIHKVIMLDDGCAFEYEGMIPSGSLWTGILGTAMNILYITAACESCGV   711
AcAV1  RNTIVRASSDWAKFWEEVKAAKVIVELDWKKFDRERPTEDLEFIIQVIISCFKPTDERERLFLRGYGIMLRRCLIDRYFITDDGGVFKIDGMVPSGSLWTGWVDTALNILYLQSVMISLGL   707
AcAV2  RNTIVRASSDWAKMWEEVKKAQVIVELDWKKFDRERPTEDLAFMIDVIISCFKPENDRERLFLLAYKTMMQRCLIERFFVTDCGGVFKVEGMVPSGSLWTGWVDTALNILYLNAALLSLGF   706
PeAV1  CNSIVRASSDWAMLWKDVSAASCVVELDWKKFDRERPADDISFMIDVIISCFEAKDDYEERLLLGYRIMLNRALIERCFVTDDGGVFHIDGMVPSGSLWTGWLDTALNILYIGAALRHVLP   700
ScAV1  KNTIIRASSDWQAIWAHVKEAEAIVELDWGKFDRERPSQDLNFIVDVVVSCFAPKNSRERRLLRAYKLMMRAALVDRLLVLDDGTVFGIEGMVPSGSLWTGWVDTALNILYLKAACLEINI   716
SpAV1  KNAVIRASSDWNEIWKGVREAEAIIELDWSKFDRERPAEDLEFIVDVVISCFKPRNERERRLLRAYGIMMRRALIERLLVMDGGGVFGIEGMVPSGSLWTGWIDTALNVLYLRSACLEVGL   703
EbAV1  KNGIVKASSDWAKIWEDVREAKAIIELDWSKFDRERPADDILFVIEVVLSCFEPTNDRERRLLRAFGLMMRRALVERIIVMDDGGVFEIDGMVPSGSLWTGWLDTALNVLYLNAACMNVGI   698
EbAV2  KNQVIRASSDWNRVWEYLRDSKVIIELDWAKFDRERPSEDLQFVIDVVLSCFAPRTPRERKLLEAYGIVMRRALIERAVVMDRGGVFTIEGMVPSGSLWTGWLDTALNILYLNAACVEAGY   702
CoAV1  RNALIRASSDWHFMWEEVQEAAVIVELDWSKFDRERPAEDIEFIINVVLSCFTPKTPREERLLRAYGIMMRRALIERLMITDNAGVFGISGMVPSGSLWTGWLDTALNILYIRAACREIGV   714
GaAV1  RNATIRASSDWGKMWEEVRQAATIVELDWSKFDRERPREDLLFIIEVILSCFLPKNRREKRLLEAYGIMLRRALVERVIVMDEGGVFTIDGMVPSGSLWTGWIDTALNILYILAACREIGV   719
CdAV1  KNGVVRASSDWGVLWRRVQQAKVIVELDWSKFDRERPREDLEFIVKVVCSCFRPRDDRGRRLLQAYQEMMRRALVERALITDSGGVFLVDGMVPSGSLWTGWIDTALNILYIKAICVELGI   718
MsAV1  KNAVVKASSDWAHIWNGVKDAAVIVELDWSKFDRERPREDLEFMVSLISSCFNPKSAREERLLHAYTTSNFRALVERPVFLDGGGVFGIEGMVPSGSLWTGWLDTALNILYMKAVCAEIGV   708
VCV-M  KNAVVKASSDWAHVWGGVKEAAVVVELDWSKFDRERPREDLEFMVNVISSCFAPEGIRQRRLLEAYTVSNQRALVERPVMLDGGGAFCISGMVPSGSLWTGWLDTALNILYIKAVCAEIGI   708
RHV-A  RNTTVRASSDWGLLYSEVERAKCIVELDWSKFDRERPSRDILFVIKVIISCFEPRNGRERRLLEAYEIMLERALVHRLLLTDAGGVLEMEGMVPSGSLWTGWLDTAMNILYIRAACLGAGI   722
AoAV1  RNTAVRASSDWSYMWEEIKEAKVCMELDWSKFDRERPRCDLEFMVDVVISCFAPKNEREQRLLRGYEVCMRRAIVERVALLDDGALFEIDGMVPSGSLWTGWLDTALNILYLSAALSEAGF   708
FpAV1  RNTAVRASSDWSYMWEEIKQAKVCMELDWSKFDRERPRDDLEFIVDLVISCFTPRTEREQRLLRAYEICMKRAIVERVAILDDGALFTIDGMVPSGSLWTGWIDTALNILYLSAALNQAGF   708
CaAV1  RNSVVRASSDWMYLWDEVKEASVCVELDWSKFDRERPSADLLFMIDVIISCFEPSNRYEVRMLEAYGICMRRALVERVLITDDGGVFEIEGMVPSGSLWTGWLDTALNILYLNAALRHLDI   709
STV    RNTVVRASSDWMHFWDEVKEAAVCVELDWSKFDRERPSEDLDFMIKVICSCFRPKDEVEAKLLRGYGVCMRRALVERRLITDDGGVIHIDGMVPSGSLWTGWLDTALNILYIKSVLRSIDI   709
PpAV1  RNYVVRASSQWRQLWDEVSSCKVLIELDWKKFDRERPPEDLLFMIDLVCSCFEPKSLREERLLAGYKVCMVRALMDRSFVLDSGSVFLVRGMVPSGSLWTGWLDTGLNALYLTHVFQDLGI   691
 *  ::**  *  .:  :  .   :**** ****:**  .: * :.:. *** .       :* .:      .::.:  . * .  :   **:********  .*.:* :*:
```

```
CiRV1  DDSQFSPKCAGDDNLTVFGKKQGRARMERLRVRLNEWFRAGIETEDFIVHYPPYYVFTEQACFPPGTHLALGTSELLDMAEWVRFDGPCPIDEAAGRSHRWRYNFNAKPKFLANYFLRDGR   822
FpAV2  DDSTFSPKCAGDDNLTCFQRRQGLDVMERLRTTLNEMFRANIDKKDFIVHYPPYAVTTVQACFEPGTDLSKGTSRMMDKAEWVPFTGPCPIDQAAGRSHRWKYVFTNKPKFLANFFLPDGK   822
FpAV3  DITSFVPKCAGDDNLTTFDRRIRKKDLEKLRLRLNSLFRAGIKEEDFIIHYPPYHVTTVQACFPPGTDLSHGTSKMLDQATWVPFEGPCDINQEEGRSHRWKYQFEGKPKFLANFFLIDGR   822
LpAV1  DITSFVPKCAGDDNLTTFDRRIRKKDLEKLRLRLNSLFRAGIKEEDFIVHYPPYHVTTVQACFPPGTDLSHGTSKMLDQATWMPFEGPCDINQEEGRSHRWKYQFEGKPKFLANFFLIDGR   822
BLV    SPLNYVPKCAGDDNLTFFSRDYGDAAFKRLRVKLNEMFRANIDEEDFIIHRPPFFVTKAQAVFPPGTDLSKGTSSILNQCKWVQFEGEIIIDEAMGLSHRWEYRFQGKPKFLANYWLPDGR   820
ZmAV1  PSDEFVPYCAGDDNLTVFKYPQRATRLGKIRGMLNEMFRAGIDPEDFIIHYPPFHVTKVQAKFPAGFDFTHGTSKFLDQCQWIPLKGNIHVDHSEGLSHRWNYVFKGKPKFLANYFMEDGR   819
ZmAV2  KRETFVPFCAGDDNLTVFDSAVDEGMLLNIRVFLNNMFRAGIEEKDFLIHYPPYHVTKVQAVFPDDFDLSLGTSKYLDKCVWVPFEGPLLVDTSRGYSHRWNYVFKGKPKFLANYFLEDGR   832
AcAV1  VSTAVSPKCAGDDNLTLFWKDFDDKRLLSIKTRLNAWFRAGIDDEDFFIHRPPYHVTREQATFPPGTDLTKGTSRKLKDAIWIPIDGEPIIDQAQGLSHRWQYVFRGKPKFLSAYWLEDGR   828
AcAV2  NMTEVFPKCAGDDNLTLFMRDVENRRLLALKDRLNSWFRAGIEAEDFLIHRPPFFVTREQAVFPRGTDLTKGTSKIIGNAVWVPIEGEMVIDQEAGRSHRWQYVFRGRPKFLSAYWLEDGR   827
PeAV1  DTSQAVAKCAGDDNLTLFYTDLPDASLLNLKKYLNEWFRAGIEDEDFIIHRPPYHITRFQATFPPGTDLSKGTSHLLDSAKWIQIHGIMNIDEAAGLSHRWKYSFAGKPKFLSCYWEENGN   821
ScAV1  PSSQYLPMCAGDDNLTLFWKDPGP-ILARLRSILNDLFRANIDAGEFKIHYPPFHVVKKQACFPPGTDLSKGTSKIMHKAFWEEFVGELHVNEDLGKSHRWEYAFEHRPKFLSFYWLPEGQ   836
SpAV1  PSQRYLPMCAGDDNLTLFWSDPGEVKLLKIRELLNEWYRAGIDDADFFIHRPPYHVVKRQACFPPGTDLKGGTSMLMKDAEWVEFEGELRVDEAAGRSHRWQYLFKGRPKFLSCYWTRDGL   824
EbAV1  GPLGFSAMCAGDDNLTLFWTDHPDHVLKRIKDELNGKFRAGISDEDFFIHRPPFHVTKQQACFPPGTDLSHGTSKLMDLVFWQVFDGEVVIDEAAGRSHRWEYVFKGRPKFLSCYWLPGGQ   819
EbAV2  GPGFFHPMCAGDDNLTLFDLDLGDRRLLKIREVLNNWFRAGISEEDFFIHRPPFHVIKKQACFPPGTDLKIGTSKRMHEAFWVEFDGELVIDEAAGRSHRWEYIFKGRPKFLSNYWLPEGQ   823
CoAV1  SDQNSFPKCAGDDNLTLFMRDPGDRALNRMKDILNKWFRAGI-KEDFFIHRPPYHVSKFQACFPPGTDLSEGTSKLIDEAEWIEFHDELEVDVSRGKSHRWEYRFKGKPKFLSCYWLENGR   834
GaAV1  PSTFCSAKCAGDDNLTLFALDPGDGALRRLRVVLNEWFRAGIDEEEFLVHRPPYHVKKVQACFPEGVDISKGTSKLLDKARWEEFEGELRVDVAAGRSHRWEYRFKGCPKFLSCYWLRDGK   840
CdAV1  ALSAVEVFCAGDDNLTLFMYDPGDVILKEFRTLLNQYFRAGIKEEDFFIHRPPFDVRKYQATFPVGSDLRGGTSRMLDRATWIEFEGEIEVDEARGRSHRWEYRFKGKPKFLSNYWLPDGR   839
MsAV1  GDDDVEVMCAGDDNLTLFKFDPGEANLRRIRDLLNDWFLAGISEDEFLIHRPPYHVTKVQACFPEGTDLSRGTSGMLKKAVWVPFEGEVVIDNARGRSHRWEYRFSGCPKFLSCYWLHDGL   829
VCV-M  GAEDIEVLCAGDDNLTLFKYDPGDSRLQTMKNLLNDWFRAGISDEDFIVHRPPFHVTKVQACFAPGTDLSHGTSKILKDAAWIPFEGELVRNDHTGRSHRWEYRFSGCPKFLSSFWLSDGL   829
RHV-A  ASDLYSSRCAGDDNLTLFWEDQSDGILGLVKQHLNNWFRAGIKDEDFIICRGPYYVERYQATFPIGTDLSEGTSRLMDSAIWVRLEGGPIIDANRGLSHRWEYRFHGKPKFLSCYWLPDGR   843
AoAV1  SSYMARPKCAGDDNLTLFMEDVPDGRLIQVRGLLNDWFLAGIKDEDFAITRAPFHVETYQAVFPPGTDLTKGTSKIIDQCEWVRFEGQLQINQEQGLSHRWQYRFKGKPKFLSCYWLADGR   829
FpAV1  SLDTARPKCAGDDNLTLFLEDVTDSRLFYVRDLLNEWFLAGIKEKDFNITRPPFYVETYQAVFPPGTDLSLGTSKIVDQCEWVRFEGQLYINQEEGHSHRWQYRFKGKPKFLSCYWLQDGR   829
CaAV1  APSSASPKCAGDDNLTLFYRDPGDEVLLRLKVVLNEWFRAGIKDEDFVITRPPYHVRTYQAVFPSGLDLSKGTSKIIHKAYWREFEDEVRVDMERGLSHRWEYRFKGCPKFLSCYWLEDGR   830
STV    LEEEAVPKCAGDDNLTVFSKDPGDEVLEEMRVKLNDYFRAGIKKEDFIITRPPFHVRTFQAVFKEGTDLSKGTSKIMKKAYWREFEDELRIDQEKGLSHRWEYRFKGAPKFLSCYWLEDGR   830
PpAV1  PRSLFCPKCAGDDNLSLFSQDYDDNILKKGRVLLNEYFNPGIEEEEFLIHRPPFHVVTEQAVFPQGLDLSKGTSKIIHQARWVPFDGMVPIDESRGFSHRWEYRFKGRPKFLSCYWLSDGR   812
              *******: *       :   :  **  : ..*   :* :   *: :   ** *  . .:  ***  :    *  : .    :   * ****.* *   ****: ::   *

CiRV1  PIRPAQDNLEKLLWPEGIHKTIEHYEAAVISMAVDNPFNHHNINHLMHRYVIIQQIKELSY-DVDPDIVMQMAAIRAKPGQPVPYPMVAFYRRQEDKVVLEEVPELSHVIKRFQKFVASVS   942
FpAV2  PIRPAHDNLEKLLWPEGIHESIEDYMAAVLSMVVDNPFNHHNVNHMMHRYLIAQQIQKQSI-FIDPAVVMELAKTRPKHGEAVPFPEIAYYRKCEGYVDIEREPEFQEFFKSFREFISSVS   942
FpAV3  PIRPAHDNLEKLLWPEGIHGTLEDYQAAVLAMVVDNPFNHHNVNHMMHRHLIAAQISRQAF-DVDPAIVMELCTSRAEPGELVPYPEIAFYRRVEGYVDLDAVPEFKEILDDFRLFVSSVS   942
LpAV1  PIRPAHDNLEKLLWPEGIHGTLEDYQAAVLAMVVDNPFNHHNVNHMMHRHLISKQISRQAF-DVDPAIVMELCTSKGEPGELIPYPEIAFYRRVDGYVDLDAVPEFKEILDNFRLFVSSVS   942
BLV    SIRPATDSLERLLYPEGLHQTIEDYEMAVLAMVVDNPWNSHNVNHMMHRFCIVQQIKRQSMVGIKAEDVMWYSKLRGKEGEGIPYPQVAYWRRQEKKVFMEEIPELMGYIKKFQTFLSGVS   941
ZmAV1  SIRPAHDNLEKLLYPENVQQKIEDYEAAVLSMAVDNPWNSHNINHLMQRFCIIQQIKRQAVHPLTAGDVLFCAKFQSISENMWMFPTVGAWRRQLLNVRMEDIEELKPTVEAFSEFVRGVT   940
ZmAV2  SIRPAHDNMEKLLFPENIQEHIEDYEAAILSMVVDNPWNSHNVNHLMQRYVIVQQIKRQSIPPLNAAEVLFCSKFREIADNMYLFPTVGYWRRQLPGVRMESIPELRPIVEEFSQFVSGVT   953
AcAV1  SIRPTHVNAEKLLFPEGIHKDIEQEYESAVLSMIVDNPYNHHNVNHCMHRFIICEQVKRQSAAGIDPVDILWFSRIRPGNPDVVPYPMVASWRRCEGYVDLEQLPFISDYVKDLKEFVAGVT   949
AcAV2  SIRPTHINSEKLLFPEGIHDSIDTYEAAVLSMVVDNPFNHHNINHCMHRFVICEQVKRQARMGLDPIDILWLSRFRSSVGEEVPYPMIASWRRQDTWVDMEQLPFVRDYVRDFREFVSGVT   948
PeAV1  SIRPAHVNLEKLLWPEGIHKTIDDYLAAVISMVVDNPFNHHNVNHMMHRFCIIQQVKRLSVAGIRDDHILSLAHIRGKEGEMVPFPMVAEWRRTQGWVDMESMPYLKKYISDFRHFAAGVM   942
ScAV1  SIRPTRDNLEKLLWPEGIHKSLDDYEAAVASMVVDNPWNHHNVNHLLMRYVIIQQIRSLAATDVKVLDLLWFSKFRPVGDEEVPCPMVAPWRRRSPHARMEDYPEVQRWVRDFKDFVAGVT   957
SpAV1  SIRPTSDNLEKLLWPEGIHKSIDDYEASVASMAVDNPFNHHNVNHMLMRYVIIQQIRRVSAGILSPEECLAFCKFRSQEGEEIPYPMVAPWRRGMHEARMEDYPETKPWVKEFRDFISGVS   945
EbAV1  SIRPTSDNMEKLLWPEGIHEDIDDYQATVMAMVVDNPWNHHCVNHLLMRYVILQQLRRVDILRGGMDDVLFLCACREKGGGPIPYPMVAPWRRSEVHGRMEDYEEVKRHIQDFSDFVTGVT   940
EbAV2  SIRPTRDNLEKLLWPEGIHKDLNEYQAALMAMVVDNPWNHHAVNHLLMRYVIVQQLRRVNSGLGREDDTLFLAGLRDMTGGVIPFPMVAPWRRGAQQGRMEDYKEVQNWCLDFQSFVTGVT   944
CoAV1  SIRPAADNLEKLLWPEGVHDKLEIYEAAVASMIVDNPWNHHNVNHLLMRYIIIQQIKRLGVGPANVEDVLFLSQFR-NEGEEIPFPQVAPWRKFKEHVVMEEYNEVMEHVENFKNFLSGVT   954
GaAV1  SIRPAADNLQKLLWPEGIHDSLDVYEAAIASMVVDNPWNHHNVNHLMSRYVIIQQVRRFSAGIVPHEMCVWLSKFRGNAGEPVPYPMIAPWRRMDTHQQLEAYPEAVVEMEVFRDFVQGVT   961
CdAV1  PVRPTRDNLEKLLWPEGVHATLEDYEGAVISMVVDNPFNHHNVNHMLMRYVIIQQLRRMSVAVGELDLLLKLGKIRAKGDEEVPMPQIAPWRRGELQAKMEDYNDAQRHLAVFEDFMRGVS   960
MsAV1  SIRPAHDNLEKLLWPEGLHDDIDVYEGAVISMVVDNPHNHHNCNHMLSRFIIIREARRLGASVEDPFIPIKHGSIRPVGDEPVPYPELAPWRRAPGGYKLEDYPENAEHIQVFRDFMQGVS   950
VCV-M  SIRPASDNLEKLLWPEGTHSDLNVYEGAVISMVVDNPHNAHNVNHLLSRYIIIQEARRLGAAVEDPMIPLRLARHRPVGDEVVPYPDLAPWRRHPEPFQLEDYPANADHIAIFKDFVQGVA   950
RHV-A  SIRPAHDCLEKLLWPEGIHGDLETYEAAVISMIVDNPFNHHNVNHMLVRYVLIQQIKRQVVSPMTAEDIVYLCKFRDREGGSVPYPMIGPWRRGQKQGRMEDYPEVGRDIKNLKGFVAGIS   964
AoAV1  SIRPASDNVEKLLFPEGIHKSVEDYISACLAMVVDNPFNQHNVNHMKHRYLIAQQIKRILVTGIPDRLVLALARIRPDGEEDIPYPMIAPWRRFKEYIDLDSYEPVQQWIKEFDDFVAGIS   950
FpAV1  SIRPASDNVEKLLFPEGIHKDIDDYIAAVLAMVVDNPFNQHNVNHMKHRYLIAQQIKRVMVSTIPYQLILALSRIRSEAGEDIPFPMIAPWRRFKEWIDLDKYPPVQEWIRDFDDFVAGIS   950
CaAV1  SIRPSSENLEKLLFPEGVHSSIDDYIASVLSMVVDNPFNDHNINHLKHRYLIAMQIKRLSAAGGRCGDIMDLARIRPRHDEEVPVPQIAVWRRVKEYIDLDEYGPTKHYIDEFNSFVTGVT   951
STV    SIRPSHENLEKLLYPEGVHANIDDYIAAVLSMVVDNPFNHHNINHLKHRYIIANQVKRLSAAGGRCEDILRLARIREREDEEVPVPQIAVWRRVKEYVDLDSYEPAKYYIQDFNAFVSGVT   951
PpAV1  SIRPTSDCQERLLFPEGIHKSFDEYLEAVMAMVVDNPFNSHTVNHMMHRFLIAHEMKRQVAGGCSADQVLFYSGMKGSPGEEVPFPSVGFWRRREEFIPLEEAYPESKWLQDFLEFAHGVS   933
       .:**:    ::**:**. :  .: *  :  :* **** * *  **   *. :  :              :    :         * :. :*:     ::          :  *  .:

CiRV1  TLYTRKAEGGIDSWKFMDIIRGEHSIGTGQFGNDVLVWCKFLSENRLTRSLRAARRFRVQAEGKVAEDDVLWKVGLIFDEVFNFCKSESVIS-PIHYADFVSKRLKEYNSK-------    1052
FpAV2  TLYARRSEGGIDAWRFMDIIRGQHSIGAGQFGNSVHEWCRFLGEHPLTRSLRKARRFRPMKEAVVADPETLGKVEKAFTWVLDLCEANDPMT-PLFYASAVSDILSSNVSAP------    1053
FpAV3  TLYARRTEGGIDSWRFMEMIRGEHSIGEGQFGNDIYEWCKFLGSNPLTRSLRATRRFKMKAPATVADEGTIRKVQEAFTWLTSICEENLIVT-PMYLAQLISDKLLL-----------    1048
LpAV1  TLYARRTEGGIDSWRFMEMIRGEHSIGEGQFGNDIYEWCKFLGSNPLTRSLRATRRFKMKTSATVVDEPTRKKVQEAFQWLTSICEENLIVT-PMYLAQLISDKLLL-----------    1048
BLV    TLYSRDSAGGIDAWMFMSILRGERDVGAGQFGNDIIEWCKFLNKNPLTSGLRAAKRFRGEVRSEEASTQVAQRWVEMISTVRERLRNPLLGS-PRSFALWVSDMYAENVFSNVS----    1054
ZmAV1  SLYSRASTGGLDSWQFMEILRGERDAGSGQYGSDIEAWCAFLNRNGLTRSLRPAKRYRKEKQTYQSAEAG-DKILVSLAAIYATSDYDLREDRDLGFIIRISNELRKQLL--------    1049
ZmAV2  SLYGRASSGGIDAWQFLDILRGDRDLGSGQYGNDIEKWCAFLNRNALSRSLRPARRFRQGTAPVEPDAKGRESLSR-LALLYEVPSSDRKDSRRLNFISRISVMLKQFVS--------    1062
AcAV1  SLYARDSIGGLDAWRFTEIIRGETDLGTGQFGNDIDVWIHWLHHHPLTKYLKPIKRHRQPPLSVMPDKELHGRIEEALQVYRDIRGTTTNAS-AEAFSLFLSNVIKHQG---------    1057
AcAV2  SLYSRQSTGGLDSWKFTSIIRGETEFGEGQFGNDIDDWISWMYRNPMTKYLRPIRRFRAVGEEVRPDRELSGSLQRCIGIYRQIRGTQENSS-SENFAIFLSNVLRQTRSYRDANAHP    1065
PeAV1  SLYARQPGGGIDAWRFMDMIRGNVLLGGEQFGNDFRDWIKFLQNNPLTSYLKPMRRFRPQKAARESTEDIISKF-KSFSACCGIGDGRTPFASTESYGSWIADLVPRMQHGKAPLATM    1059
ScAV1  SLYARSPTGGVDAYHYMDILRGYARVGEGQFGNELIHWCDWLGRHPVTKYFKAARGFRQAPVAVVLPEEELLPIRLHFEVLREKLTSGVWES-VDDFCNWLVTKHHVS----------    1064
SpAV1  SLYARKPTGGVDAYRFMEIIRGEQSPGEGQFGSDLVKWVDWLRKHPVTKFLRATGGRHHQKSTARLEGEDLRKVEDAFMALRERLGSGQVGS-TEDFSLWVSDLLRSGM---------    1053
EbAV1  SLYSRTATGGVDSWLFMNIIRGEQHVGEGQYGNDLMVWVSWIRDHPCTRYLKSVRGLRTRVEQLDVDPGLLQKATIHYSLLRETLVSGRIET-ALDFARWVRSIILGGHV--------    1049
EbAV2  SLYTREATGGMDAWQFMEIIRGDHHVGEGQFGNDLMRWLQWMASHPCSKYLRPVRGLRPGKSAVTAERELDLRAVNGFNTLRRLLSSGRLCS-SLDYALWISHSVIDSLPG-------    1054
CoAV1  SLYARRAEGGVDAWKFMEIIRGESHVGEGQFGNDLIDWLKWMHVHPMTKFLRGTRMFRTPSEEIVAGDEEMDAAVEAFNVLIERVESEEFQD-VESFAKWVSNMLRENVLPHQ-----    1066
GaAV1  ALYVRQAEGGIDAWKFMDILRGEGTVGEGQFGNDLRGWLRWMYAHPMTRHIRKVRGFTEPGTPAIADPATMQRTTYAFRILHEKLKAEEFNA-SEDFAIWLSTVIRQQKSR-------    1071
CdAV1  TLYSRAPEGGLDAWQITDIMRGMGSVGEGQFGNETLSWLKWMHHHPLTRYLRAARRYQDPSMPVELEPSQRAEALRAFEALRGRLLYDGFAS-SEEYAIWVARLVKLRYLA-------    1070
MsAV1  SLYLREATGGVDAWQFMDIIRGDAFVGEGQFGNDLRSWLNWLHHHPISKYLRETKTFRLKDGADPDAPKDLEKAGRALRALRGRLELGGFAC-VEDFVRWLDGIRSSRE---------    1058
VCV-M  SLYLRVPTGGIDAWQFMDVIRGDAFVGEGQFGNDLRTWLAWLHNHPISKYLRETKTFRLTEGIDPDQPVEVDKAARAFRALRGKAETGGFHS-VDEFVEWIHSATKET----------    1057
RHV-A  TLYARKGGGGIDAWRFMDIIRGDADLGQGQFGNEVLEWARWLGRHEVTKFLRPIKRMRAEHGEVIIEGEGRRQAEAALRALREILNEGTHRS-CLDFSFFLSDRLRASCEDPRR----    1077
AoAV1  GLYVRSTTGGIDAYKFMDFIRGDAVIGEGQWGNEMDRWIRFITEHPVSRSLRKARRHNPASVPSAQVNQYYQRAAEGLNVYRQQLVNHRLDS-SREYGLWVSDLLRR----------    1056
FpAV1  GLYVRSTTGGIDAYKFMDFIRGDSVIGEGQWGNEMTRWIRFVTEHPVSRALRKARRFNPKAAPTAGENLFYTRAQGGLNAYRRQLQNHRLES-SEEYGLWVSDLLRQK---------    1057
CaAV1  SLYARATTGGLDAYKVMDLIKGNSPIGRGQWGNDVMAWIRFVRDHPATRYLKETRRYKDEMIHIDQQTPLRPEVTGALAILSNALFNYVYED-SREFSLSIANRIRTKRSVN------    1062
STV    SLYARSSTGGLDAYKVMDLIRGNATIGRGQWGNDVMDWIRFVRDHPATKYLKGAKRFREQHNQEPTTAKPSKDARRAIKLLRNSLINEAYAD-SNSFAISISERLRRKKPTR------    1062
PpAV1  TLYVRDSAGNLDAWMFMEILRGERAVHPDQIGSDVDAWLTFLRENALTKYLRPIRRLRPEVKAKEYSEQDTSQGRAALLRLRDGVLNREWKN-GADFAMYISNLLICNVQQDLQ----    1046
        ** *   *.:*::   ..::*       * *..   *  ::  :  :  ::                                                    :
```