

Identification of *Trichosanthes* associated rhabdovirus 1, a novel member of the genus *Cytorhabdovirus* of the family *Rhabdoviridae*, in the *Trichosanthes kirilowii* transcriptome

C. J. GOH, D. PARK, Y. HAHN*

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

Received November 6, 2019; accepted November 15, 2019

Summary. – The genome sequence of a novel RNA virus, *Trichosanthes* associated rhabdovirus 1 (TrARV1), was identified in a transcriptome dataset isolated from a root sample of *Trichosanthes kirilowii*, which is a flowering plant belonging to the family *Cucurbitaceae*. The fruits, seeds, and root tubers of *T. kirilowii* have been used clinically in traditional Chinese medicine. The TrARV1 genome sequence was predicted to have six open reading frames (ORFs) encoding five canonical structural proteins of the family *Rhabdoviridae* (N ORF for nucleocapsid, P ORF for phosphoprotein, M ORF for matrix protein, G ORF for glycoprotein, and L ORF for polymerase), and an accessory protein. Sequence comparisons and phylogenetic analyses based on L and N proteins confirmed that TrARV1 is a novel member of the genus *Cytorhabdovirus* of the family *Rhabdoviridae*. TrARV1 is most closely related to Wuhan insect virus 5 and persimmon virus A. The putative *cis*-regulatory elements involved in transcription termination and polyadenylation, commonly found in the gene junction regions of rhabdoviruses, were also identified in the TrARV1 genome having the consensus sequence 3'-ACUAAAUUUUUGAUCUUU-5'. The genome sequence of TrARV1 may be useful to study the evolution and molecular biology of cytorhabdoviruses.

Keywords: *Trichosanthes* associated rhabdovirus 1; *Cytorhabdovirus*; *Rhabdoviridae*; *Trichosanthes kirilowii*

Introduction

Cytorhabdovirus is a genus of negative-sense, single-stranded RNA viruses and one of the 18 approved genera in the family *Rhabdoviridae* (Dietzgen *et al.*, 2017; Maes *et al.*, 2019). Currently, the genus *Cytorhabdovirus* comprises 11 species, including the type species, *Lettuce necrotic yellows cytorhabdovirus*, which is represented by lettuce

necrotic yellows virus (LNYV) (Dietzgen *et al.*, 2006; Maes *et al.*, 2019).

Cytorhabdoviruses are plant-infecting viruses and their replication and maturation occurs in the cytoplasm of infected cells (Jackson *et al.*, 2005; Franova *et al.*, 2019). Some cytorhabdoviruses have also been detected in insects, mainly aphids, leafhoppers, and planthoppers, which may be transmitting vectors (Ammar el *et al.*, 2009; Li *et al.*, 2015; Yang *et al.*, 2017; Whitfield *et al.*, 2018; Bejerman and Dietzgen, 2019). The virions of cytorhabdoviruses have a bullet-shaped structure with an enveloped capsid, containing a linear, negative-sense, single-stranded RNA genome, about 13 kb in length (Heim *et al.*, 2008; Walker *et al.*, 2018). The 3' and 5' terminal regions of their genomes are partially complementary to each other (Jackson *et al.*, 2005; Franova *et al.*, 2019).

The cytorhabdovirus genomes have six to 10 open reading frames (ORFs) (Walker *et al.*, 2011; Ito *et al.*, 2013; Bejer-

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

Abbreviations: LNYV = lettuce necrotic yellow virus; NCBI = National Center for Biotechnology Information; ORF(s) = open reading frame(s); PeVA = persimmon virus A; RdRp = RNA-dependent RNA polymerase; SRA = Sequence Read Archive; TrARV1 = *Trichosanthes* associated rhabdovirus 1; TTP = transcription termination polyadenylation; WhIV-5 = Wuhan insect virus 5

man *et al.*, 2015; Dietzgen *et al.*, 2017). The genome of LNYV, for example, has six ORFs encoding a nucleocapsid (N), phosphoprotein (P), 4b protein (4b), matrix protein (M), glycoprotein (G), and polymerase (L) (Dietzgen *et al.*, 2006). Five genes, N, P, M, G, and L, are canonical structural protein genes that are shared among rhabdoviruses (Walker *et al.*, 2011; Dietzgen *et al.*, 2017). The L protein contains an RNA-dependent RNA polymerase (RdRp) domain and is capable of viral RNA genome replication and viral mRNA transcription (Ammar el *et al.*, 2009).

Cytorhabdoviruses produce distinct mRNAs, one for each of the protein coding genes, which are transcribed from the negative-sense RNA genome (Jackson *et al.*, 2005; Walker *et al.*, 2011). The conserved *cis*-regulatory elements, known as the transcription termination polyadenylation (TTP) elements, are present in gene junction regions of rhabdoviruses (Jackson *et al.*, 2005; Heim *et al.*, 2008; Franova *et al.*, 2019). The TTP element regulates transcription termination and polyadenylation of upstream mRNAs and may serve as a transcription initiation site of the downstream gene. The consensus sequence of the TTP element is known as 3'-AUUNUUUU-5', where N is any nucleotide (Jackson *et al.*, 2005; Franova *et al.*, 2019).

Transcriptome data acquired from plant samples by the high-throughput RNA-seq method often contain sequences derived from RNA viruses, which can be identified by a comprehensive sequence analysis (Nibert *et al.*, 2016; Pecman *et al.*, 2017; Nibert *et al.*, 2018). Analyses of the publicly available transcriptome datasets has revealed many novel RNA viruses (Kim *et al.*, 2018; Park *et al.*, 2018; Lee *et al.*, 2019).

In this study, we analyzed a transcriptome dataset acquired from a root sample of *Trichosanthes kirilowii* (also known as gualou), of the family *Cucurbitaceae*, and identified the genome sequence of a novel member of the genus *Cytorhabdovirus*, belonging to the family *Rhabdoviridae*. Fruits, root tubers, and seeds of *T. kirilowii* contain pharmacologically active substances and have been used clinically in traditional Chinese medicine (Sha *et al.*, 2013; Ye *et al.*, 2016; Yu *et al.*, 2018). Many of these substances show promising pharmaceutical potential for treatment of cancer and other diseases. For example, trichosanthin, a ribosome-inactivating protein isolated from *T. kirilowii* root tubers, exhibits anti-tumor and anti-human immunodeficiency virus activities (Sha *et al.*, 2013; Ye *et al.*, 2016).

Materials and Methods

Transcriptome dataset. A transcriptome dataset (4.1 gigabases) obtained from the root sample of *T. kirilowii* was retrieved from the Sequence Read Archive (SRA) of the National Center

for Biotechnology Information (NCBI). The SRA Acc. No. for the RNA-seq data is SRR2084104. Raw reads were trimmed and filtered out using the Sickle program (version 1.33; <https://github.com/najoshi/sickle>), with the parameters, “-q 30 -l 50.” The resulting high-quality reads were assembled *de novo*, into contigs using the SPAdes Genome Assembler (version 3.13.1; <http://cab.spbu.ru/software/spades>), with the parameter, “--rna” (Bankevich *et al.*, 2012). The mapping of sequence reads to a contig for examination of sequencing depth was achieved using the BWA program (version 0.7.17-r1194-dirty; <https://github.com/lh3/bwa>) (Li and Durbin, 2009).

Known viral RdRp motif sequences. The Pfam database (release 32.0; <https://pfam.xfam.org>) was used to collect known virus-derived RdRp motif sequences (El-Gebali *et al.*, 2019). A keyword search was performed with the search term “RNA-dependent RNA polymerase”, and the resulting Pfam families were manually inspected to collect virus-derived sequences. A total of 2,539 RdRp motif sequences were obtained from 22 Pfam families (Pfam Acc. Nos. PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03035, PF03431, PF04196, PF04197, PF05788, PF05919, PF06317, PF07925, PF08467, PF08716, PF08717, PF12426, and PF17501). The USEARCH program (version 11.0.667; <https://drive5.com/usearch>) was used to reduce sequence redundancy, with the parameters, “-cluster_fast input -id 0.9 -centroids output -sort length” (Edgar, 2010). As a result, 2,195 non-redundant viral RdRp motif sequences were obtained.

Identification and annotation of virus genome contig. BLASTX searches of the viral RdRp motif sequences were performed using assembled transcriptome contigs as queries to identify novel viral genome sequences. The nucleotide (nt) and protein sequence similarity searches of putative viral genome contigs were performed using the BLAST web server at NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Based on the sequence comparison with related virus proteins, the ORFs were predicted. The Pfam web server was used to predict functional domains of viral proteins. Signal peptides and the transmembrane domain of proteins were predicted using SignalP (version 5.0; <http://www.cbs.dtu.dk/services/SignalP>) and TMHMM (version 2.0; <http://www.cbs.dtu.dk/services/TMHMM>), respectively (Moller *et al.*, 2001; Almagro Armenteros *et al.*, 2019).

Sequence comparison and phylogenetic analysis. Pairwise identities of protein sequences were calculated using the GGSEARCH program, in the FASTA package (version 36.3.8g; <https://github.com/wrpearson/fasta36>). Multiple alignments of viral protein sequences were produced using the MAFFT program (version 7.450; <https://mafft.cbrc.jp/alignment/software>), with the parameter, “--auto” (Nakamura *et al.*, 2018). Phylogenetic trees were constructed using the IQ-TREE program (version 1.6.12; <http://www.iqtree.org>), with the parameter, “-s input -bb 1000” (Nguyen *et al.*, 2015). Putative *cis*-regulatory elements conserved in gene junction regions, were detected using the MEME web server (version 5.0.5; <http://meme-suite.org/tools/>

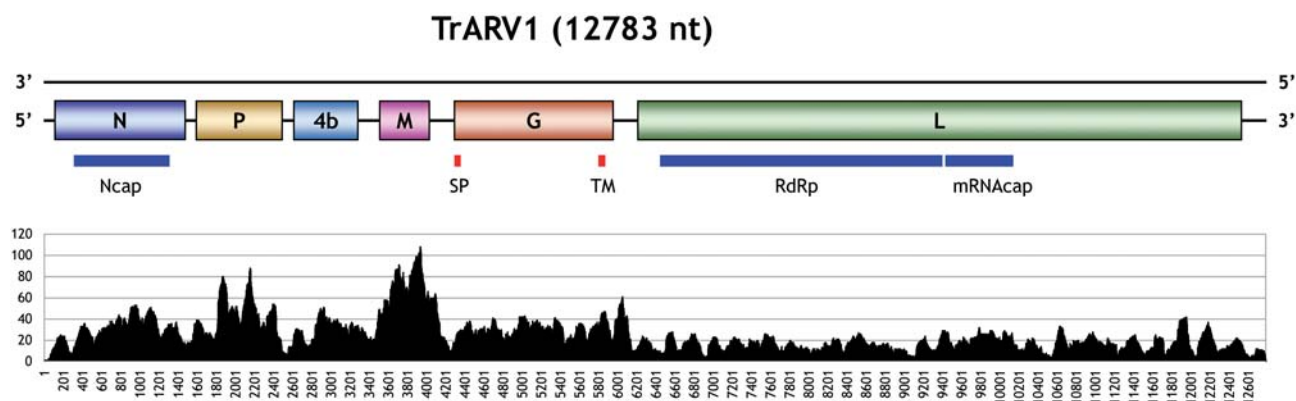


Fig. 1

Schematic representation of the TrARV1 genome organization

The genome organization of TrARV1 (12,783 nt) is schematically depicted. The six protein coding ORFs for the nucleocapsid (N), phosphoprotein (P), 4b protein (4b), matrix protein (M), glycoprotein (G), and polymerase (L), are represented as boxes. The predicted protein domains are marked by lines below the ORF: Ncap, Rhabdovirus nucleoprotein; SP, signal peptide; TM, transmembrane; RdRp, Mononegavirales RNA dependent RNA polymerase; and mRNAcap, Mononegavirales mRNA-capping region V. Sequencing depth of the TrARV1 genome contig is shown at the bottom. The X and Y axes represent the genomic position and sequencing depth, respectively.

meme) with the default setting (Bailey *et al.*, 2009). A sequence logo representation of aligned sequences was prepared using the WebLogo server (version 3; <http://weblogo.threeplusone.com>) (Schneider and Stephens, 1990; Crooks *et al.*, 2004).

Results and Discussion

A total of 4.1 Gbp of the *T. kirilowii* RNA-seq reads prepared from a root sample were assembled into contigs. The contig sequences were compared with the Pfam viral RdRp motif sequences. A 12,783 nt contig showed high sequence similarity to the RdRp region of the L protein of Wuhan insect virus 5 (WhIV-5) (Li *et al.*, 2015). WhIV-5 is a member of the genus *Cytorhabdovirus*, belonging to the family *Rhabdoviridae* (Jackson *et al.*, 2005; Li *et al.*, 2015; Dietzgen *et al.*, 2017; Walker *et al.*, 2018).

A sequence similarity search of the NCBI protein database revealed that the 12,783 nt contig contained multiple ORFs, showing sequence similarities with proteins of the known members of the genus *Cytorhabdovirus*, such as WhIV-5, persimmon virus A (PeVA), *Trifolium pratense* virus A (TpVA), tomato yellow mottle-associated virus (TYMaV), raspberry vein chlorosis virus (RVCV), and strawberry crinkle virus (SCV) (Ito *et al.*, 2013; Xu *et al.*, 2017; Koloniuk *et al.*, 2018; Franova *et al.*, 2019; Jones *et al.*, 2019). Therefore, the *T. kirilowii* contig was assumed to be a genomic sequence of a novel member of the genus *Cytorhabdovirus* in the family *Rhabdoviridae* and was named *Trichosanthes* associated rhabdovirus 1 (TrARV1). The nucleotide sequence of the TrARV1 genome is available in the Third Party Annotation (TPA) section of the DDBJ/ENA/GenBank databases under the accession number TPA: BK011194.

Table 1. Open reading frames (ORFs) of the TrARV1 genome sequence

ORF	Protein	Position (nt)	ORF length (nt)	Protein length (aa)	Pfam domain	Pfam Acc. No.	Position (aa)
N	Nucleocapsid	117-1487	1371	456	Rhabdovirus nucleoprotein	PF03216	67-399
P	Phosphoprotein	1601-2503	903	300	-	-	-
4b	4b protein	2614-3294	681	226	-	-	-
M	Matrix protein	3515-4045	531	176	-	-	-
G	Glycoprotein	4301-5962	1662	553	Signal peptide	-	1-20
-	-	-	-	-	Transmembrane	-	502-524
L	Polymerase	6216-12530	6315	2104	Mononegavirales RNA dependent RNA polymerase	PF00946	79-1062
-	-	-	-	-	Mononegavirales mRNA-capping region V	PF14318	1074-1309

The TrARV1 genome was predicted to have six ORFs (N, P, 4b, M, G, and L) based on sequence comparison with ORFs of the WhIV-5 genome (Table 1 and Fig. 1 top). These five ORFs (N, P, M, G, and L) encode canonical structural proteins commonly found in viruses of the family *Rhabdoviridae* (Walker *et al.*, 2018). The L ORF is the longest and encodes a 2,014 amino acid (aa) polymerase protein, which is involved in replication of the viral genome and transcription of viral mRNAs (Jackson *et al.*, 2005; Walker *et al.*, 2018). The L protein was predicted to have a “Mononegavirales RdRp” domain (Pfam Acc. No. PF00946), and a “Mononegavirales mRNA-capping region V” domain (Pfam Acc. No. PF14318). The N ORF encodes a 456 aa nucleocapsid protein that encapsidates the viral RNA genome. A “Rhabdovirus nucleocapsid protein” domain (Pfam Acc. No. PF03216) was predicted in the N protein. The P ORF encodes a 300 aa phosphoprotein, which is a component of the nucleocapsid core and forms complexes with N and L proteins (Jackson *et al.*, 2005; Walker *et al.*, 2018). The M ORF encodes a 176 aa matrix protein which forms a layer encasing the nucleocapsid and interacts with the viral envelope. The G ORF encodes a 553 aa glycoprotein, which is embedded in the viral envelope. The G protein was predicted to have a signal peptide at the N-terminus and a transmembrane domain near the C-terminus. The 4b ORF encodes a 226 aa accessory protein that may act as a movement protein (Ito *et al.*, 2013).

The sequencing depth of the TrARV1 genome sequence was examined by mapping high-quality reads to the con-

tig (Fig. 1 bottom). A total of 3,445 reads were mapped to the TrARV1 genome contig. Interestingly, the sequencing depth varied among ORFs. The average depth was found to be highest in the M ORF and lowest in the L ORF. This variation may indicate the transcription activity of viral ORFs, which are individually transcribed by the L protein. A similar pattern of difference in sequencing depth among ORFs was also observed in other cytorhabdoviruses (Franova *et al.*, 2019). The uneven distribution of sequencing depth indicates that the TrARV1 genome contig was assembled from a relatively larger amount of viral mRNAs together with viral genomic RNAs.

A sequence similarity search was performed against the NCBI protein database using the TrARV1 L protein sequence as a query, for the phylogenetic analysis of TrARV1. A total of 20 representative viruses from the genus *Cytorhabdovirus* were collected. The pairwise sequence identities between TrARV1 L and N proteins and their respective orthologs were calculated (Table 2). The TrARV1 L and N proteins showed the highest aa sequence identities of 60.9% and 57.8%, respectively, with orthologous WhIV-5 proteins. The PeVA L and N proteins showed the second highest similarities, with sequence identities of 49.3% and 35.3%, respectively.

A multiple sequence alignment of L proteins of TrARV1 and representative cytorhabdoviruses was generated (Supplementary Fig. S1). A maximum-likelihood phylogenetic tree was inferred from the multiple alignment of L protein sequences (Fig. 2). TrARV1 and WhIV-5 formed a

Table 2. Sequence identities of TrARV1 L and N proteins and their respective orthologs

Virus	Acronym	NCBI Acc. No. ^a	L protein	N protein
Wuhan insect virus 5	WhIV-5	NC_031227	1286/2111 (60.9%) ^b	267/462 (57.8%)
Persimmon virus A	PeVA	NC_018381	1045/2118 (49.3%)	164/464 (35.3%)
Wuhan insect virus 6	WhIV-6	NC_031232	903/2119 (42.6%)	156/468 (33.3%)
Trifolium pratense virus A	TpVA	MH982250	937/2141 (43.8%)	147/462 (31.8%)
Tomato yellow mottle-associated virus	TYMaV	NC_034240	874/2129 (41.1%)	132/464 (28.4%)
Strawberry associated virus 1	SAV1	MK159261	870/2141 (40.6%)	133/467 (28.5%)
Alfalfa dwarf virus	ADV	NC_028237	882/2124 (41.5%)	157/491 (32.0%)
Raspberry vein chlorosis virus	RVCV	MK257717	883/2136 (41.3%)	150/490 (30.6%)
Strawberry crinkle virus	SCV	MH129615	884/2132 (41.5%)	147/472 (31.1%)
Wuhan insect virus 4	WhIV-4	NC_031225	868/2157 (40.2%)	153/496 (30.8%)
Cabbage cytorhabdovirus 1	CCyV-1	KY810772	865/2123 (40.7%)	152/470 (32.3%)
Lettuce necrotic yellows virus	LNyV	NC_007642	897/2119 (42.3%)	148/464 (31.9%)
Lettuce yellow mottle virus	LYMoV	NC_011532	909/2123 (42.8%)	159/463 (34.3%)
Trifolium pratense virus B	TpVB	MH982249	893/2120 (42.1%)	158/462 (34.2%)
Bean-associated cytorhabdovirus	BaCV	MK202584	616/2188 (28.2%)	103/499 (20.6%)
Rice stripe mosaic virus	RSMV	NC_040786	618/2183 (28.3%)	110/510 (21.6%)
Colocasia bobone disease-associated virus	CBDaV	NC_034551	661/2163 (30.6%)	107/496 (21.6%)
Maize-associated cytorhabdovirus	MaCyV	KY965147	680/2180 (31.2%)	119/495 (24.0%)
Maize yellow striate virus	MYSV	KY884672	671/2164 (31.0%)	119/487 (24.4%)
Barley yellow striate mosaic virus	BYSMV	NC_028244	663/2168 (30.6%)	123/487 (25.3%)

^aNCBI Acc. No. for genomic sequence. ^bAmino acid sequence identities in the format: “identical residues/aligned length (% identity)”.

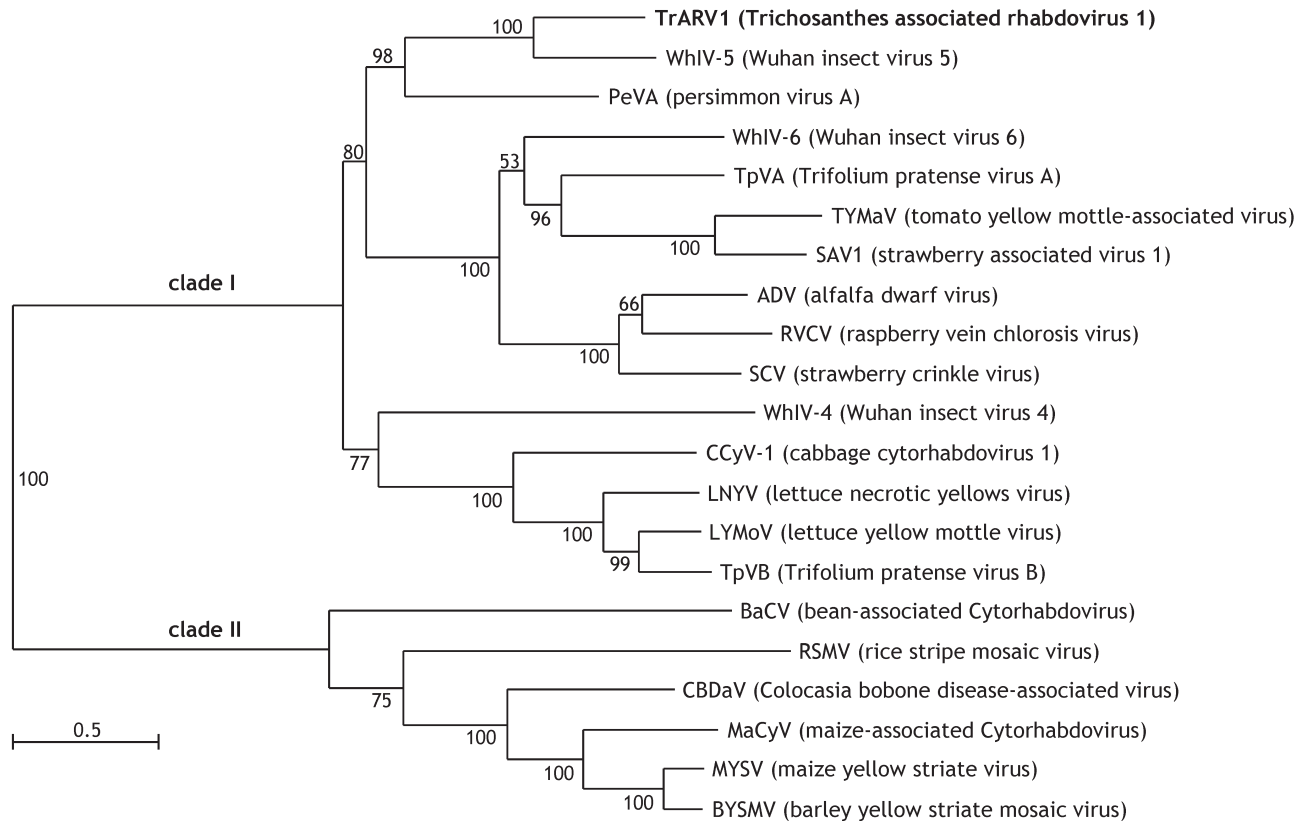


Fig. 2

Phylogenetic relationship of TrARV1 and other cytorhabdoviruses

A phylogenetic tree constructed from multiple alignment of L protein sequences of TrARV1 and other members of the genus *Cytorhabdovirus*. The WhIV-5 virus is the closest to TrARV1, and the second closest is PeVA. Cytorhabdoviruses are clustered into two clades (clade I and clade II). TrARV1 belongs to the clade I. Bootstrap branch support values, calculated from 1,000 replicates, are shown at the tree nodes.

strong subclade, with a bootstrap value of 100%. Cytorhabdoviruses analyzed in this study were clustered into two clades named here as 'clade I' and 'clade II.' TrARV1 was placed in clade I together with PeVA, TpVA, TYMaV, RVCV, and SCV. A multiple alignment of N protein sequences was also generated (Supplementary Fig. S2). A phylogenetic tree based on N proteins showed almost identical topology with the L protein tree, except a small difference in Wuhan insect virus 6 and RVCV (Supplementary Fig. S3). The phylogenetic analyses based on L and N proteins confirmed that TrARV1 is a novel species of the genus *Cytorhabdovirus*, which is most closely related to WhIV-5.

The cytorhabdoviruses have conserved cis-regulatory elements known as TTP elements in gene junction regions (Jackson *et al.*, 2005; Heim *et al.*, 2008; Franova *et al.*, 2019). These elements are known to be involved in the transcription termination and polyadenylation of viral gene mRNAs, which are separately transcribed by the L protein (Jackson *et al.*, 2005; Walker *et al.*, 2011). The gene junction regions of the TrARV1 genome were investigated

to find putative TTP elements. Two closely related viruses, WhIV-5 and PeVA, were analyzed to identify shared motifs among them. A total of 7 junction regions, 3'/N, N/P, P/4b, 4b/M, M/G, G/L, and L/5', were extracted from each genome sequence. A total of 21 junction sequences collected from TrARV1, WhIV-5, and PeVA were subjected to a motif search analysis.

A 20 nt sequence motif was identified in gene junction regions of TrARV1, which was shared in WhIV-5 and PeVA (Fig. 3). In each genome, the motif sequence was found in 6 out of 7 junction regions: N/P, P/4b, 4b/M, M/G, G/L, and L/5'. No shared motif sequence was found in the 3'/N junction. The deduced consensus sequence of putative TTP elements of TrARV1 is 3'-ACUAAAUUUUUUGAUCUUU-5', where the underlined sequence matches the known rhabdovirus consensus sequence 3'-AUUNUUUU-5' (N, any nucleotide) (Jackson *et al.*, 2005; Franova *et al.*, 2019).

To conclude, the genome sequence of a novel RNA virus named TrARV1 was identified from the *T. kirilowii* transcriptome data obtained from a root sample. The

TrARV1 genome was predicted to encode five major structural proteins of the family *Rhabdoviridae* and an accessory protein. Sequence comparison and phylogenetic analysis indicated that TrARV1 is a novel member of the genus *Cytorhabdovirus* in the family *Rhabdoviridae*. The TrARV1 genome sequence reported in this study may be useful to investigate the evolution of plant-infecting rhabdoviruses.

Acknowledgment. This research was supported by grants from the National Research Foundation of Korea, funded by the Government of Korea [grant Nos. NRF-2017R1A1B4005866 and NRF-2018R1A5A1025077].

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

References

- Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019): SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. <https://doi.org/10.1038/s41587-019-0036-z>
- Ammar el D, Tsai CW, Whitfield AE, Redinbaugh MG, Hogenhout SA (2009): Cellular and molecular aspects of rhabdovirus interactions with insect and plant hosts. *Annu. Rev. Entomol.* 54, 447–468. <https://doi.org/10.1146/annurev.ento.54.110807.090454>
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009): MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–208. <https://doi.org/10.1093/nar/gkp335>
- 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>
- Bejerman N, Dietzgen RG (2019): Letter to the Editor: Bean-associated cytorhabdovirus and papaya cytorhabdovirus are strains of the same virus. *Viruses* 11. <https://doi.org/10.3390/v11030230>
- Bejerman N, Giolitti F, de Breuil S, Trucco V, Nome C, Lenardon S, Dietzgen RG (2015): Complete genome sequence and integrated protein localization and interaction map for alfalfa dwarf virus, which combines properties of both cytoplasmic and nuclear plant rhabdoviruses. *Virology* 483, 275–283. <https://doi.org/10.1016/j.virol.2015.05.001>
- 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>
- Dietzgen RG, Callaghan B, Wetzel T, Dale JL (2006): Completion of the genome sequence of lettuce necrotic yellows

TrARV1	N/P	ACUAAAUUUUUGAUCUUU
	P/4b	ACUAAAUUUUUGAUCUUU
	4b/M	ACUAAAUUUUUGAUCUUU
	M/G	ACUAAAUUUUUGAUCGUU
	G/L	ACUAAAUUUUUGAUCAC
L/5'	CCUAAAUUUUUGAUCUCG	
WhIV-5	N/P	ACUAAAUUUUUGAUCUUU
	P/4b	ACUAAAUUUUUGAUCUAG
	4b/M	ACUAAAUUUUUGAUCAAU
	M/G	ACUAAAUUUUUGGUCGUU
	G/L	ACUAAAUUUUUGAUCAUU
L/5'	ACUAAAUUUUUGAUC CAG	
PeVA	N/P	AGUAAAUUUUUGGUCUCU
	P/4b	AGUAAAUUUUUGAUCUUU
	4b/M	ACAAAUUUUUGAUCUCU
	M/G	GGUAAAUCUUUUGAUCUUU
	G/L	GGUAAAUCUUUUGAUCUUU
L/5'	UGUAAAUUUUUGAUCACA	

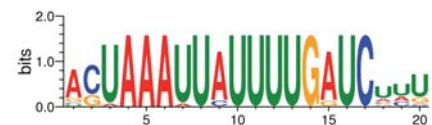


Fig. 3

Conserved sequences in gene junctions

Putative cis-regulatory elements for transcription termination and polyadenylation, identified in gene junctions of TrARV1, WhIV-5, and PeVA are shown in 3' to 5' orientation. The conserved residues in 50% or more of the aligned sequences are highlighted on black background. A sequence logo representation calculated from aligned sequences is shown at the bottom.

virus, type species of the genus *Cytorhabdovirus*. *Virus Res.* 118, 16–22. <https://doi.org/10.1016/j.virusres.2005.10.024>

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

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

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019): The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. <https://doi.org/10.1093/nar/gky995>

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

- Heim F, Lot H, Delecqle B, Bassler A, Krczal G, Wetzel T (2008): Complete nucleotide sequence of a putative new cytorhabdovirus infecting lettuce. *Arch. Virol.* 153, 81–92. <https://doi.org/10.1007/s00705-007-1071-5>
- Ito T, Suzaki K, Nakano M (2013): Genetic characterization of novel putative rhabdovirus and dsRNA virus from Japanese persimmon. *J. Gen. Virol.* 94, 1917–1921. <https://doi.org/10.1099/vir.0.054445-0>
- Jackson AO, Dietzgen RG, Goodin MM, Bragg JN, Deng M (2005): Biology of plant rhabdoviruses. *Annu. Rev. Phytopathol.* 43, 623–660. <https://doi.org/10.1146/annurev.phyto.43.011205.141136>
- Jones S, McGavin W, MacFarlane S (2019): The complete sequences of two divergent variants of the rhabdovirus raspberry vein chlorosis virus and the design of improved primers for virus detection. *Virus Res.* 265, 162–165. <https://doi.org/10.1016/j.virusres.2019.03.004>
- 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>
- Koloniuk I, Franova J, Sarkisova T, Pribylova J (2018): Complete genome sequences of two divergent isolates of strawberry crinkle virus co-infecting a single strawberry plant. *Arch. Virol.* 163, 2539–2542. <https://doi.org/10.1007/s00705-018-3860-4>
- Lee JS, Goh CJ, Park D, Hahn Y (2019): Identification of a novel plant RNA virus species of the genus Amalgavirus in the family Amalgaviridae from chia (*Salvia hispanica*). *Genes Genomics.* <https://doi.org/10.1007/s13258-019-00782-1>
- Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ (2015): Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* 4. <https://doi.org/10.7554/eLife.05378>
- 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>
- Maes P, Amarasinghe GK, Ayllon MA, Basler CF, Bavari S, Blasdel KR, Briese T, Brown PA, Bukreyev A, Balkema-Buschmann A, Buchholz UJ, Chandran K, Crozier I, de Swart RL, Dietzgen RG, Dolnik O, Domier LL, Drexler JF, Durrwald R, Dundon WG, Duprex WP, Dye JM, Easton AJ, Fooks AR, Formenty PBH, Fouchier RAM, Freitas-Astua J, Ghedin E, Griffiths A, Hewson R, Horie M, Hurwitz JL, Hyndman TH, Jiang D, Kobinger GP, Kondo H, Kurath G, Kuzmin IV, Lamb RA, Lee B, Leroy EM, Li J, Marzano SL, Muhlberger E, Netesov SV, Nowotny N, Palacios G, Palyi B, Paweska JT, Payne SL, Rima BK, Rota P, Rubbenstroth D, Simmonds P, Smither SJ, Song Q, Song T, Spann K, Stenglein MD, Stone DM, Takada A, Tesh RB, Tomonaga K, Tordo N, Towner JS, van den Hoogen B, Vasilakis N, Wahl V, Walker PJ, Wang D, Wang LF, Whitfield AE, Williams JV, Ye G, Zerbini FM, Zhang YZ, Kuhn JH (2019): Taxonomy of the order Mononegavirales: second update 2018. *Arch. Virol.* 164, 1233–1244. <https://doi.org/10.1007/s00705-018-04126-4>
- Moller S, Croning MD, Apweiler R (2001): Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17, 646–653. <https://doi.org/10.1093/bioinformatics/17.7.646>
- 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>
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015): IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>
- 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>
- Nibert ML, Vong M, Fugate KK, Debat HJ (2018): Evidence for contemporary plant mitoviruses. *Virology* 518, 14–24. <https://doi.org/10.1016/j.virol.2018.02.005>
- Park D, Goh CJ, Kim H, Hahn Y (2018): Identification of two novel amalgaviruses in the common eelgrass (*Zostera marina*) and in silico analysis of the amalgavirus +1 programmed ribosomal frameshifting sites. *Plant Pathol. J.* 34, 150–156. <https://doi.org/10.5423/PPJ.NT.11.2017.0243>
- Pecman A, Kutnjak D, Gutierrez-Aguirre I, Adams I, Fox A, Boonham N, Ravnikar M (2017): Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. *Front. Microbiol.* 8, 1998. <https://doi.org/10.3389/fmicb.2017.01998>
- 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>
- Sha O, Niu J, Ng TB, Cho EY, Fu X, Jiang W (2013): Anti-tumor action of trichosanthin, a type I ribosome-inactivating protein, employed in traditional Chinese medicine: a mini review. *Cancer Chemother. Pharmacol.* 71, 1387–1393. <https://doi.org/10.1007/s00280-013-2096-y>
- Walker PJ, Blasdel KR, Calisher CH, Dietzgen RG, Kondo H, Kurath G, Longdon B, Stone DM, Tesh RB, Tordo N, Vasilakis N, Whitfield AE, ICTV_Report_Consortium (2018): ICTV virus taxonomy profile: Rhabdoviridae. *J. Gen. Virol.* 99, 447–448. <https://doi.org/10.1099/jgv.0.001020>
- Walker PJ, Dietzgen RG, Joubert DA, Blasdel KR (2011): Rhabdovirus accessory genes. *Virus Res.* 162, 110–125. <https://doi.org/10.1016/j.virusres.2011.09.004>
- Whitfield AE, Huot OB, Martin KM, Kondo H, Dietzgen RG (2018): Plant rhabdoviruses—their origins and vector interactions. *Curr. Opin. Virol.* 33, 198–207. <https://doi.org/10.1016/j.coviro.2018.11.002>
- Xu C, Sun X, Taylor A, Jiao C, Xu Y, Cai X, Wang X, Ge C, Pan G, Wang Q, Fei Z, Wang Q (2017): Diversity, distribution, and evolution of tomato viruses in China uncovered by small RNA sequencing. *J. Virol.* 91. <https://doi.org/10.1128/JVI.00173-17>

- Yang X, Huang J, Liu C, Chen B, Zhang T, Zhou G (2017): Rice stripe mosaic virus, a novel cytorhabdovirus infecting rice via leafhopper transmission. *Front. Microbiol.* 7, 2140. <https://doi.org/10.3389/fmicb.2016.02140>
- Ye X, Ng CC, Wong JH, Ng TB, Chan GH, Guan S, Sha O (2016): Ribosome-inactivating proteins from root tubers and seeds of *Trichosanthes kirilowii* and other *Trichosanthes* species. *Protein Pept. Lett.* 23, 699-706. <https://doi.org/10.2174/0929866523666160526130220>
- Yu X, Tang L, Wu H, Zhang X, Luo H, Guo R, Xu M, Yang H, Fan J, Wang Z, Su R (2018): *Trichosanthes Fructus*: botany, traditional uses, phytochemistry and pharmacology. *J. Ethnopharmacol.* 224, 177-194. <https://doi.org/10.1016/j.jep.2018.05.034>

SUPPLEMENTARY INFORMATION

**Identification of Trichosanthes associated
rhabdovirus 1, a novel member of the genus *Cytorhabdovirus* of the family
Rhabdoviridae, in the *Trichosanthes kirilowii* transcriptome**

C. J. GOH, D. PARK, Y. HAHN*

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

Received November 6, 2019; accepted November 15, 2019

Fig. S1. Multiple sequence alignment of the TrARV1 L (polymerase) and related proteins

TrARV1	-----MDFLELND--QFGSGDGNVEVKHETDQLPDFHLRSPLHPLRGLY---LPPNRVPRRIRDSIEILRSIRSN-LVEG	68
WhIV-5:YP_009300875	-----MDFLDFAD--DFRNEDGAQRGSRDQLPDFHLRSPLKPMYLY---DDTVR-SLRNQEAIELLTEIRRD-LKEG	67
PeVA:YP_006576506	-----MDYMAFCTGGAGDGETDVKANDKFDNLPDYHLRNPLHPMKYLY---TEANRSPDIRKSLSSLRSHKKT-LTEG	70
WhIV-6:YP_009301361	-----MDFLSLI----NEDDCSQPKKSVLPDYHLRSPLKPLDWIH---MDHNRITSRNRDLNLYLRNYVRH-FIVG	65
TpVA:AYH53279	-----MSYFFGN----GDDDGAEAPVIQKTIILPDYHLRNPLKQPLWLS---DGKRKLPMRQRKDKAFLRSYYPD-AKWG	65
TYMaV:YP_009352236	-----MDFISKLM---DDTEGLAQSISETPLADYHLRNPLKRLEW----WDKQFVPSRQHKDRLRIRQSYPRAEPLK	65
SAV1:QDJ94287	-----MDFREKLE---DDISDQAIIVSETALADYHLRNPLKEITW----WEKERLPARLYKDRKIREIFGDRLTIG	65
ADV:YP_009177021	-----MDFLV-----FNDDEHQAKVKNKAPLPDYHLRNPIKPLWLH---DDSRLTSRVRDKAYLKKLGMGRIELG	63
SCV:Awk49433	-----MDFAEK-----ANDTEDTSVRNKDPLPDFHLRNVPRLDWLN---NNMRATSRNRIDQAYINKIGNQRVRVG	64
RVCV:QBS46644	-----MDYVN-----HNTGEGGNVRKEDSLPDFHMRNPITQLEWLH---DKTKLTSRNKFDLGVINSIGKHKVKVG	63
WhIV-4:YP_009300689	MFYLVKMDFH-----DINDESEFRQKYDGLPDFHLRNPLRSFPHHIIE-IKNKSTDKRVKSSLEALRGVRKT-LKGG	70
CCyV-1:ATS17313	-----MNF-----IESEGKKKLYDPLPDFHLRNPLYSLNERIQAWKNKRLPVRLYKSFHALSSASNN-LEEG	64
LNyV:YP_425092	-----MDLW-----NEETPKKNAYDSLDPYHLQNPLVAITDQLSMLKRGKRLNRYLTSYKLMKSQSTD-IKEG	64
LYMoV:YP_002308376	-----MEIY-----QDEDAPRKNIFDPLDPYHLQNPLYSISGLLDKIKKKERVNYRFHQSCRVLKSKESK-LVEG	64
TpVB:AYH53273	-----MEIF-----NEDGGRKNQFDSLDPYHLQNPLYSISDLISKVLNRQVRVNNRFYQSYRAMQRESSQ-IEES	64
BaCV:QAU20941	-----MSFFF-----EEGLVDEGDAQVFMRGLGDFHLRSALVLPKLAQ---LERGHGRKREVKSIAKVKEKK---LFAA	63
RSMV:YP_009553369	-----MDL-----DDGGLWRRRARGLDYHLRSALVTPSLER---FRSRKGRHRQLCFDRMKSLGWM-LRWV	58
CBDaV:YP_009362280	M-----ADF-----DPFDEIRRPMKGLGDFHLRSALVPIINLER---LKLGEGRWRERRAFTQMRQAFDD-ISVG	60
MaCyV:ARS22495	-----MDFG-----LDDDELKPLRGLGDFHLRSALVPIINISE---LLLQGKRREQRSFNISIRQIPN-IVAG	60
MYSV:ATN96453	-----MDF-----LDDEVKRRMRGLGDFHLRSALVPIIDING---LKEGKGRARELKSFKKIIIRNYPD-VYKG	58
BYSMV:YP_009177231	-----MDL-----LEDDVKRRMRGLGDFHLRSALVPIEINS---LRAGIGRARELRAFRRIMQYHPN-VFCG	58
	* * : * : : . . . :	* :

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

TrARV1 SPIMLRHLL -DHP -YDESTRTAEDCMSDVGVRLWKDGRLLN ---PMPVVLMEEA--LGKLSRDFPVDLWNR--- -MRFSD 137
WhIV-5: YP_009300875 DPITLRSMFTDYA -EVESTAGAEDSISDLLVRIERDSRLNF ---PLPVEMMIDA--VRQFSDNFPILWLSR--- -MRLLS 137
PeVA: YP_006576506 SPEDLITSMGDLI -PIANRSKLYTGLDIIIMRLTDEEALP ---QLNGLHMEHL--ITEITPEMNLWLDC--- -MRFWG 140
WhIV-6: YP_009301361 NPIDLISAFNDKI -PNQNTESWSKVTDGDLIRMGIDRDILP ---EETHPDFEG--- -IRDIICLDNLGNLWNR--- -MRFWN 134
TpVA: AYH53279 KPFLSNRRISPFII -PDMPLPEPYGISLSDLELRSLIDNGLFP ---PDSHIPVSS--- -IINSLKQLERRYWN--- -MRFWN 134
TYMaV: YP_009352236 DPSSLHRMTHHII -PSMDSNKFSEALGDMVYRLRMDNGILP ---CGSRLPIKS-- -VMRAASSMPKRYWDG--- -MRYWN 134
SAV1: QDJ94287 HPSSLHRITSSRL -PNMDMRQMECSLDLALRMLMDNGILP ---RESRAPIKK-- -VLANAAREIPRMYWNG--- -MRFWG 134
ADV: YP_009177021 SHIDLYL TENMKF -DTESTLPLEDAIADLIVRLKIDVDVMG ---KSRGVSVD--- -IAEAVVSAPKHLWGG--- -MRFWG 132
SCV: AWK49433 IHTDLYL TEHLDF -SHISVEGYKESVIDMITRAKIDTEILG ---QNRDIDLQD--- -ILIEEMGRVDIKFWKG--- -MRFFG 133
RVCV: QBS46644 SHIDLYL DREMDM -TTVPLSPLLESADVSLRSLIDLDILG ---KNRSFDTG--- -VITAMLKTDIVYWN--- -MRFWG 132
WhIV-4: YP_009300689 RSHDLWTDHAYL -VDDTLGEGISLQDITIERLSLDSYLSH ---YFVDTMQDL-- -IMSSSDRLLESKYMIN--- -MNNFQ 140
CCyV-1: ATS17313 NPIDLNLFLSLP -VISVHPLPAYALDDCLYKRLDLSKSDH ---NISTKLLKDC-- -WLKVCANFPARLWSS--- -MREGQ 134
LNYV: YP_425092 DPILLKEWARNWF -SETELYVVDQTLSDCENRSLDETDL ---HFDTHLLRES-- -VRMWRCEFPDDWVR--- -MRGMQ 134
LYMoV: YP_002308376 DPILLHRLSDNWF -IDTEPCIIEIFEDCSYRLQMSLNP ---QFDIKMLSA-- -WESLKNRFPVDYWVR--- -MRGMQ 134
TpVB: AYH53273 NPIKRLRELARWF -NDSEMETIQEIAAGCDFRDLDSNLGD ---HFDISTLKES-- -WEIAKTQFPIDYWIT--- -MRGLQ 134
BaCV: QAU20941 DPGRFLSLFFQLSLQSGFNDVQPYLNELEVELVEHSAIR -PLTQDQDQDQCLNLLKT -GEIKTNYGAG--- -KLMFQ 137
RSMV: YP_009553369 DQGKLLGLYLMVEA-- -NKSLPKS TANQELLVETLKEYGCLRQIIMTDGDLHDQVISYLDL -KSISTHYTHG--- -REVFQ 131
CBDaV: YP_009362280 DPAKLLSEIMSIS-- -RYEPEENLPIEETVELLQMEIDGLS -FLTSPGDPLAK-- -VLKRLKDPQTRPTSSLYGPKILFQ 134
MaCyV: ARS22495 DPADLLAQILLGV-- -KGLQPREIPKLEETKRLLNAEFEQLS -WLRGENDPMAQLMDYLDIIEGGVQSQHTM--- -KIKLQ 133
MYSV: ATN96453 DPASVLAITLKSS-- -SSGELRSILMFRETVELINAEYKGLG -FLTANQDPMKLLMLLMKVRGIQSRYSASS--- -KISYQ 131
BYSMV: YP_009177231 DPASTLAQILRTS-- -SSGELKSILMFKETVELIKAEYKGMG -FLTSNQDPMERLLMLLMKVKGIRYSRYASS--- -KIAAYQ 131

:

TrARV1 KLLTTLNAISSKRQPPGRLTKVK -----DNLYKAITRT -HYIYVTPALLGAE--PLEWQNGEAP----- 193
WhIV-5: YP_009300875 KILSTMNAISSRRSPGGYIKIK -----KNLYATMIRN -TRLMTSTILGVE--IIDESTGHG----- 192
PeVA: YP_006576506 DVLLTMNAMSSRRPAPSCSVELQ -----FGCRGIHMNDGVTCTYITGSVIAME--SADDSK----- 193
WhIV-6: YP_009301361 SVLLVLNALSSSRPVEGFIISR -----GVVKITCLKT -TEVIMVRTCIVVR--CGRG----- 184
TpVA: AYH53279 KVLII LNALSSERSPPSGVVIKN -----GMGYLTMKMR -ASLIICRTCVFV--HRD----- 183
TYMaV: YP_009352236 KVLYYLNAFASNRRCPDGSSDVP -----AVLSLLEGEN--KLLVYRSCVLLI--TEEMK----- 184
SAV1: QDJ94287 RALLDLNANVSNRKKPAGEMTIP -----SVLKFEEDRD-- -KIYIFRTCSVLSRSETPS----- 186
ADV: YP_009177021 RVL SIMNARSSKRPLPPGHSEVD -----GVVVRVNGAL-- -HVTVLNTGILIN--RDRS----- 181
SCV: AWK49433 HVLSVMNLSRRRQIPKEFAEEG -----GVVKIQLKEM-- -TVTLLTGILVS--GNTG----- 182
RVCV: QBS46644 HILRIMNALSNNRKPPTGVVVVN -----GTGRVNAGRY-- -TIIIMKTSIMIT--GDNI----- 181
WhIV-4: YP_009300689 IFVVIINALSSRRPIPEKYKEHQRS -----GVVHFTISG -NQVLLTPSFLGLI--KYEYPT----- 194
CCyV-1: ATS17313 LLLSGLNALSSRRPLPSGFVKIS -----ETLYERRTHS -VRWLISPALLGVC--VNRSN----- 185
LNYV: YP_425092 NLLIVMNAISSRRPPPRHTCII -----PGLSKITVEG -GVVLTSSLLGFQ--PEGDK----- 185
LYMoV: YP_002308376 RILVAMNAVSSRRPPPRYSTAD -----PLKAIMEVNK -GRLIITGSLGKY--QDGED----- 185
TpVB: AYH53273 NVLLSMNAVSSRRPPPRFSSPD -----PQVAQMDVLG -GRLIITGSLGFK--EDGST----- 185
BaCV: QAU20941 LLVYYSMILLSSGRTWEGGKVFQKE ---KDGTPWVSIFG -VKVWMLGEFIAIP--YSESLRFGVEDPQDINDCAQELKA 210
RSMV: YP_009553369 EALIVVMALSSGREPPDHVNNLGYEMLNEELEVPPVRYG -VIFYLFGDLIYVK--YPE----- 187
CBDaV: YP_009362280 RGLMCLNAMTSERDYSFFDLPHD -E-----SGFPMIEIGP -LRLILGGDLFGSI--IGMK----- 185
MaCyV: ARS22495 RLLMCANAMNSDRDYSAFNLPSPHN ---WDKPYQIFGP -VMIVTWGELITAS--SEYL----- 185
MYSV: ATN96453 RLLMCANAMNSQREYGVFNLEDV ---DGAPSLTVNG -ITVGIWGLITVS--SPDM----- 183
BYSMV: YP_009177231 RLLMCANAMNSQRSYSVFLDQDVT ---PGAPSLTVNG -LTVGVWGEFITIS--SADN----- 183

: * *

TrARV1 -----HLVCFDADWLRMVSVDVHSERFLVYVGCAGFTEINPLHYPEWNTISKIISWGDDILDEYEGNEGFKLLKAYEA 264
WhIV-5: YP_009300875 -----FVLCYSDWIRMISDVYTERFLLSVSGSGFGHHMNPDPHYPSQSIIEKLLWLGDSVLAEFNGDGFLLKAYEA 263
PeVA: YP_006576506 -----GLICYDADWIRMASSDVTQRFLLNLSTLIGNKLSNVVYPKVALVKDLIEWGDTVLKICGNEGYKLIKNEYEA 264
WhIV-6: YP_009301361 -----SGTIYDGDWVRMSSDLHTQRFLLFSLRSLGQKVNQFVYPGVDLISSIISWGDDVLFNFKNEGYKLLKTYEA 255
TpVA: AYH53279 -----LREVFDDGWVRMSSDINTQRFLLLSGKIGREVNNEYHYPEGDILTDLFDWGDVLEERRGNDGFKLLKAYEA 254
TYMaV: YP_009352236 -----YGAVLDGDWIRMISDLYTQRWLVRTSASIGRAVNPYHYPSDEVIEKIYHWGDQVLDLGNEGFAVIVKVES 255
SAV1: QDJ94287 -----LHILIDGDWVRMRLSDLYTQRWLVRCGAAIGRCFNPEHYPPAEVIERVYQWGDVSIADLGNTRYKIRTFEP 257
ADV: YP_009177021 -----KYLIDYDGDWVRLAADVYTYRFLYITGAYVGLHINPSQYPSIANVESVIRWGDVNLSELGNEGFKVLTKEA 252
SCV: AWK49433 -----THLLYDGDWVRLAADVYTYRFLFCIGAAIGRSINPMQYDPLRTIEAVMRWGDVLRALGNEGFKVLTKEA 253
RVCV: QBS46644 -----EYLLYDGDWVRLTSDVYTYRFLLAGANAGSINSHQYPSVDTINSVIAWGDVLRDMGNLTGFKVLTKEA 252
WhIV-4: YP_009300689 -----EILYPSDWRCSDDVHTERFLNLNTANIGNELNREHYPPWSVISKILTWGDQVLYRYGNPFGFKLIKTYEA 265
CCyV-1: ATS17313 -----KIKLIDSDWLRVSDVYTERFLVCLGLTLGSSYNLPHYPTWDRLLPVLTWGDDVLLKKGNSGFKLLKAFEA 256
LNYV: YP_425092 -----ETTVFAADWVRASDVYTERFLVLSGAILGRCLSDHYPPQVLDLEIFINWGDVTLRRKNGKGFLLKAYEA 256
LYMoV: YP_002308376 -----HLVIFASDWVRGSDVHTERFLVHIGATLGDMSKDHYPPSSAQIEYIIQWGDVLRNKAAGMTGFKLLKAYEA 256
TpVB: AYH53273 -----TLLIFASDWLRVSDVYTERFLVHFGANIGRNINRDHYPESSDIEYIINWGDVAVLGNKNGSGFKLLKAYEA 256
BaCV: QAU20941 DRRTLKPKQDKLLILTDMVRMAADKFSERENIEIANHLSGSSVCPYIEYPPDILGQLLEIMDRELWVFNHGYKIFKTYEA 290
RSMV: YP_009553369 -----EEGMITLDMFRNLTDKFSERENIVIAATQLGTEILQEIPSETVLTWVSLWDKGLLEKGNDFYTVVKTFEA 258
CBDaV: YP_009362280 -----SLEIFSLDVFIRMIVDKLTERDNVIVASLLGRSIFPQVYPPNRLIEIFFIWFDEWLMAGNAGYALLKTFEA 256
MaCyV: ARS22495 -----SPSVYMDVLRMLTDKVTERDNVLLSSHIGNKIFPSIYSPKTIKGVFNILDEWVYDQGNAGYLLKAYEA 256
MYSV: ATN96453 -----PISLFLSDVLRMIVDKLTERDNVIVSSIFGEEKIFPQIYPRVETISKIFDLFDSWLYLKGNSGYKMMKTFEA 254
BYSMV: YP_009177231 -----PISLFLSDVLRMIVDKLTERDNVIVSSIFGEEKIFPHIYPKVDVITQVFDLFDNWLYYKNGSGYRLMKTYEA 254

* * * * * : * * * * * : * * * * * : * * * * *


```

TrARV1             LTPKERELNPTPRMFALMSHLMRIYVVITEQMLSDHVLPPFPQITMTDSSLDDLTKKLYTTVKPQAQKGGARKKRATWTSR 575
WhIV-5:YP_009300875 LTPKERELNPTPRMFALMSHLMRIYVVITEQMLSDHVLPPFPQITMTDSSLDDLTKKLYTTVKPQSDGASVRRRNKTWTSK 574
PeVA:YP_006576506   LTPKERELNPTPRMFALMSHLMRIYVVITEQMLSEHILPMPFPQITMTDSSLDDLTKKTYSIARRQSPTLKLPSKGLWASR 576
WhIV-6:YP_009301361 LTPKERELNPTPRMFALMSHLMRVYVITEQMLSDHVLNMFQITMTDSSLDDLTKKMYVTVKQSSNVGRKTGVKRRWASR 566
TpVA:AYH53279       LTPKERELNPTPRMFSLMSHLRYYYVITEQMLSDHVLKMPFPQITMTDSSLDDLTKKMYATVKNQSSRVKRHNKSGNWVSR 566
TYMaV:YP_009352236 LTPKERELNRVPRMFSLMSHLRYYYVITEQMLSDHILEMFPQITMTDSSLDDLTKRMYNTRVNSLKKRHNKKEGWASK 566
SAV1:QDJ94287       LTPKERELNSVPRMFSLMSHLRYYYVITEQMLSDHILEMFPQITMTDSSLDDLTKRMYSTVKNQSSLKKNMKSERGWASR 568
ADV:YP_009177021    LTPKERELNPTPRMFALMSHLLRVYVVLTEQLISDHILKYPFQITMTDSSLDDLTKKMFSTVKHQSVQNKKRGKDKTWASK 563
SCV:AWK49433        LTPKERELNPTPRMFALMSHLLRVYVVLTEQLSDHILKYPFQITMTDSSLDDLTKKMYSTVTRHQSVLNNKRGSDRTWASK 564
RVCV:QBS46644       LTPKERELNPTPRMFALMSHLLRVYVVLTEQLISDHVLRYPFQITMTDSSLDDLTKKMYSTVKHQSIQNRKRKGNLWASR 563
WhIV-4:YP_009300689 LTPKERELNPTPRMFALMSHLMRVYVITEQMLSEHILPMPFPQITMTDSSLDDLTKKMYSTVTRHQSVLNNKRGSDRTWASK 575
CCyV-1:ATS17313     LRSKERELNPTPRMFALMSHLMRVYVITEQMLSEHILPMPFPQITMTDSSLDDLTKKTYTTRVNSANKAK---SRLMASK 564
LNYV:YP_425092      LRSKEREMNPTPRMFALMSHLMRVYVITEQMLSEHILPYPFPQITMTDSSLDDLTKKTYSTVKNQAAKIRR--SGALYDTK 564
LYMoV:YP_002308376 LRSKERELNPTPRMFALMSHLMRVYVITEQMLSEHILPYPFPQITMTDSSLDDLTKKMYSTVKNQSVRKKQ--IGAI FDSK 564
TpVB:AYH53273       LRSKERELNPTPRMFALMSHLMRVYVITEQMLSEHILPYPFPQITMTDSSLDDLTKKMYSTVKNQSVRTRA--AGSIFDSK 564
BaCV:QAU20941       LYPKERENPVARMFSLMSLMMRAYIVVITEQMLADDILRYIPGVTMTFDLLELTKSMVQATRDMFVQGEH-----SR 594
RSMV:YP_009553369   LYPKEREENMVPARMFSLMSLMMRAYIVVITEQMLADDILRYIPGVTMTFDLLELTKSMVQATRDMFVQGEH-----DSI 566
CBDaV:YP_009362280 LYPKERELNPTARMFALMSLKMRSYVVITENMLSENILRYIPGVTMTYNNLDLAKEMIHSTSSQRLQGAT-----SR 562
MaCyV:ARS22495      LYPKEREMNPIARMFALMSLKMRSYVVITENMLSENILPNVPGITMTYNNLDLAKEMIKITRAQSGSQGEF-----SR 560
MYSV:ATN96453       LYPKEREMNPVARMFALMSLKMRSYVVITENMLSENILPYIPGITMTYSMLDLAKEMIRSTRAQSKQGEF-----SR 558
BYSMV:YP_009177231 LYPKEREMNPTARMFALMSLKMRSYVVITENMLSENILPHVPGITMTYSMLDLAKEMIKSTRSQKQGEF-----SR 558
* .*** * .***.** : * : *::*::*::: . :** * * :. .

```

```

TrARV1             TVCVSLDFEKWNGHMRMESTFYVFLALGELFGMSHLYNRYEIFEQSYLYLADGSYLVSVLDG--QLV--EDPPKSFSGH 651
WhIV-5:YP_009300875 TVCVSLDFEKWNGHMRYESTYHVKFALGELFGLPNLYNRYEIFEQSLIYLADGSYLPRYSGN--EMI--MEEP LSFTGH 650
PeVA:YP_006576506   TVCM SLDFEKWNGHMRK ESTEPIFTCLGDLFGMKELYNVYD L FSESYLYLADGSYVPIVKN G--ELI--EETG LSFSGH 652
WhIV-6:YP_009301361 VVCM SLDFEKWNGHMRK TMTKPVTSLGDL FGLSELNMYTYDIFEGSYYLADGT YIPKVDSSG-NLE--VTEPYSFTGH 643
TpVA:AYH53279       VICM SLDFEKWNGHMRK DMTLGVFSALGTLFGLPELNYR YD L FECCYYLADGSYVPRVDG--TLI--VEEPMSFTGH 642
TYMaV:YP_009352236 TVCVSLDFEKWNGHMRK SMTSGVFTAI GDLFGMGLFN VTYD L FSESYLYLADGSYVPSIDDG-NLV--VDEPKSFINH 643
SAV1:QDJ94287       TVCVSLDFEKWNGHMRK EMTQGVFTAVGELFGLPELNYR YD L FSKCYYLADGSYVPAVSEDK-ELI--VEEP LSFENH 645
ADV:YP_009177021    VICM SLDFEKWNGHMRK SMTSGVFTAL GDLFGMSEL NYQTYD L FEGCYYLADGT YVPGIEGD--KLL--VENPSSFTGH 639
SCV:AWK49433        VICM SLDFEKWNGHMRK EMTLGVFTIPIGDLFGMTELYNRYDIFSECCYYLADGT YVPIHKG--DLL--VSEPF SFQH 640
RVCV:QBS46644       VVCM SLDFEKWNGHMRK EMTSGVFTAL GDLFGLTEIYNYTYDIFKDCYYLADGSYVPTCDDHVGLS--IDEPKSF IHH 641
WhIV-4:YP_009300689 TVCVSLDFEKWNGHMRK SSTYVFLGNLFGMGLFN VTYD IFEQSYLYLADGSYVPSIDHG-ELE--VSEPR SFKH 652
CCyV-1:ATS17313     TVCM SLDFEKWNGHMRQEATLHVFKSLGELFGLDDLYHV TYDIFKESYFYLADGSYV PVFNANG-DFA--PEPPLSFTGH 641
LNYV:YP_425092      TVCM SLDFEKWNGHMRK ESTFHVFALGDLFGMENLYN ETYD L FKDSYFYLADGSYVPSMDAQQ-NFT--PEPPYSFTGH 641
LYMoV:YP_002308376 TVCM SLDFEKWNGHMRK ESTYHVKFALGELFGMTDLYN MYTYDIFKDSYFYLADGSYL PKINESG-DFV--PEPPFSFTGH 641
TpVB:AYH53273       TVCM SLDFEKWNGHMRK ESTLGVFRLGELFGMPELNYM TYDIFKDSFFYLADGSYVPMINEG-DLV--PEPPYSFTGH 641
BaCV:QAU20941       TFCINIDFEKWNLFNRKEATYHVFKNLGELYGLTQLYNR TYDIFRNSVLYLSDGSYLPLFDENL-NLVD--DGKGLAKTGH 672
RSMV:YP_009553369   TFCINIDFEKWNLFNRWME TGVFREMGRLFGLPEI FNRTYDIFKKSIIYLADGSFDLL L DTEL-IEEPGTPDCAYTG H 645
CBDaV:YP_009362280 TFCINMDFEKWNLFNRK ESTQYVFENLGRMFGLPSLYN TYDIFRNSL IYLADGSYTPK LSEGL-ECME--KDPD LAYTGH 640
MaCyV:ARS22495      TFCINMDFEKWNLFNRK EATYYTFL EGR LFGMPNLYN RTYDIFRNSI IYLADGSYTPD IDEDL-NLHE--KDN NLA YEGH 638
MYSV:ATN96453       TFCINMDFEKWNLFNRK ESTYYTFL EGVGLFGLPQLYNR TYDIFRNSL IYLADGSYTPSLDDKL-ESLE--KDLDL CYTGH 636
BYSMV:YP_009177231 TFCINMDFEKWNLFNRK EATYYTFL EIGR LFGLPQLYNR TYDIFRNSL IYLADGSYTPK LDDNL-DSLE--KQNQL CVTDH 636
...:**** :* * * * * : : : * * * : . ** ** :

```

```

TrARV1             KGGMEGLRQKGWTFITVCCLNMCVSKYNTYKIMGMGDNVQLQITLYTYKIDHRG-APTAEGNEMREILLSLFDDL VYT 730
WhIV-5:YP_009300875 KGGMEGLRQKGWTFITVCCLNMCVSKYNTYKIMGMGDNVQLQITLYTYKVAASG-EATEAGSYEMKLTLES LFADLIY T 729
PeVA:YP_006576506   KGGMEGLRQKGWTFITVCCLDVICSKYNTYKIMGMGDNVQLQITVTFYSVVDGAG-VATDGKREMRGIIYTLFDLLVST 731
WhIV-6:YP_009301361 KGGMEGLRQKGWTFITVCCLEVLISKYNTYKIMGMGDNVQLQITLYTNKIDL RG-SPTVEGLSDMKTTLGLIFDDLINS 722
TpVA:AYH53279       RGGMEGLRQKGWTIYTVCCLEVLISKYDCTYKIMGMGDNVQLQITVYSKLDTEG-KVTDEGVEEMRRLIKTIFDDL VHS 721
TYMaV:YP_009352236 QGGMEGLRQKGWTFITVCCLEVLISKYDCEYRIMGMGDNVQLQITVYSNI VDES G-KATAEGL LQMSGILDDIFKDLVRS 722
SAV1:QDJ94287       RGGMEGLRQKGWTFITVCCLEVLISKYDCEYRIMGMGDNVQLQITVYSNI VDES G-KASDAGVRQMAATLQRIFSDLVES 724
ADV:YP_009177021    KGGMEGLRQKGWTFITVCCLEVLISKYDCEYKIMGMGDNVQLQITLYTNKVDNAG-LPTDEGLREIKLMSNDIFHKL IET 718
SCV:AWK49433        QGGMEGLRQKGWTFITVCALEVLSQYDCTYRIMGMGDNVQLQITLYTGK LNDG-SASDEGLSEMKETMRRIF EALVNT 719
RVCV:QBS46644       KGGMEGLRQKGWTFITVCALEVLSRHDCCYKIMGMGDNVQLQITLYTNKVTD SG-APSEEG LTEMKNMNALIFKELVST 720
WhIV-4:YP_009300689 KGGMEGLRQKGWTFITIVCLDMICKRHNCIYNSMGMDNQVQLMLTYTYKVNLDG-SIKESGIQEIKEKHRLFDLLIDV 731
CCyV-1:ATS17313     KGGQEG LRQKGWTFITVGLDWICSEHNCIYKIMGMGDNVQLQITMYTNKVDIQG-APTEQGMQMDRTLKNVFDLL ET 720
LNYV:YP_425092      KGGQEG LRQKGWTFITVCLDWICRKHNCIYKIMGMGDNVQLQITMYTYRV DASG-KATERGREDMRRVLFGFDDL LDV 720
LYMoV:YP_002308376 KGGQEG LRQKGWTFITVGLDWICRKHNCIYKSMGMGDNVQLQITMYTYQV DSSG-KATKMGLEEMKTVL FGLFDDL LDG 720
TpVB:AYH53273       KGGQEG LRQKGWTFITVGLDWICRKHNCIYKIMGMGDNVQLQITFYTYRV DASG-KATTKGIEEMRKVLFGLFDDL LDV 720
BaCV:QAU20941       LAGLEGLRQKGWTFITAVVIAQC DKLGDYKLMGQGDNVQLVVKIYSSRARFLG-IEHPRS VQEEIEEKMEQLVKDLSTE 751
RSMV:YP_009553369   VGGFEG LRQKGWTFITAVLSSICDEMGIKTHLMGQGDNVQLMTIYSRAARETGDLKSAVAPALEITNTLETFKRSLVSL 725
CBDaV:YP_009362280 IGGFEG LRQKGWTFITVVLIAVYCDLGIHYKLMGQGDNVQLMVTIYSQHAKL SG-LDSPGSISEITSS LQTLXQLIEV 719
MaCyV:ARS22495      IGGFEG LRQKGWTFITVVMIAHICDQLGIQYRLMGQGDNVQLMVTIHSKHARL SG-IDSGASISEISSTLNKLLIALQEN 717
MYSV:ATN96453       IGGFEG LRQKGWTFITVVLIAHVCDQLGIQYRLMGQGDNVQLMVTIHSKHARL CG-IDSQASKNEISDKL KILIRRLQET 715
BYSMV:YP_009177231 IGGFEG LRQKGWTFITVVLIAHVCDQLGIQYRLMGQGDNVQLMVTIHSKHARL NG-IDSPGSKNEIAGK LQMLRLSLQDT 715
* ***** :* : * . * * * * : : * . * . *

```

TrARV1	FRQLGLPLKPLETWMSEDLVLYGKYPVLLKGVPLSMDLKKIMRMFHNSNDDIMTIENGMGTVYVGNAAASATQLSCCTIVPYM	810
WhIV-5:YP_009300875	FRQLGLPLKPLETWMSEDLVLYGKYPVLLRGVPLSMDLKKIMRMFHNSNDDVMTIENGMGTVYVGNAAASATQLSCCIVPYM	809
PeVA:YP_006576506	FGKLGPLKPLETWMSEDLHLYGKYPVWRGVPLCMDLKKIMRMFAYSNDDIMTIENALGTIYVGNAAASATQGTFCCLLPYI	811
WhIV-6:YP_009301361	FTLSGLPLKPLETWLSEDLVLYGKTPVWKGVPLSMDMKLMRFTFPWSNADVMTLENALGTISGNASSATQSTSCVWLPYL	802
TpVA:AYH53279	FTQAGLPLKPLETWMSEDLVLYGKVPVWKGVPLTMDMKLMRFTFPMNEDIMTIENALGTVSSNSTAATQASPCVWTAYC	801
TYMaV:YP_009352236	FTESGLPLKPLETWMSEDLVLYGKVPVWKGVPLTMDLKKLMRFTFPMNEDIMTIENALGTVSSNSTAATQASPCVWTAYC	802
SAV1:QDJ94287	FTASGLPLKPLETWMSEDLVLYGKVPVWKGVPLTMDLKKLMRFTFAMNEDIMTIENALGTVSSNSTAATQASPCVWTAYC	804
ADV:YP_009177021	FTDAGLPLKPLETWMSEDLVYVYKVPVWRGVPLMDLKKIMRMFPFNADVMTLENALGTISGNASSATQASACIWTYPV	798
SCV:AWK49433	FYSAGLPLKPLETWMSEDLVYVYKIPVWKGVPLSMDLKKIMRMFPFNADVMTLENALGTISGNASSATQASACIWTYPV	799
RVCV:QBS46644	FTDAGLPLKPLETWMSEDLVYVYKIPVWKGVPLMDLKKIMRMFPFNADVMTLENALGTISGNASSATQASACIWTYPV	800
WhIV-4:YP_009300689	FGELGLPLKPLETWASECLFLYKGFVWKGVPLSMDMKLMRFTFPMNEDIMTIENMLNTVAGAAAATQSSPCVLSVYL	811
CCyV-1:ATS17313	FNSLGLPLKPLETWIASEDLFVYKVPVWKGVPLSMDIKIMRFTFPMNEDIMTIENMLNTVAGAAAATQSSPCVLSVYL	800
LNyV:YP_425092	FSDLGLPLKPLETWIASEDLFVYKVPVWKGVPLSMDLKKIMRFTFPMNEDIMTIENMLNTVAGAAAATQSSPCVLSVYL	800
LYMoV:YP_002308376	FAELGLPLKPLETWIASEDLFVYKVPVWKGVPLTMDLKKVMRFTFPMNEDIMTIENMLNTVAGAAAATQSSPCVLSVYL	800
TpVB:AYH53273	FSELGLPLKPLETWIASEDLFVYKVPVWKGVPLTMDLKKVMRFTFPMNEDIMTIENMLNTVAGAAAATQSSPCVLSVYL	800
BaCV:QAU20941	FLKGLPLKPLETWVSDQFFSYGKVPVWKGVPLSMDLKKIMRFTFPMNEDIMTIENMLNTVAGAAAATQSSPCVLSVYL	831
RSMV:YP_009553369	FANLGLPLKPLETWVSEELFAYGKTPVYRAVPLAMSLKRISRVFAFVSNEDLMTIDNALGAIANSQAAMSSLYVVPYI	805
CBDaV:YP_009362280	FGDVLPLKPLETWVSDTFFSYGKVPVWKGVPLSMDLKRISRVFAFVSNEDLMTIDNALGAVTANSQASAMADVHPVPIYF	799
MaCyV:ARS22495	FQMVLPLKPLETWVSDTFFSYGKVPVWKGVPLSMDLKRISRVFAFVSNEDLMTIDNALGAVTANSQASAMADVHPVPIYF	797
MYSV:ATN96453	FDCVGLPLKPLETWVSDTFFSYGKVPVWKGVPLSMDLKRISRVFAFVSNEDLMTIDNALGAVTANSQASAMADVHPVPIYF	795
BYSMV:YP_009177231	FSEVGLPLKPLETWVSDTFFSYGKVPVWKGVPLSMDLKRISRVFAFVSNEDLMTIDNALGAVTANSQASAMADVHPVPIYF	795
	* **:*:***** * : . *** * : :.* .:*** * * ** ** * . * : : . : : . *	
TrARV1	VGIFMASYCACSFLDYHPLLGYGIMKEID-----SSHTWSLYMPGERTERYEMGARH--YSIQCL-RLLMQTV	875
WhIV-5:YP_009300875	VGIFMASYCACSFLEHPLLGHGILKEVE-----AKSTWDLFMPGKPRQFIPIGRRT--FNIQCL-RLLMQLV	874
PeVA:YP_006576506	VGIMMSSLCVMDFMFYHPLFGEGLNKHMA-----SDMAWNLSGDKIKTSKIKIGEHH--MNSMI-RLMIQIV	876
WhIV-6:YP_009301361	IGLIMSSLCITDFLSYHPLIGSSLSRRGS-----ELGHWLILAMSKLNKKYRTEMRG--LSTDHI-RLVIQIM	867
TpVA:AYH53279	VAIFMSSLCIQDFGSYHPLLGNGITSISY-----DKNWVWVLSHNEKIAYKIRTDG--VNEHDL-RILIQMI	866
TYMaV:YP_009352236	IYVLMSTLCIDDFLDYHPLIGDLSLYKTLID-----KDKHWVLRSHRFAIRYQLPKESWMTMTRHSL--RRAISII	869
SAV1:QDJ94287	IYVMMTSLCIEDYLDYHPLIGRGLMKEEAW-----K--KNDSWVLSKRTSSNIYYQLQPDSSWLGREKL-RLLISMI	874
ADV:YP_009177021	MCILMNSLCIKDFMTHYHPLIGELMSGMT-----GQEGWTLRTASGERYDFSAAGLTL--PGIRSV-SLCMQMI	864
SCV:AWK49433	VAILMNSLCIRDYLEHPLIGQGLMHVGT-----HCKEWRLLRPNGERHNFINTDTV--PSRRL--TLVMQMI	865
RVCV:QBS46644	MCVMMHSLCIRDYLEHPLIGELMSVND-----GEGTWLRTSEGESFSELPKQKG--LNRHL--NMLMQII	866
WhIV-4:YP_009300689	VGIFMFLCFARTQLLHYHPLMGKSVIYQDYAMRKGKIDARMWKAISG--RLRSFGLNQQ--VQISDL-VLLMITV	887
CCyV-1:ATS17313	IGLFMMSVCADDLDYHPLIGELGLKSLA-----DQPEWVITIKKEKPRKTKLGTWR--PSRMLI--RRLMSMV	865
LNyV:YP_425092	VGIFMISLCSHDMVYHPLIAKGLREVLN-----ENRTWGLKFKGAHQVKTDIKEM--VDETNL--RRLMMNV	865
LYMoV:YP_002308376	VGLFMSSVCTEDLLTYHPLIGKGLMEVLR-----DDDCWGLKFRGGHVSQTKVGVKVR--ITRGEI--RRLMQTV	865
TpVB:AYH53273	VGLFMGGLCSADLLKYHPLIAKGLMEVLR-----DDNEWGLKFKEGRSISTPVGPVR--ISKEDL--RRLMFLV	865
BaCV:QAU20941	MLKLQSLMCIQLFQSYHPLVGDAPDFNKPG-----HFKTRRTTEARPKSVYVND--LTDPSLFRKLAHSV	895
RSMV:YP_009553369	IAKWQHLCACIIFSNYHPLCGCAPHVSGEEW-----AIKRLRPSGKKIQETSDEE---IDERDL--MKLIVTI	868
CBDaV:YP_009362280	IAKWQQLQCLNVFSKYHPLVGGPPAVLGAPF-----SFSMRTKAGEFLKFYGEPI---ENQRSL--IKVFATI	862
MaCyV:ARS22495	IAKWQHLCCLIFSKYHPLVGDMSINTDEEI-----NFYMKLKGDESWNAKETRR---YSIDIL--LMVMATV	860
MYSV:ATN96453	IAKWQHLCCLSVFSRYHPLVGEPMMSQESII-----KFRMRDLKGEWSESETDQT---FPRETL--LMVYASV	858
BYSMV:YP_009177231	IAKWQHLCCLSVFSRYHPLVGEPMMEHEGTV-----KFMKLDKGEKWESESQI---FSRETL--LMVYASV	858
	: ***: . :	
TrARV1	PRTLGGYNSLNIFEIIMRGFPDNLSDRLTYIYQITRSKG--EDSLDITLKNWMPYIMRTKNYKLLIEDVTSVNLMPRSP	954
WhIV-5:YP_009300875	PRTLGGYNSLNIFEIIMRGFPDNLSDRLTYIYQIVRLDK--RDALDEALTNWIHPYMPDKNFKLLIEDVSSVNLSPRSP	953
PeVA:YP_006576506	PRTLGGYNSLNMYELMMRGFPDNLSDRLTYIYKILMGS--PPGFENCLKNWLAPLYMPTKNYKLLIEDVSSVNLSPRSP	954
WhIV-6:YP_009301361	PRTLGTGHNGLNLYEFMMRGFPDNLTRDVSYLTAVLKCPSPAKWLKSVIENWLSPIFMPDNTYSTLLQDVTAVNLSPPRAP	947
TpVA:AYH53279	PKSLGGYNGINIEYELMMRGFPDNLSDRLCYLSKILTHGVMPDMRSCIECVWTPMYPGINYSTLLQDVTAVNLSPPRSP	946
TYMaV:YP_009352236	PKSLSGYCGANIYEMMVRGFSRDLSDRLSYLNNIVNGGNTSGETATLNNWLNPLMPECNYSMLLEDVYAVNLSPPRSP	949
SAV1:QDJ94287	PKSINGYCGANIYEMMVRGFSRDLSDRLSYLNNIVNIRLSSATPAWIEVYLERWISPIYMPKESYLLLEDVCAVNLSPRSP	954
ADV:YP_009177021	PRTLTYGYNVNIEMMVRGFPDNLSDRVSYIMSVIKSAAAPHWLAELLRWVVKPIYMPINYNATLVQDVTAVNLSPPRSP	944
SCV:AWK49433	PRTLGMGYNINILEMMVRGFPDNLSDRVSYLRSVEDCTSCPDLKIIHAWLQPIYMPHINYATLQDVTAVNLSPPRSP	945
RVCV:QBS46644	PRVLTGYNGINILEMMVRGFPDNLSDRVSYLTAVLKSDKTPSWLRELTNWIYPIYMPQINYANLVTDTVAVNLSPPRST	946
WhIV-4:YP_009300689	PRSLGGYVYTNLPSIMIRGFPDNLSDRLSYIMGMIRSSL--NLRISAYLENWSRVIFMPEINYKMMMEDILSVNHLNIPIT	966
CCyV-1:ATS17313	PRVLGGYVSFNLFGLLMRGFPDPSLALSQLYAYGVKDDADDLVLILKRWCDPIYMPDVSLLKLLIEDVSSVNLAPVTP	945
LNyV:YP_425092	PRILGGYVSFNLVGLMMRGFPDPSLALSQLYAYGVKDDADDLVLILKRWCDPIYMPDVSLLKLLIEDVSSVNLAPVTP	944
LYMoV:YP_002308376	PRILGGYVTFNLWLSLMMRGFPDPSLALSQLYAYWDLQSA--TDRTSKYLLRWVKPLFMPERSMKLLIEDVSSVNLAPVTP	944
TpVB:AYH53273	PRILGGYVSFNWGLMMRGFPDPSLALSQLYAYWVRNP--RTTTESYLLRWVKPLFMPERSMKLLIEDVSSVNLAPVTP	944
BaCV:QAU20941	PKIFGGFNANAFHMMIRGFPDRASMDASWIYAMVQQT--DTPFTRSLMNLRVVFDVSKDYSYLQDPESINIVMPPST	973
RSMV:YP_009553369	PRSLGGYNTLTYLEMIMRGFSDPVRDMCWLFAIASES--TGKLRGYLINWIKIPVSPVNAQHLIQDPTALNLLVPPNA	946
CBDaV:YP_009362280	PKTLGGLNIITYFDMIRGFSDDPPCRDQWLSLLVDHS--PADIRPYLRNWHLLLNDSVDYLHLQDPTSLNLLVPPNS	940
MaCyV:ARS22495	PKTLGGLNIVTFDDMILRGYSDDPPMKDLQFLYLMYKGT--RGELKHAIKNLRVCLSPSIDYLHLIQDPTALNLLVPPNS	938
MYSV:ATN96453	PKTLGGLNIVTYFDMILRGYSDDPPMKDYQFLTLLEGA--TGNLRRGLINWIKLLSPSVDYLHLIQDPTSLNLLVPPNS	936
BYSMV:YP_009177231	PKTLGGLNIATYFDMILRGYSDDPPMKDYQFLNLLTEGS--TGQLRQGLVNWRRVLLSPSVDYLHLIQDPTSLNLLVPPNS	936
	*: * . :*:*. * :	

TrARV1	SSGIRQAVVERFISG--GRAVKNHEFRDLMSKIKIQAERLAELKCSGKDLHIRLLHDIYASTIIGYVDGIISKVTKASTI	1032
WhIV-5:YP_009300875	SSGIRQAVEKFIGS--GRTVKNLEFKDLMSKIKHQADMLAELKCSGSDIHIRLMHDIYASTIIGYVDGILSKVTKASTI	1031
PeVA:YP_006576506	LAGIKQTVKFIISD--PKRIRNPEFRGLMSGKIRDQAQELAASLCEGDDLHIRLLHDIYATIIGYVDSILSKVTKSSTI	1032
WhIV-6:YP_009301361	TAGIKQTVRYMTD--ERLIKNEEFRDLMATKVKKEHEEYLSECLCEGKILHIRLLHDIYATIIGYVNSILSKVTKSTI	1025
TpVA:AYH53279	TSGIRQLVANYMSS--GVNITNEEFKELMSSKYKNMSDYLSECLCEGEEHLHIRLLHDIYATIIFGYVDGILSKVVKSTI	1024
TYMaV:YP_009352236	LSGVRQVVQRYLNS--GMKIENPEFMQLVRAKNDQDCRYLAELCEGRELHIRLLHDIYATIIGYVDSILSKVTKTTTI	1027
SAV1:QDJ94287	LAGVRQVVARYLSG--GMRINPEFMQLVRAKNDQDCRYLAELCEGRELHIRLLHDIYATIIGYVDSILSKVTKTTTI	1032
ADV:YP_009177021	SSGIKQVVTYLMSS--GASIKNEEFKDLMLTKDKLHEEFLSELLCTGDDLHIRLLHDIYATIIGYVDGILSKVVKNTI	1022
SCV:AWK49433	SSGIKQVVTQYLG--GATKIRNQEFKGLMEVKNKNEEFFSELLCEGEEHLHIRLLHDIYATIIGYVDSILSKVVKTTTI	1023
RVCV:QBS46644	SSGIKQVVTKYLSS--GKDIRNEEFRLNMQTKQKKEEFMAELICCGPELHKLHLDIDATIIGYVDSILSKVVKNTI	1024
WhIV-4:YP_009300689	MSHIRQTVAQFLSS--HRKVKNKEFVLDMAVSKDPNKELLSNMLCSGRKLHIRLLHDIYESTIIGYVDGIISKVTKSTI	1044
CCyV-1:ATS17313	TAGLRRVVEKYLAE--GRAIRNQEFKDLMKTRVADVEDVIADQLCKGDTLHIRLLHDIYATIIFGYIKSIVSKVTKSSTI	1023
LNYV:YP_425092	TAGLRRVVEQYLT--GRVIKNAEFRDLMSRDNELEVDISEHLCSGEHLHIRLLHDIYESTIIFGYIKSIVSKVTKSSTI	1022
LYMoV:YP_002308376	TAGLRRVVERFLSD--GRVIENSEFRDLMSRDPMEDVISEHLCSGDHLHIRLLHDIYESTIIFGYIKSIVSKVTKSSTI	1022
TpVB:AYH53273	TAGLRRVVEQFLSD--GRVIKNESEFRDLMSRDPEVEEIISEHLCSGRTLHIRLLHDIYESTIIFGYIKSIVSKVTKSSTI	1022
BaCV:QAU20941	KTIKINMVRDQLRGIKGNTAESELFNVLDHTSRTHVEDLVCRLTNSDLSNVRFLHDIYESTIIFGYIKSIVSKVTKVTL	1053
RSMV:YP_009553369	TSVIKRMIDKSLAALPKRSQFAVWFSEILEISGDKIEISKLAEALTRTDSLNRFLHDIYESTIIFGYIKSIVSKVTKVTL	1026
CBDaV:YP_009362280	NTIIRKMIQQTIAKLPKDESEFASWFQELMGISADKAVEPIVEKLTASAEINVRLLCHDILGSTLFGYADSIASKVTKVTL	1020
MaCyV:ARS22495	NTIIRKMIHSTIDSIDSESEFSTWFKELMGISAEKKEGIVAKLSAGPEINVRLLCHDILGSTLFGYADSIASKVTKVTL	1018
MYSV:ATN96453	NTLTKRMIHNTIEEMESESEFAGWFKELISISAEEKMDGIVEKLTSGEEINARLCHDIMGATLFGYSDAIASKVTKVTL	1016
BYSMV:YP_009177231	NTMIKRMIHSTIEGIDSNSSEFSGWFKELIGISSEKMDGIVEKLTSGEEINARLCHDIMGATLFGYSDAIASKVTKVTL	1016
	: : . : : * : : : : * : : : * : : * : : *	
TrARV1	QRLAVQTDQRDIMKTVAADEVNFRYFVWRSGVR--PDGNIEECPVSMAQRMRREGWGKLLRGVSTPFPMAVLSPTDCNSN	1111
WhIV-5:YP_009300875	QRLAVATNEKDIMQVVNDEINFRFRFLWRSSTR--PNNEISECPTEMAKRLRSESWGKSLRGVTPFPMAVLSPTRCGET	1110
PeVA:YP_006576506	QRLAVSKSKFSDMSAVYKDEINCFRFLWRCVSR--SDEAISPCTTHAKMKREKSWGKLRGITTPFPMSYLARTECGEN	1111
WhIV-6:YP_009301361	QNLAMAKSSFDVFDVDCRDEVNVCNYVWIRSSAK--GMQLSLSLCPTEFCKEIRLIGWGKDLRGVTPYPASYLESNCDVT	1104
TpVA:AYH53279	QKLAINDSRGAVFKSIEERDEVNFHRYFRWRVCVQM--ESAFADCEPTERCKRIRREGWRKELRGVTPFPASFDKSNCGRQ	1103
TYMaV:YP_009352236	QKLAIQSDSVKVFDTIMKDERNYFAFFVWRCSFQS--GTFPETERCATTQCKLIREGQGWKKIIRGVTPFPPLSYMQVTCSTA	1106
SAV1:QDJ94287	QKLAIQSDSVRVFETIESDETNFYKFFRWRCSFAE--GDPLVSNCACTQCKVMREKGGWGTIRGVTPFPPLSYLIETDCGRQ	1111
ADV:YP_009177021	QRLAMSQSSKDFDVIIESDEKTYFKFRWRCSAQE--GDTLRSDCPTEMCREMRQEGWKKLRGITTPHPHFSFMKETLCDHI	1101
SCV:AWK49433	QRLAMKSSSRDIFDVIDTDDDEMGYFNFKWRCTVS--GEETGIRCPTIEICRNIRASGWQKTLRGVTPHPHFSFMREQTCNEV	1102
RVCV:QBS46644	QKLAMQHSIEDIFKTIYRDEKTYKFFKWRCSCTR--GEVPSLCPDIDICREMRLIGWQKELRGITIPHPFSYMEETSCDHR	1103
WhIV-4:YP_009300689	ANLAVRHAQKINDAISNTEINRYFVWRCSFHF--GQPWETDCPTRYAKHVRFEFGWKRELGKRVTPFPPLSYMQVTCSTA	1123
CCyV-1:ATS17313	LSIAVDKSTRDPLAKVIEDERNYKFFVWRCSVE--GGNLLPDCPTDLAKSMRLNGWGKSLIGVTVAFPLSYLTKTTTCYRK	1102
LNYV:YP_425092	VSLAIGKAKGDPDLTRLMSEENYFRFFMWRCSVIN--PHYDIPACPTDLAKQVRRIGWGKELIGVTVAFVPSWFLKKTDCYES	1101
LYMoV:YP_002308376	VSLAIGKTKGDPDLRRLMLDEENYFRFFLWRCVSE--PTFELPDCPTDLAKQMYRLGWGKELIGVTVAFVPSWFLTKAQCMEN	1101
TpVB:AYH53273	VSLAIGKSRGDPDLRRLMDEENYKFFVWRCSVE--PDYDLPCPTDLCKSMRREGWGKELIGVTVAFVPSWYLRKTECFDQ	1101
BaCV:QAU20941	SRMTS--FSHNVVGLAKGKENVWDFYFKWRLLHKKGRKVKETCTEFCVFNIRDEGWKKIIRGVTPFPFHLTRDRS---	1130
RSMV:YP_009553369	SRMAL--ASKDVGALIKGEMRLYSYFGWRTLQSRGLPLTTRCPNKKVRIIRDISWQKQIKAVSVYPYTHFLSEDIS---	1101
CBDaV:YP_009362280	SRMTV--GKEDVIEALVMGERRRWNYLGWSTSSAGEAPPSSCPSFNIRYRDKGWKKRVLGVSTPFPFHLTRDRS---	1095
MaCyV:ARS22495	SRMTV--ARNDDVVTSLARGEIRMWRYFEWRTHYSRGERFQTKCPATQIRYLRDKGWKKIIGISTPYPFHYITLNER---	1093
MYSV:ATN96453	SRMTV--AQQDVVSSLVNGEKRIWNYLWRWNTYKGGQPRISECPKQIRHLRDKGWKKIIGISTPYPFHYMGGLG----	1090
BYSMV:YP_009177231	SRMTV--AQQDVVGSVLVAGEQVRWRYLWRSNRYRGGVEVIRISCPKQIRYLRDKGWKKIIGISTPYPFHYIGGIE----	1090
	: : : * : : * : : * : : * : : * : : * : : * : : *	
TrARV1	--NACSCDTGYSIHFPDRQETN--TAWNLTGTNAPYLGSIKKEKVTGLGSKISYSEPLVRRPLNLLRAINWVFPKSV	1188
WhIV-5:YP_009300875	--GNLCMDGYSIHLSDSQTEN--VRWNRTLSGNAPYLGSIKKEKVTGKSKVYSSEPLIRRPLNLLRAINWVFPKSI	1187
PeVA:YP_006576506	--KLCNCEGDYISVHFDPKQINE--YMNWNTIGRNPPYLGSMTEKKEVITGLVKGAYSGRPLRGLVGRVWVPPSSN	1188
WhIV-6:YP_009301361	--DSCMCDGYSVHFDPDDQITN--EMWFMFDIGGNPPYLGSTTKEKVVIGTGRRVYSSEPLVLRPILNLLRINWVFPDSDN	1181
TpVA:AYH53279	--GTCDCADGYMSVHFDPDQQLTN--RMWNLDIGGNPPYLGSTTKEKVIIGAGGKIYSGEPLVRRPITLLKTINWVFPDSDR	1180
TYMaV:YP_009352236	--TGCDACDGFISVHYPDKQMPN--ESWCSIDIGGNPPYLGSMTEKVVVGTGGKIYSAEPLIRRPIRLNLLRINWVFPQHS	1183
SAV1:QDJ94287	--GVCDCQDGYSSIHYPDKQLPN--DMWNEDIGGNPPYLGSMTEKVVVGGGKIYSAEPLIRRPNLLRINWVFPDSDN	1188
ADV:YP_009177021	--IGCDCSDGYMSIHLDPDQMPN--DVWLNSIGGSPYLGSMTEKVVVGGGKIYSAEPLIRRPNLLRINWVFPPESEN	1178
SCV:AWK49433	--KGCNCQDGYMSVHLDPDQQLPN--SVWMTSIGGNPPYLGSMTEKVVVGGGKIYSAEPLIKRPNLLRINWVFPPEESI	1179
RVCV:QBS46644	--KGCDCIDGYMSIHLDPDQQLPD--DMWNMSIGANPPYLGSMTEKVIIVGAGGKIYSGEPLIKRPNLLRINWVFPEDSV	1180
WhIV-4:YP_009300689	LGSTCKCNDGYSVHLSDDVNYN--EWNASIGKSLPYMGSMTEKVLVQSGIKIYSSEPLIKRPLRLMRAINWVFPPEEN	1201
CCyV-1:ATS17313	--DQGCNCEGDFISLFLPDNNVPT--KEWNATIGRNPPYLGSMTEKVLVSSGAKIYSGEPLIKRPIELMRVWVFPPESEN	1180
LNYV:YP_425092	--GCLCTCDGDFISLFLPDSPVTP--QEWDRSITGNPPYLGSMTEKVVVISTGKSVYSGEPLIKRPIELMRVIGWVFPPEESE	1179
LYMoV:YP_002308376	--GSFDCDCEDGDFISLFLPDSPVTR--DQWNLGIGTNPYLGSVTKEKVVITGSKIYSGEPLVLRPILNLLRINWVFPPEESE	1179
TpVB:AYH53273	--GRLCDCLDGCISLFLPDSPVTR--DSWNYEIGTNPYLGSVTKEKVVVITGARIYSGEPLVLRPILNLLRINWVFPPEESE	1179
BaCV:QAU20941	---IQERVGSYLYAVISEYALRDPGYLFLTLGKSLPYLGSIIKEKLRSEVSLAFGSEPLISRPVRLRTIGWFIPESSN	1207
RSMV:YP_009553369	---ETDRPDSWIECYIDDAPTSVDRSCMIYSTGKALPYLGSVTEKELTRGAKAAYGTEPLIRPILNLLRINWVFPPEESN	1178
CBDaV:YP_009362280	---KTNRPDSEYEVVMDVGVNDRNTNALLMSTAGSSLPYLGSVTEKELQASATRAAYGTEPLIARPNLRLRIGWFIPEKSN	1172
MaCyV:ARS22495	---DTRPDSYIEVVISDRVLVSHPQTLINSSGTSLPYLGSVTEKELHSTAARAAYGTEPLITRPLRLMRAIGWFIPEKSN	1170
MYSV:ATN96453	---DTRPDSYIEVVDLVLVSHPDRLLLSGASLPYLGSVTEKELHNTAARAAYGTEPLITRPLRLRIGWFIPEKSN	1167
BYSMV:YP_009177231	---DTRDTSYIEVVDLVLVSHPDLLQTAGSSLPYLGSVTEKELHNTAARAAYGTEPLITRPLRLRIGWFIPEKSN	1167
	. . : * ** : * : : * : * : : * : *	

TrARV1 TASVIRSLTYAVTDVTKVKFEGLSEGTAGA EHVHYRDMSLKHGALASSNYLYTRMHISTDNFTKYSKGGNDYDIHFQAC 1268
 WhIV-5: YP_009300875 TASVIRELTHAVTDINPTS YEGQSEGTAGA EHVHYRDMSLKHGALSSSNLYPTRMHISTDNFVRYSKGGNDYDVHFQAC 1267
 PeVA: YP_006576506 TADVIRSTVRAVTDTPD HHHEGKVEGTAGA EHVHYRDTSLKHGALTSSNLYPTRYHISTDHFTRYSKGGENDYDVHFQAC 1268
 WhIV-6: YP_009301361 AARLILSTVS AVTDLDPHPYQG ILEGTAGSEIHR YRDSCTSHGTLTSSCF LFSTRYHLS TDQLTRYSKGAENTDIHYQAV 1261
 TpVA: AYH53279 TA EVIKTCVKA VTDLDPEPYQGV SEGTAGSE VHYRDSSTH GALTSSNFLFSTRYHISTDNFTRYSKGSENTDIHYQAL 1260
 TYMaV: YP_009352236 MAKIE EACVSSVTDMETS KFKGMEEGTAGSEAHRY QDSSTFRGALS SSNLYSTRCHISTDLSVRYSKGAENTDFHYQAT 1263
 SAV1: QDJ94287 MAKVIEK CASSVTDLEIERFKGMVEGTAGTAEHR YQDSSTFRGALS SSNFLFSTRCHISTDNLTRYSKGAENTDFHYQAA 1268
 ADV: YP_009177021 TANIIRMLVASVSDIDPEPYAGVSEGAAGAEVHYRYKDSSTSHGALTSSSYLLSTRYHISSDHFHYRCRGSNDTDLHFQAL 1258
 SCV: AWK49433 VSNVIKSLNSAITDMDPTPYVGVREGSAGAE LHRYKDSSTSHGALTSSSYLLSTRYHISSDHFHYRCRGSNDTDLHFQAL 1259
 RVCV: QBS46644 TAEMIKR LVSSVSDIDPEPYQGVSEGVAGAE LHRYKDSSTSHGALTSSSYLLSTRYHISSDHFHYRCRGSNDTDLHFQAL 1260
 WhIV-4: YP_009300689 TAKVIEK CVQSVTNISTEQFKGVEEGTSGAEIHR YNDSSLSHGTLTSSNLYSTRYHMSTDLNFKYSKGGENDYDVHFQSM 1281
 CCyV-1: ATS17313 TADIIRSCV GAVSDLNPNFRGITEGSSGSEIHR YKDSLKHGALCSSNLYSTRYHVSTDTFSRYAKGQNYDMMFQAN 1260
 LNVV: YP_425092 TAKIIMSCV SAVSDVNPLIFKGMTEGTS GSEIHRFRDTS LKHGALCSSNLYSTRYHVSTDTFTRYAKGQNYDMLFQAN 1259
 LYMoV: YP_002308376 TAKII QSCVTA VSDIDPLQFRGVT EGTS GSEIHRFRDTS LKHGALCSSNLYSTRYHVSTDTFTRYAKGQNYDMLFQAN 1259
 TpVB: AYH53273 TAKVILQ CVRAVSDIDPLRFRGVT EGTS GSEIHRFKDTS LKHGALCSSNLYSTRYHVSTDTFTRYKGAQNYDMLFQAN 1259
 BaCV: QAU20941 WAELKQLVK AVADIDPEILITVPDEVKGSM EHR YQDFTSHGTFWGT LFGPCTHVMN STNKWNTNQK DGVNFTLQYQAV 1287
 RSMV: YP_009553369 FAELIKMLLGAVTDLP IEEVLYIPEMVSGS MAHRYLD MStQHGSL WMPLYGPAT FLHMSTNTFVQYLKGTENVTLHFQCV 1258
 CBDaV: YP_009362280 WSVTLRNL LLSVTDLP SVVSIPEHVKGSMAHRYLDFVLKHGSLWMSLFGAPSHLSMSTNSL TEYARGSKNVTFFHFQAA 1252
 MaCyV: ARS22495 FATSI RSLLTAVTDL PADDI IVIP EHVKGSMMHRYQDMATD HGALWMP SFGPPSHLSLSTNSFAEYAKGSKNVTLHFQAT 1250
 MYSV: ATN96453 LAKSIENLLKAVTDLDP EIVIP EHVKGSMMHRYVDMALKHGS LWMP SFGPPSHLSISTNYFAEYAKGSKNVTMHFQAL 1247
 BYSMV: YP_009177231 LAKSIENLLRAVTDLP SEV IVIP EHVKGSMMHRYVDMALKHGS LWMP SFGPPSHLSISTNYFAEYAKGSKNVTLHFQAL 1247
 : : ::: : * : ** : * :* : : . : : * : : * . . *

TrARV1 LC SLTEWSNMLLHHWFYS --GANV PKFLHFQKTC DCLCVNK IDESFV DLEDENT PYYIPSRK TNPYLFVHQESLVEVHKMK 1346
 WhIV-5: YP_009300875 LC SLTEWCNMR LMSWLKY --E EWIPKVIHFQKTCYCCVTK LDES FVDLP DPL TNRYIPSRK TNPYLFVSGKLLTEVHKMR 1345
 PeVA: YP_006576506 LCSIVEQ TNIIWISDQNRQ --CNPISK FQH YKQSCYECVNI VDETFDIDIPDKRT S DLI PHRKENNYLVKEDQM QDTLAYS 1346
 WhIV-6: YP_009301361 FSWMV ELCNMYVTNRIG --GDVVTKFKHYKQ SCYRCINPIPEDF IDIPSGKAAD TVPSQKDNPPYL YASKATIRISEKIS 1339
 TpVA: AYH53279 FCYILEMSNMMI SNKLRT --GDTVPRFDHYKQNCYCKC INPIPEDFKD LESLKA VDAIPKRVTN PYLVFVKEKIRILERIS 1338
 TYMaV: YP_009352236 FCVIL ELSNMYLSNKIRG --EEVVARFKHFQR CCEIHP IEDFV DLSSEKALSVIPSFKNP YLYTPSSRI RIVERIS 1341
 SAV1: QDJ94287 FCVLL ELSNMYITNKLR --DDVICRFKHFKQC CYECIHPVEDDF IDLP NAKAVNVLPSYKDNEYLYTSSSKIRILERIS 1346
 ADV: YP_009177021 YCYLVERTNLEIVKAI RN --QDP I KRFKHYKQC CYCVKVPVQD FIDIPDDR VL TAVPSRKS NKYL FVKEKIRILEDRS 1336
 SCV: AWK49433 YCYIVEMTNLELHKCFRS --GAVIPRFKH YHQHCNEI KTVKEDFIDLTRER IIETIPSKKSNPYLFVSSNTIRVLEDRS 1337
 RVCV: QBS46644 YCFLVEMTNLEITRCLQT --SALIPRLKHYRQRCRDIKPVKEDFVDIANDRILTAIPSRKS NIPLYVKS D KIRILEDRS 1338
 WhIV-4: YP_009300689 LCLISEFNIEV FQITSGFRSTR LPRVTHWKET C NRCTSVWDDSFSDIK FCSWR YIPSKPNQNYLFYTPSSRI RIVERIS 1361
 CCyV-1: ATS17313 LCVIVESTHLDVIDLNAS --GEITPKV THFNQSCYHCISPLDETFHDLRDGEAARV IPSNK TNRYLVYVRAEKVSLKLYLP 1338
 LNVV: YP_425092 LCAIVETTHQYV LKTNQ S --KEPQRKTHHYKQTC YSCINPLDES FYDVKS KLSQLIPS KKTNKYL YVPEAKISMTLEHV 1337
 LYMoV: YP_002308376 LCAIIEGMHQYIIATNKS --NMMQK THHYKQICYSCINPLDEEFYDIQSS RPL LIPS KKKNKYL FVPKEKISMVLEYL 1337
 TpVB: AYH53273 LCAIIEGMHQYILRTN NL --GLMQK THHYRQVCYSCINPLDEEFHD IETSR LPL LIPS KKTNKYL FVPEEKISMVLEYK 1337
 BaCV: QAU20941 LCFM ---QAVI INNSTT --PSA -TKMMKF YECSCQICV PVETDFDFPELQEPVP PPSIPE S NPYLF LQKDDVDMQEDI 1360
 RSMV: YP_009553369 MGLI ---QYAI VNKALG --ECPTKRMTRFR RSCPDCIVPIDD TLEDLPEVPSLD LIPERTTNPYL YLKKEKIELNVRHR 1332
 CBDaV: YP_009362280 LCLV ---QFASLNINMS --DLP -RKVMRFYRGCPHCIT PVDEPNNDLES PVLPEDFSRPGN PLYL FIGKEEIELIHSQT 1325
 MaCyV: ARS22495 LGLI ---QFASINRAMS --SNH -RKMLKFRSCTEC ILPV DNPKEDI EALTVDDIPSRK NPYLVYDKEQH IHLVHKL 1323
 MYSV: ATN96453 LGLV ---QYSVINRCMS --E EP -RKILRYRYTCKDCITPVDEPSEDIRE SLEQEI P SRDPNPYL VVKEKIALIHKRD 1320
 BYSMV: YP_009177231 LGLI ---QYSVINKCMS --G EP -RKILRFYRTCPDITPVDEPKEDIA EILSDKEI P SRDPNPYL VVKEKIALIHKRE 1320
 : : : : : : : * * * : : : * * ** : :

TrARV1 PWFSTALS YIRSE EYQ LKSESKY GVLVDTVADMV YLDIATSK -GEETS FNTCLTDVKAYERTMFLKLD PNDVFRSVLDR 1425
 WhIV-5: YP_009300875 PLLSRLSV MPVEE YVRMDCKDKMRWLVDVSDKI YLDIISQR -GEDNSFSLCLTDVKAYERTMFLKLSPREVYLCLIDK 1424
 PeVA: YP_006576506 PLISLKLTYL TNMQYQLMEASTKYRVL ETIADL VASDICS GN -QNESFFSTGLTDVKS YERTMFLKLNPKDLFFSVLNR 1425
 WhIV-6: YP_009301361 PLFVMSPREFSYDQYQSM TNRQLVWL HDIIADRVSDVVS NS -SDDTAVSIGLHDVRSYERTMYLKLNP KYL INKVL SR 1418
 TpVA: AYH53279 PLS ELSERSLE FEDY LKISGSDKMIWLQDVICDRIVSDLMGNS -ASDTYVS VGLLDVKS YERTMYLKFSPQYMKR LISQ 1417
 TYMaV: YP_009352236 PLYELSDRELSHEDYDNISGRRAKAVL LHR SICDRII KDI V TGO -KSETHSV VGLTSVKAYERTMYFKLDP RIMVD TVMME 1420
 SAV1: QDJ94287 PLYDR TADVV TSEEYDKMGGA EKSRLLHRTIADRVIKDIMKGR -GTDHTVTVGLTSVKSYERTMYFKLDPKVM IDTVMSG 1425
 ADV: YP_009177021 PLSRL NAIAMETQEYDAMS PRMKRLWLQDIADKIVSDIMGQS -SEDT SMEASLLDVKSFERTMYLKYEDR YIIDCV MGG 1415
 SCV: AWK49433 PISRLNAE EYTS SDYDTASEK LKRNWLQDILDKITS DIMRMS -NDETS LSNLLDN GSFERTMYLKLDPKIYIDGVMGN 1416
 RVCV: QBS46644 PLSRISP KLES S DLYFTMSGYRKR LWLQDI VADKIVAAMIGIG -DGENTMDTGLVTNPS FERTMYLKI DPKYVIDAVMGG 1417
 WhIV-4: YP_009300689 PFLSMMNRMYTEG EYNRMSDVTKYCW MIDSISDKIVS QISTTQDAGELDDQVDLHGDDIFNRVAY LKVDPHDLFE E VSNK 1441
 CCyV-1: ATS17313 PFPGWITGMTPPSEMLRMTNL ERYK TWL DESVVDN ICDIQGAT -SESNYLT TALLDIKEHNL FYLSVSP EAVYNTLCSR 1417
 LNVV: YP_425092 PAKSWEFGLTSSDGFQ LSVNLRLQWLD TAVADN VVIDL L NPA -GEESYTTTALMDIKEHNL FYLTARPKDVYDQVCNR 1416
 LYMoV: YP_002308376 PHAGWDLGK LSDDALDS MNPRTKLQWL TDAIADNIMIDVTGPA -GEESFTTVSLMDVKEHNL FYLTARPKDVYDQVCNR 1416
 TpVB: AYH53273 PHANVWFGDLSEEFERMGDVKLDWLD TDSIADNVIDITGTGA -GEESFTTVSLMDIKEHNL FYMTASPRDVFQLCNR 1416
 BaCV: QAU20941 FTSLKGRSKVSTENFAE -NSRDMY WYLVESLASKFSYSGSTAG ----NIQSDPLDMNSKNKSSFKKIRPADLFRVVLNR 1434
 RSMV: YP_009553369 LAVINEIRVIA RDEIEE -TPLANNLTDVMSLRAAQ RIFYTA -KGKE ---AQWDLQTADREGYLKLD FVSVIRKIIGH 1406
 CBDaV: YP_009362280 ANLFEEIKRLGKDDMIK -SPGLSR TLLTEV VASKSAQA ILRND -SSRQ ---DAGLSDIGGMSR T VFLKLSVRNV FVTTLKF 1401
 MaCyV: ARS22495 LESYENIP SVSKQFIVD -NPGIGR NFLMEILAIKVASNILFAN -KTE ---LGEITDVSIGIRTVFLKLNPLHFTLTVLKA 1398
 MYSV: ATN96453 LDNYEISV VNKSHLIQ -YPYIGR KFLTEVLAVRAASSILYQS -SND ---ESGVFDISGVSRTIFLKLDRVFLLLITIKM 1395
 BYSMV: YP_009177231 LDNYEISITVIHKPFIK -YPHVGRRFLTEVLSVRTANAILYQT -SLE ---DSGVLDISGVSRTIFLKLDRVFLLLITIKM 1395
 : : : : : : : . : .

TrARV1 LNVHAKMTGDGT -DSSILPDEKRRNKIMSLLEYCPASTFLGLGMFYSWETLKKMN -FSSYCHSPNTFPMTIDSMCNAA 1503
WhIV-5:YP_009300875 LLRHSDMTMKT -VSGRRVAGKIQVDRALALIHSPVSSFLGLGMFYSWSETIQKMH -FSSFCVMPNTPFITTESMCAA 1502
PeVA:YP_006576506 LRTFARWKCAS -KVS RMP TI KD IQR YLESYLVRVESDCMSGLSMFYGWEET YKMMN -FHRAIILPNTVPASLFSSCESV 1503
WhIV-6:YP_009301361 LLIVAEWGVKIS -GHVGLSMDDDIKR TLINILENANPTSFLGLGMFYCWEESSRMD -VYPEMIRPNTNPISSASACEAM 1496
TpVA:AYH53279 LWRVARWKSREIKGLVKYVGDADVLRSFNTMMSTSSDDHAFMGMIAMFYCWEETS KRIN -IYPEMVPNTNPNVSISSACQAM 1496
TYMaV:YP_009352236 LRRVSKWTVLRSHPERTDASDNEIERVMIDILNSADTHGFLGLAMFFCWEETS LAYSRIYEVVPPATNPISVFSACESV 1500
SAV1:QDJ94287 LREICGWTVLRSHPERTKPSPEISRVLINCLNSADTHCFGLGMFYCWEETRRFRSRAYAEI VPPSTNPNVSVSSACEAI 1505
ADV:YP_009177021 LSLCAEWRWLESTGHQKDC TGEISRSIIISIVRGASSMSMIGLGMFFCWEDS AARFANTYPEI VTPISNPLTIETCCSAV 1495
SCV:AWK49433 LSMLAEWNLLESTGHHKMSIGE IARSIIITL TGADSSGFIGIGMFYCWEITANRLVQTYPEITVPSNPNISVETACRAI 1496
RVCV:QBS46644 LSMCAEWRWLESTGHHKMC TPGE LARNMITIIQSADISGLMGLGMFFCWEESSR LLLTYP E M VPNTNPMTFDSCCSAL 1497
WhIV-4:YP_009300689 LIYAMG SILVK -GDFILPNAEALKERCQTMI SD CPVGNFIGLGLYSWAETR QMMD --LDANWEPDDPPSLRGACEAS 1518
CCyV-1:ATS17313 MIMMAEWRC L TS -ADWKVP TASSIERAITAMIGDMNTDKLSMGAGFFT WPEAMKRYY -FANEI VEPDTPVNVASACKAI 1495
LNVV:YP_425092 ILILA EWRCMSL -SDWKTPTEAISRAAE A IIGDTPISRWYGV TGF FS WPS S MERY Y -VYPEIQEPDTPVTALSACRSI 1494
TYMaV:YP_009308376 LYMov:YP_002308376 LYLAEWRCMSK -SDWKTPTEAISRAAE A I I L V T P V T R W Y G M T G F F S W P S S M H Y Y -AYPEIQEPDTPV TFSACRSI 1494
TpVB:AYH53273 IMILA EWRCMTK -SDWKIPTQESVIRAEEAI I L D T P V S K W Y L S G F F S W P S S M E K Y Y -FCEPMEPDTIPVTFSACRSI 1494
BaCV:QAU20941 CWVNYHNENN ---KGRRTPSWERDKIYISRLSL LPPRSFSC LGGFFWDS LKALR -QERWFTPPVSYSPASPLGIAISC 1510
RSMV:YP_009553369 LFMES E T L K ---AGSDYPTFQLQRRIMRRVRS D T S N F V H L G G F F C W Q E S I A R I Q -K L K W S V M P A T F P I T A E S V L A A 1482
CBDAV:YP_009362280 IWIGLSSGEAVV -ESGAYPSWNYMKRSLIRKLWETPLASF TLL TGFYIWEEMINEMK -SIPWVVMPLSYLPTPSSLGVAA 1479
MaCyV:ARS22495 LWIGASSGPRIS -GGEI F P S W N Y M K T I S R R I W E A P A A F L T L A G F Y L W D D T I R V L S -C S K W A V M P L S Y P A T R H S V S Q A A 1476
MYSV:ATN96453 LWIGASASARVA -SGEIYPSWEFMKGSLIRRVWEAPASAFSILSGLYLWEEINEMV -LYPWAVMPLTYPATPN SV C I A A 1473
BYSMV:YP_009177231 IWIGASSTARVA -SGDIYPSWAYMKDLSIRRVWEAPSSAFTILSGLYLWEEINEMV -RYSWAVMPLTYPATPN SV C I A A 1473

: : : * * *

TrARV1 RITLLGMVDRKEYPK -RRTVHWISHDLCL -NRPLRHK -LIA YD -ALSTFFDCQDC LHTLAKTECHLFSV --EKTNLICDKG 1577
WhIV-5:YP_009300875 RTTSLGLIEVQWGR -DRFSNLVLYDFA -KALHIHK -LFSW -VISRYSKCND CIVA IAD TDSGAFGT --DELTIMCCSG 1576
PeVA:YP_006576506 RKTLLIGMVD E D A Y M -ARENYYIPEDTK -NNRFIK -LMISD -YITTKTRCEPCI LTVFNTESSTLLE --SLYSLVCSG 1577
WhIV-6:YP_009301361 QKNMIGLSLGRKWKD -SNRIAILADEEK -YNFVIVK -MFLYN -YLKSHSQCNLCRRLTSSLQIGDVRY ---LRTLTCDRG 1569
TpVA:AYH53279 RSNLIALVNNRRWNM -KPCRNIVTDEK -YSSLVYK -MLSD -WIELNDCQEKIA CADLMNGK N L G W --EIKHTVCRNG 1570
TYMaV:YP_009352236 R T M S I L V S K R I V T G -L K R S E I I L H D E Q -N E K L I Y K -F L I L D -D L E K A T R C A C I S M I E R S E L S D W V R --T L S F A I C K R G 1574
SAV1:QDJ94287 R C S L L S L A H R D L S I G -L S R S A I I L H D E K -N E T L I Y K -I L L C E -Q I K K R T K C V S C I A E T C R L E T G E L W R --D F R L L N C H Y G 1579
ADV:YP_009177021 K T S I I S L A L K K R W S L -P D R I S I I A D D E K -S S M W W M K -K L L Y Q -F R I K Q D K L D C R R L T S K M T S T D I K R ---M R T L Q C R Q S 1568
SCV:AWK49433 K E S L M S L A L K K T W A P -K D R F P V I S E D E K -S S M F V I R -K L L Y R -D M T K I N N C L G C Y K I I S R L S N V M D M K R ---L R S I R C S S N 1569
RVCV:QBS46644 K T S L V S L S L K R W T A -S C R F Y G L A D D E K -S S L M V V K -K M L Y Q -K Y G V E S K I S C K R L I S K L Q K H D I Q R ---L R F L I C K S G 1570
WhIV-4:YP_009300689 R R V L Y S Y I S N I P A A L R P R K T F N L V D E M K -D I G L C Y K -L I A C Q K L M S Y E G Y C S M I E L M S A P V R E Y T E --Q I L H A I C K R G 1594
CCyV-1:ATS17313 K I S L I N L L S S G K T F D V K R Q H Y I L L E E T K -T S K L V L K -M M I Y E -I L R A R T S R W C C L R V I G N M S P Y D L A N S -N A S L L T C H N S 1571
LNVV:YP_425092 R N S L L G L L G S S R K F R -G R N T R I F S E D A K -A S K L S L K -L M V Y D -W V K K K K C R A C H R E I G I M S A H Q L S S T -L P N S I I C P K G 1569
TYMaV:YP_002308376 R Q S L L G L I G S T R K F P -S R Q T R I I S E D V K -T S K M T L K -L M V F D -W I K K N T T C R A C W R S V G V L S T H N L A N L -D T R T F M A C A Q R 1569
TpVB:AYH53273 R Q S L M G L V S G S R Q F P -C R K T R L F S E D S K -Q S K L V L K -L M I Y D -W K A N T E C R S C W R A V G T L S T Y N L S S I -D L R E F V C S L S 1569
BaCV:QAU20941 K S S L I A M T N L L T S L E -V P I S T T I D V T K E -R D S V T K N Q L C F A V T G E I R S K C T S C A M E A F S S R I T R K T Q E S L V K L K C E W G 1588
RSMV:YP_009553369 K M S L I G A M T S G L T P K -R C D G V L A E N L I P -D V T R Q A K -N I I C L -D K V F Q S R C D Y C Y T A A M T N R W S D S I N S D S I F N M R C E R G 1558
CBDAV:YP_009362280 K N T L I R L A E R V Q R F P -A Q G S M L I T P L I S M N F G M V L K H N L L F N -R G E F R S S C P D C I V S G M T S K L N Q R A D W K S L K A I R C D A G 1557
MaCyV:ARS22495 K N S L I R L L H M T K E L R -R P S S W L V S P L I K I D P G L L V K G N L L F W -K G L P K T K C Q S C W T S G M G C K I R A N Y E W K D I Q K I K C E Y G 1554
MYSV:ATN96453 K N S L I G F M S R L D S I V -L S S S W L V S P L I K V D P G V L L K S N L L F Y -K G K I K S E C N S C V L V G M S C R V K S S D S W S E I S R I K C Q A G 1551
BYSMV:YP_009177231 K N S L I R F M G R V E N V S -I A S S W L V S P L I K V D P G I L L K S N L L F Y -R G K I R S E C N S C V T V G M S C K I R S S D S W S E I A I K I C K S G 1551

: : : * * *

TrARV1 HTLLDRI -KYCKM SHVT LERLRKDAD DGN -SKPDRETRE -----RLVPMIRTEVRTL L DLSKA --RRTLIA Y T N D -- 1645
WhIV-5:YP_009300875 HDAMELI -RSKCKISYVT LERLRKDAENAV -SIGCEDIASKINFSYP E M S R M S R T E I T L I D F R M --RHDLVAVRCD -- 1650
PeVA:YP_006576506 HMVFEQ --KMKIYQSHVT IERLRKDCASG -EKKTSFART -----DLIPLPQNTVADLINFDRM --RQRQINWGSAL 1645
WhIV-6:YP_009301361 HVVFD R M I K V P Y S A S N V T V E R L R K D V S N T M -RANRSRAR -----ITNPMKGQYSIEMV SSTQMGROYSSVRDQDN -- 1638
TpVA:AYH53279 H I F W D R M R Q Y P W I R S S V T V E R L R K D G C A S E -E R D P R L S L -----V V P K V S F G F Y V E L I R D S S L M R M F S V R G Y Q T --- 1638
TYMaV:YP_009352236 H K M S D W M K S P I K S Y V T V E R L R K D C D N I S -N E R A A D R I R L -----N S T K K F N F N I T L L S S D T L R I R P E S N D K A --- 1643
SAV1:QDJ94287 H R T A D W I P V A P W R K S Y I T V E R L R K D C C A S S -W E D R M P A L R L -----K G T P E N N F L V R L I G S H Q I R F R G E A V N Y Q --- 1648
ADV:YP_009177021 H K P F D R A A N T P W R S Y V T I E R L R K D C D S T -I Q S S -K A V K G -----H M K M K M A F C T A L I T K G D I I M R P E A V T P P A D -- 1638
SCV:AWK49433 H S P L N R L S S M P W R Y S H V T I E R L R K D C E S G D -I A T I N H L H E I -----N R L P E G L R F V R S I M N N T D L M M R P T E V E T L R D -- 1640
RVCV:QBS46644 H S T F G A L N D I P W A S S Y V T I E R L R K D C D S Y T -R E G R T D G I A N A V --Y Q R N C P I D --V I E L L R D S D V V S R G E Q E A Y V R I -- 1642
WhIV-4:YP_009300689 H H V L E R Y -R L K V N L S Y V S M D R L R K D C M A H E V -R S I Q E V T K M M N -----R Q I N L K Y D H L I T L F R T T E H --R P K F Y R V S D N A L 1665
CCyV-1:ATS17313 H V L F P K G V E G M I N R A Q I T M D A L K S I D S E -I Q D T Y H V R R E -----L I Q P L L E T S C R I G F S S L F --R A Q L I L N P --- 1638
LNVV:YP_425092 H Y V T G L K D L D I M R S R V T L D S L R K C C S D E -I P T E P M E K I T -----E W A P L T S T C R T L F D S S S M --R S E L I P Y S --- 1637
TYMaV:YP_002308376 H F P F Q R F N S K S V M K S R I T L D S L R K S I D I K D -D G E K E S Q L R S -----T I H P L T L T T C I T L F E S S I I --R A E I I P F C --- 1636
TpVB:AYH53273 H Y P L Q N Y S D T V I K S R V T L D S L R K S C D S S E -E Y D Y G R D K V I -----D I L P L E L T T C I T L F E S G I I --R A E L I P Y S --- 1636
BaCV:QAU20941 H S V Y D N T S E K K K F T L S L D Q V A D L T P P I V M D I R A E R H E P -----V L F P P P -----F L L F S S A S S P -Q Y ----- 1647
RSMV:YP_009553369 H T I L S P R M L H Q L R R T I L P E G A L Y T L A V R V V -H Q S P P E A S Q P -----L Q I T P C R R E R Y Q I L S E A D M P -R Q ----- 1620
CBDAV:YP_009362280 H H V F L L N S W R R L Q R V M D L E T L D G L V P S I P -R T S K R K T N L V -----I V P P V V -N Q V N E I F S T S T P ----- 1615
MaCyV:ARS22495 H S V F T L K G W R K L R K I N L D V E T L A D N I P S V P -I T E D H H I A I Y -----L D L G G G R E V G S S L F G S D D I ----- 1613
MYSV:ATN96453 H S V F T L K G W R M I R K V V L D I E T L A D H V P T A T -R R V K E R L P R G -----F P L S L K S G I T Y E L T S N M N F ----- 1610
BYSMV:YP_009177231 H S V F T M K G W K L L K R L V L D I E T L A D N V P T A P -R K T L P R A D R K -----F P L V A P Y G I T H E L A N Q S T F ----- 1610

* : :

TrARV1 TPYPKGLLSFTVG---NQRVSESSLNKVVSAPTSTRYKYIEL---FSRFREHLVG--RAMVLGDGLGQTSSILAS-MSGV 1716
 WhIV-5:YP_009300875 MTLPSRSLVFCNQ---NRLVTTSSLHKVLSYPTSTAYRYVEI---LSAYGTVIRG--RVMVVDGDLGVTSELVSI-YCHA 1721
 PeVA:YP_006576506 ALSPPGSKVMSDG---QSYTNQTSRLKIYSLPTSANYKYMEL---FSRYLPLIKG-KNTFLVGDGLGSTSLLLSD-M-TG 1716
 WhIV-6:YP_009301361 ---GILEVDYPQD---YHRYSIYQLIMLTQLPTNTRYKYSIDI---LARVGKRIVN-SNVFCLGDGLGTSLLWE-L-GA 1706
 TpVA:AYH53279 ----DTGDHDYTP---PSKLIKYSLFTLKSPLTVSWYKCIDV---YNIREDIKN-KAVFCVGDGLGSSSSIIAS-M-GA 1705
 TYMaV:YP_009352236 ----IPPEIVSDV---RSGFTVYHLATVMSMPTSTSYKMQDI---IGCGGIQILG-RKCLCIGDGLGTSSTVLSA-L-GA 1710
 SAV1:QDJ94287 ----EPRDQLEET---SSGITIFHLAALGTMPTSTIYKMQDV---IGMGISLRN-KTCLCLGDGLGSSSTVLRSL-M-GC 1715
 ADV:YP_009177021 ----DPLPSFPED---FHRFSEYHLMNIDTVPTRTKSKCTVL---LNPYLDIDIKG-RDVFVVDGGLGSSGVVAL-M-GA 1705
 SCV:AWK49433 ----DNLSKFPDD---HYFSDWHLHAIDTIPKTRSKYVTI---LDPHRKAIRG-MKTFCCIGDGLGNTSDILVS-M-GV 1707
 RVCV:QBS46644 ----DYLNPFPVK---HHEFTLYHLYLTDLTPTRTKSKYMSI---LGPLRKHIIIG-KSVFLLGDGLGTSALLSN-I-GA 1709
 WhIV-4:YP_009300689 RIASDARLRFDSE---SMLYSETDIMKVIKTPSALYKYSEI---FNHLRLKEFK-SDILLGDGSGWTSLLRENIHNN 1738
 CCyV-1:ATS17313 ----NVDDHRFVS---VRPSGAGDLCKLFLSPTGAEYKYTDI---VSFLIHDIKRLKALVLGNLGGSSNVLRRL-M-WR 1706
 LNVV:YP_425092 ----PTGISSIN---VQPIPKSDLYKLSLPTNAMYKYMEV---ISRNIEGIMNCKTAFVTGNLGGTSKVLN-M-WP 1705
 LYMoV:YP_002308376 ----DPIDDDKIT---VRPIEGVDLHKLVSPLTNASYKYMEI---FSREIQELNKYKSVFITGNLGGTSQVLE-L-WG 1704
 TpVB:AYH53273 ----PLAEDDDVD---VVPVIGVDLFKIVSLPTNAAKYLEI---FSREYDNIMEYGSFAVVTGNLGGTSQVLE-L-CG 1704
 BaCV:QAU20941 QDEPTIYENVISFD---RELLLEAKLISFVQIPTRSYGRVYEI---LSGLKITQRR-DRVLVMGDGFGWSSVVTM-LNPL 1718
 RSMV:YP_009553369 TGYSGVYRRPTNQ---ICMYTEVELAKKYRLPTNSLYRILDLHDCFFSQKLMTR-GNILVVDGSGYSSLLTKC-LNPD 1695
 CBDaV:YP_009362280 KSYADTPSRLPGDEVINAPNMVNLKFSIPTKSLYRVHEA---LSHLDDCMEY-GSILCLGDGFGYSSMAAKV-ISPG 1690
 MaCyV:ARS22495 DGISSEHLRPGSEYLERPAPLLQLLEFNLPTRSLYRIYEI---MTHFHNHREI-GRILLVLDGSGYSSMMAKM-FCPK 1688
 MYSV:ATN96453 PGSSNVSYQMPGEEFLLRPKYDSDNYEYSLPTRALYRVYEG---ISEVTDLFNY-GSVLVMGDGFGYSSLVCKM-VNPN 1685
 BYSMV:YP_009177231 SGASEVSYQMPGEEFLLRPKDDYSDKYEYALPTKAIYRVFEG---VSEVPEFKNF-GSILVLDGDFGYSSLVCKM-INPD 1685

* * : : : * : * * * .

TrARV1 SKIFVSTLAETGNAAPQTYPHAVQPY--NTSGYSLIDNKTIGKHNHILSQDFLEEWSDVFR---QTDILISDIEIIGED 1791
 WhIV-5:YP_009300875 SRVTISTLADTGLAAPQTYPHSVQPI--VRSSPYKIDGSTMIDKHNHILSESYTEEWRPVMA---SLNLSLSDIEIIGPE 1796
 PeVA:YP_006576506 NNILVSTLLETDDAMPQSYPHLMQPIA-VRESKTNVDRKTMINRVNDLSSESWAKDWKDAAS---NCEIIVVSDIEILGKR 1792
 WhIV-6:YP_009301361 SKIICSTLLTPDDAIPQTYVHNVPPM--LSDLDGDDVDFKHMNLNPNILDSRWHDWSKSVVS---NCDMVLCDVEILGRN 1781
 TpVA:AYH53279 SEVIVSTILSPDDAIPQTYVHNVSPVS-MEYKSDKINDKLMINRNNITDDTWSRDWQVDM---NCSVLVSDAEVIVRE 1781
 TYMaV:YP_009352236 ESVTSSTMLPEDEAIPHAYSHNVLPVP-QFYGIGNIDATKAANRHNDRVNRQAQWSSDWAEEELR---SCDVLVSDAEVNVND 1786
 SAV1:QDJ94287 ASVVSSTMIPEDEAIPHAYSHNVLPVP-QVYGMTGIECAKSIQRHNDIRNANWERDWHHEELS---SADVVSDAEIVNPD 1791
 ADV:YP_009177021 KKVITSTILDPERAIPQTYVHNSIPISLKFDLNDVLDTKTMINKMNNIMDKGWTDSWKGTTL---DCRALVSDIEIFRRE 1782
 SCV:AWK49433 SAVISSSTILDPSKAIQTYVHNVYVGVLGSGNKGLIDTVQSIDKTNNILDDNWMISWKTTE---TCDLLVSDIEIIGSD 1784
 RVCV:QBS46644 AHVVTSTLLDPGRAIPQTYVHNVPTNIVASCNPLIDSSTMDIKINDVLDTRWSESWEGSVQ---DCDICISDIEIFKRE 1786
 WhIV-4:YP_009300689 STIYVSTLISSESVMPTMPLFDHTM---NLSNIDKTSVMNKVNNILDDRWDGEDWEPAS---VCGILISDIEIIGEN 1811
 CCyV-1:ATS17313 GKLIISTLVDGTGESIPQAYPMCNNAF--KFLSDPDVDSMNRVNDISHEGWKSWANIIP-P-DVDFCVSDIEIINPT 1782
 LNVV:YP_425092 GRIITSTLLDGTDAIPQVYVPCDKGS--SSYARGTVISDLMVTRVNDVNHLLWGEDWKPVFQ--SYETDLCSIDIEINGEL 1782
 LYMoV:YP_002308376 GRIIISTLLDGTSAIPQVYVPHCDAL--KNTGAASIISHLMVDRANDVLDHRWETDWNVPFT--SYNIQVLISDIEITGE- 1780
 TpVB:AYH53273 GFIVISTLLDGTSTIPQIYPHCDSSM--KNPNPGRIESSWMVDRVNDVMHDNWEWDKPIFD-DNDIRFLVSDIEIIGK- 1780
 BaCV:QAU20941 SEVWSWDLIDISQSPPHILHQSSPPT--HYKFEIMEIKNELSYTISDIFHDGFEKSVTDLDDIAGFHLVISETELCNPN 1796
 RSMV:YP_009553369 RNVVSWTYIEPSEALPHSLRISKPPM--HYKADVQIDSSPSIDRISDIHNSSPYDEFKAVVT--KNGITALISDIEIVVVS 1772
 CBDaV:YP_009362280 ASVYGTMTIDTSTSVQHCLRLSRPPT--HYRLDVGVDSSLSIDRVSDVYSPRFPKFEQEVIR-EKRVDLVISEIEFRYSG 1767
 MaCyV:ARS22495 AKVFSWTLIDTSASVQHCLRQFKPPT--HYSLGFIDICSASIERVSDITNIEFPGELAEFVD-ENKIDLISEVELMYSN 1765
 MYSV:ATN96453 AKVLSWTLIDTSSGIQHCLRLSRPPT--HYMSNLEIDNSPTIELPSDVSDARFPDSMKYIVR-EKKIDIVISEVELLYSG 1762
 BYSMV:YP_009177231 AKIISWTLIDPSSGIQHCLRLSRPPT--HYMSNMQIDNSPTIEYPSDVSDPKFKEALKEVVE-KRKVDLVISEVELLYSG 1762

: : : : : : : : *

TrARV1 RA--IDRDVAIEKMFKERSWRFFLVKDLYLTLLEEFTRRSSFIIPKCGKLYSVVTTNMRSLLAPEIWWGGPTASEY---- 1865
 WhIV-5:YP_009300875 RS--DDRTSAIRKLLLEICDWEFIIKDYIYSLREFTDRCSIIILPKMNGDIKLVTSKMRSVRAPEVWVWIGQNSKAEY---- 1870
 PeVA:YP_006576506 NC--FQRDNVNLNKILTKDWDFAIKDYIYFIEIYDLRLYIKLVSKVGKMTLVSLGSKTPGVEVWIRKTDHSR---- 1866
 WhIV-6:YP_009301361 KA--KERLILMEKILRINNWRATAVKIDYLFSPDDMCNQIRRIASSQPKGWSLVTSKFRSAHHPVWVWILDRSREVTG--- 1856
 TpVA:AYH53279 RG--EIRQEIFENLLSLKDWCAVVKDYLFSPADMVNRISSMHNS-CVSWKLVTRFRSSFYPEVWVWLVGCVGSKV--- 1855
 TYMaV:YP_009352236 DH--ESRFELVKKIALSGRRPVTVIKDYIWSAEELSNIKIGIMWASRARRWELITTRFRSHNYPEVWVWLVHDAVSPTS--- 1861
 SAV1:QDJ94287 EH--RTRYEVLKSKILLSGKREITIVKDYIYVSEELANKIGVHSCGSSQVSLVTRFRSHHYPEVWVWVIRDSNPKRG--- 1866
 ADV:YP_009177021 DR--AMRSRALRQLLSLKEWFAVKDYIYVSEELANKIGVHSCGSSQVSLVTRFRSHHYPEVWVWVIRDSNPKRG--- 1857
 SCV:AWK49433 TE--ILRENAVMMMLKLSKWKFCIKDYIYSCHYLSNMIKLIGSSRPKSWSLVTTTELRSVHYPEVWVWVIRDSNPKRG--- 1860
 RVCV:QBS46644 DN--ESRRTLLTSVLCCLKSWKMCILKDYIYASDLSNSMKIISSSLPRRWLVTTPLRSASYPECVWVWVIRDSNPKRG--- 1861
 WhIV-4:YP_009300689 RY--QDRELMFRKLLSVTSWKIACIKDYIYVSEELANKIGVHSCGSSQVSLVTRFRSHHYPEVWVWVIRDSNPKRG--- 1887
 CCyV-1:ATS17313 QN--YDRNQVMRKVLALKSWKMLLTKDYVYSAELESRLSIALQH-SDDVKMLFSGARQVVPFVWVWVIRDSNPKRG--- 1856
 LNVV:YP_425092 NG--ESRQMTIATVMAHDWKMVMZMKDYIYVSEELANKIGVHSCGSSQVSLVTRFRSHHYPEVWVWVIRDSNPKRG--- 1854
 LYMoV:YP_002308376 EI--ETRSVLSKMIHAHEWKFALIKDYIYRSLEENRSLIILGL-YAKVEIITCNTRQRTMPVWVWVIRDSNPKRG--- 1853
 TpVB:AYH53273 GS--NLRDDVIRKLIITAHPEYAVIKDYIYRSLEELRSLYLLGF-YDHVELITCNTRQRTMPVWVWVIRDSNPKRG--- 1853
 BaCV:QAU20941 VEGDQSYVDVIKRLMLIRS-NSFLVKVVDQSPRLMEIVNFVANH-FSTWAVFNSLSPAYSIGFVWVWVIRDSNPKRG--- 1874
 RSMV:YP_009553369 GE--KSVASLINALWNNQI-QLGALKLELMT-DPLEKVEYAHNA-YQRWELFTLPGANLGGVLYIGFYGRREKLTGYI 1847
 CBDaV:YP_009362280 KR--DDPDLMIGLFYQAGV-LRFLIKVEVDGLEIINKYVNCAYKM-YSFVEVFTSLCGLHTGDVWVWVIRDSNPKRG--- 1839
 MaCyV:ARS22495 ST--MEPGKILIRLYWESKA-SAVILKFEVDCFSNLKIAIEHVWRY-FRQWKVIQTPNAGLHTGDVFIIMYDRITSS--- 1837
 MYSV:ATN96453 VS--QSSEELIELYWGTE-SKILHXYQKFSSEIKNLIEVVRGY-YNSWKIFISGSVNFKNGSFVWVWVIRDSNPKRG--- 1834
 BYSMV:YP_009177231 VE--MTELDMLVQYWESGV-QHVHLKFFQFTQVDRDIESVRS-LYGSWKIFISGTVNFKSGEFVWVWVIRDSNPKRG--- 1834

* : : : : : : : *

TrARV1 -----SVRRCYDPRSMRVEWNLVKTLMF--KHEYGDKTIIQDLQNRAMSDSTLAYMYSVILDWSSIIPGLGKTLPTNGNFTD 1940
 WhIV-5:YP_009300875 -----TVRRCYDPRSIRELWDDCVRSLLH-DTEYQDPTVLEDLTRRVLSRANLASVHSVLEWSSLAVVGKMLPSDGSYTE 1945
 PeVA:YP_006576506 -----TVQLSYDPRSIKEEWNDAWYMLLEDLDRDGVPKYIMDDMNERLFINNGLDTMHNVRYWSTLPMVGNALPKKGNFTD 1942
 WhIV-6:YP_009301361 -----SISVGFVGNVLSTWWTVRRELMESYNDLILSTDMKTLAFINSNGHARRMETYVRSWLIFFPVIGSMPLPQKGSFTQ 1932
 TpVA:AYH53279 -QKKGVGISVDWRNCLGVWVNIKRELEERYDHEVIDHSVAEKIRMMSPSGGHLKMMTYLRTWANFPITGSLLEPKGAFTK 1934
 TYMaV:YP_009352236 -----DTILYPIPNKVQCLWIGIERALK--SHEYSISGEDNALISSLHROMLYKMSRVRWAADFVLDIGSLLPKNGSYTS 1935
 SAV1:QD94287 -----RVPVSPIASNILNIWENIKRDLSDSSSYGVFTSEENHRISMTPSIVTEKMISYVRAWATFQMVGNLLPSGGTFT 1942
 ADV:YP_009177021 -----KLGLSIETGKLNRVWNNYLEVMTQVELEDTEREEIELRSLHGEVLESSLSHVRNWLSLPIVGLLVPNGGNFTR 1933
 SCV:AWK49433 -----PLQLGMRPDRLSRIWNRFMEIMIGGEGELHCINREEIMILRSMTPHDIQRRVLSGVRAWLSLPLIGLVFPDKGRMTR 1936
 RVCV:QBS46644 -----TMVLGVDPKSLQRHWNKILETLTAYPEPEDVSDAVKYYLLSFMDESADAKMLSYLSRWMTPISLGLVYPDKGTFT 1937
 WhIV-4:YP_009300689 -DDNETCLNFVPEIMTSFWSNTMSGLC--KRNAIEWPLTEIDHLLNKSMYDSMIQRIARYCTVPTVGLCPHQNNFTR 1964
 CCyV-1:ATS17313 -----KAETGYHKRVQMEIWSDFEHLN--YSDDLPEILTNINHLIMDEGRLAAITGRIKLWATLPIAGSALPHKGSYTR 1930
 LNYV:YP_425092 -----GKLLGYHRSVIRQIWDGVEKGIN--TADWAMESVFSEINRTIASTADMIAITIRLKAFFSLPIVGSVLPYKGSYTR 1928
 LYMoV:YP_002308376 -----FKRLGYHRSVIRQQWECFKNIN--QRDWALGEVLTENLHRLASDDKLISSMMVRTKSAFSLPIVGCVPFHKGNYTR 1927
 TpVB:AYH53273 -----PRLGYHRAVIRQQWECLKVNI--QKDWALESVLSDLNERLATKDKLISMMIKAKTAFSLPIVGCAPFHKGNYTR 1927
 BaCV:QAU20941 LSDECAELMIDSVKEQLLSPKIELMGML----RPFKSHHELLALSSIGTSSYDIASMTNEW--LSQVGMVSVRSNFT 1948
 RSMV:YP_009553369 IPHSGVETLMDRLATEVEDEDRGRRAED-RDRWEQLDTME-----SKIHLLQYHRMRLLDLW--FGSAYLSGYLSEDMTE 1918
 CBDaV:YP_009362280 -----QMNRCLDQNSTIQLYQSLRHSSSV-DDQWIDISESVKIINRVF SKTSLTVDIQIMLNIW--FQDGHIIHWKEEDFSR 1912
 MaCyV:ARS22495 -----NNWEYLSYNNMTLLYERLRIAHET-MNDWYPGDYCNTINQTLQGTLETFRQQLLDVW--FQDSTIQYWREKDCQ 1910
 MYSV:ATN96453 -----LANTKLTYNVSVISLYDLTRFNHER-MLMVSPEGYCHDINKYLDNTFLKHYREQLLDIWI--FQDASIIHWSRDFTQ 1907
 BYSMV:YP_009177231 -----ISKSKLTYNSTILLYDLTRLNHNEN-MLLVSPGGYCSDLNKHLDATFLKGYREQMLDIWI--FQDASIIHWRAKDFQ 1907

TrARV1 LFYIIQYKRPKEKVPENVDSSKLLHMSDYLLKREVLFGLVSMVADIRVVR--LMN-ESEKWLVDWEP-----NNTRT 2012
 WhIV-5:YP_009300875 AFYKLYGKRPKIVSDDRDNRADIRLHLSDYKREILFGLAVSMIADIKMRQR-LIN-ESELWLDWEP-----NQKYT 2017
 PeVA:YP_006576506 LFYRLSSGKRPAIIHVERENKQLRMYVKDHYKIREVLFGMAVAMLANIKDREK-ILN-ESEQWILEWEL-----GKESS 2014
 WhIV-6:YP_009301361 LFYYLKKTKRPGKIHTVRGNSNRLYSSDFSLRIKMFALAVSMLADINDRLT-MIK-ESSEWLDVWY-----EDKYD 2003
 TpVA:AYH53279 MFLYMKKSKRPEHVKLRQDSSKLLYDSDYKLEKIFAMAVSMLADVEDRHK-MLN-ESENHLEWY-----EKRRK 2005
 TYMaV:YP_009352236 LYYYILKTKRPFKIKAS-ESSSHKLYSDYLELRGKLFATIAVSMMAVDNRLS-MIN-RSHHWSLIWK-----KYSD 2004
 SAV1:QD94287 VYYYILKTKRPPTVKQS-DGPDKLLYSDYLELRSKLFAIAASMIAPIDLRLT-MIK-ESHKWLWY-----NYGS 2011
 ADV:YP_009177021 IYLYMKKYPQAWVKTRVDSALKLYDKDYRRLDILLCLALGLCESDQVVT-ELK-RTENWYLNWV-----EKETG 2004
 SCV:AWK49433 VYYYLRKSKMPAYAKFQRDNSKLLYDKDYRRLDILLCLALSMCEDISMLSK-EMS-MTEYWFLDWE-----EKETG 2007
 RVCV:QBS46644 VYYYLRKKNKQPAYAKLQRSEKRLKLYDKDYRRLDILLALALSMCSVDQMVMV-EMA-RSDMYLEWE-----EKEIG 2008
 WhIV-4:YP_009300689 VLGLLQSGFRPAKVSFRWERSVKLYKSNEIKLLEILIIAASMIANVKTRVD-FIN-NARFMMWVWSG-----KPNTN 2036
 CCyV-1:ATS17313 FFGYLQRGKPKADIRWEKDDLGRKLYMSDYDQLREILLGLAASMIAPLDRNQ-FVD-STQYVALIWK-----PSRTG 2001
 LNYV:YP_425092 LLGYLQRGKPKEDISLITSDDGKRLYLYSDYEKVRVSVLFGLVAGMCSATERDR-MLD-ESEYWAIDWI-----PSGPH 1999
 LYMoV:YP_002308376 LLGYLQRGKPKIDISIISTQSNRKYSSDFEIKIRVWVFGIACSMCAKITDRER-MLD-QSDRWMLDWK-----PAGER 1998
 TpVB:AYH53273 LLGYLQRGKRPIDISIVAATSGRKLYTSDYEKVRVWVFLGACSMCGRVADREK-LMN-ESENHMLDWK-----PSGSK 1998
 BaCV:QAU20941 MLNELRHFKLYELLDLGRDRKYFFKEDHERLRLRLICLAGSMLTSQYLQAE-FL---TTRWKLWSTKGNKGLDDY 2024
 RSMV:YP_009553369 FFYSVKTSYRPPAISWEKGNQARYLYGSREEILFMSMITVALSSYLEDDQVLEFL--SSSGWKLQWKD----KKNPR 1991
 CBDaV:YP_009362280 LYGYIKTGRPKVEVLDTTGHGVVYLHFDQSQKMFIRLMTLALSLMENRYTISS-ILNSQDNQVWQKWKQGIHGGSRHNY 1991
 MaCyV:ARS22495 LYYSRLTGRPPQSVDAAAGNTVYVYHGDLEEKIFERLFTLAMASIKDIEEKIARII---AQNWRLKWERTSSIVGTREY 1987
 MYSV:ATN96453 MYALRTGKRPAASVLDTQGNVYVYHSDMAEKIFERLTLALSLKPENRNIG-LL---DLPWRLKWEKSQETVSVRSTY 1983
 BYSMV:YP_009177231 MYALRTGKRPAHVLDTQGNVYVYHADMSKIFERLTLALSLKSESIFIG-IL---DLPWRLKWEKSQETVSVRSKY 1983

TrARV1 RWVPYLYLSDGRKE----LPKIDVSDYVYPYLSM--MSLQKLLFRGVKNSVFRSLGTS-----REFVGFVPS 2074
 WhIV-5:YP_009300875 VWVPYLYKSDGRFD----LPQINVDYVYPYLSM--MAKRRLLFREVKKSIRFRSVGGV-----RSHLVFPVS 2079
 PeVA:YP_006576506 DWYLYLYRSDSLS-----SKINVDYVYPYLSM--MIRDRLLFVGKFGKSIKFNENMK-----KKDVFPPIS 2075
 WhIV-6:YP_009301361 DWRPKLTRSLDNQL----PRCEVINYPMLSSI--MLREGLIFRSHTDDVMFRP-DRK-----REELCFPIS 2063
 TpVA:AYH53279 SWSCMLERAPGSPD----IKCNVSDYIPVLCVL--MRQQLIFRKPSPKILFAPQNSN-----RDSIFFPIS 2066
 TYMaV:YP_009352236 KWDVSLLEEKEINE----PPCDVIKYVPALNLM--MKEHGLLFKNCNSKVEFRS-TRS-----REELCFPIT 2064
 SAV1:QD94287 QWEVKLVRESETE-----PECDVIKYIPTLSIL--MKQQLLFEKIGSAVSFGI-NRK-----RDEVYFPVT 2071
 ADV:YP_009177021 VWDCELTRSTTEG-----MRADIEDYLPYRIM--MIKDRLLFCCKYINCDTIKFRPSSRG-----DQDIHFSIS 2065
 SCV:AWK49433 IWDCVLRKSLTEPEG----MRASIDDYLPYIRSI--MFQRMTFREIGHEVRFKHMSK-----EHEERKIVFPIS 2072
 RVCV:QBS46644 HWDVCLIRSLPEPEG----PTAKIEDYLPVLRVL--MEKDGLSFSNIPDTRFVWLSRA-----E-RDKDYVFPIS 2072
 WhIV-4:YP_009300689 TWPMPRLMNTERNP----GFEVNI-DYIGVNLRF--FNKQQLCFREVSFPKIFAY-SKK-----RGTLCPYVA 2096
 CCyV-1:ATS17313 IWMVPMKKMERSL-----APAHIYDMVPLSIF--MRRDRLLFKYEGNTIEFKP-DRN-----RKKLCFPIT 2060
 LNYV:YP_425092 IWLPYLFKGVERS-----TLIHVYDIPIMLTLI--MKRERLLFKSSSDIEFKY-TNN-----RDSCCFPIT 2058
 LYMoV:YP_002308376 NWAPYLWYSRHKS-----TPIHVCDYIPIILSV--MKKERLLFHGGSGQTIKFK-SLS-----REKCCFPVT 2057
 TpVB:AYH53273 NWVPYLVWKSDEKS-----SPIHVYDIPIILSLW--MKKEKLLFECVDWEIEFYH-TKS-----RKECCFPIT 2057
 BaCV:QAU20941 YYNPVL TANGTEDD--NDIIRV-LSRHVKSIVLRCRITTPRAIVFSKFGKVTQYLPKHESENDDGTLSSASLLCFPIS 2101
 RSMV:YP_009553369 HWSPLYLERSDSQLV--WKKNKAMIFKLVPLRGS--KPREYKMGKASDSEIVFRYIPRK-----VREGEVPLCFPIS 2059
 CBDaV:YP_009362280 QWSPRELIRVYKSRFWSSAAVNEIKRYLPPVKQI--RPIATRQIIEVPRIRFMYVRR-----TQAMKLCFPIS 2059
 MaCyV:ARS22495 KWAPSLVSGGRQAI--TDRMRNNIYRYPGVIAV--IKRQRPKQYDDVSQYIRFEPDAKK-----RGEKKQKARFPIT 2056
 MYSV:ATN96453 RWAPALIQDGNSTI--PDRVKSNIARYLPGVAIV--RRSLGSLYEVPDRVRFQNPGR-----SKVLKFEIT 2047
 BYSMV:YP_009177231 KWAPALVQEGNVLII--SDRVRNSIIRYLPVSAV--RREMGSYETVPRKIRFCPPGKN-----TKMLRFDIIT 2047

: * : : : : *

TrARV1	GVAYGSFTAMKHGKRGIEGKMKKIQPKCK	2104
WhIV-5:YP_009300875	GSAYATFTGFNSARRKHN-----D	2098
PeVA:YP_006576506	ATANLKVKPYKKNR-----	2089
WhIV-6:YP_009301361	VVADKNKVAYYPVVRQ-----	2079
TpVA:AYH53279	KNASS-----	2072
TYMaV:YP_009352236	SLSERLRREAPRKRRTDKIT-----T	2085
SAV1:QDJ94287	QLSLTLRRSQAGNGKPNRGG-----S	2092
ADV:YP_009177021	KTSYMKPSKLIKDKRMFGT-----D	2085
SCV:AWK49433	RIAHMTSKHLYKDKRRS-----Q	2090
RVCV:QBS46644	RTASMRSHVLIKDKRS-----	2088
WhIV-4:YP_009300689	KNMVIRLPK-----	2105
CCyV-1:ATS17313	KAAFIRIEREL-----	2071
LNyV:YP_425092	KTAKIKFNIK-----	2068
LYMoV:YP_002308376	KTAAIKFGCLK-----	2068
TpVB:AYH53273	KTAEIKFKVK-----	2067
BaCV:QAU20941	KNSSYLAFIA-----D	2113
RSMV:YP_009553369	KVASMSV-----	2066
CBDaV:YP_009362280	KLSSYTVA-----R	2068
MaCyV:ARS22495	KQAFSTPM-----G	2066
MYSV:ATN96453	RGASFTTP-----Y	2056
BYSMV:YP_009177231	KGASFTTP-----Y	2056

Fig. S2. Multiple sequence alignment of the TrARV1 N (nucleocapsid) and related proteins

TrARV1	MA-----STTPEKQGLKERLAAAKGLMH-----QESHKNTMEHN	34
WhIV-5:YP_009300870	MA-----SSSVSDRL-EALARAK-----MAGKSVTPDT	28
PeVA:YP_006576501	-----MSTVAERL-AALRAAKNTP-----PPPTKNIDEKTEKDDK	36
WhIV-6:YP_009301356	-----MSTLRSLRNSDIPSTSKNPV-----QTPRDGSKDKGKGIS	35
TpVA:AYH53274	MRYLDLFLISTVLSFTISPSLVLYLSFVCGFESIANMTTKLLNAL-----	44
TYMaV:YP_009352242	-----MNTLALQL-----	8
SAV1:QDJ94294	-----MAALHTAM-----	8
ADV:YP_009177015	MD-----AETKAAAI-RALNAAKADTR-----RRKLSNPQPTGEPVK	36
SCV:AWK49426	MT-----TKEARRKLRAALSSSKGSGS-----	22
RVCV:QBS46637	-----MNVNRAELRKALLSLRGEPSQDRGKSIALTNNERRMSDGGPGRPFK	48
WhIV-4:YP_009300684	MDT-----DRIKQLEEMKRKRALLKQANEAKSQQA-----GTSIGVKTDTMTAQ	45
CCyV-1:ATS17308	MT-----SQKTEQQLQDEIQRIRMERA-----RKGNIEAGPSNVPPP	38
LNyV:YP_425087	-----MTTSAEKL-AKLEQLRKERA-----AVIQKPTTVQSS	31
LYMoV:YP_002308371	M-----AETMAEKL-ARLQALRGGKS-----EVKSTPQEQEP	32
TpVB:AYH53268	M-----SETTEQKM-ERLRKIKEGMA-----NASTKPAQETAK	32
BaCV:QAU20935	M-----	1
RSMV:YP_009553363	M-----	1
CBDaV:YP_009362275	M-----	1
MaCyV:ARS22490	MRP-----RSLGFLHLRI-----LRVKMA-----	19
MYSV:ATN96444	M-----ATQ-----	4
BYSMV:YP_009177222	M-----	1
TrARV1	QQV-----ATGKKQGGNAKLFKTVHNMTPGKTIPKVS-DNELSKIPIYSLTQLSVSEAVLLGNHMMTMLVG	101
WhIV-5:YP_009300870	SQV-----RPKSSKKPGENGALFSKVKNMVTPGKSMPEWK-DSELEKVSTYDIDVLDAAEAVLMGQAMIRMLAG	97
PeVA:YP_006576501	DQVKK-----DNNVSERRAMNTNRYSAVSGLTGGGKMITKTWNDDTELPNIPIYSLSEITVDQLCVGAATIKRIKN	108
WhIV-6:YP_009301356	NQVLL-----NDGKNSSNTKVNRYAELGSTVPSKHKIKIWD-DSHIKDLKVYGLIQLTEDKIITYGKALRDYLN	106
TpVA:AYH53274	-----AAQTKYADVDDIGVNSGRPVAVWE-IGHEKKIKIYDVGILNGQDANKYGEYMMGCMRF	101
TYMaV:YP_009352242	-----SNTDYDDLASITVAPGGSNVAWN-DEDVLSIRRYSLAVMDSPTMVLHGQYVVFESLNN	64
SAV1:QDJ94294	-----INKDYDDLVDVSVLPGGSIAVWR-DADVATIRRYTTTTMDVALAVQHGDYVIGCLTD	64
ADV:YP_009177015	VEIPKASGSNTVKTIPIASGPTANSKKYTDIDSVRTGDSKTTLTWR-NDSFSKIKVYEVTLNDDDCVLVYGRAVFAINT	115
SCV:AWK49426	-EIPVQAVVKSS-ETPIPTDRNQMRKYAELDNINVLSRSTSEWK-DSDFNLSVYQIEHLNTDDAVLFGAAMMTAIAE	99
RVCV:QBS46637	-QMSSS-----ADKQERKSPPKQYQDIDNISLSVGREASKWE-DSHFALQIYDITALNSDDGVLVYGRAVKAINE	119
WhIV-4:YP_009300684	EKIVVDKTTKSK-GKETVKPTTPKVDKYHDLAETRVSLKSTPKDWK-DEDIKNVNIYGVKQLSISEIVRLGNSVVFESISK	123
CCyV-1:ATS17308	V-----KPRTVLPKQTSNQRYLEIDSVSV-GKLTSPWS-DTELSKIPYRVNAINAACKCLTLGRTVFENLNA	104
LNyV:YP_425087	E-----PVVTEKPRVRNTAYDSLAVGTI-GKRASRKWS-DADIKSIPIYDVHVQVPAQVIALGKDLLTQIQN	96
LYMoV:YP_002308371	-----KIARVANKLYDAVDDVTV-GKRASKRWD-DNDLANVVCYDVAQLTAGNMISIGKTLNLSNIAS	92
TpVB:AYH53268	-----TVNRVNTLYDAVDDVTV-GKRAAIKWS-DAELSKVVCYDVQVQTAGSMISMGKLLTHLTS	92
BaCV:QAU20935	-----PKAYKDNSAYAKEIFTPITKEISHGRIQKQAWT-DQVFDQMVYKINKYTAEDLERYGMVMVDSISK	67
RSMV:YP_009553363	-----ATDKSFEKLSLVPENTKLYSISPEAYS-DDKFDKANCYKLEKRSEYELTRLYKGLVRDLGN	62
CBDaV:YP_009362275	-----AQHTIDAKITALFDGIPEDVCLGPVTGVKFT-NDAFDALTCYRLEAVADAKVRYMAEKAMSEIFQ	65
MaCyV:ARS22490	-----SGSEQHIDKDLTDAFVSNVPELVAVNTVPDETFSS-DDAFKSHSVYEINELTIKQVHHAYKLFNRNMGQ	85
MYSV:ATN96444	-----TQDDHGLDKSISNMFEVDPNDISTSLIPGVVFS-DEEFDKVPAYTTPRPNDSQLVRASSKLLDLSLGG	70
BYSMV:YP_009177222	-----AKEDHGLDRENYDLYEDVPEIDISTSLIPGIEFS-DDEFDRIPVYTPAELNDNLVAAATKFLLES LGK	67

TrARV1	N--KSNKDLIDILFLAVSLRDPND---TSKLLTPPADGFLGK--TTIGKPSVS--AVDKKTTDDD-----ADQQ	163
WhIV-5:YP_009300870	K--TASRDLIDIVLFLAVSMRDPD---LDKHLTKPAKGYGTQ--KAISKPTVQKSPQTATSSGA-----STSA	161
PeVA:YP_006576501	G--DTSNLTVDGLIYLATSRLRDPN---PEVALLTAPTAKFGAP--AALTIDL TN-----TDSQ	160
WhIV-6:YP_009301356	G--DVNANTVNIMLYLAVSLKSMEN---IQKHLISPNAILTKP--FEMSDIVLE--DQR-----VEET	161
TpVA:AYH53274	G--TDINSDLIKVMLSLAVSMRNPED---IKTHLLSKPARD-----TKYTMPHVL--KAITVVNDGG-----GEPN	160
TYMaV:YP_009352242	N--ARVDTNVLMSVHMLAANLRDPDN---ITQCLLTPPHSRPAI--IQTLPTIR--AIQ-----EDIT	120
SAV1:QDJ94294	G--AGIDSSLVVSMLNLAJNLRDPDN---PERKLLRDVATDRSL--IRMELPTIT--VVN-----NAMD	120
ADV:YP_009177015	G--NITSDVVFMMLYLAISIRSTGDK--SENYLLAAPTG-MTDL--MGLVRPTVS--DQAMATAGGSV-----ADVD	178
SCV:AWK49426	D--IINADTLNMMLRLAVSLRSTAA---TEDYMLVSPNNVLTNS--ITSSKPTVN--NNRNASEDAD-----PELQ	161
RVCV:QBS46637	G--TITSDILFMMLYLAVSCRSTMD---VSKYLLSKPTG-MTAM--MDHTPPTAA--ASDEHIEDAT-----TDRN	180
WhIV-4:YP_009300684	S--QFTSSHVDILLSLAVSLLEPGSTAETEKYVLIPLPSTVGNP--IKNVISITAS--EAK-----SKEV	181
CCyV-1:ATS17308	G--TVTAALADMCLALAVSLPKPALA--TFEHLTPIPATIGTG--VAFNQPEVN--DAA-----PSLT	160
LNYV:YP_425087	N--SVNTVTVDYCLVLAIVSIPKAMT--SFEHLTPPSPEQGTG--LDFTQPQAG--VSNR-----SGLT	153
LYMoV:YP_002308371	G--AITSKTIDACLALAVSIPKPTS--KFSAMLSPPPDGVGKK--ITFEQPSAS--SVAR-----VGLT	149
TpVB:AYH53268	G--NVNSSTVDICLTLAISIPTPATR--AFSYMLTPLPENIGRK--IDFELPTAS--ASVV-----KGLS	149
BaCV:QAU20935	K--EVSMDTCYAILILACHLHKIAP---NKDFTLFELPLKGLTP-----LPKME-----	111
RSMV:YP_009553363	S--SPSTYAVERRLLVLAHLIYETKKG--SSNFFLDYLPKTTSTANLDAGFLAKL--KETPKASDPDVSVDTEVKAKAT	136
CBDaV:YP_009362275	SAATCKDYHPYMLTLASQLYTPGKS--ATTKMVVDYFPTSQTG-----IPTAT--SDEWKLK-----TKNT	123
MaCyV:ARS22490	K--TCHAIIPYSILAMALRLPTPKK--STTSFLTSHKPSGGTA--LEQGIASKL--YPSLDLS-----TEEA	145
MYSV:ATN96444	K--NCPSGMQYFILVLAENLAKPQRG--QFETFLTDDLPSQGSQ--LDESIKTKI--YASVELTA-----AQRE	131
BYSMV:YP_009177222	K--NCPAGMQYIILVLAQLRSKTERG--QYEFSLAEDRPTGGSD--LDNSIKSHI--YANVTVAD-----SEKE	128

: : *

:

TrARV1	QGSSVAAS-----LKE--RLAARKG-----QKMVNEIIT----KTT-----A	194
WhIV-5:YP_009300870	PEPDKNLS-----VAE--RLAAKKA-----QSKASEIVT----SAS-----I	192
PeVA:YP_006576501	QATNPKIS-----VKA--QKLIDRR-----KKEASQIVD----EEE-----	190
WhIV-6:YP_009301356	ISETENLS-----FRE--KLKRRKE-----ASKAVREDV----EAT-----A	192
TpVA:AYH53274	ENEAEALA-----KLQOI-----DAGTQLADNVNKNIN-----E	188
TYMaV:YP_009352242	EEERNTLALLNVDVNI--ENQDNGN-----NQQGENV-----	151
SAV1:QDJ94294	ENERRLLR----NLNQ--HGADQGG-----PDQVAEEVL----QND-----	151
ADV:YP_009177015	NIFEAD-----KQPRKRGKGGKSTKTKGVGDELV----NNI-----E	210
SCV:AWK49426	EMLEVEMR-----KNSSKSGKKGKSTAADVAI-----KMA-----Q	195
RVCV:QBS46637	EDVISGIF----ANASSGKRTNKQG--GNKTRNPESDIALR----QAQ-----N	219
WhIV-4:YP_009300684	SLFETRLK----AARR--NYETADE--SKKEKLAGMIEQLLEADIATSQS-----S	223
CCyV-1:ATS17308	ATQQMALN----RARE--RLQETDAERQADLQRTIDRL-----QQV-----N	198
LNYV:YP_425087	AIEKMTLN----ATRK--NLLTETDEEKRYEAIKIKME----DQE-----A	191
LYMoV:YP_002308371	QLQQQLV----KSRE--AFAAETDEERKTALDKIIGLE----AQE-----A	187
TpVB:AYH53268	EIQKKQLE----KSRV--AIGEETDEEKKVLQTIINRL----AQE-----S	187
BaCV:QAU20935	-----VFAAWNK--EQAEEGGEEDGKESESEGESE-----QEE-----ISVDPK	151
RSMV:YP_009553363	LDSATTDA----DTKK--AAEYEAIGDEDSKKAEKATANTA--WIAAQEAQKKAQSAYDKAVSNAKKASRKTTSKGLF	209
CBDaV:YP_009362275	SDTDTTQI----DDEA--AVF-----	139
MaCyV:ARS22490	AKPDAEKK----RLTA--TKV-----F	161
MYSV:ATN96444	GKNETEVA----KLKA--DTI-----F	147
BYSMV:YP_009177222	GKSTEEVA----KLRA--DKI-----F	144

TrARV1	G-----SSNAPIEGISSENQA---AVYSFLAAFLRLRAHSRQPED--LIDALPKMVERCASWYEGAEIVIGNLTVEKLTIE	264
WhIV-5:YP_009300870	S-----GVPQQDGCSCSAEEQA---AIYSFIAAFMLRLHSRQTD--FSASLESMTQTRCGGWYEKADTILKQLNLETEILN	262
PeVA:YP_006576501	-----KKNDDVSVSKEGQA---SAYCFIAAYLMLRYSRTAES--FCASLDLMRSRFGSWYEGSSILDSFEMEETIAD	259
WhIV-6:YP_009301356	GIS----QAPADAENDPKEYQA---AAYAFIAAYLRLRSTRQAKP--TLEKMGTKERFTSFYDKGHATLDTFNIQSETME	264
TpVA:AYH53274	G-----ETKQEERDFSEEVKA---SFYSFVAGYLLRIQTRQVDN--VHSNFTKSKDRYAGWYDEGQELFDSNDIPKDSME	258
TYMaV:YP_009352242	-----QVDQGSIDMKA---AAYAYVCAFMLRLQCRSAQN--VSGGLRRAIDRFKTYWDGSDTIDFDEISFSIDALN	216
SAV1:QDJ94294	-----DEEAEDRGSAAAMRS---AAYSICYAYLMLRLQCRNAKN--VHDGLERAERFRRAWYDDKHGILDDLFSFVSDSLN	220
ADV:YP_009177015	GVRNNHREADAELVGENKTYQA---AAYSFAAFLRLRQCKREDN--TVPAFEKAVTRYNGFYDSGGETLHGLALTLESCR	286
SCV:AWK49426	RVRA---TPKQDNIQSDKTQQA---AAYSIAAYLMLRMCQCRQPEQ--VVGSI TNVINRYNGFYDGGSDVFNLDLTVSALE	268
RVCV:QBS46637	SRRN---NTPLQVEDSDKEHMA---AAYSISAFMLRLQCRQPDSCMMTAIEKARARYNGFYDEGQGFDSMNVSEESLT	293
WhIV-4:YP_009300684	T-----KGAKEVKPSNAEDA---FVYSYMAAYFLRLYKNTSEA--VVTKLEVAKKAATAYDLPNSALDEIDLSDAQAS	293
CCyV-1:ATS17308	G-----ERVNAPANHVNESDA---TAYCFIAAFIMKLNKAEDA--FQEGIAKMKIRYPAWYEGGSQVLLNFNPTLETLK	268
LNYV:YP_425087	G-----LGTSKATVVTSETEA---AAYGFLAATLLKLYAKSTES--YVAGLAQIRNRFAAWYDCPKAVLDAFQPTAALV	261
LYMoV:YP_002308371	G-----TPSTSRQVVDTEEA---AAYGFLAIIKLCAKTAES--FIEGLPRVRDRFSSWYDTSQVIKTFNPTESVNL	257
TpVB:AYH53268	G-----VIGTREQEMGDQSDA---AAYSFLAASLILKLCAKTAES--YIEGLPRMRDRFSSWYDTSNINLRTFTPTETALN	257
BaCV:QAU20935	GKKT--GDPQSAITAIDKERAANCVRFCCFAAFYLRLLCKIKDIQD--LSTPYTRMKERYTRFYQGVIPSTSGEDIPPSSESVK	228
RSMV:YP_009553363	GETV---TDESKVVEEVGEGKK---KFGPFLAAYLMLRLTKIASN--VTESWEHMKGMKNFYGYDAPS--DLNCPAEGFLE	281
CBDaV:YP_009362275	GRTA---KAGNATV-----KEVCFASAFLLRGLVKTVTN--ITKAFELKGGFRVSLYNLPKDD--TFVCPAAAWLT	203
MaCyV:ARS22490	GSEE---DKFEKAVRELEEEIV---FQFGCFLSAFLKLLCKNPEN--IITGWDAMRGRYSTFFSESAPS--EVKRPSEWLT	234
MYSV:ATN96444	GADS---TSPLKTK--NDEELV---KNATFIAAFLRLILTKSVDN--VTAAWTGMSRFTNFYGDPMMT--EVKTPGRDYL	217
BYSMV:YP_009177222	GAAN---TSPLRTQ--KDEDLV---KNATFIAAYLRLALTKSADN--VGAAWTGMMKARFLSFYGMSELLQ--SITLLDRNYVV	214

: . . : :

:

TrARV1 SMKNNLLARREEVMSWVMWVSYNENEVI -- LNKNNYGMLSYLAGQIYQYTGHLHAVVQVLAIQQVTVQVPMQDQLLQELNHR 342
 WhIV-5:YP_009300870 SFKVMLGRRREEVMSWVMWVAYNENSNT -- LSKNNMGLMSYIAGQIYQYTGHLHAVVQVLAIQQVTVQVPMQDQLLQELNHR 340
 PeVA:YP_006576501 DLKGLMSKKNLNTWVMWVAFNENASV -- LDRNNRGLMEYLAGQIFSYTSLHVVTQT LAIQYTKCDMAMLLSELESPM 337
 WhIV-6:YP_009301356 KLRGIISRKKEAVGTWVWLSYNNENSNS - ELTSNELGLMEYVGLQIYAYMGHVAVTQTLALRQVSKLPLDVLTLQLCCQL 343
 TpVA:AYH53274 KLVAVMIRKPEITGTWVMWLAHTENECE -- LTKQPRGLLEIYGLQIFAYQGLHIMNQLLHLKALSASGMLLRQLDCPI 336
 TYMaV:YP_009352242 QLKEAISRKPEITSTWVLYLATTENEKT -- LLKQSKGMIYGLQVFSYQGMHALTQVLAHQMSKVPRLDMMELD SPL 294
 SAV1:QDJ94294 KLDKDAISRKPEVITWVHLAVTENEKQ -- LLQPKGMLEYGLQVFSYQGMHALTQVLAHQMSKVPRLDMMELD SPL 298
 ADV:YP_009177015 ALREVI GRKPELIGTWAVWVAYNENERKAGMLKQDAGLLEYLAIQVFAFQGMHVVTQT LAIHQLSAVPLGKLLREMDQCM 366
 SCV:AWK49426 KMRDVIARKPEITGTWVAVWVAYNENTQK -- LVKQDRGLLEYLAIQVFAFQGMHVVTQTLAIHQITKIPMGLLRREMDQCM 346
 RVCV:QBS46637 MIREVLARKPEITSTWVAWTAYNENEVT -- LGRQDKGLLEYLATQVFAFQGMHVVTQTLAIHQITKVPIGKLLAQMDQCM 371
 WhIV-4:YP_009300684 IIREAMSKKPEINKTWLWCAYNENENK - NLSQNAIGMLRYLAIQMFAYTSMHAYSFVVQMQTETGVSRFDILTELCCSA 372
 CCyV-1:ATS17308 ALRTIFNRPEILSTWVMTVAVNENREG - VMLPTHQGLLNLVLCQQYSYFGMHAYSLLSIHEATGIKLGQLREMDCP 347
 LNVV:YP_425087 SLRAAFARRPEVLSWTWVAVNENRTP - GLLVTQGLLNLVLCQQFAYPGMHAYTLLIEIHEHTGMKFSDDLVEMDCPA 340
 LYMoV:YP_002308371 TLRTGFGRPEILSTWVWVAYNENRED - PLLITQGLLTYLAGQQFSYPMHAYTLLIEIHEHTGIKFGQLREMDCPA 336
 TpVB:AYH53268 SLRTAFSRRPEILSTWVWVAYNENKAE - KLLVTHQGLLTYLAGQQFSYPMHAYTLLIEIHEHTGIKFGQLREMDCPA 336
 BaCV:QAU20935 SLKTFAGDRIRSRWVWVAVNENKAE -- VGSPOQGMVYIYLAFLPFAFSGMHAMKLMDDVIMETHKLDLTIYQMMHPR 306
 RSMV:YP_009553363 QLKSELNKDRRAATSWVKIVAEADNKLD - QSTAEAGILRYVAVLPLAYSGMHAMKLMFMDVIMETHKLDLTIYQMMHPR 359
 CBDaV:YP_009362275 SYKEFLTADPMIARTWIKIVAAEQGLD - AGSNDMGVLRFLACQPLSYSGMHAMKLYLTIKDKTKLSHKWLEKVMMPA 281
 MaCyV:ARS22490 ELKNLFSSDPMISRTWVRTFSEADHLP - AGEQQVGMRLRYLALMPFSYTLGHYKFLFDVSRASNL SNKWLLEMTSPM 312
 MYSV:ATN96444 DLKNNLASDPKIGHTWVKVVASAEHRLD - VSENTAGMIRFLAVLPLSLTGMHAYKFLFLEIKRQSNLSEKWLQELITPR 295
 BYSMV:YP_009177222 DLKNNLASDPKVGHTWVRTVSAEHRSL - VNDNTAGMIRFLVLP LLSMTGMHAYKFLFLEIKRQSNLSEKWLRELTSPR 292
 : : : * . . * : : : * : : : :

TrARV1 TRQPLLALYDMLQHHEIVKEKPGKRTYFRYARIWDSGYFHELQSKQCPDLVYLAAILKEVSPTGAR - SDPTKIYAIQDL 421
 WhIV-5:YP_009300870 TRQPLQALYKMLQHELVKEKPGRTTYFRYARVWDSGYFHDLSKQVCPDLVYLAAILKEVSPTGAR - SDPTKIFAVKDL 419
 PeVA:YP_006576501 TRGTVQELYKLRDYEVTTLHPDRRTTFRYARNWDAGYFAPLQTKKCTNLVYLTASVIKLIAPTGAR - SDPTKIYGLSDL 416
 WhIV-6:YP_009301356 TRVAVDEIYNIVKNEYITDRHPNRTTYFRYARAWDSGYFGKQVSKCEPQLVYLAAILKEVSPTGAR - LK - SDPTQIYAVLSI 421
 TpVA:AYH53274 TRDGVREYVNIKNEVTTKAPTRKTYFRYARVWDSGYFTQIRSQSCAPLLYTVAKVKEVDN - AK - SDPTQIYAIQNI 414
 TYMaV:YP_009352242 TRDGLREISNILKNHERTNRPARTTYFRYARVWDTKYFAQLQSKTCVPLLYVAAVAVRDISAN - TT - SDPTQIYALQNI 372
 SAV1:QDJ94294 TRDGLREIANILRNEYQTRHPDRKTYFRYARVWSPKYFAQLQSKICVPLLYVAAVTVRDISPN - SI - ADPTQIYALQNI 376
 ADV:YP_009177015 TRRAVEEVYGIKHNQATTAPHERKTYFRYARVWNEGYFSAVQSKSCTHLLYLASKVVKQLNPS - AG - SDPTQIYALKDM 444
 SCV:AWK49426 TRSAVMEIFKILRDFQPNVDVHQRNRTTYFRYARVWNEGYFARVQSKNCPQLLYLAAILKEVSPTGAR - ST - SDPTQIFAVKMM 424
 RVCV:QBS46637 TRAAVTEIYYIVRDYQKNEHQDRKSYRYARVWSDGYFSKVQSKACGQLVYLAAILKEVSPTGAR - TN - SDPTEIYAIKDL 449
 WhIV-4:YP_009300684 TRAGVQAKILREYELTEGHPDRKTYFRYARVWSDGYFTSLQTSNCKMLAYVAAKTMKNLSST - NM - SDPTEIYALQSV 450
 CCyV-1:ATS17308 TRAGVMAAFDLIKNHEITSKNPARTTYFRYARVWNSNYFRALQSSNCTLLVYVAAKVAKITSAQKVG - GDPMEIYALKNI 426
 LNVV:YP_425087 TRAGVREALELIRDYEITKDHPRKTYFRYARNWDPKYFGALQSTQCKTLVYVAASVSKKVAQGAN - GDPMEIFAIKNL 419
 LYMoV:YP_002308371 TRAGVKEVLDIIEKDFEITKHPKRTTYFRYARNWDPKYFGTLQSTQCKTLVYVAASVCKKISAQGEK - GDPTEIYAIKTL 415
 TpVB:AYH53268 TRDGVKEVLEIIRDYEVTKAHPNRTTYFRYARNWHDYKFGNLQSTQCKTLVYVAASVCKKISAQGLS - GDPTEIYAIKSL 415
 BaCV:QAU20935 HNSALASISEIYAKHY - PEEGSKNTYVYARVWVGPQFFPALQTKKCKGLVLCVLANLVVHQQKTTTEETGDPRTSITAGIDDI 385
 RSMV:YP_009553363 TKDALDAIMDILISFE - STTKTKKSEKFRFARIVSTQFFQSLQTKNCKELVYLMVQIIAEYRKAEGV - RDPMNIAGLDDI 437
 CBDaV:YP_009362275 TTPALTEIANLLKNFE - STESAKKPTKFRYARLGSVYFQKLTQKNCLELVFLEVCILNHFHTTFAEDYQNPTKIVGVERI 360
 MaCyV:ARS22490 TSKALKLIAHILINFE - SRIGEKKSAFKYARMSTSYFQDLQTKNCAGLVYLVKILKKEAFAGTN - SNPENIIGIEKV 390
 MYSV:ATN96444 TLPALIEIKKILVNFE - DRTEKKSCKGFRYARLMSPAFFQELQTKNCPDLVYVEVCILNRYEAFPT - QDPNKIIGIEKV 373
 BYSMV:YP_009177222 TKPGLTEIRKILINFE - DRVNEKSGKFRYARLMSPAFFQELQTKNCPDLVYVEVCILNRYEAFPT - QDPNKIIGIEKV 370
 : : : : * . : * : : * * : * . * : :

TrARV1 GETKKEFLDRVATNISRWL -----VEADDDTGDSGASWAM----- 456
 WhIV-5:YP_009300870 GEVKKELLDGVANGISQWL -----VAADDNSSESGAFW----- 452
 PeVA:YP_006576501 GEMTKTKLDKVKAMKLSDDL -----TRVEGDENTGSCWIE----- 450
 WhIV-6:YP_009301356 GESIMEKLNKVSNDLVSYI -----MDSMAEDEDAGDIWK----- 455
 TpVA:AYH53274 GDALKARLDVAANKLVDMLE -----LNDTTQDVESGDVWQATE----- 451
 TYMaV:YP_009352242 GSAMKASLDRVAAALTSYL -----VDKSYRDAKSGSVWDAATPTE----- 412
 SAV1:QDJ94294 GASMKELNVRVAARLVAFV -----MERTLSDDKSGSIWDAAVEENHTEAS----- 421
 ADV:YP_009177015 GEDQKARLDKAVANKLLDFI -----WTQANDPEAGSIWKDIV-----G 482
 SCV:AWK49426 SASMREVLDSVSKLLALI -----SQHADNDKEAGSIWKGHKFGQGG----- 466
 RVCV:QBS46637 SAAARERLDRVSAKLIDYI -----WLANNDEDAGGIWKT----- 484
 WhIV-4:YP_009300684 SEKMRSTLDGVADNLDYIL -----MAKMTQDSDAGNSWKI----- 485
 CCyV-1:ATS17308 DEVMLTRLNKVAAMKSELI -----LTAMMEDEIAGVAWQ----- 460
 LNVV:YP_425087 DATIKARLDVPAENMAGKI -----LDQMLMDEMSGASWATKASTQ----- 459
 LYMoV:YP_002308371 DATIKERLEAVANRMAHKI -----IDQMLVDAMSGEAWVGGN----- 452
 TpVB:AYH53268 DPTLRGRILDVAVDKMAHMI -----IDQMLVDTLSGKAWT----- 449
 BaCV:QAU20935 SDELMMKFGKRMAMIMKNAPKTASYMYSSTAKEA ---EEGGDSTDEESEEDIDPEPLKVGKRIYKP--- 451
 RSMV:YP_009553363 SSRNKKLKNKAVRIILAEAPKASAGEYSSAMKKAFLDDEEDDTAKTRS-----IFQTKA 491
 CBDaV:YP_009362275 PEAMKTKLKAANAIVSRAPVLNPSLYSDIMSGVFLSKTEASTSGTAAQTSGLK-----TEQAIFG--- 422
 MaCyV:ARS22490 PAEHKAMLSKAADFIVSAAPQRNLGRYSESMRKAMIEPHKPKHQVAKKE-----DIFG--- 445
 MYSV:ATN96444 PESMKERLKNAAYNIFVAAPQRNAGKYS TSMKKAFLPTEKAQSARTEVKKT-----TADNIFT--- 431
 BYSMV:YP_009177222 PESMKDQLKEVANNIVTAAPQRNAGMYS TSMKKAFLPTEKKAQSARTEVKKT-----ADQIFS--- 427
 : : : : * . : * : : * * : * . * : :

Fig. S3. Phylogenetic relationship of TrARV1 and other cytorhabdoviruses based on N proteins