

A novel tepovirus, Agave virus T, identified by the analysis of the transcriptome data of blue agave (*Agave tequilana*)

Chul Jun Goh, Dongbin Park, Yoonsoo Hahn*

Department of Life Science, Chung-Ang University, Seoul 06974, Republic of Korea

Received January 8, 2021; accepted January 19, 2021

Summary. – The genome sequence of a novel RNA virus was identified by analyzing transcriptome data obtained from the stem sample of a blue agave (*Agave tequilana*) plant. Sequence comparison and phylogenetic analysis showed that the RNA virus, Agave virus T (AgVT), was a new member of the genus *Tepovirus* in the family *Betaflexiviridae*. AgVT genome had three open reading frames: a 1605-amino acid (aa) replicase (REP), 355-aa movement protein (MP), and 220-aa coat protein (CP). Phylogenetic analyses based on the REP, MP, and CP sequences of AgVT, previously reported tepoviruses, and other *Betaflexiviridae* viruses revealed that tepoviruses could be classified into two subclades: “potato virus T (PVT)-clade” and “Prunus virus T (PrVT)-clade.” PVT, the type species and founding member of the genus *Tepovirus*, belong to “PVT-clade” along with AgVT, while the other five tepoviruses belong to “PrVT-clade.” The genome sequence of AgVT may be useful for studying the phylogenetic relationships between tepoviruses and other closely related viruses.

Keywords: Agave virus T; Tepovirus; Betaflexiviridae; blue agave; Agave tequilana

Introduction

Tepoviruses (the genus *Tepovirus*) are plant-infecting RNA viruses of the family *Betaflexiviridae*. The members of this family have a monopartite, positive-sense, single-stranded RNA genome of size 6–9 kb, and the virions exist as flexuous filamentous particles (Adams *et al.*, 2012; Rubino *et al.*, 2012). Thirteen genera in the family *Betaflexiviridae* have been reported and were classified into two subfamilies: *Trivirinae* and *Quinvirinae* (<https://talk.ictvonline.org>, last accessed on December 5, 2020).

The subfamily *Trivirinae* comprises 10 genera, including *Capillovirus*, *Chordovirus*, *Citrivirus*, *Divavirus*,

Prunevirus, *Ravavirus*, *Tepovirus*, *Trichovirus*, *Vitivirus*, and *Wamavirus*. Viruses of the subfamily *Trivirinae* have three common open reading frames (ORFs) that encode a replicase (REP), movement protein (MP), and coat protein (CP) (Adams *et al.*, 2012; Rubino *et al.*, 2012). Members of the genera *Capillovirus* and *Divavirus* have a fused REP-CP ORF, while those of the other genera have independent REP and CP ORFs (Yoshikawa *et al.*, 1992; Goh *et al.*, 2018). The subfamily *Quinvirinae* consists of three genera: *Carlavirus*, *Foveavirus*, and *Robigovirus*. Viruses of the subfamily *Quinvirinae* have five common ORFs encoding an REP, three triple gene block proteins (TGB1, TGB2, and TGB3), and a CP (Morozov and Solovyev, 2003; Prosser *et al.*, 2015). Members of some genera have one or more additional ORFs (Martelli *et al.*, 1997; Adams *et al.*, 2012; Veerakone *et al.*, 2018).

Blue agave (*Agave tequilana*), also known as tequila agave, belongs to the genus *Agave*, which are succulent monocotyledonous plants native to the arid regions of North America (Coleman-Derr *et al.*, 2016). *Agave* species are highly tolerant to drought and heat stress, because

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

Abbreviations: AgVT = Agave virus T; CP = coat protein; GVA = grapevine virus A; MP = movement protein; ORF = open reading frame; PrVT = Prunus virus T; PVT = potato virus T; RdRp = RNA-dependent RNA polymerase; REP = replicase

they employ the crassulacean acid metabolism in photosynthesis, which enhances water-use efficiency (Borland *et al.*, 2009). Blue agave is an economically important plant in Mexico used in the production of the popular distilled spirit tequila (Cedeno, 1995). Because of its economic importance, blue agave has been subjected to various molecular genetic studies, including transcriptome analyses (Gross *et al.*, 2013; Coleman-Derr *et al.*, 2016; Huang *et al.*, 2018).

Transcriptome data obtained from samples of plant tissue infected with RNA viruses often contain virus-derived reads (Nibert *et al.*, 2016). Contig assembly of plant transcriptome data and comprehensive sequence analysis have yielded numerous complete genome sequences of novel RNA viruses (Kim *et al.*, 2018; Park *et al.*, 2018, 2020; Goh *et al.*, 2019, 2020). In this study, we identified the genome sequence of a novel virus belonging to the genus *Tepovirus* of the family *Betaflexiviridae* in the transcriptome dataset acquired from the stem sample of a blue agave plant (Gross *et al.*, 2013).

Materials and Methods

The transcriptome data (a total of 71.8 gigabases) obtained using the samples from the stem of a blue agave plant (Gross *et al.*, 2013) were downloaded from the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI). SRA Acc. Nos. are SRR789714, SRR789715, SRR789716, SRR789717, SRR789718, SRR789719, SRR789720, SRR789724, SRR789725, SRR789726, SRR789727, and SRR789728. Low-quality reads were filtered out using the sickle program (version 1.33; <https://github.com/najoshi/sickle>) with the parameter “-q 30 -l 55.” High-quality reads from all 12 sequencing runs were pooled and subjected to *de novo* transcriptome assembly using the SPAdes Genome Assembler (version 3.14.1; <http://cab.spbu.ru/software/spades>) with the parameter “-rna.”

The transcriptome contigs were compared with known viral RNA-dependent RNA polymerase (RdRp) domain sequences using the DIAMOND program (version 2.0.4; <http://www.diamondsearch.org/index.php>). A total of 2565 viral RdRp domain sequences were selected from 22 families (PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03035, PF03431, PF04196, PF04197, PF05788, PF05919, PF06317, PF07925, PF08467, PF08716, PF08717, PF12426, and PF17501) available in the Pfam database (release 33.1; <https://pfam.xfam.org>).

Sequencing depth of a putative viral contig was assessed by mapping the RNA-seq reads to the contig sequence using the URMAP program (version 1.0.1480; <https://drive5.com/urmap>). Putative ORFs in the viral genome sequence were inferred based on BLASTX search results against all known viral protein sequences. Conserved domains in protein sequences were

predicted using the InterPro web server (version 82.0; <https://www.ebi.ac.uk/interpro>).

Pairwise identities of protein sequences were examined using the needle program of the EMBOSS package (version 6.6.0.0; <http://emboss.open-bio.org>) using default parameters. Multiple sequence alignments were generated using the MAFFT program (version 7.475; <https://mafft.cbrc.jp/alignment/software>) with the parameter “--auto.” Phylogenetic analysis was performed using the neighbor-joining method in the ClustalW2 program (version 2.1; <http://www.clustal.org/clustal2>), after removing gaps in the sequence alignment.

Results and Discussion

RNA-seq reads obtained from the blue agave stem sample were assembled into contigs (Gross *et al.*, 2013). When these contig sequences were compared with known viral RdRp sequences, numerous contigs were identified to contain an RdRp domain. Among them, a 6408-bp contig showed strong sequence similarity to that of the RdRp domain of the potato virus T (PVT) REP sequence (UniProt Acc. No. B5ACE2) (Russo *et al.*, 2009). PVT is the type species of the genus *Tepovirus* of the family *Betaflexiviridae*, suggesting that the contig was derived from a virus related to the genus *Tepovirus*.

A BLASTX search of the NCBI protein database using the 6408-bp contig as a query confirmed that it had ORFs with the highest sequence similarities to proteins encoded by the PVT genome sequence. Therefore, this contig was considered the genome sequence of a novel virus belonging to the genus *Tepovirus* of the family *Betaflexiviridae*, and it was tentatively named as Agave virus T (AgVT). The genome sequence of this virus was deposited in the NCBI database (Acc. No. MW323519).

AgVT genome was predicted to have three complete protein-coding ORFs (Fig. 1 and Table 1): ORF1, encoding a 1605-amino acid (aa) REP; ORF2, a 355-aa MP; and ORF3, a 220-aa CP. The MP ORF overlapped with the REP and CP ORFs. AgVT proteins were predicted to possess conserved domains that are typically present in other members of the family *Betaflexiviridae*. REP had a viral methyltransferase domain (InterPro Acc. No. IPR002588) at aa positions 43–329, a “viral RNA helicase core domain” (IPR027351) at aa positions 761–1077, and an “viral RNA-dependent RNA polymerase domain” (IPR001788) at aa positions 1252–1493. MP had a “viral movement protein domain” (IPR028919) at aa positions 11–194. CP had a “viral coat protein domain” (IPR008879) at aa positions 40–215.

For sequence comparison and phylogenetic analysis, genome and protein sequences of previously reported tepoviruses, including four PVT, two Prunus virus T (PrVT), two Zostera virus T (ZoVT), one cherry virus T

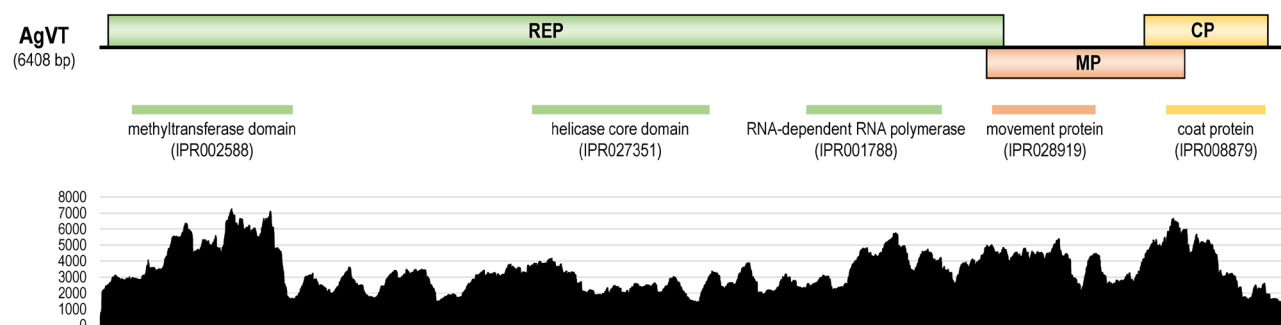


Fig. 1

Schematic representation of the AgVT genome organization

A schematic figure of the AgVT genome sequence is presented at the top. Three ORFs corresponding to the replicase (REP), the movement protein (MP), and the coat protein (CP) are depicted as boxes. Predicted InterPro domains are marked using lines below the ORFs with corresponding InterPro Acc. Nos. The sequencing depth is shown at the bottom. See Table 1 for the coordinates of the ORFs and InterPro domains.

(ChVT), one *Ficus tepovirus A* (FiTA), and one *Trichosanthes tepovirus A* (TrTA), were collected (Russo *et al.*, 2009; Marais *et al.*, 2015, 2020; Goh *et al.*, 2019). Sequences of representative members of the other 12 *Betaflexiviridae* genera were also retrieved, including grapevine virus A (GVA), Actinidia virus B (AcVB), apple stem grooving virus A, *Ribes americanum* virus A, grapevine Pinot gris virus (GPGV), carrot Ch virus 1, *Ocimum basilicum* RNA virus 1, citrus leaf blotch virus, apricot vein clearing associated virus, watermelon virus A, cherry rusty mottle associated virus, Phlox virus B, and apple stem pitting virus (Yoshikawa *et al.*, 1992; Nakaune *et al.*, 2008; Goh *et al.*, 2018). Thirty-four known *Betaflexiviridae* virus genome sequences were retrieved for sequence comparison and phylogenetic analyses (Table 2).

Pairwise comparisons of protein sequences revealed that AgVT REP had 25.7–43.8% aa identity with those of other *Betaflexiviridae* viruses. Four PVT REP sequences exhibited the highest similarity to AgVT REP sequence with 43.6–43.8% aa identity. Other tepovirus REP sequences showed 32.3–33.1% aa identity with AgVT REP sequence. REP sequences of viruses belonging to other

Betaflexiviridae genera showed 24.7–30.6% aa identity with AgVT REP sequence. Comparison between the REP protein sequences confirmed that PVT was the most closely related to AgVT, among the currently known tepoviruses.

Multiple alignments of 35 REP sequences from AgVT, 11 other tepovirus, and 23 other *Betaflexiviridae* virus genomes were generated (Supplementary Fig. S1). The phylogenetic position of AgVT within the family *Betaflexiviridae* was inferred using the neighbor-joining method (Fig. 2). The subfamily *Quinvirinae* was used as the out-group. AgVT was placed as a sister taxon of the four PVT sequences within the genus *Tepovirus* in accordance with the sequence comparison results. Among the genera of the subfamily *Trivirinae*, the genus *Vitivirus* was inferred to be the sister genus of *Tepovirus*.

PVT and PrVT are two currently approved species of the genus *Tepovirus* (<https://talk.ictvonline.org>, last accessed on December 5, 2020). According to the phylogenetic analysis, 12 REP sequences from seven tepoviruses were classified into two subclades: “PVT-clade” and “PrVT-clade.” AgVT belonged to “PVT-clade” as the second member after PVT, which is the type species and found-

Table 1. ORFs of AgVT genome sequence

ORF	ORF position (nt)	Protein length (aa)	Domain position (aa)	InterPro domain name	InterPro Acc. No.
Replicase (REP)	43–4860	1605	43–329	alphavirus-like methyltransferase (MT) domain	IPR002588
			761–1077	(+) RNA virus helicase core domain	IPR027351
			1252–1493	tymovirus, RNA-dependent RNA polymerase	IPR001788
Movement protein (MP)	4769–5836	355	11–194	viral movement protein	IPR028919
Coat protein (CP)	5616–6278	220	40–215	coat protein, trichovirus/vitivirus	IPR008879

Table 2. Sequence comparison of the REP of AgVT and representative *Betaflexiviridae* viruses

No	Subfamily	Genus	Virus	Acronym	NCBI	Identity ^a
1	Trivirinae	Tepovirus	Potato virus T	PVT	YP_002019748.1	723/1657 (43.6%)
2			Potato virus T	PVT	ADX41471.1	724/1657 (43.7%)
3			Potato virus T	PVT	AFU55321.1	723/1649 (43.8%)
4			Potato virus T	PVT	AXK90539.1	723/1652 (43.8%)
5			Prunus virus T	PrVT	YP_009051684.1	620/1881 (33.0%)
6			Prunus virus T	PrVT	AHM92766.1	620/1875 (33.1%)
7			Zostera virus T	ZoVT	QBS17025.1	621/1905 (32.6%)
8			Zostera virus T	ZoVT	QBS17031.1	616/1904 (32.4%)
9			Cherry virus T	ChVT	QNG41875.1	618/1893 (32.6%)
10			Ficus tepovirus A	FiTA	QED42804.1	605/1874 (32.3%)
11			Trichosanthes tepovirus A	TrTA	QED42832.1	607/1865 (32.5%)
12	Vitivirus	Grapevine virus A	GVA	NP_619662.1	560/1830 (30.6%)	
13			Actinidia virus B	AcVB	YP_004935358.1	531/1813 (29.3%)
14	Capillovirus	Apple stem grooving virus	ASGV	NP_044335.1	536/1785 (30.0%) ^b	
15			Yacon virus A	YaVA	YP_009268859.1	524/1784 (29.4%) ^b
16	Ravavirus	Ribes americanum virus A	RAVA	YP_009553496.1	514/1971 (26.1%)	
17	Trichovirus	Apple chlorotic leaf spot virus	ACLSV	NP_040551.1	568/1976 (28.7%)	
18			Grapevine Pinot gris virus	GPGV	YP_004732978.2	551/1937 (28.4%)
19	Chordovirus	Carrot Ch virus 1	CtChV-1	YP_009103999.1	572/1959 (29.2%)	
20			Carrot Ch virus 2	CtChV-2	YP_009103996.1	568/1946 (29.2%)
21	Divavirus	Ocimum basilicum RNA virus 1	DiVA	YP_006905850.1	552/1857 (29.7%) ^b	
22			ObRV1	YP_009408144.1	547/1837 (29.8%) ^b	
23	Citrivirus	Citrus leaf blotch virus	CLBV	NP_624333.1	581/2042 (28.5%)	
24			CLBV	AFA43536.1	565/2080 (27.2%)	
25	Prunevirus	Apricot vein clearing associated virus	AVCaV	YP_008997790.1	532/1825 (29.2%)	
26			Caucasus prunus virus	CPrV	YP_009505632.1	569/2121 (26.8%)
27	Wamavirus	Watermelon virus A	WVA	YP_009357235.1	552/1907 (28.9%)	
28			WVA	QEA69426.1	558/1922 (29.0%)	
29	Quinvirinae	Robigovirus	Cherry rusty mottle associated virus	CRMaV	YP_007761581.1	548/2121 (25.8%)
30			Cherry twisted leaf associated virus	CTLaV	YP_009046478.1	560/2163 (25.9%)
31	Carlavirus	Phlox virus B	PhlVB	YP_001552317.1	554/2155 (25.7%)	
32			Garlic common latent virus	GCLV	YP_004936159.1	544/2060 (26.4%)
33	Foveavirus	Apple stem pitting virus	ASPV	NP_604464.1	564/2279 (24.7%)	
34			Asian prunus virus 1	APV1	YP_009094347.1	572/2112 (27.1%)

^aProtein sequence identity in the form of “number of identical residues/aligned length (percent identity).” ^bCP regions were removed from the capillovirus and divavirus REP-CP protein sequences.

ing member of the genus *Tepovirus*. Bootstrap analysis strongly supported the placement of both subclades with 100% bootstrap values. However, the *Tepovirus* clade, and hence the monophyletic origin of “PVT-clade” and “PrVT-clade” viruses, was only marginally supported with a bootstrap value of 72.1%. This suggested the possibility

that “PrVT-clade” might be classified as a novel genus different from “PVT-clade.”

The monophyletic relationship between “PVT-clade” and “PrVT-clade” was further investigated using MP and CP sequences of all tepoviruses, two vitiviruses (GVA and AcVB), and a trichovirus (GPGV). Multiple alignments of

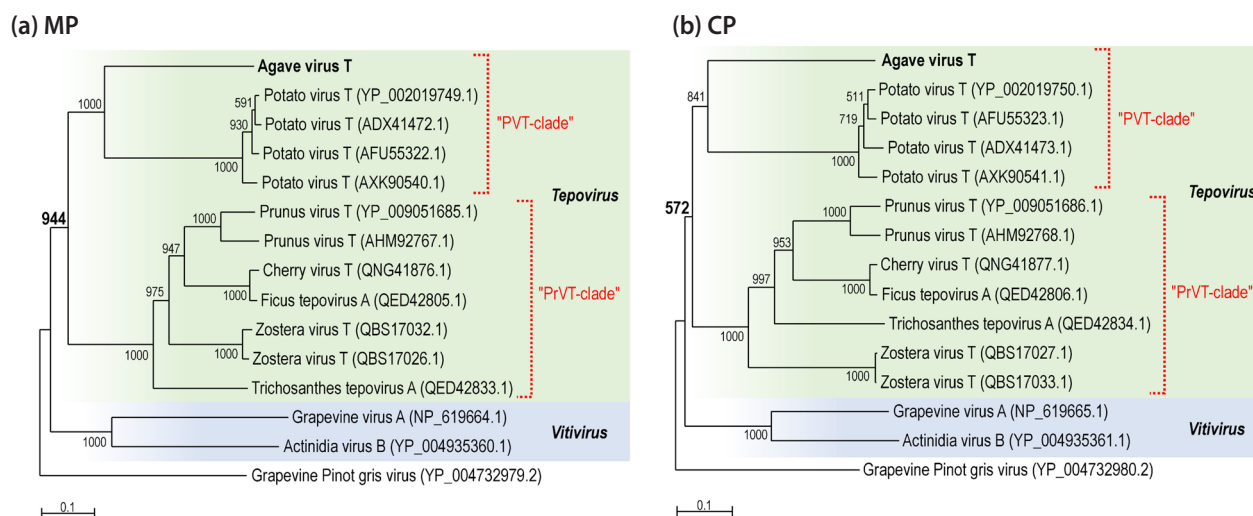


Fig. 3

Phylogenetic relationships of tepoviruses inferred from MP and CP sequences

Neighbor-joining phylogenetic trees were inferred from MP (a) and CP (b) sequences of tepoviruses, two vitiviruses, and a trichovirus. Two subclades ("PVT-clade" and "PrVT-clade") were reproduced within the genus *Tepovirus*. The sequence from the grapevine Pinot gris virus, a trichovirus, was used as the outgroup. The NCBI protein sequence Acc. Nos. are in parentheses. Bootstrap supporting values calculated from 1000 replicates are presented at the nodes. The bootstrap value for the *Tepovirus* clade is indicated in bold.

subclade. The AgVT genome sequence may be useful for studying the phylogenetic relationships of tepoviruses and other closely related *Betaflexiviridae* viruses.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) funded by the Government of Korea (grant Nos. 2018R1A5A1025077 and 2020R1A2C1013403).

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

References

- Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN, Ryu KH, Saldarelli P, N. Y (2012): Family - Betaflexiviridae. In King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (Eds): Virus Taxonomy: Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, pp. 920–941. <https://doi.org/10.1016/B978-0-12-384684-6.00078-1>
- Borland AM, Griffiths H, Hartwell J, Smith JA (2009): Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *J. Exp. Bot.* 60, 2879–2896. <https://doi.org/10.1093/jxb/erp118>
- Cedeno M (1995): Tequila production. *Crit. Rev. Biotechnol.* 15, 1–11. <https://doi.org/10.3109/07388559509150529>
- Coleman-Derr D, Desgarenes D, Fonseca-Garcia C, Gross S, Clingenpeel S, Woyke T, North G, Visel A, Partida-Martinez LP, Tringe SG (2016): Plant compartment and biogeography affect microbiome composition in cultivated and native Agave species. *New Phytol.* 209, 798–811. <https://doi.org/10.1111/nph.13697>
- Goh CJ, Park D, Hahn Y (2020): Identification of Trichosanthes associated rhabdovirus 1, a novel member of the genus Cytorhabdovirus of the family Rhabdoviridae, in the Trichosanthes kirilowii transcriptome. *Acta Virol.* 64, 36–43. https://doi.org/10.4149/av_2020_105
- Goh CJ, Park D, Kim H, Sebastiani F, Hahn Y (2018): Novel Divavirus (the family Betaflexiviridae) and Mitovirus (the family Narnaviridae) species identified in basil (*Ocimum basilicum*). *Acta Virol.* 62, 304–309. https://doi.org/10.4149/av_2018_224
- Goh CJ, Park D, Lee JS, Davey PA, Pernice M, Ralph PJ, Hahn Y (2019): Zostera virus T - a novel virus of the genus Tepovirus identified in the eelgrass, *Zostera muelleri*. *Acta Virol.* 63, 366–372. https://doi.org/10.4149/av_2019_409
- Gross SM, Martin JA, Simpson J, Abraham-Juarez MJ, Wang Z, Visel A (2013): De novo transcriptome assembly of drought tolerant CAM plants, Agave deserti and Agave tequilana. *BMC Genomics* 14, 563. <https://doi.org/10.1186/1471-2164-14-563>
- Huang X, Wang B, Xi J, Zhang Y, He C, Zheng J, Gao J, Chen H, Zhang S, Wu W, Liang Y, Yi K (2018): Transcriptome comparison reveals distinct selection patterns in domesticated and wild Agave species, the important CAM plants. *Int. J. Genomics* 2018, 5716518. <https://doi.org/10.1155/2018/5716518>

- 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>
- Marais A, Faure C, Mustafayev E, Barone M, Alioto D, Candresse T (2015): Characterization by deep sequencing of Prunus virus T, a novel tepevirus infecting Prunus species. *Phytopathology* 105, 135-140. <https://doi.org/10.1094/PHYTO-04-14-0125-R>
- Marais A, Šafářová D, Navrátil M, Faure C, Cornaggia D, Brans Y, Suchá J, Candresse T (2020): Complete genome sequence of cherry virus T, a novel cherry-infecting tepevirus. *Arch. Virol.* 165, 1711-1714. <https://doi.org/10.1007/s00705-020-04656-w>
- Martelli GP, Minafra A, Saldarelli P (1997): Vitivirus, a new genus of plant viruses. *Arch. Virol.* 142, 1929-1932. <https://doi.org/10.1007/s007050050088>
- Morozov SY, Solovyev AG (2003): Triple gene block: modular design of a multifunctional machine for plant virus movement. *J. Gen. Virol.* 84, 1351-1366. <https://doi.org/10.1099/vir.0.18922-0>
- Nakaune R, Toda S, Mochizuki M, Nakano M (2008): Identification and characterization of a new vitivirus from grapevine. *Arch. Virol.* 153, 1827-1832. <https://doi.org/10.1007/s00705-008-0188-5>
- Nibert ML, Pyle JD, Firth AE (2016): A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. *Virology* 498, 201-208. <https://doi.org/10.1016/j.virol.2016.07.002>
- Park D, Goh CJ, Kim H, Hahn Y (2018): Identification of two novel amalgaviruses in the common eelgrass (*Zostera marina*) and in silico analysis of the amalgavirus +1 programmed ribosomal frameshifting sites. *Plant Pathol. J.* 34, 150-156.
- Park D, Goh CJ, Lee JS, Sebastiani F, Hahn Y (2020): Identification of Pistacia-associated flexivirus 1, a putative mycovirus of the family Gammaflexiviridae, in the mastic tree (*Pistacia lentiscus*) transcriptome. *Acta Virol.* 64, 28-35. https://doi.org/10.4149/av_2020_104
- Prosser SW, Xiao H, Li C, Nelson RS, Meng B (2015): Subcellular localization and membrane association of the replicase protein of grapevine rupestris stem pitting-associated virus, family Betaflexiviridae. *J. Gen. Virol.* 96, 921-932. <https://doi.org/10.1099/jgv.0.000019>
- Rubino L, Russo M, De Stradis A, Martelli GP (2012): Tepevirus, a novel genus in the family Betaflexiviridae. *Arch. Virol.* 157, 1629-1633. <https://doi.org/10.1007/s00705-012-1342-7>
- Russo M, Rubino L, De Stradis A, Martelli GP (2009): The complete nucleotide sequence of potato virus T. *Arch. Virol.* 154, 321-325. <https://doi.org/10.1007/s00705-008-0300-x>
- Veerakone S, Liefting LW, Tang J, Ward LI (2018): The complete nucleotide sequence and genome organisation of a novel member of the family Betaflexiviridae from *Actinidia chinensis*. *Arch. Virol.* 163, 1367-1370. <https://doi.org/10.1007/s00705-017-3701-x>
- Yoshikawa N, Sasaki E, Kato M, Takahashi T (1992): The nucleotide sequence of apple stem grooving capillovirus genome. *Virology* 191, 98-105. [https://doi.org/10.1016/0042-6822\(92\)90170-T](https://doi.org/10.1016/0042-6822(92)90170-T)

A novel tepovirus, Agave virus T, identified by the analysis of the transcriptome data of blue agave (*Agave tequilana*)

Chul Jun Goh, Dongbin Park, Yoonsoo Hahn*

Department of Life Science, Chung-Ang University, Seoul 06974, Republic of Korea

Received January 8, 2021; accepted January 19, 2021

Summary. – The genome sequence of a novel RNA virus was identified by analyzing transcriptome data obtained from the stem sample of a blue agave (*Agave tequilana*) plant. Sequence comparison and phylogenetic analysis showed that the RNA virus, Agave virus T (AgVT), was a new member of the genus *Tepovirus* in the family *Betaflexiviridae*. AgVT genome had three open reading frames: a 1605-amino acid (aa) replicase (REP), 355-aa movement protein (MP), and 220-aa coat protein (CP). Phylogenetic analyses based on the REP, MP, and CP sequences of AgVT, previously reported tepoviruses, and other *Betaflexiviridae* viruses revealed that tepoviruses could be classified into two subclades: “potato virus T (PVT)-clade” and “Prunus virus T (PrVT)-clade.” PVT, the type species and founding member of the genus *Tepovirus*, belong to “PVT-clade” along with AgVT, while the other five tepoviruses belong to “PrVT-clade.” The genome sequence of AgVT may be useful for studying the phylogenetic relationships between tepoviruses and other closely related viruses.

Keywords: Agave virus T; Tepovirus; Betaflexiviridae; blue agave; Agave tequilana

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

AgVT PVT:YP 002019748.1 PVT:ADX41471.1 PVT:AFU5321.1 PVT:AKX90539.1 PrVt:YP 009051684.1 PrVt:AHM92766.1 ZQvT:QBS17025.1 ZQvT:QBS17031.1 ChVt:QNG41875.1 FiTA:QED42804.1 TrTA:QED42832.1 GVA:NP 619662.1 AcVb:YP 004935358.1 ASGV:NP 044335.1 YaVa:YP 009268859.1 RAva:YP 009553496.1 ACLSv:NP 040551.1 PGVv:YP 004732978.2 CtChv-1:YP 009103999.1 CtChv-2:YP 009103996.1 DiVa:YP 006905850.1 OBrV1:YP 009408144.1 ClBv:NP 624333.1 ClBv:AF443536.1 AVcAv:YP 008997790.1 CrVv:YP 009505632.1 WVa:QEA69426.1 CrMv:YP 007761581.1 CtLAv:YP 009046478.1 PhLbV:YP 001552317.1 GCLV:YP 004936159.1 ASPV:NP 604464.1 APV1:YP 009094347.1

VIQRC-LLTQSQVAPSTAMSFSWLLG---FGTGVAVRINIFKLLITYLMSLKKPDPVSAVAKVRLQCNSEIFDLELFAVHLGKLIAMEMG-DNFGROGIVKFLVRGLA 370
MIKGS-FDVSERFDRSSALTTSLMLPTR---AGKALRIRKFLRLRIIVLFSKLPDPHSAIAKTRQSSDDSIFCDEIMLADHVGGKIFEKLDP-ASPFGVKGVFDLLTSTIFK 369
MIKGS-FDVSERFDRSSALTTSLMLPTR---AGKALRIRKFLRLRIIVLFSKLPDPHSAIAKTRQSSDDSIFCDEIMLADHVGGKIFEKLDP-ASPFGVKGVFDLLTSTIFK 369
MIKGS-FDVSERFDRSSALTTSLMLPTR---AGKALRIRKFLRLRIIVLFSKLPDPHSAIAKTRQSSDDSIFCDEIMLADHVGGKIFEKLDP-ASPFGVKGVFDLLTSTIFK 369
LIEKGG-LETNDYISDEAEASLRFMGTA-AGK-ITAPTRVEVLKREIMVLLSLKSKDMNSAAAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 373
LIERGG-LKTNPFVSDFAECINAKMIGFSR-ISM-ESTPITRELVFKREIIVLMSLKKPDPMSATAKRLMSSSEDYTTGELFFNALACEIDSV---KGLHADISLLKHVQAM 373
LIEKGG-LKTNPFVSDFAECINAKMIGFSR-ISM-ESTPITRELVFKREIIVLMSLKKPDPMSATAKRLMSSSEDYTTGELFFNALACEIDSV---KGLHADISLLKHVQAM 376
LIEKGN-FETSPYVSDYSEICISARLFGFTR-YN-SFMPITREVLKREIIVLMSLKKCPDNPSAAAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 371
LVEKGN-FETSPYVSDYSEICISARLFGFTR-YN-SFMPITREVLKREIIVLMSLKKCPDNPSAAAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 371
LQKGN-LKTSERVISDSEICLSSVFDTSR-IKRGLPTIRMKIFKSEIIVLMSLKKPDPMSATAKRLMSSSEDYTTGELFFNALACEIDSV---KGFVADLIPAFIDSI 375
SALKGS-ITTEDEYTYDDYTPRGLSRRG-RDY-SGMPLRSRYQAVLLYLALMALKKPPDSSAAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 381
VASLGS-ITTEDEYTYDDYTPRGLSRRG-RDY-SGMPLRSRYQAVLLYLALMALKKPPDSSAAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 382
SFDKGG-ACNEFNH-FDKPSCLEAEMLRLT-KRF-DKAVINRSVSSLSYMAKLTANAASAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 386
SFDKGG-ACNEFNH-FDKPSCLEAEMLRLT-KRF-DKAVINRSVSSLSYMAKLTANAASAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 386
YLRKGD-FPTNDVFSVNSSELVDLSSMRQLG-LRY-PSKGRFVSVLKLVMYALSLKKTIDNNSFAKCRQDPENLILLSALLVNDLTKWLEVGT---GSLSEDDLLRRLGDSLF 376
HIYPAEDLMKEEVRFFSDYDLFDVGLFVKP-VRV-PIQDFPVLVFKKIFIVYMSLKKPDPVSAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 375
KISYKESIEQLSHRAFPGYDVLVDGSLFRGK-VRV-SIEGVSITFKKIFIVYMSLKKPDPVSAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 371
IVYKGR-RISRSIRLRFKGFYDVLVDGSLFRGK-VRV-SIEGVSITFKKIFIVYMSLKKPDPVSAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 390
AITKGG-RIVKEIRRFREFDILDLSDFSGTE-YKL-PVVDFVSVFFKIVYMSLKKPDPVSAVAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 390
YISRGS-LASESRFRFDFNLDLKVYAKNLLKMKLIRNLFNNFMKIVSYKSLKPKDPMSATAKRLMSSSEDYTTGELFFNALACEIDSV---KGIKDRGLVSTVTDGFA 389
FLSRGG-RVDESKRFFDFNLDLIPSRFNFHQTFRDVMNLIIRANFTHKISYKSLKPKDPMSATAKRLMSSSEDYTTGELFFNALACEIDSV---KGIKDRGLVSTVTDGFA 389
IESIGE-LLCEFRFSDYDYSIMSKIFLDR-FRSEYVFPITAEHLKYVSVYLLCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 400
EVMSEG-LISDSKIFSDYDYSIMSKIFLDR-FRSEYVFPITAEHLKYVSVYLLCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 402
EVPVNG-YFTDEIRFRFDFNLDLQCFIKSR-RY-FCRDFVPLVSVYVYLLCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 387
QISKGE-KITDSVRFADFNTDYSIMSKIFLDR-FRSEYVFPITAEHLKYVSVYLLCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 388
EITKGR-FVTNDPRHFEFEFCIDLFLKQRR-WRRNEYITFKKTLTKVYTYLQCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 384
EITKGR-FVTNDPRHFEFEFCIDLFLKQRR-WRRNEYITFKKTLTKVYTYLQCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 384
INRNDL-LLCEFRFSDYDYSIMSKIFLDR-FRSEYVFPITAEHLKYVSVYLLCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 394
INRNDL-LLCEFRFSDYDYSIMSKIFLDR-FRSEYVFPITAEHLKYVSVYLLCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 394
AITRGE-SMGNPRHFEFEFCIDLFLKQRR-WRRNEYITFKKTLTKVYTYLQCLKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 393
SITAGE-AVRAPTNSFNDFATCRGLPNLAFSLGPIAVPYPVRYVRYLRLQKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 394
SPVNG-AIQGMKRFNGFEAVAMKLNPLR-RKVESCLPNTKILVRYLRLQKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 395
SITKGD-LPGLERRFNSFEATGVKHLGSLG-YSVRDCTPVSSLEISKLVRYLRLQKPPDLSEGLAKLRLQSQEDYHQELCFLLNLAGRIESV---KGIKDRGLVSTVTDGFA 394

AgVT PVT:YP 002019748.1 PVT:ADX41471.1 PVT:AFU5321.1 PVT:AKX90539.1 PrVt:YP 009051684.1 PrVt:AHM92766.1 ZQvT:QBS17025.1 ZQvT:QBS17031.1 ChVt:QNG41875.1 FiTA:QED42804.1 TrTA:QED42832.1 GVA:NP 619662.1 AcVb:YP 004935358.1 ASGV:NP 044335.1 YaVa:YP 009268859.1 RAva:YP 009553496.1 ACLSv:NP 040551.1 PGVv:YP 004732978.2 CtChv-1:YP 009103999.1 CtChv-2:YP 009103996.1 DiVa:YP 006905850.1 OBrV1:YP 009408144.1 ClBv:NP 624333.1 ClBv:AF443536.1 AVcAv:YP 008997790.1 CrVv:YP 009505632.1 WVa:QEA69426.1 CrMv:YP 007761581.1 CtLAv:YP 009046478.1 PhLbV:YP 001552317.1 GCLV:YP 004936159.1 ASPV:NP 604464.1 APV1:YP 009094347.1

SFLPARVAATLN-AKMAN-SRKMGLLHFLDEPSVYVNTY---KFKQGCVRDY---SFTHFIFDNDVYAEF-----DVAQDFEERF----- 445
DIF-LLDGLFNMSDRRKS-EKFVEFMRALDYQTNKVVCT---TFSGGMRS-----GFLAEFFLDNDCEASEG-----LDEIVSRFDTFF----- 444
DIF-LLDGLFNMSDRRKS-EKFVEFMRALDYQTNKVVCT---TFSGGMRS-----GFLAEFFLDNDCEASEG-----LDEIVSRFDTFF----- 444
DIF-LLDGLFNMSDRRKS-EKFVEFMRALDYQTNKVVCT---TFSGGMRS-----GFLAEFFLDNDCEASEG-----LDEIVSRFDTFF----- 444
DIF-LLDGLFNMSDRRKS-EKFVEFMRALDYQTNKVVCT---TFSGGMRS-----GFLAEFFLDNDCEASEG-----LDEIVSRFDTFF----- 444
QTFPDRIRARMI-SKTTIDE-NSFLRLISSMEGMLKLVRE-SFP-----VENLFEFDVKDGIS-----EQFILDLMKXMSGGT----- 443
KAFPNRLARMI-SNTIDE-NSFLQLISSMEGMLKLVRE-TFP-----VENLFEFDVKDGIS-----EQFILDLMKXMSGGT----- 443
RSPFKRVRNH-EKCYKE-SDFEQMLNSNGFQVQFKR-EYPMNL-----VETLESFLSDELESES-----EDDASIEDKGRKEQL----- 452
RSPFKRVRNH-EKCYKE-SDFEQMLNSNGFQVQFKR-EYPMNL-----VETLESFLSDELESES-----EDDASIEDKGRKEQL----- 452
ELFPERVDNLT-SCSYAS-RKLEKFLAEAPVYKVRHT-----SNLEQEELNATVFDTO-----RFVITKLDMEF-GSMFEF----- 442
ELFPERVDNLT-SCSYAS-RKLEKFLAEAPVYKVRHT-----SNLEQEELNATVFDTO-----RFVITKLDMEF-GSMFEF----- 442
SIGDRTRLNL-SSYVND-LSVENLITDSKPYRVEFKR---VRP-----NWNDSWLDVDDQEGLE-----EENRNLNLLDKFSDGSSFL----- 452
DIAGSLTRFLRN-RPEYDA-RCLEKFRIRSCETIHEHRR-YMEGIRRGA-----SFKVQVNMVNDVDESAN-----FRFLNGFSCSDPLPCDALLYK----- 445
GALGDSITLYAL-KEQYK-SMEKFLIRLCSARITITRT-FRDYQRR-----KLGPELGTSTWTEDEE-----LDKVNIE----- 446
QVMPHFIAASFF-EPTFEH-LNMRKLNLDLTKGIEVPLSVII-----LDQVRWIE-----LDQVRWIE----- 445
KALPDMQLRFL---PMKRHREVFQKFLVLSPTISFSDMLR---A-----VDSNSFEGIS-----KLFCAADGPFDLSPIGLVEGVR----- 437
DKLP-YREVL-EKIGLA-DDFTRRLMKIKPLAFDTHTT---DRP-----LTVRMIDQIWEERLS-----SFDNINSIVFYGRREWLNGLV----- 451
DLCP-FKKNF---ERYHLIDDYRMLSKPLSFCNS-DD-----LDVYAFEPED-----LDFIDDFW----- 429
SILPEFVRML---GSFGR-DNLLKOISEAKFAVERITCT---S-----INRNVNDNAL-----WRVYKDLAFHHDVNELOSKDLFL----- 459
SILPEFVRML---GSFGR-DNIMKOIAMAEPFVSITET---R-----VKNVNDNSV-----FRFLNGFSCSDPLPCDALLYK----- 460
DCLPDLHSAAMF-RSHFKA-QNLDLNMNKTLSIVTTE---DFYPSK-----VDCIKETKEY----- 444
DLPPSITHALF-RKDFKA-RENLMLDLKPLTILPTA---DYHMSI-----IHRSVQIDQI----- 444
SKMNPVARTIF---PQMKK-KNTFEFLSLGTLVVDER---VCFEHLVEWGFVVITDENAYLDPLSIFAIN-----ENFNEERVDV---GYL----- 481
SKMNPVARTIF---PQMKK-KNTFEFLSLGTLVVDER---VCFEHLVEWGFVVITDENAYLDPLSIFAIN-----ENFNEERVDV---GYL----- 483
DFAPDMLLNGF---MTWKS-GNFIKKLEGEAHEA-QVEG---A-----YPEGMINFL----- 433
MALPNLSACFF---DRMHW-LNIFHFLTLTDTRVVKVCT---V-----VDRYRQTMFDDLEVKDDV---MSTKLPPEAIDTLF6GGQVVK-----G----- 465
KLFPMITQRLS---SDFKD-KNFFEVLFNENLRLVKTIR---S-----YDSRGIISMA-----EHIKIDFLEF----- 441
KLFPMITQRLS---SDFKD-KNFFEVLFNENLRLVKTIR---S-----YDSRGIISMA-----EHIKIDFLEF----- 441
NLLPRYMRFF---NSFKG-YSLGKFEIETEPFSFTRLCS---TYSRFG-----FKTSFIEEEE---AAVADKDLPLCLLKANMKPSCF----- 469
EVLPRYMRFF---KSFKG-YSLGKFEIETEPFSFTRLCS---TYSRFG-----FKTSFIEEEE---ARERDGDPYTKQSQSFSN----- 467
RKLKPLAQRFF---KTVAE-VSLDDFISIMPEPFSFQELC---D-----VDMNHYFHTFB---FFFEETADFFEMAALMD----- 458
TFPYPKYAARF---DVAIR-RLSDEFVEHRCPTYFKTAKV---V-----HDMRHLQLLF---PRDEMPRDEEDFDTRALERMQGGGAN----- 468
KLFPMITQRLS---KVVQO-LHLDNFIEITLLEEFNFSINT---SLS-----LNKMDLLEFV-NLTFGDTD---FNVEDSFAEAWGTKKDVV----- 470
SFLPFKMARFF---DCYKA-CSLDKFIHDLPEPFSFVETK---V-----INSRNPMPFE-AVTLGETMSSEEVILKILAKLDM----- 463

AgVT PVT:YP 002019748.1 PVT:ADX41471.1 PVT:AFU5321.1 PVT:AKX90539.1 PrVt:YP 009051684.1 PrVt:AHM92766.1 ZQvT:QBS17025.1 ZQvT:QBS17031.1 ChVt:QNG41875.1 FiTA:QED42804.1 TrTA:QED42832.1 GVA:NP 619662.1 AcVb:YP 004935358.1 ASGV:NP 044335.1 YaVa:YP 009268859.1 RAva:YP 009553496.1 ACLSv:NP 040551.1 PGVv:YP 004732978.2 CtChv-1:YP 009103999.1 CtChv-2:YP 009103996.1 DiVa:YP 006905850.1 OBrV1:YP 009408144.1 ClBv:NP 624333.1 ClBv:AF443536.1 AVcAv:YP 008997790.1 CrVv:YP 009505632.1 WVa:QEA69426.1 CrMv:YP 007761581.1 CtLAv:YP 009046478.1 PhLbV:YP 001552317.1 GCLV:YP 004936159.1 ASPV:NP 604464.1 APV1:YP 009094347.1

-----FFGQTSKQRMK-----KEKIDGYSYPTENPYAAS-----ERFNRFCKSPVKI-----ALKAKEGAK-----LNVLSNPTF----- 507
-----DPKKEYSAHALRVNI-----KDRTPSPYGMKARTPS-----EFSKAKK-----VNLHREYITKV-----VNLKERTGT----- 505
-----DPKKEYSAHALRVNI-----KDRTPSPYGMKARTPS-----EFSKAKK-----VNLHREYITKV-----VNLKERTGT----- 505
-----DPKKEYSAHALRVNI-----KDRTPSPYGMKARTPS-----EFSKAKK-----VNLHREYITKV-----VNLKERTGT----- 505
-----DPKKEYSAHALRVNI-----KDRTPSPYGMKARTPS-----EFSKAKK-----VNLHREYITKV-----VNLKERTGT----- 505
-----DAKMDTSYVATS-----DAKMDTSYVATS-----RNKNFVLLDFKQKQKARLAEKEEPEQRIT----- 488
-----DPQDPTSYSVKVS-----DPQDPTSYSVKVS-----RNKNVILDFKAKKARLADKESKEITRVDI----- 488
-----RIDNRRREOYLNMKEALFNKN-----ELPNKR---TRDDMIK---KKEDEYITIKNN---RRIELDRRRCSS----- 512
-----RIDSREGOYLNMKEALFNKN-----ELPNKR---TRDIRKE---RIKKEENIYKNN---RRIELDRRRCSS----- 512
-----SGEADTYSVSKSTLVC-----TOAD---LNSIQGLV---SYASTQRSNTRNLSSD----- 486
-----PGEADTYSVSKSTLVC-----TOAD---LNSIQGLV---SYASTQRSNTRNLSSD----- 486
-----KDRSSOAYYKTKLDR-----MTSE---INEVFINE---KYKVIERRIKFEKFD----- 497
-----ALSEVNFV---DISWDRVSEPYGIE-----CITHGEGSR-----TRVPLSRILRAH----- 488
-----LQEIYHV---VVEKEDRERTPSYK-----GDE-----ILDNKERRSD-----LMLQKARM----- 489
-----TRFHARMFIAEAIAVGN-----FNYNKRIFIAEAIAVGN-----TVVRNFDK-----DYRSDPNKR-----LDDLKGRF----- 460
-----PKVKKKGLAKLI-----PGREVDSHNPRIEYSDL---LSSTIWRSYDD-----CFHRHSASPL-----VILRSNRA----- 502
-----SEKKMTRDPELKMIE-----MGREDSHNPRIEYSDL---LSSTIWRSYDD-----CFHRHSASPL-----VILRSNRA----- 502
-----KEREKGEVKS-----RDRRNVGVGSPVTSYQQ---AAVCIFSMDEN-----CTGMRNRGKV-----VILRSNRA----- 490
-----FLNSCDYLRNDIKIIRS-----SDRANLVHKFTRPFSS-----KSSDKMSKDFYR---KIYDERTDITKGS---FILKSNRS----- 531
-----YLNDEQMEFKVQEDLMNS-----KSRGAFIFGSRNRYHIE-----KSSDKMSKDFYR---KIYDERTDITKGS---FILKSNRS----- 488
-----ERIRLFPWNLNDYDL-----KRKRVAIYANILSVYFEERKI-----ESAQKGNPKMLQIEWYG----- 534
-----DRVKLFPWNLNDYDL-----KRKRVAIYANILSVYFEERKI-----DLRERGHKMLQIEWYG----- 536
-----GDRRTVVIKSLDGLIK-----FSRSKNLYKMHITFLLN-----PSLIRGN---IRNFCSNG----- 513
-----FSWIPKNYDALIK-----KSSSEHV-----PSLLRCA----- 468
-----FSWIPKDYDALVK-----RSESEHV-----PSLLRCA----- 468
-----ESYPDLIFNAHTLVFASHSPSITLMLVKSFINMMV---GKTNDRYQSLVALRQ---ALNOKGAKL-----FMLHNEYS----- 537
-----TAYPDCLFHASQVSNFPHNIRLRLVTLFISTWV---GKTEADYITSLLSLK---SLSQKGLK---FRHLHDYR----- 535
-----LEQSQYSLGVNKGVLKPLCCTECLIKLGFATYAIKLDIDVYAKIDVSCWAVDVHEHRLNNTNVCVIDGAFKLLLEPNHW--- 552
-----GERTYGAIVTGP-----FC---DYPVDAIQVFLRLVRSREMSFSPMVQYECERDYTFMQASKL----- 528
-----NITWHSPVLSKFSFYHQHSTLSVKSISALTRIAKIVLSYDPCV---VEAFSESRVTLNANV---VITANLRAC----- 544
-----GAPTSLEG---LTKGKYSSEAYHGLHVLDCGGPHVFRRLVRKIADSF---NEFGIYRKYSSVSSVAITV-----LQRKTSNLSF----- 540

AgVT	-----	-----	617
PVT:YP_002019748.1	-----	-----	618
PVT:ADX41471.1	-----	-----	618
PVT:AFU55321.1	-----	-----	619
PVT:AXK90539.1	-----	-----	619
PrVT:YP_009051684.1	-----	KRS	786
PrVT:AHM92766.1	-----	KRT	786
ZoVT:QBS17025.1	-----	KVR	826
ZoVT:QBS17031.1	-----	KVR	826
ChVT:QNG41875.1	-----	RSD	796
FiTA:QED42804.1	-----	RSD	796
TrTA:QED42832.1	-----	RSD	808
GVA:NP_619662.1	-----	EED	755
AcVB:YP_004935358.1	-----	EET	750
ASGV:NP_044335.1	-----	-----	600
YaVA:YP_009268859.1	-----	-----	600
RAVA:YP_009553496.1	-----	-----	844
ACLSV:NP_040551.1	-----	QSEGLNE	847
GPV:YP_004732978.2	-----	TDSMSGV	832
CtChv-1:YP_009103999.1	-----	-----	854
CtChv-2:YP_009103996.1	-----	-----	850
DiVA:YP_006905850.1	-----	-----	665
ObrV1:YP_009408144.1	-----	-----	665
CLBV:NP_624333.1	-----	-----	974
CLBV:AF443536.1	-----	-----	1000
AVCaV:YP_008997790.1	-----	-----	649
CPrV:YP_009505632.1	-----	-----	1001
WVA:YP_009357235.1	-----	KED	778
WVA:QEA69426.1	-----	KED	778
CRMaV:YP_007761581.1	-----	SDDGGGFYFEEINK	906
CTLav:YP_009046478.1	-----	SGNEEDDSYEE MNK	916
PhIbV:YP_001552317.1	-----	QEPESPDPDEENELDKELSEDEAGGSEEKE	922
GCLV:YP_004936159.1	-----	-----	864
ASPV:NP_604464.1	-----	-----	993
APV1:YP_009094347.1	-----	-----	923

AgVT	-----	-----	617
PVT:YP_002019748.1	-----	-----	618
PVT:ADX41471.1	-----	-----	618
PVT:AFU55321.1	-----	-----	619
PVT:AXK90539.1	-----	-----	619
PrVT:YP_009051684.1	-----	-----	786
PrVT:AHM92766.1	-----	-----	786
ZoVT:QBS17025.1	-----	-----	826
ZoVT:QBS17031.1	-----	-----	826
ChVT:QNG41875.1	-----	-----	796
FiTA:QED42804.1	-----	-----	796
TrTA:QED42832.1	-----	-----	808
GVA:NP_619662.1	-----	-----	755
AcVB:YP_004935358.1	-----	-----	750
ASGV:NP_044335.1	-----	-----	600
YaVA:YP_009268859.1	-----	-----	600
RAVA:YP_009553496.1	-----	-----	844
ACLSV:NP_040551.1	-----	-----	847
GPV:YP_004732978.2	-----	-----	832
CtChv-1:YP_009103999.1	-----	-----	854
CtChv-2:YP_009103996.1	-----	-----	850
DiVA:YP_006905850.1	-----	-----	665
ObrV1:YP_009408144.1	-----	-----	665
CLBV:NP_624333.1	-----	-----	974
CLBV:AF443536.1	-----	-----	1000
AVCaV:YP_008997790.1	-----	-----	649
CPrV:YP_009505632.1	-----	-----	1001
WVA:YP_009357235.1	-----	-----	778
WVA:QEA69426.1	-----	-----	778
CRMaV:YP_007761581.1	-----	-----	937
CTLav:YP_009046478.1	-----	-----	947
PhIbV:YP_001552317.1	-----	-----	953
GCLV:YP_004936159.1	-----	-----	900
ASPV:NP_604464.1	-----	-----	1108
APV1:YP_009094347.1	-----	-----	957

AgVT	-----	-----	637
PVT:YP_002019748.1	-----	-----	638
PVT:ADX41471.1	-----	-----	638
PVT:AFU55321.1	-----	-----	639
PVT:AXK90539.1	-----	-----	639
PrVT:YP_009051684.1	-----	-----	812
PrVT:AHM92766.1	-----	-----	812
ZoVT:QBS17025.1	-----	-----	852
ZoVT:QBS17031.1	-----	-----	852
ChVT:QNG41875.1	-----	-----	823
FiTA:QED42804.1	-----	-----	823
TrTA:QED42832.1	-----	-----	834
GVA:NP_619662.1	-----	-----	781
AcVB:YP_004935358.1	-----	-----	776
ASGV:NP_044335.1	-----	-----	613
YaVA:YP_009268859.1	-----	-----	613
RAVA:YP_009553496.1	-----	-----	866
ACLSV:NP_040551.1	-----	-----	860
GPV:YP_004732978.2	-----	-----	844
CtChv-1:YP_009103999.1	-----	-----	874
CtChv-2:YP_009103996.1	-----	-----	870
DiVA:YP_006905850.1	-----	-----	686
ObrV1:YP_009408144.1	-----	-----	686
CLBV:NP_624333.1	-----	-----	989
CLBV:AF443536.1	-----	-----	1015
AVCaV:YP_008997790.1	-----	-----	667
CPrV:YP_009505632.1	-----	-----	1016
WVA:YP_009357235.1	-----	-----	815
WVA:QEA69426.1	-----	-----	815
CRMaV:YP_007761581.1	-----	-----	1050
CTLav:YP_009046478.1	-----	-----	1060
PhIbV:YP_001552317.1	-----	-----	1055
GCLV:YP_004936159.1	-----	-----	1002
ASPV:NP_604464.1	-----	-----	1209
APV1:YP_009094347.1	-----	-----	1062

AgVT	-----IKVKTWELGLKS-----ITK-----AQVVIDEISLYPPGYLDLCLALKKKE-----AQVVLGDPLQTRYHKSDDALTL-KGQADV	914
PVT:YP_002019748.1	-----VKVMTFEAGLRR-----VOK-----SSLVIDELSLMNPNGVLDMLINM-----NEEATFITLFDPLQARVHAKSDVLRV-SPENDV	912
PVT:ADX41471.1	-----VKVMTFEAGLRR-----VOK-----SSLVIDELSLMNPNGVLDMLINM-----NEEATFITLFDPLQARVHAKSDVLRV-SPENDV	912
PVT:AFU55321.1	-----VKVMTFEAGLRR-----VOK-----SSLVIDELSLMNPNGVLDMLINM-----NEEATFITLFDPLQARVHAKSDVLRV-SPENDV	913
PVT:AXK90539.1	-----VKVMTFEAGLRR-----VOK-----SSLVIDELSLMNPNGVLDMLINM-----NREATFITLFDPLQARVHAKSDVLRV-SPENDV	913
PrVt:YP_009051684.1	-----IKLRTFELAVSAIT-RMVKKGDDGKLTVIDEATLLPGGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1088
PrVt:AHM92766.1	-----IKLRTFELAVSAIT-RMVKKGDDGKLTVIDEATLLPGGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1088
ZoVt:QBS17025.1	-----IKLRTVEGITEINKRRDYTND-KPTIIDEVSLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1128
ZoVt:QBS17031.1	-----IKLRTVEGITEINKRRDYTND-KPTIIDEVSLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1128
ChVt:QMG41875.1	-----VKLRTFEGISALS-RLAIRGE-PLNIVVDEVTLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1097
FiTA:QED42804.1	-----VKLRTFEGISALS-RLAIRGE-PLNIVVDEVTLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1097
FiTA:QED42832.1	-----IKLRTFELAVSAIT-RMVKKGDDGKLTVIDEATLLPGGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1128
GVA:NP_619662.1	-----HRVTFEVAFAFD-DYG-----KTFIDVIDEIGLLPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1040
AcVb:YP_004935358.1	-----AIVRTFESAFNG-----RKG-----YENIIDEVGILLPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1035
ASGV:NP_044335.1	-----CQVHTMETALK-----IDGT-----FMEVFDVEIGLLPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	917
YaVa:YP_009268859.1	-----CHCETLEVALGK-----VDLS-----YTEIVIDEIGLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	917
RAVA:YP_009553496.1	-----VRVNTYERCKL-----SFAF-----YDVIIVDEITLPPSGFEDLLTKWASA-----KFEAFEMKPLPRFNCLGDPLQCRYCETDNALL-DKTHDI	1149
ACLSV:NP_040551.1	-----DKDITFESALKS-----DWKG-----KRLFTLDEISLLPKGFTDMLMKMME-----GANEKMMIFVGDPLQAGYSPKDHRL--VARD	1180
PGPV:YP_004732978.2	-----LKTMTFESAIK-----CLAG-----KWIILDEVTLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1158
CtChv-1:YP_009103999.1	-----VTFKTYESALKE-----NLAK-----FSLIIDEFPLTRGYTDVIAKSKYVD-----NLTCREKVKVTLIGDPLQASHYSEDDLL-AQGGEL	1193
CtChv-2:YP_009103996.1	-----LTLTKYESALKE-----NFAE-----FSMIIDEFLLVPRGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1190
DiVa:YP_006905850.1	-----IRFSTYKALTL-----SYEE-----PLNIVVDEVTLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1002
ObRv1:YP_009408144.1	-----IRFLTFRALTV-----TYQE-----SELVIDEIGLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1002
CLBV:NP_624333.1	-----RKVSTFEKFIKT-----DKSK-----LDLVIDELTLFPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1285
CLBV:AF443536.1	-----RKVATFESFIKM-----DKSK-----LDLVIDELTLFPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1310
AVCav:YP_008997790.1	-----CSVTFEVLFAK-----SISK-----LDMIVDELTLFPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	993
CPvV:YP_009505632.1	-----HKVCTFEVALQ-----NLRS-----VELVIDEIGLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1313
WVA:YP_009357235.1	-----WVKTTFESALKM-----NLGK-----SDIIVDELTLFPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1143
WVA:QEA69426.1	-----WVKTTFESALKM-----NLGK-----SDIIVDELTLFPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1143
CRMaV:YP_007761581.1	-----TDVTFEVALKKN-----GLLK-----KRFILDEISLLPKGFTDMLMKMME-----GANEKMMIFVGDPLQAGYSPKDHRL--VARD	1345
CTLav:YP_009046478.1	-----TEVATFEVALKKT-----GMLK-----KRFILDEISLLPKGFTDMLMKMME-----GANEKMMIFVGDPLQAGYSPKDHRL--VARD	1355
PhLbV:YP_001552317.1	-----SQAERMKRKNMLVCTFEIILKKH-----LWPK-----GMALIDEITFQLLPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1375
GCLV:YP_004936159.1	-----VTILTFEFLHQM-----NVKE-----GSTIIFDEIQLVPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1285
ASPV:NP_604464.1	-----LKNVRKTFELFILLHD-----SIKE-----HTIIVDEIGLLMPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1506
APV1:YP_009094347.1	-----VRVLTLESFIKAV-----FTFK-----AASVLDVEQLVPPGYLDLNSLVPFG-----STIILLFDPLQASHYHKSMDVDR-VNLPGPV	1351

AgVT	DRF---KIERYLRLSHRLSSE---LSYM-----FEFFCLSEKLEHHLH-GKIVRQEAL---SVLDKGSQVQLVASQNMKRYK-----SHRGVPTKFEVQGLTFNF-C	1003
PVT:YP_002019748.1	DRI---KVPKYLFFSKRMSSELD---FDVRCSSDQKQWELH-GKQVREPAL---FRDIKQGEFTLSPSFETAREMSKYAD-IDGCKSMTFGESQGLTNV-K-A	1006
PVT:ADX41471.1	DRI---KVPKYLFFSKRMSSELD---FDVRCSSDQKQWELH-GKQVREPAL---FRDIKQGEFTLSPSFETAREMSKYAD-IDGCKSMTFGESQGLTNV-K-A	1006
PVT:AFU55321.1	DRI---KVPKYLFFSKRMSSELD---FDVRCSSDQKQWELH-GKQVREPAL---FRDIKQGEFTLSPSFETAREMSKYAD-IDGCKSMTFGESQGLTNV-K-A	1007
PVT:AXK90539.1	DRI---KVPKYLFFSKRMSSELD---FDVRCSSDQKQWELH-GKQVREPAL---FRDIKQGEFTLSPSFETAREMSKYAD-ERNCKSMTFGESQGLTNV-K-A	1007
PrVt:YP_009051684.1	LTPIFG-QEFRYRGSYRFPFLFDLE---DFF-YGHGDEIDPNH-MRIFAQOQA---VREAIKRPVFLCPSEDDKREEL---SNFGEAYTFTGSQGLTFDF-V	1176
PrVt:AHM92766.1	LTPIFG-QEFRYRGSYRFPFLFDLE---DFF-YGHGDEIDPNH-MRIFAQOQA---VREAIKRPVFLCPSEDDKREEL---SNFGEAYTFTGSQGLTFDF-V	1176
ZoVt:QBS17025.1	FEPLYG-QSFDYKYSYRFGDLPFKIE---GLSMGMEISEHH-MKIFIKOPEA---VKKVSDPTIFSPSEAKANEL---RRVDSYTFGTSGQLTFDF-V	1217
ZoVt:QBS17031.1	FEPLYG-QSFDYKYSYRFGDLPFKIE---GLSMGMEISEHH-MKIFIKOPEA---VKKVSDPTIFSPSEAKANEL---RRVDSYTFGTSGQLTFDF-V	1217
ChVt:QMG41875.1	MDPLNN-CEFPYLFTRTHRFPKLFDFE---GLQF-LGKQELSEFH-MHEFGSPEA---VLKKIEKPTFLCPSEAKRAEL---SHYGDATYFTGSQGLTFDF-V	1186
FiTA:QED42804.1	MDPLNN-CEFPYLFTRTHRFPKLFDFE---GLQF-LGKQELSEFH-MHEFGSPEA---VLKKIEKPTFLCPSEAKRAEL---SHYGDATYFTGSQGLTFDF-V	1186
FiTA:QED42832.1	FEPLYG-TSPKYLQYTRFKNFQNF---NLVMSMAALDENH-MKFFVLOPEA---VRAAIRRPTFLCPSEAKREEL---KRYGDATYFTGSQGLTFDF-V	1198
GVA:NP_619662.1	FNRRVG---LFLVLYCSHRLPRNCKL---FEIECEGSEKRVV---YRSNRLKDEPTICATRAMKE---EKESGDMYVTFGTAGLTFDF-S-C	1118
AcVb:YP_004935358.1	FDRLMG---KKNVLYKTHRLSPNQKL---FVMPKGERSENVLYK---GAEGDANYDLITASRAEKKE---KROKQSGITGESQGLTVRR-V	1113
ASGV:NP_044335.1	DLMIKT-IKHKYLFGQYRFGQV---FOEL---FNDPPTREDESKFS---RKFFADIS---SVKTEYDGLILVAKREKGVF---AGRPVATVYTESQGMITSKRV	1005
YaVa:YP_009268859.1	DFTRKSKFNKYLFGQYRFGQV---FSEI---VNIPTRDDGSHS---RKFFPDMS---KVDISYKAVLVGPREAKISL---ASGLPVYTMVESQGLTFNFRV	1006
RAVA:YP_009553496.1	GRLIA---KEQPYLSTRNNGE---IERI---FDVESREDEGHA---NIVWSYG---DFKSVMDKIDVVLVASELEKFFY---SNQCKITYGESQGLTFDF-V	1243
ACLSV:NP_040551.1	KRLFKG---GVNYKYYSYRINPK---IAAK---LAIETMNFIDGIDEQ---SSIVDMPSSAHL---FMKKGNIHVEILVASWKEKLY---SNVGNMTFGESQGLTFDF-V	1274
PGPV:YP_004732978.2	DSVF---KDVLYQYQSYRIPAN---VAGR---FDVETREDEGIDCH---GTFYSDLSAKH---HAKRCKNDIDVVLVASELEKFFY---SNQCKITYGESQGLTFDF-V	1251
CtChv-1:YP_009103999.1	SSLET---DYPVLYLSHRLPKG---MKMS---MDLNMPLGSPFEGEKT---MKLYNSAAA---AFSEKAFDVLVAGROKTF---SNFTVMTFGESQGLTFDF-V	1280
CtChv-2:YP_009103996.1	ETLET---RKPPLYYSYRHLSSS---LGGI---LDLNMPLGSPFEGEKT---MKLYNSAAA---AFSEKAFDVLVAGROKTF---SNFTVMTFGESQGLTFDF-V	1277
DiVa:YP_006905850.1	DFTMEN---EETLVLNYSHRLNKMHFYK---PGVMELEDENSI---SRKSNMVFV---AKKTIPEAQVLVASELEKFFY---KELDAKTFGESQGLTFDF-V	1089
ObRv1:YP_009408144.1	WVLMEN---EETIIVLNSHRMSRHHYK---PGVMELEDENSI---SRKSNMVFV---AKKTIPEAQVLVASELEKFFY---KELDAKTFGESQGLTFDF-V	1089
CLBV:NP_624333.1	DRILIG---QNTIYVYTHRMSRY---FNRF---FDVPCFNQADRTEQORLWIFDDVYSIYS---ICSDRQPCDVLVLESDELKAAF---SPINVMTFGESQGLTFDF-V	1381
CLBV:AF443536.1	DRILIG---QNTIYVYTHRMSRY---FNRF---FDVPCFNQADRTEQORLWIFDDVYSIYS---ICSDRQPCDVLVLESDELKAAF---SPINVMTFGESQGLTFDF-V	1406
AVCav:YP_008997790.1	DLILDG---SEVDYMQSKRFSEELFNLFDLKKNEVDAESRETKGAKFR---PRMYTLMKVV---EENQGNPDIIVLVGSDFAEGLF---ASIXTKMTFGESQGLTFDF-V	1097
CPvV:YP_009505632.1	DRITSG---AKINYLVSYRHLSSS---LGGI---VDIIEELSGSYVYVE---LDFDNHNMVA---EAKKRGFPIDLILVASELEKFFY---AGKVNVLTFGEAQGLTVNH-S	1409
WVA:YP_009357235.1	DRIFKNGAINTYLAFLSHRLGTV---FNCV---FEGIEELGSEEMEGGS---INVFKSNFAAIA---WSEKQEFDDLILVDSREKKAAY---SGLINVLTFGEAQGLTVNH-S	1240
WVA:QEA69426.1	DRIFKNGAINTYLAFLSHRLGTV---FNCV---FEGIEELGSEEMEGGS---INVFKSNFAAIA---WSEKQEFDDLILVDSREKKAAY---SGLINVLTFGEAQGLTVNH-S	1240
CRMaV:YP_007761581.1	DRILLEG---KVVYLYSESKFRFNPVFG---RLPCTFDSSRLTLE---KEEYAVDFSFKAFKADYLSPIKTLVLSVSTFTEKTVKANAN---GRNVSLFTFGESTGMNFY-V	1444
CTLav:YP_009046478.1	DVLLDN---KVVYLYSESKFRFNPVFG---RLPCTFDSSRLTLE---KEEYAVDFSFKAFKADYLSPIKTLVLSVSTFTEKTVKANAN---GRNVSLFTFGESTGMNFY-V	1454
PhLbV:YP_001552317.1	DSVLEG---AEKYVYKSRFKNMGVFG---RLPCEFTQMGQEA---TEETHLLYSGLH---LHVPTQESKVLVLSVSEKKEKVAHEAF---GNSPVTFTGESTGLNFYK-G	1474
GCLV:YP_004936159.1	QRLTQIG---IQFYTRYSRFPNMFVFG---RLPCAINTNEDD---FEDFLELEGI---EQVQEDQEVKVLVLSVSEKKEKVAHEAF---GNSPVTFTGESTGLNFYK-G	1379
ASPV:NP_604464.1	MRLTSG---RSYKFNLLSQRFPNMFVFG---RLPCNLNKRTRTLD---EEYTLWDSIQE---FSMMGRKDCPVLSVSEKKEKVAHEAF---GNSPVTFTGESTGLNFYK-G	1604
APV1:YP_009094347.1	FEVLSG---KYYKFNVSSRRFQSEMFGV---RLPCRMDTKAMTE---NENFHLESIES---AAEVSNTYDVLVLSVSEKKEKVAHEAF---GRDLEVLFTGESTGLNFYK-G	1448

AgVT	VVVICDQHLVSNFHMIVALTRSGRGCFLVDSVNDKEAVRINLQRLIEKAMSKKVTNTFLRAMAGVSLTAEFTED---VETFKTTESVEEKLGGDPLKGLVLPVLEPEAF	1115
PVT:YP_002019748.1	VVVICDQHLVSNFHMIVALTRSGRGCFLVDSVNDKEAVRINLQRLIEKAMSKKVTNTFLRAMAGVSLTAEFTED---VETFKTTESVEEKLGGDPLKGLVLPVLEPEAF	1118
PVT:ADX41471.1	VVVICDQHLVSNFHMIVALTRSGRGCFLVDSVNDKEAVRINLQRLIEKAMSKKVTNTFLRAMAGVSLTAEFTED---VETFKTTESVEEKLGGDPLKGLVLPVLEPEAF	1118
PVT:AFU55321.1	VVVICDQHLVSNFHMIVALTRSGRGCFLVDSVNDKEAVRINLQRLIEKAMSKKVTNTFLRAMAGVSLTAEFTED---VETFKTTESVEEKLGGDPLKGLVLPVLEPEAF	1119
PVT:AXK90539.1	VVVICDQHLVSNFHMIVALTRSGRGCFLVDSVNDKEAVRINLQRLIEKAMSKKVTNTFLRAMAGVSLTAEFTED---VETFKTTESVEEKLGGDPLKGLVLPVLEPEAF	1119
PrVt:YP_009051684.1	CISIDMDGVSDFHFMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1296
PrVt:AHM92766.1	CISIDMDGVSDFHFMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1296
ZoVt:QBS17025.1	VISIDMDGVLVSNHMMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1296
ZoVt:QBS17031.1	VISIDMDGVLVSNHMMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1296
ChVt:QMG41875.1	VISIDMDGVLVSNHMMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1296
FiTA:QED42804.1	VISIDMDGVLVSNHMMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1296
FiTA:QED42832.1	VISIDMDGVLVSNHMMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1296
GVA:NP_619662.1	LIYDEHMAKKEDEDMVALTRSRGEGIGHVIT-PALKKLLTNAKSTLLKVVKGTETRRSEIVAMVRKHHPETDVLTEK---ESRLAETVDEYEARLADGDPYLSLALYDETFE	1229
AcVb:YP_004935358.1	KLVIDRDHGLNDRKAMVALTRARNTLSVEVD-KSMKHLVKAHKSSTLMLFKVIRKAGNLLIGKVLVAVKVSDFMRASGGKIMDHANLIGD---SKKGRTRREEFDLENDPWIKTQLVLEPELQ	1329
ASGV:NP_044335.1	LICLDOQLFAGGANAIVAITRSKGVDFDILK-GNSLKEVQRAQKTIWQIFIEGKSPMERIVNL---NPGASFYESPL---DVGNSSIQDQKASNDLFTMPFLIAEEFVDP	1111
YaVa:YP_009268859.1	LICLDOQLFAGGANAIVAITRSKGVDFDILK-GNSLKEVQRAQKTIWQIFIEGKSPMERIVNL---NPGASFYESPL---DVGNSSIQDQKASNDLFTMPFLIAEEFVDP	1112
RAVA:YP_009553496.1	VVMVSDTKKVSFGHLLVALTRSVRPLINTRRTFDL---VMRGNPLDATRERRITKVSVM---VKDMGFKVDF---EKIGRLE-FAFEKLGHEHLLCINDNESALPE	1344
ACLSV:NP_040551.1	VIVLSEAKLCSDAHMTVAITFRFRGFCFALGSKGSKEDYMRSMSKGLQRTCSGVGSKFEFTLGSVSSV---NLTLSEKDIAGGIDEMDREARLGGDPLKGLVLPVLEPEAF	1386
PGPV:YP_004732978.2	LISLSEEDRLCSDNHNYVALTRFKKGGFFQFNRGDLTFKYSKNSLQSKLGRYINLNDLKPFFMQMLDI---NLDMDDD---RNQVAGATEMENKSGDPLKGLVLPVLEPEAF	1360
CtChv-1:YP_009103999.1	CIALTEDSLASDNHMMIVALTRAKETINIF3KGGVLFNFYKVKAGNLLIGKVLVAVKVSDFMRASGGKIMDHANLIGD---SKKGRTRREEFDLENDPWIKTQLVLEPELQ	1387
CtChv-2:YP_009103996.1	AIALSEEDRLCSDNHNYVALTRAKETINIF3KGGVLFNFYKVKAGNLLIGKVLVAVKVSDFMRASGGKIMDHANLIGD---SKKGRTRREEFDLENDPWIKTQLVLEPELQ	1384
DiVa:YP_006905850.1	IIVLSPANNVCSNFMNIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1198
ObRv1:YP_009408144.1	IIVLSPANNVCSNFMNIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1198
CLBV:NP_624333.1	CILLSESSAANSFRMIVALTRAKTRFLSFCSTLGGMDEFKIKRREGLVSTLLEKQKQIFTFERSNMVV---KCLNLTQEK-KNGCSDEVDREARLGGDPLKGLVLPVLEPEAF	1515
CLBV:AF443536.1	CILLSESSAANSFRMIVALTRAKTRFLSFCSTLGGMDEFKIKRREGLVSTLLEKQKQIFTFERSNMVV---KCLNLTQEK-KNGCSDEVDREARLGGDPLKGLVLPVLEPEAF	1515
AVCav:YP_008997790.1	AILLSENSALSDHMMIVALTRARKGFALNCPTIRRDDF-LRQSEGKILHKLKQKVSMDFTRFMGGKVMKKAQML---EAVGRTREEFEACEGDPMKQAQSLFLENTIPE	1207
CPvV:YP_009505632.1	CIVLSEYAEKQDYYRMMIVALTRAKETISFISLHRSGLTGFMSMIGRPHAFITGLPFTSNRNMVV---NCLVECHER-ATGGRDREVDREARLGGDPLKGLVLPVLEPEAF	1518
WVA:YP_009357235.1	CIVLSEYAEKQDYYRMMIVALTRAKETISFISLHRSGLTGFMSMIGRPHAFITGLPFTSNRNMVV---NCLVECHER-ATGGRDREVDREARLGGDPLKGLVLPVLEPEAF	1518
WVA:QEA69426.1	LIVLSENSAANSFRMIVALTRARKTRFLSFCSTLGGMDEFKIKRREGLVSTLLEKQKQIFTFERSNMVV---KCLNLTQEK-KNGCSDEVDREARLGGDPLKGLVLPVLEPEAF	1515
CRMaV:YP_007761581.1	CVLLTQDSMLVDERRMMIVALSRKINISFISLHRSGLTGFMAETDGMKINALLKGEKVNKLEKSKKGF---NLNFEFNEIKNGNDQDREARLGGDPLKGLVLPVLEPEAF	1552
CTLav:YP_009046478.1	CVLLTQDSMLVDERRMMIVALSRKINISFISLHRSGLTGFMAETDGMKINALLKGEKVNKLEKSKKGF---NLNFEFNEIKNGNDQDREARLGGDPLKGLVLPVLEPEAF	1563
PhLbV:YP_001552317.1	TIITNYSKDEKRRMMIVALSRFSENICFVNLVNSLSEALARMYATLGGVGRFKRAKLSLDLLEHLPG---AVFTDSDY---ENIGKQDREVDREARLGGDPLKGLVLPVLEPEAF	1584
GCLV:YP_004936159.1	AIVLSEYKALASERRMMIVALTRARKTRVITLNLGCKSLHAEFISNLRALGRFLSCTASDNLRLPGE---PNFVEELV-PTIGANLGVVEKLVGGDPLKGLVLPVLEPEAF	1489
ASPV:NP_604464.1	AIVLSEYKALASERRMMIVALSRFSENICFVNLVNSLSEALARMYATLGGVGRFKRAKLSLDLLEHLPG---AVFTDSDY---ENIGKQDREVDREARLGGDPLKGLVLPVLEPEAF	1714
APV1:YP_009094347.1	IILSHESTLTSERRMMIVALSRFSENICFVNLVNSLSEALARMYATLGGVGRFKRAKLSLDLLEHLPG---AVFTDSDY---ENIGKQDREVDREARLGGDPLKGLVLPVLEPEAF	1558

AgVT YWNV-RNTRLALAGDDMYAVGKLLQ--LRDRDRELLDK--FTLKKAVQFTESPMMFCGHWYMPFGIIEKPLRVLERWLIAEFEKGLKCEIYINAIIEVSYGYRLGEYQYEVIK---NIE 1547
PVT:ADX41471.1 YNIT-RDVTVAFAGDDMYASGKLE--IRKDREDLAH--LTLKAKVQTEKPMFCGHWYIKKMGIVKEPRLVLERWLIARERKVIDQCFINYSIEVSYGYRLGEYLWEYFD---NLE 1549
PVT:AFU55321.1 YNIT-RDVTVAFAGDDMYASGKLE--IRKDREDLAH--LTLKAKVQTEKPMFCGHWYIKKMGIVKEPRLVLERWLIARERKVIDQCFINYSIEVSYGYRLGEYLWEYFD---NLE 1550
PVT:AKX90539.1 YNIT-RDVTVAFAGDDMYASGKLE--IRKDREDLAH--LTLKAKVQTEKPMFCGHWYIKKMGIVKEPRLVLERWLIARERKVIDQCFINYSIEVSYGYRLGEYLWEYFD---NLE 1550
PVT:AHM92766.1 YKIN-RRTPTAFAGDDMFSPGRLE--VRDRDFLLNR--FSLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1721
PVT:AHM92766.1 YKIN-RRTPTAFAGDDMFSPGRLE--VRDRDFLLNR--FSLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1721
ZQVT:QBS17025.1 YKFD-SKTPTAFAGDDMFSPKRLD--VRDRDFLLNR--FSLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1759
CHVT:QMG41875.1 YKIN-RRTPTAFAGDDMFAPGLK--IRKDREDLNR--FSLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1724
FiTA:QED42804.1 YKIN-RRTPTAFAGDDMFAPGLK--IRKDREDLNR--FSLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1724
FiTA:QED42832.1 YKID-CKTPTAFAGDDMFSPGRLE--LRDRDFLLNR--FSLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1740
GVA:NP_619662.1 YHID-RNTPMCFAGDDMYSPGILR--VKDKYEATLQD--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1656
ACVB:YF_004935358.1 YKIN-KDPTCYAGDDMYAPGHLI--ISKEHEGTLQD--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---NIEQ 1651
ASGV:NP_044335.1 YKIDPRRHRILFAGDDMCSLSLKRKRGERATRLMKS--FSLTAVEEVRKFPFMMFCGHWYLSYGVIIKSPKLLWARIKMSERQLKCEVDNYLFEATFAYRLGERLYVLIK---EEDFE 1550
YAVA:YF_009268859.1 YKIDPRRHRILFAGDDMCSLTLRKRKNSKESQRLSQ--FSLTAVEEVRKFPFMMFCGHWYLSYGVIIKSPKLLWARIKMSERQLKCEVDNYLFEATFAYRLGERLYVLIK---EEDFE 1551
RAVA:YF_009553496.1 YSGA-EGFPTLIFAGDDMCIIGTIS--ESKGGKLLDRCLRLKSKTFRKYDPEFCGWRLLPGLIFKDPILMRYTRTKLHHEGQRLKEVINSYALELAFAGYRLGDYCYEFMT--EQMG 1790
PGPV:YF_004732978.2 YQMS-GHEPTCFAGDDMICALDLK--ESDEYNAPFGS--FSLKAKVCRVTXKPLFCGWRLLTKFGLYKPELVYERLKLIAIEKDKLDLVIDSYVLFECYAYLGSWLDNWLDF--EQAD 1793
CtChv-1:YF_009183999.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1822
CtChv-2:YF_009183996.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1822
DiVA:YF_006905850.1 YDLN-GTESICFAGDDMCCNRGKARVDGKVFHLLNR--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1636
OBRV1:YF_009408144.1 YDLN-GTESICFAGDDMCCNRGKARVDGKVFHLLNR--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1636
CLBV:NP_624333.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1646
CLBV:AF443536.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1646
AVCAv:YF_008997790.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1646
CPV:YF_009505632.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1646
WVA:YF_009357235.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1646
WVA:QEA69426.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1646
CRMAv:YF_007761581.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1646
CTLav:YF_009046478.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1999
PhLVB:YF_001552317.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 2020
GCLV:YF_004936159.1 YKIN-RDVTVAFAGDDMCIILNAK--IRNEMNDIFGS--LTLKAKVNFSEKPEMFCGWRMTYGVIVKEPKLVLERFKIAEERGCFCLECLINCYLEVSFAAYRLGERLYVLIK---EEDFE 1925
ASPV:NP_604464.1 YDLN-GSEATCFAGDDMCANRRL--VSKNNENFGK--IKLAKAVQTEKPMFCGHWYLSYGVIIKSPKLLWARIKMSERQLKCEVDNYLFEATFAYRLGERLYVLIK---EEDFE 2150
APV1:YF_009094347.1 YDLN-GSEATCFAGDDMCANRRL--VSKNNENFGK--IKLAKAVQTEKPMFCGHWYLSYGVIIKSPKLLWARIKMSERQLKCEVDNYLFEATFAYRLGERLYVLIK---EEDFE 1994

AgVT DQQAIVRRIKIKKSKFLPLSVRFLFHTV-----DNGFGSSEHFSEEMG---VESEHHGSCGLWC-HLH----- 1605
PVT:ADX41471.1 DFQATVRLVTKKKKQLPPAIRRFETS-----NGVDFSGEVQETMG---GEGEHGSCGLWC-NLH----- 1606
PVT:AFU55321.1 DFQATVRLVTKKKKQLPPAIRRFETS-----NGVDFSGEVQETMG---GEGEHGSCGLWC-NLH----- 1606
PVT:AKX90539.1 DFQATVRLVTKKKKQLPPAIRRFETS-----NGVDFSGEVQETMG---REGEHHGSCGLWC-NLH----- 1607
PVT:AHM92766.1 DFQATVRLVTKKKKQLPPAIRRFETS-----NGVDFSGEVQETMG---GEGEHGSCGLWC-NLH----- 1607
PVT:AHM92766.1 DQALVRLVIVKMKKFLPKKIRKFEFESF-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1778
ZQVT:QBS17025.1 DQALVRLVIVKMKKFLPKKIRKFEFESF-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1778
CHVT:QMG41875.1 DQALVRLVIVKMKKFLPKKIRKFEFESF-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1816
FiTA:QED42804.1 DQALVRLVIVKMKKFLPKKIRKFEFESF-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1816
FiTA:QED42832.1 DQALVRLVIVKMKKFLPKKIRKFEFESF-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1781
GVA:NP_619662.1 DQALVRLVIVKMKKFLPKKIRKFEFESF-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1781
ACVB:YF_004935358.1 AQQELVREIIVKIKKHLPKKISDLFSED-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1797
ASGV:NP_044335.1 AQQELVREIIVKIKKHLPKKISDLFSED-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1797
YAVA:YF_009268859.1 YHVMVRRFVFNKSKLLRGASKSILILEIG-EG--IGS--KWLSSSTASRRSNLQ--TSKL----- 1604
RAVA:YF_009553496.1 YHVMVRRFVFNKSKLLRGASKSILILEIG-EG--IGS--KWLSSSTASRRSNLQ--TSKL----- 1604
ACLSV:NP_040551.1 YKMTVDLFLRHNKMLFEVSLKGLRDSLVEGDV-----DHEFGGCSVQEEGE---CDRVHRFGGGLC-NI----- 1826
PGPV:YF_004732978.2 YHVMVRRFVFNKSKLLRGASKSILILEIG-EG--IGS--KWLSSSTASRRSNLQ--TSKL----- 1884
CtChv-1:YF_009183999.1 YHVMVRRFVFNKSKLLRGASKSILILEIG-EG--IGS--KWLSSSTASRRSNLQ--TSKL----- 1855
CtChv-2:YF_009183996.1 YHVMVRRFVFNKSKLLRGASKSILILEIG-EG--IGS--KWLSSSTASRRSNLQ--TSKL----- 1857
DiVA:YF_006905850.1 FHQLVIRRMIAKAKHLMKSSVEILKEI-----QDCFSQDGVDD----- 1856
OBRV1:YF_009408144.1 AHYCCIRLVHHLKDLRGOQSLDKYISN-----RRFKHCKSWSIQRPSYR-----SSTMETETLIASGSVRCQ 1699
CLBV:NP_624333.1 AHYCCIRLVHHLKDLRGOQSLDKYISN-----RRFKHCKSWSIQRPSYR-----SSTMETETLIASGSVRCQ 1682
CLBV:AF443536.1 YHQAIVRIVVTHIDKDKLTKRVKDLFLEQ-----SSDEDI----- 1962
AVCAv:YF_008997790.1 YHQAIVRIVVTHIDKDKLTKRVKDLFLEQ-----SSDEDI----- 1987
CPV:YF_009505632.1 YHQAIVRIVVTHIDKDKLTKRVKDLFLEQ-----SSDEDI----- 1679
WVA:YF_009357235.1 YHQAIVRIVVTHIDKDKLTKRVKDLFLEQ-----SSDEDI----- 1986
WVA:QEA69426.1 YHQAIVRIVVTHIDKDKLTKRVKDLFLEQ-----SSDEDI----- 1826
CRMAv:YF_007761581.1 YHQAIVRIVVTHIDKDKLTKRVKDLFLEQ-----SSDEDI----- 1826
CTLav:YF_009046478.1 AHYFCVTRFLQNKLLFSSNALEFFESG-----EGCLSPER-----NFG----- 2027
PhLVB:YF_001552317.1 AHYFCVTRFLQNKLLFSSNALEFFESG-----EGCLSPER-----NFG----- 2037
GCLV:YF_004936159.1 AFYNCVRIIVKMKKLLKSDIKNLFTR-----ESCKSPDR-----NFG----- 2051
ASPV:NP_604464.1 SHYNCVRIILQNKLLKSNVLTYSFA-----ALEE----- 1951
APV1:YF_009094347.1 AHYNCVRIIVQHNHLLKSNIRDLFKGE-----SLPAS-----S 2183
NHYNCVRIIHKSHLLKSSVRDLFLSGM----- 2022

