Acta virologica **66**: 149 – 156, 2022 doi: 10.4149/av_2022_204

Artemisia capillaris nucleorhabdovirus 1, a novel member of the genus *Alphanucleorhabdovirus*, identified in the *Artemisia capillaris* transcriptome

Dongjin Choi[1], Chaerim Shin[1], Ken Shirasu[2,3], Yasunori Ichihashi[4], Yoonsoo Hahn[1*]

[1]Department of Life Science, Chung-Ang University, Seoul 06974, South Korea; [2]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan; [3]Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan; [4]RIKEN BioResource Research Center, 3-1-1 Koyadai, Tsukuba, Ibaraki, 305-0074, Japan

Received February 17, 2022; accepted April 5, 2022

Summary. – A novel, negative-sense, single-stranded RNA virus, Artemisia capillaris nucleorhabdovirus 1 (AcNRV1), was identified in the transcriptome data of *Artemisia capillaris* (commonly known as capillary wormwood) root tissue. The AcNRV1 genome contains six open reading frames encoding a nucleocapsid (N), phosphoprotein, movement protein P3, matrix protein, glycoprotein, and polymerase (L). Sequence comparison and phylogenetic analysis using L and N protein sequences revealed that AcNRV1 is a novel member of the genus *Alphanucleorhabdovirus*, one of the six plant-infecting rhabdovirus genera of the family *Rhabdoviridae*. Wheat yellow striate virus and rice yellow stunt virus were identified as the closest known rhabdoviruses of AcNRV1. The conserved regulatory sequences involved in transcription termination/polyadenylation (TTP) and transcription initiation (TI) of individual genes were identified in the AcNRV1 genome with the consensus sequence 3'-(A/U)UUAUUUUU–GGG–UUG-5' (in the negative-sense genome), whereby dashes separate the TTP, untranscribed intergenic spacer, and TI elements. The AcNRV1 genome sequence will contribute to further understanding the genome structural evolution of plant rhabdoviruses.

Keywords: Artemisia capillaris nucleorhabdovirus 1; plant virus; Alphanucleorhabdovirus; Rhabdoviridae

Introduction

Rhabdoviruses (the family *Rhabdoviridae*) infect animals and plants and have negative-sense, single-stranded RNA genomes (Dietzgen *et al*., 2017; Walker *et al*., 2018; Bejerman *et al*., 2021). The family *Rhabdoviridae* comprises three subfamilies, 40 genera, and more than 200 species currently approved by the International Committee on Taxonomy of Viruses (https://talk.ictvonline.org, last accessed on January 20, 2022). Plant-infecting rhabdoviruses of the subfamily *Betarhabdovirinae* are classified into six genera: *Alphanucleorhabdovirus*, *Betanucleorhabdovirus*, *Cytorhabdovirus*, *Dichorhavirus*, *Gammanucleorhabdovirus*, and *Varicosavirus*. They have either monopartite (*Alphanucleorhabdovirus*, *Betanucleorhabdovirus*, *Cytorhabdovirus*, and *Gammanucleorhabdovirus*) or bipartite (*Dichorhavirus* and *Varicosavirus*) genomes (Dietzgen *et al*., 2017). Insects, including aphids, leafhoppers, and planthoppers, serve as transmission vectors for plant rhabdoviruses (Ammar *et al*., 2009; Liu *et al*., 2018; Whitfield *et al*., 2018; Bhat *et al*., 2020).

Plant rhabdoviruses with a monopartite genome have six shared open reading frames (ORFs) encoding a nucle-

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

Abbreviations: AcNRV1 = Artemisia capillaris nucleorhabdovirus 1; ORF = open reading frame; RdRp = RNA-dependent RNA polymerase; RYSV = rice yellow stunt virus; SRA = Sequence Read Archive; TI = transcription initiation; TTP = transcription termination/polyadenylation; UIS = untranscribed intergenic spacer; WYSV = wheat yellow striate virus

ocapsid (N), phosphoprotein (P), movement protein 3 (P3), matrix protein (M), surface glycoprotein (G), and large multi-functional protein (L), in the order of 3'-N-P-P3-M-G-L-5' (Walker *et al.*, 2018). The L protein, which contains an RNA-directed RNA polymerase (RdRp) domain, is involved in the genome replication and transcription of individual ORFs (Jackson *et al.*, 2005; Walker *et al.*, 2018). The N protein encloses the viral RNA genomic molecule in the protein shell. The P protein acts as a cofactor of the L protein during transcription and replication and helps N proteins during encapsidation of genomic molecules (Fang *et al.*, 2019). The M protein is required for viral maturation and budding (Sun *et al.*, 2018). The G protein functions in receptor binding and membrane fusion (Coll, 1995; Mann *et al.*, 2016). The P3 protein enables viruses to move from one cell to another (Zhou *et al.*, 2019). Some plant rhabdoviruses may have additional ORFs apart from the abovementioned six (Walker *et al.*, 2011). For example, wheat yellow striate virus (WYSV) and rice yellow stunt virus (RYSV) have a seventh ORF, P6, which encodes a small protein between the G and L ORFs (Huang *et al.*, 2003; Liu *et al.*, 2018).

Rhabdovirus genes are transcribed into separate mRNAs from the negative-sense RNA genome through a transcriptase complex composed of the L and P proteins (Jackson *et al.*, 2005; Walker *et al.*, 2011). The gene junction regions of rhabdoviruses contain conserved regulatory sequences called transcription termination/polyadenylation (TTP) and transcription initiation (TI) elements that mediate polyadenylation of the preceding gene transcript and transcriptional initiation of the next gene, respectively (Jackson *et al.*, 2005; Goh *et al.*, 2020; Bejerman *et al.*, 2021; Shin *et al.*, 2021). One or several nucleotides located between the TTP and TI elements are not transcribed; this region is called the untranscribed intergenic spacer (UIS). For example, the consensus sequences of the WYSV and RYSV conserved regulatory sequences are presented as 3'-UAAAUUUUU–GGGG–UUG-5' and 3'-AUUAUUUUU–GGG–UUG-5', respectively, whereby dashes (–) separate the TTP, UIS, and TI elements (Huang *et al.*, 2003; Liu *et al.*, 2018; Bejerman *et al.*, 2021).

RNA-sequencing (RNA-seq) data are frequently obtained from plant tissues to investigate gene expression patterns during interaction with other organisms and plant development and differentiation (Ichihashi *et al.*, 2018; Yoshida *et al.*, 2019). When plant tissues are latently infected with RNA viruses, full-length virus genome sequences can be obtained by assembling transcriptome reads and identifying virus genome contigs (Bejerman *et al.*, 2020; Choi *et al.*, 2021; Goh *et al.*, 2021; Park *et al.*, 2021; Park and Hahn, 2021). Many novel rhabdovirus genome sequences, including those of Trichosanthes associated rhabdovirus 1, Agave tequilana virus 1, and Zostera associated varicosavirus 1, have previously been identified in plant transcriptome data (Goh *et al.*, 2020; Bejerman *et al.*, 2021; Shin *et al.*, 2021). In the present study, the genome sequence of a novel rhabdovirus belonging to the genus *Alphanucleorhabdovirus* (the family *Rhabdoviridae*) was identified in *Artemisia capillaris* (commonly known as capillary wormwood) root transcriptome data (Ichihashi *et al.*, 2018).

Materials and Methods

Transcriptome data analysis. RNA-seq data were previously obtained from the root parasitic plant *Thesium chinense* and its host plants (*A. capillaris*, *Eragrostis curvula*, and *Lespedeza juncea*) to understand transcriptional regulation during haustorium formation (Ichihashi *et al.*, 2018). Sequence data are deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under SRA Project Acc. No. SRP114897. RNA-seq data (14 sequencing runs from *T. chinense* and six from its hosts) were filtered to collect high-quality reads using the sickle program (version 1.33; https://github.com/najoshi/sickle), with the parameter "-q 30 -l 55." High-quality reads from the 20 sequencing runs were separately assembled into contigs using the SPAdes Genome Assembler (version 3.15.3; http://cab.spbu.ru/software/spades), with the parameter "--rnaviral" (Bushmanova *et al.*, 2019).

The assembled transcriptome contigs were compared with known viral RdRp domain sequences. A total of 2620 RdRp domain sequences were obtained from the Pfam database (release 34.0; https://pfam.xfam.org). The Pfam Acc. Nos. are PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03035, PF03431, PF04196, PF04197, PF05788, PF05919, PF06317, PF06478, PF07925, PF08467, PF08716, PF08717, PF12426, and PF17501. For sequence similarity searches, the DIAMOND program (version 2.0.4.142; https://github.com/bbuchfink/diamond) in the blastx mode was used (Buchfink *et al.*, 2015).

Viral genome annotation. RNA-seq read depth of the virus genome was examined by mapping high-quality reads to the viral contig sequence using the bwa-mem2 program (version 2.0pre2; https://github.com/bwa-mem2/bwa-mem2). Putative ORFs were predicted using the getorf program of the EMBOSS package (version 6.6.0.0; http://emboss.open-bio.org), with the parameter "-find 1." Putative conserved domains in the viral protein sequences were identified using the InterPro web server (release 87.0; https://www.ebi.ac.uk/interpro). Signal peptides and transmembrane domains were predicted using SignalP (version 6.0; https://services.healthtech.dtu.dk/service.php?SignalP-6.0) and TMHMM (version 2.0; https://services.healthtech.dtu.dk/service.php?TMHMM-2.0), respectively.

Conserved regulatory sequences in gene junction regions were predicted using the MEME webtool (version 5.4.1; https://

meme-suite.org/meme/tools/meme). The WebLogo application (version 3; http://weblogo.threeplusone.com) was used to create a sequence logo representation of the conserved sequences.

Phylogenetic analysis. Known viral protein sequences were obtained using NCBI's BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The needle program of the EMBOSS package was used to calculate pairwise identities of orthologous viral proteins. The MAFFT program (version 7.475; https://mafft.cbrc.jp/alignment/software) with the parameter "--auto" was used to generate multiple sequence alignments. Gap-rich segments of multiple alignments were removed using trimAl (version 1.4.rev22; http://trimal.cgenomics.org) with the parameter "-automated1" (Capella-Gutierrez *et al.*, 2009). Maximum-likelihood phylogenetic trees were generated using the IQ-TREE program (version 2.1.3; http://www.iqtree.org) using the parameter "-B 1000" to calculate bootstrap support values from 1000 replicates (Minh *et al.*, 2020).

Results and Discussion

High-quality RNA-seq reads obtained from the parasitic plant *T. chinense* and its host plants, *A. capillaris*, *E. curvula*, and *L. juncea*, were assembled to generate transcriptome contigs (Ichihashi *et al.*, 2018). Sequence comparisons with known viral RdRp sequences revealed that several plant transcriptome contigs may contain a viral RdRp domain. One contig that was assembled from transcriptome data (SRA Acc. No. SRR5917890) derived from *A. capillaris* root tissue showed substantial sequence similarities with the RdRp domains of WYSV and RYSV (Huang *et al.*, 2003; Maurino *et al.*, 2018). Sequence similarity search of the *A. capillaris* contig against the NCBI protein database confirmed that it encodes multiple proteins similar to those of known members of the genus *Alphanucleorhabdovirus*; the highest sequence similarity scores with WYSV and RYSV of this genus. Therefore, the *A. capillaris* contig was assumed to be the genome sequence of a novel member of *Alphanucleorhabdovirus* of the family *Rhabdoviridae* and tentatively named Artemisia capillaris nucleorhabdovirus 1 (AcNRV1). The AcNRV1 genome sequence has been deposited in NCBI under the Acc. No. OM372677.

The AcNRV1 genome is 13939 nt in length and has six complete ORFs: N, encoding a 479-amino acid (aa) nucleocapsid protein; P, a 365-aa phosphoprotein; P3, a 295-aa movement protein; M, a 269-aa matrix protein; G, a 655-aa glycoprotein; and L, a 1960-aa polymerase (Fig. 1). N, P, M, G, and L are canonical rhabdovirus proteins shared by members of the family *Rhabdoviridae*, whereas P3 is present in members of the five plant rhabdovirus genera *Alphanucleorhabdovirus*, *Betanucleorhabdovirus*, *Cytorhabdovirus*, *Dichorhavirus*, and *Gammanucleorhabdovirus* (Walker *et al.*, 2011; Walker *et al.*, 2018; Shin *et al.*, 2021). AcNRV1 proteins have functional domains commonly found in other rhabdoviruses. The N protein has a "Rhabdovirus nucleoprotein" domain (InterPro Acc. No. IPR004902). The L protein has a "Mononegavirales RNA-directed RNA poly-

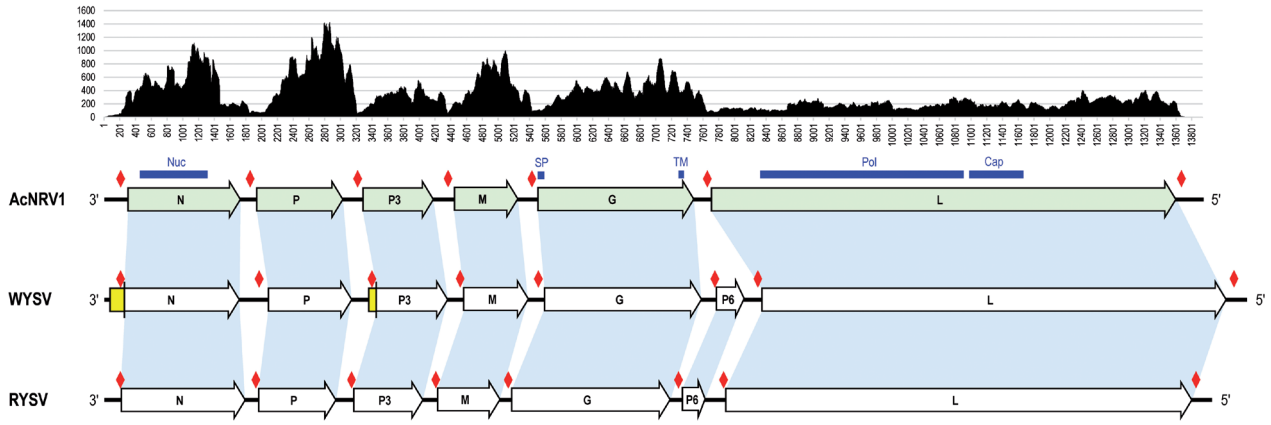

Fig. 1

Comparison of the genome organization of AcNRV1 and related viruses

Genome organization of AcNRV1 and two closely related viruses (WYSV and RYSV) are presented. The RNA-seq read depth of the AcNRV1 genome contig is shown at the top. Protein-coding ORFs are indicated with light green (AcNRV1) or white (WYSV and RYSV) arrows in the 3'-to-5' direction. Yellow regions in the WYSV N and P3 ORFs indicate genomic segments that were incorrectly predicted as being part of the ORF. Light blue shades connect orthologous ORFs of the three viruses (note that AcNRV1 lacks the P6 ORF). Predicted functional domains of the AcNRV1 proteins are marked by blue lines above the ORF: Nuc, "Rhabdovirus nucleoprotein" (InterPro Acc. No. IPR004902); SP, signal peptide; TM, transmembrane domain; Pol, "Mononegavirales RNA-directed RNA polymerase catalytic domain" (IPR014023); Cap, "Mononegavirales mRNA-capping region V" (IPR026890). Red diamonds indicate putative conserved regulatory sequences in the gene junction regions.

merase catalytic domain" (IPR014023) and a "Mononegavirales mRNA-capping region V" domain (IPR026890). The G protein has a signal peptide and a transmembrane domain near the N- and C-termini, respectively.

WYSV and RYSV are the closest known rhabdoviruses based on comparisons of the L and N protein sequences (Table 1). Further sequence comparisons using the other four proteins of AcNRV1 and their orthologs in WYSV and RYSV confirmed that they are closely related. Multiple alignments of all six proteins are presented in Supplementary Data S1. The WYSV and RYSV genomes have an additional ORF (called P6) located between the G and L ORFs, in the order of 3'-N-P-P3-M-G-P6-L-5' (Fig. 1) (Huang *et al.*, 2003; Liu *et al.*, 2018). The P6 proteins of WYSV and RYSV share 29.5% aa sequence identity, implying that they descended from a common ancestral protein. However, no ORF equivalent to P6 was found present in the genome of AcNRV1 or that of any other known rhabdovirus, indicating that the P6 ORF is specific to WYSV and RYSV.

In rhabdovirus genomes, the gene junction regions have a conserved regulatory sequence involved in the transcription termination and polyadenylation of the preceding gene transcript (the TTP element) as well as the transcriptional initiation and capping of the following gene transcript (the TI element) (Jackson *et al.*, 2005; Ogino and Green, 2019; Bejerman *et al.*, 2021; Shin *et al.*, 2021). The gene junction regions of the AcNRV1 genome, together with those of the WYSV and RYSV genomes, were analyzed to predict conserved sequences. Seven (3'–N, N–P, P–P3, P3–M, M–G, G–L, and L–5' for AcNRV1) or eight (3'–N, N–P, P–P3, P3–M, M–G, G–P6, P6–L, and L–5' for WYSV and RYSV) gene junction regions were extracted. A motif search analysis using 23 gene junction region sequences revealed shared motif sequences 20- or 21-nt long (Fig. 2). The deduced consensus sequence of the AcNRV1 gene junction regions was 3'-(A/U)UUAUUUUU–GGG–UUG-5', where dashes (–) separate TTP, UIS, and TI elements.

The TTP element in the gene junction region induces transcription termination of the preceding gene. Since the AcNRV1 genome sequence was assembled from a large number of viral transcripts and a small number of viral genomic RNAs, the sequencing depths near the conserved sequences in gene junction regions would be lower than in other regions. The read depths of the AcNRV1 genome sequence sharply decreased in regions where conserved regulatory sequences were predicted (Fig. 1). Similar sharp drops in sequencing depths at the conserved sequences were observed in other rhabdoviruses (Shin *et al.*, 2021).

The TI element of a conserved gene junction sequence mediates the transcriptional initiation of the downstream gene. In AcNRV1, WYSV, and RYSV gene junction regions, the TI element has a 3'-UUG-5' sequence. When the positions of ORFs and conserved regulatory sequences of

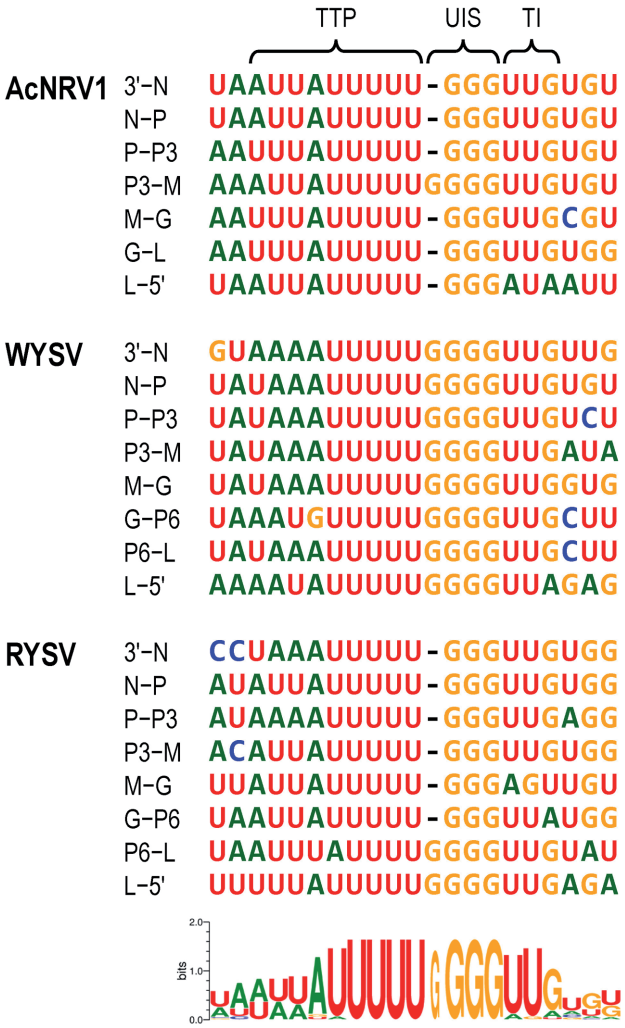

Fig. 2

Conserved sequences in the gene junction regions
Putative conserved regulatory sequences for polyadenylation and transcriptional initiation, identified in the gene junction regions of AcNRV1, WYSV, and RYSV, are shown in the 3'-to-5' orientation. A sequence logo representation generated from 23 sequences is shown at the bottom. TTP, transcription termination/polyadenylation; UIS, untranscribed intergenic spacer; TI, transcription initiation.

AcNRV1, WYSV, and RYSV were compared, two conserved sequences of WYSV associated with the N and P3 ORFs were found located within the respective ORF (Fig. 1). The predicted TI element possibly associated with the N ORF is located at positions 209–211 of the WYSV genome sequence, whereas the start codon AUG is at positions 77–79. The second in-frame AUG codon is present at positions 230–232, which is closely positioned after the predicted TI element, supporting the possibility that the second AUG codon is the actual start codon. Sequence comparison of N proteins of AcNRV1, WYSV, and RYSV showed that the

Table 1. Sequence comparison of the L and N proteins of AcNRV1 and representative *Rhabdoviridae* viruses

No	Genus	Virus	Acronym	Genome[a]	L[b]	N[b]
1	*Alphanucleo-rhabdovirus*	Wheat yellow striate virus	WYSV	NC_055484.1	937/1992 (47.0%)	199/588 (33.8%)
2		Rice yellow stunt virus	RYSV	NC_003746.1	932/2012 (46.3%)	200/525 (38.1%)
3		Joa yellow blotch-associated virus	JYBaV	MW014292.1	706/2097 (33.7%)	122/535 (22.8%)
4		Potato yellow dwarf virus	PYDV	NC_016136.1	689/2085 (33.0%)	140/510 (27.5%)
5		Physostegia chlorotic mottle virus	PhCMoV	NC_055466.1	717/2096 (34.2%)	119/504 (23.6%)
6		Eggplant mottled dwarf virus	EMDV	NC_025389.1	710/2080 (34.1%)	119/521 (22.8%)
7		Peach virus 1	PeV1	MN520414.1	720/2102 (34.3%)	152/505 (30.1%)
8		Maize Iranian mosaic virus	MIMV	NC_036390.1	659/2102 (31.4%)	129/502 (25.7%)
9		Maize mosaic virus	MMV	NC_005975.1	655/2131 (30.7%)	129/523 (24.7%)
10		Taro vein chlorosis virus	TaVCV	NC_006942.1	679/2106 (32.2%)	135/546 (24.7%)
11		Morogoro maize-associated virus	MMaV	NC_055512.1	654/2078 (31.5%)	130/513 (25.3%)
12		Agave tequilana virus 1	ATV1	BK014297.1	677/2113 (32.0%)	138/507 (27.2%)
13	*Gammanucleo-rhabdovirus*	Maize fine streak virus	MFSV	NC_005974.1	590/2213 (26.7%)	121/554 (21.8%)
14	*Betanucleo-rhabdovirus*	Alfalfa-associated nucleorhabdovirus	AaNV	MG948563.1	595/2261 (26.3%)	102/524 (19.5%)
15		Sonchus yellow net virus	SYNV	NC_001615.3	577/2291 (25.2%)	111/524 (21.2%)
16		Cardamom vein clearing virus	CdVCV	MN273311.1	605/2321 (26.1%)	116/543 (21.4%)
17		Black currant-associated rhabdovirus	BCaRV	MF543022.1	583/2356 (24.7%)	99/545 (18.2%)
18		Datura yellow vein virus	DYVV	NC_028231.1	607/2343 (25.9%)	98/524 (18.7%)
19		Sowthistle yellow vein virus	SYVV	MT185675.1	624/2335 (26.7%)	121/523 (23.1%)
20	*Dichorhavirus*	Orchid fleck virus	OFV	NC_009609.1, NC_009608.1	590/2244 (26.3%)	110/544 (20.2%)
21		Citrus leprosis virus N	CiLV-N	NC_052231.1, NC_052230.1	591/2132 (27.7%)	104/522 (19.9%)
22		Citrus chlorotic spot virus	CiCSV	NC_055209.1, NC_055208.1	586/2149 (27.3%)	116/524 (22.1%)
23		Clerodendrum chlorotic spot virus	ClCSV	NC_043649.1, NC_043648.1	579/2121 (27.3%)	113/507 (22.3%)
24		Coffee ringspot virus	CoRSV	NC_038755.1, NC_038756.1	561/2146 (26.1%)	115/522 (22.0%)
25	*Cytorhabdovirus*	Northern cereal mosaic virus	NCMV	NC_002251.1	543/2340 (23.2%)	88/574 (15.3%)
26		Barley yellow striate mosaic virus	BYSMV	NC_028244.1	561/2333 (24.0%)	93/526 (17.7%)
27		Rice stripe mosaic virus	RSMV	NC_040786.1	537/2320 (23.1%)	83/613 (13.5%)
28		Tomato yellow mottle-associated virus	TYMaV	NC_034240.1	519/2303 (22.5%)	86/528 (16.3%)
29		Lettuce necrotic yellows virus	LNYV	NC_007642.1	544/2293 (23.7%)	89/583 (15.3%)
30		Trichosanthes associated rhabdovirus 1	TrARV1	BK011194.1	554/2391 (23.2%)	72/566 (12.7%)
31	*Varicosavirus*	Lettuce big-vein associated virus	LBVaV	NC_011558.1, NC_011568.1	516/2273 (22.7%)	95/559 (17.0%)
32		Alopecurus myosuroides varicosavirus 1	AMVV1	NC_026801.1, NC_026798.1	483/2345 (20.6%)	98/626 (15.7%)
33		Red clover associated varicosavirus	RCaVV	MF918568.1, MF918569.1	492/2278 (21.6%)	94/552 (17.0%)
34		Zostera associated varicosavirus 1	ZaVV1	BK014484.1, BK014485.1	529/2259 (23.4%)	89/562 (15.8%)

[a]NCBI Acc. No. for genome sequences (note that dichorhaviruses and varicosaviruses have two segments); [b]amino acid sequence identity to the AcNRV1 L and N proteins in the format of "identical residues/aligned length (percent identity)."

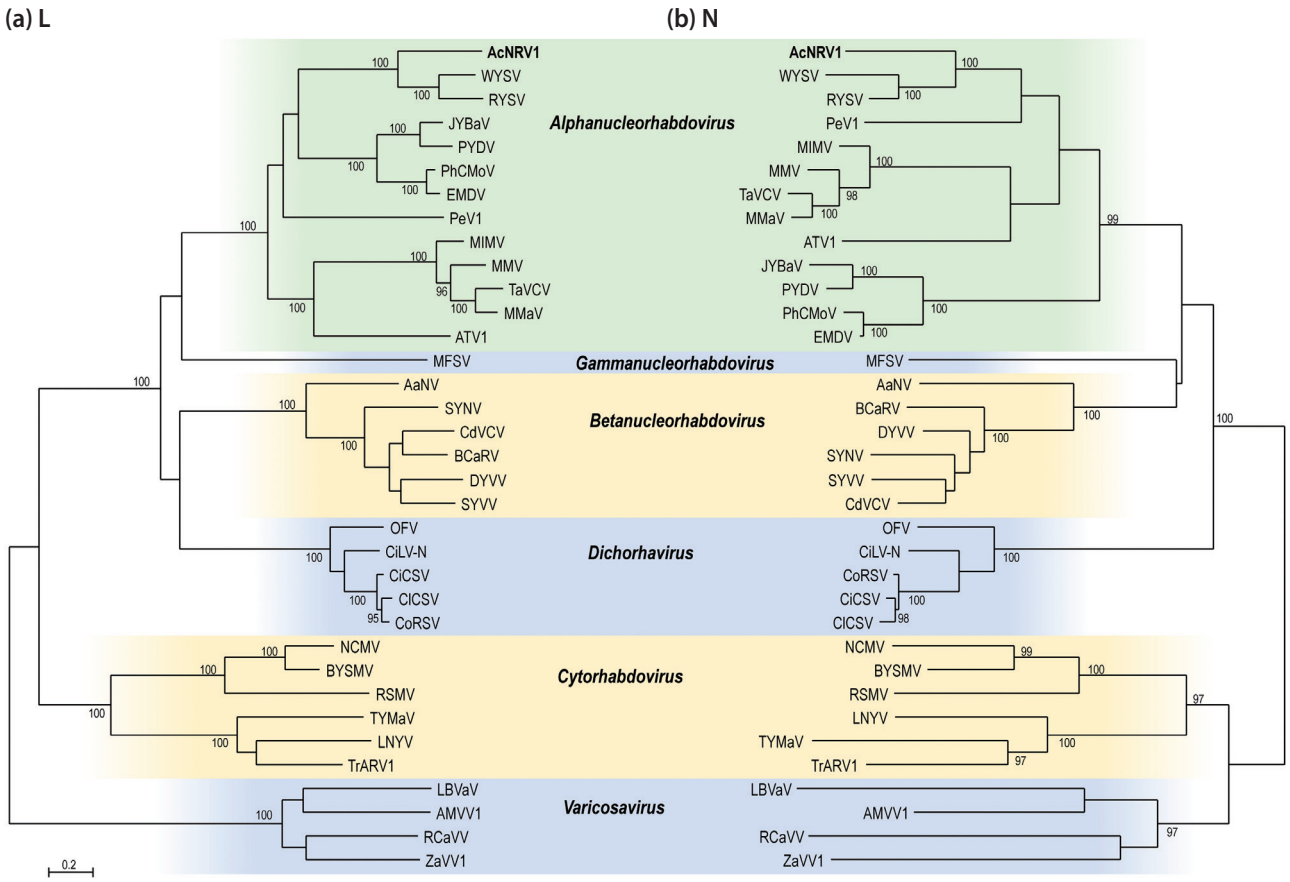

Fig. 3

Phylogenetic relationships of AcNRV1 and known plant rhabdoviruses

Maximum-likelihood phylogenetic trees were inferred based on the L **(a)** and N **(b)** protein sequences of AcNRV1 and 34 representative plant rhabdoviruses. AcNRV1 formed a subclade with WYSV and RYSV within the genus *Alphanucleorhabdovirus*. Bootstrap support percentages, 95% or greater, calculated from 1000 replicates, are shown at the respective nodes.

WYSV N protein has an extended region at its N-terminus compared with the AcNRV1 and RYSV N proteins, supporting that the WYSV N ORF was overpredicted (yellow-highlighted regions in Fig. 1 and Supplementary Data S1a) and that the second methionine is the true N-terminus.

Similarly, the predicted TI element associated with the WYSV P3 ORF is located at positions 3395–3397, whereas its annotated start codon is at positions 3348–3350. The P3 ORF has a second AUG codon at positions 3468–3470, which is in good agreement with the N-termini of the AcNRV1 and RYSV P3 proteins, indicating that the second methionine is the actual N-terminus (Supplementary Data S1c).

Overprediction of an ORF may hinder the identification of any associated conserved regulatory sequence. Indeed, the originally deduced regulatory sequences in the WYSV 3'–N and P–P3 regions were incorrect and markedly different from the typical consensus sequence because they were predicted from incorrectly defined untranslated gene junction regions (Liu *et al.*, 2018). Therefore, for the correct annotation of ORFs and conserved regulatory sequences in a rhabdovirus genome sequence, the locational relationships between ORFs and conserved gene junction sequences must be carefully examined.

To determine the phylogenetic position of AcNRV1, the L and N protein sequences of 34 viruses belonging to the six plant rhabdovirus genera were collected (Table 1). The AcNRV1 L and N proteins showed 30.7%–47.0% and 22.8%–38.1% aa sequence identities, respectively, with those of other members of the genus *Alphanucleorhabdovirus*. The L and N proteins of other rhabdoviruses showed 20.6%–27.7% and 12.7%–23.1% aa sequence identities with the AcNRV1 L and N proteins, respectively. Maximum-likelihood phylogenetic trees constructed from the L and N protein sequences confirmed that AcNRV1 belongs to the genus *Alphanucleorhabdovirus* and is closely related

to WYSV and RYSV, whereby the three form a strong subclade (Fig. 3).

In conclusion, the genome sequence of AcNRV1, a novel, negative-sense, single-stranded RNA virus, was identified in *A. capillaris* transcriptome data. The AcNRV1 genome has six protein-coding ORFs that are shared with members of the five plant rhabdovirus genera. Phylogenetic analysis revealed that AcNRV1 is a novel species of the genus *Alphanucleorhabdovirus* of the family *Rhabdoviridae*. The AcNRV1 genome sequence may be a useful resource for studying the evolution of genomic organization of plant rhabdoviruses.

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

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

References

Ammar E-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

Bejerman N, Debat H, Dietzgen RG (2020): The plant negative-sense RNA virosphere: Virus discovery through new eyes. Front. Microbiol. 11, 588427. https://doi.org/10.3389/fmicb.2020.588427

Bejerman N, Dietzgen RG, Debat H (2021): Illuminating the plant rhabdovirus landscape through metatranscriptomics data. Viruses 13, 1304. https://doi.org/10.3390/v13071304

Bhat AI, Pamitha NS, Naveen KP, Biju CN (2020): Identification and characterization of cardamom vein clearing virus, a novel aphid-transmitted nucleorhabdovirus. Eur. J. Plant Pathol. 156, 1053–1062. https://doi.org/10.1007/s10658-020-01958-2

Buchfink B, Xie C, Huson DH (2015): Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD (2019): rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. Gigascience 8. https://doi.org/10.1093/gigascience/giz100

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009): trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Choi D, Shin C, Shirasu K, Hahn Y (2021): Two novel poty-like viruses identified from the transcriptome data of purple witchweed (Striga hermonthica). Acta Virol. 65, 365–372. https://doi.org/10.4149/av_2021_402

Coll JM (1995): The glycoprotein G of rhabdoviruses. Arch. Virol. 140, 827–851. https://doi.org/10.1007/BF01314961

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

Fang XD, Yan T, Gao Q, Cao Q, Gao DM, Xu WY, Zhang ZJ, Ding ZH, Wang XB (2019): A cytorhabdovirus phosphoprotein forms mobile inclusions trafficked on the actin/ER network for viral RNA synthesis. J. Exp. Bot. 70, 4049–4062. https://doi.org/10.1093/jxb/erz195

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, Hahn Y (2021): A novel tepovirus, Agave virus T, identified by the analysis of the transcriptome data of blue agave (Agave tequilana). Acta Virol. 65, 68–71. https://doi.org/10.4149/av_2021_107

Huang Y, Zhao H, Luo Z, Chen X, Fang RX (2003): Novel structure of the genome of rice yellow stunt virus: identification of the gene 6-encoded virion protein. J. Gen. Virol. 84, 2259–2264. https://doi.org/10.1099/vir.0.19195-0

Ichihashi Y, Kusano M, Kobayashi M, Suetsugu K, Yoshida S, Wakatake T, Kumaishi K, Shibata A, Saito K, Shirasu K (2018): Transcriptomic and metabolomic reprogramming from roots to haustoria in the parasitic plant, Thesium chinense. Plant Cell Physiol. 59, 729–738. https://doi.org/10.1093/pcp/pcx200

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

Liu Y, Du Z, Wang H, Zhang S, Cao M, Wang X (2018): Identification and characterization of wheat yellow striate virus, a novel leafhopper-transmitted nucleorhabdovirus infecting wheat. Front. Microbiol. 9, 468. https://doi.org/10.3389/fmicb.2018.00468

Mann KS, Bejerman N, Johnson KN, Dietzgen RG (2016): Cytorhabdovirus P3 genes encode 30 K-like cell-to-cell movement proteins. Virology 489, 20–33. https://doi.org/10.1016/j.virol.2015.11.028

Maurino F, Dumon AD, Llauger G, Alemandri V, de Haro LA, Mattio MF, Del Vas M, Laguna IG, Gimenez Pecci MP (2018): Complete genome sequence of maize yellow striate virus, a new cytorhabdovirus infecting maize and wheat crops in Argentina. Arch. Virol. 163, 291–295. https://doi.org/10.1007/s00705-017-3579-7

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020): IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Ogino T, Green TJ (2019): RNA synthesis and capping by non-segmented negative strand RNA viral polymerases:

Lessons from a prototypic virus. Front. Microbiol. 10, 1490. https://doi.org/10.3389/fmicb.2019.01490

Park D, Goh CJ, Hahn Y (2021): Two novel closteroviruses, fig virus A and fig virus B, identified by the analysis of the high-throughput RNA-sequencing data of fig (Ficus carica) latex. Acta Virol. 65, 42–48. https://doi.org/10.4149/av_2021_104

Park D, Hahn Y (2021): Identification of genome sequences of novel partitiviruses in the quinoa (Chenopodium quinoa) transcriptome datasets. J. Gen. Plant Pathol. 87, 236–241. https://doi.org/10.1007/s10327-021-01002-z

Shin C, Choi D, Hahn Y (2021): Identification of the genome sequence of Zostera associated varicosavirus 1, a novel negative-sense RNA virus, in the common eelgrass (Zostera marina) transcriptome. Acta Virol. 65, 373–380. https://doi.org/10.4149/av_2021_404

Sun K, Zhou X, Lin W, Zhou X, Jackson AO, Li Z (2018): Matrix-glycoprotein interactions required for budding of a plant nucleorhabdovirus and induction of inner nuclear membrane invagination. Mol. Plant Pathol. 19, 2288–2301. https://doi.org/10.1111/mpp.12699

Walker PJ, Blasdell 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, Blasdell KR (2011): Rhabdovirus accessory genes. Virus Res. 162, 110–25. 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

Yoshida S, Kim S, Wafula EK, Tanskanen J, Kim YM, Honaas L, Yang Z, Spallek T, Conn CE, Ichihashi Y, Cheong K, Cui S, Der JP, Gundlach H, Jiao Y, Hori C, Ishida JK, Kasahara H, Kiba T, Kim MS, Koo N, Laohavisit A, Lee YH, Lumba S, McCourt P, Mortimer JC, Mutuku JM, Nomura T, Sasaki-Sekimoto Y, Seto Y, Wang Y, Wakatake T, Sakakibara H, Demura T, Yamaguchi S, Yoneyama K, Manabe RI, Nelson DC, Schulman AH, Timko MP, dePamphilis CW, Choi D, Shirasu K (2019): Genome sequence of Striga asiatica provides insight into the evolution of plant parasitism. Curr. Biol. 29, 3041–052.e4. https://doi.org/10.1016/j.cub.2019.07.086

Zhou X, Lin W, Sun K, Wang S, Zhou X, Jackson AO, Li Z (2019): Specificity of plant rhabdovirus cell-to-cell movement. J. Virol. 93, e00296–19. https://doi.org/10.1128/JVI.00296-19

Acta virologica **66**: 138 – 145, 2022 doi: 10.4149/av_2022_204

SUPPLEMENTARY INFORMATION

Artemisia capillaris nucleorhabdovirus 1, a novel member of the genus *Alphanucleorhabdovirus*, identified in the *Artemisia capillaris* transcriptome

Dongjin Choi[1], Chaerim Shin[1], Ken Shirasu[2,3], Yasunori Ichihashi[4], Yoonsoo Hahn[1*]

[1]Department of Life Science, Chung-Ang University, Seoul 06974, South Korea; [2]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan; [3]Graduate School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan; [4]RIKEN BioResource Research Center, 3-1-1 Koyadai, Tsukuba, Ibaraki, 305-0074, Japan

Received February 17, 2022; accepted April 5, 2022

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

Data S1. Multiple sequence alignments of the homologous proteins of AcNRV1, WYSV, and RYSV

(a) N (potentially overpredicted region is highlighted in yellow)

```
AcNRV1  ----------------------------------------------------MSDLSAYISTPLSGTFPLFRNRDNVPTVGNEAQEEPYIYQSYITWL    46
WYSV    MDTSLSGSQKYLGPRKRRHIPNHGKNITHRQGASCSSISHFKNPNNNFTGTMAQNPNVANYANAAPLPRFEGLGDRENLAPIGNEAVEIPYQKEAYLAWI   100
RYSV    ----------------------------------------------------MANDNVSDYANAAPFARFANLQNRETLNPIGNEAKEIPYNRDQYLTWL    48
                                                                ::: * .:.    *  : :*:.: .:**** * **  : *::*:

AcNRV1  NDGNTFALERLTDDQVIASWTAIEASMEGNTFSESNMRALVQCALNLRGVIGDNRPLYRELPENAARRYAPAPSTAPIYGGEAMQAGMLMPDMQIRQ--D   144
WYSV    NEGRVFQVNQLTDEQMIQMWETVKTSMQGNTFSEQHMRDIVQMACNLKGVDPATKPLYRQYEMPENGRWADAPSQDPIFSGQQ-VAGVIVPLQEAQP--L   197
RYSV    AEGKLFQIGALTDAEIVAAWTTIKTAMGNNTFSETHMRSIVKIACNLRGITPGSTPLLVTYNPPQSATWAPAPSTDAIYSGTP-VAGVIIPQNTGAGGED   147
         :*. * :  *** :::  * :::::* .***** :** :*: * **:*:   . **           :* ***  .*:.*    **:::*

AcNRV1  QESNTSRAKAICFICAYLVRLITKMDTHLEKSIEDIKKQYGKLYNTQSSLLNNWSPNITWAKRVKTAFDTYPWLKSTVAFAVGEADSSLERNTLDYGLCR   244
WYSV    VEDVSGKARAIGFICGFLLRFIVKTEEHLNNSLANLKLQFSRIYGVQSATINQWNPTTTWASRIKLAFDTYLTLRATVALHVALADGNLNADNVNFGLCR   297
RYSV    TETEASKARAIAFICCYLLRFIVKTEEHLTNSLGNLKLQYSRLYSAQSATLSNWNPSNTWASRVKLGFDTYLTLRATVAYNIASADALLVPENVNYGLCR   247
         *  :.:*:** *** :*:*:*.* : ** :*: ::* *:.::*..**: :.:*.*. ***.*:* .****  *::***  :. **. *  :.:::****

AcNRV1  MLLFQHLELSGMQIYKMTMTLIAHLNLIEPERFLEWICLPASTETVQKIYQIGITYDGPNHQANKHWKYAKLAKPNYWLSVNVRHNKFLAYLLGDLQVAY   344
WYSV    MLVFQHLELSGLQLYKMTMTLISHLNLISPAKFLSWVYDPLAEKPITQIYTIATTHDTRDRQDQKHWKYAKLARGQYWLDTTVKRNQFFAYVLADLEVRY   397
RYSV    MLVFQHLELSGLQLYKMAMTLIAHFKLIEPNKFLSWIYDPLSEASIDQIYKIAVNYDNVNSKTHKHWKYAKLARGQYWLNTTVKRNQFLAYILADLELKY   347
        **:********:*:***:****:*::**.* :**.*:  * :  .: :** *. .:*  : : :*********: :***...*::*:*:**:*.**:: *


AcNRV1  GLGGKTEYSDPRNMGALEGMPENQKATAIKVSDTIVKFYEAIEKQQSGGSGPAWRAAMGLAAA---------------------PMRAQDRQD-------   416
WYSV    GLGGKTEYSNPKRMKALDGTPVETRNDAESVAAAFQQMYRVIEDEKRREAGAAFRLARGMPPP----------PN-------PPPAAGGQGGG-------   473
RYSV    GLAGKSDYSSPKRMKALSGMPVERMTEAETISKAVEQMYTAIESAKRVDAGAAYRLAKKLGPPRANAHSRRKEPNNSRQHRDKQPNSKQQGRDKRNKHQV   447
        **.**::**.*:.* **.* * :    * .:: :. ::* .**. :   :*.*:* *  : ..                    *    :  .

AcNRV1  -PTPEEIRLQQEEAERARILEEERLRQERAANNQGGEGGGGGDVNMADPQ----------PDDRALLADMLGIQ   479
WYSV    VGAQGLGNVQAPPGAGGQVIPPGNLPPPPAAPPAGAAGQAPG---QQDQQINPPGQVPMDLDQRARDAAALGLV   544
RYSV    LGRHSKPQAQGPPNKQQDLGPHNNKPKEASRPPQQDRQQLAQPWRLTRRQRGARGHRTQTLSGMFCKGTCQCIQ   521
              . *       :    .    :                  *          .    .    :
```

(b) P

```
AcNRV1  MSGENNTQPKSNANVN-------TRSKTKHDKNAYDTAPPPPPIPPVTGRYSGFNPGSAQIISNDNRDSNAKTFETGAPPTLPPPPDQTLKRQEALAAPL    93
WYSV    MADSLDTDRGRTTRSGSKLKAQTTHRVGTSSGSVRGGKPYEKDFKAVEARFHNFDPISASIIRGDQEGTKADLIMSDSSVTKATAP------AEAAAPPP    94
RYSV    MSGSGSEQTPRLTRSSS-----RSTLTGVASGRVEKIRSSPKSLDRIAKKYKDFDPETAKIIRADLEETQADKTMEVGGSTQESA---------------    80
        *:.. . :    :. .      :      .  .    .    :  :  :: .*:* :*.**  * . ::*.     .  *  ..

AcNRV1  PANSSSEESGASSEERNQARQSPGDKRAKSASQSSKADISNKRFKMDTEMASLNLAFRDQEMLPMDKKDGASIVKTFCDTIKTQQKEITVADIVNIFAAG   193
WYSV    PPQPAAVPLTPPASATTQGQKRPGDPETLEEGNAAKRVQRAN---------KVNNLLTTQGIGEPSKSAISQYVVARLNANNVEHDDAMVAECANMAVYG   185
RYSV    --------------QQITGAKRPNEEDQGGAQEAAKRVNRSN---------KVNSLLTSNGVTDPAKSKISNYIVGRLNANNIEADSVMVAECTNIAIHA   157
                        . : *.:       :::*     :         .:*  :  : :    *.  :. :    :: : : ..  **: .*:   .

AcNRV1  YKLARKSIPTDILHQISGTLPELITELRTVSGSISLGVGNLDKSIHDMTTHMSNSFDLLLENDRIKNACETTLLATAQQIYQNSVKETKVDFINNYIQKS   293
WYSV    WKEGKKFVSTQILNQATTVIPDLITSMVTNANLITNAANALNSIPDKVAGAIRTNIEHVAFQTGSKATKRDTLVRTAESIYQNAAAESKVDFINNFIISA   285
RYSV    WKEGKKYLDDKIISQATTTIPTLITNLVSNANTLSNVIASLNNVPDKLVSDIRTQVENVSNQTGQKAAKRDVLLKSSESIYNNAVKESKVDFINNYLTSS   257
        :* .:* :  .*: * : .:* ***.: : :. ::     *:.  ..:.  : ...: :  :   * : . .*: :::.**:*:. *:*******:: .:

AcNRV1  KIDLSILKTNIDTYKAVYHAIIPEYTVLILSGEGESYPGLVSKIKDNMALIMNIARSSLPRSVYQATTSYAR   365
WYSV    GINIQEVKRNVANYRTIANAIIKRNTVLNIINETTEHQALLTQVQGNKDDIRNIARGLSASYVI--------   349
RYSV    GVNVDELRKDSHHYRTVVSRIEKKYTVLVMMPEHEEHHTLKEKVATNRTFVKESAQTLSQKYVTQ-------   322
         :::. :: :   *:::   *  . *** :  *  .:  *  ::  *   : : *:     *
```

(c) P3 (potentially overpredicted region is highlighted in yellow)

```
AcNRV1  ---------------------------------MEDSKYFEGPIKAGKITLLNGKASQINLKLRGSFFSNMSLIGTKKRQSWSVGDTT-------TITN    59
WYSV    MYSKFTYVSLLYLKTPTEIYSLHSRQPNSEERPKDRKTTSMASEAKNSQKFSFRNTEEEVDLSISKFALFKLKLKQSRVVTMFGNDDKVDPANCYINMTS   100
RYSV    ----------------------------------------MGEGKNHQSFSFKNADDEIDLSLSKFSLFKLKMAKSKIIKVFGQNDPVDPNNCYINMRS    59
                                                 .  *  :   :..  .:::*.:    : ::.:  ::    :. .* .       .: .

AcNRV1  VTLKVRSILHVLSDEYIKATITLDSSGKTTRVGWALWKSSFSGNIVFQTELGICKTHQGVYTPYSINIETSTGEELKGVSIDMKVVIREPGHNKLNSGEG   159
WYSV    IKIVTSSVLPESDPRFMVWEMSYKTDEENHKLGQLAWKASYNGSFVIKSTYAMMVVNGDLYTPYTASITTSDGSEIKGVKVNVILNWTPSDVRPSNAKMG   200
RYSV    IKITTSSVLPESDPKYLIWEMSYKTDEEDHTLGQLAWKASYNGTFIVTTTYAMMVTGGELYTPYTAVIRSSDGNEIKGVKVKVTLSWDPANDRPSKARMG   159
        :.: . *:*   . .::   :: .:. :   :*   **:*:.*.::. :  .:  .   :****:  * :* *.*:***.:.: :    .. .  ::  *

AcNRV1  SAILDVYS-LMTGDNSHYHPVLAWYVNKEGYAKQRILSTEPEAIMAGYIPVFEAFPTLQGQEPILHYENLLYFILEDEELKVFNNIIGDRRRGMSPRQSD   258
WYSV    GFMSDIYCNNPSNGKTQIPPMVAWYMGKDEQRYCKIINKLATELSSENIYPMMELISSPQTSLNPILNRLISSSLEKGDRDRISQMTKAVGGGMSLNKAD   300
RYSV    GFIQDMYCKTITNGKTQISPMVGWYIGQDERRYCKVLNKSALEFSSEGIYPLMELVSGADSVINPLINKLISGMLNDEEKRRVSLYTSTVGAGTSLTQSE   259
        . : *:*.   :..:::  *::.**:.::     :::.. .  : :  *  :  : :          :.*:   *:. :   ..        * *  :::

AcNRV1  IEHGMRILTSQKLTCLIEFLKRASPKIKGDVDFFHKV   295
WYSV    ISYLKGIIEKKSSSALVTFLMRASTELGVEVYIDG--   335
RYSV    KLLLKKLVESKTGSGLVQFLMRACKELGTDVYLEA--   294
              :: .:. : *: ** **. ::  :* :
```

(d) M

```
AcNRV1  M-TSEQEITIIRPTSQKSNAGIENDGVVVDSQYHLPPIVNYSLSFMIEISFYDQESCKMFLDKKMKTVELFGALKENWINSTGAVKYHTPVEPTDEQAHL    99
WYSV    MAHSKIRITGAGLDGVSYLERIRKSLSLNPADYHDDSAVRISLSFYIKISFHDVSDYDVFIREGVTTTELFDALKTSWKRNPENVRYIDGMTHTDDEIDT   100
RYSV    MANKKIRVTGAEAEQPSILKRISGALTLNPLDYHLDYSKLISLNFMVKISFHDASDYDLFVREGITPVELFEALASNWSTDSGEVHYVDGNTHDKDEIDT   100
        *  .: .:*       .    *     :   :**       **.* ::***:* .. .:*: : :...*** **  .*  ..  *:*       .:: .

AcNRV1  VIDICCLIAVLRELPFVEGIIQTHFHRNTTVLRMSFGNVLPQTVHTSIVPISSQLQLPLIFHGNLRYEFVEDTGKIHASVMMGIRATAFPKGNSPPVMDT   199
WYSV    TILLCELVTILKDLSLHRSDERTHYSLMSTSLTLGFGDQITQPHDNAVIPIVTTKVIPRYMHTVIQYEYPRVSGGIAASVCAGICIRSPPIGKCPPIMKT   200
RYSV    SVKLCELITIIKDLPFHKSEESTHFSILSTSLTLGFGDQILQKHDNSVIPIITERSLPQYMHAIIQYEYPRVSGGIAATICAGICIRSPPIGNCPPIMKP   200
         : :* *::::::*.: ..   **:   :* * :.**: : *  ..:::** :   :*  :*  ::**: . :* * *::  **   : * *:.**:*..

AcNRV1  VHLEHLLPHHGIREIQPQLPSSTSSNQAQDGDETGSTERRRKREGNDAGALVPSMKKAATSLANRLLRRT-   269
WYSV    VNVEVLAFHYGLDAGQGPQLEETNATPKEEELKKINGFKRMTTLNKSSRVIKKPSKEGVGAALSRMLSWK-   270
RYSV    LHLELLCYHYGLKMSG------DAPSPAEG---KIGRIKRPTERKEDTPSMTKRLKGGVGATISRMLSWKE   262
        :::* *  *:*:           ..  :    . .  :* .   :.:  :    * .. :  .*:*  .
```

(e) G

```
AcNRV1  MSYARVLVSLTTLIILTQVFLLLEEACSEISYWE----------KDGKWYAQWVPSPATSVEVPAPSPERIPAPPSPTLDKLPAVHQTTYG--TIHVG-V    87
WYSV    MRTTTRMLTIIICMLFGLYMIMLGENMILAFGPVDGDDD----SQENRPFTLHPAHGAIPVLPSSQPSQKVSPDDTNSLIRLTTSEQLWHN-HPQDIGPK    95
RYSV    MNIIKRAGVFIGLVYFTVLLIIYGEHIVRSYATEDCMPAGLYYYAEGAAYTNLPVCKDEPKSSYKGTELFPTSPNIDKPLRMTTDIMVHPISSTSDLGPM   100
        *        :   : :   :::  *                    :.  ::        .      .    ..    .  ::.:         . .:*

AcNRV1  PTDTDLYPIYYCPNVSIYDLLPVWYSGCLSMCSITSVQDLVDISVFQINSTLGSTDLYKVSVRSVTKYSHVSVFGSCTVSVTADDYVSPPSDKLSEWVTK   187
WYSV    DVPADMYPIYSCPNLSNAYLLPIWYGSCLDACQITTPKHTVNVKLWTINSSVTDVDGYQIDVYYDTKFSHVGPFGGCSVSLSESVTKEPKQEDIMIWKSR   195
RYSV    DVPPDMYPMYSCPNLSNSSPLPLWYGSCINSCQITSSKETMAVTVWAINSTVDIIMGYKLTTFFETKFSHVGPFGGCSISLVGSEPRSAPEALISGWRSK   200
         . .*:**:* ***:*    **:**..*:. *.**: :. : :.:: ***::     *:: .   **:***. **.*::*:  .   .. .  :  * ::

AcNRV1  LSIDRPSGSQEWDYSSDPPCNYFSDEYSTGYKLYVTHYKVEISVDSAQNRYLRDPTNGLYMSYEQDYYVDGSTFYHWVQNSDVSQEVCYFREIGEDSCTV   287
WYSV    LVSKPVNDVESWIMYDEPSCNYFSDEYSSGFRLVITRTKLKLMIDSVGNLYIADLRPGSYDTYKTGYAIHGSTAWIWDTDDSMNHGMCYFKQTDDTYCDY   295
RYSV    ILSQGNIQDAAWYLYNDPECNYFSDIYSSGFFVRIDRIKLTVLIDAVGNLYLSDLVAGSYDSYDKGFATHGNNVWIWDTDRVKVDNNCYFKQTDDTYCDY   300
        :  .       *   .:* ****** **:*: : : : *: : :*:. * *: *   * * :*. .:  .*.. : *  :    .  ***:: .:  *
```

```
AcNRV1  DSSSGYISCRVLGITFKKDIKNIITSSCAGSLNVSSDGVIYQVLSAYNKTSLSDRLSGLLSS--NINMNQLIHVINDALISIEDAYCSSTCDVMEIILSN   385
WYSV    DNNTKYMFCKMSGVSFDTTVQQRITSSCAGDLNISTDGVIYQIGDSGDTASTQQRLSDILHQNVELGMQSLVSLINDVFINIESSYCTGVCDIMEVIVSN   395
RYSV    DNTTGYILCKTIGVSFRSELQTRVESPCAGHLNISTDGVIYRLEGTADTPSTQDRLSSILKNNVNLGMESLISLINDVFTNIESSYCTGMCDLMEVILSN   400
        *..: *: *:  *::* . ::  : *.*** **:*:*****:: .: :..* .:***.:* .  ::.*:.*: :***.: .**.:**:. **:**:*:**

AcNRV1  YPTATTVIETPIGPWLPLVAKGNTVMTACSADTEWYLKLPIEYCAEKNMMKVVHNSTKSEEWWRIMNTYVVSGESCNSSDSGWDEILVEKMQSKREIVFH   485
WYSV    YPTATTVLETPIGPWLPITSDGHTIMTPCMADVNWIIQTPIVYCFSKEMIKVINKDTRKEAWWRIVNSYIILNETCSDTNSTALEILRDRMSKRRDIVYS   495
RYSV    YPTATTVLETPIGPWLPITTNGHTVMTACSSDPNWVIKTPVSYCYSKKMIKVVNKDTMREAWWRIVNSYIILNETCTDENTTSFNLVKERMELRQDITYS   500
        *******:*********:.:.*:*:**.* :* :* :: *: ** .*:*:**:::.*  * ****:*:*:: .*:*.. ::   ::: ::*. :::*.:

AcNRV1  FWRGDLTLVYPYNVSHWTLYKDEKIQRSSKWFPKLQYMMLDKEIYIDNVTQALHNYTQDIYNHYMGNQTASTGLGSIIPRILDRITLSVGNAVKLVVKIS   585
WYSV    FWRGDLIVSYPYNKSRWITYKDEKIQRSSKWFDKLVDLKYKHPITLDNITSQLVNHTADLYEWHMGDKNGTAGQ-TTFSDLLGRVEKAGTNVIKGCVKMT   594
RYSV    FWRGDLIVSYPYNKSRWITYKDEKIQRSSKWFDKLLPLRYNHPITLDNITMELINHTRDIYDVHVYPSQGSTSK-RPMSDVIGRVGAAGSNLIKHLVKGI   599
        ****** : **** *:*  ************* **  :  .: * :**:*  * *:* *:*: ::  . .::.    :. ::.*:  :  * :*  **

AcNRV1  GEVILFIMSHIHLIGDIIITVVCMVIGYYLIYIPTRWVRKRRGD-TSARITAVSNSYEMVPSSRTTRKDIL   655
WYSV    GNLLIWITSHIEMIGDMLIIIVCLIGGYYVLIIPYGFLRRGRGPGTVTEVQQSNSQFRSPLIPRTYL----   661
RYSV    GETFFWVTQHIELICDILIIIVCVVIGYCVVLKPILYIT-MRGREKQQPVIVMRDEGRSHLFSRSTATNAL   669
        *: :::: .**.:* *::* :**:: ** ::  *  ::   **  .   :    .. .    .*:
```

(f) L

```
AcNRV1  --------MSDRDSDDDGADYSGRYEEDDEDSNYKRHGEKADMVEGGDYHLKSALKCEADLEMDHVYLKEKNAMMAELGDLGDLIYPHSVFQFIEKISHM    92
WYSV    MDVEEPTYWAHDEDDDYWWDADEYLGEEEEDEIYEDTED--ELVEGGDFHLKSALRGEEDMLQNPIYKKERDSMVEELGSLGTLMNHIDVVSMLNLIGTR    98
RYSV    MDDEGHGYWQDYDEDESWLDAEN---DVFDDDIFEEAEDNEHLVEGGDFHLKSALRGEADMLTNPIYEKEREKLQEDVGDLGVALGHLNVRPFLEKMGER    97
                  . :.*:   * .    :  :*. ::   :  .:*****:******: * *:  : :* **:: :  ::*.**  :   .*  ::: :.

AcNRV1  WSTDR-ASGLSRWVTVNPDSRLD-MSNTKRLLGAEISLVCPDVNPTDVGGWLTSGYDAFAARFSHLEHTMKILLILLPILNNIKNLRNGMGVDNAINEYI   190
WYSV    VAQDKRPSGGHTFLHEGGFNIIP-LQENVKLIKAELMSVLPEMAGMNIDQVLSGSYHLFMADHPYVTCTISTVLFILSVLNNVKNIRQGIEVGTLIGRIL   197
RYSV    ISTDHTNTPLIENIKQGGFGAIHHMKSTARLIVAEAATVTETMLTSGVDTTLAGAYHFFEQNHPHMKCTVNMVLMFLTILNNVKNIRLNVDLHQILKNNI   197
         : *:  :     :  .  . :  :... :*: **   *   :   .:.  *:..*. *   ..::  *:. :*::*.:***:**:* .: :   : . :

AcNRV1  TINSGLVSVDMNSVSVVRMNTDIVVFDCPDERYYLPKDYFLNMCDKLQERVNIYLYTEVCSQTWVPLAPKIELVSRVIAWGDEVLKKLGNDGYRVIGIYE   290
WYSV    YRKGNLVCLSLAAATLCYLSTDILVFDVDGSRHYMPKTYFLNGCDKVQERFNIMLYSYMAESLGVPGSCPSQIVKRTISWGDNVLANMGNEGYNVIGLYE   297
RYSV    SIKGSTVCMYITVATVAYFSTDIVVFDIDGQKYNTPKTYFLNACDKIQERFNVILYSYLAEGLSIPGSPPTYIVNRIIDWGDSILYQMGNDGYDVIALYE   297
          :.. *.: :  .::  :.***:***  ..::  ** **** ***:***.*: **: :..   :* :    :*.* * ***.:* ::**:** **.:**

AcNRV1  ATLVGIILNRDDPSLITNPDRPGFLEALKEDLGGVEIRFIDQLVNMLSTVDVRTIIDLHGLYRIWGHPIIDIDGGVKKLQKVTQEKKEVDPDSEESINAR   390
WYSV    AILVGMILDRDDEALSPTPRSESFLSNILTGLSSEEQGYAKRLITTLQDMSPAQLADLHGLYRIWGHPIIDIDGGVRKLQRVTREEKGDINAKPESRETV   397
RYSV    ATVVGVILSRDDKDLSPTSGGGDFLENIQQDLGPVQQNHIRYLIALLHDMTPHQLADLHGLYRIWAHPIIDIDGGVKKLQKVTQSLKGDINSKPESQETV   397
        * :**:**.***  * ...   .**. :  .*.  :  .   *:  *  :    : *********.**********:***:**:. *     . ** ::

AcNRV1  RSFRKIFCKNYFKKNGFYPPLKLIDRTTPCYLSRCLLEGVEVMENHMDYKFQQWDNVALQKAFSIPYSWNIVHLAKDKSISPTRSELYSLLSKTGNIFSA   490
WYSV    RSFRRLFTTEYFKKHSIYPPMEITSKFN-TYLGNCIRTGKEIDEKHINYQFSDWDYIELQETFSIPYSWNVLHLAKDKAISPTRNEIYNMLLSKGRIFNA   496
RYSV    RSFRRLFVMDYFVKHQFYPPITLPEKSN-CYIGNCIRTSKKIDESHINYNFSDWDLVELGGAFSIPYSWNVLHLAKDKAISPTRSEMYTMLCKTKRVFNA   496
        ****::*  :** *: :***: : .: .  *:..*:  . :: *.*::*:*.:** : *  :********::******:*****.*:*.:* .. .:*.*

AcNRV1  NLRRGILKLLDSKLIPLRDFLTKVAAEGLHVDDCIIGLFPKEREIKILARFFSLMSFNMRLYFTSTESLLGDKVLKYFPQITMNLNLLDMQEKMSSMSKS   590
WYSV    ELRRGVLKLMTTTLIPLRDFLKKVASDGLDIDDCIIGLFPKERELKILARFFALLSFNMRLYFTSTEELLGSKLLKYFPQITMSSNLLEMQEKMASMSKE   596
RYSV    ELRRGVLKLMNTTLTPLREFLTEVAEHGLDIDDCIIGLFPKERELKIMARFFALLSFKMRLYFTATEELLGSKLLRYFPQITMSSNLLDMQEKMSSMSRD   596
        :****:***: :.* ***:**.:** .**.:**************:**:****:*:**:******:**.***.*:*:*******. ***:*****:***:.

AcNRV1  LKEQSDSVTYVVNMDFVKWNQQMREEICRSVFAELDAIFGLPGLYTRSHQIFKQSILYLADGTRELHSDPHTGVHVDGKYCWRDDGAGKEGIRQKAWTIM   690
WYSV    LRTQNRSVTYVINMDFVKWNQQMRESTCRGVFTELDKLFGLKGLYTRSHQIFKDSVLYIADGTRRILPDPISGVMIDDNACWTDDGAGKEGIRQKAWTIM   696
RYSV    LESQNKSVTYVINMDFVKWNQQMRESTCEGFLKNWSKLFGLPGLYSRSHQIFRDSILYIADGTRDLTPDPDTGIMVDNNVCWIDDGAGKEGIRQKAWTIM   696
        *. *. *****:**************. *...: : . :*** ***:******::*:**:***** : .** :*: :*.: ** ****************
```

```
AcNRV1  TVCDIETVAQGHTGQFQLVGGGDNQVLTVTYKTNNIDGQGDITEEGKAKIKRKMENFLADLERHFQRRGLPLKTSETWCSTSLFMYNKHIYYKGAATRAP   790
WYSV    TVCDIAAVARHHPGKFHLVGGGDNQVLTITYHTNQVDTQGVITEAGKQKIKSKVKRFLTDLEKHFSERGLPLKTTETWCSTSLFMYNKYMYYQGSPLRSP   796
RYSV    TVCDIAAVARHHPGVFHLVGGGDNQVLTVTYHTNQIDTDGNITEEGKSKIKAKVKKFIEALETHFAARGLPLKTSETWCSTSLFMYNKFMYYKGVPLRSP   796
        ***** :**: *.* *:**********:**:**::* :* *** ** *** *::.*:  ** **  *******:*************.:**:* . *:*

AcNRV1  LKQVSRMFPYSNNTSMSISSMSQCLGTTLRSICQKDYVPFPALFMRNYWAQLLCWIISYCHPMMMSIS-KNIKQGVGGICRRGRRLEVPVEGSCSKHLML   889
WYSV    LKQVSRLFPFSNNTSMTLQSMAQCLGTGLRSVAQKELSHIPALMMRNIWGSILTWIVTVCHPMLLSISNQDSLKGDGIITRGKKEIKMKVESVESAPLAL   896
RYSV    LKQVSRLFPYSNNTSMTLQSMAQCLGTGLRSAAQKEVSHIALLFMRNIWGSVLGWIILYCHPMLPSISSETCNSGVSTIVRGRKAINMKTHRIDLVALIL   896
        ******:**:******::.**:***** *** .**:   :. *:*** *..:* **:  ****: *** :   .* . * *  : ::: ..      * *

AcNRV1  RLAYLPGHFGGPGMINIYQAMMRGFPDPITEGICFLRNMTQALKRRDDILYSFFRDLSGVSFSLSRSMEPLIEDVTSLNNDAPRTGASEHREAAKEILTK   989
WYSV    KIIYLPGHFGGPGLVNLLQMTMRGFPDPITEGIYFLMAMKKNTARLGSAYTSLFQRMAGVSFSRSRNYEPLVEDVCSLNTDAPRSGTSEQREVARKILLK   996
RYSV    KILYLPGQLGGPGLVNIYQMTMRGFPDPVTEAICFLRKFKNYLICVGSSYSSHLARMAGVSFSASRSYEPLIEDVCSLNLDTPRSGTGGKREVPRKILLK   996
        :: ****::****::*: *  *******:**.* **  :.:     ..   * :  ::***** **. ***:*** *** *:**:*:. :**..::** *

AcNRV1  SKLGKNLELKSLISIMSGDLETQFYRSLASGPSLDVKVLHEIASATLYSITNTLTSRIDKTSTLSRLTHKISLLKKMALAEESYLKYLMIRDSRKHDLSF  1089
WYSV    SKLGGNENLRDLLRIMEGDNEKNFYKALTASRVLDIRVLHEIASATLYAITNTFTSRVDRTATLKRLTLRYSMIKSLAESERKFLRYLLVRDLKQHDIMF  1096
RYSV    SRLGGNQHLKELLGIMKGPSEGEFYRAISSGKVLDVRVMHEITSSTLYAVTNTFTSRVDKTATLKRLTFKFSMLESLADAELKYIRYLSVRDDKTHDIMF  1096
        *:** * .*:.*: **.*  * :**:::::.  **::*:***:*:***::***:***:*:*:**.*** : *:::.:* :* .:::** :** : **: *

AcNRV1  DACSRVLADKARSTSWRKQIIGVTVPTPFEYLVCKSGQDHVCNGDAIIARVDEV-DRTEMMTELGPCKVYLGTYTREKFKMTEIAAAYGEEDILAKSIQL  1188
WYSV    EKCSRVTADECRTLGWGKQIIGVTVATPFEYLSVSLREKHVCDGNNIIVRLSSSGNKKQLGEVLGPCKPYLGTYTKEKFKMTEVAAAYGDEDVLTKALRI  1196
RYSV    DGCSRIIADECRTKGWGKPVLGVTVPTPFEYLQISWTDEHICDNNHITVRISNQ-DRQVTTETLGPCKPYLGAYTKEKFKMTEVAAAYGDEDVLSKSLRI  1195
        : ***: **:.*: .* * ::****.******  .  :.*:*:.: * .*:..  ::      ***** ***:**:*******:*****:**:*:*::::

AcNRV1  MKLINWRYPEGSQMAEILKCPLRAVTDLNPELLIQNEVLTKGDYDHRRKLDSRVHGGIPNFVTTSTSHTSLCTSTWVSHARGGKNENIHFQACLISSQFR  1288
WYSV    MKIINWRFPEGSTMSEILKAPFRAVTDIDPQRMIQESSVTKGDYDHRRKMDSRVHGGIPNFVITPLSHMSICTSTWYKHARGGKNENIHFQACIIQTMYK  1296
RYSV    LKIINWRYQDGSTMSEIIKAPFRAVTDIDPQRMVQESTVTKGDYDHRRKMDARVHGGIPNFVTTPLSHISISTSTWYKHARG-KNENIHFQACIIQTMYQ  1294
        :*:****: :** *:**:*.*:*****::*: ::*:. :**********:*:********** *. ** *:.**** .**** **********:*.: ::

AcNRV1  ILCMRYWGILTSRVLHFHESCQTCIQALEQP--APGTTTPSIIPAFPTLRANRLVFVPSQEVEFDFEDQERVTRAQFLGLVDCEKKNVYDWADLYSSLAW  1386
WYSV    IVMLLMNNGKCDEIIHAHESCSTCICEITEP--VLDLVTPMMDLIFPQLQNNSLVYIPEQAISFDHSRMKEVDYARRVGMLDCHRDTDYTWLDTYSSLSW  1394
RYSV    IIIRTMSNIHSQEKLHVHECCQTSISEIQEPRANLDLTTPQMIPTFPTLMGNPLVYIPEQSITFDYSRTQEVDYSRRVGSLDCQRSTDYGWVDAYSSLSW  1394
        *:     .   .. :* **.*.*.*  : :*    . .** :   ** *  * **::*.* : **..  :.*  :: :* :**.:.. * * * ****:*

AcNRV1  TIICNMLGITSMSPSFYLIMRDNVNHHCLGLYLAMFWRCLRKLGALESTPNINWSPMIYTFSCEEGMRVLQEEMGLVITGSVEGGFVMGTGDSCDTLDLL  1486
WYSV    LIMCDVIGWTRMPHSFYFMIQNDVDHYLLSMYLICLWKVMRKEFELINT-QLDWEPMIKVYSTQEGITVLSNQMGLVVGGSLEGGLEVDMKDICSAFNTL  1493
RYSV    LIMCDVIGWTKMPDSFYIMQQERVDHYLLSMYIISIWKVLKSEFDLHKT-MLDWGPLLKVYSSDTGVQVLSQSMGIVVLGKLEGGLSVNMMDICNSLTGL  1493
         *:*:::* * *. ***:: :: *:*: *.:*:  :*: ::.   * .*  ::* *:: .:* : *: **.:.**:*: *.:***: :.  * *.::  *

AcNRV1  LIDEMPPMTCNIFCPMIKWQVCVWWYSQSSFVSGCILCMRNIKSMWMSSNMLHRVDEDLRCSSHMTNEMKPVMINAHIENLRQYGKYSTEEPHTDIDTVC  1586
WYSV    TITQIPPFSVNLALPKIHKQVAAWWMLMDDNTLLCIDCHNNLNGYWDGTSINRRVDNDLRCYIHADGDVSPRIINAHISNLSQG----LIRLTPGGTSQS  1589
RYSV    TSNQIPPFSVNLALPRIHKQVATWWKLTCSDTMYCIQCSNILKGYWEGTLINTRIDNDLRCSECADGGISPRIVNAHISNLSDH----LIKIIPDREEVQ  1589
           ::**:: *:  * *: **..**    . .  ** * . ::. * .: :  *:*:****     . :.* ::****.** :       .  ..

AcNRV1  LPISKDLVEEEVGMIRRMISLPSEYPDEGSVPPVIRYGGDVDNVTLFRMILESLSLLNPVHIVLGPDINMLRNVVAGCSLTSLEMEIDTLMIDIVISALS  1686
WYSV    HPIIIG--GDKLNNLETVPSVEDEWPDVESITLLNQIGETNDIDGRLRFIIESMSVIMPDIVVVRPDMLDITLVKTARELLKCDDGSK-LIIYILVEEPS  1686
RYSV    LPLLAG--GEDLNNLEVVLSLEDEWPDSDDITILNRIGEMNDIDGRMKNYLEVMALLNPDVVIIRPEILELEILRRVCEVVRLASEGK-LIIHILVEEPS  1686
         *:  .   :.:. :. : *: .*:**  .:. : : *   *    ::  :* :::: *  ::: *::  :  :    .:       . *:* *::.  *

AcNRV1  YQESDTFYELKAVVPKRMSRRIRFCMNAPQEGFCRFFIKPDRKLLKYLIAGDWVLCDCHELIEMNNTLKPDEVYYLPCTPLRW-LLLVKCGTGVFGPGDL  1785
WYSV    YLESGTIYEMERAFPNSETVDVQFDYRKAPGGLHKLWINPREESIQRVEVGDWICISFTDLIKHHDSILADTVLTMKERRLGR-QFLVNKGVGSFKPGDL  1785
RYSV    YLESGAIYEMERAFPRHNLFHVQFSFRKPPEGLHKLWLLPQDITLARADVGDWLVIPFTQLLDFHDTITPDTRLHHQREAIRQDVFWSTKEQGGLGLVVL  1786
        * **.::**:: ..*.     ::*  . .  *: :::: *    :    .***:     :*:. :::: .*        :    :  .   * :    *
```

```
AcNRV1  IETLSKPHYEP---RFWKIDHCIISAFSCGKLSGGGFTELRRKLSWSLFSNKYELLTICHRVTNQMNRSKAYGGLKQWSRNYREGIIVFILSVLMRSENH  1882
WYSV    GKYIFSSELNMSYKQSRQVDYEINKVMIQNPTLGGKFMILKRRCSWALFSSRYEMLRAVERIKKNMTVEGRSTKTKEWSKTHQDDVLILIISMLMRADEK  1885
RYSV    QHICTKTELSYRLDRSK-VDVEINKMMIKDPVMGGKYVILKRRCSWAIFSTKYELLCSAERIKGNLSADGKKITSKDFSRTYKEDIIIFIISVLMRSLDC  1885
         .   ... .    :   :*  * . :  .   ** :  *:*: **::**.:**:*   .*:. ::. .      *::*:.:::.::::*:*:***: :

AcNRV1  DLHLITTLSYAEMDLTHLTITPKFNKAGVVSIRYLPYIWFFKRYYDTKRCHVDIEFATAIKSMNLGSPRGSGLNIIML----  1960
WYSV    DESEIFTLGYVNCDVENLVMYPKFSSTQVVSLRYLDYFWFFKRMYEYQEKTFAIPYTEIRTSLNIGSPKRSGYNTINSSE--  1965
RYSV    DEKRVMTLSYVECLPELLTLKPRFNKSGVISVRYLDYFWFFKRVYSYNNPEVAIPYSAVIKGVNLGSPRKEGYSLASSDGAP  1967
        *   : **.*.:     *.: *:*..: *:*:*** *:***** *. :.  . * ::   ..:*:***: .* .
```