Studies on Two Alternaviruses, which Proposed to be a New Mycoviral Family—Alternaviridae, and Identification of a Novel Deltaflexivirus

2022.9

Department of Biological Production Science
United Graduate School of Agriculture Science
Tokyo University of Agriculture and Technology

CHIEN-FU WU
(吳建甫)

Abstract

Mycovirus was first discovered on cultivated mushrooms in 1962. Since then, more and more fungal viruses have been reported. In addition to infecting fungi, mycovirus-related viruses have been found on insects, plants, and oomycetes. The types of the genome of mycoviruses are double-stranded RNA, single-stranded RNA, circular single/double-stranded DNA, and single-stranded RNA reverse-transcribing.

The family name Alternaviridae was proposed in 2013 by a research group in the UK based on the paper by Aoki et al. (2009). Nowadays, we know that the fungal hosts of alternaviruses are *Alternaria* sp, *Aspergillus* spp. *Fusarium* spp., and *Diaporthe* sp. There are ten alternaviruses reported now. Alternaria alternata virus 1 (AaV1) was identified in the saprophytic fungus *A. alternata* strain EGS 35-193 (Aoki et al., 2009). AaV1 has four genomic double-stranded (ds) RNA segments (dsRNA1–4) packaged in isometric particles. The 3′ end of each positive strand within the particles is polyadenylated (36–50 nt), but the presence of a cap structure at each 5′ end was unknown.

In this study, I investigated that the dsRNA segments of alternaviruses have unique properties of terminal structures. The protein composition of the viral particle of alternaviruses and the purified viral particles of deltaflexivirus were also determined. It is expected to provide taxonomic criteria for virus taxonomy and to expand new knowledge in virology and molecular biology. Firstly, I characterized the AaV1 genome and found that it has unique features among the mycoviruses. The existence of cap structures at the 5′ ends of the AaV1 genomic dsRNAs was confirmed using RNA dot blots with anti-cap antibodies. Polyclonal antibodies against purified AaV1 particles specifically bound to an 82 kDa protein, suggesting that this protein is the major capsid component. Subsequent analysis of LC-MS/MS indicated that the AaV1 dsRNA3 segment encodes the major coat protein. And the smaller size proteins of empty AaV1 particles are also related to dsRNA3-encoded protein.

I also investigated the two kinds of defective AaV1 dsRNA2, which is 2,794 bp (844 aa) in length when intact, appeared in EGS 35-193 during subculturing, as confirmed by RT-PCR and northern hybridization. Sequence analysis revealed that one of the two defective dsRNA2s contained a 231 bp deletion, while the other carried both the 231 bp deletion and an additional 465 bp deletion in the open reading frame (ORF). Both deletions occurred in-frame, resulting in 767 aa and 612 aa hypothetical proteins. The fungal isolates carrying virions with the defective dsRNA2s showed impaired growth and abnormal pigmentation. To the best of my knowledge, AaV1 is the first dsRNA virus that has both 5′cap and 3' poly (A) tail which is accompanied with complementary poly (U) structure at the ends of genomic segments and some isolates have defective dsRNA2s.

Secondly, I identified a novel dsRNA mycovirus—Diaporthe alternavirus 1 (DAV1) from the *Diaporthe* aff. *acuta* strain IbSTRPmp18001, isolated from a crown rot of strawberry. The isometric virions (ca. 35-40 nm in diameter, buoyant density: 1.349-1.374 g/cm^3) of DAV1 consist of four double-stranded RNA (dsRNA) segments and 79.3 kDa coat proteins. These four dsRNA segments are dsRNA1 (3.7 kbp), dsRNA2 (2.7 kbp), dsRNA3 (2.5 kbp), and dsRNA4 (1.7 kbp); each segment has a 5' cap structure and a 3′ poly (A: U) structure. Each of the four-dsRNA segments of DAV1 has a single ORF, dsRNA1 encodes an RNA-dependent RNA polymerase (RdRp), like other alternaviruses, the glycine residue is replaced by an alanine in the most conserved GDD motif. dsRNA3 encodes the coat protein. dsRNA2- and dsRNA4-encoded proteins are hypothetical proteins. The phylogenetic analysis of the amino acid sequence of RdRp indicates that DAV1 is classified as a member of Alternaviridae but has only about 30-40% sequence identity with other alternaviruses. A virus-free strain was obtained during subculture on the media. Comparisons of the hyphal morphologies and fungal growth between the

DAV1-infected and the DAV1-free strains showed no significant differences, suggesting that DAV1 latently infects the host fungus. DAV1 is the first alternavirus found in *Diaporthe* sp. These results of alternaviruses may contribute to the fundamental molecular and biochemical characteristics and classification of the proposed Alternaviridae.

Finally, I also identified a novel deltaflexivirus which was isolated from *Fusarium oxysporum* f. sp. *melonis*. The full-length viral genome was sequenced and the phylogenetic analysis was performed based on the replication proteins of the viruses. This novel virus is a positive ssRNA virus (8125 nt in length with five ORFs), classified into *Deltaflexiviridae*, named Fusarium deltaflexivirus 2 (FDFV2). In the purified virus suspension, isometric and filamentous particles were observed and were associated with the FDFV2 ORF4-encoded 18 kDa protein, which should be the major viral protein. To the best of my knowledge, this is the first report of particle and protein analysis observed in association with the deltaflexivirus. The results of 5’ RACE showed that FDFV2 might have subgenomic RNAs to express the viral proteins. I will perform northern hybridization with riboprobes to confirm the organization of the subgenomes.

Table of Contents

List of Tables

List of Figures

1. Chapter-1 General Introduction

1.1 Mycoviruses

Mycoviruses are viruses that can infect fungi. So far, mycoviruses are ubiquitous in most of the fungal genera. In 1962, the first mycovirus was found in the cultivated mushrooms (Hollings, 1962), then nowadays, the number of reports for mycoviruses has been increasing (Ghabrial and Suzuki, 2009; Ghabrial *et al*., 2015). Viruses related to mycoviruses also have been isolated from insects (Koyama *et al*., 2015; Liu *et al*., 2016), plants (Li *et al*, 2009, Nibert *et al*., 2014), and other lower eukaryotic organisms, such as oomycetes (Hacker *et al*., 2005; Cai and Hillman, 2013; Uchida *et al*., 2021). According to the International Committee on Taxonomy of Viruses (https://ictv.global/taxonomy/), most of the common mycoviruses are identified as linear double-stranded (ds) RNA viruses: *Chrysoviridae, Megabirnaviridae, Partitiviridae, Quadriviridae, Reoviridae, Totiviridae, Polymycoviridae* followed by linear positive/negative single-stranded (+/-ss) RNA viruses: *Alphaflexiviridae, Barnaviridae, Botourimiaviridae, Deltaflexiviridae, Endornaviridae, Gammaflexiviridae, Hypoviridae, Narnaviridae, Mymonaviridae, Yadokariviridae, Hadakaviridae,* circular single/double-stranded (ss/ds) DNA viruses: *Genomoviridae,* Rhizidiovirus, and single-stranded RNA reverse-transcribing (ssRNA RT) viruses: *Metaviridae, Pseudoviridae.* There are also the proposals of new mycoviral families: Alternaviridae, Fusariviridae, Mycomonoviridae (Kozlakidis *et al*., 2013; Zhang *et al*., 2014; Nerva *et al*., 2019).

1.2 Presence of the 5′ cap and 3′ poly (A) tail in dsRNA mycoviruses

Most of the mycoviral families are reported to have no 3′ poly (A) tails and 5′ cap structures, although positive ssRNA genomes of *Alphaflexiviridae* and *Pseudoviridae* have both 5′ cap structures and 3′ poly (A) tails (Lefkowitz *et al*., 2018). *Metaviridae* and *Hypoviridae* have only 3′ poly (A) tails (Lefkowitz *et al*., 2018; Li et al., 2015), whereas

Reoviridae has only 5′ cap, but no 3′ poly (A) tails (Furuichi *et al*., 1976). In this study, I reported that Alternaria alternata virus 1 (AaV1) (Wu *et al*., 2021) and Diaporthe alternavirus 1 (DAV1) having both 5′ cap and 3′ poly (A: U) tail. Fusarium solani alternavirus 1 (FsAV1) infecting *Fusarium solani,* also has been proved the presence of 5′ cap / 3′ poly (A: U) structure in its dsRNA genome (Lutz *et al*., 2022).

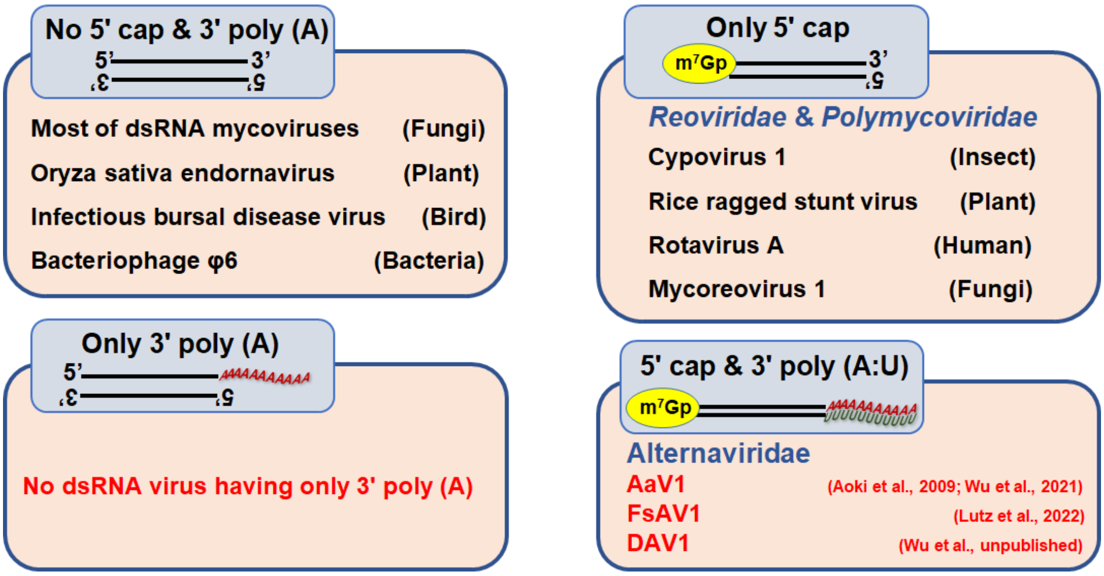


Figure 1-1. Presence of 5′ cap and 3′ poly (A) tail of dsRNA virus genomes.

1.3 Genome alteration of viruses

Rearrangements of dsRNA genomes or deletion events were also observed in the major genera of *Reoviridae* (Desselberger, 1996). The inducible genomes rearrangement reported in mycoreovirus (Sun and Suzuki, 2008; Eusebio-Cope *et al*., 2010; Kanematsu *et al*., 2014), although the mechanisms and significances of genome alterations remain to be further investigated. Defective interfering (DI) RNAs of Saccharomyces cerevisiae L-A virus (ScV-L-A), X dsRNA, automatically replicate in yeast *Saccharomyces cerevisiae* depending on RNA-dependent RNA polymerase (RdRp) of ScV-L-A, which causes lower the copy number of ScV-L-A according to the “head-full replication model” (Esteban and Wickner, 1988; Wickner *et al*, 2013).

1.4 Affect of mycoviruses on fungal hosts

Many mycoviruses are latent in their hosts, but a growing number of reports indicate that mycoviruses can be causal agents to modulate traits of host fungi (Tavantzis *et al.*, 2002). Cryphonectria parasitica hypovirus 1 (CHV1) is the most well-known mycovirus, which leads to the reduction in both pigmentation and sporulation of the host, and attenuates the host virulence (McCabe and Van Alfen, 2002; Nuss, 2005). In *Rosellinia necatrix*, white root rot fungus, there are at least five virus families, which are also well studied especially for virus-host and virus-virus interactions (Kondo *et al.*, 2013). Magnaporthe oryzae chrysovirus 1-D (MoCV1-D) causes the morphology changing and abnormal pigmentation of the host by decreasing the accumulation of melanin biosynthesis intermediate scylatone (Higashiura et al., 2019). The high-level RNA accumulation of Alternaria alternata chrysovirus 1 (AaCV1) in the host, results in down-regulation of host growth, and up-regulation of host virulence through increasing AK-toxin production during spore germination (Okada *et al.*, 2018). A unique ssDNA mycovirus, Sclerotinia sclerotiorum hypovirulence-associated DNA virus (SsHADV-1), contributes hypovirulence to the plant pathogenic fungus and has the potential to be a biological control agent of plant fungal diseases (Yu *et al.*, 2010).

1.5 Objectives of this study

The objectives of this study are to further invest the molecular characteristics of two alternaviruses of the proposed Alternaviridae and to identify a novel deltaflexivirus. The major focus is on the properties of dsRNA molecules of alternaviruses and the composition of viral particle-associated proteins. In addition, the presence of virus particles of deltaflexivirus was investigated. It is expected that this study will expand the unknown fields of virology and molecular biology.

2. Chapter 2 - Molecular Characterization of Alternaria Alternata Virus 1 (AaV1)

2.1 Introduction

2.1.1 *Alternaria alternata*

Alternaria alternata is a fungus belonging to Ascomycota. *A. alternata* is a common plant pathogen with a wide range of hosts. According to the USDA National Fungus Collections Fungal Database (https://nt.ars-grin.gov/fungaldatabases/), there are currently about 780 plant species recorded to be infected by *A. alternata.*

2.1.2 *Alternaria* spp.-infecting viruses

So far, many mycoviruses that can infect *Alternaria* spp. have been reported, such as *Endoraviridae, Totiviridae, Mitoviridae, Partitiviridae, Chrysoviridae, Fusariviridae,* and Botybirnavirus (Shang *et al*., 2015; Komatsu *et al*., 2016; Chen *et al*., 2017; Xiang *et al*., 2017; Okada *et al*., 2018; Xavier *et al*., 2018; Shamsi *et al*., 2019). The AaV1 in this study was first reported by our laboratory in 2009 and classified into proposed Alternaviridae.

2.2.3 Proposed Alternaviridae and Alternaria alternata virus 1 (AaV1)

A new viral family, Alternaviridae, was proposed in 2013 (Kozlakidis *et al*., 2013) and currently accommodates ten species, including Alternaria alternata virus 1 (AaV1) (Table2-1). AaV1 was the first alternavirus to be completely sequenced. It has four genomic dsRNA segments (dsRNA1–4; 3.6–1.4 kbp in size), packaged in an isometric virion about 33 nm in diameter. The dsRNA1 encodes a protein with the conserved motifs of an RNA-dependent RNA polymerase (RdRp), however, the glycine residue is replaced by an alanine in the most conserved GDD motif (Aoki *et al*., 2009, Moriyama *et al*., 2021). A property of AaV1, the presence of intact poly (A: U) tails (36 to 50 nt) at the 3′ terminal regions of all four dsRNA molecules, was described in Aoki's doctoral dissertation (Aoki, 2009). The presence of capping structures at the 5′ ends has not previously been investigated (Aoki *et al*., 2009). AaV1 infection leads to phenotypic alterations in the saprophytic fungus *Alternaria alternata* (strain EGS 35-193), including irregular pigmentation, decreased mycelial growth, collapsed aerial hyphae, and cytolysis in the hyphae (Aoki *et al*., 2009).

In this chapter, the results show that AaV1 has a 7-methylguanosine (m7G)-cap structure on the 5′ end of each dsRNA segment. Additionally, spontaneous in-frame deletions have been detected in the AaV1 genome, and these may be related to the impaired growth of the fungal host. The main results of this study were published in *Frontiers in Microbiology* in October 2021 (Wu *et al*., 2021). The proposal of establishing Alternaviridae has been submitted to ICTV at the end of May, 2022.

2.2 Materials and methods

2.2.1 Fungal isolates and culture conditions

The AaV1-infected *A. alternata* isolate EGS 35-193 was described previously (Johnson *et al*., 2001; Aoki *et al*., 2009). We realized that the original isolate EGS 35-193 contained not only an intact dsRNA2 segment but a dsRNA2 segment with an in-frame deletion. Therefore, we named the original EGS 35-193 as EGS 35-193-1d. During the subculturing of EGS 35-193-1d on YGA plates (0.5% yeast extract, 2% glucose, and 1% agar), we found two other isolates, EGS 35-193-0d and EGS 35-193-2d. EGS 35-193-0d has only the intact dsRNA2, while EGS 35-193-2d has the intact dsRNA2 and two other kinds of dsRNA2 segments with either one or two internal deletions. All isolates were cultured on YGA plates, H-PDA plates (200 g/L unpeeled potato slices, 20 g/L dextrose, and 15 g/L agar), D-PDA (24 g/L Difico PDA powder and 15 g/L agar) and V8A plates (200 ml/L V8 juice, 3 g/L $CaCO_3$, 15 g/L agar) at 25°C for one week. For liquid cultures, mycelial plugs were used to inoculate YG broth (0.5% yeast extract, 2% glucose) and grown at 25°C for two weeks with shaking (60 strokes per min).

2.2.2 Curing of an AaV1-infected *A. alternata* isolate

The isolate EGS 35-193 was cured of AaV1 infection using a modified fragmentation treatment (Figures 2-1A and 2-1B; Kim *et al*., 2013) in Hamada's master thesis (Hamada, 2015). Mycelia were collected after one week of incubation on YGA plates, suspended in autoclaved distilled water, chilled on ice for 40 s, then fragmented twice for 5 s each time using a tissue homogenizer (Precellys®24., Bertin Corp, MD, USA) set at 5000 rpm. After chilling on ice for a further 10 s, the fragmented mycelia were spread on YGA plates and cultured for 1–2 days at 25°C, and then in YG broth for two weeks at 25°C. The process was repeated until no AaV1 was detectable.

2.2.3 Purification of virus particles and antibody production

Purification of AaV1 virions was conducted as described previously (Aoki *et al*., 2009), with minor modifications. The entire process was carried out at 4°C. Briefly, 10 g (fresh weight) of mycelia were homogenized in 100 ml buffer A (0.1 M sodium phosphate, 0.2 M KCl, pH 7.4) with a mixer and a French Press (One Shot A Model, Constant Systems, U.K.; 35Kpsi). Next, the homogenate was mixed for 1 h with 40% (v/v) chloroform/n-butanol (1:1), and the mixture was centrifuged at 8,000 × *g* (TOMY Suprema 21, NA-8 rotor, Japan) for 20 min. The supernatant was adjusted to 8% (w/v) polyethylene glycol 6,000 and 1% (w/v) NaCl, and gently stirred for 3 h to overnight. The solution was then centrifuged at 10,000 × *g* (TOMY Suprema 21, NA-8 rotor) for 5 min, and the pellet was resuspended in 8 ml 0.05 M sodium phosphate buffer (pH 7.0) and left at 4°C overnight. Finally, the suspension was layered onto 15 ml of a 45% sucrose cushion and centrifuged at 69,260 × g (Hitachi CP80WX, P28S swing rotor) for 16 h at 4°C, and the pellet was resuspended in 0.05 M sodium phosphate buffer (pH 7.0).

Saccharomyces cerevisiae virus L-A (ScV-L-A) virus particles were purified from *Saccharomyces cerevisiae* strain YPH499 and Mycoreovirus 1 (MyRV1) virus particles were purified from chestnut blight fungus using the methods described by Powilleit *et al*. (2007) and Hillman *et al*. (2004), respectively.

The purified virus particles were negatively stained with 2% uranyl acetate, and then observed by transmission electron microscope (TEM; JEM 1400 Plus, JOEL, Japan) with an acceleration voltage of 80 kV.

Anti-AaV1 antiserum production was described in Aoki *et al*., 2009, about 1.5 mg of partially purified AaV1 proteins was obtained from 100 g (fresh weight) of EGS 35-193-1d grown in YG broth. These purified proteins were injected into rabbits (about 0.2 mg per injection) every week for 3 weeks, and then the rabbits were given further injections (about 0.05 mg per injection) every week for another 3 weeks to ensure the success of the immunization (Protein Purify, Isesaki, Japan). The antiserum was collected after the fifth injection over a period of three weeks. The immunoglobulin G (IgG) against AaV1 was then purified from the anti-AaV1 antiserum using protein A agarose (Funakoshi, Japan), then stored at -80°C.

Anti-AaV1 ORF2 and AaV1 ORF3 were produced by using synthesized peptides as antigens. The predictions and synthesis of peptide sequences and the productions of antibodies were ordered to the company (Eurofins, Japan).

2.2.4 Protein analysis.

Purified virus particles were analyzed by 8% SDS-PAGE and either stained with Coomassie Brilliant Blue (CBB) (EzStainAQua, ATTO, Japan) or transferred to PVDF membrane (ATTO, Japan) for western blotting assays. For the western assays, the proteins were first exposed to the anti-AaV1 primary anti-serum (1:5000 dilution) and then to a secondary HRP-conjugated goat anti-rabbit polyclonal antibody (Bio-Rad, CA, USA) (1:10000 dilution). After washing, antibody-bound proteins were detected by luminescence using the EzWestLumi plus and EZ-Capture MG system (ATTO, Japan).

The major 82 kDa viral protein was isolated for sequence analysis. Purified AaV1 virus particles (50 μg protein) were resolved by 8% SDS-PAGE, and the band corresponding to the 82 kDa protein was excised from the gel. The collected protein was digested with Lysyl endopeptidase at 37°C for 16 h, then with trypsin at 37°C for 4 h. The digested sample was resolved with reverse-phase HPLC, and two peptide fragments were selected for amino acid sequencing using the Edman degradation method (Toray Research Center, Inc., Kamakura, Japan).

The AaV1-associated proteins of empty viral particle fraction and full viral particle fraction in CsCl isopycnic gradient centrifugation, were collected after SDS-PAGE electrophoresis and treated with In-gel digestion of trypsin, then analyzed by LC-MS/MS (LTQ XL^{TM}, Thermo Fisher) in Gene Research Center of Tokyo University of Agriculture and Technology.

2.2.5 Purification and detection of dsRNA

Viral dsRNA was extracted from 0.2 g (dry weight) of fungal mycelia using a micro-spin column method (Okada *et al*., 2015). Briefly, virus particles were isolated as described above, and 0.2 ml of viral suspension was mixed with 0.2 ml of 2× STE buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 200 mM NaCl) containing 1% SDS and 0.2 ml of phenol: chloroform: isoamyl alcohol (25:24:1). The mixture was vortexed for 10 min at room temperature, centrifuged at 15,000 × *g* for 5 min, and the aqueous phase was collected. The purified dsRNA was subjected to agarose gel electrophoresis (1%) containing ethidium bromide (EtBr, 0.5 μg/ml).

2.2.6 Detection of 5′ cap structures

In Takesita' s master thesis (Takeshita, 2011), RNA ligase-mediated rapid amplification of the cDNA ends (RLM-RACE) with the GeneRacer™ Kit (Thermo Fisher Scientific, Waltham, MA, USA) to confirm the presence of the 5′ cap structures on each of the AaV1 dsRNA segments (Figure 2-13). First, 500 ng of AaV1 dsRNA were resuspended in 10 μl distilled water, then DMSO (90% v/v) was added and the RNAs were denatured at 65°C for 15 min. Next, the denatured dsRNAs were recovered by ethanol precipitation and sequentially treated with calf intestinal phosphatase and tobacco acid pyrophosphatase, following the manufacturer's protocol. The GeneRacer™ oligo RNA (Table 2-2) was then ligated to the 5′ ends of the treated dsRNAs using T4 RNA ligase. Oligo $(dT)_{36}$ primers were used to create first-strand cDNA from the oligo RNA-ligated dsRNAs, then GoTaq® Green Master Mix (Promega, WI, USA) was used in PCR with the GeneRacer™ 5′ primer (Table 2-2) and specific 3′ reverse primers (Table 2-2) to amplify the target fragments.

In this study, I conducted RNA dot blot assays to look for m^7G cap structures at the 5′ ends of the AaV1 dsRNA segments (Figure 2-17). The AaV1, ScV-L-A, and MyRV1 dsRNAs (1000, 500, and 250 ng/μl respectively) were heat-denatured at 95°C for 5 min and chilled on ice for 5 min, then 1 μl of each solution was spotted onto Zeta-Probe Membrane (Bio-Rad). After UV cross-linking twice with 120,000 $\mu J/cm^2$ for 1 min each time in a UV crosslinker (UVC500, Hoefer Inc., Holliston, MA), the membrane was agitated in 20 ml blocking buffer containing 1× TBS-T buffer (0.02 M Tris, 0.15 M NaCl, 0.05 % Tween-20, pH 7.4) and 5% skim milk powder at room temperature for 1 h. Then, the membrane was gently rinsed in 10 ml of 1× TBS-T buffer three times for 5 min each.

Subsequently, the membrane was probed in 10 ml of primary antibody solution containing 1× TBS-T buffer, 1% skim milk powder, and 10 μg of an anti-m^7G-cap monoclonal antibody (mAb) (Code No. RN016M, MBL®, Woburn, MA, USA) at room temperature for 2 h with gentle agitation. The membrane was rinsed three times as described above, then probed in 10 ml of secondary antibody solution containing 1× TBS-T buffer, 1% skim milk powder, and 2 μg of a goat anti-mouse IgG (Code No. 401215, Merck, Darmstadt, Germany) at room temperature for 1 h with gentle agitation. After rinsing again, the antibody-bound spots were detected using the EzWestLumi plus and EZ-Capture MG system (ATTO).

2.2.7 Northern hybridization analysis

The AaV1 dsRNAs were separated by electrophoresis in 1% agarose gels, denatured in 10% (v/v) formaldehyde at 60˚C for 1 h, chilled in 20×SSC buffer for at least 15 min, then blotted onto nylon membranes (Zeta-Probe, Bio-Rad) using the capillary method. After cross-linking in a UV crosslinker (UVC500), the membranes were probed with a digoxygenin (DIG) labeled DNA probe. The probe (499 nt) was synthesized as a PCR product amplified from full-length AaV1 dsRNA2 using dsRNA2-specific primers (Table 2-2). Northern hybridization was conducted using the DIG DNA Labeling and Detection Kit (Roche, Germany) following the manufacturer's protocols.

2.2.8 RT-PCR, cloning, and sequencing

AaV1 dsRNA was heat-denatured at 98˚C for 5 min and immediately chilled on ice for at least 5 min. The SuperScript III First-strand synthesis system (Invitrogen, MA, USA) was used for first-strand cDNA synthesis, and then PCR was performed using the KOD One™ PCR Master Mix (TOYOBO, Japan). The PCR conditions were: 95˚C for 3 min followed by 35 cycles of 95˚C for 45 s, 55˚C for 30 s, and 72˚C for 45 s. The PCR products were then analyzed by electrophoresis in 1% agarose gels containing EtBr (0.5 μg/ml). The primer pairs used are listed in Table 2-2.

After electrophoresis, the predicted PCR bands were extracted from the agarose gels and purified using the GENECLEAN II Kit (MP Biomedical). EX-*Taq* was used for A-tailing and the PCR products were then cloned into the pCR™ 4-TOPO™ TA-cloning Vector (Invitrogen). The cloned PCR products were sequenced using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, MA, USA) and the Applied Biosystems 3130xl Genetic Analyzer (Applied Biosystems) according to the manufacturer's protocols. The sequences were analyzed with MegAlign software (Lasergene7, DNA-STAR®, WI, USA).

2.2.9 Phylogenetic analysis

The amino sequences of Altenariaviridae, *Totiviridae*, and *Chrysoviridae* were retrieved from NCBI (Table 2-3), then aligned using MUSCLE program. The phylogenetic tree was constructed using the Maximum Likelihood method and Poisson correction model by MEGA X (Kumar *et al*., 2018). The bootstrap values are marked next to the branches (1000 replicates).

2.3 Results

2.3.1 Curing *A. alternata* isolate EGS 35-193 of AaV1

In Hamada's master thesis (Hamada, 2015), AaV1 was hypothesized that would propagate and be distributed heterogeneously in the EGS 35-193 hyphae (Figure 2-1B). Therefore, AaV1-infected hyphae were cut into small fragments, spread out on YGA, then screened for reduced virus titer or virus absence by dsRNA extraction followed by agarose gel electrophoresis, and primer-specific RT-PCR (only dsRNA1).

Three isolates with reduced AaV1 titers were detected among fifteen isolates assessed (nos. 4, 7, and 10, Figure 2-2A). One of them (no. 10) was selected and the fragmentation process was repeated. Finally, two isolates cured of AaV1 were obtained among ten isolates assessed (nos. 10-9 and 10-10, Figure 2-2B), as confirmed by RT-PCR using specific primer sets for each of the four dsRNA segments (Figure 2-2C). In this study, existence of dsRNA2-dsRNA4 was confirmed by primer-specific RT-PCR (Table 2-2), the stability of virus-free isolates was confirmed after subculturing several times (data not shown).

2.3.2 Morphological effects of variations in AaV1 dsRNAs in EGS 35-193 isolates

During subculturing of the original AaV1-infected *A. alternata* EGS 35-193-1d on YGA plates, we noted sectors with distinct mycelial morphology: one with a flatter mycelium and intense pigmentation (named EGS 35-193-0d), and the other with slow growth (named EGS 35-193-2d) (Figure 2-3A). Compared with the virus-free isolate of EGS 35-193 (named EGS 35-193-VF), the three AaV1-infected isolates showed impaired and abnormal growth phenotypes. EGS 35-193-0d showed the highest growth rate among the three AaV1-infected isolates, followed by EGS 35-193-1d, and lastly by EGS 35-193-2d (Figure 2-4). High levels of pigmentation could only be observed in EGS 35-193-0d, while the other AaV1-infected isolates and the virus-free isolate showed modest pigmentation on YGA plates (Figure 2-3A). These different phenotypes were also seen when the isolates were grown on other media (Figures 2-5A, B and C).

To confirm the presence of AaV1 in the three EGS 35-193 variants, we purified dsRNAs from the fungal mycelia (Figure 2-3B) and from virions isolated from the mycelia of the three variants (Figures 2-3C). The purified virions were analyzed by SDS-

PAGE and with a western blot probed with the anti-AaV1 antibody (Figures 2-9A ad B). Although all three AaV1-infected isolates harbored the four dsRNA elements, the ratio of dsRNA2 (2794 bp) to dsRNA3 (2576 bp) fluctuated in preparations from both the fungal mycelia and the virions (Figures 2-3B and C). In preparations from EGS 35-193-0d, the dsRNA2 and dsRNA3 bands had similar intensities, while the dsRNA3 bands were much more intense than the dsRNA2 bands in preparations from both EGS 35-193-1d and EGS 35-193-2d (Figures 2-3B and C). In addition, a 2.1 kbp dsRNA was discovered in EGS 35-193-2d (Figures 2-3B and C, lane 2d, black arrows). I investigated the origin of the 2.1 kbp dsRNA in the following experiments.

2.3.3 Detection of defective dsRNA2 segments in AaV1-infected *A. alternata*

To explain the varying amounts of dsRNA2 and dsRNA3 in the three EGS 35-193 isolates, and to identify the origin of the 2.1 kbp dsRNA segment in EGS 35-193-2d (Figures 2-3B and C), I conducted RT-PCR with specific primers designed to amplify the full length of each AaV1 segment excluding the poly (A) tail (Figure 2-14C, Table 2-2). I used dsRNAs from AaV1 virions isolated from each EGS 35-193 isolate as templates. In the sample from EGS 35-193-0d, I obtained full-sized PCR bands derived from dsRNA1 (3.6 kbp), dsRNA2 (2.8 kbp), dsRNA3 (2.6 kbp), and dsRNA4 (1.4 kbp) (Figure 2-6B, lanes 1–4). However, in the samples from EGS 35-193-1d and EGS 35-193-2d I obtained multiple PCR-bands, ranging from 2.1–2.8 kbp, with the dsRNA2-specific primer pair (Figures 2-6A and C, lane 2). These results suggested that the fluctuating ratios of AaV1 dsRNA2 to dsRNA3 and the 2.1-kbp dsRNA segment may be attributed to these dsRNA2 variants, which potentially carried internal deletions. To confirm this hypothesis, northern hybridization was performed using dsRNAs from each AaV1-infected isolate with a DIG-labeled dsRNA2-specific probe (Figures 2-6E and F, Table 2-

2). Only one signal corresponding to the intact, 2794 bp dsRNA2 was detected in the EGS 35-193-0d isolate (Figures. 2-6E and F). In the EGS35-193-1d isolate, I detected two signals corresponding to the intact dsRNA2 (2794 bp) and dsRNA2 del-1, similar in size to the dsRNA3 segment (2576 bp) (Figures 2-6E and F). In the EGS 35-193-2d isolate, I detected three signals corresponding to the intact dsRNA, dsRNA2 del-1, and dsRNA2 del-2, which was similar in size to the additional 2.1 kbp segment detected by agarose gel electrophoresis, were observed (Figures 2-6E and F).

2.3.4 Characterization of the deletions in AaV1 dsRNA2 del-1 and dsRNA2 del-2

To identify the deleted regions in dsRNA2 del-1 and dsRNA2 del-2, I sequenced the PCR products shown in Figure 2-6H. The two amplified dsRNA2 products from EGS 35-193-1d, the intact dsRNA2 and the dsRNA2 del-1, were separated by extended electrophoresis (1% agarose, 18 V, 20 h; Figure 2-6H), extracted from the gel, cloned, and sequenced. I found an in-frame deletion site (D1) near the middle of the dsRNA2 segment: the region from nt 1,275 to nt 1,505 in the intact dsRNA2 was deleted in dsRNA2 del-1. The exact size of the dsRNA2 del-1 segment was 2,513 bp excluding the poly $(A)_{50}$, and the length of D1 was 231 bp (Figures 2-7A and 2-8B).

A similar analysis was performed for EGS 35-195-2d (Figure 2-6H). The dsRNA2 del-1 segment from EGS 35-195-2d had only one deletion site, D1, which was identical to that in dsRNA2 del-1 from EGS 35-193-1d (Figures 2-7A and 2-8C). The dsRNA2 del-2 segment had two deletions: D1, which again was identical to the deletion in EGS 35-193-1d, and D2, a deletion in the 5′ region of the segment spanning from nt 113 to nt 577 in the intact dsRNA2 (Figures 2-7B and 2-8C). The D1 and D2 deletions shortened the length of dsRNA2 del-2 from 2794 bp to 2098 bp. Both the defective dsRNA2 del-1 and dsRNA2 del-2 segments had open reading frames encoding putative proteins of 767 aa and 612 aa, respectively (Figures 2-8B and C). The full-sized dsRNA2 segment had an open reading frame encoding a putative protein of 844 aa (Figure 2-6H and 2-8A).

2.3.5 Analysis of the AaV1 major structural protein

In Aoki's thesis (Aoki, 2009), The purified AaV1 particle proteins were used as antigens to immunize rabbits, and an immunoglobulin G (IgG) against AaV1 was purified from the anti-AaV1 antiserum. Western blot analysis of the purified AaV1 virions showed that the purified anti-AaV1 IgG specifically detected an 82 kDa protein band, also visible following SDS-PAGE with CBB staining (Figure 2-10A). In the former report on AaV1 (Aoki et al., 2009), claimed the molecular weight of the AaV1 major protein at 97 kDa, however, my further analysis clarified that the major protein band was approximately 82 kDa.

Edman degradation was used to investigate the primary structure of the purified 82 kDa protein, and to clarify which dsRNA segment encoded the AaV1 structural protein in Aoki's thesis (Aoki, 2009). Since no phenylthiohydantoin-amino acid derivatives were observed after five cycles of reaction, indicated that the N-terminus of the 82 kDa protein was blocked (data not shown). In-gel digestions were applied to treat the major protein band with lysyl endopeptidase and trypsin. This resulted in two clear peaks (peaks 2 and 3) in reverse-phase HPLC (Figure 2-10B). Both peak 2 (ASVQAIMAR, blue-colored) and peak 3 (SVDWSPVGLT, red-colored) corresponded to the internal sequences of the predicted protein encoded by dsRNA3 (Figure 2-10C).

In this study, I tried to purify the AaV1 virus particles by CsCl isopycnic gradient centrifugation. Two visible virus particle-associated fractions formed after CsCl isopycnic gradient centrifugation (Figure 2-11A). After confirmation by agarose gel electrophoresis and SDS-PAGE electrophoresis (Figure 2-11B), the fractions 2, 3, 4 were empty particle fractions, while fractions 7, 8, 9 were full partilce fractions. The AaV1-associated proteins (Figure 2-11C) were subjected to anlysis of LC-MS/MS. The results showed that all the AaV1-associated proteins (Figure 2-11C) corresponded to AaV1 ORF3-encoded protein (Figure 2-12A-D).

2.3.6 Presence of 5′ cap structures on AaV1 dsRNAs

In Takeshita's thesis (Takeshita, 2011), four dsRNA segments of AaV1 were extracted from the agarose gel (Figure 2-14A) and used as templates in RLM-RACE experiments, which would yield amplicons only when dsRNA was capped (Figure 2-13). For each segment, the results revealed amplicons of the predicted size based on the designed primers, as shown in Figure 2-14B: 108 bp, 242 bp, 279 bp, and 307 bp for dsRNA1, dsRNA2, dsRNA3, and dsRNA4, respectively (Table 2-2, Figure 2-14C).

Viral dsRNA molecules were extracted from purified AaV1, ScV-L-A (Figure 2-15), and MyRV1 (Figure 2-16) virions (Figure 2-18A). Following quantification, an RNA dot blot assay with anti-m^{7}G-Cap mAb was performed for detecting the presence of 5' cap structures. The dsRNA dots of AaV1 and MyRV1 (positive control) showed signals, while the dots from an AaV1-free isolate, ScV-L-A, and DW (negative and no template controls) showed no signal (Figure 2-18B).

2.4 Discussion

In previous report, AaV1 was the first dsRNA virus to be identified with a poly (A:U) structure at the 3' end of each segment of its genome by 5' RACE on minus strands of dsRNA segments (Aoki, 2009; Aoki *et al.*, 2009). The 5′ m^{7}G-cap structures were first found on the viral dsRNAs of vaccinia virus (Wei and Moss, 1975) and cytoplasmic polyhedrosis virus, belonging to the family *Reoviridae* (Furuichi and Miura, 1975). In this study, I used RNA dot blot assays to demonstrate that each of the AaV1 dsRNA segments has a capping structure, 7-methylguanosine (m^7G), at its 5' end. This result can correspond the RLM-RACE detection (Figure 2-14B; Takeshita, 2011) To the best of our knowledge, AaV1 is the first dsRNA virus to be discovered with both the 3' poly (A) tail and the 5' cap structure on each genomic segment.

To our knowledge, translation of particular RNA viruses other than reoviruses is independent of the cap structure. It is known that potyviruses use a viral protein (VPg) that is covalently linked to the 5’ end of the RNA (Zhang *et al*., 2015), and carmoviruses use the 3'-untranslated region (Simon, 2015) as alternatives for assisting the formation of the translation initiation complex. The influenza virus and the yeast L-A virus have mechanisms for snatching cellular mRNA caps and using them to assist translation of the viral RNA (Plotch S.J, 1979; Fujimura and Esteban, 2011). The cap at the 5′ end is essential for translation initiation of cellular mRNAs and is associated with the eucaryotic initiation factors eIF4E, eIF4G, and eIF4A, which recruit the 40S ribosomal subunit (Lindqvist *et al*., 2008). In addition, the cap structure also serves as a defense against *SKI*/*XRN1* exoribonuclease, which explicitly degrades mRNAs with no cap (Masison *et al*., 1995). Thus, the cap structures of AaV1 are considered advantageous for effective viral propagation.

The 5' cap structure and the 3' poly (A) tail have the functions of protecting the genome, transporting the RNA, and enhancing translation (Hocine *et al*., 2010). Cap structures also increase the accuracy and efficiency of mRNA splicing, which generally occurs in the nucleus (Inoue *et al*., 1989). It seems more efficient for mycoviral RNA genomes to have a 5' cap and a 3' poly (A) tail for propagation in the cytoplasm of eukaryotic cells, since these modifications act synergistically to enhance translation (Gallie, 1991). However, few RNA viruses have a cap and a poly (A) tail, and the reason for this is unknown. Indeed, ribosomal RNA, which accounts for 60% of the total RNA in cells (Woolford and Baserga, 2013), also does not have a cap and a poly (A) structure and is not translated. It may be favorable for the mycoviruses to propagate in the cytoplasm without these terminal structures as this might lead to high copy numbers.

In our previous study, the viral proteins were analyzed by 7% SDS-PAGE and the protein size was determined as 97 kDa by comparison with the low molecular weight marker (GE Healthcare, UK) (Aoki *et al.*, 2009). However, in this study, we analyzed the viral proteins by 10% SDS-PAGE and found that the size of the major protein was 82 kDa, using DynaMarker® Protein MultiColor (Funakoshi Co., Ltd., Japan). This corresponds with the predicted size of the protein encoded by dsRNA3. Edman degradation demonstrated that the AaV1 dsRNA3 encodes the 82 kDa protein, which is the major structural protein of the AaV1 virion. While carrying out the Edman analysis, we found that the N-terminus of the 82 kDa protein was blocked (data not shown). It is possible that the AaV1 82 kDa protein is N-terminally acetylated since this is one of the most common protein modifications in eukaryotes (Arnesen *et al.*, 2009). *MAK3* N-acetyltransferase modification of Gag is necessary for virion assembly of the yeast L-A virus (Tercero and Wickner., 1992; Tercero *et al.*, 1993). N-terminal acetylation may also be necessary for AaV1 virion formation.

During the subculturing of the AaV1-infected EGS 35-193 strain, I found fungal isolates carrying virions with defective dsRNA2 segments, dsRNA2 del-1 and dsRNA2 del-2, which occurred by in-frame deletion events (Figures 2-6 to 2-8). Based on agarose gel electrophoresis of the AaV1 dsRNA genomes purified from virions (Figure 2-3C), the defective dsRNA2 segments appeared to have no significant effects on accumulation of the other dsRNA segments. To confirm the relative quantification of dsRNA2 segments in three isolates, I analyzed the results of agarose gel electrophoresis (Figure 2-6E) and northern hybridization (Figure 2-6F), by Fiji/ImageJ software (Schneider *et al*., 2012). The results showed that the relative quantification of dsRNA2 segments in three isolates has no significant difference (Figure2-6G), and the detailed results will be obtained by real-time RT-PCR in the future. The defective dsRNA2 segments were responsible for altered phenotypes in the fungal host, including reduced hyphal growth rates and irregular pigmentation (Figure 2-3 and 2-5 A-C). No deletions were found in dsRNA1, dsRNA3, or dsRNA4. I speculate that this is because the AaV1 open reading frame ORF1 (RdRp) and ORF3 (coat protein) are essential for viral replication or virion packaging. Similar selective deletion events were found in mycoreoviruses. The inducible genome deletions of MyRV1 dsRNA S4 and S10 are related to changes in vertical transmission efficiency and host colony morphology, but not to viral replication (Sun and Suzuki, 2008; Eusebio-Cope *et al*., 2010; Kanematsu *et al*., 2014). Several other findings of RNA genome deletions in the hypoviruses also demonstrated their encoded in-frame fusion proteins affecting host growth negatively (Hillman *et al*., 2000; Xie *et al*., 2011; You *et al*., 2019). Interestingly, both the deletions in AaV1 dsRNA2 were in-frame (Figures 6B, C). These in-frame deletions might have protected the deleted dsRNA2 segments from nonsense-mediated mRNA decay, since AaV1 dsRNAs with 5' cap structures would recruit decapping enzymes, such as Dcp1p/Dcp2, and the major cytoplasmic 5'-3' exonuclease (Ski1/Xrn1), which target nonsense-containing mRNAs (Peltz *et al*., 1993; Maderazo *et al*., 2003; Celik *et al*., 2017). Similar to the full-length ORF2, the two defective ORF2 segments contained no conserved domains based on searches of the NCBI database (data not shown). These in-frame fusion ORF2 proteins negatively affect host growth, but the exact mechanisms by which they effect host growth remain unclear. In the future, I will investigate the connection between the defective genomes and host growth using a heterologous expression system in yeast.

In the previous phylogenetic analysis, AaV1 was shown to be related to *Chrysoviridae* and *Totiviridae* (Aoki *et al*., 2009). The increased number of available mycovirus sequences has now allowed us to construct a more precise phylogenetic tree (Figure 2-19A, Table 2-3). The new tree shows that the Alternaviridae family is more closely related to *Totiviridae* than to *Chrysoviridae*. Interestingly, AaV1 and the other nine alternaviruses have an ADD motif instead of GDD in the conserved motif VI of RdRp (Figure 2-19B) (Aoki *et al*., 2009; Kamer and Argos, 1984; Koonin, 1991). The GDD motif sometimes shows flexible glycine residue requirements, such as IDD in infectious bursal disease virus or SDD in phage Φ6, even though these variants possessed the same function as the GDD motif (Shwed *et al*., 2002). Therefore, I expect that the ADD sequence of RdRp motif VI of alternaviruses is functionally active.

In the future, I would like to investigate what advantages are provided by the cap and poly (A) structures on the AaV1 dsRNA genomic segments. The *SKI* genes are involved in 5′-3′ and 3′-5′ mRNA degradation pathways (Toh-E *et al*., 1978; Widner and Wickner, 1993; Zhang *et al*., 2019). Therefore, I will utilize *SKI*-deficient mutants or *SKI*-overexpressing strains of *S. cerevisiae*, since this might provide valuable insights into the roles of the cap and poly (A) structures in RNA degradation.

Table 2-1. Current members of Alternaviridae.

Virus	Genome	Particle size	5' cap	3' poly (A) tail	Reference
Alternaria alternata virus 1 (AaV1)	dsRNA1 (3617 nt, RdRp) dsRNA2 (2794 nt, P2) dsRNA3 (2576 nt, Coat protein) dsRNA4 (1420 nt, P4)	33 nm	+	+[a]	Aoki et al., 2009; Wu et al., 2021
Diaporthe alternavirus 1 (DAV1)	dsRNA1 (3676 nt, RdRp) dsRNA2 (2683 nt, P2) dsRNA3 (2480 nt, Coat protein) dsRNA4 (1698 nt, P4)	35-40 nm	+	+[a]	this study
Aspergillus mycovirus 341 (AsV341)	dsRNA1 (3588 nt, RdRp)	nd	nd	+	Hammond et al., 2008
Aspergillus foetidus mycovirus (AfV-F)	dsRNA1 (3588 nt, RdRp) dsRNA2 (2770 nt, P2) dsRNA3 (2466 nt, P3) dsRNA4 (2005 nt, P4)	nd	nd	+	Kozlakidis et al., 2013
Fusarium poae alternavirus 1 (FpAV1)	dsRNA1 (3559 nt, RdRp) dsRNA2 (2496 nt, P2) dsRNA3 (2482 nt, P3)	nd	nd	+	Osaki et al., 2016
Fusarium graminearum alternavirus 1 (FgAV1)	dsRNA1 (3524 nt, RdRp) dsRNA2 (2470 nt, P2) dsRNA3 (2485 nt, P3)	nd	nd	+	He et al., 2018
Fusarium incarnatum alternavirus 1 (FiAV1)	dsRNA1 (3548 nt, RdRp) dsRNA2 (2514 nt, P2) dsRNA3 (2498 nt, P3)	nd	nd	+	Zhang et al., 2019
Aspergillus heteromorphus alternavirus 1 (AheAV1)	dsRNA1 (3576 nt, RdRp) dsRNA2 (2742 nt, P2) dsRNA3 (2427 nt, P3)	nd	nd	+	Gilbert et al., 2019
Cordyceps chanhua alternavirus 1 (CcAV1)	dsRNA1 (3512 nt, RdRp) dsRNA2 (2655 nt, P2) dsRNA3 (2415 nt, P3)	nd	nd	+	Zhang et al., 2022
Fusarium solani alternavirus 1 (FsAV1)	dsRNA1 (3522 nt, RdRp) dsRNA2 (2633 nt, P2) dsRNA3 (2403 nt, Coat protein) dsRNA4 (1721 nt, P4)	31 nm	+	+[a]	Lutz et al., 2022

nd: no data

a: poly (A:U) structure

Table 2-2. List of the primers used in Chpter 2.

Primer name	Sequence	Purpose
GeneRacer$^{\mathrm{TM}}$ oligo	5′-rCrGrArCrUrGrGrArGrCrArCrGrArGrGrArCrArCrUrGrArC rArUrGrGrArCrUrGrArArGrGrArGrUrArGrArArA	RLM-RACE of 5′ cap detection
GeneRacer$^{\mathrm{TM}}$ 5′ primer	5′-CGA CTG GAG CAC GAG GAC ACT GA	
RLM-RACE-dsRNA1-3′	5′-TTC CTG AAC CCC AGC AAT AGG	
RLM-RACE-dsRNA2-3′	5′-ATG GGG AAA CGG GAA CGC C	
RLM-RACE-dsRNA3-3′	5′-AAA TGT CAT CGG CCG AAC CC	
RLM-RACE-dsRNA4-3′	5′-TTG TCA ACA TGG ACG TGA TGG	
AaV1 dsRNA1-F	5′-GAG GGA TCC TGG CAA AAA GCT ATC GTA TCT CG	AaV1 full-length genomes detection and cloning
AaV1 dsRNA1-R	5′-TTC GGA TCC GGA TAT TCC GAC TAA ATC ACG	
AaV1 dsRNA2-F	5′-CCG AAT TCC ACA AAG CAA TCG TAT CGC CAG	
AaV1 dsRNA2-R	5′-CGC GAA TTC ATA TTT GTT CCA CTA CAC CAG ACC	
AaV1 dsRNA3-F	5′-TGG AGC TCC ATA AAG CTA TCG TAT CTC GAG	
AaV1 dsRNA3-R	5′-TAC CGA GCT CGT TCC ACG TTC GAG ACA CGC	
AaV1 dsRNA4-F	5′-CCG AAT TCC ATA AAG CAA TCG TAT CGC CAG	
AaV1 dsRNA4-R	5′-GCT TGA ATT CGT TGT TAT CCT CAC AGC ACC	
AaV1 dsRNA2-F	5′-CCG AAT TCC ACA AAG CAA TCG TAT CGC CAG	DIG DNA probe synthesis
AAM2RA-C2	5′-TCT CGC GCA TAT GCG TGC TC	
AaV1-RdRp-F	5′-CTT AAC CGC GAG CTC TCG GCA A	AaV1 detection of curing experiment
AaV1-RdRp-R	5′-AGG CTC CAC AAC AAG CCT TGT A	
dsRNA2-check-5′	5′-TAG CAA GCG TGG GGT ATC G	
dsRNA2-check-3′	5′-ATT ACG GTA TTG AGT TGG C	
AaV1 ORF3-F	5′-ATA GTC GAC ATG GCG ACG TTT GGA AGT G	
AaV1 ORF3-R	5′-GAG TCG ACT CAA ATG ACA CCA GAA GCT C	
AaV1 ORF4-F	5′-GGG AAT TCA TGT TTG ATT CCT TTT GTT CC	
AaV1 ORF4-R	5′-AAG AAT TCT TAT GCA GAC GTA CCT ACT CC	

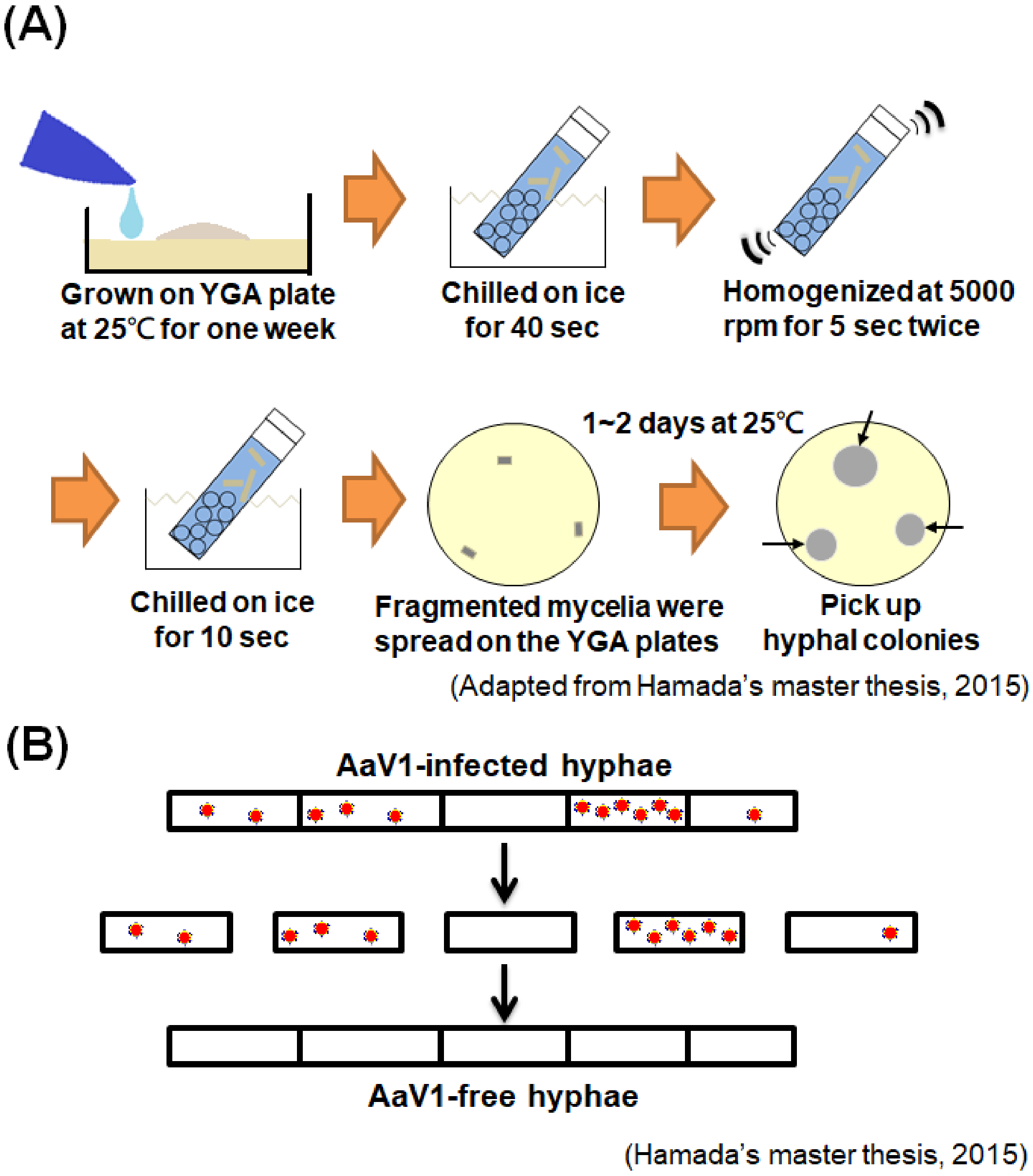
*Underlined sequences indicate viral sequences.

Table 2-3. List of the amino acid sequences of viral RdRps used in the phylogenetic analysis of Alternaviridae.

Virus name	Abbreviation	Accession No.
Chrysoviridae		
Alphachrysovirus		
Amasya cherry disease associated chrysovirus	ACDACV	YP_001531163
Anthurium mosaic-associated virus	AMAV	YP_009667023
Aspergillus fumigatus chrysovirus	AfuCV	YP_009508104
Brassica campestris chrysovirus 1	BcCV1	YP_009667006
Colletotrichum gloeosporioides chrysovirus 1	CgCV1	YP_009667012
Chrysothrix chrysovirus 1	CcCV1	QGR26538
Cryphonectria nitschkei chrysovirus 1	CnCV1	YP_009507942
Fusarium oxysporum chrysovirus 1	FoCV1	YP_009665200
Helminthosporium victoriae virus 145S	HvV145S	YP_052858
Isaria javanica chrysovirus 1	IjCV1	YP_009337840
Macrophomina phaseolina chrysovirus 1	MpCV1	YP_009667008
Penicillium chrysogenum virus	PcV	YP_392482
Persea americana chrysovirus	PaCV	YP_009666328
Raphanus sativus chrysovirus 1	RsCV1	YP_009667003
Shuangao insect-associated chrysovirus	SCLV	ASA47445
Verticillium dahliae chrysovirus 1	VdCV1	YP_009507948
Zea mays chrysovirus 1	ZMCV1	YP_009551655
Betachrysovirus		
Alternaria alternata chrysovirus 1	AaCV1	YP_009553287
Aspergillus thermomutatus chrysovirus 1	AthCV1	AWC67507
Botryosphaeria dothidea chrysovirus 1	BdCV1	YP_009353026
Colletotrichum fructicola chrysovirus 1	CfCV1	YP_009551629
Coniothyrium diplodiella chrysovirus 1	CdCV1	QDB74971
Fusarium graminearum dsRNA mycovirus 2	FgV2	ADW08802
Fusarium oxysporum f. sp. dianthi mycovirus	FodV	YP_009158913
Magnaporthe oryzae chrysovirus 1-A	MoCV1-A	YP_003858286
Neofusicoccum parvum chrysovirus 1	NpCV1	QDB74975
Penicillium janczewskii chrysovirus 1	PjCV1	YP_009182332
Penicillium janczewskii chrysovirus 2	PjCV2	YP_009667018

Virus name	**Abbreviation**	**Accession No.**
Totiviridae		
Totivirus		
Saccharomyces cerevisiae virus L-A	ScV-L-A	NP_620495
Saccharomyces cerevisiae virus L-BC	ScV-L-BC	NP_042581
Ustilago maydis virus H1	UmVH1	NP_620728
Scheffersomyces segobiensis virus L	SSVL	YP_009507831
Xanthophyllomyces dendrorhous virus L1A	XdV-L1A	YP_007697651
Xanthophyllomyces dendrorhous virus L1b	XdV-L1B	YP_009507835
Tuber aestivum virus 1	TaV1	YP_009507833
Victovirus		
Helminthosporium victoriae virus 190S	Hv190SV	NP_619670
Sphaeropsis sapinea RNA virus 1	SsRV1	NP_047558
Sphaeropsis sapinea RNA virus 2	SsRV2	NP_047560
Coniothyrium minitans RNA virus	CmRV	YP_392467
Magnaporthe oryzae virus 1	MoV1	YP_122352
Magnaporthe oryzae virus 2	MoV2	YP_001649206
Alternaria arborescens victorivirus 1	AaVV1	YP_009553478
Gremmeniella abientina RNA virus L1	GaRV-L1	NP_624332
Aspergillus foetidus slow virus 1	AfV-S1	YP_009508249
Beauveria bassiana victorivirus 1	BbVV1	YP_009508251
Chalara elegans RNA Virus 1	CeRV1	YP_024728
Helicobasidium mompa totivirus 1-17	HmV1-17	NP_898833
Rosellinia necatrix victorivirus 1	RnVV1	YP_008130308
Tolypocladium cylindrosporum virus 1	TcV1	YP_004089630
Leishmaniavirus		
Leishmania RNA virus 1-1	LRV1	NP_041191
Leishmania RNA virus 2-1	LRV2	NP_043465
Giardiavirus		
Giardia lamblia virus	GLV	NP_620070
Trichomonasvirus		
Trichomonas vaginalis virus 1	TVV1	AAA62868
Trichomonas vaginalis virus 2	TVV2	NP_624323
Trichomonas vaginalis virus 3	TVV3	NP_659390
Trichomonas vaginalis virus 4	TVV4	YP_009507836

Virus name	**Abbreviation**	**Accession No.**
Alternaviridae		
Alternavirus		
Alternaria alternata virus 1	AaV1	YP_001976142
Aspergillus mycovirus 341	AsV341	ABX79997
Aspergillus foetidus mycovirus	AfV-F	YP_007353985
Fusarium poae alternavirus 1	FpAV1	YP_009272952
Fusarium graminearum alternavirus 1	FgAV1	YP_009667012
Fusarium incarnatum alternavirus 1	FiAV1	AYJ09265
Aspergillus heteromorphus alternavirus	AheAV1	AZT88575
Cordyceps chanhua alternavirus 1	CcAV1	UPH33984
Fusarium solani alternavirus 1	FsAV1	OM326757
Diaporthe alternavirus 1	DAV1	-



(Adapted from Hamada's master thesis, 2015)

(Hamada's master thesis, 2015)

Figure 2-1. Curing of AaV1-infected *A. alternata* isolate.

(A) Flowchart of procedures to isolate AaV1-free isolates, and **(B)** Paradigm of AaV1-distribution in hyphae of the strain EGS 35-193. (Hamada, 2015).

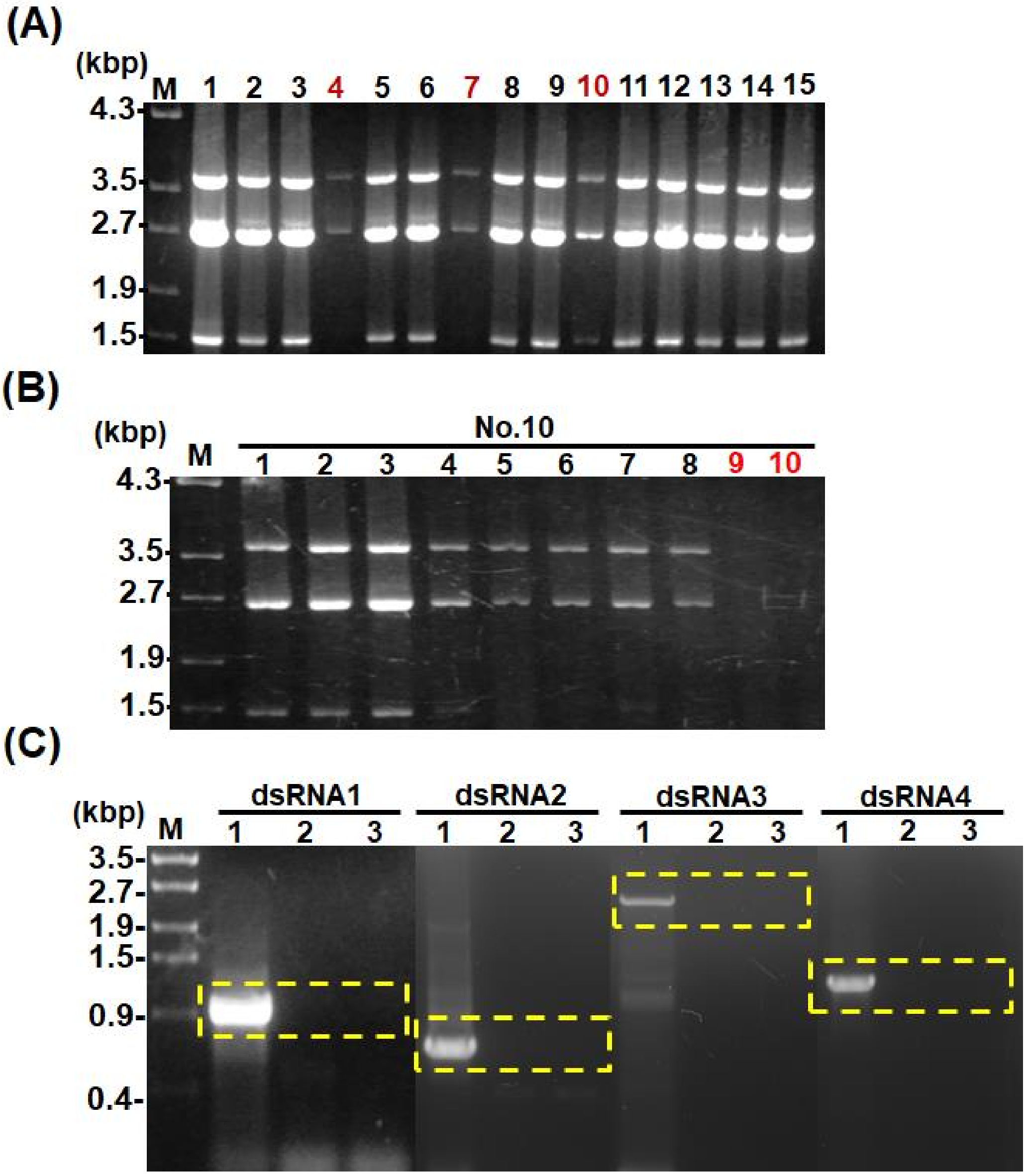


Figure 2-2. Selection of AaV1-free isolates.

(A, B) Step-wised selection of Aav1-free isolates after curing procedures. (Hamada, 2015). **(C)** The AaV1 specific RT-PCR bands amplified for the AaV1 positive control (dsRNA1–937 bp; dsRNA2–733 bp; dsRNA3–2297 bp; dsRNA4–1198 bp), while no RT-PCR band was amplified from the isolates No.10-9 and No.10-10. These two isolates were used as AaV1-free isolates, EGS 35-193-VF in this study.

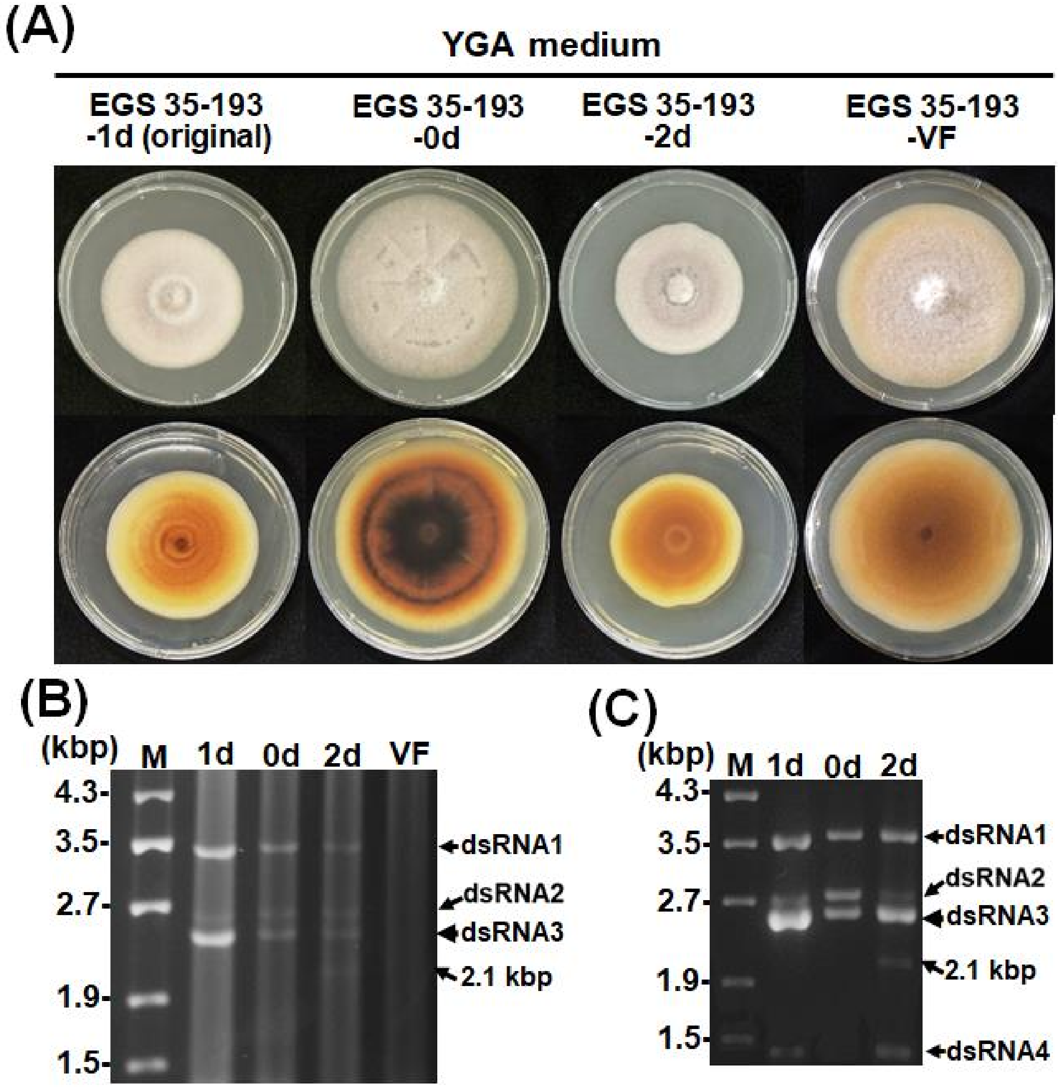

(A)
YGA medium
EGS 35-193 -1d (original)
EGS 35-193 -0d
EGS 35-193 -2d
EGS 35-193 -VF
(B)
(kbp)
M
1d
0d
2d
VF
4.3
3.5
2.7
1.9
1.5
dsRNA1
dsRNA2
dsRNA3
2.1 kbp
(C)
(kbp)
M
1d
0d
2d
4.3
3.5
2.7
1.9
1.5
dsRNA1
dsRNA2
dsRNA3
2.1 kbp
dsRNA4

Figure 2-3. Phenotypic changes in EGS 35-193 mycelia caused by rearrangements of AaV1 dsRNA2.

(A) Four types of colony morphologies were exhibited by the EGS 35-193 derivatives EGS 35-193-1d (the original strain), EGS 35-193-0d, EGS 35-193-2d, and EGS 35-195-VF (AaV1-free), grown on YGA plates for 7 days at 25˚C. **(B, C)** Agarose gel electrophoresis of dsRNAs purified from mycelia (20 mg) of the four EGS 35-193 derivatives, purified by the spin column method **(B)**, and dsRNAs extracted from purified virus particles of the three EGS 35-193 AaV1-infected isolates **(C)**. The dsRNAs were separated in 1.0% agarose gels with EtBr (0.5 μg/ml) at 18 V for 20 h. Lane designation: M, 250 ng of λ-EcoT14I-digested DNA marker; 1d, EGS 35-193-1d; 0d, EGS 35-193-0d; 2d, EGS 35-193-2d; VF, EGS 35-193-VF.

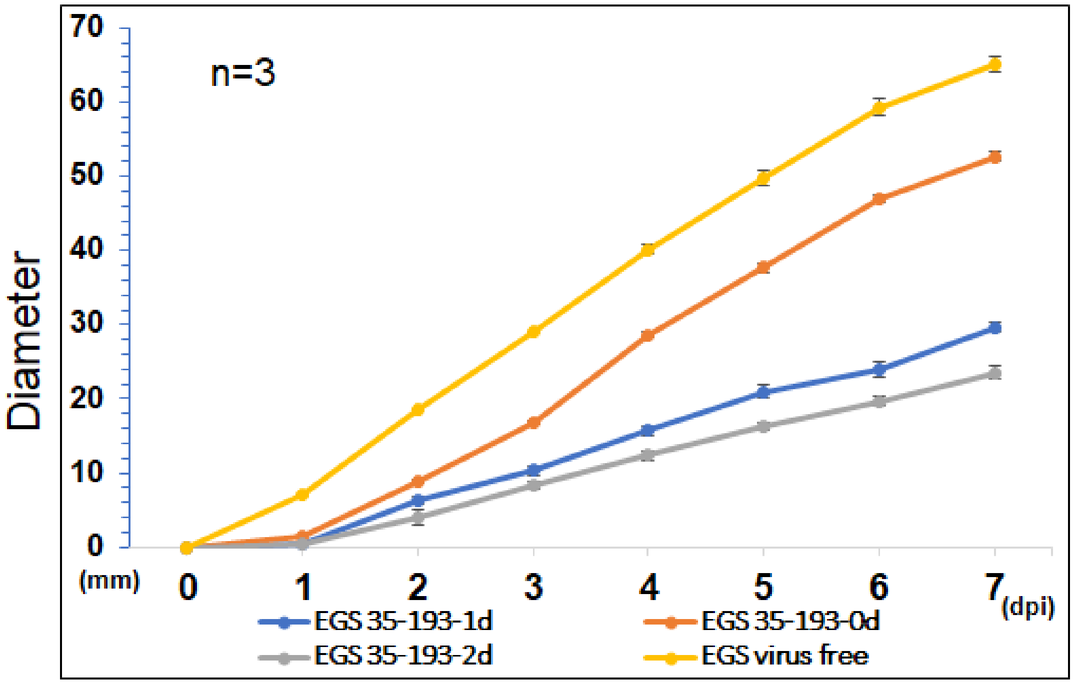
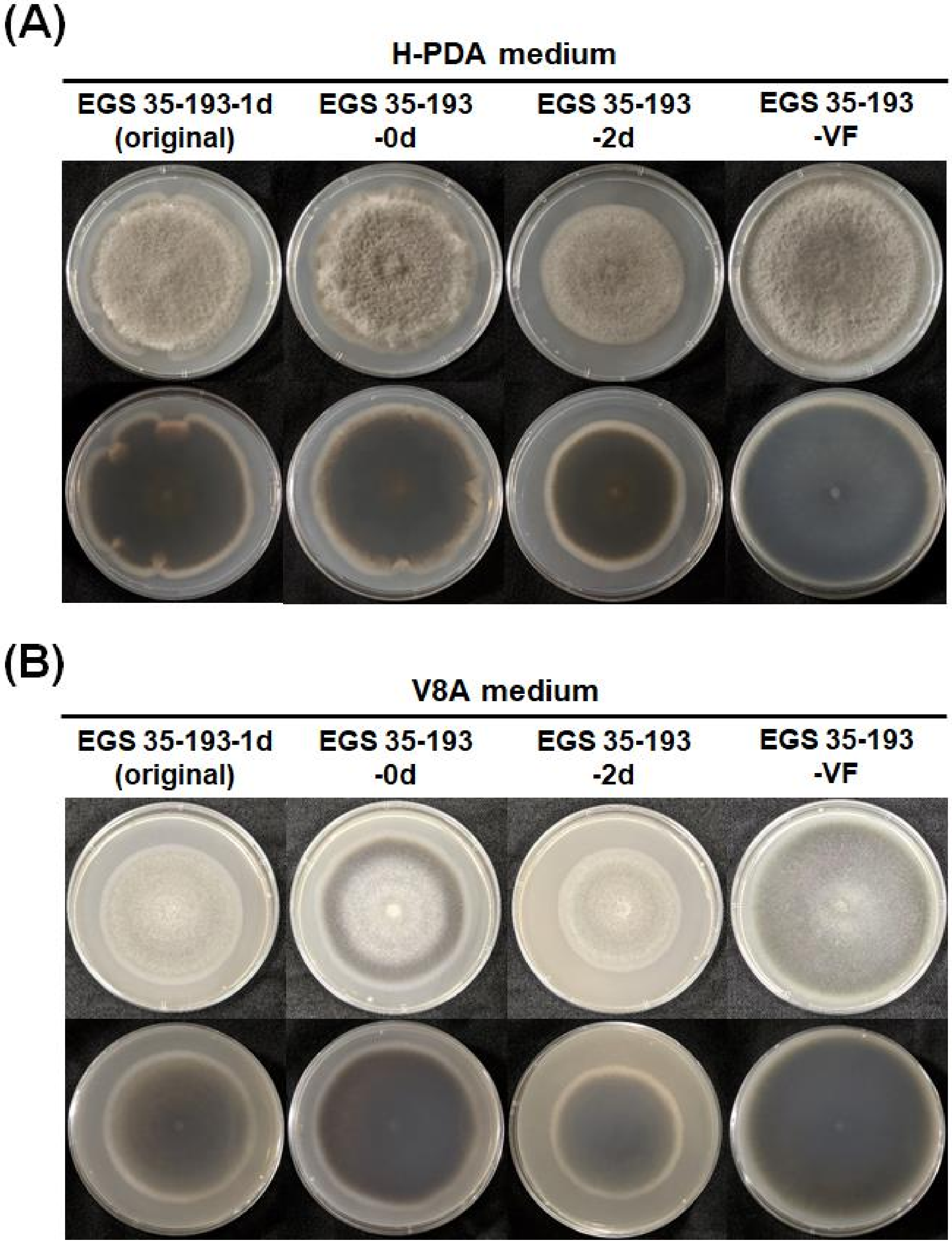


Figure 2-4. Mycelia growth rates of AaV1-infected isolates and AaV1-free isolate on YGA plates (n = 3 replicates).

Each isolate was cultured on YGA plates at 25 ℃ for 7 days and measured the diameter of colonies every day (excluding the agar plug with a diameter of 5mm). dpi: days post inoculation.


(A)
H-PDA medium
EGS 35-193-1d (original)
EGS 35-193 -0d
EGS 35-193 -2d
EGS 35-193 -VF
(B)
V8A medium
EGS 35-193-1d (original)
EGS 35-193 -0d
EGS 35-193 -2d
EGS 35-193 -VF

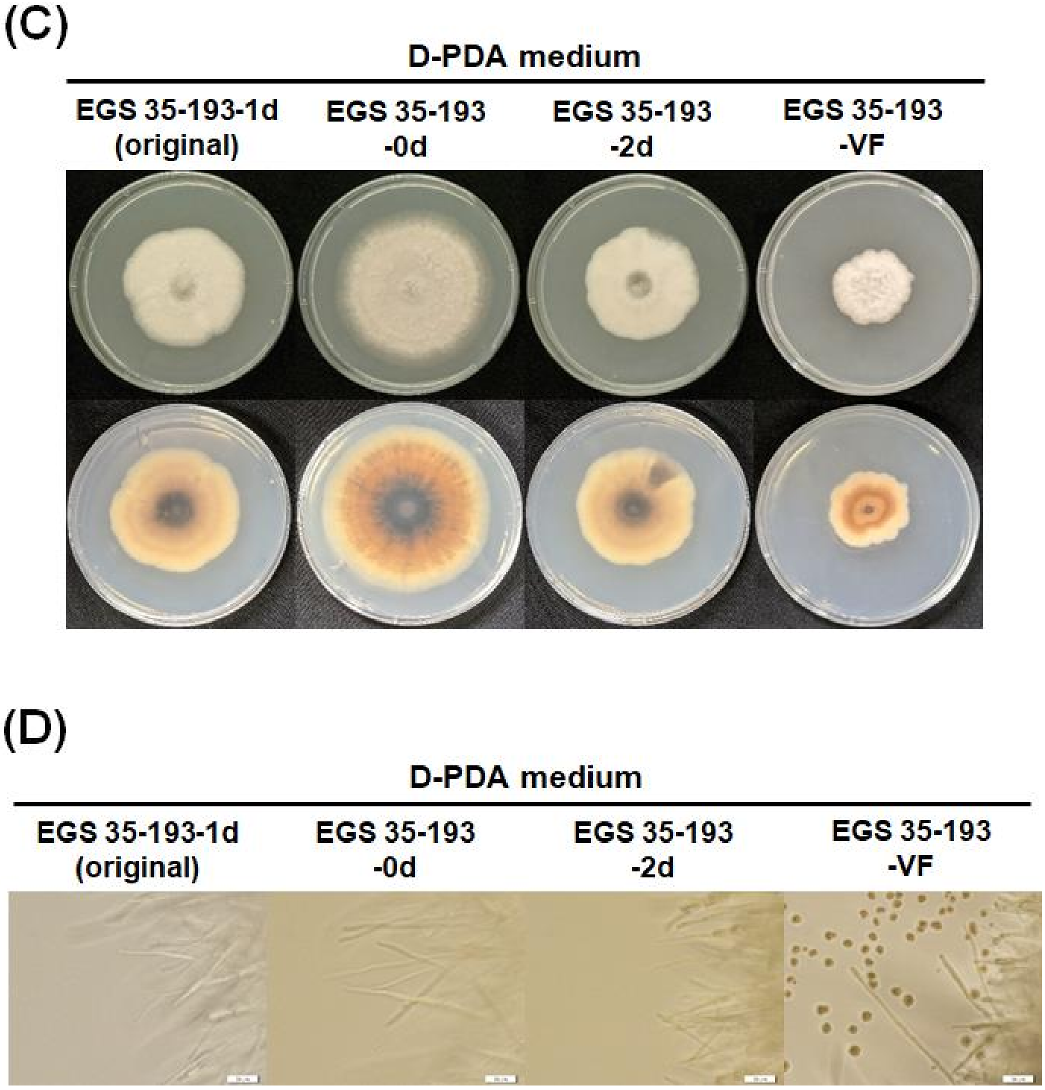


Figure 2-5. Colony morphologies of AaV1-infected isolates and AaV1-free isolate on different media.

(A) H-PDA, **(B)** V8A, and **(C)** D-PDA media. **(D)** EGS 35-193 virus-free isolate yielded unknown products while culturing on D-PDA medium.

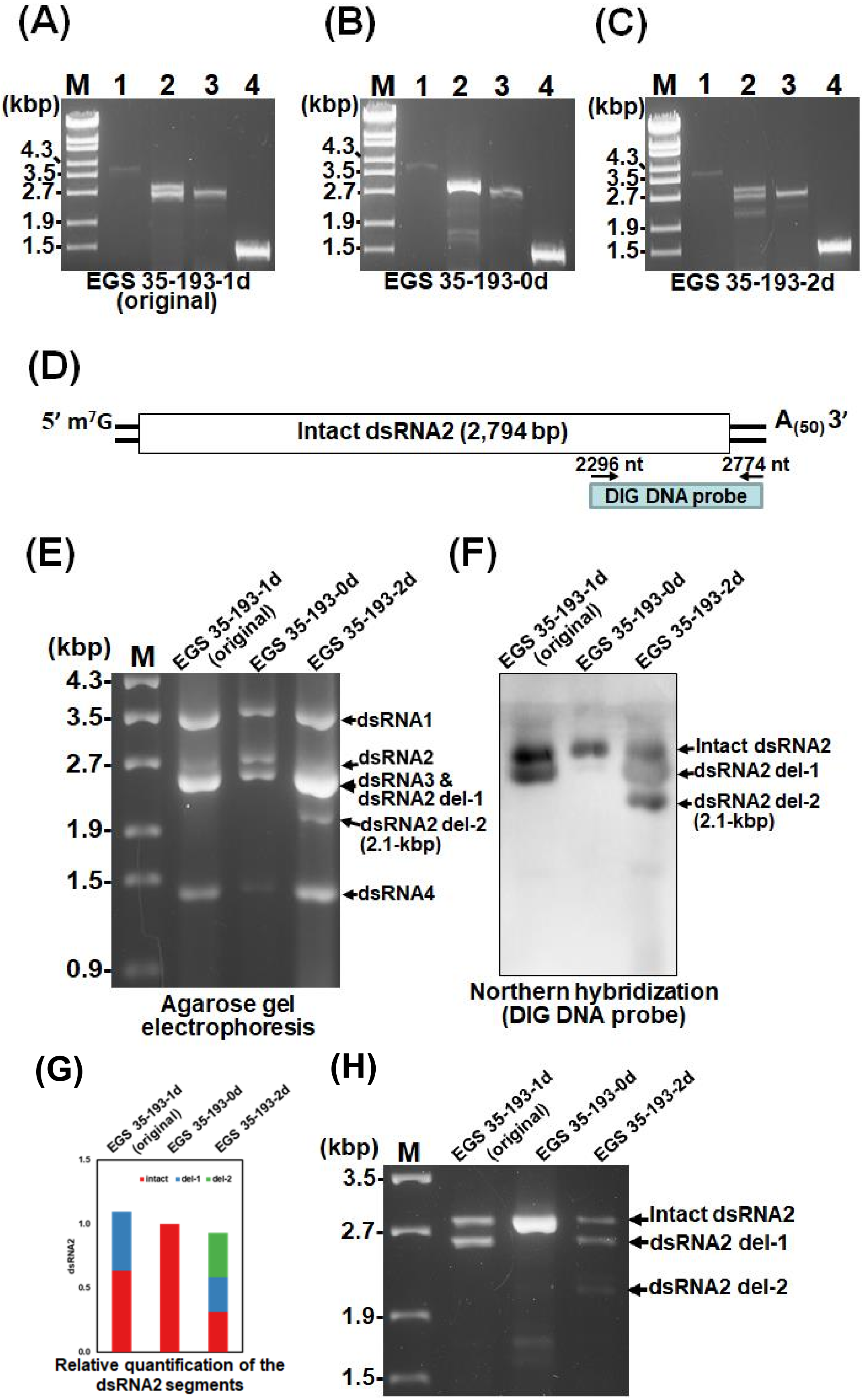
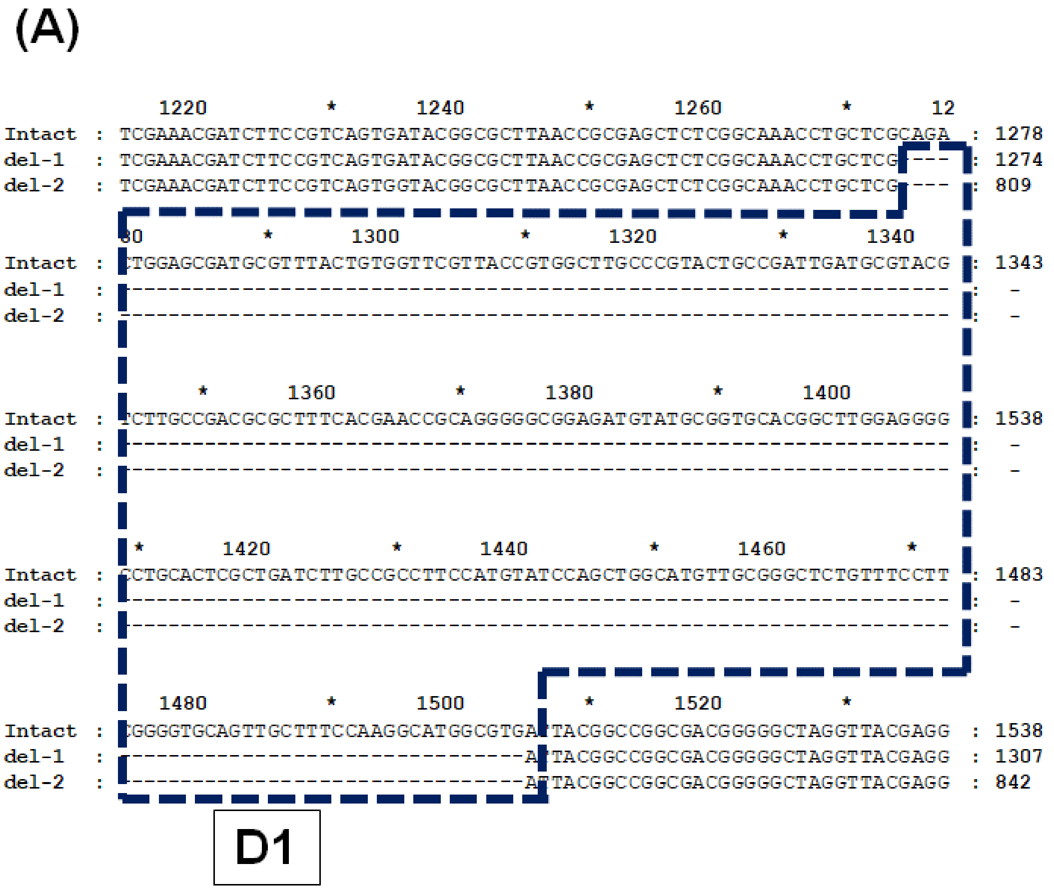

(A)
M 1 2 3 4
(kbp)
4.3
3.5
2.7
1.9
1.5
EGS 35-193-1d
(original)
(B)
M 1 2 3 4
(kbp)
4.3
3.5
2.7
1.9
1.5
EGS 35-193-0d
(C)
M 1 2 3 4
(kbp)
4.3
3.5
2.7
1.9
1.5
EGS 35-193-2d
(D)
5′ m7G
Intact dsRNA2 (2,794 bp)
A(50) 3′
2296 nt
2774 nt
DIG DNA probe
(E)
(kbp)
M
EGS 35-193-1d (original)
EGS 35-193-0d
EGS 35-193-2d
4.3
3.5
2.7
1.9
1.5
0.9
dsRNA1
dsRNA2
dsRNA3 &
dsRNA2 del-1
dsRNA2 del-2
(2.1-kbp)
dsRNA4
Agarose gel
electrophoresis
(F)
EGS 35-193-1d (original)
EGS 35-193-0d
EGS 35-193-2d
Intact dsRNA2
dsRNA2 del-1
dsRNA2 del-2
(2.1-kbp)
Northern hybridization
(DIG DNA probe)
(G)
EGS 35-193-1d (original)
EGS 35-193-0d
EGS 35-193-2d
1.5
1.0
0.5
0.0
dsRNA2
intact
del-1
del-2
Relative quantification of the
dsRNA2 segments
(H)
(kbp)
M
EGS 35-193-1d (original)
EGS 35-193-0d
EGS 35-193-2d
3.5
2.7
1.9
1.5
Intact dsRNA2
dsRNA2 del-1
dsRNA2 del-2

Figure 2-6. Analysis of the AaV1 dsRNA2-associated segments in virions purified from the three AaV1-infected isolates.

(A-C) RT-PCR detection of the full-length dsRNA1–4 segments in EGS 35-193-1d **(A)**, EGS 35-193-0d **(B)**, and EGS 35-193-2d **(C)**. RT-PCR was performed with the four primer pairs (Table 2-1) designed to amplify the full-length dsRNA1–4 segments. The RT-PCR products were separated in 1.0% agarose gels with EtBr (0.5 μg/ml) at 50 V for 1 h. Lane designation: M, 250 ng of λ-EcoT14I-digested DNA marker; 1, dsRNA1; 2, dsRNA2; 3, dsRNA3; 4, dsRNA4. **(D)** Position of the DIG DNA probe used to detect the 3′ region of AaV1 dsRNA2 in the northern hybridization experiment. The DIG DNA probe was synthesized using the probe synthesis primer pair (Table 2-1). **(E)** Agarose gel electrophoresis of dsRNA genomes extracted from the purified virus particles of the three AaV1-infected isolates. These AaV1 dsRNAs were separated in a 1.0% agarose gel with EtBr (0.5 μg/ml) at 18 V for 20 h. Lane M, 250 ng of λ-EcoT14I-digested DNA marker. **(F)** Northern hybridization to detect the AaV1 dsRNA2-associated segments. After agarose gel electrophoresis, the dsRNA genomes were denatured and blotted onto a nylon membrane, and probed with the DIG DNA probe. **(G)** Relative quantification of the three dsRNA2 segments, intact, del-1 and del-2, following northern hybridization **(F)**. The total signal in each lane was normalized using the amount of dsRNA 1, following agarose gel electrophoresis **(E)**. Quantification of individual bands was performed using Fiji/ImageJ. **(H)** Agarose gel electrophoresis of RT-PCR products amplified with the primer pair AaV1 dsRNA2-F and AaV1 dsRNA2-R, which is designed to amplify full-length AaV1 dsRNA2.



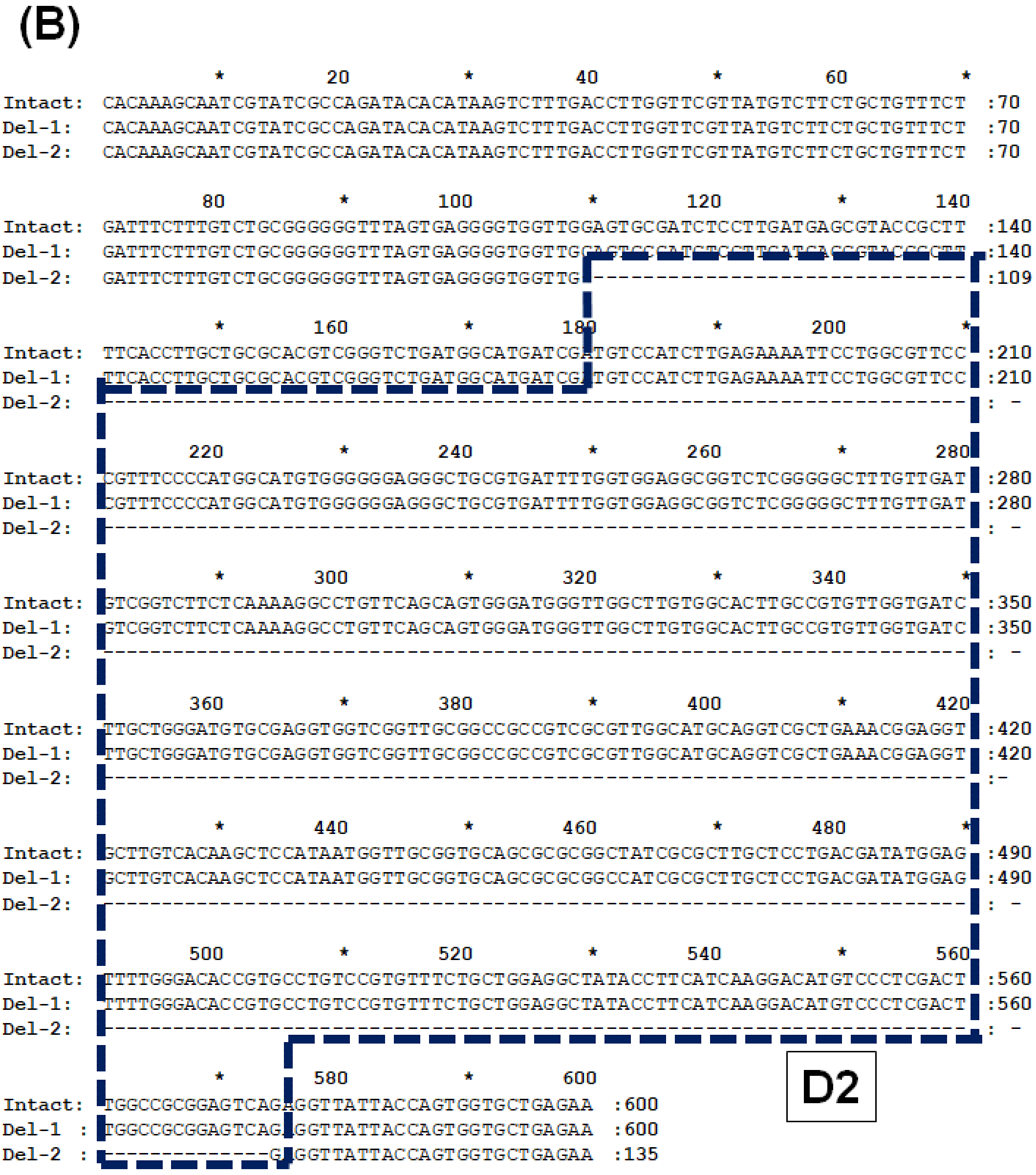
```
                    1220        *       1240        *       1260        *       12
Intact  :  TCGAAACGATCTTCCGTCAGTGATACGGCGCTTAACCGCGAGCTCTCGGCAAACCTGCTCGCAGA  : 1278
del-1   :  TCGAAACGATCTTCCGTCAGTGATACGGCGCTTAACCGCGAGCTCTCGGCAAACCTGCTCG----  : 1274
del-2   :  TCGAAACGATCTTCCGTCAGTGGTACGGCGCTTAACCGCGAGCTCTCGGCAAACCTGCTCG----  : 809

           80       *       1300        *       1320        *       1340
Intact  :  CTGGAGCGATGCGTTTACTGTGGTTCGTTACCGTGGCTTGCCCGTACTGCCGATTGATGCGTACG  : 1343
del-1   :  -----------------------------------------------------------------  :  -
del-2   :  -----------------------------------------------------------------  :  -

                *       1360        *       1380        *       1400
Intact  :  TCTTGCCGACGCGCTTTCACGAACCGCAGGGGGCGGAGATGTATGCGGTGCACGGCTTGGAGGGG  : 1538
del-1   :  -----------------------------------------------------------------  :  -
del-2   :  -----------------------------------------------------------------  :  -

            *       1420        *       1440        *       1460        *
Intact  :  CCTGCACTCGCTGATCTTGCCGCCTTCCATGTATCCAGCTGGCATGTTGCGGGCTCTGTTTCCTT  : 1483
del-1   :  -----------------------------------------------------------------  :  -
del-2   :  -----------------------------------------------------------------  :  -

              1480        *       1500        *       1520        *
Intact  :  CGGGGTGCAGTTGCTTTCCAAGGCATGGCGTGATTACGGCCGGCGACGGGGCTAGGTTACGAGG  : 1538
del-1   :  --------------------------------ATTACGGCCGGCGACGGGGCTAGGTTACGAGG  : 1307
del-2   :  --------------------------------ATTACGGCCGGCGACGGGGCTAGGTTACGAGG  : 842
```



```
                   *        20         *        40         *        60         *
Intact:   CACAAAGCAATCGTATCGCCAGATACACATAAGTCTTTGACCTTGGTTCGTTATGTCTTCTGCTGTTTCT :70
Del-1:    CACAAAGCAATCGTATCGCCAGATACACATAAGTCTTTGACCTTGGTTCGTTATGTCTTCTGCTGTTTCT :70
Del-2:    CACAAAGCAATCGTATCGCCAGATACACATAAGTCTTTGACCTTGGTTCGTTATGTCTTCTGCTGTTTCT :70

                  80         *       100         *       120         *       140
Intact:   GATTTCTTTGTCTGCGGGGGGTTTAGTGAGGGGTGGTTGGAGTGCGATCTCCTTGATGAGCGTACCGCTT :140
Del-1:    GATTTCTTTGTCTGCGGGGGGTTTAGTGAGGGGTGGTTGGAGTGCGATCTCCTTGATGAGCGTACCGCTT :140
Del-2:    GATTTCTTTGTCTGCGGGGGGTTTAGTGAGGGGTGGTTG------------------------------- :109

                   *       160         *       180         *       200         *
Intact:   TTCACCTTGCTGCGCACGTCGGGTCTGATGGCATGATCGATGTCCATCTTGAGAAAATTCCTGGCGTTCC :210
Del-1:    TTCACCTTGCTGCGCACGTCGGGTCTGATGGCATGATCGATGTCCATCTTGAGAAAATTCCTGGCGTTCC :210
Del-2:    ---------------------------------------------------------------------- : -

                 220         *       240         *       260         *       280
Intact:   CGTTTCCCCATGGCATGTGGGGGGAGGGCTGCGTGATTTTGGTGGAGGCGGTCTCGGGGGCTTTGTTGAT :280
Del-1:    CGTTTCCCCATGGCATGTGGGGGGAGGGCTGCGTGATTTTGGTGGAGGCGGTCTCGGGGGCTTTGTTGAT :280
Del-2:    ---------------------------------------------------------------------- : -

                   *       300         *       320         *       340         *
Intact:   GTCGGTCTTCTCAAAAGGCCTGTTCAGCAGTGGGATGGGTTGGCTTGTGGCACTTGCCGTGTTGGTGATC :350
Del-1:    GTCGGTCTTCTCAAAAGGCCTGTTCAGCAGTGGGATGGGTTGGCTTGTGGCACTTGCCGTGTTGGTGATC :350
Del-2:    ---------------------------------------------------------------------- : -

                 360         *       380         *       400         *       420
Intact:   TTGCTGGGATGTGCGAGGTGGTCGGTTGCGGCCGCCGTCGCGTTGGCATGCAGGTCGCTGAAACGGAGGT :420
Del-1:    TTGCTGGGATGTGCGAGGTGGTCGGTTGCGGCCGCCGTCGCGTTGGCATGCAGGTCGCTGAAACGGAGGT :420
Del-2:    ---------------------------------------------------------------------- :-

                   *       440         *       460         *       480         *
Intact:   GCTTGTCACAAGCTCCATAATGGTTGCGGTGCAGCGCGCGGCTATCGCGCTTGCTCCTGACGATATGGAG :490
Del-1:    GCTTGTCACAAGCTCCATAATGGTTGCGGTGCAGCGCGCGGCCATCGCGCTTGCTCCTGACGATATGGAG :490
Del-2:    ---------------------------------------------------------------------- : -

                 500         *       520         *       540         *       560
Intact:   TTTTGGGACACCGTGCCTGTCCGTGTTTCTGCTGGAGGCTATACCTTCATCAAGGACATGTCCCTCGACT :560
Del-1:    TTTTGGGACACCGTGCCTGTCCGTGTTTCTGCTGGAGGCTATACCTTCATCAAGGACATGTCCCTCGACT :560
Del-2:    ---------------------------------------------------------------------- : -

                   *       580         *       600
Intact:   TGGCCGCGGAGTCAGAGGTTATTACCAGTGGTGCTGAGAA :600
Del-1  :  TGGCCGCGGAGTCAGAGGTTATTACCAGTGGTGCTGAGAA :600
Del-2  :  --------------GAGGTTATTACCAGTGGTGCTGAGAA :135
```

Figure 2-7. In-frame deletion sites of AaV1 dsRNA2 genome.

(A) D1 deletion site. From nt 1,275 to nt 1,505, length: 231 bp. **(B)** D2 deletion site. From nt 113 to nt 577, length: 465 bp.

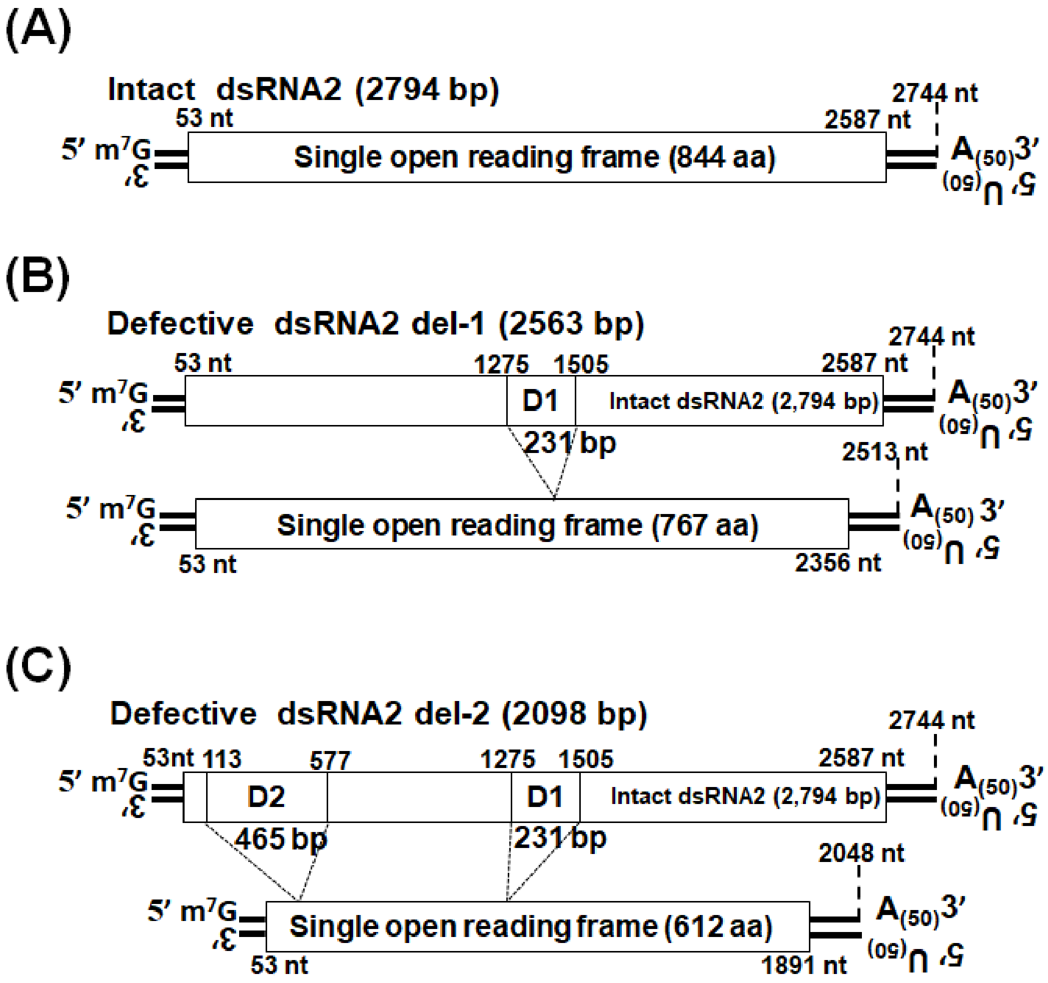


Figure 2-8. Schematic diagrams showing the genome rearrangements of the AaV1 dsRNA2 derivatives.

(A) Intact dsRNA2 has no deletion site. **(B)** Defective dsRNA2 del-1 has an in-frame deletion site (D1, 231 bp) located at nt 1275–1505 of the intact AaV1 dsRNA2. **(C)** Defective dsRNA2 del-2 has two in-frame deletion sites, D1 and D2 (465 bp), located at nt 113–577 of the intact AaV1 dsRNA2.

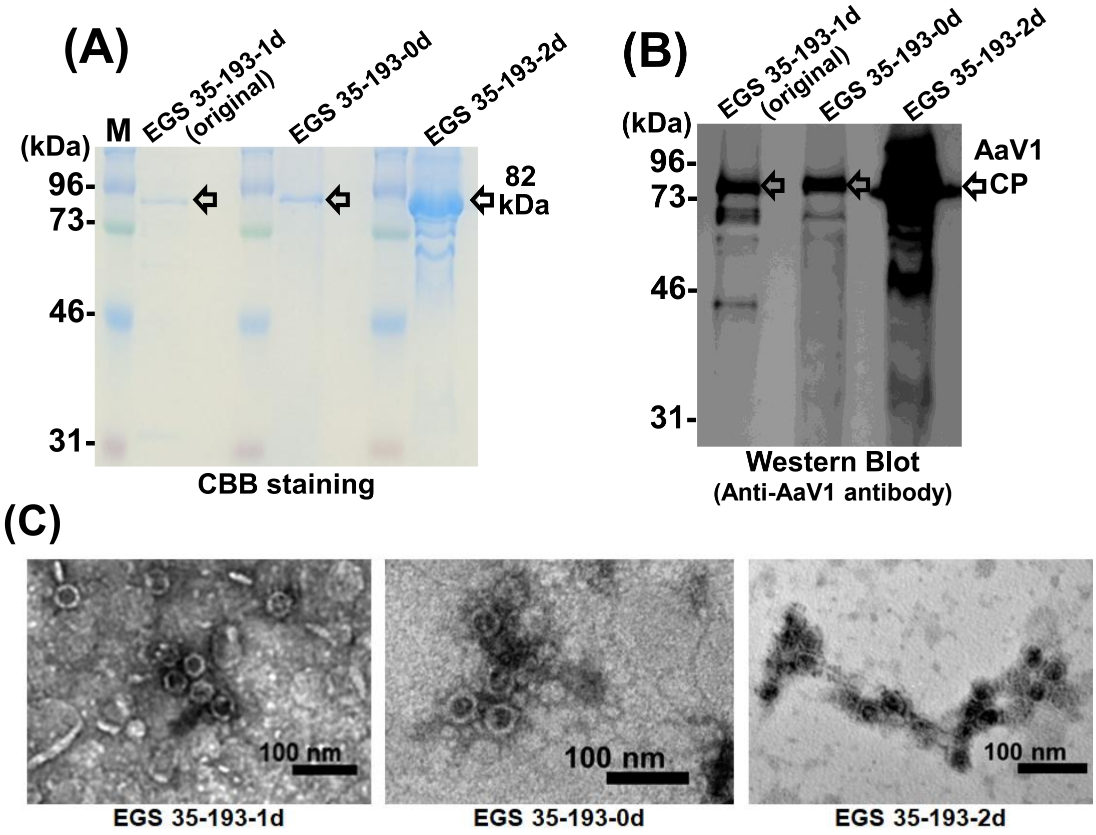
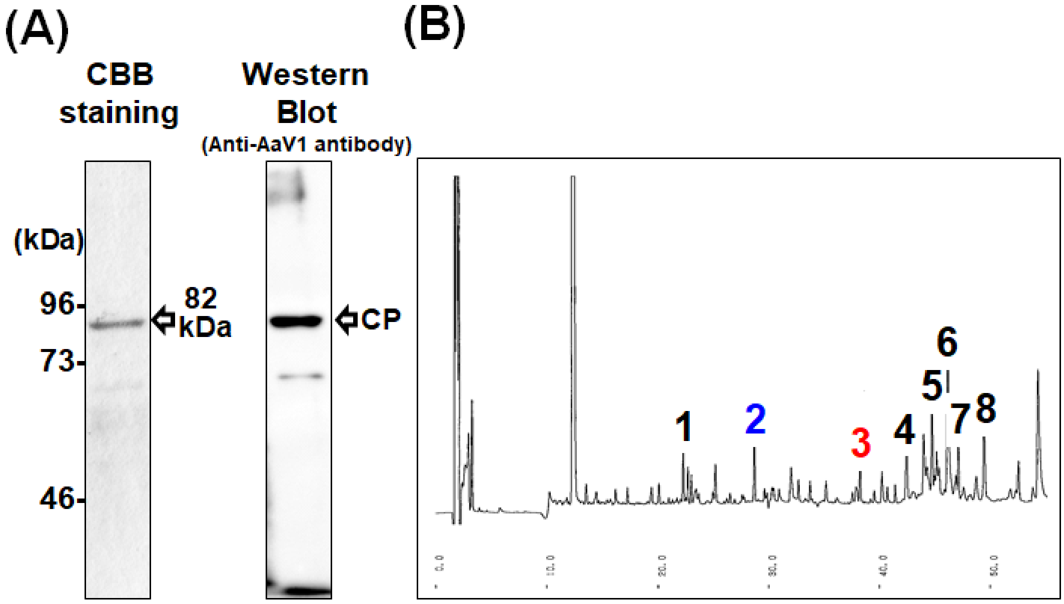
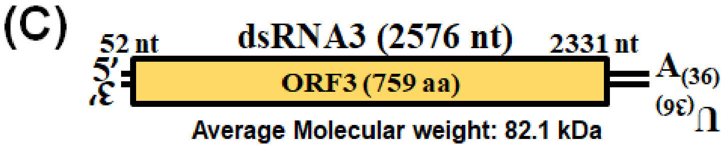


Figure 2-9. Purified virus particles from AaV1-infected isolates (EGS 35-193-1d, EGS 35-193-0d, and EGS 35-193-2d).

(A) SDS-PAGE of purified virus particles from EGS 35-193-1d, -0d, and -2d. The purified viral proteins were separated in an 8% polyacrylamide gel at 120V for 2 h, then stained with CBB. Lane M, prestained protein marker. **(B)** Western blot analysis of purified virus particles from EGS 35-193-1d, -0d, and -2d, with antiserum raised against the AaV1 virus particles from EGS 35-193-1d. **(C)** Purified virus particles of EGS 35-193-1d, EGS 35-193-0d, and EGS35-193-2d isolates. The purified virus particles were stained with 2% uranyl acetate and observed by TEM.





```
1    MATFGSAEAY RAAQLAAIDL LTRGDWGDFG GLGETSDVFD EHERFPEDPF DAPDHIDPGS 60
61   SWASVATGSA DDISREVTPT VVDYVPRREP YSVRGVDIVE PVTAFCSPAD LGMANFPGNV 120
121  SKGSATTLRR AGSAAANAGA ILGTDMLMAP KASVQAIMAR VVSALEMLQS GWDVGGPPDV 180
181  GLDVRHAARD DFDDASARYP QHAWLYIPSD WTEQEVAALV SLMVEGGPAA YRWGYARGDP 240
241  GGDEGNGQAR ERVMPAGAAW RWPGGWSNYL LIGERDRGWN VAFGGDALSV ASLSAVLRRM 300
301  VEAYGQRIYL DACRAAAVAN RAYCPPCYQA GSKERDIGTA FTSDRVVVRD GNGGHMRQIV 360
361  PARNMVKDEP GVLPRPAEGW DPVDDGGVVQ GGCGGSVWAL PVFPSQRDAD RVGRFYAPSI 420
421  DDRGPGGAGR APPHFRSVDW SPVGLTVVTN HGRRLFPWAW DPPPMAEDRV NEEGVEEADE 480
481  VADAWVRMAA CYLRKVEIVR ALEGDHGVHT REGDTRTVFG TSAHYTSYQM PRLNLDGWWP 540
541  ALIGLSVLRH DRVVPKLDRR LLRPAFTKFA ADVHLLTHRT LFESGNSVAD LSDALVGAKV 600
601  VSRFPPAYRA GVWPHVFSSV NMPYGNYECL ESGVLLGGGN ETEGVGFNVP GSWKWDGVQR 660
661  KAELDGSDAP AIRQSLRALD SVARKLYFYG GTLRLDVHPD RPVYVVRPAG SRLYHPYFVP 720
721  VRVLEDRLPS GVRYTAIGSA AHLLSPGRPT DVGRASGVI                       759
```

Figure 2-10. Characterization of the AaV1 major coat protein.

(A) SDS-PAGE and western blot of the purified AaV1 particles. Purified viral proteins from strain EGS 35-193-1d were stained with CBB (left) or immunoblotted with antiserum against the AaV1 virus particles (right). The arrows indicate the viral structural protein. (B) Reverse-phase HPLC of the 82 kDa major protein after digestion with Lysyl endopeptidase and trypsin. Peaks 2 and 3 were subjected to Edman degradation. (C) The deduced amino acid sequence of ORF3 (759 aa) written in one-letter code. The peptide sequences of peak 2 (blue) and peak 3 (red) were the same as the two regions in the predicted ORF3 peptide sequence.

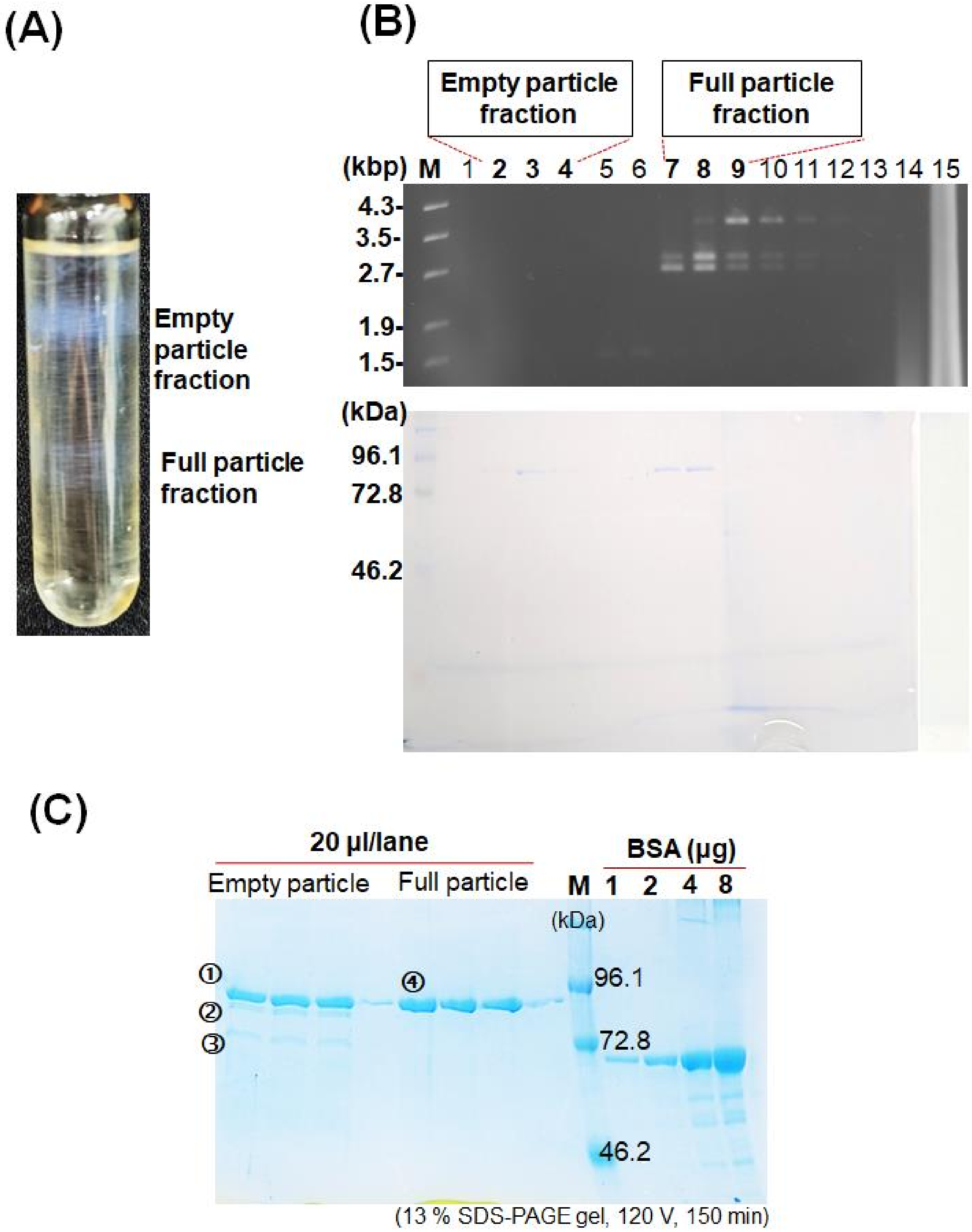
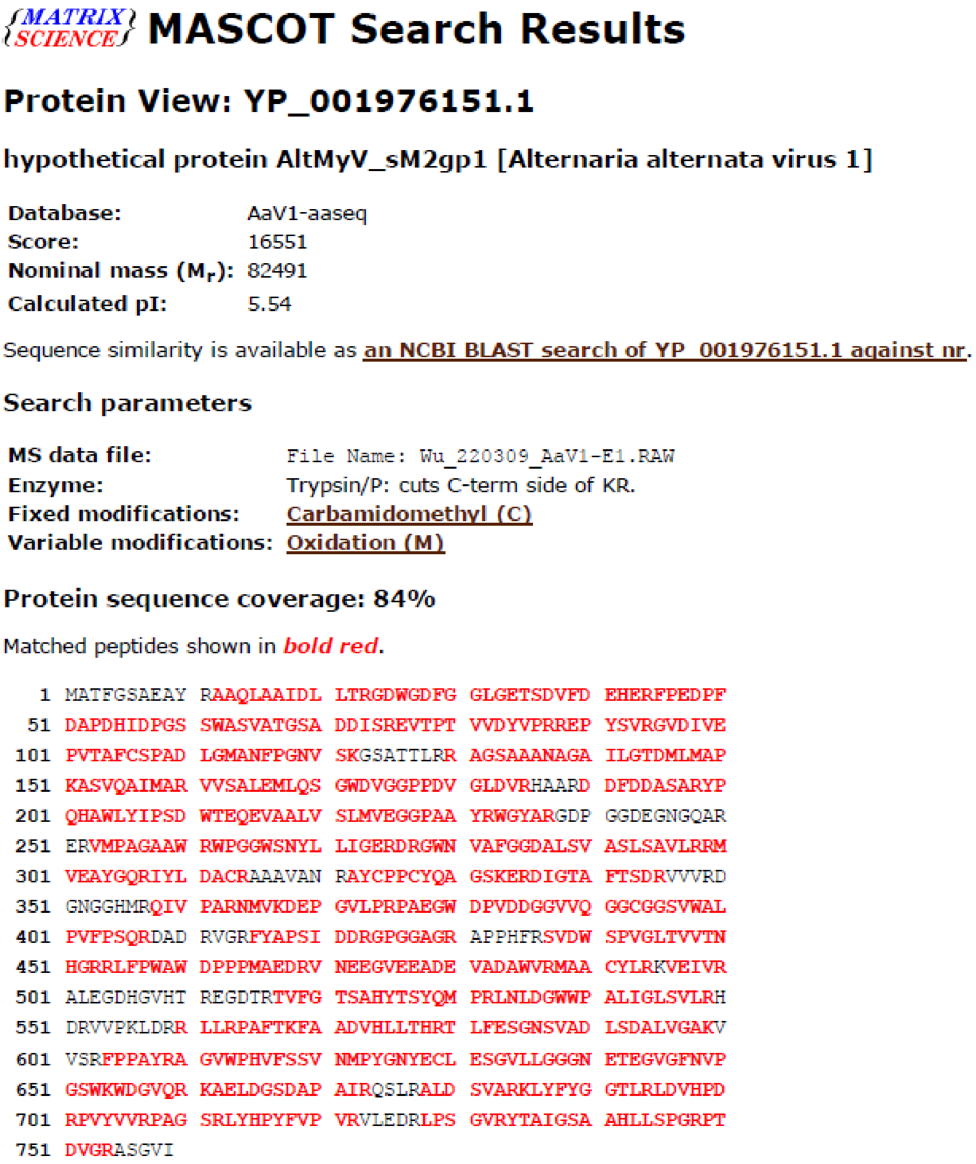


Figure 2-11. Purification of AaV1 virus particles with CsCl.

(A) CsCl isopycnic gradient centrifugation of AaV1. **(B)** Results of SDS-PAGE and dsRNA agarose electrophoresis of 15 fractions in **(A)**. **(C)** Estimation of AaV1 associated protein for LC-MS/MS. Protein band①: 6.5 μg in total; Protein band②: 0.5 μg in total; Protein band③: 0.1 μg in total; Protein band④: 10.57 μg in total.

(A)



MASCOT Search Results

Protein View: YP_001976151.1

hypothetical protein AltMyV_sM2gp1 [Alternaria alternata virus 1]

Database: AaV1-aaseq
Score: 16551
Nominal mass (M_r): 82491
Calculated pI: 5.54

Sequence similarity is available as an NCBI BLAST search of YP_001976151.1 against nr.

Search parameters

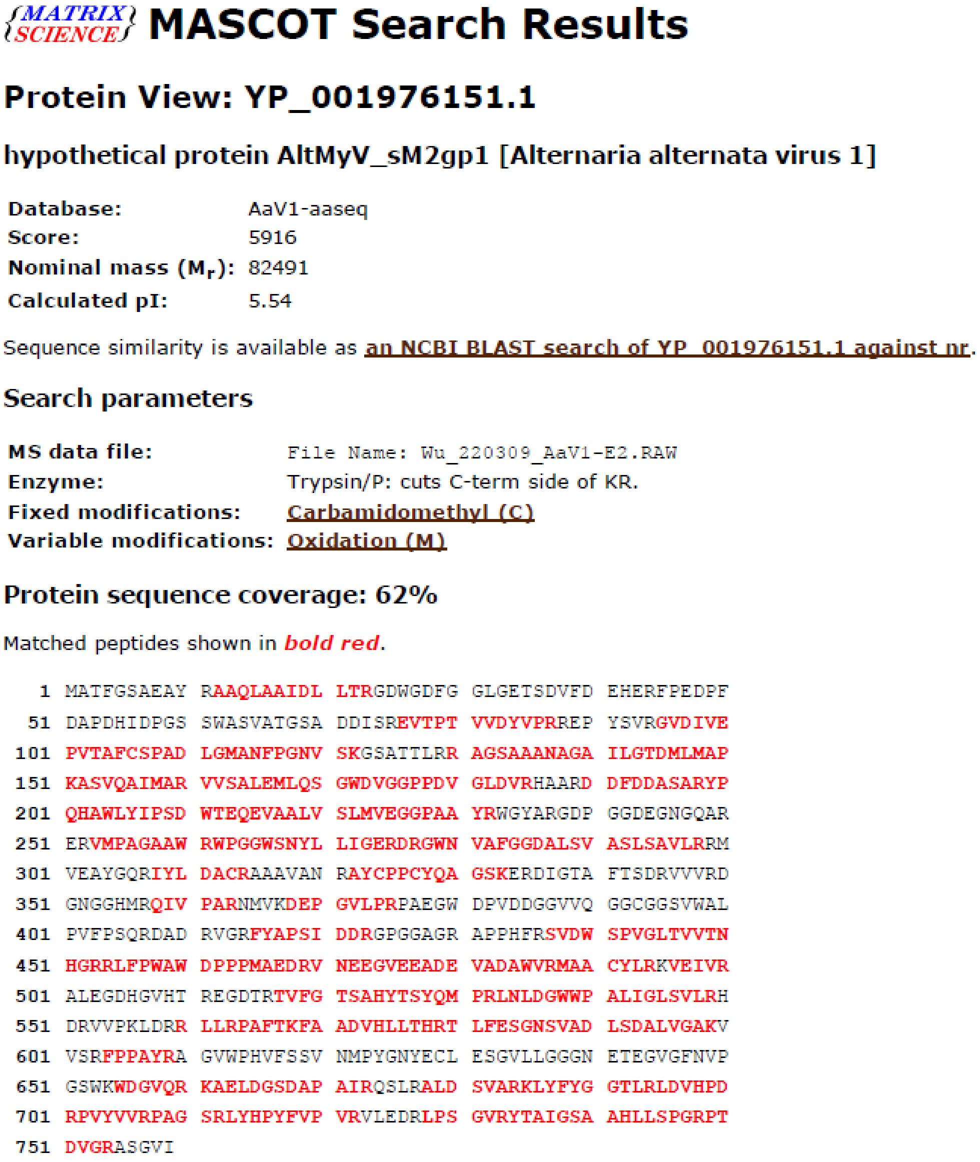
MS data file: `File Name: Wu_220309_AaV1-E1.RAW`
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 84%

Matched peptides shown in ***bold red***.

```
  1 MATFGSAEAY RAAQLAAIDL LTRGDWGDFG GLGETSDVFD EHERFPEDPF
 51 DAPDHIDPGS SWASVATGSA DDISREVTPT VVDYVPRREP YSVRGVDIVE
101 PVTAFCSPAD LGMANFPGNV SKGSATTLRR AGSAAANAGA ILGTDMLMAP
151 KASVQAIMAR VVSALEMLQS GWDVGGPPDV GLDVRHAARD DFDDASARYP
201 QHAWLYIPSD WTEQEVAALV SLMVEGGPAA YRWGYARGDP GGDEGNGQAR
251 ERVMPAGAAW RWPGGWSNYL LIGERDRGWN VAFGGDALSV ASLSAVLRRM
301 VEAYGQRIYL DACRAAAVAN RAYCPPCYQA GSKERDIGTA FTSDRVVVRD
351 GNGGHMRQIV PARNMVKDEP GVLPRPAEGW DPVDDGGVVQ GGCGGSVWAL
401 PVFPSQRDAD RVGRFYAPSI DDRGPGGAGR APPHFRSVDW SPVGLTVVTN
451 HGRRLFPWAW DPPPMAEDRV NEEGVEEADE VADAWVRMAA CYLRKVEIVR
501 ALEGDHGVHT REGDTRTVFG TSAHYTSYQM PRLNLDGWWP ALIGLSVLRH
551 DRVVPKLDRR LLRPAFTKFA ADVHLLTHRT LFESGNSVAD LSDALVGAKV
601 VSRFPPAYRA GVWPHVFSSV NMPYGNYECL ESGVLLGGGN ETEGVGFNVP
651 GSWKWDGVQR KAELDGSDAP AIRQSLRALD SVARKLYFYG GTLRLDVHPD
701 RPVYVVRPAG SRLYHPYFVP VRVLEDRLPS GVRYTAIGSA AHLLSPGRPT
751 DVGRASGVI
```

(B)



{MATRIX SCIENCE} MASCOT Search Results

Protein View: YP_001976151.1

hypothetical protein AltMyV_sM2gp1 [Alternaria alternata virus 1]

Database:	AaV1-aaseq
Score:	5916
Nominal mass (M_r):	82491
Calculated pI:	5.54

Sequence similarity is available as an NCBI BLAST search of YP_001976151.1 against nr.

Search parameters

MS data file:	File Name: Wu_220309_AaV1-E2.RAW
Enzyme:	Trypsin/P: cuts C-term side of KR.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Oxidation (M)

Protein sequence coverage: 62%

Matched peptides shown in ***bold red***.

1 MATFGSAEAY R**AAQLAAIDL** **LTR**GDWGDFG GLGETSDVFD EHERFPEDPF

51 DAPDHIDPGS SWASVATGSA DDISR**EVTPT** **VVDYVPR**REP YSVR**GVDIVE**

101 **PVTAFCSPAD** **LGMANFPGNV** **SK**GSATTLRR **AGSAAANAGA** **ILGTDMLMAP**

151 **KASVQAIMAR** **VVSALEMLQS** **GWDVGGPPDV** **GLDVR**HAARD **DFDDASARYP**

201 **QHAWLYIPSD** **WTEQEVAALV** **SLMVEGGPAA** **YR**WGYARGDP GGDEGNGQAR

251 ER**VMPAGAAW** **RWPGGWSNYL** **LIGERDRGWN** **VAFGGDALSV** **ASLSAVLR**RM

301 VEAYGQR**IYL** **DACR**AAAVAN R**AYCPPCYQA** **GSK**ERDIGTA FTSDRVVVRD

351 GNGGHMR**QIV** **PAR**NMVK**DEP** **GVLPR**PAEGW DPVDDGGVVQ GGCGGSVWAL

401 PVFPSQRDAD RVGR**FYAPSI** **DDR**GPGGAGR APPHFR**SVDW** **SPVGLTVVTN**

451 **HGRRLFPWAW** **DPPPMAEDRV** **NEEGVEEADE** **VADAWVRMAA** **CYLR**K**VEIVR**

501 ALEGDHGVHT REGDTR**TVFG** **TSAHYTSYQM** **PRLNLDGWWP** **ALIGLSVLR**H

551 DRVVPKLDR**R** **LLRPAFTKFA** **ADVHLLTHRT** **LFESGNSVAD** **LSDALVGAK**V

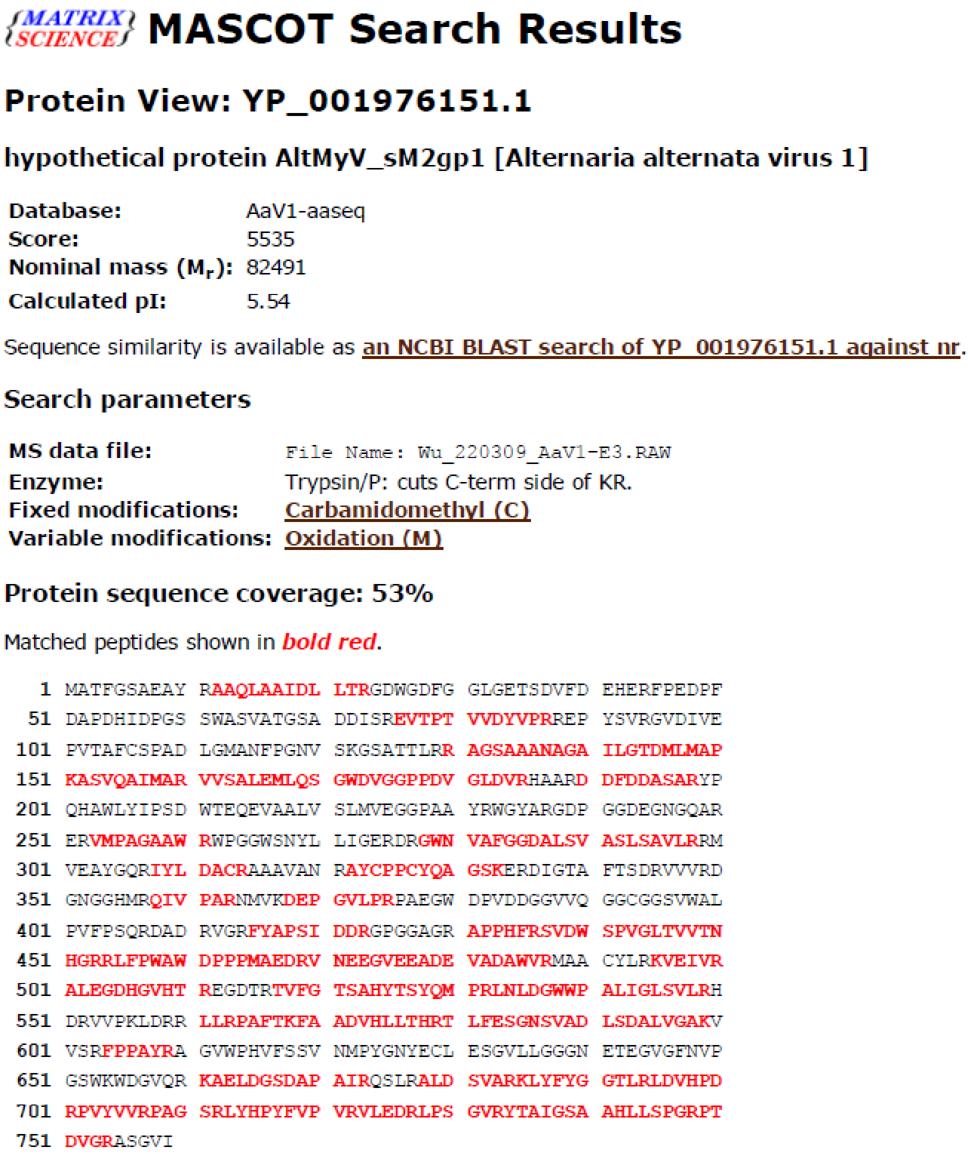
601 VSR**FPPAYR**A GVWPHVFSSV NMPYGNYECL ESGVLLGGGN ETEGVGFNVP

651 GSWK**WDGVQR** **KAELDGSDAP** **AIR**QSLR**ALD** **SVARKLYFYG** **GTLRLDVHPD**

701 **RPVYVVRPAG** **SRLYHPYFVP** **VR**VLEDR**LPS** **GVRYTAIGSA** **AHLLSPGRPT**

751 **DVGR**ASGVI

(C)



MASCOT Search Results

Protein View: YP_001976151.1

hypothetical protein AltMyV_sM2gp1 [Alternaria alternata virus 1]

Database:	AaV1-aaseq
Score:	5535
Nominal mass (M_r):	82491
Calculated pI:	5.54

Sequence similarity is available as an NCBI BLAST search of YP_001976151.1 against nr.

Search parameters

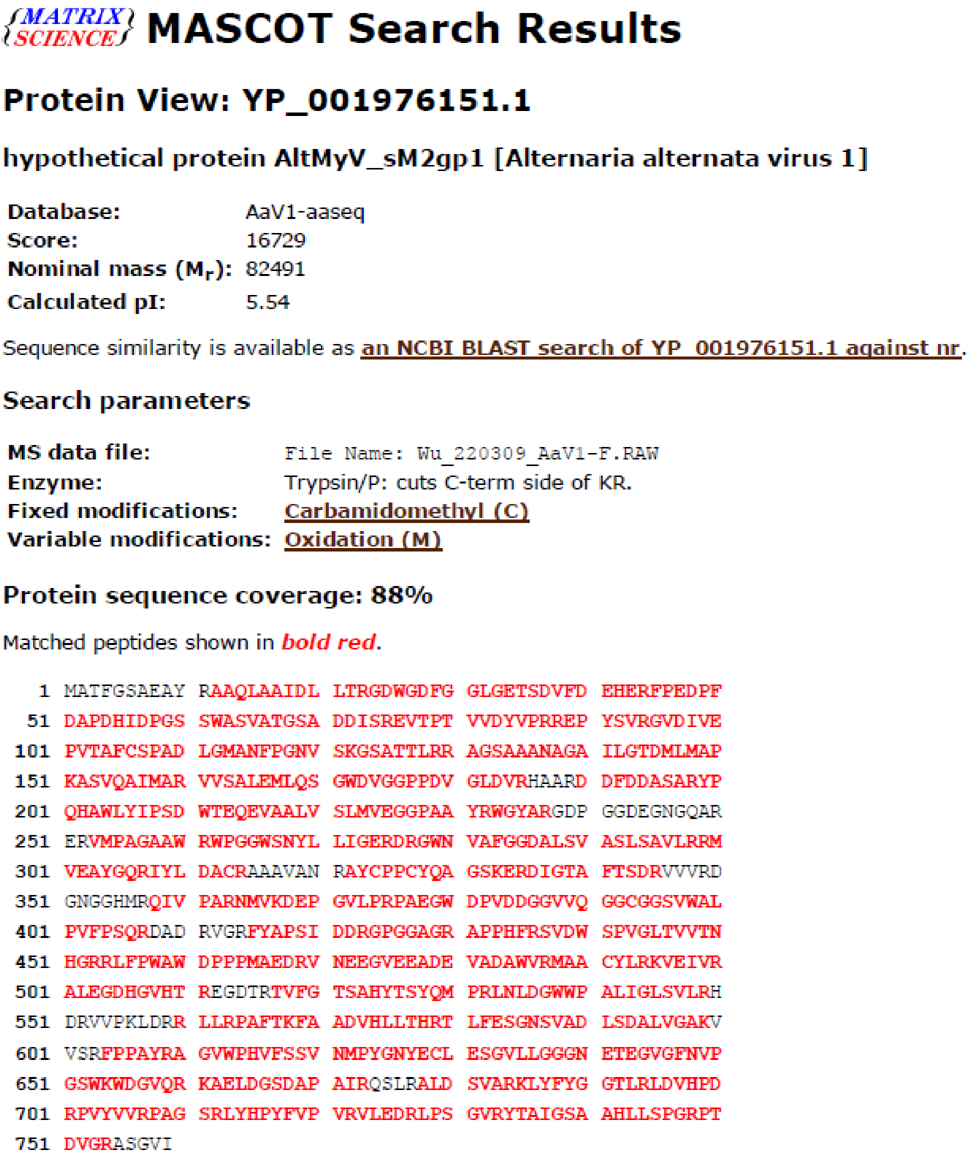
MS data file:	File Name: Wu_220309_AaV1-E3.RAW
Enzyme:	Trypsin/P: cuts C-term side of KR.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Oxidation (M)

Protein sequence coverage: 53%

Matched peptides shown in ***bold red***.

1 MATFGSAEAY R**AAQLAAIDL LTR**GDWGDFG GLGETSDVFD EHERFPEDPF
51 DAPDHIDPGS SWASVATGSA DDISR**EVTPT VVDYVPR**REP YSVRGVDIVE
101 PVTAFCSPAD LGMANFPGNV SKGSATTLR**R AGSAAANAGA ILGTDMLMAP**
151 **KASVQAIMAR VVSALEMLQS GWDVGGPPDV GLDVR**HAAR**D DFDDASAR**YP
201 QHAWLYIPSD WTEQEVAALV SLMVEGGPAA YRWGYARGDP GGDEGNGQAR
251 ER**VMPAGAAW R**WPGGWSNYL LIGERDR**GWN VAFGGDALSV ASLSAVLR**RM
301 VEAYGQR**IYL DACR**AAAVAN R**AYCPPCYQA GSK**ERDIGTA FTSDRVVVRD
351 GNGGHMR**QIV PAR**NMVK**DEP GVLPR**PAEGW DPVDDGGVVQ GGCGGSVWAL
401 PVFPSQRDAD RVGR**FYAPSI DDR**GPGGAGR **APPHFRSVDW SPVGLTVVTN**
451 **HGRRLFPWAW DPPPMAEDRV NEEGVEEADE VADAWVR**MAA CYLR**KVEIVR**
501 **ALEGDHGVHT** REGDTR**TVFG TSAHYTSYQM PRLNLDGWWP ALIGLSVLR**H
551 DRVVPKLDRR **LLRPAFTKFA ADVHLLTHRT LFESGNSVAD LSDALVGAK**V
601 VSR**FPPAYR**A GVWPHVFSSV NMPYGNYECL ESGVLLGGGN ETEGVGFNVP
651 GSWKWDGVQR **KAELDGSDAP AIR**QSLR**ALD SVARKLYFYG GTLRLDVHPD**
701 **RPVYVVRPAG SRLYHPYFVP VRVLEDRLPS GVRYTAIGSA AHLLSPGRPT**
751 **DVGR**ASGVI

(D)



MASCOT Search Results

Protein View: YP_001976151.1

hypothetical protein AltMyV_sM2gp1 [Alternaria alternata virus 1]

Database: AaV1-aaseq
Score: 16729
Nominal mass (M_r): 82491
Calculated pI: 5.54

Sequence similarity is available as an NCBI BLAST search of YP_001976151.1 against nr.

Search parameters

MS data file: File Name: Wu_220309_AaV1-F.RAW
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 88%

Matched peptides shown in ***bold red***.

```
  1 MATFGSAEAY RAAQLAAIDL LTRGDWGDFG GLGETSDVFD EHERFPEDPF
 51 DAPDHIDPGS SWASVATGSA DDISREVTPT VVDYVPRREP YSVRGVDIVE
101 PVTAFCSPAD LGMANFPGNV SKGSATTLRR AGSAAANAGA ILGTDMLMAP
151 KASVQAIMAR VVSALEMLQS GWDVGGPPDV GLDVRHAARD DFDDASARYP
201 QHAWLYIPSD WTEQEVAALV SLMVEGGPAA YRWGYARGDP GGDEGNGQAR
251 ERVMPAGAAW RWPGGWSNYL LIGERDRGWN VAFGGDALSV ASLSAVLRRM
301 VEAYGQRIYL DACRAAAVAN RAYCPPCYQA GSKERDIGTA FTSDRVVVRD
351 GNGGHMRQIV PARNMVKDEP GVLPRPAEGW DPVDDGGVVQ GGCGGSVWAL
401 PVFPSQRDAD RVGRFYAPSI DDRGPGGAGR APPHFRSVDW SPVGLTVVTN
451 HGRRLFPWAW DPPPMAEDRV NEEGVEEADE VADAWVRMAA CYLRKVEIVR
501 ALEGDHGVHT REGDTRTVFG TSAHYTSYQM PRLNLDGWWP ALIGLSVLRH
551 DRVVPKLDRR LLRPAFTKFA ADVHLLTHRT LFESGNSVAD LSDALVGAKV
601 VSRFPPAYRA GVWPHVFSSV NMPYGNYECL ESGVLLGGGN ETEGVGFNVP
651 GSWKWDGVQR KAELDGSDAP AIRQSLRALD SVARKLYFYG GTLRLDVHPD
701 RPVYVVRPAG SRLYHPYFVP VRVLEDRLPS GVRYTAIGSA AHLLSPGRPT
751 DVGRASGVI
```

Figure 2-12. Results of LC-MS/MS of AaV1 associated proteins analysis.

(A) Protein band① **(B)** Protein band② **(C)** Protein band③ **(D)** Protein band④ in Fig. 2-11C. All the proteins correspond to deduced amino acids of AaV1 ORF3.products.

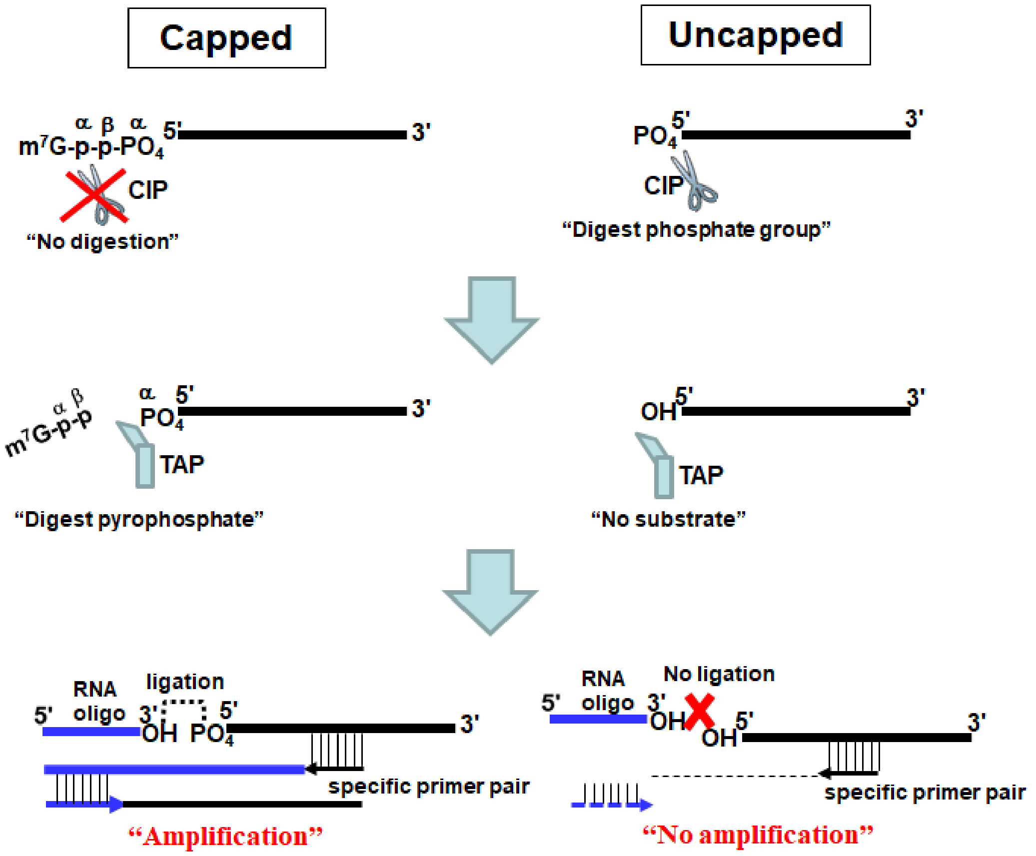
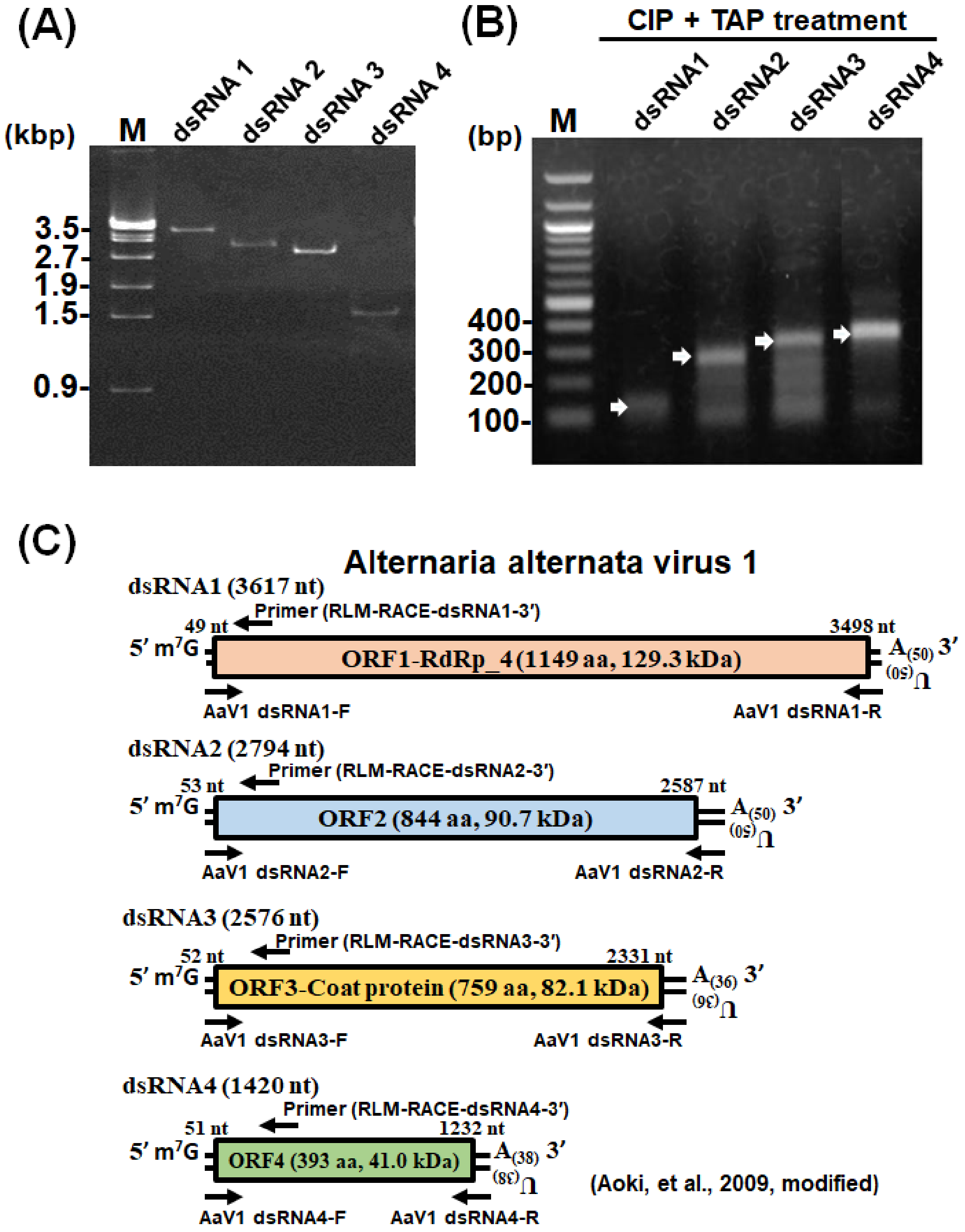


Figure 2-13. Flowchart of RLM-RACE for detecting m^7G-cap structures in AaV1 dsRNA genomes.



(Aoki, et al., 2009, modified)

Figure 2-14. Detection of 5′ cap structures on each AaV1 dsRNA segment by RLM-RACE.

(A) Electrophoresis of the separately excised AaV1 dsRNA segments in a 1% agarose gel with EtBr (0.5 µg/ml) at 50 V for 1 h (Mupid-2plus, Takara Bio, Japan). **(B)** Results of the RLM-RACE analysis, confirming the presence of 5′ cap structures on each AaV1 dsRNA segment. The separately excised dsRNAs were subjected to the RLM-RACE procedure (Figure 2-13), then electrophoresed in a 1% agarose gel with EtBr (0.5 µg/ml) at 100 V for 0.5 h. Lane M, 100 bp DNA ladder. The arrows indicate the amplified target bands. **(C)** Diagrams of the AaV1 dsRNA1, 2, 3, and 4 segments showing the primer pairs used for RLM-RACE and for amplification of each full-length dsRNA segment.

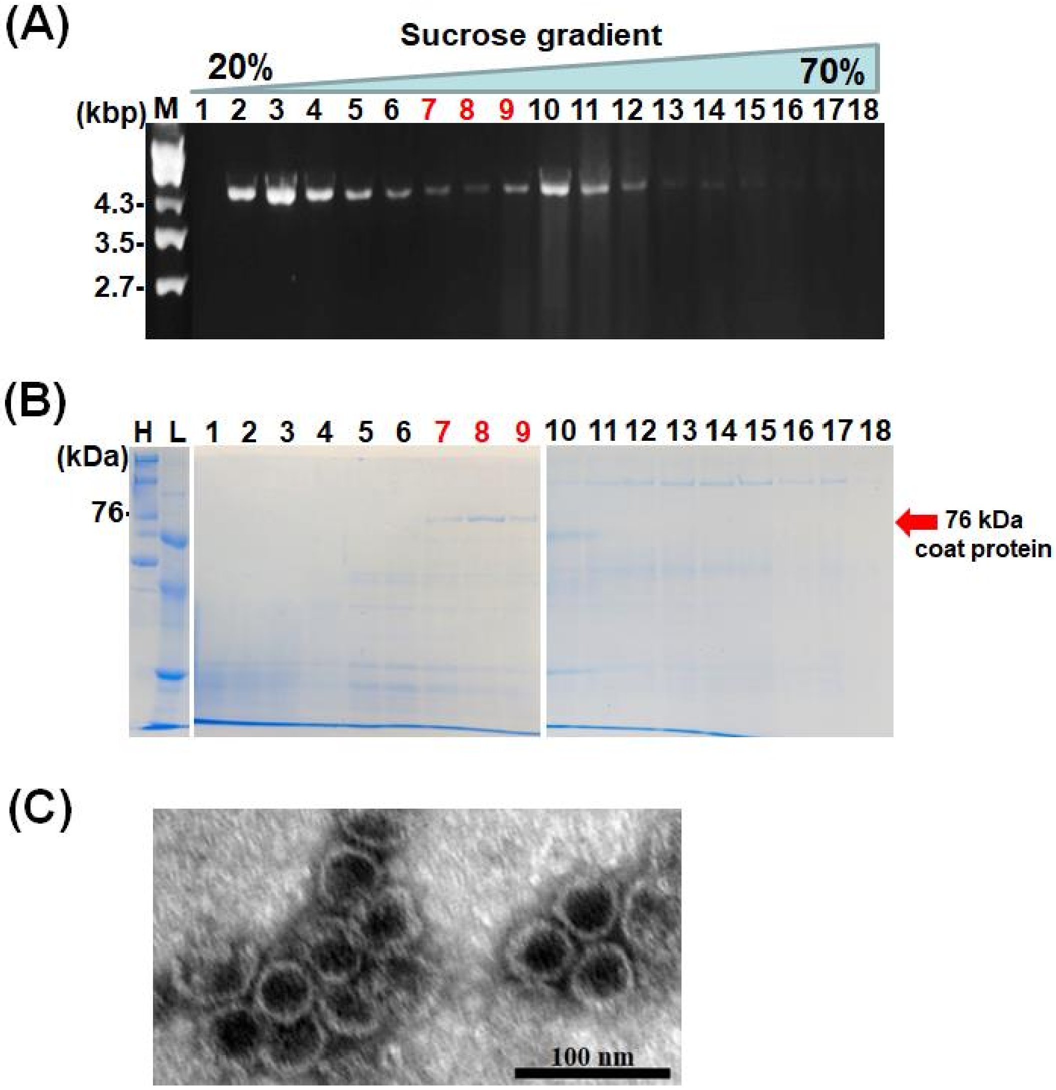


Figure 2-15. ScV-L-A virus purification.

The virus purification followed the published article (Powilleit *et al*., 2007), and resolved with 20-70% sucrose gradient, then fractioned into 18 fractions. The ScV-L-A viral particles **(C)** from fractions 7 to 9 (lanes 7-9), which showed 4.6 kb dsRNA genome **(A)** and 76 kDa coat protein **(B)**, were stained with 2% uranyl acetate and observed by TEM.

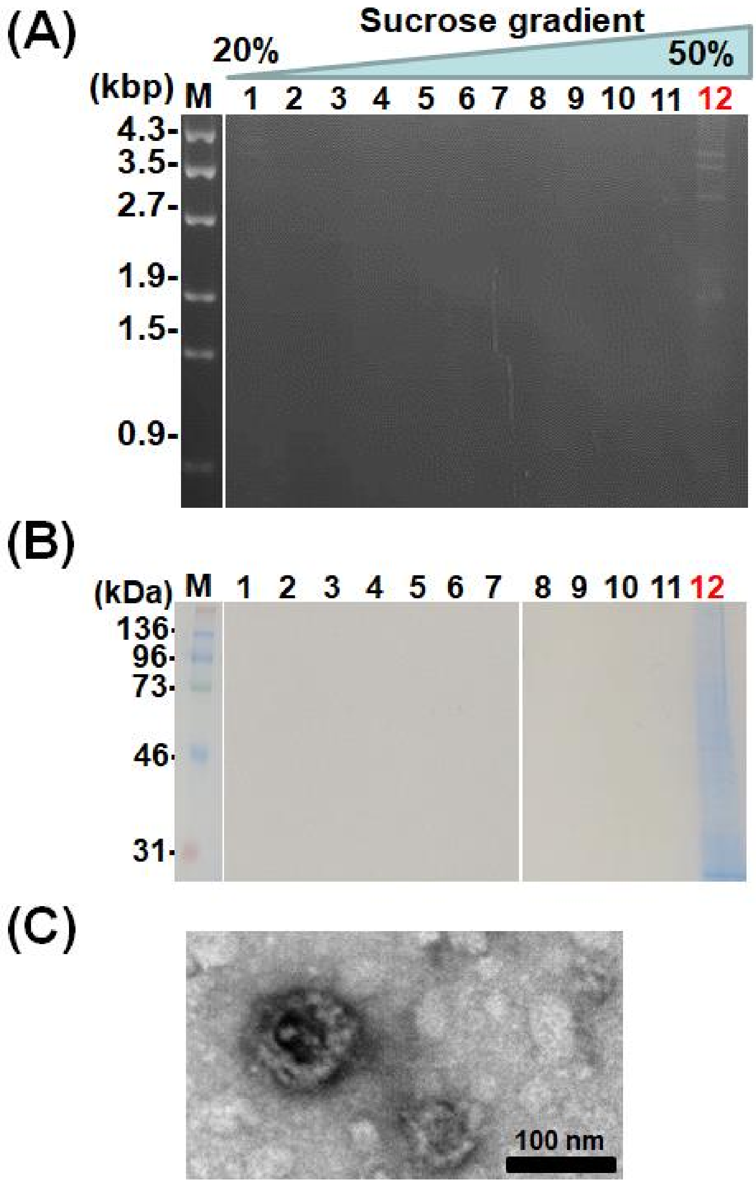


Figure 2-16. MyRV1 virus purification.

The virus purification followed the published article (Hillman *et al*., 2004), and resolved with 20-50% sucrose gradient, then fractioned into 12 fractions. The MyRV1 viral particles **(C)** from fractions 12 (lanes 12), which showed dsRNA genomes **(A)**, were stained with 2% uranyl acetate and observed by TEM. The specific protein bands were not obvious in the SDS-PAGE **(B)**.

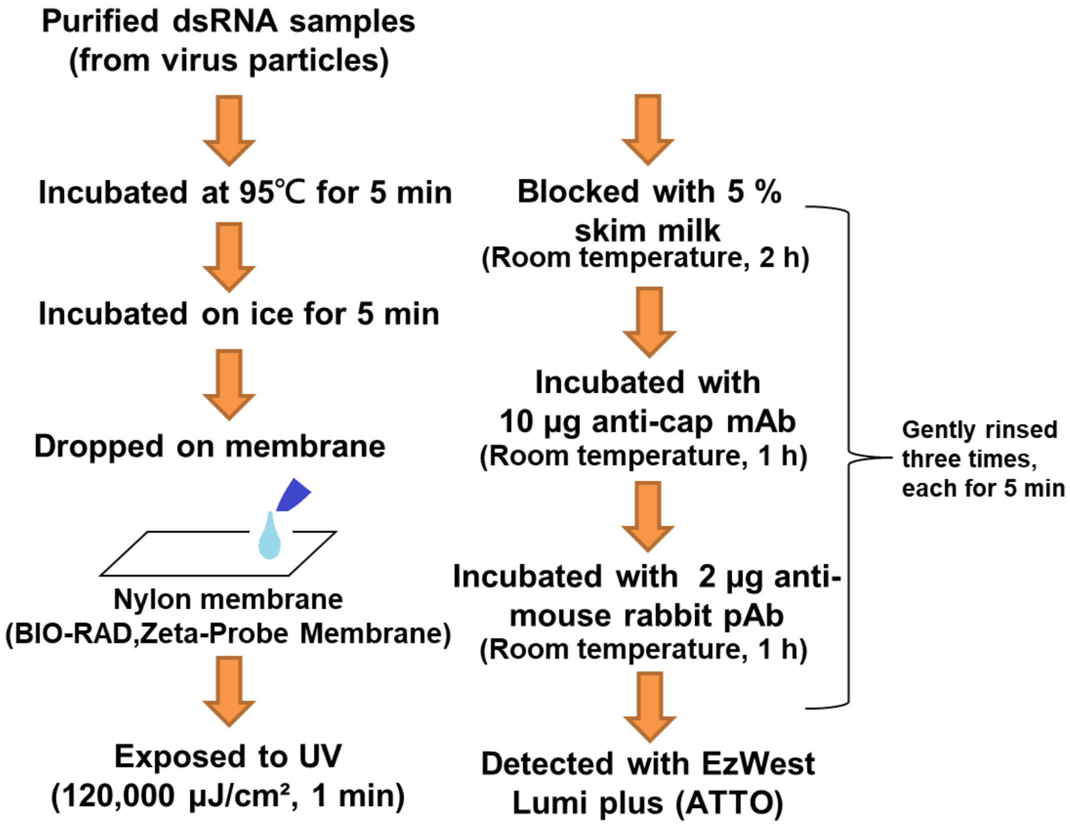


Figure 2-17. Flowchart of RNA dot blot for detecting m^{7}G-cap structures in AaV1 dsRNA genomes.

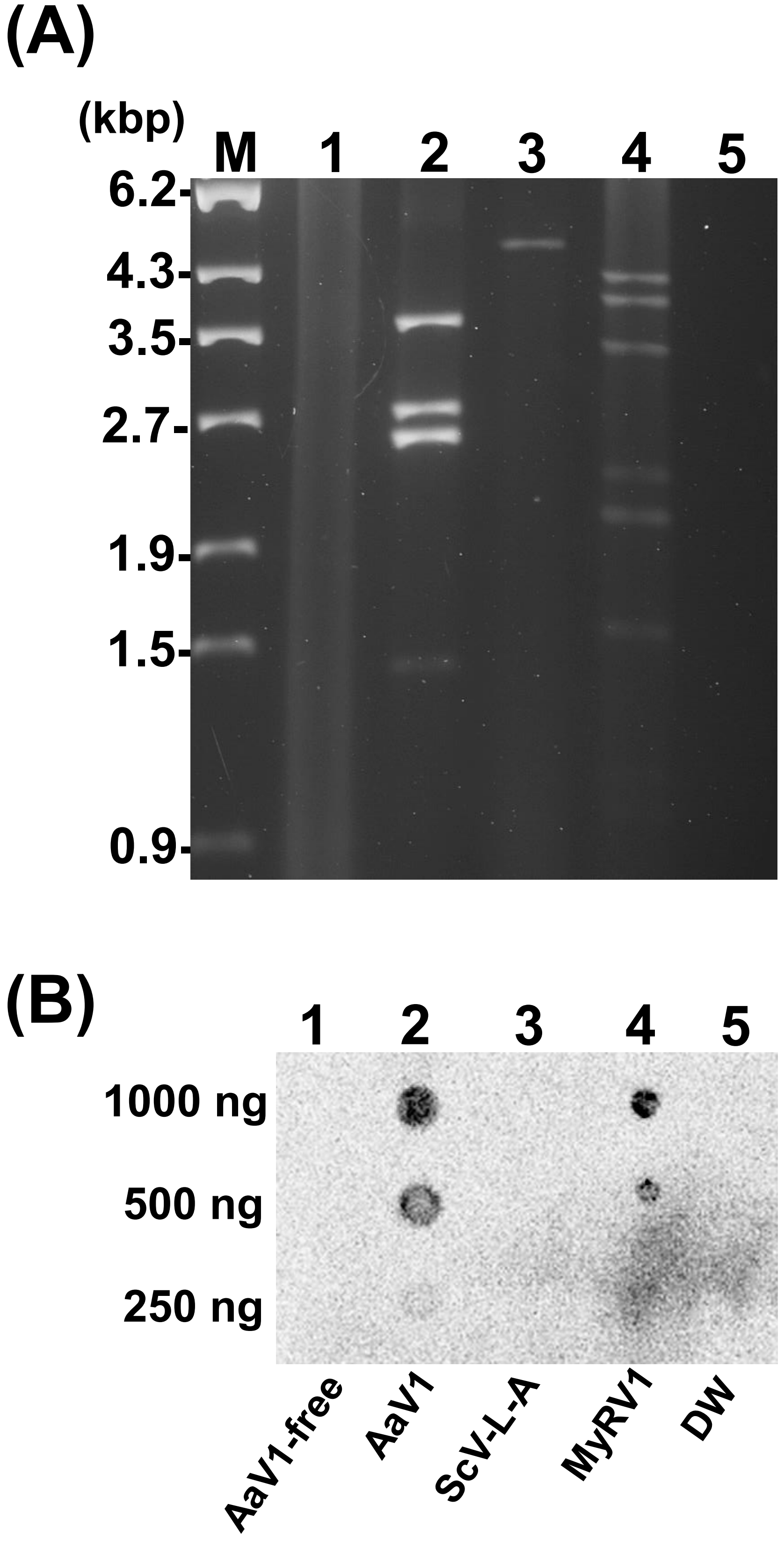
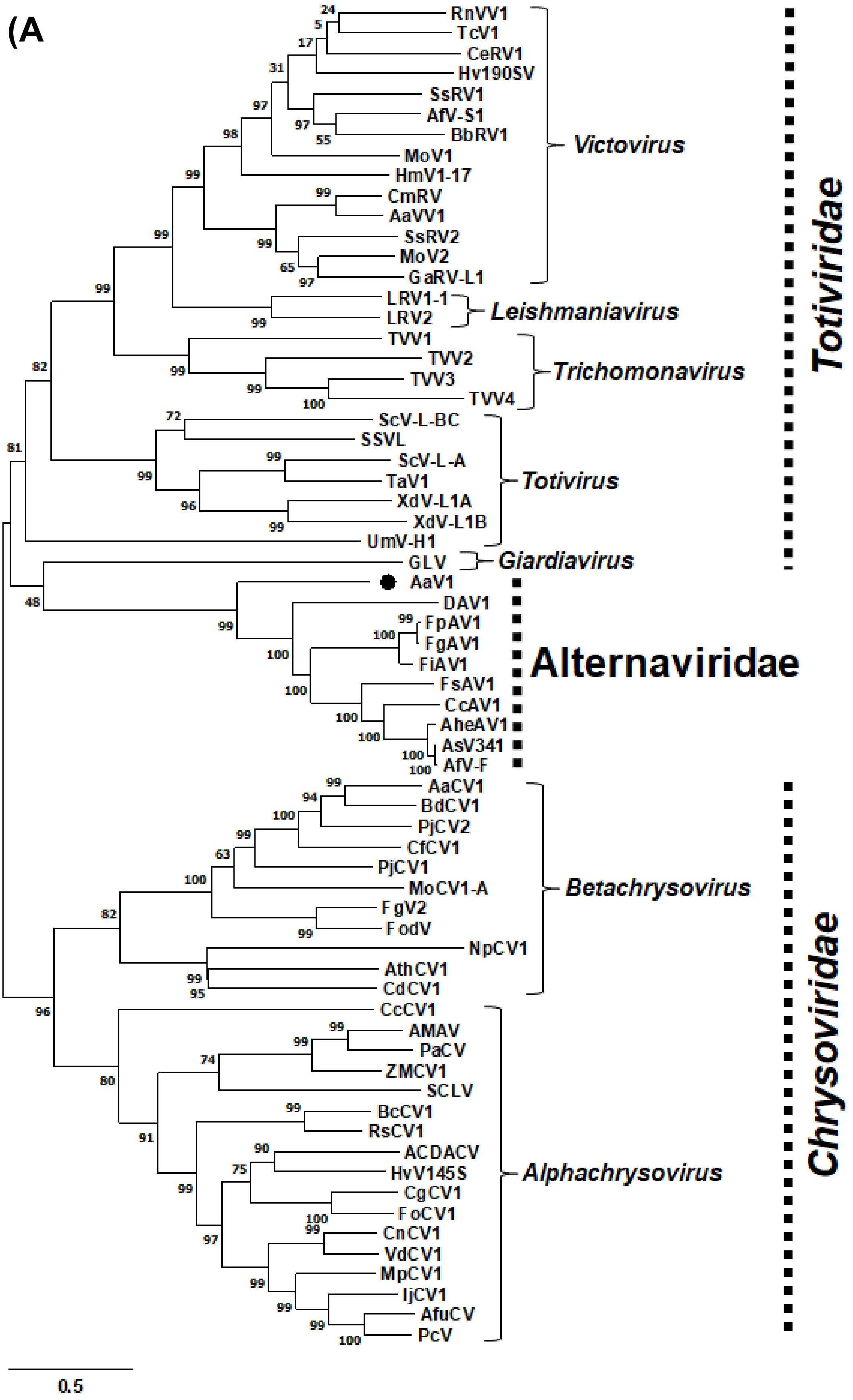

(A)
(kbp)
M
1
2
3
4
5
6.2-
4.3-
3.5-
2.7-
1.9-
1.5-
0.9-
(B)
1
2
3
4
5
1000 ng
500 ng
250 ng
AaV1-free
AaV1
ScV-L-A
MyRV1
DW

Figure 2-18. Detection of the 5′ cap structures in the AaV1 dsRNA genome by RNA dot blot with the anti-m^{7}G-Cap mAb.

(A) Viral dsRNAs from purified virus particles. The dsRNAs were isolated from virus particles, then electrophoresed in 1 % agarose gel with EtBr (0.5 µg/ml) at 18 V for 20 h. Lane designation: M, 250 ng of λ-EcoT14I-digested DNA marker; 1, AaV1-free; 2, AaV1 dsRNAs; 3, ScV-L-A dsRNA; 4, MyRV1 dsRNAs; 5, DW (distilled water, no template control). **(B)** RNA dot blot assay with the anti-m^{7}G-Cap mAb. The dsRNA solutions (1000, 500, and 250 ng/µl) were dropped on the nylon membrane then probed with anti-m^{7}G-Cap mAb. The AaV1 dsRNAs and the MyRV1 dsRNAs (positive control) showed positive signals, while the AaV1-free sample, the ScV-L-A dsRNA (negative control) and DW showed no signal.


(A
RnVV1
TcV1
CeRV1
Hv190SV
SsRV1
AfV-S1
BbRV1
MoV1
HmV1-17
CmRV
AaVV1
SsRV2
MoV2
GaRV-L1
Victovirus
LRV1-1
LRV2
Leishmaniavirus
TVV1
TVV2
TVV3
TVV4
Trichomonavirus
ScV-L-BC
SSVL
ScV-L-A
TaV1
XdV-L1A
XdV-L1B
UmV-H1
Totivirus
GLV
Giardiavirus
Totiviridae
AaV1
DAV1
FpAV1
FgAV1
FiAV1
FsAV1
CcAV1
AheAV1
AsV341
AfV-F
Alternaviridae
AaCV1
BdCV1
PjCV2
CfCV1
PjCV1
MoCV1-A
FgV2
FodV
NpCV1
AthCV1
CdCV1
Betachrysovirus
CcCV1
AMAV
PaCV
ZMCV1
SCLV
BcCV1
RsCV1
ACDACV
HvV145S
CgCV1
FoCV1
CnCV1
VdCV1
MpCV1
IjCV1
AfuCV
PcV
Alphachrysovirus
Chrysoviridae
24
5
17
31
97
97
55
98
99
99
99
99
65
97
99
99
82
99
99
100
72
81
99
99
96
99
48
99
99
100
100
100
100
100
100
100
99
94
100
99
63
100
82
99
99
95
96
99
99
74
80
99
91
90
75
99
100
99
97
99
99
99
100
0.5

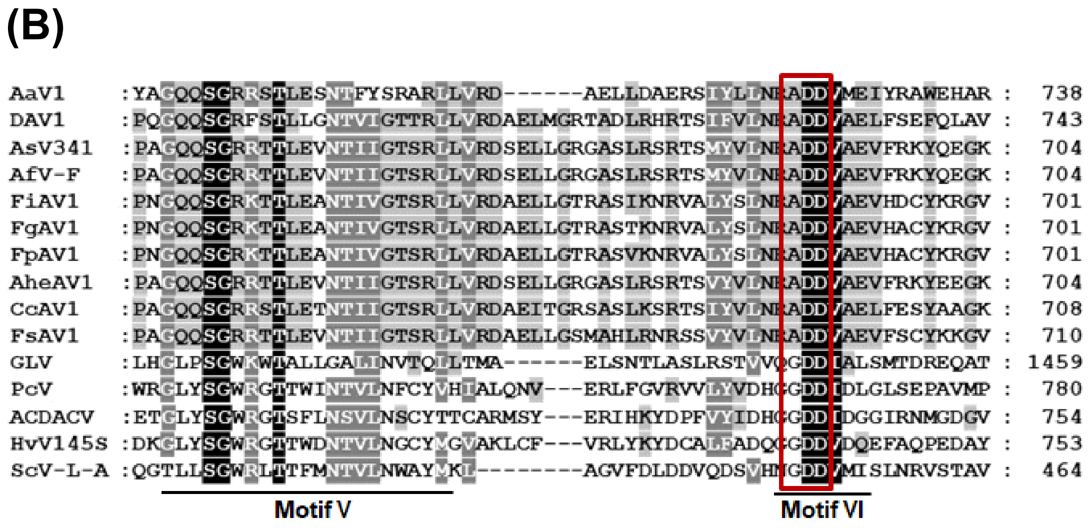


Figure 2-19. Phylogenetic analysis of Alternaviridae.

(A) Phylogenetic analysis of Alternaviridae, *Totiviridae* and *Chrysoviridae* according to the amino acid sequences of RdRps. **(B)** Multiple alignment of the motif VI of the RdRp of ten alternaviruses, three chrysoviruses, one totivirus, and one giardiavirus. Virus abbreviation: AaV1, Alternaria alternata virus 1; DAV1, Diaporthe alternavirus 1; AsV341, Aspergillus mycovirus 341; AfV-F, Aspergillus foetidus mycovirus; FpAV1, Fusarium poae alternavirus 1; FgAV1, Fusarium graminearum alternavirus 1; FiAV1, Fusarium incarnatum alternavirus 1; AheAV1, Aspergillus heteromorphus alternavirus 1; GLV, Giardia lamblia virus; PcV, Penicillium chrysogenum virus; ACDACV, Amasya cherry disease-associated chrysovirus; Hv145SV, Helminthosporium victoriae 145S virus; ScV-L-A, Saccharomyces cerevisiae virus L-A.

2.5 Appendix

2.5.1 AaV1 dsRNA1 sequence (NC_010984)

3567nt excluding poly (A) tail

```
5'  UGGCAAAAAGCUAUCGUAUCUCGUGGUACACAUAAGUCUUUGGCCUUGAUGCGGAGUGACAAAACCCUCC
o                                                                              70
3'  ACCGUUUUUCGAUAGCAUAGAGCACCAUGUGUAUUCAGAAACCGGAACUACGCCUCACUGUUUUGGGAGG
1    Trp Gln Lys Ala Ile Val Ser Arg Gly Thr His Lys Ser Leu Ala Leu Met Arg Ser Asp Lys Thr Leu
2     Gly Lys Lys Leu Ser Tyr Leu Val Val His Ile Ser Leu Trp Pro . Cys Gly Val Thr Lys Pro Ser
3   Leu Ala Lys Ser Tyr Arg Ile Ser Trp Tyr Thr . Val Phe Gly Leu Asp Ala Glu . Gln Asn Pro Pro
o

5'  UAUUGCUGGGGUUCAGGAACGCUGCUUGGUUGCGUGCACAGGGGUCCCACCCGGAUCUCGGCAGCAUUCC
o                                                                              140
3'  AUAACGACCCCAAGUCCUUGCGACGAACCAACGCACGUGUCCCCAGGGUGGGCCUAGAGCCGUCGUAAGG
1   Leu Leu Leu Gly Phe Arg Asn Ala Ala Trp Leu Arg Ala Gln Gly Ser His Pro Asp Leu Gly Ser Ile Pro
2    Tyr Cys Trp Gly Ser Gly Thr Leu Leu Gly Cys Val His Arg Gly Pro Thr Arg Ile Ser Ala Ala Phe
3     Ile Ala Gly Val Gln Glu Arg Cys Leu Val Ala Cys Thr Gly Val Pro Pro Gly Ser Arg Gln His Ser
o

5'  AAAGCGAGCGUUCGUCCCGGAACCAAAGGAGAUUGAUGCCUGCCUGCAGGGUAUCGAGGACGGAUGGGGU
o                                                                              210
3'  UUUCGCUCGCAAGCAGGGCCUUGGUUUCCUCUAACUACGGACGGACGUCCCAUAGCUCCUGCCUACCCCA
1     Lys Arg Ala Phe Val Pro Glu Pro Lys Glu Ile Asp Ala Cys Leu Gln Gly Ile Glu Asp Gly Trp Gly
2   Gln Ser Glu Arg Ser Ser Arg Asn Gln Arg Arg Leu Met Pro Ala Cys Arg Val Ser Arg Thr Asp Gly Val
3    Lys Ala Ser Val Arg Pro Gly Thr Lys Gly Asp . Cys Leu Pro Ala Gly Tyr Arg Gly Arg Met Gly
o

5'  GACGGCGCCUUCUGGUGCCAGACAGUCGUAGGUAGCGGUUUUCCUGUUUUCUUCGUUCUUUCUGAUGUUU
o                                                                              280
3'  CUGCCGCGGAAGACCACGGUCUGUCAGCAUCCAUCGCCAAAAGGACAAAAGAAGCAAGAAAGACUACAAA
1    Asp Gly Ala Phe Trp Cys Gln Thr Val Val Gly Ser Gly Phe Pro Val Phe Phe Val Leu Ser Asp Val
2     Thr Ala Pro Ser Gly Ala Arg Gln Ser . Val Ala Val Phe Leu Phe Ser Ser Phe Phe Leu Met Phe
3   . Arg Arg Leu Leu Val Pro Asp Ser Arg Arg . Arg Phe Ser Cys Phe Leu Arg Ser Phe . Cys Phe
o

5'  CUCGUAAUGUUUUUUGUCAUUCUUUUGCUGAGGUUGGUGGUGAUGUCUCUGUUGUGAGUUUCGGGCUGAC
o                                                                              350
3'  GAGCAUUACAAAAAACAGUAAGAAAACGACUCCAACCACCACUACAGAGACAACACUCAAAGCCCGACUG
1   Ser Arg Asn Val Phe Cys His Ser Phe Ala Glu Val Gly Gly Asp Val Ser Val Val Ser Phe Gly Leu Thr
2    Leu Val Met Phe Phe Val Ile Leu Leu Leu Arg Leu Val Val Met Ser Leu Leu . Val Ser Gly .
3     Ser . Cys Phe Leu Ser Phe Phe Cys . Gly Trp Trp . Cys Leu Cys Cys Glu Phe Arg Ala Asp
o

5'  GGAAGCGCGUUCUCAGUUGUACCGCAAAGCGCAAGUCUGUGUUGUGCCUAAAGCACGUCCCGCUGUAGCA
o                                                                              420
3'  CCUUCGCGCAAGAGUCAACAUGGCGUUUCGCGUUCAGACACAACACGGAUUUCGUGCAGGGCGACAUCGU
1     Glu Ala Arg Ser Gln Leu Tyr Arg Lys Ala Gln Val Cys Val Val Pro Lys Ala Arg Pro Ala Val Ala
2   Arg Lys Arg Val Leu Ser Cys Thr Ala Lys Arg Lys Ser Val Leu Cys Leu Lys His Val Pro Leu . His
3    Gly Ser Ala Phe Ser Val Val Pro Gln Ser Ala Ser Leu Cys Cys Ala . Ser Thr Ser Arg Cys Ser
o

5'  UACCAUCGCGAUCUCUCGGCAGUGGCCAGGGUCUUGCGUUUUGGCGUGCUCAAUGCGAAUAAUAAACUGU
o                                                                              490
3'  AUGGUAGCGCUAGAGAGCCGUCACCGGUCCCAGAACGCAAAACCGCACGAGUUACGCUUAUUAUUUGACA
1    Tyr His Arg Asp Leu Ser Ala Val Ala Arg Val Leu Arg Phe Gly Val Leu Asn Ala Asn Asn Lys Leu
2     Thr Ile Ala Ile Ser Arg Gln Trp Pro Gly Ser Cys Val Leu Ala Cys Ser Met Arg Ile Ile Asn Cys
3   Ile Pro Ser Arg Ser Leu Gly Ser Gly Gln Gly Leu Ala Phe Trp Arg Ala Gln Cys Glu . . Thr Val
o

5'  GCGAUGCAUAUACAGACUUUGUUCGGUCACCUGUUGCGCGGACGGAGGAGACUUGGGGUUUGCUCGGUGG
o                                                                              560
3'  CGCUACGUAUAUGUCUGAAACAAGCCAGUGGACAACGCGCCUGCCUCCUCUGAACCCCAAACGAGCCACC
1   Cys Asp Ala Tyr Thr Asp Phe Val Arg Ser Pro Val Ala Arg Thr Glu Glu Thr Trp Gly Leu Leu Gly Gly
2    Ala Met His Ile Gln Thr Leu Phe Gly His Leu Leu Arg Gly Arg Arg Arg Leu Gly Val Cys Ser Val
3     Arg Cys Ile Tyr Arg Leu Cys Ser Val Thr Cys Cys Ala Asp Gly Gly Asp Leu Gly Phe Ala Arg Trp
o
```

```
5'   CUAUCCGUGGAGGGAUCUGGAUGCGUUGGUGGCGUCGUCGACGCGCGGAGUAUGAUCGUGAUGCAGAC
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   630
3'   GAUAGGCACCUCCCUAGACCUACGCAACCACCGCAGCAGCUGCGCGCCUCAUACUAGCACUACGUCUG
1     Tyr Pro Trp Arg Asp Leu Asp Ala Leu Val Ala Ser Ser Thr Arg Ala Glu Tyr Asp Arg Asp Ala Asp
2    Ala Ile Arg Gly Gly Ile Trp Met Arg Trp Trp Arg Arg Arg Arg Ala Arg Ser Met Ile Val Met Gln Thr
3     Leu Ser Val Glu Gly Ser Gly Cys Val Gly Gly Val Val Asp Ala Arg Gly Val . Ser . Cys Arg
o
5'   ACGAAGCUGGUCGGUUUCCUUGACCGCUUGUGGGGCAAGAAGCGUGGGAAGGGGUUGGGCAGGCAAGACA
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   700
3'   UGCUUCGACCAGCCAAAGGAACUGGCGAACACCCCGUUCUUCGCACCCUUCCCCAACCCGUCCGUUCUGU
1     Thr Lys Leu Val Gly Phe Leu Asp Arg Leu Trp Gly Lys Lys Arg Gly Lys Gly Leu Gly Arg Gln Asp
2      Arg Ser Trp Ser Val Ser Leu Thr Ala Cys Gly Ala Arg Ser Val Gly Arg Gly Trp Ala Gly Lys Thr
3    His Glu Ala Gly Arg Phe Pro . Pro Leu Val Gly Gln Glu Ala Trp Glu Gly Val Gly Gln Ala Arg Gln
o
5'   AGAAGACCCGUGCCAAGGCGGCAAUGGCUUGGGGCGAUGCUAAUCUCUGGACUCGCGAGGUGUCUUUUAA
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   770
3'   UCUUCUGGGCACGGUUCCGCCGUUACCGAACCCCGCUACGAUUAGAGACCUGAGCGCUCCACAGAAAAUU
1    Lys Lys Thr Arg Ala Lys Ala Ala Met Ala Trp Gly Asp Ala Asn Leu Trp Thr Arg Glu Val Ser Phe Lys
2     Arg Arg Pro Val Pro Arg Arg Gln Trp Leu Gly Ala Met Leu Ile Ser Gly Leu Ala Arg Cys Leu Leu
3      Glu Asp Pro Cys Gln Gly Gly Asn Gly Leu Gly Arg Cys . Ser Leu Asp Ser Arg Gly Val Phe .
o
5'   GCGGCGCCACUAUUAUGAUCUUGCUACGUUUUUGUGUACGAUGGACGAAGUGUGGGGGGACCGGGCUGUG
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   840
3'   CGCCGCGGUGAUAAUACUAGAACGAUGCAAAAACACAUGCUACCUGCUUCACACCCCCCUGGCCCGACAC
1      Arg Arg His Tyr Tyr Asp Leu Ala Thr Phe Leu Cys Thr Met Asp Glu Val Trp Gly Asp Arg Ala Val
2    Ser Gly Ala Thr Ile Met Ile Leu Leu Arg Phe Cys Val Arg Trp Thr Lys Cys Gly Gly Thr Gly Leu Trp
3     Ala Ala Pro Leu Leu . Ser Cys Tyr Val Phe Val Tyr Asp Gly Arg Ser Val Gly Gly Pro Gly Cys
o
5'   GCUAUGCAGGGGGGCGCAACUUCUAAGGAAGCUGAUGUGCAGUGUCUUUCCUUUAGGAAGGCGGUUCGUA
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   910
3'   CGAUACGUCCCCCCGCGUUGAAGAUUCCUUCGACUACACGUCACAGAAAGGAAAUCCUUCCGCCAAGCAU
1     Ala Met Gln Gly Gly Ala Thr Ser Lys Glu Ala Asp Val Gln Cys Leu Ser Phe Arg Lys Ala Val Arg
2      Leu Cys Arg Gly Ala Gln Leu Leu Arg Lys Leu Met Cys Ser Val Phe Pro Leu Gly Arg Arg Phe Val
3    Gly Tyr Ala Gly Gly Arg Asn Phe . Gly Ser . Cys Ala Val Ser Phe Leu . Glu Gly Gly Ser Tyr
o
5'   UGGUUUUGUCGGAUUUGAUCUUUGGCCCUGCUACGUAUGAGCGGAUCUACUUGACUGUGUUCUGGAUCUU
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   980
3'   ACCAAAACAGCCUAAACUAGAAACCGGGACGAUGCAUACUCGCCUAGAUGAACUGACACAAGACCUAGAA
1    Met Val Leu Ser Asp Leu Ile Phe Gly Pro Ala Thr Tyr Glu Arg Ile Tyr Leu Thr Val Phe Trp Ile Phe
2     Trp Phe Cys Arg Ile . Ser Leu Ala Leu Leu Arg Met Ser Gly Ser Thr . Leu Cys Ser Gly Ser
3      Gly Phe Val Gly Phe Asp Leu Trp Pro Cys Tyr Val . Ala Asp Leu Leu Asp Cys Val Leu Asp Leu
o
5'   CUCAACGCGGUGGUGGCCGAGCUUGGUGCCUGUUUUGUUGCACUUUGGCUCCCUUGGAAUGUCUGAUGAC
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   1050
3'   GAGUUGCGCCACCACCGGCUCGAACCACGGACAAAACAACGUGAAACCGAGGGAACCUUACAGACUACUG
1      Ser Thr Arg Trp Trp Pro Ser Leu Val Pro Val Leu Leu His Phe Gly Ser Leu Gly Met Ser Asp Asp
2    Ser Gln Arg Gly Gly Gly Arg Ala Trp Cys Leu Phe Cys Cys Thr Leu Ala Pro Leu Glu Cys Leu Met Thr
3     Leu Asn Ala Val Val Ala Glu Leu Gly Ala Cys Phe Val Ala Leu Trp Leu Pro Trp Asn Val . .
o
5'   GAAUACACUGCCGUACACAAGGAAAUGACUGCCGUUGUGACUUCGACUUGGAUGGUGCCCGGCACUUGCA
o    +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   1120
3'   CUUAUGUGACGGCAUGUGUUCCUUUACUGACGGCAACACUGAAGCUGAACCUACCACGGGCCGUGAACGU
1     Glu Tyr Thr Ala Val His Lys Glu Met Thr Ala Val Val Thr Ser Thr Trp Met Val Pro Gly Thr Cys
2      Asn Thr Leu Pro Tyr Thr Arg Lys . Leu Pro Leu . Leu Arg Leu Gly Trp Cys Pro Ala Leu Ala
3    Arg Ile His Cys Arg Thr Gln Gly Asn Asp Cys Arg Cys Asp Phe Asp Leu Asp Gly Ala Arg His Leu Gln
o
```

```
5'  GGCAGCACCAGUUCUCUGCCAAUUUCCUUAACGCGGAGGACCUUACAGGGUGGUCUGACCGGGACUCGUU
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1190
3'  CCGUCGUGGUCAAGAGACGGUUAAAGGAAUUGCGCCUCCUGGAAUGUCCCACCAGACUGGCCCUGAGCAA
1   Arg Gln His Gln Phe Ser Ala Asn Phe Leu Asn Ala Glu Asp Leu Thr Gly Trp Ser Asp Arg Asp Ser Leu
2    Gly Ser Thr Ser Ser Leu Pro Ile Ser Leu Thr Arg Arg Thr Leu Gln Gly Gly Leu Thr Gly Thr Arg
3     Ala Ala Pro Val Leu Cys Gln Phe Pro . Arg Gly Gly Pro Tyr Arg Val Val . Pro Gly Leu Val
o
5'  GAAGGGGGGUGUGUGCGCUGAGAUCAUCAAAUUUGCUCUCGCUACUUUUGAGUAUAAGGCAGAUGUGCGG
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1260
3'  CUUCCCCCCACACACGCGACUCUAGUAGUUUAAACGAGAGCGAUGAAAACUCAUAUUCCGUCUACACGCC
1     Lys Gly Gly Val Cys Ala Glu Ile Ile Lys Phe Ala Leu Ala Thr Phe Glu Tyr Lys Ala Asp Val Arg
2    . Arg Gly Val Cys Ala Leu Arg Ser Ser Asn Leu Leu Ser Leu Leu Leu Ser Ile Arg Gln Met Cys Gly
3   Glu Gly Gly Cys Val Arg . Asp His Gln Ile Cys Ser Arg Tyr Phe . Val . Gly Arg Cys Ala
o
5'  GAUGGGGGCGGCGAAGCGGGCAUACCUAUUCGUGGCAAAGAAGGUUCGCCCGAGGAAUACCUCGGGCGCU
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1330
3'  CUACCCCCGCCGCUUCGCCCGUAUGGAUAAGCACCGUUUCUUCCAAGCGGGCUCCUUAUGGAGCCCGCGA
1    Asp Gly Gly Gly Glu Ala Gly Ile Pro Ile Arg Gly Lys Glu Gly Ser Pro Glu Glu Tyr Leu Gly Arg
2     Met Gly Ala Ala Lys Arg Ala Tyr Leu Phe Val Ala Lys Lys Val Arg Pro Arg Asn Thr Ser Gly Ala
3   Gly Trp Gly Arg Arg Ser Gly His Thr Tyr Ser Trp Gln Arg Arg Phe Ala Arg Gly Ile Pro Arg Ala Leu
o
5'  ACCGCCAGGCGAUGUAUGAUUUGCUCCGUCCUAUGUACGCACGUUAUGCGCCCAAUAUGGUGGAUCUUCA
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1400
3'  UGGCGGUCCGCUACAUACUAAACGAGGCAGGAUACAUGCGUGCAAUACGCGGGUUAUACCACCUAGAAGU
1   Tyr Arg Gln Ala Met Tyr Asp Leu Leu Arg Pro Met Tyr Ala Arg Tyr Ala Pro Asn Met Val Asp Leu Gln
2    Thr Ala Arg Arg Cys Met Ile Cys Ser Val Leu Cys Thr His Val Met Arg Pro Ile Trp Trp Ile Phe
3     Pro Pro Gly Asp Val . Phe Ala Pro Ser Tyr Val Arg Thr Leu Cys Ala Gln Tyr Gly Gly Ser Ser
o
5'  GGCGCACAUGGAGAAGCGGAUGGCGUGGAUGUCUGGUGGUGCCGUGGGCCGUCGGGCCAAAGAGUUGCUG
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1470
3'  CCGCGUGUACCUCUUCGCCUACCGCACCUACAGACCACCACGGCACCCGGCAGCCCGGUUUCUCAACGAC
1     Ala His Met Glu Lys Arg Met Ala Trp Met Ser Gly Gly Ala Val Gly Arg Arg Ala Lys Glu Leu Leu
2   Arg Arg Thr Trp Arg Ser Gly Trp Arg Gly Cys Leu Val Val Pro Trp Ala Val Gly Pro Lys Ser Cys Trp
3    Gly Ala His Gly Glu Ala Asp Gly Val Asp Val Trp Trp Cys Arg Gly Pro Ser Gly Gln Arg Val Ala
o
5'  GGGCCGGGGAUUGCUCCGCCCGGCAGUUCGAAAGCCUAUGUGGCGGCGCGCGCUGAUAUCAGCCAGCUCA
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1540
3'  CCCGGCCCCUAACGAGGCGGGCCGUCAAGCUUUCGGAUACACCGCCGCGCGCGACUAUAGUCGGUCGAGU
1    Gly Pro Gly Ile Ala Pro Pro Gly Ser Ser Lys Ala Tyr Val Ala Ala Arg Ala Asp Ile Ser Gln Leu
2     Gly Arg Gly Leu Leu Arg Pro Ala Val Arg Lys Pro Met Trp Arg Arg Ala Leu Ile Ser Ala Ser Ser
3   Gly Ala Gly Asp Cys Ser Ala Arg Gln Phe Glu Ser Leu Cys Gly Gly Ala Arg . Tyr Gln Pro Ala His
o
5'  CCAGAGACUGGGGGGAGAUGAGGAUUGAGGUUGGGGGGAAGGGUAAUGAGAGGGGUUCCGAGCGUACGUU
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1610
3'  GGUCUCUGACCCCCCUCUACUCCUAACUCCAACCCCCCUUCCCAUUACUCUCCCCAAGGCUCGCAUGCAA
1   Thr Arg Asp Trp Gly Glu Met Arg Ile Glu Val Gly Gly Lys Gly Asn Glu Arg Gly Ser Glu Arg Thr Leu
2    Pro Glu Thr Gly Gly Arg . Gly Leu Arg Leu Gly Gly Arg Val Met Arg Gly Val Pro Ser Val Arg
3     Gln Arg Leu Gly Gly Asp Glu Asp . Gly Trp Gly Glu Gly . . Glu Gly Phe Arg Ala Tyr Val
o
5'  GUUGGCGACUGAUUUGCGAGAUCAGGUGAGCGAAUCGUACCUUUUGCAUGCGUUUAAGAAUAGGUACGGC
o   +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  1680
3'  CAACCGCUGACUAAACGCUCUAGUCCACUCGCUUAGCAUGGAAAACGUACGCAAAUUCUUAUCCAUGCCG
1     Leu Ala Thr Asp Leu Arg Asp Gln Val Ser Glu Ser Tyr Leu Leu His Ala Phe Lys Asn Arg Tyr Gly
2   Cys Trp Arg Leu Ile Cys Glu Ile Arg . Ala Asn Arg Thr Phe Cys Met Arg Leu Arg Ile Gly Thr Ala
3    Val Gly Asp . Phe Ala Arg Ser Gly Glu Arg Ile Val Pro Phe Ala Cys Val . Glu . Val Arg
o
```

```
5'   CUAAUAGGAGUUGACGUGGGCGAGACGCCCGUCGAGAUGUUUAAGCGUCACGUGACCGUAGCGUCGGCUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1750
3'   GAUUAUCCUCAACUGCACCCGCUCUGCGGGCAGCUCUACAAAUUCGCAGUGCACUGGCAUCGCAGCCGAU
1     Leu Ile Gly Val Asp Val Gly Glu Thr Pro Val Glu Met Phe Lys Arg His Val Thr Val Ala Ser Ala
2      .  .  Glu Leu Thr Trp Ala Arg Arg Pro Ser Arg Cys Leu Ser Val Thr .  Pro .  Arg Arg Leu
3    Pro Asn Arg Ser .  Arg Gly Arg Asp Ala Arg Arg Asp Val .  Ala Ser Arg Asp Arg Ser Val Gly Tyr
o
5'   CGGACGUGCCACUGGUUCAUCAUGGGCGUGACAAGAAGGUCCUUGCAGCUUGGGAUUACUCCAAGUGGGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1820
3'   GCCUGCACGGUGACCAAGUAGUACCCGCACUGUUCUUCCAGGAACGUCGAACCCUAAUGAGGUUCACCCU
1    Thr Asp Val Pro Leu Val His His Gly Arg Asp Lys Lys Val Leu Ala Ala Trp Asp Tyr Ser Lys Trp Asp
2     Arg Thr Cys His Trp Phe Ile Met Gly Val Thr Arg Arg Ser Leu Gln Leu Gly Ile Thr Pro Ser Gly
3      Gly Arg Ala Thr Gly Ser Ser Trp Ala .  Gln Glu Gly Pro Cys Ser Leu Gly Leu Leu Gln Val Gly
o
5'   CCACCACGUUAUGCUUGCUGAAAGGCUCAUUCUGGUCGAGGUCAUGAGGAAGCUUGUGUUGGAAUUCGUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1890
3'   GGUGGUGCAAUACGAACGACUUUCCGAGUAAGACCAGCUCCAGUACUCCUUCGAACACAACCUUAAGCAG
1      His His Val Met Leu Ala Glu Arg Leu Ile Leu Val Glu Val Met Arg Lys Leu Val Leu Glu Phe Val
2    Thr Thr Thr Leu Cys Leu Leu Lys Gly Ser Phe Trp Ser Arg Ser .  Gly Ser Leu Cys Trp Asn Ser Ser
3     Pro Pro Arg Tyr Ala Cys .  Lys Ala His Ser Gly Arg Gly His Glu Glu Ala Cys Val Gly Ile Arg
o
5'   CAGCGGCCCGACGUUCGGGAGGAUAUGUUGCGCGAGCUAGAGGUUCUUGAAUCUAGCCAUCGCACUGCAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1960
3'   GUCGCCGGGCUGCAAGCCCUCCUAUACAACGCGCUCGAUCUCCAAGAACUUAGAUCGGUAGCGUGACGUU
1     Gln Arg Pro Asp Val Arg Glu Asp Met Leu Arg Glu Leu Glu Val Leu Glu Ser Ser His Arg Thr Ala
2      Ser Gly Pro Thr Phe Gly Arg Ile Cys Cys Ala Ser .  Arg Phe Leu Asn Leu Ala Ile Ala Leu Gln
3    Pro Ala Ala Arg Arg Ser Gly Gly Tyr Val Ala Arg Ala Arg Gly Ser .  Ile .  Pro Ser His Cys Asn
o
5'   UAUACCGAUCGCGGGCGUUCGCUGAUGCGAAAUAUACCGACCAGGUGGAUUCUUUAAUCCGUGAAGGCGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2030
3'   AUAUGGCUAGCGCCCGCAAGCGACUACGCUUUAUAUGGCUGGUCCACCUAAGAAAUUAGGCACUUCCGCA
1    Ile Tyr Arg Ser Arg Ala Phe Ala Asp Ala Lys Tyr Thr Asp Gln Val Asp Ser Leu Ile Arg Glu Gly Val
2     Tyr Thr Asp Arg Gly Arg Ser Leu Met Arg Asn Ile Pro Thr Arg Trp Ile Leu .  Ser Val Lys Ala
3      Ile Pro Ile Ala Gly Val Arg .  Cys Glu Ile Tyr Arg Pro Gly Gly Phe Phe Asn Pro .  Arg Arg
o
5'   CGCCGCAGGUUUCAAGGGUGAGGUGCAGCGUUUGUCGGCGGACCAGGUUCGCAUCACAAACUACGCGGGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2100
3'   GCGGCGUCCAAAGUUCCCACUCCACGUCGCAAACAGCCGCCUGGUCCAAGCGUAGUGUUUGAUGCGCCCC
1      Ala Ala Gly Phe Lys Gly Glu Val Gln Arg Leu Ser Ala Asp Gln Val Arg Ile Thr Asn Tyr Ala Gly
2    Ser Pro Gln Val Ser Arg Val Arg Cys Ser Val Cys Arg Arg Thr Arg Phe Ala Ser Gln Thr Thr Arg Gly
3     Arg Arg Arg Phe Gln Gly .  Gly Ala Ala Phe Val Gly Gly Pro Gly Ser His His Lys Leu Arg Gly
o
5'   CAGCAGUCUGGGCGCAGGAGUACACUUGAGUCAAACACCUUUUACUCCAGGGCGCGUCUGCUAGUGCGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2170
3'   GUCGUCAGACCCGCGUCCUCAUGUGAACUCAGUUUGUGGAAAAUGAGGUCCCGCGCAGACGAUCACGCGC
1     Gln Gln Ser Gly Arg Arg Ser Thr Leu Glu Ser Asn Thr Phe Tyr Ser Arg Ala Arg Leu Leu Val Arg
2      Ser Ser Leu Gly Ala Gly Val His Leu Ser Gln Thr Pro Phe Thr Pro Gly Arg Val Cys .  Cys Ala
3    Ala Ala Val Trp Ala Gln Glu Tyr Thr .  Val Lys His Leu Leu Leu Gln Gly Ala Ser Ala Ser Ala Arg
o
5'   AUGCCGAACUUCUGGAUGCUGAGCGCAGCAUAUACCUGUUGAAUAGGGCCGAUGACGUCAUGGAGAUUUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2240
3'   UACGGCUUGAAGACCUACGACUCGCGUCGUAUAUGGACAACUUAUCCCGGCUACUGCAGUACCUCUAAAU
1    Asp Ala Glu Leu Leu Asp Ala Glu Arg Ser Ile Tyr Leu Leu Asn Arg Ala Asp Asp Val Met Glu Ile Tyr
2     Met Pro Asn Phe Trp Met Leu Ser Ala Ala Tyr Thr Cys .  Ile Gly Pro Met Thr Ser Trp Arg Phe
3      Cys Arg Thr Ser Gly Cys .  Ala Gln His Ile Pro Val Glu .  Gly Arg .  Arg His Gly Asp Leu
o
```

```
5'   CCGGGCGUGGGAGCACGCGAGGAAUGCGAUCGAUGUGAUGCUGCUGCAGGGCCAUAAGGCAAAUAAGAAG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2310
3'   GGCCCGCACCCUCGUGCGCUCCUUACGCUAGCUACACUACGACGACGUCCCGGUAUUCCGUUUAUUCUUC
1      Arg Ala Trp Glu His Ala Arg Asn Ala Ile Asp Val Met Leu Leu Gln Gly His Lys Ala Asn Lys Lys
2    Thr Gly Arg Gly Ser Thr Arg Gly Met Arg Ser Met . Cys Cys Cys Arg Ala Ile Arg Gln Ile Arg Arg
3     Pro Gly Val Gly Ala Arg Glu Glu Cys Asp Arg Cys Asp Ala Ala Ala Gly Pro . Gly Lys . Glu
o
5'   AAGCAGGUUGUCCAGGUGCGAACUGGAGUAUACUUCCGAAUCUUGUAUGCCAAUGGUAGUAUGCGCGGCU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2380
3'   UUCGUCCAACAGGUCCACGCUUGACCUCAUAUGAAGGCUUAGAACAUACGGUUACCAUCAUACGCGCCGA
1     Lys Gln Val Val Gln Val Arg Thr Gly Val Tyr Phe Arg Ile Leu Tyr Ala Asn Gly Ser Met Arg Gly
2      Ser Arg Leu Ser Arg Cys Glu Leu Glu Tyr Thr Ser Glu Ser Cys Met Pro Met Val Val Cys Ala Ala
3    Glu Ala Gly Cys Pro Gly Ala Asn Trp Ser Ile Leu Pro Asn Leu Val Cys Gln Trp . Tyr Ala Arg Leu
o
5'   UCCCUCCUCGUGCUGUUUAUGCUUGCGCUUCAGCCGGGCCUUCAAUGGCCGCGAGCGGGGGUUUUGAUCC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2450
3'   AGGGAGGAGCACGACAAAUACGAACGCGAAGUCGGCCCGGAAGUUACCGGCGCUCGCCCCCAAAACUAGG
1    Phe Pro Pro Arg Ala Val Tyr Ala Cys Ala Ser Ala Gly Pro Ser Met Ala Ala Ser Gly Gly Phe Asp Pro
2     Ser Leu Leu Val Leu Phe Met Leu Ala Leu Gln Pro Gly Leu Gln Trp Pro Arg Ala Gly Val Leu Ile
3      Pro Ser Ser Cys Cys Leu Cys Leu Arg Phe Ser Arg Ala Phe Asn Gly Arg Glu Arg Gly Phe . Ser
o
5'   GGUGGAGCGCCUCUCUUCUCUUUCGGGCGCGUUGGAUAGGUUAGCGCGGCGUGGCAGCGGUUACUACGUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2520
3'   CCACCUCGCGGAGAGAAGAGAAAGCCCGCGCAACCUAUCCAAUCGCGCCGCACCGUCGCCAAUGAUGCAG
1      Val Glu Arg Leu Ser Ser Leu Ser Gly Ala Leu Asp Arg Leu Ala Arg Arg Gly Ser Gly Tyr Tyr Val
2    Arg Trp Ser Ala Ser Leu Leu Phe Arg Ala Arg Trp Ile Gly . Arg Gly Val Ala Ala Val Thr Thr Ser
3     Gly Gly Ala Pro Leu Phe Ser Phe Gly Arg Val Gly . Val Ser Ala Ala Trp Gln Arg Leu Leu Arg
o
5'   GCACGCGCUCUUUAUUUCGAGGCUGAGGAUUAUUACCGCGAUGUUCGCGUGCAGUUAGCGGCAAAGGCUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2590
3'   CGUGCGCGAGAAAUAAAGCUCCGACUCCUAAUAAUGGCGCUACAAGCGCACGUCAAUCGCCGUUUCCGAC
1     Ala Arg Ala Leu Tyr Phe Glu Ala Glu Asp Tyr Tyr Arg Asp Val Arg Val Gln Leu Ala Ala Lys Ala
2      His Ala Leu Phe Ile Ser Arg Leu Arg Ile Ile Thr Ala Met Phe Ala Cys Ser . Arg Gln Arg Leu
3    Arg Thr Arg Ser Leu Phe Arg Gly . Gly Leu Leu Pro Arg Cys Ser Arg Ala Val Ser Gly Lys Gly .
o
5'   ACAAAUACACUCGUUUUACCAUUCCCCGCGAGGUGUUGCGUGCUUCGCCCGAUUUGGGUGGGUGUGGAGU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2660
3'   UGUUUAUGUGAGCAAAAUGGUAAGGGGCGCUCCACAACGCACGAAGCGGGCUAAACCCACCCACACCUCA
1    Asp Lys Tyr Thr Arg Phe Thr Ile Pro Arg Glu Val Leu Arg Ala Ser Pro Asp Leu Gly Gly Cys Gly Val
2     Thr Asn Thr Leu Val Leu Pro Phe Pro Ala Arg Cys Cys Val Leu Arg Pro Ile Trp Val Gly Val Glu
3      Gln Ile His Ser Phe Tyr His Ser Pro Arg Gly Val Ala Cys Phe Ala Arg Phe Gly Trp Val Trp Ser
o
5'   GCUGCCCCCUGGGUGUUAUGACUAUGACUACACGAUCAAGUGCUCUAACAUCAAAGGUCCUAAUGAGCAG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2730
3'   CGACGGGGGACCCACAAUACUGAUACUGAUGUGCUAGUUCACGAGAUUGUAGUUUCCAGGAUUACUCGUC
1      Leu Pro Pro Gly Cys Tyr Asp Tyr Asp Tyr Thr Ile Lys Cys Ser Asn Ile Lys Gly Pro Asn Glu Gln
2    Cys Cys Pro Leu Gly Val Met Thr Met Thr Thr Arg Ser Ser Ala Leu Thr Ser Lys Val Leu Met Ser Ser
3     Ala Ala Pro Trp Val Leu . Leu . Leu His Asp Gln Val Leu . His Gln Arg Ser . . Ala
o
5'   CUCUGGCAGGCGAUUCGUGAGCGUUUAGAGAAGCGGCACAAGUUCAGGGGUCUUGCCGACUUGCAGGGCU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2800
3'   GAGACCGUCCGCUAAGCACUCGCAAAUCUCUUCGCCGUGUUCAAGUCCCCAGAACGGCUGAACGUCCCGA
1     Leu Trp Gln Ala Ile Arg Glu Arg Leu Glu Lys Arg His Lys Phe Arg Gly Leu Ala Asp Leu Gln Gly
2      Ser Gly Arg Arg Phe Val Ser Val . Arg Ser Gly Thr Ser Ser Gly Val Leu Pro Thr Cys Arg Ala
3    Ala Leu Ala Gly Asp Ser . Ala Phe Arg Glu Ala Ala Gln Val Gln Gly Ser Cys Arg Leu Ala Gly Leu
o
```

```
5'  CGGCUGGGCGUGCCUUCGCUCACGACGUUCCGUUGGAGGUGCCUGACAAGGCUUGGCAGCAGUGGUCGAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2870
3'  GCCGACCCGCACGGAAGCGAGUGCUGCAAGGCAACCUCCACGGACUGUUCCGAACCGUCGUCACCAGCUU
1   Ser Ala Gly Arg Ala Phe Ala His Asp Val Pro Leu Glu Val Pro Asp Lys Ala Trp Gln Gln Trp Ser Lys
2    Arg Leu Gly Val Pro Ser Leu Thr Thr Phe Arg Trp Arg Cys Leu Thr Arg Leu Gly Ser Ser Gly Arg
3     Gly Trp Ala Cys Leu Arg Ser Arg Arg Ser Val Gly Gly Ala  .  Gln Gly Leu Ala Ala Val Val Glu
o
5'  GCGUUGGAGGGGCGAUAGGGCGGCUCAGGAUGGGCGCGGCGAGCUACUUCGUGCAAAGAUGAUGGCAAAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2940
3'  CGCAACCUCCCCGCUAUCCCGCCGAGUCCUACCCGCGCCGCUCGAUGAAGCACGUUUCUACUACCGUUUC
1     Arg Trp Arg Gly Asp Arg Ala Ala Gln Asp Gly Arg Gly Glu Leu Leu Arg Ala Lys Met Met Ala Lys
2   Ser Val Gly Gly Ala Ile Gly Arg Leu Arg Met Gly Ala Ala Ser Tyr Phe Val Gln Arg  .  Trp Gln Ser
3    Ala Leu Glu Gly Arg  .  Gly Gly Ser Gly Trp Ala Arg Arg Ala Thr Ser Cys Lys Asp Asp Gly Lys
o
5'  CUGUGGCGCUACAAAAGGAGCUGGCACCGUGACAAGAAGGGCGAUACCUGGGGUUUGGAAGGCUUCGGCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3010
3'  GACACCGCGAUGUUUUCCUCGACCGUGGCACUGUUCUUCCCGCUAUGGACCCCAAACCUUCCGAAGCCGG
1    Leu Trp Arg Tyr Lys Arg Ser Trp His Arg Asp Lys Lys Gly Asp Thr Trp Gly Leu Glu Gly Phe Gly
2     Cys Gly Ala Thr Lys Gly Ala Gly Thr Val Thr Arg Arg Ala Ile Pro Gly Val Trp Lys Ala Ser Ala
3   Ala Val Ala Leu Gln Lys Glu Leu Ala Pro  .  Gln Glu Gly Arg Tyr Leu Gly Phe Gly Arg Leu Arg Pro
o
5'  AUUUACUGUCUUUACCGGUUGAUUUCUUCAAUAUUGCUUGGAGGAUGUUACGCGACUGUCCUGGUACGGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3080
3'  UAAAUGACAGAAAUGGCCAACUAAAGAAGUUAUAACGAACCUCCUACAAUGCGCUGACAGGACCAUGCCC
1   His Leu Leu Ser Leu Pro Val Asp Phe Phe Asn Ile Ala Trp Arg Met Leu Arg Asp Cys Pro Gly Thr Gly
2    Ile Tyr Cys Leu Tyr Arg Leu Ile Ser Ser Ile Leu Leu Gly Gly Cys Tyr Ala Thr Val Leu Val Arg
3     Phe Thr Val Phe Thr Gly  .  Phe Leu Gln Tyr Cys Leu Glu Asp Val Thr Arg Leu Ser Trp Tyr Gly
o
5'  UGAGGUUGAUCGUUUGAUGAAGGGCGUCCAGGGUCCACCAGCCGAAGGUAUGUUGGCCAAAUUGUGGUAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3150
3'  ACUCCAACUAGCAAACUACUUCCCGCAGGUCCCAGGUGGUCGGCUUCCAUACAACCGGUUUAACACCAUA
1     Glu Val Asp Arg Leu Met Lys Gly Val Gln Gly Pro Pro Ala Glu Gly Met Leu Ala Lys Leu Trp Tyr
2   Val Arg Leu Ile Val  .   .  Arg Ala Ser Arg Val His Gln Pro Lys Val Cys Trp Pro Asn Cys Gly Met
3    .  Gly  .  Ser Phe Asp Glu Gly Arg Pro Gly Ser Thr Ser Arg Arg Tyr Val Gly Gln Ile Val Val
o
5'  GGGCUUGGGCAGGACGUCAUCUCUGUUCUCGGUCGAGAUGUGUUGGAGGUUAUUCAGUCUGCUUCGGCUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3220
3'  CCCGAACCCGUCCUGCAGUAGAGACAAGAGCCAGCUCUACACAACCUCCAAUAAGUCAGACGAAGCCGAG
1    Gly Leu Gly Gln Asp Val Ile Ser Val Leu Gly Arg Asp Val Leu Glu Val Ile Gln Ser Ala Ser Ala
2     Gly Leu Gly Arg Thr Ser Ser Leu Phe Ser Val Glu Met Cys Trp Arg Leu Phe Ser Leu Leu Arg Leu
3   Trp Ala Trp Ala Gly Arg His Leu Cys Ser Arg Ser Arg Cys Val Gly Gly Tyr Ser Val Cys Phe Gly Ser
o
5'  CGGGACGCGAGUUUGCUGCUGAGUGGGGUUGGUUACCGCUUGAUGUGCGUGAGAAAUUCUUGCGCGGCCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3290
3'  GCCCUGCGCUCAAACGACGACUCACCCCAACCAAUGGCGAACUACACGCACUCUUUAAGAACGCGCCGGU
1   Pro Gly Arg Glu Phe Ala Ala Glu Trp Gly Trp Leu Pro Leu Asp Val Arg Glu Lys Phe Leu Arg Gly His
2    Arg Asp Ala Ser Leu Leu Leu Ser Gly Val Gly Tyr Arg Leu Met Cys Val Arg Asn Ser Cys Ala Ala
3     Gly Thr Arg Val Cys Cys  .  Val Gly Leu Val Thr Ala  .  Cys Ala  .  Glu Ile Leu Ala Arg Pro
o
5'  UUUGGGGCCUGCUGGGGCGUGGAUGAAUUUAAUACCUGCUAGUUUUGCCCCGUGGCUGAACCAGGCUGUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3360
3'  AAACCCCGGACGACCCCGCACCUACUUAAAUUAUGGACGAUCAAAACGGGGCACCGACUUGGUCCGACAA
1     Leu Gly Pro Ala Gly Ala Trp Met Asn Leu Ile Pro Ala Ser Phe Ala Pro Trp Leu Asn Gln Ala Val
2   Ile Trp Gly Leu Leu Gly Arg Gly  .  Ile  .  Tyr Leu Leu Val Leu Pro Arg Gly  .  Thr Arg Leu Leu
3    Phe Gly Ala Cys Trp Gly Val Asp Glu Phe Asn Thr Cys  .  Phe Cys Pro Val Ala Glu Pro Gly Cys
o
```

```
5'  AAUUUCGUCCUGUUUCAGGCAUUUGCCACCGGUGUCCGGUAUAGAGGAUCUCAUUGGCUUUUGUGCGUAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3430
3'  UUAAAGCAGGACAAAGUCCGUAAACGGUGGCCACAGGCCAUAUCUCCUAGAGUAACCGAAAACACGCAUG
1    Asn Phe Val Leu Phe Gln Ala Phe Ala Thr Gly Val Arg Tyr Arg Gly Ser His Trp Leu Leu Cys Val
2     Ile Ser Ser Cys Phe Arg His Leu Pro Pro Val Ser Gly Ile Glu Asp Leu Ile Gly Phe Cys Ala Tyr
3    . Phe Arg Pro Val Ser Gly Ile Cys His Arg Cys Pro Val . Arg Ile Ser Leu Ala Phe Val Arg Thr
o
5'  GACAAGCAAUAACCGGCGCGAUUAGCAGGCAGUUCGUGAACUCACGGCCUUGGUUGAUGCUACACUAAGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  3500
3'  CUGUUCGUUAUUGGCCGCGCUAAUCGUCCGUCAAGCACUUGAGUGCCGGAACCAACUACGAUGUGAUUCA
1   Arg Gln Ala Ile Thr Gly Ala Ile Ser Arg Gln Phe Val Asn Ser Arg Pro Trp Leu Met Leu His . Val
2    Asp Lys Gln . Pro Ala Arg Leu Ala Gly Ser Ser . Thr His Gly Leu Gly . Cys Tyr Thr Lys
3     Thr Ser Asn Asn Arg Arg Asp . Gln Ala Val Arg Glu Leu Thr Ala Leu Val Asp Ala Thr Leu Ser
o
5'  CGAUCAGUUGUCAUGGGCGGUUCUACUUGCCGGCGGGCCGUCUUCACGUGAUUUAGUCGGAAUAUCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++     3567
3'  GCUAGUCAACAGUACCCGCCAAGAUGAACGGCCGCCCGGCAGAAGUGCACUAAAUCAGCCUUAUAGG
1     Asp Gln Leu Ser Trp Ala Val Leu Leu Ala Gly Gly Pro Ser Ser Arg Asp Leu Val Gly Ile Ser
2   Ser Ile Ser Cys His Gly Arg Phe Tyr Leu Pro Ala Gly Arg Leu His Val Ile . Ser Glu Tyr Pro
3    Arg Ser Val Val Met Gly Gly Ser Thr Cys Arg Arg Ala Val Phe Thr . Phe Ser Arg Asn Ile
o
```

2.5.2 AaV1 dsRNA2 sequence (NC_010989)

2744 nt excluding poly (A) tail

```
5'  CACAAAGCAAUCGUAUCGCCAGAUACACAUAAGUCUUUGACCUUGGUUCGUUAUGUCUUCUGCUGUUUCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  70
3'  GUGUUUCGUUAGCAUAGCGGUCUAUGUGUAUUCAGAAACUGGAACCAAGCAAUACAGAAGACGACAAAGA
1    His Lys Ala Ile Val Ser Pro Asp Thr His Lys Ser Leu Thr Leu Val Arg Tyr Val Phe Cys Cys Phe
2     Thr Lys Gln Ser Tyr Arg Gln Ile His Ile Ser Leu . Pro Trp Phe Val Met Ser Ser Ala Val Ser
3   Pro Gln Ser Asn Arg Ile Ala Arg Tyr Thr . Val Phe Asp Leu Gly Ser Leu Cys Leu Leu Leu Phe Leu
o
5'  GAUUUCUUUGUCUGCGGGGGGUUUAGUGAGGGGUGGUUGGAGUGCGAUCUCCUUGAUGAGCGUACCGCUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  140
3'  CUAAAGAAACAGACGCCCCCCAAAUCACUCCCCACCAACCUCACGCUAGAGGAACUACUCGCAUGGCGAA
1    . Phe Leu Cys Leu Arg Gly Val . . Gly Val Val Gly Val Arg Ser Pro . . Ala Tyr Arg Phe
2    Asp Phe Phe Val Cys Gly Gly Phe Ser Glu Gly Trp Leu Glu Cys Asp Leu Leu Asp Glu Arg Thr Ala
3     Ile Ser Leu Ser Ala Gly Gly Leu Val Arg Gly Gly Trp Ser Ala Ile Ser Leu Met Ser Val Pro Leu
o
5'  UUCACCUUGCUGCGCACGUCGGGUCUGAUGGCAUGAUCGAUGUCCAUCUUGAGAAAAUUCCUGGCGUUCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  210
3'  AAGUGGAACGACGCGUGCAGCCCAGACUACCGUACUAGCUACAGGUAGAACUCUUUUAAGGACCGCAAGG
1     Ser Pro Cys Cys Ala Arg Arg Val . Trp His Asp Arg Cys Pro Ser . Glu Asn Ser Trp Arg Ser
2   Phe His Leu Ala Ala His Val Gly Ser Asp Gly Met Ile Asp Val His Leu Glu Lys Ile Pro Gly Val Pro
3    Phe Thr Leu Leu Arg Thr Ser Gly Leu Met Ala . Ser Met Ser Ile Leu Arg Lys Phe Leu Ala Phe
o
5'  CGUUUCCCCAUGGCAUGUGGGGGGAGGGCUGCGUGAUUUUGGUGGAGGCGGUCUCGGGGGCUUUGUUGAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  280
3'  GCAAAGGGGUACCGUACACCCCCCUCCCGACGCACUAAAACCACCUCCGCCAGAGCCCCCGAAACAACUA
1    Arg Phe Pro Met Ala Cys Gly Gly Arg Ala Ala . Phe Trp Trp Arg Arg Ser Arg Gly Leu Cys .
2     Val Ser Pro Trp His Val Gly Gly Gly Leu Arg Asp Phe Gly Gly Gly Gly Leu Gly Gly Phe Val Asp
3   Pro Phe Pro His Gly Met Trp Gly Glu Gly Cys Val Ile Leu Val Glu Ala Val Ser Gly Ala Leu Leu Met
o
5'  GUCGGUCUUCUCAAAAGGCCUGUUCAGCAGUGGGAUGGGUUGGCUUGUGGCACUUGCCGUGUUGGUGAUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  350
3'  CAGCCAGAAGAGUUUUCCGGACAAGUCGUCACCCUACCCAACCGAACACCGUGAACGGCACAACCACUAG
1   Cys Arg Ser Ser Gln Lys Ala Cys Ser Ala Val Gly Trp Val Gly Leu Trp His Leu Pro Cys Trp . Ser
2    Val Gly Leu Leu Lys Arg Pro Val Gln Gln Trp Asp Gly Leu Ala Cys Gly Thr Cys Arg Val Gly Asp
3     Ser Val Phe Ser Lys Gly Leu Phe Ser Ser Gly Met Gly Trp Leu Val Ala Leu Ala Val Leu Val Ile
o
5'  UUGCUGGGAUGUGCGAGGUGGUCGGUUGCGGCCGCCGUCGCGUUGGCAUGCAGGUCGCUGAAACGGAGGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  420
3'  AACGACCCUACACGCUCCACCAGCCAACGCCGGCGGCAGCGCAACCGUACGUCCAGCGACUUUGCCUCCA
1     Cys Trp Asp Val Arg Gly Gly Arg Leu Arg Pro Pro Ser Arg Trp His Ala Gly Arg . Asn Gly Gly
2   Leu Ala Gly Met Cys Glu Val Val Gly Cys Gly Arg Arg Arg Val Gly Met Gln Val Ala Glu Thr Glu Val
3    Leu Leu Gly Cys Ala Arg Trp Ser Val Ala Ala Ala Val Ala Leu Ala Cys Arg Ser Leu Lys Arg Arg
o
5'  GCUUGUCACAAGCUCCAUAAUGGUUGCGGUGCAGCGCGCGGCUAUCGCGCUUGCUCCUGACGAUAUGGAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  490
3'  CGAACAGUGUUCGAGGUAUUACCAACGCCACGUCGCGCGCCGAUAGCGCGAACGAGGACUGCUAUACCUC
1    Ala Cys His Lys Leu His Asn Gly Cys Gly Ala Ala Arg Gly Tyr Arg Ala Cys Ser . Arg Tyr Gly
2     Leu Val Thr Ser Ser Ile Met Val Ala Val Gln Arg Ala Ala Ile Ala Leu Ala Pro Asp Asp Met Glu
3   Cys Leu Ser Gln Ala Pro . Trp Leu Arg Cys Ser Ala Arg Leu Ser Arg Leu Leu Leu Thr Ile Trp Ser
o
5'  UUUUGGGACACCGUGCCUGUCCGUGUUUCUGCUGGAGGCUAUACCUUCAUCAAGGACAUGUCCCUCGACU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  560
3'  AAAACCCUGUGGCACGGACAGGCACAAAGACGACCUCCGAUAUGGAAGUAGUUCCUGUACAGGGAGCUGA
1   Val Leu Gly His Arg Ala Cys Pro Cys Phe Cys Trp Arg Leu Tyr Leu His Gln Gly His Val Pro Arg Leu
2    Phe Trp Asp Thr Val Pro Val Arg Val Ser Ala Gly Gly Tyr Thr Phe Ile Lys Asp Met Ser Leu Asp
3     Phe Gly Thr Pro Cys Leu Ser Val Phe Leu Leu Glu Ala Ile Pro Ser Ser Arg Thr Cys Pro Ser Thr
o
```

```
5'  UGGCCGCGGAGUCAGAGGUUAUUACCAGUGGUGCUGAGAAGUUGCUCGCGGACUUCUUUCGUAGCACGGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   630
3'  ACCGGCGCCUCAGUCUCCAAUAAUGGUCACCACGACUCUUCAACGAGCGCCUGAAGAAAGCAUCGUGCCG
1     Gly Arg Gly Val Arg Gly Tyr Tyr Gln Trp Cys  .  Glu Val Ala Arg Gly Leu Leu Ser  .  His Gly
2   Leu Ala Ala Glu Ser Glu Val Ile Thr Ser Gly Ala Glu Lys Leu Leu Ala Asp Phe Phe Arg Ser Thr Ala
3    Trp Pro Arg Ser Gln Arg Leu Leu Pro Val Val Leu Arg Ser Cys Ser Arg Thr Ser Phe Val Ala Arg
o
5'  CGAGCCCGACACGGAGGUCGUGAUGUUGUCGCAGGACGAUGAAGAGGCUGUUGUGGCCUCCUGCGCAGAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   700
3'  GCUCGGGCUGUGCCUCCAGCACUACAACAGCGUCCUGCUACUUCUCCGACAACACCGGAGGACGCGUCUA
1    Arg Ala Arg His Gly Gly Arg Asp Val Val Ala Gly Arg  .  Arg Gly Cys Cys Gly Leu Leu Arg Arg
2     Glu Pro Asp Thr Glu Val Val Met Leu Ser Gln Asp Asp Glu Glu Ala Val Val Ala Ser Cys Ala Asp
3   Pro Ser Pro Thr Arg Arg Ser  .  Cys Cys Arg Arg Thr Met Lys Arg Leu Leu Trp Pro Pro Ala Gln Ile
o
5'  CACGAGGUCGGCUGGGGUAACCUGGCAACACUUUCACCCUGUGUGGCAGUGGAUGGCGCCGAUGAUGAUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   770
3'  GUGCUCCAGCCGACCCCAUUGGACCGUUGUGAAAGUGGGACACACCGUCACCUACCGCGGCUACUACUAC
1   Ser Arg Gly Arg Leu Gly  .  Pro Gly Asn Thr Phe Thr Leu Cys Gly Ser Gly Trp Arg Arg  .   .  Cys
2    His Glu Val Gly Trp Gly Asn Leu Ala Thr Leu Ser Pro Cys Val Ala Val Asp Gly Ala Asp Asp Asp
3     Thr Arg Ser Ala Gly Val Thr Trp Gln His Phe His Pro Val Trp Gln Trp Met Ala Pro Met Met Met
o
5'  UUGUUUCUGUUGUGUCCGGGCGGGGUGACGACGUAGGGUUUGAGGUCGAUACGACUCAAGGUGACGGGCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   840
3'  AACAAAGACAACACAGGCCCGCCCCACUGCUGCAUCCCAAACUCCAGCUAUGCUGAGUUCCACUGCCCGA
1     Cys Phe Cys Cys Val Arg Ala Gly  .  Arg Arg Arg Val  .  Gly Arg Tyr Asp Ser Arg  .  Arg Ala
2   Val Val Ser Val Val Ser Gly Arg Gly Asp Asp Val Gly Phe Glu Val Asp Thr Thr Gln Gly Asp Gly Leu
3    Leu Phe Leu Leu Cys Pro Gly Gly Val Thr Thr  .  Gly Leu Arg Ser Ile Arg Leu Lys Val Thr Gly
o
5'  GGGGGAGCACCUUGACGACUUCCUCUUUCUGAGGGGCGGAACUGCGGUGGAGAAGAGUAGAUCGGUGGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   910
3'  CCCCCUCGUGGAACUGCUGAAGGAGAAAGACUCCCCGCCUUGACGCCACCUCUUCUCAUCUAGCCACCGC
1    Gly Gly Ala Pro  .  Arg Leu Pro Leu Ser Glu Gly Arg Asn Cys Gly Gly Glu Glu  .  Ile Gly Gly
2     Gly Glu His Leu Asp Asp Phe Leu Phe Leu Arg Gly Gly Thr Ala Val Glu Lys Ser Arg Ser Val Ala
3   Trp Gly Ser Thr Leu Thr Thr Ser Ser Phe  .  Gly Ala Glu Leu Arg Trp Arg Arg Val Asp Arg Trp Arg
o
5'  GGCAACCUUCUUGCCUUGUUGCGACGCGUGGGUGGCACUGGUGUUGAUGCGGAGGGAGUCGUUGUGAUCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   980
3'  CCGUUGGAAGAACGGAACAACGCUGCGCACCCACCGUGACCACAACUACGCCUCCCUCAGCAACACUAGA
1   Gly Gln Pro Ser Cys Leu Val Ala Thr Arg Gly Trp His Trp Cys  .  Cys Gly Gly Ser Arg Cys Asp Leu
2    Gly Asn Leu Leu Ala Leu Leu Arg Arg Val Gly Gly Thr Gly Val Asp Ala Glu Gly Val Val Val Ile
3     Ala Thr Phe Leu Pro Cys Cys Asp Ala Trp Val Ala Leu Val Leu Met Arg Arg Glu Ser Leu  .  Ser
o
5'  UCGGGGACUCGCCGGGCGUUGUGGCCCGCGAGCUAGCAAGCGUGGGGUAUCGCGUGCUGGGUGUCGACAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1050
3'  AGCCCCUGAGCGGCCCGCAACACCGGGCGCUCGAUCGUUCGCACCCCAUAGCGCACGACCCACAGCUGUU
1     Arg Gly Leu Ala Gly Arg Cys Gly Pro Arg Ala Ser Lys Arg Gly Val Ser Arg Ala Gly Cys Arg Gln
2   Phe Gly Asp Ser Pro Gly Val Val Ala Arg Glu Leu Ala Ser Val Gly Tyr Arg Val Leu Gly Val Asp Lys
3    Ser Gly Thr Arg Arg Ala Leu Trp Pro Ala Ser  .  Gln Ala Trp Gly Ile Ala Cys Trp Val Ser Thr
o
5'  GGAUCCGGCACACGCGGCACCCCCUGGCUGGCUCGACAAGUAUAGGACUGUCGUGGCUGAGGUCACUGAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1120
3'  CCUAGGCCGUGUGCGCCGUGGGGGACCGACCGAGCUGUUCAUAUCCUGACAGCACCGACUCCAGUGACUG
1    Gly Ser Gly Thr Arg Gly Thr Pro Trp Leu Ala Arg Gln Val  .  Asp Cys Arg Gly  .  Gly His  .
2     Asp Pro Ala His Ala Ala Pro Pro Gly Trp Leu Asp Lys Tyr Arg Thr Val Val Ala Glu Val Thr Asp
3   Arg Ile Arg His Thr Arg His Pro Leu Ala Gly Ser Thr Ser Ile Gly Leu Ser Trp Leu Arg Ser Leu Thr
o
```

```
5'  GGGCUGACCAGUUCUGAAGUGGAGGGGUGGCUGGCGCAGGUGGGUUGGGGGGGUAAGCGUGUUUUGGCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1190
3'  CCCGACUGGUCAAGACUUCACCUCCCCACCGACCGCGUCCACCCAACCCCCCCAUUCGCACAAAACCGAC
1   Arg Ala Asp Gln Phe  .  Ser Gly Gly Val Ala Gly Ala Gly Gly Leu Gly Gly  .  Ala Cys Phe Gly Cys
2    Gly Leu Thr Ser Ser Glu Val Glu Gly Trp Leu Ala Gln Val Gly Trp Gly Gly Lys Arg Val Leu Ala
3     Gly  .  Pro Val Leu Lys Trp Arg Gly Gly Trp Arg Arg Trp Val Gly Gly Val Ser Val Phe Trp Leu
o
5'  CCCUGAUGGACAUAGAUCAAGGUUCGAAACGAUCUUCCGUCAGUGAUACGGCGCUUAACCGCGAGCUCUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1260
3'  GGGACUACCUGUAUCUAGUUCCAAGCUUUGCUAGAAGGCAGUCACUAUGCCGCGAAUUGGCGCUCGAGAG
1     Pro Asp Gly His Arg Ser Arg Phe Glu Thr Ile Phe Arg Gln  .  Tyr Gly Ala  .  Pro Arg Ala Leu
2   Ala Leu Met Asp Ile Asp Gln Gly Ser Lys Arg Ser Ser Val Ser Asp Thr Ala Leu Asn Arg Glu Leu Ser
3    Pro  .  Trp Thr  .  Ile Lys Val Arg Asn Asp Leu Pro Ser Val Ile Arg Arg Leu Thr Ala Ser Ser
o
5'  GGCAAACCUGCUCGCAGACUGGAGCGAUGCGUUUACUGUGGUUCGUUACCGUGGCUUGCCCGUACUGCCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1330
3'  CCGUUUGGACGAGCGUCUGACCUCGCUACGCAAAUGACACCAAGCAAUGGCACCGAACGGGCAUGACGGC
1    Gly Lys Pro Ala Arg Arg Leu Glu Arg Cys Val Tyr Cys Gly Ser Leu Pro Trp Leu Ala Arg Thr Ala
2     Ala Asn Leu Leu Ala Asp Trp Ser Asp Ala Phe Thr Val Val Arg Tyr Arg Gly Leu Pro Val Leu Pro
3   Arg Gln Thr Cys Ser Gln Thr Gly Ala Met Arg Leu Leu Trp Phe Val Thr Val Ala Cys Pro Tyr Cys Arg
o
5'  AUUGAUGCGUACGUCUUGCCGACGCGCUUUCACGAACCGCAGGGGGCGGAGAUGUAUGCGGUGCACGGCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1400
3'  UAACUACGCAUGCAGAACGGCUGCGCGAAAGUGCUUGGCGUCCCCCGCCUCUACAUACGCCACGUGCCGA
1   Asp  .  Cys Val Arg Leu Ala Asp Ala Leu Ser Arg Thr Ala Gly Gly Gly Asp Val Cys Gly Ala Arg Leu
2    Ile Asp Ala Tyr Val Leu Pro Thr Arg Phe His Glu Pro Gln Gly Ala Glu Met Tyr Ala Val His Gly
3     Leu Met Arg Thr Ser Cys Arg Arg Ala Phe Thr Asn Arg Arg Gly Arg Arg Cys Met Arg Cys Thr Ala
o
5'  UGGAGGGGCCUGCACUCGCUGAUCUUGCCGCCUUCCAUGUAUCCAGCUGGCAUGUUGCGGGCUCUGUUUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1470
3'  ACCUCCCCGGACGUGAGCGACUAGAACGGCGGAAGGUACAUAGGUCGACCGUACAACGCCCGAGACAAAG
1     Gly Gly Ala Cys Thr Arg  .  Ser Cys Arg Leu Pro Cys Ile Gln Leu Ala Cys Cys Gly Leu Cys Phe
2   Leu Glu Gly Pro Ala Leu Ala Asp Leu Ala Ala Phe His Val Ser Ser Trp His Val Ala Gly Ser Val Ser
3    Trp Arg Gly Leu His Ser Leu Ile Leu Pro Pro Ser Met Tyr Pro Ala Gly Met Leu Arg Ala Leu Phe
o
5'  CUUCGGGGUGCAGUUGCUUUCCAAGGCAUGGCGUGAUUACGGCCGGCGACGGGGGCUAGGUUACGAGGGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1540
3'  GAAGCCCCACGUCAACGAAAGGUUCCGUACCGCACUAAUGCCGGCCGCUGCCCCCGAUCCAAUGCUCCCA
1    Leu Arg Gly Ala Val Ala Phe Gln Gly Met Ala  .  Leu Arg Pro Ala Thr Gly Ala Arg Leu Arg Gly
2     Phe Gly Val Gln Leu Leu Ser Lys Ala Trp Arg Asp Tyr Gly Arg Arg Arg Gly Leu Gly Tyr Glu Gly
3   Pro Ser Gly Cys Ser Cys Phe Pro Arg His Gly Val Ile Thr Ala Gly Asp Gly Gly  .  Val Thr Arg Val
o
5'  GGUGGUAAAGGGGAUGUGUCACCGGAGGGGGGCGAGCGGAAGGGGUUUACGGGCAAAUUCCGUGUUUCCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1610
3'  CCACCAUUUCCCCUACACAGUGGCCUCCCCCCGCUCGCCUUCCCCAAAUGCCCGUUUAAGGCACAAAGGC
1   Trp Trp  .  Arg Gly Cys Val Thr Gly Gly Gly Arg Ala Glu Gly Val Tyr Gly Gln Ile Pro Cys Phe Arg
2    Gly Gly Lys Gly Asp Val Ser Pro Glu Gly Gly Glu Arg Lys Gly Phe Thr Gly Lys Phe Arg Val Ser
3     Val Val Lys Gly Met Cys His Arg Arg Gly Ala Ser Gly Arg Gly Leu Arg Ala Asn Ser Val Phe Pro
o
5'  GGGCGACUGGUUACUUGUCGCUGCUGGGCCUGAGGAGUGGUGUUGAGGAGCUGCGACUGGCGCAUACUCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1680
3'  CCCGCUGACCAAUGAACAGCGACGACCCGGACUCCUCACCACAACUCCUCGACGCUGACCGCGUAUGAGG
1     Gly Asp Trp Leu Leu Val Ala Ala Gly Pro Glu Glu Trp Cys  .  Gly Ala Ala Thr Gly Ala Tyr Ser
2   Gly Ala Thr Gly Tyr Leu Ser Leu Leu Gly Leu Arg Ser Gly Val Glu Glu Leu Arg Leu Ala His Thr Pro
3    Gly Arg Leu Val Thr Cys Arg Cys Trp Ala  .  Gly Val Val Leu Arg Ser Cys Asp Trp Arg Ile Leu
o
```

```
5'    GUUUAUGCGGGCUGGGAAGCGUGCGCGGGCUUUUGUUGAGCUGCUUGGCCAACUCAAGACCGUAAUGGGG
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1750
3'    CAAAUACGCCCGACCCUUCGCACGCGCCCGAAAACAACUCGACGAACCGGUUGAGUUCUGGCAUUACCCC
1      Val Tyr Ala Gly Trp Glu Ala Cys Ala Gly Phe Cys . Ala Ala Trp Pro Thr Gln Asp Arg Asn Gly
2       Phe Met Arg Ala Gly Lys Arg Ala Arg Ala Phe Val Glu Leu Leu Gly Gln Leu Lys Thr Val Met Gly
3     Arg Leu Cys Gly Leu Gly Ser Val Arg Gly Leu Leu Leu Ser Cys Leu Ala Asn Ser Arg Pro . Trp Gly
o
5'    AAGGUUCGGCCUGAACGUGUGGCCGAGCUGACCGCCAAAUUCCAGCCUGCGGACAGAGUUUUGCUGCGGC
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1820
3'    UUCCAAGCCGGACUUGCACACCGGCUCGACUGGCGGUUUAAGGUCGGACGCCUGUCUCAAAACGACGCCG
1     Glu Gly Ser Ala . Thr Cys Gly Arg Ala Asp Arg Gln Ile Pro Ala Cys Gly Gln Ser Phe Ala Ala Ala
2      Lys Val Arg Pro Glu Arg Val Ala Glu Leu Thr Ala Lys Phe Gln Pro Ala Asp Arg Val Leu Leu Arg
3       Arg Phe Gly Leu Asn Val Trp Pro Ser . Pro Pro Asn Ser Ser Leu Arg Thr Glu Phe Cys Cys Gly
o
5'    ACGUGUCGCAAGCGACGUUGAGUGUGUCACCCACACUGCUGCAUACGAUGCAAUCCCGGUACGACGGAUU
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1890
3'    UGCACAGCGUUCGCUGCAACUCACACAGUGGGUGUGACGACGUAUGCUACGUUAGGGCCAUGCUGCCUAA
1       Arg Val Ala Ser Asp Val Glu Cys Val Thr His Thr Ala Ala Tyr Asp Ala Ile Pro Val Arg Arg Ile
2     His Val Ser Gln Ala Thr Leu Ser Val Ser Pro Thr Leu Leu His Thr Met Gln Ser Arg Tyr Asp Gly Leu
3      Thr Cys Arg Lys Arg Arg . Val Cys His Pro His Cys Cys Ile Arg Cys Asn Pro Gly Thr Thr Asp
o
5'    GCUGAAGUAUUGUAGCGCAGCCGAGUUGCCGGAUCAAGGUCCGGAUUUGGGUGAUUUCACCAAGCAUCCC
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1960
3'    CGACUUCAUAACAUCGCGUCGGCUCAACGGCCUAGUUCCAGGCCUAAACCCACUAAAGUGGUUCGUAGGG
1      Ala Glu Val Leu . Arg Ser Arg Val Ala Gly Ser Arg Ser Gly Phe Gly . Phe His Gln Ala Ser
2       Leu Lys Tyr Cys Ser Ala Ala Glu Leu Pro Asp Gln Gly Pro Asp Leu Gly Asp Phe Thr Lys His Pro
3     Cys . Ser Ile Val Ala Gln Pro Ser Cys Arg Ile Lys Val Arg Ile Trp Val Ile Ser Pro Ser Ile Pro
o
5'    GGGUUUGGCGCCCUGUUUCACUGGGGCUUCUCCAAGGUGCGCGAUGAGAUGGGGGUUGUCUUCCCCUACA
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2030
3'    CCCAAACCGCGGGACAAAGUGACCCCGAAGAGGUUCCACGCGCUACUCUACCCCCAACAGAAGGGGAUGU
1     Arg Val Trp Arg Pro Val Ser Leu Gly Leu Leu Gln Gly Ala Arg . Asp Gly Gly Cys Leu Pro Leu His
2      Gly Phe Gly Ala Leu Phe His Trp Gly Phe Ser Lys Val Arg Asp Glu Met Gly Val Val Phe Pro Tyr
3       Gly Leu Ala Pro Cys Phe Thr Gly Ala Ser Pro Arg Cys Ala Met Arg Trp Gly Leu Ser Ser Pro Thr
o
5'    CGCAGUAUGCGCGUGUGUUGCCGCUCGCGGCAAUGCCGAUACUUGCGAGUAGGAGCUGGGUGCGCAUGGU
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2100
3'    GCGUCAUACGCGCACACAACGGCGAGCGCCGUUACGGCUAUGAACGCUCAUCCUCGACCCACGCGUACCA
1       Ala Val Cys Ala Cys Val Ala Ala Arg Gly Asn Ala Asp Thr Cys Glu . Glu Leu Gly Ala His Gly
2     Thr Gln Tyr Ala Arg Val Leu Pro Leu Ala Ala Met Pro Ile Leu Ala Ser Arg Ser Trp Val Arg Met Val
3      Arg Ser Met Arg Val Cys Cys Arg Ser Arg Gln Cys Arg Tyr Leu Arg Val Gly Ala Gly Cys Ala Trp
o
5'    CGCCUGGCUGCUGAAAGCGUAUGAUCGCUUGAUGGGGAAGCCUCUCCAUACGUGGGAGCUACAAGGCUUG
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2170
3'    GCGGACCGACGACUUUCGCAUACUAGCGAACUACCCCUUCGGAGAGGUAUGCACCCUCGAUGUUCCGAAC
1      Arg Leu Ala Ala Glu Ser Val . Ser Leu Asp Gly Glu Ala Ser Pro Tyr Val Gly Ala Thr Arg Leu
2       Ala Trp Leu Leu Lys Ala Tyr Asp Arg Leu Met Gly Lys Pro Leu His Thr Trp Glu Leu Gln Gly Leu
3     Ser Pro Gly Cys . Lys Arg Met Ile Ala . Trp Gly Ser Leu Ser Ile Arg Gly Ser Tyr Lys Ala Cys
o
5'    UUGUGGAGCCUGUCUCACGUAGGCACGCAGGAGGAGCGCGAAGUUUACUUCUGGGGCAAGCUGCAGGGGG
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2240
3'    AACACCUCGGACAGAGUGCAUCCGUGCGUCCUCCUCGCGCUUCAAAUGAAGACCCCGUUCGACGUCCCCC
1     Val Val Glu Pro Val Ser Arg Arg His Ala Gly Gly Ala Arg Ser Leu Leu Leu Gly Gln Ala Ala Gly Gly
2      Leu Trp Ser Leu Ser His Val Gly Thr Gln Glu Glu Arg Glu Val Tyr Phe Trp Gly Lys Leu Gln Gly
3       Cys Gly Ala Cys Leu Thr . Ala Arg Arg Arg Ser Ala Lys Phe Thr Ser Gly Ala Ser Cys Arg Gly
o
```

```
5'  CAGCUGAUCUACUCCUGGCCUCGAGGCGUCGAGGUCAGCGCACGCAGGGUGCCGCUUCUGCGGACCUGAA
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   2310
3'  GUCGACUAGAUGAGGACCGGAGCUCCGCAGCUCCAGUCGCGUGCGUCCCACGGCGAAGACGCCUGGACUU
1    Ser  .  Ser  Thr  Pro  Gly  Leu  Glu  Ala  Ser  Arg  Ser  Ala  His  Ala  Gly  Cys  Arg  Phe  Cys  Gly  Pro  Glu
2   Ala  Ala  Asp  Leu  Leu  Leu  Ala  Ser  Arg  Arg  Arg  Gly  Gln  Arg  Thr  Gln  Gly  Ala  Ala  Ser  Ala  Asp  Leu  Asn
3    Gln  Leu  Ile  Tyr  Ser  Trp  Pro  Arg  Gly  Val  Glu  Val  Ser  Ala  Arg  Arg  Val  Pro  Leu  Leu  Arg  Thr  .
o
5'  UCGCAUUCUGAGCGCGGCUGAAGCUUUGGGGAGUCGGAAGCUGAAGUUGUCUCCGGCGGAACGUAUGGUG
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   2380
3'  AGCGUAAGACUCGCGCCGACUUCGAAACCCCUCAGCCUUCGACUUCAACAGAGGCCGCCUUGCAUACCAC
1    Ser  His  Ser  Glu  Arg  Gly  .  Ser  Phe  Gly  Glu  Ser  Glu  Ala  Glu  Val  Val  Ser  Gly  Gly  Thr  Tyr  Gly
2     Arg  Ile  Leu  Ser  Ala  Ala  Glu  Ala  Leu  Gly  Ser  Arg  Lys  Leu  Lys  Leu  Ser  Pro  Ala  Glu  Arg  Met  Val
3   Ile  Ala  Phe  .  Ala  Arg  Leu  Lys  Leu  Trp  Gly  Val  Gly  Ser  .  Ser  Cys  Leu  Arg  Arg  Asn  Val  Trp  Cys
o
5'  UUCGGCGCCGCUUAUGGGGUCGUCGGACGUGGCCAAUUCCGCCAUGACUCUGAUUACAGCCGCAAUCCUC
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   2450
3'  AAGCCGCGGCGAAUACCCCAGCAGCCUGCACCGGUUAAGGCGGUACUGAGACUAAUGUCGGCGUUAGGAG
1   Val  Arg  Arg  Arg  Leu  Trp  Gly  Arg  Arg  Thr  Trp  Pro  Ile  Pro  Pro  .  Leu  .  Leu  Gln  Pro  Gln  Ser  Ser
2    Phe  Gly  Ala  Ala  Tyr  Gly  Val  Val  Gly  Arg  Gly  Gln  Phe  Arg  His  Asp  Ser  Asp  Tyr  Ser  Arg  Asn  Pro
3     Ser  Ala  Pro  Leu  Met  Gly  Ser  Ser  Asp  Val  Ala  Asn  Ser  Ala  Met  Thr  Leu  Ile  Thr  Ala  Ala  Ile  Leu
o
5'  CCGGCGGGUUGCGGUACGCGGAACGUCCCACGCUAGGGCGUAUGGCGUCCACUUCCGAUGGUACAAGCCA
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   2520
3'  GGCCGCCCAACGCCAUGCGCCUUGCAGGGUGCGAUCCCGCAUACCGCAGGUGAAGGCUACCAUGUUCGGU
1     Arg  Arg  Val  Ala  Val  Arg  Gly  Thr  Ser  His  Ala  Arg  Ala  Tyr  Gly  Val  His  Phe  Arg  Trp  Tyr  Lys  Pro
2   Pro  Gly  Gly  Leu  Arg  Tyr  Ala  Glu  Arg  Pro  Thr  Leu  Gly  Arg  Met  Ala  Ser  Thr  Ser  Asp  Gly  Thr  Ser  His
3    Pro  Ala  Gly  Cys  Gly  Thr  Arg  Asn  Val  Pro  Arg  .  Gly  Val  Trp  Arg  Pro  Leu  Pro  Met  Val  Gln  Ala
o
5'  UCAUCGUACGCUGUCCAGCAGUACAAAUGGAAGCCGCGGUGUUUACACUCCGCGCAGAAUCGGCUAGCUU
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   2590
3'  AGUAGCAUGCGACAGGUCGUCAUGUUUACCUUCGGCGCCACAAAUGUGAGGCGCGUCUUAGCCGAUCGAA
1    Ser  Ser  Tyr  Ala  Val  Gln  Gln  Tyr  Lys  Trp  Lys  Pro  Arg  Cys  Leu  His  Ser  Ala  Gln  Asn  Arg  Leu  Ala
2     His  Arg  Thr  Leu  Ser  Ser  Ser  Thr  Asn  Gly  Ser  Arg  Gly  Val  Tyr  Thr  Pro  Arg  Arg  Ile  Gly  .  Leu
3   Ile  Ile  Val  Arg  Cys  Pro  Ala  Val  Gln  Met  Glu  Ala  Ala  Val  Phe  Thr  Leu  Arg  Ala  Glu  Ser  Ala  Ser  Leu
o
5'  AAAUUAAGUCGUUCCAUGAGUCAUCCGUUCUUACGAGCUGGACGUGUUUUUAGAAAGAAGCUACAUAGCG
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   2660
3'  UUUAAUUCAGCAAGGUACUCAGUAGGCAAGAAUGCUCGACCUGCACAAAAAUCUUUCUUCGAUGUAUCGC
1    .  Ile  Lys  Ser  Phe  His  Glu  Ser  Ser  Val  Leu  Thr  Ser  Trp  Thr  Cys  Phe  .  Lys  Glu  Ala  Thr  .  Arg
2    Lys  Leu  Ser  Arg  Ser  Met  Ser  His  Pro  Phe  Leu  Arg  Ala  Gly  Arg  Val  Phe  Arg  Lys  Lys  Leu  His  Ser
3     Asn  .  Val  Val  Pro  .  Val  Ile  Arg  Ser  Tyr  Glu  Leu  Asp  Val  Phe  Leu  Glu  Arg  Ser  Tyr  Ile  Ala
o
5'  GCGGGGCCUCCGUUUUCUCUUUAGGCCCUUGAGUCAGUCGGAUCUCCUCGGCGGUGGGGUGGUCUGGUGU
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|   2730
3'  CGCCCCGGAGGCAAAAGAGAAAUCCGGGAACUCAGUCAGCCUAGAGGAGCCGCCACCCCACCAGACCACA
1     Arg  Gly  Leu  Arg  Phe  Leu  Phe  Arg  Pro  Leu  Ser  Gln  Ser  Asp  Leu  Leu  Gly  Gly  Gly  Val  Val  Trp  Cys
2   Gly  Gly  Ala  Ser  Val  Phe  Ser  Leu  Gly  Pro  .  Val  Ser  Arg  Ile  Ser  Ser  Ala  Val  Gly  Trp  Ser  Gly  Val
3    Ala  Gly  Pro  Pro  Phe  Ser  Leu  .  Ala  Leu  Glu  Ser  Val  Gly  Ser  Pro  Arg  Arg  Trp  Gly  Gly  Leu  Val
o
5'  AGUGGAAC
o   ++++|+++                                                                 2738
3'  UCACCUUG
1    Ser  Gly  Thr
2     Val  Glu
3   .  Trp  Asn
o
```

2.5.3 AaV1 dsRNA2 del-1 sequence

2513 nt excluding poly (A) tail

```
5' CACAAAGCAAUCGUAUCGCCAGAUACACAUAAGUCUUUGACCUUGGUUCGUUAUGUCUUCUGCUGUUUCU
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  70
3' GUGUUUCGUUAGCAUAGCGGUCUAUGUGUAUUCAGAAACUGGAACCAAGCAAUACAGAAGACGACAAAGA
1  His Lys Ala Ile Val Ser Pro Asp Thr His Lys Ser Leu Thr Leu Val Arg Tyr Val Phe Cys Cys Phe
2  Thr Lys Gln Ser Tyr Arg Gln Ile His Ile Ser Leu . Pro Trp Phe Val Met Ser Ser Ala Val Ser
3  Ser Gln Ser Asn Arg Ile Ala Arg Tyr Thr . Val Phe Asp Leu Gly Ser Leu Cys Leu Leu Leu Phe Leu
o
5' GAUUUCUUUGUCUGCGGGGGGUUUAGUGAGGGGUGGUUGGAGUGCGAUCUCCUUGAUGAGCGUACCGCUU
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  140
3' CUAAAGAAACAGACGCCCCCCAAAUCACUCCCCACCAACCUCACGCUAGAGGAACUACUCGCAUGGCGAA
1  . Phe Leu Cys Leu Arg Gly Val . . Gly Val Val Gly Val Arg Ser Pro . . Ala Tyr Arg Phe
2  Asp Phe Phe Val Cys Gly Gly Phe Ser Glu Gly Trp Leu Glu Cys Asp Leu Leu Asp Glu Arg Thr Ala
3  Ile Ser Leu Ser Ala Gly Gly Leu Val Arg Gly Gly Trp Ser Ala Ile Ser Leu Met Ser Val Pro Leu
o
5' UUCACCUUGCUGCGCACGUCGGGUCUGAUGGCAUGAUCGAUGUCCAUCUUGAGAAAAUUCCUGGCGUUCC
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  210
3' AAGUGGAACGACGCGUGCAGCCCAGACUACCGUACUAGCUACAGGUAGAACUCUUUUAAGGACCGCAAGG
1  Ser Pro Cys Cys Ala Arg Arg Val . Trp His Asp Arg Cys Pro Ser . Glu Asn Ser Trp Arg Ser
2  Phe His Leu Ala Ala His Val Gly Ser Asp Gly Met Ile Asp Val His Leu Glu Lys Ile Pro Gly Val Pro
3  Phe Thr Leu Leu Arg Thr Ser Gly Leu Met Ala . Ser Met Ser Ile Leu Arg Lys Phe Leu Ala Phe
o
5' CGUUUCCCCAUGGCAUGUGGGGGGAGGGCUGCGUGAUUUUGGUGGAGGCGGUCUCGGGGGCUUUGUUGAU
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  280
3' GCAAAGGGGUACCGUACACCCCCCUCCCGACGCACUAAAACCACCUCCGCCAGAGCCCCCGAAACAACUA
1  Arg Phe Pro Met Ala Cys Gly Gly Arg Ala Ala . Phe Trp Trp Arg Arg Ser Arg Gly Leu Cys .
2  Val Ser Pro Trp His Val Gly Gly Gly Leu Arg Asp Phe Gly Gly Gly Gly Leu Gly Gly Phe Val Asp
3  Pro Phe Pro His Gly Met Trp Gly Glu Gly Cys Val Ile Leu Val Glu Ala Val Ser Gly Ala Leu Leu Met
o
5' GUCGGUCUUCUCAAAAGGCCUGUUCAGCAGUGGGAUGGGUUGGCUUGUGGCACUUGCCGUGUUGGUGAUC
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  350
3' CAGCCAGAAGAGUUUUCCGGACAAGUCGUCACCCUACCCAACCGAACACCGUGAACGGCACAACCACUAG
1  Cys Arg Ser Ser Gln Lys Ala Cys Ser Ala Val Gly Trp Val Gly Leu Trp His Leu Pro Cys Trp . Ser
2  Val Gly Leu Leu Lys Arg Pro Val Gln Gln Trp Asp Gly Leu Ala Cys Gly Thr Cys Arg Val Gly Asp
3  Ser Val Phe Ser Lys Gly Leu Phe Ser Ser Gly Met Gly Trp Leu Val Ala Leu Ala Val Leu Val Ile
o
5' UUGCUGGGAUGUGCGAGGUGGUCGGUUGCGGCCGCCGUCGCGUUGGCAUGCAGGUCGCUGAAACGGAGGU
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  420
3' AACGACCCUACACGCUCCACCAGCCAACGCCGGCGGCAGCGCAACCGUACGUCCAGCGACUUUGCCUCCA
1  Cys Trp Asp Val Arg Gly Gly Arg Leu Arg Pro Pro Ser Arg Trp His Ala Gly Arg . Asn Gly Gly
2  Leu Ala Gly Met Cys Glu Val Val Gly Cys Gly Arg Arg Arg Val Gly Met Gln Val Ala Glu Thr Glu Val
3  Leu Leu Gly Cys Ala Arg Trp Ser Val Ala Ala Ala Val Ala Leu Ala Cys Arg Ser Leu Lys Arg Arg
o
5' GCUUGUCACAAGCUCCAUAAUGGUUGCGGUGCAGCGCGCGGCUAUCGCGCUUGCUCCUGACGAUAUGGAG
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  490
3' CGAACAGUGUUCGAGGUAUUACCAACGCCACGUCGCGCGCCGAUAGCGCGAACGAGGACUGCUAUACCUC
1  Ala Cys His Lys Leu His Asn Gly Cys Gly Ala Ala Arg Gly Tyr Arg Ala Cys Ser . Arg Tyr Gly
2  Leu Val Thr Ser Ser Ile Met Val Ala Val Gln Arg Ala Ala Ile Ala Leu Ala Pro Asp Asp Met Glu
3  Cys Leu Ser Gln Ala Pro . Trp Leu Arg Cys Ser Ala Arg Leu Ser Arg Leu Leu Leu Thr Ile Trp Ser
o
5' UUUUGGGACACCGUGCCUGUCCGUGUUUCUGCUGGAGGCUAUACCUUCAUCAAGGACAUGUCCCUCGACU
o  ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  560
3' AAAACCCUGUGGCACGGACAGGCACAAAGACGACCUCCGAUAUGGAAGUAGUUCCUGUACAGGGAGCUGA
1  Val Leu Gly His Arg Ala Cys Pro Cys Phe Cys Trp Arg Leu Tyr Leu His Gln Gly His Val Pro Arg Leu
2  Phe Trp Asp Thr Val Pro Val Arg Val Ser Ala Gly Gly Tyr Thr Phe Ile Lys Asp Met Ser Leu Asp
3  Phe Gly Thr Pro Cys Leu Ser Val Phe Leu Leu Glu Ala Ile Pro Ser Ser Arg Thr Cys Pro Ser Thr
o
```

```
5'  UGGCCGCGGAGUCAGAGGUUAUUACCAGUGGUGCUGAGAAGUUGCUCGCGGACUUCUUUCGUAGCACGGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   630
3'  ACCGGCGCCUCAGUCUCCAAUAAUGGUCACCACGACUCUUCAACGAGCGCCUGAAGAAAGCAUCGUGCCG
1     Gly Arg Gly Val Arg Gly Tyr Tyr Gln Trp Cys . Glu Val Ala Arg Gly Leu Leu Ser . His Gly
2   Leu Ala Ala Glu Ser Glu Val Ile Thr Ser Gly Ala Glu Lys Leu Leu Ala Asp Phe Phe Arg Ser Thr Ala
3    Trp Pro Arg Ser Gln Arg Leu Leu Pro Val Val Leu Arg Ser Cys Ser Arg Thr Ser Phe Val Ala Arg
o
5'  CGAGCCCGACACGGAGGUCGUGAUGUUGUCGCAGGACGAUGAAGAGGCUGUUGUGGCCUCCUGCGCAGAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   700
3'  GCUCGGGCUGUGCCUCCAGCACUACAACAGCGUCCUGCUACUUCUCCGACAACACCGGAGGACGCGUCUA
1    Arg Ala Arg His Gly Gly Arg Asp Val Val Ala Gly Arg . Arg Gly Cys Cys Gly Leu Leu Arg Arg
2     Glu Pro Asp Thr Glu Val Val Met Leu Ser Gln Asp Asp Glu Glu Ala Val Val Ala Ser Cys Ala Asp
3   Pro Ser Pro Thr Arg Arg Ser . Cys Cys Arg Arg Thr Met Lys Arg Leu Leu Trp Pro Pro Ala Gln Ile
o
5'  CACGAGGUCGGCUGGGGUAACCUGGCAACACUUUCACCCUGUGUGGCAGUGGAUGGCGCCGAUGAUGAUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   770
3'  GUGCUCCAGCCGACCCCAUUGGACCGUUGUGAAAGUGGGACACACCGUCACCUACCGCGGCUACUACUAC
1   Ser Arg Gly Arg Leu Gly . Pro Gly Asn Thr Phe Thr Leu Cys Gly Ser Gly Trp Arg Arg . . Cys
2    His Glu Val Gly Trp Gly Asn Leu Ala Thr Leu Ser Pro Cys Val Ala Val Asp Gly Ala Asp Asp Asp
3     Thr Arg Ser Ala Gly Val Thr Trp Gln His Phe His Pro Val Trp Gln Trp Met Ala Pro Met Met Met
o
5'  UUGUUUCUGUUGUGUCCGGGCGGGGUGACGACGUAGGGUUUGAGGUCGAUACGACUCAAGGUGACGGGCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   840
3'  AACAAAGACAACACAGGCCCGCCCCACUGCUGCAUCCCAAACUCCAGCUAUGCUGAGUUCCACUGCCCGA
1     Cys Phe Cys Cys Val Arg Ala Gly . Arg Arg Arg Val . Gly Arg Tyr Asp Ser Arg . Arg Ala
2   Val Val Ser Val Val Ser Gly Arg Gly Asp Asp Val Gly Phe Glu Val Asp Thr Thr Gln Gly Asp Gly Leu
3    Leu Phe Leu Leu Cys Pro Gly Gly Val Thr Thr . Gly Leu Arg Ser Ile Arg Leu Lys Val Thr Gly
o
5'  GGGGGAGCACCUUGACGACUUCCUCUUUCUGAGGGGCGGAACUGCGGUGGAGAAGAGUAGAUCGGUGGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   910
3'  CCCCCUCGUGGAACUGCUGAAGGAGAAAGACUCCCCGCCUUGACGCCACCUCUUCUCAUCUAGCCACCGC
1    Gly Gly Ala Pro . Arg Leu Pro Leu Ser Glu Gly Arg Asn Cys Gly Gly Glu Glu . Ile Gly Gly
2     Gly Glu His Leu Asp Asp Phe Leu Phe Leu Arg Gly Gly Thr Ala Val Glu Lys Ser Arg Ser Val Ala
3   Trp Gly Ser Thr Leu Thr Thr Ser Ser Phe . Gly Ala Glu Leu Arg Trp Arg Arg Val Asp Arg Trp Arg
o
5'  GGCAACCUUCUUGCCUUGUUGCGACGCGUGGGUGGCACUGGUGUUGAUGCGGAGGGAGUCGUUGUGAUCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   980
3'  CCGUUGGAAGAACGGAACAACGCUGCGCACCCACCGUGACCACAACUACGCCUCCCUCAGCAACACUAGA
1   Gly Gln Pro Ser Cys Leu Val Ala Thr Arg Gly Trp His Trp Cys . Cys Gly Gly Ser Arg Cys Asp Leu
2    Gly Asn Leu Leu Ala Leu Leu Arg Arg Val Gly Gly Thr Gly Val Asp Ala Glu Gly Val Val Val Ile
3     Ala Thr Phe Leu Pro Cys Cys Asp Ala Trp Val Ala Leu Val Leu Met Arg Arg Glu Ser Leu . Ser
o
5'  UCGGGGACUCGCCGGGCGUUGUGGCCCGCGAGCUAGCAAGCGUGGGGUAUCGCGUGCUGGGUGUCGACAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1050
3'  AGCCCCUGAGCGGCCCGCAACACCGGGCGCUCGAUCGUUCGCACCCCAUAGCGCACGACCCACAGCUGUU
1     Arg Gly Leu Ala Gly Arg Cys Gly Pro Arg Ala Ser Lys Arg Gly Val Ser Arg Ala Gly Cys Arg Gln
2   Phe Gly Asp Ser Pro Gly Val Val Ala Arg Glu Leu Ala Ser Val Gly Tyr Arg Val Leu Gly Val Asp Lys
3    Ser Gly Thr Arg Arg Ala Leu Trp Pro Ala Ser . Gln Ala Trp Gly Ile Ala Cys Trp Val Ser Thr
o
5'  GGAUCCGGCACACGCGGCACCCCCUGGCUGGCUCGACAAGUAUAGGACUGUCGUGGCUGAGGUCACUGAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1120
3'  CCUAGGCCGUGUGCGCCGUGGGGGACCGACCGAGCUGUUCAUAUCCUGACAGCACCGACUCCAGUGACUG
1    Gly Ser Gly Thr Arg Gly Thr Pro Trp Leu Ala Arg Gln Val . Asp Cys Arg Gly . Gly His .
2     Asp Pro Ala His Ala Ala Pro Pro Gly Trp Leu Asp Lys Tyr Arg Thr Val Val Ala Glu Val Thr Asp
3   Arg Ile Arg His Thr Arg His Pro Leu Ala Gly Ser Thr Ser Ile Gly Leu Ser Trp Leu Arg Ser Leu Thr
o
```

```
5'   GGGCUGACCAGUUCUGAAGUGGAGGGGUGGCUGGCGCAGGUGGGUUGGGGGGGUAAGCGUGUUUUGGCUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1190
3'   CCCGACUGGUCAAGACUUCACCUCCCCACCGACCGCGUCCACCCAACCCCCCCAUUCGCACAAAACCGAC
1   Arg Ala Asp Gln Phe  .  Ser Gly Gly Val Ala Gly Ala Gly Gly Leu Gly Gly  .  Ala Cys Phe Gly Cys
2    Gly Leu Thr Ser Ser Glu Val Glu Gly Trp Leu Ala Gln Val Gly Trp Gly Gly Lys Arg Val Leu Ala
3     Gly  .  Pro Val Leu Lys Trp Arg Gly Gly Trp Arg Arg Trp Val Gly Gly Val Ser Val Phe Trp Leu
o
5'   CCCUGAUGGACAUAGAUCAAGGUUCGAAACGAUCUUCCGUCAGUGAUACGGCGCUUAACCGCGAGCUCUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1260
3'   GGGACUACCUGUAUCUAGUUCCAAGCUUUGCUAGAAGGCAGUCACUAUGCCGCGAAUUGGCGCUCGAGAG
1     Pro Asp Gly His Arg Ser Arg Phe Glu Thr Ile Phe Arg Gln  .  Tyr Gly Ala  .  Pro Arg Ala Leu
2   Ala Leu Met Asp Ile Asp Gln Gly Ser Lys Arg Ser Ser Val Ser Asp Thr Ala Leu Asn Arg Glu Leu Ser
3    Pro  .  Trp Thr  .  Ile Lys Val Arg Asn Asp Leu Pro Ser Val Ile Arg Arg Leu Thr Ala Ser Ser
o
5'   GGCAAACCUGCUCGAUUACGGCCGGCGACGGGGGCUAGGUUACGAGGGUGGUGGUAAAGGGGAUGUGUCA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1330
3'   CCGUUUGGACGAGCUAAUGCCGGCCGCUGCCCCCGAUCCAAUGCUCCCACCACCAUUUCCCCUACACAGU
1    Gly Lys Pro Ala Arg Leu Arg Pro Ala Thr Gly Ala Arg Leu Arg Gly Trp Trp  .  Arg Gly Cys Val
2     Ala Asn Leu Leu Asp Tyr Gly Arg Arg Arg Gly Leu Gly Tyr Glu Gly Gly Gly Lys Gly Asp Val Ser
3   Arg Gln Thr Cys Ser Ile Thr Ala Gly Asp Gly Gly  .  Val Thr Arg Val Val Val Lys Gly Met Cys His
o
5'   CCGGAGGGGGGCGAGCGGAAGGGGUUUACGGGCAAAUUCCGUGUUUCCGGGGCGACUGGUUACUUGUCGC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1400
3'   GGCCUCCCCCCGCUCGCCUUCCCCAAAUGCCCGUUUAAGGCACAAAGGCCCCGCUGACCAAUGAACAGCG
1   Thr Gly Gly Gly Arg Ala Glu Gly Val Tyr Gly Gln Ile Pro Cys Phe Arg Gly Asp Trp Leu Leu Val Ala
2    Pro Glu Gly Gly Glu Arg Lys Gly Phe Thr Gly Lys Phe Arg Val Ser Gly Ala Thr Gly Tyr Leu Ser
3     Arg Arg Gly Ala Ser Gly Arg Gly Leu Arg Ala Asn Ser Val Phe Pro Gly Arg Leu Val Thr Cys Arg
o
5'   UGCUGGGCCUGAGGAGUGGUGUUGAGGAGCUGCGACUGGCGCAUACUCCGUUUAUGCGGGCUGGGAAGCG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1470
3'   ACGACCCGGACUCCUCACCACAACUCCUCGACGCUGACCGCGUAUGAGGCAAAUACGCCCGACCCUUCGC
1     Ala Gly Pro Glu Glu Trp Cys  .  Gly Ala Ala Thr Gly Ala Tyr Ser Val Tyr Ala Gly Trp Glu Ala
2   Leu Leu Gly Leu Arg Ser Gly Val Glu Glu Leu Arg Leu Ala His Thr Pro Phe Met Arg Ala Gly Lys Arg
3    Cys Trp Ala  .  Gly Val Val Leu Arg Ser Cys Asp Trp Arg Ile Leu Arg Leu Cys Gly Leu Gly Ser
o
5'   UGCGCGGGCUUUUGUUGAGCUGCUUGGCCAACUCAAGACCGUAAUGGGGAAGGUUCGGCCUGAACGUGUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1540
3'   ACGCGCCCGAAAACAACUCGACGAACCGGUUGAGUUCUGGCAUUACCCCUUCCAAGCCGGACUUGCACAC
1    Cys Ala Gly Phe Cys  .  Ala Ala Trp Pro Thr Gln Asp Arg Asn Gly Glu Gly Ser Ala  .  Thr Cys
2     Ala Arg Ala Phe Val Glu Leu Leu Gly Gln Leu Lys Thr Val Met Gly Lys Val Arg Pro Glu Arg Val
3   Val Arg Gly Leu Leu Leu Ser Cys Leu Ala Asn Ser Arg Pro  .  Trp Gly Arg Phe Gly Leu Asn Val Trp
o
5'   GCCGAGCUGACCGCCAAAUUCCAGCCUGCGGACAGAGUUUUGCUGCGGCACGUGUCGCAAGCGACGUUGA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1610
3'   CGGCUCGACUGGCGGUUUAAGGUCGGACGCCUGUCUCAAAACGACGCCGUGCACAGCGUUCGCUGCAACU
1   Gly Arg Ala Asp Arg Gln Ile Pro Ala Cys Gly Gln Ser Phe Ala Ala Ala Arg Val Ala Ser Asp Val Glu
2    Ala Glu Leu Thr Ala Lys Phe Gln Pro Ala Asp Arg Val Leu Leu Arg His Val Ser Gln Ala Thr Leu
3     Pro Ser  .  Pro Pro Asn Ser Ser Leu Arg Thr Glu Phe Cys Cys Gly Thr Cys Arg Lys Arg Arg  .
o
5'   GUGUGUCACCCACACUGCUGCAUACGAUGCAAUCCCGGUACGACGGAUUGCUGAAGUAUUGUAGCGCAGC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1680
3'   CACACAGUGGGUGUGACGACGUAUGCUACGUUAGGGCCAUGCUGCCUAACGACUUCAUAACAUCGCGUCG
1     Cys Val Thr His Thr Ala Ala Tyr Asp Ala Ile Pro Val Arg Arg Ile Ala Glu Val Leu  .  Arg Ser
2   Ser Val Ser Pro Thr Leu Leu His Thr Met Gln Ser Arg Tyr Asp Gly Leu Leu Lys Tyr Cys Ser Ala Ala
3    Val Cys His Pro His Cys Cys Ile Arg Cys Asn Pro Gly Thr Thr Asp Cys  .  Ser Ile Val Ala Gln
o
```

```
5'  CGAGUUGCCGGAUCAAGGUCCGGAUUUGGGUGAUUUCACCAAGCAUCCCGGGUUUGGCGCCCUGUUUCAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1750
3'  GCUCAACGGCCUAGUUCCAGGCCUAAACCCACUAAAGUGGUUCGUAGGGCCCAAACCGCGGGACAAAGUG
1    Arg Val Ala Gly Ser Arg Ser Gly Phe Gly  .  Phe His Gln Ala Ser Arg Val Trp Arg Pro Val Ser
2     Glu Leu Pro Asp Gln Gly Pro Asp Leu Gly Asp Phe Thr Lys His Pro Gly Phe Gly Ala Leu Phe His
3   Pro Ser Cys Arg Ile Lys Val Arg Ile Trp Val Ile Ser Pro Ser Ile Pro Gly Leu Ala Pro Cys Phe Thr
o
5'  UGGGGCUUCUCCAAGGUGCGCGAUGAGAUGGGGGUUGUCUUCCCCUACACGCAGUAUGCGCGUGUGUUGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1820
3'  ACCCCGAAGAGGUUCCACGCGCUACUCUACCCCCAACAGAAGGGGAUGUGCGUCAUACGCGCACACAACG
1   Leu Gly Leu Leu Gln Gly Ala Arg  .  Asp Gly Gly Cys Leu Pro Leu His Ala Val Cys Ala Cys Val Ala
2    Trp Gly Phe Ser Lys Val Arg Asp Glu Met Gly Val Val Phe Pro Tyr Thr Gln Tyr Ala Arg Val Leu
3     Gly Ala Ser Pro Arg Cys Ala Met Arg Trp Gly Leu Ser Ser Pro Thr Arg Ser Met Arg Val Cys Cys
o
5'  CGCUCGCGGCAAUGCCGAUACUUGCGAGUAGGAGCUGGGUGCGCAUGGUCGCCUGGCUGCUGAAAGCGUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1890
3'  GCGAGCGCCGUUACGGCUAUGAACGCUCAUCCUCGACCCACGCGUACCAGCGGACCGACGACUUUCGCAU
1     Ala Arg Gly Asn Ala Asp Thr Cys Glu  .  Glu Leu Gly Ala His Gly Arg Leu Ala Ala Glu Ser Val
2   Pro Leu Ala Ala Met Pro Ile Leu Ala Ser Arg Ser Trp Val Arg Met Val Ala Trp Leu Leu Lys Ala Tyr
3    Arg Ser Arg Gln Cys Arg Tyr Leu Arg Val Gly Ala Gly Cys Ala Trp Ser Pro Gly Cys  .  Lys Arg
o
5'  UGAUCGCUUGAUGGGGAAGCCUCUCCAUACGUGGGAGCUACAAGGCUUGUUGUGGAGCCUGUCUCACGUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1960
3'  ACUAGCGAACUACCCCUUCGGAGAGGUAUGCACCCUCGAUGUUCCGAACAACACCUCGGACAGAGUGCAU
1    .  Ser Leu Asp Gly Glu Ala Ser Pro Tyr Val Gly Ala Thr Arg Leu Val Val Glu Pro Val Ser Arg
2     Asp Arg Leu Met Gly Lys Pro Leu His Thr Trp Glu Leu Gln Gly Leu Leu Trp Ser Leu Ser His Val
3   Met Ile Ala  .  Trp Gly Ser Leu Ser Ile Arg Gly Ser Tyr Lys Ala Cys Cys Gly Ala Cys Leu Thr  .
o
5'  GGCACGCAGGAGGAGCGCGAAGUUUACUUCUGGGGCAAGCUGCAGGGGGCAGCUGAUCUACUCCUGGCCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2030
3'  CCGUGCGUCCUCCUCGCGCUUCAAAUGAAGACCCCGUUCGACGUCCCCCGUCGACUAGAUGAGGACCGGA
1   Arg His Ala Gly Gly Ala Arg Ser Leu Leu Leu Gly Gln Ala Ala Gly Gly Ser  .  Ser Thr Pro Gly Leu
2    Gly Thr Gln Glu Glu Arg Glu Val Tyr Phe Trp Gly Lys Leu Gln Gly Ala Ala Asp Leu Leu Leu Ala
3     Ala Arg Arg Arg Ser Ala Lys Phe Thr Ser Gly Ala Ser Cys Arg Gly Gln Leu Ile Tyr Ser Trp Pro
o
5'  CGAGGCGUCGAGGUCAGCGCACGCAGGGUGCCGCUUCUGCGGACCUGAAUCGCAUUCUGAGCGCGGCUGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2100
3'  GCUCCGCAGCUCCAGUCGCGUGCGUCCCACGGCGAAGACGCCUGGACUUAGCGUAAGACUCGCGCCGACU
1     Glu Ala Ser Arg Ser Ala His Ala Gly Cys Arg Phe Cys Gly Pro Glu Ser His Ser Glu Arg Gly  .
2   Ser Arg Arg Arg Gly Gln Arg Thr Gln Gly Ala Ala Ser Ala Asp Leu Asn Arg Ile Leu Ser Ala Ala Glu
3    Arg Gly Val Glu Val Ser Ala Arg Arg Val Pro Leu Leu Arg Thr  .  Ile Ala Phe  .  Ala Arg Leu
o
5'  AGCUUUGGGGAGUCGGAAGCUGAAGUUGUCUCCGGCGGAACGUAUGGUGUUCGGCGCCGCUUAUGGGGUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2170
3'  UCGAAACCCCUCAGCCUUCGACUUCAACAGAGGCCGCCUUGCAUACCACAAGCCGCGGCGAAUACCCCAG
1    Ser Phe Gly Glu Ser Glu Ala Glu Val Val Ser Gly Gly Thr Tyr Gly Val Arg Arg Arg Leu Trp Gly
2     Ala Leu Gly Ser Arg Lys Leu Lys Leu Ser Pro Ala Glu Arg Met Val Phe Gly Ala Ala Tyr Gly Val
3   Lys Leu Trp Gly Val Gly Ser  .  Ser Cys Leu Arg Arg Asn Val Trp Cys Ser Ala Pro Leu Met Gly Ser
o
5'  GUCGGACGUGGCCAAUUCCGCCAUGACUCUGAUUACAGCCGCAAUCCUCCCGGCGGGUUGCGGUACGCGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2240
3'  CAGCCUGCACCGGUUAAGGCGGUACUGAGACUAAUGUCGGCGUUAGGAGGGCCGCCCAACGCCAUGCGCC
1   Arg Arg Thr Trp Pro Ile Pro Pro  .  Leu  .  Leu Gln Pro Gln Ser Ser Arg Arg Val Ala Val Arg Gly
2    Val Gly Arg Gly Gln Phe Arg His Asp Ser Asp Tyr Ser Arg Asn Pro Pro Gly Gly Leu Arg Tyr Ala
3     Ser Asp Val Ala Asn Ser Ala Met Thr Leu Ile Thr Ala Ala Ile Leu Pro Ala Gly Cys Gly Thr Arg
o
```

```
5'  AACGUCCCACGCUAGGGCGUAUGGCGUCCACUUCCGAUGGUACAAGCCAUCAUCGUACGCUGUCCAGCAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2310
3'  UUGCAGGGUGCGAUCCCGCAUACCGCAGGUGAAGGCUACCAUGUUCGGUAGUAGCAUGCGACAGGUCGUC
1     Thr Ser His Ala Arg Ala Tyr Gly Val His Phe Arg Trp Tyr Lys Pro Ser Ser Tyr Ala Val Gln Gln
2   Glu Arg Pro Thr Leu Gly Arg Met Ala Ser Thr Ser Asp Gly Thr Ser His His Arg Thr Leu Ser Ser Ser
3    Asn Val Pro Arg . Gly Val Trp Arg Pro Leu Pro Met Val Gln Ala Ile Ile Val Arg Cys Pro Ala
o
5'  UACAAAUGGAAGCCGCGGUGUUUACACUCCGCGCAGAAUCGGCUAGCUUAAAUUAAGUCGUUCCAUGAGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2380
3'  AUGUUUACCUUCGGCGCCACAAAUGUGAGGCGCGUCUUAGCCGAUCGAAUUUAAUUCAGCAAGGUACUCA
1    Tyr Lys Trp Lys Pro Arg Cys Leu His Ser Ala Gln Asn Arg Leu Ala . Ile Lys Ser Phe His Glu
2     Thr Asn Gly Ser Arg Gly Val Tyr Thr Pro Arg Arg Ile Gly . Leu Lys Leu Ser Arg Ser Met Ser
3   Val Gln Met Glu Ala Ala Val Phe Thr Leu Arg Ala Glu Ser Ala Ser Leu Asn . Val Val Pro . Val
o
5'  CAUCCGUUCUUACGAGCUGGACGUGUUUUUAGAAAGAAGCUACAUAGCGGCGGGGCCUCCGUUUUCUCUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2450
3'  GUAGGCAAGAAUGCUCGACCUGCACAAAAAUCUUUCUUCGAUGUAUCGCCGCCCCGGAGGCAAAAGAGAA
1   Ser Ser Val Leu Thr Ser Trp Thr Cys Phe . Lys Glu Ala Thr . Arg Arg Gly Leu Arg Phe Leu Phe
2    His Pro Phe Leu Arg Ala Gly Arg Val Phe Arg Lys Lys Leu His Ser Gly Gly Ala Ser Val Phe Ser
3     Ile Arg Ser Tyr Glu Leu Asp Val Phe Leu Glu Arg Ser Tyr Ile Ala Ala Gly Pro Pro Phe Ser Leu
o
5'  UAGGCCCUUGAGUCAGUCGGAUCUCCUCGGCGGUGGGGUGGUCUGGUGUAGUGGAACAAAUAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|+++          2513
3'  AUCCGGGAACUCAGUCAGCCUAGAGGAGCCGCCACCCCACCAGACCACAUCACCUUGUUUAUA
1     Arg Pro Leu Ser Gln Ser Asp Leu Leu Gly Gly Gly Val Val Trp Cys Ser Gly Thr Asn Ile
2   Leu Gly Pro . Val Ser Arg Ile Ser Ser Ala Val Gly Trp Ser Gly Val Val Glu Gln Ile
3    . Ala Leu Glu Ser Val Gly Ser Pro Arg Arg Trp Gly Gly Leu Val . Trp Asn Lys Tyr
o
```

2.5.4 AaV1 dsRNA2 del-2 sequence

2048 nt excluding poly (A) tail

```
5'   CACAAAGCAAUCGUAUCGCCAGAUACACAUAAGUCUUUGACCUUGGUUCGUUAUGUCUUCUGCUGUUUCU
o                                                                               70
3'   GUGUUUCGUUAGCAUAGCGGUCUAUGUGUAUUCAGAAACUGGAACCAAGCAAUACAGAAGACGACAAAGA
1     His Lys Ala Ile Val Ser Pro Asp Thr His Lys Ser Leu Thr Leu Val Arg Tyr Val Phe Cys Cys Phe
2       Thr Lys Gln Ser Tyr Arg Gln Ile His Ile Ser Leu . Pro Trp Phe Val Met Ser Ser Ala Val Ser
3    Ser Gln Ser Asn Arg Ile Ala Arg Tyr Thr . Val Phe Asp Leu Gly Ser Leu Cys Leu Leu Leu Phe Leu
o
5'   GAUUUCUUUGUCUGCGGGGGGUUUAGUGAGGGGUGGUUGGAGGUUAUUACCAGUGGUGCUGAGAAGUUGC
o                                                                               140
3'   CUAAAGAAACAGACGCCCCCCAAAUCACUCCCCACCAACCUCCAAUAAUGGUCACCACGACUCUUCAACG
1    . Phe Leu Cys Leu Arg Gly Val . . Gly Val Val Gly Gly Tyr Tyr Gln Trp Cys . Glu Val Ala
2     Asp Phe Phe Val Cys Gly Gly Phe Ser Glu Gly Trp Leu Glu Val Ile Thr Ser Gly Ala Glu Lys Leu
3       Ile Ser Leu Ser Ala Gly Gly Leu Val Arg Gly Gly Trp Arg Leu Leu Pro Val Val Leu Arg Ser Cys
o
5'   UCGCGGACUUCUUUCGUAGCACGGCCGAGCCCGACACGGAGGUCGUGAUGUUGUCGCAGGACGAUGAAGA
o                                                                               210
3'   AGCGCCUGAAGAAAGCAUCGUGCCGGCUCGGGCUGUGCCUCCAGCACUACAACAGCGUCCUGCUACUUCU
1       Arg Gly Leu Leu Ser . His Gly Arg Ala Arg His Gly Gly Arg Asp Val Val Ala Gly Arg . Arg
2    Leu Ala Asp Phe Phe Arg Ser Thr Ala Glu Pro Asp Thr Glu Val Val Met Leu Ser Gln Asp Asp Glu Glu
3     Ser Arg Thr Ser Phe Val Ala Arg Pro Ser Pro Thr Arg Arg Ser . Cys Cys Arg Arg Thr Met Lys
o
5'   GGCUGUUGUGGCCUCCUGCGCAGAUCACGAGGUCGGCUGGGGUAACCUGGCAACACUUUCACCCUGUGUG
o                                                                               280
3'   CCGACAACACCGGAGGACGCGUCUAGUGCUCCAGCCGACCCCAUUGGACCGUUGUGAAAGUGGGACACAC
1     Gly Cys Cys Gly Leu Leu Arg Arg Ser Arg Gly Arg Leu Gly . Pro Gly Asn Thr Phe Thr Leu Cys
2       Ala Val Val Ala Ser Cys Ala Asp His Glu Val Gly Trp Gly Asn Leu Ala Thr Leu Ser Pro Cys Val
3    Arg Leu Leu Trp Pro Pro Ala Gln Ile Thr Arg Ser Ala Gly Val Thr Trp Gln His Phe His Pro Val Trp
o
5'   GCAGUGGAUGGCGCCGAUGAUGAUGUUGUUUCUGUUGUGUCCGGGCGGGGUGACGACGUAGGGUUUGAGG
o                                                                               350
3'   CGUCACCUACCGCGGCUACUACUACAACAAAGACAACACAGGCCCGCCCCACUGCUGCAUCCCAAACUCC
1    Gly Ser Gly Trp Arg Arg . . Cys Cys Phe Cys Cys Val Arg Ala Gly . Arg Arg Arg Val . Gly
2     Ala Val Asp Gly Ala Asp Asp Asp Val Val Ser Val Val Ser Gly Arg Gly Asp Asp Val Gly Phe Glu
3       Gln Trp Met Ala Pro Met Met Met Leu Phe Leu Leu Cys Pro Gly Gly Val Thr Thr . Gly Leu Arg
o
5'   UCGAUACGACUCAAGGUGACGGGCUGGGGGAGCACCUUGACGACUUCCUCUUUCUGAGGGGCGGAACUGC
o                                                                               420
3'   AGCUAUGCUGAGUUCCACUGCCCGACCCCCUCGUGGAACUGCUGAAGGAGAAAGACUCCCCGCCUUGACG
1       Arg Tyr Asp Ser Arg . Arg Ala Gly Gly Ala Pro . Arg Leu Pro Leu Ser Glu Gly Arg Asn Cys
2    Val Asp Thr Thr Gln Gly Asp Gly Leu Gly Glu His Leu Asp Asp Phe Leu Phe Leu Arg Gly Gly Thr Ala
3     Ser Ile Arg Leu Lys Val Thr Gly Trp Gly Ser Thr Leu Thr Thr Ser Ser Phe . Gly Ala Glu Leu
o
5'   GGUGGAGAAGAGUAGAUCGGUGGCGGGCAACCUUCUUGCCUUGUUGCGACGCGUGGGUGGCACUGGUGUU
o                                                                               490
3'   CCACCUCUUCUCAUCUAGCCACCGCCCGUUGGAAGAACGGAACAACGCUGCGCACCCACCGUGACCACAA
1     Gly Gly Glu Glu . Ile Gly Gly Gly Gln Pro Ser Cys Leu Val Ala Thr Arg Gly Trp His Trp Cys
2       Val Glu Lys Ser Arg Ser Val Ala Gly Asn Leu Leu Ala Leu Leu Arg Arg Val Gly Gly Thr Gly Val
3    Arg Trp Arg Arg Val Asp Arg Trp Arg Ala Thr Phe Leu Pro Cys Cys Asp Ala Trp Val Ala Leu Val Leu
o
5'   GAUGCGGAGGGAGUCGUUGUGAUCUUCGGGGACUCGCCGGGCGUUGUGGCCCGCGAGCUAGCAAGCGUGG
o                                                                               560
3'   CUACGCCUCCCUCAGCAACACUAGAAGCCCCUGAGCGGCCCGCAACACCGGGCGCUCGAUCGUUCGCACC
1    . Cys Gly Gly Ser Arg Cys Asp Leu Arg Gly Leu Ala Gly Arg Cys Gly Pro Arg Ala Ser Lys Arg Gly
2     Asp Ala Glu Gly Val Val Val Ile Phe Gly Asp Ser Pro Gly Val Val Ala Arg Glu Leu Ala Ser Val
3       Met Arg Arg Glu Ser Leu . Ser Ser Gly Thr Arg Arg Ala Leu Trp Pro Ala Ser . Gln Ala Trp
o
```

```
5'   GGUAUCGCGUGCUGGGUGUCGACAAGGAUCCGGCACACGCGGCACCCCCUGGCUGGCUCGACAAGUAUAG
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   630
3'   CCAUAGCGCACGACCCACAGCUGUUCCUAGGCCGUGUGCGCCGUGGGGGACCGACCGAGCUGUUCAUAUC
1      Val Ser Arg Ala Gly Cys Arg Gln Gly Ser Gly Thr Arg Gly Thr Pro Trp Leu Ala Arg Gln Val  .
2    Gly Tyr Arg Val Leu Gly Val Asp Lys Asp Pro Ala His Ala Ala Pro Pro Gly Trp Leu Asp Lys Tyr Arg
3     Gly Ile Ala Cys Trp Val Ser Thr Arg Ile Arg His Thr Arg His Pro Leu Ala Gly Ser Thr Ser Ile
o
5'   GACUGUCGUGGCUGAGGUCACUGACGGGCUGACCAGUUCUGAAGUGGAGGGGUGGCUGGCGCAGGUGGGU
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   700
3'   CUGACAGCACCGACUCCAGUGACUGCCCGACUGGUCAAGACUUCACCUCCCCACCGACCGCGUCCACCCA
1     Asp Cys Arg Gly  .  Gly His  .  Arg Ala Asp Gln Phe  .  Ser Gly Gly Val Ala Gly Ala Gly Gly
2      Thr Val Val Ala Glu Val Thr Asp Gly Leu Thr Ser Ser Glu Val Glu Gly Trp Leu Ala Gln Val Gly
3    Gly Leu Ser Trp Leu Arg Ser Leu Thr Gly  .  Pro Val Leu Lys Trp Arg Gly Gly Trp Arg Arg Trp Val
o
5'   UGGGGGGGUAAGCGUGUUUUGGCUGCCCUGAUGGACAUAGAUCAAGGUUCGAAACGAUCUUCCGUCAGUG
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   770
3'   ACCCCCCCAUUCGCACAAAACCGACGGGACUACCUGUAUCUAGUUCCAAGCUUUGCUAGAAGGCAGUCAC
1    Leu Gly Gly  .  Ala Cys Phe Gly Cys Pro Asp Gly His Arg Ser Arg Phe Glu Thr Ile Phe Arg Gln  .
2     Trp Gly Gly Lys Arg Val Leu Ala Ala Leu Met Asp Ile Asp Gln Gly Ser Lys Arg Ser Ser Val Ser
3      Gly Gly Val Ser Val Phe Trp Leu Pro  .  Trp Thr  .  Ile Lys Val Arg Asn Asp Leu Pro Ser Val
o
5'   AUACGGCGCUUAACCGCGAGCUCUCGGCAAACCUGCUCGAUUACGGCCGGCGACGGGGGCUAGGUUACGA
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   840
3'   UAUGCCGCGAAUUGGCGCUCGAGAGCCGUUUGGACGAGCUAAUGCCGGCCGCUGCCCCCGAUCCAAUGCU
1      Tyr Gly Ala  .  Pro Arg Ala Leu Gly Lys Pro Ala Arg Leu Arg Pro Ala Thr Gly Ala Arg Leu Arg
2    Asp Thr Ala Leu Asn Arg Glu Leu Ser Ala Asn Leu Leu Asp Tyr Gly Arg Arg Arg Gly Leu Gly Tyr Glu
3     Ile Arg Arg Leu Thr Ala Ser Ser Arg Gln Thr Cys Ser Ile Thr Ala Gly Asp Gly Gly  .  Val Thr
o
5'   GGGUGGUGGUAAAGGGGAUGUGUCACCGGAGGGGGGCGAGCGGAAGGGGUUUACGGGCAAAUUCCGUGUU
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   910
3'   CCCACCACCAUUUCCCCUACACAGUGGCCUCCCCCCGCUCGCCUUCCCCAAAUGCCCGUUUAAGGCACAA
1     Gly Trp Trp  .  Arg Gly Cys Val Thr Gly Gly Gly Arg Ala Glu Gly Val Tyr Gly Gln Ile Pro Cys
2      Gly Gly Gly Lys Gly Asp Val Ser Pro Glu Gly Gly Glu Arg Lys Gly Phe Thr Gly Lys Phe Arg Val
3    Arg Val Val Val Lys Gly Met Cys His Arg Arg Gly Ala Ser Gly Arg Gly Leu Arg Ala Asn Ser Val Phe
o
5'   UCCGGGGCGACUGGUUACUUGUCGCUGCUGGGCCUGAGGAGUGGUGUUGAGGAGCUGCGACUGGCGCAUA
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   980
3'   AGGCCCCGCUGACCAAUGAACAGCGACGACCCGGACUCCUCACCACAACUCCUCGACGCUGACCGCGUAU
1    Phe Arg Gly Asp Trp Leu Leu Val Ala Ala Gly Pro Glu Glu Trp Cys  .  Gly Ala Ala Thr Gly Ala Tyr
2     Ser Gly Ala Thr Gly Tyr Leu Ser Leu Leu Gly Leu Arg Ser Gly Val Glu Glu Leu Arg Leu Ala His
3      Pro Gly Arg Leu Val Thr Cys Arg Cys Trp Ala  .  Gly Val Val Leu Arg Ser Cys Asp Trp Arg Ile
o
5'   CUCCGUUUAUGCGGGCUGGGAAGCGUGCGCGGGCUUUUGUUGAGCUGCUUGGCCAACUCAAGACCGUAAU
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1050
3'   GAGGCAAAUACGCCCGACCCUUCGCACGCGCCCGAAAACAACUCGACGAACCGGUUGAGUUCUGGCAUUA
1      Ser Val Tyr Ala Gly Trp Glu Ala Cys Ala Gly Phe Cys  .  Ala Ala Trp Pro Thr Gln Asp Arg Asn
2    Thr Pro Phe Met Arg Ala Gly Lys Arg Ala Arg Ala Phe Val Glu Leu Leu Gly Gln Leu Lys Thr Val Met
3     Leu Arg Leu Cys Gly Leu Gly Ser Val Arg Gly Leu Leu Leu Ser Cys Leu Ala Asn Ser Arg Pro  .
o
5'   GGGGAAGGUUCGGCCUGAACGUGUGGCCGAGCUGACCGCCAAAUUCCAGCCUGCGGACAGAGUUUUGCUG
o    +++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1120
3'   CCCCUUCCAAGCCGGACUUGCACACCGGCUCGACUGGCGGUUUAAGGUCGGACGCCUGUCUCAAAACGAC
1     Gly Glu Gly Ser Ala  .  Thr Cys Gly Arg Ala Asp Arg Gln Ile Pro Ala Cys Gly Gln Ser Phe Ala
2      Gly Lys Val Arg Pro Glu Arg Val Ala Glu Leu Thr Ala Lys Phe Gln Pro Ala Asp Arg Val Leu Leu
3    Trp Gly Arg Phe Gly Leu Asn Val Trp Pro Ser  .  Pro Pro Asn Ser Ser Leu Arg Thr Glu Phe Cys Cys
o
```

```
5'  CGGCACGUGUCGCAAGCGACGUUGAGUGUGUCACCCACACUGCUGCAUACGAUGCAAUCCCGGUACGACG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1190
3'  GCCGUGCACAGCGUUCGCUGCAACUCACACAGUGGGUGUGACGACGUAUGCUACGUUAGGGCCAUGCUGC
1   Ala  Ala  Arg  Val  Ala  Ser  Asp  Val  Glu  Cys  Val  Thr  His  Thr  Ala  Ala  Tyr  Asp  Ala  Ile  Pro  Val  Arg  Arg
2    Arg  His  Val  Ser  Gln  Ala  Thr  Leu  Ser  Val  Ser  Pro  Thr  Leu  Leu  His  Thr  Met  Gln  Ser  Arg  Tyr  Asp
3     Gly  Thr  Cys  Arg  Lys  Arg  Arg  .  Val  Cys  His  Pro  His  Cys  Cys  Ile  Arg  Cys  Asn  Pro  Gly  Thr  Thr
o
5'  GAUUGCUGAAGUAUUGUAGCGCAGCCGAGUUGCCGGAUCAAGGUCCGGAUUUGGGUGAUUUCACCAAGCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1260
3'  CUAACGACUUCAUAACAUCGCGUCGGCUCAACGGCCUAGUUCCAGGCCUAAACCCACUAAAGUGGUUCGU
1     Ile  Ala  Glu  Val  Leu  .  Arg  Ser  Arg  Val  Ala  Gly  Ser  Arg  Ser  Gly  Phe  Gly  .  Phe  His  Gln  Ala
2   Gly  Leu  Leu  Lys  Tyr  Cys  Ser  Ala  Ala  Glu  Leu  Pro  Asp  Gln  Gly  Pro  Asp  Leu  Gly  Asp  Phe  Thr  Lys  His
3    Asp  Cys  .  Ser  Ile  Val  Ala  Gln  Pro  Ser  Cys  Arg  Ile  Lys  Val  Arg  Ile  Trp  Val  Ile  Ser  Pro  Ser
o
5'  UCCCGGGUUUGGCGCCCUGUUUCACUGGGGCUUCUCCAAGGUGCGCGAUGAGAUGGGGGUUGUCUUCCCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1330
3'  AGGGCCCAAACCGCGGGACAAAGUGACCCCGAAGAGGUUCCACGCGCUACUCUACCCCCAACAGAAGGGG
1    Ser  Arg  Val  Trp  Arg  Pro  Val  Ser  Leu  Gly  Leu  Leu  Gln  Gly  Ala  Arg  .  Asp  Gly  Gly  Cys  Leu  Pro
2     Pro  Gly  Phe  Gly  Ala  Leu  Phe  His  Trp  Gly  Phe  Ser  Lys  Val  Arg  Asp  Glu  Met  Gly  Val  Val  Phe  Pro
3   Ile  Pro  Gly  Leu  Ala  Pro  Cys  Phe  Thr  Gly  Ala  Ser  Pro  Arg  Cys  Ala  Met  Arg  Trp  Gly  Leu  Ser  Ser  Pro
o
5'  UACACGCAGUAUGCGCGUGUGUUGCCGCUCGCGGCAAUGCCGAUACUUGCGAGUAGGAGCUGGGUGCGCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1400
3'  AUGUGCGUCAUACGCGCACACAACGGCGAGCGCCGUUACGGCUAUGAACGCUCAUCCUCGACCCACGCGU
1   Leu  His  Ala  Val  Cys  Ala  Cys  Val  Ala  Ala  Arg  Gly  Asn  Ala  Asp  Thr  Cys  Glu  .  Glu  Leu  Gly  Ala  His
2    Tyr  Thr  Gln  Tyr  Ala  Arg  Val  Leu  Pro  Leu  Ala  Ala  Met  Pro  Ile  Leu  Ala  Ser  Arg  Ser  Trp  Val  Arg
3     Thr  Arg  Ser  Met  Arg  Val  Cys  Cys  Arg  Ser  Arg  Gln  Cys  Arg  Tyr  Leu  Arg  Val  Gly  Ala  Gly  Cys  Ala
o
5'  UGGUCGCCUGGCUGCUGAAAGCGUAUGAUCGCUUGAUGGGGAAGCCUCUCCAUACGUGGGAGCUACAAGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1470
3'  ACCAGCGGACCGACGACUUUCGCAUACUAGCGAACUACCCCUUCGGAGAGGUAUGCACCCUCGAUGUUCC
1     Gly  Arg  Leu  Ala  Ala  Glu  Ser  Val  .  Ser  Leu  Asp  Gly  Glu  Ala  Ser  Pro  Tyr  Val  Gly  Ala  Thr  Arg
2   Met  Val  Ala  Trp  Leu  Leu  Lys  Ala  Tyr  Asp  Arg  Leu  Met  Gly  Lys  Pro  Leu  His  Thr  Trp  Glu  Leu  Gln  Gly
3    Trp  Ser  Pro  Gly  Cys  .  Lys  Arg  Met  Ile  Ala  .  Trp  Gly  Ser  Leu  Ser  Ile  Arg  Gly  Ser  Tyr  Lys
o
5'  CUUGUUGUGGAGCCUGUCUCACGUAGGCACGCAGGAGGAGCGCGAAGUUUACUUCUGGGGCAAGCUGCAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1540
3'  GAACAACACCUCGGACAGAGUGCAUCCGUGCGUCCUCCUCGCGCUUCAAAUGAAGACCCCGUUCGACGUC
1    Leu  Val  Val  Glu  Pro  Val  Ser  Arg  Arg  His  Ala  Gly  Gly  Ala  Arg  Ser  Leu  Leu  Leu  Gly  Gln  Ala  Ala
2     Leu  Leu  Trp  Ser  Leu  Ser  His  Val  Gly  Thr  Gln  Glu  Glu  Arg  Glu  Val  Tyr  Phe  Trp  Gly  Lys  Leu  Gln
3   Ala  Cys  Cys  Gly  Ala  Cys  Leu  Thr  .  Ala  Arg  Arg  Arg  Ser  Ala  Lys  Phe  Thr  Ser  Gly  Ala  Ser  Cys  Arg
o
5'  GGGGCAGCUGAUCUACUCCUGGCCUCGAGGCGUCGAGGUCAGCGCACGCAGGGUGCCGCUUCUGCGGACC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1610
3'  CCCCGUCGACUAGAUGAGGACCGGAGCUCCGCAGCUCCAGUCGCGUGCGUCCCACGGCGAAGACGCCUGG
1   Gly  Gly  Ser  .  Ser  Thr  Pro  Gly  Leu  Glu  Ala  Ser  Arg  Ser  Ala  His  Ala  Gly  Cys  Arg  Phe  Cys  Gly  Pro
2    Gly  Ala  Ala  Asp  Leu  Leu  Leu  Ala  Ser  Arg  Arg  Arg  Gly  Gln  Arg  Thr  Gln  Gly  Ala  Ala  Ser  Ala  Asp
3     Gly  Gln  Leu  Ile  Tyr  Ser  Trp  Pro  Arg  Gly  Val  Glu  Val  Ser  Ala  Arg  Arg  Val  Pro  Leu  Leu  Arg  Thr
o
5'  UGAAUCGCAUUCUGAGCGCGGCUGAAGCUUUGGGGAGUCGGAAGCUGAAGUUGUCUCCGGCGGAACGUAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1680
3'  ACUUAGCGUAAGACUCGCGCCGACUUCGAAACCCCUCAGCCUUCGACUUCAACAGAGGCCGCCUUGCAUA
1     Glu  Ser  His  Ser  Glu  Arg  Gly  .  Ser  Phe  Gly  Glu  Ser  Glu  Ala  Glu  Val  Val  Ser  Gly  Gly  Thr  Tyr
2   Leu  Asn  Arg  Ile  Leu  Ser  Ala  Ala  Glu  Ala  Leu  Gly  Ser  Arg  Lys  Leu  Lys  Leu  Ser  Pro  Ala  Glu  Arg  Met
3    .  Ile  Ala  Phe  .  Ala  Arg  Leu  Lys  Leu  Trp  Gly  Val  Gly  Ser  .  Ser  Cys  Leu  Arg  Arg  Asn  Val
o
```

```
5'  GGUGUUCGGCGCCGCUUAUGGGGUCGUCGGACGUGGCCAAUUCCGCCAUGACUCUGAUUACAGCCGCAAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1750
3'  CCACAAGCCGCGGCGAAUACCCCAGCAGCCUGCACCGGUUAAGGCGGUACUGAGACUAAUGUCGGCGUUA
1    Gly Val Arg Arg Arg Leu Trp Gly Arg Arg Thr Trp Pro Ile Pro Pro . Leu . Leu Gln Pro Gln
2     Val Phe Gly Ala Ala Tyr Gly Val Val Gly Arg Gly Gln Phe Arg His Asp Ser Asp Tyr Ser Arg Asn
3   Trp Cys Ser Ala Pro Leu Met Gly Ser Ser Asp Val Ala Asn Ser Ala Met Thr Leu Ile Thr Ala Ala Ile
o
5'  CCUCCCGGCGGGUUGCGGUACGCGGAACGUCCCACGCUAGGGCGUAUGGCGUCCACUUCCGAUGGUACAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1820
3'  GGAGGGCCGCCCAACGCCAUGCGCCUUGCAGGGUGCGAUCCCGCAUACCGCAGGUGAAGGCUACCAUGUU
1   Ser Ser Arg Arg Val Ala Val Arg Gly Thr Ser His Ala Arg Ala Tyr Gly Val His Phe Arg Trp Tyr Lys
2    Pro Pro Gly Gly Leu Arg Tyr Ala Glu Arg Pro Thr Leu Gly Arg Met Ala Ser Thr Ser Asp Gly Thr
3     Leu Pro Ala Gly Cys Gly Thr Arg Asn Val Pro Arg . Gly Val Trp Arg Pro Leu Pro Met Val Gln
o
5'  GCCAUCAUCGUACGCUGUCCAGCAGUACAAAUGGAAGCCGCGGUGUUUACACUCCGCGCAGAAUCGGCUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1890
3'  CGGUAGUAGCAUGCGACAGGUCGUCAUGUUUACCUUCGGCGCCACAAAUGUGAGGCGCGUCUUAGCCGAU
1     Pro Ser Ser Tyr Ala Val Gln Gln Tyr Lys Trp Lys Pro Arg Cys Leu His Ser Ala Gln Asn Arg Leu
2   Ser His His Arg Thr Leu Ser Ser Ser Thr Asn Gly Ser Arg Gly Val Tyr Thr Pro Arg Arg Ile Gly .
3    Ala Ile Ile Val Arg Cys Pro Ala Val Gln Met Glu Ala Ala Val Phe Thr Leu Arg Ala Glu Ser Ala
o
5'  GCUUAAAUUAAGUCGUUCCAUGAGUCAUCCGUUCUUACGAGCUGGACGUGUUUUUAGAAAGAAGCUACAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1960
3'  CGAAUUUAAUUCAGCAAGGUACUCAGUAGGCAAGAAUGCUCGACCUGCACAAAAAUCUUUCUUCGAUGUA
1    Ala . Ile Lys Ser Phe His Glu Ser Ser Val Leu Thr Ser Trp Thr Cys Phe . Lys Glu Ala Thr
2     Leu Lys Leu Ser Arg Ser Met Ser His Pro Phe Leu Arg Ala Gly Arg Val Phe Arg Lys Lys Leu His
3   Ser Leu Asn . Val Val Pro . Val Ile Arg Ser Tyr Glu Leu Asp Val Phe Leu Glu Arg Ser Tyr Ile
o
5'  AGCGGCGGGGCCUCCGUUUUCUCUUUAGGCCCUUGAGUCAGUCGGAUCUCCUCGGCGGUGGGGUGGUCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2030
3'  UCGCCGCCCCGGAGGCAAAAGAGAAAUCCGGGAACUCAGUCAGCCUAGAGGAGCCGCCACCCCACCAGAC
1   . Arg Arg Gly Leu Arg Phe Leu Phe Arg Pro Leu Ser Gln Ser Asp Leu Leu Gly Gly Gly Val Val Trp
2    Ser Gly Gly Ala Ser Val Phe Ser Leu Gly Pro . Val Ser Arg Ile Ser Ser Ala Val Gly Trp Ser
3     Ala Ala Gly Pro Pro Phe Ser Leu . Ala Leu Glu Ser Val Gly Ser Pro Arg Arg Trp Gly Gly Leu
o
5'  GUGUAGUGGAACAAAUAU
o   ++++|++++|++++|+++                                                       2048
3'  CACAUCACCUUGUUUAUA
1     Cys Ser Gly Thr Asn Ile
2   Gly Val Val Glu Gln Ile
3    Val . Trp Asn Lys Tyr
o
```

2.5.5 AaV1 dsRNA3 sequence (NC_010990)

2540 nt excluding poly (A) tail

```
5'   CAUAAAGCUAUCGUAUCUCGAGAUACACAUAAGUCUUUAUCCUGGCGAGAGAUGGCGACGUUUGGAAGUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   70
3'   GUAUUUCGAUAGCAUAGAGCUCUAUGUGUAUUCAGAAAUAGGACCGCUCUCUACCGCUGCAAACCUUCAC
1     His Lys Ala Ile Val Ser Arg Asp Thr His Lys Ser Leu Ser Trp Arg Glu Met Ala Thr Phe Gly Ser
2      Ile Lys Leu Ser Tyr Leu Glu Ile His Ile Ser Leu Tyr Pro Gly Glu Arg Trp Arg Arg Leu Glu Val
3    Pro . Ser Tyr Arg Ile Ser Arg Tyr Thr . Val Phe Ile Leu Ala Arg Asp Gly Asp Val Trp Lys Cys
o
5'   CUGAGGCGUAUCGUGCGGCGCAGCUUGCUGCGAUUGAUCUUUUGACACGUGGAGAUUGGGGUGAUUUCGG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   140
3'   GACUCCGCAUAGCACGCCGCGUCGAACGACGCUAACUAGAAAACUGUGCACCUCUAACCCCACUAAAGCC
1    Ala Glu Ala Tyr Arg Ala Ala Gln Leu Ala Ala Ile Asp Leu Leu Thr Arg Gly Asp Trp Gly Asp Phe Gly
2     Leu Arg Arg Ile Val Arg Arg Ser Leu Leu Arg Leu Ile Phe . His Val Glu Ile Gly Val Ile Ser
3      . Gly Val Ser Cys Gly Ala Ala Cys Cys Asp . Ser Phe Asp Thr Trp Arg Leu Gly . Phe Arg
o
5'   AGGUUUGGGUGAGACGAGUGAUGUUUUCGACGAGCAUGAGCGGUUCCCUGAGGAUCCGUUCGAUGCGCCC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   210
3'   UCCAAACCCACUCUGCUCACUACAAAAGCUGCUCGUACUCGCCAAGGGACUCCUAGGCAAGCUACGCGGG
1     Gly Leu Gly Glu Thr Ser Asp Val Phe Asp Glu His Glu Arg Phe Pro Glu Asp Pro Phe Asp Ala Pro
2    Glu Val Trp Val Arg Arg Val Met Phe Ser Thr Ser Met Ser Gly Ser Leu Arg Ile Arg Ser Met Arg Pro
3     Arg Phe Gly . Asp Glu . Cys Phe Arg Arg Ala . Ala Val Pro . Gly Ser Val Arg Cys Ala
o
5'   GAUCACAUUGAUCCGGGAAGUUCGUGGGCCUCGGUGGCCACGGGUUCGGCCGAUGACAUUUCGCGGGAGG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   280
3'   CUAGUGUAACUAGGCCCUUCAAGCACCCGGAGCCACCGGUGCCCAAGCCGGCUACUGUAAAGCGCCCUCC
1     Asp His Ile Asp Pro Gly Ser Ser Trp Ala Ser Val Ala Thr Gly Ser Ala Asp Asp Ile Ser Arg Glu
2      Ile Thr Leu Ile Arg Glu Val Arg Gly Pro Arg Trp Pro Arg Val Arg Pro Met Thr Phe Arg Gly Arg
3    Arg Ser His . Ser Gly Lys Phe Val Gly Leu Gly Gly His Gly Phe Gly Arg . His Phe Ala Gly Gly
o
5'   UGACGCCAACCGUUGUGGACUACGUUCCGCGACGGGAGCCUUACUCCGUGCGAGGUGUUGACAUCGUGGA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   350
3'   ACUGCGGUUGGCAACACCUGAUGCAAGGCGCUGCCCUCGGAAUGAGGCACGCUCCACAACUGUAGCACCU
1    Val Thr Pro Thr Val Val Asp Tyr Val Pro Arg Arg Glu Pro Tyr Ser Val Arg Gly Val Asp Ile Val Glu
2     . Arg Gln Pro Leu Trp Thr Thr Phe Arg Asp Gly Ser Leu Thr Pro Cys Glu Val Leu Thr Ser Trp
3      Asp Ala Asn Arg Cys Gly Leu Arg Ser Ala Thr Gly Ala Leu Leu Arg Ala Arg Cys . His Arg Gly
o
5'   GCCAGUGACUGCUUUUUGUUCCCCAGCGGACUUGGGGAUGGCCAACUUUCCGGGCAAUGUGUCGAAGGGG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   420
3'   CGGUCACUGACGAAAAACAAGGGGUCGCCUGAACCCCUACCGGUUGAAAGGCCCGUUACACAGCUUCCCC
1     Pro Val Thr Ala Phe Cys Ser Pro Ala Asp Leu Gly Met Ala Asn Phe Pro Gly Asn Val Ser Lys Gly
2    Ser Gln . Leu Leu Phe Val Pro Gln Arg Thr Trp Gly Trp Pro Thr Phe Arg Ala Met Cys Arg Arg Gly
3     Ala Ser Asp Cys Phe Leu Phe Pro Ser Gly Leu Gly Asp Gly Gln Leu Ser Gly Gln Cys Val Glu Gly
o
5'   UCGGCAACCACACUGCGGAGGGCCGGGUCUGCUGCCGCUAAUGCGGGUGCCAUUCUGGGGACGGAUAUGU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   490
3'   AGCCGUUGGUGUGACGCCUCCCGGCCCAGACGACGGCGAUUACGCCCACGGUAAGACCCCUGCCUAUACA
1     Ser Ala Thr Thr Leu Arg Arg Ala Gly Ser Ala Ala Ala Asn Ala Gly Ala Ile Leu Gly Thr Asp Met
2      Arg Gln Pro His Cys Gly Gly Pro Gly Leu Leu Pro Leu Met Arg Val Pro Phe Trp Gly Arg Ile Cys
3    Val Gly Asn His Thr Ala Glu Gly Arg Val Cys Cys Arg . Cys Gly Cys His Ser Gly Asp Gly Tyr Val
o
5'   UAAUGGCGCCCAAAGCUAGUGUGCAGGCCAUUAUGGCGCGGUUGUAUCAGCUCUGGAGAUGCUGCAGUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   560
3'   AUUACCGCGGGUUUCGAUCACACGUCCGGUAAUACCGCGCCCAACAUAGUCGAGACCUCUACGACGUCAG
1    Leu Met Ala Pro Lys Ala Ser Val Gln Ala Ile Met Ala Arg Val Val Ser Ala Leu Glu Met Leu Gln Ser
2     . Trp Arg Pro Lys Leu Val Cys Arg Pro Leu Trp Arg Gly Leu Tyr Gln Leu Trp Arg Cys Cys Ser
3      Asn Gly Ala Gln Ser . Cys Ala Gly His Tyr Gly Ala Gly Cys Ile Ser Ser Gly Asp Ala Ala Val
o
```

```
5'  GGGUUGGGACGUGGGUGGUCCGCCUGAUGUUGGGCUAGAUGUGAGGCACGCAGCGCGACGAUUUCGAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  630
3'  CCCAACCCUGCACCCACCAGGCGGACUACAACCCGAUCUACACUCCGUGCGUCGCGCUGCUAAAGCUA
1     Gly Trp Asp Val Gly Gly Pro Pro Asp Val Gly Leu Asp Val Arg His Ala Ala Arg Asp Asp Phe Asp
2   Arg Val Gly Thr Trp Val Val Arg Leu Met Leu Gly . Met . Gly Thr Gln Arg Ala Thr Ile Ser Met
3    Gly Leu Gly Arg Gly Trp Ser Ala . Cys Trp Ala Arg Cys Glu Ala Arg Ser Ala Arg Arg Phe Arg
o
5'  GACGCGAGUGCCCGCUAUCCGCAGCAUGCUUGGCUGUACAUUCCGUCAGAUUGGACGGAGCAGGAGGUGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  700
3'  CUGCGCUCACGGGCGAUAGGCGUCGUACGAACCGACAUGUAAGGCAGUCUAACCUGCCUCGUCCUCCACC
1    Asp Ala Ser Ala Arg Tyr Pro Gln His Ala Trp Leu Tyr Ile Pro Ser Asp Trp Thr Glu Gln Glu Val
2     Thr Arg Val Pro Ala Ile Arg Ser Met Leu Gly Cys Thr Phe Arg Gln Ile Gly Arg Ser Arg Arg Trp
3   . Arg Glu Cys Pro Leu Ser Ala Ala Cys Leu Ala Val His Ser Val Arg Leu Asp Gly Ala Gly Gly Gly
o
5'  CAGCGCUUGUGUCGCUCAUGGUGGAGGGGGGCCUGCUGCAUACAGAUGGGGGUAUGCGCGGGGCGAUCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  770
3'  GUCGCGAACACAGCGAGUACCACCUCCCCCCGGACGACGUAUGUCUACCCCCAUACGCGCCCCGCUAGG
1   Ala Ala Leu Val Ser Leu Met Val Glu Gly Gly Pro Ala Ala Tyr Arg Trp Gly Tyr Ala Arg Gly Asp Pro
2    Gln Arg Leu Cys Arg Ser Trp Trp Arg Gly Gly Leu Leu His Thr Asp Gly Gly Met Arg Gly Ala Ile
3     Ser Ala Cys Val Ala His Gly Gly Gly Gly Ala Cys Cys Ile Gln Met Gly Val Cys Ala Gly Arg Ser
o
5'  UGGUGGUGAUGAGGGUAACGGACAGGCGCGUGAGCGCGUUAUGCCUGCCGGCGCCGCUUGGCGCUGGCCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  840
3'  ACCACCACUACUCCCAUUGCCUGUCCGCGCACUCGCGCAAUACGGACGGCCGCGGCGAACCGCGACCGGG
1     Gly Gly Asp Glu Gly Asn Gly Gln Ala Arg Glu Arg Val Met Pro Ala Gly Ala Ala Trp Arg Trp Pro
2   Leu Val Val Met Arg Val Thr Asp Arg Arg Val Ser Ala Leu Cys Leu Pro Ala Pro Leu Gly Ala Gly Pro
3    Trp Trp . . Gly . Arg Thr Gly Ala . Ala Arg Tyr Ala Cys Arg Arg Arg Leu Ala Leu Ala
o
5'  GGCGGAUGGUCUAACUAUCUGCUGAUCGGUGAGCGUGAUCGCGGCUGGAACGUUGCUUUUGGGGGCGAUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  910
3'  CCGCCUACCAGAUUGAUAGACGACUAGCCACUCGCACUAGCGCCGACCUUGCAACGAAAACCCCCGCUAC
1    Gly Gly Trp Ser Asn Tyr Leu Leu Ile Gly Glu Arg Asp Arg Gly Trp Asn Val Ala Phe Gly Gly Asp
2     Ala Asp Gly Leu Thr Ile Cys . Ser Val Ser Val Ile Ala Ala Gly Thr Leu Leu Leu Gly Ala Met
3   Arg Arg Met Val . Leu Ser Ala Asp Arg . Ala . Ser Arg Leu Glu Arg Cys Phe Trp Gly Arg Cys
o
5'  CCCUAUCUGUUGCUUCUUUGUCGGCUGUUUUGCGGAGGAUGGUUGAAGCGUAUGGGCAGCGCAUCUAUCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  980
3'  GGGAUAGACAACGAAGAAACAGCCGACAAAACGCCUCCUACCAACUUCGCAUACCCGUCGCGUAGAUAGA
1   Ala Leu Ser Val Ala Ser Leu Ser Ala Val Leu Arg Arg Met Val Glu Ala Tyr Gly Gln Arg Ile Tyr Leu
2    Pro Tyr Leu Leu Leu Leu Cys Arg Leu Phe Cys Gly Gly Trp Leu Lys Arg Met Gly Ser Ala Ser Ile
3     Pro Ile Cys Cys Phe Phe Val Gly Cys Phe Ala Glu Asp Gly . Ser Val Trp Ala Ala His Leu Ser
o
5'  UGAUGCCUGCCGCGCAGCAGCCGUCGCCAAUCGGGCAUAUUGCCCGCCGUGUUAUCAAGCUGGCUCUAAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1050
3'  ACUACGGACGGCGCGUCGUCGGCAGCGGUUAGCCCGUAUAACGGGCGGCACAAUAGUUCGACCGAGAUUC
1     Asp Ala Cys Arg Ala Ala Ala Val Ala Asn Arg Ala Tyr Cys Pro Pro Cys Tyr Gln Ala Gly Ser Lys
2   Leu Met Pro Ala Ala Gln Gln Pro Ser Pro Ile Gly His Ile Ala Arg Arg Val Ile Lys Leu Ala Leu Arg
3    . Cys Leu Pro Arg Ser Ser Arg Arg Gln Ser Gly Ile Leu Pro Ala Val Leu Ser Ser Trp Leu .
o
5'  GAGAGGGAUAUAGGCACAGCCUUCACCUCGGAUCGCGUUGUGGUGCGUGAUGGGAACGGGGGUCAUAUGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1120
3'  CUCUCCCUAUAUCCGUGUCGGAAGUGGAGCCUAGCGCAACACCACGCACUACCCUUGCCCCCAGUAUACG
1    Glu Arg Asp Ile Gly Thr Ala Phe Thr Ser Asp Arg Val Val Val Arg Asp Gly Asn Gly Gly His Met
2     Arg Gly Ile . Ala Gln Pro Ser Pro Arg Ile Ala Leu Trp Cys Val Met Gly Thr Gly Val Ile Cys
3   Gly Glu Gly Tyr Arg His Ser Leu His Leu Gly Ser Arg Cys Gly Ala . Trp Glu Arg Gly Ser Tyr Ala
o
```

```
5'  GGCAGAUCGUGCCUGCGCGCAAUAUGGUGAAGGACGAGCCCGGGGUGCUUCCGCGCCCUGCGGAGGGCUG
o                                                                            1190
3'  CCGUCUAGCACGGACGCGCGUUAUACCACUUCCUGCUCGGGCCCCACGAAGGCGCGGGACGCCUCCCGAC
1   Arg Gln Ile Val Pro Ala Arg Asn Met Val Lys Asp Glu Pro Gly Val Leu Pro Arg Pro Ala Glu Gly Trp
2    Gly Arg Ser Cys Leu Arg Ala Ile Trp  .  Arg Thr Ser Pro Gly Cys Phe Arg Ala Leu Arg Arg Ala
3     Ala Asp Arg Ala Cys Ala Gln Tyr Gly Glu Gly Arg Ala Arg Gly Ala Ser Ala Pro Cys Gly Gly Leu
o
5'  GGAUCCUGUGGACGACGGGGGCGUGGUGCAAGGCGGUUGCGGCGGCAGCGUCUGGGCACUACCCGUCUUC
o                                                                            1260
3'  CCUAGGACACCUGCUGCCCCCGCACCACGUUCCGCCAACGCCGCCGUCGCAGACCCGUGAUGGGCAGAAG
1     Asp Pro Val Asp Asp Gly Gly Val Val Gln Gly Gly Cys Gly Gly Ser Val Trp Ala Leu Pro Val Phe
2   Gly Ile Leu Trp Thr Thr Gly Ala Trp Cys Lys Ala Val Ala Ala Ala Ala Ser Gly His Tyr Pro Ser Ser
3    Gly Ser Cys Gly Arg Arg Gly Arg Gly Ala Arg Arg Leu Arg Arg Gln Arg Leu Gly Thr Thr Arg Leu
o
5'  CCGAGCCAGCGUGAUGCUGAUCGCGUGGGGCGUUUCUAUGCCCCUUCAAUUGAUGACCGUGGACCUGGUG
o                                                                            1330
3'  GGCUCGGUCGCACUACGACUAGCGCACCCCGCAAAGAUACGGGGAAGUUAACUACUGGCACCUGGACCAC
1    Pro Ser Gln Arg Asp Ala Asp Arg Val Gly Arg Phe Tyr Ala Pro Ser Ile Asp Asp Arg Gly Pro Gly
2     Arg Ala Ser Val Met Leu Ile Ala Trp Gly Val Ser Met Pro Leu Gln Leu Met Thr Val Asp Leu Val
3   Pro Glu Pro Ala  .  Cys  .  Ser Arg Gly Ala Phe Leu Cys Pro Phe Asn  .   .  Pro Trp Thr Trp Trp
o
5'  GCGCGGGCAGAGCCCCGCCGCAUUUCCGUUCCGUCGACUGGAGUCCGGUAGGACUAACGGUUGUCACAAA
o                                                                            1400
3'  CGCGCCCGUCUCGGGGCGGCGUAAAGGCAAGGCAGCUGACCUCAGGCCAUCCUGAUUGCCAACAGUGUUU
1   Gly Ala Gly Arg Ala Pro Pro His Phe Arg Ser Val Asp Trp Ser Pro Val Gly Leu Thr Val Val Thr Asn
2    Ala Arg Ala Glu Pro Arg Arg Ile Ser Val Pro Ser Thr Gly Val Arg  .  Asp  .  Arg Leu Ser Gln
3     Arg Gly Gln Ser Pro Ala Ala Phe Pro Phe Arg Arg Leu Glu Ser Gly Arg Thr Asn Gly Cys His Lys
o
5'  CCAUGGCCGGCGUCUUUUCCCGUGGGCGUGGGACCCGCCUCCGAUGGCUGAGGAUAGGGUGAAUGAGGAG
o                                                                            1470
3'  GGUACCGGCCGCAGAAAAGGGCACCCGCACCCUGGGCGGAGGCUACCGACUCCUAUCCCACUUACUCCUC
1     His Gly Arg Arg Leu Phe Pro Trp Ala Trp Asp Pro Pro Pro Met Ala Glu Asp Arg Val Asn Glu Glu
2   Thr Met Ala Gly Val Phe Ser Arg Gly Arg Gly Thr Arg Leu Arg Trp Leu Arg Ile Gly  .  Met Arg Arg
3    Pro Trp Pro Ala Ser Phe Pro Val Gly Val Gly Pro Ala Ser Asp Gly  .  Gly  .  Gly Glu  .  Gly
o
5'  GGUGUGGAGGAGGCGGACGAGGUUGCGGAUGCUUGGGUGAGGAUGGCCGCGUGUUACCUACGCAAGGUGG
o                                                                            1540
3'  CCACACCUCCUCCGCCUGCUCCAACGCCUACGAACCCACUCCUACCGGCGCACAAUGGAUGCGUUCCACC
1    Gly Val Glu Glu Ala Asp Glu Val Ala Asp Ala Trp Val Arg Met Ala Ala Cys Tyr Leu Arg Lys Val
2     Val Trp Arg Arg Arg Thr Arg Leu Arg Met Leu Gly  .  Gly Trp Pro Arg Val Thr Tyr Ala Arg Trp
3   Gly Cys Gly Gly Gly Gly Arg Gly Cys Gly Cys Leu Gly Glu Asp Gly Arg Val Leu Pro Thr Gln Gly Gly
o
5'  AGAUUGUUCGUGCGCUUGAGGGCGAUCAUGGUGUGCAUACUCGUGAGGGCGAUACGCGUACUGUGUUCGG
o                                                                            1610
3'  UCUAACAAGCACGCGAACUCCCGCUAGUACCACACGUAUGAGCACUCCCGCUAUGCGCAUGACACAAGCC
1   Glu Ile Val Arg Ala Leu Glu Gly Asp His Gly Val His Thr Arg Glu Gly Asp Thr Arg Thr Val Phe Gly
2    Arg Leu Phe Val Arg Leu Arg Ala Ile Met Val Cys Ile Leu Val Arg Ala Ile Arg Val Leu Cys Ser
3     Asp Cys Ser Cys Ala  .  Gly Arg Ser Trp Cys Ala Tyr Ser  .  Gly Arg Tyr Ala Tyr Cys Val Arg
o
5'  CACCUCCGCGCAUUACACGAGCUAUCAAAUGCCGCGGCUGAAUUUGGACGGUUGGUGGCCUGCGUUGAUC
o                                                                            1680
3'  GUGGAGGCGCGUAAUGUGCUCGAUAGUUUACGGCGCCGACUUAAACCUGCCAACCACCGGACGCAACUAG
1     Thr Ser Ala His Tyr Thr Ser Tyr Gln Met Pro Arg Leu Asn Leu Asp Gly Trp Trp Pro Ala Leu Ile
2   Ala Pro Pro Arg Ile Thr Arg Ala Ile Lys Cys Arg Gly  .  Ile Trp Thr Val Gly Gly Leu Arg  .  Ser
3    His Leu Arg Ala Leu His Glu Leu Ser Asn Ala Ala Ala Glu Phe Gly Arg Leu Val Ala Cys Val Asp
o
```

```
5'  GGCCUGUCCGUUUUGCGCCACGAUCGUGUGGUGCCAAAAUUGGAUCGCCGGCUUCUGCGGCCUGCGUUCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1750
3'  CCGGACAGGCAAAACGCGGUGCUAGCACACCACGGUUUUAACCUAGCGGCCGAAGACGCCGGACGCAAGU
1    Gly Leu Ser Val Leu Arg His Asp Arg Val Val Pro Lys Leu Asp Arg Arg Leu Leu Arg Pro Ala Phe
2     Ala Cys Pro Phe Cys Ala Thr Ile Val Trp Cys Gln Asn Trp Ile Ala Gly Phe Cys Gly Leu Arg Ser
3   Arg Pro Val Arg Phe Ala Pro Arg Ser Cys Gly Ala Lys Ile Gly Ser Pro Ala Ser Ala Ala Cys Val His
o
5'  CCAAGUUUGCAGCCGACGUUCACUUGCUCACGCAUCGGACGUUGUUUGAGUCGGGGAAUUCCGUGGCCGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1820
3'  GGUUCAAACGUCGGCUGCAAGUGAACGAGUGCGUAGCCUGCAACAAACUCAGCCCCUUAAGGCACCGGCU
1   Thr Lys Phe Ala Ala Asp Val His Leu Leu Thr His Arg Thr Leu Phe Glu Ser Gly Asn Ser Val Ala Asp
2    Pro Ser Leu Gln Pro Thr Phe Thr Cys Ser Arg Ile Gly Arg Cys Leu Ser Arg Gly Ile Pro Trp Pro
3     Gln Val Cys Ser Arg Arg Ser Leu Ala His Ala Ser Asp Val Val . Val Gly Glu Phe Arg Gly Arg
o
5'  CCUCUCUGAUGCACUGGUCGGGGCGAAGGUGGUUUCUCGGUUUCCGCCUGCUUAUAGAGCAGGCGUGUGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1890
3'  GGAGAGACUACGUGACCAGCCCCGCUUCCACCAAAGAGCCAAAGGCGGACGAAUAUCUCGUCCGCACACC
1     Leu Ser Asp Ala Leu Val Gly Ala Lys Val Val Ser Arg Phe Pro Pro Ala Tyr Arg Ala Gly Val Trp
2   Thr Ser Leu Met His Trp Ser Gly Arg Arg Trp Phe Leu Gly Phe Arg Leu Leu Ile Glu Gln Ala Cys Gly
3    Pro Leu . Cys Thr Gly Arg Gly Glu Gly Gly Phe Ser Val Ser Ala Cys Leu . Ser Arg Arg Val
o
5'  CCGCAUGUGUUCUCUUCGGUGAACAUGCCUUAUGGGAACUACGAAUGCCUCGAGAGCGGCGUGCUGUUGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1960
3'  GGCGUACACAAGAGAAGCCACUUGUACGGAAUACCCUUGAUGCUUACGGAGCUCUCGCCGCACGACAACC
1    Pro His Val Phe Ser Ser Val Asn Met Pro Tyr Gly Asn Tyr Glu Cys Leu Glu Ser Gly Val Leu Leu
2     Arg Met Cys Ser Leu Arg . Thr Cys Leu Met Gly Thr Thr Asn Ala Ser Arg Ala Ala Cys Cys Trp
3   Ala Ala Cys Val Leu Phe Gly Glu His Ala Leu Trp Glu Leu Arg Met Pro Arg Glu Arg Arg Ala Val Gly
o
5'  GGGGUGGUAACGAGACGGAAGGGGUGGGCUUCAAUGUGCCGGGUUCGUGGAAGUGGGACGGAGUGCAGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2030
3'  CCCCACCAUUGCUCUGCCUUCCCCACCCGAAGUUACACGGCCCAAGCACCUUCACCCUGCCUCACGUCGC
1   Gly Gly Gly Asn Glu Thr Glu Gly Val Gly Phe Asn Val Pro Gly Ser Trp Lys Trp Asp Gly Val Gln Arg
2    Gly Val Val Thr Arg Arg Lys Gly Trp Ala Ser Met Cys Arg Val Arg Gly Ser Gly Thr Glu Cys Ser
3     Gly Trp . Arg Asp Gly Arg Gly Gly Leu Gln Cys Ala Gly Phe Val Glu Val Gly Arg Ser Ala Ala
o
5'  CAAGGCCGAGUUGGAUGGCUCGGAUGCGCCGGCUAUUCGCCAAAGCUUGAGGGCUUUGGACUCUGUUGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2100
3'  GUUCCGGCUCAACCUACCGAGCCUACGCGGCCGAUAAGCGGUUUCGAACUCCCGAAACCUGAGACAACGC
1     Lys Ala Glu Leu Asp Gly Ser Asp Ala Pro Ala Ile Arg Gln Ser Leu Arg Ala Leu Asp Ser Val Ala
2   Ala Arg Pro Ser Trp Met Ala Arg Met Arg Arg Leu Phe Ala Lys Ala . Gly Leu Trp Thr Leu Leu Arg
3    Gln Gly Arg Val Gly Trp Leu Gly Cys Ala Gly Tyr Ser Pro Lys Leu Glu Gly Phe Gly Leu Cys Cys
o
5'  CGGAAGUUGUACUUCUAUGGUGGAACCCUGCGUCUAGAUGUUCACCCCGACCGUCCGGUCUAUGUGGUUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2170
3'  GCCUUCAACAUGAAGAUACCACCUUGGGACGCAGAUCUACAAGUGGGGCUGGCAGGCCAGAUACACCAAG
1    Arg Lys Leu Tyr Phe Tyr Gly Gly Thr Leu Arg Leu Asp Val His Pro Asp Arg Pro Val Tyr Val Val
2     Gly Ser Cys Thr Ser Met Val Glu Pro Cys Val . Met Phe Thr Pro Thr Val Arg Ser Met Trp Phe
3   Ala Glu Val Val Leu Leu Trp Trp Asn Pro Ala Ser Arg Cys Ser Pro Arg Pro Ser Gly Leu Cys Gly Ser
o
5'  GGCCGGCUGGUAGUAGGCUCUACCAUCCAUAUUUUGUCCCUGUCAGGGUGCUCGAGGAUAGGCUCCCUAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2240
3'  CCGGCCGACCAUCAUCCGAGAUGGUAGGUAUAAAACAGGGACAGUCCCACGAGCUCCUAUCCGAGGGAUC
1   Arg Pro Ala Gly Ser Arg Leu Tyr His Pro Tyr Phe Val Pro Val Arg Val Leu Glu Asp Arg Leu Pro Ser
2    Gly Arg Leu Val Val Gly Ser Thr Ile His Ile Leu Ser Leu Ser Gly Cys Ser Arg Ile Gly Ser Leu
3     Ala Gly Trp . . Ala Leu Pro Ser Ile Phe Cys Pro Cys Gln Gly Ala Arg Gly . Ala Pro .
o
```

```
5'    CGGUGUGAGGUAUACUGCCAUUGGCAGCGCCGCGCACCUGUUGUCGCCUGGCAGGCCAACUGAUGUUGGU
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2310
3'    GCCACACUCCAUAUGACGGUAACCGUCGCGGCGCGUGGACAACAGCGGACCGUCCGGUUGACUACAACCA
1       Gly Val Arg Tyr Thr Ala Ile Gly Ser Ala Ala His Leu Leu Ser Pro Gly Arg Pro Thr Asp Val Gly
2     Ala Val . Gly Ile Leu Pro Leu Ala Ala Pro Arg Thr Cys Cys Arg Leu Ala Gly Gln Leu Met Leu Val
3      Arg Cys Glu Val Tyr Cys His Trp Gln Arg Arg Ala Pro Val Val Ala Trp Gln Ala Asn . Cys Trp
o
5'    AGAGCUUCUGGUGUCAUUUGAGAUUUCGUUAUAGUUGCAUUUUCUCGCGCAUAUGCGUGCUCACACUCUU
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2380
3'    UCUCGAAGACCACAGUAAACUCUAAAGCAAUAUCAACGUAAAAGAGCGCGUAUACGCACGAGUGUGAGAA
1      Arg Ala Ser Gly Val Ile . Asp Phe Val Ile Val Ala Phe Ser Arg Ala Tyr Ala Cys Ser His Ser
2       Glu Leu Leu Val Ser Phe Glu Ile Ser Leu . Leu His Phe Leu Ala His Met Arg Ala His Thr Leu
3     . Ser Phe Trp Cys His Leu Arg Phe Arg Tyr Ser Cys Ile Phe Ser Arg Ile Cys Val Leu Thr Leu Phe
o
5'    UAACGCGAUUCGAAUAAGUCUGCCGUAGUCAUCGGCCCCUUUUCUGUGGGCUCGACGUGUUUUUUAGAAA
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2450
3'    AUUGCGCUAAGCUUAUUCAGACGGCAUCAGUAGCCGGGGAAAAGACACCCGAGCUGCACAAAAAAUCUUU
1     Leu Thr Arg Phe Glu . Val Cys Arg Ser His Arg Pro Leu Phe Cys Gly Leu Asp Val Phe Phe Arg Lys
2      . Arg Asp Ser Asn Lys Ser Ala Val Val Ile Gly Pro Phe Ser Val Gly Ser Thr Cys Phe Leu Glu
3       Asn Ala Ile Arg Ile Ser Leu Pro . Ser Ser Ala Pro Phe Leu Trp Ala Arg Arg Val Phe . Lys
o
5'    GAAGCGUUUAGCGGCGGAACCUCCGCAUUUACUGUAGGUUCUUGAGUCAGUGGGGCACGCUCGGCGGUGG
o     ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2520
3'    CUUCGCAAAUCGCCGCCUUGGAGGCGUAAAUGACAUCCAAGAACUCAGUCACCCCGUGCGAGCCGCCACC
1       Lys Arg Leu Ala Ala Glu Pro Pro His Leu Leu . Val Leu Glu Ser Val Gly His Ala Arg Arg Trp
2     Arg Ser Val . Arg Arg Asn Leu Arg Ile Tyr Cys Arg Phe Leu Ser Gln Trp Gly Thr Leu Gly Gly Gly
3      Glu Ala Phe Ser Gly Gly Thr Ser Ala Phe Thr Val Gly Ser . Val Ser Gly Ala Arg Ser Ala Val
o
5'    GCGUGUCUCGAACGUGGAAC
o     ++++|++++|++++|++++|                                                     2540
3'    CGCACAGAGCUUGCACCUUG
1      Ala Cys Leu Glu Arg Gly Thr
2       Arg Val Ser Asn Val Glu
3     Gly Val Ser Arg Thr Trp Asn
o
```

2.5.6 AaV1 dsRNA4 sequence (NC_010991)

1382 nt excluding poly (A) tail

```
5'  CAUAAAGCAAUCGUAUCGCCAGAUACACAUAAGUCUUUGCCCUCCGUGUGAUGUUUGAUUCCUUUUGUUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   70
3'  GUAUUUCGUUAGCAUAGCGGUCUAUGUGUAUUCAGAAACGGGAGGCACACUACAAACUAAGGAAAACAAG
1    His Lys Ala Ile Val Ser Pro Asp Thr His Lys Ser Leu Pro Ser Val  .  Cys Leu Ile Pro Phe Val
2     Ile Lys Gln Ser Tyr Arg Gln Ile His Ile Ser Leu Cys Pro Pro Cys Asp Val  .  Phe Leu Leu Phe
3   Pro  .  Ser Asn Arg Ile Ala Arg Tyr Thr  .  Val Phe Ala Leu Arg Val Met Phe Asp Ser Phe Cys Ser
o
5'  CUUCCCCGCUUUUGUCCCUUUUGUUCUUGAUGAUUUUUCCGGUCGGGUGAGCGCCCCCGCGCUGCUCACU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   140
3'  GAAGGGGCGAAAACAGGGAAAACAAGAACUACUAAAAAGGCCAGCCCACUCGCGGGGGCGCGACGAGUGA
1   Pro Ser Pro Leu Leu Ser Leu Leu Phe Leu Met Ile Phe Pro Val Gly  .  Ala Pro Pro Arg Cys Ser Leu
2    Leu Pro Arg Phe Cys Pro Phe Cys Ser  .   .  Phe Phe Arg Ser Gly Glu Arg Pro Arg Ala Ala His
3     Phe Pro Ala Phe Val Pro Phe Val Leu Asp Asp Phe Ser Gly Arg Val Ser Ala Pro Ala Leu Leu Thr
o
5'  GUGCAGGCUACGGGUGGCAUGAAAGAUGGGCGCCAAAUGGCGACCGCCUCGAUUGCUUUCUCCGGGGCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   210
3'  CACGUCCGAUGCCCACCGUACUUUCUACCCGCGGUUUACCGCUGGCGGAGCUAACGAAAGAGGCCCCGAC
1     Cys Arg Leu Arg Val Ala  .  Lys Met Gly Ala Lys Trp Arg Pro Pro Arg Leu Leu Ser Pro Gly Leu
2   Cys Ala Gly Tyr Gly Trp His Glu Arg Trp Ala Pro Asn Gly Asp Arg Leu Asp Cys Phe Leu Arg Gly Cys
3    Val Gln Ala Thr Gly Gly Met Lys Asp Gly Arg Gln Met Ala Thr Ala Ser Ile Ala Phe Ser Gly Ala
o
5'  UUCCAAGCGUGGAUGCGUUGGAUGACGACACCCUUUCCUGUGGUGUCGGCCGUUAUAUCCAUCACGUCCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   280
3'  AAGGUUCGCACCUACGCAACCUACUGCUGUGGGAAAGGACACCACAGCCGGCAAUAUAGGUAGUGCAGGU
1    Phe Gln Ala Trp Met Arg Trp Met Thr Thr Pro Phe Pro Val Val Ser Ala Val Ile Ser Ile Thr Ser
2     Ser Lys Arg Gly Cys Val Gly  .  Arg His Pro Phe Leu Trp Cys Arg Pro Leu Tyr Pro Ser Arg Pro
3   Val Pro Ser Val Asp Ala Leu Asp Asp Asp Thr Leu Ser Cys Gly Val Gly Arg Tyr Ile His His Val His
o
5'  UGUUGACAAGGUGAGUUUUUGCGGGGGGCCGUGUGUGAUAAGGAGACGUGUUGUAGUGGGUGUGGCGUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   350
3'  ACAACUGUUCCACUCAAAAACGCCCCCCGGCACACACUAUUCCUCUGCACAACAUCACCCACACCGCAA
1   Met Leu Thr Arg  .  Val Phe Ala Gly Gly Pro Cys Val Ile Arg Arg Arg Val Val Val Gly Val Ala Phe
2    Cys  .  Gln Gly Glu Phe Leu Arg Gly Gly Arg Val  .   .  Gly Asp Val Leu  .  Trp Val Trp Arg
3     Val Asp Lys Val Ser Phe Cys Gly Gly Ala Val Cys Asp Lys Glu Thr Cys Cys Ser Gly Cys Gly Val
o
5'  UCCAGCGUGCGUCCUGAUCCGGAGGGGGGUUUCUUCUCCUCCGCGUCUUUGGGUGCGCUGGUUUCCGCGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   420
3'  AGGUCGCACGCAGGACUAGGCCUCCCCCCAAAGAAGAGGAGGCGCAGAAACCCACGCGACCAAAGGCGCC
1     Pro Ala Cys Val Leu Ile Arg Arg Gly Val Ser Ser Pro Pro Arg Leu Trp Val Arg Trp Phe Pro Arg
2   Phe Gln Arg Ala Ser  .  Ser Gly Gly Gly Phe Leu Leu Leu Arg Val Phe Gly Cys Ala Gly Phe Arg Gly
3    Ser Ser Val Arg Pro Asp Pro Glu Gly Gly Phe Phe Ser Ser Ala Ser Leu Gly Ala Leu Val Ser Ala
o
5'  GUUUGGUCGAGGCCGGUCUGGAAGACGAGCGAUGCACCGUGAUGGUGUCCUUUGGGGGCUUACGUAUCCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   490
3'  CAAACCAGCUCCGGCCAGACCUUCUGCUCGCUACGUGGCACUACCACAGGAAACCCCCGAAUGCAUAGGG
1    Val Trp Ser Arg Pro Val Trp Lys Thr Ser Asp Ala Pro  .  Trp Cys Pro Leu Gly Ala Tyr Val Ser
2     Phe Gly Arg Gly Arg Ser Gly Arg Arg Ala Met His Arg Asp Gly Val Leu Trp Gly Leu Thr Tyr Pro
3   Gly Leu Val Glu Ala Gly Leu Glu Asp Glu Arg Cys Thr Val Met Val Ser Phe Gly Gly Leu Arg Ile Pro
o
5'  UUUGGAUGCCACGGGUGGCGUCGUGCCGCAUUCGCGCGCGCUGGAUGAGGCUGCUUUGGGGGUUUUCCUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   560
3'  AAACCUACGGUGCCCACCGCAGCACGGCGUAAGCGCGCGCGACCUACUCCGACGAAACCCCCAAAAGGAA
1   Leu Trp Met Pro Arg Val Ala Ser Cys Arg Ile Arg Ala Arg Trp Met Arg Leu Leu Trp Gly Phe Ser Phe
2    Phe Gly Cys His Gly Trp Arg Arg Ala Ala Phe Ala Arg Ala Gly  .  Gly Cys Phe Gly Gly Phe Pro
3     Leu Asp Ala Thr Gly Gly Val Val Pro His Ser Arg Ala Leu Asp Glu Ala Ala Leu Gly Val Phe Leu
o
```

```
5'  CCCCCUCCGCAGACCUCCGAUCCAGGCGAUGUGGCUGUACCUAACGAGCCUUGCGAUGAAGAGAUCAAGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  630
3'  GGGGGAGGCGUCUGGAGGCUAGGUCCGCUACACCGACAUGGAUUGCUCGGAACGCUACUUCUCUAGUUCU
1    Pro Leu Arg Arg Pro Pro Ile Gln Ala Met Trp Leu Tyr Leu Thr Ser Leu Ala Met Lys Arg Ser Arg
2   Ser Pro Ser Ala Asp Leu Arg Ser Arg Arg Cys Gly Cys Thr  .  Arg Ala Leu Arg  .  Arg Asp Gln Asp
3    Pro Pro Pro Gln Thr Ser Asp Pro Gly Asp Val Ala Val Pro Asn Glu Pro Cys Asp Glu Glu Ile Lys
o
5'  CCCUCGCUGACGGGUUGUUGUCCCAUGCCGUUGAAUUGGGAAUGCCCGAGCAGGCCGCAGAAGCCUCAGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  700
3'  GGGAGCGACUGCCCAACAACAGGGUACGGCAACUUAACCCUUACGGGCUCGUCCGGCGUCUUCGGAGUCA
1    Pro Ser Leu Thr Gly Cys Cys Pro Met Pro Leu Asn Trp Glu Cys Pro Ser Arg Pro Gln Lys Pro Gln
2    Pro Arg  .  Arg Val Val Val Pro Cys Arg  .  Ile Gly Asn Ala Arg Ala Gly Arg Arg Ser Leu Ser
3   Thr Leu Ala Asp Gly Leu Leu Ser His Ala Val Glu Leu Gly Met Pro Glu Gln Ala Ala Glu Ala Ser Val
o
5'  CGUUGUCCCGUUGCCCCCUCCACCGCCUGAUCCGGAGGAUCUGCCUGGCAGUGCUAUGAACCUCGGCACG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  770
3'  GCAACAGGGCAACGGGGGAGGUGGCGGACUAGGCCUCCUAGACGGACCGUCACGAUACUUGGAGCCGUGC
1   Ser Leu Ser Arg Cys Pro Leu His Arg Leu Ile Arg Arg Ile Cys Leu Ala Val Leu  .  Thr Ser Ala Arg
2    Arg Cys Pro Val Ala Pro Ser Thr Ala  .  Ser Gly Gly Ser Ala Trp Gln Cys Tyr Glu Pro Arg His
3    Val Val Pro Leu Pro Pro Pro Pro Pro Asp Pro Glu Asp Leu Pro Gly Ser Ala Met Asn Leu Gly Thr
o
5'  UUGGGUGGCGUUCUUGCCGCCAUUGCAGAUGGGUCUCUCGUUCGCGAGAUCAAGACUUAUGAUCCACCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  840
3'  AACCCACCGCAAGAACGGCGGUAACGUCUACCCAGAGAGCAAGCGCUCUAGUUCUGAAUACUAGGUGGAC
1    Trp Val Ala Phe Leu Pro Pro Leu Gln Met Gly Leu Ser Phe Ala Arg Ser Arg Leu Met Ile His Leu
2   Val Gly Trp Arg Ser Cys Arg His Cys Arg Trp Val Ser Arg Ser Arg Asp Gln Asp Leu  .  Ser Thr Cys
3    Leu Gly Gly Val Leu Ala Ala Ile Ala Asp Gly Ser Leu Val Arg Glu Ile Lys Thr Tyr Asp Pro Pro
o
5'  CUGGCGGCUUCGAAGGCCCUAGAAGUGAGGUGCUUGAGCGGCGUUUGCAGUGGCUGAAUGCUUUGGGCAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  910
3'  GACCGCCGAAGCUUCCGGGAUCUUCACUCCACGAACUCGCCGCAAACGUCACCGACUUACGAAACCCGUC
1    Leu Ala Ala Ser Lys Ala Leu Glu Val Arg Cys Leu Ser Gly Val Cys Ser Gly  .  Met Leu Trp Ala
2    Trp Arg Leu Arg Arg Pro  .  Lys  .  Gly Ala  .  Ala Ala Phe Ala Val Ala Glu Cys Phe Gly Gln
3   Ala Gly Gly Phe Glu Gly Pro Arg Ser Glu Val Leu Glu Arg Arg Leu Gln Trp Leu Asn Ala Leu Gly Ser
o
5'  CCUUUCGGAUUCUUGCUACUCAGUUGUAGCAGAUGCCGGCUGGUAUCGCCUUCUUUUGCCUUUCGCUCAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  980
3'  GGAAAGCCUAAGAACGAUGAGUCAACAUCGUCUACGGCCGACCAUAGCGGAAGAAAACGGAAAGCGAGUA
1   Ala Phe Arg Ile Leu Ala Thr Gln Leu  .  Gln Met Pro Ala Gly Ile Ala Phe Phe Cys Leu Ser Leu Ile
2    Pro Phe Gly Phe Leu Leu Leu Ser Cys Ser Arg Cys Arg Leu Val Ser Pro Ser Phe Ala Phe Arg Ser
3    Leu Ser Asp Ser Cys Tyr Ser Val Val Ala Asp Ala Gly Trp Tyr Arg Leu Leu Leu Pro Phe Ala His
o
5'  UGGCAAAAGCAGUUGGGUUUGUUGUCUGGGCGUGAGAUGGAUCUGCUCCUGGGGGGUCUAUCUCGUGUCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1050
3'  ACCGUUUUCGUCAACCCAAACAACAGACCCGCACUCUACCUAGACGAGGACCCCCCAGAUAGAGCACAGU
1    Gly Lys Ser Ser Trp Val Cys Cys Leu Gly Val Arg Trp Ile Cys Ser Trp Gly Val Tyr Leu Val Ser
2   Leu Ala Lys Ala Val Gly Phe Val Val Trp Ala  .  Asp Gly Ser Ala Pro Gly Gly Ser Ile Ser Cys Gln
3    Trp Gln Lys Gln Leu Gly Leu Leu Ser Gly Arg Glu Met Asp Leu Leu Leu Gly Gly Leu Ser Arg Val
o
5'  AGGAUCCCCGUCUGCGUGCGGAUAGGUAUUUGCAUUUCUUGCGUACUGUGGUUGGUGCUUCCGCCGGUGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1120
3'  UCCUAGGGGCAGACGCACGCCUAUCCAUAAACGUAAAGAACGCAUGACACCAACCACGAAGGCGGCCACC
1    Arg Ile Pro Val Cys Val Arg Ile Gly Ile Cys Ile Ser Cys Val Leu Trp Leu Val Leu Pro Pro Val
2    Gly Ser Pro Ser Ala Cys Gly  .  Val Phe Ala Phe Leu Ala Tyr Cys Gly Trp Cys Phe Arg Arg Trp
3   Lys Asp Pro Arg Leu Arg Ala Asp Arg Tyr Leu His Phe Leu Arg Thr Val Val Gly Ala Ser Ala Gly Gly
o
```

```
5'  CGUGGGUGAUCCUCCUUGGGCUUACGGCUCUGUAGCCAAUUUUCAAGUUGGGAUGUAUCGCCUGUUGUGG
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  1190
3'  GCACCCACUAGGAGGAACCCGAAUGCCGAGACAUCGGUUAAAAGUUCAACCCUACAUAGCGGACAACACC
1   Ala Trp Val Ile Leu Leu Gly Leu Thr Ala Leu  .  Pro Ile Phe Lys Leu Gly Cys Ile Ala Cys Cys Gly
2    Arg Gly  .  Ser Ser Leu Gly Leu Arg Leu Cys Ser Gln Phe Ser Ser Trp Asp Val Ser Pro Val Val
3     Val Gly Asp Pro Pro Trp Ala Tyr Gly Ser Val Ala Asn Phe Gln Val Gly Met Tyr Arg Leu Leu Trp
o
5'  CCGGCUCGUGGCGCCAAGGUUGGAGUAGGUACGUCUGCAUAAGUCGCCGGGCGAGUCACCUCUCUGUAUA
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  1260
3'  GGCCGAGCACCGCGGUUCCAACCUCAUCCAUGCAGACGUAUUCAGCGGCCCGCUCAGUGGAGAGACAUAU
1     Arg Leu Val Ala Pro Arg Leu Glu  .  Val Arg Leu His Lys Ser Pro Gly Glu Ser Pro Leu Cys Ile
2   Ala Gly Ser Trp Arg Gln Gly Trp Ser Arg Tyr Val Cys Ile Ser Arg Arg Ala Ser His Leu Ser Val  .
3    Pro Ala Arg Gly Ala Lys Val Gly Val Gly Thr Ser Ala  .  Val Ala Gly Arg Val Thr Ser Leu Tyr
o
5'  GAGUUAGGCGUGUUUUUUAGAACCUCCAGGAUCCUUAUUGUCCUCCGCUCGUACUUUAGGGCAAUACUUA
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  1330
3'  CUCAAUCCGCACAAAAAAUCUUGGAGGUCCUAGGAAUAACAGGAGGCGAGCAUGAAAUCCCGUUAUGAAU
1    Glu Leu Gly Val Phe Phe Arg Thr Ser Arg Ile Leu Ile Val Leu Arg Ser Tyr Phe Arg Ala Ile Leu
2     Ser  .  Ala Cys Phe Leu Glu Pro Pro Gly Ser Leu Leu Ser Ser Ala Arg Thr Leu Gly Gln Tyr Leu
3   Arg Val Arg Arg Val Phe  .  Asn Leu Gln Asp Pro Tyr Cys Pro Pro Leu Val Leu  .  Gly Asn Thr Tyr
o
5'  CUUCGCCCCUCCGCCGGCGGUAGCGGGAGGGUGGUGCUGUGAGGAUAACAAC
o   ----+----|----+----|----+----|----+----|----+----|--                    1382
3'  GAAGCGGGGAGGCGGCCGCCAUCGCCCUCCCACCACGACACUCCUAUUGUUG
1   Thr Ser Pro Leu Arg Arg Arg  .  Arg Glu Gly Gly Ala Val Arg Ile Thr Thr
2    Leu Arg Pro Ser Ala Gly Gly Ser Gly Arg Val Val Leu  .  Gly  .  Gln
3     Phe Ala Pro Pro Pro Ala Val Ala Gly Gly Trp Cys Cys Glu Asp Asn Asn
o
```

3. Chapter 3 – Molecular Characterization of a Novel Alternavirus, Diaporthe alternavirus 1 (DAV1)

3.1 Introduction

3.1.1 *Diaporthe* spp.

The genus *Diaporthe* (asexual morph *Phomopsis*) is a member of Ascomycota, having a worldwide distribution and broad host ranges. They are identified as plant pathogens, saprobes, endophytes even pathogens of humans and animals. *Diaporthe* spp. of plant pathogens cause rot, wilt, decay, blights dieback, cankers, and leaf spots of many economically important plants, such as *Citrus* and *Pyrus*. The identification and classification of *Diaporthe* species are difficult. At present, species of *Diaporthe* is classified by multi-locus analysis of internal transcribed spacer (ITS) region, partial translation elongation factor 1-alpha (*TEF*1), beta-tubulin (*TUB*), histone H3 (*HIS*) and calmodulin (*CAL*) genes, and the difference of colony and spore morphology (Gomes *et al.*, 2013).

3.1.2 *Diaporthe* spp.-infecting mycoviruses

Until now, only one mycovirus has been identified in the genus *Diaporthe*. Diaporthe RNA virus 1 (DRV1), a 4.1 kbp positive single-stranded RNA virus, found in a plant pathogen *Diaporthe perjuncta* (Preisig *et al.*, 2000; Moleleki *et al.*, 2003). The DRV1-infected strain showed hypovirulence, and lack of sporulation (Moleleki *et al.*, 2003; Moleleki *et al.*, 2011). The transfection experiment of the in-vitro transcribed DRV1 RNAs indicated that DRV1 were related to morphological changes and pathogenicity of *Diaporthe* strains (Moleleki *et al.*, 2011).

In this study, I identified a new alternavirus, Diaporthe alternavirus 1 (DAV1). The four dsRNA segments were packaged in an isometric particle which consist of ORF3-encoded 79.4 kDa proteins. Like AaV1, the dsRNA1 segments of DAV1 have both 5' caps and 3' poly (A:U) structures, which are expected to provide more information for the classification of Alternaviridae.

3.2 Materials and methods

3.2.1 Fungal strain and culture conditions

Two fungal isolates of *Diaporthe* affinis (abbreviated aff.) *acuta*, IbSTRPmp18001 and IbSTR18001-VF, were used in this study. IbSTRPmp18001 was isolated from crown rot of strawberry by Dr. Ryo Okada in Ibaraki of Japan in 2018. IbSTRPmp18001-VF is the isolate, cured of virus-infected strain IbSTRPmp18001 in this study. All isolates were grown on D- PDA media (24 g/L Difico^{TM} Potato dextrose powder and 15 g/L agar) and OMA media (72.5 g/L Difico^{TM} Oatmeal agar powder) at 25°C for 5 days, and the mycelial discs were stored at -80°C in glycerol. The types of conidia were observed under microscope after at least 30 days culture on D-PDA media. For liquid culture, mycelial plugs were inoculated in PDB (24 g/L Difico^{TM} Potato dextrose powder) and cultured with reciprocal shaking (60 strokes per min) at 25°C for one week.

The species ascription of fungal strain was based on cloning, sequencing, and multilocus analysis of partial internal transcribed spacer (ITS) region, *translation elongation factor 1-α* (*TEF1*), *calmodulin* (*CAL*), *histone h-3* (*HIS*), and *β-tubulin* (*TUB*) genes. The primer pairs used are listed in Table 3-1. For the phylogenetic analyses, species closely related to IbSTRPmp18001 were included from previous studies (Ozawa *et al*., 2019; Guo *et al*., 2020) and the sequences were downloaded from NCBI Genbank (Table 3-2). *D. citri* CBS 135422 was used as an outgroup. The sequences generated in this study and the sequences downloaded from Genbank were merged and then aligned with MAFFT (Katoh and Standley, 2013). Alignments were manually modified where necessary and the poorly aligned regions were removed with Gblocks program (Talavera and Castresana, 2007). The maximum-likelihood tree and Bayesian inference tree were inferred with RAxML 8.2.4 and MrBayes v. 3.2.6 programs under Linux Mint 20.2 operating system, respectively. Random seeds used throughout the analysis were set to 5566. For maximum-likelihood analyses, the bootstrap support values were calculated from 1000 bootstrap replicates. For Bayesian inference, two MCMC chains were run simultaneously from a random starting tree for 1,000,000 generations. Every 1000th generation was sampled, resulting in 1000 trees. Bayesian posterior probabilities were calculated from the last three quarters of the 1000 trees sampled. (This work was in cooperation with Dr. Jie-Hao Ou of National Chung Hsing University)

3.2.2 Curing of DAV1-infected isolates

Two methods were applied to obtain the DAV1-free isolates. The first method is continuous subculturing on the medium. Another is protoplast method which adopted the protocol of the published article (De Guido *et al.*, 2003; Anco *et al.*, 2009) with modifications.

IbSTR18001 strain (virus-infected) was cultured in 100 ml PDB with 120 rpm shaking at 25°C for 36 h. About 1 g of mycelia were collected by filtration using double layers of sterilized Miracloth (Calbiochem®, San Diego, CA), then wash with sterilized 0.6 M MgSO4 solution three times. The washed mycelia were digested in the 10 ml enzyme buffer (0.5 g lysing enzyme in 10 ml 1.2 M $MgSO_4$ solution) and slowly rotated at room temperature for 3 h. The suspension was filtered through a sterilized Miracloth to remove the mycelial debris and the filtrate was collected in a 15 ml sterilized falcon tube. 2 ml sterilized 0.6M KCl solution was gently layered on the filtrate, then centrifuged at 2,800 × g (TOMY Suprema 21, TS-38N) for 15 min at 4°C, then the interface of two layers was collected and the number of protoplasts was confirmed by the cytometer. The diluted protoplast solution (2×10^3 protoplasts/100μg) was mixed with 45°C regeneration media (0.6 M sucrose, 5 mM HEPES, 1 mM $NH_4H_2PO_4$, 35 mg/L, and 10 g/L agar; with or without 100 μM Ribavirin), then poured into petridish (20 ml/plate). After 48 h incubation at 25°C, the single colonies were selected and subcultured on the PDA media at 25°C for 2 days.

The selected isolates were subjected to the liquid culture, then the existence of viral dsRNAs was confirmed by dsRNA extraction which is described in sections 2.2.5.

3.2.3 Determination of viral dsRNA sequence and phylogenetic analysis

The viral dsRNA molecules were isolated from fungal strain IbSTR18001 by spin column method described in section 2.2.5. The draft sequences were determined by random RT-PCR, cDNA library cloning, then sequencing. The 5′ and 3′ terminal sequences were obtained by RACE method. The above works were performed by Dr. Okada.

The determined sequences were analyzed by BLAST on NCBI. Then the phylogenetic tree was built by the same procedures described in section 2.2.9.

3.2.4 Detection of 5′ cap structure and 3′ poly (A: U) structure

To detect the 5′ m^7G-cap structures of DAV1 dsRNA segments, the viral dsRNAs were extracted from purified virus particle suspensions of DAV1, ScV-L-A and AaV1. The concentrations of dsRNA samples were adjusted to 1000, 500, and 250 ng/μl, respectively, then subjected to RNA dot blot assay with anti-m^7G-cap antibody. The procedures of RNA dot blot are described in section 2.2.6.

In Aoki's doctoral thesis (Aoki, 2009), the presence and length of 3′ poly (A:U) of AaV1 dsRNA segments were determined by the 5′ RACE method on 5′ terminuses of each AaV1 dsRNA minus strands. The presence of 3′ poly (A:U) structures of DAV1 dsRNA genome was preliminarily confirmed by RT-PCR (Figure 3-9A) which described in 2.2.8 with specific primers (Table 3-1) and oligo $dT_{(20)}$ primer. The PCR products were analyzed by electrophoresis in 1% agarose gels containing EtBr (0.5 μg/ml).

3.2.5 Purification of virus particle and analysis of virus-associated proteins

The virus particle purification procedure of DAV1 is the same with that of AaV1, which is described in section 2.2.3 The DAV1-associated proteins were collected after SDS-PAGE electrophoresis, followed by the analysis of LC-MS/MS as described in section 2.2.4.

3.3 Results

3.3.1 Identification of *Diaporthe* sp.

The sequences obtained in this study were deposited to the DNA Data Bank of Japan (DDBJ; LC663742 (ITS), LC663743 (CAL), LC663744 (HIS), LC663745 (TUB), LC663746 (TEF)). In multi-locus phylogenetic analyses, strain IbSTRPmp18001 formed a sister group with *D. acuta* with moderate statistical support (69/0.81) (Figure 3-1F). Slight differences were observed in pairwise alignments between IbSTRPmp18001 and the type strain PSCG 047 of *D. acuta* (ITS, 470/505 (93%); *CAL*, 452/456 (99%); *HIS* 429/430 (99%); *TUB* 455/456 (99%); *TEF* 283/286 (99%)).

The colonies of IbSTRPmp18001 were white, effuse (Figure 3-2A), 80 mm in diameter after 4 days on D-PDA. Dark brown to black conidiomata were observed after 4 weeks on D-PDA. Alpha conidia (Figure 3-3A) were hyaline, aseptate, fusiform to oval, 5.5–7.5 × 2–3 μm (mean±SD = 6.64±0.53 ×2.63 ±0.32 μm, L/W ratio = 2.56, n=30). Beta conidia (Figure 3-3B) were hyaline, aseptate, filiform, 14–31 × 1–2 μm (mean±SD. =

24.10±4.27 × 1.50 ±0.28 µm, L/W ratio = 16.8, n=25). In addition, the isolates cultured on D-PDA media produced more alpha-conidia, while those cultured on OMA media produced more beta-conidia (data not shown). We tentatively identified IbSTRPmp18001 strain as “*Diaporthe* aff. *acuta*”.

3.3.2 Curing of virus-infected *D.* aff. *acuta* isolate IbSTR18001

After 6 months of continuous succession culture on D-PDA, I obtained one isolate that did not contain viral dsRNA segments (Figure 3-6A). The virus-free strain was confirmed by specific RT-PCR, and no specific band was detected (Figure 3-4).

No significant difference in colony growth and hyphae form was observed between virus-infected isolate and virus-free isolate after five days of incubation on D-PDA (Figure 3-2A and C). There was also no significant difference in the growth of colonies on OMA media commonly used for culturation of *Diaporthe* spp. (Figure 3-2B).

In addition, experiments to cure virus-infected strain by protoplast method are still in progress. The resultant protoplasts were treated with 100 µM lysing enzyme for 3 hours, and about 80 % of the hyphae were digested (Figure 3-5), and finally 2 ml of protoplast resuspension with a concentration of about 1.38 × 10^4 cells/ul was obtained. The resultant protoplasts grew on regeneration medium for two days, then were transferred to D-PDA on the third day. The further confirmation of virus-free treatment will be applied in the future.

3.3.3 Genome organization and classification of DAV1

Four dsRNA segments, dsRNA1, dsRNA2, daRNA3, and dsRNA4, were isolated from *Diarpothe* aff. *acuta* strain IbSTRPmp18001 by spin column method. The results of full-length sequencing showed that the lengths of dsRNA1-4 were 3676 bp, 2683 bp, 2480 bp, and 1698 bp excluding 3' poly (A: U), respectively (Appendix 3.5.6-3.5.9). The sequences of DAV1 will be deposited in Genbank of NCBI.

Each four dsRNA segments have ORF (Figure 3-6B). ORF1 is predicted to be the viral RdRp, while ORF3 was shown to translate the 79.4 kDa viral major protein in the following experiment, which should be the viral coat protein. ORF2 and ORF3 encode hypothetical proteins. The deduced amino acid sequence of ORF1 was close to the RdRp of members of Alternaviridae after BLAST search on NCBI, and the identities were about 30-40 %. The maximum-likelihood phylogenetic analysis depending on the RdRp amino acid sequences, also showed that this virus could be classified into Alternaviridae (Figure 3-7). Like other alternaviruses, motif VI of RdRp contains ADD instead of GDD (Figure 2-19). Due to the above characteristics, it was named “Diaporthe alternavirus 1 (DAV1)”.

3.3.4 5′ cap and 3′ poly (A: U) terminal structures of DAV1 dsRNAs

To confirm whether the dsRNA segments of this new alternavirus, DAV1, have same 5' cap and 3' poly (A: U) with those of AaV1, I performed the RNA dot blot assay as desribed in chapter 2, to isolate dsRNA molecules from viral particles of DAV1, ScV-L-A, and AaV1 (Figure 3-8A). And, the 5' cap was detected with anti-m7G-cap antibody. (Figure3-8B). This indicates that dsRNA segments from DAV1 virus particles have 5' cap structures.

The specific amplicons (dsRNA1: ~3.6 kbp; dsRNA2: ~2.6 kbp; dsRNA3: ~2.4 kbp; dsRNA4: ~1.6 kbp) were amplified from in four dsRNA segments of DAV1 by RT-PCR (Figure3-9A, Table 3-1), indicating that the 3' ends of the positive strands and 5' ends of the negative strands of DAV1 dsRNA segments have poly (A: U) sequences.

3.3.5 Virus particle and virus-associated proteins

I tried the 45% sucrose cushion method used for purification of AaV1 to purify DAV1. In the results of SDS-PAGE electrophoresis, two protein bands of similar size of about 80 kDa and one protein band of about 70 kDa were observed in pellet fraction (Figure 3-10A), and dsRNA segments of DAV1 were present (Figure 3-10B). In the results of SDS-PAGE electrophoresis of the two different purifications, the pellet fractions were observed to have three protein bands of the same size as described above (Figure 3-10C). The isometric particles (~35-40 nm in diameter) were observed in the resultant pellet fractions by TEM (Figure 3-10D).

I applied CsCl isopycnic gradient centrifugation to further purify the virus particles and viral proteins. Two protein bands of about 80 kDa (Figure 3-11A) and DAV1 dsRNA segments (Figure 3-11B) were observed in fractions 5, 6, and 7 of the CsCl isopycnic gradient centrifugation. In fractions 2 and 3, about 70 kDa protein bands were observed (Figure 3-11A), but no DAV1 dsRNA segments were present (Figure 3-11B). The isometric particles (~35-40 nm in diameter, buoyant density: 1.349-1.374 g/cm^3) were observed in the resultant pellet fractions by TEM (Figure 3-11C)

After SDS-PAGE electrophoresis, I performed in-gel digestion of the three protein bands above and analyzed them by LC-MS/MS. Protein band ② in Figure 3-12, corresponded to the DAV1 ORF3-encoded 79.4 kDa protein (Figure 3-13B). The protein bands ① and ② corresponded to the host proteins (Figure 3-13A and C).

3.4 Discussion

For the identification of species of *Diaporthe,* currently, most studies used ITS region, *CAL*, *HIS*, *TUB*, and *TEF* genes for multi-locus phylogenetic analyses and then delimited species according to the Genealogical Concordance Phylogenetic Species

Recognition (GCPSR) principle (Santos *et al*., 2017). However, many studies have used only a few strains and minor differences between these five regions to delimit the species, which is not in accordance with the principle of GCPSR that requires the use of representative and extensive sampling to obtain reliable result (Taylor *et al*., 2000). For example, Hilário *et al*. (2021) demonstrated that several "new species" found in the *Diaporthe amygdali* "species complex" in recent years with multi-locus phylogenetic analyses were actually all the same species. The reason for this overestimation of the number of species is due to inappropriate sampling, failing to notice inconsistencies among analyses, and incorrect interpretation of the results. Although IbSTRPmp18001 isolated in this study formed a sister group with *D. acuta* in multi-locus phylogenetic analyses (Figure 3-1F), it did not form a sister group with *D. acuta* in the ITS (Figure 3-1A) and *HIS* (Figure 3-1C) single-region phylogenetic analyses. This inconsistency suggests that this group may require further studies to revalidate the taxonomic status. However, this is beyond the scope of this study, so it is tentatively named as *Diaporthe* aff. *acuta*.

I identified a novel alternavirus, DAV1, in *Diaporthe* sp. Four dsRNA segments are packaged in an isometric particle, consisted of ORF-encoded 79.4 kDa proteins. The DAV1-free isolate obtained in this study was not observed to be different from the DAV1-infected strain in terms of phenotype. The DAV1-free isolates obtained in this study were not observed to be different from the DAV1-infected isolates (Figure 3-1), and whether the DAV1-free isolates have other characteristics, such as drug resistance, needs to be further investigated.

Through RNA dot blot with anti-m^{7}G antibody (Figure 3-8) and specific RT-PCR (Figure 3-9), I confirmed that the dsRNA segments of DAV1 have 5'-m^{7}G caps and 3' poly (A: U) structures similar with those of AaV1 or FsAV1(Aoki *et al*., 2009; Wu *et al*., 2021; Lutz *et al*., 2022). These characteristics are as a member of Alternaviridae.

Unlike AaV1 (section 2.3.3; Wu *et al*., 2021), I did not observe a specific deletion of DAV1 dsRNA segments. In the CsCl isopycnic gradient centrifugation, four dsRNA segments of DAV1 coexisted with the major viral coat protein (Figure 3-1). In contrast to AaV1, which has empty particles, dsRNA segments are packaged in different particles (Figure 2-11; Aoki *et al*., 2009). The interaction between dsRNA genome and virus protein should be further investigated.

According to the RdRp sequence analysis, DAV1 is a new alternavirus and has unique properties of dsRNA genomes that help to establish the classification criteria of Alternaviridae. In the future, the molecular properties of the dsRNA genome and proteins of alternvirus can be further explored as mentioned in chapter 2.

Table 3-1. List of primers used in chapter 3.

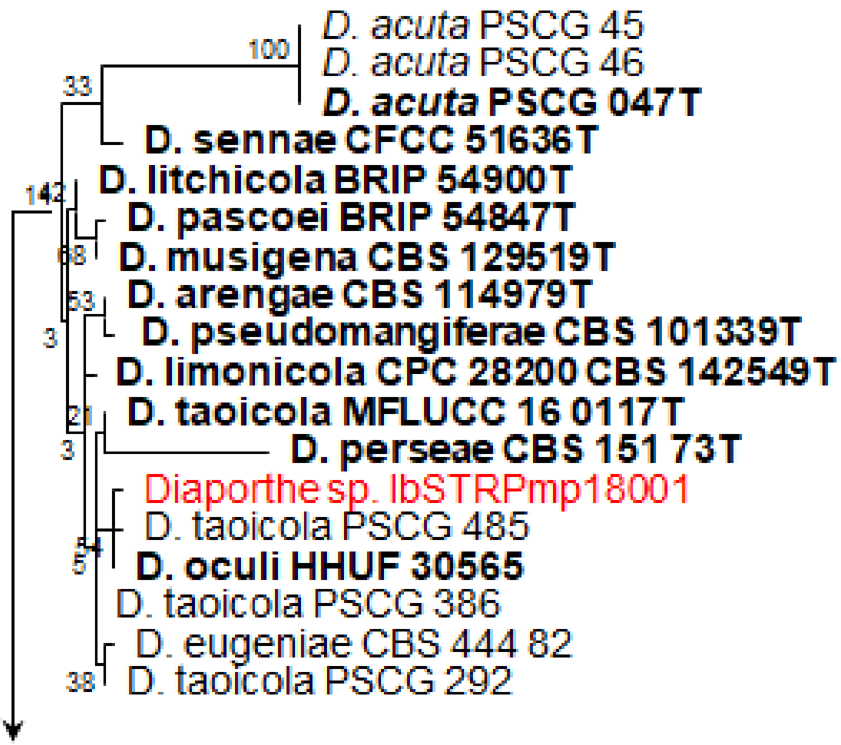
Primer name	Sequence	Purpose
ITS1	5’-TCCGTAGGTGAACCTGCGG	ITS region sequence cloning
ITS4	5’-TCCTCCGCTTATTGATATGC	
EF1-728F	5'-CATCGAGAAGTTCGAGAAGG	*TEF* gene cloning
EF1-986R	5'-TACTTGAAGGAACCCTTACC	
T1	5’-AACATGCGTGAGATTGTAAGT	*TUB* gene cloning
Bt2b	5’-ACCCTCAGTGTAGTGACCCTTGGC	
Cal228-F	5’-FGAGTTCAAGGAGGCCTTCTCCC	*CAL* gene cloning
CAL2Rd-R	5’-TGRTCNGCCTCDCGGATCATCTC	
CYLH3F-F	5'-AGGTCCACTGGTGGCAAG	*HIS* gene cloning
H3-1b-R	5’-GCGGGCGAGCTGGATGTCCTT	
DAV1-ORF1-F	5'-AAGTTAACATGAGGCTCGAGGAGGTGG	DAV1 RT-PCR detection
DAV1-ORF1-R	5'- CCGTTAACCTAGTGAATAAACAGCTGAG	
DAV1-ORF2-F	5'- CCGAATTCATGGAGACTGTCAAGATACC	
DAV1-ORF2-R	5'- CCGAATTCTCAACCTCGAAAGAAACCAG	
DAV1-ORF3-F	5’-CCGAATTCATGAAAATGGCGACCGACG	
DAV1-ORF3-R	5’-ACGAATTCCTAGTACGCGTTGACCGTG	
DAV1-ORF4-F	5’-CCGAATTCATGTCTCGGTCATTCGGAC	
DAV1-ORF4-R	5’-TCGAATTCCTAAGCAGGAACCACCACG	
DAV1-ORF1-F	5'-AAGTTAACATGAGGCTCGAGGAGGTGG	poly (A:U) detection
DAV1-ORF2-F	5'- CCGAATTCATGGAGACTGTCAAGATACC	
DAV1-ORF3-F	5’-CCGAATTCATGAAAATGGCGACCGACG	
DAV1-ORF4-F	5’-CCGAATTCATGTCTCGGTCATTCGGAC	
oligo dT (20)	oligo dT (20)	

Table 3-2. Sequences retrieved from GenBank and used in the phylogenetic analyses of *Diaporthe* spp.

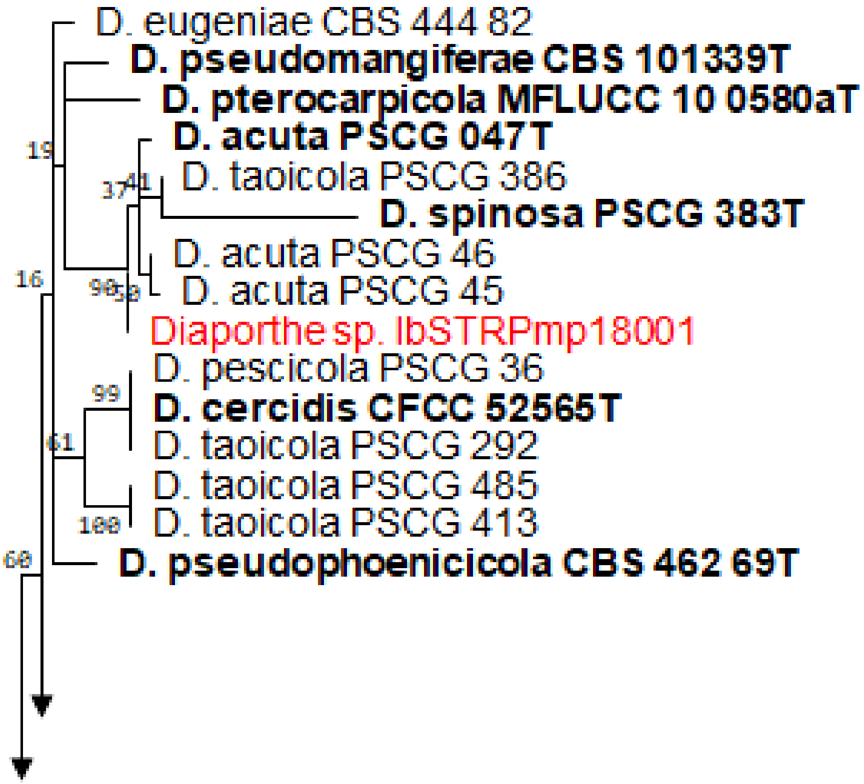
Species	Strain	Origin	ITS	*CAL*	*HIS*	*TEF*	*TUB*
D. acuta	PSCG 047^{T}	Pyrus pyrifolia	MK626957	MK691125	MK726161	MK654802	MK691225
D. acuta	PSCG 045	Pyrus pyrifolia	MK626956	MK691123	MK726160	MK654809	MK691223
D. acuta	PSCG 046	Pyrus pyrifolia	MK626958	MK691124	MK726162	MK654803	MK691224
D. arecae	CBS 161.64	Areca catechu	KC343032	KC343274	KC343516	KC343758	KC344000
D. arengae	CBS 114979^{T}	Arenga engleri	KC343034	KC343276	KC343518	KC343760	KC344002
D. cercidis	CFCC 52565^{T}	Cercis chinensis	MH121500	MH121424	MH121460	MH121542	MH121582
D. cercidis	PSCG 259	Pyrus pyrifolia	MK626847	MK691170	MK726154	MK654795	MK691218
D. eugeniae	CBS 444.82	Eugenia aromatica	KC343098	KC343340	KC343582	KC343824	KC344066
D. fraxini-angustifoliae	MFLUCC 15-0748	Vitis vinifera	KT459428	KT459462	–	KT459446	–
D. fulvicolor	PSCG 051^{T}	Pyrus pyrifolia	MK626859	MK691132	MK726163	MK654806	MK691236
D. hongkongensis	CBS 115448^{T}	Dichroa febr?fuga	KC343119	KC343361	KC343603	KC343845	KC344087
D. hongkongensis	PSCG 001	Pyrus pyrifolia	MK626846	MK691103	MK726150	MK654788	MK691240
D. limonicola	CPC 28200 = CBS 142549^{T}	Citrus limon	MF418422	MF418256	MF418342	MF418501	MF418582
D. litchicola	BRIP 54900^{T}	Litchi chinensis	JX862533	–	–	JX862539	KF170925
D. lithocarpus	CGMCC 3.15175^{T}	Lithocarpus glabra	KC153104	KF576235	–	KC153095	KF576311
D. melitensis	CPC 27873 = CBS 142551	Citrus limon	MF418424	MF418258	MF418344	MF418503	MF418584
D. musigena	CBS 129519^{T}	Musa sp.	KC343143	KC343385	KC343627	KC343869	KC344111
D. pascoei	BRIP 54847^{T}	Persea americana	JX862532	–	–	JX862538	KF170924
D. perseae	CBS 151.73^{T}	Persea gratissima	KC343173	KC343415	KC343657	KC343899	KC344141
D. pescicola	MFLUCC 16-0105^{T}	Prunus persica	KU557555	KU557603	–	KU557623	KU557579
D. pescicola	PSCG 036	Pyrus bretschneideri	MK626855	MK691116	MK726159	MK654796	MK691226
D. podocarpi-macrophylli	LC6200	Podocarpus macrophyllus	KX986769	KX999276	KX999240	KX999161	KX999201
D. pseudomangiferae	CBS 101339^{T}	Mangifera indica	KC343181	KC343423	KC343665	KC343907	KC344149
D. pseudophoenicicola	CBS 462.69^{T}	Phoenix dactylifera	KC343183	KC343425	KC343667	KC343909	KC344151
D. pterocarpicola	MFLUCC 10-0580a^{T}	Pterocarpus indicus	JQ619887	JX197433	–	JX275403	JX275441
D. sennae	CFCC 51636^{T}	Senna bicapsularis	KY203724	KY228875	KY228879	KY228885	KY228891
D. spinosa	PSCG 383^{T}	Pyrus pyrifolia	MK626849	MK691129	MK726156	MK654811	MK691234
D. taoicola	MFLUCC 16-0117^{T}	Prunus persica	KU557567	–	–	KU557635	KU557591
D. taoicola	PSCG 292	Pyrus pyrifolia	MK626871	MK691115	MK726168	MK654800	MK691232
D. taoicola	PSCG 386	Pyrus pyrifolia	MK626868	MK691122	MK726166	MK654797	MK691222
D. taoicola	PSCG 413	Pyrus pyrifolia	MK626890	MK691119	MK726167	MK654814	MK691238
D. taoicola	PSCG 485	Pyrus pyrifolia	MK626869	MK691120	MK726173	MK654812	MK691227
D. tectonigena	LC6512	Camellia sinensis	KX986782	KX999284	KX999254	KX999174	KX999215
D. virgiliae	CMW40748	Virgilia oroboides	KP247566	–	–	–	KP247575
D. oculi	HHUF:30565^{T}	Diseased human eye	LC373514	–	–	LC373516	LC373518
D. pseudooculi	HHUF:30617	Diseased human eye	LC373515	–	–	LC373517	LC373519
D. citri	CBS 135422^{T}	Citrus sp.	KC843311	KC843157	KJ490523	KC843071	KC843187

T: type strain

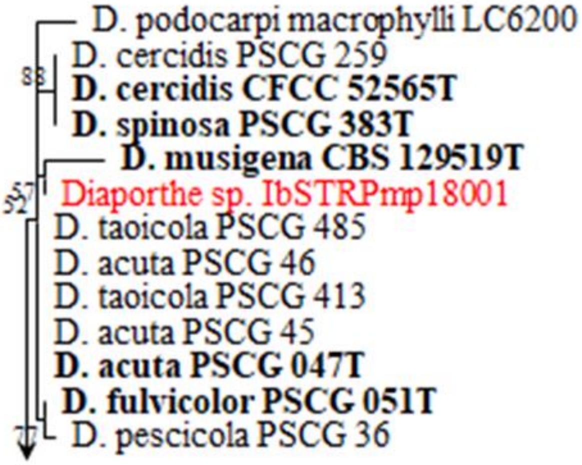
(A)


D. acuta PSCG 45
D. acuta PSCG 46
***D. acuta* PSCG 047T**
D. sennae CFCC 51636T
D. litchicola BRIP 54900T
D. pascoei BRIP 54847T
D. musigena CBS 129519T
D. arengae CBS 114979T
D. pseudomangiferae CBS 101339T
D. limonicola CPC 28200 CBS 142549T
D. taoicola MFLUCC 16 0117T
D. perseae CBS 151 73T
Diaporthe sp. IbSTRPmp18001
D. taoicola PSCG 485
D. oculi HHUF 30565
D. taoicola PSCG 386
D. eugeniae CBS 444 82
D. taoicola PSCG 292
100
33
142
68
53
3
21
3
54
5
38

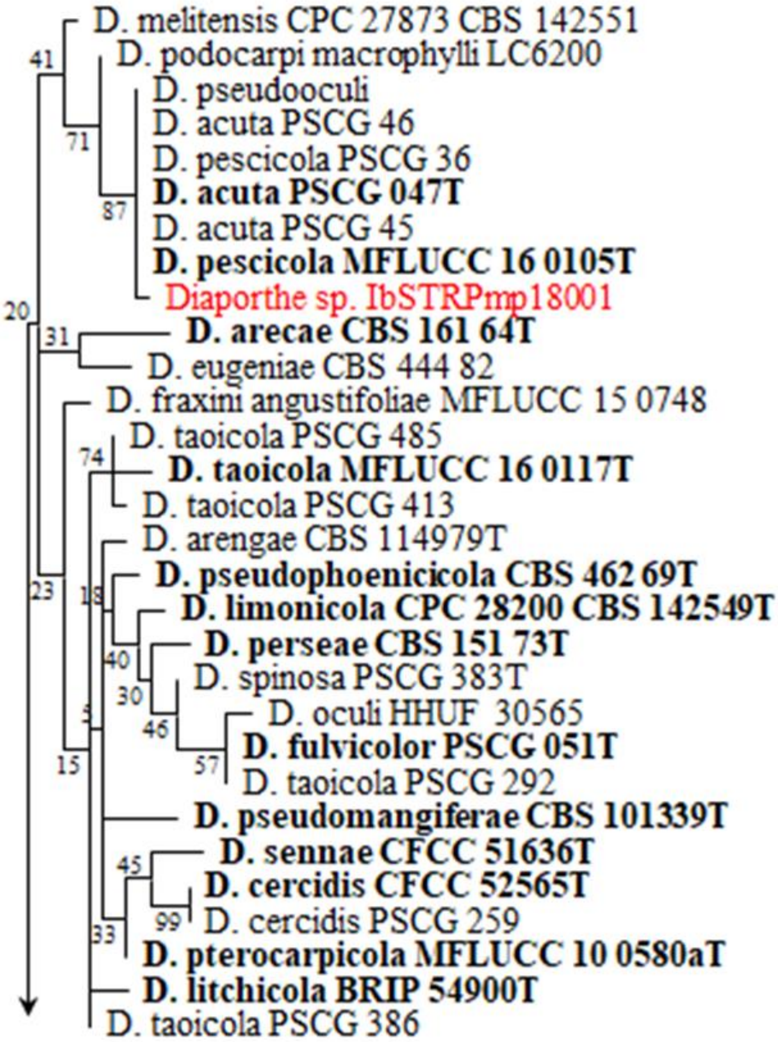
(B)


D. eugeniae CBS 444 82
D. pseudomangiferae CBS 101339T
D. pterocarpicola MFLUCC 10 0580aT
D. acuta PSCG 047T
D. taoicola PSCG 386
D. spinosa PSCG 383T
D. acuta PSCG 46
D. acuta PSCG 45
Diaporthe sp. IbSTRPmp18001
D. pescicola PSCG 36
D. cercidis CFCC 52565T
D. taoicola PSCG 292
D. taoicola PSCG 485
D. taoicola PSCG 413
D. pseudophoenicicola CBS 462 69T
19
37
41
16
90
58
99
61
100
60

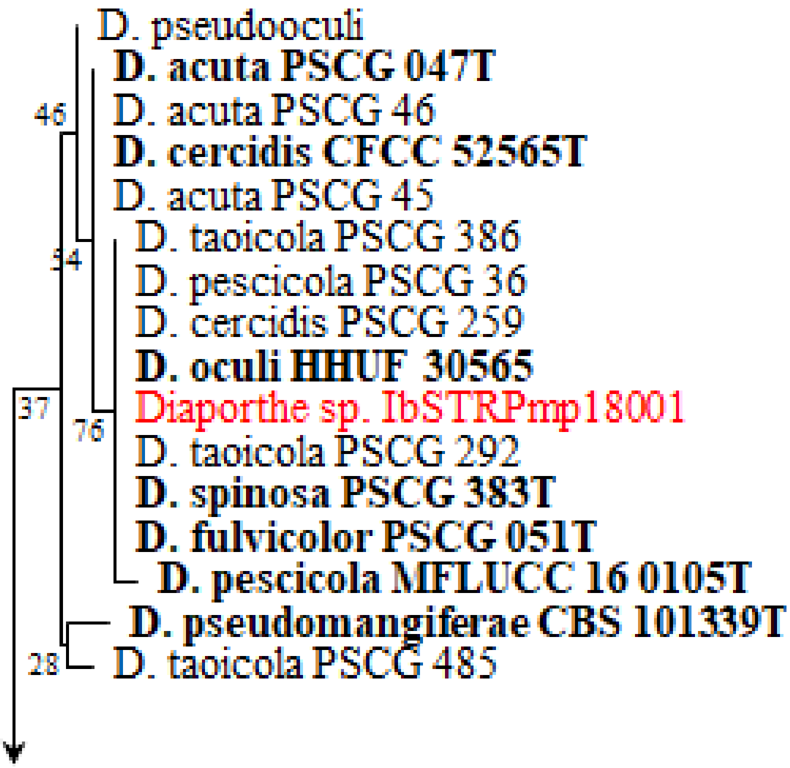
(C)


D. podocarpi macrophylli LC6200
D. cercidis PSCG 259
D. cercidis CFCC 52565T
D. spinosa PSCG 383T
D. musigena CBS 129519T
Diaporthe sp. IbSTRPmp18001
D. taoicola PSCG 485
D. acuta PSCG 46
D. taoicola PSCG 413
D. acuta PSCG 45
D. acuta PSCG 047T
D. fulvicolor PSCG 051T
D. pescicola PSCG 36
88
57
52
77

(D)


D. melitensis CPC 27873 CBS 142551
D. podocarpi macrophylli LC6200
D. pseudooculi
D. acuta PSCG 46
D. pescicola PSCG 36
D. acuta PSCG 047T
D. acuta PSCG 45
D. pescicola MFLUCC 16 0105T
Diaporthe sp. IbSTRPmp18001
D. arecae CBS 161 64T
D. eugeniae CBS 444 82
D. fraxini angustifoliae MFLUCC 15 0748
D. taoicola PSCG 485
D. taoicola MFLUCC 16 0117T
D. taoicola PSCG 413
D. arengae CBS 114979T
D. pseudophoenicicola CBS 462 69T
D. limonicola CPC 28200 CBS 142549T
D. perseae CBS 151 73T
D. spinosa PSCG 383T
D. oculi HHUF 30565
D. fulvicolor PSCG 051T
D. taoicola PSCG 292
D. pseudomangiferae CBS 101339T
D. sennae CFCC 51636T
D. cercidis CFCC 52565T
D. cercidis PSCG 259
D. pterocarpicola MFLUCC 10 0580aT
D. litchicola BRIP 54900T
D. taoicola PSCG 386
41
71
87
20
31
74
23
18
40
30
46
57
5
15
45
99
33

(E)


D. pseudooculi
D. acuta PSCG 047T
D. acuta PSCG 46
D. cercidis CFCC 52565T
D. acuta PSCG 45
D. taoicola PSCG 386
D. pescicola PSCG 36
D. cercidis PSCG 259
D. oculi HHUF 30565
Diaporthe sp. IbSTRPmp18001
D. taoicola PSCG 292
D. spinosa PSCG 383T
D. fulvicolor PSCG 051T
D. pescicola MFLUCC 16 0105T
D. pseudomangiferae CBS 101339T
D. taoicola PSCG 485
46
54
37
76
28

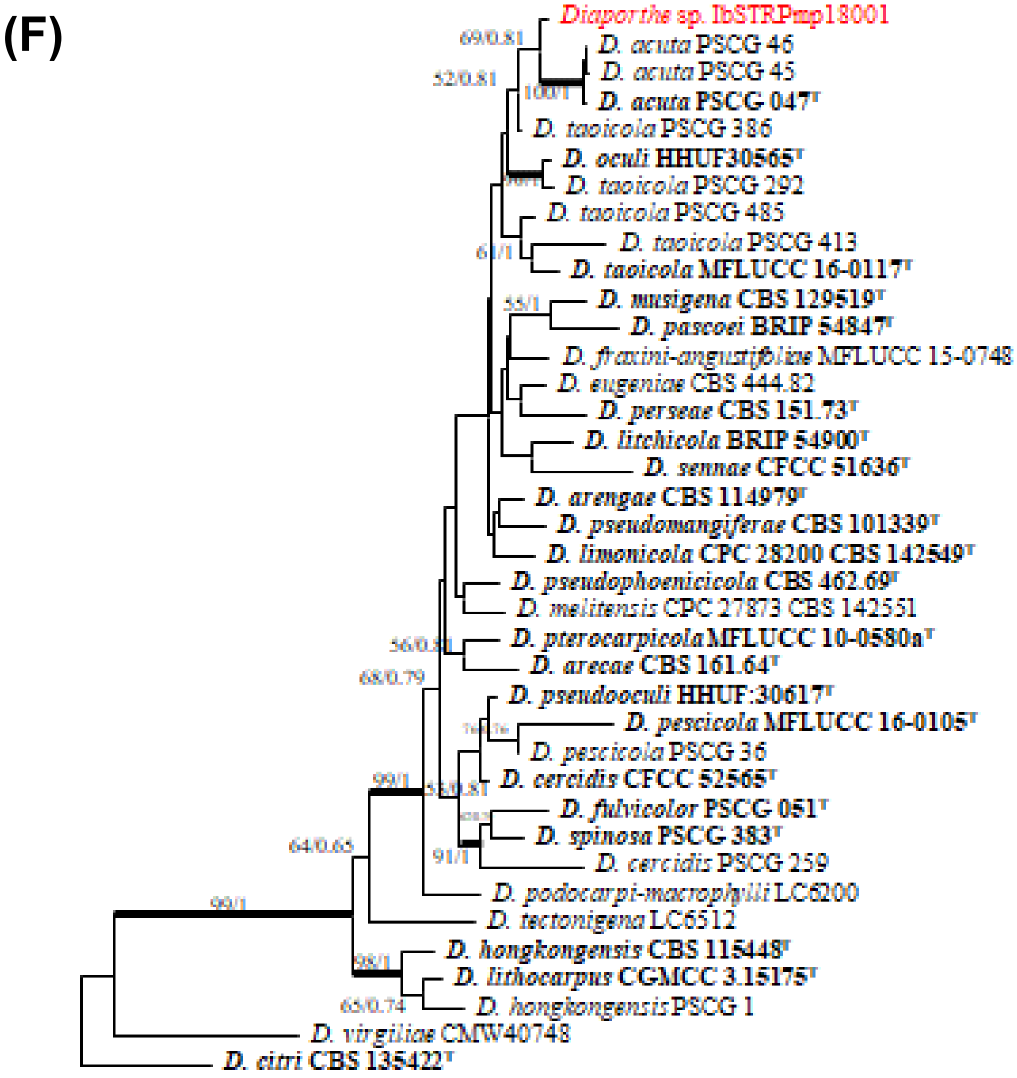


Figure 3-1. Phylogenetic analysis of *Diaporthe* spp.

Phylogenetic maximum-likelihood tree of **(A)** ITS region, **(B)** *CAL* gene, **(C)** *HIS* gene, **(D)** *TUB* gene, **(E)** *TEF* gene, **(F)** multi-locus analysis of *Diaporthe* spp.

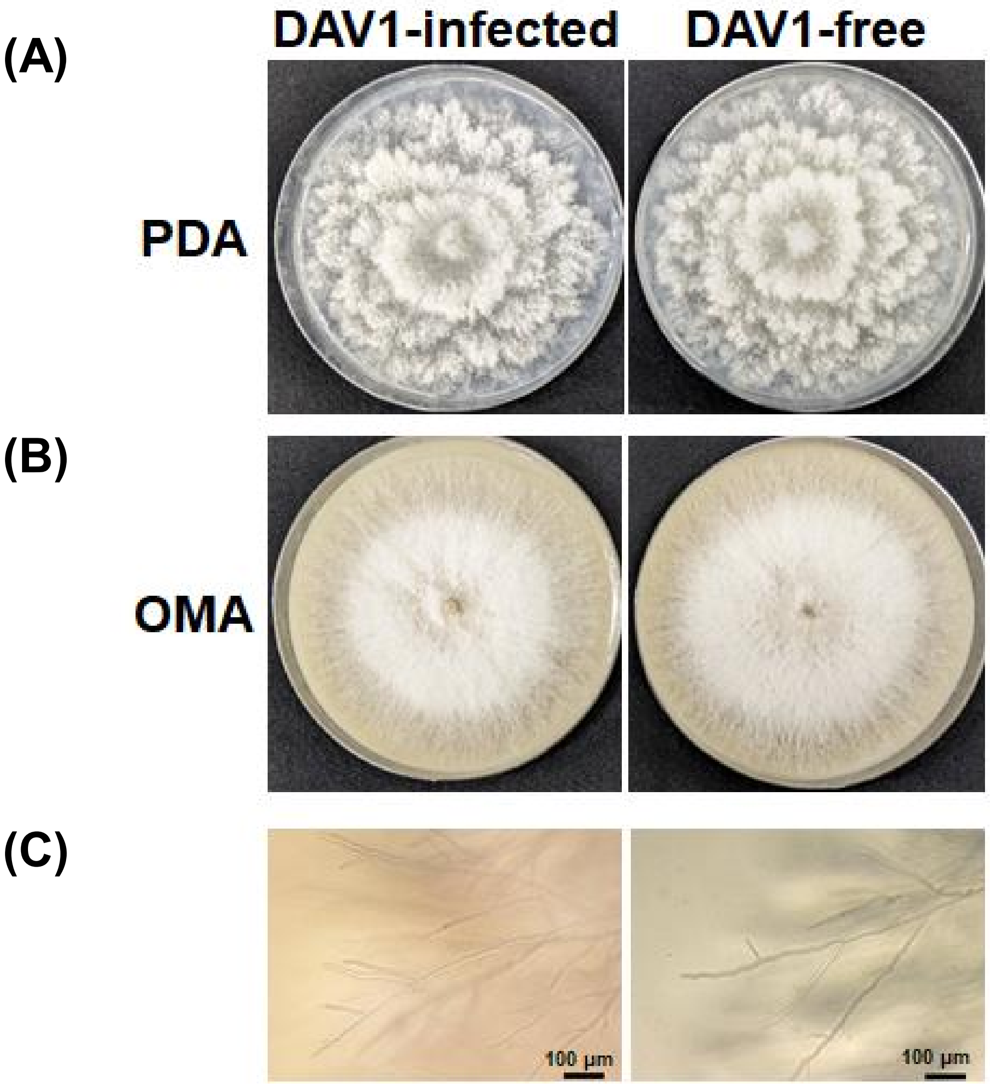


Figure 3-2. Colony morphologies of DAV1-infected isolates and DAV1-free isolate on different media.

(A) D-PDA, **(B)** OMA media, and **(C)** Mycelial morphology of DAV1-infected/-free isolates on D-PDA medium.

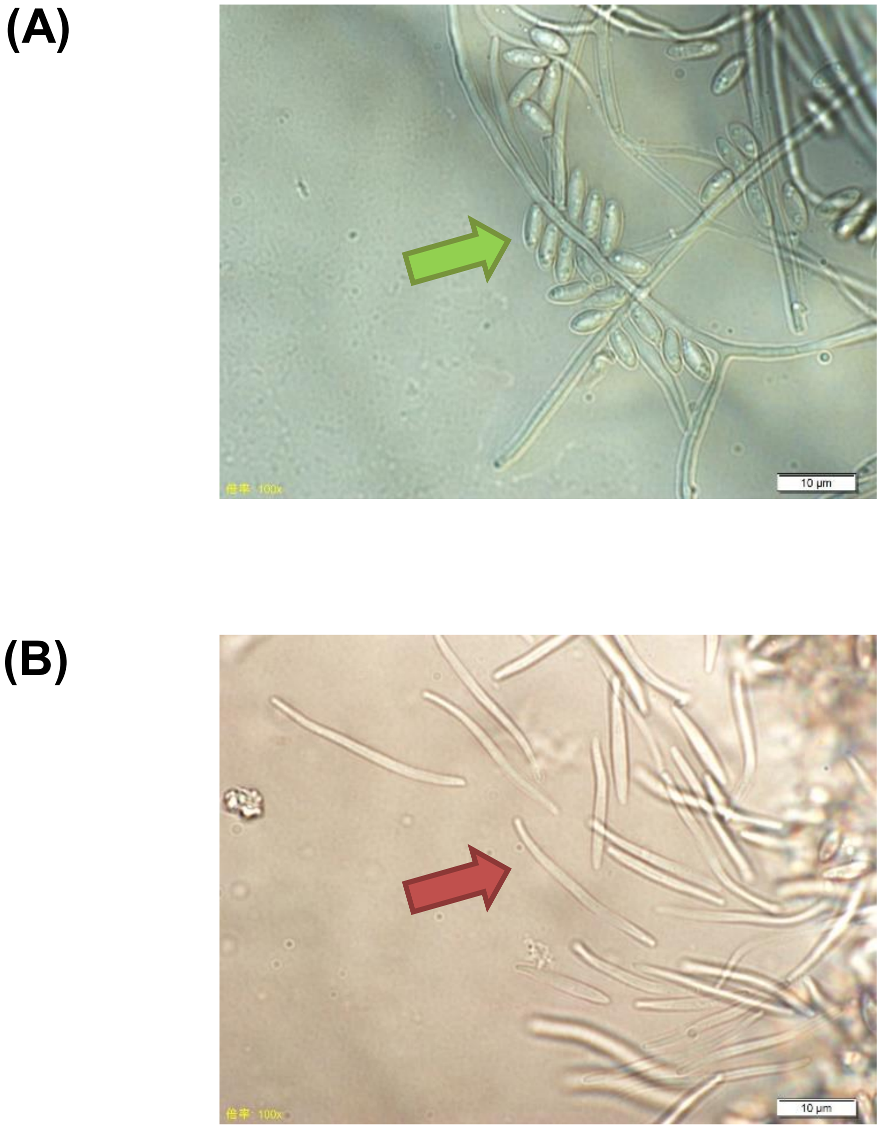


Figure 3-3 Conidia of *Diaporthe* aff. *acuta* strain IbSTR18001.

(A) Alpha-conidia **(B)** Beta-conidia.

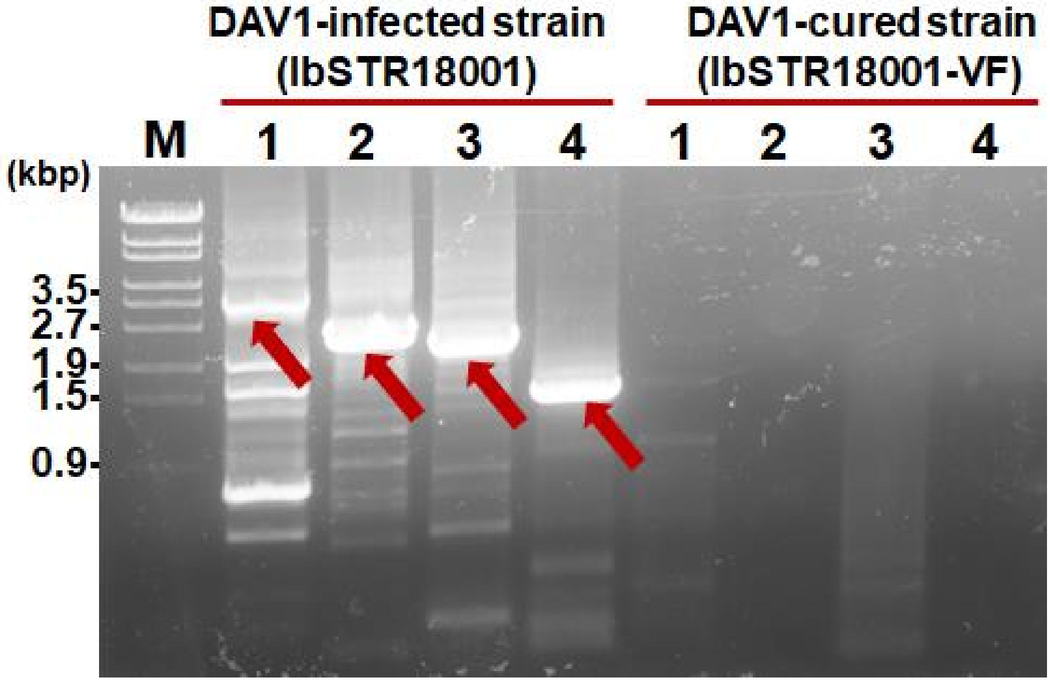


Figure 3-4. Curing of DAV1-infected isolate.

Confirmation of the existence of DAV1 using specific RT-PCR.

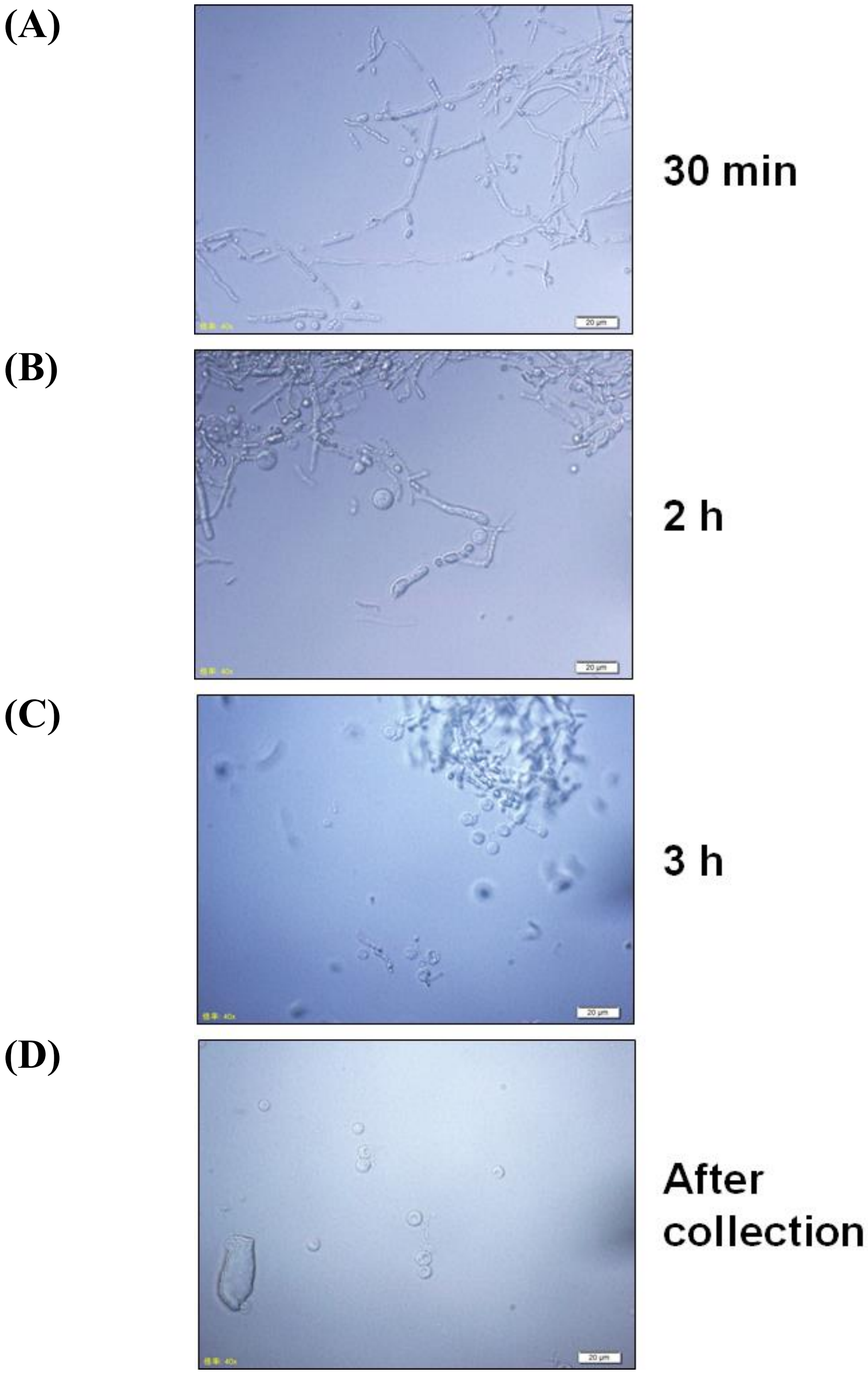


Figure 3-5. Curing of DAV1-infected strain by protoplast method.

The condition of mycelium, **(A)** 30 min **(B)** 2 h **(C)** 3h after treatment of lysing enzyme, and the resultant protoplasts **(D)**.

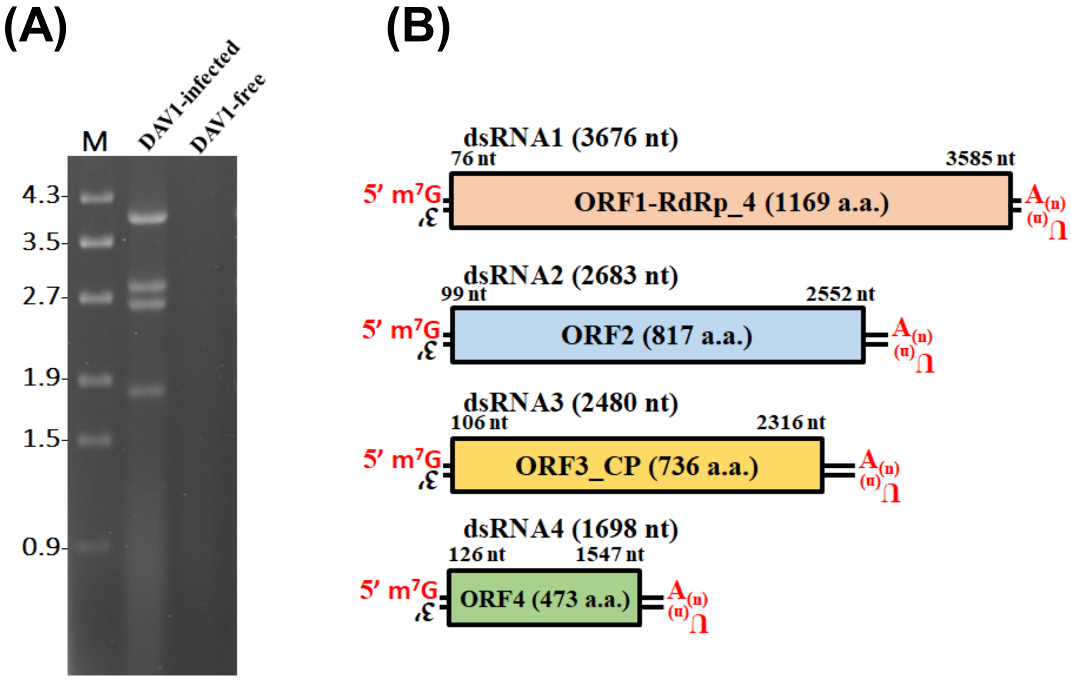


Figure 3-6. Genome of DAV1.

(A) Agarose gel electrophoresis of dsRNA genomes extracted from the mycelia of DAV1-infected/-free isolates. **(B)** Schematic diagram of DAV1 dsRNA segments.

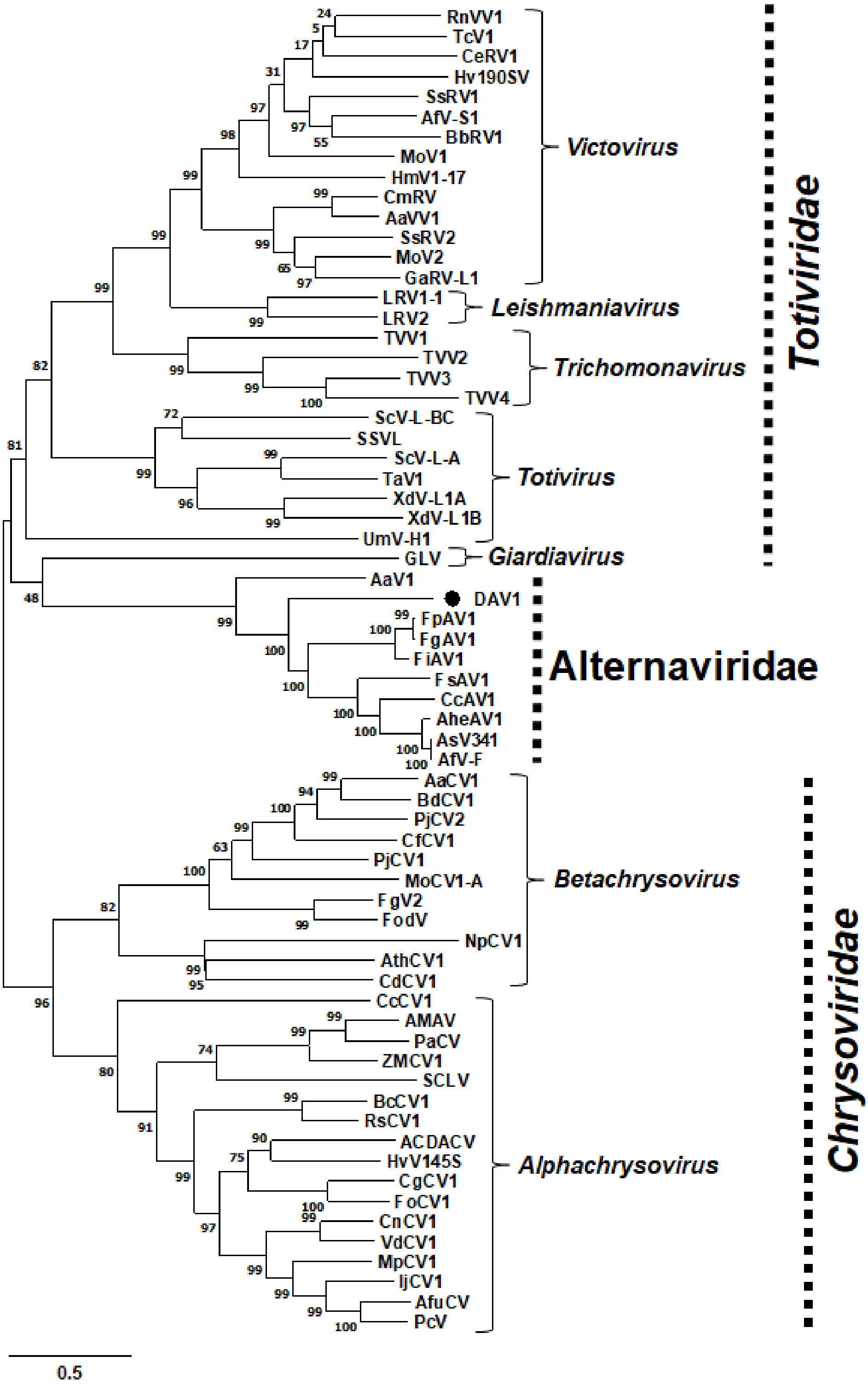


Figure 3-7. Phylogenetic analysis of DAV1.

Phylogenetic analysis of Alternaviridae, *Totiviridae,* and *Chrysoviridae* according to the amino acid sequences of RdRps.

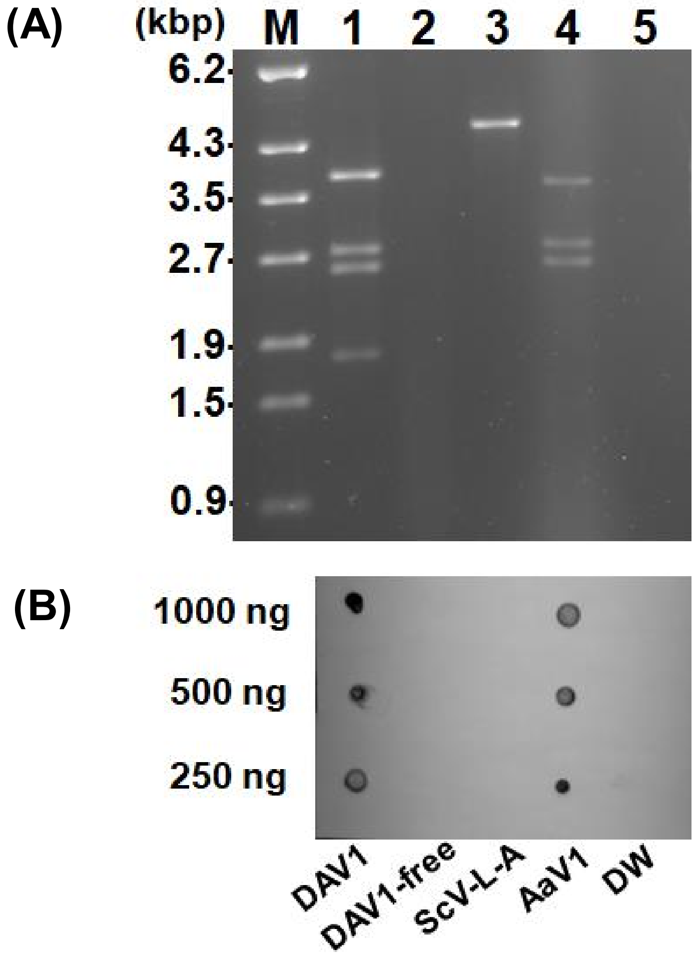


Figure 3-8. Detection of the 5′ cap structures in the DAV1 dsRNA genome by RNA dot blot with the anti-m^{7}G-Cap mAb.

(A) Viral dsRNAs from purified virus particles. The dsRNAs were isolated from virus particles, then electrophoresed in 1 % agarose gel with EtBr (0.5 μg/ml) at 18 V for 20 h. Lane designation: M, 250 ng of λ-EcoT14I-digested DNA marker; 1, DAV1 RNAs; 2, DAV1-free; 3, ScV-L-A dsRNA; 4, AaV1 dsRNAs; 5, DW (distilled water, no template control). **(B)** RNA dot blot assay with the anti-m^{7}G-Cap mAb. The DAV1 dsRNAs and the AaV1 dsRNAs (positive control) showed positive signals, while the DAVV1-free sample, the ScV-L-A dsRNA (negative control) and DW showed no signal.

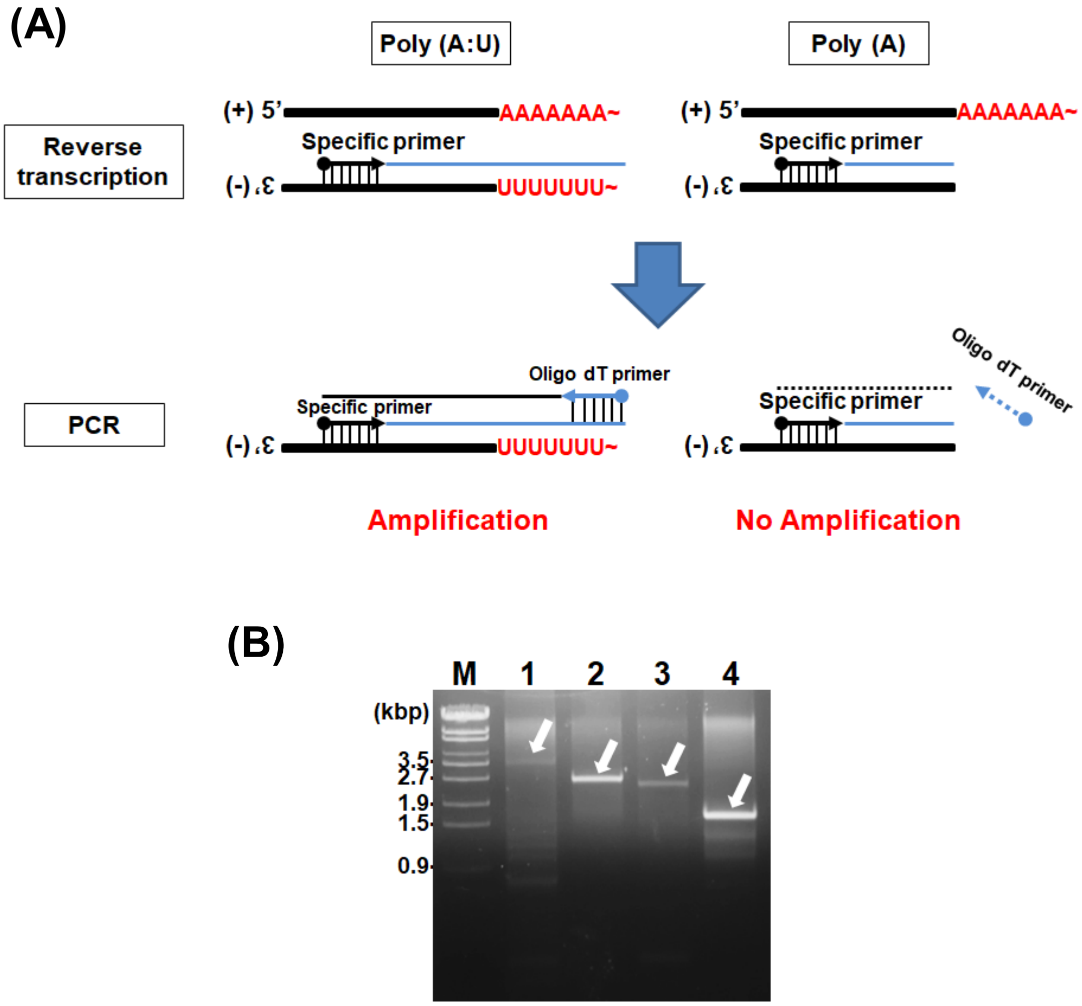


Figure 3-9. Confirmation of poly (A:U) in DAV1 dsRNA segments.

(A) Flowchart of poly (A:U) detection by specific RT-PCR. **(B)** Result of specific RT-PCR detection of poly (A:U) in DAV1 dsRNA segments.

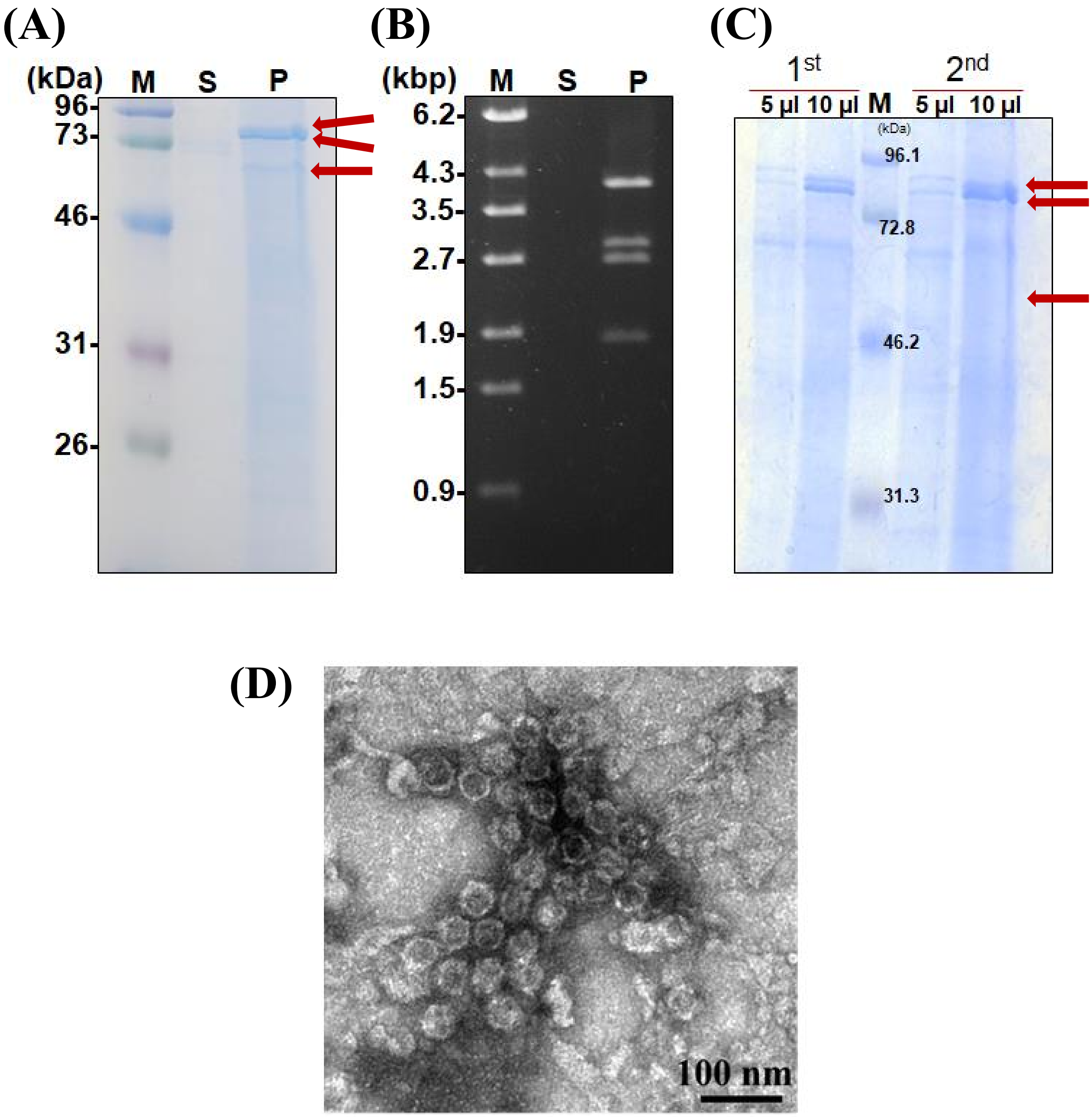


Figure 3-10. Purification of DAV1 virus particles with 45 % sucrose cushion.

(A) SDS-PAGE and **(B)** dsRNA agarose gel electrophoresis of DAV1 purification by 45 % sucrose cushion. S: Supernatant fraction; P: pellet fraction. **(C)** SDS-PAGE electrophoresis of two times of pellet fractions in two different purifications. **(D)** Purified virus particles were observed by TEM. The size of the particles are about 35-40 nm in diameter.

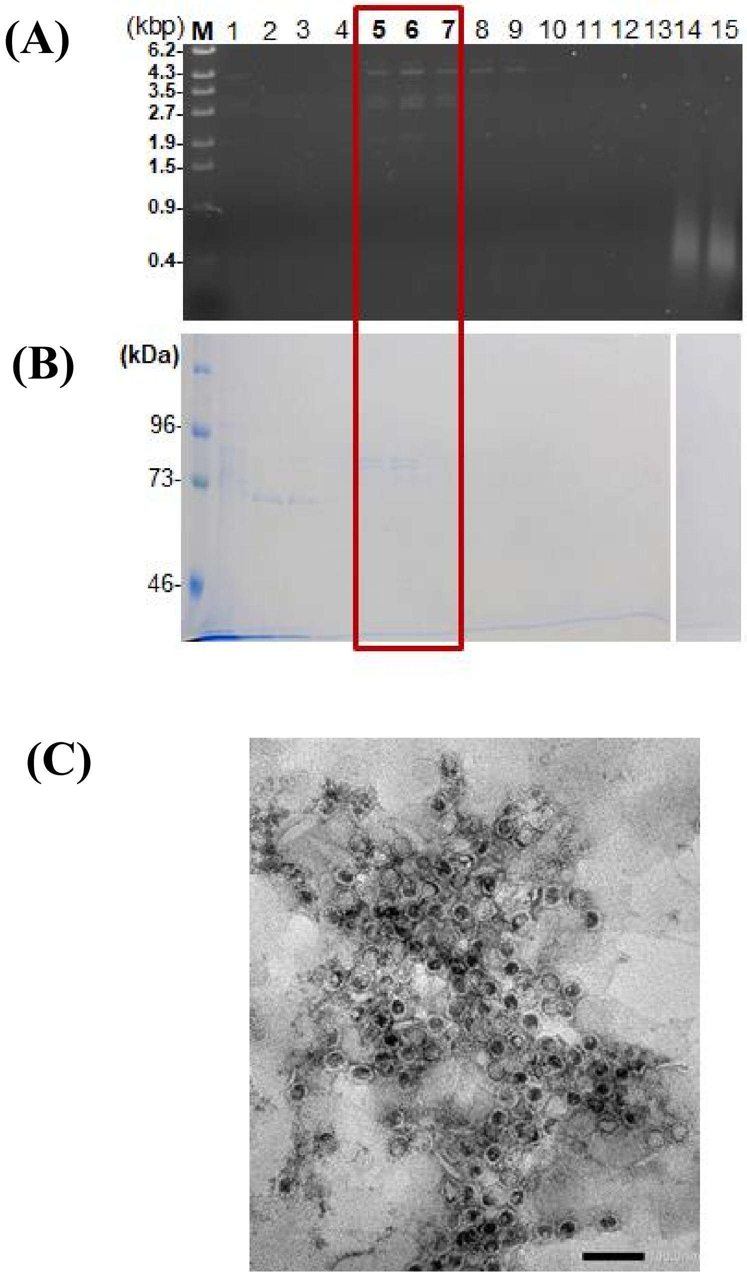


Figure 3-11. Purification of DAV1 virus particles with CsCl isopycnic gradient centrifugation.

(A) SDS-PAGE electrophoresis (8% SDS-PAGE, 120V, 150 min) and **(B)** dsRNA agarose electrophoresis of 15 fractions of CsCl isopycnic gradient centrifugation.

(C) Purified virus particles (mixture of fractions 5, 6, and 7) were observed by TEM. The size of purified particles is about 35-40 nm in diameter. Buoyant density: 1.349-1.374 g/cm^3. Scale bar=100 nm.

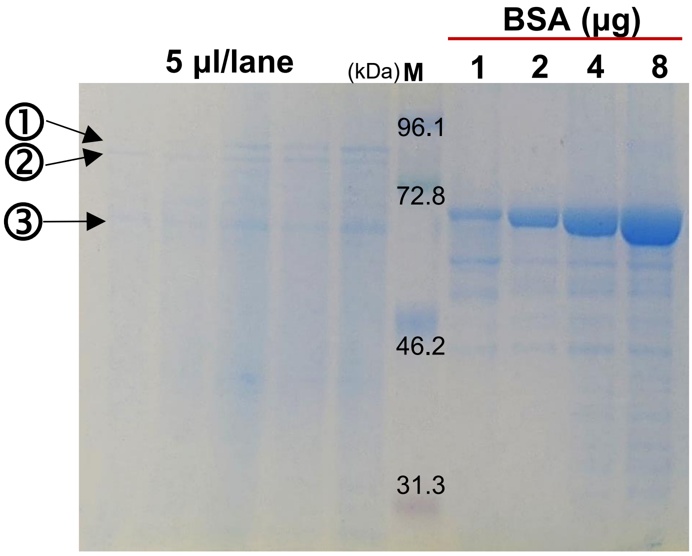
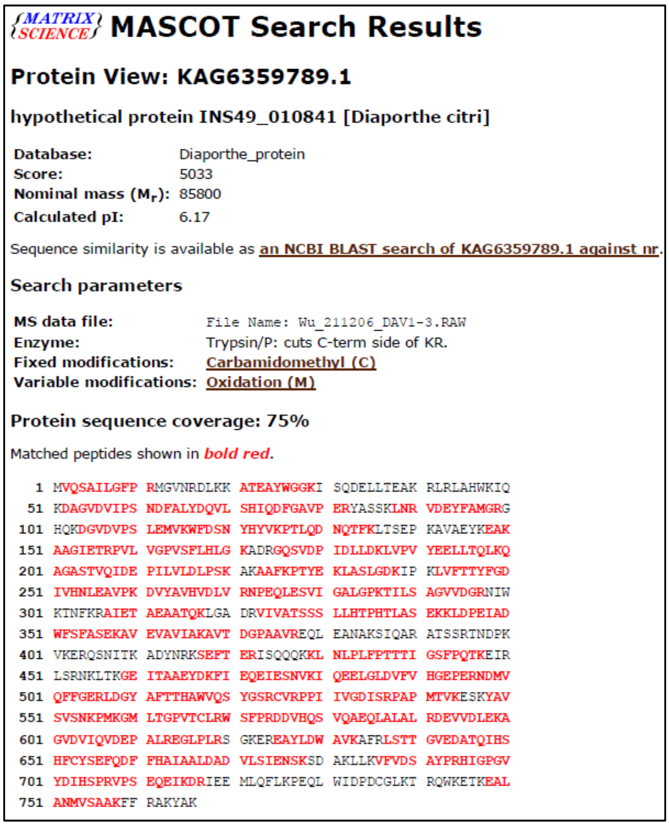


Figure 3-12. Estimation of DAV1 associated protein for LC-MS/MS.

Protein band①: 0.1 μg in total; Protein band②: 0.13 μg in total; Protein band③: 0.39 μg in total.

(A)



MASCOT Search Results

Protein View: KAG6359789.1

hypothetical protein INS49_010841 [Diaporthe citri]

Database: Diaporthe_protein
Score: 5033
Nominal mass (M_r): 85800
Calculated pI: 6.17

Sequence similarity is available as an NCBI BLAST search of KAG6359789.1 against nr.

Search parameters

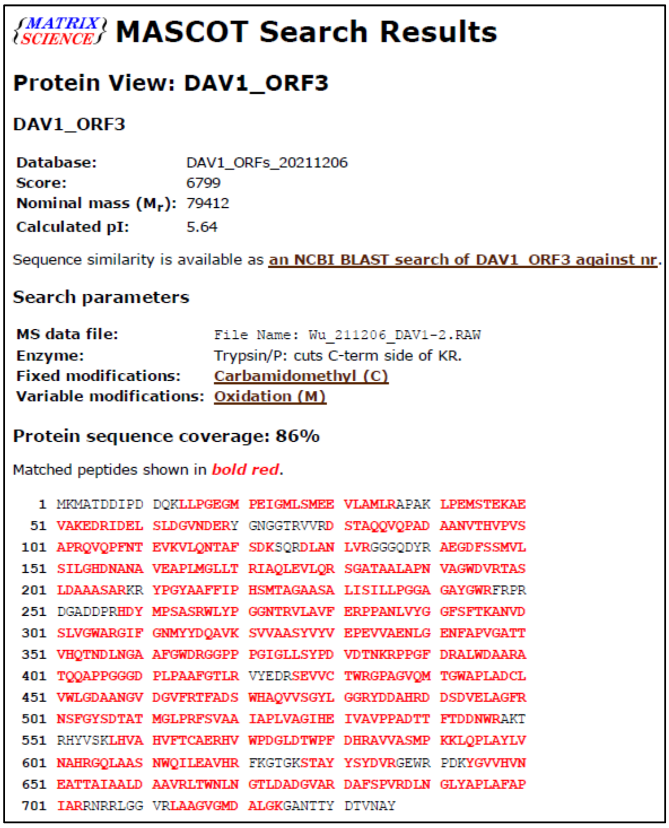
MS data file: File Name: Wu_211206_DAV1-3.RAW
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 75%

Matched peptides shown in ***bold red***.

```
  1 MVQSAILGFP RMGVNRDLKK ATEAYWGGKI SQDELLTEAK RLRLAHWKIQ
 51 KDAGVDVIPS NDFALYDQVL SHIQDFGAVP ERYASSKLNR VDEYFAMGRG
101 HQKDGVDVPS LEMVKWFDSN YHYVKPTLQD NQTFKLTSEP KAVAEYKEAK
151 AAGIETRPVL VGPVSFLHLG KADRGQSVDP IDLLDKLVPV YEELLTQLKQ
201 AGASTVQIDE PILVLDLPSK AKAAFKPTYE KLASLGDKIP KLVFTTYFGD
251 IVHNLEAVPK DVYAVHVDLV RNPEQLESVI GALGPKTILS AGVVDGRNIW
301 KTNFKRAIET AEAATQKLGA DRVIVATSSS LLHTPHTLAS EKKLDPEIAD
351 WFSFASEKAV EVAVIAKAVT DGPAAVREQL EANAKSIQAR ATSSRTNDPK
401 VKERQSNITK ADYNRKSEFT ERISQQQKKL NLPLFPTTTI GSFPQTKEIR
451 LSRNKLTKGE ITAAEYDKFI EQEIESNVKI QEELGLDVFV HGEPERNDMV
501 QFFGERLDGY AFTTHAWVQS YGSRCVRPPI IVGDISRPAP MTVKESKYAV
551 SVSNKPMKGM LTGPVTCLRW SFPRDDVHQS VQAEQLALAL RDEVVDLEKA
601 GVDVIQVDEP ALREGLPLRS GKEREAYLDW AVKAFRLSTT GVEDATQIHS
651 HFCYSEFQDF FHAIAALDAD VLSIENSKSD AKLLKVFVDS AYPRHIGPGV
701 YDIHSPRVPS EQEIKDRIEE MLQFLKPEQL WIDPDCGLKT RQWKETKEAL
751 ANMVSAAKFF RAKYAK
```

(B)



MASCOT Search Results

Protein View: DAV1_ORF3

DAV1_ORF3

Database: DAV1_ORFs_20211206
Score: 6799
Nominal mass (M_r): 79412
Calculated pI: 5.64

Sequence similarity is available as an NCBI BLAST search of DAV1_ORF3 against nr.

Search parameters

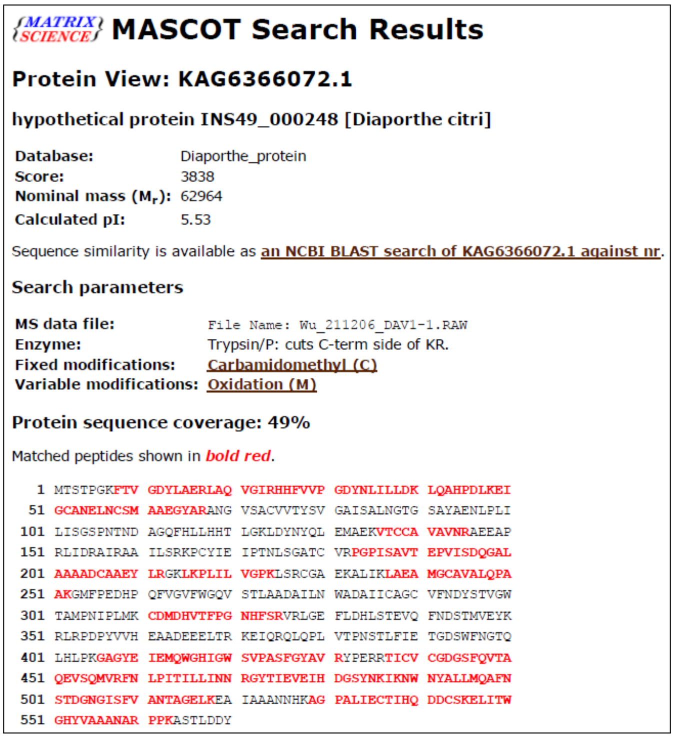
MS data file: File Name: Wu_211206_DAV1-2.RAW
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 86%

Matched peptides shown in ***bold red***.

```
  1 MKMATDDIPD DQKLLPGEGM PEIGMLSMEE VLAMLRAPAK LPEMSTEKAE
 51 VAKEDRIDEL SLDGVNDERY GNGGTRVVRD STAQQVQPAD AANVTHVPVS
101 APRQVQPFNT EVKVLQNTAF SDKSQRDLAN LVRGGGQDYR AEGDFSSMVL
151 SILGHDNANA VEAPLMGLLT RIAQLEVLQR SGATAALAPN VAGWDVRTAS
201 LDAAASARKR YPGYAAFFIP HSMTAGAASA LISILLPGGA GAYGWRFRPR
251 DGADDPRHDY MPSASRWLYP GGNTRVLAVF ERPPANLVYG GFSFTKANVD
301 SLVGWARGIF GNMYYDQAVK SVVAASYVYV EPEVVAENLG ENFAPVGATT
351 VHQTNDLNGA AFGWDRGGPP PGIGLLSYPD VDTNKRPPGF DRALWDAARA
401 TQQAPPGGGD PLPAAFGTLR VYEDRSEVVC TWRGPAGVQM TGWAPLADCL
451 VWLGDAANGV DGVFRTFADS WHAQVVSGYL GGRYDDAHRD DSDVELAGFR
501 NSFGYSDTAT MGLPRFSVAA IAPLVAGIHE IVAVPPADTT FTDDNWRAKT
551 RHYVSKLHVA HVFTCAERHV WPDGLDTWPF DHRAVVASMP KKLQPLAYLV
601 NAHRGQLAAS NWQILEAVHR FKGTGKSTAY YSYDVRGEWR PDKYGVVHVN
651 EATTAIAALD AAVRLTWNLN GTLDADGVAR DAFSPVRDLN GLYAPLAFAP
701 IARRNRRLGG VRLAAGVGMD ALGKGANTTY DTVNAY
```

(C)



MASCOT Search Results

Protein View: KAG6366072.1

hypothetical protein INS49_000248 [Diaporthe citri]

Database: Diaporthe_protein
Score: 3838
Nominal mass (M_r): 62964
Calculated pI: 5.53

Sequence similarity is available as an NCBI BLAST search of KAG6366072.1 against nr.

Search parameters

MS data file: `File Name: Wu_211206_DAV1-1.RAW`
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 49%

Matched peptides shown in ***bold red***.

```
  1 MTSTPGKFTV GDYLAERLAQ VGIRHHFVVP GDYNLILLDK LQAHPDLKEI
 51 GCANELNCSM AAEGYARANG VSACVVTYSV GAISALNGTG SAYAENLPLI
101 LISGSPNTND AGQFHLLHHT LGKLDYNYQL EMAEKVTCCA VAVNRAEEAP
151 RLIDRAIRAA ILSRKPCYIE IPTNLSGATC VRPGPISAVT EPVISDQGAL
201 AAAADCAAEY LRGKLKPLIL VGPKLSRCGA EKALIKLAEA MGCAVALQPA
251 AKGMFPEDHP QFVGVFWGQV STLAADAILN WADAIICAGC VFNDYSTVGW
301 TAMPNIPLMK CDMDHVTFPG NHFSRVRLGE FLDHLSTEVQ FNDSTMVEYK
351 RLRPDPYVVH EAADEEELTR KEIQRQLQPL VTPNSTLFIE TGDSWFNGTQ
401 LHLPKGAGYE IEMQWGHIGW SVPASFGYAV RYPERRTICV CGDGSFQVTA
451 QEVSQMVRFN LPITILLINN RGYTIEVEIH DGSYNKIKNW NYALLMQAFN
501 STDGNGISFV ANTAGELKEA IAAANNHKAG PALIECTIHQ DDCSKELITW
551 GHYVAAANAR PPKASTLDDY
```

Figure 3-13. Results of LC-MS/MS of DAV1-associated proteins analysis.

(A) Protein band①: 85.8 kDa **(B)** Protein band②: 79.4 kDa **(C)** Protein band③: 63 kDa in Fig. 3-12. Protein band① and ③ are host proteins. Protein band② corresponds to DAV1 ORF3.

3.5 Appendix

3.5.1 *Diaporthe* aff. *acuta* ITS region sequence (LC663742)

CGAATTCGCCCTTTCCGTAGGTGAACCTGCGGAGGGATCATTGCTGGAAC
GCGCCCCAGGCGCACCCAGAAACCCTTTGTGAACTTATACCTTTTGTTGC
CTCGGCGCATGCTGGTCTCTAGTAGGCCCCTCACCCCGGTGAGGAGACGG
CACGCCGGCGGCCAAGTTAACTCTTGTTTTTACACTGAAACTCTGAGAAA
AAAACACAAATGAATCAAAACTTTCAACAACGGATCTCTTGGTTCTGGCA
TCGATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTC
AGTGAATCATCGAATCTTTGAACGCACATTGCGCCCTCTGGTATTCCGGAG
GGCATGCCTGTTCGAGCGTCATTTCAACCCTCAAGCATTGCTTGGTGTTGG
GGCACTGCTTTTTACCAAGCAGGCCCTGAAATCTAGTGGCGAGCTCGCCA
GGACCCCGAGCGCAGTAGTTAAACCCTCGCTTTGGAAGGCCCTGGCGGT
GCCCTGCCGTTAAACCCCCAACTTTTGAAAATTTGACCTCGGATCAGGTA
GGAATACCCGCTGAACTTAAGCATATCAATAAGCGGAGGAAAGGGCGAAT

3.5.2 *Diaporthe* aff. *acuta CAL* gene sequence (LC663743)

TGAGTTCAAGGAGGCCTTCTCCCTCTTTGTAAGTACCATCTCCAACCAGCGGACA
CGAGTTCAAGCTCCCCGCCCTCCTCTGCTGTTGCGCATGATGCTAACGGACCGTTT
TCGGTTTGTAGGATAAGGATGGCGATGGTTAGTGCGGCCGCCCTCGTTCTTCCTTC
TCACACACGCACGCGTCATGCTCGATCCTCCGCGACGGCCTGCGCGTGCCTAATTT
CCAACCAAGCGATTATCACATCTATTGCGAGTACCATGCTGAGATATGGCGTGCAG
GACAAATCACCACCAAGGAGCTCGGCACCGTCATGCGATCCCTGGGCCAGAACCC
TTCCGAGTCTGAGCTGCAAGATATGATTAACGAGGTCGACGCCGACAACAACGGC
ACCATTGACTTCCCTGGTACGTCCAGATATACGCCCATTGAGGTGGGGGGGGGGG
CGCGGCCTGCTGTTACCCCTCTGCCGCAGCCTTGAATCACGTACTGACGCGCCCA
ACCAGAGTTCCTGACCATGATGGCAAGGAAGATGAAGGACACCGACTCCGAAGA
GGAAATCCGCGAGGCCTTCAAGGTTCGATCTACTCTACTCATACTCTGCCCCAACG
CCATGGGAATGGGACGGGCTTCTAACAACATCACCAGGTCTTCGACCGCGACAAC
AACGGCTTCATCTCCGCGGCTGAGCTGCGTCACGTCATGACCTCCATCGGCGAGA
AGCTTACGGACGACGAGGTCGACGAGATGATCCGCGAGGCCGACCA

3.5.3 *Diaporthe* aff. *acuta HIS* gene sequence (LC663744)

AGGTCCACTGGTGGCAAGGCTCCCCGTAAGCAGCTCGCCTCCAAGGCTGCCCGCA
AGTCCGCGCCCTCCACCGGAGGTGTCAAGAAGCCTCACCGCTACAAGCCTGGTAC
CGTCGCTCTGCGTGAGATCCGTCGCTACCAGAAGAGCACTGAGCTGCTGATCCGC
AAGCTCCCCTTCCAGCGTCTGGTATGCCGCCGCCCCCGTCTTCTGAATGCCCACCC
GCGTACCTTTGCTGACCATCGCCTTCCACTTCCAGGTCCGTGAGATCGCCCAGGAC
TTCAAGTCCGACCTCCGCTTCCAGTCCTCCGCCATCGGTGCCCTGCAGGAGTCCG
TCGAGTCTTACCTCGTCTCCCTCTTCGAGGACACCAACCTGTGCGCCATCCACGCC
AAGCGTGTCACCATCCAGTCGGTACGTAACATATCTGCTTGACCCACCCACCCGCG
ACGCGCCTTGTGGCATACGCTAACAACCTCACAGAAGGACATCCAGCTCGCCCGC

3.5.4 *Diaporthe* aff. *acuta TUB* gene sequence (LC663745)

ATTCGCCCTTAACATGCGTGAGATTGTAAGTCGTCCTCGTCCCCaAAAAAACCAC
CACCGAGACGCGTCGTCGGCCCATGCTGCTTTCGCATCCTCTGCCCCTGAGCCTG
AGCCTACCCCACCATCGCGACCGCACCCACGGTCGGGCCTCAAAAACATCACCA
AACACCCTGGGAAAAGCACCCAAATGTCCTTGAAAAACGCGTCAGATTGCTAAC
GTGACCTTTTTCTCGACTACAGGTTCACCTTCAGACCGGCCAATGCGTAAGTTGC
TCCTGTCAACACCGACGGACCTTATCATCGCCACCCGTAGCTGACACGTTTCCCA
GGGTAACCAAATCGGTGCTGCTTTCTGGTGCGTCCACCACCGCGACGCTCGACG
CGCGACAACATGACCTCGAGCCATCGTTACTGACCTCGACTCTCAGGCAAACCAT
CTCTGGCGAGCACGGCCTCGACAGCAATGGCGTGTATGCACCTCCTATTTCCTGC
CTTCAGATCTCGTCCTCCCTGCCGGCTTGGCACTGACAATCACACAGTTACAACG
GCACTTCCGAGCTCCAGCTCGAGCGCATGAACGTCTACTTCAACGAGGTAAGTC
AACAGCCACGTCGTTGGTCATCTGCAGCACGATTTTCTGCCGTCGCCGGGGCCTT
GCTAACGCGTTATCGCCCAGGCCTCCGGCAACAAGTATGTCCCTCGCGCCGTCCT
CGTCGATCTCGAGCCCGGTACCATGGACGCCGTCCGTGCCGGTCCCTTTGGCCAG
CTCTTCCGCCCCGACAACTTCGTCTTCGGCCAGTCCGGTGCTGGAAACAACTGG
GCCAAGGGTCACTACACTGAGGGT

3.5.5 *Diaporthe* aff. *acuta TEF* gene sequence (LC663746)

TTCATCGAGAAGTTCGAGAAGGAAGGTTAGTTATGATCACCATCATAGAACATGC
CTCTCTGCCCTCAGCCTTTTGCCTCAAATCATCAGTGTGCGGCCGTTGGTACGGC
TTCGTCATACTGAGGGGCGCATTTTCACCCCTCGCTCTGGATTTTCCATTTTCAGT
GCGGGTGCGGGGTGCGCTTATCAGGGAGCTTATCTCCTACACCCAAAACCCTGTT
GCAACAACACACCCACCCTACCGCTACTTCATTCATCCTCAACCCCCATCGACAT
TCCTGAATCAATCAAGCAAACCGCCAGGAAATCATGCTGACTTTCTATCTACATA
GCCGCCGAGCTTGGTAAGGGTTCCTTCAAGTAAAGGGCGA

3.5.6 DAV1 dsRNA1 sequence

3676 nt excluding poly (A) tail

```
5'  GGCGAGCGAUCCGUUUUGAGGAGAGAGGAACCUGGAUUACCAGAUCAAUUUCCCCUGCAGACGUGACCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   70
3'  CCGCUCGCUAGGCAAAACUCCUCUCUCCUUGGACCUAAUGGUCUAGUUAAAGGGGACGUCUGCACUGGAC
1    Gly Glu Arg Ser Val Leu Arg Arg Glu Glu Pro Gly Leu Pro Asp Gln Phe Pro Leu Gln Thr . Pro
2     Ala Ser Asp Pro Phe . Gly Glu Arg Asn Leu Asp Tyr Gln Ile Asn Phe Pro Cys Arg Arg Asp Leu
3   Arg Arg Ala Ile Arg Phe Glu Glu Arg Gly Thr Trp Ile Thr Arg Ser Ile Ser Pro Ala Asp Val Thr Trp
o
5'  GUGAGAUGAGGCUCGAGGAGGUGGUGUUGCGGAACGCCGAGACGCUGGCGCGACGCAUUGGAGCGUGGGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   140
3'  CACUCUACUCCGAGCUCCUCCACCACAACGCCUUGCGGCUCUGCGACCGCGCUGCGUAACCUCGCACCCU
1   Gly Glu Met Arg Leu Glu Glu Val Val Leu Arg Asn Ala Glu Thr Leu Ala Arg Arg Ile Gly Ala Trp Glu
2    Val Arg . Gly Ser Arg Arg Trp Cys Cys Gly Thr Pro Arg Arg Trp Arg Asp Ala Leu Glu Arg Gly
3     . Asp Glu Ala Arg Gly Gly Gly Val Ala Glu Arg Arg Asp Ala Gly Ala Thr His Trp Ser Val Gly
o
5'  GCAGGAUGUGGGGCGCUGGAACGCAGUGCACUCAAAAGACAAUCGCUUCAGUAGUCUGUACUCGGACCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   210
3'  CGUCCUACACCCCGCGACCUUGCGUCACGUGAGUUUUCUGUUAGCGAAGUCAUCAGACAUGAGCCUGGAC
1     Gln Asp Val Gly Arg Trp Asn Ala Val His Ser Lys Asp Asn Arg Phe Ser Ser Leu Tyr Ser Asp Leu
2   Ser Arg Met Trp Gly Ala Gly Thr Gln Cys Thr Gln Lys Thr Ile Ala Ser Val Val Cys Thr Arg Thr Cys
3    Ala Gly Cys Gly Ala Leu Glu Arg Ser Ala Leu Lys Arg Gln Ser Leu Gln . Ser Val Leu Gly Pro
o
5'  CCAAGCGUCGAAAAAUGGUGGAGGCCAGAUGCAUUGUGGAUACCGGCACACAAACCACACGGUAUGCCUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   280
3'  GGUUCGCAGCUUUUUACCACCUCCGGUCUACGUAACACCUAUGGCCGUGUGUUUGGUGUGCCAUACGGAA
1    Pro Ser Val Glu Lys Trp Trp Arg Pro Asp Ala Leu Trp Ile Pro Ala His Lys Pro His Gly Met Pro
2     Gln Ala Ser Lys Asn Gly Gly Gly Gln Met His Cys Gly Tyr Arg His Thr Asn His Thr Val Cys Leu
3   Ala Lys Arg Arg Lys Met Val Glu Ala Arg Cys Ile Val Asp Thr Gly Thr Gln Thr Thr Arg Tyr Ala Phe
o
5'  UGCUUGCGGUCUCGGAUCACACAACUGGGAAGGAAUGGCUGGCGGAGCUUACCAGCCGGGGUGUGAAGUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   350
3'  ACGAACGCCAGAGCCUAGUGUGUUGACCCUUCCUUACCGACCGCCUCGAAUGGUCGGCCCCACACUUCAC
1   Leu Leu Ala Val Ser Asp His Thr Thr Gly Lys Glu Trp Leu Ala Glu Leu Thr Ser Arg Gly Val Lys Cys
2    Cys Leu Arg Ser Arg Ile Thr Gln Leu Gly Arg Asn Gly Trp Arg Ser Leu Pro Ala Gly Val . Ser
3     Ala Cys Gly Leu Gly Ser His Asn Trp Glu Gly Met Ala Gly Gly Ala Tyr Gln Pro Gly Cys Glu Val
o
5'  UGUCUUUGCGGAAAAUCUGGAUGAGGCGAGGAAGACGCUUUUCGAAUGUGUAACCAUCGCAAGGAUCGAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   420
3'  ACAGAAACGCCUUUUAGACCUACUCCGCUCCUUCUGCGAAAAGCUUACACAUUGGUAGCGUUCCUAGCUG
1     Val Phe Ala Glu Asn Leu Asp Glu Ala Arg Lys Thr Leu Phe Glu Cys Val Thr Ile Ala Arg Ile Asp
2   Val Ser Leu Arg Lys Ile Trp Met Arg Arg Gly Arg Arg Phe Ser Asn Val . Pro Ser Gln Gly Ser Thr
3    Cys Leu Cys Gly Lys Ser Gly . Gly Glu Glu Asp Ala Phe Arg Met Cys Asn His Arg Lys Asp Arg
o
5'  CGCGACAGAUCGAUACGCACAGUAUCAUCACCGACUCUAUGCCGCGAAUUGUGCGACUGGAUUCAGAUGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   490
3'  GCGCUGUCUAGCUAUGCGUGUCAUAGUAGUGGCUGAGAUACGGCGCUUAACACGCUGACCUAAGUCUACG
1    Arg Asp Arg Ser Ile Arg Thr Val Ser Ser Pro Thr Leu Cys Arg Glu Leu Cys Asp Trp Ile Gln Met
2     Ala Thr Asp Arg Tyr Ala Gln Tyr His His Arg Leu Tyr Ala Ala Asn Cys Ala Thr Gly Phe Arg Cys
3   Pro Arg Gln Ile Asp Thr His Ser Ile Ile Thr Asp Ser Met Pro Arg Ile Val Arg Leu Asp Ser Asp Ala
o
5'  CGACGUCUCUCGGCACUCCACAUUCUGAUGCGUAUGCAAAAUAUGUCAACUUUGAGCUCGGCAGGAAUGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   560
3'  GCUGCAGAGAGCCGUGAGGUGUAAGACUACGCAUACGUUUUAUACAGUUGAAACUCGAGCCGUCCUUACU
1   Pro Thr Ser Leu Gly Thr Pro His Ser Asp Ala Tyr Ala Lys Tyr Val Asn Phe Glu Leu Gly Arg Asn Glu
2    Arg Arg Leu Ser Ala Leu His Ile Leu Met Arg Met Gln Asn Met Ser Thr Leu Ser Ser Ala Gly Met
3     Asp Val Ser Arg His Ser Thr Phe . Cys Val Cys Lys Ile Cys Gln Leu . Ala Arg Gln Glu .
o
```

```
5'   GGGCAGUGGGUAAGGAGACAACCGCGAUGGUUUCGAAAUCGUGGUUCUUCCCACUCACCAAUUGCCAAAC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1190
3'   CCCGUCACCCAUUCCUCUGUUGGCGCUACCAAAGCUUUAGCACCAAGAAGGGUGAGUGGUUAACGGUUUG
1    Arg Ala Val Gly Lys Glu Thr Thr Ala Met Val Ser Lys Ser Trp Phe Phe Pro Leu Thr Asn Cys Gln Thr
2     Gly Gln Trp Val Arg Arg Gln Pro Arg Trp Phe Arg Asn Arg Gly Ser Ser His Ser Pro Ile Ala Lys
3      Gly Ser Gly . Gly Asp Asn Arg Asp Gly Phe Glu Ile Val Val Leu Pro Thr His Gln Leu Pro Asn
o
5'   GAUCGCUUCAGCAUACUUCCUGAACGUCGAGGAUUUGCUCGGCUAUCCGGAUUCGGCGACGGAGACAGGC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1260
3'   CUAGCGAAGUCGUAUGAAGGACUUGCAGCUCCUAAACGAGCCGAUAGGCCUAAGCCGCUGCCUCUGUCCG
1     Ile Ala Ser Ala Tyr Phe Leu Asn Val Glu Asp Leu Leu Gly Tyr Pro Asp Ser Ala Thr Glu Thr Gly
2    Arg Ser Leu Gln His Thr Ser . Thr Ser Arg Ile Cys Ser Ala Ile Arg Ile Arg Arg Arg Arg Gln Ala
3      Asp Arg Phe Ser Ile Leu Pro Glu Arg Arg Gly Phe Ala Arg Leu Ser Gly Phe Gly Asp Gly Asp Arg
o
5'   GCGGUGGCACUGGAAGUACUCGAUUUCGUCACAGACAAUUACGAGUACAGCUAUCCGACAGGUACGGGCU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1330
3'   CGCCACCGUGACCUUCAUGAGCUAAAGCAGUGUCUGUUAAUGCUCAUGUCGAUAGGCUGUCCAUGCCCGA
1     Ala Val Ala Leu Glu Val Leu Asp Phe Val Thr Asp Asn Tyr Glu Tyr Ser Tyr Pro Thr Gly Thr Gly
2      Arg Trp His Trp Lys Tyr Ser Ile Ser Ser Gln Thr Ile Thr Ser Thr Ala Ile Arg Gln Val Arg Ala
3    Arg Gly Gly Thr Gly Ser Thr Arg Phe Arg His Arg Gln Leu Arg Val Gln Leu Ser Asp Arg Tyr Gly Leu
o
5'   GUGAGAGGGACUUCGAACAGUACCGGCGGCAAUUCAUGGAUGGCGUGCUAUCACUACUGGAGCCGGUGUA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1400
3'   CACUCUCCCUGAAGCUUGUCAUGGCCGCCGUUAAGUACCUACCGCACGAUAGUGAUGACCUCGGCCACAU
1    Cys Glu Arg Asp Phe Glu Gln Tyr Arg Arg Gln Phe Met Asp Gly Val Leu Ser Leu Leu Glu Pro Val Tyr
2     Val Arg Gly Thr Ser Asn Ser Thr Gly Gly Asn Ser Trp Met Ala Cys Tyr His Tyr Trp Ser Arg Cys
3      . Glu Gly Leu Arg Thr Val Pro Ala Ala Ile His Gly Trp Arg Ala Ile Thr Thr Gly Ala Gly Val
o
5'   UCGCGAAUUCGGUGCUCGUGCCGAGACUUACGACGAUGUUUUGGAGAUGCGCGUGGCCAAGUUGGCCGGC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1470
3'   AGCGCUUAAGCCACGAGCACGGCUCUGAAUGCUGCUACAAAACCUCUACGCGCACCGGUUCAACCGGCCG
1     Arg Glu Phe Gly Ala Arg Ala Glu Thr Tyr Asp Asp Val Leu Glu Met Arg Val Ala Lys Leu Ala Gly
2    Ile Ala Asn Ser Val Leu Val Pro Arg Leu Thr Thr Met Phe Trp Arg Cys Ala Trp Pro Ser Trp Pro Ala
3      Ser Arg Ile Arg Cys Ser Cys Arg Asp Leu Arg Arg Cys Phe Gly Asp Ala Arg Gly Gln Val Gly Arg
o
5'   GGUGUUUCCGGGCGCGCUGCCCGUGAACGCUUCGGUGAGGACGUGGCACCUGCCGGCAGUACGAAGAAGU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1540
3'   CCACAAAGGCCCGCGCGACGGGCACUUGCGAAGCCACUCCUGCACCGUGGACGGCCGUCAUGCUUCUUCA
1     Gly Val Ser Gly Arg Ala Ala Arg Glu Arg Phe Gly Glu Asp Val Ala Pro Ala Gly Ser Thr Lys Lys
2      Val Phe Pro Gly Ala Leu Pro Val Asn Ala Ser Val Arg Thr Trp His Leu Pro Ala Val Arg Arg Ser
3    Arg Cys Phe Arg Ala Arg Cys Pro . Thr Leu Arg . Gly Arg Gly Thr Cys Arg Gln Tyr Glu Glu Val
o
5'   AUGUUGGCGCCCGUACCACGAAGGAAGUCUGGAAACAUAAGUGGGGCACUACAAUCGUCGAAGUUGCAGU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1610
3'   UACAACCGCGGGCAUGGUGCUUCCUUCAGACCUUUGUAUUCACCCCGUGAUGUUAGCAGCUUCAACGUCA
1    Tyr Val Gly Ala Arg Thr Thr Lys Glu Val Trp Lys His Lys Trp Gly Thr Thr Ile Val Glu Val Ala Val
2     Met Leu Ala Pro Val Pro Arg Arg Lys Ser Gly Asn Ile Ser Gly Ala Leu Gln Ser Ser Lys Leu Gln
3      Cys Trp Arg Pro Tyr His Glu Gly Ser Leu Glu Thr . Val Gly His Tyr Asn Arg Arg Ser Cys Ser
o
5'   CAAGGUCGAUGAGCGCGGAGGCGUGCGUACCAUAAUGGCCACAGAGAUGCGCGAUCAGUUGUCCGAAAUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1680
3'   GUUCCAGCUACUCGCGCCUCCGCACGCAUGGUAUUACCGGUGUCUCUACGCGCUAGUCAACAGGCUUUAC
1     Lys Val Asp Glu Arg Gly Gly Val Arg Thr Ile Met Ala Thr Glu Met Arg Asp Gln Leu Ser Glu Met
2    Ser Arg Ser Met Ser Ala Glu Ala Cys Val Pro . Trp Pro Gln Arg Cys Ala Ile Ser Cys Pro Lys Trp
3      Gln Gly Arg . Ala Arg Arg Arg Ala Tyr His Asn Gly His Arg Asp Ala Arg Ser Val Val Arg Asn
o
```

```
5'  GUCGCCAUGAAGCCGAUAACCAAUCGGCUUUCGGCAGUUGGUAUGGACGUGGGGGAAUCACCCAUCCAGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1750
3'  CAGCGGUACUUCGGCUAUUGGUUAGCCGAAAGCCGUCAACCAUACCUGCACCCCCUUAGUGGGUAGGUCC
1    Val Ala Met Lys Pro Ile Thr Asn Arg Leu Ser Ala Val Gly Met Asp Val Gly Glu Ser Pro Ile Gln
2     Ser Pro  .  Ser Arg  .  Pro Ile Gly Phe Arg Gln Leu Val Trp Thr Trp Gly Asn His Pro Ser Arg
3   Gly Arg His Glu Ala Asp Asn Gln Ser Ala Phe Gly Ser Trp Tyr Gly Arg Gly Gly Ile Thr His Pro Gly
o
5'  CAAUGGAACGGCAUCUCAGGACGGUUGGGAUGGCCGACGACUACGAUGGUUACAGAUUCGAUGGGCAGGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1820
3'  GUUACCUUGCCGUAGAGUCCUGCCAACCCUACCGGCUGCUGAUGCUACCAAUGUCUAAGCUACCCGUCCA
1   Ala Met Glu Arg His Leu Arg Thr Val Gly Met Ala Asp Asp Tyr Asp Gly Tyr Arg Phe Asp Gly Gln Val
2    Gln Trp Asn Gly Ile Ser Gly Arg Leu Gly Trp Pro Thr Thr Thr Met Val Thr Asp Ser Met Gly Arg
3     Asn Gly Thr Ala Ser Gln Asp Gly Trp Asp Gly Arg Arg Leu Arg Trp Leu Gln Ile Arg Trp Ala Gly
o
5'  GCUGGUGUGUUGGGAUUGGCGGGCGUUCGAUCACUUCGUACACUGCGCAGAGCGUAAACUGAUGCUGCAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1890
3'  CGACCACACAACCCUAACCGCCCGCAAGCUAGUGAAGCAUGUGACGCGUCUCGCAUUUGACUACGACGUC
1     Leu Val Cys Trp Asp Trp Arg Ala Phe Asp His Phe Val His Cys Ala Glu Arg Lys Leu Met Leu Gln
2   Cys Trp Cys Val Gly Ile Gly Gly Arg Ser Ile Thr Ser Tyr Thr Ala Gln Ser Val Asn  .  Cys Cys Arg
3    Ala Gly Val Leu Gly Leu Ala Gly Val Arg Ser Leu Arg Thr Leu Arg Arg Ala  .  Thr Asp Ala Ala
o
5'  GCGAUGGUGCAGCUUACCGAGCGGUUCUGCAGGGGGGAGGCACGACGCGAUAUGUUGCACGAACUCGAUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1960
3'  CGCUACCACGUCGAAUGGCUCGCCAAGACGUCCCCCCUCCGUGCUGCGCUAUACAACGUGCUUGAGCUAG
1    Ala Met Val Gln Leu Thr Glu Arg Phe Cys Arg Gly Glu Ala Arg Arg Asp Met Leu His Glu Leu Asp
2     Arg Trp Cys Ser Leu Pro Ser Gly Ser Ala Gly Gly Arg His Asp Ala Ile Cys Cys Thr Asn Ser Ile
3   Gly Asp Gly Ala Ala Tyr Arg Ala Val Leu Gln Gly Gly Gly Thr Thr Arg Tyr Val Ala Arg Thr Arg Ser
o
5'  AGCUUAUCGCCGGGACGGAGGAGUUGAUCUUCCGGUCCAGAGCAUUCGCUGAGGACAAGUACAAGGAAGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2030
3'  UCGAAUAGCGGCCCUGCCUCCUCAACUAGAAGGCCAGGUCUCGUAAGCGACUCCUGUUCAUGUUCCUUCG
1   Gln Leu Ile Ala Gly Thr Glu Glu Leu Ile Phe Arg Ser Arg Ala Phe Ala Glu Asp Lys Tyr Lys Glu Ala
2    Ser Leu Ser Pro Gly Arg Arg Ser  .  Ser Ser Gly Pro Glu His Ser Leu Arg Thr Ser Thr Arg Lys
3     Ala Tyr Arg Arg Asp Gly Gly Val Asp Leu Pro Val Gln Ser Ile Arg  .  Gly Gln Val Gln Gly Ser
o
5'  CGUCGAACAGAUACUCAGCAAGGAGCCAGGCAGAGCGACCAGGCUAGCCGGGCAUGCUGGUCAAUACGCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2100
3'  GCAGCUUGUCUAUGAGUCGUUCCUCGGUCCGUCUCGCUGGUCCGAUCGGCCCGUACGACCAGUUAUGCGG
1     Val Glu Gln Ile Leu Ser Lys Glu Pro Gly Arg Ala Thr Arg Leu Ala Gly His Ala Gly Gln Tyr Ala
2   Pro Ser Asn Arg Tyr Ser Ala Arg Ser Gln Ala Glu Arg Pro Gly  .  Pro Gly Met Leu Val Asn Thr Pro
3    Arg Arg Thr Asp Thr Gln Gln Gly Ala Arg Gln Ser Asp Gln Ala Ser Arg Ala Cys Trp Ser Ile Arg
o
5'  AUUGCAGUCAAGAAUCCCCAGGGGCAGCAAUCGGGAAGGUUCAGCACGUUGCUGGGCAAUACAGUCAUCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2170
3'  UAACGUCAGUUCUUAGGGGUCCCCGUCGUUAGCCCUUCCAAGUCGUGCAACGACCCGUUAUGUCAGUAGC
1    Ile Ala Val Lys Asn Pro Gln Gly Gln Gln Ser Gly Arg Phe Ser Thr Leu Leu Gly Asn Thr Val Ile
2     Leu Gln Ser Arg Ile Pro Arg Gly Ser Asn Arg Glu Gly Ser Ala Arg Cys Trp Ala Ile Gln Ser Ser
3   His Cys Ser Gln Glu Ser Pro Gly Ala Ala Ile Gly Lys Val Gln His Val Ala Gly Gln Tyr Ser His Arg
o
5'  GGACGACGCGGCUGCUCGUCAGGGAUGCAGAGCUGAUGGGACGCACCGCCGAUUUGCGACACCGAACAUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2240
3'  CCUGCUGCGCCGACGAGCAGUCCCUACGUCUCGACUACCCUGCGUGGCGGCUAAACGCUGUGGCUUGUAG
1   Gly Thr Thr Arg Leu Leu Val Arg Asp Ala Glu Leu Met Gly Arg Thr Ala Asp Leu Arg His Arg Thr Ser
2    Gly Arg Arg Gly Cys Ser Ser Gly Met Gln Ser  .  Trp Asp Ala Pro Pro Ile Cys Asp Thr Glu His
3     Asp Asp Ala Ala Ala Arg Gln Gly Cys Arg Ala Asp Gly Thr His Arg Arg Phe Ala Thr Pro Asn Ile
o
```

```
5'  CAUUUUCGUACUCAACAGGGCCGACGACGUCGCAGAGCUAUUUUCUGAAUUUCAACUGGCAGUCAAGGCU
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2310
3'  GUAAAAGCAUGAGUUGUCCCGGCUGCUGCAGCGUCUCGAUAAAAGACUUAAAGUUGACCGUCAGUUCCGA
1     Ile Phe Val Leu Asn Arg Ala Asp Asp Val Ala Glu Leu Phe Ser Glu Phe Gln Leu Ala Val Lys Ala
2   Pro Phe Ser Tyr Ser Thr Gly Pro Thr Thr Ser Gln Ser Tyr Phe Leu Asn Phe Asn Trp Gln Ser Arg Leu
3    His Phe Arg Thr Gln Gln Gly Arg Arg Arg Arg Arg Ala Ile Phe  .  Ile Ser Thr Gly Ser Gln Gly
o
5'  AUACAGACCAUGUUGGACCAGGGACAUAAGGCCAAUCCGAAGAAGCAGCUCGUGCAAUGGCGUGCAUGCA
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2380
3'  UAUGUCUGGUACAACCUGGUCCCUGUAUUCCGGUUAGGCUUCUUCGUCGAGCACGUUACCGCACGUACGU
1    Ile Gln Thr Met Leu Asp Gln Gly His Lys Ala Asn Pro Lys Lys Gln Leu Val Gln Trp Arg Ala Cys
2     Tyr Arg Pro Cys Trp Thr Arg Asp Ile Arg Pro Ile Arg Arg Ser Ser Ser Cys Asn Gly Val His Ala
3   Tyr Thr Asp His Val Gly Pro Gly Thr  .  Gly Gln Ser Glu Glu Ala Ala Arg Ala Met Ala Cys Met His
o
5'  UUUACUUACGCAUACUGUAUGCGGGCGGUACGAUGCGCGCUUUUCCGGCGCGCGCCAUCUACGCAGCGGC
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2450
3'  AAAUGAAUGCGUAUGACAUACGCCCGCCAUGCUACGCGCGAAAAGGCCGCGCGCGGUAGAUGCGUCGCCG
1   Ile Tyr Leu Arg Ile Leu Tyr Ala Gly Gly Thr Met Arg Ala Phe Pro Ala Arg Ala Ile Tyr Ala Ala Ala
2    Phe Thr Tyr Ala Tyr Cys Met Arg Ala Val Arg Cys Ala Leu Phe Arg Arg Ala Pro Ser Thr Gln Arg
3     Leu Leu Thr His Thr Val Cys Gly Arg Tyr Asp Ala Arg Phe Ser Gly Ala Arg His Leu Arg Ser Gly
o
5'  GACAGCGAGUCCGCCAAAGGGCACCGGCGGAGACAUGACAAUAGCGGAGAAGUUGGGGUCUAUCGCGAAG
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2520
3'  CUGUCGCUCAGGCGGUUUCCCGUGGCCGCCUCUGUACUGUUAUCGCCUCUUCAACCCCAGAUAGCGCUUC
1     Thr Ala Ser Pro Pro Lys Gly Thr Gly Gly Asp Met Thr Ile Ala Glu Lys Leu Gly Ser Ile Ala Lys
2   Arg Gln Arg Val Arg Gln Arg Ala Pro Ala Glu Thr  .  Gln  .  Arg Arg Ser Trp Gly Leu Ser Arg Arg
3    Asp Ser Glu Ser Ala Lys Gly His Arg Arg Arg His Asp Asn Ser Gly Glu Val Gly Val Tyr Arg Glu
o
5'  GGGUUGGACAUGUACGUGCGGCGCGGUUUUUGUCGAAGAUGGCCGGUGCGUUGUACGAUGAUGCGCGCC
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2590
3'  CCCAACCUGUACAUGCACGCCGCGCCAAAAACAGCUUCUACCGGCCACGCAACAUGCUACUACGCGCGG
1    Gly Leu Asp Met Tyr Val Arg Arg Gly Phe Leu Ser Lys Met Ala Gly Ala Leu Tyr Asp Asp Ala Arg
2     Gly Trp Thr Cys Thr Cys Gly Ala Gly Phe Cys Arg Arg Trp Pro Val Arg Cys Thr Met Met Arg Ala
3   Gly Val Gly His Val Arg Ala Ala Arg Val Phe Val Glu Asp Gly Arg Cys Val Val Arg  .  Cys Ala Pro
o
5'  GGUUCUUCUCGCGCGGCAGGUUGUGGCUACCCCCGGACUCGCGAGGCCGCCGGCCGAAGCAACGUUUCAU
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2660
3'  CCAAGAAGAGCGCGCCGUCCAACACCGAUGGGGGCCUGAGCGCUCCGGCGGCCGGCUUCGUUGCAAAGUA
1   Arg Phe Phe Ser Arg Gly Arg Leu Trp Leu Pro Pro Asp Ser Arg Gly Arg Arg Pro Lys Gln Arg Phe Met
2    Gly Ser Ser Arg Ala Ala Gly Cys Gly Tyr Pro Arg Thr Arg Glu Ala Ala Gly Arg Ser Asn Val Ser
3     Val Leu Leu Ala Arg Gln Val Val Ala Thr Pro Gly Leu Ala Arg Pro Pro Ala Glu Ala Thr Phe His
o
5'  GCGCGUGCAGAUAGCGCGUGAGGUGCUCGAGGGUGCGACAGAGCACGGCGGUAUGGGUGUGCUGCCACCG
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2730
3'  CGCGCACGUCUAUCGCGCACUCCACGAGCUCCCACGCUGUCUCGUGCCGCCAUACCCACACGACGGUGGC
1     Arg Val Gln Ile Ala Arg Glu Val Leu Glu Gly Ala Thr Glu His Gly Gly Met Gly Val Leu Pro Pro
2   Cys Ala Cys Arg  .  Arg Val Arg Cys Ser Arg Val Arg Gln Ser Thr Ala Val Trp Val Cys Cys His Arg
3    Ala Arg Ala Asp Ser Ala  .  Gly Ala Arg Gly Cys Asp Arg Ala Arg Arg Tyr Gly Cys Ala Ala Thr
o
5'  GGCCAGUACGACUACGAUUACUCCAUCAAGUGCAGCAAGCGCAAAUACAUGGAGGAGAUCGCAGACGGUU
o   +++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|+++++++++|  2800
3'  CCGGUCAUGCUGAUGCUAAUGAGGUAGUUCACGUCGUUCGCGUUUAUGUACCUCCUCUAGCGUCUGCCAA
1    Gly Gln Tyr Asp Tyr Asp Tyr Ser Ile Lys Cys Ser Lys Arg Lys Tyr Met Glu Glu Ile Ala Asp Gly
2     Ala Ser Thr Thr Thr Ile Thr Pro Ser Ser Ala Ala Ser Ala Asn Thr Trp Arg Arg Ser Gln Thr Val
3   Gly Pro Val Arg Leu Arg Leu Leu His Gln Val Gln Gln Ala Gln Ile His Gly Gly Asp Arg Arg Arg Leu
o
```

```
5'  GGGACACACAGCUACGCCGCCACGUUGCCAAAUUGUCACGUCGUGCACCUGGUGUGGACGACCUUGAGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2870
3'  CCCUGUGUGUCGAUGCGGCGGUGCAACGGUUUAACAGUGCAGCACGUGGACCACACCUGCUGGAACUCGC
1   Trp Asp Thr Gln Leu Arg Arg His Val Ala Lys Leu Ser Arg Arg Ala Pro Gly Val Asp Asp Leu Glu Arg
2    Gly Thr His Ser Tyr Ala Ala Thr Leu Pro Asn Cys His Val Val His Leu Val Trp Thr Thr Leu Ser
3     Gly His Thr Ala Thr Pro Pro Arg Cys Gln Ile Val Thr Ser Cys Thr Trp Cys Gly Arg Pro  .  Ala
o
5'  CAACGCGGUGGCGCGGUUUCAGGACGCUGCGGGGGUCAGGGUUGACACGAAGACACGAGGUGAUUUCCGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2940
3'  GUUGCGCCACCGCGCCAAAGUCCUGCGACGCCCCCAGUCCCAACUGUGCUUCUGUGCUCCACUAAAGGCC
1     Asn Ala Val Ala Arg Phe Gln Asp Ala Ala Gly Val Arg Val Asp Thr Lys Thr Arg Gly Asp Phe Arg
2   Ala Thr Arg Trp Arg Gly Phe Arg Thr Leu Arg Gly Ser Gly Leu Thr Arg Arg His Glu Val Ile Ser Gly
3    Gln Arg Gly Gly Ala Val Ser Gly Arg Cys Gly Gly Gln Gly  .  His Glu Asp Thr Arg  .  Phe Pro
o
5'  UUGAAGUGGUGUGAGGCGGUAGUGCAUGGGGGAGCGGAUCCGGACACGCGAUUCGCAGAGAGGACUCGAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3010
3'  AACUUCACCACACUCCGCCAUCACGUACCCCCUCGCCUAGGCCUGUGCGCUAAGCGUCUCUCCUGAGCUU
1    Leu Lys Trp Cys Glu Ala Val Val His Gly Gly Ala Asp Pro Asp Thr Arg Phe Ala Glu Arg Thr Arg
2     .  Ser Gly Val Arg Arg  .  Cys Met Gly Glu Arg Ile Arg Thr Arg Asp Ser Gln Arg Gly Leu Glu
3   Val Glu Val Val  .  Gly Gly Ser Ala Trp Gly Ser Gly Ser Gly His Ala Ile Arg Arg Glu Asp Ser Lys
o
5'  GCCUGGCACGCAGCAUCAAGUCAUGGCAUAGGGAUACUCGUGAUUCCUACAGCCGCAUUCCUGCGGUGGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3080
3'  CGGACCGUGCGUCGUAGUUCAGUACCGUAUCCCUAUGAGCACUAAGGAUGUCGGCGUAAGGACGCCACCC
1   Ser Leu Ala Arg Ser Ile Lys Ser Trp His Arg Asp Thr Arg Asp Ser Tyr Ser Arg Ile Pro Ala Val Gly
2    Ala Trp His Ala Ala Ser Ser His Gly Ile Gly Ile Leu Val Ile Pro Thr Ala Ala Phe Leu Arg Trp
3     Pro Gly Thr Gln His Gln Val Met Ala  .  Gly Tyr Ser  .  Phe Leu Gln Pro His Ser Cys Gly Gly
o
5'  CAAUCUGAUCAAUGAGGCAUUGAUCAGGUUCGAGAGGGCCAAGAGGUAUGCGAAAAUGAGCAUCAACGGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3150
3'  GUUAGACUAGUUACUCCGUAACUAGUCCAAGCUCUCCCGGUUCUCCAUACGCUUUUACUCGUAGUUGCCG
1     Asn Leu Ile Asn Glu Ala Leu Ile Arg Phe Glu Arg Ala Lys Arg Tyr Ala Lys Met Ser Ile Asn Gly
2   Ala Ile  .  Ser Met Arg His  .  Ser Gly Ser Arg Gly Pro Arg Gly Met Arg Lys  .  Ala Ser Thr Ala
3    Gln Ser Asp Gln  .  Gly Ile Asp Gln Val Arg Glu Gly Gln Glu Val Cys Glu Asn Glu His Gln Arg
o
5'  GGUGAGGCCCGUUUCGAGCUGGAGCGUAUGCGCGGACAGCCAGGCUACGGCAUUCUGGACCACGAGUGGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3220
3'  CCACUCCGGGCAAAGCUCGACCUCGCAUACGCGCCUGUCGGUCCGAUGCCGUAAGACCUGGUGCUCACCA
1    Gly Glu Ala Arg Phe Glu Leu Glu Arg Met Arg Gly Gln Pro Gly Tyr Gly Ile Leu Asp His Glu Trp
2     Val Arg Pro Val Ser Ser Trp Ser Val Cys Ala Asp Ser Gln Ala Thr Ala Phe Trp Thr Thr Ser Gly
3   Arg  .  Gly Pro Phe Arg Ala Gly Ala Tyr Ala Arg Thr Ala Arg Leu Arg His Ser Gly Pro Arg Val Val
o
5'  UCGGUUAUGCUGACCUGUACUUGAAGGAAUUGUCCGCUGCACGUAGGAUGGACGCGCUUUACUCCAUCCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3290
3'  AGCCAAUACGACUGGACAUGAACUUCCUUAACAGGCGACGUGCAUCCUACCUGCGCGAAAUGAGGUAGGA
1   Phe Gly Tyr Ala Asp Leu Tyr Leu Lys Glu Leu Ser Ala Ala Arg Arg Met Asp Ala Leu Tyr Ser Ile Leu
2    Ser Val Met Leu Thr Cys Thr  .  Arg Asn Cys Pro Leu His Val Gly Trp Thr Arg Phe Thr Pro Ser
3     Arg Leu Cys  .  Pro Val Leu Glu Gly Ile Val Arg Cys Thr  .  Asp Gly Arg Ala Leu Leu His Pro
o
5'  GCCGGCGACGAAGCACGGUACUGAACUGCUGUCAAACACGCGCACGUGGCCACAUCGUGCACUGUUUAUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3360
3'  CGGCCGCUGCUUCGUGCCAUGACUUGACGACAGUUUGUGCGCGUGCACCGGUGUAGCACGUGACAAAUAC
1     Pro Ala Thr Lys His Gly Thr Glu Leu Leu Ser Asn Thr Arg Thr Trp Pro His Arg Ala Leu Phe Met
2   Cys Arg Arg Arg Ser Thr Val Leu Asn Cys Cys Gln Thr Arg Ala Arg Gly His Ile Val His Cys Leu Cys
3    Ala Gly Asp Glu Ala Arg Tyr  .  Thr Ala Val Lys His Ala His Val Ala Thr Ser Cys Thr Val Tyr
o
```

```
5'  CACCUGUCAGGCAAGCUUGGUCCUGCAGGCGCGUGGGAUAAGUUGAUGCCACCUUCUUGGUCCGGUUUUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3430
3'  GUGGACAGUCCGUUCGAACCAGGACGUCCGCGCACCCUAUUCAACUACGGUGGAAGAACCAGGCCAAAAG
1    His Leu Ser Gly Lys Leu Gly Pro Ala Gly Ala Trp Asp Lys Leu Met Pro Pro Ser Trp Ser Gly Phe
2     Thr Cys Gln Ala Ser Leu Val Leu Gln Ala Arg Gly Ile Ser . Cys His Leu Leu Gly Pro Val Phe
3   Ala Pro Val Arg Gln Ala Trp Ser Cys Arg Arg Val Gly . Val Asp Ala Thr Phe Leu Val Arg Phe Ser
o
5'  UCGCGAAUACGUUGCAAUACUGUUUGGAUUUCUGUUUCAGGCACUAUCCCACCAAAUUUUCGACAACAUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3500
3'  AGCGCUUAUGCAACGUUAUGACAAACCUAAAGACAAAGUCCGUGAUAGGGUGGUUUAAAAGCUGUUGUAU
1   Leu Ala Asn Thr Leu Gln Tyr Cys Leu Asp Phe Cys Phe Arg His Tyr Pro Thr Lys Phe Ser Thr Thr Tyr
2    Ser Arg Ile Arg Cys Asn Thr Val Trp Ile Ser Val Ser Gly Thr Ile Pro Pro Asn Phe Arg Gln His
3     Arg Glu Tyr Val Ala Ile Leu Phe Gly Phe Leu Phe Gln Ala Leu Ser His Gln Ile Phe Asp Asn Ile
o
5'  UAACAUACUGCGACUGCGGGCCGAUGUUACACGUUCGGUUGUUGGGGGUUGUUCUGCCGCCACUCUCAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3570
3'  AUUGUAUGACGCUGACGCCCGGCUACAAUGUGCAAGCCAACAACCCCCAACAAGACGGCGGUGAGAGUC
1     Asn Ile Leu Arg Leu Arg Ala Asp Val Thr Arg Ser Val Val Gly Gly Leu Phe Cys Arg His Ser Gln
2   Ile Thr Tyr Cys Asp Cys Gly Pro Met Leu His Val Arg Leu Leu Gly Gly Cys Ser Ala Ala Thr Leu Ser
3     . His Thr Ala Thr Ala Gly Arg Cys Tyr Thr Phe Gly Cys Trp Gly Val Val Leu Pro Pro Leu Ser
o
5'  CUGUUUAUUCACUAGUCAGGUAUGAGAGAACGGGACAACAAAGUACACCGAAUUCAAAGCCAUGAGACGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   3640
3'  GACAAAUAAGUGAUCAGUCCAUACUCUCUUGCCCUGUUGUUUCAUGUGGCUUAAGUUUCGGUACUCUGCG
1    Leu Phe Ile His . Ser Gly Met Arg Glu Arg Asp Asn Lys Val His Arg Ile Gln Ser His Glu Thr
2     Cys Leu Phe Thr Ser Gln Val . Glu Asn Gly Thr Thr Lys Tyr Thr Glu Phe Lys Ala Met Arg Arg
3   Ala Val Tyr Ser Leu Val Arg Tyr Glu Arg Thr Gly Gln Gln Ser Thr Pro Asn Ser Lys Pro . Asp Ala
o
5'  GAGACGUAACGCAUGCUACGCGAACUCGGGUCUGCC
o   ++++|++++|++++|++++|++++|++++|++++|+                                     3676
3'  CUCUGCAUUGCGUACGAUGCGCUUGAGCCCAGACGG
1   Arg Asp Val Thr His Ala Thr Arg Thr Arg Val Cys
2    Glu Thr . Arg Met Leu Arg Glu Leu Gly Ser Ala
3     Arg Arg Asn Ala Cys Tyr Ala Asn Ser Gly Leu Pro
o
```

3.5.7 DAV1 dsRNA2 sequence

2683 nt excluding poly (A) tail

```
5'  GGCGAGCGAUCCGUUUUGAGGAGAAUUGAACAGGGAAUACCUAAUCAAUUCUACCUGCAGACGAGGCGUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   70
3'  CCGCUCGCUAGGCAAAACUCCUCUUAACUUGUCCCUUAUGGAUUAGUUAAGAUGGACGUCUGCUCCGCAC
1    Gly Glu Arg Ser Val Leu Arg Arg Ile Glu Gln Gly Ile Pro Asn Gln Phe Tyr Leu Gln Thr Arg Arg
2     Ala Ser Asp Pro Phe  .  Gly Glu Leu Asn Arg Glu Tyr Leu Ile Asn Ser Thr Cys Arg Arg Gly Val
3   Gly Arg Ala Ile Arg Phe Glu Glu Asn  .  Thr Gly Asn Thr  .  Ser Ile Leu Pro Ala Asp Glu Ala  .
o
5'  AAAGGUGUAGAGAGGAGAAUCGACUGAGAUGGAGACUGUCAAGAUACCGUUUCGCCUGGCGGCUGACACU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   140
3'  UUUCCACAUCUCUCCUCUUAGCUGACUCUACCUCUGACAGUUCUAUGGCAAAGCGGACCGCCGACUGUGA
1   Glu Arg Cys Arg Glu Glu Asn Arg Leu Arg Trp Arg Leu Ser Arg Tyr Arg Phe Ala Trp Arg Leu Thr Leu
2    Lys Gly Val Glu Arg Arg Ile Asp  .  Asp Gly Asp Cys Gln Asp Thr Val Ser Pro Gly Gly  .  His
3     Lys Val  .  Arg Gly Glu Ser Thr Glu Met Glu Thr Val Lys Ile Pro Phe Arg Leu Ala Ala Asp Thr
o
5'  UACGAUGCCCCAGAGCGCCUAGACGAUACCAGCAAUUUGUGGCUUGCUUGCAUUUCCAGCAAGGCCAAAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   210
3'  AUGCUACGGGGUCUCGCGGAUCUGCUAUGGUCGUUAAACACCGAACGAACGUAAAGGUCGUUCCGGUUUC
1     Thr Met Pro Gln Ser Ala  .  Thr Ile Pro Ala Ile Cys Gly Leu Leu Ala Phe Pro Ala Arg Pro Lys
2   Leu Arg Cys Pro Arg Ala Pro Arg Arg Tyr Gln Gln Phe Val Ala Cys Leu His Phe Gln Gln Gly Gln Ser
3    Tyr Asp Ala Pro Glu Arg Leu Asp Asp Thr Ser Asn Leu Trp Leu Ala Cys Ile Ser Ser Lys Ala Lys
o
5'  CCACCGACUUCCACGUCACCUUGUCCAGAACCGAGGACCAAGUCAAAGACGUGACACCCGAAAUCCUCCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   280
3'  GGUGGCUGAAGGUGCAGUGGAACAGGUCUUGGCUCCUGGUUCAGUUUCUGCACUGUGGGCUUUAGGAGGC
1    Pro Pro Thr Ser Thr Ser Pro Cys Pro Glu Pro Arg Thr Lys Ser Lys Thr  .  His Pro Lys Ser Ser
2     His Arg Leu Pro Arg His Leu Val Gln Asn Arg Gly Pro Ser Gln Arg Arg Asp Thr Arg Asn Pro Pro
3   Ala Thr Asp Phe His Val Thr Leu Ser Arg Thr Glu Asp Gln Val Lys Asp Val Thr Pro Glu Ile Leu Arg
o
5'  CAAGCACGAUGUCAUGACAAGGCGCGGCCAACUGUCAGGCAUUUUCGACAGACUCGGAGAAGCGCUCGGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   350
3'  GUUCGUGCUACAGUACUGUUCCGCGCCGGUUGACAGUCCGUAAAAGCUGUCUGAGCCUCUUCGCGAGCCG
1   Ala Ser Thr Met Ser  .  Gln Gly Ala Ala Asn Cys Gln Ala Phe Ser Thr Asp Ser Glu Lys Arg Ser Ala
2    Gln Ala Arg Cys His Asp Lys Ala Arg Pro Thr Val Arg His Phe Arg Gln Thr Arg Arg Ser Ala Arg
3     Lys His Asp Val Met Thr Arg Arg Gly Gln Leu Ser Gly Ile Phe Asp Arg Leu Gly Glu Ala Leu Gly
o
5'  GGUAGCAUGGAUGGUGUGGGGGUUGAUUUGAGGAUGCUACCGGUCGACGGUCGCGCAGGGCAAACCUGUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   420
3'  CCAUCGUACCUACCACACCCCCAACUAAACUCCUACGAUGGCCAGCUGCCAGCGCGUCCCGUUUGGACAA
1     Val Ala Trp Met Val Trp Gly Leu Ile  .  Gly Cys Tyr Arg Ser Thr Val Ala Gln Gly Lys Pro Val
2   Arg  .  His Gly Trp Cys Gly Gly  .  Phe Glu Asp Ala Thr Gly Arg Arg Ser Arg Arg Ala Asn Leu Phe
3    Gly Ser Met Asp Gly Val Gly Val Asp Leu Arg Met Leu Pro Val Asp Gly Arg Ala Gly Gln Thr Cys
o
5'  UUGCGACAACCACGAGUCUCGCCAUCAUUGCGAAGCGGAUGGCCACAUUGGUGGAGCGGUACGAUAUCCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   490
3'  AACGCUGUUGGUGCUCAGAGCGGUAGUAACGCUUCGCCUACCGGUGUAACCACCUCGCCAUGCUAUAGGA
1    Leu Arg Gln Pro Arg Val Ser Pro Ser Leu Arg Ser Gly Trp Pro His Trp Trp Ser Gly Thr Ile Ser
2     Cys Asp Asn His Glu Ser Arg His His Cys Glu Ala Asp Gly His Ile Gly Gly Ala Val Arg Tyr Pro
3   Phe Ala Thr Thr Thr Ser Leu Ala Ile Ile Ala Lys Arg Met Ala Thr Leu Val Glu Arg Tyr Asp Ile Leu
o
5'  GCGUGAGAUGGCCUUCGUCAAAUACCAUAUCACGAUUGAGUAUGCUGGUGGGGCGACAAUGAAGUUUGUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   560
3'  CGCACUCUACCGGAAGCAGUUUAUGGUAUAGUGCUAACUCAUACGACCACCCCGCUGUUACUUCAAACAG
1   Cys Val Arg Trp Pro Ser Ser Asn Thr Ile Ser Arg Leu Ser Met Leu Val Gly Arg Gln  .  Ser Leu Ser
2    Ala  .  Asp Gly Leu Arg Gln Ile Pro Tyr His Asp  .  Val Cys Trp Trp Gly Asp Asn Glu Val Cys
3     Arg Glu Met Ala Phe Val Lys Tyr His Ile Thr Ile Glu Tyr Ala Gly Gly Ala Thr Met Lys Phe Val
o
```

```
5'  ACGACUAUGGCCGUUGACGACUCAGACCCCGCGCAAUUCGAUCGCACAGAUUUGCGGGCCAUGGAACGUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   630
3'  UGCUGAUACCGGCAACUGCUGAGUCUGGGGCGCGUUAAGCUAGCGUGUCUAAACGCCCGGUACCUUGCAA
1     Arg Leu Trp Pro Leu Thr Thr Gln Thr Pro Arg Asn Ser Ile Ala Gln Ile Cys Gly Pro Trp Asn Val
2   His Asp Tyr Gly Arg  .  Arg Leu Arg Pro Arg Ala Ile Arg Ser His Arg Phe Ala Gly His Gly Thr Phe
3    Thr Thr Met Ala Val Asp Asp Ser Asp Pro Ala Gln Phe Asp Arg Thr Asp Leu Arg Ala Met Glu Arg
o
5'  UGGCCGCGGUGGAGCUGGAUACCACGCCGCCGGCACCACGCGAGGUUACGCCGUACGCGCCACCUGUUUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   700
3'  ACCGGCGCCACCUCGACCUAUGGUGCGGCGGCCGUGGUGCGCUCCAAUGCGGCAUGCGCGGUGGACAAAG
1    Trp Pro Arg Trp Ser Trp Ile Pro Arg Arg Arg His His Ala Arg Leu Arg Arg Thr Arg His Leu Phe
2     Gly Arg Gly Gly Ala Gly Tyr His Ala Ala Gly Thr Thr Arg Gly Tyr Ala Val Arg Ala Thr Cys Phe
3   Leu Ala Ala Val Glu Leu Asp Thr Thr Pro Pro Ala Pro Arg Glu Val Thr Pro Tyr Ala Pro Pro Val Ser
o
5'  AUUUGAGGAUGUCACUGAAGUGAAGGGUGACGUCGCAUCGCAGACCGGCACUAUUCUGGUCGAACAAGAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   770
3'  UAAACUCCUACAGUGACUUCACUUCCCACUGCAGCGUAGCGUCUGGCCGUGAUAAGACCAGCUUGUUCUA
1   His Leu Arg Met Ser Leu Lys  .  Arg Val Thr Ser His Arg Arg Pro Ala Leu Phe Trp Ser Asn Lys Met
2    Ile  .  Gly Cys His  .  Ser Glu Gly  .  Arg Arg Ile Ala Asp Arg His Tyr Ser Gly Arg Thr Arg
3     Phe Glu Asp Val Thr Glu Val Lys Gly Asp Val Ala Ser Gln Thr Gly Thr Ile Leu Val Glu Gln Asp
o
5'  GCACCUGAGCCGCAGGUGCCGUUUGAGCCUGCGGACUUACAUGCCAGCUUGCUGGGCAACGCGCUGCUGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   840
3'  CGUGGACUCGGCGUCCACGGCAAACUCGGACGCCUGAAUGUACGGUCGAACGACCCGUUGCGCGACGACG
1     His Leu Ser Arg Arg Cys Arg Leu Ser Leu Arg Thr Tyr Met Pro Ala Cys Trp Ala Thr Arg Cys Cys
2   Cys Thr  .  Ala Ala Gly Ala Val  .  Ala Cys Gly Leu Thr Cys Gln Leu Ala Gly Gln Arg Ala Ala Ala
3    Ala Pro Glu Pro Gln Val Pro Phe Glu Pro Ala Asp Leu His Ala Ser Leu Leu Gly Asn Ala Leu Leu
o
5'  AUCCGGACGUUUCCCCUACGGAGAAGUAUACAGCGUUGGCAAACAUGAUCGCACCGAUGCGCGACAAGCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   910
3'  UAGGCCUGCAAAGGGGAUGCCUCUUCAUAUGUCGCAACCGUUUGUACUAGCGUGGCUACGCGCUGUUCGU
1    Ile Arg Thr Phe Pro Leu Arg Arg Ser Ile Gln Arg Trp Gln Thr  .  Ser His Arg Cys Ala Thr Ser
2     Ser Gly Arg Phe Pro Tyr Gly Glu Val Tyr Ser Val Gly Lys His Asp Arg Thr Asp Ala Arg Gln Ala
3   His Pro Asp Val Ser Pro Thr Glu Lys Tyr Thr Ala Leu Ala Asn Met Ile Ala Pro Met Arg Asp Lys Gln
o
5'  GUCAGGAGUGAUCAUUUACGGCGACAACCCUGGGACUCUAGCUCAAGCGUUGGUUGCUGUUGGUCACAAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   980
3'  CAGUCCUCACUAGUAAAUGCCGCUGUUGGGACCCUGAGAUCGAGUUCGCAACCAACGACAACCAGUGUUG
1   Ser Gln Glu  .  Ser Phe Thr Ala Thr Thr Leu Gly Leu  .  Leu Lys Arg Trp Leu Leu Leu Val Thr Thr
2    Val Arg Ser Asp His Leu Arg Arg Gln Pro Trp Asp Ser Ser Ser Ser Val Gly Cys Cys Trp Ser Gln
3     Ser Gly Val Ile Ile Tyr Gly Asp Asn Pro Gly Thr Leu Ala Gln Ala Leu Val Ala Val Gly His Asn
o
5'  GUGACGGGCGUCGAUCCGAAGAAUAUGACGACGACGCAACGAGGCAGAGGCGGUCGUGGGCAGUACAGGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1050
3'  CACUGCCCGCAGCUAGGCUUCUUAUACUGCUGCUGCGUUGCUCCGUCUCCGCCAGCACCCGUCAUGUCCU
1     .  Arg Ala Ser Ile Arg Arg Ile  .  Arg Arg Arg Asn Glu Ala Glu Ala Val Val Gly Ser Thr Gly
2   Arg Asp Gly Arg Arg Ser Glu Glu Tyr Asp Asp Asp Ala Thr Arg Gln Arg Arg Ser Trp Ala Val Gln Asp
3    Val Thr Gly Val Asp Pro Lys Asn Met Thr Thr Thr Gln Arg Gly Arg Gly Gly Arg Gly Gln Tyr Arg
o
5'  UGGUGUGUGGAUCAGUUGCGCUAGUCGACGAUGAGGUGCAAGUGUUGGACAAGGCCUCCCGGCAGGUGGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1120
3'  ACCACACACCUAGUCAACGCGAUCAGCUGCUACUCCACGUUCACAACCUGUUCCGGAGGGCCGUCCACCU
1    Trp Cys Val Asp Gln Leu Arg  .  Ser Thr Met Arg Cys Lys Cys Trp Thr Arg Pro Pro Gly Arg Trp
2     Gly Val Trp Ile Ser Cys Ala Ser Arg Arg  .  Gly Ala Ser Val Gly Gln Gly Leu Pro Ala Gly Gly
3   Met Val Cys Gly Ser Val Ala Leu Val Asp Asp Glu Val Gln Val Leu Asp Lys Ala Ser Arg Gln Val Glu
o
```

```
5'   GGGGAUCGGAACUCAAUUCGGUGCUUUUGUGGUGGACACGUCCCUGGACGGCGAGAGAGCUGAGGAUGCC
o                                                                              1190
3'   CCCCUAGCCUUGAGUUAAGCCACGAAAACACCACCUGUGCAGGGACCUGCCGCUCUCUCGACUCCUACGG
1    Arg Gly Ser Glu Leu Asn Ser Val Leu Leu Trp Trp Thr Arg Pro Trp Thr Ala Arg Glu Leu Arg Met Pro
2     Gly Asp Arg Asn Ser Ile Arg Cys Phe Cys Gly Gly His Val Pro Gly Arg Arg Glu Ser . Gly Cys
3      Gly Ile Gly Thr Gln Phe Gly Ala Phe Val Val Asp Thr Ser Leu Asp Gly Glu Arg Ala Glu Asp Ala
o
5'   ACGAAGCGCAACCUCGAUAUUGCGUGGCGAUUAAAAAAGGACAACAUCAACGCUACUGUGAUAUGUCAGC
o                                                                              1260
3'   UGCUUCGCGUUGGAGCUAUAACGCACCGCUAAUUUUUUCCUGUUGUAGUUGCGAUGACACUAUACAGUCG
1      Arg Ser Ala Thr Ser Ile Leu Arg Gly Asp . Lys Arg Thr Thr Ser Thr Leu Leu . Tyr Val Ser
2    His Glu Ala Gln Pro Arg Tyr Cys Val Ala Ile Lys Lys Gly Gln His Gln Arg Tyr Cys Asp Met Ser Ala
3     Thr Lys Arg Asn Leu Asp Ile Ala Trp Arg Leu Lys Lys Asp Asn Ile Asn Ala Thr Val Ile Cys Gln
o
5'   UGCGCUCGGCGCCAUUGGCAAAGGGGGACAUGGUCAUCCUGGACCUUCCAGGUCACAGCAAUCAGGGGUG
o                                                                              1330
3'   ACGCGAGCCGCGGUAACCGUUUCCCCCUGUACCAGUAGGACCUGGAAGGUCCAGUGUCGUUAGUCCCCAC
1     Cys Ala Arg Arg His Trp Gln Arg Gly Thr Trp Ser Ser Trp Thr Phe Gln Val Thr Ala Ile Arg Gly
2      Ala Leu Gly Ala Ile Gly Lys Gly Gly His Gly His Pro Gly Pro Ser Arg Ser Gln Gln Ser Gly Val
3    Leu Arg Ser Ala Pro Leu Ala Lys Gly Asp Met Val Ile Leu Asp Leu Pro Gly His Ser Asn Gln Gly Cys
o
5'   CGAGGCAUAUGCGCUACUGAUUGACAAGGGCAGUAUGAAUUUCGGUCGGGCCAGGCCGGAUUGGCAACGC
o                                                                              1400
3'   GCUCCGUAUACGCGAUGACUAACUGUUCCCGUCAUACUUAAAGCCAGCCCGGUCCGGCCUAACCGUUGCG
1    Ala Arg His Met Arg Tyr . Leu Thr Arg Ala Val . Ile Ser Val Gly Pro Gly Arg Ile Gly Asn Ala
2     Arg Gly Ile Cys Ala Thr Asp . Gln Gly Gln Tyr Glu Phe Arg Ser Gly Gln Ala Gly Leu Ala Thr
3      Glu Ala Tyr Ala Leu Leu Ile Asp Lys Gly Ser Met Asn Phe Gly Arg Ala Arg Pro Asp Trp Gln Arg
o
5'   UACGGGUGGGCCUUCGUGGAGGAGGAGCAGAGUGACGAUACGCAAAUGGAGGGUGCGAUCGACCUGGCGG
o                                                                              1470
3'   AUGCCCACCCGGAAGCACCUCCUCCUCGUCUCACUGCUAUGCGUUUACCUCCCACGCUAGCUGGACCGCC
1      Thr Gly Gly Pro Ser Trp Arg Arg Ser Arg Val Thr Ile Arg Lys Trp Arg Val Arg Ser Thr Trp Arg
2    Leu Arg Val Gly Leu Arg Gly Gly Gly Ala Glu . Arg Tyr Ala Asn Gly Gly Cys Asp Arg Pro Gly Gly
3     Tyr Gly Trp Ala Phe Val Glu Glu Glu Gln Ser Asp Asp Thr Gln Met Glu Gly Ala Ile Asp Leu Ala
o
5'   ACGAGUUUCUGUUUAGCCUGAGCGACAGAGCAGCGGAGUGUUCGGACGUUUUCGCAGCGUACUUUUCGAC
o                                                                              1540
3'   UGCUCAAAGACAAAUCGGACUCGCUGUCUCGUCGCCUCACAAGCCUGCAAAAGCGUCGCAUGAAAAGCUG
1     Thr Ser Phe Cys Leu Ala . Ala Thr Glu Gln Arg Ser Val Arg Thr Phe Ser Gln Arg Thr Phe Arg
2      Arg Val Ser Val . Pro Glu Arg Gln Ser Ser Gly Val Phe Gly Arg Phe Arg Ser Val Leu Phe Asp
3    Asp Glu Phe Leu Phe Ser Leu Ser Asp Arg Ala Ala Glu Cys Ser Asp Val Phe Ala Ala Tyr Phe Ser Thr
o
5'   UGUGGACGAGCCGAUUCAAAGCUGCAGCAAGGGUUACGACGGUGACAUGUACGACCACGGCGGCAAGACC
o                                                                              1610
3'   ACACCUGCUCGGCUAAGUUUCGACGUCGUUCCCAAUGCUGCCACUGUACAUGCUGGUGCCGCCGUUCUGG
1    Leu Trp Thr Ser Arg Phe Lys Ala Ala Ala Arg Val Thr Thr Val Thr Cys Thr Thr Thr Ala Ala Arg Pro
2     Cys Gly Arg Ala Asp Ser Lys Leu Gln Gln Gly Leu Arg Arg . His Val Arg Pro Arg Arg Gln Asp
3      Val Asp Glu Pro Ile Gln Ser Cys Ser Lys Gly Tyr Asp Gly Asp Met Tyr Asp His Gly Gly Lys Thr
o
5'   ACAUACCGCAGCUCCUAUCGGUUUUGGUAUCACUUCAUUGUCACGCGAUGCUUCUUGAGAGCUGCUACAG
o                                                                              1680
3'   UGUAUGGCGUCGAGGAUAGCCAAAACCAUAGUGAAGUAACAGUGCGCUACGAAGAACUCUCGACGAUGUC
1      His Thr Ala Ala Pro Ile Gly Phe Gly Ile Thr Ser Leu Ser Arg Asp Ala Ser . Glu Leu Leu Gln
2    His Ile Pro Gln Leu Leu Ser Val Leu Val Ser Leu His Cys His Ala Met Leu Leu Glu Ser Cys Tyr Arg
3     Thr Tyr Arg Ser Ser Tyr Arg Phe Trp Tyr His Phe Ile Val Thr Arg Cys Phe Leu Arg Ala Ala Thr
o
```

```
5'   AGCAGAAGAAGGCGCUCUCGCGGUCAGCCGUUGCGGCUGCAAAUCCGGUCGAACUCCAGGCUCUCUGGGA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1750
3'   UCGUCUUCUUCCGCGAGAGCGCCAGUCGGCAACGCCGACGUUUAGGCCAGCUUGAGGUCCGAGAGACCCU
1     Ser Arg Arg Arg Arg Ser Arg Gly Gln Pro Leu Arg Leu Gln Ile Arg Ser Asn Ser Arg Leu Ser Gly
2      Ala Glu Glu Gly Ala Leu Ala Val Ser Arg Cys Gly Cys Lys Ser Gly Arg Thr Pro Gly Ser Leu Gly
3    Glu Gln Lys Lys Ala Leu Ser Arg Ser Ala Val Ala Ala Ala Asn Pro Val Glu Leu Gln Ala Leu Trp Glu
o
5'   ACAUGAACUCGGUGGACAGGCGGACAAUAUUGUCAAUGAGGCGAUUUUCGAAGCGCAUGCCGCUGAAUCG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1820
3'   UGUACUUGAGCCACCUGUCCGCCUGUUAUAACAGUUACUCCGCUAAAAGCUUCGCGUACGGCGACUUAGC
1    Asn Met Asn Ser Val Asp Arg Arg Thr Ile Leu Ser Met Arg Arg Phe Ser Lys Arg Met Pro Leu Asn Arg
2     Thr . Thr Arg Trp Thr Gly Gly Gln Tyr Cys Gln . Gly Asp Phe Arg Ser Ala Cys Arg . Ile
3      His Glu Leu Gly Gly Gln Ala Asp Asn Ile Val Asn Glu Ala Ile Phe Glu Ala His Ala Ala Glu Ser
o
5'   CUCACGGGCACGAGUCCGGGGUUGUACAAGCUGCACAAAUCGUUGCGGCCAGUGACUCUAGCUCGGUACG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1890
3'   GAGUGCCCGUGCUCAGGCCCCAACAUGUUCGACGUGUUUAGCAACGCCGGUCACUGAGAUCGAGCCAUGC
1      Ser Arg Ala Arg Val Arg Gly Cys Thr Ser Cys Thr Asn Arg Cys Gly Gln . Leu . Leu Gly Thr
2    Ala His Gly His Glu Ser Gly Val Val Gln Ala Ala Gln Ile Val Ala Ala Ser Asp Ser Ser Ser Val Arg
3     Leu Thr Gly Thr Ser Pro Gly Leu Tyr Lys Leu His Lys Ser Leu Arg Pro Val Thr Leu Ala Arg Tyr
o
5'   AAGCACUGGUCGAAAUGACAACAUACGAGGACGCAGUGACCCCCGACGCCAUCAAGCGUCUGUUCACCCA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1960
3'   UUCGUGACCAGCUUUACUGUUGUAUGCUCCUGCGUCACUGGGGGCUGCGGUAGUUCGCAGACAAGUGGGU
1     Lys His Trp Ser Lys . Gln His Thr Arg Thr Gln . Pro Pro Thr Pro Ser Ser Val Cys Ser Pro
2      Ser Thr Gly Arg Asn Asp Asn Ile Arg Gly Arg Ser Asp Pro Arg Arg His Gln Ala Ser Val His Pro
3    Glu Ala Leu Val Glu Met Thr Thr Tyr Glu Asp Ala Val Thr Pro Asp Ala Ile Lys Arg Leu Phe Thr Gln
o
5'   GCCGGCAUCCGGAGUCUUGUUGAGGGGUUUGUCGAACAUUCGGGGGUUGUACGCGGAGGAAUGGGGAGGU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2030
3'   CGGCCGUAGGCCUCAGAACAACUCCCCAAACAGCUUGUAAGCCCCCAACAUGCGCCUCCUUACCCCUCCA
1    Ser Arg His Pro Glu Ser Cys . Gly Val Cys Arg Thr Phe Gly Gly Cys Thr Arg Arg Asn Gly Glu Val
2     Ala Gly Ile Arg Ser Leu Val Glu Gly Phe Val Glu His Ser Gly Val Val Arg Gly Gly Met Gly Arg
3      Pro Ala Ser Gly Val Leu Leu Arg Gly Leu Ser Asn Ile Arg Gly Leu Tyr Ala Glu Glu Trp Gly Gly
o
5'   CCUGUCUUGGCGAAGUACAAACGGGUGCUUCCGUUGGAGUUCAUGUCGGUGCUGGCUUAUAGACCGUUGU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2100
3'   GGACAGAACCGCUUCAUGUUUGCCCACGAAGGCAACCUCAAGUACAGCCACGACCGAAUAUCUGGCAACA
1      Leu Ser Trp Arg Ser Thr Asn Gly Cys Phe Arg Trp Ser Ser Cys Arg Cys Trp Leu Ile Asp Arg Cys
2    Ser Cys Leu Gly Glu Val Gln Thr Gly Ala Ser Val Gly Val His Val Gly Ala Gly Leu . Thr Val Val
3     Pro Val Leu Ala Lys Tyr Lys Arg Val Leu Pro Leu Glu Phe Met Ser Val Leu Ala Tyr Arg Pro Leu
o
5'   ACAUUGCGAUCUAUAAUUACGUCUUCGCAUGCAUGCAAGGAAUGGGCAAGAAGAUGCACUCCUGGGAGCU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2170
3'   UGUAACGCUAGAUAUUAAUGCAGAAGCGUACGUACGUUCCUUACCCGUUCUUCUACGUGAGGACCCUCGA
1     Thr Leu Arg Ser Ile Ile Thr Ser Ser His Ala Cys Lys Glu Trp Ala Arg Arg Cys Thr Pro Gly Ser
2      His Cys Asp Leu . Leu Arg Leu Arg Met His Ala Arg Asn Gly Gln Glu Asp Ala Leu Leu Gly Ala
3    Tyr Ile Ala Ile Tyr Asn Tyr Val Phe Ala Cys Met Gln Gly Met Gly Lys Lys Met His Ser Trp Glu Leu
o
5'   GCAGCAACUUCUCUGGACCGUCACUAUGCACGGUACGACGACCGAGAAGUUCAAGUUCAUUGCCCAGCGC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2240
3'   CGUCGUUGAAGAGACCUGGCAGUGAUACGUGCCAUGCUGCUGGCUCUUCAAGUUCAAGUAACGGGUCGCG
1    Cys Ser Asn Phe Ser Gly Pro Ser Leu Cys Thr Val Arg Arg Pro Arg Ser Ser Ser Ser Leu Pro Ser Ala
2     Ala Ala Thr Ser Leu Asp Arg His Tyr Ala Arg Tyr Asp Asp Arg Glu Val Gln Val His Cys Pro Ala
3      Gln Gln Leu Leu Trp Thr Val Thr Met His Gly Thr Thr Thr Glu Lys Phe Lys Phe Ile Ala Gln Arg
o
```

```
5'   AUGAGGAGCGUUCUCUUUCCGACUGCGACUGGUCGGAAGCCGCAGGCACUGCGACUAGCGCGAGAUGAUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2310
3'   UACUCCUCGCAAGAGAAAGGCUGACGCUGACCAGCCUUCGGCGUCCGUGACGCUGAUCGCGCUCUACUAG
1      .  Gly Ala Phe Ser Phe Arg Leu Arg Leu Val Gly Ser Arg Arg His Cys Asp  .  Arg Glu Met Ile
2    His Glu Glu Arg Ser Leu Ser Asp Cys Asp Trp Ser Glu Ala Ala Gly Thr Ala Thr Ser Ala Arg  .  Ser
3     Met Arg Ser Val Leu Phe Pro Thr Ala Thr Gly Arg Lys Pro Gln Ala Leu Arg Leu Ala Arg Asp Asp
o
5'   AUCUUGCAAACUUGGAACGGAACCUGCGAGUACUUGGGGUGGCGCACGAUUAUGAUGCGAUCGACCCCAC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2380
3'   UAGAACGUUUGAACCUUGCCUUGGACGCUCAUGAACCCCACCGCGUGCUAAUACUACGCUAGCUGGGGUG
1     Ile Leu Gln Thr Trp Asn Gly Thr Cys Glu Tyr Leu Gly Trp Arg Thr Ile Met Met Arg Ser Thr Pro
2      Ser Cys Lys Leu Gly Thr Glu Pro Ala Ser Thr Trp Gly Gly Ala Arg Leu  .  Cys Asp Arg Pro His
3    His Leu Ala Asn Leu Glu Arg Asn Leu Arg Val Leu Gly Val Ala His Asp Tyr Asp Ala Ile Asp Pro Thr
o
5'   CUAUGCCGCUGCGCGGCAGGCAAUCUUCGCGGUGACGAGGGACCGUGCAUACAACCCAGCCAUGCCUUUU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2450
3'   GAUACGGCGACGCGCCGUCCGUUAGAAGCGCCACUGCUCCCUGGCACGUAUGUUGGGUCGGUACGGAAAA
1    Pro Met Pro Leu Arg Gly Arg Gln Ser Ser Arg  .  Arg Gly Thr Val His Thr Thr Gln Pro Cys Leu Leu
2     Leu Cys Arg Cys Ala Ala Gly Asn Leu Arg Gly Asp Glu Gly Pro Cys Ile Gln Pro Ser His Ala Phe
3      Tyr Ala Ala Ala Arg Gln Ala Ile Phe Ala Val Thr Arg Asp Arg Ala Tyr Asn Pro Ala Met Pro Phe
o
5'   GAGCCCGGAGCUGCACAUGCGCGAUCUUUGUCGGGAAGUUCGGUGUCUGGUAUUGGAGGAGGAAGGAGGA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2520
3'   CUCGGGCCUCGACGUGUACGCGCUAGAAACAGCCCUUCAAGCCACAGACCAUAACCUCCUCCUUCCUCCU
1      Ser Pro Glu Leu His Met Arg Asp Leu Cys Arg Glu Val Arg Cys Leu Val Leu Glu Glu Glu Gly Gly
2     .  Ala Arg Ser Cys Thr Cys Ala Ile Phe Val Gly Lys Phe Gly Val Trp Tyr Trp Arg Arg Lys Glu Asp
3     Glu Pro Gly Ala Ala His Ala Arg Ser Leu Ser Gly Ser Ser Val Ser Gly Ile Gly Gly Gly Arg Arg
o
5'   CGGUGCGAACUGCUGGUUUCUUUCGAGGUUGAUUUGCGAUGAUUCUCUCACGGAAUGUGUCACGCGUUAG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2590
3'   GCCACGCUUGACGACCAAAGAAAGCUCCAACUAAACGCUACUAAGAGAGUGCCUUACACAGUGCGCAAUC
1     Arg Cys Glu Leu Leu Val Ser Phe Glu Val Asp Leu Arg  .  Phe Ser His Gly Met Cys His Ala Leu
2      Gly Ala Asn Cys Trp Phe Leu Ser Arg Leu Ile Cys Asp Asp Ser Leu Thr Glu Cys Val Thr Arg  .
3    Thr Val Arg Thr Ala Gly Phe Phe Arg Gly  .  Phe Ala Met Ile Leu Ser Arg Asn Val Ser Arg Val Ser
o
5'   UACAGUAAGACCGCCCCUACUUGCCGGGCGGCAGCAAAAGCAAUGAGUGUGCAGACGUAGCGCCUGCGUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2660
3'   AUGUCAUUCUGGCGGGGAUGAACGGCCCGCCGUCGUUUUCGUUACUCACACGUCUGCAUCGCGGACGCAG
1    Val Gln  .  Asp Arg Pro Tyr Leu Pro Gly Gly Ser Lys Ser Asn Glu Cys Ala Asp Val Ala Pro Ala Ser
2     Tyr Ser Lys Thr Ala Pro Thr Cys Arg Ala Ala Ala Lys Ala Met Ser Val Gln Thr  .  Arg Leu Arg
3      Thr Val Arg Pro Pro Leu Leu Ala Gly Arg Gln Gln Lys Gln  .  Val Cys Arg Arg Ser Ala Cys Val
o
5'   GCGAACUACCACACUGCCAAGGG
o    ++++|++++|++++|++++|+++                                                   2683
3'   CGCUUGAUGGUGUGACGGUUCCC
1      Arg Thr Thr Thr Leu Pro Arg
2    Arg Glu Leu Pro His Cys Gln Gly
3     Ala Asn Tyr His Thr Ala Lys Gly
o
```

3.5.8 DAV1 dsRNA3 sequence

2480 nt excluding poly (A) tail

```
5'   GGCGAGCGAUCUGUUUUGAGGAAGAGUCGACCAGGAUUACCUGAUCGAUUUUCUACCUGCAGACAGAGUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   70
3'   CCGCUCGCUAGACAAAACUCCUUCUCAGCUGGUCCUAAUGGACUAGCUAAAAGAUGGACGUCUGUCUCAC
1     Gly Glu Arg Ser Val Leu Arg Lys Ser Arg Pro Gly Leu Pro Asp Arg Phe Ser Thr Cys Arg Gln Ser
2      Ala Ser Asp Leu Phe . Gly Arg Val Asp Gln Asp Tyr Leu Ile Asp Phe Leu Pro Ala Asp Arg Val
3    Gly Arg Ala Ile Cys Phe Glu Glu Glu Ser Thr Arg Ile Thr . Ser Ile Phe Tyr Leu Gln Thr Glu Cys
o
5'   UGGAGAACGAAGAGGCCGUGCAAGACCGGCAUAACAUGAAAAUGGCGACCGACGACAUUCCCGACGACCA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   140
3'   ACCUCUUGCUUCUCCGGCACGUUCUGGCCGUAUUGUACUUUUACCGCUGGCUGCUGUAAGGGCUGCUGGU
1    Val Glu Asn Glu Glu Ala Val Gln Asp Arg His Asn Met Lys Met Ala Thr Asp Asp Ile Pro Asp Asp Gln
2     Trp Arg Thr Lys Arg Pro Cys Lys Thr Gly Ile Thr . Lys Trp Arg Pro Thr Thr Phe Pro Thr Thr
3      Gly Glu Arg Arg Gly Arg Ala Arg Pro Ala . His Glu Asn Gly Asp Arg Arg His Ser Arg Arg Pro
o
5'   GAAGUUGCUGCCAGGCGAAGGCAUGCCCGAGAUCGGAAUGUUGAGCAUGGAGGAGGUUUUGGCGAUGUUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   210
3'   CUUCAACGACGGUCCGCUUCCGUACGGGCUCUAGCCUUACAACUCGUACCUCCUCCAAAACCGCUACAAC
1      Lys Leu Leu Pro Gly Glu Gly Met Pro Glu Ile Gly Met Leu Ser Met Glu Glu Val Leu Ala Met Leu
2    Arg Ser Cys Cys Gln Ala Lys Ala Cys Pro Arg Ser Glu Cys . Ala Trp Arg Arg Phe Trp Arg Cys Cys
3     Glu Val Ala Ala Arg Arg Arg His Ala Arg Asp Arg Asn Val Glu His Gly Gly Gly Phe Gly Asp Val
o
5'   CGUGCACCCGCGAAGCUCCCGGAGAUGAGCACGGAGAAGGCAGAAGUGGCGAAAGAGGACCGGAUCGAUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   280
3'   GCACGUGGGCGCUUCGAGGGCCUCUACUCGUGCCUCUUCCGUCUUCACCGCUUUCUCCUGGCCUAGCUAC
1     Arg Ala Pro Ala Lys Leu Pro Glu Met Ser Thr Glu Lys Ala Glu Val Ala Lys Glu Asp Arg Ile Asp
2      Val His Pro Arg Ser Ser Arg Arg . Ala Arg Arg Arg Gln Lys Trp Arg Lys Arg Thr Gly Ser Met
3    Ala Cys Thr Arg Glu Ala Pro Gly Asp Glu His Gly Glu Gly Arg Ser Gly Glu Arg Gly Pro Asp Arg .
o
5'   AGCUGAGUCUGGACGGAGUAAACGACGAGCGAUAUGGCAACGGAGGCACCCGCGUGGUGCGUGACUCAAC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   350
3'   UCGACUCAGACCUGCCUCAUUUGCUGCUCGCUAUACCGUUGCCUCCGUGGGCGCACCACGCACUGAGUUG
1    Glu Leu Ser Leu Asp Gly Val Asn Asp Glu Arg Tyr Gly Asn Gly Gly Thr Arg Val Val Arg Asp Ser Thr
2     Ser . Val Trp Thr Glu . Thr Thr Ser Asp Met Ala Thr Glu Ala Pro Ala Trp Cys Val Thr Gln
3      Ala Glu Ser Gly Arg Ser Lys Arg Arg Ala Ile Trp Gln Arg Arg His Pro Arg Gly Ala . Leu Asn
o
5'   UGCACAGCAAGUGCAGCCCGCAGAUGCGGCCAACGUCACCCAUGUGCCUGUUUCCGCGCCUCGUCAGGUA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   420
3'   ACGUGUCGUUCACGUCGGGCGUCUACGCCGGUUGCAGUGGGUACACGGACAAAGGCGCGGAGCAGUCCAU
1      Ala Gln Gln Val Gln Pro Ala Asp Ala Ala Asn Val Thr His Val Pro Val Ser Ala Pro Arg Gln Val
2    Leu His Ser Lys Cys Ser Pro Gln Met Arg Pro Thr Ser Pro Met Cys Leu Phe Pro Arg Leu Val Arg Tyr
3     Cys Thr Ala Ser Ala Ala Arg Arg Cys Gly Gln Arg His Pro Cys Ala Cys Phe Arg Ala Ser Ser Gly
o
5'   CAACCGUUCAACACCGAGGUUAAGGUGCUGCAGAAUACGGCCUUUUCCGACAAGAGCCAGCGGGAUUUGG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   490
3'   GUUGGCAAGUUGUGGCUCCAAUUCCACGACGUCUUAUGCCGGAAAAGGCUGUUCUCGGUCGCCCUAAACC
1     Gln Pro Phe Asn Thr Glu Val Lys Val Leu Gln Asn Thr Ala Phe Ser Asp Lys Ser Gln Arg Asp Leu
2      Asn Arg Ser Thr Pro Arg Leu Arg Cys Cys Arg Ile Arg Pro Phe Pro Thr Arg Ala Ser Gly Ile Trp
3    Thr Thr Val Gln His Arg Gly . Gly Ala Ala Glu Tyr Gly Leu Phe Arg Gln Glu Pro Ala Gly Phe Gly
o
5'   CAAACCUGGUGCGGGGUGGAGGACAAGACUAUCGUGCGGAGGGCGACUUUUCGAGCAUGGUGUUGAGCAU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   560
3'   GUUUGGACCACGCCCCACCUCCUGUUCUGAUAGCACGCCUCCCGCUGAAAAGCUCGUACCACAACUCGUA
1    Ala Asn Leu Val Arg Gly Gly Gly Gln Asp Tyr Arg Ala Glu Gly Asp Phe Ser Ser Met Val Leu Ser Ile
2     Gln Thr Trp Cys Gly Val Glu Asp Lys Thr Ile Val Arg Arg Ala Thr Phe Arg Ala Trp Cys . Ala
3      Lys Pro Gly Ala Gly Trp Arg Thr Arg Leu Ser Cys Gly Gly Arg Leu Phe Glu His Gly Val Glu His
o
```

```
5'   CCUCGGUCACGACAACGCAAACGCUGUGGAGGCGCCGUUGAUGGGACUGCUCACGCGAAUUGCGCAAUUG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   630
3'   GGAGCCAGUGCUGUUGCGUUUGCGACACCUCCGCGGCAACUACCCUGACGAGUGCGCUUAACGCGUUAAC
1     Leu Gly His Asp Asn Ala Asn Ala Val Glu Ala Pro Leu Met Gly Leu Leu Thr Arg Ile Ala Gln Leu
2    Ser Ser Val Thr Thr Thr Gln Thr Leu Trp Arg Arg Arg . Trp Asp Cys Ser Arg Glu Leu Arg Asn Trp
3      Pro Arg Ser Arg Gln Arg Lys Arg Cys Gly Gly Ala Val Asp Gly Thr Ala His Ala Asn Cys Ala Ile
o
5'   GAAGUUCUCCAGCGGUCUGGCGCAACGGCGGCACUCGCGCCGAAUGUGGCAGGCUGGGAUGUACGGACCG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   700
3'   CUUCAAGAGGUCGCCAGACCGCGUUGCCGCCGUGAGCGCGGCUUACACCGUCCGACCCUACAUGCCUGGC
1     Glu Val Leu Gln Arg Ser Gly Ala Thr Ala Ala Leu Ala Pro Asn Val Ala Gly Trp Asp Val Arg Thr
2      Lys Phe Ser Ser Gly Leu Ala Gln Arg Arg His Ser Arg Arg Met Trp Gln Ala Gly Met Tyr Gly Pro
3    Gly Ser Ser Pro Ala Val Trp Arg Asn Gly Gly Thr Arg Ala Glu Cys Gly Arg Leu Gly Cys Thr Asp Arg
o
5'   CGAGUCUCGACGCGGCUGCAUCCGCACGCAAGCGUUACCCAGGAUACGCUGCAUUCUUCAUCCCGCACAG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   770
3'   GCUCAGAGCUGCGCCGACGUAGGCGUGCGUUCGCAAUGGGUCCUAUGCGACGUAAGAAGUAGGGCGUGUC
1    Ala Ser Leu Asp Ala Ala Ala Ser Ala Arg Lys Arg Tyr Pro Gly Tyr Ala Ala Phe Phe Ile Pro His Ser
2     Arg Val Ser Thr Arg Leu His Pro His Ala Ser Val Thr Gln Asp Thr Leu His Ser Ser Ser Arg Thr
3      Glu Ser Arg Arg Gly Cys Ile Arg Thr Gln Ala Leu Pro Arg Ile Arg Cys Ile Leu His Pro Ala Gln
o
5'   CAUGACUGCAGGCGCAGCAUCGGCACUGAUAUCGAUUCUUUUGCCGGGCGGAGCUGGAGCAUACGGCUGG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   840
3'   GUACUGACGUCCGCGUCGUAGCCGUGACUAUAGCUAAGAAAACGGCCCGCCUCGACCUCGUAUGCCGACC
1     Met Thr Ala Gly Ala Ala Ser Ala Leu Ile Ser Ile Leu Leu Pro Gly Gly Ala Gly Ala Tyr Gly Trp
2    Ala . Leu Gln Ala Gln His Arg His . Tyr Arg Phe Phe Cys Arg Ala Glu Leu Glu His Thr Ala Gly
3     His Asp Cys Arg Arg Ser Ile Gly Thr Asp Ile Asp Ser Phe Ala Gly Arg Ser Trp Ser Ile Arg Leu
o
5'   CGCUUUCGACCACGCGACGGGGCUGACGACCCACGACACGAUUACAUGCCGUCAGCAAGCCGUUGGCUCU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   910
3'   GCGAAAGCUGGUGCGCUGCCCCGACUGCUGGGUGCUGUGCUAAUGUACGGCAGUCGUUCGGCAACCGAGA
1     Arg Phe Arg Pro Arg Asp Gly Ala Asp Asp Pro Arg His Asp Tyr Met Pro Ser Ala Ser Arg Trp Leu
2      Ala Phe Asp His Ala Thr Gly Leu Thr Thr His Asp Thr Ile Thr Cys Arg Gln Gln Ala Val Gly Ser
3    Ala Leu Ser Thr Thr Arg Arg Gly . Arg Pro Thr Thr Arg Leu His Ala Val Ser Lys Pro Leu Ala Leu
o
5'   AUCCGGGAGGGAACACGCGCGUGUUGGCCGUCUUUGAACGGCCCCCGCAAACUUGGUCUACGGAGGCUU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   980
3'   UAGGCCCUCCCUUGUGCGCGCACAACCGGCAGAAACUUGCCGGGGGCGUUUGAACCAGAUGCCUCCGAA
1    Tyr Pro Gly Gly Asn Thr Arg Val Leu Ala Val Phe Glu Arg Pro Pro Ala Asn Leu Val Tyr Gly Gly Phe
2     Ile Arg Glu Gly Thr Arg Ala Cys Trp Pro Ser Leu Asn Gly Pro Pro Gln Thr Trp Ser Thr Glu Ala
3      Ser Gly Arg Glu His Ala Arg Val Gly Arg Leu . Thr Ala Pro Arg Lys Leu Gly Leu Arg Arg Leu
o
5'   CUCCUUCACGAAGGCAAAUGUCGACAGUUUGGUCGGCUGGGCUCGAGGAAUCUUUGGCAACAUGUAUUAC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1050
3'   GAGGAAGUGCUUCCGUUUACAGCUGUCAAACCAGCCGACCCGAGCUCCUUAGAAACCGUUGUACAUAAUG
1     Ser Phe Thr Lys Ala Asn Val Asp Ser Leu Val Gly Trp Ala Arg Gly Ile Phe Gly Asn Met Tyr Tyr
2    Ser Pro Ser Arg Arg Gln Met Ser Thr Val Trp Ser Ala Gly Leu Glu Glu Ser Leu Ala Thr Cys Ile Thr
3     Leu Leu His Glu Gly Lys Cys Arg Gln Phe Gly Arg Leu Gly Ser Arg Asn Leu Trp Gln His Val Leu
o
5'   GACCAAGCCGUGAAGUCGGUGGUAGCUGCGAGCUAUGUUUACGUCGAACCAGAGGUGGUAGCCGAGAAUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1120
3'   CUGGUUCGGCACUUCAGCCACCAUCGACGCUCGAUACAAAUGCAGCUUGGUCUCCACCAUCGGCUCUUAG
1     Asp Gln Ala Val Lys Ser Val Val Ala Ala Ser Tyr Val Tyr Val Glu Pro Glu Val Val Ala Glu Asn
2      Thr Lys Pro . Ser Arg Trp . Leu Arg Ala Met Phe Thr Ser Asn Gln Arg Trp . Pro Arg Ile
3    Arg Pro Ser Arg Glu Val Gly Gly Ser Cys Glu Leu Cys Leu Arg Arg Thr Arg Gly Gly Ser Arg Glu Ser
o
```

```
5'  UGGGCGAGAACUUUGCCCCCGUUGGUGCAACGACGGUGCACCAGACAAACGAUCUCAAUGGCGCCGCCUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1190
3'  ACCCGCUCUUGAAACGGGGGCAACCACGUUGCUGCCACGUGGUCUGUUUGCUAGAGUUACCGCGGCGGAA
1   Leu Gly Glu Asn Phe Ala Pro Val Gly Ala Thr Thr Val His Gln Thr Asn Asp Leu Asn Gly Ala Ala Phe
2   Trp Ala Arg Thr Leu Pro Pro Leu Val Gln Arg Arg Cys Thr Arg Gln Thr Ile Ser Met Ala Pro Pro
3   Gly Arg Glu Leu Cys Pro Arg Trp Cys Asn Asp Gly Ala Pro Asp Lys Arg Ser Gln Trp Arg Arg Leu
o
5'  CGGCUGGGAUAGAGGCGGACCUCCUCCGGGUAUUGGUUUGCUAUCCUACCCUGACGUCGACACGAACAAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1260
3'  GCCGACCCUAUCUCCGCCUGGAGGAGGCCCAUAACCAAACGAUAGGAUGGGACUGCAGCUGUGCUUGUUC
1   Gly Trp Asp Arg Gly Gly Pro Pro Pro Gly Ile Gly Leu Leu Ser Tyr Pro Asp Val Asp Thr Asn Lys
2   Ser Ala Gly Ile Glu Ala Asp Leu Leu Arg Val Leu Val Cys Tyr Pro Thr Leu Thr Ser Thr Arg Thr Ser
3   Arg Leu Gly . Arg Arg Thr Ser Ser Gly Tyr Trp Phe Ala Ile Leu Pro . Arg Arg His Glu Gln
o
5'  CGGCCACCGGGCUUUGAUCGCGCAUUGUGGGAUGCCGCACGUGCGACGCAGCAGGCGCCGCCAGGAGGAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1330
3'  GCCGGUGGCCCGAAACUAGCGCGUAACACCCUACGGCGUGCACGCUGCGUCGUCCGCGGCGGUCCUCCUC
1   Arg Pro Pro Gly Phe Asp Arg Ala Leu Trp Asp Ala Ala Arg Ala Thr Gln Gln Ala Pro Pro Gly Gly
2   Gly His Arg Ala Leu Ile Ala His Cys Gly Met Pro His Val Arg Arg Ser Arg Arg Arg Gln Glu Glu
3   Ala Ala Thr Gly Leu . Ser Arg Ile Val Gly Cys Arg Thr Cys Asp Ala Ala Gly Ala Ala Arg Arg Arg
o
5'  GUGAUCCGCUCCCUGCAGCAUUCGGGACCUUGCGGGUGUAUGAGGAUCGGUCAGAGGUGGUGUGUACUUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1400
3'  CACUAGGCGAGGGACGUCGUAAGCCCUGGAACGCCCACAUACUCCUAGCCAGUCUCCACCACACAUGAAC
1   Gly Asp Pro Leu Pro Ala Ala Phe Gly Thr Leu Arg Val Tyr Glu Asp Arg Ser Glu Val Val Cys Thr Trp
2   Val Ile Arg Ser Leu Gln His Ser Gly Pro Cys Gly Cys Met Arg Ile Gly Gln Arg Trp Cys Val Leu
3   . Ser Ala Pro Cys Ser Ile Arg Asp Leu Ala Gly Val . Gly Ser Val Arg Gly Gly Val Tyr Leu
o
5'  GAGAGGGCCUGCUGGUGUGCAGAUGACGGGAUGGGCGCCCUUGGCUGACUGCUUGGUGUGGCUGGGUGAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1470
3'  CUCUCCCGGACGACCACACGUCUACUGCCCUACCCGCGGGAACCGACUGACGAACCACACCGACCCACUA
1   Arg Gly Pro Ala Gly Val Gln Met Thr Gly Trp Ala Pro Leu Ala Asp Cys Leu Val Trp Leu Gly Asp
2   Gly Glu Gly Leu Leu Val Cys Arg . Arg Asp Gly Arg Pro Trp Leu Thr Ala Trp Cys Gly Trp Val Met
3   Glu Arg Ala Cys Trp Cys Ala Asp Asp Gly Met Gly Ala Leu Gly . Leu Leu Gly Val Ala Gly .
o
5'  GCAGCCAACGGAGUGGACGGUGUGUUUCGCACCUUUGCGGAUUCUUGGCAUGCGCAAGUGGUCUCGGGAU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1540
3'  CGUCGGUUGCCUCACCUGCCACACAAAGCGUGGAAACGCCUAAGAACCGUACGCGUUCACCAGAGCCCUA
1   Ala Ala Asn Gly Val Asp Gly Val Phe Arg Thr Phe Ala Asp Ser Trp His Ala Gln Val Val Ser Gly
2   Gln Pro Thr Glu Trp Thr Val Cys Phe Ala Pro Leu Arg Ile Leu Gly Met Arg Lys Trp Ser Arg Asp
3   Cys Ser Gln Arg Ser Gly Arg Cys Val Ser His Leu Cys Gly Phe Leu Ala Cys Ala Ser Gly Leu Gly Ile
o
5'  ACCUGGGUGGACGCUACGACGACGCUCAUCGCGACGACAGCGAUGUCGAACUUGCGGGUUUCCGCAACUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1610
3'  UGGACCCACCUGCGAUGCUGCUGCGAGUAGCGCUGCUGUCGCUACAGCUUGAACGCCCAAAGGCGUUGAG
1   Tyr Leu Gly Gly Arg Tyr Asp Asp Ala His Arg Asp Asp Ser Asp Val Glu Leu Ala Gly Phe Arg Asn Ser
2   Thr Trp Val Asp Ala Thr Thr Thr Leu Ile Ala Thr Thr Ala Met Ser Asn Leu Arg Val Ser Ala Thr
3   Pro Gly Trp Thr Leu Arg Arg Arg Ser Ser Arg Arg Gln Arg Cys Arg Thr Cys Gly Phe Pro Gln Leu
o
5'  CUUCGGCUAUUCAGACACUGCCACGAUGGGUCUGCCACGUUUCAGCGUGGCAGCGAUUGCUCCACUGGUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1680
3'  GAAGCCGAUAAGUCUGUGACGGUGCUACCCAGACGGUGCAAAGUCGCACCGUCGCUAACGAGGUGACCAA
1   Phe Gly Tyr Ser Asp Thr Ala Thr Met Gly Leu Pro Arg Phe Ser Val Ala Ala Ile Ala Pro Leu Val
2   Pro Ser Ala Ile Gln Thr Leu Pro Arg Trp Val Cys His Val Ser Ala Trp Gln Arg Leu Leu His Trp Leu
3   Leu Arg Leu Phe Arg His Cys His Asp Gly Ser Ala Thr Phe Gln Arg Gly Ser Asp Cys Ser Thr Gly
o
```

```
5'   GCAGGGAUACACGAGAUUGUUGCGGUGCCGCCUGCCGAUACCACGUUCACAGACGACAACUGGCGUGCGA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1750
3'   CGUCCCUAUGUGCUCUAACAACGCCACGGCGGACGGCUAUGGUGCAAGUGUCUGCUGUUGACCGCACGCU
1     Ala Gly Ile His Glu Ile Val Ala Val Pro Pro Ala Asp Thr Thr Phe Thr Asp Asp Asn Trp Arg Ala
2      Gln Gly Tyr Thr Arg Leu Leu Arg Cys Arg Leu Pro Ile Pro Arg Ser Gln Thr Thr Thr Gly Val Arg
3    Cys Arg Asp Thr Arg Asp Cys Cys Gly Ala Ala Cys Arg Tyr His Val His Arg Arg Gln Leu Ala Cys Glu
o
5'   AGACCCGGCACUACGUCUCAAAGCUUCAUGUCGCGCACGUGUUCACUUGUGCUGAGCGACAUGUCUGGCC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1820
3'   UCUGGGCCGUGAUGCAGAGUUUCGAAGUACAGCGCGUGCACAAGUGAACACGACUCGCUGUACAGACCGG
1    Lys Thr Arg His Tyr Val Ser Lys Leu His Val Ala His Val Phe Thr Cys Ala Glu Arg His Val Trp Pro
2     Arg Pro Gly Thr Thr Ser Gln Ser Phe Met Ser Arg Thr Cys Ser Leu Val Leu Ser Asp Met Ser Gly
3      Asp Pro Ala Leu Arg Leu Lys Ala Ser Cys Arg Ala Arg Val His Leu Cys . Ala Thr Cys Leu Ala
o
5'   CGACGGCUUAGACACGUGGCCUUUCGAUCACAGGGCAGUUGUGGCGAGCAUGCCGAAGAAGUUGCAACCG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1890
3'   GCUGCCGAAUCUGUGCACCGGAAAGCUAGUGUCCCGUCAACACCGCUCGUACGGCUUCUUCAACGUUGGC
1      Asp Gly Leu Asp Thr Trp Pro Phe Asp His Arg Ala Val Val Ala Ser Met Pro Lys Lys Leu Gln Pro
2    Pro Thr Ala . Thr Arg Gly Leu Ser Ile Thr Gly Gln Leu Trp Arg Ala Cys Arg Arg Ser Cys Asn Arg
3     Arg Arg Leu Arg His Val Ala Phe Arg Ser Gln Gly Ser Cys Gly Glu His Ala Glu Glu Val Ala Thr
o
5'   CUGGCGUACCUUGUCAAUGCUCAUCGCGGACAAUUGGCAGCCUCGAAUUGGCAGAUCUUGGAGGCGGUGC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1960
3'   GACCGCAUGGAACAGUUACGAGUAGCGCCUGUUAACCGUCGGAGCUUAACCGUCUAGAACCUCCGCCACG
1     Leu Ala Tyr Leu Val Asn Ala His Arg Gly Gln Leu Ala Ala Ser Asn Trp Gln Ile Leu Glu Ala Val
2      Trp Arg Thr Leu Ser Met Leu Ile Ala Asp Asn Trp Gln Pro Arg Ile Gly Arg Ser Trp Arg Arg Cys
3    Ala Gly Val Pro Cys Gln Cys Ser Ser Arg Thr Ile Gly Ser Leu Glu Leu Ala Asp Leu Gly Gly Gly Ala
o
5'   ACAGGUUUAAGGGUACUGGUAAGAGUACGGCGUACUACUCUUAUGACGUGCGGGGCGAGUGGCGGCCGGA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2030
3'   UGUCCAAAUUCCCAUGACCAUUCUCAUGCCGCAUGAUGAGAAUACUGCACGCCCCGCUCACCGCCGGCCU
1    His Arg Phe Lys Gly Thr Gly Lys Ser Thr Ala Tyr Tyr Ser Tyr Asp Val Arg Gly Glu Trp Arg Pro Asp
2     Thr Gly Leu Arg Val Leu Val Arg Val Arg Arg Thr Thr Leu Met Thr Cys Gly Ala Ser Gly Gly Arg
3      Gln Val . Gly Tyr Trp . Glu Tyr Gly Val Leu Leu Leu . Arg Ala Gly Arg Val Ala Ala Gly
o
5'   CAAAUAUGGCGUGGUUCACGUGAAUGAAGCGACCACGGCAAUCGCCGCUCUCGACGCCGCCGUCAGGCUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2100
3'   GUUUAUACCGCACCAAGUGCACUUACUUCGCUGGUGCCGUUAGCGGCGAGAGCUGCGGCGGCAGUCCGAG
1      Lys Tyr Gly Val Val His Val Asn Glu Ala Thr Thr Ala Ile Ala Ala Leu Asp Ala Ala Val Arg Leu
2    Thr Asn Met Ala Trp Phe Thr . Met Lys Arg Pro Arg Gln Ser Pro Leu Ser Thr Pro Pro Ser Gly Ser
3     Gln Ile Trp Arg Gly Ser Arg Glu . Ser Asp His Gly Asn Arg Arg Ser Arg Arg Arg Arg Gln Ala
o
5'   ACGUGGAACCUCAACGGCACGCUGGAUGCCGAUGGAGUCGCGCGUGACGCCUUCAGUCCUGUGCGCGACC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2170
3'   UGCACCUUGGAGUUGCCGUGCGACCUACGGCUACCUCAGCGCGCACUGCGGAAGUCAGGACACGCGCUGG
1     Thr Trp Asn Leu Asn Gly Thr Leu Asp Ala Asp Gly Val Ala Arg Asp Ala Phe Ser Pro Val Arg Asp
2      Arg Gly Thr Ser Thr Ala Arg Trp Met Pro Met Glu Ser Arg Val Thr Pro Ser Val Leu Cys Ala Thr
3    His Val Glu Pro Gln Arg His Ala Gly Cys Arg Trp Ser Arg Ala . Arg Leu Gln Ser Cys Ala Arg Pro
o
5'   UGAACGGACUGUAUGCACCGCUGGCGUUCGCACCGAUCGCGAGGAGGAACAGGAGGCUCGGAGGGGUUCG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2240
3'   ACUUGCCUGACAUACGUGGCGACCGCAAGCGUGGCUAGCGCUCCUCCUUGUCCUCCGAGCCUCCCCAAGC
1    Leu Asn Gly Leu Tyr Ala Pro Leu Ala Phe Ala Pro Ile Ala Arg Arg Asn Arg Arg Leu Gly Gly Val Arg
2     . Thr Asp Cys Met His Arg Trp Arg Ser His Arg Ser Arg Gly Gly Thr Gly Gly Ser Glu Gly Phe
3      Glu Arg Thr Val Cys Thr Ala Gly Val Arg Thr Asp Arg Glu Glu Glu Gln Glu Ala Arg Arg Gly Ser
o
```

```
5'  UCUGGCUGCAGGUGUUGGAAUGGAUGCUCUGGGGAAGGGAGCGAACACAACGUACGACACGGUCAACGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2310
3'  AGACCGACGUCCACAACCUUACCUACGAGACCCCUUCCCUCGCUUGUGUUGCAUGCUGUGCCAGUUGCGC
1    Leu Ala Ala Gly Val Gly Met Asp Ala Leu Gly Lys Gly Ala Asn Thr Thr Tyr Asp Thr Val Asn Ala
2   Val Trp Leu Gln Val Leu Glu Trp Met Leu Trp Gly Arg Glu Arg Thr Gln Arg Thr Thr Arg Ser Thr Arg
3    Ser Gly Cys Arg Cys Trp Asn Gly Cys Ser Gly Glu Gly Ser Glu His Asn Val Arg His Gly Gln Arg
o
5'  UACUAGUUCGCACGUCAUUUUCAUGCUAUGUCCGUCGAUUCAACCUUCUUUCCGCGUCAUACAUUGAAAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2380
3'  AUGAUCAAGCGUGCAGUAAAAGUACGAUACAGGCAGCUAAGUUGGAAGAAAGGCGCAGUAUGUAACUUUU
1    Tyr . Phe Ala Arg His Phe His Ala Met Ser Val Asp Ser Thr Phe Phe Pro Arg His Thr Leu Lys
2     Thr Ser Ser His Val Ile Phe Met Leu Cys Pro Ser Ile Gln Pro Ser Phe Arg Val Ile His . Lys
3   Val Leu Val Arg Thr Ser Phe Ser Cys Tyr Val Arg Arg Phe Asn Leu Leu Ser Ala Ser Tyr Ile Glu Asn
o
5'  CGGUCAGAACAACGAGAAGAUCGGCGCCUAGCGCAAAACGAUAAAAGCCAUAGCGGCGAGACGUAACGCA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  2450
3'  GCCAGUCUUGUUGCUCUUCUAGCCGCGGAUCGCGUUUUGCUAUUUUCGGUAUCGCCGCUCUGCAUUGCGU
1   Thr Val Arg Thr Thr Arg Arg Ser Ala Pro Ser Ala Lys Arg . Lys Pro . Arg Arg Asp Val Thr His
2    Arg Ser Glu Gln Arg Glu Asp Arg Arg Leu Ala Gln Asn Asp Lys Ser His Ser Gly Glu Thr . Arg
3     Gly Gln Asn Asn Glu Lys Ile Gly Ala . Arg Lys Thr Ile Lys Ala Ile Ala Ala Arg Arg Asn Ala
o
5'  UGCUACGCGAACUCGCUUGCUGAUGAAGGG
o   ++++|++++|++++|++++|++++|++++|                                          2480
3'  ACGAUGCGCUUGAGCGAACGACUACUUCCC
1    Ala Thr Arg Thr Arg Leu Leu Met Lys Gly
2   Met Leu Arg Glu Leu Ala Cys . . Arg
3    Cys Tyr Ala Asn Ser Leu Ala Asp Glu Gly
o
```

3.5.9 DAV1 dsRNA4 sequence

1698 nt excluding poly (A) tail

```
5'  GGCGAGCGAUCCGUUUUGAGGUGCAAUUGCUGCGGAUUACCGCAUCAAUGCACCUGCAGACGAGGCGUGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   70
3'  CCGCUCGCUAGGCAAAACUCCACGUUAACGACGCCUAAUGGCGUAGUUACGUGGACGUCUGCUCCGCACC
1    Gly Glu Arg Ser Val Leu Arg Cys Asn Cys Cys Gly Leu Pro His Gln Cys Thr Cys Arg Arg Gly Val
2     Ala Ser Asp Pro Phe  .  Gly Ala Ile Ala Ala Asp Tyr Arg Ile Asn Ala Pro Ala Asp Glu Ala Trp
3   Arg Arg Ala Ile Arg Phe Glu Val Gln Leu Leu Arg Ile Thr Ala Ser Met His Leu Gln Thr Arg Arg Gly
o
5'  UGAACAAGAAUAAAUCCCACUUUCAUCACGAACAAACGACAAGACGACACCGAGUAUGUCUCGGUCAUUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   140
3'  ACUUGUUCUUAUUUAGGGUGAAAGUAGUGCUUGUUUGCUGUUCUGCUGUGGCUCAUACAGAGCCAGUAAG
1   Val Asn Lys Asn Lys Ser His Phe His His Glu Gln Thr Thr Arg Arg His Arg Val Cys Leu Gly His Ser
2     .  Thr Arg Ile Asn Pro Thr Phe Ile Thr Asn Lys Arg Gln Asp Asp Thr Glu Tyr Val Ser Val Ile
3    Glu Gln Glu  .  Ile Pro Leu Ser Ser Arg Thr Asn Asp Lys Thr Thr Pro Ser Met Ser Arg Ser Phe
o
5'  GGACAGAACAACAUGGAGGCAGAGAUGCGAAGCUACAUCGUGGACUCCGGCAUUCGACAAGGCUAUCAGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   210
3'  CCUGUCUUGUUGUACCUCCGUCUCUACGCUUCGAUGUAGCACCUGAGGCCGUAAGCUGUUCCGAUAGUCC
1     Asp Arg Thr Thr Trp Arg Gln Arg Cys Glu Ala Thr Ser Trp Thr Pro Ala Phe Asp Lys Ala Ile Arg
2   Arg Thr Glu Gln His Gly Gly Arg Asp Ala Lys Leu His Arg Gly Leu Arg His Ser Thr Arg Leu Ser Gly
3    Gly Gln Asn Asn Met Glu Ala Glu Met Arg Ser Tyr Ile Val Asp Ser Gly Ile Arg Gln Gly Tyr Gln
o
5'  GACCGGUUCUCGCCUCGCUAGCUAAGUUCGUGGCAGAUAUUGAAGGAAUCGCGGACGACUUCAGUACCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   280
3'  CUGGCCAAGAGCGGAGCGAUCGAUUCAAGCACCGUCUAUAACUUCCUUAGCGCCUGCUGAAGUCAUGGAC
1    Asp Arg Phe Ser Pro Arg  .  Leu Ser Ser Trp Gln Ile Leu Lys Glu Ser Arg Thr Thr Ser Val Pro
2     Thr Gly Ser Arg Leu Ala Ser  .  Val Arg Gly Arg Tyr  .  Arg Asn Arg Gly Arg Leu Gln Tyr Leu
3   Gly Pro Val Leu Ala Ser Leu Ala Lys Phe Val Ala Asp Ile Glu Gly Ile Ala Asp Asp Phe Ser Thr Trp
o
5'  GACACUCGCGCAGGGAGUGGACCUGCAGACGGAAGAGGAUGAUGCAGUGAUCAACGGUGCACUGCUGGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   350
3'  CUGUGAGCGCGUCCCUCACCUGGACGUCUGCCUUCUCCUACUACGUCACUAGUUGCCACGUGACGACCGC
1   Gly His Ser Arg Arg Glu Trp Thr Cys Arg Arg Lys Arg Met Met Gln  .  Ser Thr Val His Cys Trp Arg
2    Asp Thr Arg Ala Gly Ser Gly Pro Ala Asp Gly Arg Gly  .  Cys Ser Asp Gln Arg Cys Thr Ala Gly
3     Thr Leu Ala Gln Gly Val Asp Leu Gln Thr Glu Glu Asp Asp Ala Val Ile Asn Gly Ala Leu Leu Ala
o
5'  UACACUGAGGACGUGUUCUUGCAAUUUGCCAUUGCACACGACCCCACGAUUAGAGACGCUGGUGGGAAUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   420
3'  AUGUGACUCCUGCACAAGAACGUUAAACGGUAACGUGUGCUGGGGUGCUAAUCUCUGCGACCACCCUUAC
1     Thr Leu Arg Thr Cys Ser Cys Asn Leu Pro Leu His Thr Thr Pro Arg Leu Glu Thr Leu Val Gly Met
2   Val His  .  Gly Arg Val Leu Ala Ile Cys His Cys Thr Arg Pro His Asp  .  Arg Arg Trp Trp Glu Cys
3    Tyr Thr Glu Asp Val Phe Leu Gln Phe Ala Ile Ala His Asp Pro Thr Ile Arg Asp Ala Gly Gly Asn
o
5'  CUGUGGUGUACGGGCCUGAGCCUCCACCGAAGCAGUCAGCCACAUCGCUGGUUUCGUCAUCGGACAGUUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   490
3'  GACACCACAUGCCCGGACUCGGAGGUGGCUUCGUCAGUCGGUGUAGCGACCAAAGCAGUAGCCUGUCAAG
1    Leu Trp Cys Thr Gly Leu Ser Leu His Arg Ser Ser Gln Pro His Arg Trp Phe Arg His Arg Thr Val
2     Cys Gly Val Arg Ala  .  Ala Ser Thr Glu Ala Val Ser His Ile Ala Gly Phe Val Ile Gly Gln Phe
3   Ala Val Val Tyr Gly Pro Glu Pro Pro Pro Lys Gln Ser Ala Thr Ser Leu Val Ser Ser Ser Asp Ser Ser
o
5'  UGGAGGCGCGCCACUGGAUGCGGAAAAGGCGGACGACAACGAGUGUGCGGUUGAGGAGGAUGUUGCUCCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   560
3'  ACCUCCGCGCGGUGACCUACGCCUUUUCCGCCUGCUGUUGCUCACACGCCAACUCCUCCUACAACGAGGC
1   Leu Glu Ala Arg His Trp Met Arg Lys Arg Arg Thr Thr Thr Ser Val Arg Leu Arg Arg Met Leu Leu Arg
2    Trp Arg Arg Ala Thr Gly Cys Gly Lys Gly Gly Arg Gln Arg Val Cys Gly  .  Gly Gly Cys Cys Ser
3     Gly Gly Ala Pro Leu Asp Ala Glu Lys Ala Asp Asp Asn Glu Cys Ala Val Glu Glu Asp Val Ala Pro
o
```

```
5'  AGCGUCGGACCGGCUGGUUUCGAUGGUAUCGUAGUUGAGGAGCGUGGCUCUAUGCUCCCGAACAAAGCGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   630
3'  UCGCAGCCUGGCCGACCAAAGCUACCAUAGCAUCAACUCCUCGCACCGAGAUACGAGGGCUUGUUUCGCU
1     Ala Ser Asp Arg Leu Val Ser Met Val Ser  .  Leu Arg Ser Val Ala Leu Cys Ser Arg Thr Lys Arg
2   Glu Arg Arg Thr Gly Trp Phe Arg Trp Tyr Arg Ser  .  Gly Ala Trp Leu Tyr Ala Pro Glu Gln Ser Glu
3    Ser Val Gly Pro Ala Gly Phe Asp Gly Ile Val Val Glu Glu Arg Gly Ser Met Leu Pro Asn Lys Ala
o
5'  AGACAUACGACGCCAUCAACUACGCCGCCCAGCAAGAGAGCAAGAUCGGGGAUGCGACUGCACGUGCGGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   700
3'  UCUGUAUGCUGCGGUAGUUGAUGCGGCGGGUCGUUCUCUCGUUCUAGCCCCUACGCUGACGUGCACGCCU
1    Arg His Thr Thr Pro Ser Thr Thr Pro Pro Ser Lys Arg Ala Arg Ser Gly Met Arg Leu His Val Arg
2     Asp Ile Arg Arg His Gln Leu Arg Arg Pro Ala Arg Glu Gln Asp Arg Gly Cys Asp Cys Thr Cys Gly
3   Lys Thr Tyr Asp Ala Ile Asn Tyr Ala Ala Gln Gln Glu Ser Lys Ile Gly Asp Ala Thr Ala Arg Ala Glu
o
5'  GGUCGGACGCGAUAAGCCGAGUGUCGAGGAGGAGUUUGGCGUCUUGCGGCUUGAGGUUCCGCCUGACAAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   770
3'  CCAGCCUGCGCUAUUCGGCUCACAGCUCCUCCUCAAACCGCAGAACGCCGAACUCCAAGGCGGACUGUUU
1   Arg Ser Asp Ala Ile Ser Arg Val Ser Arg Arg Ser Leu Ala Ser Cys Gly Leu Arg Phe Arg Leu Thr Lys
2    Gly Arg Thr Arg  .  Ala Glu Cys Arg Gly Gly Val Trp Arg Leu Ala Ala  .  Gly Ser Ala  .  Gln
3     Val Gly Arg Asp Lys Pro Ser Val Glu Glu Glu Phe Gly Val Leu Arg Leu Glu Val Pro Pro Asp Lys
o
5'  GGCGCUGUGCGCAGGAGACUCAUGUUCAUCAUGGAGCAAGAACAGAUGUGCGACGUGUACGUCUGUGGUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   840
3'  CCGCGACACGCGUCCUCUGAGUACAAGUAGUACCUCGUUCUUGUCUACACGCUGCACAUGCAGACACCAC
1     Ala Leu Cys Ala Gly Asp Ser Cys Ser Ser Trp Ser Lys Asn Arg Cys Ala Thr Cys Thr Ser Val Val
2   Arg Arg Cys Ala Gln Glu Thr His Val His His Gly Ala Arg Thr Asp Val Arg Arg Val Arg Leu Trp Cys
3    Gly Ala Val Arg Arg Arg Leu Met Phe Ile Met Glu Gln Glu Gln Met Cys Asp Val Tyr Val Cys Gly
o
5'  UCAUCGAAGCGUUGUUGGACGCUAAGGGGAUACCUAGGCCUGAACUCGUUCUGGCCUCAAUUCUCAUGAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   910
3'  AGUAGCUUCGCAACAACCUGCGAUUCCCCUAUGGAUCCGGACUUGAGCAAGACCGGAGUUAAGAGUACUG
1    Ser Ser Lys Arg Cys Trp Thr Leu Arg Gly Tyr Leu Gly Leu Asn Ser Phe Trp Pro Gln Phe Ser  .
2     His Arg Ser Val Val Gly Arg  .  Gly Asp Thr  .  Ala  .  Thr Arg Ser Gly Leu Asn Ser His Asp
3   Val Ile Glu Ala Leu Leu Asp Ala Lys Gly Ile Pro Arg Pro Glu Leu Val Leu Ala Ser Ile Leu Met Thr
o
5'  GAGCAAGGAGUACGAAGUGCGGCGUGCAGUGGCGCAAACCCUCAAGUCGUUUGUCAUGGAUGAUUUCGAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   980
3'  CUCGUUCCUCAUGCUUCACGCCGCACGUCACCGCGUUUGGGAGUUCAGCAAACAGUACCUACUAAAGCUC
1   Arg Ala Arg Ser Thr Lys Cys Gly Val Gln Trp Arg Lys Pro Ser Ser Arg Leu Ser Trp Met Ile Ser Ser
2    Glu Gln Gly Val Arg Ser Ala Ala Cys Ser Gly Ala Asn Pro Gln Val Val Cys His Gly  .  Phe Arg
3     Ser Lys Glu Tyr Glu Val Arg Arg Ala Val Ala Gln Thr Leu Lys Ser Phe Val Met Asp Asp Phe Glu
o
5'  UGCACGAAUGGUUCGCUUUGGCGUGAGCGCGGUCAACAUUACCUCAUUGCCACAUCUGCGGAGUUGGAGG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1050
3'  ACGUGCUUACCAAGCGAAACCGCACUCGCGCCAGUUGUAAUGGAGUAACGGUGUAGACGCCUCAACCUCC
1     Ala Arg Met Val Arg Phe Gly Val Ser Ala Val Asn Ile Thr Ser Leu Pro His Leu Arg Ser Trp Arg
2   Val His Glu Trp Phe Ala Leu Ala  .  Ala Arg Ser Thr Leu Pro His Cys His Ile Cys Gly Val Gly Gly
3    Cys Thr Asn Gly Ser Leu Trp Arg Glu Arg Gly Gln His Tyr Leu Ile Ala Thr Ser Ala Glu Leu Glu
o
5'  AGGGCACAGACGACAAGGCUCUGUCGUUUGACGAACGGAAGAAGUUGUCCAGGCUCAGCCGCGCGGUUUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1120
3'  UCCCGUGUCUGCUGUUCCGAGACAGCAAACUGCUUGCCUUCUUCAACAGGUCCGAGUCGGCGCGCCAAAC
1    Arg Ala Gln Thr Thr Arg Leu Cys Arg Leu Thr Asn Gly Arg Ser Cys Pro Gly Ser Ala Ala Arg Phe
2     Gly His Arg Arg Gln Gly Ser Val Val  .  Arg Thr Glu Glu Val Val Gln Ala Gln Pro Arg Gly Leu
3   Glu Gly Thr Asp Asp Lys Ala Leu Ser Phe Asp Glu Arg Lys Lys Leu Ser Arg Leu Ser Arg Ala Val Cys
o
```

```
5'  UGUCACGGCGAUAUGCGUGCGACAGAAGGAUACGGGCUUCGGUAGGCUGCUGCAGCACUCUACGCGCAUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1190
3'  ACAGUGCCGCUAUACGCACGCUGUCUUCCUAUGCCCGAAGCCAUCCGACGACGUCGUGAGAUGCGCGUAC
1   Val Ser Arg Arg Tyr Ala Cys Asp Arg Arg Ile Arg Ala Ser Val Gly Cys Cys Ser Thr Leu Arg Ala  .
2    Cys His Gly Asp Met Arg Ala Thr Glu Gly Tyr Gly Leu Arg  .  Ala Ala Ala Ala Leu Tyr Ala His
3     Val Thr Ala Ile Cys Val Arg Gln Lys Asp Thr Gly Phe Gly Arg Leu Leu Gln His Ser Thr Arg Met
o
5'  AACCACUUCUGGCGCCGUGAAUCCUCGAAAGCGCGCGAGGGUGGACUUGAGGCCAUGAGUGGCCUUGACG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1260
3'  UUGGUGAAGACCGCGGCACUUAGGAGCUUUCGCGCGCUCCCACCUGAACUCCGGUACUCACCGGAACUGC
1     Thr Thr Ser Gly Ala Val Asn Pro Arg Lys Arg Ala Arg Val Asp Leu Arg Pro  .  Val Ala Leu Thr
2   Glu Pro Leu Leu Ala Pro  .  Ile Leu Glu Ser Ala Arg Gly Trp Thr  .  Gly His Glu Trp Pro  .  Arg
3    Asn His Phe Trp Arg Arg Glu Ser Ser Lys Ala Arg Glu Gly Gly Leu Glu Ala Met Ser Gly Leu Asp
o
5'  ACUUUCGGUUGGAUUUGUGCUGCCGUACCAGCGCUACGCGGGUGGGGUACAUAUGGGCUUUGAGGCCCGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1330
3'  UGAAAGCCAACCUAAACACGACGGCAUGGUCGCGAUGCGCCCACCCCAUGUAUACCCGAAACUCCGGGCU
1    Thr Phe Gly Trp Ile Cys Ala Ala Val Pro Ala Leu Arg Gly Trp Gly Thr Tyr Gly Leu  .  Gly Pro
2     Leu Ser Val Gly Phe Val Leu Pro Tyr Gln Arg Tyr Ala Gly Gly Val His Met Gly Phe Glu Ala Arg
3   Asp Phe Arg Leu Asp Leu Cys Cys Arg Thr Ser Ala Thr Arg Val Gly Tyr Ile Trp Ala Leu Arg Pro Glu
o
5'  GGAACGCGUGACGGCACUGGGGAUGCAUGGUUGUUCGAUGACCUUUCGGGAGUUGGCGAACUCGGAUGUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1400
3'  CCUUGCGCACUGCCGUGACCCCUACGUACCAACAAGCUACUGGAAAGCCCUCAACCGCUUGAGCCUACAA
1   Arg Asn Ala  .  Arg His Trp Gly Cys Met Val Val Arg  .  Pro Phe Gly Ser Trp Arg Thr Arg Met Phe
2    Gly Thr Arg Asp Gly Thr Gly Asp Ala Trp Leu Phe Asp Asp Leu Ser Gly Val Gly Glu Leu Gly Cys
3     Glu Arg Val Thr Ala Leu Gly Met His Gly Cys Ser Met Thr Phe Arg Glu Leu Ala Asn Ser Asp Val
o
5'  CGUAUCAGGAAGCGGAACCCCAGAUUGGCAAUUGAGUUCGGCGACGACGGGAGGCCUGGCCUGGUGUCAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1470
3'  GCAUAGUCCUUCGCCUUGGGGUCUAACCGUUAACUCAAGCCGCUGCUGCCCUCCGGACCGGACCACAGUC
1     Val Ser Gly Ser Gly Thr Pro Asp Trp Gln Leu Ser Ser Ala Thr Thr Gly Gly Leu Ala Trp Cys Gln
2   Ser Tyr Gln Glu Ala Glu Pro Gln Ile Gly Asn  .  Val Arg Arg Arg Arg Glu Ala Trp Pro Gly Val Ser
3    Arg Ile Arg Lys Arg Asn Pro Arg Leu Ala Ile Glu Phe Gly Asp Asp Gly Arg Pro Gly Leu Val Ser
o
5'  CGCGGCGCAACGGCGCGUUUUGGAGCCAGGUGAACGUUGUGGCAGUCAAUUGGGGGACGUGGUGGUUCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1540
3'  GCGCCGCGUUGCCGCGCAAAACCUCGGUCCACUUGCAACACCGUCAGUUAACCCCCUGCACCACCAAGG
1    Arg Gly Ala Thr Ala Arg Phe Gly Ala Arg  .  Thr Leu Trp Gln Ser Ile Gly Gly Thr Trp Trp Phe
2     Ala Ala Gln Arg Arg Val Leu Glu Pro Gly Glu Arg Cys Gly Ser Gln Leu Gly Gly Arg Gly Gly Ser
3   Ala Arg Arg Asn Gly Ala Phe Trp Ser Gln Val Asn Val Val Ala Val Asn Trp Gly Asp Val Val Val Pro
o
5'  UGCUUAGGUGGUUGCAGCAUUCAGAUUCCGUGUCACGCACUGGUUUUGUGCUUCCUCAACACUCCGGGAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1610
3'  ACGAAUCCACCAACGUCGUAAGUCUAAGGCACAGUGCGUGACCAAAACACGAAGGAGUUGUGAGGCCCUC
1   Leu Leu Arg Trp Leu Gln His Ser Asp Ser Val Ser Arg Thr Gly Phe Val Leu Pro Gln His Ser Gly Ser
2    Cys Leu Gly Gly Cys Ser Ile Gln Ile Pro Cys His Ala Leu Val Leu Cys Phe Leu Asn Thr Pro Gly
3     Ala  .  Val Val Ala Ala Phe Arg Phe Arg Val Thr His Trp Phe Cys Ala Ser Ser Thr Leu Arg Glu
o
5'  CACCCAGACUCUUCGCCAGUCUGUAAACAGGAGUUCAUAAGCCAUGGGCAGCUAGACGUAGCAUUCACGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|  1680
3'  GUGGGUCUGAGAAGCGGUCAGACAUUUGUCCUCAAGUAUUCGGUACCCGUCGAUCUGCAUCGUAAGUGCU
1     Thr Gln Thr Leu Arg Gln Ser Val Asn Arg Ser Ser  .  Ala Met Gly Ser  .  Thr  .  His Ser Arg
2   Ala Pro Arg Leu Phe Ala Ser Leu  .  Thr Gly Val His Lys Pro Trp Ala Ala Arg Arg Ser Ile His Asp
3    His Pro Asp Ser Ser Pro Val Cys Lys Gln Glu Phe Ile Ser His Gly Gln Leu Asp Val Ala Phe Thr
o
5'   UGUGAACUAGGUGUCGCC
o   ++++|++++|++++|+++
3'   ACACUUGAUCCACAGCGG
1     Cys Glu Leu Gly Val Ala
2      Val Asn  .  Val Ser Pro
3   Met  .  Thr Arg Cys Arg
o
```

4. Chapter 4 - Identification of a Novel Deltaflexivirus, Fusarium Deltaflexivirus 2 (FDFV2), Infecting *Fusarium oxysporum* f. sp. *melon*

4.1 Introduction

4.1.1 *Fusarium oxsporum*

Fusarium oxysporum is an Ascomycete fungus selected as no. 5 of the top ten plant pathogenic fungi in the world in 2012 in the journal *Molecular Plant Pathology* (Dean *et al*., 2012). *F. oxysporum* is a common soil-borne plant pathogen having a wide host range, caused agricultural economic damage. The main symptoms include wilting, browning of vascular bundles and even plant death.

4.1.2 *Deltaflexiviridae*

According to ICTV (https://ictv.global/taxonomy/), *Deltaflexiviridae* was established in 2017, with three deltaflexviruses—Sclerotinia sclerotiorum deltaflexivirus 1 (SsDFV1) (Li *et al*., 2016), Soybean leaf-associated mycoflexivirus 1 (SlaMFV1) (Marzano and Domier, 2016), and Fusarium graminearum deltaflexivirus 1 (FgDFV1) (Chen *et al*., 2016). Until now, more and more deltaflexiviruses have been reported (Table 4-1). In general, all the deltaflexiviruses are mycoviruses, with linear (+) ssRNA genome of 6-8 kb in size. However, no virions have been observed yet (Li *et al*., 2016; Hamid *et al*., 2018).

In this study, I identified a novel deltaflexivirus, Fusarium deltaflexivirus 2 (FDFV2) in *F. oxysporum* f.sp. *melonis*. Through virus particle purification and protein analysis by LC-MS/MS, the results suggested that the 8.2 kbp ssRNA genomes were packaged in a isometric or filamentous particles consisted with the 18 kDa proteins. To my best knowledge, this is the first discovery regarding the virions of *Deltaflexivirdae*.

4.2 Materials and methods

4.2.1 Fungal strains and culture conditions

Fusarium oxysporum f. sp. *melonis* isolate Fom 405 (race 1, 2y) was discovered from Fusarium wilt of melon in Ibaraki prefecture of Japan in 2006. The isolation and identification were conducted by Dr. Okada. The isolates Fom 405-W and Fom 405-P were isolated during subculture of Fom 405. All the isolate grown on D-PDA media (24 g/L Difico^{TM} Potato dextrose powder and 15 g/L agar) at 25°C for 7 days, and the mycelial discs were stored at -80°C in glycerol. For liquid culture of dsRNA extraction and virus particle purification, mycelial discs were inoculated in PDB (24 g/L Difico^{TM} Potato dextrose powder) and cultured with reciprocal shaking (60 strokes per min) at 25°C for one week.

4.2.2 Determination of viral RNA sequence and phylogenetic analysis

The viral dsRNA was purified from strain Fom 405 by the spin column method described in section 2.2.5. The complete viral sequence was determined by Dr. Okada with the same procedures in section 3.2.3.

The sequences of *Tymovirales* were retrieved from Genbank (Table 4-1), and followed the procedures in section 2.2.9 to build the phylogenetic tree.

4.2.3 Detection of subgenomic RNA

The total RNAs were purified from 0.2 g dried mycelium of strain Fom 405 by Trizol® reagent (Thermo Fisher) following the manufacturer's protocol.

For DIG northern hybridization, 10 μg of total RNA per sample was subjected to northern hybridization analysis following the procedures in section 2.2.7. The primers for

DIG DNA probes synthesis were listed in Table 4-2. The positions of DIG probes were shown in Figur 4-3A.

The 5' rapid amplification of cDNA ends (RACE) was performed with SMARTer® RACE 5'/3' Kit (Takara Bio USA, Inc., CA, USA), and 1 μg total RNA was subjected to the analysis. The primers (Table 4-2)— FDFV2-NB-ORF1-R, FDFV2-NB-ORF2-R, FDFV2-NB-ORF3-R, FDFV2-NB-ORF44-R, and FDFV2-NB-ORF5-R with UPM primer of SMARTer® RACE 5'/3' Kit were used for specific PCR detection following manufacturer's protocol.

4.2.4 Purification of virus particle and analysis of virus-associated proteins

The following purification procedure was conducted under 4°C. For partial purification, the protocol followed the published article (Howitt *et al.*, 1995) with modification. In brief, 20 g dried mycelium was ground to a fine powder with liquid nitrogen, then mixed well with 40 ml buffer A (0.1 M sodium phosphate, 0.2 M KCl, pH 7.4). The mixture was mixed with 20 ml chloroform and vortexed, then centrifuged at 11,000 × g (TOMY Suprema 21, NA-8 rotor, Japan) for 20 min. The upper aqueous phase of the supernatant was ultracentrifuged at 120,000 × g (Hitachi CP80WX, P80AT rotor) for 90 min. The resultant pellet was resuspended in 3 ml buffer A with gentle rotation overnight. The suspension was transferred to three new 1.5 ml tubes and centrifuged at 2000 × g for 10 min to remove debris. The supernatant was filled up to 10 ml with buffer A, then ultracentrifuged at 120,000 × g (Hitachi CP80WX, P80AT rotor) for 90 min. The resultant pellet was resuspended with 0.5 ml buffer A in a 1.5 ml tube overnight. Finally, the suspension was centrifuged at 1,700 × g for 5 min to remove debris, then stored at -20°C.

For purification of sucrose gradient, the protocol modified from published article (Jia *et al*., 2017). Ten grams of dried mycelium was ground to a fine powder with liquid nitrogen, and mixed with 100 ml buffer A, then centrifuged at 12,100 × g for 30 min. The resultant supernatant was subjected to ultracentrifugation at 110,000 × g (Hitachi CP80WX, P45AT rotor) for 1 h, then the pellet was resuspended in 2 ml buffer A overnight with gentle rotation. The suspension was layered on 20 %-60 % sucrose gradient and centrifuged at 50,200 × g (Hitachi CP80WX, P28S rotor) for 3h, then fractioned into 20 fractions. All the fractions were stored at -20°C.

The purified virus suspensions were confirmed by TEM following the procedures in section 2.2.3. The virus-associated proteins were analyzed by SDS-PAGE and LC-MS/MS described in section 2.2.4. The presence of viral RNAs in virus suspensions were confirmed by RT-PCR with specific primer pair for FDFV2 (Table 4-2)

4.3 Results

4.3.1 Morphologies of colonies and genome organization of FDFV2

F. oxysporum strain Fom 405, a dsRNA segment of about 8.2 kbp in size was isolated by the spin column method (Figure 4-1A). After full-length sequencing, blast search, and phylogenetic analysis (Figure 4-2), it was identified as a new deltaflexivirus (8125 bp excluding poly (A) tail), called "Fusarium deltaflexivirus 2 (FDFV2)". Two isolates with different colony morphologies, Fom 405-W and Fom 405-P, were isolated during the subculture process.

Fom 405 (original) colony were the fastest growing with no visible pigment accumulation (Figure 4-1B). Fom 405-W colony grew slowly, with irregular colony edges and white hyphae in the center of the colony, and had the highest amount of red pigment accumulation among the three isolates (Figure 4-1B). Fom 405-P colony was also slow growing, with irregular colony edges, medium accumulation of red pigment, and no white hyphae in the center of the colony (Figure 4-1B). 8.2 kbp dsRNA segments were detected from both Fom 405-W and Fom 405-P (Figure 4-1A).

The genome of FDFV2 has five ORFs, of which ORF1 encodes a 220 kDa polyprotein related to replication, including domains of methyltransferase, helicase and RdRp (Figure 4-1C). ORF4 encodes a protein of about 18 kDa (Figure 4-1C), which was shown to be a possible component of the viral particles in this study. ORF2-, ORF3-, and ORF5-encoded proteins are hypothetical proteins (Figure 4-1C). The nucleotide sequence of FDFV2 will be deposited in Genbank of NCBI (Appendix 4.5.1).

4.3.2 Detection of subgenomic RNA of FDFV2

I failed to extract the total nucleic acid with the RNeasy Plant Mini kit (Qiagen, Germany) (data not shown), while using the Trizol® reagent (Thermo Fisher) successfully extracted the total nucleic acid in high concentration from Fom 405 mycelia.

In northern hybridization, the DIG DNA probes for each ORFs and mixed probes of FDFV2 detected only 8125 nt full-length segments of FDFV2 (Figure 4-3B). Only the full-length segment of FDFV2 was detected in detection of mixing five types of probes (Figure 4-3B). None of the above northern hybridization analysis detected any subgenomic RNA in FDFV2 (Figure 4-3B).

The result of 5’ RACE (Figure 4-3C) revealed that the specific primers amplified the amplicons which might correspond to the hypothesized subgenomes (Figure 4-3D).

4.3.3 Virus-associated particle and virus-associated proteins

To purify the viral particles of FDFV2 or virus-related proteins, I first performed a partial purification, and the resultant purified suspension was confirmed by SDS-PAGE and three protein bands were observed (Figure 4-4A). The purified suspension was used for TEM analysis and two kinds of particle form were observed, one was isometric particles (~35 nm in diameter; Figure 4-4B, yellow arrows) and the other was filamentous particles (~800-1000 nm in length; Figure 4-4B, red arrows).

To separate the two kinds of virus-like particles (VLPs), I used 20-60% sucrose gradient for further purification. Total nucleic acid of each fractions purified by sucrose gradient was extracted and visualized by agarose gel electrophoresis, and only dsRNA bands of about 8.2 kbp in size were observed in fraction 2 (Figure 4-5A). Since FDFV2 is a ssRNA virus, RT-PCR was performed for all the fractions to detect FDFV2 using virus specific primers. The virus was detected from all fractions with a target band of about 550 bp (Figure 4-5B). In the SDS-PAGE electrophoresis, two protein bands were detected in fraction 2 (Figure 4-5C), which were the same size as the bands in the partial purification (Figure 4-4A). No significant bands were observed in fractions 3-20 (Figure 4-5C), but the RNA genome of FDFV2 could be detected by RT-PCR (Figure 4-4B). Therefore, I mixed the fractions 3-20 and concentrated for SDS-PAGE. A protein band of about 18 kDa was observed in SDS-PAGE electrophoresis (Figure 4-5D), which was the same size as that of partial purification (Figure 4-4A).

I used TEM to observe resuspensions of fraction 2 and mixed fraction. In fraction 2, it was observed that most of the particles were isometric and a few were filamentous (Figure 4-5E). In mixed fraction, clear filamentous particles and very few isometric particles were observed (Figure 4-5F).

To clarify the relation between the protein bands from SDS-PAGE (Figures 4-4A and 4-5D) and virus particles, I collected the protein bands and subjected to the in-gel digestion, then analyzed by LC-MS/MS. Comparison with the database showed that the

18.2 kDa protein was encoded by FDFV2 ORF4 (Figure 4-7C). To the best of my knowledge, this is the first finding related to virus particle and viral protein of deltaflexiviruses. The 49.7 kDa and 43 kDa proteins were from the fungal host *F. oxysporum* (Figures 4-7A and B).

4.4 Discussion

To the best of my knowledge, no viral particles of deltaflexivirus have been found so far. In this study, I identified a new deltaflexivirus, FDFV2, and observed two VLPs that may associate with FDFV2. Virus purification results showed that spherical particles of 30 nm in diameter coexisted with FDFV2 dsRNA segment (Figure 4-5), while filamentous particles were associated with ORF4-ecoded 18.2 kDa protein (Figures 4-5 and 4-7C). I'm going to perform immunogold TEM with antibodies against FDFV2 ORF4-encoded protein to further confirmation.

In 2021, Sclerotinia sclerotiorum deltaflexivirus 3 (SsDFV3) was reported as a virus with a tripartite genome (Mu *et al.*, 2021). In this study, northern blot results showed that no subgenome was observed (Figure 4-3A and B). But, the 5'RACE (Figure 4-3C) indicated that FDFV2 should have subgenome to express the viral proteins. The organization of subgenomes need to be further confirmed by northern hybridization with riboprobes in the future.

In this study, a new deltaflexivirus, FDFV2, was identified. This is the first report of deltaflexivirus that may associate VLPs, Protein functions and the strategy of gene expression need to be further confirmed.

Table 4-1. List of the amino acid sequences of viral RdRps used in the phylogenetic analysis of *Deltaflexiviridae*.

Virus name	Abbreviation	Accession No.
Alphaflexiviridae		
Allexivirus		
Shallot virus X	ShVX	NP_620648
Botrexvirus		
Botrytis virus X	BotV-X	NP_932306
Lolavirus		
Lolium latent virus	LoLV	YP_001718499
Sclerodarnavirus		
Sclerotinia sclerotiorum debilitation-associated RNA virus	SsDRV	YP_325662
Platypuvirus		
Donkey orchid symptomless virus	DOSV	YP_008828152
Potexvirus		
Potato virus X	PVX	YP_002332929
Betaflexiviridae		
Carlavirus		
Carnation latent virus	CLV	QJX15400
Foveavirus		
Apple stem pitting virus	ASPV	NP_604464
Robigovirus		
Cherry necrotic rusty mottle virus	CNRMV	NP_059937
Capillovirus		
Apple stem grooving virus	ASGV	NP_044335
Chordovirus		
Carrot Ch virus 1	CChV–1	YP_009103999
Citrivirus		
Citrus leaf blotch virus	CLBV	NP_624333
Divavirus		
Diuris virus A	DVA	YP_006905850
Prunevirus		
Apricot vein clearing associated virus	AVCaV	YP_008997790
Ravavirus		
Ribes americanum virus A	RAVA	YP_009553496
Tepovirus		
Potato virus T	PVT	YP_002019748
Trichovirus		
Apple chlorotic leaf spot virus	ACLSV	NP_040551
Vitivirus		
Grapevine virus A	GVA	NP_619662
Wamavirus		
Watermelon virus A	WVA	YP_009357235

Deltaflexiviridae		
Deltaflexivirus		
Fusarium graminearum deltaflexivirus 1	FgDFV1	YP_009268710.1
Sclerotinia sclerotiorum deltaflexivirus 1	SsDFV1	YP_009508363
Soybean leaf-associated mycoflexivirus 1	SlaMFV1	YP_009508374
Sclerotinia sclerotiorum deltaflexivirus 2	SsDFV2	YP_009552771.1
Sclerotinia sclerotiorum deltaflexivirus 3	SsDFV3	UOJ41056.1
Fusarium deltaflexivirus 2	FDFV2	this study
Gammaflexiviridae		
Mycoflexivirus		
Botrytis virus F	BotV-F	NP_068549
Tymoviridae		
Maculavirus		
Grapevine fleck virus	GFkV	NP_542612
Marafivirus		
Maize rayado fino virus	MRFV	NP_115454
Tymovirus		
Turnip yellow mosaic virus	TYMV	NP_663297

Table 4-2. List of the primers used in chapter 4.

Primer name	Sequence	Purpose
FDFV2 detF4978-4997	5'-ACTGTTCTAAGCGCACGGAA	RT-PCR detection
FDFV2 detR5548-5529	5'-TCTTCCGGTTTCACATCGCA	
FDFV2-NB-ORF1-F	5'-GGCTTACCCCTTACTTATTTGTG	DIG DNA probe synthesis
FDFV2-NB-ORF1-R	5'-GGTAATAGTCCACCAAATGACG	
FDFV2-NB-ORF2-F	5'-ATGGCAACTGTCTTCGATAAC	
FDFV2-NB-ORF2-R	5'-CTACAATTTCTTGATGGAGCG	
FDFV2-NB-ORF3-F	5'-ATGTTTAGCCACCCTTTTGTC	
FDFV2-NB-ORF3-R	5'-AGACATGCTGGAAGGAAAACC	
FDFV2-NB-ORF4-F	5'-ATGTCTGACCTGGTTGGTG	
FDFV2-NB-ORF4-R	5'-TCATGATTTCCAAGTCTGCAC	
FDFV2-NB-ORF5-F	5'-ATGAACGCTCTGCCGCCG	
FDFV2-NB-ORF5-R	5'-TCAAACCAATGTGAAATAATAG	

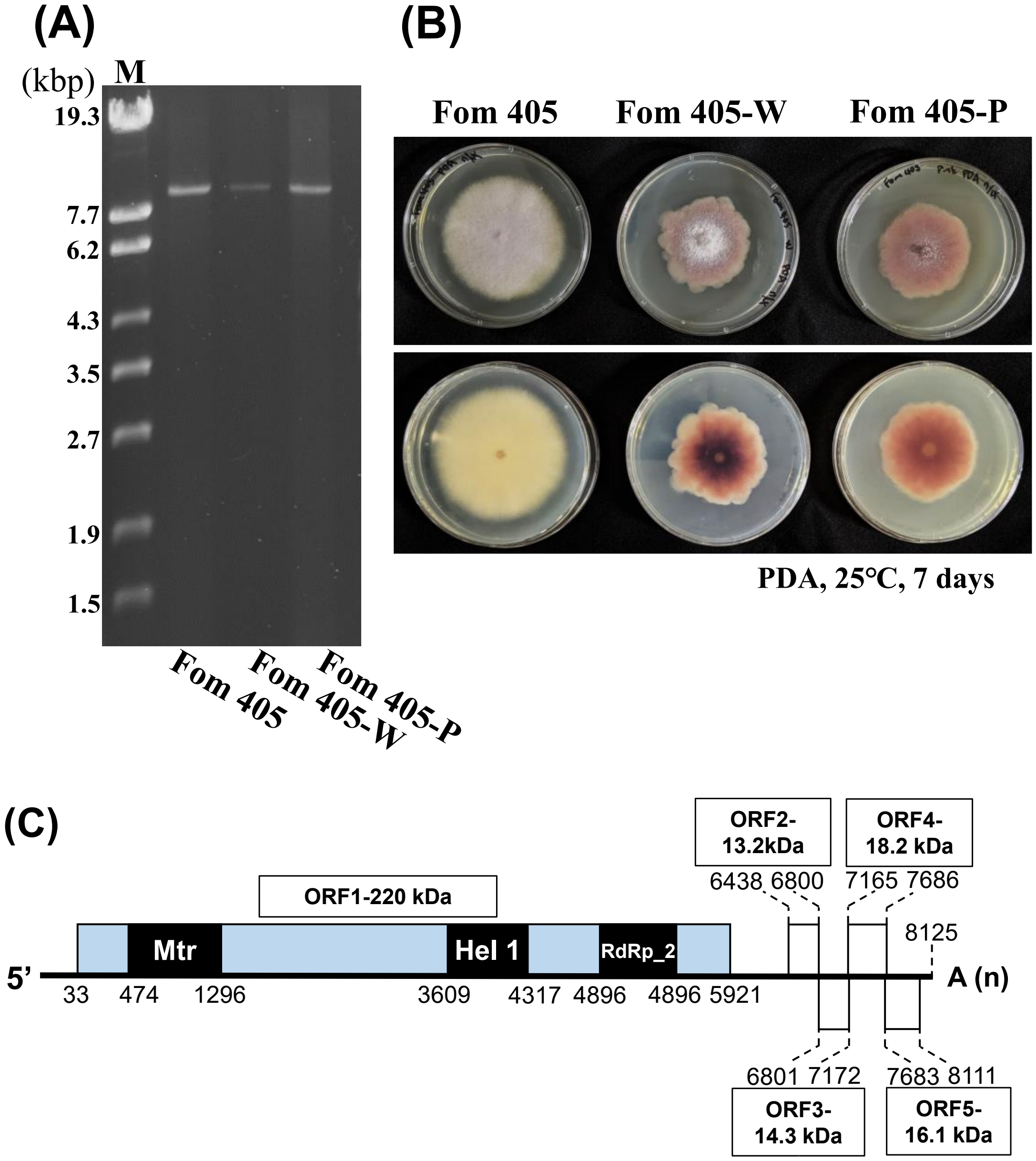


Figure 4-1. FDFV2-infected Fom 405 isolates.

(A) Agarose gel electrophoresis of dsRNA genomes extracted from the mycelia of FDFV2-infected isolates. **(B)** Colony morphorlogies of FDFV2-infected isolates on D-PDA media. **(C)** Schematic diagram of FDFV2 genome.

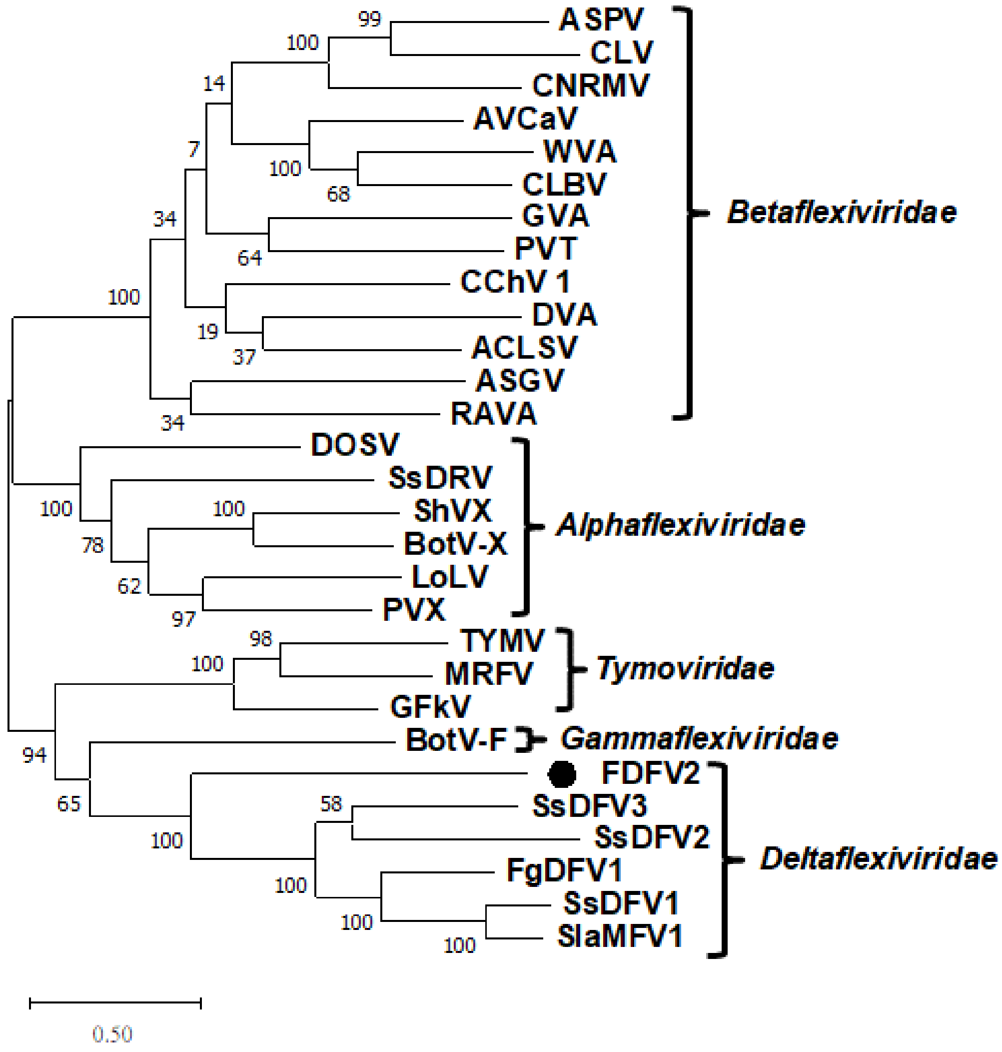
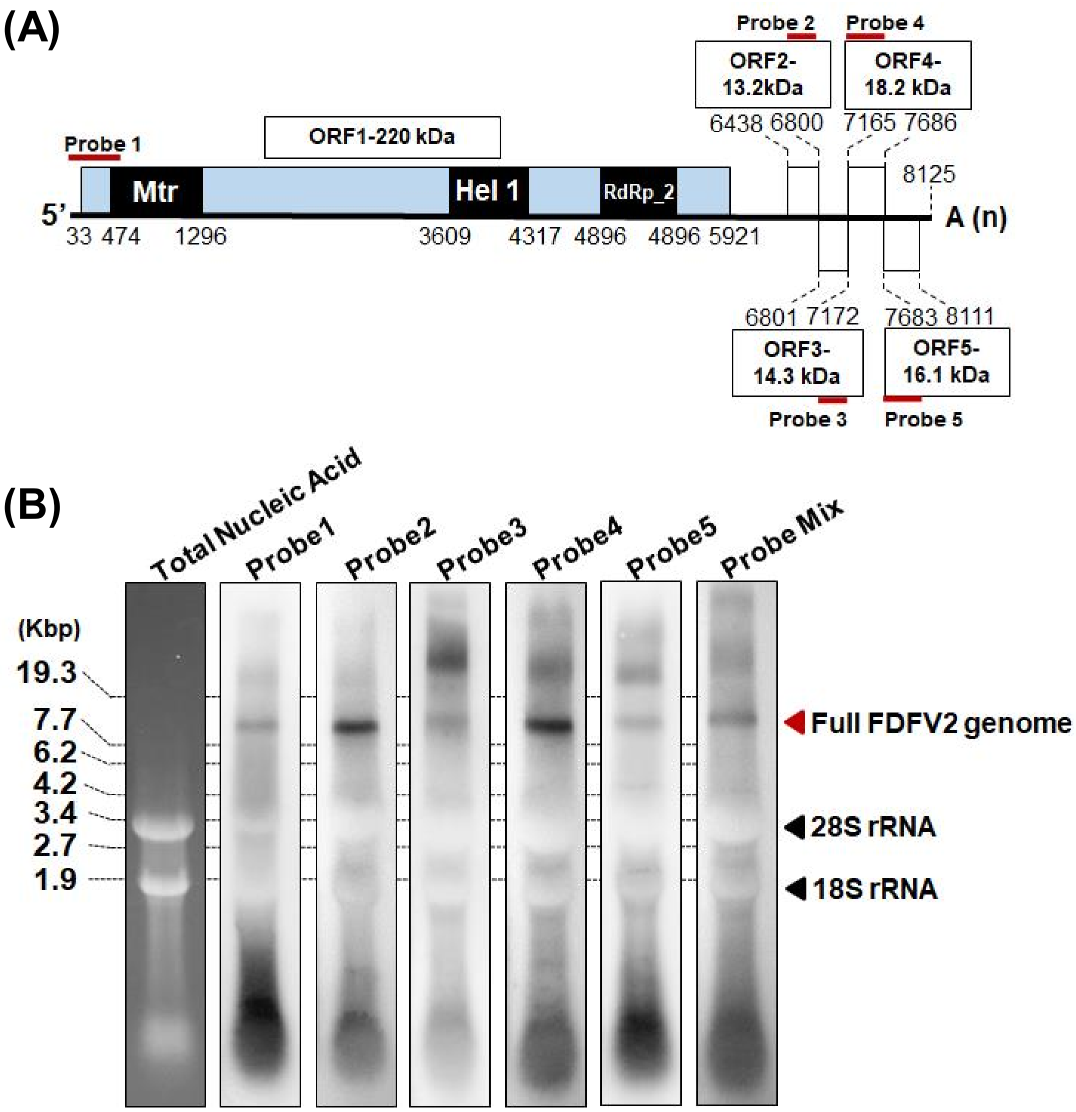


Figure 4-2. Phylogenetic analysis of *Tymovirales.*

Amino acid sequence of RdRps of the members in *Alphaflexiviridae, Betaflexiviridae, Gammaflexiviridae, Deltaflexiviridae,* and *Tymoviridae* were used.


(A)
Probe 2
Probe 4
ORF2-
13.2kDa
ORF4-
18.2 kDa
6438 6800
7165 7686
Probe 1
ORF1-220 kDa
8125
5'
Mtr
Hel 1
RdRp_2
A (n)
33 474
1296
3609
4317 4896
4896 5921
6801 7172
7683 8111
ORF3-
14.3 kDa
ORF5-
16.1 kDa
Probe 3
Probe 5
(B)
Total Nucleic Acid
Probe1
Probe2
Probe3
Probe4
Probe5
Probe Mix
(Kbp)
19.3
7.7
6.2
4.2
3.4
2.7
1.9
Full FDFV2 genome
28S rRNA
18S rRNA

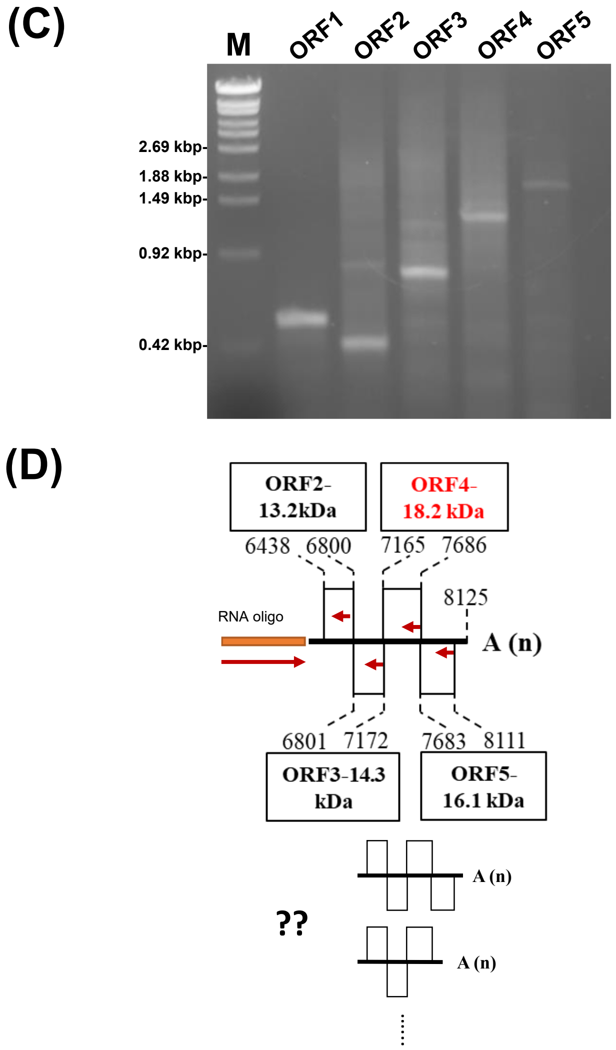


Figure 4-3. Detection of subgenomic RNAs of FDFV2.

(A) Positions of the DIG DNA probe used to detect each FDFV2 ORFs in the northern hybridization experiment. Red lines indicate the position of probes. **(B)** Total RNAs extracted from Fom 405 mycelia were subjected to northern hybridization to detect the subgenomic RNAs of FDFV2 with the DIG DNA probe. **(C)** The result of 5'RACE showed that specific primers detected the predicted bands. ORF1: UPM primer/ FDFV2-NB-ORF1-R; ORF2: UPM primer/ FDFV2-NB-ORF2-R; ORF3: UPM primer/ FDFV2-NB-ORF3-R; ORF4: UPM primer/ FDFV2-NB-ORF4-R; ORF5: UPM primer/ FDFV2-NB-ORF5-R **(D)** Schematics of hypothesized subgenomes.

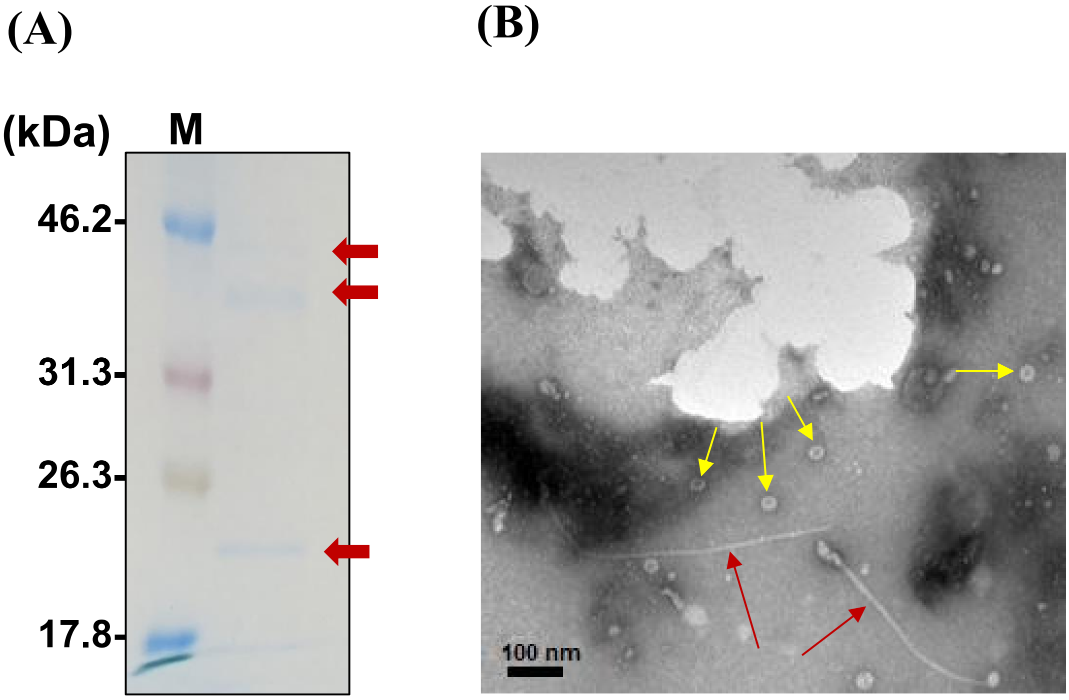


Figure 4-4. Partial purification of FDFV2 virus particles

(A) SDS-PAGE electrophoresis of partial purified FDFV2 virus particles. Red arrows indicate the protein bands. 13% SDS-PAGE, 120V, 150 min. **(B)** Virus-like particlesilces were observed by TEM. Yellow arrows indicate the isometric particles, while red arrows indicate the filamentous particles.

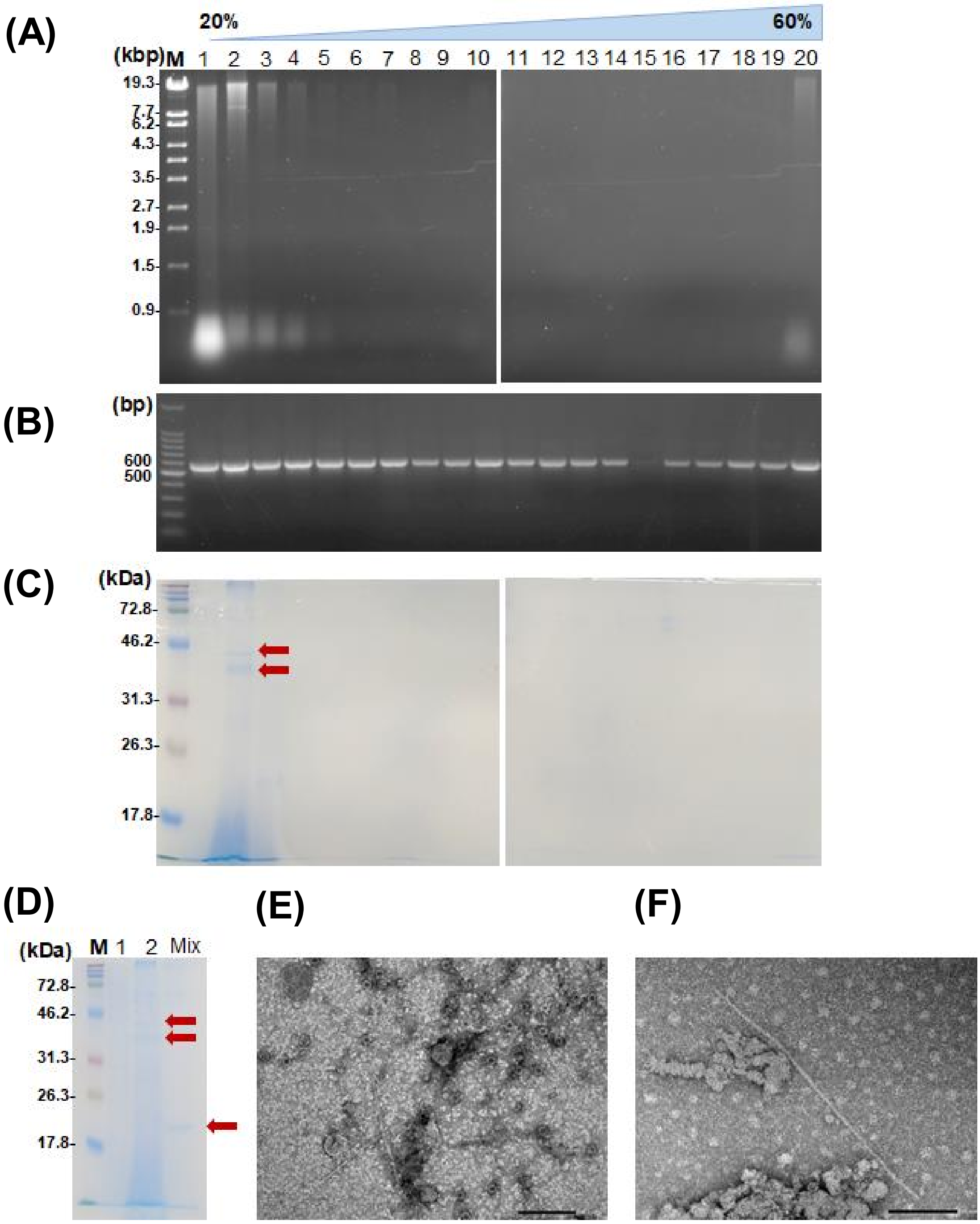

(A)
20%
60%
(kbp) M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
19.3-
7.7-
6.2-
4.3-
3.5-
2.7-
1.9-
1.5-
0.9-
(B)
(bp)
600
500
(C)
(kDa)
72.8-
46.2-
31.3-
26.3-
17.8-
(D)
(kDa) M 1 2 Mix
72.8-
46.2-
31.3-
26.3-
17.8-
(E)
(F)

Figure 4-5. Virus particle prurification by sucros gradient ultracentrifugation. (A) Agarose gel electrophoresis of total nucleic acids of purified fractions. **(B)** RT-PCR using FDFV2 specific primers. (C) SDS-PAGE electrophoresis of purified fractions. **(D)** SDS-PAGE electrophoresis of the concentrated fractions. 1: fraction 1 in **(A)**; 2: fraction 2 in **(A)**; Mix: mixed and concentrated fractions 3-20 in **(A)**. **(E)** Isometric particles (~30 nm in diameter) of fraction 2, and **(F)** filamentous particles (~800 nm in diameter) of mixed fraction were observed byTEM. Scale bar=200 nm.

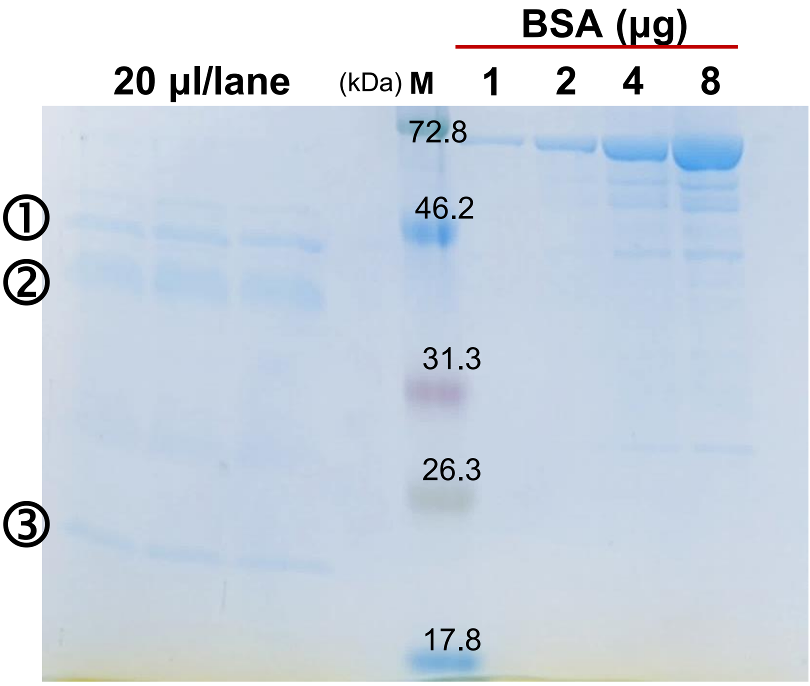
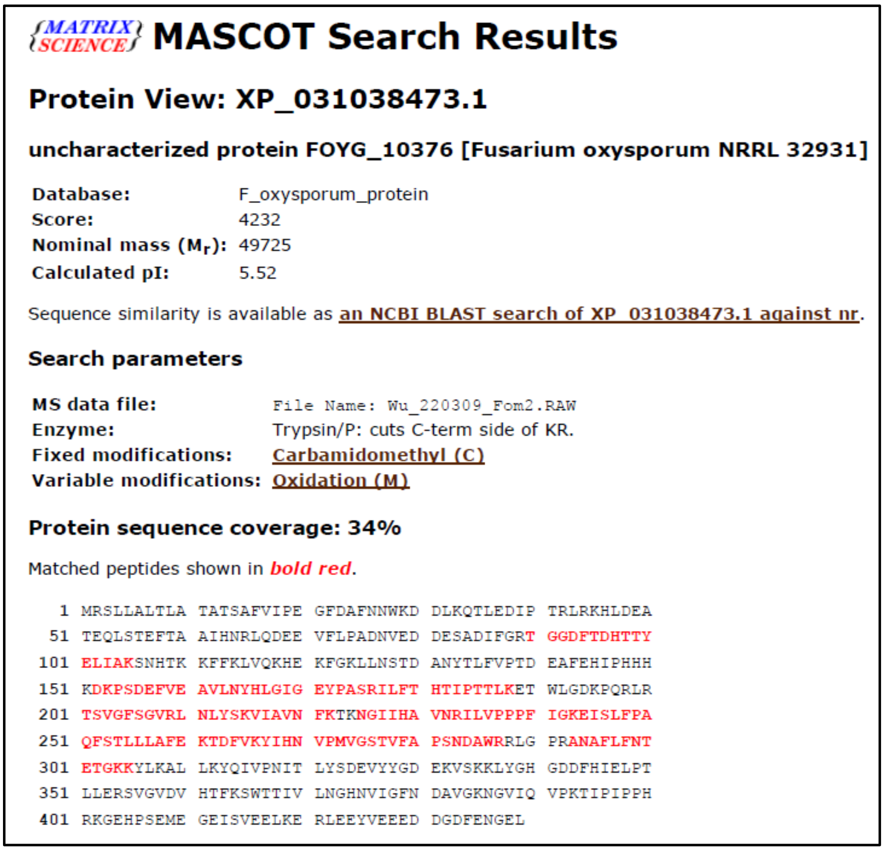


Figure 4-6. Estimation of FDFV2 associated protein for LC-MS/MS.

Protein band①: 1.7 µg in total; Protein band②: 3.3 µg in total; Protein band③: 0.9 µg in total.

(A)



MASCOT Search Results

Protein View: XP_031038473.1

uncharacterized protein FOYG_10376 [Fusarium oxysporum NRRL 32931]

Database:	F_oxysporum_protein
Score:	4232
Nominal mass (M_r):	49725
Calculated pI:	5.52

Sequence similarity is available as an NCBI BLAST search of XP_031038473.1 against nr.

Search parameters

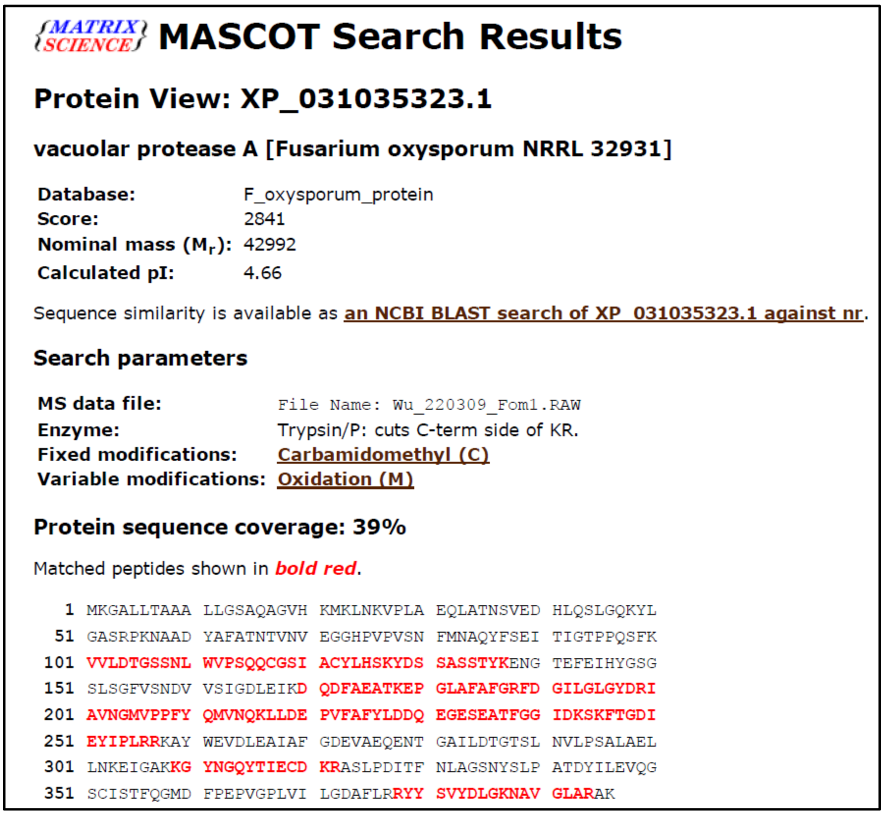
MS data file:	File Name: Wu_220309_Fom2.RAW
Enzyme:	Trypsin/P: cuts C-term side of KR.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Oxidation (M)

Protein sequence coverage: 34%

Matched peptides shown in ***bold red***.

1 MRSLLALTLA TATSAFVIPE GFDAFNNWKD DLKQTLEDIP TRLRKHLDEA
51 TEQLSTEFTA AIHNRLQDEE VFLPADNVED DESADIFGR**T** **GGDFTDHTTY**
101 **ELIAK**SNHTK KFFKLVQKHE KFGKLLNSTD ANYTLFVPTD EAFEHIPHHH
151 K**DKPSDEFVE** **AVLNYHLGIG** **EYPASRILFT** **HTIPTTLK**ET WLGDKPQRLR
201 **TSVGFSGVRL** **NLYSKVIAVN** **FK**TK**NGIIHA** **VNRILVPPPF** **IGKEISLFPA**
251 **QFSTLLLAFE** **KTDFVKYIHN** **VPMVGSTVFA** **PSNDAWR**RLG PR**ANAFLFNT**
301 **ETGKK**YLKAL LKYQIVPNIT LYSDEVYYGD EKVSKKLYGH GDDFHIELPT
351 LLERSVGVDV HTFKSWTTIV LNGHNVIGFN DAVGKNGVIQ VPKTIPIPPH
401 RKGEHPSEME GEISVEELKE RLEEYVEEED DGDFENGEL

(B)



MASCOT Search Results

Protein View: XP_031035323.1

vacuolar protease A [Fusarium oxysporum NRRL 32931]

Database:	F_oxysporum_protein
Score:	2841
Nominal mass (M_r):	42992
Calculated pI:	4.66

Sequence similarity is available as an NCBI BLAST search of XP_031035323.1 against nr.

Search parameters

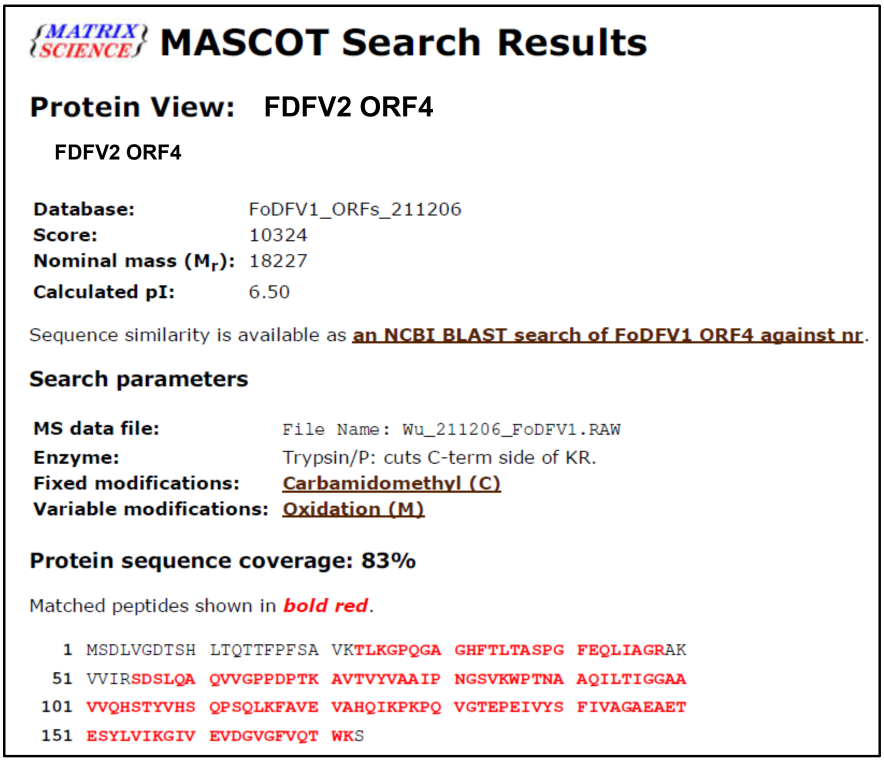
MS data file:	File Name: Wu_220309_Fom1.RAW
Enzyme:	Trypsin/P: cuts C-term side of KR.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Oxidation (M)

Protein sequence coverage: 39%

Matched peptides shown in ***bold red***.

1 MKGALLTAAA LLGSAQAGVH KMKLNKVPLA EQLATNSVED HLQSLGQKYL
51 GASRPKNAAD YAFATNTVNV EGGHPVPVSN FMNAQYFSEI TIGTPPQSFK
101 **VVLDTGSSNL** **WVPSQQCGSI** **ACYLHSKYDS** **SASSTYK**ENG TEFEIHYGSG
151 SLSGFVSNDV VSIGDLEIK**D** **QDFAEATKEP** **GLAFAFGRFD** **GILGLGYDRI**
201 **AVNGMVPPFY** **QMVNQKLLDE** **PVFAFYLDDQ** **EGESEATFGG** **IDKSKFTGDI**
251 **EYIPLRR**KAY WEVDLEAIAF GDEVAEQENT GAILDTGTSL NVLPSALAEL
301 LNKEIGAK**KG** **YNGQYTIECD** **KR**ASLPDITF NLAGSNYSLP ATDYILEVQG
351 SCISTFQGMD FPEPVGPLVI LGDAFLR**RYY** **SVYDLGKNAV** **GLAR**AK

(C)



MASCOT Search Results

Protein View: FDFV2 ORF4

FDFV2 ORF4

Database: FoDFV1_ORFs_211206
Score: 10324
Nominal mass (M_r): 18227
Calculated pI: 6.50

Sequence similarity is available as an NCBI BLAST search of FoDFV1 ORF4 against nr.

Search parameters

MS data file: File Name: Wu_211206_FoDFV1.RAW
Enzyme: Trypsin/P: cuts C-term side of KR.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)

Protein sequence coverage: 83%

Matched peptides shown in ***bold red***.

```
  1 MSDLVGDTSH LTQTTFPFSA VKTLKGPQGA GHFTLTASPG FEQLIAGRAK
 51 VVIRSDSLQA QVVGPPDPTK AVTVYVAAIP NGSVKWPTNA AQILTIGGAA
101 VVQHSTYVHS QPSQLKFAVE VAHQIKPKPQ VGTEPEIVYS FIVAGAEAET
151 ESYLVIKGIV EVDGVGFVQT WKS
```

Figure 4-7. Results of LC-MS/MS of FDFV2 associated proteins analysis.

(A) Protein band①: 49.7 kDa **(B)** Protein band②: 43 kDa **(C)** Protein band③: 18.2 kDa in Fig. 4-5. Protein band① and ② are host proteins. Protein band③ corresponds to the protein from FDFV2 ORF4.

4.5 Appendix

4.5.1 FDFV2 genome sequence

```
5'  GGCUUACCCCUUACUUAUUUGUGGCUGACGGAAUGUCGACCACAGAAACUCGUAAGCCCUCCUGGUGGAG
o                                                                          70
3'  CCGAAUGGGGAAUGAAUAAACACCGACUGCCUUACAGCUGGUGUCUUUGAGCAUUCGGGAGGACCACCUC
1    Gly Leu Pro Leu Thr Tyr Leu Trp Leu Thr Glu Cys Arg Pro Gln Lys Leu Val Ser Pro Pro Gly Gly
2     Ala Tyr Pro Leu Leu Ile Cys Gly . Arg Asn Val Asp His Arg Asn Ser . Ala Leu Leu Val Glu
3   Trp Leu Thr Pro Tyr Leu Phe Val Ala Asp Gly Met Ser Thr Thr Glu Thr Arg Lys Pro Ser Trp Trp Ser
o
5'  UCUGAUUGUAGGCAAGCGCGGUAUGCGUAAGCCCAACAAACAGAUUCGCCUACCUGGACCCAAUUGCUCU
o                                                                          140
3'  AGACUAACAUCCGUUCGCGCCAUACGCAUUCGGGUUGUUUGUCUAAGCGGAUGGACCUGGGUUAACGAGA
1   Val . Leu . Ala Ser Ala Val Cys Val Ser Pro Thr Asn Arg Phe Ala Tyr Leu Asp Pro Ile Ala Leu
2    Ser Asp Cys Arg Gln Ala Arg Tyr Ala . Ala Gln Gln Thr Asp Ser Pro Thr Trp Thr Gln Leu Leu
3     Leu Ile Val Gly Lys Arg Gly Met Arg Lys Pro Asn Lys Gln Ile Arg Leu Pro Gly Pro Asn Cys Ser
o
5'  GAGACCAUCACCGAAUUCAAAGCUGAUUUCCAGCCCAGCCAAGCCCAUCAGUGGGACUGUGACUGUGUGG
o                                                                          210
3'  CUCUGGUAGUGGCUUAAGUUUCGACUAAAGGUCGGGUCGGUUCGGGUAGUCACCCUGACACUGACACACC
1     Arg Pro Ser Pro Asn Ser Lys Leu Ile Ser Ser Pro Ala Lys Pro Ile Ser Gly Thr Val Thr Val Trp
2   . Asp His His Arg Ile Gln Ser . Phe Pro Ala Gln Pro Ser Pro Ser Val Gly Leu . Leu Cys Gly
3    Glu Thr Ile Thr Glu Phe Lys Ala Asp Phe Gln Pro Ser Gln Ala His Gln Trp Asp Cys Asp Cys Val
o
5'  AUCUCAAUUCAUCCUCUUGUGCCAAAGCGCCUGAAGCUGAGGCCAUCCGACUUACAUUAGACCUUGAGCA
o                                                                          280
3'  UAGAGUUAAGUAGGAGAACACGGUUUCGCGGACUUCGACUCCGGUAGGCUGAAUGUAAUCUGGAACUCGU
1    Ile Ser Ile His Pro Leu Val Pro Lys Arg Leu Lys Leu Arg Pro Ser Asp Leu His . Thr Leu Ser
2     Ser Gln Phe Ile Leu Leu Cys Gln Ser Ala . Ser . Gly His Pro Thr Tyr Ile Arg Pro . Ala
3   Asp Leu Asn Ser Ser Ser Cys Ala Lys Ala Pro Glu Ala Glu Ala Ile Arg Leu Thr Leu Asp Leu Glu Gln
o
5'  GGCAAUCUCGCCACCCUGCCAAAAUCCCUAUGUCUGUCUUUGUAAUUCUCCUGGUGGCCGCAAUGCAUGU
o                                                                          350
3'  CCGUUAGAGCGGUGGGACGGUUUUAGGGAUACAGACAGAAACAUUAAGAGGACCACCGGCGUUACGUACA
1   Arg Gln Ser Arg His Pro Ala Lys Ile Pro Met Ser Val Phe Val Ile Leu Leu Val Ala Ala Met His Val
2    Gly Asn Leu Ala Thr Leu Pro Lys Ser Leu Cys Leu Ser Leu . Phe Ser Trp Trp Pro Gln Cys Met
3     Ala Ile Ser Pro Pro Cys Gln Asn Pro Tyr Val Cys Leu Cys Asn Ser Pro Gly Gly Arg Asn Ala Cys
o
5'  CUGCAUGCUUGGGCCCAUUAUCUUGUCCUGAACUCCACCAUAGGUGCAUCAUUUAACCUUGACCACUUUA
o                                                                          420
3'  GACGUACGAACCCGGGUAAUAGAACAGGACUUGAGGUGGUAUCCACGUAGUAAAUUGGAACUGGUGAAAU
1     Cys Met Leu Gly Pro Ile Ile Leu Ser . Thr Pro Pro . Val His His Leu Thr Leu Thr Thr Leu
2   Ser Ala Cys Leu Gly Pro Leu Ser Cys Pro Glu Leu His His Arg Cys Ile Ile . Pro . Pro Leu .
3    Leu His Ala Trp Ala His Tyr Leu Val Leu Asn Ser Thr Ile Gly Ala Ser Phe Asn Leu Asp His Phe
o
5'  AUGGAUCGCCACAACAGCACGCCUUGCUCAAUCGUCAUUUGGUGGACUAUUACCGAGACCGCACCCGGUC
o                                                                          490
3'  UACCUAGCGGUGUUGUCGUGCGGAACGAGUUAGCAGUAAACCACCUGAUAAUGGCUCUGGCGUGGGCCAG
1    Met Asp Arg His Asn Ser Thr Pro Cys Ser Ile Val Ile Trp Trp Thr Ile Thr Glu Thr Ala Pro Gly
2     Trp Ile Ala Thr Thr Ala Arg Leu Ala Gln Ser Ser Phe Gly Gly Leu Leu Pro Arg Pro His Pro Val
3   Asn Gly Ser Pro Gln Gln His Ala Leu Leu Asn Arg His Leu Val Asp Tyr Tyr Arg Asp Arg Thr Arg Ser
o
5'  UCAGGAGACCACCCCUUACAACAUUCCAAAUUCACAACGCCAUCUCCUUCAACCACUUGGAAUGGAAUUA
o                                                                          560
3'  AGUCCUCUGGUGGGGAAUGUUGUAAGGUUUAAGUGUUGCGGUAGAGGAAGUUGGUGAACCUUACCUUAAU
1   Leu Arg Arg Pro Pro Leu Thr Thr Phe Gln Ile His Asn Ala Ile Ser Phe Asn His Leu Glu Trp Asn Tyr
2    Ser Gly Asp His Pro Leu Gln His Ser Lys Phe Thr Thr Pro Ser Pro Ser Thr Thr Trp Asn Gly Ile
3     Gln Glu Thr Thr Pro Tyr Asn Ile Pro Asn Ser Gln Arg His Leu Leu Gln Pro Leu Gly Met Glu Leu
o
```

```
5'   CCCCAACCCGACGCCCCUGAAACUCCACAUGCUCUUCAUAAGUGCAUUGAGGAAUUUCAGUUGCGUAGAA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   630
3'   GGGGUUGGGCUGCGGGGACUUUGAGGUGUACGAGAAGUAUUCACGUAACUCCUUAAAGUCAACGCAUCUU
1     Pro Asn Pro Thr Pro Leu Lys Leu His Met Leu Phe Ile Ser Ala Leu Arg Asn Phe Ser Cys Val Glu
2    Thr Pro Thr Arg Arg Pro . Asn Ser Thr Cys Ser Ser . Val His . Gly Ile Ser Val Ala . Asn
3     Pro Gln Pro Asp Ala Pro Glu Thr Pro His Ala Leu His Lys Cys Ile Glu Glu Phe Gln Leu Arg Arg
o
5'   CUGCUCGACUGCUGCUACAACCAAAUAAUUUUGGCGUCAUCUCGGUAAAGAAUUCCAAAUUGACUCUACU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   700
3'   GACGAGCUGACGACGAUGUUGGUUUAUUAAAACCGCAGUAGAGCCAUUUCUUAAGGUUUAACUGAGAUGA
1     Leu Leu Asp Cys Cys Tyr Asn Gln Ile Ile Leu Ala Ser Ser Arg . Arg Ile Pro Asn . Leu Tyr
2     Cys Ser Thr Ala Ala Thr Thr Lys . Phe Trp Arg His Leu Gly Lys Glu Phe Gln Ile Asp Ser Thr
3    Thr Ala Arg Leu Leu Leu Gln Pro Asn Asn Phe Gly Val Ile Ser Val Lys Asn Ser Lys Leu Thr Leu Leu
o
5'   UCCCACUCCCGGUUCCGUACAGAAUCCUGUGUUCGAGGCCAAGGACGUCACCCGCUAUCCUGGGUCCGCG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   770
3'   AGGGUGAGGGCCAAGGCAUGUCUUAGGACACAAGCUCCGGUUCCUGCAGUGGGCGAUAGGACCCAGGCGC
1    Phe Pro Leu Pro Val Pro Tyr Arg Ile Leu Cys Ser Arg Pro Arg Thr Ser Pro Ala Ile Leu Gly Pro Arg
2     Ser His Ser Arg Phe Arg Thr Glu Ser Cys Val Arg Gly Gln Gly Arg His Pro Leu Ser Trp Val Arg
3     Pro Thr Pro Gly Ser Val Gln Asn Pro Val Phe Glu Ala Lys Asp Val Thr Arg Tyr Pro Gly Ser Ala
o
5'   AUCAAGCAUGCGCAUUUCAAAGAUUUUCCGGUGCAUCUCCUCCAUGACGUCACUUCUGAGGUCAGCCCAC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   840
3'   UAGUUCGUACGCGUAAAGUUUCUAAAAGGCCACGUAGAGGAGGUACUGCAGUGAAGACUCCAGUCGGGUG
1     Ser Ser Met Arg Ile Ser Lys Ile Phe Arg Cys Ile Ser Ser Met Thr Ser Leu Leu Arg Ser Ala His
2    Asp Gln Ala Cys Ala Phe Gln Arg Phe Ser Gly Ala Ser Pro Pro . Arg His Phe . Gly Gln Pro Thr
3     Ile Lys His Ala His Phe Lys Asp Phe Pro Val His Leu Leu His Asp Val Thr Ser Glu Val Ser Pro
o
5'   AUGAAUUGGUUGAAAAGCUGGUCGCCGAAAAUCCCGAAGGACACCUAAUAGCAACUGGAAUGAAUCCGAU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   910
3'   UACUUAACCAACUUUUCGACCAGCGGCUUUUAGGGCUUCCUGUGGAUUAUCGUUGACCUUACUUAGGCUA
1     Met Asn Trp Leu Lys Ser Trp Ser Pro Lys Ile Pro Lys Asp Thr . . Gln Leu Glu . Ile Arg
2     . Ile Gly . Lys Ala Gly Arg Arg Lys Ser Arg Arg Thr Pro Asn Ser Asn Trp Asn Glu Ser Asp
3    His Glu Leu Val Glu Lys Leu Val Ala Glu Asn Pro Glu Gly His Leu Ile Ala Thr Gly Met Asn Pro Ile
o
5'   AGAGGUUCUCGAUCGUGCCGCAUCUUUUGAACCAGCUAGUCACACAAUUGAGUAUGACUUGGAUUCUUUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   980
3'   UCUCCAAGAGCUAGCACGGCGUAGAAAACUUGGUCGAUCAGUGUGUUAACUCAUACUGAACCUAAGAAAG
1     . Arg Phe Ser Ile Val Pro His Leu Leu Asn Gln Leu Val Thr Gln Leu Ser Met Thr Trp Ile Leu Ser
2     Arg Gly Ser Arg Ser Cys Arg Ile Phe . Thr Ser . Ser His Asn . Val . Leu Gly Phe Phe
3     Glu Val Leu Asp Arg Ala Ala Ser Phe Glu Pro Ala Ser His Thr Ile Glu Tyr Asp Leu Asp Ser Phe
o
5'   AAUUUUGUCUUCACCGGUUCUGAGUCCGAAUCUUACUUCACCCCUGCUUGGGUCACCAACACGUGGUUGC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1050
3'   UUAAAACAGAAGUGGCCAAGACUCAGGCUUAGAAUGAAGUGGGGACGAACCCAGUGGUUGUGCACCAACG
1     Ile Leu Ser Ser Pro Val Leu Ser Pro Asn Leu Thr Ser Pro Leu Leu Gly Ser Pro Thr Arg Gly Cys
2    Gln Phe Cys Leu His Arg Phe . Val Arg Ile Leu Leu His Pro Cys Leu Gly His Gln His Val Val Ala
3     Asn Phe Val Phe Thr Gly Ser Glu Ser Glu Ser Tyr Phe Thr Pro Ala Trp Val Thr Asn Thr Trp Leu
o
5'   GCACCUCUAGUGUUUGUGCGUCUAAUGGUCGCGUUUACCAUGUUACAUUGUUGGACUACAAGUUGGGACA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1120
3'   CGUGGAGAUCACAAACACGCAGAUUACCAGCGCAAAUGGUACAAUGUAACAACCUGAUGUUCAACCCUGU
1     Ala Pro Leu Val Phe Val Arg Leu Met Val Ala Phe Thr Met Leu His Cys Trp Thr Thr Ser Trp Asp
2     His Leu . Cys Leu Cys Val . Trp Ser Arg Leu Pro Cys Tyr Ile Val Gly Leu Gln Val Gly Thr
3    Arg Thr Ser Ser Val Cys Ala Ser Asn Gly Arg Val Tyr His Val Thr Leu Leu Asp Tyr Lys Leu Gly His
o
```

```
5' CUGUGUUUGGCACAUUUACUGUGGCAACGCCGCUGAUCAACACACUCGAACAUUUUCAACUGGCUCUUAC
o                                                                          1190
3' GACACAAACCGUGUAAAUGACACCGUUGCGGCGACUAGUUGUGUGAGCUUGUAAAAGUUGACCGAGAAUG
1  Thr Val Phe Gly Thr Phe Thr Val Ala Thr Pro Leu Ile Asn Thr Leu Glu His Phe Gln Leu Ala Leu Thr
2   Leu Cys Leu Ala His Leu Leu Trp Gln Arg Arg . Ser Thr His Ser Asn Ile Phe Asn Trp Leu Leu
3    Cys Val Trp His Ile Tyr Cys Gly Asn Ala Ala Asp Gln His Thr Arg Thr Phe Ser Thr Gly Ser Tyr
o
5' GUGAGGAUCCCCGCUGUCGUUUCGGGCACACUCACCGACGAAUACCUCCCAGUGAAUUUACUUUCAGGCA
o                                                                          1260
3' CACUCCUAGGGGCGACAGCAAAGCCCGUGUGAGUGGCUGCUUAUGGAGGGUCACUUAAAUGAAAGUCCGU
1   . Gly Ser Pro Leu Ser Phe Arg Ala His Ser Pro Thr Asn Thr Ser Gln . Ile Tyr Phe Gln Ala
2  Arg Glu Asp Pro Arg Cys Arg Phe Gly His Thr His Arg Arg Ile Pro Pro Ser Glu Phe Thr Phe Arg His
3   Val Arg Ile Pro Ala Val Val Ser Gly Thr Leu Thr Asp Glu Tyr Leu Pro Val Asn Leu Leu Ser Gly
o
5' UCCUUGACUUCGUUGGUCGAACACCGGAUUUAAGCACCAGAAACCUCUCUGCAAAAGUUUCACAGCUGGC
o                                                                          1330
3' AGGAACUGAAGCAACCAGCUUGUGGCCUAAAUUCGUGGUCUUUGGAGAGACGUUUUCAAAGUGUCGACCG
1   Ser Leu Thr Ser Leu Val Glu His Arg Ile . Ala Pro Glu Thr Ser Leu Gln Lys Phe His Ser Trp
2    Pro . Leu Arg Trp Ser Asn Thr Gly Phe Lys His Gln Lys Pro Leu Cys Lys Ser Phe Thr Ala Gly
3  Ile Leu Asp Phe Val Gly Arg Thr Pro Asp Leu Ser Thr Arg Asn Leu Ser Ala Lys Val Ser Gln Leu Ala
o
5' CACUGGCAUCAACCCCCGGGCCACCGCUCGUGACCGUUGGGUUGCGACUUAUGUGGCCGGGCUUCUUGCC
o                                                                          1400
3' GUGACCGUAGUUGGGGGCCCGGUGGCGAGCACUGGCAACCCAACGCUGAAUACACCGGCCCGAAGAACGG
1  Pro Leu Ala Ser Thr Pro Gly Pro Pro Leu Val Thr Val Gly Leu Arg Leu Met Trp Pro Gly Phe Leu Pro
2   His Trp His Gln Pro Pro Gly His Arg Ser . Pro Leu Gly Cys Asp Leu Cys Gly Arg Ala Ser Cys
3    Thr Gly Ile Asn Pro Arg Ala Thr Ala Arg Asp Arg Trp Val Ala Thr Tyr Val Ala Gly Leu Leu Ala
o
5' CCAAACAAAGACUUUUGGUGGUACUUCCGUAAGACAUUUUGGAAUUUUAUGUACGCGGUCACUUUUCAAU
o                                                                          1470
3' GGUUUGUUUCUGAAAACCACCAUGAAGGCAUUCUGUAAAACCUUAAAAUACAUGCGCCAGUGAAAAGUUA
1    Gln Thr Lys Thr Phe Gly Gly Thr Ser Val Arg His Phe Gly Ile Leu Cys Thr Arg Ser Leu Phe Asn
2  Pro Lys Gln Arg Leu Leu Val Val Leu Pro . Asp Ile Leu Glu Phe Tyr Val Arg Gly His Phe Ser Met
3   Pro Asn Lys Asp Phe Trp Trp Tyr Phe Arg Lys Thr Phe Trp Asn Phe Met Tyr Ala Val Thr Phe Gln
o
5' GGCACAUGCUCAAACCAUUGCCUGACAUCUACUCCUAUGUAGAUGAGCGCAAACGAACCCGUGUCAUACA
o                                                                          1540
3' CCGUGUACGAGUUUGGUAACGGACUGUAGAUGAGGAUACAUCUACUCGCGUUUGCUUGGGCACAGUAUGU
1   Gly Thr Cys Ser Asn His Cys Leu Thr Ser Thr Pro Met . Met Ser Ala Asn Glu Pro Val Ser Tyr
2    Ala His Ala Gln Thr Ile Ala . His Leu Leu Leu Cys Arg . Ala Gln Thr Asn Pro Cys His Thr
3  Trp His Met Leu Lys Pro Leu Pro Asp Ile Tyr Ser Tyr Val Asp Glu Arg Lys Arg Thr Arg Val Ile His
o
5' UCCCACACCAGGUGGUGGUUGGAGUCCUCGAACUUCCUUCAAAGUUCGUAAAGCAUCAAUUCCUAACAUG
o                                                                          1610
3' AGGGUGUGGUCCACCACCAACCUCAGGAGCUUGAAGGAAGUUUCAAGCAUUUCGUAGUUAAGGAUUGUAC
1  Ile Pro His Gln Val Val Val Gly Val Leu Glu Leu Pro Ser Lys Phe Val Lys His Gln Phe Leu Thr Cys
2   Ser His Thr Arg Trp Trp Leu Glu Ser Ser Asn Phe Leu Gln Ser Ser . Ser Ile Asn Ser . His
3    Pro Thr Pro Gly Gly Gly Trp Ser Pro Arg Thr Ser Phe Lys Val Arg Lys Ala Ser Ile Pro Asn Met
o
5' CCUUCGGCACUUCAACGACUGUCCGCCUUCACAGCUAGCGUUUUCGUGUUCAUGAUUCCGAAGAUUCUUA
o                                                                          1680
3' GGAAGCCGUGAAGUUGCUGACAGGCGGAAGUGUCGAUCGCAAAAGCACAAGUACUAAGGCUUCUAAGAAU
1    Leu Arg His Phe Asn Asp Cys Pro Pro Ser Gln Leu Ala Phe Ser Cys Ser . Phe Arg Arg Phe Leu
2  Ala Phe Gly Thr Ser Thr Thr Val Arg Leu His Ser . Arg Phe Arg Val His Asp Ser Glu Asp Ser Tyr
3   Pro Ser Ala Leu Gln Arg Leu Ser Ala Phe Thr Ala Ser Val Phe Val Phe Met Ile Pro Lys Ile Leu
o
```

```
5'   UAGGUGAGGUGAUCACUAACGUUAUACUUAAACUGGACUACUUCGGGUUUUUCAAAACACUUUGGCGCCU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1750
3'   AUCCACUCCACUAGUGAUUGCAAUAUGAAUUUGACCUGAUGAAGCCCAAAAAGUUUUGUGAAACCGCGGA
1     .  Val  Arg  .  Ser  Leu  Thr  Leu  Tyr  Leu  Asn  Trp  Thr  Thr  Ser  Gly  Phe  Ser  Lys  His  Phe  Gly  Ala
2      Arg  .  Gly  Asp  His  .  Arg  Tyr  Thr  .  Thr  Gly  Leu  Leu  Arg  Val  Phe  Gln  Asn  Thr  Leu  Ala  Pro
3    Ile  Gly  Glu  Val  Ile  Thr  Asn  Val  Ile  Leu  Lys  Leu  Asp  Tyr  Phe  Gly  Phe  Phe  Lys  Thr  Leu  Trp  Arg  Leu
o
5'   CGCUGAAAUUGACAUUGGUCGCUUCUUAUUAACACUUGCCGUGAUUGUCGUCACCGCAAUCAUCCCCGGG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1820
3'   GCGACUUUAACUGUAACCAGCGAAGAAUAAUUGUGAACGGCACUAACAGCAGUGGCGUUAGUAGGGGCCC
1    Ser  Leu  Lys  Leu  Thr  Leu  Val  Ala  Ser  Tyr  .  His  Leu  Pro  .  Leu  Ser  Ser  Pro  Gln  Ser  Ser  Pro  Gly
2     Arg  .  Asn  .  His  Trp  Ser  Leu  Leu  Ile  Asn  Thr  Cys  Arg  Asp  Cys  Arg  His  Arg  Asn  His  Pro  Arg
3      Ala  Glu  Ile  Asp  Ile  Gly  Arg  Phe  Leu  Leu  Thr  Leu  Ala  Val  Ile  Val  Val  Thr  Ala  Ile  Ile  Pro  Gly
o
5'   UCUGUAGUUAAGGUUUUUUCCCGAUUGGCCGGUCACUUUUGGCGCCAGCUCUGGAUGCCUGGAUGGAUCA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1890
3'   AGACAUCAAUUCCAAAAAAGGGCUAACCGGCCAGUGAAAACCGCGGUCGAGACCUACGGACCUACCUAGU
1     Leu  .  Leu  Arg  Phe  Phe  Pro  Asp  Trp  Pro  Val  Thr  Phe  Gly  Ala  Ser  Ser  Gly  Cys  Leu  Asp  Gly  Ser
2    Val  Cys  Ser  .  Gly  Phe  Phe  Pro  Ile  Gly  Arg  Ser  Leu  Leu  Ala  Pro  Ala  Leu  Asp  Ala  Trp  Met  Asp  Gln
3     Ser  Val  Val  Lys  Val  Phe  Ser  Arg  Leu  Ala  Gly  His  Phe  Trp  Arg  Gln  Leu  Trp  Met  Pro  Gly  Trp  Ile
o
5'   AAUGCGGUUUUGAGCUUCUGGUGCACGAAGUUGUCGGCGCCCCAGGAUACAAAUUUCUAUUUACACUGCC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   1960
3'   UUACGCCAAAACUCGAAGACCACGUGCUUCAACAGCCGCGGGGUCCUAUGUUUAAAGAUAAAUGUGACGG
1     Asn  Ala  Val  Leu  Ser  Phe  Trp  Cys  Thr  Lys  Leu  Ser  Ala  Pro  Gln  Asp  Thr  Asn  Phe  Tyr  Leu  His  Cys
2      Met  Arg  Phe  .  Ala  Ser  Gly  Ala  Arg  Ser  Cys  Arg  Arg  Pro  Arg  Ile  Gln  Ile  Ser  Ile  Tyr  Thr  Ala
3    Lys  Cys  Gly  Phe  Glu  Leu  Leu  Val  His  Glu  Val  Val  Gly  Ala  Pro  Gly  Tyr  Lys  Phe  Leu  Phe  Thr  Leu  Pro
o
5'   AGGCAGAGGGUGGUUCUACCAAGCCUACCUGUGGAUAGUUGCCUCAUACUCAGUGCUUCCAGGACUACCG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2030
3'   UCCGUCUCCCACCAAGAUGGUUCGGAUGGACACCUAUCAACGGAGUAUGAGUCACGAAGGUCCUGAUGGC
1    Gln  Ala  Glu  Gly  Gly  Ser  Thr  Lys  Pro  Thr  Cys  Gly  .  Leu  Pro  His  Thr  Gln  Cys  Phe  Gln  Asp  Tyr  Arg
2     Arg  Gln  Arg  Val  Val  Leu  Pro  Ser  Leu  Pro  Val  Asp  Ser  Cys  Leu  Ile  Leu  Ser  Ala  Ser  Arg  Thr  Thr
3      Gly  Arg  Gly  Trp  Phe  Tyr  Gln  Ala  Tyr  Leu  Trp  Ile  Val  Ala  Ser  Tyr  Ser  Val  Leu  Pro  Gly  Leu  Pro
o
5'   UUCAUGUGGGCUUUUCCUUGGUACUGUCUCACUGUCCCACAUGGUUGGCUGUUCUUUGCAAUUUGGGCAG
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2100
3'   AAGUACACCCGAAAAGGAACCAUGACAGAGUGACAGGGUGUACCAACCGACAAGAAACGUUAAACCCGUC
1      Ser  Cys  Gly  Leu  Phe  Leu  Gly  Thr  Val  Ser  Leu  Ser  His  Met  Val  Gly  Cys  Ser  Leu  Gln  Phe  Gly  Gln
2    Val  His  Val  Gly  Phe  Ser  Leu  Val  Leu  Ser  His  Cys  Pro  Thr  Trp  Leu  Ala  Val  Leu  Cys  Asn  Leu  Gly  Ser
3     Phe  Met  Trp  Ala  Phe  Pro  Trp  Tyr  Cys  Leu  Thr  Val  Pro  His  Gly  Trp  Leu  Phe  Phe  Ala  Ile  Trp  Ala
o
5'   CUAUCAUCCUUGAGAAUUUGAUCGCCUGCGCGCCAAAAUGUCCCGUCCCGAGCUCGUUCCCCAAGUACCA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2170
3'   GAUAGUAGGAACUCUUAAACUAGCGGACGCGCGGUUUUACAGGGCAGGGCUCGAGCAAGGGGUUCAUGGU
1     Leu  Ser  Ser  Leu  Arg  Ile  .  Ser  Pro  Ala  Arg  Gln  Asn  Val  Pro  Ser  Arg  Ala  Arg  Ser  Pro  Ser  Thr
2      Tyr  His  Pro  .  Glu  Phe  Asp  Arg  Leu  Arg  Ala  Lys  Met  Ser  Arg  Pro  Glu  Leu  Val  Pro  Gln  Val  Pro
3    Ala  Ile  Ile  Leu  Glu  Asn  Leu  Ile  Ala  Cys  Ala  Pro  Lys  Cys  Pro  Val  Pro  Ser  Ser  Phe  Pro  Lys  Tyr  His
o
5'   UUAUAAUCUGCUCAACACUGUGAAACAUGUGUGGAAUAAGCCUCGGGAAUCCUACGUCUGGUCUUUCUUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   2240
3'   AAUAUUAGACGAGUUGUGACACUUUGUACACACCUUAUUCGGAGCCCUUAGGAUGCAGACCAGAAAGAAG
1    Ile  Ile  Ile  Cys  Ser  Thr  Leu  .  Asn  Met  Cys  Gly  Ile  Ser  Leu  Gly  Asn  Pro  Thr  Ser  Gly  Leu  Ser  Ser
2     Leu  .  Ser  Ala  Gln  His  Cys  Glu  Thr  Cys  Val  Glu  .  Ala  Ser  Gly  Ile  Leu  Arg  Leu  Val  Phe  Leu
3      Tyr  Asn  Leu  Leu  Asn  Thr  Val  Lys  His  Val  Trp  Asn  Lys  Pro  Arg  Glu  Ser  Tyr  Val  Trp  Ser  Phe  Phe
o
```

```
5'  CUCACUGCUACUGCUUGCUGGACGGCGUGGACUUUUGCCAUUCUGAUGUGCCGCACCUACAACGCCAUCC
o                                                                             2310
3'  GAGUGACGAUGACGAACGACCUGCCGCACCUGAAAACGGUAAGACUACACGGCGUGGAUGUUGCGGUAGG
1    Ser Leu Leu Leu Leu Ala Gly Arg Arg Gly Leu Leu Pro Phe  .  Cys Ala Ala Pro Thr Thr Pro Ser
2   Pro His Cys Tyr Cys Leu Leu Asp Gly Val Asp Phe Cys His Ser Asp Val Pro His Leu Gln Arg His Pro
3    Leu Thr Ala Thr Ala Cys Trp Thr Ala Trp Thr Phe Ala Ile Leu Met Cys Arg Thr Tyr Asn Ala Ile
o
5'  GAUCGAUUUACAUUCACCCAUCCAUGCGCGGUAACUGUGACCGGGCUGGCUGCUUGCCGGCAGUUAACGU
o                                                                             2380
3'  CUAGCUAAAUGUAAGUGGGUAGGUACGCGCCAUUGACACUGGCCCGACCGACGAACGGCCGUCAAUUGCA
1    Asp Arg Phe Thr Phe Thr His Pro Cys Ala Val Thr Val Thr Gly Leu Ala Ala Cys Arg Gln Leu Thr
2     Ile Asp Leu His Ser Pro Ile His Ala Arg  .  Leu  .  Pro Gly Trp Leu Leu Ala Gly Ser  .  Arg
3   Arg Ser Ile Tyr Ile His Pro Ser Met Arg Gly Asn Cys Asp Arg Ala Gly Cys Leu Pro Ala Val Asn Val
o
5'  CGCUCAGCCCCAGACAGUUCGCAAGCGGUAUGUUGACCUACCGCUACCGUCAGUCACUGUCAUACCUGAC
o                                                                             2450
3'  GCGAGUCGGGGUCUGUCAAGCGUUCGCCAUACAACUGGAUGGCGAUGGCAGUCAGUGACAGUAUGGACUG
1   Ser Leu Ser Pro Arg Gln Phe Ala Ser Gly Met Leu Thr Tyr Arg Tyr Arg Gln Ser Leu Ser Tyr Leu Thr
2    Arg Ser Ala Pro Asp Ser Ser Gln Ala Val Cys  .  Pro Thr Ala Thr Val Ser His Cys His Thr  .
3     Ala Gln Pro Gln Thr Val Arg Lys Arg Tyr Val Asp Leu Pro Leu Pro Ser Val Thr Val Ile Pro Asp
o
5'  CGCAUCCCAGUACAACAUGCCGUUGGUGGUUUGGCUGUUGAUCCAAUAGGCCUCAAUCUACCAGACUGGC
o                                                                             2520
3'  GCGUAGGGUCAUGUUGUACGGCAACCACCAAACCGACAACUAGGUUAUCCGGAGUUAGAUGGUCUGACCG
1     Ala Ser Gln Tyr Asn Met Pro Leu Val Val Trp Leu Leu Ile Gln  .  Ala Ser Ile Tyr Gln Thr Gly
2   Pro His Pro Ser Thr Thr Cys Arg Trp Trp Phe Gly Cys  .  Ser Asn Arg Pro Gln Ser Thr Arg Leu Ala
3    Arg Ile Pro Val Gln His Ala Val Gly Gly Leu Ala Val Asp Pro Ile Gly Leu Asn Leu Pro Asp Trp
o
5'  GGGACCAGGUUGAGGCAGCUUAUCGAAACCAACCCAACGCGUACCCACAAUUAACACCAGGACAGUCGUG
o                                                                             2590
3'  CCCUGGUCCAACUCCGUCGAAUAGCUUUGGUUGGGUUGCGCAUGGGUGUUAAUUGUGGUCCUGUCAGCAC
1    Gly Thr Arg Leu Arg Gln Leu Ile Glu Thr Asn Pro Thr Arg Thr His Asn  .  His Gln Asp Ser Arg
2     Gly Pro Gly  .  Gly Ser Leu Ser Lys Pro Thr Gln Arg Val Pro Thr Ile Asn Thr Arg Thr Val Val
3   Arg Asp Gln Val Glu Ala Ala Tyr Arg Asn Gln Pro Asn Ala Tyr Pro Gln Leu Thr Pro Gly Gln Ser Cys
o
5'  UUUCUGGGAUUGUCUCGCCAAGUUUGGUGGCACGCCCCACAUGUGGUACAGUUGGUUUAUGGCAUUCACU
o                                                                             2660
3'  AAAGACCCUAACAGAGCGGUUCAAACCACCGUGCGGGGUGUACACCAUGUCAACCAAAUACCGUAAGUGA
1   Val Ser Gly Ile Val Ser Pro Ser Leu Val Ala Arg Pro Thr Cys Gly Thr Val Gly Leu Trp His Ser Leu
2    Phe Leu Gly Leu Ser Arg Gln Val Trp Trp His Ala Pro His Val Val Gln Leu Val Tyr Gly Ile His
3     Phe Trp Asp Cys Leu Ala Lys Phe Gly Gly Thr Pro His Met Trp Tyr Ser Trp Phe Met Ala Phe Thr
o
5'  CGACGCACCCCAAAUCCCAAUGACCCGGUCGUGGGCAAUGUCACCCUUCCUGAAAUUCAGGAAUUUGCUG
o                                                                             2730
3'  GCUGCGUGGGGUUUAGGGUUACUGGGCCAGCACCCGUUACAGUGGGAAGGACUUUAAGUCCUUAAACGAC
1     Asp Ala Pro Gln Ile Pro Met Thr Arg Ser Trp Ala Met Ser Pro Phe Leu Lys Phe Arg Asn Leu Leu
2   Ser Thr His Pro Lys Ser Gln  .  Pro Gly Arg Gly Gln Cys His Pro Ser  .  Asn Ser Gly Ile Cys Cys
3    Arg Arg Thr Pro Asn Pro Asn Asp Pro Val Val Gly Asn Val Thr Leu Pro Glu Ile Gln Glu Phe Ala
o
5'  UUGUGUCCAGAUUCGGAGUUGCAAUUUCCGGACUUGUCUCUGAGGUCAUUCAGCCUCAAGGGGCUGACCG
o                                                                             2800
3'  AACACAGGUCUAAGCCUCAACGUUAAAGGCCUGAACAGAGACUCCAGUAAGUCGGAGUUCCCCGACUGGC
1    Leu Cys Pro Asp Ser Glu Leu Gln Phe Pro Asp Leu Ser Leu Arg Ser Phe Ser Leu Lys Gly Leu Thr
2     Cys Val Gln Ile Arg Ser Cys Asn Phe Arg Thr Cys Leu  .  Gly His Ser Ala Ser Arg Gly  .  Pro
3   Val Val Ser Arg Phe Gly Val Ala Ile Ser Gly Leu Val Ser Glu Val Ile Gln Pro Gln Gly Ala Asp Arg
o
```

```
5' CCCAACCUUAAAUUUACACCUCAGGCGAAGUGCCAUCCGCGGACAAUUGCAUAUAGAGUGGGUCGAUCCU
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  2870
3' GGGUUGGAAUUUAAAUGUGGAGUCCGCUUCACGGUAGGCGCCUGUUAACGUAUAUCUCACCCAGCUAGGA
1  Ala Gln Pro . Ile Tyr Thr Ser Gly Glu Val Pro Ser Ala Asp Asn Cys Ile . Ser Gly Ser Ile Leu
2   Pro Asn Leu Lys Phe Thr Pro Gln Ala Lys Cys His Pro Arg Thr Ile Ala Tyr Arg Val Gly Arg Ser
3    Pro Thr Leu Asn Leu His Leu Arg Arg Ser Ala Ile Arg Gly Gln Leu His Ile Glu Trp Val Asp Pro
o
5' ACUACCAGCACACAACCGGUCGCCAAUCUAGCCCGUAUUUUAUCCACUAUUUUGCGGGUCCAUCCAAACU
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  2940
3' UGAUGGUCGUGUGUUGGCCAGCGGUUAGAUCGGGCAUAAAAUAGGUGAUAAAACGCCCAGGUAGGUUUGA
1   Leu Pro Ala His Asn Arg Ser Pro Ile . Pro Val Phe Tyr Pro Leu Phe Cys Gly Ser Ile Gln Thr
2  Tyr Tyr Gln His Thr Thr Gly Arg Gln Ser Ser Pro Tyr Phe Ile His Tyr Phe Ala Gly Pro Ser Lys Leu
3   Thr Thr Ser Thr Gln Pro Val Ala Asn Leu Ala Arg Ile Leu Ser Thr Ile Leu Arg Val His Pro Asn
o
5' GGGACAUCGCAAUACAAAAUGCCUUUAACCAAGCUCCGGUGGACCAAAUUCCCCGUCCCACCCCAAUUCU
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  3010
3' CCCUGUAGCGUUAUGUUUUACGGAAAUUGGUUCGAGGCCACCUGGUUUAAGGGGCAGGGUGGGGUUAAGA
1   Gly Thr Ser Gln Tyr Lys Met Pro Leu Thr Lys Leu Arg Trp Thr Lys Phe Pro Val Pro Pro Gln Phe
2    Gly His Arg Asn Thr Lys Cys Leu . Pro Ser Ser Gly Gly Pro Asn Ser Pro Ser His Pro Asn Ser
3  Trp Asp Ile Ala Ile Gln Asn Ala Phe Asn Gln Ala Pro Val Asp Gln Ile Pro Arg Pro Thr Pro Ile Leu
o
5' UUGCGGCUUCGCUGGUGAACAUGAACUGCCUAAAACUCGGGAAGACAUUGCAGAUGCCAUCGUCGCCUCA
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  3080
3' AACGCCGAAGCGACCACUUGUACUUGACGGAUUUUGAGCCCUUCUGUAACGUCUACGGUAGCAGCGGAGU
1  Phe Ala Ala Ser Leu Val Asn Met Asn Cys Leu Lys Leu Gly Lys Thr Leu Gln Met Pro Ser Ser Pro Gln
2   Leu Arg Leu Arg Trp . Thr . Thr Ala . Asn Ser Gly Arg His Cys Arg Cys His Arg Arg Leu
3    Cys Gly Phe Ala Gly Glu His Glu Leu Pro Lys Thr Arg Glu Asp Ile Ala Asp Ala Ile Val Ala Ser
o
5' AGGACUGCUCUGCCCAUCCAGGCGGCUCCUAACAACGAGGGCUUCGGAUUUGACCGAGCUGUUAUCUACA
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  3150
3' UCCUGACGAGACGGGUAGGUCCGCCGAGGAUUGUUGCUCCCGAAGCCUAAACUGGCUCGACAAUAGAUGU
1    Gly Leu Leu Cys Pro Ser Arg Arg Leu Leu Thr Thr Arg Ala Ser Asp Leu Thr Glu Leu Leu Ser Thr
2  Lys Asp Cys Ser Ala His Pro Gly Gly Ser . Gln Arg Gly Leu Arg Ile . Pro Ser Cys Tyr Leu Gln
3   Arg Thr Ala Leu Pro Ile Gln Ala Ala Pro Asn Asn Glu Gly Phe Gly Phe Asp Arg Ala Val Ile Tyr
o
5' AUGCCCCGUAUCCCCCAUUGUUCACUUUCGACCCACCUAUCGUCAACCUACCAAUUGCCAGAACAGAACC
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  3220
3' UACGGGGCAUAGGGGGUAACAAGUGAAAGCUGGGUGGAUAGCAGUUGGAUGGUUAACGGUCUUGUCUUGG
1   Met Pro Arg Ile Pro His Cys Ser Leu Ser Thr His Leu Ser Ser Thr Tyr Gln Leu Pro Glu Gln Asn
2    Cys Pro Val Ser Pro Ile Val His Phe Arg Pro Thr Tyr Arg Gln Pro Thr Asn Cys Gln Asn Arg Thr
3  Asn Ala Pro Tyr Pro Pro Leu Phe Thr Phe Asp Pro Pro Ile Val Asn Leu Pro Ile Ala Arg Thr Glu Pro
o
5' AUCUGUGAUGUGGCGGCGGUUCACUUCACUUGCUCGUACUGCUGCCGGUAGACUUCGACUUAAGGUUACC
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  3290
3' UAGACACUACACCGCCGCCAAGUGAAGUGAACGAGCAUGACGACGGCCAUCUGAAGCUGAAUUCCAAUGG
1  His Leu . Cys Gly Gly Gly Ser Leu His Leu Leu Val Leu Leu Pro Val Asp Phe Asp Leu Arg Leu Pro
2   Ile Cys Asp Val Ala Ala Val His Phe Thr Cys Ser Tyr Cys Cys Arg . Thr Ser Thr . Gly Tyr
3    Ser Val Met Trp Arg Arg Phe Thr Ser Leu Ala Arg Thr Ala Ala Gly Arg Leu Arg Leu Lys Val Thr
o
5' GAGCCACACCCCCUCACUGACAAGUCCAUGAAAAUUGGAGCAUCCCGUGAACAACACCGUUCUAAUGCCA
o  +++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|  3360
3' CUCGGUGUGGGGGAGUGACUGUUCAGGUACUUUUAACCUCGUAGGGCACUUGUUGUGGCAAGAUUACGGU
1    Ser His Thr Pro Ser Leu Thr Ser Pro . Lys Leu Glu His Pro Val Asn Asn Thr Val Leu Met Pro
2  Arg Ala Thr Pro Pro His . Gln Val His Glu Asn Trp Ser Ile Pro . Thr Thr Pro Phe . Cys His
3   Glu Pro His Pro Leu Thr Asp Lys Ser Met Lys Ile Gly Ala Ser Arg Glu Gln His Arg Ser Asn Ala
o
```

```
5'  UGCAAGGAAAUGUCCGACCGGAACCGCCUAUGUGGACCGUUUUGCGUAAUGAACUCGAGCAGGAAUUAAA
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3430
3'  ACGUUCCUUUACAGGCUGGCCUUGGCGGAUACACCUGGCAAAACGCAUUACUUGAGCUCGUCCUUAAUUU
1    Cys Lys Glu Met Ser Asp Arg Asn Arg Leu Cys Gly Pro Phe Cys Val Met Asn Ser Ser Arg Asn  .
2     Ala Arg Lys Cys Pro Thr Gly Thr Ala Tyr Val Asp Arg Phe Ala  .   .  Thr Arg Ala Gly Ile Lys
3   Met Gln Gly Asn Val Arg Pro Glu Pro Pro Met Trp Thr Val Leu Arg Asn Glu Leu Glu Gln Glu Leu Lys
o
5'  GGAAUUUAAACAAUUACGCCUACCACCGGUAAAACUGCAGCAAGAGACCAUCCAGUAUACUGCCGAUCUC
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3500
3'  CCUUAAAUUUGUUAAUGCGGAUGGUGGCCAUUUUGACGUCGUUCUCUGGUAGGUCAUAUGACGGCUAGAG
1   Arg Asn Leu Asn Asn Tyr Ala Tyr His Arg  .  Asn Cys Ser Lys Arg Pro Ser Ser Ile Leu Pro Ile Ser
2    Gly Ile  .  Thr Ile Thr Pro Thr Thr Gly Lys Thr Ala Ala Arg Asp His Pro Val Tyr Cys Arg Ser
3     Glu Phe Lys Gln Leu Arg Leu Pro Pro Val Lys Leu Gln Gln Glu Thr Ile Gln Tyr Thr Ala Asp Leu
o
5'  CAGAGAGCCAUGCGUUUAGUGGCUGAUUUGAAGGCGCACCCUUCUGUUCUCGAAUCCCGAGCCUCCCCUG
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3570
3'  GUCUCUCGGUACGCAAAUCACCGACUAAACUUCCGCGUGGGAAGACAAGAGCUUAGGGCUCGGAGGGGAC
1     Arg Glu Pro Cys Val  .  Trp Leu Ile  .  Arg Arg Thr Leu Leu Phe Ser Asn Pro Glu Pro Pro Leu
2   Pro Glu Ser His Ala Phe Ser Gly  .  Phe Glu Gly Ala Pro Phe Cys Ser Arg Ile Pro Ser Leu Pro Cys
3    Gln Arg Ala Met Arg Leu Val Ala Asp Leu Lys Ala His Pro Ser Val Leu Glu Ser Arg Ala Ser Pro
o
5'  UUGUUCUACAAAGUCUCGACUCCAUUCUGGACGCUAUCAAAGUCACACACACCACAGUCACCAAGCCAGU
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3640
3'  AACAAGAUGUUUCAGAGCUGAGGUAAGACCUGCGAUAGUUUCAGUGUGUGUGGUGUCAGUGGUUCGGUCA
1    Leu Phe Tyr Lys Val Ser Thr Pro Phe Trp Thr Leu Ser Lys Ser His Thr Pro Gln Ser Pro Ser Gln
2     Cys Ser Thr Lys Ser Arg Leu His Ser Gly Arg Tyr Gln Ser His Thr His His Ser His Gln Ala Ser
3   Val Val Leu Gln Ser Leu Asp Ser Ile Leu Asp Ala Ile Lys Val Thr His Thr Thr Val Thr Lys Pro Val
o
5'  CCACGCUUACCUUGGUGUUUGGGGCUCAGGUAAAACAACUGCCACUGUGGCUUUUCUGCACACCCUGACC
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3710
3'  GGUGCGAAUGGAACCACAAACCCCGAGUCCAUUUUGUUGACGGUGACACCGAAAAGACGUGUGGGACUGG
1   Ser Thr Leu Thr Leu Val Phe Gly Ala Gln Val Lys Gln Leu Pro Leu Trp Leu Phe Cys Thr Pro  .  Pro
2    Pro Arg Leu Pro Trp Cys Leu Gly Leu Arg  .  Asn Asn Cys His Cys Gly Phe Ser Ala His Pro Asp
3     His Ala Tyr Leu Gly Val Trp Gly Ser Gly Lys Thr Thr Ala Thr Val Ala Phe Leu His Thr Leu Thr
o
5'  CCUGAGCAACGUCGCAAUGUGAGAAUCGUGUCGCAUACUGAGUCCCUUCGCGCACAGGCCAAGUCUAAGC
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3780
3'  GGACUCGUUGCAGCGUUACACUCUUAGCACAGCGUAUGACUCAGGGAAGCGCGUGUCCGGUUCAGAUUCG
1     Leu Ser Asn Val Ala Met  .  Glu Ser Cys Arg Ile Leu Ser Pro Phe Ala His Arg Pro Ser Leu Ser
2   Pro  .  Ala Thr Ser Gln Cys Glu Asn Arg Val Ala Tyr  .  Val Pro Ser Arg Thr Gly Gln Val  .  Ala
3    Pro Glu Gln Arg Arg Asn Val Arg Ile Val Ser His Thr Glu Ser Leu Arg Ala Gln Ala Lys Ser Lys
o
5'  UUGAUUUCCCUGAGCUUCGUGGCUACAACUUCCCCACCAUUGCCUCCAUCAUCGCUGAGCCUUCCACUGG
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3850
3'  AACUAAAGGGACUCGAAGCACCGAUGUUGAAGGGGUGGUAACGGAGGUAGUAGCGACUCGGAAGGUGACC
1    Leu Ile Ser Leu Ser Phe Val Ala Thr Thr Ser Pro Pro Leu Pro Pro Ser Ser Leu Ser Leu Pro Leu
2     .  Phe Pro  .  Ala Ser Trp Leu Gln Leu Pro His His Cys Leu His His Arg  .  Ala Phe His Trp
3   Leu Asp Phe Pro Glu Leu Arg Gly Tyr Asn Phe Pro Thr Ile Ala Ser Ile Ile Ala Glu Pro Ser Thr Gly
o
5'  GGGCGUCAUAUUUGACGAUGCUGGUAAGUACUGGGGGGGUGUUCUGGAUCUGGUCAUGCUCACUAAUCCU
o   +++++|++++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|+++++|++++++|   3920
3'  CCCGCAGUAUAAACUGCUACGACCAUUCAUGACCCCCCCACAAGACCUAGACCAGUACGAGUGAUUAGGA
1   Gly Ala Ser Tyr Leu Thr Met Leu Val Ser Thr Gly Gly Val Phe Trp Ile Trp Ser Cys Ser Leu Ile Leu
2    Gly Arg His Ile  .  Arg Cys Trp  .  Val Leu Gly Gly Cys Ser Gly Ser Gly His Ala His  .  Ser
3     Gly Val Ile Phe Asp Asp Ala Gly Lys Tyr Trp Gly Gly Val Leu Asp Leu Val Met Leu Thr Asn Pro
o
```

```
5'  UUAGCAGAAUUCUUCGUCGUGAACGGUGACCCAUGCCAGACCACCUCCAGGUUUCCGGUUCCUGGAUCCC
o                                                                          3990
3'  AAUCGUCUUAAGAAGCAGCACUUGCCACUGGGUACGGUCUGGUGGAGGUCCAAAGGCCAAGGACCUAGGG
1     .   Gln Asn Ser Ser Ser  .  Thr Val Thr His Ala Arg Pro Pro Pro Gly Phe Arg Phe Leu Asp Pro
2   Phe Ser Arg Ile Leu Arg Arg Glu Arg  .  Pro Met Pro Asp His Leu Gln Val Ser Gly Ser Trp Ile Pro
3    Leu Ala Glu Phe Phe Val Val Asn Gly Asp Pro Cys Gln Thr Thr Ser Arg Phe Pro Val Pro Gly Ser
o
5'  AGAGUGAGUUUGAUCCCAGCCCCAUAACCUGUAUAGCCAACCUUGCUACUCGUUAUGCCACUGUCAGCCA
o                                                                          4060
3'  UCUCACUCAAACUAGGGUCGGGGUAUUGGACAUAUCGGUUGGAACGAUGAGCAAUACGGUGACAGUCGGU
1    Arg Val Ser Leu Ile Pro Ala Pro  .  Pro Val  .  Pro Thr Leu Leu Leu Val Met Pro Leu Ser Ala
2     Glu  .  Val  .  Ser Gln Pro His Asn Leu Tyr Ser Gln Pro Cys Tyr Ser Leu Cys His Cys Gln Pro
3   Gln Ser Glu Phe Asp Pro Ser Pro Ile Thr Cys Ile Ala Asn Leu Ala Thr Arg Tyr Ala Thr Val Ser His
o
5'  CAGGUCUUUCGGUUUGCUUGCAAACACUCUUGGUGUACACACCACCAAUCCAAACAACGGACACAUCACC
o                                                                          4130
3'  GUCCAGAAAGCCAAACGAACGUUUGUGAGAACCACAUGUGUGGUGGUUAGGUUUGUUGCCUGUGUAGUGG
1   Thr Gly Leu Ser Val Cys Leu Gln Thr Leu Leu Val Tyr Thr Pro Pro Ile Gln Thr Thr Asp Thr Ser Pro
2    Gln Val Phe Arg Phe Ala Cys Lys His Ser Trp Cys Thr His His Gln Ser Lys Gln Arg Thr His His
3     Arg Ser Phe Gly Leu Leu Ala Asn Thr Leu Gly Val His Thr Thr Asn Pro Asn Asn Gly His Ile Thr
o
5'  CACACAGUAACUGGCAAGACUGGCAUCCCUGUCUGUACAGCGUCCCCACGGUACGUGCAGGUGCUUGGCA
o                                                                          4200
3'  GUGUGUCAUUGACCGUUCUGACCGUAGGGACAGACAUGUCGCAGGGGUGCCAUGCACGUCCACGAACCGU
1    Thr Gln  .  Leu Ala Arg Leu Ala Ser Leu Ser Val Gln Arg Pro His Gly Thr Cys Arg Cys Leu Ala
2   Pro His Ser Asn Trp Gln Asp Trp His Pro Cys Leu Tyr Ser Val Pro Thr Val Arg Ala Gly Ala Trp Gln
3    His Thr Val Thr Gly Lys Thr Gly Ile Pro Val Cys Thr Ala Ser Pro Arg Tyr Val Gln Val Leu Gly
o
5'  GUUAUGGCCGACAGGCUUACACUUACUCAAGCAUUCAAGGUGAAGACUUUGAUCAGGAUGUGGAAAUUGA
o                                                                          4270
3'  CAAUACCGGCUGUCCGAAUGUGAAUGAGUUCGUAAGUUCCACUUCUGAAACUAGUCCUACACCUUUAACU
1    Val Met Ala Asp Arg Leu Thr Leu Thr Gln Ala Phe Lys Val Lys Thr Leu Ile Arg Met Trp Lys Leu
2     Leu Trp Pro Thr Gly Leu His Leu Leu Lys His Ser Arg  .  Arg Leu  .  Ser Gly Cys Gly Asn  .
3   Ser Tyr Gly Arg Gln Ala Tyr Thr Tyr Ser Ser Ile Gln Gly Glu Asp Phe Asp Gln Asp Val Glu Ile Asp
o
5'  CAUGACCGGCCUGGAAGGGGCCGUUCUUGAUUCAGCAGCUUAUGUCGCACUUACUCGCAGUAAGACCGGU
o                                                                          4340
3'  GUACUGGCCGGACCUUCCCCGGCAAGAACUAAGUCGUCGAAUACAGCGUGAAUGAGCGUCAUUCUGGCCA
1   Thr  .  Pro Ala Trp Lys Gly Pro Phe Leu Ile Gln Gln Leu Met Ser His Leu Leu Ala Val Arg Pro Val
2    His Asp Arg Pro Gly Arg Gly Arg Ser  .  Phe Ser Ser Leu Cys Arg Thr Tyr Ser Gln  .  Asp Arg
3     Met Thr Gly Leu Glu Gly Ala Val Leu Asp Ser Ala Ala Tyr Val Ala Leu Thr Arg Ser Lys Thr Gly
o
5'  GUCUAUGUCCACAUGGAAGCUAUGGACCCACGCUCAACUAUCCGCAAGCCACCAACCGGAAGUGAAAUUA
o                                                                          4410
3'  CAGAUACAGGUGUACCUUCGAUACCUGGGUGCGAGUUGAUAGGCGUUCGGUGGUUGGCCUUCACUUUAAU
1    Ser Met Ser Thr Trp Lys Leu Trp Thr His Ala Gln Leu Ser Ala Ser His Gln Pro Glu Val Lys Leu
2   Cys Leu Cys Pro His Gly Ser Tyr Gly Pro Thr Leu Asn Tyr Pro Gln Ala Thr Asn Arg Lys  .  Asn Tyr
3    Val Tyr Val His Met Glu Ala Met Asp Pro Arg Ser Thr Ile Arg Lys Pro Pro Thr Gly Ser Glu Ile
o
5'  UCAAUGCCUUGGUCUACUGUGUACGCGCUGGUAAUUCCAAUCAGCUGUUGCAGCCUGAUUGGCUUAUUAA
o                                                                          4480
3'  AGUUACGGAACCAGAUGACACAUGCGCGACCAUUAAGGUUAGUCGACAACGUCGGACUAACCGAAUAAUU
1    Ser Met Pro Trp Ser Thr Val Tyr Ala Leu Val Ile Pro Ile Ser Cys Cys Ser Leu Ile Gly Leu Leu
2     Gln Cys Leu Gly Leu Leu Cys Thr Arg Trp  .  Phe Gln Ser Ala Val Ala Ala  .  Leu Ala Tyr  .
3   Ile Asn Ala Leu Val Tyr Cys Val Arg Ala Gly Asn Ser Asn Gln Leu Leu Gln Pro Asp Trp Leu Ile Lys
o
```

```
5'  GGCCGCCUUUUAUCGCCACAUUAAGCGCUGCAUUCCCUGCUUGCCCUGGUUUGCCGACAUUGGGGCUAGU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   4550
3'  CCGGCGGAAAAUAGCGGUGUAAUUCGCGACGUAAGGGACGAACGGGACCAAACGGCUGUAACCCCGAUCA
1   Arg Pro Pro Phe Ile Ala Thr Leu Ser Ala Ala Phe Pro Ala Cys Pro Gly Leu Pro Thr Leu Gly Leu Val
2    Gly Arg Leu Leu Ser Pro His . Ala Leu His Ser Leu Leu Ala Leu Val Cys Arg His Trp Gly .
3     Ala Ala Phe Tyr Arg His Ile Lys Arg Cys Ile Pro Cys Leu Pro Trp Phe Ala Asp Ile Gly Ala Ser
o
5'  GUAGGUGUGGAGCAUUUUCAGCAAGUCUUGCCCGUGUCCAAUUAUCACGUCAUUGACGAAGGAAGCACUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   4620
3'  CAUCCACACCUCGUAAAAGUCGUUCAGAACGGGCACAGGUUAAUAGUGCAGUAACUGCUUCCUUCGUGAA
1    . Val Trp Ser Ile Phe Ser Lys Ser Cys Pro Cys Pro Ile Ile Thr Ser Leu Thr Lys Glu Ala Leu
2   Cys Arg Cys Gly Ala Phe Ser Ala Ser Leu Ala Arg Val Gln Leu Ser Arg His . Arg Arg Lys His Phe
3    Val Gly Val Glu His Phe Gln Gln Val Leu Pro Val Ser Asn Tyr His Val Ile Asp Glu Gly Ser Thr
o
5'  CUGAAAUGCUGGCUAAUGACGCGGUACAGCCCAGUGACGGACCCAUUGAAAACCUCGUGCCAGAAACACA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   4690
3'  GACUUUACGACCGAUUACUGCGCCAUGUCGGGUCACUGCCUGGGUAACUUUUGGAGCACGGUCUUUGUGU
1    Leu Lys Cys Trp Leu Met Thr Arg Tyr Ser Pro Val Thr Asp Pro Leu Lys Thr Ser Cys Gln Lys His
2     . Asn Ala Gly . . Arg Gly Thr Ala Gln . Arg Thr His . Lys Pro Arg Ala Arg Asn Thr
3   Ser Glu Met Leu Ala Asn Asp Ala Val Gln Pro Ser Asp Gly Pro Ile Glu Asn Leu Val Pro Glu Thr His
o
5'  CUUUAUUGCUAAGGAGCACCGUGAAGAGCGCGUUCGUGGUGGUGGAACUGAUCAAUUUAAAGAAACCGCG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   4760
3'  GAAAUAACGAUUCCUCGUGGCACUUCUCGCGCAAGCACCACCACCUUGACUAGUUAAAUUUCUUUGGCGC
1   Thr Leu Leu Leu Arg Ser Thr Val Lys Ser Ala Phe Val Val Val Glu Leu Ile Asn Leu Lys Lys Pro Arg
2    Leu Tyr Cys . Gly Ala Pro . Arg Ala Arg Ser Trp Trp Trp Asn . Ser Ile . Arg Asn Arg
3     Phe Ile Ala Lys Glu His Arg Glu Glu Arg Val Arg Gly Gly Gly Thr Asp Gln Phe Lys Glu Thr Ala
o
5'  UUUGUUAACCCACAUGUACACAAACGCAAUGACACCGCCACUUACUUUCUUUCGGUUGAACAAAGGCUGA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   4830
3'  AAACAAUUGGGUGUACAUGUGUUUGCGUUACUGUGGCGGUGAAUGAAAGAAAGCCAACUUGUUUCCGACU
1     Leu Leu Thr His Met Tyr Thr Asn Ala Met Thr Pro Pro Leu Thr Phe Phe Arg Leu Asn Lys Gly .
2   Val Cys . Pro Thr Cys Thr Gln Thr Gln . His Arg His Leu Leu Ser Phe Gly . Thr Lys Ala Glu
3    Phe Val Asn Pro His Val His Lys Arg Asn Asp Thr Ala Thr Tyr Phe Leu Ser Val Glu Gln Arg Leu
o
5'  AACCUUCUUCUUAUGAAGCAAAUCUCGCACGCAUGGCCUCGGUUCAGCGUCGAGACAUGUGUGAAGCAUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   4900
3'  UUGGAAGAAGAAUACUUCGUUUAGAGCGUGCGUACCGGAGCCAAGUCGCAGCUCUGUACACACUUCGUAU
1    Asn Leu Leu Leu Met Lys Gln Ile Ser His Ala Trp Pro Arg Phe Ser Val Glu Thr Cys Val Lys His
2     Thr Phe Phe Leu . Ser Lys Ser Arg Thr His Gly Leu Gly Ser Ala Ser Arg His Val . Ser Ile
3   Lys Pro Ser Ser Tyr Glu Ala Asn Leu Ala Arg Met Ala Ser Val Gln Arg Arg Asp Met Cys Glu Ala Tyr
o
5'  CGACAAGCUUGUGCCCCACCCCCCAAAGUGGUCUGCAGAAAAGCAUCAACACUAUGUCGAUCAAUGCAUC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   4970
3'  GCUGUUCGAACACGGGGUGGGGGGUUUCACCAGACGUCUUUUCGUAGUUGUGAUACAGCUAGUUACGUAG
1   Thr Thr Ser Leu Cys Pro Thr Pro Gln Ser Gly Leu Gln Lys Ser Ile Asn Thr Met Ser Ile Asn Ala Ser
2    Arg Gln Ala Cys Ala Pro Pro Pro Lys Val Val Cys Arg Lys Ala Ser Thr Leu Cys Arg Ser Met His
3     Asp Lys Leu Val Pro His Pro Pro Lys Trp Ser Ala Glu Lys His Gln His Tyr Val Asp Gln Cys Ile
o
5'  AGCGAAUACUGUUCUAAGCGCACGGAAUCUGCUGUCCUGUCCAAACUCAAAGCACAUGACCCCACGCGUA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5040
3'  UCGCUUAUGACAAGAUUCGCGUGCCUUAGACGACAGGACAGGUUUGAGUUUCGUGUACUGGGGUGCGCAU
1     Ala Asn Thr Val Leu Ser Ala Arg Asn Leu Leu Ser Cys Pro Asn Ser Lys His Met Thr Pro Arg Val
2   Gln Arg Ile Leu Phe . Ala His Gly Ile Cys Cys Pro Val Gln Thr Gln Ser Thr . Pro His Ala Tyr
3    Ser Glu Tyr Cys Ser Lys Arg Thr Glu Ser Ala Val Leu Ser Lys Leu Lys Ala His Asp Pro Thr Arg
o
```

```
5'  CCGGGUCUGACAUCGUCAUCUCACUUAAGAAUCAGGUCAUCAAGAAAGACGAAAAACGCCACAAAAUCAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5110
3'  GGCCCAGACUGUAGCAGUAGAGUGAAUUCUUAGUCCAGUAGUUCUUUCUGCUUUUUGCGGUGUUUUAGUU
1    Pro Gly Leu Thr Ser Ser Ser His Leu Arg Ile Arg Ser Ser Arg Lys Thr Lys Asn Ala Thr Lys Ser
2     Arg Val . His Arg His Leu Thr . Glu Ser Gly His Gln Glu Arg Arg Lys Thr Pro Gln Asn Gln
3   Thr Gly Ser Asp Ile Val Ile Ser Leu Lys Asn Gln Val Ile Lys Lys Asp Glu Lys Arg His Lys Ile Lys
o
5'  AGCCAUACCAGGACAGCUCAUUCACGAGUAUGACAUCUCCAUCACUUUGGGCGAUGCGCCUUAUGCUCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5180
3'  UCGGUAUGGUCCUGUCGAGUAAGUGCUCAUACUGUAGAGGUAGUGAAACCCGCUACGCGGAAUACGAGAC
1   Lys Pro Tyr Gln Asp Ser Ser Phe Thr Ser Met Thr Ser Pro Ser Leu Trp Ala Met Arg Leu Met Leu Cys
2    Ser His Thr Arg Thr Ala His Ser Arg Val . His Leu His His Phe Gly Arg Cys Ala Leu Cys Ser
3     Ala Ile Pro Gly Gln Leu Ile His Glu Tyr Asp Ile Ser Ile Thr Leu Gly Asp Ala Pro Tyr Ala Leu
o
5'  UUUCUGGAGAAUGAAAUCAUCCCGGCCUUCCCAGACAACUAUCUGUUCUACAGACGCAUGUCCCCCCCUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5250
3'  AAAGACCUCUUACUUUAGUAGGGCCGGAAGGGUCUGUUGAUAGACAAGAUGUCUGCGUACAGGGGGGGAC
1     Phe Trp Arg Met Lys Ser Ser Arg Pro Ser Gln Thr Thr Ile Cys Ser Thr Asp Ala Cys Pro Pro Leu
2   Val Ser Gly Glu . Asn His Pro Gly Leu Pro Arg Gln Leu Ser Val Leu Gln Thr His Val Pro Pro .
3    Phe Leu Glu Asn Glu Ile Ile Pro Ala Phe Pro Asp Asn Tyr Leu Phe Tyr Arg Arg Met Ser Pro Pro
o
5'  AGUUCAUCAAGGCUUACAAGUCCAAGUGGCGUGUGAACAAUGGUGCAUACUCCUCUGAUGUGACAAGGUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5320
3'  UCAAGUAGUUCCGAAUGUUCAGGUUCACCGCACACUUGUUACCACGUAUGAGGAGACUACACUGUUCCAC
1    Ser Ser Ser Arg Leu Thr Ser Pro Ser Gly Val . Thr Met Val His Thr Pro Leu Met . Gln Gly
2     Val His Gln Gly Leu Gln Val Gln Val Ala Cys Glu Gln Trp Cys Ile Leu Leu . Cys Asp Lys Val
3   Glu Phe Ile Lys Ala Tyr Lys Ser Lys Trp Arg Val Asn Asn Gly Ala Tyr Ser Ser Asp Val Thr Arg Trp
o
5'  GGACGUGGGUUGCGAUGCCGGCAUGUUGAAUUUUGAUGUCCACGUCAUGCAACGCUCCGGAUUCCCAGCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5390
3'  CCUGCACCCAACGCUACGGCCGUACAACUUAAAACUACAGGUGCAGUACGUUGCGAGGCCUAAGGGUCGA
1   Gly Thr Trp Val Ala Met Pro Ala Cys . Ile Leu Met Ser Thr Ser Cys Asn Ala Pro Asp Ser Gln Leu
2    Gly Arg Gly Leu Arg Cys Arg His Val Glu Phe . Cys Pro Arg His Ala Thr Leu Arg Ile Pro Ser
3     Asp Val Gly Cys Asp Ala Gly Met Leu Asn Phe Asp Val His Val Met Gln Arg Ser Gly Phe Pro Ala
o
5'  GAUUACAUUGAGGCCUACAUCACGCGCAGACUGUCGUCCAAAUCCCAACAUGGUGUUAUGGCCACCAUGC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5460
3'  CUAAUGUAACUCCGGAUGUAGUGCGCGUCUGACAGCAGGUUUAGGGUUGUACCACAAUACCGGUGGUACG
1     Ile Thr Leu Arg Pro Thr Ser Arg Ala Asp Cys Arg Pro Asn Pro Asn Met Val Leu Trp Pro Pro Cys
2   . Leu His . Gly Leu His His Ala Gln Thr Val Val Gln Ile Pro Thr Trp Cys Tyr Gly His His Ala
3    Asp Tyr Ile Glu Ala Tyr Ile Thr Arg Arg Leu Ser Ser Lys Ser Gln His Gly Val Met Ala Thr Met
o
5'  AGAACUCUGGUGAUAGGUACACCUGGCCUCUAAACACCGUUAGACGCGCCGUGGUCACGAGUAUCGUUUG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5530
3'  UCUUGAGACCACUAUCCAUGUGGACCGGAGAUUUGUGGCAAUCUGCGCGGCACCAGUGCUCAUAGCAAAC
1    Arg Thr Leu Val Ile Gly Thr Pro Gly Leu . Thr Pro Leu Asp Ala Pro Trp Ser Arg Val Ser Phe
2     Glu Leu Trp . . Val His Leu Ala Ser Lys His Arg . Thr Arg Arg Gly His Glu Tyr Arg Leu
3   Gln Asn Ser Gly Asp Arg Tyr Thr Trp Pro Leu Asn Thr Val Arg Arg Ala Val Val Thr Ser Ile Val Cys
o
5'  CGAUGUGAAACCGGAAGACACGGUUGCCGUUAACGGCGAUGACGCUGCCAUUGAUCGUAACUGUAGUGCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   5600
3'  GCUACACUUUGGCCUUCUGUGCCAACGGCAAUUGCCGCUACUGCGACGGUAACUAGCAUUGACAUCACGG
1   Ala Met . Asn Arg Lys Thr Arg Leu Pro Leu Thr Ala Met Thr Leu Pro Leu Ile Val Thr Val Val Pro
2    Arg Cys Glu Thr Gly Arg His Gly Cys Arg . Arg Arg . Arg Cys His . Ser . Leu . Cys
3     Asp Val Lys Pro Glu Asp Thr Val Ala Val Asn Gly Asp Asp Ala Ala Ile Asp Arg Asn Cys Ser Ala
o
```

```
5'   ACACAGUUCCCUGACUCACCUUGGGUCUUCAAGGACUGUAACGGUUACCGCGUUGAGUUCUCUGGCUUUG
o                                                                             5670
3'   UGUGUCAAGGGACUGAGUGGAACCCAGAAGUUCCUGACAUUGCCAAUGGCGCAACUCAAGAGACCGAAAC
1      His Ser Ser Leu Thr His Leu Gly Ser Ser Arg Thr Val Thr Val Thr Ala Leu Ser Ser Leu Ala Leu
2    His Thr Val Pro  .  Leu Thr Leu Gly Leu Gln Gly Leu  .  Arg Leu Pro Arg  .  Val Leu Trp Leu  .
3     Thr Gln Phe Pro Asp Ser Pro Trp Val Phe Lys Asp Cys Asn Gly Tyr Arg Val Glu Phe Ser Gly Phe
o
5'   AACUCGGCGGCCCUGAGCCAACCUAUUCCGCUUCUGGGAUUUGGUACAGGACUGCCAUUUUGAUGUCCAG
o                                                                             5740
3'   UUGAGCCGCCGGGACUCGGUUGGAUAAGGCGAAGACCCUAAACCAUGUCCUGACGGUAAAACUACAGGUC
1     Asn Ser Ala Ala Leu Ser Gln Pro Ile Pro Leu Leu Gly Phe Gly Thr Gly Leu Pro Phe  .  Cys Pro
2      Thr Arg Arg Pro  .  Ala Asn Leu Phe Arg Phe Trp Asp Leu Val Gln Asp Cys His Phe Asp Val Gln
3    Glu Leu Gly Gly Pro Glu Pro Thr Tyr Ser Ala Ser Gly Ile Trp Tyr Arg Thr Ala Ile Leu Met Ser Arg
o
5'   AGACCCUUCCGCACAGGACAAAUGGGUUAAUUAUUUAGACCUGCUGCAGUACACGGACCUCAACGAUCCU
o                                                                             5810
3'   UCUGGGAAGGCGUGUCCUGUUUACCCAAUUAAUAAAUCUGGACGACGUCAUGUGCCUGGAGUUGCUAGGA
1    Glu Thr Leu Pro His Arg Thr Asn Gly Leu Ile Ile  .  Thr Cys Cys Ser Thr Arg Thr Ser Thr Ile Leu
2     Arg Pro Phe Arg Thr Gly Gln Met Gly  .  Leu Phe Arg Pro Ala Ala Val His Gly Pro Gln Arg Ser
3      Asp Pro Ser Ala Gln Asp Lys Trp Val Asn Tyr Leu Asp Leu Leu Gln Tyr Thr Asp Leu Asn Asp Pro
o
5'   CAUGCCAUUGAUGUGGCUAGGGCCGCUCAUCAACACAUGAAGCCCGUCAGCCGUUUCGCUGAAUGCUUAC
o                                                                             5880
3'   GUACGGUAACUACACCGAUCCCGGCGAGUAGUUGUGUACUUCGGGCAGUCGGCAAAGCGACUUACGAAUG
1      Met Pro Leu Met Trp Leu Gly Pro Leu Ile Asn Thr  .  Ser Pro Ser Ala Val Ser Leu Asn Ala Tyr
2    Ser Cys His  .  Cys Gly  .  Gly Arg Ser Ser Thr His Glu Ala Arg Gln Pro Phe Arg  .  Met Leu Thr
3     His Ala Ile Asp Val Ala Arg Ala Ala His Gln His Met Lys Pro Val Ser Arg Phe Ala Glu Cys Leu
o
5'   CUGAACCACUACGCCCCCACUUCCCCACAGUAGUUUUUUAGUGGAGUCUUAUUGUACUUCCCUUCUCUUG
o                                                                             5950
3'   GACUUGGUGAUGCGGGGGUGAAGGGGUGUCAUCAAAAAAUCACCUCAGAAUAACAUGAAGGGAAGAGAAC
1     Leu Asn His Tyr Ala Pro Thr Ser Pro Gln  .  Phe Phe Ser Gly Val Leu Leu Tyr Phe Pro Ser Leu
2      .  Thr Thr Thr Pro Pro Leu Pro His Ser Ser Phe Leu Val Glu Ser Tyr Cys Thr Ser Leu Leu Leu
3    Pro Glu Pro Leu Arg Pro His Phe Pro Thr Val Val Phe  .  Trp Ser Leu Ile Val Leu Pro Phe Ser  .
o
5'   AUGCAACUUCUCUUCCCUUUCCCUUGCUUAGCUUAACCUAACUCGACUCCCAAUUUUUUUUCUACUUCGA
o                                                                             6020
3'   UACGUUGAAGAGAAGGGAAAGGGAACGAAUCGAAUUGGAUUGAGCUGAGGGUUAAAAAAAAGAUGAAGCU
1    Asp Ala Thr Ser Leu Pro Phe Pro Leu Leu Ser Leu Thr  .  Leu Asp Ser Gln Phe Phe Phe Tyr Phe Asp
2     Met Gln Leu Leu Phe Pro Phe Pro Cys Leu Ala  .  Pro Asn Ser Thr Pro Asn Phe Phe Ser Thr Ser
3      Cys Asn Phe Ser Ser Leu Ser Leu Ala  .  Leu Asn Leu Thr Arg Leu Pro Ile Phe Phe Leu Leu Arg
o
5'   CCUCAGCAAGUCGAUAAACUGCUAUAGCGUUUGUCAUCACGCUUCUCAUGGCCCAGGGGCGGGCCUAGAC
o                                                                             6090
3'   GGAGUCGUUCAGCUAUUUGACGAUAUCGCAAACAGUAGUGCGAAGAGUACCGGGUCCCCGCCCGGAUCUG
1      Leu Ser Lys Ser Ile Asn Cys Tyr Ser Val Cys His His Ala Ser His Gly Pro Gly Ala Gly Leu Asp
2    Thr Ser Ala Ser Arg  .  Thr Ala Ile Ala Phe Val Ile Thr Leu Leu Met Ala Gln Gly Arg Ala  .  Thr
3     Pro Gln Gln Val Asp Lys Leu Leu  .  Arg Leu Ser Ser Arg Phe Ser Trp Pro Arg Gly Gly Pro Arg
o
5'   CGAAAACUUGGACCUUGAUCCCACCGCGAGGUGGCCAGUGCAUGCGUGCCAAGCAUGUGCGUGUCAGCUG
o                                                                             6160
3'   GCUUUUGAACCUGGAACUAGGGUGGCGCUCCACCGGUCACGUACGCACGGUUCGUACACGCACAGUCGAC
1     Arg Lys Leu Gly Pro  .  Ser His Arg Glu Val Ala Ser Ala Cys Val Pro Ser Met Cys Val Ser Ala
2      Glu Asn Leu Asp Leu Asp Pro Thr Ala Arg Trp Pro Val His Ala Cys Gln Ala Cys Ala Cys Gln Leu
3    Pro Lys Thr Trp Thr Leu Ile Pro Pro Arg Gly Gly Gln Cys Met Arg Ala Lys His Val Arg Val Ser Cys
o
```

```
5'  CUCAGUCCAGGAGCGUCCCAAAGCUGGCUGUCCGUGUCGAUCUUCUAGGGCGUAGGCCUGCUACGAUGAA
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6230
3'  GAGUCAGGUCCUCGCAGGGUUUCGACCGACAGGCACAGCUAGAAGAUCCCGCAUCCGGACGAUGCUACUU
1   Ala Gln Ser Arg Ser Val Pro Lys Leu Ala Val Arg Val Asp Leu Leu Gly Arg Arg Pro Ala Thr Met Lys
2    Leu Ser Pro Gly Ala Ser Gln Ser Trp Leu Ser Val Ser Ile Phe . Gly Val Gly Leu Leu Arg .
3     Ser Val Gln Glu Arg Pro Lys Ala Gly Cys Pro Cys Arg Ser Ser Arg Ala . Ala Cys Tyr Asp Glu
o
5'  GAAGAGGAGUUCAGGGAAGUUUGCCUGACAGGUCCCACAGUCAAGGACCAGUGUGACAACACUUUUCACA
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6300
3'  CUUCUCCUCAAGUCCCUUCAAACGGACUGUCCAGGGUGUCAGUUCCUGGUCACACUGUUGUGAAAAGUGU
1    Lys Arg Ser Ser Gly Lys Phe Ala . Gln Val Pro Gln Ser Arg Thr Ser Val Thr Thr Leu Phe Thr
2   Arg Arg Gly Val Gln Gly Ser Leu Pro Asp Arg Ser His Ser Gln Gly Pro Val . Gln His Phe Ser His
3     Glu Glu Glu Phe Arg Glu Val Cys Leu Thr Gly Pro Thr Val Lys Asp Gln Cys Asp Asn Thr Phe His
o
5'  CACUGUAGAGACUGCACACACGGCGGGCCUAUAGCACCAAUCCACUCACUCGUUUGGGAUUUGUGUUGGG
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6370
3'  GUGACAUCUCUGACGUGUGUGCCGCCCGGAUAUCGUGGUUAGGUGAGUGAGCAAACCCUAAACACAACCC
1    His Cys Arg Asp Cys Thr His Gly Gly Pro Ile Ala Pro Ile His Ser Leu Val Trp Asp Leu Cys Trp
2     Thr Val Glu Thr Ala His Thr Ala Gly Leu . His Gln Ser Thr His Ser Phe Gly Ile Cys Val Gly
3   Thr Leu . Arg Leu His Thr Arg Arg Ala Tyr Ser Thr Asn Pro Leu Thr Arg Leu Gly Phe Val Leu Gly
o
5'  UCUGAAAGGAACAGUCCACCCACCUCUGAGAAACCGCAUUUCGAUCGUCUCAAGUGGUUGCAUCAUUAUG
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6440
3'  AGACUUUCCUUGUCAGGUGGGUGGAGACUCUUUGGCGUAAAGCUAGCAGAGUUCACCAACGUAGUAAUAC
1   Val . Lys Glu Gln Ser Thr His Leu . Glu Thr Ala Phe Arg Ser Ser Gln Val Val Ala Ser Leu Trp
2    Ser Glu Arg Asn Ser Pro Pro Thr Ser Glu Lys Pro His Phe Asp Arg Leu Lys Trp Leu His His Tyr
3     Leu Lys Gly Thr Val His Pro Pro Leu Arg Asn Arg Ile Ser Ile Val Ser Ser Gly Cys Ile Ile Met
o
5'  GCAACUGUCUUCGAUAACAACAUUAACAACAUUCGUAACAUCUUGGAUCAAGUUUCAGGAACUACACUCG
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6510
3'  CGUUGACAGAAGCUAUUGUUGUAAUUGUUGUAAGCAUUGUAGAACCUAGUUCAAAGUCCUUGAUGUGAGC
1     Gln Leu Ser Ser Ile Thr Thr Leu Thr Thr Phe Val Thr Ser Trp Ile Lys Phe Gln Glu Leu His Ser
2   Gly Asn Cys Leu Arg . Gln His . Gln His Ser . His Leu Gly Ser Ser Phe Arg Asn Tyr Thr Arg
3    Ala Thr Val Phe Asp Asn Asn Ile Asn Asn Ile Arg Asn Ile Leu Asp Gln Val Ser Gly Thr Thr Leu
o
5'  AAGCCGGUCACUUAAAAGGACACGUCGUCACCAAAGCUCAGUUGUCUGCGGUCAACGAGUUUCUUGACAA
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6580
3'  UUCGGCCAGUGAAUUUUCCUGUGCAGCAGUGGUUUCGAGUCAACAGACGCCAGUUGCUCAAAGAACUGUU
1    Lys Pro Val Thr . Lys Asp Thr Ser Ser Pro Lys Leu Ser Cys Leu Arg Ser Thr Ser Phe Leu Thr
2     Ser Arg Ser Leu Lys Arg Thr Arg Arg His Gln Ser Ser Val Val Cys Gly Gln Arg Val Ser . Gln
3   Glu Ala Gly His Leu Lys Gly His Val Val Thr Lys Ala Gln Leu Ser Ala Val Asn Glu Phe Leu Asp Lys
o
5'  GUUGGCUGGUGAAUCUCCUCGCUGUCUCACACCCACGACGGAGUCGUUUGAACUUCCCGAACCAUUAGCC
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6650
3'  CAACCGACCACUUAGAGGAGCGACAGAGUGUGGGUGCUGCCUCAGCAAACUUGAAGGGCUUGGUAAUCGG
1   Ser Trp Leu Val Asn Leu Leu Ala Val Ser His Pro Arg Arg Ser Arg Leu Asn Phe Pro Asn His . Pro
2    Val Gly Trp . Ile Ser Ser Leu Ser His Thr His Asp Gly Val Val . Thr Ser Arg Thr Ile Ser
3     Leu Ala Gly Glu Ser Pro Arg Cys Leu Thr Pro Thr Thr Glu Ser Phe Glu Leu Pro Glu Pro Leu Ala
o
5'  CGCUCCAAGCUAUCCCUAUUGGACCGCAUCUCGGUUCUUGAAACUAAGCUCACGCUAGCUGAUCGUGCAG
o   ----+----|----+----|----+----|----+----|----+----|----+----|----+----|  6720
3'  GCGAGGUUCGAUAGGGAUAACCUGGCGUAGAGCCAAGAACUUUGAUUCGAGUGCGAUCGACUAGCACGUC
1     Ala Pro Ser Tyr Pro Tyr Trp Thr Ala Ser Arg Phe Leu Lys Leu Ser Ser Arg . Leu Ile Val Gln
2   Pro Leu Gln Ala Ile Pro Ile Gly Pro His Leu Gly Ser . Asn . Ala His Ala Ser . Ser Cys Arg
3    Arg Ser Lys Leu Ser Leu Leu Asp Arg Ile Ser Val Leu Glu Thr Lys Leu Thr Leu Ala Asp Arg Ala
o
```

```
5'  AACGUACGGAAUUGAACCACGAGAUUGGCAAGGCUCGAUCAAAGCUUGCCGCUAUUGAACGCUCCAUCAA
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   6790
3'  UUGCAUGCCUUAACUUGGUGCUCUAACCGUUCCGAGCUAGUUUCGAACGGCGAUAACUUGCGAGGUAGUU
1    Asn Val Arg Asn . Thr Thr Arg Leu Ala Arg Leu Asp Gln Ser Leu Pro Leu Leu Asn Ala Pro Ser
2     Thr Tyr Gly Ile Glu Pro Arg Asp Trp Gln Gly Ser Ile Lys Ala Cys Arg Tyr . Thr Leu His Gln
3   Glu Arg Thr Glu Leu Asn His Glu Ile Gly Lys Ala Arg Ser Lys Leu Ala Ala Ile Glu Arg Ser Ile Lys
o
5'  GAAAUUGUAGAUGUUUAGCCACCCUUUUGUCUGUGUUUCACUCCUCAACUUGUUUUGGGCGUGUUUUCUU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   6860
3'  CUUUAACAUCUACAAAUCGGUGGGAAAACAGACACAAAGUGAGGAGUUGAACAAAACCCGCACAAAAGAA
1   Arg Asn Cys Arg Cys Leu Ala Thr Leu Leu Ser Val Phe His Ser Ser Thr Cys Phe Gly Arg Val Phe Leu
2    Glu Ile Val Asp Val . Pro Pro Phe Cys Leu Cys Phe Thr Pro Gln Leu Val Leu Gly Val Phe Ser
3     Lys Leu . Met Phe Ser His Pro Phe Val Cys Val Ser Leu Leu Asn Leu Phe Trp Ala Cys Phe Leu
o
5'  GGAUUGCUGGGCCGAAUCCACAGCUUCGAUCAAUUUGCUCGCAGUGUUGUUUUGUUUCCCUAUUGGGCAG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   6930
3'  CCUAACGACCCGGCUUAGGUGUCGAAGCUAGUUAAACGAGCGUCACAACAAAACAAAGGGAUAACCCGUC
1     Asp Cys Trp Ala Glu Ser Thr Ala Ser Ile Asn Leu Leu Ala Val Leu Phe Cys Phe Pro Ile Gly Gln
2   Trp Ile Ala Gly Pro Asn Pro Gln Leu Arg Ser Ile Cys Ser Gln Cys Cys Phe Val Ser Leu Leu Gly Ser
3    Gly Leu Leu Gly Arg Ile His Ser Phe Asp Gln Phe Ala Arg Ser Val Val Leu Phe Pro Tyr Trp Ala
o
5'  UUUUGAUUUGUUUUGUUCUGUUUUGGCUUAUUCGAUUUGUUUUGGACAGGAGAGCUUACGAACUCCAACG
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   7000
3'  AAAACUAAACAAAACAAGACAAAACCGAAUAAGCUAAACAAAACCUGUCCUCUCGAAUGCUUGAGGUUGC
1    Phe . Phe Val Leu Phe Cys Phe Gly Leu Phe Asp Leu Phe Trp Thr Gly Glu Leu Thr Asn Ser Asn
2     Phe Asp Leu Phe Cys Ser Val Leu Ala Tyr Ser Ile Cys Phe Gly Gln Glu Ser Leu Arg Thr Pro Thr
3   Val Leu Ile Cys Phe Val Leu Phe Trp Leu Ile Arg Phe Val Leu Asp Arg Arg Ala Tyr Glu Leu Gln Arg
o
5'  UCGGUUUCACGUUAUCCAUUCGCCUGCCCAUCCACGCUUUCCCCGCAACAUUGGGCAACAACGUGACCAC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   7070
3'  AGCCAAAGUGCAAUAGGUAAGCGGACGGGUAGGUGCGAAAGGGGCGUUGUAACCCGUUGUUGCACUGGUG
1   Val Gly Phe Thr Leu Ser Ile Arg Leu Pro Ile His Ala Phe Pro Ala Thr Leu Gly Asn Asn Val Thr Thr
2    Ser Val Ser Arg Tyr Pro Phe Ala Cys Pro Ser Thr Leu Ser Pro Gln His Trp Ala Thr Thr . Pro
3     Arg Phe His Val Ile His Ser Pro Ala His Pro Arg Phe Pro Arg Asn Ile Gly Gln Gln Arg Asp His
o
5'  CAAACUCAGCCGGCUCCGCAAGGCUCAACACCGUCUGGUGAUAUAACUUCAGCAUCACCUAUCACACCCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   7140
3'  GUUUGAGUCGGCCGAGGCGUUCCGAGUUGUGGCAGACCACUAUAUUGAAGUCGUAGUGGAUAGUGUGGGG
1     Lys Leu Ser Arg Leu Arg Lys Ala Gln His Arg Leu Val Ile . Leu Gln His His Leu Ser His Pro
2   Pro Asn Ser Ala Gly Ser Ala Arg Leu Asn Thr Val Trp . Tyr Asn Phe Ser Ile Thr Tyr His Thr Pro
3    Gln Thr Gln Pro Ala Pro Gln Gly Ser Thr Pro Ser Gly Asp Ile Thr Ser Ala Ser Pro Ile Thr Pro
o
5'  AACGCACGCGGUUUUCCUUCCAGCAUGUCUGACCUGGUUGGUGACACUUCUCACCUGACUCAGACAACCU
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   7210
3'  UUGCGUGCGCCAAAAGGAAGGUCGUACAGACUGGACCAACCACUGUGAAGAGUGGACUGAGUCUGUUGGA
1    Asn Ala Arg Gly Phe Pro Ser Ser Met Ser Asp Leu Val Gly Asp Thr Ser His Leu Thr Gln Thr Thr
2     Thr His Ala Val Phe Leu Pro Ala Cys Leu Thr Trp Leu Val Thr Leu Leu Thr . Leu Arg Gln Pro
3   Gln Arg Thr Arg Phe Ser Phe Gln His Val . Pro Gly Trp . His Phe Ser Pro Asp Ser Asp Asn Leu
o
5'  UCCCUUUUAGUGCCGUCAAGACUUUGAAGGACCCCAAGGGGCUGGUCAUUUCACAUUGACCGCCAGUCC
o   ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   7280
3'  AGGGAAAAUCACGGCAGUUCUGAAACUUCCUGGGGUUCCCCGACCAGUAAAGUGUAACUGGCGGUCAGG
1   Phe Pro Phe Ser Ala Val Lys Thr Leu Lys Gly Pro Gln Gly Ala Gly His Phe Thr Leu Thr Ala Ser Pro
2    Ser Leu Leu Val Pro Ser Arg Leu . Arg Asp Pro Lys Gly Leu Val Ile Ser His . Pro Pro Val
3     Pro Phe . Cys Arg Gln Asp Phe Glu Gly Thr Pro Arg Gly Trp Ser Phe His Ile Asp Arg Gln Ser
o
```

```
5'  AGGUUUUGAGCAGCUUAUUGCUGGUCGUGCCAAGGUUGUGAUAAGAUCCGAUUCCCUCCAGGCUCAAGUU
o                                                                            7350
3'  UCCAAAACUCGUCGAAUAACGACCAGCACGGUUCCAACACUAUUCUAGGCUAAGGGAGGUCCGAGUUCAA
1     Gly Phe Glu Gln Leu Ile Ala Gly Arg Ala Lys Val Val Ile Arg Ser Asp Ser Leu Gln Ala Gln Val
2   Gln Val Leu Ser Ser Leu Leu Leu Val Val Pro Arg Leu  .   .  Asp Pro Ile Pro Ser Arg Leu Lys Leu
3    Arg Phe  .  Ala Ala Tyr Cys Trp Ser Cys Gln Gly Cys Asp Lys Ile Arg Phe Pro Pro Gly Ser Ser
o
5'  GUUGGACCUCCCGACCCCACGAAAGCUGUCACCGUUUACGUGGCUGCGAUACCCAAUGGUUCCGUCAAGU
o                                                                            7420
3'  CAACCUGGAGGGCUGGGGUGCUUUCGACAGUGGCAAAUGCACCGACGCUAUGGGUUACCAAGGCAGUUCA
1    Val Gly Pro Pro Asp Pro Thr Lys Ala Val Thr Val Tyr Val Ala Ala Ile Pro Asn Gly Ser Val Lys
2     Leu Asp Leu Pro Thr Pro Arg Lys Leu Ser Pro Phe Thr Trp Leu Arg Tyr Pro Met Val Pro Ser Ser
3   Cys Trp Thr Ser Arg Pro His Glu Ser Cys His Arg Leu Arg Gly Cys Asp Thr Gln Trp Phe Arg Gln Val
o
5'  GGCCCACCAACGCAGCUCAGAUUCUCACGAUUGGUGGUGCUGCUGUGGUGCAGCAUUCUACUUACGUGCA
o                                                                            7490
3'  CCGGGUGGUUGCGUCGAGUCUAAGAGUGCUAACCACCACGACGACACCACGUCGUAAGAUGAAUGCACGU
1   Trp Pro Thr Asn Ala Ala Gln Ile Leu Thr Ile Gly Gly Ala Ala Val Val Gln His Ser Thr Tyr Val His
2    Gly Pro Pro Thr Gln Leu Arg Phe Ser Arg Leu Val Val Leu Leu Trp Cys Ser Ile Leu Leu Thr Cys
3     Ala His Gln Arg Ser Ser Asp Ser His Asp Trp Trp Cys Cys Cys Gly Ala Ala Phe Tyr Leu Arg Ala
o
5'  UUCUCAACCUUCACAGCUUAAGUUUGCUGUGGAAGUUGCUCACCAAAUUAAGCCAAAGCCGCAAGUAGGC
o                                                                            7560
3'  AAGAGUUGGAAGUGUCGAAUUCAAACGACACCUUCAACGAGUGGUUUAAUUCGGUUUCGGCGUUCAUCCG
1     Ser Gln Pro Ser Gln Leu Lys Phe Ala Val Glu Val Ala His Gln Ile Lys Pro Lys Pro Gln Val Gly
2   Ile Leu Asn Leu His Ser Leu Ser Leu Leu Trp Lys Leu Leu Thr Lys Leu Ser Gln Ser Arg Lys  .  Ala
3    Phe Ser Thr Phe Thr Ala  .  Val Cys Cys Gly Ser Cys Ser Pro Asn  .  Ala Lys Ala Ala Ser Arg
o
5'  ACUGAACCUGAGAUUGUUUACAGCUUCAUCGUGGCUGGGGCUGAGGCCGAAACUGAAUCCUACUUGGUCA
o                                                                            7630
3'  UGACUUGGACUCUAACAAAUGUCGAAGUAGCACCGACCCCGACUCCGGCUUUGACUUAGGAUGAACCAGU
1    Thr Glu Pro Glu Ile Val Tyr Ser Phe Ile Val Ala Gly Ala Glu Ala Glu Thr Glu Ser Tyr Leu Val
2     Leu Asn Leu Arg Leu Phe Thr Ala Ser Ser Trp Leu Gly Leu Arg Pro Lys Leu Asn Pro Thr Trp Ser
3   His  .  Thr  .  Asp Cys Leu Gln Leu His Arg Gly Trp Gly  .  Gly Arg Asn  .  Ile Leu Leu Gly His
o
5'  UCAAGGGUAUAGUGGAGGUUGAUGGUGUAGGAUUUGUGCAGACUUGGAAAUCAUGAACGCUCUGCCGCCG
o                                                                            7700
3'  AGUUCCCAUAUCACCUCCAACUACCACAUCCUAAACACGUCUGAACCUUUAGUACUUGCGAGACGGCGGC
1   Ile Lys Gly Ile Val Glu Val Asp Gly Val Gly Phe Val Gln Thr Trp Lys Ser  .  Thr Leu Cys Arg Arg
2    Ser Arg Val  .  Trp Arg Leu Met Val  .  Asp Leu Cys Arg Leu Gly Asn His Glu Arg Ser Ala Ala
3     Gln Gly Tyr Ser Gly Gly  .  Trp Cys Arg Ile Cys Ala Asp Leu Glu Ile Met Asn Ala Leu Pro Pro
o
5'  CUCCACUCUGAUUUGGGAGUUCAUUUCCUUGCGUUAGCUGCCACUUUUGUCUCUAAGAACGAGAAGACUA
o                                                                            7770
3'  GAGGUGAGACUAAACCCUCAAGUAAAGGAACGCAAUCGACGGUGAAAACAGAGAUUCUUGCUCUUCUGAU
1     Ser Thr Leu Ile Trp Glu Phe Ile Ser Leu Arg  .  Leu Pro Leu Leu Ser Leu Arg Thr Arg Arg Leu
2   Ala Pro Leu  .  Phe Gly Ser Ser Phe Pro Cys Val Ser Cys His Phe Cys Leu  .  Glu Arg Glu Asp Tyr
3    Leu His Ser Asp Leu Gly Val His Phe Leu Ala Leu Ala Ala Thr Phe Val Ser Lys Asn Glu Lys Thr
o
5'  CAACACGUCCAGUUUCUUCACUCCAAAUCCCUCCUCUCACAACAGGGGAUGCAGCAAGUUCACCUGUUGC
o                                                                            7840
3'  GUUGUGCAGGUCAAAGAAGUGAGGUUUAGGGAGGAGAGUGUUGUCCCCUACGUCGUUCAAGUGGACAACG
1    Gln His Val Gln Phe Leu His Ser Lys Ser Leu Leu Ser Gln Gln Gly Met Gln Gln Val His Leu Leu
2     Asn Thr Ser Ser Phe Phe Thr Pro Asn Pro Ser Ser His Asn Arg Gly Cys Ser Lys Phe Thr Cys Cys
3   Thr Thr Arg Pro Val Ser Ser Leu Gln Ile Pro Pro Leu Thr Thr Gly Asp Ala Ala Ser Ser Pro Val Ala
o
```

```
5'   UGUUGGAUUUUGGAUUCAACAUAAAGAGUUCUUUGACUUUUGCAUUGCUUGCCACUCUAAGUUUUCCCUA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   7910
3'   ACAACCUAAAACCUAAGUUGUAUUUCUCAAGAAACUGAAAACGUAACGAACGGUGAGAUUCAAAAGGGAU
1   Leu Leu Asp Phe Gly Phe Asn Ile Lys Ser Ser Leu Thr Phe Ala Leu Leu Ala Thr Leu Ser Phe Pro .
2    Cys Trp Ile Leu Asp Ser Thr . Arg Val Leu . Leu Leu His Cys Leu Pro Leu . Val Phe Pro
3     Val Gly Phe Trp Ile Gln His Lys Glu Phe Phe Asp Phe Cys Ile Ala Cys His Ser Lys Phe Ser Leu
o
5'   ACCGCCAAACUGAUCCCGAAGUUUCCUGUUUCAGUCAUAGGCGAUGAUUUCUUUCGUUUAGGAACCUAUC
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   7980
3'   UGGCGGUUUGACUAGGGCUUCAAAGGACAAAGUCAGUAUCCGCUACUAAAGAAAGCAAAUCCUUGGAUAG
1     Pro Pro Asn . Ser Arg Ser Phe Leu Phe Gln Ser . Ala Met Ile Ser Phe Val . Glu Pro Ile
2   Asn Arg Gln Thr Asp Pro Glu Val Ser Cys Phe Ser His Arg Arg . Phe Leu Ser Phe Arg Asn Leu Ser
3    Thr Ala Lys Leu Ile Pro Lys Phe Pro Val Ser Val Ile Gly Asp Asp Phe Phe Arg Leu Gly Thr Tyr
o
5'   GAAUGGAUUUAGACUCAAUUGAGAUCUAUACUAAAGGUCGUAAGUGGGAGACCGUUUACACCUUUGUUGA
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   8050
3'   CUUACCUAAAUCUGAGUUAACUCUAGAUAUGAUUUCCAGCAUUCACCCUCUGGCAAAUGUGGAAACAACU
1    Glu Trp Ile . Thr Gln Leu Arg Ser Ile Leu Lys Val Val Ser Gly Arg Pro Phe Thr Pro Leu Leu
2     Asn Gly Phe Arg Leu Asn . Asp Leu Tyr . Arg Ser . Val Gly Asp Arg Leu His Leu Cys .
3   Arg Met Asp Leu Asp Ser Ile Glu Ile Tyr Thr Lys Gly Arg Lys Trp Glu Thr Val Tyr Thr Phe Val Asp
o
5'   CGACGAAGAUGAGUUUGAUUUCAUGCCCGAUGGCACCUACUAUUAUUUCACAUUGGUUUGACCUGUUGGU
o    ++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|++++|   8120
3'   GCUGCUUCUACUCAAACUAAAGUACGGGCUACCGUGGAUGAUAAUAAAGUGUAACCAAACUGGACAACCA
1   Thr Thr Lys Met Ser Leu Ile Ser Cys Pro Met Ala Pro Thr Ile Ile Ser His Trp Phe Asp Leu Leu Val
2    Arg Arg Arg . Val . Phe His Ala Arg Trp His Leu Leu Leu Phe His Ile Gly Leu Thr Cys Trp
3     Asp Glu Asp Glu Phe Asp Phe Met Pro Asp Gly Thr Tyr Tyr Tyr Phe Thr Leu Val . Pro Val Gly
o
5'   CUUUU
o    ++++|                                                                    8125
3'   GAAAA
1     Phe
2   Ser Phe
3    Leu Leu
o
```

5. References

Anco, D. J., Kim, S., Mitchell, T. K., Madden, L. V., and Ellis, M. A. (2009). Transformation of *Phomopsis viticola* with the green fluorescent protein. *Mycologia* 101(6), 853–858. DOI: 10.3852/09-007.

Aoki, N., Moriyama, H., Kodama, M., Arie, T., Teraoka, T., and Fukuhara, T. (2009). A novel mycovirus associated with four double-stranded RNAs affects host fungal growth in Alternaria alternata. *Virus Res* 140:179-187. DOI: 10.1016/j.virusres.2008.12.003.

Arnesen, T., P. Van Damme, B. Polevoda, K. Helsens, R. Evjenth, N. Colaert, J. Varhaug et al (2009). Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci* 106:8157–8162 DOI: 10.1073/pnas.0901931106.

Cai, G., and Hillman, B.I. (2013). "Phytophora viruses", in Advances in Virus Research, ed. Said Ghabrial (Cambridge, MA: Academic Press), 86: 327-350. DOI: org/10.1016/B978-0-12-394315-6.00012-X

Celik, A., Baker, R., He, F., and Jacobson, A. (2017) High-resolution profiling of NMD targets in yeast reveals translational fidelity as a basis for substrate selection. *RNA* 23(5):735-748. DOI: 10.1261/rna.060541.116.

Chen X., He, H., Yang, X., Zeng, H., Qiu, D., and Guo, L. (2016). The complete sequence of a novel *Fusarium graminearum* RNA virus in a new proposed family within the order *Tymovirales*. *Arch Virol* 161:2899-2903. DOI: 10.1007/s00705-016-2961-1

Chen, Y., Shang, H.-H., Yang, H.-Q., Gao, B.-D., and Zhong, J. (2017). A mitovirus isolated from the phytopathogenic fungus *Alternaria brassicicola*. *Arch Virol* 162 2869–2874. DOI: 10.1007/s00705-017-3371-8.

Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D et al. (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol 13*(4), 414–430. DOI: 10.1111/j.1364-3703.2011.00783.x.

De Guido, M. A., Pollastro, S., Carlucci, A., De Miccolis Angelini, R. M., and Faretra, F. (2003). *Phomopsis viticola* is easily transformed with *hph* and *Bml*^{r} genes. *J Plant Pathol* 43-52. DOI: 10.4454/jpp.v85i1.1008.

Desselberger, U. (1996) Genome rearrangements of rotaviruses. *Arch Virol Suppl* 12:37-51. DOI: 10.1007/978-3-7091-6553-9_5.

Esteban, R., Wickner, R.B. (1988) A deletion mutant of L-A double-stranded RNA replicates like M1 double-stranded RNA. *J Virol* 62:1278-1285.

Eusebio-Cope, A., Sun, L., Hillman, B.I., and Suzuki, N. (2010) Mycoreovirus 1 S4-coded protein is dispensable for viral replication but necessary for efficient vertical transmission and normal symptom induction. *Virology* 397:399-408. DOI: org/10.1016/j.virol.2009.11.035.

Fujimura, T., and Esteban, R. (2011) Cap-snatching mechanism in yeast L-A double-stranded RNA virus. *Proc Natl Acad Sci USA* 108: 17667-17671. DOI: org/10.1073/pnas.1111900108

Furuichi, Y., and Miura K.-I. (1975) A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* 253: 374-375.

Furuichi, Y., Muthukrishnan, S., Tomasz, J., and Shatkin, A.J. (1976) Mechanism of formation of reovirus mRNA 5′-terminal blocked and methylated sequence, m^{7}GpppGmpC. *J Biol Chem* 251:5043-53.

Gallie D.R. (1991) The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes & Development* 5:2108-2116.

Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., and Suzuki, N. (2015). 50-plus years of fungal viruses. *Virology* 479-480: 356-368. DOI: 10.1016/j.virol.2015.02.034.

Ghabrial, S. A., and Suzuki, N. (2009). Viruses of plant pathogenic fungi. Annu. Rev. *Phytopathol*. 47, 353–384. DOI: 10.1146/annurev-phyto-080508-081932

Gilbert, K. B., Holcomb, E. E., Allscheid, R. L., and Carrington, J. C. (2019). Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. *PloS One, 14* (7), e0219207. DOI: 10.1371/journal.pone. 0219207.

Gomes, RR., Glienke, C., Videira, SI,, Lombard, L., Groenewald, JZ., and Crous, PW. (2013). *Diaporthe*: a genus of endophytic, saprobic and plant pathogenic fungi. *Persoonia*. 31:1-41. DOI: 10.3767/003158513X666844.

Guo, Y. S., Crous, P. W., Bai, Q., Fu, M., Yang, M. M., Wang, X. H et al. (2020). High diversity of *Diaporthe* species associated with pear shoot canker in China. *Persoonia* 45, 132–162. DOI: 10.3767/persoonia.2020.45.05

Hacker, C.V., Brasier, C.M., and Buck, K.W. (2005) A double-stranded RNA from a Phytophthora species is related to the plant endornaviruses and contains a putative UDP glycosyltransferase gene. *J Gen Virol* 86:1561-1570. DOI: org/10.1099/vir.0.80808-0.

Hamid, MR., Xie, J., Wu, S., Maria, SK., Zheng, D., Assane Hamidou, A., Wang, Q et al. (2018). A Novel Deltaflexivirus that Infects the Plant Fungal Pathogen, Sclerotinia sclerotiorum, Can Be Transmitted Among Host Vegetative Incompatible Strains. *Viruses* 10(6):295. DOI: 10.3390/v10060295.

Hammond, T.M., Andrewski, M.D., Roossinck, M.J., and Keller, N.P. (2008). *Aspergillus* Mycoviruses Are Targets and Suppressors of RNA Silencing. *Eukaryot Cell* 7:350-357. DOI: 10.1128/ec.00356-07.

He, H., Chen, X., Li, P., Qiu, D., and Guo, L. (2018) Complete Genome Sequence of a

Fusarium graminearum Double-Stranded RNA Virus in a Newly Proposed Family, Alternaviridae. *Genome Announc* 6:e00064-18. DOI: 10.1128/genomeA.00064-18.

Higashiura, T., Katoh, Y., Urayama, S.-i., Hayashi, O., Aihara, M., Fukuhara, T et al (2019) Magnaporthe oryzae chrysovirus 1 strain D confers growth inhibition to the host fungus and exhibits multiform viral structural proteins. *Virology* 535:241-254. DOI: 10.1016/j.virol.2019.07.014.

Hilário, S., Santos, L., and Alves, A. (2021). *Diaporthe amygdali*, a species complex or a complex species? *Fungal biol* 125(7), 505–518. DOI: 10.1016/j.funbio. 2021.01.006

Hillman, B. I., Foglia, R., and Yuan, W. (2000). Satellite and defective RNAs of Cryphonectria hypovirus 3-grand haven 2, a virus species in the family Hypoviridae with a single open reading frame. *Virology* 276(1), 181-189. DOI: 10.1006/viro.2000.0548.

Hillman, B. I., Supyani, S., Kondo, H., and Suzuki, N. (2004). A reovirus of the fungus Cryphonectria parasitica that is infectious as particles and related to the coltivirus genus of animal pathogens. *J Virol* 78 (2), 892-898. DOI:10.1128/jvi.78.2.892-898.2004

Hocine, S., Singer, R.H., Grünwald, D. (2010) RNA processing and export. *Cold Spring Harbor perspectives in biology* 2:a000752-a000752. DOI: 10.1101/cshperspect. a000752.

Hollings, M. (1962) Viruses Associated with A Die-Back Disease of Cultivated Mushroom. *Nature* 196:962-965. DOI: 10.1038/196962a0.

Howitt, R. L., Beever, R. E., Pearson, M. N., and Forster, R. L. (1995). Presence of double-stranded RNA and virus-like particles in *Botrytis cinerea*. *Mycol Res*

99(12), 1472-1478. DOI: 10.1016/S0953-7562(09)80795-8.

Inoue, K., Ohno, M., Sakamoto, H., Shimura, Y. (1989) Effect of the cap structure on pre-mRNA splicing in Xenopus oocyte nuclei. *Genes* Dev 3:1472-9. DOI: 10.1101/gad.3.9.1472.

Jia, H., Dong, K., Zhou, L., Wang, G., Hong, N., Jiang, D., et al. (2017). A dsRNA virus with filamentous viral particles. *Nat Commun* 8(1), 1-12. DOI: 10.1038/s41467-017-00237-9.

Johnson, L. J., Johnson, R. D., Akamatsu, H., Salamiah, A., Otani, H., Kohmoto, K., and Kodama, M. (2001). Spontaneous loss of a conditionally dispensable chromosome from the Alternaria alternata apple pathotype leads to loss of toxin production and pathogenicity. *Curr Gnet*, 40(1), 65–72. DOI: 10.1007/s002940100233

Kamer, G., and Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res 12*(18), 7269–7282. DOI: 10.1093/nar/12.18.7269

Kanematsu, S., Shimizu, T., Salaipeth, L., Yaegashi, H., Sasaki, A., Ito, T et al (2014) Genome rearrangement of a mycovirus Rosellinia necatrix megabirnavirus 1 affecting its ability to attenuate virulence of the host fungus. *Virology* 450-451:308-15. DOI: 10.1016/j.virol.2013.12.002.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4), 772–780. DOI: 10.1093/molbev/mst010

Kim, J. M., Jung, J. E., Park, J. A., Park, S. M., Cha, B. J., and Kim, D. H. (2013). Biological function of a novel chrysovirus, CnV1-BS122, in the Korean Cryphonectria nitschkei BS122 strain. *J Biosci Bioeng* 115(1), 1-3. DOI:10.1016/j.jbiosc.2012.08.007.

Komatsu K., Katayama Y., Omatsu T., Mizutani T., Fukuhara T., Kodama M., et al. (2016). Genome sequence of a novel victorivirus identified in the phytopathogenic fungus *Alternaria arborescens*. *Arch Virol* 161 1701–1704. DOI: 10.1007/s00705-016-2796-9

Kondo, H., Kanematsu, S., and Suzuki, N. (2013), "Viruses of the white root rot fungus, Rosellinia necatrix", in Advances in Virus Research, ed. Said Ghabrial (Cambridge, MA: Academic Press), 86: 177-214. DOI: 10.1016/B978-0-12-394315-6.00007-6.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72(9), 2197-2206. DOI: 10.1099/0022-1317-72-9-2197

Koyama, S., Urayama, S., Ohmatsu, T., Sassa, Y., Sakai, C., Takata, M., Hayashi, S., Nagai, M., Furuya, T., Moriyama, H., Satoh, T., Ono, S., and Mizutani, T. (2015). Identification, characterization and full-length sequence analysis of a novel dsRNA virus isolated from the arboreal ant Camponotus yamaokai. The Journal of general virology, 96(Pt 7), 1930–1937. DOI: 10.1099/vir.0.000126

Kozlakidis, Z., Herrero, N., Ozkan, S., Bhatti, M.F., and Coutts, R.H. (2013). A novel dsRNA element isolated from the Aspergillus foetidus mycovirus complex. *Arch Virol* 158:2625-2628. DOI: 10.1073/pnas.1419225112

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*, 35(6), 1547–1549. DOI: 10.1093/molbev/msy096

Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., and Smith, D.B. (2018). Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res* 46: 708-717. DOI:

10.1093/nar/gkx932.

Li, K., Zheng, D., Cheng, J., Chen, T., Fu, Y., Jiang, D., and Xie, J. (2016). Characterization of a novel *Sclerotinia sclerotiorum* RNA virus as the prototype of a new proposed family within the order *Tymovirales*. *Virus Res* 219: 92-99. DOI: 10.1016/j.virusres.2015.11.019

Li, L, Tian, Q, Du, Z, Duns, G. J., and Chen, J. (2009). A novel double-stranded RNA virus detected in *Primula malacoides* is a plant-isolated partitivirus closely related to partitivirus infecting fungal species. *Arch Virol* 154: 565-572. DOI: 10.1007/s00705-009-0342-8

Li, P., Zhang, H., Chen, X., Qiu, D., and Guo, L. (2015). Molecular characterization of a novel hypovirus from the plant pathogenic fungus Fusarium graminearum. *Virology* 481:151-160. DOI: 10.1016/j.virol.2015.02.047.

Lindqvist, I., Imataka, H., and Pelletier, J. (2008). Cap-dependent eukaryotic initiation factor-mRNA interactions probed by cross-linking. *RNA* 14: 960–969. DOI/10.1261/rna.971208.

Liu, S., Xie, J., Cheng, J., Li, B., Chen, T., Fu, Y et al. (2016) Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. *Proc Natl Acad Sci USA* 113:12803-12808. DOI: 10.1073/pnas.1608013113.

Lutz, T., Japić, E., Bien, S., Langer, G. J., and Heinze, C. (2022). Characterization of a novel Alternavirus infecting the fungal pathogen Fusarium solani. *Virus Res* DOI: 10.1016/j.virusres.2022.198817

McCabe, P.M., and Van Alfen, N.K. (2002). “Molecular basis of symptom expression by the Cryphonectria hypovirus”, in DsRNA Genetic Elements*,* ed. Tanvantzis S.M (Boca Raton, FL: CRC Press), 125-143.

Maderazo, A.B., Belk J.P., He, F., and Jacobson, A. (2003). Nonsense-containing mRNAs that accumulate in the absence of a functional nonsense-mediated mRNA decay pathway are destabilized rapidly upon its restitution. *Mol Cell Biol* 23(3):842-51. DOI: 10.1128/mcb.23.3.842-851.2003.

Masison, D.C., Blanc, A., Ribas, J.C., Carroll, K., Sonenberg, N., and Wickner, R. (1995). Decoying the cap- mRNA degradation system by a double-stranded RNA virus and poly(A)- mRNA surveillance by a yeast antiviral system. *Mol Cell Biol* 15: 2763–2771. DOI: 10.1128/mcb.15.5.2763.

Marzano, S-Y.L. and Domier, L.L. (2016). Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Res* 213: 332-342. DOI: 10.1016/j.virusres.2015.11.002

Moleleki, N., van Heerden, S. W., Wingfield, M. J., Wingfield, B. D., and Preisig, O. (2003). Transfection of Diaporthe perjuncta with Diaporthe RNA virus. *Appl Environ Microbiol* 69(7), 3952-3956. DOI: 10.1128/AEM.69.7.3952-3956.2003.

Moleleki, N., Wingfield, M. J., Wingfield, B. D., and Preisig, O. (2011). Effect of Diaporthe RNA virus 1 (DRV1) on growth and pathogenicity of different Diaporthe species. *Eur J Plant Pathol* 131, 261–268. DOI: 10.1007/s10658-011-9804-4.

Moriyama, H., Aoki, N., Fuke, K., Urayama-Takeshita, K., Takeshita, N., and Wu, Chien-Fu. (2021). "Alternaviruses (Unassigned)", in Encyclopedia of Virology, 4th Edition, Vol. 4: 544-548. DOI:10.1016/B978-0-12-814515-9.00031-X.

Mu, F., Li, B., Cheng, S., Jia, J., Jiang, D., Fu, Y., Cheng, J., Lin, Y., Chen, T., Xie, J. (2021). Nine viruses from eight lineages exhibiting new evolutionary modes that co-infect a hypovirulent phytopathogenic fungus. *PLoS Pathog* 24;17(8):e1009823. DOI: 10.1371/journal.ppat.1009823.

Nerva, L., Turina, M., Zanzotto, A., Gardiman, M., Gaiotti, F., Gambino, G., and Chitarra,

W. (2019) Isolation, molecular characterization and virome analysis of culturable wood fungal endophytes in esca symptomatic and asymptomatic grapevine plants. *Environ Microbiol* 21:2886-2904. DOI: 10.1111/1462-2920.14651.

Nibert, M.L., Ghabrial, S.A., Maiss, E., Lesker, T., Vainio, E.J., Daohong, J., and Suzuki, N. (2014) Taxonomic reorganization of family *Partitiviridae* and other recent progress in partitivirus research. *Virus Res* 188: 128-141. DOI: 10.1016/j.virusres.2014.04.007

Nuss, D.L. (2005) Hypovirulence: Mycoviruses at the fungal–plant interface. *Nat Rev Microbiol* 3:632-642. DOI: 10.1038/nrmicro1206.

Okada, R., Kiyota, E., Moriyama, H., Fukuhara, T., and Natsuaki, T. (2015). A simple and rapid method to purify viral dsRNA from plant and fungal tissue. *J Gen Plant Pathol* 81(2), 103-107. DOI: 10.1007/s10327-014-0575-6.

Okada, R., Ichinose, S., Takeshita, K., Urayama, S.I., Fukuhara, T., Komatsu, K et al. (2018) Molecular characterization of a novel mycovirus in Alternaria alternata manifesting two-sided effects: Down-regulation of host growth and up-regulation of host plant pathogenicity. *Virology* 519:23-32. DOI: 10.1016/j.virol.2018.03.027.

Osaki, H., Sasaki, A., Nomiyama, K., and Tomioka, K. (2016). Multiple virus infection in a single strain of Fusarium poae shown by deep sequencing. *Virus Genes* 52:835-847. DOI: 10.1007/s11262-016-1379-x.

Ozawa, K., Mochizuki, K., Takagi, D., Ishida, K., Sunada, A., Ohkusu, K et al. (2019). Identification and antifungal sensitivity of two new species of Diaporthe isolated. *J Infect Chemother* 25(2), 96–103. DOI: 10.1016/j.jiac.2018.10.008

Peltz, S.W., Brown, A.H., and Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes & Development* 7:1737-1754. ISSN

0890-9369/93.

Plotch, S.J., Bouloy, M., and Krug, R.M. (1979). Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. *Proc Natl Acad Sci USA* 76: 1618-1622.

Powilleit, F., Breinig, T., and Schmitt, M.J. (2007). Exploiting the yeast L-A viral capsid for the in vivo assembly of chimeric VLPs as platform in vaccine development and foreign protein expression. *PloS One* 2:e415-e415. DOI: 10.1371/journal.pone.0000415.

Preisig, O., Moleleki, N., Smit, WA., Wingfield, BD., and Wingfield, MJ. (2000). A novel RNA mycovirus in a hypovirulent isolate of the plant pathogen *Diaporthe ambigua. J Gen Virol* 81(12):3107-3114. DOI: 10.1099/0022-1317-81-12-3107.

Santos, L., Alves, A., and Alves, R. (2017). Evaluating multi-locus phylogenies for species boundaries determination in the genus *Diaporthe*. *PeerJ* 5:e3120 DOI: 10.7717/peerj.3120

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, 9(7), 671–675. DOI:10.1038/nmeth.2089

Shamsi, W., Sato, Y., Jamal, A., Shahi, S., Kondo, H., Suzuki, N et al. (2019). Molecular and biological characterization of a novel botybirnavirus identified from a Pakistani isolate of *Alternaria alternata*. *Virus Res* DOI: 10.1016/ j.virusres.2019.01.006.

Shang, H.-H., Zhong, J., Zhang, R.-J., Chen, C.-Y., Gao, B.-D-, and Zhu, H.-J. (2015).Genome sequence of a novel endornavirus from the phytopathogenic fungus *Alternaria brassicicola. Arch Virol* 160(7):1827-30. DOI: 10.1007/s00705-015-2426-y.

Shwed, P. S., Dobos, P., Cameron, L. A., Vakharia, V. N., and Duncan, R. (2002).

Birnavirus VP1 proteins form a distinct subgroup of RNA-dependent RNA polymerases lacking a GDD motif. *Virology* 296(2), 241–250. DOI: 10.1006/viro.2001.1334

Simon, A.E. (2015) 3' UTR of Carmoviruses. *Virus Res* 206: 27-36. DOI: 10.1016/j.virusres.2015.01.023

Sun, L., and Suzuki, N. (2008). Intragenic rearrangements of a mycoreovirus induced by the multifunctional protein p29 encoded by the prototypic hypovirus CHV1-EP713. *RNA* 14:2557-2571. DOI: 10.1261/rna.1125408.

Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S et al. (2000). Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol* 31(1), 21–32. DOI: 10.1006/fgbi.2000.1228

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56(4), 564–577. DOI:10.1080/10635150701472164

Tavantzis, S.M., Lakshman, D.K., and Liu, C. (2002). “Double-stranded RNA elements modulating virulence in Rizoctonia solani”, in DsRNA Genetic Elements, ed. Tanvantzis S.M (Boca Raton, FL: CRC Press), 191-212.

Tercero, J. C., Dinman, J. D., and Wickner, R. B. (1993). Yeast MAK3 N-acetyltransferase recognizes the N-terminal four amino acids of the major coat protein (gag) of the L-A double-stranded RNA virus. *J Bacteriol* 175(10), 3192–3194. DOI: 10.1128/jb.175.10.3192-3194.1993

Tercero, J. C., and Wickner, R. B. (1992). MAK3 encodes an N-acetyltransferase whose modification of the L-A gag NH2 terminus is necessary for virus particle assembly. *J Biol Chem* 267(28), 20277–20281. DOI: 10.1016/S0021-9258(19)88697-0

Toh-E, A., Guerry, P., and Wickner, R.B. (1978) Chromosomal superkiller mutants of

Saccharomyces cerevisiae. *J Bacteriol* 136: 1002 – 1007

Uchida, K., Sakuta, K., Ito, A., Takahashi, Y., Katayama, Y., Omatsu, T et al. (2021). Two novel endornaviruses co-infecting a Phytophthora pathogen of Asparagus officinalis modulate the developmental stages and fungicide sensitivities of the host oomycete. *Front Microbiol* DOI: /10.3389/fmic b.2021.633502

Wei, C.M., and Moss, B. (1975). Methylated Nucleotides Block 5'-Terminus of Vaccinia Virus Messenger RNA. *Proc Natl Acad Sci USA* 72: 318-322.

Wickner, R.B., Fujimura, T., Esteban, R. (2013). Viruses and prions of Saccharomyces cerevisiae. *Adv Virus Res* 86:1-36. DOI: 10.1016/b978-0-12-394315-6.00001-5.

Widner, W.R., and Wickner, R.B. (1993). Evidence that the *SKI* antiviral system of *Saccharomyces cerevisiae* acts by blocking expression of viral mRNA. *Mol Cell Biol.* 13: 4331 – 4341.

Woolford, J. L., Jr, and Baserga, S. J. (2013). Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics* 195(3), 643–681. DOI: 10.1534/ genetics.113.153197

Wu, C.-F., Aoki, N., Takeshita, N., Fukuhara, T., Chiura, X. H., Arie, T., Kotta-Loizou, I., Okada, R., Komatsu, K., and Moriyama, H. (2021). Unique terminal regions and specific deletions of the segmented double-stranded RNA genome of Alternaria alternata virus 1, in the proposed family Alternaviridae. *Front Microbiol -Virology* DOI: 10.3389/fmicb.2021.773062.

Xavier A. D., de Barros A. P. O., Godinho M. T., Zerbini F. M., Souza F. D., Bruckner F. P., et al. (2018). A novel mycovirus associated to Alternaria alternata comprises a distinct lineage in *Partitiviridae*. *Virus Res* DOI: 10.1016/j.virusres. 2017.10.007.

Xiang, J., Fu, M., Hong, N., Zhai, L.-F., Xiao, F., and Wang, G.-P. (2017). Characterization of a novel botybirnavirus isolated from a phytopathogenic *Alternaria fungus*. *Arch Virol* 162 3907–3911. DOI: 10.1007/s00705-017-3543-6.

Xie, J., Xiao, X., Fu, Y., Liu, H., Cheng, J., Ghabrial, S. A et al. (2011). A novel mycovirus closely related to hypoviruses that infects the plant pathogenic fungus Sclerotinia sclerotiorum. *Virology* 418, 49–56. DOI: 10.1016/j.virol.2011.07.008.

You, J., Zhou, K., Liu, X., Wu, M., Yang, L., Zhang, J., et al. (2019). Defective RNA of a novel mycovirus with high transmissibility detrimental to biocontrol properties of Trichoderma spp. *Microorganisms* 7(11), 507. DOI: 10.3390/microorganisms7110507.

Yu, X., Li, B., Fu, Y., Jiang, D., Ghabrial, S.A., Li, G et al. (2010). A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc Natl Acad Sci USA* 107:8387-92. DOI: 10.1073/pnas.0913535107.

Zhang, E., Khanna, V., Dacheux, E., Namane, A., Doyen, A., Gomard, M. et al. (2019). A specialised *SKI* complex assists the cytoplasmic RNA exosome in the absence of direct association with ribosomes. *EMBO J* 38(14):e100640. DOI: 10.15252/embj.2018100640.

Zhang, R., Liu, S., Chiba, S., Kondo, H., Kanematsu, S., and Suzuki, N. (2014). A novel single-stranded RNA virus isolated from a phytopathogenic filamentous fungus, Rosellinia necatrix, with similarity to hypo-like viruses. *Front Microbiol* DOI: 10.3389/fmicb.2014.00360.

Zhang, J, Roberts, R., and Rakotondrafara, A.M. (2015). The role of the 5' untranslated regions of Potyviridae in translation. *Virus Res* 206: 74-81. DOI: 10.1016/j.virusres.2015.02.005

Zhang, X., Xie, Y., Zhang, F., Sun, H., Zhai, Y., Zhang, S et al. (2019). Complete genome sequence of an alternavirus from the phytopathogenic fungus Fusarium incarnatum. *Arch Virol* 164:923-925. DOI: 10.1007/s00705-018-04128-2.

Zhang, Y., Shi, N., Wang, P., Zhu, Q., Yang, G., and Huang, B. (2022). Molecular characterization of a novel alternavirus infecting the entomopathogenic fungus *Cordyceps chanhua*. *Arch Virol* 167, 1467–1470. DOI: 10.1007/s00705-022-05446-2

Aoki, N. 2009. A novel mycovirus, *Alternaria alternata* virus 1, affects host fungal growth in *Alternaria alternata.* Tokyo University of Agriculture and Technology, United Graduated School of Agriculture Science, Doctoral thesis.

Hamada, K. 2015. Particle purification of Alternaria alternata mycovirus AaV1, and establishment of virus-free strain by mycelial cell fragment method. Tokyo University of Agriculture and Technology, Faculty of Agriculture, Master thesis.

Takeshita, K. 2011. Sstructural and functional analysis of two mycovirues Alternaria alternate virus 1 (AaV1) and Alternaria alternate virus-N18, which are attenuing host fungus. Tokyo University of Agriculture and Technology, Faculty of Agriculture, Master thesis.

6. Acknowledgement

I thank my supervisor, Prof. Hiromitsu Moriyama for giving me much support not on my research but my life in Japan during the pandemic. I thank Prof. Toshiyuki Fukuhara (Tokyo University of Agriculture and Technology), Dr. Ioly Kotta-Loizou (Imperial College London), Dr. Ryo Okada (Ibaraki Horticultural Research Center), and the committee members for giving me kind suggestions and help on my studies. I also thank Mr. Kohei Sakuta, Mr. Yuho Hayashi, Mr. Kentaro Kitaura, Mr. Yohei Maruyama, and the other lab members for the discussion and sharing life. Finally, I thank my family for supporting me to finish my doctoral course.

7. Publications and Presentations

➢ Article

Wu, C.-F., Aoki, N., Takeshita, N., Fukuhara, T., Chiura, X. H., Arie, T., Kotta-Loizou, I., Okada, R., Komatsu, K., and Moriyama, H. (2021) Unique terminal regions and specific deletions of the segmented double-stranded RNA genome of Alternaria alternata virus 1, in the proposed family Alternaviridae. *Front. Microbiol. -Virology.* DOI: 10.3389/fmicb.2021.773062

Wu, C.-F., Okada, R., Ogawara, T., Ou, J.-H., Nakamura, K., Kitaura, K., Moriyama, H. A Novel Alternavirus, Diaporthe Alternavirus 1, Furnishes 5′ Cap and 3′ Poly (A: U) structures in its Double-stranded RNA genomes, Latently Infects in *Diaporthe* sp. (Preparing)

Wu, C.-F., Okada, R., Ogawara, T., Kitaura, K., Moriyama, H. Identification of a novel deltaflexivirus infecting in *Fusarium oxysporum* f. sp. *melonis.* (Preparing)

➢ Chapter

Moriyama, H., Aoki, N., Fuke, K., Urayama, K. T., Takeshita, N. & **Wu, C.-F.**, (2021) Alternaviruses. Encyclopedia of Virology (Fourth Edition), Academic Press. DOI: 10.1016/B978-0-12-814515-9.00031-X

➢ Presentation

岡田亮・北浦健太朗・**呉建甫**・宮本拓也・林可奈子・小河原孝司・森山裕充．複数の分化型の *Fusarium oxysporum* から見つかった新規デルタフレキシウイルスの塩基配列の解析．日本植物病理学会大会．2020年3月19日～21日．

Wu, C.-F., Fukuhara, T., Okada, R. & Moriyama, H. 5′ terminal structure identification and specific genome alterations of segmented dsRNA mycovirus, Altenaria alternata virus 1 (AaV1), causing the morphologies changing in saprophyte A.

alternata. Annual Meeting of the Phytopathological Society of Japan. Online meeting. March 17th ~19th, 2021.

Wu, C.-F., Nakamura, K., Fukuhara, T., Ogawara, T., & Moriyama, H. "A new alternavirus, Diaporthe alternavirus 1 (DAV1), isolated from Diaporthe aff. acuta of strawberry". Kanto meeting of the Phytopathological Society of Japan. Online meeting. September 21st ~22nd, 2021.

Wu, C.-F., Ito, Y., Sasaki, N., Ogawara, T., Moriyama, H. & Okada, R. "Analysis of the viral particle-associated proteins of a novel mycovirus, Diaporthe alternavirus 1 (DAV1)". Annual Meeting of the Phytopathological Society of Japan. Online meeting. 2022.

Wu, C.-F., Fukuhara, T., Kotta-Loizou, I., Okada, R., and Moriyama, H. The unique molecular characteristics of Alternaria alternata virus 1 (AaV1): 5' m^{7}G-cap / 3' Poly (A:U) structure, specific defective dsRNA genomes. 5th International Mycovirus Symposium. Garganon, Italy. May 30th ~June 1st , 2022.

➢ Poster

Wu, C.-F., Fukuhara, T., Okada, R. & Moriyama, H., "Specific genome alterations of segmented dsRNA mycovirus, Alternaria alternata virus 1 (AaV1), causing the morphologies changing in saprophyte *A. alternata*", The Phytopathological Society of Japan (PSJ), Kanto Online Meeting, Japan, September 28th~29th, 2020.

Chien-Fu Wu, Ryo Okada, Toshiyuki Fukuhara and Hiromitsu Moriyama, "Identification of 5' terminal structure and specific genome alterations of segmented dsRNA mycovirus, Alternaria alternata virus 1 (AaV1), causing the morphologies changing in saprophyte A. alternata", The American Society of Virology (ASV), 40th annual meeting, July 19th -23rd, 2021.