Virus Detection in a Wild Plant, *Arctium tomentosum*, by siRNA Deep-Sequencing

Yaqi Bi
Master's thesis
University of Helsinki
Department of Agricultural Sciences
Plant production science/ Plant pathology
December 2011

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta/Osasto — Fakultet/Sektion — Faculty	Laitos — Institution — Department
Faculty of Agriculture and Forestry	Department of Agricultural Sciences

Tekijä — Författare — Author

Yaqi Bi

Työn nimi — Arbetets titel — Title

Viruses Detection in a Wild Plant, *Arctium tomentosum*, by siRNA Deep-Sequencing

Oppiaine —Läroämne — Subject

Plant Pathology (Plant Virology)

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	December, 2011	82

Tiivistelmä — Referat — Abstract

Plant is able to recognize dsRNA, and cleave the dsRNA into siRNA in the cell. This mechanism helps plant to against virus. A novel method of virus detection based on siRNA deep-sequencing has been developed. The method does not require any prior supposition, and it provides an unbiased view for detecting of all viruses. Thus it was used to detect viruses from a wild plant (*Arctium tomentosum*) with viral symptoms in Helsinki, Finland.

Overlapping siRNA reads were used to build contigs using the program Velvet. Programs MAQ and Novoalign were used to align the siRNA reads to a reference sequence for the viral sequence recovery. In this study, two viruses, *Alstroemeria virus X* (AlsVX, genus *Potexvirus*, family Alphaflexiviridae) and *Fig mosaic virus*-Hel (FMV, unassigned genus *Emaravirus*) were identified. This is the first report for the occurrences of both viruses in Finland. The siRNA deep-sequencing detection results were confirmed by RT-PCR. The distributions of the viruses in Helsinki were also studied.

Partial sequences of AlsVX-Hel and FMV-Hel were compared with related viruses in NCBI. The amino acid identity of the coat protein gene between AlsVX-Hel detected from Helsinki and AlsVX-Jap from Japan is 90%, and the amino acid identities of the putative nucleocapsid protein gene between FMV-Hel and other FMV strains were about 78%. The differences indicate that the AlsVX-Hel in Helsinki might be a new strain of AlsVX, and FMV-Hel might be a new strain of FMV, or a new virus.

Avainsanat — Nyckelord — Keywords

Wild plant, siRNA deep-sequencing, virus detection, *Fig mosaic virus, Alstroemeria virus X, Arctium tomentosum.*

Säilytyspaikka — Förvaringsställe — Where deposited

Department of Agricultural Sciences & Viikki Science Library

Muita tietoja — Övriga uppgifter — Further information

Supervised by: Prof. Jari Valkonen

TABLE OF CONTENTS

1 INTRODUCTION

Research on plant diseases has been carried out for decades all over the world. In order to ensure good yields and farmers' profits, farmers and scientists try their best to control plant diseases which cause yield losses. Plant viruses are among the most significant pathogens that threaten plant growth and production. For instance, *Cotton leaf curl virus* (genus: *Begomovirus*) caused a loss estimated at 5 billion US$ in Pakistan between 1992 to 1997 (Briddon & Markham, 2000). *Sweetpotato chlorotic stunt virus* (genus *Crinivirus*) reduces yields by approximately 30 to 80% in sweatpotato (Gibson & Aritua, 2002). *Tomato yellow leaf curl virus* (genus: *Begomovirus*) causes diseases in tomatoes in USA, Greece, Morocco and other countries; and in many tropical and subtropical regions, the losses are up to 100% (Agrama & Scott, 2006).

Viruses which exist in non-cultivated plant communities are generally considered not to be harmful, however, the relationships between viruses infecting wild plants and cultivated crops may be underestimated. Wild plants may be a source of agriculture-related viral pathogens. Viruses are transmitted by aphids and other vector organisms, which can bring viruses to around fields, and pose threats to agricultural production. Furthermore, in wild plants, viruses retain large biodiversities and abundance, which suggests their importance in ecosystems.

1.1 Plant viruses

The earliest written record about an almost certain plant viral disease was mentioned in a Japanese poem in 752 AD. Paintings and drawings in 17th century from the Netherlands demonstrated flower viral disease symptoms in tulips, known as the leading "tulipomania" (Hull, 2002). In 1898, Martinus Willem Beijerinck separated a non-bacteria infective agent, which was considered as the birth of virology (Zaitlin, 1998; Bos, 1995). In 1955, Fraenkel-Conrat and Williams isolated naked RNA and protective coat protein of *Tobacco mosaic virus* (TMV) (Fraenkel-Conrat & Williams, 1955). Then they demonstrated that only the RNA molecule is sufficient for viral infectivity, and RNA carries information for coat protein synthesis (Fraenkel-Conrat, 1956). The discovery initiated modern virology. Till 2009, 2284 species of viruses have been discovered (ICTV: http://ictvonline.org/virusTaxInfo.asp), and almost half of them can attack plants (Agrios, 2005).

With the development of virology, taxonomy of viruses has changed. Till 2009, 87 viral families were classified by ICTV (http://www.ictvonline.org/virusTaxonomy.asp?version=2009), and viruses from at least 31 families can infect plants (http://www.dpvweb.net/notes/families.php). The criteria used in viral classification include seven characteristics: a) structure of the viral particle; b) physicochemical properties of viral particle; c) properties of viral nucleic acids; d) viral proteins; e) serological relationships; f) activities in the plant and g) means of transmission (Hull, 2002).

The typical plant viruses consist of a genomic nucleic acid protected by a protein coat. The genome can consist of one or more nucleic acids, which could be either DNA or RNA, and could be either single-stranded or double-stranded. Plant viral genome(s) encode 1 to 12 proteins, which have functions in replication, movement, transmission, structure and infectivity (Hull, 2002). Viruses are obligate parasites, which means that they cannot survive independently out of a living host. The multiplication of plant viruses consist of five main steps. 1) The virus enters the host cell. Most plant viruses enter the host through wounds, or are transmitted by vectors that feed on the plant. 2) Uncoating and release of the viral genomic nucleic acid. 3) Replication of viral genome and synthesis of viral protein. 4) Self-assembly of viral particles and 5) spread of viruses to neighbouring cells through plasmodesmata, and transport to other parts of the host through vascular tissues.

1.1.1 Viral genomes and replication

Viruses with different nucleic acid types differ in the viral genome replication process. Most plant viruses have a positive single-stranded RNA (+ssRNA) genome. After protein uncoating, the viral RNA is utilized as template, for synthesis of a complementary RNA strand by the viral RNA polymerase. The complementary RNA, which is in negative sense, is dissociated from the original positive RNA, and serves as the template for viral RNA synthesis (Figure 1). Some viruses with negative single-stranded RNA (-ssRNA) have to be first transcribed into positive-strand RNA by virus-carried transcriptase, and then replicate as described above. For some of the plant viruses with double-stranded RNA (dsRNA), genome replication occurs in a subviral particle, which is called core. Core includes the viral genome and RNA dependent RNA polymerase (RdRp) molecules. The genomic nucleic acids are transcribed with the help

of RdRp, and then the positive-strand RNA is exported to cytoplasm where it acts similar to mRNA. The viral mRNA is translated to a protein and assembles with the new synthesized proteins to become an immature virion. Inside this virion, mRNA is transcribed to produce a negative-strand, and a new virion is formed (Ball, 2007).

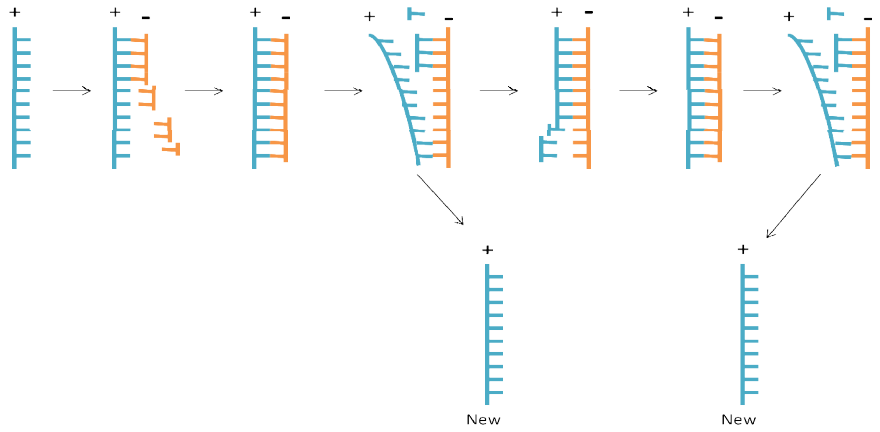


Figure 1. Schematic representation of +ssRNA viral replication.

Some plant viruses have single-stranded DNA (ssDNA); the replication mechanism of these viruses relies on using cellular DNA replication proteins. ssDNA first has to be converted into double-stranded DNA, and forming a circle and producing a negative-strand used as template to multiple positive-strands which are cleaved to the genome unit (Gutierrez, 1999) (Figure 2).

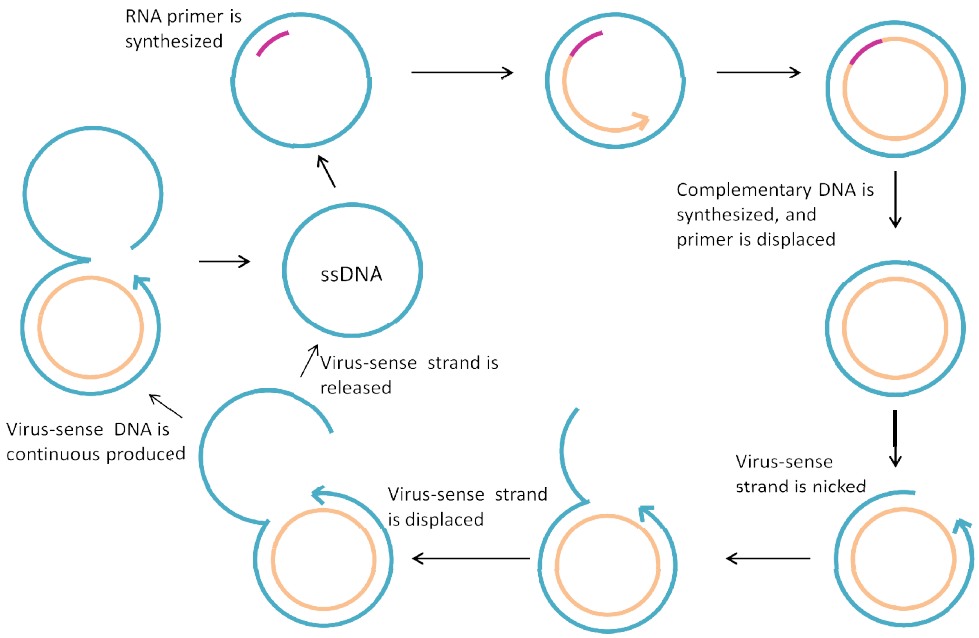


Figure 2. Schematic representation of replication of ssDNA virus.

For a few viruses with double-stranded DNA (dsDNA), the viral genome enters host nucleus and forms a mini-chromosome by twisting and supercoiling. The mini-chromosome is transcribed into two ssRNA: the smaller one is used for protein translation, and the bigger one is encapsulated by coat protein and is served as the template for reverse transcription to produce a dsDNA virus (Agrios, 2005).

1.1.2 Virus movement

Viruses complete the replication process in the initially infected host cell, and then spread to other cells. Viral movement between host cells relies on specific movement proteins. Viral movement is divided to two phases: cell-to-cell movement and long distance transportation. Cell-to-cell movement is operated by movement protein(s) (MP). There are a range of types of MP encoded by the viral genome. To complete a movement, one or several MPs are required to mediate intracellular movement and intercellular viral spread via plasmodesmata depending on virus species (Lucas, 2006).

By cell-to-cell movement, viruses spread to the vascular tissue in which they are carried together with photosynthates, and move rapidly through the host plant systemically. Viral long distance transport in vascular tissue mostly takes place in the sieve tube system of phloem (Gilbertson & Lucas, 1996). Some beetle-transmitted viruses can be transported through xylem (Fulton *et al.*, 1987). Viruses require several days to spread toward growing regions or other energy-utilizing parts since the infection started (Figure 3).

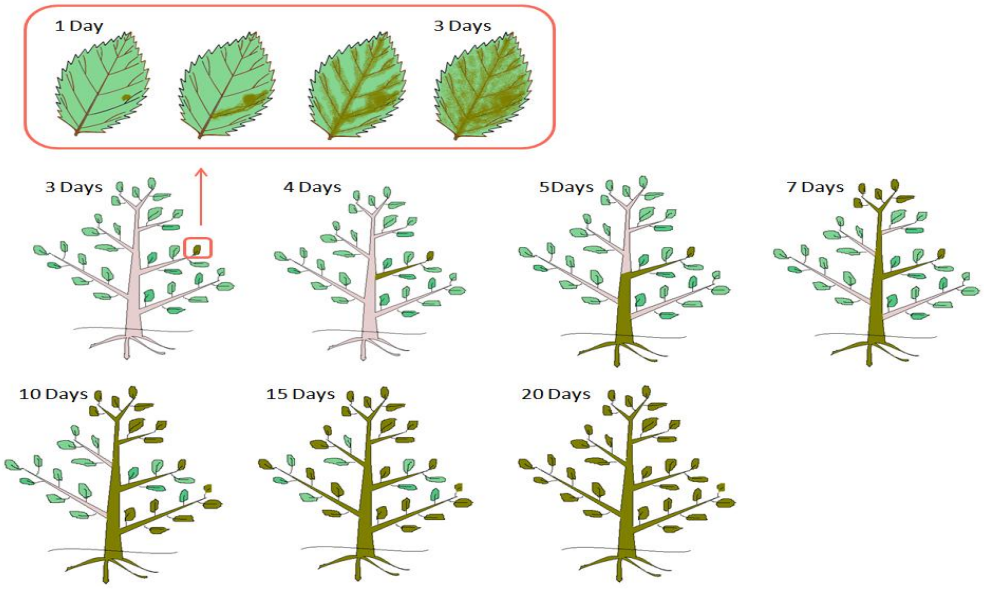


Figure 3. Schematic representation of the direction and rate of translocation of a virus in a plant.

1.1.3 Virus transmission

Viruses have no movable reproductive structure, such as spores in fungi. Therefore viral transmission between host plants depends on contact between an infected plant and a healthy plant, or transmission by vectors. Contact infection includes vegetative propagation and mechanical transmission in sap, which is commonly used for experimental inoculation in laboratory. The other way for viral transmission between host individuals is with the help of other organisms, called vectors. Viral vectors include insects, mites, nematodes and fungus-like organisms. Usually, a vector has an ability to transmit only a few types of viruses, and a virus can be transmitted by only a few kinds of vectors. Some viruses exist in plant pollen and seeds, and appear in the next generation of plant. The mechanisms of these cross-generation transmissions are called pollen-borne and seed-borne transmission, respectively.

The aforementioned examples reveal that the mechanisms of viral infection and viral diseases in plants are different from diseases caused by other pathogens. Bacteria or fungal pathogens cause diseases by producing enzymes, toxins or other pathogenic substances which have negative effects on the host, or consume nutrients of the host cells as parasites. By contrast, viruses affect plants in an indirect way. For instance, viruses cause symptoms such as leaf mosaic, ring spots, dwarfing, or stunting by interfering with the normal host metabolism. Many viruses cause visible symptoms in the infected plant, however, some infected plants are symptomless. Viruses in symptomless plants are called latent viruses.

1.2 Virus detection

Several methods have been developed to detect plant viruses, such as the enzyme-linked immunosorbent assay (ELISA), polymerase chain reaction (PCR), molecular beacons, immune-capillary zone electrophoresis (I-CZE), liquid chromatography, nucleic acid sequence based amplification (NASBA) (Wong, 2002) and loop-mediated isothermal amplification (LAMP) (Notomi *et al.*, 2000).

ELISA is a biochemical tool used mainly in medical immunology and plant pathology for pathogen diagnostics (Clark & Adams, 1977). It is performed in an ELISA plate made of polyvinyl chloride (PVC). The wells of the ELISA plate contain

immobilized antibodies, which can bind antigens from plant sample. Combined antigens capture the second specific antibodies, which can be detected by florescent techniques. ELISA test is rapid and sensitive, thus viruses in low titres may be detected (Stynen *et al.*, 1995).

PCR has been used in plant virus detection since early 1990s (Wong, 2002). For RNA viruses, viral RNA has to be reverse transcribed to complementary DNA (cDNA) which is used as template for PCR. PCR techniques, such as multiplex PCR, allow simultaneous detection of several viruses (Mumford *et al.*, 2000); and real-time PCR enables quantification, which can be used to determine viral concentration.

Molecular beacons are used to detect target sequence by a single-stranded nucleic acid probe which contains a hairpin structure. The probe sequence, which is complementary to the target nucleic acid, locates in the loop portion of hairpin structure. The two arm sequences are unrelated to the target sequence, and are complementary with each other to form the stem of hairpin. A fluorescent moiety is attached to the end of one arm of the hairpin structure, and a fluorescent quencher moiety is attached to the other arm. The two moieties are at a close proximity due to the two complementary arms of the stem. When the probe at the loop hybridizes to a target sequence, the hairpin structure is destructed and the two arms are separated. The separation leads to the detachment of the two moieties, and appearance of fluorescence (Tyagi & Kramer, 1996). Eun and Wong successfully detected *Cymbidium mosaic virus* (CymMV: genus *Potexvirus*) and *Odontoglossum ringspot virus* (ORSV: genus *Tobamovirus*) by molecular beacons (Eun & Wong, 2000).

I-CZE was also used by Eun and Wong (1999) to detect the above-mentioned two viruses. I-CZE combines the specificity afforded by serological assays with CZE. Based on the principle of CZE which separates components in samples by applying a voltage across a hollow capillary filled with electrolyte solution, I-CZE migrates antigen-antibody complexes, and the migration of complexes can be detected in real time.

Liquid chromatography is used in protein chemistry in conjunction with electrospray ionization (ESI) and matrix-assisted laser desorption/ ionization (MALDI) techniques. The method was proposed for identification of mutants of *Tobacco mosaic virus* (TMV: genus *Tobamovirus*) (Lewis *et al.*, 1998). All viruses with known coat protein molecular weights can be detected by this rapid, accurate and sensitive technique. The method was later used in some other studies (Koo *et al.*, 1999, Yao *et al.*, 2002).

NASBA is an isothermal amplification method for amplifying RNA. The technique integrates reverse transcription into the amplification process, and produces single stranded RNA. *Potato leafroll virus* (PLRV: genus *Polerovirus*) in potato tubers was detected by this method (Leone *et al*., 1997).

LAMP was developed by Notomi *et al.* (2000) to amplify DNA at a constant temperature (60 °C). On the target DNA, four specific primers recognize six distinct sequences. A complementary strand is synthesised by inner primer, after which the out primer releases the single-stranded DNA which serves as a template for the next cycle of DNA synthesis. The new strand hybridizes to the target DNA, and a stem-loop structure is produced by the hybridization. As the reaction continues, more stem-loop structures and DNA strands are produced (Notomi *et al.,* 2000). Because of the high selectivity and relatively more rapid reaction process of the LAMP method, it is widely used for DNA amplifications (Fukuta *et al.*, 2003, Fukuta *et al.*, 2004, Nie, 2005).

However, all techniques mentioned above rely on some presumption of the virus in order to design the primers or probes. Therefore most of them suit to detect viruses in cultivated plants. For wild plants which grow outside fields and gardens, there is little information about viral diseases, and it is difficult to suspect specific viral pathogens. Recently, a novel method was developed to provide an effective solution to detect viruses from plants without previous study. The method assembles viral small RNA reads from infected plants to contigs, and the contigs are used to search hits with Basic local alignment search tool (BLAST) in the nucleotide and protein databases of the National Center for Biotechnology Information (NCBI). In the study conducted by Kreuze *et al.* (2009), sweetpotato was infected by *Sweetpotato feathery mottle virus* (SPFMV: genus *Potyvirus*) and *Sweetpotato chlorotic stunt virus* (SPCSV: genus *Crinivirus*). Surprisingly, besides SPFMV and SPCSV RNAs, mastrevirus- and badnavirus-specific RNAs were also detected (Kreuze *et al*., 2009). The outcomes indicated that there exist more kinds of viruses than expected, even in crops which are well studied. The small RNA deep-sequencing method makes it possible to detect viruses without any prior knowledge, and it can be used for discovering viruses from the wild nature.

1.3 RNA silencing

1.3.1 Gene silencing

Gene silencing is generally described as suppression of gene activity at the mRNA expression level. It includes two types of mechanisms: transcriptional gene silencing (TGS) and posttranscriptional gene silencing (PTGS), or RNA silencing. RNA silencing refers to the process of down-regulation of gene expression triggered by dsRNA. The mechanism was first mentioned by Napoli *et al.* (1990) when they tried to overexpress chalcone synthase in colourful *Petunia* petals. However, they unexpectedly got variegated white flower instead of a dark-purple one in transgenic *Petunia*, because RNA silencing was induced and reduced mRNA accumulation. There are three types of RNA silencing in plants: cytoplasmic RNA silencing pathway, endogenous mRNA silencing pathway and DNA methylation and suppression of transcription (Baulcombe, 2004). In a virus-infected plant, cytoplasmic RNA silencing pathway provides an efficient virus resistance mechanism.

1.3.2 Virus-induced RNA silencing in plants

Virus-induced RNA silencing occurs in three main steps: initiation, amplification and spreading. The initiation is triggered by dsRNA, which is derived from replicating viral dsRNA or stem-loop structures in viral ssRNA, or it may be synthesized by one or more cellular RNA-dependent RNA polymerases (RdRp) that use viral ssRNA as a template. The dsRNA is recognized and accurately cleaved into viral small interfering RNAs (vsiRNAs) by Dicer-like (DCL) endonucleases (Voinnet, 2008). Unlike animals and fungi which possess one or two *DCL* genes, there are four *DCL* genes reported from rice and *Arabidopsis thaliana* (Margis *et al.*, 2006). The Dicers hierarchically process dsRNA into vsiRNAs of different sizes, which direct a series of gene-silencing pathways (Fusaro *et al.*, 2006). DCL4 is the primary sensor of viral dsRNA, and it produces 21 nt long vsiRNAs which are RNA components of the cell-to-cell silencing signals in plants (Dunoyer *et al.*, 2005). DCL2 acts as surrogate of DCL3 or DCL4, and produces 22 nt long vsiRNAs (Xie *et al.*, 2004). DCL3 generates 24 nt vsiRNAs in nucleus (Matzke & Birchler, 2005). DCL1, the minor contributor, accesses viral dsRNA when other DCL activities are compromised, and produces 21 nt long siRNA (Kurihara

& Watanabe, 2004).

vsiRNA are stabilized by methylation at their 3'-terminal nucleotide by a methyltransferase HEN1 (Chen, 2005). The methylated siRNAs are associated with different ARGONAUTEs (AGOs) according to the 5' terminal nucleotide of siRNA, and form RNA-induced silencing complexes (RISCs) (Mi *et al.*, 2008). The passenger strand, which is one of the vsiRNA strands, is degraded by small RNA degrading nucleases. The other strand called guide strand utilizes a sequence homology-dependent mechanism to guide RISC to attach to the target ssRNA by sequence-specific recognition. RISC serves as a slicer to cleave the target ssRNA (Hutvagner & Simard, 2008). Most of the cleaved RNA products are degraded by exoribonucleases, however, some RNAs are recruited by RdRp and synthesized new dsRNAs, which are the preferential substrates for DCL4, DCL3 and DCL2 to produce secondary vsiRNAs (Figure 4). Chen *et al.* (2010) suggested that secondary vsiRNA biogenesis is triggered by 22-nucleotide RNAs (Chen *et al.*, 2010), however, the latest study provides inconsistent results indicating that 21-nucleotide long RNAs direct the downstream pathway (Wang *et al.*, 2011). The secondary pool of vsiRNA spreads throughout the plant and acts as mobile silencing signals between cells to support plant systemic resistance of virus (Dunoyer *et al.*, 2010).

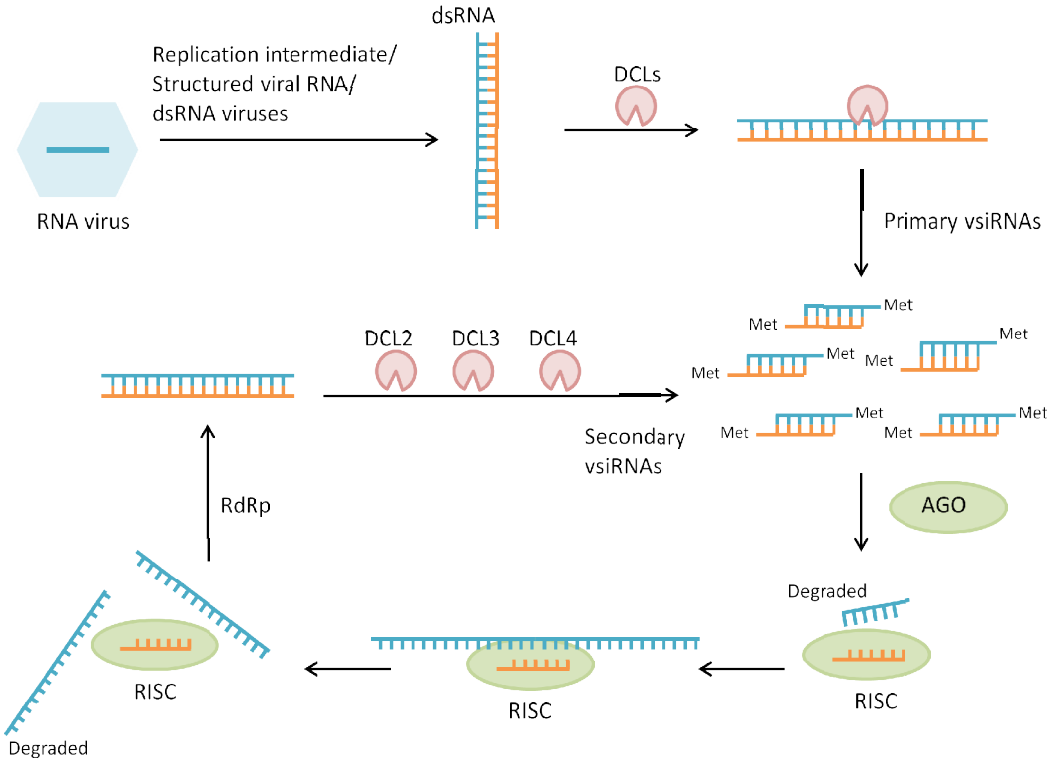


Figure 4. Schematic representation of virus-induced RNA silencing as plant antiviral

1.3.3 Viral counter-defence against silencing

The plant operates with RNA silencing as an antiviral mechanism; correspondingly, viruses generate solutions to overcome host antiviral silencing. There are three groups of proteins which help viruses to inhibit host antiviral immunity through a) reducing siRNA accumulation, b) blocking the silencing signal transport, and c) interfering the systemic signalling of RNA silencing (Li & Ding, 2001). These kinds of proteins encoded by virus are called viral suppressors of RNA silencing (VSRs). RNA silencing process might be blocked because the upstream reactions are interfered in the infected cell. For examples, potyviral helper component-proteinase (HC-Pro) interferes with the HEN1-mediated methylation vsiRNAs (Yu *et al.*, 2006); and poleroviral P0 degrades AGO which serves as the catalytic component in the mechanism (Baumberger *et al.*, 2007). Suppression of RNA silencing enhances viral accumulation in the host, and helps viruses to counter the antiviral defence.

1.4 Deep-sequencing of siRNA

1.4.1 Advantages

Compared with other virus detection methods, the advantages of virus detection by deep-sequencing of siRNA are apparent. First, deep-sequencing of siRNA does not require any supposition of the pathogen, thus it provides an excellent solution in studying viral species with limited available information. Second, it offers an unbiased view to all viral species and is not limited by the predictions, and helps to discover new viral pathogens even in well-studied plant species. Furthermore, deep-sequencing of siRNA is sensitive enough to detect viruses which only have small population in the hosts.

1.4.2 Applications

Scientists first started to use siRNA deep-sequencing to detect low-frequency viral mutations in *Human immunodeficiency virus* (HIV) because of the high sensitivity of the approach (Wang *et al.*, 2007). Then it was used in studies on drug-resistant viral mutations, for instance, the worldwide spread flu pandemic in 2009 (H1N1 influenza virus) (Greninger *et al.*, 2010, Ghedin *et al.*, 2011). The first application of the method

utilized in plant disease diagnostic was conducted by Kreuze *et al.* (2009). After that, Wu *et al.* (2010) used a similar technology to detect viruses from invertebrates (fruit flies, mosquitoes and nematodes) and discovered five previously undescribed viruses, among which four only exhibited low sequence similarities to known viruses (Wu *et al.*, 2010). These two studies suggested that deep-sequencing of vsiRNA is suitable for viral detection in both plants and animals. Then, many studies based on the method were done in different countries. Coetzee *et al.* (2010) identified four known viral pathogens including *Grapevine virus E* that had never been reported in South Africa vineyards, and built a viral census of South African vineyards using the deep-sequencing method. Besides viruses, viroids are also elicitors and targets of the RNA silencing mechanism, and can be detected by this method (Di Serio *et al.*, 2009, Bolduc *et al.*, 2010, Navarro *et al.*, 2009, Ling *et al.*, 2011).

In this study, siRNA deep-sequencing method is used to detect viruses in a wild plant, *Arctium tomentosum* Mill. (family Asteraceae, genus *Arctium*), showing vein yellowing and mosaic symptoms in Helsinki, Finland. *Arctium tomentosum*, which is also called woolly burdock or downy burdock, is a biennial herb that grows frequently in southern and central Finland. It has been observed throughout Europe except Portugal, Ireland and Albania (Flora Italiana: http://luirig.altervista.org/schedeit/ae/arctium_tomentosum.htm), and it is quite common in southern Sweden, central and eastern Europe and Russia. It is also has been found in several states of USA and several provinces of Canada, and in a small part of northern China (Den Virtuella Floran: http://linnaeus.nrm.se/flora/di/astera/arcti/arcttomv.jpg). The plant is usually found by roadsides and fence lines, and on waste ground. *A.tomentosum* completes its life cycle biennially. In the first year, a large fleshy taproot is developed, and heart-shaped leaves are usually 45 cm long by 40 cm wide with beneath white and woolly hair, and grow as a rosette. In the second year, the plants grow up to 2 m high with flower heads at the ends of branches. The leaves are alternated and smaller in size toward the top of a mature plant. The floral clusters consist of purple disk florets with overlapping hooked bracts. After flowering, the whole shoots of the plant die, and the flower heads turn brown. With bristling hook structure, the seeds are spread by attaching to animal skin or human cloth.

2 OBJECTIVES

The aim of this study was to detect viruses in the wild plant, *Arctium tomentosum*, displaying viral disease symptoms in Helsinki.

The study also aimed to find out how prevalent the possibly detected viruses are in the plants with similar symptoms.

3 MATERIALS AND METHODS

3.1 Viruses and plants material

The plants showing vein yellowing and mosaic symptoms were suspected to be infected by plant viruses. At first, six samples (Pe10-15) with symptoms were collected from Pihlajamäki (Helsinki, Finland) and total RNA from the samples was extracted by Trizol (Appendix 1). RNA from the six samples was pooled for siRNA deep-sequencing. The works mentioned above were done by other researchers in KPAT group (University of Helsinki, Finland). After siRNA deep-sequencing, the data were used in this study.

After viruses had been identified in the first samples, more samples with similar disease symptoms were collected in Helsinki (Figure 5) to study the distribution of viruses. The collected samples were marked as Pih 1-19 (Pihlajamäki), Vi 1-8 (Viikki), Pa 1-4 (Pasila), H 1-8 (Hämeenlinnanväylä) and R 1-8 (Rapakiventie) according to the sampling site. Fresh leaves were put into liquid nitrogen immediately after collection and stored at -80°C till use.

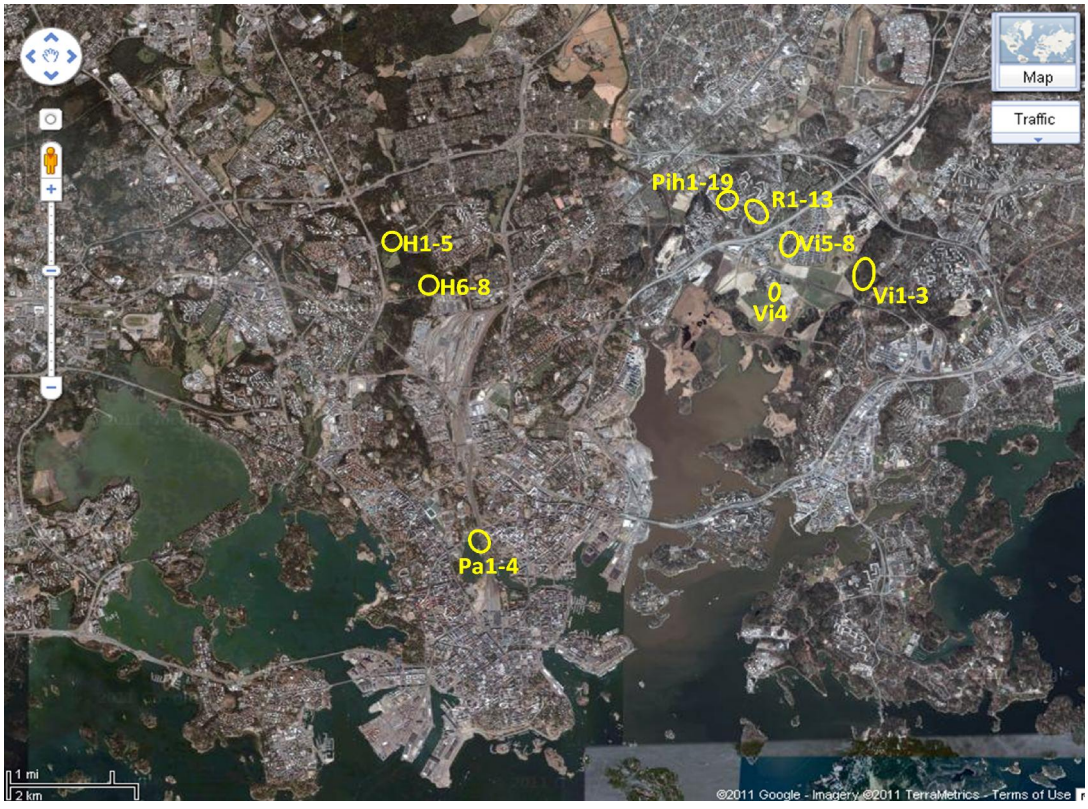


Figure 5. Sample collection sites in Helsinki.

3.2 Plant species identification

3.2.1 Host identification based on the *matK* gene sequence

"Barcode" is a short piece of DNA from organisms used to distinguish the species from others. Barcoding has been used in animal taxonomy since 2003, and was developed to be used in plant taxonomy since 2007 (Pennisi, 2007). The maturase K (*matK*) gene is one of the seven leading candidate genes of plant "barcode" (Hollingsworth *et al.*, 2009). It was used for identifying the plant species inspected for viruses in this study.

Total DNA was extracted from leaves of two plants using the cetyltrimethylammonium bromide (CTAB) method (Allen *et al.*, 2006) and used in polymerase chain reaction (PCR). About 200 mg of plant leaf tissue was ground in liquid nitrogen, and transferred into a 2 ml Eppendorf tube. Before the leaf tissue powder melted, 1.2 ml of pre-heated extraction buffer (65°C) (Appendix 1) was added into the tubes. The tubes were vortexed for 10 seconds to mix thoroughly, and incubated at 65°C for 30 min with inversion every 10 min. Samples were centrifuged at 13,500 *g* for 10 min at room temperature. The supernatants were transferred into new 1.5 ml Eppendorf tubes that contained 800 µl of phenol (pH 7.5): chloroform: isoamyl alcohol (25:24:1). Subsequently, samples were mixed by inverting tubes for 20 min at room temperature, and were centrifuged at 13,500 *g* for 10 min to separate the phases. The upper aqueous phase was transferred to a new Eppendorf tube containing 800 µl of cold ethanol. DNA was precipitated by mixing the solution and incubating for 10 min at room temperature, and collected by centrifuging the mixture at 13,500 *g* for 10 min. Supernatants were removed and DNA pellets were resuspended in 250 µl of Tris-ethylenediaminetetraacetic acid (TE) buffer (Appendix 1). Resuspended DNAs were treated by incubating with 2.5 µl DNase-free RNase at 37°C for 30 min, and precipitated with 600 µl of precooled (-20°C) ethanol with 25 µl of 3 M sodium acetate (NaAc) at -20°C for 20 min. Then, samples were centrifuged and supernatants were removed as described above. DNA pellets were washed by 500 µl of cold 70% ethanol with 10 min centrifugation at 13,500 *g*. Ethanol was removed carefully without disturbing the pellets, and the pellets were dried in air by keeping the tubes upside down for at least one hour. Finally, DNA was resuspended in 25 µl of nucleotide-free water (Sigma W4502), and stored at -80°C till use.

A pair of universal primers for land plants, matK2.1af (5'- ATCCATCTGGAAA

TCTTAGTTC-3') and matK5r (5'- GTTCTAGCACAAGAAAGTCG-3') from Kew Royal Botanic Garden's DNA Barcoding website (http://www.kew.org/barcoding/protocols.html), were ordered from Oligomer Oy (Helsinki, Finland) for amplifying the *matK* gene. The PCR reaction (25 µl) consisted of 18 µl of nucleotide-free water, 2.5 µl of 10x Dynazyme buffer (F511), 0.3 µl of 10 mM dNTPs, 0.7 µl of forward primer, 0.7 µl of reverse primer, 0.3 µl of DynazymeTM II DNA Polymerase (Finnzymes, Espoo, Finland) and 2.5 µl of extracted DNA. The PCR program was 94°C for 3 min, followed by 35 cycles of 94°C for 25 seconds, 49.5°C for 25 seconds, 72°C for 75 seconds, and final extension at 72°C for 10 min. PCR products were analyzed by agarose gel electrophoresis by running the gel at 140 V for 40 min, and visualized under UV light. The 1% agarose gel was prepared in 1x Tris-acetate-ethylenediaminetetraacetic acid (TAE) buffer. Final concentration of 1 µg/ml ethidium bromide was added when the agarose solution had cooled down to 60°C.

PCR products of the expected size were purified from the agarose gel using E.Z.N.A Gel Extraction Spin Protocol (E.Z.N.A. TM gel extraction Kit D2500-02, Omega Bio-Tek, Norcross GA, USA) according to the manufacturer's instructions. Specific DNA bands in the gel were rapidly cut by using a wide, clean and sharp scalpel under UV light (UV Hoefer Macrovue transilluminator UV25; American Biosciences, Blauvelt, NY, USA). The volume of gel slice was determined by weighing it in a 1.5 ml Eppendorf tube. Equal volume of Binding Buffer (XP2) was added into the tube containing the gel slice, and the tube was incubated at 60°C for 10 min or until the gel slice melted completely, and mixed thoroughly by shaking for 2 min. The DNA/agarose solution was transferred into a HiBind DNA Mini Column in a 2 ml tube, and the tube was centrifuged at 10,000 *g* for 1 min at room temperature. The flow-through liquid was discarded, and 300 µl of Binding Buffer (XP2) was added into the column. The 2 ml tube which comprises the colomn was centrifuged again as described, and flow-through liquid was discarded. After that, HiBind DNA Mini Column was washed by 700 µl of SPW Washing Buffer by centrifuging at 13,000 *g* for 1 min. The liquid was discarded again, and the empty column was centrifuged at the maximal speed for 2 min to dry the column matrix. Subsequently, the HiBind DNA Mini Column was placed into a clean 1.5 ml Eppendorf tube. About 30 µl of nucleotide-free water was added onto the column matrix, and the column was incubated for 2 min at room temperature. DNA was eluted by centrifuging at the maximal speed for 1 min.

Concentration and quality of the recovered DNA was determined by

spectrophotometer at the wavelength of 260 nm. Samples were sequenced at Haartman Institute (Helsinki, Finland) and analysed by Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) of the National Center for Biotechnology Information (NCBI).

3.2.2 Identification based on morphological characteristics

Besides identification of the "barcode" gene, morphological observation and identification was also done according to a Swedish Flora (Mossberg & Stenberg, 1997) with help of botanist Ilpo Kuokka.

3.3 RNA extraction

Single-stranded RNA is easily degraded by ribonucleases (RNase). Therefore, it is important to keep plant leaves at -80°C before use. Gloves are required during the whole experimental process in order to prevent RNase contamination.

Total RNA was extracted from *A.tomentosum* samples for reverse transcription polymerase chain reaction (RT-PCR) according to the Trizol protocol. In this protocol, plant leaf samples were ground into powder in liquid nitrogen, and about 0.4 g leaf powder was quickly transferred into a 1.5 ml Eppendorf tube. Prepared Trizol (1ml for each sample) (Appendix 1) was added before leaf powder melted, mixed well by vortexing and incubated for 5 min at room temperature. Samples were centrifuged at 12 000 *g* for 10 min at +4°C, and the supernatants were transferred to new 1.5 ml Eppendorf tubes containing 0.2 ml of chloroform. The samples were mixed well vigorously shaking by hand. Mixtures were incubated at room temperature for 5-15 min and centrifuged for 15 min at 12 000 *g* at +4°C. The upper aqueous phase was carefully transferred into a new 1.5 ml Eppendorf tube without any contamination from the lower phase. The same volume of isopropanol was added and mixed well, followed by centrifuging at 12 000 *g* for 20 min at +4°C. All transfers mentioned above were done in a fume hood. After centrifugation, supernatant was removed carefully without loss of the RNA pellet. The pellets were then washed with 1 ml of 75% ethanol, and samples were centrifuged for 15 min to collect the pellets. Tubes with RNA pellets were kept upside down on a paper tissue for drying the RNA pellets until they became clear at the bottom. Pellets were dissolved in 70 µl of nuclease-free water by heating the tubes at

55-60°C for 10 min to make sure the RNA pellets were dissolved completely. The concentration and purity of RNA samples were determined by a spectrophotometer at the wavelength of 260 nm and the absorbance ratio at 260 nm/280 nm (A260/280) and 260nm/230 nm (A260/230). Agarose gel electrophoresis was used to check whether RNA was intact.

3.4 siRNA sequencing

Small RNA deep-sequencing required 10 µg of total RNA which was sent to Fasteris SA (Plan-les-Ouates, Switzerland). The main steps of small RNA sequencing include separation of RNA by electrophoresis in an acrylamide gel, isolation of small RNA (less than 30 nt) from the gel, and ligation with a single-stranded 3' adapter and a bar-coded 5' adapter. Ligated RNA is reverse-transcribed and amplified by PCR to generate the DNA Colony Template Library. After quality control of the DNA Colony Template Library, it is purified and diluted to concentration of 10 nM. Then, Illumina Genome Analyzer was applied for high-throughput DNA sequencing. Data were analyzed by GA Pipeline to convert images into sequences.

3.5 Sequence analysis

3.5.1 Velvet assembly

Velvet (http://www.csc.fi/english/research/sciences/bioscience/programs/velvet/index_html) is a program used for producing contiguous sequences (contigs) by assembling short RNA reads based on De Bruijn graphs. In the De Bruijn graphs, data elements are organized around words of *k* nucleotides (*k*-mer). Reads are mapped and go through the graph in a determined order from one word to the next (Zerbino & Birney, 2008). Reads are hashed according to a predefined *k*-mer, which represents the numbers of nucleotides overlapping between reads. Thus, the procedure with large *k*-mer brings out more specific contigs but decreases the read coverage, while small *k*-mer produces contigs with high coverage but in lower specificity (Zerbino, 2008).

Velvet carried out error removing, which means that it focuses on analyzing topological features of sequences instead of correcting errors made by sequencing process or biological reason such as polymorphisms. There are three structural types of erroneous data: "tips" meaning errors at the edges of reads; "bulges" meaning errors at

internal reads; and erroneous connections meaning errors due to cloning or distant merging tips. Velvet corrects the three errors mentioned above after graph creation and allows simultaneous operations (Zerbino & Birney, 2008).

There were several parameters which had to be determined before running the program. Parameter *coverage cutoff* is used to eliminate reads with low coverage, which are supposed to be errors. A large *cutoff* corrects the output, but with a risk that reads in low concentration will be wrongly eliminated.

There were two steps to construct short reads into contigs by Velvet. The first step was to produce a hashtable, and create two new files which were necessary for the following step. For instance, command:

velveth *output_dir k-mer* –short *read.fastq*

was used for achieving the first stage. Multiple read files could be added into the same command by separating them with a space. *K*-mer must be an odd number inferior reads length, and in maximum it could be 31. Experience showed that the best *k*-mer for these data were 15 and 17 (Zerbino, 2008).

The second step produces contigs by typing command:

velvetg *output_dir* –min_contig_lgth *50* -cov_cutoff *5* –unused_reads yes

This command is used to output the results with determined minimum contigs length and the *cutoff* value (*cutoff* = 50 in this study). Since the concentration of some of the virus might be relatively low in the host plant, *coverage cutoff* was set to 5 to prevent loss of small amount of viral reads.

3.5.2 BLAST

Contigs assembled by Velvet were used to search similar sequences in databases by BLAST. The *Organism* option of BLAST program was limited to viruses in order to prevent mismatches, especially for short contigs. Viral sequences from BLAST results were used as references for siRNA reads alignment with MAQ and Novoalign.

3.5.3 siRNA alignment by MAQ

MAQ (http://maq.sourceforge.net/index.shtml) was used to build assemblies by mapping shotgun reads to a reference sequences. MAQ searches ungapped match with lowest mismatch score, and it defines the mismatched bases as the sum of qualities at

the alignment stage. In order to speed up the alignment, MAQ only aligns reads to positions which have two or fewer mismatches in the first 28 bp (default parameters) (Li *et al.*, 2008).

Generally, selected viral genomic sequences as references can be divided into two groups. The first group contains viruses which have been well-studied, and these viruses in certain plant might be expected according to prior experiences and studies. The efficient presumption significantly saves time for running the program. However, in cases where it is difficult to make valuable assumptions, such as with viruses that have been little studied, or a host whose viruses have not been studied, the better method is to use all viral genome data available in the database as reference. The latter approach definitely increases workload, but provides higher chance to detect out novel viruses.

MAQ process includes several stages. The first thing is to load reference sequences (download from database e.g. NCBI) and siRNA reads into the program, and convert them into binary fasta format for alignment. After that, statistics of alignment is executed and mapping assemblies are built. Finally, consensus sequences and assembling qualities are extracted and displayed (MAQ user's manual, Release 0.5.0).

MAQ offers an easy solution to run the program by using the EASYRUN command. The command does most of procedures in batch if references are in fasta format and reads are in fastq format. Command of easyrun script is:

maq.pl easyrun -d *outdir ref.fasta reads.fastq*

In this command, "*outdir*" stands for directory which stores the outputs, while "*ref.fasta*" stands for the sequence used as reference in fasta format. "*reads.fastq*" is the fastq file containing siRNA reads. Output of MAQ indicates numbers of aligned reads among the total amount of siRNA reads.

3.5.4 siRNA alignment by Novoalign

Novoalign (http://www.novocraft.com/main/index.php) is a program similar as MAQ and used to align short reads to reference sequences. The difference between the two programs is that Novoalign optimizes the algorithm to produce an alignment score more accurately. It alleviates the limitations of MAQ that allows two mismatches per read by being able to align reads having as many as eight mismatches.

For running the program, Novoalign requires indexes of formatting before alignment. Command

novocraft/novoindex *ref-formatting-name ref.fasta*

was used for formatting reference into fasta file.

Alignment was conducted by command:

novocraft/novoalign -d *ref-formatting-name* -f *reads.fastq* -F ILMFQ

In the procedure, "*ref-formatting-name*" represents the formatted reference file, "*reads.fastq*" represents siRNA reads files in fastq format that need to be aligned, and "ILMFQ" is the fastq file with qualified Illunimar coding. More options about the command are available in Novoalign user guide (http://www.novocraft.com/wiki/tiki-index.php).

3.5.5 Alignment analysis

Viral sequences aligned by the programs might be non-contiguous. The coverage of alignment can be tested through comparison between the recovered sequence and the sequence in database. MultAlin (http://multalin.toulouse.inra.fr/multalin/multalin.html) and ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) are programs to align multiple sequences, and used to analyze coverage quality.

3.6 Virus detection by RT-PCR

Viruses detected in the samples by siRNA deep-sequencing were confirmed by other assays in the laboratory. The individual samples were tested to discover the incidence and geographical distribution of the viruses in *A.tomentosum*.

3.6.1 Reverse transcription reaction

To synthesize cDNA on *Alstroemeria virus X* (AlsVX, genus *Potexvirus*, family Alphaflexiviridae), poly-T oligonucleotide primer (23 T+N) (Oligomer Oy) was used because the virus contains a 3'-terminal poly-A tail. cDNA of *Fig mosaic virus*-Hel (FMV, genus *Emaravirus*) was synthesized using random hexamer primers (Oligomer Oy).

For the reverse transcription reaction, 5 µl of RNA and 2.5 µl of primers (10 mM) were added into a 0.5 ml PCR tube. The tube was incubated at 70°C for 5 minutes to allow primers attach, and was chilled on ice immediately to stop the reaction. After cooling down, pre-mixed 17.5 µl of the reaction components (mastermix) was added.

The mix contained 9 µl of nuclease-free water, 2 µl of 10 mM deoxyribonucleoside triphosphates (dNTPs) (Finnzymes, Espoo, Finland), 1 µl of *Moloney murine leukemia virus* (M-MLV) reverse transcriptase (Promega, Madison, WI, USA), 5 µl of 5x M-MLV reverse transcription buffer and 0.5 µl of ribonuclease inhibitor enzyme (Promega, Madison, WI, USA). The RNA and reaction components were mixed well and the reaction mixes were incubated at 37°C for 60 minutes for synthesis cDNA. After that, cDNA were diluted by 2 times and stored at -20°C till use.

3.6.2 PCR reaction

Qualities of cDNA were tested by a positive control PCR, which used the universal primers to amplify *matK* gene of plant samples. Primers for viral detection were designed by program Primer3 (http://frodo.wi.mit.edu/primer3/). Primers were designed to amplify a fragment of the coat protein (CP) gene of AlsVX (Figure 6) and the putative nucleocapsid protein (NP) gene of the FMV-Hel (Figure 7). Since the amount of PCR products in AlsVX-infected samples were too small to be visualized clearly in gel electrophoresis, two pairs of primers were designed for amplifying the CP gene of AlsVX by nested PCR. The second round reaction (amplified by inner primers) uses the products from the first round reaction (amplified by outer primers) as templates. Nested PCR promotes PCR results to achieve very high sensitivity.

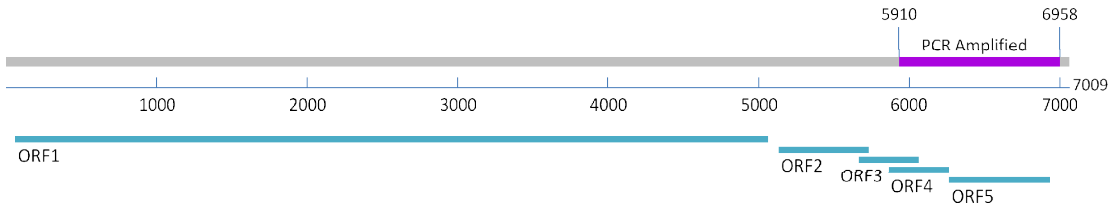


Figure 6. Genomic RNA of AlsVX. The five open reading frames (ORF) encode the replicase protein (ORF1: nt 98-5038), the triple gene block protein 1 (ORF2: nt 5067-5771), the triple gene block protein 2 (ORF3: nt 5746-6078), the triple gene block protein 3 (ORF4: nt 5903-6202) and the coat protein (ORF5: nt 6240-6932).

The PCR mastermix contained all necessary reagents except template (Tables 1-3) was divided into PCR tubes. Mastermix helps to control contaminations during pipetting. Details of PCR reagents are shown in Tables 1-3.

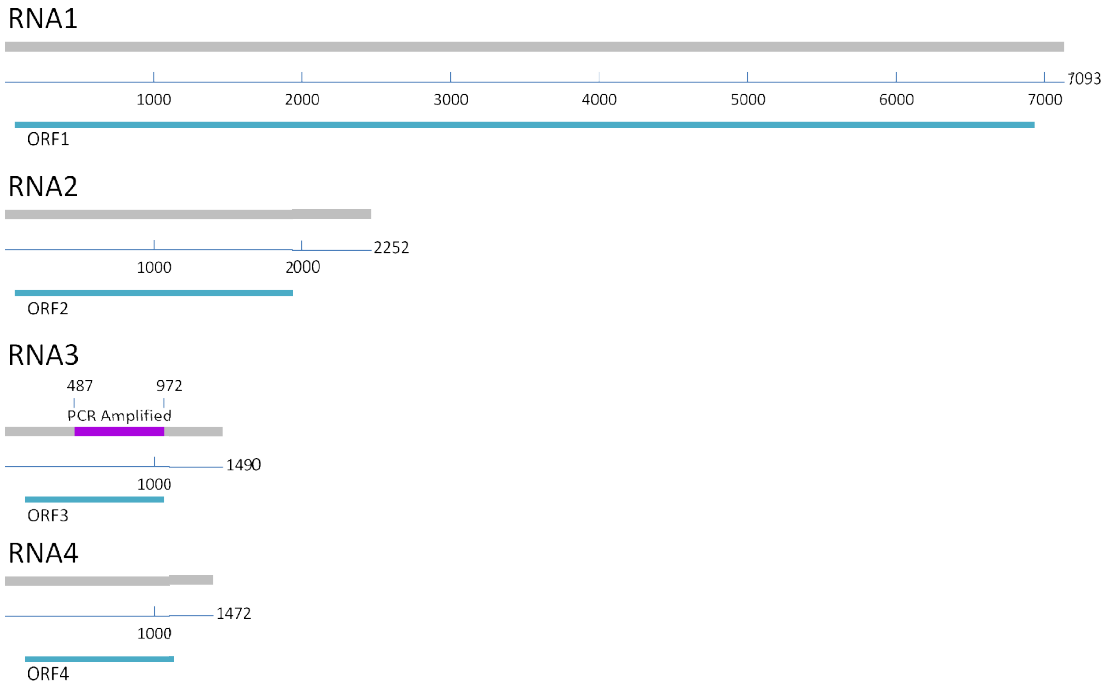


Figure 7. The four genomic RNAs of FMV each contain a single open reading frame (ORF). The ORFs encode the RNA-dependent RNA polymerase (RdRp) (ORF1: nt 40-6993), the putative glycoprotein precursor (ORF2: nt 53-1978), the putative nucleocapsid protein (ORF3: nt 100-1047) and the unknown protein P4 (ORF4: nt 82-1167).

Table 1. The reaction for the first round of PCR for AlsVX detection.

PCR mix		Numbers of PCR cycles		
Mastermix: 10xDynazyme buffer (F511)	2.5 µl			
dNTPs 10 mM	0.3 µl	1x	3 min	94°C
Outer forward primer 10 µM	0.7 µl	5x	30 sec	94°C
(5'CCACCGCCTAACTACGAAAA-3')			30 sec	50°C
Outer reverse primer 10 µM	0.7 µl		75 sec	72°C
(5'TTTTTTTTTTTTTTTTTTTTTTTN-3')				
Dynazyme	0.3 µl	30x	30 sec	94°C
H_2O	18 µl		30 sec	52°C
Template: cDNA	2.5 µl		75 sec	72°C
Total	25 µl	1x	10 min	72°C
Expected product size (nt): 1 348				

Table 2. The reaction for the second round of PCR for AlsVX detection.

PCR mix		Numbers of PCR cycles		
Mastermix: 10xDynazyme buffer (F511)	2.5 µl			
dNTPs 10 mM	0.3 µl	1x	3 min	94°C
Inner forward primer 10 µM	0.7 µl	35x	30 sec	94°C
(5'-GGCGACCAACAACATTCTCT-3')			30 sec	56°C
Inner reverse primer 10 µM	0.7 µl		75 sec	72°C
(5'-CAAACTGTAGAAAAACACCTCC-3')				
Dynazyme	0.3 µl	1x	10 min	72°C
H_2O	18 µl			
Template: PCR product from the first round	2.5 µl			
Total	25 µl			
Expected product size (nt): 1 120				

Table 3. The reaction of PCR for the FMV-Hel detection.

PCR mix		Numbers of PCR cycles		
Mastermix: 10xDynazyme buffer (F511)	2.5 µl			
dNTPs 10 mM	0.3 µl	1x	3 min	94°C
Forward primer 10 µM	0.7 µl	5x	30 sec	94°C
(5'-TCAACAAACGACTTCCTCACTG-3')			30 sec	50°C
Reverse primer 10 µM	0.7 µl		75 sec	72°C
(5'-AGGGTGCTGATCTGTCTGCT -3')				
Dynazyme	0.3 µl	30x	30 sec	94°C
H_2O	18 µl		30 sec	53°C
Template: cDNA	2.5 µl		75 sec	72°C
Total	25 µl	1x	10 min	72°C
Expected product size (nt): 604				

3.6.3 DNA sequencing and sequence alignment

PCR products of several samples collected from different sites were purified by E.Z.N.A Gel Extraction Spin Protocol and sent to Haartman Institute (University of Helsinki) for sequencing. Sequences were analyzed by program ClustalW2 and BLAST.

3.7 Phylogenetic analysis

Phylogenetic analysis was carried out to study the evolutionary relationships among FMV-Hel isolates and other viruses in the genus *Emaravirus.* The phylogenic tree was constructed by program MEGA 4. In the program, viral nucleotide sequences were first translated into amino acid sequences in order to align the proteins among the viruses. The alignment was done by ClustalW. After that, viral sequences retrieved from databases were trimmed to the same length with the sequences determined from the FMV-Hel in this study. The tree was built under the model Kimura 2-parameter by Neighbour-joining method with default bootstrap.

4. RESULTS

4.1 Plant species

The sequences of the *matK* gene amplified from two plants were subjected to BLAST search. They showed 98% identity at nucleotide level (nucleotide BLAST) and 96% identity in amino acid level (BLASTx) with the *matK* gene of *Arctium lappa* (Accession NO: HM989769.1). However, morphological characteristics including plant size, the leaf shape, the primary root and the densely hairy underside of leaves suggested that the plant belonged to *Arctium tomentosum* (Mossberg & Sterberg, 1997). Until now, no *matK* gene of *A.tomentosum* has been published.

4.2 Virus identification based on the contigs

Program Velvet produced a large amount of contigs (Table 4). Many of them showed high identity to sequences in databases when subjected to BLAST. BLAST results on the 15 nt and 17 nt hash-length contigs suggested that the viruses in the samples belonged to genus *Potexvirus* (family *Alphaflexiviridae*), and to genus *Emaravirus* (Table 5).

The lengths of specific contigs and the coverage of identity fit to known viruses are shown in Table 5. With hash-length 15, most contigs which could be aligned were between 100-230 nt, and the shortest and largest ones (47 nt and 227 nt) were aligned to *European mountain ash ringspot-associated virus* (EMARAV) and AlsVX, respectively. With hash-length 17, longer contigs could be aligned with sequences in databases. The largest aligned contig (488 nt) and shortest aligned contig (43 nt) matched to *Fig mosaic associated virus* (FMAV) and AlsVX, respectively.

Table 4. Number of contigs in different size under 15 nt and 17 nt hash-lengths.

k-mer	No. of contigs	Contigs Size (nt)					Maximum size	
		<100	100-299	300-599	600-1000	>1000		
15 nt	1055	951	97	5	2	0	36	642
17 nt	943	860	75	5	2	1	34	1090

Table 5. Assembled contigs and their alignment with sequences retrieved from the NCBI database by nucleotide BLAST.

					Hash-length 17 nt		Hash length 15 nt	
Virus [a]	Family	Genus	Country/ strain	Gene [b]	Length of contig (nt)	Identity	Length of contig (nt)	Identity
AlsVX	Alphaflexiviridae	*Potexvirus*	Japan	RP gene	314	271/320 (82%)	227	227/240 (95%)
				RP gene	271	273/287 (95%)	167	151/179 (84%)
				CP gene	108	109/116 (94%)	106	104/111 (94%)
				RP gene	43	56/61(92%)		
FMAV	Unassigned ssRNA viruses	*Emaravirus*	USA/ JJW2008	NP-like gene	488	305/374 (82%)	108	81/97 (84%)
FMV	Unassigned ssRNA viruses	*Emaravirus*	Canada/ Can01	Putative NP gene	488	307/375 (82%)	108	82/92 (85%)
			Canada/ Can01	RdRp gene	51	48/55 (87%)	(Not determined)	
NMV	Alphaflexiviridae	*Potexvirus*	Netherlands	RdRp gene	162	41/44 (93%)		
			New Zealand	RdRp gene	162	47/50 (94%)		
EMARAV	Unassigned ssRNA viruses	*Emaravirus*	Germany	RdRp gene	(Not determined)		47	45/47 (96%)

[a] NMV: *Narcissus mosaic virus*.

[b] RP gene: Replicase protein gene. CP gene: coat protein gene. NP gene: putative nucleocapsid protein gene. RdRp gene: RNA-dependent RNA polymerase gene.

4.3 Virus identification based on the vsiRNA reads

Both programs, MAQ and Novoalign, gave out similar alignment scores. Novoalign performed better than MAQ, meaning that it aligned more reads to the references. However, the results of both programs were the same when the suspected viruses were sorted by alignment scores (Table 6). According to the output, it is clear that AlsVX had the highest alignment score in all reads with different sizes. MAQ recovered 40% of the genomic nucleotide sequences of AlsVX by aligning siRNA reads. FMAV and FMV showed the next highest fit, which MAQ recovered 18% and 17% of their genomeic sequences by siRNA alignment, respectively. Generally, reads could be aligned with 10 viruses, which included AlsVX, FMAV, FMV, *Asparagus virus 3* (AV3), *Lettuce virus X* (LeVX), *Scallion virus X* (ScaVX), EMARAV, NMV, *Chenopodium mosaic virus X* (CMVX, also called *Malva mosaic virus*) and *Pepino mosaic virus* (PeMV). Scores of the alignment varied between different viruses. FMAV, FMV and EMARAV were taxonomically placed into the unassigned genus *Emaravirus* (Elbeaino et al. 2009). The remaining seven viruses all belonged to genus *Potexvirus* in family Alphaflexiviridae.

Table 6. Numbers of aligned reads by MAQ and Novoalign.

	21 nt reads (n=603,783)		22 nt reads (n=341,407)		23 nt reads (n=373,102)		24 nt reads (n=2,440,554)	
Virus	MAQ	NOVO	MAQ	NOVO	MAQ	NOVO	MAQ	NOVO
AlsVX	3 142	3 167	2 304	2 393	386	460	731	920
FMAV	1 261	1 271	340	344	109	111	209	232
FMV	1 074	1 105	318	341	72	85	143	198
AV3	668	687	386	409	62	72	159	221
LeVX	511	525	348	366	67	74	61	123
ScaVX	411	420	256	270	42	48	75	127
EMARAV	377	391	148	151	20	29	46	67
NMV	366	380	229	251	45	52	60	137
CMVX	257	276	140	169	39	45	43	104
PeMV	141	141	49	61	10	20	26	51

4.4 Virus identification based on RT-PCR and sequencing

4.4.1 Detection of AlsVX

In the first-collected six samples, AlsVX was only detected in sample Pe15 (Figure 8). The CP gene and the triple gene block protein 3 of AlsVX are 692 nt (230 aa) and 299 nt (99 aa) long, respectively. The amplified fragment of AlsVX was about 1100 nt long, which contains complete CP gene and partial of the triple gene block protein 3 (Figure 6). The sequence of the PCR product was aligned with the sequence of AlsVX CP gene in NCBI. The CP sequences were 82% identical at the nucleotide level (Appendix 2) and 90% identical at protein level. According to BLASTx results, the virus in sample Pe15 also had 73% amino acid identity to the CP of NMV strain New Zealand, 70% identity to CP of ScaVX, 69% identity to CP of AV3, 66% identity to CP of CMVX, 64% identity to CP of PeMV, and 64% identity to CP of LeVX (Table 7; Appendix 3 for accession number).

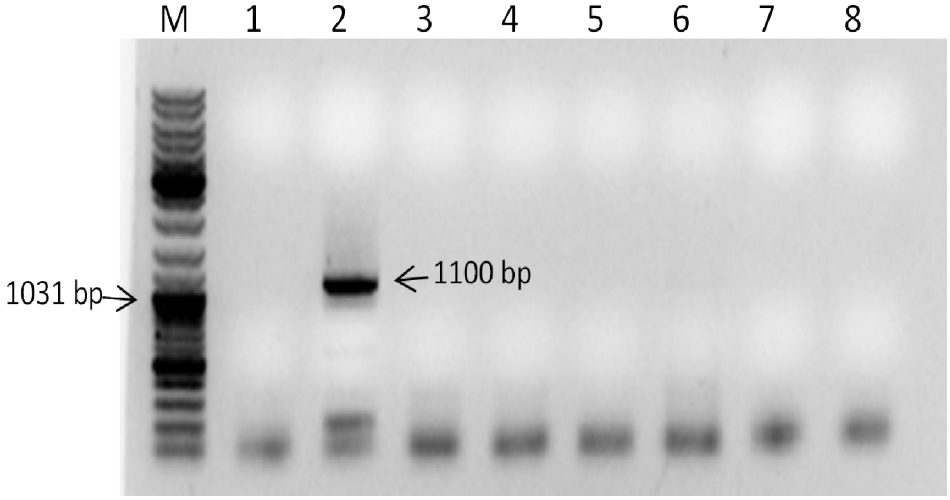


Figure 8. AlsVX detection in *A.tomentosum* Pe15 by RT-PCR. Lane 1: negative control, lane 2: Pe15, lane 3: Pe14, lane 4: Pe13, lane 5: negative control, lane 6: Pe12, lane 7: Pe11, lane 8: Pe10, M: molecular weight marker (GeneRuler^{TM} DNA Ladder Mix).

Table 7. Percentage of nucleotide identity (upper triangle, 1033 nt) and amino acid identity (lower triangle) of AlsVX-Hel and other potexviruses (accession numbers in Appendix 3).

nt / aa	AlsVX-Hel	AlsVX-Jap	NMV	ScaVX	AV3	CMVX	PeMV	LeVX
AlsVX-Hel	*	82	72	74	72	70	66	68
AlsVX-Jap	90	*	71	70	70	68	68	72
NMV	73	74	*	75	73	73	67	68
ScaVX	70	73	80	*	80	71	68	67
AV3	69	69	83	91	*	73	67	68
CMVX	66	70	78	75	78	*	66	64
PeMV	64	65	71	66	66	64	*	68
LeVX	64	63	65	63	61	59	58	*

4.4.2 Detection of FMV-Hel

The FMV-Hel was detected in all six *A.tomentosum* plants of the first collection. The sequences of the six PCR products were identical. The sequence of FMV-Hel described in this study differs from other FMV strains at the nucleotide level (Appendix 4). The putative nucleocapsid protein gene of the FMV-Hel included in the amplified region was 78% identical with the FMV strain Can01, 77% identical with FMV strain Gr10, 78% identical with FMAV-JJW2008, 69% identical with *Rose rosette virus* (RRV), and 59% identical with EMARAV. At protein level, the FMV-Hel had 78%, 78%, 75%, 74% and 52% identity with the putative nucleocapsid protein of FMV-Can01, FMV-Gr10, FMAV-JJW2008, RRV and EMARAV, respectively (Table 8).

Table 8. Percentage of nucleotides identity (upper triangle, 518 nt) and amino acid identity (lower triangle) of FMV-Hel and other related viruses (accession numbers in Appendix 3).

nt / aa	FMV-Hel	FMV-Can01	FMV-Gr10	FMAV-JJW2008	RRV	EMARAV
FMV-Hel	*	78	77	78	69	59
FMV- Can01	78	*	94	98	70	66
FMV-Gr10	78	100	*	94	69	64
FMAV-JJW2008	75	99	99	*	70	63
RRV	74	65	62	75	*	58
EMARAV	52	41	41	50	42	*

FMV shares high amino acid identities with EMARAV, *Pigeonpea sterility mosaic virus* (PPSMV) and *Maize red stripe virus* (MRSV) (Elbeaino *et al.*, 2009). A phylogenetic tree was constructed to demonstrate relatedness of members in genus *Emaravirus* (Figure 9). According to the phylogenetic analysis, the FMV-Hel from Helsinki is only distantly related to FMV strains from Canada (FMV-Can01), Italy (FMV-Gr10) and USA (FMAV-JJW2008), and has even further evolutionary distance to other *Emaravirus* members (Figure 9).

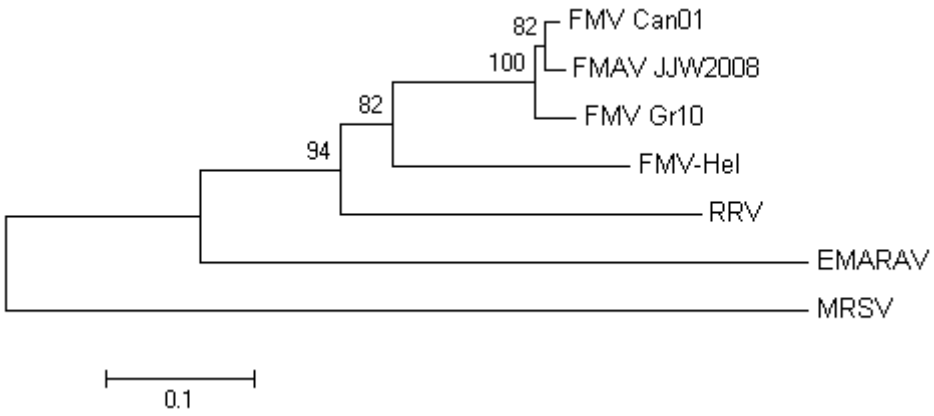


Figure 9. Phylogenetic analysis of the RNA3 putative nucleocapsid encoding region of members of genus *Emaravirus* (accession numbers in Appendix 3).

4.5 Distribution and variability of the FMV-Hel in Helsinki area

The FMV-Hel was detected in *A.tomentosum* plants with virus-like disease symptoms in different places of Helsinki (Table 9; Figures 10-14). However, all these samples except Pe15 were AlsVX-free. Symptoms and the results of PCR-based detection of the FMV-Hel are shown in Table 9. The typical symptoms of *A.tomentosum* infected with FMV-Hel were yellowing of veins in the leaves. Furthermore, leaf mosaic appearing along to veins or distributed on the whole leaf was also a typical symptom in the severely diseased plants (Figures 10-14).

Table 9. PCR-based detection of the FMV-Hel, and the foliar symptoms observed in *A.tomentosum* at different sampling sites.

Sampling site	Sample	PCR test	Vein yellowing	Mosaic	Chlorotic rings	Leaf rolling	Crinkle	Dark green spots
Pasila	Pa1	+		+				
	Pa2	+	+	+		+		
	Pa3	+	+	+				
	Pa4	+	+	+				
Viikki	Vi1	-					+	
	Vi2 (ns[a])	-						
	Vi3	-					+	
	Vi4	+	+					
	Vi5	+	+				+	
	Vi6	-					+	
	Vi7 (ns)	-						
	Vi8	+	+				+	
Pihlajamäki	Pih1	+	+	+				
	Pih2	+		+				
	Pih3	-						+
	Pih4 (ns)	-						
	Pih5	-						+
	Pih6	-					+	
	Pih7	+	+	+			+	
	Pih8	+	+	+				
	Pih9	+	+					
	Pih10	+	+					
	Pih11	+	+					
	Pih12	+	+					
	Pih13	+	+	+				
	Pih14	+	+	+				
	Pih15	+	+					
	Pih16	+	+	+				
	Pih17	+	+				+	
	Pih18	+	+	+				
	Pih19	+		+				
Hämeenlinnanväylä	H1	+	+					
	H2	-			+			
	H3	-			+			
	H4	+	+					
	H5	+		+				
	H6	+	+	+				
	H7	+	+	+				
	H8	+	+				+	+
Rapakiventie	R1	+	+					
	R2	+	+					
	R3	+		+			+	
	R4	+	+					
	R5	+	+				+	
Rapakiventie	R6	+	+	+				
	R7	+						
	R8	+	+	+				
	R9	+		+				
	R10	+	+				+	
	R11	+	+					
	R12	+	+			+		
	R13	+	+					

[a]ns: No symptoms.

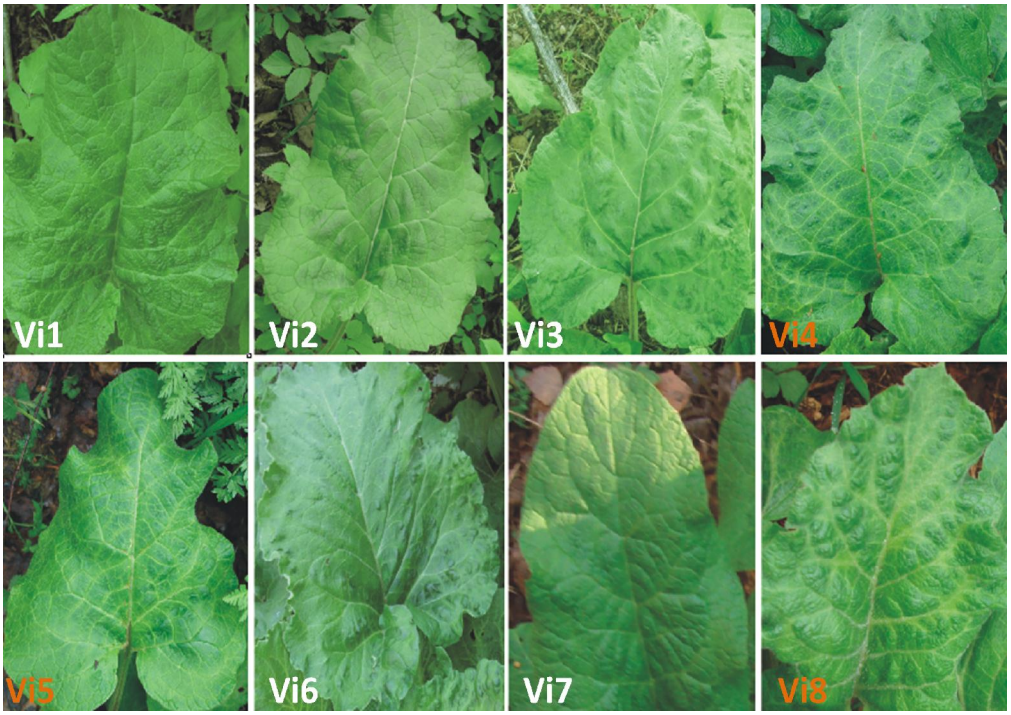


Figure 10. Leaves of *A.tomentosum* plants Vi1-8. Sampled in Viikki on 18th May, 2011. Red color indicates that the leaf was infected with the FMV-Hel.

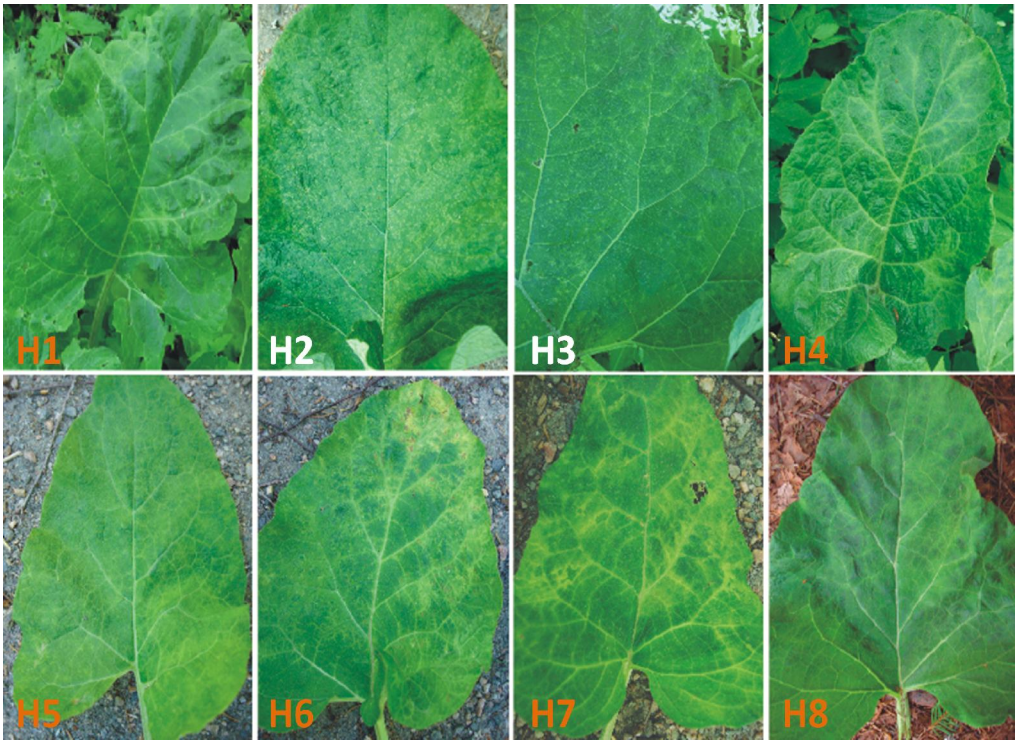


Figure 11. Leaves of *A.tomentosum* plants H1-8. Sampled in Hämeenlinnanväylä on 3rd June, 2011. Red color indicates that the leaf was infected with the FMV-Hel.

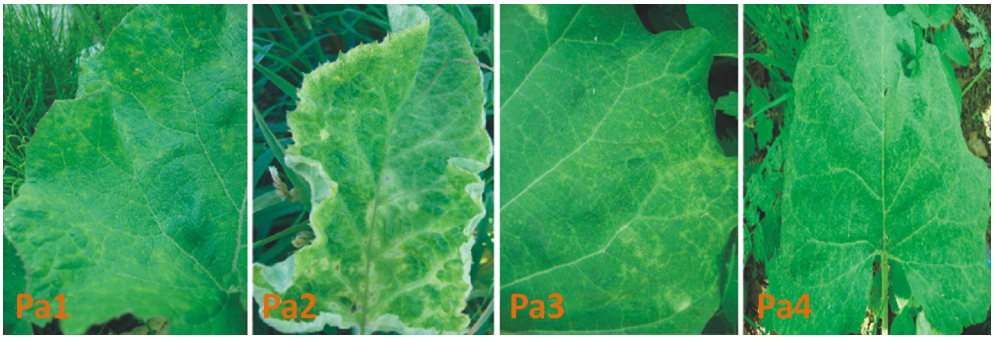


Figure 12. Leaves of *A.tomentosum* plants Pa1-4. Sampled in Pasila on 6th June, 2011. Red color indicates that the leaf was infected with the FMV-Hel.

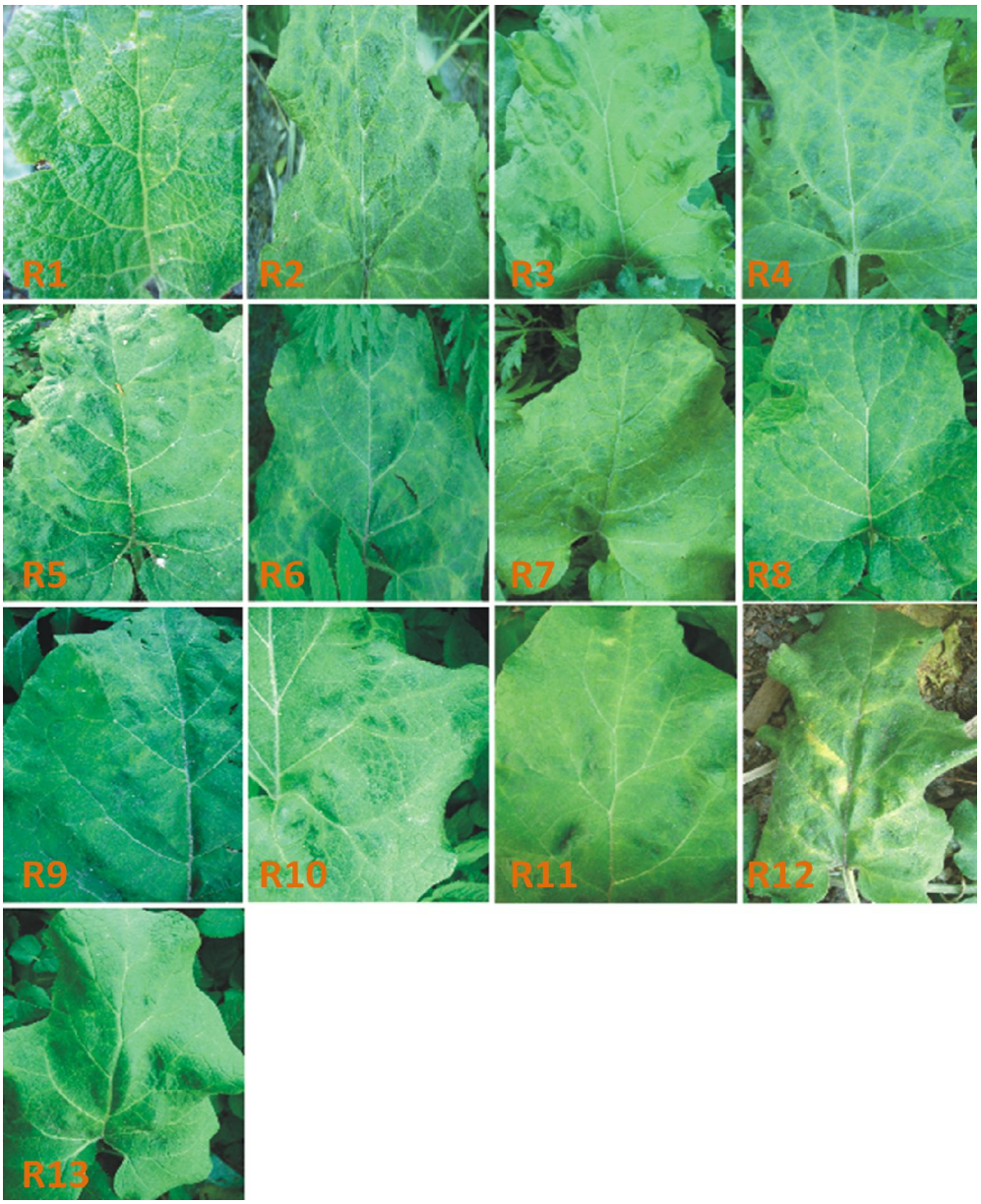


Figure 13. Leaves of *A.tomentosum* plantsR1-13. Sampled in Rapakiventie on 26th May, 2011. Red color indicates that the leaf was infected with the FMV-Hel.

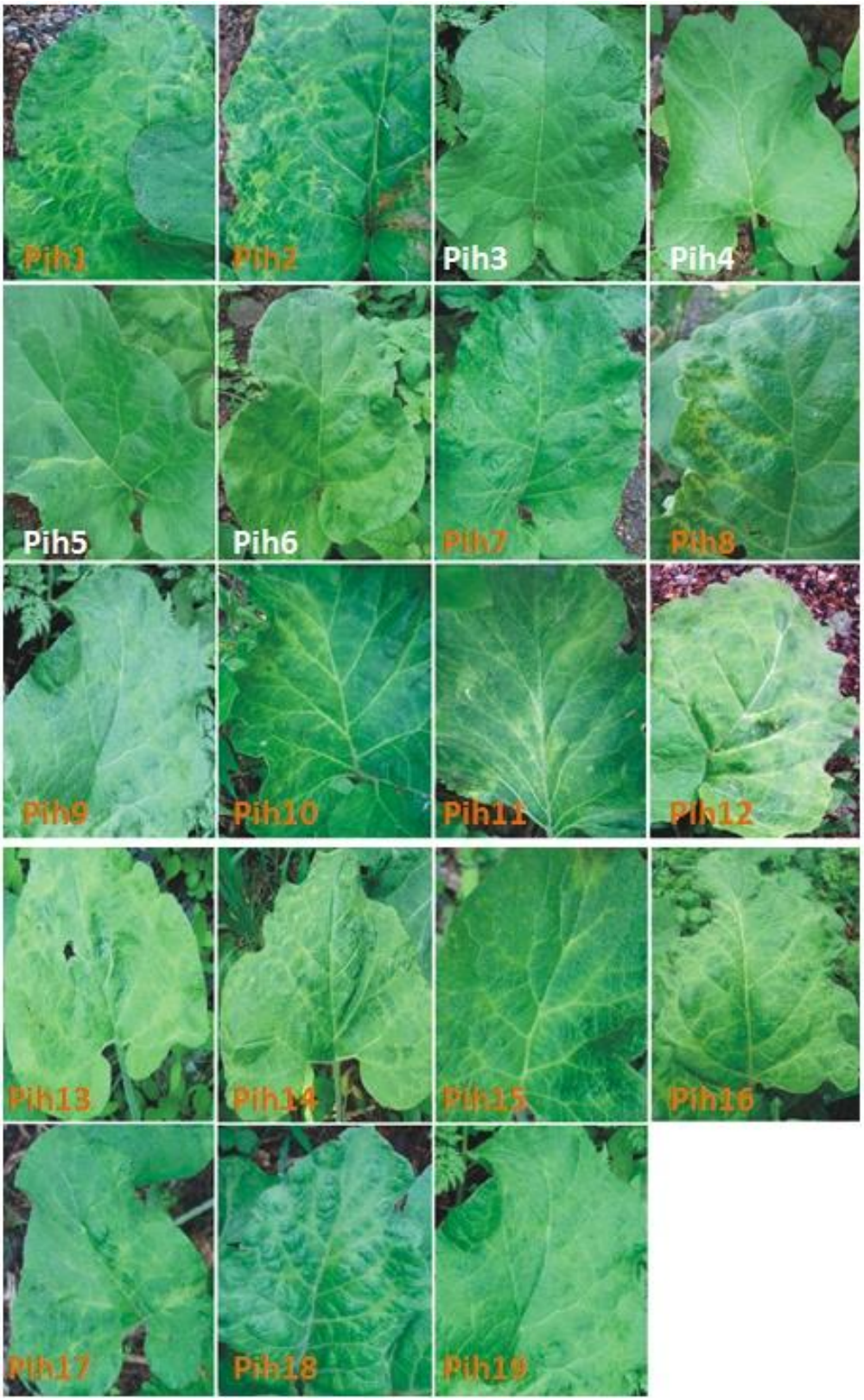


Figure 14. Leaves of *A.tomentosum* plants Pih1-19. Sampled in Pihlajamäki on 16th May, 2011. Red color indicates that the leaf was infected with the FMV-Hel.

In order to study the differences of nucleotide sequences between virus isolates collected from different places, PCR products of two samples from each sampling site were randomly chosen and sent to Haartman Institute for sequencing. The viral sequences from different individual samples were identical, which suggested that the FMV-Hel in Helsinki might be the same strain (Appendix 5).

5 DISCUSSION

AlsVX had only been isolated previously from *Alstroemeria* (*Alstroemeria* spp., Alstroemeriaceae) in Japan. This study is the first report of the virus occurrence outside Japan. The viral isolate detected in *A.tomentosum* in Helsinki showed 82% CP nucleotide sequences identity and 90% CP amino acid sequence identity with the isolate described in Japan, which is well above the species demarcation limit (76.3% nt identity and 88.8% aa identity) in genus *Potexvirus* (Adams *et al.*, 2004). AlsVX was first described in 2005 (Fuji *et al.*, 2005) from an alstroemeria plant (cultivar Virginia) in Hokkaido. The virus caused systemic symptoms of mosaic and necrotic streaks in *Alstroemeria*. AlsVX was placed into family Flexiviridae, genus *Potexvirus*, according to particle morphology and physical properties. It has flexuous viral particles which resemble potexviruses, and contains a 3'-terminal poly(A) tail. The genomic RNA of AlsVX-Jap is 7009 nucleotides excluding the poly(A) tail. As other members of the genus, it contains 5 open reading frames (ORFs) (Figure 6) (Fuji *et al.*, 2005).

Fuji *et al.* (2005) concluded that AlsVX was closely related to NMV and ScaVX. This suggests that reads and contigs of AlsVX-Hel are possible to be aligned to the conserved region of sequences of potexviruses (Table 5; Table 6). The motif "GGAAAA" of AlsVX is identical to ScaVX (Fuji *et al.*, 2005). The motif had been proposed to be an element of RNA replication or protein translation, and it is in the beginning of the AlsVX's 5' untranslated region (97 nt of 5'-UTR). Other conserved motifs in potexviruses such as "ACUUAA", which is supposed to be a *cis*-acting element affecting viral RNA synthesis, and a putative polyadenylation signal sequence "AAUAAG" (Cotillon *et al.*, 2002), were found in the 3'-UTR (nt 6955-6960) of AlsVX (Fuji *et al.*, 2005). Comparison of amino acid similarity and identity of AlsVX-Jap and other potexvirus demonstrated that ORFs 2, 3 and 5 of AlsVX shared the highest similarity and identity with the corresponding ORFs of NMV, and ORFs 1 and 4 were highly similar to ScaVX.

With limited information about AlsVX, it is hard to explain why the virus existed in *A.tomentosum* in Helsinki. In a total of 58 tested plants, only one plant contained AlsVX.

This study is also the first report of the occurrence of the FMV-Hel in Finland. FMV has been reported in Italy, Canada, United States, Turkey, Iran and Egypt since it was characterized in 2009. The FMV-Hel which exists in *A.tomentosum* was supposed to be FMV because it has 78% nucleotide identity with other FMV strains (Table 8).

However, a recently characterized negative-strand RNA virus (RRV) is also closely similar to the FMV-Hel according to the results of BLASTx. Four of the genomic RNAs of RRV share 68%, 51%, 60%, 59% identity with the corresponding FMV RNAs, and the RNA1 and RNA2 share 49 % and 40 % identity with the EMARAV respectively (Laney *et al.*, 2011). Therefore, RRV was placed into genus *Emaravirus*, which already contains EMARAV, FMV, *Maize red stripe virus* (MRSV) and *Pigeonpea sterility mosaic virus* (PPSMV) (http://www.dpvweb.net/notes/showgenusmembers.php?genus=Emaravirus). RRV causes excessive lateral shoot growth, excessive thorniness, witches' broom, leaf proliferation and malformation, mosaic, red pigmentation and plant death in *Rosa multiflora* in USA.

The FMV-Hel was mostly found in *A.tomentosum* growing at the edges of woodland and road. The virus transmission mechanism of FMV is not clear yet, but according to the results of this study, two means of transmission can be suggested. One presumption is that the virus could be transmitted by a vector. Elberaino *et al.* (2009) indicated that viruses from genus *Emaravirus* are all vectored by eriophyid mites. Eriophyid mites are among the smallest mites, with less than 0.3 mm long (Graham, 2004). The tiny size facilitates eriophyid mites travelling between hosts by wind, water, animals and humans.

The second possibility is that the FMV-Hel is seed-borne, since isolates from different areas of Helsinki are highly identical (Appendix 5). The seeds of the plant mature in the second growth year after flowering and are equipped with hooked bristles, which help seeds to adhere to people's clothes and animals' skin for long distance dispersal.

Deep-sequencing of siRNA has been proven to be very useful for detection of unknown viruses, strains and isolates. It can be used not only for viral detection, but also for recovery of viral genomic RNA sequences. Kreuze *et al.* (2009) successfully recovered viral genomes by assembling contigs from siRNA. However, in this study, neither AlsVX-Hel nor FMV-Hel viral genomes were recovered completely (Appendixes 6 & 7). The reason might be that less siRNA than expectation was obtained. It is possible that the plants did not produce siRNA abundantly, or the protocol used is not good enough for siRNA extraction from *A.tomentosum*.

Analysis with Velvet program indicated that the largest numbers of siRNA reads aligned to AlsVX, which means that AlsVX should have the highest concentration in plant samples. However, by RT-PCR experiments, AlsVX was detected from only one

sample by nested-PCR after RNA was reverse transcribed. In comparison, FMV-Hel which had lower alignment scores was detected in all individual samples by normal RT-PCR. The conflict between computational analysis and laboratory practice might be that siRNAs derived from AlsVX were more abundant and the viral RNA was highly degraded in the plant, whereas the case with the FMV-Hel was opposite to AlsVX. Therefore, in siRNA deep-sequencing results, the siRNA of AlsVX were in higher concentration than FMV-Hel.

MAQ and Novoalign are efficient tools for aligning siRNA with existing viral genomes. However, they are not powerful enough if the tested sequence greatly differs from the reference sequence. In this study, FMV-Hel differed from other FMV strains. The differences between the reference sequence and the sequence of the virus in the sample also increased the difficulties in primer design, so the whole region of the FMV putative nucleocapsid protein gene could not be amplified. The amplified partial fragment of FMV-Hel putative nucleocapsid gene (525 nt) was compared with FMV-Can01, FMV-Gr10 and FMAV-JJW2008 (Appendix 4). The differences (75% -78% at nucleotide level) suggested that the FMV-Hel in Helsinki is a new strain of FMV, or a new virus.

Research on discovery of viruses from wild plants enriches the knowledge on viral biodiversity. Deep-sequencing of vsiRNA is simple to be conducted and effectively works for viral detection from wild plants. However, the limitation is that it can be used only for detecting viruses which are related to known viruses, because known sequences are necessary to be used as reference. For detecting a novel virus completely unrelated to known viruses, this method is not suitable.

6 CONCLUSIONS

siRNA deep-sequencing method was successfully applied to virus detection in diseased plants of *A.tomentosum* in Helsinki. The plants growing in nature with yellow vein and mosaic symptoms contained two viruses: *Alstroemeria virus X* and *Fig mosaic virus*-Hel, which have never been reported in Finland. AlsVX-Hel only existed in one plant sampled in Pihlajamäki; however, the FMV-Hel had a wide distribution in Helsinki.

This study also revealed differences between the known FMV strains and the FMV-Hel detected in Helsinki. The partial sequence of the putative nucleocapsid protein gene was different from other FMV strains, which hinted that the FMV-Hel might be a new strain of FMV, or a new virus.

7 ACKNOWLEDGEMENTS

I would like to express my sincerest thanks to Professor Jari Valkonen, who offered me such a good opportunity to work in this field, and supervised me during the whole research period including research preparation, experimental practices and thesis writing.

Thanks to Jie Xiong and MSc. Hanna Susi, who gave me a lot of guidance in siRNA data analysis.

Thanks to botanist Ilpo Kuokka, who helped me to confirm the identification of *A.tomentosum*, and also helped me to understand some references in Swedish.

I gratefully acknowledge Dr. Minna Rajamäki and Dr. Shadid Siddiqui, researchers Yanping Tian, Anssi Vuorinen, Johanna Santala, Mikko Lehtonen, Xianbao Deng and Katrin Wiegmann in KPAT lab for help to all of my questions, both theoretical and practical.

Finally, thanks to my parents for all supports during my whole study life.

8 REFERENCES

Adams, M.J., Antoniw, J.F., Bar-Joseph, M., Brunt, A.A., Candresse, T., Foster, G.D., Martelli, G.P., Milne, R.G, Zavriev, S.K. & Fauquet, C.M. 2004. The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation. Archives of Virology 149: 1045-1060.

Agrama, H.A. & Scott, J.W. 2006. Quantitative trait loci for Tomato yellow leaf curl virus and Tomato mottle virus resistance in tomato. Journal of the American Society for Horticultural Science 131 (2): 267-272.

Agrios, G.N. 2005. Plant Pathology. Fifth Edition. Elsevier Academic Press. San Diego. CA.

Allen, G. C., Flores-Vergata, M. A., Krasynanski, S., Kumar, S. & Thomson, W. F. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nature Protocols 1: 2320-2325.

Ball, L.A. 2007. Virus replication strategies. In: Knipe, D.M. & Howley, P. M. (ed.). Fields Virology, Volume One. Fifth Edition. PA, USA: Lippincott Williams & Wilkins, a Wolters Kluwer Business. pp. 135.

Baulcombe, D. 2004. RNA silencing in plants. Nature 431: 356.

Baumberger, N., Tsai, C.H., Lie, M., Havecker, E., & Baulcombe, D.C. 2007. The polerovirus silencing suppressor P0 targets Argonaute proteins for degradation. Current Biology 17: 1609–1614.

Bolduc, F., Hoareau, C., St-Pierre, P. & Perreault, J.P. 2010. In-depth sequencing of the siRNAs associated with peach latent mosaic viroid infection. BMC Molecular Biology 11: 16.

Bos, L. 1995. The embryonic beginning of virology: unbiased thinking and dogmatic stagnation. Archives of Virology 140: 613-619.

Briddon, R. W. & Markham, P. J. 2000. Cotton leaf curl virus disease. Virus Research 71: 151-159.

Chen, H.M., Chen, L.T., Patel, K., Li, Y.H., Baulcombe, D.C. & Wu, S.H. 2010. 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proceedings of the National Academy of Science of the United States of America 107: 15269–15274.

Chen, X. 2005. MicroRNA biogenesis and function in plants. FEBS Letters 579:5923-5931.

Clark, M.F. & Adams, A.N. 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. Journal of General Virology 34 (3): 475-483.

Coetzee, B., Freeborough M.J., Maree, H.J., Celton, J.M., Rees, D.J.G. & Burger, J.T. 2010. Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. Virology 400 (2): 157-163.

Cotillon, A.C., Girard, M. & Ducouret, S. 2002. Complete nucleotide sequence of the genomic RNA of a French isolate of *Pepino mosaic virus* (PepMV). Archives of Virology 147:2231–2238.

Den Virtuella Floran. Arctium tomentosum Mill. http://linnaeus.nrm.se/flora/di/astera/arcti/arcttomv.jpg ; accessed 29, 11, 2011.

Descriptions of plant virus. Members of Genus: Emaravirus. http://www.dpvweb.net/notes/showgenusmembers.php?genus=Emaravirus; accessed 29.09.2011.

Descriptions of Plant Viruses. List of Families sorted by name. http://www.dpvweb.net/notes/families.php; accessed 17.11.2011.

Di Serio, F., Gisel, A., Navarro, B., Delgado, S., Martinez de Alba, A., Donvito, G. & Flores, R. 2009. Deep sequencing of the small RNAs derived from two symptomatic variants of a chloroplastic viroid: implications for their genesis and for pathogenesis. PLoS One 4(10): e7539.

Dunoyer, P., Himber, C. & Voinnet, O. 2005. Dicer-like 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. Nature Genetics 37: 1356-1360.

Dunoyer, P., Schott, G., Himber, C., Meyer, D., Takeda, A., Carrington, J.C. & Voinnet, O. 2010. Small RNA duplexes function as mobile silencing signals between plant cells. Science 328: 912–916.

Elbeaino, T., Digiaro, M. & Martelli, G. P. 2009. Complete nucleotide sequence of four RNA segments of fig mosaic virus. Archives of Virology 154: 1719-1727.

Eun, A.J.C. & Wong, S.M. 1999. Detection of cymbidium mosaic potexvirus and odontoglossum ringspot tobamovirus using immuno-capillary electrophoresis. Phytopathology 89: 522-528.

Eun, A.J.C. & Wong, S.M. 2000. Molecular beacons: a new approach to plant virus detection. Phytopathology 90: 269-275.

Flora Italiana. Plant profile: Arctium tomentosum. http://luirig.altervista.org/schedeit/ae/arctium_tomentosum.htm; accessed 17. 11. 2011.

Fraenkel-Conrat, H. 1956. The role of the nucleic acid in the reconstitution of active Tobacco mosaic virus. Journal of the American Chemical Society 78: 882-883.

Fraenkel-Conrat, H., & Williams, R.C. 1955. Reconstitution of active Tobacco mosaic virus from its inactive protein and nucleic acid components. Proceedings of the National Academy of Sciences of the United States of America 41: 690-698.

Fuji, S., Shinoda, K., Ikeda, M., Furuya, H., Naito, H & Fukumoto, F. 2005. Complete nucleotide sequence of the new potexvirus “Alstroemeria virus X”. Achieve of Virology 150: 2377-2385.

Fukuta, S., Kato, S., Yoshida, K., Mizukami, Y., Ishida, A., Ueda, J., Kanbe, M. & Ishimoto, Y. 2003. Detection of Tomato yellow leaf curl virus by loop-mediated isothermal amplification reaction. Journal of Virological methods 112: 35-40.

Fukuta, S., Yoshida, K., Mizukami, Y., Ishida, A. & Kanbe, M. 2004. Development of immunocapture reverse transcription loop-mediated isothermal amplification for the detection of Tomato spotted wilt virus from chrysanthemum. Journal of Virological methods 121: 49-55.

Fulton, J.P., Gergerich, R.C. & Scott, H.A. 1987. Beetle transmission of plant viruses. Annual Review of Phytopathology 25 (1): 11-23.

Fusaro, A.F., Matthew, L., Smith, N.A., Curtin, S.J., Dedic-Hagan, J., Ellacott, G.A., Watson, J.M., Wang, M.B., Brosnan, C., Carroll, B.J. & Waterhouse, P.M. 2006. RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. EMBO Reports 7: 1168-1175.

Ghedin, E., Laplante, J., DePasse, J., Wentworth, D.E., Santos, R.P., Lepow, M.L., Porter, J., Stellrecht, K., Lin, X., Operario, D., Griesemer, S., Fitch, A., Halpin, R.A., Stockwell, T.B., Spiro, D.J., Holmes, E.C. & St. George, K. 2011. Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of Oseltamivir resistance. Journal of Infectious Disease 203:168–174.

Gibson, R.W. & Aritua, V. 2002. The perspective of Sweetpotato chlorotic stunt virus in sweetpotato production in Africa: a review. African Crop Science Journal 10: 281-310.

Gilbertson, R.L. & Lucas, W.J. 1996. How do viruses traffic on the ‘vascular highway’? Trends in Plant Science 1: 260-268.

Graham, J. 2004. Ornamental plant damage by eriophyid mites. University of Nevada. Cooperative Extension.

Greninger, A.L., Chen, E.C., Sittler, T., Scheinerman, A., Roubinian, N., Yu, G., Kim, E., Pillai, D.R., Guyard, C., Mazzulli, T., Isa, P., Arias, C.F., Hackett Jr, J., Schochetman, G., Miller, S., Tang, P., & Chiu, C.Y. 2010. A metagenomic analysis

of pandemic influenza A (2009 H1N1) infection in patients from North America. PLoS One 5: e13381.

Gutierrez, C. 1999. Geminivirus DNA replication. Cellular and Molecular Life Sciences 56: 313-329.

Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L., Fazekas, A.J., Graham, S.W., James, K.E., Kim, K.J., Kress, W.J., Schneider, H., van AlphenStahl, J., Barrett, S.C.H., van den Berg, C., Bogarin, D., Burgess, K.S., Cameron, K.M., Carine, M., Chacón, J., Clark, A., Clarkson, J.J., Conrad, F., Devey, D.S., Ford, C.S., Hedderson, T.A.J., Hollingsworth, M.L., Husband, B.C., Kelly, L.J., Kesanakurti, P.R., Kim, J.S., Kim, Y.D., Lahaye, R., Lee, H.L., Long, D.G., Madrinán, S., Maurin, O., Meusnier, I., Newmaster, S.G., Park, C.W., Percy, D.M., Petersen, G., Richardson, J.E., Salazar, G.A., Savolainen, V., Senerg, O., Wilkinson, M.J., Yi, D.K. & Little, D.P. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences of United States of America 106: 12794–12797.

Hull, R. 2002. Matthews' Plant Virology. 4th Edition. Academic Press, Bath, UK.

Hutvagner, G. & Simard, M.J. 2008. Argonaute proteins: key players in RNA silencing. Nature Reviews Molecular Cell Biology 9: 22-32.

International committee on taxonomy of viruses. Virus taxonomy: 2009 release. http://www.ictvonline.org/virusTaxonomy.asp?version=2009; accessed 17.11.2011.

International committee on taxonomy of viruses. Virus taxonomy assignments. http://ictvonline.org/virusTaxInfo.asp; accessed 17. 11. 2011.

Koo, M., Bendahmane, M., Lettieri, G.A., Paoletti, A.D., Lane, T.E., Fitchen, J.H., Buchmeier, M.J. & Beachy, R.N. 1999. Protective immunity against murine hepatitis virus (MHV) induced by intranasal or subcutaneous administration of

hybrids of Tobacco mosaic virus that carries an MHV epitope. Proceedings of the National Academy of Sciences of the United States of America 96 (14): 7774-7779.

Kreuze, J. F., Perez, A., Untiveros, M., Quispe D., Fuentes, S., Barker, I. & Simon, R. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. Virology 388: 1-7.

Kurihara, Y. & Watanabe, Y. 2004. Arabidoposis micro-RNA biogenesis through Dicer-like 1 protein functions. Proceedings of the National Academy of Sciences of the United States of America 101 (34): 12753-12758.

Laney, .G., Keller, K.E., Martin, R.R. & Tzanetakis, I.E. 2011. A discovery 70 years in the making: characterization of the Rose rosette virus. Journal of General Virology 92 (7): 1727-1732.

Leone, G., van Schijndel, H.B., van Gemen, B. & Schoen, C.D. 1997. Direct detection of Potato leafroll virus in potato tubers by immunocapture and the isothermal nucleic acid amplification method NASBA. Journal of Virological Methods 66: 19-27.

Lewis, K.J., Bendahmane, M., Smith, T.J., Beachy, R.N. & Siuzdak, G. 1998. Identification of viral mutants by mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 95: 8596-8601.

Li, H., Ruan, J. & Durbin, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research 18: 1851-1858.

Li, W.X. & Ding, S.W. 2001. Viral suppressors of RNA silencing. Current Opinion in Biotechnology 12: 150-154.

Ling, K., Fei, Z., Wechter, W.P. & Hernamdez, A.G. 2011. Deep sequencing of small RNAs for virus and viroid identification in tomatoes. Phytopathology 101: S106.

Lucas, W.J. 2006. Plant viral movement proteins: agents for cell-to-cell trafficking of

viral genomes. Virology 344: 169-184.

MAQ User's Manual (Release 0.5.0). Maq: Mapping and Assembly with Qualities. http://maq.sourceforge.net/maq-man.shtml; accessed 17.09.2011.

Margis, R. Fusaro, A.F., Smith, N.A., Curtin, S.J., Watson, J.M., Finnegan, E.J. & Waterhouse, P.M. 2006. The evolution and diversification of Dicers in plants. FEBS Letters 580: 2442–2450.

Matzke, M.A. & Birchler, J.A. 2005. RNAi-mediated pathways in the nucleus. Nature Reviews Genetics 6: 24-35.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J. & Qi, Y. 2008. Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. Cell 133:116-127.

Mossberg, B. & Stenberg, L. 1997. Den Nordiska Floran. Stockholm, Sweden: Wahlström & Widstrand Förlag. 472-473 pp.

Mumford, R.A., Walsh, K., Barker, I. & Boonham, N. 2000. Detection *of Potato mop top virus and Tobacco rattle virus* using a multiplex real-time fluorescent reverse-trasncription polymerase chain reaction assay. Phytopathology 90: 448-453.

Napoli, C., Lemieux, C. & Jorgensen. R. 1990. Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous gene *in trans*. The Plant Cell 2: 279-289.

Navarro, B., Pantaleo, V., Gisel A., Moxon, S., Dalmay, T., Bisztray, G., Di Serio, F. & Burgyán, J. 2009. Deep sequencing of viroid-derived small RNAs from grapevine provides new insights on the role of RNA silencing in plant-viroid interaction. PLoS One 4: e7686.

Nie, X. 2005. Reverse transcription loop-mediated isothermal amplification of DNA for detection of Potato virus Y. Plant Diseases 89: 605-610.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, K. &

Hase, T. 2000. Loop-mediated isothermal amplification of DNA. Nucleic Acids Research 28 (12): e63.

Pennisi, E. 2007. Wanted: a barcode for plants. Science 318: 190-191.

Stynen, D., Goris, A., Sarfati, J. & Latge, J.P. 1995. A new sensitive sandwich enzyme-linked immunosorbent assay to detect galactofuran in patients with invasive aspergillosis. Journal of Clinical Microbiology 33: 497-500.

Tyagi, S. & Kramer, F.R. 1996. Molecular beacons: probes that fluoresce upon hybridization. Nature Biotechnology 14: 303-308.

Voinnet, O. 2008. Post-transcriptional RNA silencing in plant–microbe interactions: a touch of robustness and versatility. Current Opinion in Plant Biology 11: 464–470.

Wang, C., Mitsuya, Y., Gharizadeh, B., Ronaghi, M. & Shafer, R.W. 2007. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Research 17:1195-1201.

Wang, X.B., Jovel, J., Udomporn, P., Wang, Y., Wu, Q., Li, W.X., Gasciolli, V., Vaucheret, H. & Ding, S.W. 2011. The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative Argonautes in Arabidopsis thaliana. Plant Cell 23: 1625-1638.

Wong, S.M. 2002. Recent advances in plant virus detection. Acta Horticulturae (ISHS) 568: 113-116.

Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E.C., Li, W.X. & Ding, S.W. 2010. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. Proceedings of the National Academy of Sciences of the United States of America 107 (4): 1606-1611.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., & Carrington, J.C. 2004. Genetic and functional diversification of small RNA pathways in plants. PLoS Biology 2 (5): e104.

Yao, Z.P., Demirev, P.A. & Fenselau, C. 2002. Mass spectrometry-based proteolytic mapping for rapid virus identification. Analytical Chemistry 74 (11): 2529-2534.

Yu, B., Chapman, E. J., Yang, Z., Carrington, J. C., & Chen, X. 2006. Transgenically expressed viral RNA silencing suppressors interfere with microRNA methylation in Arabidopsis. FEBS Letters 580: 3117-3120.

Zaitlin, M. 1998. The discovery of the causal agent of the tobacco mosaic disease. In: Kung, S.D. & Yang, S.F. (eds). Discoveries in Plant Biology. World Scientific Publishing Co. Pte. Ltd. Singapore. pp. 106-110.

Zerbino, D. 2008. Velvet Manual - version 1.1. http://helix.nih.gov/Applications/velvet_manual.pdf ; accessed 15.09.2011.

Zerbino, D.R. & Birney, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18: 821-829.

APPENDIX 1:

INGREDIENTS OF BUFFERS AND SOLUTIONS

Extraction buffer for DNA (250 ml)

25 ml Tris-HCl, pH 8.0 (1 M stock)

70 ml Sodium chloride (5 M stock)

10 ml Ethylenediaminetetraacetic acid (EDTA) (0.5 M stock)

5 g Cetyl trimethylammonium bromide (CTAB)

Add sterile H_2O to 250 ml

0.5-1% β-mercaptoethanol (add just before use)

Tris-ethylenediaminetetraacetic acid (TE) buffer (1L)

10 ml Tris-HCl, pH 8.0 (1 M stock)

2 ml Sodium EDTA (0.5 M stock)

Add sterile H_2O to one litre.

Homemade Trizol (1L)

380 ml Phenol pH 4.3 (38%, Sigma P4682)

118.6 g Guanidine thiocyanate (Sigma G9277 or Fluka 50981)

76.12 g Ammonium thiocyanate (0.4 M, Fluka O9938)

33.4 ml Sodium acetate pH 5.0 (3 M stock)

50 ml Glycerol

Add sterile H_2O to one litre.

APPENDIX 2:

COMPARISON OF THE COAT PROTEIN GENE BETWEEN AlsVX-Hel AND AlsVX-Jap

```
AlsVX-Hel       TGGGACAAGACTATTTCTTACTTCCGCCCTTCTGACCCTATCACCCACTCAAATAAATGG 60
AlsVX-Jap       GGCACCAAAACTGTCTCTTACTTTCGCCCTTCTGACCCTATCACCCACTCTAATAAATGG 60
                 * ..***.***.* ******** **************************:*********

AlsVX-Hel       CTGGCCTTCTCATCAGTCCTCACCATCAGCCTCATCATTTGGCTTTGCTCTAAATTTTCT 120
AlsVX-Jap       CTGGCCTTCTCATCAGTCATCACCATCAGCCTCATCATTTGGCTATGTTCTAAATTTTCT 120
                ******************.**************************:** ************

AlsVX-Hel       CCTGGTGATCGGCGGCCTCTACCTCCTGTATGTACTCACTGCCCCAGTCGCTAGAATTGG 180
AlsVX-Jap       CCTGGTGATAGGCGGCCTCTACCTCCTGTATGCTCTCACTGCACCTCCCACTAAGGCCGG 180
                *********.********************** :********.**:  *.***...  **

AlsVX-Hel       GTGCGTGGTCCGTGTTACCGGAGAAAGTGTGACAATAGATAATTGCCCGGACACGGCCCA 240
AlsVX-Jap       TTGCATCGTACGCGTGACGGGCGAAAGTGTGACCATAGATAATTGCCCGGACACAGCTTT 240
                 ***.* **.** ** ** **.***********.*******************.**  :

AlsVX-Hel       AATAATGGAAAAGTTTAACCTCGCTCCTTGGAACGGGGTTAAGTTTCCAAAAGATTGAAA 300
AlsVX-Jap     AATAATGGAAAAATTCAATCTCGCTCCTTGGAACGGGGTTAAGTTTCCAAAAGAATGAAA 300
                ************.** ** **********************************:******

                                                      ↓Coat protein start
AlsVX-Hel       ATTTTACT---ATACCT--CACCT-CCAATTACCTATGGCAGGAGATAACCCTAACCCCC 354
AlsVX-Jap       ATTTTAATTATAAACCAGACACCTTTCTTAATTAAATGGCAGGAGAAAACACACCACCCA 360
                ******.*   *:***:  *****  *::::: .:***********:***.*:...***.

AlsVX-Hel       AAGCAACCAAAGCACCCACCGGGTCAGCAGCCGACCTCGCCAACCCCAACCGCAGCCCCA 414
AlsVX-Jap       CCCCCCCCGACAGCACCAGCCGGCTCCGCTGCCGACCTGGCTAACCCCAACCGCAGCCCCA 420
                .. *..* *.******..**** **.**:******** ** ******************

AlsVX-Hel       CCGTAGCTGACCTGGCCCAACTCAAGTACCAATCGACTACCACATCAGTTGCCACACCCC 474
AlsVX-Jap       CAGTTTCCGACCTAGCTAAACTCAACTACCAATCGACCACCACCTCAGTCGCCACCCCAG 480
                *.**: * *****.** .******* ********** *****.***** *****.**.

AlsVX-Hel       AAGAAATAGCCGCACTCGGTGTACTATTCACCAACGCCGGCGTCGACCCCAACTCTGTAC 534
AlsVX-Jap       AGGAGATCACTGCCCTTGGTGTGCTCTTCACTAGCGCCGGCGTAGACCCCAACTCAGTTC 540
                *.**.**..* **.** *****.**.***** *.*********.**********:**:*

AlsVX-Hel       CCCTCGCCATGTGGGACCTTGCCCGCGCCTACGCTGACGTCCAGTCAAGCCGCAGTGCTA 594
AlsVX-Jap       CACTGGCCATGTGGGATTTAGCTCGGGCCTATGCTGACGTTCAGTCCAGCCGCAGTGCCA 600
                *.** **********  *:** ** ***** ******** *****.********** *

AlsVX-Hel       TGCTCGTAGGTTCCACCCCTTCCAACTCAGCCATCACTCGCCAAACTCTAGCCCGGAAAC 654
AlsVX-Jap       TGCTAGTGGGGTCCACTCCATCCAACAAATCCATCACCCGTCAGACCCTTGCCCGCAAGC 660
                ****.**.** ***** **:******:.* ******* ** **.** **:***** **.*

AlsVX-Hel       TCGAACAGGTCAACATCACCCCCCGCCAGTTCTGCATGTACTTCGCTAAGGTAGTGTGGA 714
AlsVX-Jap       TGGATCAAGTCAACATCACGCCACGCCAATTCTGCATGTACTTCGCAAAAGTCGTGTGGA 720
                * **:**.********** **.*****.*****************:**.**.*******

AlsVX-Hel       ATATCCTCATCACCTCCAACGTAGCACCTGCTGGTTGGGCTAAACTAGGGTACCCTGAGG 774
AlsVX-Jap       ATATCCTGATCAGCAGCAACACCGCACCCGCCGGCTGGGCAAAGTTGGGGTACCCCGAAG 780
                ******* **** *: ****. .***** ** ** *****:**. *.******** **.*

AlsVX-Hel       AGTGCAAATTCGCCGCTTTCGATTTCTTCGACGGCGTCCTATCCCCGGCAGCACTCGAAC 834
AlsVX-Jap       AATGCAAGTTCGCTGCATTCGACTTCTTCGACGGTGTCCTCTCCCCAGCCGCACTTGATC 840
                *.*****.***** **:***** ********** *****.*****.**.***** **:*

AlsVX-Hel       CCGCTGATGGCCTTGTTCGACAACCCAACACCCGTGAGATACAGGCACATTCCACCGCCA 894
AlsVX-Jap       CCGCTGACGGGCTCATCAGACAACCCAACACCCGGGAGATCCAAGCCCACTCAACGGCTA 900
                ******* ** ** .* .**************** *****.**.**.** **.** ** *

AlsVX-Hel       AATTTGGAGCTCTAGCTCGCCAACGCATCAGCAATGGCAACTATGTCTCCACTCTTGCCG 954
AlsVX-Jap       AGTTTGGAGCTCTCGCCCGCCAACGCATCAGCAATGGCAATTTCGTCTCAACACTCGCCG 960
                *.**********.** ********************** *: *****.**:** ****
```

```
AlsVX-Hel       AAGTCACCCACGGCAGAGTTGGTGGGGTCAATGCCATGTACTCCATCGAAGCCCCCCCAG 1014
AlsVX-Jap       AGGTCACCCACGGCAGAGTCGGAGGTGTCAATGCCATGTACTCCATCGAAGCCCCCCCAG 1020
                *.***************** **:** **********************************

                       ↓Coat protein end
AlsVX-Hel       AAGCCTAAGGCCTTTCTCCTTGCCCAACTTAAATAGAGATAGAACTGGTGTGGTGTGTGT 1074
AlsVX-Jap       AAGCCTAAG-CCTTTCTCCTTGCCCAACTTAAATA-AGGAGGTGTTTTTCTACAGT-TTT 1077
                ********* ************************ **.:.*:. *  * *. :** * *

AlsVX-Hel       GGAGAAGCCAGATTAAGTCGAAAGCGAT 1102
AlsVX-Jap       ATTTTCACTAATATACTATGTGTGTGTT 1105
                . : :..* *.::**. : *:.:* *:*
```

APPENDIX 3:

ACCESSION NUMBERS OF THE VIRAL SEQUENCES AND PROTEINS OBTAINED FROM THE NCBI DATABASE

Viruses	Accession No. of nucleotides	Accession No. of proteins
Alstroemeria virus X-Japan (AlsVX-Jap)	AB206396.1	BAE44215.1
Narcissus mosaic virus (NMV)	AY225449.1	AAP51016.1
Scallion virus X (ScaVX)	AJ316085.1	CAC87090.1
Asparagus virus 3 (AV3)	AB304848.1	BAG12162.1
Chenopodium mosaic virus X (CMVX)	DQ660333.1	ABG48664.1
Pepino mosaic virus (PeMV)	AY509926.1	AAS79813.1
Lettuce virus X (LeVX)	AM745758.1	CAN88812.1
Fig mosaic virus-Can01 (FMV-Can01)	HQ703345.1	AEI98678.1
Fig mosaic virus-Gr10 (FMV-Gr10)	FM991954.1	CAX21211.1
Fig mosaic-associated virus (FMAV-JJW2008)	FJ211073.1	Not accessible
Rose rosette virus (RRV)	HQ891906.1	ADZ55370.1
European mountain ash ringspot-associated virus (EMARAV)	GU563319.1	ADM35031.1
Maize red stripe virus (MRSV)	DQ324466.1	ABC58222.1

APPENDIX 4:

COMPARISON OF AMPLIFIED PARTIAL SEQUENCE OF NUCLEOCAPSID GENE OF FMV-Hel WITH FMV STRAINS Gr10, Can01 AND WITH FMAV JJW2008

```
FMV_Gr10          GTTGTTAAAGAGCTTAATGATAGTACAG---------TTGATAATGTTGTTTCATTTAAC 51
FMAV_JJW2008      GTTGTTAAAGAGCTTAATGATAGTACAG---------TTGACAATGTTGTTTCCTTTAAC 51
FMV_Can01         GTTGTAAAGGACATGACAGAAAGTAACGTTAAGAATGTTGATAATGTTGTTTCTTTTAAC 60
FMV-Hel           GTTGTAAAGGACATGACAGAAAGTAACG---------TTAAGAATGTTGTATCCTTCAAT 51
                  *****:**.** .* *.:**:****..*         **.* ********:** ** **

FMV_Gr10          AAAGCATGTGCTATCATGTCAGCTGGTATACTAAAACATACCTTTGATGAGGAATTTGAC 111
FMAV_JJW2008      AAAGCATGTGCTATCATGTCAGCTGGTATACTCAAGCATACCTTTGATGAGGAATTTGAC 111
FMV_Can01         AAAGCATGTGCTATCATGTCAGCTGGTATACTCAAACATACCTTTGATGAGGAATTTGAC 120
FMV-Hel           AAAGCATGTGCTATTATGTCAGCTGGTATACTCAAACACACTTTCGACGAGGAATTCGAC 111
                  ************** *****************.**.** ** ** ** ******** ***

FMV_Gr10          TGGAAACTATCAAAATATGTAAAGACGAACAACACTACCAAAGTGATTCCTGATGTGAAG 171
FMAV_JJW2008      TGGAAACTATCAAAATATGTAAAAACGAACAATACTACTAAGGTGATTCCTGATGTAAAA 171
FMV_Can01         TGGAAATTATCAAAATATGTAAAGACGAACAATACTACTAAGGTAGTTCCTGATGTGAAA 180
FMV-Hel           TGGAAGACTGGAAAATATGTTAGTACTGAGAACAAAGCAAAAGTTGTTCCTAATGCTCAG 171
                  *****.  :  *********:*. ** .* ** *.:.* **.** .*****.***  .*.

FMV_Gr10          ATCATAAATCGACTAGCAGGTCAGATGGGCTTGTCAGCTGGTAATCCATATTACTGGATG 231
FMAV_JJW2008      ATCATAAATAGACTAGCGGGTCAGATGGGCTTATCAGCTGGTAATCCATATTACTGGATG 231
FMV_Can01         ATCATAAACAGACTAGCGGGTCAGATGGGCTTATCAGCTGGTAATCCATATTACTGGATG 240
FMV-Hel           ATTATAAACAGACTTGCTGGCCAAATGGGTTTAGCACCTAGTTCACCTTATTACTGGATG 231
                  ** ***** .****:** ** **.***** **. ** **.**:.:**:************

FMV_Gr10          ATTGTGCCAGGTTATGAATTCTTATATGAACTTTATCCAGCTGAGGTATTAGCATACACT 291
FMAV_JJW2008      ATTGTGCCTGGTTATGAATTCTTGTATGAACTGTATCCAGCTGAGGTATTAGCATACACT 291
FMV_Can01         ATTGTGCCTGGTTATGAATTCTTGTATGAACTGTATCCAGCTGAGGTATTAGCATACACT 300
FMV-Hel           ATAGTACCTGGTTATGAATTCTTGTATGAACTGTACCCAGCTGAGGTATTAGCATACACT 291
                  **:**.**:***************.******** ** ************************

FMV_Gr10          CTGGTGAGACTCCAATACAGGAAAAACCTAAACATACCAGATAGTATGACTGATGCTGAT 351
FMAV_JJW2008      CTGGTAAGGCTCCAATACAGGAAAAACCTAAATATACCAGATAGTATGACTGATGCTGAT 351
FMV_Can01         CTGGTAAGGCTCCAATACAGGAAAAACCTAAACATACCAGATAGTATGACTGATGCTGAT 360
FMV-Hel           CTTGTGAGACTTGAATTCAGGAAGAACTTAAACATTCCTGATAAGACAACTGATGCTGAC 351
                  ** **.**.**  ***:******.*** **** **:**:****. * .***********

FMV_Gr10          ATTGTCAGTTCTCTTGTCATGAAAATGAATAGAATTCATAAACTTGAACAGACCAGCTTT 411
FMAV_JJW2008      ATTGTCAGTTCTCTTGTTATGAAAATGAATAGAATTCATAAACTTGAACAGACCAGCTTT 411
FMV_Can01         ATTGTCAGTTCTCTTGTTATGAAAATGAATAGAATTCATAAACTTGAACAGACCAGCTTT 420
FMV-Hel           ATTGTTAGTTCACTTGTTATGAAAATGAACAGGATTCATAAACTTGATGTCACAACATTT 411
                  ***** *****:***** ********** **.**************: : **.* .***

FMV_Gr10          GATGAAGCCCTAAATCTTATAGGCAAGGATAACGTATCCGAAGCTTATGTGGAACTTGCA 471
FMAV_JJW2008      GATGAAGCCCTAAATCTTATAGGCAAGGATAACGTATCTGAAGCTTATGTAGAACTTGCA 471
FMV_Can01         GATGAAGCCCTAAATCTTATAGGCAAGGATAACGTATCTGAAGCTTATGTAGAACTTGCA 480
FMV-Hel           GATGATGCTATCAACACAATAGGTCAGAACAATGTGTCTGATGCTTATATTGAGTTAGCA 471
                  *****:** .*.** . :***** .**.* ** **.** **:******.* **. *:***

FMV_Gr10          AGAGATATAGGAAGCACGAGCAAGACAAAGAGAAATGATGAAGCAATTCTTAAG 525
FMAV_JJW2008      AGAGATATAGGAAGCACGAGCAAGACAAAGAGAAATGATGAAGCAGTTCTTAAG 525
FMV_Can01         AGAGATATAGGAAGCACGAGCAAGACAAAGAGAAATGATGAAGCAATTCTTAAG 534
FMV-Hel           AGAGACATAGGTACTACGAGCAAGAGCAAGAGGAATGATGAAGCTATTATGAAA 525
                  ***** *****:*  ********** .*****.***********:.**.* **.
```

APPENDIX 5:

SEQUENCES OF FMV-Hel ISOLATES FROM DIFFERENT PLANTS OF *A.tomentosum* IN HELSINKI

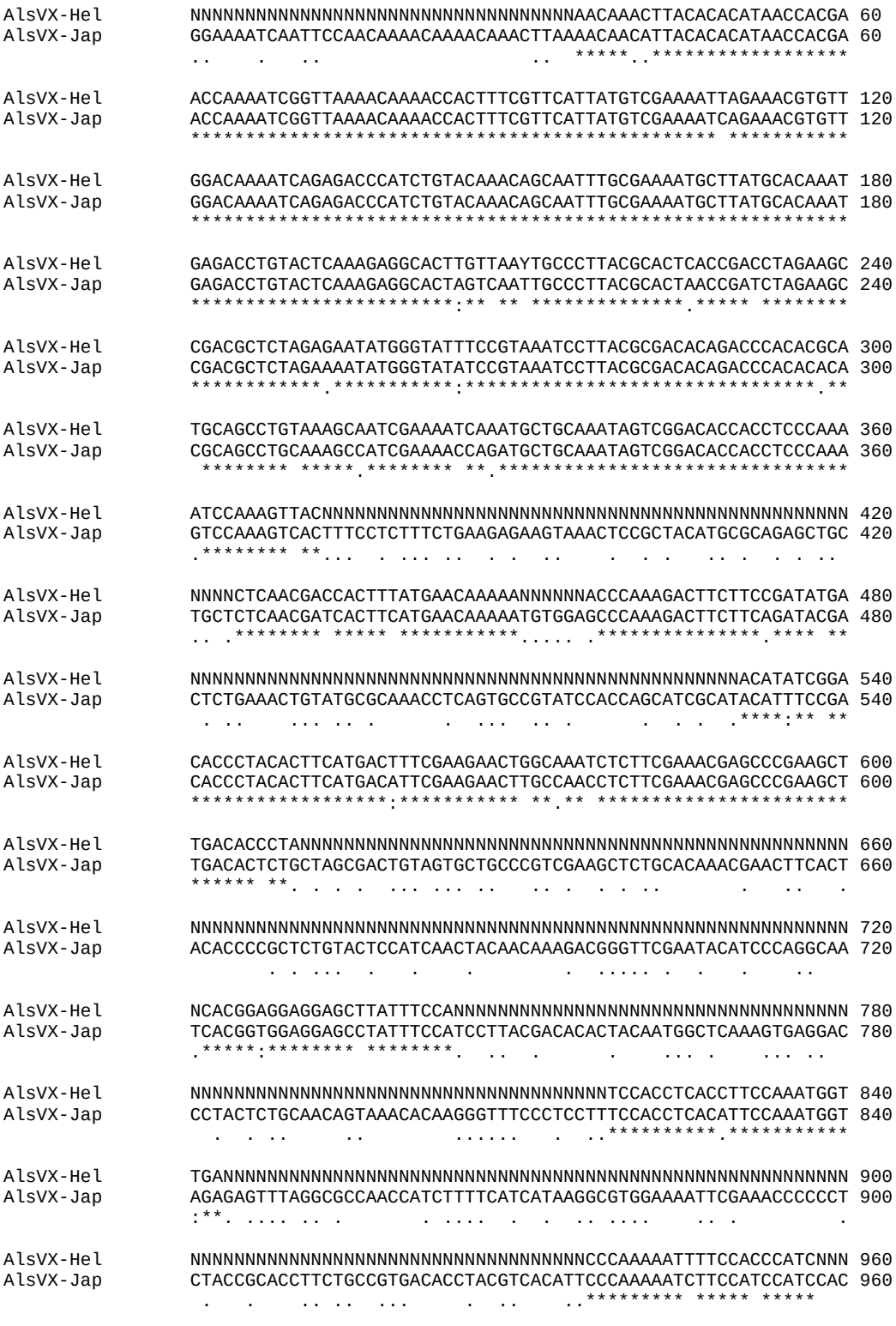
```
H5              CGTTAAGA-TGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 59
Pih15           CGTTAAGAATGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 60
H4              CGTTAAGAATGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 60
Pa3             CGTTAAGAATGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 60
Pih1            CGTTAAGAATGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 60
Vi5             CGTTAAGAATGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 60
Vi4             CGTTAAGAATGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 60
Pa2             CGTTAAGA-TGTTGTATCCTTCAATAAAGCATGTGCTATTATGTCAGCTGGTATACTCAA 59
                ******** **************************************************

H5              ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 119
Pih15           ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 120
H4              ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 120
Pa3             ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 120
Pih1            ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 120
Vi5             ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 120
Vi4             ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 120
Pa2             ACACACTTTCGACGAGGAATTCGACTGGAAGACTGGAAAATATGTTAGTACTGAGAACAA 119
                ************************************************************

H5              AGCAAAAGTTGTTCC-TAATGCTCAGATTATAAACAGACTTGCTGGCCAAATGGG-TTTA 177
Pih15           AGCAAAAGTTGTTCC-TAATGCTCAGATTATAAACAGACTTGCTGGCCAAATGGG-TTTA 178
H4              AGCAAAAGTTGTTCC-TAATGCTCAGATTATAAACAGACTTGCTGGCCAAATGGG-TTTA 178
Pa3             AGCAAAAGTTGTTCC-TAATGCTCAGATTATAAACAGACTTGCTGGCCAAATGGG-TTTA 178
Pih1            AGCAAAAGTTGTTCC-TAATGCTCAGATTATAAACAGACTTGCTGGCCAAATGGG-TTTA 178
Vi5             AGCAAAAGTTGTTCC-TAATGCTCAGATTATAAACAGACTTGCTGGCCAAATGGG-TTTA 178
Vi4             AGCAAAAGTTGTTCC-TAATGCTCAGATTATAAACAGACTTGCTGGCCAAATGGGGTTTA 179
Pa2             -GCAAA-GTTGTTCCCTAATGCTCAAATTATAAACAGACTTGCTGGCCAAATGGG-TTTA 176
                 ***** ******** ********* ***************************** ****
```

```
H5              GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 237
Pih15           GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 238
H4              GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 238
Pa3             GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 238
Pih1            GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 238
Vi5             GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 238
Vi4             GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 239
Pa2             GCACCTAGTTCACCTTATTACTGGATGATAGTACCTGGTTATGAATTCTTGTATGAACTG 236
                ************************************************************

H5              TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 297
Pih15           TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 298
H4              TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 298
Pa3             TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 298
Pih1            TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 298
Vi5             TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 298
Vi4             TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 299
Pa2             TACCCAGCTGAGGTATTAGCATACACTCTTGTGAGACTTGAATTCAGGAAGAACTTAAAC 296
                ************************************************************

H5              ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 357
Pih15           ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 358
H4              ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 358
Pa3             ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 358
Pih1            ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 358
Vi5             ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 358
Vi4             ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 359
Pa2             ATTCCTGATAAGACAACTGATGCTGACATTGTTAGTTCACTTGTTATGAAAATGAACAGG 356
                ************************************************************

H5              ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 417
Pih15           ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 418
H4              ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 418
Pa3             ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 418
Pih1            ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 418
Vi5             ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 418
Vi4             ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 419
Pa2             ATTCATAAACTTGATGTCACAACATTTGATGATGCTATCAACACAATAGGTCAGAACAAT 416
                ************************************************************
```

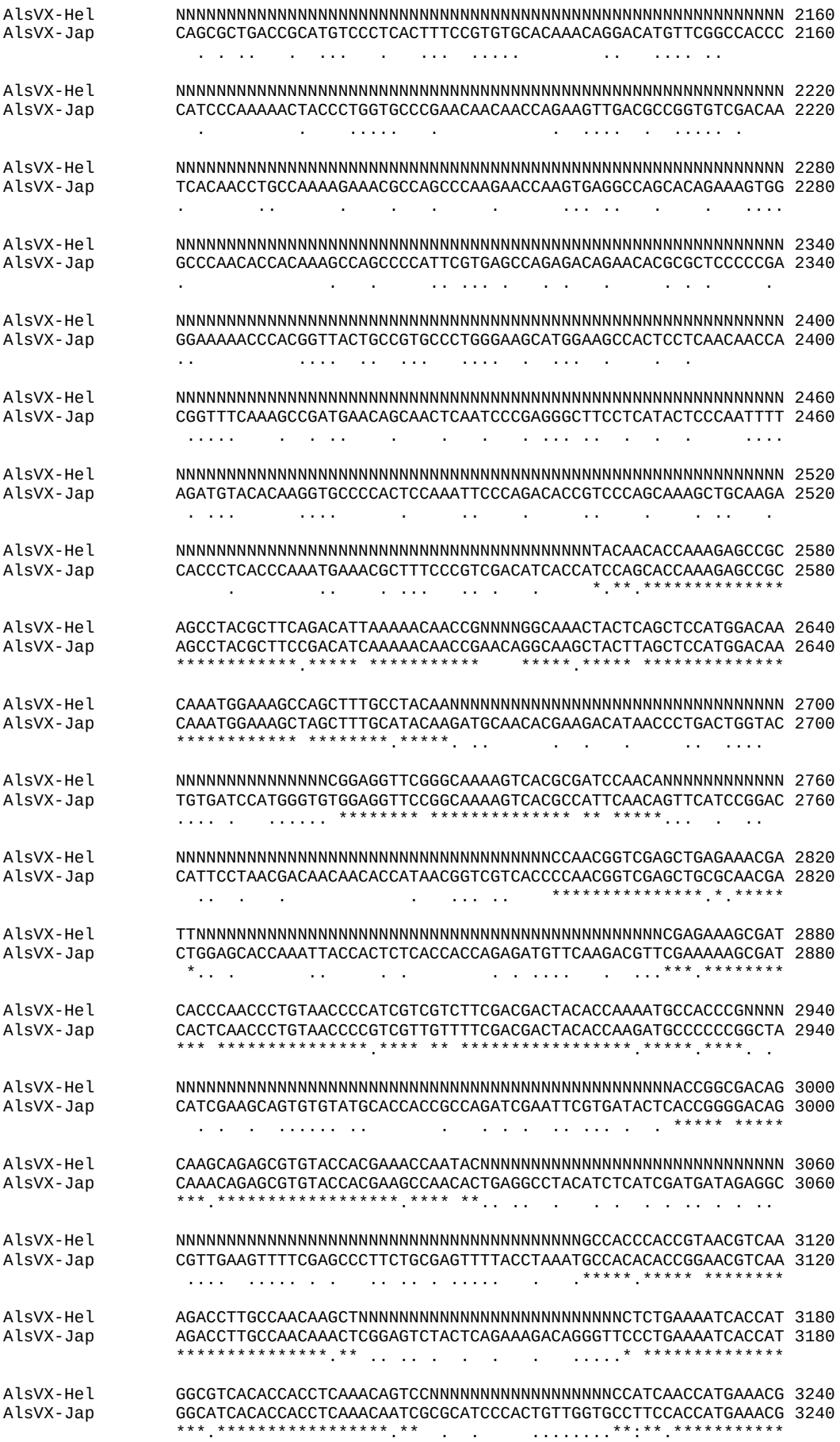
```
H5              GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACGAGCAAGAGCAAGAGG 477
Pih15           GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACCAGCAAGAACAAGAAG 478
H4              GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACGAGCAAGAGCAAGAGG 478
Pa3             GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACGAGCAAGAGCAAGAGG 478
Pih1            GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACGAGCAAGAGCAAGAGG 478
Vi5             GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACGAGCAAGAGCAAGAGG 478
Vi4             GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACGAGCAAGAGCAAGAGG 479
Pa2             GTGTCTGATGCTTATATTGAGTTAGCAAGAGACATAGGTACTACCAGCAAGAGCAAGAGG 476
                ******************************************* ******* ***** *

H5              AATGATGAAGCTATTATGAAA 498
Pih15           AATGATGAAGCTATTATGAAA 499
H4              AATGATGAAGCTATTATGAAA 499
Pa3             AATGATGAAGCTATTATGAAA 499
Pih1            AATGATGAAGCTATTATGAAA 499
Vi5             AATGATGAAGCTATTATGAAA 499
Vi4             AATGATGAAGCTATTATGAAA 500
Pa2             AATGATGAAGCTATTATGAAA 497
                ********************
```

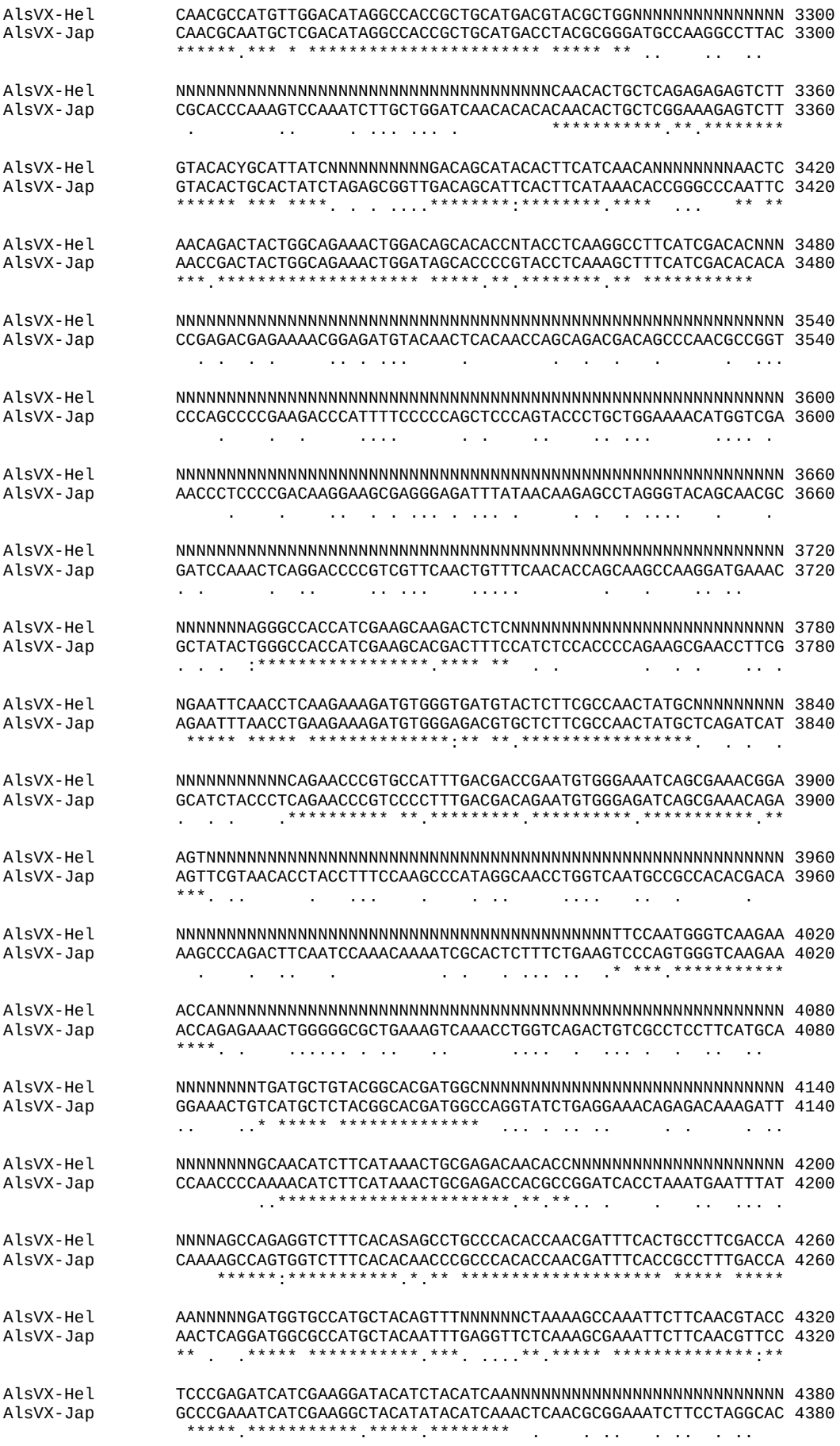
APPENDIX 6:

MAQ-RECOVERED GENOMIC SEQUENCE OF AlsVX-Hel

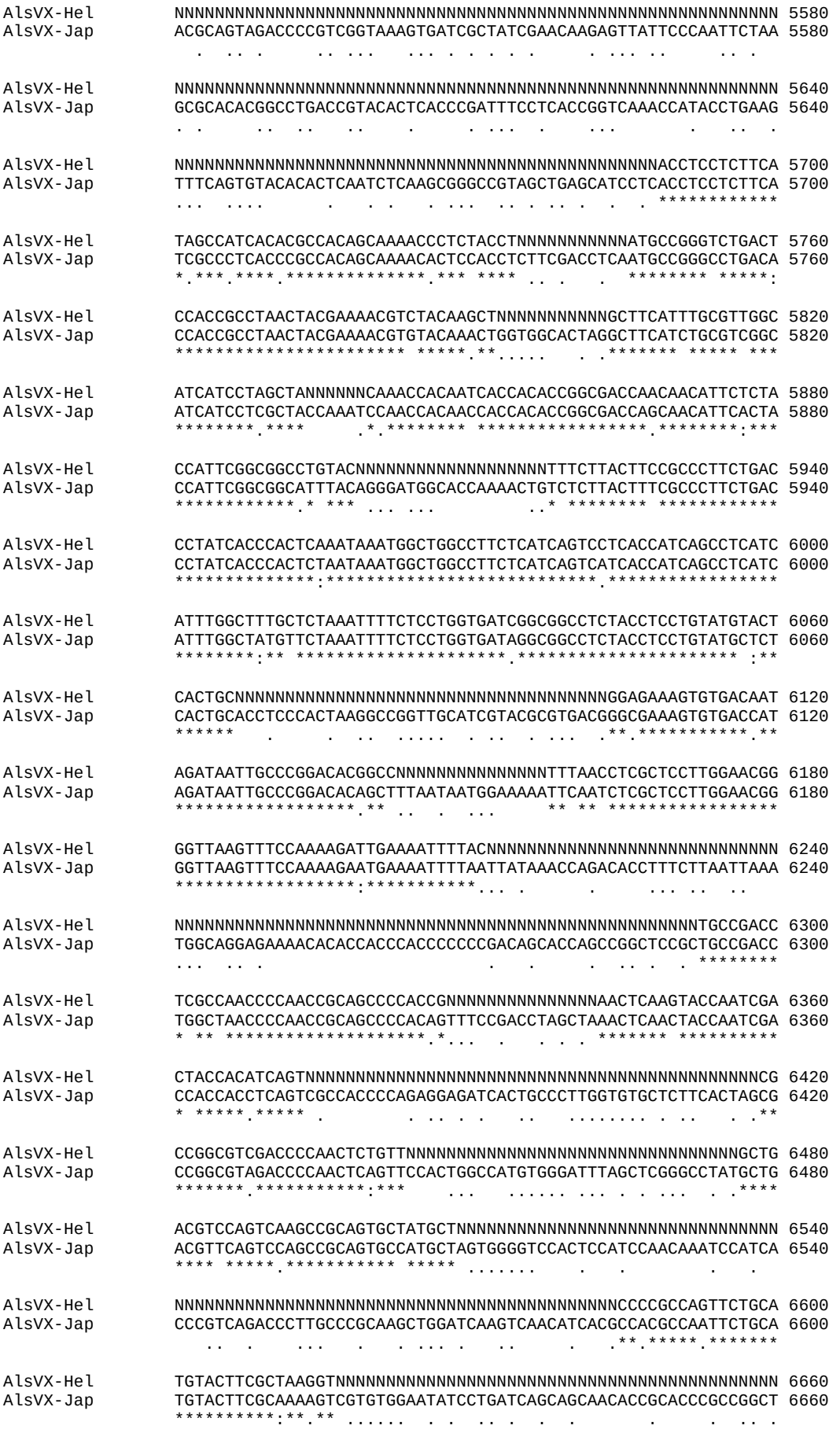


```
AlsVX-Hel       CGCCAACGCCAATAAACCGATNNNNNNNNNNNNNNNNATGCAGCTTTGGCTTTACTGCAA 1020
AlsVX-Jap       CGCCAACGCCAACAAACCCATCCAAAAACAAAAAGCCATGCAGCTCTGGCTATACTGCAA 1020
                ************ ***** **             .  ******** *****:********

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTACGCAC 1080
AlsVX-Jap       GTCAGTCAAAGCAGTAACCGAACGGGACATCTATGCCAAAATAAGGCAAGTACTGCGAAC 1080
                ..  ..    .  ..    .   ...   . . ..      .  ..   ..  *.**.**

AlsVX-Hel       CGAAGAACTTGAGCTTTATGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
AlsVX-Jap       CGAAGAACTTGAGCTGTACGAGCCGGCGGAGGTGGCACACATAGTAAACTACTTCTACTT 1140
                *************** ** **.  .. .. .....      . ..    .  .. . ..

AlsVX-Hel       NNNNNNNNNNNNNNACTCCATCACACGCTACGAAGACCTGCTATCTGGCTCWCTACTCAA 1200
AlsVX-Jap       CATAGCGAACCTCCACTCCATCACTTGCTACGAAGACCTGCTGTCAGGGTCACTACTCAA 1200
                  . . .    .  **********: ****************.**:** ** ********

AlsVX-Hel       ACGGATGACTGTCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGAGGAAA 1260
AlsVX-Jap       AAGGATGACAGTCCCAATACGCACGGCCATCACAAAGTTCGTCGAGGTCTTTCGCGGAAA 1260
                *.*******:*****  .  .   ..   .      ... .. . ... ... *.*****

AlsVX-Hel       GAATGACTTCGTCAAACTCNNNNNNNNNNNNNNNNGGGACAACTTCTCCTACAGCATCAA 1320
AlsVX-Jap       GAATGACTTTGTCAAACTTTTGGATGCTCTCAAGTGGGACAACTTTACCTACAGCATCAA 1320
                ********* ******** .... .. . .   ..********** :**************

AlsVX-Hel       CCCGGAAGANNNNNNNNNNNNAATGGACTCACATCAGGTAAACCACTTGGTGGAAGAGTA 1380
AlsVX-Jap       ACCGGAGGACATGACCACCAAAATGGACTCACACCAGGTCAACCACCTCGTGGAAGAGTA 1380
                .*****.**  ..         ************ *****.****** * ***********

AlsVX-Hel       TGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1440
AlsVX-Jap       TGGCATACCCATCAAAGAAGCTGAAGCGACCATCGCTCAGCTACACCGCGAACCAGAACT 1440
                ***  .    .   .  . ..  . .   . . .  . .    . .    .   .

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1500
AlsVX-Jap       GCTCAACGACGCGGACTTGGACTCCCCACCCATGGTGGATTTCCTCGGACGCACTACATA 1500
                . .    .  . ..  ....  .         ..... ...  . ..  .   .   .

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1560
AlsVX-Jap       CGAGCCCTTTAACTATCCCAGAACCACAACCCCGACAGACCCCAAGCCAGCCGCCACAAC 1560
                 . .   ...   . .    .           .   .      .   .  .

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1620
AlsVX-Jap       CGCTGACGACTCCGCCAGCTGTTCCAAACCGCCCAACTCCGCAGCTGGGACCAGCCCGAG 1620
                 . ..  .  .  .   . ....      .     .  .  . ....   .  . .

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1680
AlsVX-Jap       CCCCAACCACACGCAACCCGAGAAGCTGAAACCCGATGCACAACCCACTCAGGACTCCGC 1680
                            .     . .  . ..     . ..         .  ..  .  .

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1740
AlsVX-Jap       CACAATCCCCGTCCACGTCTCAGGCTGGGAAAGAGAGCCCTTCGACTATGTGTTCCACAC 1740
                     .    ..    .. .  .. ....   . . .   .. .  . ......

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCAGCTTTCTTAAGCACCCGACAGGACAT 1800
AlsVX-Jap       ATCGCAGAGCCAACAACTACGGGGACGCAAAGCAGCCTTCCTAAGCACCCGACAAGACAT 1800
                 . .  . .        .  ....  .    ***** *** *************.*****

AlsVX-Hel       CGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1860
AlsVX-Jap       AGACTACGGCCATGACAAAATCCGATATGACACAAACCCCTGGAACGAGATACTGACTGA 1860
                .*  .  ..   ..      .  . . . ..          ...   . . .  ..  ..

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1920
AlsVX-Jap       CTTTCTCAGAACTGTGGGTTTCCAAGCCAACGCCGTCCTGATTCAAGTCTATGAAGCTAA 1920
                 ... .  .   .........    .     .  ..  .. ..   .. . ..  . .

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1980
AlsVX-Jap       CGCAGGCATACCATGGCACAGAGACGACGAAAAGTGCTACGAAAAAGATGCCATTCTCAC 1980
                 .  ..  .    ...    . .  .  .    ... .  .     . ..   .. .

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2040
AlsVX-Jap       GCTCAACTTCGGCACAGCCACCTTTGAGTTCTCAAATGGCACTAAGCTGCACCTGACCGA 2040
                . .    .. ..    .     .... ... .    ...   .  . . ..    ..   .

AlsVX-Hel       NNNNTGCCACTTCATGATGTCAGGCAAATGGCTNNNNNNNNNNNNNNNNNNNNNNNNNN 2100
AlsVX-Jap       CGGGTGTCACTTCATAATGTCAGGCGAATGGCTGGACCTCAAACACCGGGTAACGAACTG 2100
                 ...** ********.*********.*******..   .       ....   .   ..
```


AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2160
AlsVX-Jap CAGCGCTGACCGCATGTCCCTCACTTTCCGTGTGCACAAACAGGACATGTTCGGCCACCC 2160
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2220
AlsVX-Jap CATCCCAAAAACTACCCTGGTGCCCGAACAACAACCAGAAGTTGACGCCGGTGTCGACAA 2220
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2280
AlsVX-Jap TCACAACCTGCCAAAAGAAACGCCAGCCCAAGAACCAAGTGAGGCCAGCACAGAAAGTGG 2280
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
AlsVX-Jap GCCCAACACCACAAAGCCAGCCCCATTCGTGAGCCAGAGACAGAACACGCGCTCCCCCGA 2340
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2400
AlsVX-Jap GGAAAAACCCACGGTTACTGCCGTGCCCTGGGAAGCATGGAAGCCACTCCTCAACAACCA 2400
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2460
AlsVX-Jap CGGTTTCAAAGCCGATGAACAGCAACTCAATCCCGAGGGCTTCCTCATACTCCCAATTTT 2460
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2520
AlsVX-Jap AGATGTACACAAGGTGCCCCACTCCAAATTCCCAGACACCGTCCCAGCAAAGCTGCAAGA 2520
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTACAACACCAAAGAGCCGC 2580
AlsVX-Jap CACCCTCACCCAAATGAAACGCTTTCCCGTCGACATCACCATCCAGCACCAAAGAGCCGC 2580
AlsVX-Hel AGCCTACGCTTCAGACATTAAAAACAACCGNNNNGGCAAACTACTCAGCTCCATGGACAA 2640
AlsVX-Jap AGCCTACGCTTCCGACATCAAAAACAACCGAACAGGCAAGCTACTTAGCTCCATGGACAA 2640
AlsVX-Hel CAAATGGAAAGCCAGCTTTGCCTACAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2700
AlsVX-Jap CAAATGGAAAGCTAGCTTTGCATACAAGATGCAACACGAAGACATAACCCTGACTGGTAC 2700
AlsVX-Hel NNNNNNNNNNNNNNNCGGAGGTTCGGGCAAAAGTCACGCGATCCAACANNNNNNNNNNNN 2760
AlsVX-Jap TGTGATCCATGGGTGTGGAGGTTCCGGCAAAAGTCACGCCATTCAACAGTTCATCCGGAC 2760
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCAACGGTCGAGCTGAGAAACGA 2820
AlsVX-Jap CATTCCTAACGACAACAACACCATAACGGTCGTCACCCCAACGGTCGAGCTGCGCAACGA 2820
AlsVX-Hel TTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGAGAAAGCGAT 2880
AlsVX-Jap CTGGAGCACCAAATTACCACTCTCACCACCAGAGATGTTCAAGACGTTCGAAAAAGCGAT 2880
AlsVX-Hel CACCCAACCCTGTAACCCCATCGTCGTCTTCGACGACTACACCAAAATGCCACCCGNNNN 2940
AlsVX-Jap CACTCAACCCTGTAACCCCGTCGTTGTTTTCGACGACTACACCAAGATGCCCCCCGGCTA 2940
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACCGGCGACAG 3000
AlsVX-Jap CATCGAAGCAGTGTGTATGCACCACCGCCAGATCGAATTCGTGATACTCACCGGGGACAG 3000
AlsVX-Hel CAAGCAGAGCGTGTACCACGAAACCAATACNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3060
AlsVX-Jap CAAACAGAGCGTGTACCACGAAGCCAACACTGAGGCCTACATCTCATCGATGATAGAGGC 3060
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCCACCCACCGTAACGTCAA 3120
AlsVX-Jap CGTTGAAGTTTTCGAGCCCTTCTGCGAGTTTTACCTAAATGCCACACACCGGAACGTCAA 3120
AlsVX-Hel AGACCTTGCCAACAAGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCTGAAAATCACCAT 3180
AlsVX-Jap AGACCTTGCCAACAAACTCGGAGTCTACTCAGAAAGACAGGGTTCCCTGAAAATCACCAT 3180
AlsVX-Hel GGCGTCACACCACCTCAAACAGTCCNNNNNNNNNNNNNNNNNNCCATCAACCATGAAACG 3240
AlsVX-Jap GGCATCACACCACCTCAAACAATCGCGCATCCCACTGTTGGTGCCTTCCACCATGAAACG 3240


AlsVX-Hel CAACGCCATGTTGGACATAGGCCACCGCTGCATGACGTACGCTGGNNNNNNNNNNNNNNN 3300
AlsVX-Jap CAACGCAATGCTCGACATAGGCCACCGCTGCATGACCTACGCGGGATGCCAAGGCCTTAC 3300
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAACACTGCTCAGAGAGAGTCTT 3360
AlsVX-Jap CGCACCCAAAGTCCAAATCTTGCTGGATCAACACACACAACACTGCTCGGAAAGAGTCTT 3360
AlsVX-Hel GTACACYGCATTATCNNNNNNNNNNGACAGCATACACTTCATCAACANNNNNNNNAACTC 3420
AlsVX-Jap GTACACTGCACTATCTAGAGCGGTTGACAGCATTCACTTCATAAACACCGGGCCCAATTC 3420
AlsVX-Hel AACAGACTACTGGCAGAAACTGGACAGCACACCNTACCTCAAGGCCTTCATCGACACNNN 3480
AlsVX-Jap AACCGACTACTGGCAGAAACTGGATAGCACCCCGTACCTCAAAGCTTTCATCGACACACA 3480
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3540
AlsVX-Jap CCGAGACGAGAAAACGGAGATGTACAACTCACAACCAGCAGACGACAGCCCAACGCCGGT 3540
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3600
AlsVX-Jap CCCAGCCCCGAAGACCCATTTTCCCCCAGCTCCCAGTACCCTGCTGGAAAACATGGTCGA 3600
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3660
AlsVX-Jap AACCCTCCCCGACAAGGAAGCGAGGGAGATTTATAACAAGAGCCTAGGGTACAGCAACGC 3660
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3720
AlsVX-Jap GATCCAAACTCAGGACCCCGTCGTTCAACTGTTTCAACACCAGCAAGCCAAGGATGAAAC 3720
AlsVX-Hel NNNNNNNAGGGCCACCATCGAAGCAAGACTCTCNNNNNNNNNNNNNNNNNNNNNNNNNN 3780
AlsVX-Jap GCTATACTGGGCCACCATCGAAGCACGACTTTCCATCTCCACCCCAGAAGCGAACCTTCG 3780
AlsVX-Hel NGAATTCAACCTCAAGAAAGATGTGGGTGATGTACTCTTCGCCAACTATGCNNNNNNNNN 3840
AlsVX-Jap AGAATTTAACCTGAAGAAAGATGTGGGAGACGTGCTCTTCGCCAACTATGCTCAGATCAT 3840
AlsVX-Hel NNNNNNNNNNNCAGAACCCGTGCCATTTGACGACCGAATGTGGGAAATCAGCGAAACGGA 3900
AlsVX-Jap GCATCTACCCTCAGAACCCGTCCCCTTTGACGACAGAATGTGGGAGATCAGCGAAACAGA 3900
AlsVX-Hel AGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3960
AlsVX-Jap AGTTCGTAACACCTACCTTTCCAAGCCCATAGGCAACCTGGTCAATGCCGCCACACGACA 3960
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTCCAATGGGTCAAGAA 4020
AlsVX-Jap AAGCCCAGACTTCAATCCAAACAAAATCGCACTCTTTCTGAAGTCCCAGTGGGTCAAGAA 4020
AlsVX-Hel ACCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4080
AlsVX-Jap ACCAGAGAAACTGGGGGCGCTGAAAGTCAAACCTGGTCAGACTGTCGCCTCCTTCATGCA 4080
AlsVX-Hel NNNNNNNNTGATGCTGTACGGCACGATGGCNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4140
AlsVX-Jap GGAAACTGTCATGCTCTACGGCACGATGGCCAGGTATCTGAGGAAACAGAGACAAAGATT 4140
AlsVX-Hel NNNNNNNNGCAACATCTTCATAAACTGCGAGACAACACCNNNNNNNNNNNNNNNNNNNN 4200
AlsVX-Jap CCAACCCCAAAACATCTTCATAAACTGCGAGACCACGCCGGATCACCTAAATGAATTTAT 4200
AlsVX-Hel NNNNAGCCAGAGGTCTTTCACASAGCCTGCCCACACCAACGATTTCACTGCCTTCGACCA 4260
AlsVX-Jap CAAAAGCCAGTGGTCTTTCACACAACCCGCCCACACCAACGATTTCACCGCCTTTGACCA 4260
AlsVX-Hel AANNNNNGATGGTGCCATGCTACAGTTTNNNNNNCTAAAAGCCAAATTCTTCAACGTACC 4320
AlsVX-Jap AACTCAGGATGGCGCCATGCTACAATTTGAGGTTCTCAAAGCGAAATTCTTCAACGTTCC 4320
AlsVX-Hel TCCCGAGATCATCGAAGGATACATCTACATCAANNNNNNNNNNNNNNNNNNNNNNNNNNNN 4380
AlsVX-Jap GCCCGAAATCATCGAAGGCTACATATACATCAAACTCAACGCGGAAATCTTCCTAGGCAC 4380

```
AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4440
AlsVX-Jap        CCTCGGGATCATGAGACTCAGCGGTGAAGGCCCAACTTTCGACGCTAACACGGAATGCAG 4440
                   . ... .  .. .  .  . ....  ..      ... .  . .     ..  ..   .

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACAAGTGTACGCCGG 4500
AlsVX-Jap        CATCGCCTACAACGCAACACGATTTTTCGTAGACGACTCAGTCGCACAAGTCTACGCCGG 4500
                   . .  .     .      . ..... .. .  .  .  .. . ****** ********

AlsVX-Hel        CGACGACATGGCTTTAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4560
AlsVX-Jap        CGATGACATGGCATTAGATCGCGTGGTGCAGGAGAAACCTAGCTTCAAAAGGCTGGAACA 4560
                 *** ********:*****. . ......  .. .     . . ..     .. ...

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4620
AlsVX-Jap        TCAGCTCAAGCTCACATCCAAACCGCAATACCCAAGCCAAACAAAAGGTGACTACGCCGA 4620
                 .  . .   . .    .        .  .      .          ....  .  .  .

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCAAGATGCAGGC 4680
AlsVX-Jap        ATTCTGCGGGTGGATATTCACACCTAGCGGCATCATGAAGCATAGTCTCAAAATGCAGGC 4680
                  .. .. ...... . ..      . . ..  .  ..  .  . .. ****.********

AlsVX-Hel        CTCAATCCAACTCCAGAAGAAGATCAACAACATTGCCCAATCCGCNNNNNNNNNNNNNNN 4740
AlsVX-Jap        CTCCATCCAGCTCCAAAAGAAAATCAACAACATCGCCCAATCAGCTCGTTCATACGCACT 4740
                 ***.*****.*****.*****.*********** ********.**. ...  .  .   .

AlsVX-Hel        NNNNNNNNNNNNNGCCTACGACATGGGTGACGCGCTACAAGATCANNNNNNNNNNNNNNN 4800
AlsVX-Jap        GGATCTGAAGTACGCTTACGACATGGGTGACGCACTACAAGACCATCTAACTGAAGCTGA 4800
                 .. . ..  ..  ** *****************.******** **. .   ..  . ..

AlsVX-Hel        NNNNNNNNNNNNNNNNCAATCCGTCCGAGACATGCACCTACTCCACCAGAATGANNNNNN 4860
AlsVX-Jap        GTGTGAACTGCACGCTCAATCGGTCCGCGACATGCACCTACTCCACCAACATGACGTCCT 4860
                 .....   ..   . .***** *****.********************..**** ..  .

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4920
AlsVX-Jap        AGTCAACGGTGCTTCAAGCCCCCCCCATCCGACCCAAACAACGTCAGCCGCTGACATCCC 4920
                  ..     .... ..    .         .  .           ..  .  . ..    .

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4980
AlsVX-Jap        CACTGGTAGCGCCAAGACTAAGAAAAGAAACCTATCCAAAAGGAAAGCTAAAGTAGCCAA 4980
                    .... . .    .  .  .     .     . .       ..   . .    .. .

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5040
AlsVX-Jap        AATCGACAGTAGCGGAGATTTCAATTTCCGTTGCCCCAGTTCCCCCCCACGCCAATAGGA 5040
                   . .   .. . .. . ...   ...  ....      ...          .    . ..

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNA 5100
AlsVX-Jap        GCCGGGTTAAGTGTCCTGATTGTGAAATGGCTTTGTTCTACCTCCAAACTTTATTACTCA 5100
                 .  .....  ....  .. .....   ... ...... .   .      ... ..  . *

AlsVX-Hel        ACAGCGATTTCACCCGGACCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5160
AlsVX-Jap        ACAGTGATTTCACCAGGACCACAACCCCCACAGCTAGCCCCCTCATCATCTTTGGTACTG 5160
                 **** *********.******           . . .     .  .  . ......  ..

AlsVX-Hel        NNNNNNNNNNNNNNNNNNCAATAATCCGCGCCCTAGCTTNNNNNNNNNNNNNNNNNNNCT 5220
AlsVX-Jap        CTGGCTGTGGCAAAACCTCAATAATCCGCGCTCTCGCTTCCCAATTCCCCGAACTCCACT 5220
                  ... .....       .************* **.****      ..     .   . **

AlsVX-Hel        TCTCYAGCTTCCACCCCATCGTTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5280
AlsVX-Jap        TCTCTAGCTTTCACCCCATCGTGCTCCTACCCAACATCCGTAAGAAACAGCACCTAGCCA 5280
                 **** ***** *********** *.  .         .  .. .     .    . .

AlsVX-Hel        NNNNNNNNNNNNNNACCGACGTGTTAGACGAATTCCTCGCAGGCCCCAACNNNNNNNNNN 5340
AlsVX-Jap        ACCCCAACGAAGCCACTGACGTGTTGGACGAATTCCTTGCTGGCCCCAACCCCGAAGTCC 5340
                         .  .  ** ********.*********** **:*********    .  ..

AlsVX-Hel        NNCTAGCCAAATTCTGTGACCCCNNNNNNNNNNNNNNNNNNNNNNNNCCCGAACCACACT 5400
AlsVX-Jap        GAATAGCCAAATTCTGCGACCCCCTTCAATACAACTGTGAGACCTTGCCAGAACCCCACT 5400
                 . .************* ****** ..  .      .... .   ...**.*****.****

AlsVX-Hel        TCATAAGTGAAACCACTTACCGATTCTGCCCACGTACCTGCGAACNNNNNNNNNNNNNNN 5460
AlsVX-Jap        TCATAAGTGAAACCACCTACCGCTTCTGCCCACGCACCTGTGAACTACTCAACGACATCT 5460
                 **************** *****.*********** ***** ****.  .    .   . .

AlsVX-Hel        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5520
AlsVX-Jap        TCAAAACAACTCTCAAGAGCAAGGTTCTAGAGATTTGCAAAGTTGCGTGCGTCGACCCAT 5520
                 .         . .    . .    .... . . . ....     .... ... .. .     .
```

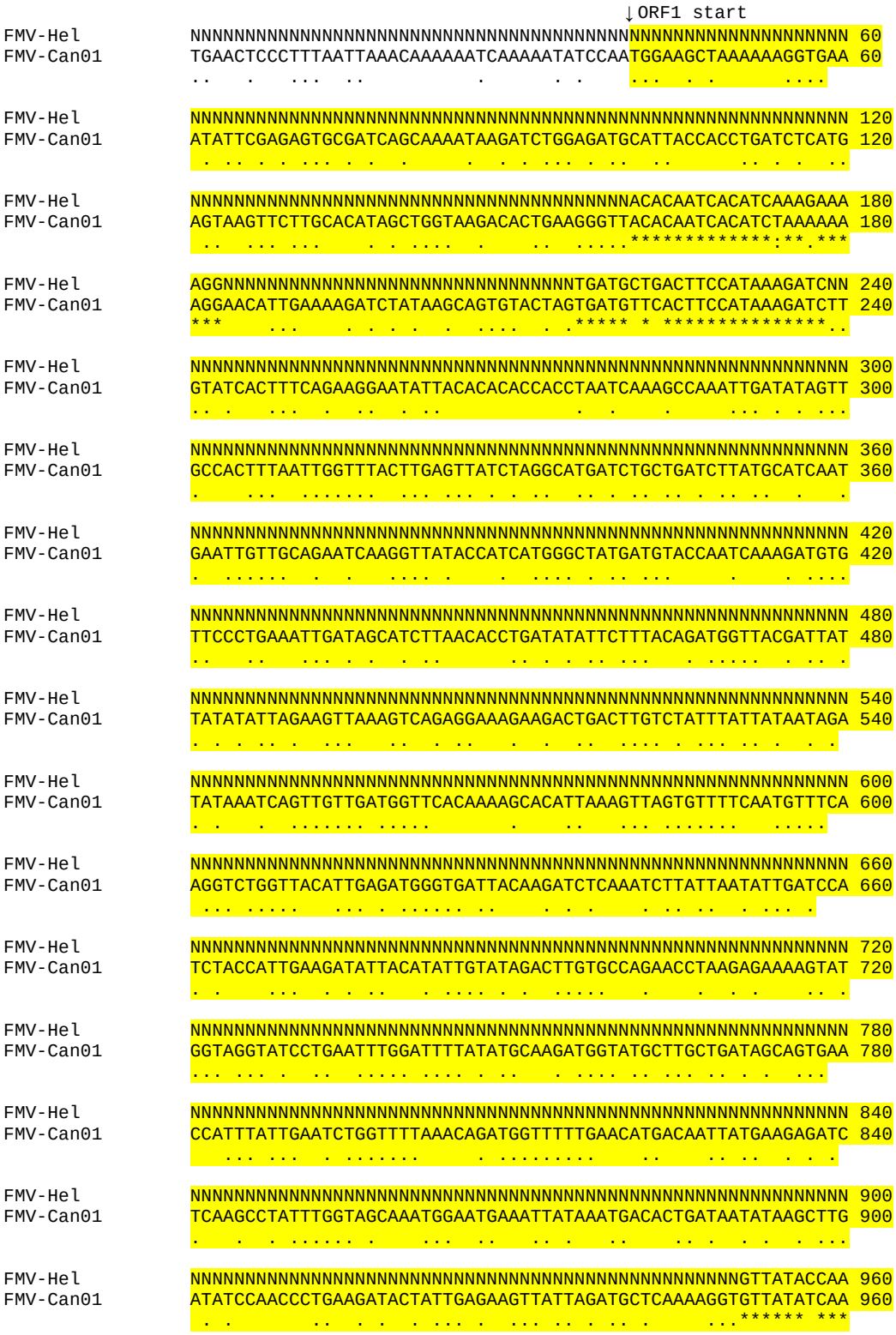

AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5580
AlsVX-Jap ACGCAGTAGACCCCGTCGGTAAAGTGATCGCTATCGAACAAGAGTTATTCCCAATTCTAA 5580
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5640
AlsVX-Jap GCGCACACGGCCTGACCGTACACTCACCCGATTTCCTCACCGGTCAAACCATACCTGAAG 5640
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACCTCCTCTTCA 5700
AlsVX-Jap TTTCAGTGTACACACTCAATCTCAAGCGGGCCGTAGCTGAGCATCCTCACCTCCTCTTCA 5700
AlsVX-Hel TAGCCATCACACGCCACAGCAAAACCCTCTACCTNNNNNNNNNNNNATGCCGGGTCTGACT 5760
AlsVX-Jap TCGCCCTCACCCGCCACAGCAAAACACTCCACCTCTTCGACCTCAATGCCGGGCCTGACA 5760
AlsVX-Hel CCACCGCCTAACTACGAAAACGTCTACAAGCTNNNNNNNNNNNGCTTCATTTGCGTTGGC 5820
AlsVX-Jap CCACCGCCTAACTACGAAAACGTGTACAAACTGGTGGCACTAGGCTTCATCTGCGTCGGC 5820
AlsVX-Hel ATCATCCTAGCTANNNNNNCAAACCACAATCACCACACCGGCGACCAACAACATTCTCTA 5880
AlsVX-Jap ATCATCCTCGCTACCAAATCCAACCACAACCACCACACCGGCGACCAGCAACATTCACTA 5880
AlsVX-Hel CCATTCGGCGGCCTGTACNNNNNNNNNNNNNNNNNNNTTTCTTACTTCCGCCCTTCTGAC 5940
AlsVX-Jap CCATTCGGCGGCATTTACAGGGATGGCACCAAAACTGTCTCTTACTTTCGCCCTTCTGAC 5940
AlsVX-Hel CCTATCACCCACTCAAATAAATGGCTGGCCTTCTCATCAGTCCTCACCATCAGCCTCATC 6000
AlsVX-Jap CCTATCACCCACTCTAATAAATGGCTGGCCTTCTCATCAGTCATCACCATCAGCCTCATC 6000
AlsVX-Hel ATTTGGCTTTGCTCTAAATTTTCTCCTGGTGATCGGCGGCCTCTACCTCCTGTATGTACT 6060
AlsVX-Jap ATTTGGCTATGTTCTAAATTTTCTCCTGGTGATAGGCGGCCTCTACCTCCTGTATGCTCT 6060
AlsVX-Hel CACTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGAGAAAGTGTGACAAT 6120
AlsVX-Jap CACTGCACCTCCCACTAAGGCCGGTTGCATCGTACGCGTGACGGGCGAAAGTGTGACCAT 6120
AlsVX-Hel AGATAATTGCCCGGACACGGCCNNNNNNNNNNNNNNNTTTAACCTCGCTCCTTGGAACGG 6180
AlsVX-Jap AGATAATTGCCCGGACACAGCTTTAATAATGGAAAAATTCAATCTCGCTCCTTGGAACGG 6180
AlsVX-Hel GGTTAAGTTTCCAAAAGATTGAAAATTTTACNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6240
AlsVX-Jap GGTTAAGTTTCCAAAAGAATGAAAATTTTAATTATAAACCAGACACCTTTCTTAATTAAA 6240
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGCCGACC 6300
AlsVX-Jap TGGCAGGAGAAAACACACCACCCACCCCCCCGACAGCACCAGCCGGCTCCGCTGCCGACC 6300
AlsVX-Hel TCGCCAACCCCAACCGCAGCCCCACCGNNNNNNNNNNNNNNNNAACTCAAGTACCAATCGA 6360
AlsVX-Jap TGGCTAACCCCAACCGCAGCCCCACAGTTTCCGACCTAGCTAAACTCAACTACCAATCGA 6360
AlsVX-Hel CTACCACATCAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCG 6420
AlsVX-Jap CCACCACCTCAGTCGCCACCCCAGAGGAGATCACTGCCCTTGGTGTGCTCTTCACTAGCG 6420
AlsVX-Hel CCGGCGTCGACCCCAACTCTGTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCTG 6480
AlsVX-Jap CCGGCGTAGACCCCAACTCAGTTCCACTGGCCATGTGGGATTTAGCTCGGGCCTATGCTG 6480
AlsVX-Hel ACGTCCAGTCAAGCCGCAGTGCTATGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6540
AlsVX-Jap ACGTTCAGTCCAGCCGCAGTGCCATGCTAGTGGGGTCCACTCCATCCAACAAATCCATCA 6540
AlsVX-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCCCGCCAGTTCTGCA 6600
AlsVX-Jap CCCGTCAGACCCTTGCCCGCAAGCTGGATCAAGTCAACATCACGCCACGCCAATTCTGCA 6600
AlsVX-Hel TGTACTTCGCTAAGGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6660
AlsVX-Jap TGTACTTCGCAAAAGTCGTGTGGAATATCCTGATCAGCAGCAACACCGCACCCGCCGGCT 6660

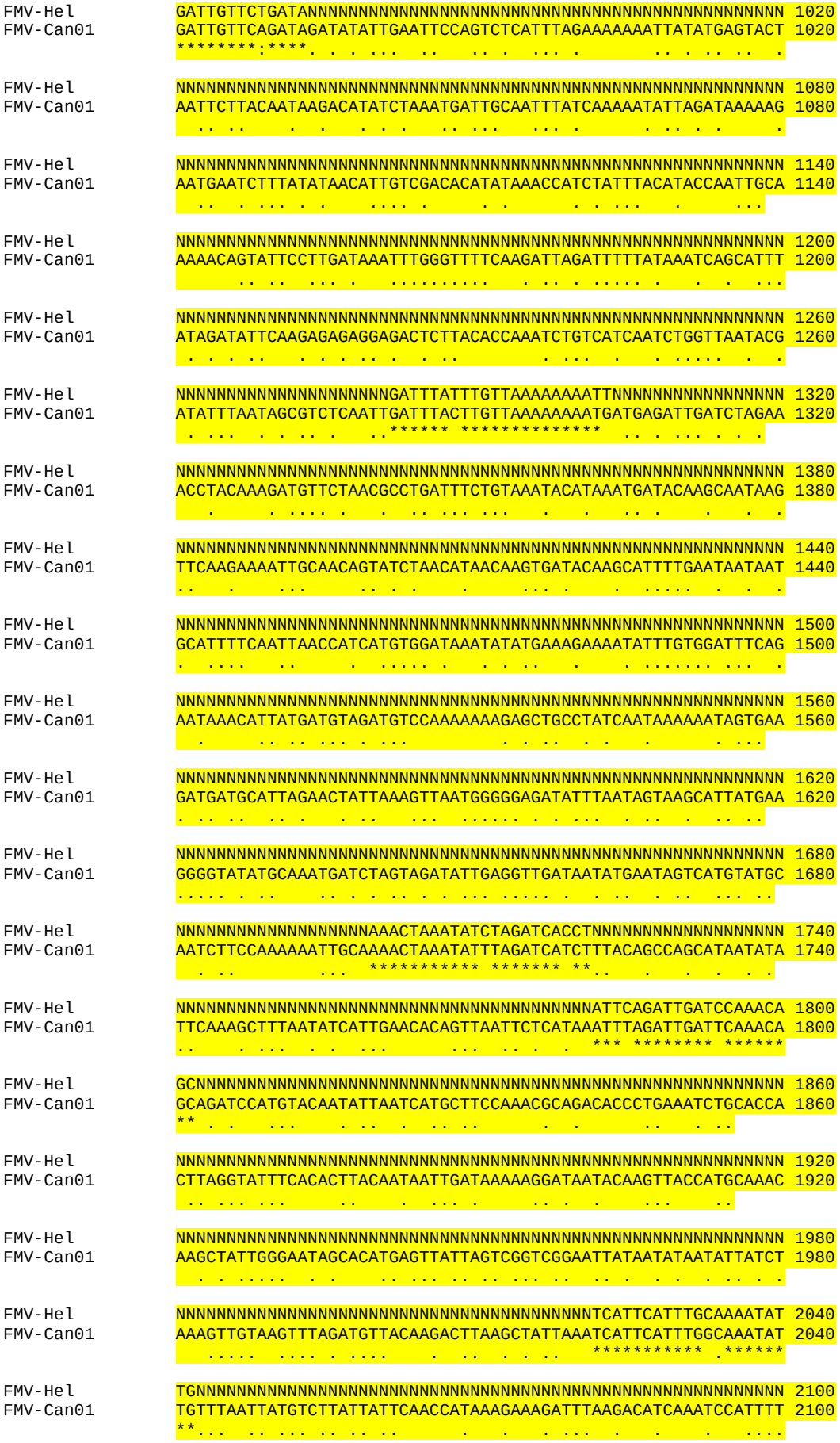
```
AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCTGCTTTCGATTTCTTCGACGGCG 6720
AlsVX-Jap       GGGCAAAGTTGGGGTACCCCGAAGAATGCAAGTTCGCTGCATTCGACTTCTTCGACGGTG 6720
                ...    ........     .  .  ..   ... *****:***** ********** *

AlsVX-Hel       TCCTATCCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6780
AlsVX-Jap       TCCTCTCCCCAGCCGCACTTGATCCCGCTGACGGGCTCATCAGACAACCCAACACCCGGG 6780
                ****.****  .  .   ... .   . ..  ... .  .  .             ...

AlsVX-Hel       NNNNNNNNNNNNNNNNNNNNNNNNNNNNTGGAGCTCTAGCTCGCCAACGCATCAGCAATG 6840
AlsVX-Jap       AGATCCAAGCCCACTCAACGGCTAAGTTTGGAGCTCTCGCCCGCCAACGCATCAGCAATG 6840
                 . .    .     .    .. .  ...*********.** ******************

AlsVX-Hel       GCAACTANNNNNNNACTCTTGCCGAAGTCACCCACGGCAGAGTTGGTGGGGTCAATGCCA 6900
AlsVX-Jap       GCAATTTCGTCTCAACACTCGCCGAGGTCACCCACGGCAGAGTCGGAGGTGTCAATGCCA 6900
                **** *: .. .  **:** *****.**************** **:** **********

AlsVX-Hel       TGTACTCCATCGAAGCCCCCCCAGAAGCCTAAGCCTTTCTCCTTGCCCAACTTAAATAAG 6960
AlsVX-Jap       TGTACTCCATCGAAGCCCCCCCAGAAGCCTAAGCCTTTCTCCTTGCCCAACTTAAATAAG 6960
                ************************************************************

AlsVX-Hel       GAGGTGTTTTTCTACAGTTTGTTTTTNNNNNNNNNNNNNNNNNNNNN 7009
AlsVX-Jap       GAGGTGTTTTTCTACAGTTTTATTTTCACTAATATACTATGTGTGTGTT 7009
                ******************* :****   .  . .  . ..........
```

APPENDIX 7:

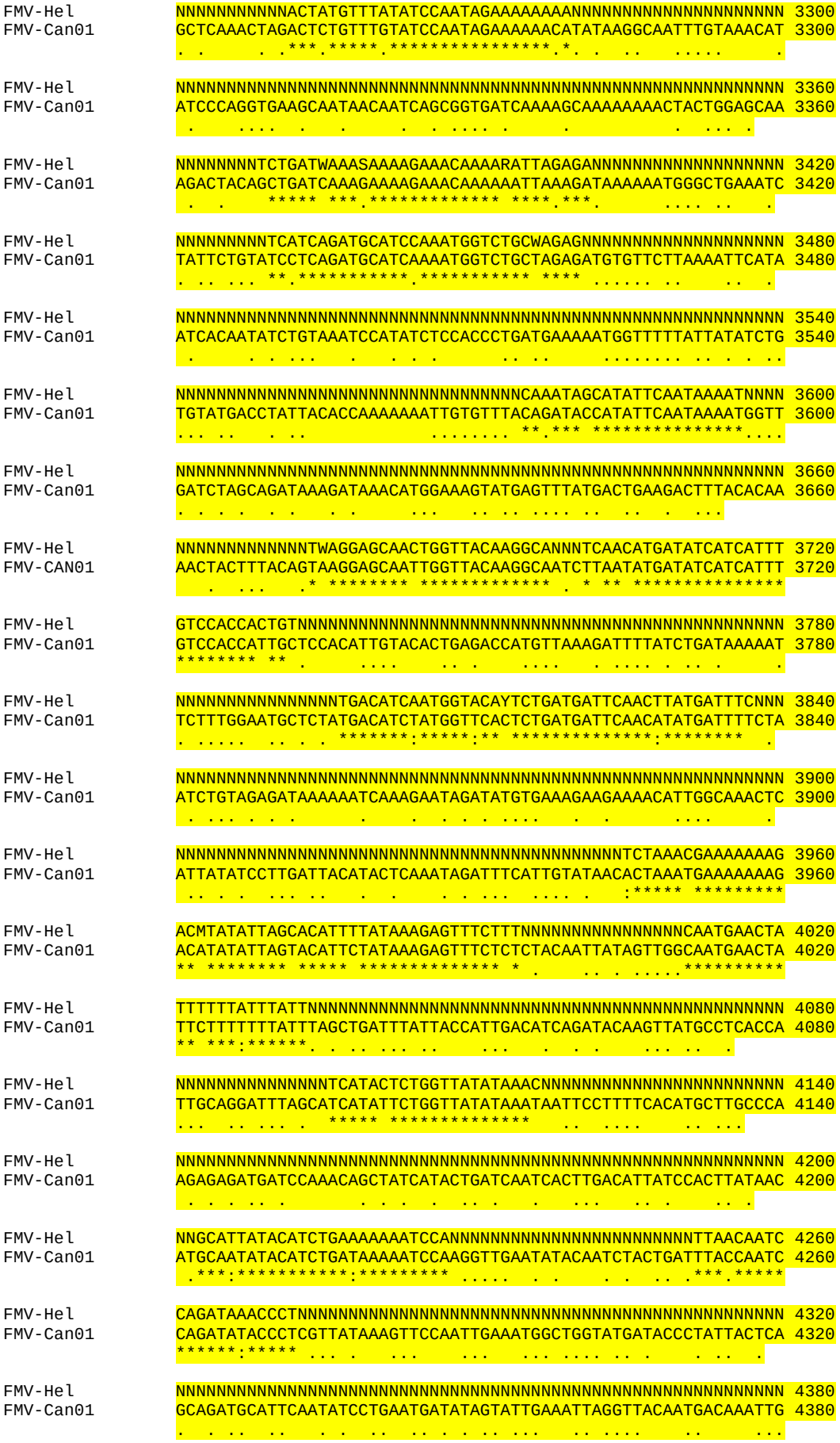
MAQ-RECOVERED GENOMIC SEQUENCE OF FMV-Hel

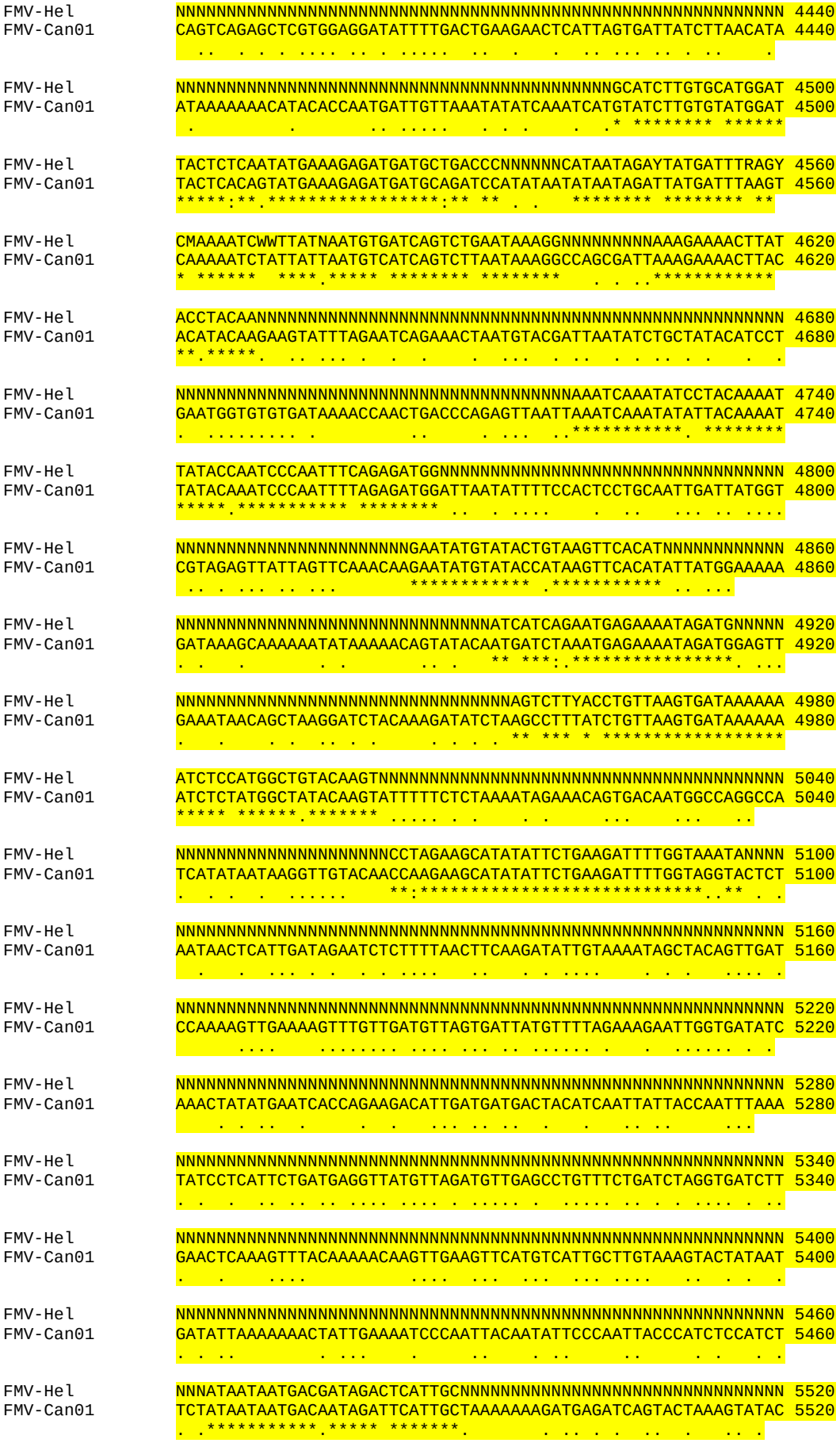
FMV RNA1-RdRp gene

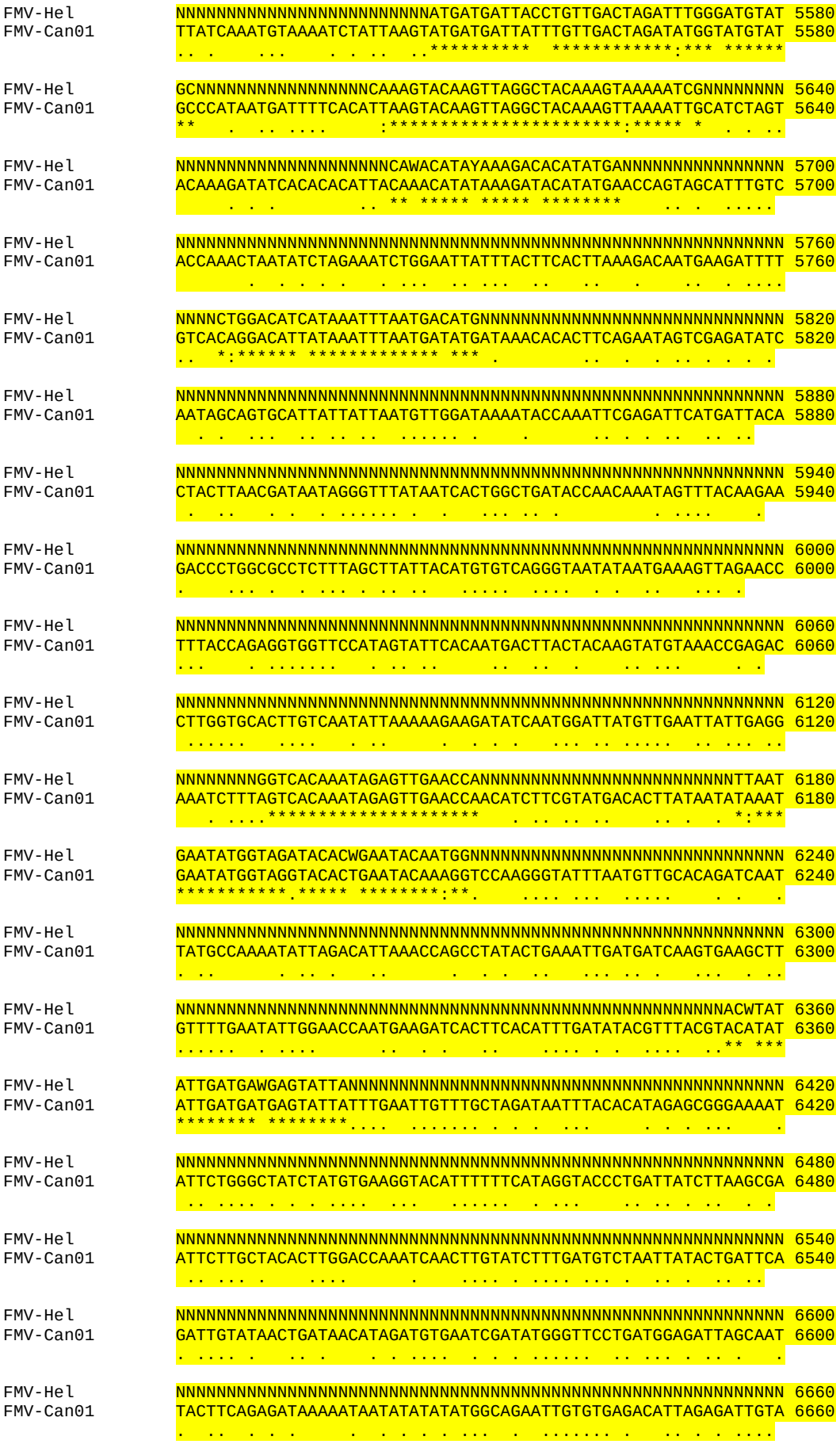



FMV-Hel GATTGTTCTGATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
FMV-Can01 GATTGTTCAGATAGATATATTGAATTCCAGTCTCATTTAGAAAAAAATTATATGAGTACT 1020
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1080
FMV-Can01 AATTCTTACAATAAGACATATCTAAATGATTGCAATTTATCAAAAATATTAGATAAAAAG 1080
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
FMV-Can01 AATGAATCTTTATATAACATTGTCGACACATATAAACCATCTATTTACATACCAATTGCA 1140
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1200
FMV-Can01 AAAACAGTATTCCTTGATAAATTTGGGTTTTCAAGATTAGATTTTTATAAATCAGCATTT 1200
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1260
FMV-Can01 ATAGATATTCAAGAGAGAGGAGACTCTTACACCAAATCTGTCATCAATCTGGTTAATACG 1260
FMV-Hel NNNNNNNNNNNNNNNNNNNNNGATTTATTTGTTAAAAAAAATTNNNNNNNNNNNNNNNNN 1320
FMV-Can01 ATATTTAATAGCGTCTCAATTGATTTACTTGTTAAAAAAAATGATGAGATTGATCTAGAA 1320
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1380
FMV-Can01 ACCTACAAAGATGTTCTAACGCCTGATTTCTGTAAATACATAAATGATACAAGCAATAAG 1380
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1440
FMV-Can01 TTCAAGAAAATTGCAACAGTATCTAACATAACAAGTGATACAAGCATTTTGAATAATAAT 1440
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1500
FMV-Can01 GCATTTTCAATTAACCATCATGTGGATAAATATATGAAAGAAAATATTTGTGGATTTCAG 1500
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1560
FMV-Can01 AATAAACATTATGATGTAGATGTCCAAAAAAAGAGCTGCCTATCAATAAAAAATAGTGAA 1560
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1620
FMV-Can01 GATGATGCATTAGAACTATTAAAGTTAATGGGGGAGATATTTAATAGTAAGCATTATGAA 1620
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1680
FMV-Can01 GGGGTATATGCAAATGATCTAGTAGATATTGAGGTTGATAATATGAATAGTCATGTATGC 1680
FMV-Hel NNNNNNNNNNNNNNNNNNNAAACTAAATATCTAGATCACCTNNNNNNNNNNNNNNNNNNN 1740
FMV-Can01 AATCTTCCAAAAAATTGCAAAACTAAATATTTAGATCATCTTTACAGCCAGCATAATATA 1740
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATTCAGATTGATCCAAACA 1800
FMV-Can01 TTCAAAGCTTTAATATCATTGAACACAGTTAATTCTCATAAATTTAGATTGATTCAAACA 1800
FMV-Hel GCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1860
FMV-Can01 GCAGATCCATGTACAATATTAATCATGCTTCCAAACGCAGACACCCTGAAATCTGCACCA 1860
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1920
FMV-Can01 CTTAGGTATTTCACACTTACAATAATTGATAAAAAGGATAATACAAGTTACCATGCAAAC 1920
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1980
FMV-Can01 AAGCTATTGGGAATAGCACATGAGTTATTAGTCGGTCGGAATTATAATATAATATTATCT 1980
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCATTCATTTGCAAAATAT 2040
FMV-Can01 AAAGTTGTAAGTTTAGATGTTACAAGACTTAAGCTATTAAATCATTCATTTGGCAAATAT 2040
FMV-Hel TGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2100
FMV-Can01 TGTTTAATTATGTCTTATTATTCAACCATAAAGAAAGATTTAAGACATCAAATCCATTTT 2100

```
FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNCCATATCATCACTGAGCATAACTGATACATAT 2160
FMV-Can01       ATGGTGTGGATATTATGTCAATTTATTACTATATCATCACTAAGCATAACTGACACCTAT 2160
                 ........ . .. ...   ... .. * ***********.********** **.***

FMV-Hel         AAGAATTTTATAATGGCTATATATTCAGATTACTCAAATATTGANNNNNNNNNNNNNNNN 2220
FMV-Can01       AAGAATTTTGTCATGGCTATATATTCTGATTATTCAAATATTGACAATCTGATTGATGAC 2220
                *********.*.**************:***** ***********   . .. ... ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2280
FMV-Can01       AAGTTAGAGTCTAGGCCTACAACATTAGGGCATATAGTAATCTTACAAAGGTGTTTCCAT 2280
                  ... . .. . ..  .      .. ...  . . ..  . ..     .......   .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2340
FMV-Can01       GGTCTACTCCAGGCTTCTACACAACTCAAGGTAATTAACAAAAACAAAAACAGCGAAGAG 2340
                ... .  .   .. .. .       .   ...  ..                . .  . .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2400
FMV-Can01       ATTGATGAGAGGGGTGATTTAATCGAAACTGGATTTAATAGCAACTTATCATTAAAGCTG 2400
                 ... .. . ..... ...  . .    ... ... . .    .. .  ..   . ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2460
FMV-Can01       CCAATATCAGGTATTAGTGTCAATAATCCTAGGGAGATTATACATGAAGCATTTATGTTA 2460
                    . .  ... .. ....  .  .  . ... . .. .   ..  .  ... ....

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2520
FMV-Can01       TTTTATCTAGGTAATAAAGGCCTTCATGGTTCTCCACAGGAGTTGTTAAACTTGTATTAC 2520
                .... . . ...  .   ..  ..  ..... .    .. ......    .... ..

FMV-Hel         NNNNNNNNNNNNTTCGAAAAAGAATATACTGATNNNNNNNNNNNNNNNNNNNNNNNNNN 2580
FMV-Can01       ACACCATACAAATTTGAACAAGAATATACTGATGTAATACAGAGATATAAAACTATTTTA 2580
                      .     ** ***.***************..  .   . . . .    . ....

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2640
FMV-Can01       CAAGAGGATGGAAATATTTCTAATATGTCAATCTCATATGATGTTATGAAAAAAACATCA 2640
                   . .. ...   . ... .  . ...   . .  . .. .... ..         .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2700
FMV-Can01       TTAGCAGCTTATTCAAATATTTATAATAAAAAAGAAAACATAAGGCAATCTATATGCAGT 2700
                .. .  . .. ..    . ... .  .      .      .  ..   . . . ..  ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2760
FMV-Can01       GAATTAAGTATTAATAAGCCATTACTAGCATTGAAACAATTCTCTAGTACAAAATCTATG 2760
                .  ..  .. ..  .  .   ..  . .  ...      .. . . ..      . . ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2820
FMV-Can01       GTGTCTAACACAAAGCAGGTCGAAATAGAATTACCAAAAAAATTACCTGATAATCTAGAT 2820
                .... .        .  ... .   . .  ..          ..   .. .  . . . .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2880
FMV-Can01       CTATATGGGTTAGAGAATATCATATCGAATTGCAAGATTGATGAAGATATGGAGTTCATG 2880
                 . . ..... . .  .  .  . . .  ...   . ... ..  . . ... ...  ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2940
FMV-Can01       AGATTTATAAATTCAGAAATTTACAGAATTAATTCAAAGAGATTAAAAAATGAAATGAGT 2940
                 . ... .   ..  .   ...   .  ..  ..   . . ..      ..   .. ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATAGAGACA 3000
FMV-Can01       AGTTCCACAAGAGAAGAAAAAAAGGAAGTTACTTTATTACCTGAGATATATATAGAAACA 3000
                 ...      . .  .       ..  ...  ... ..   .. .***********.***

FMV-Hel         GAGAAAGGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3060
FMV-Can01       AAGAAAGGTGTGAAATTTTTAAGGATAAAAAGGCATGAATACACTAGATTGATCAATGGT 3060
                .********...   .....  .. .     ..  ..  .    . . ... .   ....

FMV-Hel         NNNNNNNNNNNNNNNNNNNATGCTAAGGTCTTTGATGAATTNNNNNNNNNNNNNNNNNN 3120
FMV-CAN01       GACTATATTAGACAGAACAATGCAAAGGTTTTTGATGAATTTTACAAATTATCTGAAGAG 3120
                .  . . .. .   .     ****.***** ***********..     .. . ..  . .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTTAATACATGAGGATGAATTANNN 3180
FMV-Can01       TATGAGCTCGTCTCACTCAAGGACGCATACTCTAAGTTAATACATGATGATGAATTACTG 3180
                . .. . . .. .   .   ..  .  .  . .    *********** ********* ..

FMV-Hel         NNNNNNNNNNNNNNNNAAAGACCAAAGAACTGCTGAKGATAGAGAAATNNNNNNNNNNNNN 3240
FMV-Can01       ATTCGAATATTTTATAAAGATCAGAGAACTGCTGATGACAGAGAAATATATACAGGTAAT 3240
                 .. .  . .... .***** **.***********.** ********  . .   ...  .
```


FMV-Hel NNNNNNNNNNNACTATGTTTATATCCAATAGAAAAAAAANNNNNNNNNNNNNNNNNNN 3300
FMV-Can01 GCTCAAACTAGACTCTGTTTGTATCCAATAGAAAAAACATATAAGGCAATTTGTAAACAT 3300
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3360
FMV-Can01 ATCCCAGGTGAAGCAATAACAATCAGCGGTGATCAAAAGCAAAAAAAACTACTGGAGCAA 3360
FMV-Hel NNNNNNNNTCTGATWAAASAAAAGAAACAAAARATTAGAGANNNNNNNNNNNNNNNNNNN 3420
FMV-Can01 AGACTACAGCTGATCAAAGAAAAGAAACAAAAAATTAAAGATAAAAAATGGGCTGAAATC 3420
FMV-Hel NNNNNNNNNTCATCAGATGCATCCAAATGGTCTGCWAGAGNNNNNNNNNNNNNNNNNNN 3480
FMV-Can01 TATTCTGTATCCTCAGATGCATCAAAATGGTCTGCTAGAGATGTGTTCTTAAAATTCATA 3480
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3540
FMV-Can01 ATCACAATATCTGTAAATCCATATCTCCACCCTGATGAAAAATGGTTTTTATTATATCTG 3540
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAATAGCATATTCAATAAAATNNNN 3600
FMV-Can01 TGTATGACCTATTACACCAAAAAAATTGTGTTTACAGATACCATATTCAATAAAATGGTT 3600
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3660
FMV-Can01 GATCTAGCAGATAAAGATAAACATGGAAAGTATGAGTTTATGACTGAAGACTTTACACAA 3660
FMV-Hel NNNNNNNNNNNNNNTWAGGAGCAACTGGTTACAAGGCANNNTCAACATGATATCATCATTT 3720
FMV-CAN01 AACTACTTTACAGTAAGGAGCAATTGGTTACAAGGCAATCTTAATATGATATCATCATTT 3720
FMV-Hel GTCCACCACTGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3780
FMV-Can01 GTCCACCATTGCTCCACATTGTACACTGAGACCATGTTAAAGATTTTATCTGATAAAAAT 3780
FMV-Hel NNNNNNNNNNNNNNNNNTGACATCAATGGTACAYTCTGATGATTCAACTTATGATTTCNNN 3840
FMV-Can01 TCTTTGGAATGCTCTATGACATCTATGGTTCACTCTGATGATTCAACATATGATTTTCTA 3840
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 3900
FMV-Can01 ATCTGTAGAGATAAAAAATCAAAGAATAGATATGTGAAAGAAGAAAACATTGGCAAACTC 3900
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCTAAACGAAAAAAAG 3960
FMV-Can01 ATTATATCCTTGATTACATACTCAAATAGATTTCATTGTATAACACTAAATGAAAAAAAG 3960
FMV-Hel ACMTATATTAGCACATTTTATAAAGAGTTTCTTTNNNNNNNNNNNNNNNNCAATGAACTA 4020
FMV-Can01 ACATATATTAGTACATTCTATAAAGAGTTTCTCTCTACAATTATAGTTGGCAATGAACTA 4020
FMV-Hel TTTTTTATTTATTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4080
FMV-Can01 TTCTTTTTTTATTTAGCTGATTTATTACCATTGACATCAGATACAAGTTATGCCTCACCA 4080
FMV-Hel NNNNNNNNNNNNNNNNTCATACTCTGGTTATATAAACNNNNNNNNNNNNNNNNNNNNNNNN 4140
FMV-Can01 TTGCAGGATTTAGCATCATATTCTGGTTATATAAATAATTCCTTTTCACATGCTTGCCCA 4140
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4200
FMV-Can01 AGAGAGATGATCCAAACAGCTATCATACTGATCAATCACTTGACATTATCCACTTATAAC 4200
FMV-Hel NNGCATTATACATCTGAAAAAAATCCANNNNNNNNNNNNNNNNNNNNNNNTTAACAATC 4260
FMV-Can01 ATGCAATATACATCTGATAAAAATCCAAGGTTGAATATACAATCTACTGATTTACCAATC 4260
FMV-Hel CAGATAAACCCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4320
FMV-Can01 CAGATATACCCTCGTTATAAAGTTCCAATTGAAATGGCTGGTATGATACCCTATTACTCA 4320
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4380
FMV-Can01 GCAGATGCATTCAATATCCTGAATGATATAGTATTGAAATTAGGTTACAATGACAAATTG 4380


FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4440
FMV-Can01 CAGTCAGAGCTCGTGGAGGATATTTTGACTGAAGAACTCATTAGTGATTATCTTAACATA 4440
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCATCTTGTGCATGGAT 4500
FMV-Can01 ATAAAAAAACATACACCAATGATTGTTAAATATATCAAATCATGTATCTTGTGTATGGAT 4500
FMV-Hel TACTCTCAATATGAAAGAGATGATGCTGACCCNNNNNNCATAATAGAYTATGATTTRAGY 4560
FMV-Can01 TACTCACAGTATGAAAGAGATGATGCAGATCCATATAATATAATAGATTATGATTTAAGT 4560
FMV-Hel CMAAAATCWWTTATNAATGTGATCAGTCTGAATAAAGGNNNNNNNNNAAAGAAAACTTAT 4620
FMV-Can01 CAAAAATCTATTATTAATGTCATCAGTCTTAATAAAGGCCAGCGATTAAAGAAAACTTAC 4620
FMV-Hel ACCTACAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4680
FMV-Can01 ACATACAAGAAGTATTTAGAATCAGAAACTAATGTACGATTAATATCTGCTATACATCCT 4680
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAATCAAATATCCTACAAAAT 4740
FMV-Can01 GAATGGTGTGTGATAAAACCAACTGACCCAGAGTTAATTAAATCAAATATATTACAAAAT 4740
FMV-Hel TATACCAATCCCAATTTCAGAGATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 4800
FMV-Can01 TATACAAATCCCAATTTTAGAGATGGATTAATATTTTCCACTCCTGCAATTGATTATGGT 4800
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNGAATATGTATACTGTAAGTTCACATNNNNNNNNNNNNN 4860
FMV-Can01 CGTAGAGTTATTAGTTCAAACAAGAATATGTATACCATAAGTTCACATATTATGGAAAAA 4860
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATCATCAGAATGAGAAAATAGATGNNNNN 4920
FMV-Can01 GATAAAGCAAAAAATATAAAAACAGTATACAATGATCTAAATGAGAAAATAGATGGAGTT 4920
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGTCTTYACCTGTTAAGTGATAAAAAA 4980
FMV-Can01 GAAATAACAGCTAAGGATCTACAAAGATATCTAAGCCTTTATCTGTTAAGTGATAAAAAA 4980
FMV-Hel ATCTCCATGGCTGTACAAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5040
FMV-Can01 ATCTCTATGGCTATACAAGTATTTTTCTCTAAAATAGAAACAGTGACAATGGCCAGGCCA 5040
FMV-Hel NNNNNNNNNNNNNNNNNNNNNCCTAGAAGCATATATTCTGAAGATTTTGGTAAATANNNN 5100
FMV-Can01 TCATATAATAAGGTTGTACAACCAAGAAGCATATATTCTGAAGATTTTGGTAGGTACTCT 5100
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5160
FMV-Can01 AATAACTCATTGATAGAATCTCTTTTAACTTCAAGATATTGTAAAATAGCTACAGTTGAT 5160
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5220
FMV-Can01 CCAAAAGTTGAAAAGTTTGTTGATGTTAGTGATTATGTTTTAGAAAGAATTGGTGATATC 5220
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5280
FMV-Can01 AAACTATATGAATCACCAGAAGACATTGATGATGACTACATCAATTATTACCAATTTAAA 5280
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5340
FMV-Can01 TATCCTCATTCTGATGAGGTTATGTTAGATGTTGAGCCTGTTTCTGATCTAGGTGATCTT 5340
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5400
FMV-Can01 GAACTCAAAGTTTACAAAAACAAGTTGAAGTTCATGTCATTGCTTGTAAAGTACTATAAT 5400
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5460
FMV-Can01 GATATTAAAAAAACTATTGAAAATCCCAATTACAATATTCCCAATTACCCATCTCCATCT 5460
FMV-Hel NNNATAATAATGACGATAGACTCATTGCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5520
FMV-Can01 TCTATAATAATGACAATAGATTCATTGCTAAAAAAAGATGAGATCAGTACTAAAGTATAC 5520


FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNATGATGATTACCTGTTGACTAGATTTGGGATGTAT 5580
FMV-Can01 TTATCAAATGTAAAATCTATTAAGTATGATGATTATTTGTTGACTAGATATGGTATGTAT 5580
FMV-Hel GCNNNNNNNNNNNNNNNNCAAAGTACAAGTTAGGCTACAAAGTAAAAATCGNNNNNNNNN 5640
FMV-Can01 GCCCATAATGATTTTCACATTAAGTACAAGTTAGGCTACAAAGTTAAAATTGCATCTAGT 5640
FMV-Hel NNNNNNNNNNNNNNNNNNNNCAWACATAYAAAGACACATATGANNNNNNNNNNNNNNNN 5700
FMV-Can01 ACAAAGATATCACACACATTACAAACATATAAAGATACATATGAACCAGTAGCATTTGTC 5700
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5760
FMV-Can01 ACCAAACTAATATCTAGAAATCTGGAATTATTTACTTCACTTAAAGACAATGAAGATTTT 5760
FMV-Hel NNNNCTGGACATCATAAATTTAATGACATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5820
FMV-Can01 GTCACAGGACATTATAAATTTAATGATATGATAAACACACTTCAGAATAGTCGAGATATC 5820
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5880
FMV-Can01 AATAGCAGTGCATTATTATTAATGTTGGATAAAATACCAAATTCGAGATTCATGATTACA 5880
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5940
FMV-Can01 CTACTTAACGATAATAGGGTTTATAATCACTGGCTGATACCAACAAATAGTTTACAAGAA 5940
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6000
FMV-Can01 GACCCTGGCGCCTCTTTAGCTTATTACATGTGTCAGGGTAATATAATGAAAGTTAGAACC 6000
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6060
FMV-Can01 TTTACCAGAGGTGGTTCCATAGTATTCACAATGACTTACTACAAGTATGTAAACCGAGAC 6060
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6120
FMV-Can01 CTTGGTGCACTTGTCAATATTAAAAAGAAGATATCAATGGATTATGTTGAATTATTGAGG 6120
FMV-Hel NNNNNNNNGGTCACAAATAGAGTTGAACCANNNNNNNNNNNNNNNNNNNNNNNTTAAT 6180
FMV-Can01 AAATCTTTAGTCACAAATAGAGTTGAACCAACATCTTCGTATGACACTTATAATATAAAT 6180
FMV-Hel GAATATGGTAGATACACWGAATACAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6240
FMV-Can01 GAATATGGTAGGTACACTGAATACAAAGGTCCAAGGGTATTTAATGTTGCACAGATCAAT 6240
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6300
FMV-Can01 TATGCCAAAATATTAGACATTAAACCAGCCTATACTGAAATTGATGATCAAGTGAAGCTT 6300
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACWTAT 6360
FMV-Can01 GTTTTGAATATTGGAACCAATGAAGATCACTTCACATTTGATATACGTTTACGTACATAT 6360
FMV-Hel ATTGATGAWGAGTATTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6420
FMV-Can01 ATTGATGATGAGTATTATTTGAATTGTTTGCTAGATAATTTACACATAGAGCGGGAAAAT 6420
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6480
FMV-Can01 ATTCTGGGCTATCTATGTGAAGGTACATTTTTTCATAGGTACCCTGATTATCTTAAGCGA 6480
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6540
FMV-Can01 ATTCTTGCTACACTTGGACCAAATCAACTTGTATCTTTGATGTCTAATTATACTGATTCA 6540
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6600
FMV-Can01 GATTGTATAACTGATAACATAGATGTGAATCGATATGGGTTCCTGATGGAGATTAGCAAT 6600
FMV-Hel NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6660
FMV-Can01 TACTTCAGAGATAAAAATAATATATATATGGCAGAATTGTGTGAGACATTAGAGATTGTA 6660

```
FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6720
FMV-Can01        GCAATGAGCACTAATGTGAACATACTAAATGTTAGAAACCCTGAAAAATTTATAAACAAC 6720
                 .   .. .   .  ....    .  .   .... .       ..     ... .

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6780
FMV-Can01        TGCAATAAAATTAAGCTCCATCCATTATGCCATGAATATGTGATACATAAATATCGTGCA 6780
                 ..   .    ..  . .   .   .. ..   ..  . .... .   .   . . ...

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6840
FMV-Can01        GATGAAGAAGCACCATATTATAGATTATTTAGATTAGTTATGAAATACAGATACAACAAG 6840
                 . ..  .  .    . .. . . .. ... . .. ... ..   .   . .       .

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6900
FMV-Can01        AGCCCTGTAGAGAAAATCATCCTATTCATCATCTTGATATTTAGATTTTACATTCATGAA 6900
                  .   ... . .   .  .  . ..  .  . ... . ... . ....   ..  ..

                                                 ↓ORF1 end
FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6932
FMV-Can01        TACATCGATGTTGAAGCTGAGTTTGAATATTG 6932
                 .   . . .....  . .. .....  . ...
```

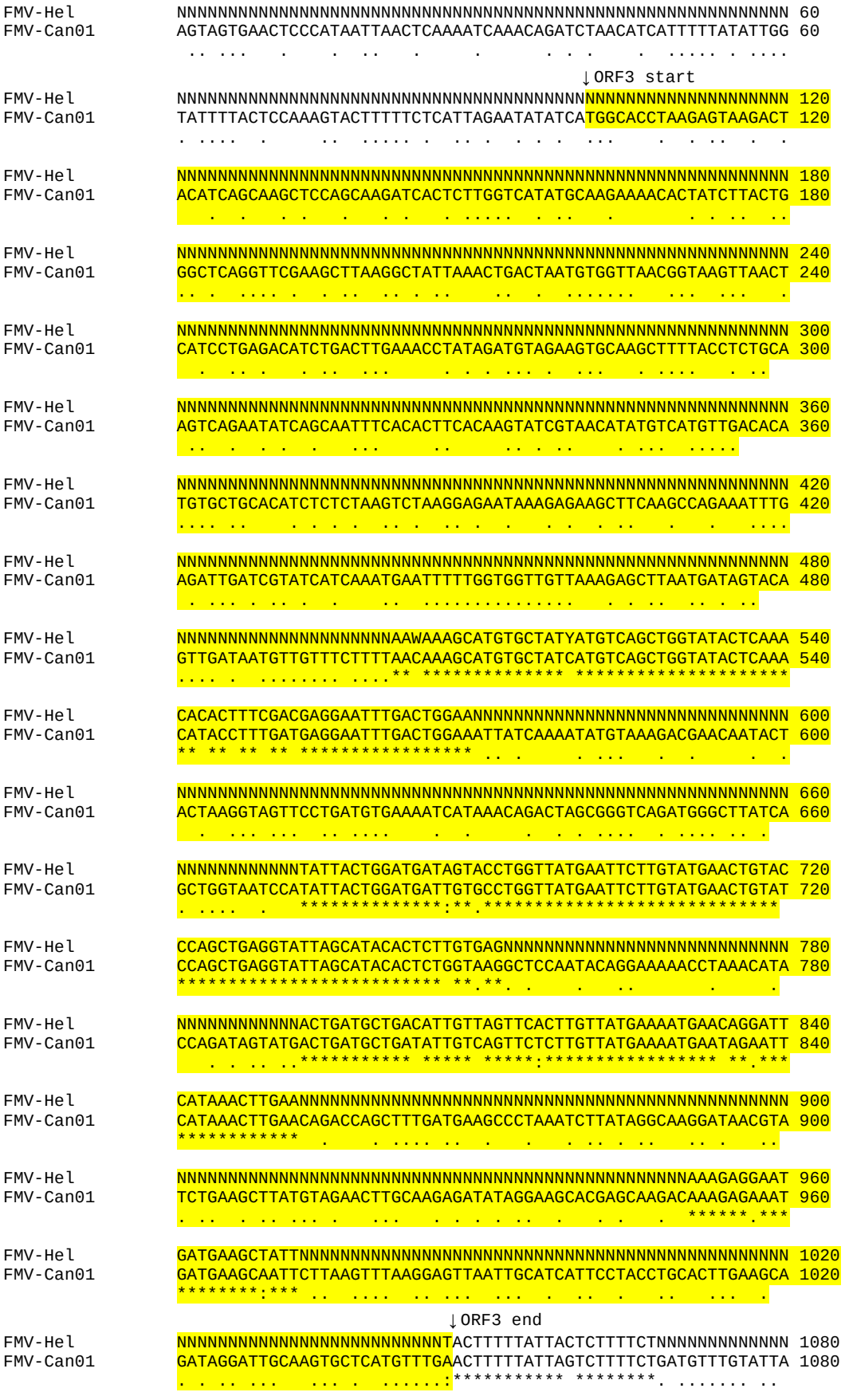
FMV RNA2-putative glycoprotein precursor gene

```
                                                                          ↓ORF2 start
FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 60
FMV-Can01    AGTAGTGAACTCCTCTTAAACAAATCAATCTGATAAAAATCTAAAAGCAAATATGATGAG 60
              .. ...    .  . ..        .   . .. .       . .     .    . .. .. .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120
FMV-Can01    TATTATTAACTCAGTAATAGCGTTACTTATAGTTGCAATCCTGTGCAACACTGTGAAGAA 120
             . .. ..    .   ..  .  . . ...  .. . ....   .   ....      ....  .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 180
FMV-Can01    TGCAGTCTACATAGAAGAAGAATTATGTGACTGTCAGCCTTCAATATTAAAACTTGGCAA 180
             ..   .. .   .  .  .   .  .. ....  ...  .   ..   . ..     ....

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 240
FMV-Can01    AGGGTTTATAGTGTGTTTTAAAGGTTGTGATGTAAAGCCAGTAAACTCTAAGCTCTACAA 240
              ...... . .........   ....... ...   .    ..    .  .   .  .  .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 300
FMV-Can01    CACAACTTGCAGTCACATGAAAGAAGTAACAATAACAATGTGCAAAGGTAATCGATATGT 300
                   ...  ..    ..   .  ..      .     ....    ...  .  .  .  ...

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 360
FMV-Can01    CACATCATTACCTACTGCAACTGTCCATGAAGAATACACAATCACATCAATCATCCATAA 360
                 .  ..   .  ..   ...   ..  .  .      .    .   .  .   .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
FMV-Can01    GATATGGAAAGTGGTTATCACAATTACAGTTTGGCTGATCTTTATTGTATTAAAGGTCCC 420
             . . ...   ...... .     ..    ...... .. . ... .... ..   ...

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTTCCATGAACTCCAAGCT 480
FMV-Can01    AATTCTGTGCTTACTATCAGTTGGGAATCAGGTATTCAATGGTTTCATGAAATCCAAGCT 480
               .. .... ..  .  .  ......  .   ... ..   ..*** ******.********

FMV-Hel      GANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 540
FMV-Can01    GAGAGAGTGTGACACATGCAAAACCAAATATTCTGTATCGCATGTTGATTGTCCTACACC 540
             **. . .....    ..           . .. ... . .  ..... ....   .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
FMV-Can01    AAGTTTCAAGCTAAGGACAGATTATAATCTTTTTTTCTACACTTTCCTCACTGTGTTACT 600
              ....  . .  ..   . .. .  . ....... .    ...  .   ......  .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
FMV-Can01    CTTTGTGACACTTGCAAAGGCAGATGATAATGAATTCAATTTTTATCATCATGGAAATCA 660
              ......    ...    ..  . .. .  ..  ..   ..... .  .   ...   .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 720
FMV-Can01    GACTGAAGTACAAATCTTAGACAGGGAACATTATAAACAAGACTTCGATGTCCAAGGCTA 720
             .  ..  ..     . .. .   ...    .. .        .  .. . ...    .. .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 780
FMV-Can01    CCTGTACACTGTTACTGTCTTAAATTCGCATCTTGAGATACAAACTATAAATGTTTCTGA 780
               ...     ....  ... ..    .. .  . ... . .        . .   ..... ..

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 840
FMV-Can01    AATATTGATTCCAACAAAGCATATTTTGACACATGAACACTTCAGTTGTGATGGCTCTGA 840
               . ... ..         .  . .....     ..      ..  ...... ... . ..

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 900
FMV-Can01    AGGCTGTAAAAAAGAGTGTCAAGCTAGGACTGGTTCCATGCCTATGTATTACTTAAAAAA 900
              .. ...      . ....   . . ..  .....   ..  .  ... ..   ..

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 960
FMV-Can01    GGCTTATGATGGATTCTCATGTCTTTTTACGAGTGCAACTTTGTGTGGTCTTTGCAAGAG 960
             .. .. .. ... .. .   ... .....  . ...     ......... ....    . .

FMV-Hel      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1020
FMV-Can01    TGAAATGAAGACAATAGGCTACAAAGTAACTACAACAAAGGTTTCACCATATATTGATAT 1020
             ..   ..  .    . .. .      ..   .        .....     . . ... . .

FMV-Hel      NNNNNNNNNGCATGGCAATAAAACTGATATCATAAAGATAAGAGAATTCTCTAAGTTNNN 1080
FMV-Can01    TGAAGTTGTTCATGGCAATAAAACTGAAATGATAAAGATAAGAGACTTTTCTAAGTTTAT 1080
             ..  .....  ****************:** ************** ** ********  .
```

```
FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
FMV-Can01       ACATGAACCTTATTATGTAAAACCAATAGAGCCAGTGTGGGTTGAATCTGTAGATATGTT 1140
                   ..    .. .. ...        . . .   .........  . ... . . ....

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATGCCRAGTTATGGATGCTT 1200
FMV-Can01       TGTTACTGGTACAGATGTTTATGTTGGTCAGATTTGTAATATGCCTAGTTATGGTTGCTT 1200
                ....  ....   . ..... .......  . .....  .***** ********:*****

FMV-Hel         TGGTCCCAATTATATAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1260
FMV-Can01       TGGTCCAAATTATAAAAAAGATAATAAAACTTTTGTTATAAGTGCACCTAAGGTTAAAGA 1260
                ******.*******:****. .  .     ....... .  ...    .  ....   .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTGTATAGATCCAGGTAATAGTGA 1320
FMV-Can01       TCCTATGACACATGATCGTGAATTGATATTAGAACACTGTGTTGATCCAGGTAATAGTGA 1320
                .  . ..     .. . ...  ... . .. .    ****.*:*****************

FMV-Hel         TATCAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTAATCAGACCATA 1380
FMV-Can01       CATAAACAGTCTGCAAAAAACTCAATCAATTTACCATGATGGTGTAATTATTAGACCATA 1380
                 **.**  .. ..        .   .   ...    .. ......  *:** ********

FMV-Hel         TGAATTTGGATTGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1440
FMV-Can01       TGAATTTGGTATGTTATCTATTGGGTTTCCCATAGTGGGCAAGCTGATAGGTAATTTCTG 1440
                **********:.:***. . . ........    . .....   . .. . ...  ... ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGTTGTTATGATTGTCAATCAGG 1500
FMV-Can01       TGAAAAGCCTGCTGATGTTAAAGAAATAGTGGTTAATGGTTGCTTTGATTGTCAATCAGG 1500
                ..    .  .. .. ....   .   . ......  .***** *:***************

FMV-Hel         TATAGARGCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1560
FMV-Can01       TTTAGAGGCTAAAATATCATATAAATTAACTGAAAGGTGCGGGAAAATCATATGTTACTT 1560
                *:**** ***    . .  . .   ..   ..   .... ...    .  . ....  ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1620
FMV-Can01       TGGTCAAGTGAAGTATGAATATTTTGTAGATCATGACTATAATGACATAAGTATTCATAC 1620
                ....   ...  .. ..  . ...... . .  ..  . .  ..   .  .. ..  .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1680
FMV-Can01       ATGTTATGATAAGAAAAATATCATAATCAAGTGTAATGACTACTCAAACACATTTTCATT 1680
                 .... .. .  .     . .  .  .   ....  ..  .  .        ....  ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATGGWGTTCATGGTTCWGCTGA 1740
FMV-Can01       AGATCACAGTAAAGATACAAATTATTATGGTACTAGTAATGAAGTTCATGGTTCAGCTGA 1740
                 . .    ..   . .     .. .. ....  . .. ***. ********** *****

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1800
FMV-Can01       TATGGAATTCAACATGCTGAAACACTTGCCCAATCTATTATTTAATCCTAAAGTTGTTGT 1800
                . ...  ..     .. ..      ...     . . .. ...  .  .   ........

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1860
FMV-Can01       AACCACTATGTTAATAATAGCTATGACTATTTATATGGCTTATTGTATATCTAAGCAATT 1860
                      . ....  .  . . . ..  . ... . ... .. .... . . .  .   ..

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1920
FMV-Can01       ATATAAGCATTATGTTAAAGTTAAGATGGACAGGATTATAAGGTATCATAAAAAAACAGA 1920
                 . .  .  .. ....   ...  . ...   .. .. .  ... .  .         .
                                                                    ↓ORF2 end
FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1980
FMV-Can01       TACCATCACAGATGAATCAGCAAATGAATTCATTGTAGTGACTGGTTCTGCACAGTAATA 1980
                .    .    . .. .  .    ..  ..  .... ...  ..... ..    ..  .

FMV-Hel         NNNNNNNNNNNNNNNNAATTCCCTAAATCCAAATATCNNNNNNNNNNNNNNNNNNNNNNNN 2040
FMV-Can01       ATATTTACAATATAAAAATCCCTAAATCCAAATTTCTTTCTAAAGCTTTTTTTTATTTCT 2040
                 . ...    . .  **:***************:**... .   . ........ ... .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2100
FMV-Can01       GATGCGTTGCATCACTCATCCAGCAAGCTGTATGGCTATACATGGTGAAAAGCCTGCAAG 2100
                . .. ....  .   .  .   .   . ... ... . .   .....    .  ..   .

FMV-Hel         NNNNNNNNNNNCTCTCAAGGTTGATGCATATGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2160
FMV-Can01       TTGGTCTATGCCTCTCAAGATTGACGCATATGTACCCATTGTATGCAACTTTTTCCTAAA 2160
                ..... . ..  ********.**** *******.     .... ..    .....  .

FMV-Hel         NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2220
FMV-Can01       AGTTGTATACCTGCTGGATAGAGATGTATACTCTAATCCTTGATTTCTTTAAGCTTTTTT 2220
                 ..... .   .. ... . . . ... .  . .  .  ... ... ...  . ......
```

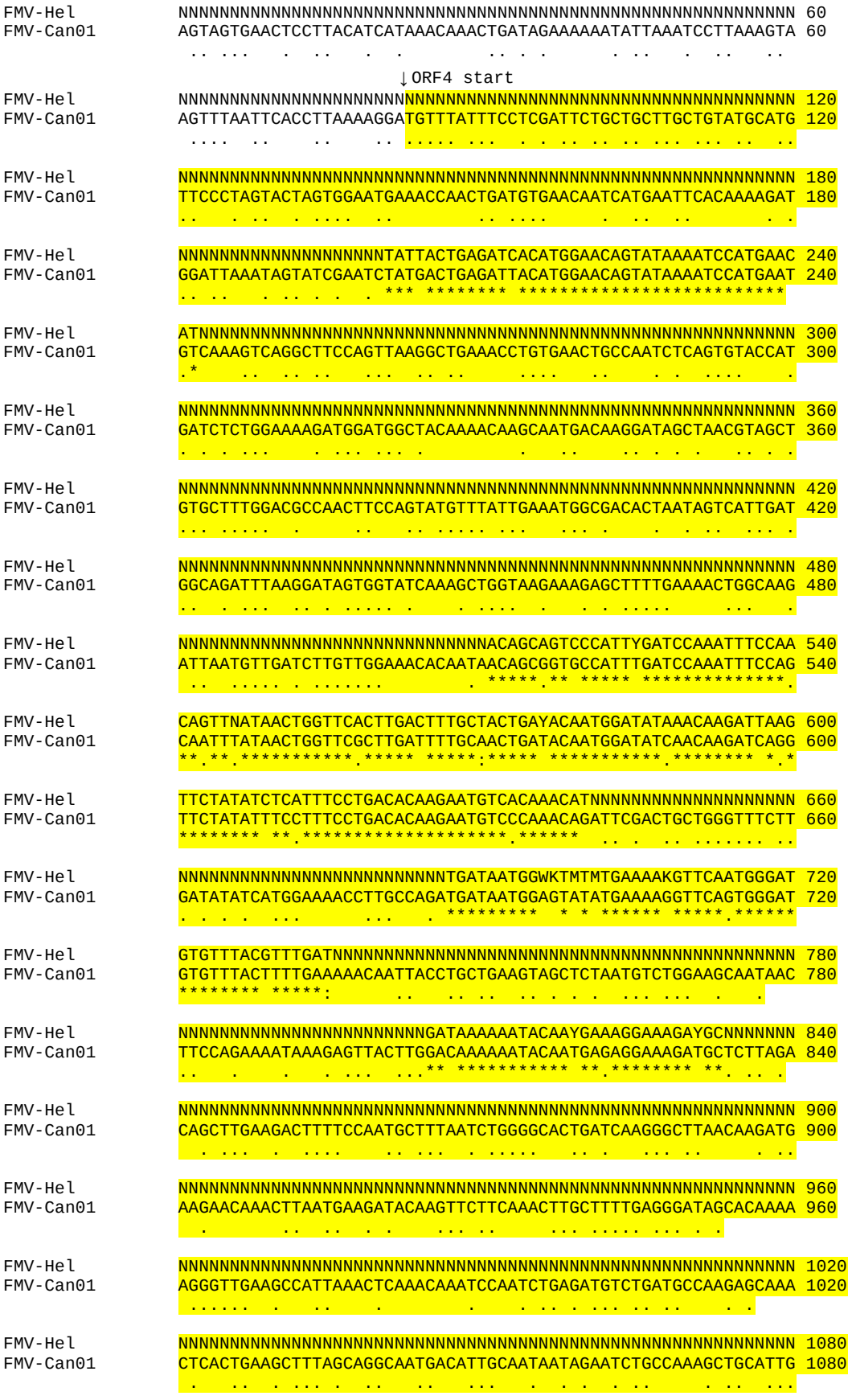
```
FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNN 2248
FMV-Can01        TGATTTTGTTATGTTAAGAGGAGAACAC 2248
                 .. ....... ....  . .. .
```

FMV RNA3-putative nucleocapsid protein gene



```
FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140

FMV-Can01        GTTAAATAATAAATTTAATATAATTAATAGTATATCTTAGTATATATTAATGTAATCTAA 1140
                 ...   .  .   ...  . .  ..  . .. . . .. .. . . .. ...  . .

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1200
FMV-Can01        GTGTGTTATTAGCAATATATATCTATGTGTGTACTTGCTAAGAATACAAAATTAAAAAAA 1200
                 ....... .. .   . . . . . ........  ... .  .  .      ..

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1260
FMV-Can01        CAAAAATATATAAGATGATGTAAGATGTGTAAGTGTGTATATATGTTCTTCAAGTTTCTT 1260
                       . . .  . .. ...  . .....  ...... . . .... ..   .... ..

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAAA 1320
FMV-Can01        AAATATCAACAGTGTTCAATATATGTATTGTGATCTATTATATGTTTATCATGAAAAAAA 1320
                    . .     .....   . . ... ..... . . .. . ..... .  ..  *****

FMV-Hel          AMCAAAAAAAWSAAAAAAATCATAAAAAGACAAAAAAAAGAWATATATRTATATATATGT 1380
FMV-Can01        ACCAAAAAAAGCAAAAAAATCACAAAAAGACAAAAAAAAGACATATATGTATATATATGT 1380
                 * ******** .********** ****************** ****** **********

FMV-Hel          GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1440
FMV-Can01        CGAAACTCCGTCTTGAAAATCTATCTGACACGTATTTTTCTTTATTGATACCATTGTTAG 1440
                  .    .  .. ...    . . . ..    .. ..... ... ... .    ..... .

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1486
FMV-Can01        ATTTGTTAAGATTTGACTGATTTAGTTTTAAATTATGGGAGAACAC 1486
                  ......  . ....  .. ... .....   .. .... .
```

FMV RNA4-P4 gene



```
FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1140
FMV-Can01        GTAGCAGTATTCAACAAGTATAACATTCCATTACCAGTGTCTGAAAACACTGAGGAAGGA 1140
                 .. .  .. ..      .. .    ..   ..    .... ..       .. ..  ..

                                             ↓ORF4 end
FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATTTTTTTGTTTTTTTCTCTTTNNNNNNNN 1200
FMV-Can01        TATGTTATCATTGGTAATGCTCAGTAATCTATGTTTTTGTTTTTTTCTATTTTAAACTTA 1200
                 . .... .  .....  .. .  ..  . *** ***************.***.    ..

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1260
FMV-Can01        TCTGTCTATTGCTGTGTTTTGTTTACCTATTTTATTGTAATGTTGTGTCTATCTATATAA 1260
                 . ... . ... ............   . .... ....  ........ . . . . .

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAA 1320
FMV-Can01        ATTATGTACTTCTTTCACTTTGTTTAAATGTTTCAATTATGTTATCTTAGAAAATCCAAA 1320
                  .. ...  .. ...   .......   .....   .. .... . .. .    .   **

FMV-Hel          AAAAAAAAAACCATAAAAAGYGGTTGCTTGCAAYCCNNNNNNNNNNNNNNNNNNNNNNNN 1380
FMV-Can01        AAAAAAAAAAAACCTAAAAAGAGGTTGCTTGCAACCCATCTAAGTCCGATATACTAAGGAA 1380
                 **********.*.******* ************ ** . .  ..  . . .  .  ..

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1440
FMV-Can01        CTCTGCAAGTTTTTAGGATGTTGATTACTTAACTTTTGATATTTTTCTTAGCTTTTGATT 1440
                  . ..   ...... .. ..... ..  ..   ..... . ..... .. . ..... ..

FMV-Hel          NNNNNNNNNNNNNNNNNNNNNNNNNN 1466
FMV-Can01        GAACAATTTTGTTATGTAAGGAGAAC 1466
                 .     ....... ...  .. .
```