Genome sequence and variation in the Australian native *Velvet tobacco mottle virus*

by

Kieren Arthur

MSc (Hons), The University of Auckland, NZ

A thesis submitted for the degree of
Doctor of Philosophy

at
The University of Adelaide,
School of Agriculture, Food and Wine

February, 2011

- Table of Contents -

- Thesis Summary -

Velvet tobacco mottle virus (VTMoV; genus Sobemovirus) infects *Nicotiana velutina* (Velvet tobacco), a native of the arid region of central Australia. In the field, the virus, mirid vector and native host plant together comprise a unique plant virus pathosystem which is well adapted to its ecological niche, and independent of anthropogenic influences. The purpose of this research was to describe the sequence variation amongst VTMoV isolates and relate this to ecological factors.

The full genome sequence of VTMoV was obtained using a genome walking strategy with both degenerate and specific primers. Sequence and genome organisation confirm that VTMoV is a unique sobemovirus and phylogenetic analysis groups it separately from other sequenced Australian sobemoviruses. This is consistent with the hypothesis that VTMoV is not a recently introduced sobemovirus, but rather a product of evolution within a unique Australian ecosystem, representing a novel plant virus lineage. The genome sequences of two isolates of VTMoV, K1 and R17 were compared and a limited amount of variation was observed between these isolates.

Sequence diversity was observed in the RNA dependent RNA polymerase (RdRp) gene from 15 isolates of the virus. Analysis determined mutations were limited to maintain protein function, which is indicative of purifying selection. In addition, the first evidence of recombination in the RdRp of a sobemovirus was detected in three VTMoV sequences.

Sequence variation associated with transmission of VTMoV by the mirid *Cyrtopeltis nicotianae* [Hemiptera; Miridae] was also investigated. Isolates K1 and R17 were serially passaged monthly either through obligatory mirid transmission or via mechanical inoculation. After two years, sequences were compared from two regions of the VTMoV genome associated with movement (open reading frames of protein P1 and the coat protein). Several different trends were observed in the sequences, but only one difference could be associated with the mode of transmission. The coat protein region sequence from mirid transmission had a higher mutation rate that sequence from the mechanical mode of transmission.

This thesis contains the first complete genome sequence of VTMoV. It describes natural variation amongst a range of VTMoV isolates, and assesses sequence variation in parts of the genome after long term mirid transmission. The sequence of VTMoV is discussed in the context of the unique nature of the virus and the evolutionary mechanisms that may have played a role the evolution of the virus.

- Statement -

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Kieren Arthur and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works.

List of published works:

Arthur, K., S. Dogra and J. W. Randles (2010). "Complete nucleotide sequence of *Velvet tobacco mottle virus isolate K1*." Archives of Virology 155 (11): 1893-1896.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Kieren Arthur

- Acknowledgements -

The process of scientific discovery is, in effect, a continual flight from wonder.

Albert Einstein

Firstly, thank you to my supervisors Prof. John Randles, Dr Nick Collins and Dr Satish Dogra for all of your advice, mentorship and support during this project.

A gigantic thank you to all the morning tea, coffee and lunch buddies that I had, especially Steve, Ellena, Robin, Christine, Hayley, Dale, Cathy, Amanda, as well as the other numerous people on Waite campus. Your support through the tough times and the celebrations of the small successes was what kept me going some days.

Thanks also to me friends and family back home in New Zealand, your love and support was felt across an ocean.

Thank you to Dr Allan Gould for the supply of some primers for the sequencing work and finally thank you to the C. J Everard scholarship for their funding support.

Dedicated to my grandfather,

Whose adventurous spirit encouraged me to dream

- List of Abbreviations -

Abbreviation guidelines of the *Journal of General Virology* were followed for this thesis.

CP- Coat protein

DIBA- Dot immunobinding assay

d.p.i – days post inoculation

gRNA- genomic RNA

LB- Luria broth

mg- milligrams

MP- Movement protein

nm-nanometres

sgRNA – subgenomic RNA

VPg- viral genome-linked protein

WINC – Waite Institute culture collection

Virus abbreviations

Names not in italics signify species yet to be officially recognised by the International Committee on Taxonomy of Viruses (ICTV)

BBTV- *Banana bunchy top virus*

BSSV- *Blueberry shoestring virus*

BYDV- *Barley yellow dwarf virus*

CMV- *Cucumber mosaic virus*

CoMV- *Cocksfoot mottle virus*

IYMV- Imperata yellow mottle virus

LTSV – *Lucerne transient streak virus*

PLRV- *Potato leaf roll virus*

PLYV- Papaya lethal yellowing virus

PVY – *Potato virus Y*

RGMoV- *Ryegrass mottle virus*

RuCMoV – Rubus chlorotic mottle virus

RYMV- *Rice yellow mottle virus*

SBMV- *Southern bean mosaic virus*

SCMoV- *Subterranean clover mottle virus*

SCPMV – *Southern cowpea mosaic virus*

SeMV- *Sesbania mosaic virus*

SNMoV- *Solanum nodiflorum mottle virus*

SoMV- *Sowbane mosaic virus*

TMV- *Tobacco mosaic virus*

TuMV- *Turnip mosaic virus*

TRoV- *Turnip rosette virus*

VTMoV- *Velvet tobacco mottle virus*

WSMV- *Wheat streak mosaic virus*

- Chapter 1 -
General Introduction

1.1 *Velvet tobacco mottle virus*

Velvet tobacco mottle virus (VTMoV) is a virus that has adapted to survive in a unique ecological niche and provides the opportunity to study the unique features of this native Australian plant virus. VTMoV is also classified as a wild plant-adapted virus (WILPAD) (Harrison 1981) because it is a virus that infects a single native host species adapted to the central arid region of Australia.

The virus was discovered after observing virus-like symptoms on leaves of the Australian native *Nicotiana velutina* (velvet tobacco) plant. *N. velutina* is a native that grows in sand dunes and dry river bank regions throughout semi-arid regions of central Australia (Botanic Gardens Trust; 2007). Virus symptoms were first observed in May 1979 at Cobblers Sandhill (29°46';139°57') and then again in November 1979 near Innamincka (27°46';140°42'). Both locations are on the Strzelecki track (NE South Australia) (Randles et al. 1981). The natural host of the virus seemed to be limited, as it was not detected in *N. glauca* plants growing at Cobblers Sandhill and the experimental host range of VTMoV was limited to *Nicotiana* spp (Randles et al. 1981). In *N. clevelandii* symptoms included prominent rugosity of leaves, and a yellow mosaic.

1.1.1 VTMoV taxonomy

VTMoV was classified as a species in the genus *Sobemovirus* [type species *Southern bean mosaic virus* (SBMV)], a 'floating' genus that is yet to be assigned to any existing plant virus family (Fauquet et al. 2005). VTMoV was classified based on gross particle morphology, an icosahedral shape with an approximate size of 30nm (indistinguishable from SBMV), the size of the genome, and its distribution throughout infected cells (Randles et al. 1981). In addition, VTMoV is serologically related to *Solanum nodiflorum mottle virus* (SNMoV), another species of the genus (Randles et al. 1981), which is endemic to sub-tropical eastern Australia (Greber & Randles 1986).

1.1.2 The genus *Sobemovirus*

There are currently thirteen recognised species within the genus and three recently discovered tentative species: Imperata yellow mottle virus (IYMV) (Sérémé et al. 2008); Rubus chlorotic mottle virus (RuCMV) (McGavin & MacFarlane 2009); and Papaya lethal yellowing virus (PLYV) (Amaral et al. 2006). All sobemovirus species are readily transmitted mechanically and natural vectors include beetle and aphid species. Details of host range, geographic distribution and modes of transmission of sobemovirus species are listed in Table 1.1.

Table 1.1: List of species in the genus *Sobemovirus* with their acronyms; GenBank accession number; natural host range; geographical distribution; modes of transmission and the presence of a virusoid RNA sequence with its GenBank accession number. A '+' indicates that this feature has been proven for the species, and a '-'indicates that the feature was proven negative or not detected, and the use of 'nd' indicates where no data about this feature could be found.

				Transmission				
Species & Acronym	GenBank accession number	Natural host range	Geographic distribution	Vector	Pollen or seed	Soil	Virusoid RNA	Reference
Blueberry shoestring virus (BSSV)	Nd	*Vaccinium corymbosum, V. angustifolium*	North America	Aphid	-	-	-	(Ramsdell 1979)
Cocksfoot mottle virus (CoMV)*	AB040447	*Dactylis glomerata, Triticum aestivum*	France, UK, Germany, Japan, New Zealand	Beetle	-	-	-	(Catherall 1970)
Imperata yellow mottle virus (IYMV)	AM990928	*Zea mays, Imperata cylindrical, Rottboellia exaltata*	Africa	nd	-	nd	nd	(Sérémé et al. 2008)
Lucerne transient streak virus (LTSV)	U31286	*Medicago sativa*	New Zealand, Australia, Canada	nd	-	-	NC_003798	(Forster & Jones 1980)
Papaya lethal yellowing virus (PLYV)	Partial GU066876	*Carica papaya L. Jacaratia heterophylla, J. spinosa, Vasconcella cauliflora, V. quercifolia, and V. monoica*	NE Brazil	nd	nd	nd	nd	(Nascimento et al. 2010).
Rice yellow mottle virus (RYMV)	U23124	*Oryza sativa, O. longistaminata*	Africa	Beetle	-	-	NC_003380	(Bakker 1975)
Rubus chlorotic mottle virus (RuCMV)	AM940437	*Rubus idaeus and Rubus subgenus C. quinoa, C. amaranticolor*	NE Scotland	nd	nd	nd	nd	(McGavin & MacFarlane 2009)

				Transmission				
Species & Acronym	GenBank accession number	Natural host range	Geographic distribution	Vector	Pollen or seed	Soil	Virusoid RNA	Reference
Ryegrass mottle virus (RgMoV)	AB40446	*Dactylis glomerata, Lolium multiflorum*	Japan	nd	nd	-	-	(Toriyama et al. 1983)
Sesbania mosaic virus (SeMV)	AY004291	*Sesbania grandiflora*	India	nd	nd	nd	nd	(Subramanya et al. 1993)
Solanum nodiflorum mottle virus (SNMoV)	nd	*Solanum nodiflorum, S. nitidibaccatum & S. nigrum*	Australia	Beetle	-	-	NC_003850	(Greber & Randles 1986)
Southern bean mosaic virus (SBMV)	AF05587	*Phaseolus vulgaris, Vigna unguiculata, V. mungo and Glycine max*	Africa, The Americas, France.	Beetle	+	+	-	(Hull 2004)
Southern cowpea mosaic virus (SCPMV)	M23021	*Vigna unguiculata*	nd	Beetle	+	-	-	(Lee et al. 2001)
Sowbane mosaic virus (SoMV)	GQ845002	*Chenopodium spp., Vitis sp., Prunus domestica, Atriplex suberecta,*	The Americas, Europe, Australia, Japan	Aphid, Leafhopper, Leafminer, Fleahopper	+	-	-	(Kado 1971)
Subterranean clover mottle virus (SCMoV)	AF8001	*Trifolium subterraneum*	Australia	-	+	-	M33000, M33001	(Jones et al. 2001)
Turnip rosette virus (TRoV)	AY177608	*Brassica campestris ssp. Napus and rapa*	UK	-	-	-	nd	(Hollings & Stone 1973)
Velvet tobacco mottle virus (VTMoV)		*Nicotiana velutina N. clevelandii*	Australia	Mirid	-	-	NC003906	(Randles & Francki 1986)

* both CoMV and CfMV acronyms are used in the literature, but CoMV will be used throughout this thesis as this is the official taxonomic acronym (Fauquet et al. 2005).

Genome organisation

Sobemovirus genomes consist of a single positive sense RNA genome 4.0 - 4.5 kb in size, which until recently were divided into two groups based on genome organisation (Tamm & Truve 2000 ; Fauquet et al. 2005). However, recent work re-sequencing several sobemovirus genomes (Meier & Truve 2007) proved that all species have a CoMV-like genome organisation (Figure 1.1) (Meier & Truve 2007).

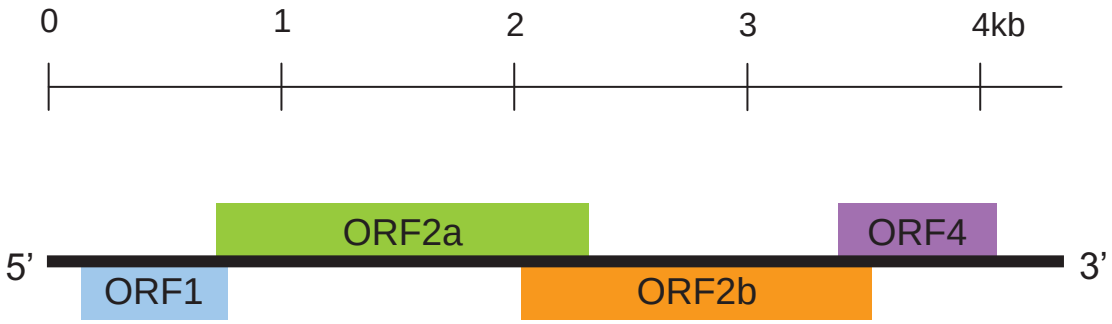


Figure 1.1. Sobemovirus common CoMV – like genome organisation model (Meier & Truve 2007) with four distinct ORFs: ORF1 in blue, ORF2a in green which overlaps with ORF2b shown in orange, and the purple ORF4 which also overlaps with ORF2b. There are two UTRs at both the 5' and 3' ends of the gene shown as horizontal lines. In some species ORF1 and ORF2a do not overlap, and instead there is a small third intergenic UTR.

Sobemoviruses have four ORFs and both 5' and 3' untranslated regions (UTR), but no 3' poly-A tail. At the 5' end of the genome a viral genome-linked protein (VPg) (Meier et al. 2008) is covalently attached in a species specific manner (Olspert et al. 2011). ORF1 encodes protein P1 thought to be involved in movement and silencing (Tamm & Truve 2000). ORF2a encodes a polyprotein containing both serine protease and VPg motifs. ORF2b encodes the RdRp gene translated by a -1 frameshift as part of a P2a2b polyprotein (Tamm & Truve 2000). ORF4 encodes the coat protein, translated from a subgenomic RNA (sgRNA), which in SBMV was also shown to have a VPg attached at the 5' end (Ghosh et al. 1981). The first aim of this thesis was to determine the genome sequence of VTMoV and compare it to the genome model (Figure 1.1) for other sobemoviruses.

1.2 Unique features of VTMoV

VTMoV has several features that make it unique; the presence of a virusoid sequence; its mode of transmission; and its ability to survive with a host adapted to an arid environment.

1.2.1 VTMoV virusoid

VTMoV was found to contain a viroid-like RNA (Randles et al. 1981), which was the first time this type of RNA had been described. These RNA species were described (Gould 1981) and sequencing determined that the RNA was 366 nt in size (Haseloff & Symons 1982). Since the discovery in VTMoV, viroid-like RNA or virusoids have now been found in four other *Sobemovirus* species, three present in Australasia: LTSV, SNMoV and SCMoV - where multiple virusoids were detected (Francki et al. 1983 ; Davies et al. 1990 ; Symons & Randles 1999) and in RYMV from Africa (Collins et al. 1998). The virusoids have all been sequenced and range in size from 220 to 390 nt (Fauquet et al. 2005). Phylogenetic analyses of these virusoids cluster them all together on a tree, distinct from viroid sequences (Elena et al. 2001). Although they were first thought to be required for infection and transmission (Gould et al. 1981), the status of the virusoid as a satellite RNA was confirmed when isolates of LTSV (Jones et al. 1983) and VTMoV (Francki et al. 1986a) were discovered without detectable virusoid (Symons & Randles 1999). This proved that the viruses could replicate without the virusoid and that virusoid RNA was not infective when inoculated by itself (Jones et al. 1983); defining the virusoid as a satellite RNA (Roossinck et al. 1992). Experimental evidence suggests that the virusoids replicate via a rolling-circle mechanism similar to that described for viroids (Symons & Randles 1999). This replication allows virusoid RNA to replicate in large quantities and VTMoV virusoid RNA can account for up to 50% of total RNA content of infected leaves (Randles et al. 1981). Although the VTMoV virusoid RNA accumulates at a high level in the leaf, only a small proportion is encapsidated by the virus (Hanada & Francki 1989).

1.2.2 VTMoV Transmission

When the virus was discovered in *N. velutina,* the plants were also infested with nymphs and adults of the mirid bug, *Cyrtopeltis nicotianae* [Hemiptera; Miridae]. These mirids were shown to transmit VTMoV to *N. velutina* and to *N. clevelandii* (Randles et al. 1981). Subsequent studies of mirid transmission determined that it had features of both non-persistent and persistent modes of transmission. Mirid transmission has an acquisition threshold of less than 1 minute and no detectable latent period, which are both indicative of a non-persistent mode of transmission. However, the virus is also translocated from the gut to the haemolymph, but not detected in salivary glands (Gibb & Randles 1990). It is transmitted transtadially (through moulting) and individual mirids can remain intermittently infective for up to 9 days. These three features are all characteristic of a persistent mode of transmission (Gibb & Randles 1990).

Gibb & Randles (1990) proposed an ingestion-egestion model for mirid transmission. In their model mirids acquire the virus by feeding via a stylet bundle that sucks up infected plant sap. The virus then accumulates in the gut (also translocates to the haemolymph) where it persists for up to six days without propagation. The virus is then slowly eliminated from the gut via defecation and virus particles in the faeces remain infective for up to three days. Infected faecal matter can either then contaminate the mirid mouth parts or alternatively, the mirid infects the plant via regurgitation during probing (Gibb & Randles 1990). These regurgitant and intermittent transmission features are also observed in beetle transmission (Gibb & Randles 1990).

1.2.3 VTMoV RNA

The sequence of the VTMoV genomic RNA has not previously been studied. The only genetic information available for VTMoV is a description of the physical properties of the RNA species encapsidated in partially purified virus (Figure 1.2).

Figure 1.2. VTMoV RNA species 1, 1a, 1b, 2 and 3 separated on a denaturing polyacrylamide gel containing 7M urea. The transmission electron microscope images on the right are rotary-shadowed molecules of RNA species 1, 2 and 3 spread after denaturation. Bar represents 500 nm. Image copied from (Randles & Francki 1986).

VTMoV RNA can be separated into five distinct bands. The slowest running species, RNA 1, is likely to be the genomic RNA (gRNA) of VTMoV, with a molecular mass of 1.5 x 10^6 (Randles et al. 1981). RNAs 1a and 1b are thought to be either degraded forms of RNA1 or sgRNAs

because they hybridise to cDNA probes synthesised against RNA1 (Gould 1981). RNAs 2 and 3 are virusoid RNA, and are present in both a circular (RNA2) and linear (RNA3) forms.

1.3 Sequence variation in VTMoV

A collection of isolates of VTMoV is maintained in the Waite Institute collection (WINC). This collection includes some of the original 1979 samples, as well as samples collected in other years and locations. Little is known about plant viruses in natural systems (Xu et al. 2008), and WILPADs are poorly studied. This collection provides the opportunity to discover the inherent sequence variation present in a WILPAD.

Similar types of studies have been undertaken with RYMV using samples collected over ~40 years. Analysis of the sequence variation of RYMV, initially with the CP (Abubakar et al. 2003) and then with 16 full genome sequences (Fargette et al. 2004), showed a strong correlation between the genetic diversity and geographic distance of samples. This pronounced spatial structure was consistent with a theory in which RYMV moved westward across the African continent, originating in East Africa (Fargette et al. 2004). The VTMoV collection provides an opportunity to investigate VTMoV for a similar spatial and temporal pattern.

1.4 Genetic bottlenecks during vector transmission

Genetic bottlenecks occur when only a small number of virus particles, known as the effective population (*Ne*), start the virus population in a new plant. This reduced population size leads to a reduction in the genetic variation of the progeny population (Ali & Roossinck 2008). These types of genetic bottlenecks are referred to as founder events and happen frequently in plant RNA viruses (Li & Roossinck 2004 ; de la Iglesia & Elena 2007). Studies with mutant strains of *Cucumber mosaic virus* (CMV) transmitted by aphids have shown that vector transmission can impose significant bottlenecks on CMV populations, especially during the inoculation phase (Ali et al. 2006). These repeated reductions in population intensify the magnitude of genetic drift (Ali & Roossinck 2008), which leads to an increase in inter-population diversity and a decrease in intra-population diversity (García-Arenal et al. 2001). A comparison of sequence from VTMoV subjected to possible bottlenecks via mirid transmission could show whether any genetic changes are associated with this mode of transmission.

In potyviruses, the DAGX motif in the CP is conserved for transmission by aphid vectors (Atreya et al. 1991 ; Flasinski & Cassidy 1998). Repeated mechanical transmission of potyviruses can lead to mutations in this DAGX motif and can result in the loss of aphid transmissibility of the virus (Ng & Perry 1999). VTMoV sequence will also be examined for sequence regions that may be conserved for mirid transmission.

1.5 Scope of this thesis

The thesis concentrates on some of the unique features of VTMoV. The first aim was to sequence the RNA genome of VTMoV. The second aim was to determine the inherent sequence variation present in VTMoV, initially between two isolates VTMoV K1 and R17, which have different symptoms, and then through describing the inherent sequence variation present in a collection of VTMoV samples across the RNA dependent RNA polymerase gene. The third aim of this work was to evaluate sequence variation in VTMoV sequences after long term mirid transmission and to search for evidence of genetic bottlenecks.

Part of the work described in this thesis has been published as,

Arthur, K., Dogra S. and J. W. Randles (2010). "Complete nucleotide sequence of *Velvet tobacco mottle virus isolate K1*." Archives of Virology 155 (11): 1893-1896. (See Appendix One).

- Chapter 2 -
General Materials & Methods

2.2 Materials

2.2.1 Virus

VTMoV isolates were sourced from the WINC (listed in Appendix 2) refer to specific chapters for isolates used.

2.2.2 Plants

N. clevelandii were raised from seed (pre-soaked in 1 % Tween 20 for 24 h). Healthy and virus inoculated plants were maintained in a greenhouse at 21-24 ºC without supplementary light.

2.2.3 Solutions, buffers, media

All solutions, buffers and media preparations are provided as concentrations of final reaction volumes. See Appendix 3 for the buffer and media recipes and a list of suppliers.

2.2.4 Primers

All primers were sourced from Geneworks (Adelaide, Australia) and 100 µM stocks stored at -20 ºC. Working stocks of 10 µM were prepared by dilution using DEPC treated water. A list of all primers is provided in Appendix 4.

2.3 Methods

2.3.1 Inoculation of plants

At approximately the 3-5 leaf stage, *N. clevelandii* seedlings were mechanically inoculated. Virus inoculum was prepared by grinding infected leaf material (either dried or fresh) in water and wiped by hand across the leaf surface covered in Carborundum powder. Systemic symptoms appeared at approximately 7 days post inoculation (d.p.i) and systemically infected leaves were used for experiments.

2.3.2 Harvest and storage of leaf

Leaf tissue was stored long term as dried leaf material. Symptomatic leaves were harvested and cut into strips. Leaf material was desiccated under vacuum with activated silica gel for three days. Dried leaf was stored in capped 20 ml glass bottles, over anhydrous calcium chloride and activated silica gel at 4 ºC until analysis.

2.3.3 Partial purification of VTMoV

As described by Arthur et al. (2010); ca. 80 g of symptomatic systemically infected leaves were blended in a mixture of 100 mM sodium potassium phosphate buffer (pH 7.4) containing 0.1 % monothiolglycerol (2:1 w/v), chloroform (2:1 w/v) and *n*-butanol (2:1 w/v), strained through muslin, and clarified by centrifugation for 10 min at 12000 *g*. The supernatant was mixed with 8 % (w/v) PEG 6000 and 2 % (w/v) sodium chloride at 4 °C for 1 hour, before centrifugation for 10 min at 12000 *g*. The resulting pellet was resuspended in water for 1 hour then centrifuged for 10 min at 12000 *g*. The supernatant was centrifuged in a TLA100.3 rotor (Beckman, USA) for 30 min at 220000 *g* and the pellet was resuspended in 200 µl of 10 mM Tris-HCl (pH 7.5). Virions were fractionated by isopycnic density gradient centrifugation on a preformed 30 - 60 % Nycodenz® gradient for 19 hours at 45000 rpm in a TLA100.3 rotor (Beckman, USA). The light scattering band at a density of 1.27 g l^{-1} was recovered and diluted before centrifugation for 45 min at 220000 *g*. The pellet was resuspended and dialysed against 10 mM Tris-HCl (pH 7.5). All centrifugation was performed at 5 ºC unless otherwise stated. The UV absorption spectra of virion preparations were used to determine yield. Virion concentration in mg ml^{-1} was calculated by dividing the A260 nm value (corrected for 310nm light scattering) by an absorption coefficient of $A_{260}^{0.1\%} = 5$ (Randles & Francki 1986), before adding 20 % glycerol and storing at -20 ºC. Virus yield was calculated and virion preparations were tested using dot immunobinding assays (DIBA) to confirm the presence of VTMoV.

2.3.4 Serological testing - DIBA

The DIBA method was modified from Hibi & Saito (1985). Samples from stored passages were prepared in 1X PBS (2:1 w/v) and 1 µl aliquots were spotted onto nitrocellulose membrane with positive VTMoV, healthy leaf sap and buffer controls. The membrane was initially submerged in blocking buffer B at room temperature for 30 min before incubating with a 1: 2000 (v/v) dilution of VTMoV rabbit polyclonal antibody in blocking buffer B for 30 min at room temperature. The

membrane was washed three times with blocking buffer A before incubating in a 1:2000 (v/v) dilution of goat anti-rabbit antibody (Sigma Aldrich) in 1X PBS with 1% BSA. The membrane was incubated again for 30 min at room temperature, before first washing twice for 3 min in wash buffer AP 7.5, followed by washing twice for 3 min again in wash buffer AP 9.5. Membranes were submerged in substrate solution containing bromo-4-chloro-3-indolyl phosphate (0.165 µg µl^{-1}) and nitro blue tetrazolium [0.33 µg µl^{-1}] and allowed to develop. The reaction was stopped by submerging membranes in stop buffer for 2 min at room temperature, before membranes were dried and scored.

2.3.5 Total nucleic acid extraction

Total nucleic acids were extracted using a modified protocol of Chomczynski & Sacchi (1987). Infected leaves were ground in liquid nitrogen and nucleic acids were extracted by adding 10 vol (w/v) of guanidine thiocyanate extraction buffer and shaking for 20 min at room temperature. One fifth volume of chloroform - isoamyl alcohol (49:1) was added to the extract, and shaking was continued for a further 20 min at room temperature. Extracts were then centrifuged to separate the aqueous supernatant. Nucleic acids were precipitated with 3 volumes of ethanol in the presence of 0.3 M sodium acetate, or with 0.9 volume of isopropanol. Nucleic acids were pelleted by centrifugation and then washed in ethanol before dissolving in 0.01 M sodium acetate.

2.3.6 Virus RNA extraction

Virion preparations were pre-treated with 10 mM EDTA (Hull 1977), and were then digested by either one of two alternative protein digestion protocols. Method one used the proteinase K method of Veerisetty and Sehgal (1980). The second method was based on Murant *et al.* (1972) where virion preparations were mixed 1:1 with a protease solution (0.25 mg ml$^{-1)}$ and incubated overnight at 37 ºC.

Virus RNA was extracted using either guanidine thiocyanate extraction buffer as in section 2.3.5, or phenol - chloroform. For the phenol - chloroform extraction, 100 µl digested virion preparations were mixed vigorously with 50 µl of phenol (90 % aqueous solution containing 0.1 % 8-hydroxy quinoline) for 1 to 2 min and then centrifuged at 10000 *g* for 10 min. The supernatant was mixed 2:1 (v/v) with phenol - chloroform (1:1) and centrifuged at 10000 *g* for

10 min. RNA was precipitated with ethanol, pelleted and washed as described in section 2.3.5, and resuspended in 0.01 M sodium acetate.

Nucleic acid concentration

The UV spectra of nucleic acid samples were measured using a Nanodrop (ND-1000) spectrophotometer. The concentrations of nucleic acids were calculated using $\frac{A260}{A^{0.1\%}_{260}}$. $A^{0.1\%}_{260}$ = 25 for RNA, and $A^{0.1\%}_{260} = 20$ for DNA. Nucleic acid solutions were then stored at either - 80 ºC (RNA) or - 20 ºC (DNA).

2.3.7 Polyacrylamide Gel Electrophoresis

Three and a half percent polyacrylamide gels (19:1 acrylamide: *bis*-acrylamide) were prepared and submerged in 0.5X TBE running buffer. For denaturing polyacrylamide gels urea was added to 7 M, and gels were run for 30 min at 60 V before loading. Nucleic acid samples were mixed 2:1 with 6X loading dye (containing bromophenol blue and xylene cyanol). Polyacrylamide gels were stained with either 0.2 µg ml^{-1} ethidium bromide (in 0.5X TBE buffer) and visualised using UV fluorescence with the GeneFlash system (Syngene, USA) or with silver (Sambrook & Russell 2001). Stained gels were stored in 10 % ethanol containing 0.7 % glycerol.

2.3.8 cDNA synthesis

To synthesise cDNA a mixture containing template RNA, water and primer - either 0.5 µM specific reverse primer or 5 ng $µl^{-1}$ random hexamer - was prepared and heated at 65 ºC for 5 min before chilling on ice for 1 min. A second reaction mixture was added containing; 0.5 mM of each dNTP, 0.4 - 0.8 U $µl^{-1}$ Transcriptor reverse transcriptase, 0.8 U $µl^{-1}$ RNase OUT and 1X reaction buffer (50 mM Tris/HCl, 30 mM KCl, 8 mM $MgCl_2$, pH 8). The following conditions were then used (i) 25 ºC for 10 min, (ii) 45-50 ºC for 30-45 min and (iii) 85 ºC for 5 min in an Eppendorf mastercycler ep gradient S (Eppendorf GmbH. When specific primers were used step (i) was omitted. Following cDNA synthesis, 0.1U $µl^{-1}$ of RNase H was added and the mixture was incubated at 37 ºC for 20 min. cDNA was stored at -20 °C when not in use.

2.3.9 PCR Amplification

PCR reaction mixtures contained: 1-2 µl of template cDNA, 1X reaction buffer B with 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM each of a forward and reverse primer, and 0.05 U μl^{-1} of Taq polymerase. PCR cycling conditions were: initial denaturation at 94 °C for 2 min, (ii) denaturation for 15 - 30 s at 94 °C, annealing for 30 s at 45 - 60 °C, extension for 30 s at 60 - 72 °C (25-30 cycles) and (iii) final extension step of 5- 10 min at 60 - 72 °C. PCR products were amplified in either an Eppendorf mastercycler ep gradient S (Eppendorf, GmbH) or GeneAmp PCR system 2400 (Perkin-Elmer Ltd.).

2.3.10 Agarose gel electrophoresis

PCR products were separated by agarose gel electrophoresis. Gels of 1.5 % (w/v) were prepared using 0.5X TBE containing 0.5 µg ml^{-1} ethidium bromide. Products were loaded with 1X loading dye. PCR products were visualised with UV fluorescence in the GeneFlash system (Syngene, USA) and compared to either 1Kb or $1Kb^{+}$ DNA ladder size standards (Invitrogen).

2.3.11 Purification of PCR products

When PCR products were present as single bands the QIAquick® PCR Purification Kit (QIAGEN GmbH) was used according to the manufacturer's instructions. If multiple bands were present PCR products were first separated using agarose gel electrophoresis and specific bands were excised from the gel. DNA was extracted using the QIAquick® Gel Extraction Kit (QIAGEN GmbH) according to the manufacturer's instructions. DNA concentrations were then determined as in section 2.3.6.

2.3.12 Cloning

DNA was cloned into the pCR®4-TOPO® vector using the TOPO® TA Cloning Kit, with either 80 µg ml^{-1} Ampicillin or 50 µg ml^{-1} Kanamycin antibiotic selection in growth media. Ligation reactions solutions contained 4 µl of template DNA, 10 ng of vector, 0.2 M NaCl and 0.01 M $MgCl_2$. Ligation reactions were then incubated for 30 min at room temperature.

Transformation

Ligation reaction products (2-3 µl) were added to 50 µl of thawed Mach™1-T1® chemically competent cells and mixtures were gently tapped 2-3X before leaving on ice for 10 min. Cells

were heat shocked at 42 ºC in a water bath for 30 s. Cells were immediately transferred to ice for 2 min before adding 250 µl of SOC medium and incubating at 37 ºC with shaking for 1 h. Cells were plated onto LB agar plates with antibiotic selection and incubated at 37 ºC overnight to identify transformed colonies

Screening transformants

Transformants were screened by PCR. Clone screen PCR reactions contained; 1X reaction buffer B with 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 µM each of the forward primer MI3F and the reverse primer M13R, and 0.05 U $µl^{-1}$ of Taq polymerase. Transformant colonies were picked using a sterile pipette tip and directly added to PCR tubes. PCR cycling conditions were: initial denaturation at 94 °C for 10 min, (ii) denaturation for 15 s at 94 °C, annealing for 30 s at 50 °C, extension for 30 s at 72 °C (25 cycles) and (iii) final extension for 10 min at 72 °C.

Liquid culture of clones

The pipette tips used in PCR clone screening were also stab inoculated onto LB agar plates in a grid pattern and incubated overnight at 37 ºC. Colonies were selected based on screening results and incubated overnight at 37 ºC with constant shaking in 4 ml Luria broth with either 80 µg ml^{-1} Ampicillin or 50 µg ml^{-1} Kanamycin antibiotic selection.

Plasmid DNA extraction

Transformant plasmid DNA was extracted from liquid culture using either the QIAprep™ Spin Miniprep kit (QIAGEN, GmbH) or the HiYield™ plasmid mini kit (Real Biotech Corporation, Taiwan) according to the manufacturer's instructions. Final plasmid DNA yield was determined as in section 2.3.6.

Restriction enzyme digest

Plasmid DNA was digested with *Eco*RI, which cuts the pCR®4-TOPO® vector on both sides of the fragment insertion point. Less than 1 µg of plasmid DNA was incubated with 1U $µl^{-1}$ of *EcoRI* and 2X Tango™ Buffer at 37 ºC for 1 h, before heating at 65 ºC for 15 min. Digestion products were visualised by agarose gel electrophoresis (section 2.3.10).

2.3.13 Sequence analysis

DNA products were sequenced at the Australian Genome Research Facility Ltd., with either universal or gene specific primers. Sequences are discussed and listed as RNA with uracil (U) in descriptions throughout this thesis but are presented as GenBank sequence (DNA) with thymine (T) replacing U in figures.

Alignment and editing of sequence

Each sequencing result was edited to remove vector sequence and regions of poor sequence quality. Overlapping sequences were then combined using BioEdit Sequence Alignment Editor v.7.0.5.3© software (Hall 1999) to resolve sequence discrepancies between electropherograms, and create a consensus sequence from multiple sequencing results. Edited sequence was analysed using a BLAST search with either the BLASTn (nucleotide to nucleotide search) or the tBLASTx (translated nucleotide to translated nucleotide) options (Altschul et al. 1997) in GenBank (Benson et al. 2008).

Phylogenetic analyses

Multiple sequences were aligned using ClustalX v2. software (Larkin et al. 2007) with default parameters. Aligned sequence datasets were analysed using a range of protocols, software and evolutionary models as described in specific chapters of this thesis.

- Chapter 3 -
Sequencing the VTMoV genome

3.1 Introduction

This chapter describes the sequencing of the RNA genome of VTMoV, which has a molecular mass of ca. 1.5 x 10^6 (Randles et al. 1981), equivalent to ca. 4.4 kb. It was hypothesised that because of its classification as a sobemovirus, VTMoV would have the same genome organisation as previously sequenced sobemoviruses. Degenerate primers were therefore targeted to conserved regions in these genomes. After collecting initial sequence data, a genome walking approach combining specific and degenerate primers was used to sequence the remainder of the genome. VTMoV isolate K1 (isolate 65781) was selected for sequencing as it was apparently free of the virusoid RNA (Francki et al. 1986a).

3.2 Materials and Methods

3.2.1 Sequence fragment collection

Polyadenylation of viral RNA at 3' terminus

E. coli poly (A) polymerase (New England Biotech) was incubated with 1 mM ATP, 1X reaction buffer and viral RNA before incubating at 37 ºC for 20 min, according to the manufacturer's instructions.

Rapid amplification of cDNA ends (RACE)

The 5' and 3' ends of the genome were amplified using the SMART™ RACE cDNA amplification kit (Clontech Laboratories Inc., CA, USA), according to the manufacturer's instructions.

RT-PCR

Specific PCR primer combinations used to amplify VTMoV genomic fragments are listed in Table 3.1 and pictured relative to each other showing overlaps in Figure 3.1. For a complete list of primers (primer sequence, source, and melting temperatures) refer to Appendix 4.

Primer PA14 was a degenerate primer developed by Dr. Allan Gould. The PCR conditions for this primer were as follows; initial denaturation at 94 °C for 2 min, denaturation for 1 min at 94 °C, annealing for 1 min at 37 °C, extension for 1 min at 60 °C (30 cycles) followed by a final extension for 10 min at 60 °C. The sequence of this primer is unknown because of a confidentiality agreement with Dr Gould.

Virusoid PCR

Fragment A in Table 3.1 was amplified using 0.5 μM each of primers AGSATF and AGSATR designed by Dr. Allan Gould. PCR conditions were; initial denaturation at 94 °C for 2 min, followed by 10 cycles of denaturation for 15 s at 94 °C, annealing for 15 s at 50 °C, extension for 30 s at 72 °C, with a final extension for 5 min at 72 °C.

Single primer PCR

Fragments 2 and 4 (Table 3.1) were amplified using single primer PCR (SP-PCR) in which reaction mixtures contained 0.5 mM of a single primer. PCR conditions were; initial denaturation at 94 °C for 2 min, followed by 10 cycles of denaturation for 15 s at 94 °C, annealing for 15 s at 42 °C, extension for 3 min at 72 °C, followed by 25 cycles of 15 s at 94 °C, annealing for 15 s at 52 °C, and extension for 3 min at 72 °C, with a final extension for 10 min at 72 °C.

Sequencing fragments

DNA from PCR fragments was isolated as described in section 2.3.11, before cloning as in section 2.3.12 and sequenced as in section 2.3.13. Consensus sequences were determined from multiple clones of each PCR fragment.

3.2.2 Sequence analysis

Consensus sequences for each fragment were aligned using BioEdit (Hall 1999) and a final consensus sequence was determined. Where sequence mismatches occurred, sequences near the middle of fragments were given precedence over those near the ends of fragments to account for reduction in sequence read quality and primer sequence.

ORF analysis

The complete genome sequence of VTMoV was analysed using Frameplot 2.3.2 software (Ishikawa & Hotta 1999) to identify putative ORFs. The default settings displayed all ORFs with a minimum size of 20 aa. When putative ORFs had alternative start codon positions, only the largest ORF was selected. The mass of putative ORF products was calculated using the pI/Mw tool (Gasteiger et al. 2005) on the ExPASy Proteomics Server (Gasteiger et al. 2003). Putative ORF sequences were used to query GenBank (Benson et al. 2008) to identify similarities with described sobemoviruses. Nucleotide sequence was used to perform BLASTn (Zhang et al. 2000), tBLASTn (Altschul et al. 1997), and tBLASTx (Altschul et al. 1997) searches while protein sequence was used to perform BLASTx (Altschul et al. 1997) and BLASTp (Altschul et al. 1997 ; Altschul et al. 2005) searches.

RNA structural analysis

RNA sequences were analysed using Mfold (Zuker 2003) at 37 ºC. Default ionic conditions were 1 M NaCl with no divalent ions.

3.3 Results

3.3.1 VTMoV genome

Sequenced fragments from VTMoV isolate K1 (Table 3.1) were assembled (Figure 3.1) to produce a single contiguous sequence.

Table 3.1. List of RT-PCR fragments used to assemble the VTMoV genome sequence, showing the primers used in cDNA synthesis and PCR, the size of each product and their positions on the genome. Nucleotide sequences are in Appendix 5.

Fragment	Primers used		PCR product size (bp)	Map position (nt)
	cDNA synthesis	PCR		
1	Random hexamers	SOBDF2 SOBDR2	**303**	2645-2947
2	CPF1	CPF1	**884**	2389-3273
3	VTMoV-Rf	SOBDF4 VTMoV-Rf	**1223**	1609-2790
4	VTMoV-Ra	VTMoV-Ra	**971**	2807-3779
5	VTMoV-CPR2	VTMoV2abF1 VTMoVCPR2	**1436**	2305-3595
6	VTMoV-2aR3	ORF2aDF3 VTMoV2aR2	**653**	1062-1721
7	VTMoV-2aR3	PA14 VTMoV2aR2	**705**	1003-1749
8	VTMoV-2aR3	PA14 VTMoV2aR2	**1182**	439-1662
9	Poly A primer	VTMoV-CPF3 Poly A primer	**507**	3527-4060
10	3' RACE	VTMoV-CPF4 NUP	**261**	3986-4248
11	5' RACE	NUP ORF2aR4	**429**	159-596
12	5' RACE	NUP ORF1R1	**383**	1-382
A	AGSATR	AGSATF AGSATR	**366**	Virusoid sequence

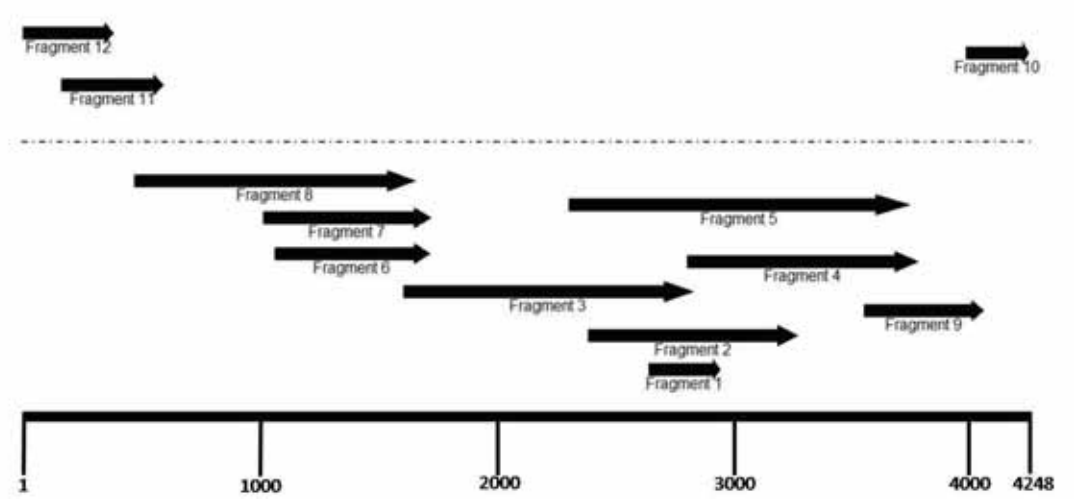


Figure 3.1. Map showing positions of the 12 fragments (Table 3.1) used to assemble the nucleotide sequence of the VTMoV genome. Fragments above the dashed line were amplified using RACE and those below were amplified using a mixture of degenerate and sequence specific primers. Scale bar shows nucleotide positions.

The complete genome sequence of VTMoV isolate K1 was 4248 nt in length with a base composition of 25.4 % adenine (1081 nt), 25.5 % guanine (1085 nt), 26.7 % uracil (1137 nt) and 22.2 % cytosine (945 nt), and an overall G+C component of 47.7 %. This sequence has been submitted to GenBank and assigned accession number HM754263 (Figure 3.2).

>ACAAAATATATTGAAGGGAAAGTCTTGCACTCAGTACCCATTATATTTTGATATGCCCAGCATTG
ATGTTGAAGTAGAAAAGATCCTGCACTTGAGCAACAGAAAGAGACTAAGGTCTGTCTGTGGCGAGG
AAGAAGACGTTCGTAGACAAGATATCCGAAGGATTTGTCTGCACTCTTTACGGAGAGTACGATCACGA
CTACGTCACCAGTGTGCATTTGCACATTGTTTGCTCGTGCGGTAGAGCTTTCTTTGATTTTGTTGAGT
TCAAGGACATAAAACTCCGAGATTTCCAATATCAAAGTCAGTGTATCTGTCCCACCAGGAATTGGTCG
ACGGTCGCTTACACTGATTTTACTACCGTCAACTGCCAAGTGGAAGGTTGTGATTACTGTGAGGGGAT
TGAGTCAGACTCTGATTCGGATTCTGAGGCCATCATAGAGGAATTTCTTCAAAAGTTTTCCGAGGTTG
GCATCTCTGGATCGTCCAGTTCTCCCCAGACTAACTAGGACTCGGTTCCGTTTATTTTCTTACAAATC
ATTTAATTGTTTTGTAATTGAGAAAGGATGTTGAGCGAGTTAGTCCAGTTGTGTCTTTCAGCAACCAT
GACTCTAGCGCTGGTAGTGTCAATGATGCTGGACCCGAGCTACGGGTGGCTGATCACCATGTCTGTGA
TGTTAGTCTCCTTAGAGCTGACAATCAGGCCGTTCAGGCGCTCTCTAGACTACATGAAAATTGTGGTC
CGAGATGAGCCCAATGAGCCTGAAGTTGCTAGGATAGTAGGCGCGCCAACTTACCACCCAATTCATGG
CGTTGTTTGCAAGTGTCTGTGGAACTCGAAAGAGTTCAATGTGGTTGTGCAACCCAGTTTGTGGCCTC
TAGCCTTTCCCCCAAACATATCACAAGGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACA
TCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCTTATACAACGATGGAGTCAGAATCGGTAT
GGGCTCTAGAGTTAATTGGCAAGGCGCAGATTACTTGTTGACAGCCTCCCACGTTTGGTCTCTCGTAA
CTGGTGATTTTCAAATGGCAAAAGCCAGTAAGATGGTCTCAGTGAAGGACGCCAAGTGTTATGTTGAG
GCTGCTCACGCTAAGCTGGACTTTGCCCTAATTAAAGTTCCGAACAAATATTGGAGCTCGATAGGGGT
TGGTTCGGCTAAATTGTTGTGGCATAAGCCCGGGCAAGTGGTCAAGGTTTACGGAGGTAGATCTGATG
AATTGGTCTCTTCTGTGGGCCGAGCCGAAAAGGATCCTGATCTGAGCCTACGTCTGACCCATAACGCT
TCAACTGCTCCAGGATGGAGCGGCTCCCCGCTTTACAACTCTGAAAATTTCGTAGTGGGATTGCATAC
TGGCTTTTCCGCGGCTGAGCAGAGAAATGAGGCGGTTGATGTAGCGAAGTTGCTGCACCTCGCCCTTA
GAACAAAGGAAACGACGTTTTCGGAGATCGGTGTCTCTCTTATCGATGAAGATGAGATTGAGTCTCGC
GGTTACCAATTCGACGACTTTGAGCTTAGGGGTGAGGTAAACGTTAAAGGGAAAATGGCGCGGAATGA
GATATCTCTCATCGCCTCTAAGAATAAAGGGAAACCTTGCTATCTTCAAGAGGAAGGTGATGATGAAT
TCTATGACTCGATCCGTGAGAAAGATTTCCTTGCCCGCTTCCGTGAACAAACTGGTAAAGAGACAGTG
GGGAATTTAAACTGCCAGAGGGCGGCCCAAACATTGGAGCCGCCCTTCGAGAACTTGAGGCCGTGCGA
TGGGAAGAACCCAGAGCTATTCAAACCGGCGGGATGGGATTCAACTATGTTGGAGTCTCGTCTTGCAA
GTTTAGAGAGGGCCCTAAGCACGCTGCTCGCGGAGCAATCAGTTCTGCTGAGCAAGTTTTCCCAGAAC
TCAAACAGTATGATTGGCCAGAAAGAGGCTCTAAAGCCGAGTTCGATTCCCTCCTCCTCCAAGCCAGC
CGTTTCCGGAGAACCTCCTGTCCAGAGCAAACCGAGCGTAAATGCGAAGTCCTCGCTGAGAAGTACCC
CAAAAGTAGAGCCCACCGTTGCTTCAGGAGAGAAAACTTCTGTCAACGCCAGCTCCTCAGGGAACAAA
TCGAAGCGACGACGACGTCGCCGGAAATCAACGACAAAGCCAGCCCCGGCTCCCCCTGGTCAAGACTT
GAAAAGACCAATGGTGAGCTCATCTCCCGTTTCAAAGACCTCTTAATTGAGGCAGTGTTGAGGCGAGT
TTTATTGCTTGCCTCCACTCCAACCAGTGAGATTTTATGCATGTCTGCTTCAGATCTCGTTAGAGCTA
ACTTAGTGGATCCTGTGCGATTGTTCGTCAAGCAGGAACCCCATACGAAGAAGAAGCTGAATGAGCGT
CGCTTTAGGCTTATTTCGTCTGTGTCTATAGTTGATCAGATAATCGAGAGATTATTGTTTGGTCCTCA
GAACAGGCTTGAGATAGCCTTATGGCATCAGATACCTTCCAAACCTGGAATGGGACTGAGCGCTCGTA
CGCAAGCTGATTTGTTGTGGAATGAGTTATTCGCAAAGAGCGAGATTGCCCCTGCGGCCGAGGCTGAC
ATTTCAGGATTTGACTGGTCTGTGCAGGAATGGGAGTTATGGGCCGATCTGAGCATGAGGATTTCTTT
GTGCGAAGATATGCATGATGGTCTACGGAGGTTGATGGTTAACAGATACCGCTGCTTTATGCTCTCTT
GCTTCCAATTGTCGAATGGGGAATTGTATGAACAAGTTGAGCCTGGACTCATGAAGTCTGGTTCCTAC
TGCACTTCCTCTTCCAACTCCAGAATCAGGTGCCTAATGGGTTATCTAATTGGAGCCCCCTGGATAAT
AGCCATGGGGGATGATTCTGTTGAGGGTTATGTGAGAGACGCGAAAGGCAAGTATGAGGAATTAGGAC
ACACTTGTAAGGAATACGAGTTGTGTGATGTTGATTCAGACGGCGCGTTGAGATCTGTGAACTTTTGT
TCACATTTGATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGA
CTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTTGGCTCATGTCCCAAGTGGGCCAAGATAA
AGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGAC
TATGTCGAAGAAACTCACCAAGAATCAGGTCAAGCAGATGATACAGGCAACCTTGCCGAAAGAGCAGA
CTTCTGCTCGGTCCCGACGACGGAGACGCCGAAGGTCGACGCAGCAAGGGCAGAGTTCTACAGTTATG
GCCCCAATGGCTGGAGCTGTGATATACCGGAAGCGACCCATGTTAATCAATGGCCGCTCTGGGGTTAC
AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCTGGGACTACTAACTTCTCGGCCACTTCTGTTA
CTCTCGCGCCTAACACTTTCACCTGGTTAGCTGTACAAGGATCCCTCTATTCTAAATGGAGGTGGATA
TCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTTTCCA
GTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAG
GCGCTCCGTGGTCAGGCTTTAGTGGTTCTAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCT
GGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAGTCTGTTAT
ACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTT
CAACTTTGAGATATGGTGAAGTGCACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCG
GTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGAAAGGAAAGGTGA
GAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCGATATAACGCCACCATTGACGTCTGGAGAGA
CTTAAATGGGTTTTGCATAACTACTCCGGTTGCTACGTCATAGCACGTTACCAAGAAGTTCTCTTGGT
CCCTCTACATGGTTGGGTGAACACCATGTGTTCAT

Figure 3.2. Complete nucleotide sequence of the genome of VTMoV isolate K1. The four start codons of the putative ORFs (Table 3.2) are boxed.

3.3.2 Virusoid sequence

It was found that VTMoV- K1 does contain a virusoid. Fragment A (Table 3.1) was a 366 bp PCR product faintly amplified (data not shown) by primers AGSATF and AGSATR which are adjacent to each other but do not overlap. These primers target the VTMoV virusoid sequenced by Haseloff and Symons (1982). Sequencing the VTMoV-K1 product (Fragment A, Table 3.1) confirmed the identity as a VTMoV virusoid, with 99 % identity BLASTn match to the previously published sequence (NC003906). The K1 virusoid has a base composition of 20.7 % adenine (76 nt), 23.1 % uracil (84 nt) 30.6 % guanine (112 nt), and 25.6 % cytosine (94 nt), with an overall G+C composition of 56.2% (Figure 3.3). Virusoid sequence from isolate K1 was submitted to GenBank and assigned accession number HQ680397.

GTGGATGTGTATCCACTCTGATGAGTCCGAAAGGACGAAACGGATGTACCGCTTCTTGTCTCGACCTC
GACCTGGACTAGTGATCGAGGGAGGCTCAACCTCACGCCCGCTGGGTAGATGTAGTCTCATACTCCAA
TGACTTGGGGTCACTGTGTAAAGGTACTACAGAGCTACGACCATGTGATAGGCGGGGAGCTGGACCCT
CTCACCACCTAGGTAGTGTTGAAGGTCGCGAGGGAGTCAAGGACGCCCGGCATCAGAGGATTGCACGC
CACCGGTATCACGGAGGGCAACTGCTTTCCAGGCTGGCAGGTAACGTTCCTGCCCTTGGGGACTGATT
TTTGGTTCGCCTGGTCCGTGTCCGTA

Figure 3.3. Complete nucleotide sequence of virusoid RNA from VTMoV isolate K1. Points of difference from previously described sequence (NC003906) are boxed. The putative virusoid self cleavage site is shown in red letters (Keese & Symons 1987). Primer sequences are underlined.

When compared to the previously sequenced virusoid, the K1 sequence has 2 substitutions 234T>G and 271A>G. These substitutions have no effect on the putative secondary structure (Figure 3.4), as the new sequence gave a secondary structure identical to that of virusoid sequence NC003906 with the Mfold program (data not shown). The primer design allowed both linear and circular forms of the virusoid to serve as template for RT-PCR amplification.

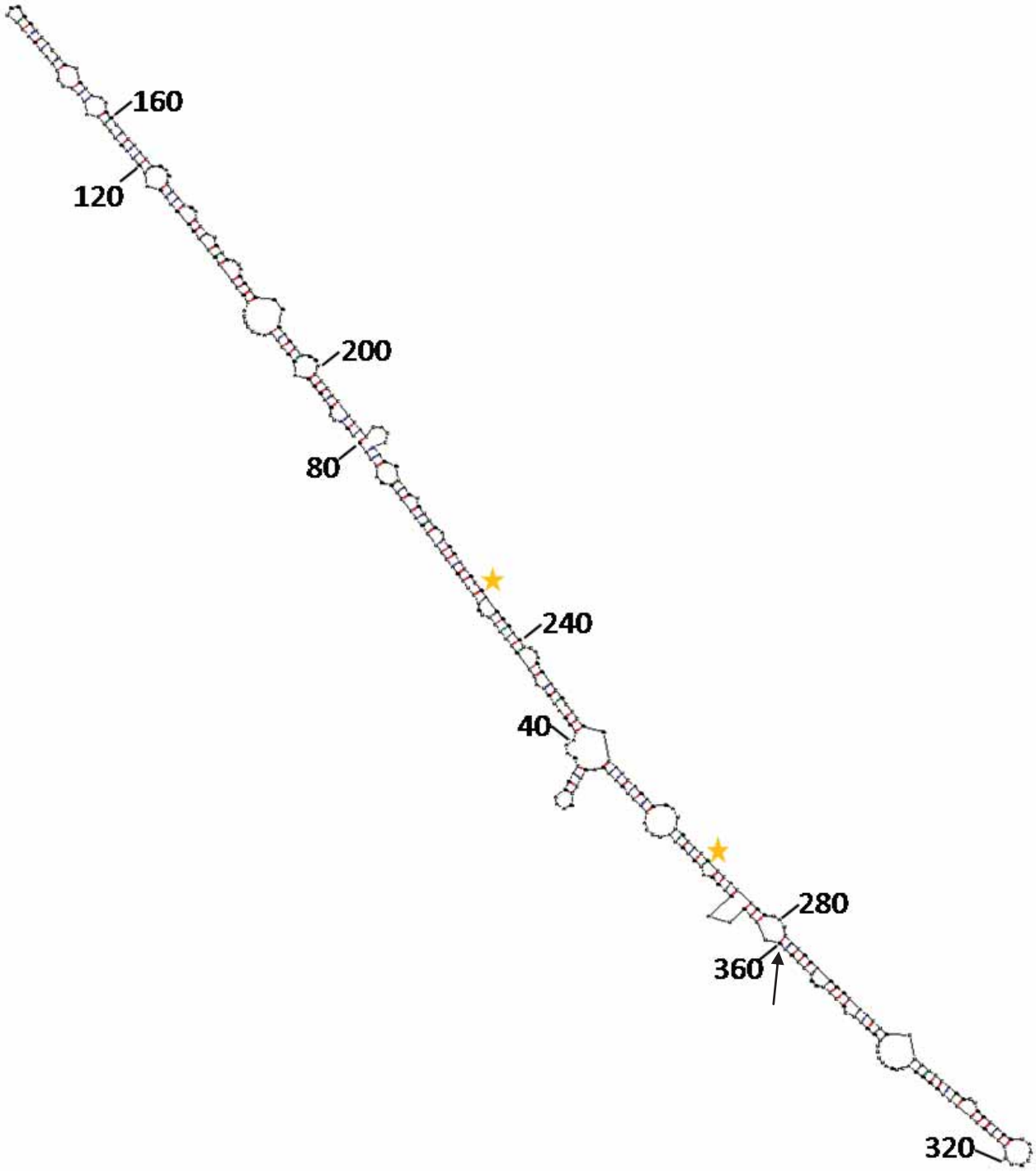


Figure 3.4. Predicted secondary structure of VTMoV-K1 virusoid RNA, with points of difference to VTMoV virusoid sequence NC003906 (gold stars) and nucleotide positions. The putative self cleavage site from nucleotides 358 – 365 is marked with an arrow.

3.3.3 ORF analysis

The VTMoV genome was assessed using Frameplot 2.3.2 for the presence and location of putative ORFs. The output of this analysis is shown in Figure 3.5.

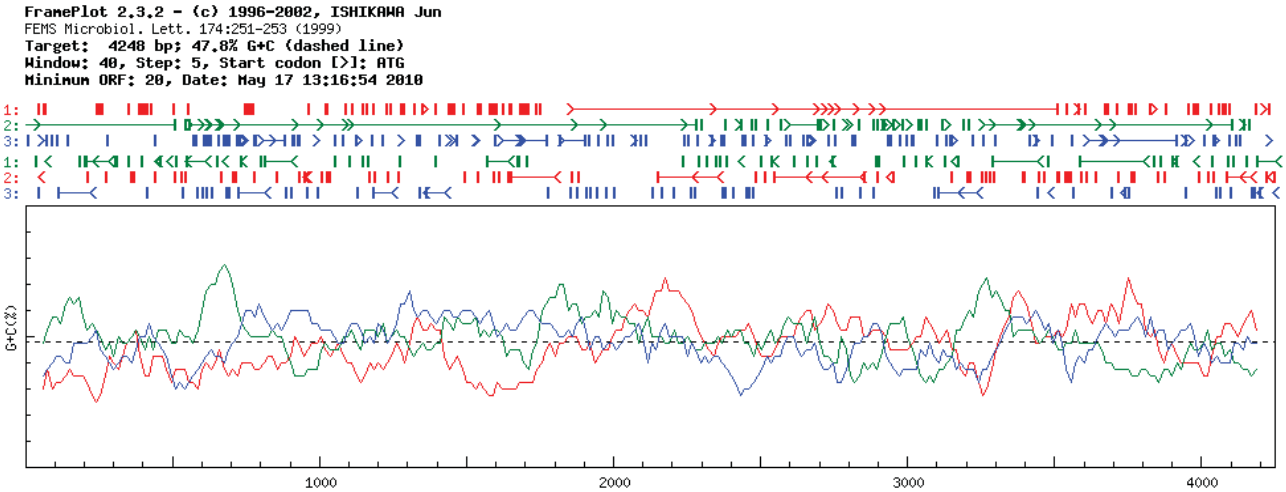


Figure 3.5. Frameplot 2.3.2 output of the VTMoV-K1 genome sequence analysis. Predicted ORFs are shown as horizontal lines with start codons (ATG) indicated by arrows, and stop codons (TAA, TAG, TGA) by vertical lines. Colours indicate the three frames in each of the sense and anti-sense strands. The G+C plot refers to the G+C content across the sequence and numbers refer to nucleotide positions on the input sequence.

Frameplot analysis (Figure 3.5) identified 25 potential ORFs in the VTMoV sequences that are listed in Table 3.2.

Table 3.2. List of putative ORFs from Frameplot 2.3.2 analysis describing genome location, orientation, the size and mass of predicted proteins. Those ORFs which are similar to sobemovirus ORFs in both position and size are in red text.

Strand	Frame	ORF	Genome position (nt)	Peptide size (aa)	Molecular mass (kDa)
Sense	**1**	**A**	1867-3513	549	62.5
	2	**B**	53-511	152	17.3
		C	569-2287	572	62.6
		D	2603-2699	32	3.67
		E	3263-4108	282	30.7
	3	**F**	810-882	24	2.57
		G	1626-1773	49	5.70
		H	1833-1905	24	2.81
		I	3618-3915	99	10.88
Antisense	**1**	**J**	4248-4185	21	2.36
		K	3798-3582	72	7.56
		L	3438-3288	50	5.61
		M	1641-1566	25	2.89
		N	900-810	30	3.37
		O	609-546	21	2.48
		P	279-207	25	2.78
	2	**Q**	4163-4085	26	2.94
		R	2837-2543	99	10.71
		S	2351-2150	67	7.23
		T	1799-1649	50	3.75
	3	**U**	3235-3103	44	5.17
		V	1423-1360	21	2.43
		W	1246-1183	21	2.58
		X	811-724	29	3.52
		Y	220-112	36	4.10

Of the four largest ORFs (bold text, Table 3.2) three showed sequence homology to sobemovirus proteins; ORFA to sobemovirus 2b proteins; ORFC to sobemovirus 2a proteins and ORFE to sobemovirus coat protein, suggesting that they these ORFs may be translated as VTMoV proteins. The fourth (ORFB) showed no sequence homology to any database sequence. However, when it was mapped (Figure 3.6) it had the same genome position as the sobemovirus P1 proteins (Meier & Truve 2007). In addition, putative ORFs C, A and E all have favourable context for translation of A[A]CA*ATG*[G]C; where the A from ATG start codon is at position +1. ORFC and ORFE have an A at positions -3, and ORFA has a G at +4 (Futterer & Hohn 1996). ORFB does not have the start codon in a

favourable translation context. Mapping of these four ORFs (Figure 3.6) shows similar genome positioning and organisation as ORFs in other sobemovirus species seen in Figure 1.1.

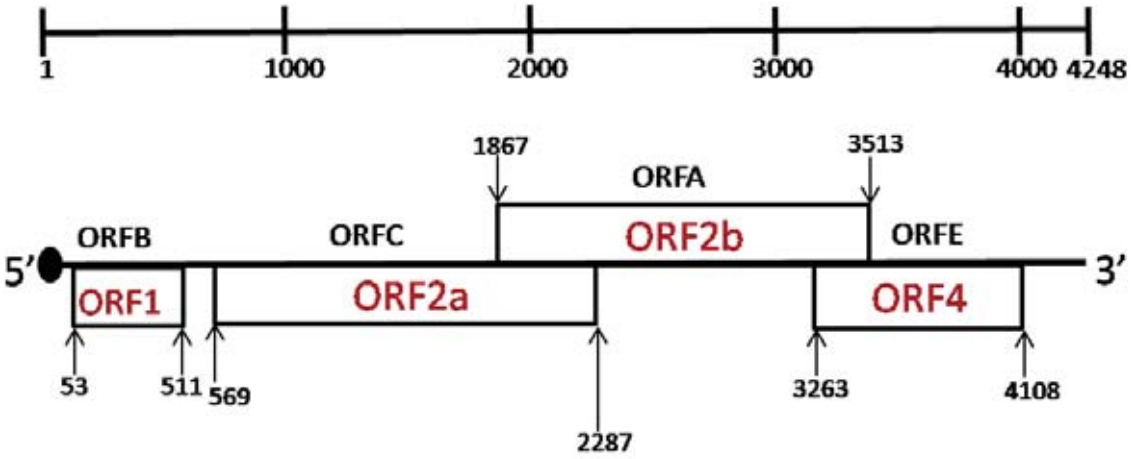


Figure 3.6. Organisation of the putative ORFs from the VTMoV genome sequence. Each putative ORF from the Frameplot analysis (Table 3.2) are shown in black and the start and stop codons are indicated by numbers and arrows. Sobemovirus ORFs of Figure 1.1 are shown in red. The location of attachment of the putative VPg (black circle) is also indicated. Scale bar indicates number of nucleotides.

Figure 3.6 also shows regions of sequence which appear to be untranslated (UTR). These are at the 5' end (nucleotides 1-52), between ORFs 1 and 2 (nucleotides 512-568), and at the 3' end (nucleotides 4109- 4248). The lengths of these UTRs and putative ORFs are compared to those of the corresponding regions in other sequenced sobemovirus species in Table 3.3.

Table 3.3. Comparison of the ORFs and UTRs from VTMoV with described sobemoviruses.* indicates genome sequences with corrections. Multiple sequences from both RYMV and SBMV are available through GenBank; therefore two sequences were selected for use throughout this work, those that have been used in published sobemovirus analysis. RYMV-Ni is an isolate from Nigeria, and SBMV-Bar is the wild type isolate of SBMV isolated from Arkansas (Lee & Anderson 1998).

			Sizes (nt)					
Species	**Accession Number**	**Total genome size (nt)**	**5' UTR**	**ORF1**	**ORF2a**	**ORF2b**	**ORF4**	**3' UTR**
CoMV	AB040447	4082	70	326	1706	1493	764	225
IYMV	AM990928	4547	45	641	1796	1592	821	166
LTSV*[a]	U31286	4277	78	646	1628	1526	836	127
RGMoV*[a]	AB40446	4209	100	401	1661	1595	707	198
RuCMoV	AM940437	3983	79	494	1196	1592	725	99
RYMV-Ni*[b]	U23142	4449	81	473	1817	1354	584	283
SBMV-Bar*[a]	AF055887	4136	92	446	1727	1569	800	129
SCMoV	AF08001	4258	69	539	1742	1550	761	174
SCPMV*[a]	M23021	4193	49	557	1718	1544	839	138
SeMV*[c]	AY004291	4148	77	482	1768	1565	806	125
SoMV	GQ845002.2	3983	78	494	1658	1598	725	98
TRoV	AY177608	4037	68	1a -437 1b-233	1427	1565	767	72
VTMoV	HM754263	4248	53	458	1717	1647	846	139

*Genomes were corrected as outlined in a-(Meier & Truve 2007), b-(Fargette et al. 2004) and c- (Lokesh et al. 2006).

Compared with the sequenced sobemoviruses species, VTMoV has the largest ORF2b (1647 nt) and ORF4 (846 nt) (Table 3.3), but does not have the largest genome.

3.4 Discussion

Strategy adopted for sequencing of the VTMoV-K1 RNA

Isolate K1 was chosen for sequencing because it had no detectable virusoid RNA. The abundant nature of virusoid RNA in other VTMoV isolates may have resulted in non-specific amplification of the virusoid because of the low specificity PCR conditions required for some of the degenerate primers. The sequenced sobemovirus genomes were aligned and degenerate primers were designed to target conserved regions. These degenerate primers were used to screen nucleic acid extracts from both VTMoV infected and healthy control *N. clevelandii* plants, to identify potential virus derived fragments, which were unique to the VTMoV infected plants. This healthy plant control identified plant derived fragments and allowed these to be removed from evaluation when compared with those from infected plants.

The initial sequence (Fragment 1, Table 3.1) was identified as having a virus origin by analysis with BLAST. Sequence specific primers were designed to target to this sequence and these were combined with degenerate primers in a genome walking approach to sequence in the 5' direction. To ensure that no gaps occurred in the sequence, the primers were designed to produce overlaps between fragments. Degenerate primers were not successful as reverse primers to walk towards the 3' end of the genome, but single primer PCR produced some sequence in this direction. This single primer success was probably due to the low specificity of the PCR conditions, which allowed mis-priming of the single primer in unsequenced regions towards the 3' end of the genome. The 3' and 5' ends of the genome were amplified using RACE. The 5' end of the genome was confirmed by the presence of a conserved sobemovirus sequence motif ACAAAA (see Chapter 4). The 3' end of the genome sequence was obtained from the longest clone sequences of Fragment 10 (Table 3.1).

The 4248 nt genome of VTMoV is within the range of genome sizes of other sobemovirus species (Table 3.3). Analysis of the four largest ORFs showed that VTMoV has a genome organisation similar to those of other sobemoviruses (Meier & Truve 2007). When the surrounding sequence contexts of the four VTMoV ORFs are examined, three of the start codons are in a favourable context for efficient translation. Although the start codon for ORF1 is not in a favourable context for translation, this is expected for a leaky scanning mode of translation. When the first start codon is in an unfavourable

context, this allows scanning ribosomes to bypass the first start codon (ORF1) and continue to the second start codon (ORF2a) which is in a more favourable context (Koramova et al. 2009). In addition, the size of the ORFs and UTRs in VTMoV are comparable to other sobemovirus species, although VTMoV has the largest ORF4 (CP) and ORF2b (RdRp) within the genus. Annotation of this genome is presented in Chapter 4.

Presence of a virusoid

RT-PCR has detected a virusoid sequence in VTMoV isolate K1, which has not been detected previously by denaturing PAGE analysis (Francki et al. 1986a). RT-PCR is a much more sensitive technique because it can amplify from a single copy of sequence. This difference in detection suggests that a trace of virusoid sequence remained in isolate K1, and that it may have a function required by the virus. This is a hypothesis that needs to be tested in the future.

- Chapter 4 -
Annotation of the VTMoV genome

4.1. Introduction

This chapter outlines the annotation of features of the VTMoV genome sequence described in Chapter 3. This annotation includes features of both the nucleotide and amino acid sequences. Putative functions of the ORFs are described and compared to those identified in other sobemoviruses. A phylogenetic analysis is presented which classifies the taxonomy of VTMoV based on nucleotide sequence.

4.2. Materials and methods

The VTMoV genome was compared with those of published sobemoviruses. RNA secondary structure was predicted using Mfold version 2.3 (Zuker 2003) at temperatures of 37 ºC (default) and 47 ºC and ionic conditions of 1 M NaCl with no divalent ions.

Amino acid analyses

Amino acid sequences from translation products of the predicted ORFs were annotated by comparison with published sobemovirus proteins (Table 4.2.). Bioinformatics tools included Nucleotide localisation signal (NLS) (Nair et al. 2003) and TargetP (Emanuelsson et al. 2007) for signal peptide prediction, and Jpred (Cole et al. 2008) for predicting secondary structure. Tertiary structure and protein function were predicted using I-TASSER, an automated protein prediction server (Roy et al. 2010).

Phylogenetic analyses

Data sets of both nucleotide and amino acid sequences for sobemovirus species were accessed from GenBank (accession numbers Table 3.3) and compared with VTMoV ORFs (Table 3.2). Sequence data sets were aligned using Clustal X2 (Larkin et al. 2007), before analysis with phylogenetics software Mega 4.0 (Kumar et al. 2008). Default settings were used for both neighbour joining and maximum parsimony analyses. Bootstrap analysis of trees with a replication of 1000 assessed the accuracy of phylogenetic trees. Refer to Appendices for alignments used in analysis for Figure 4.9 (Appendix 1), Figure 4.10 (Appendix 7).

4.3. Results

4.3.1 Nucleotide sequence annotation

VTMoV nucleotide sequence annotation presented in Table 4.1 and in Figures 4.1 – 4.4 was based on comparison with described sobemovirus sequence and predicted RNA secondary structure analysis.

Table 4.1. Nucleotide sequence motifs identified in the VTMoV genome

Nucleotide sequence motifs			
Motif	**Nucleotide position**	**Description**	**Reference**
ACAAAA	1-6	Precedes ORF1 start codon	(Tamm & Truve 2000 ; McGavin & MacFarlane 2009)
"	3226-3231	Precedes ORF4 start codon	"
AAGGGAAA	14-21	Polypurine tract	(Hacker & Sivakumaran 1997)
(see Figure 4.1)	1-49	Stem loop structure	Mfold prediction (Zuker 2003)
(see Figure 4.1)	527-568	"	"
UUUAAAC	1771-1777	-1 frameshift signal	(Makinen et al. 1995a)
(see Figure 4.1)	1783 -1813	Stem loop structure	"
(see Figure 4.2)	4109 - 4248	3' UTR multiple stem loop structures (see Figure 4.2)	(Govind & Savithri 2010) Mfold prediction (Zuker 2003)

Figure 4.1 shows three sequence features from Table 4.1.

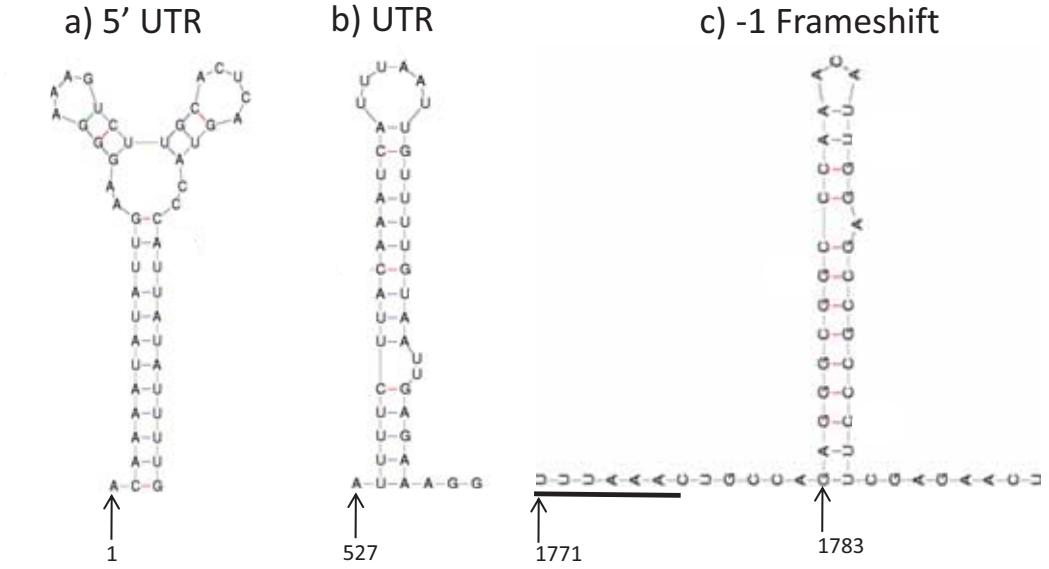


Figure 4.1. RNA secondary structures identified in VTMoV genome, predicted using Mfold with default settings. a) 5' UTR stem loop structure; b) stem loop structure in UTR between ORF1 and ORF2a, and c) -1 frameshift heptanucleotide signal (underlined) and associated stem loop structure. Numbers identify nucleotide positions.

As well as the stem loop structures in Figure 4.1, the 3' UTR of VTMoV contains multiple stem loop structures (Figure 4.2). The structures were identical, regardless of whether the predictions were made at a 37 ºC or 47 ºC folding temperature (data not shown), confirming the stability of these structures.

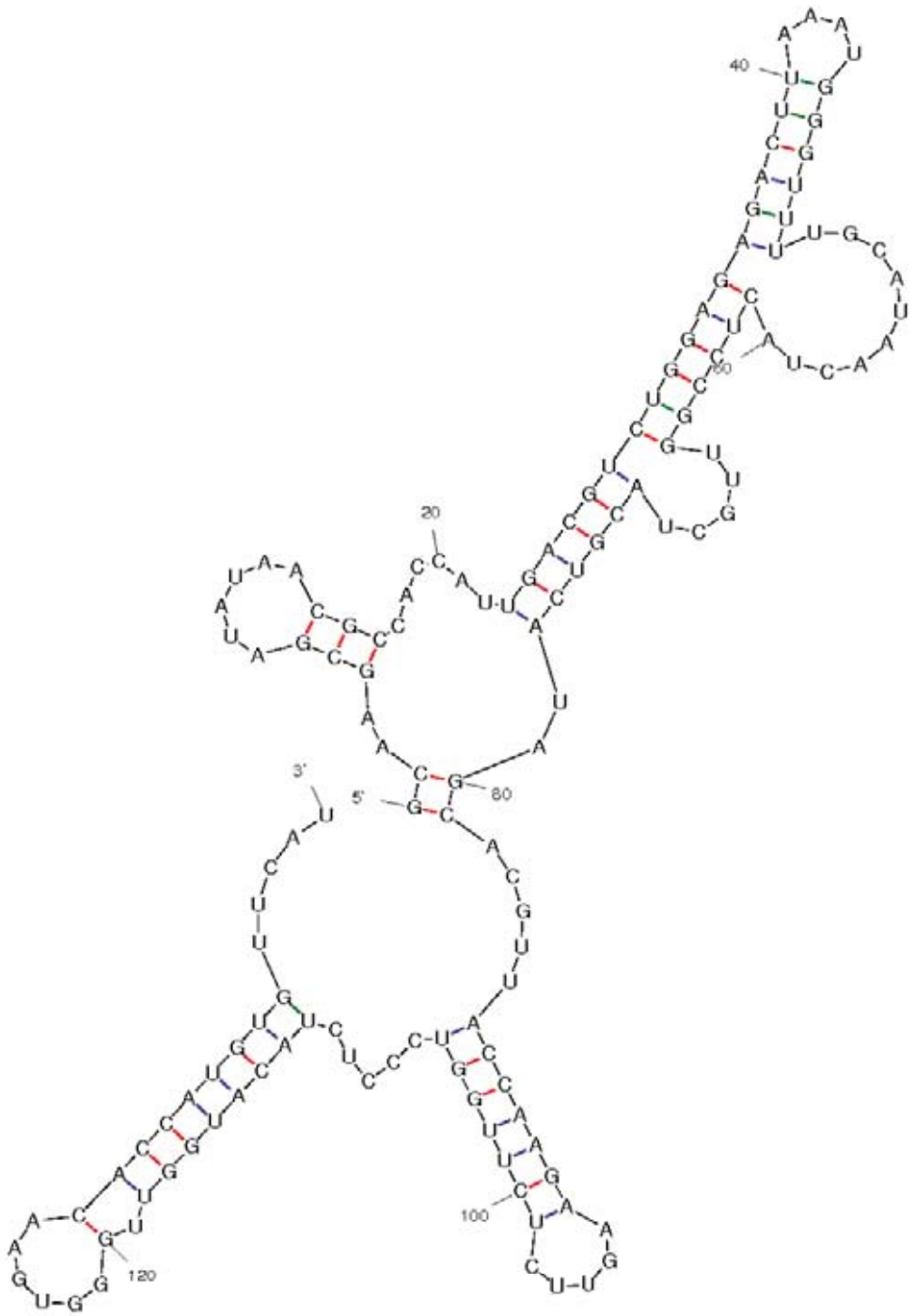


Figure 4.2. RNA secondary structure of 139 nt at the 3' terminus of the VTMoV genome (complete 3' UTR) predicted using Mfold with default settings. Numbers refer to nucleotide positions 4109 -4248; 1= 4109. Red lines indicate bonds between G-C nucleotides

To check whether these secondary features are present in the context of the full RNA secondary structure of the genomic RNA, the VTMoV RNA genome was analysed using Mfold at 37 ºC (Figure 4.3).

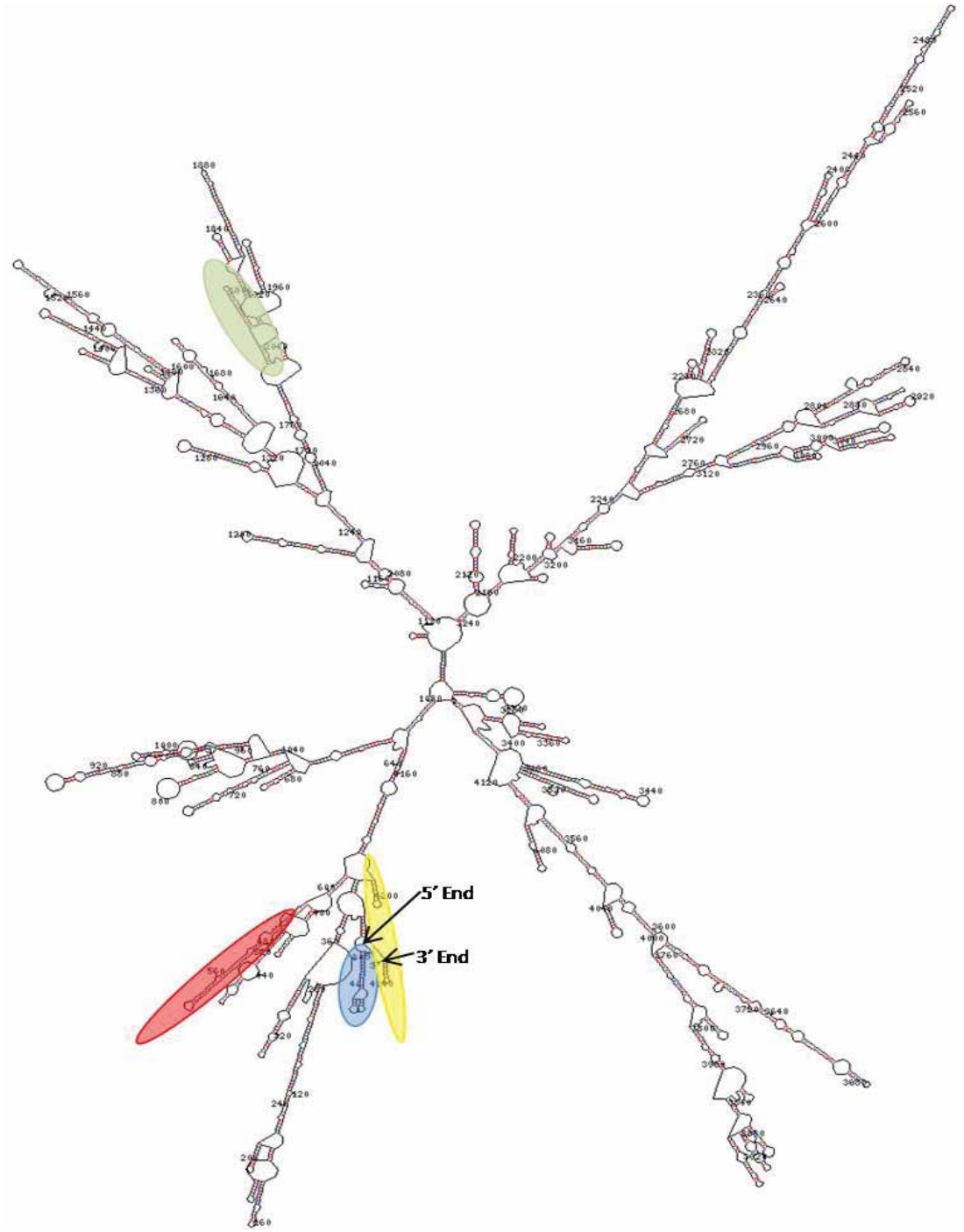


Figure 4.3. Predicted secondary structure of complete VTMoV genome at 37 ºC. Secondary structure motifs from Figure 4.1 and Figure 4.2 are highlighted; the 5' UTR stem loop structure is in blue; the stem loop structure in the UTR between ORF1 and ORF2a is in red; the -1 frameshift is in green, and the 3' UTR is in yellow. The 5' and 3' ends of the genome are marked. Red lines indicate bonds between G-C nucleotides

The RNA genome of VTMoV was therefore predicted to contain secondary structures and multiple stem loop regions. To assess the stability of these regions, the Mfold analysis was repeated at 47 ºC (Figure 4.4).

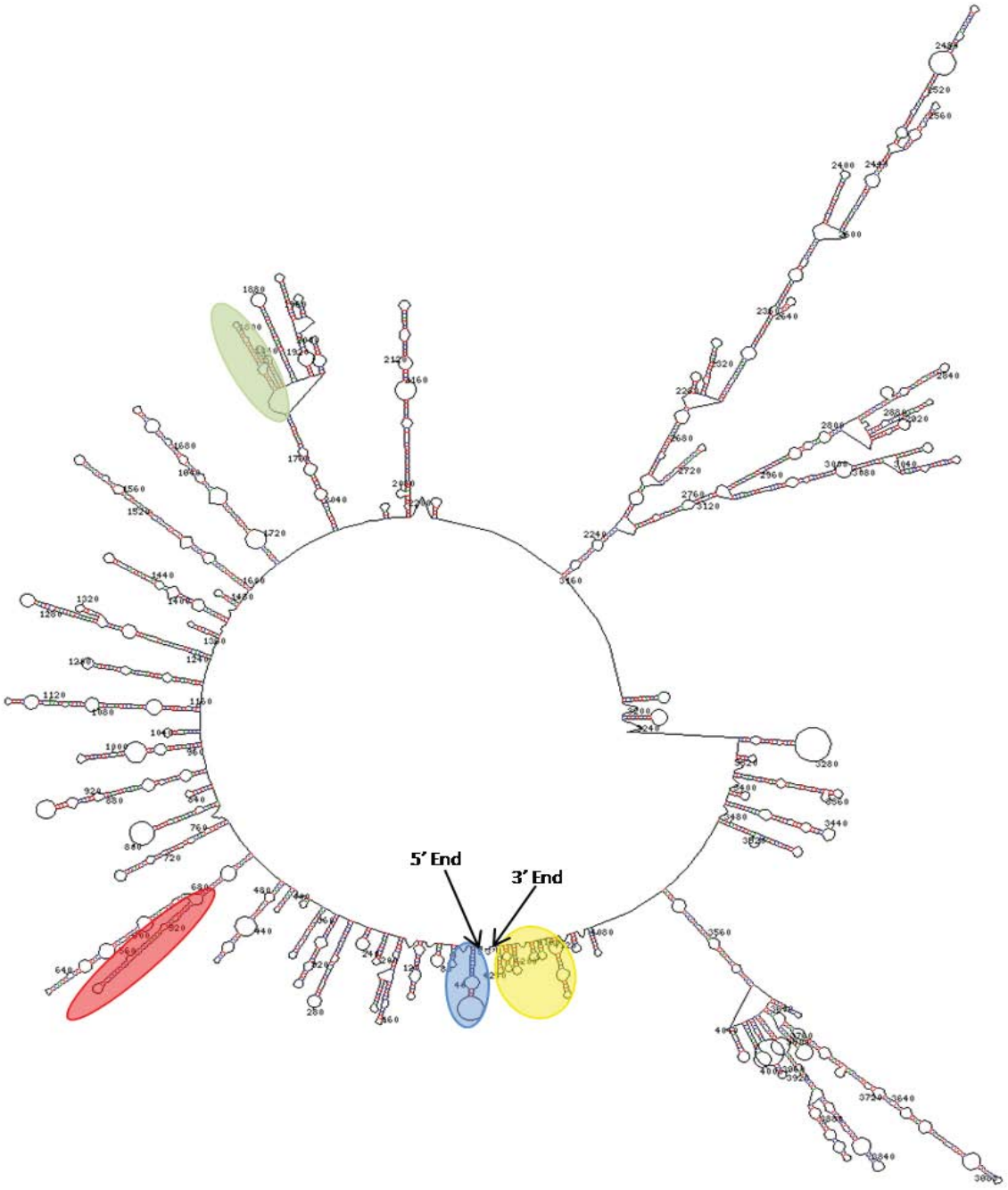


Figure 4.4. Predicted secondary structure of complete VTMoV genome at 47 ºC.Secondary structure motifs from Figure 4.1 and Figure 4.2 are highlighted; the 5' UTR stem loop structure is in blue; the stem loop structure in the UTR between ORF1 and ORF2a is in red; the -1 frameshift is in green, and the 3' UTR is in yellow. The 5' and 3' ends of the genome are marked. Red lines indicate bonds between G-C nucleotides.

At the higher temperature, the predicted structures from Figure 4.1 were stable whereas many other secondary structures predicted at 37 ºC (Figure 4.3) were not predicted at 47 ºC due to their lower stability. Moreover, the secondary structures of the 3' UTR (Figure 4.2) are also seen more clearly at 47 ºC than at 37 ºC.

4.3.2 Annotation of putative translation products

The amino acid sequences of translation products from predicted ORFs (Chapter 3) were annotated by referring to described sobemovirus sequence and protein analysis methods. Results are listed in Table 4.2.

Table 4. 2. Protein motifs identified from the VTMoV genome.

Protein sequence motifs			
Motif	**Amino acid positions**	**Description**	**Reference**
ORF1 (P1)		**Unknown**	
No motifs			
ORF2a (P2a)		**Serine protease-VPg polyprotein**	
	1-30	Signal Peptide	(Emanuelsson et al. 2007)
$H(x_{34})D(x_{62})TxxGxxGS$	160-272	Serine protease	(Gorbalenya et al. 1988 ; Gorbalenya et al. 1989)
EEGDDE	371-377	VPg: ED rich region	(Makinen et al. 1995b)
EN	278-279, 416-417	Potential polyprotein dipeptide cleavage site	(Makinen et al. 2000)
ES	329-330, 440-441	"	(Gorbalenya et al. 1988)
ET	312-313, 396-398	"	"
ORF2b (P2b)		**Replicase**	
EN	90-91, 423-424	Potential polyprotein dipeptide cleavage site	(Makinen et al. 2000)
ES	472-473	"	(Gorbalenya et al. 1988)
ET	468-469, 495-496	"	"
GxxxTxxxN(x19)GDD	326-357	Conserved feature in RdRp	(Kamer & Argos 1984)
RdRp Motifs I-VII	I -172-183 II-196-221 III- 230-238 IV- 260-273 V- 319-343 VI- 350-360 VII- 394-403	Conserved Motifs in RdRp used for taxonomic groupings	(Koonin & Dolja 1993)
ORF4 (CP)		**Coat Protein**	
R (Random) domain	1-67	Random domain of protein subunits	(Hermodson et al. 1982)
KNQVKQMIQ TSARSRRRR	7-15 23-31	α-helices in random domain	(Lee et al. 2001 ; Lee & Hacker 2001)
RSRRRRRRRR		Arginine Rich Motif (ARM) required for RNA encapsidation	(Satheshkumar et al. 2005)
	25-34	Nuclear localisation signal	(Nair et al. 2003 ; Emanuelsson et al. 2007 ; Olspert et al. 2010)
STVMAPMAGAVIYR	42-55	β-annulus region	(Hermodson et al. 1982)
DxxD Y NT	137-141 200 251-252	Calcium binding sites	(Qu et al. 2000)
S (Shell) domain	68-282	Shell domain involved in CP subunit –subunit interactions	(Hermodson et al. 1982)

4.3.3. Tertiary structure and function predictions of VTMoV proteins

Tertiary structural analysis of VTMoV P1

VTMoV ORF1 is in the same genome position as P1 proteins from other sobemovirus species. Within the genus, the P1 amino acid sequences from different species show no sequence similarity. P1 of VTMoV also shows no sequence similarity (nucleotide or amino acid) to any sequences on GenBank. For this reason, the tertiary structure of the P1 protein was predicted using the I-TASSER program (detailed output in Appendix 6). This analysis produced a prediction of the tertiary structure (3D model) as shown in Figure 4.5. The confidence score was -2.08 which is below the generally accepted accuracy range of -1.5 (Roy et al. 2010).

Figure 4.5. 3D structure of P1 predicted by I-TASSER analysis. The α-helices are in pink, β-sheets are in yellow and coiled regions are in blue.

Structural analogues of the VTMoV P1 identified by I-TASSER analysis all belong to the Lipocalin protein family (Pfam PF0061) (Finn et al. 2010). I-TASSER analysis also identified a ligand binding site that may bind ligands such as heme, retinoic acid, Fluorescin, biliverdin IX gamma chromophore, palmitic acid, or 3,6-bis(methylene) decanoic acid.

Polyproteins 2a and 2b

The motifs of P2a and P2b noted in Table 4.2 are mapped in Figure 4.6.

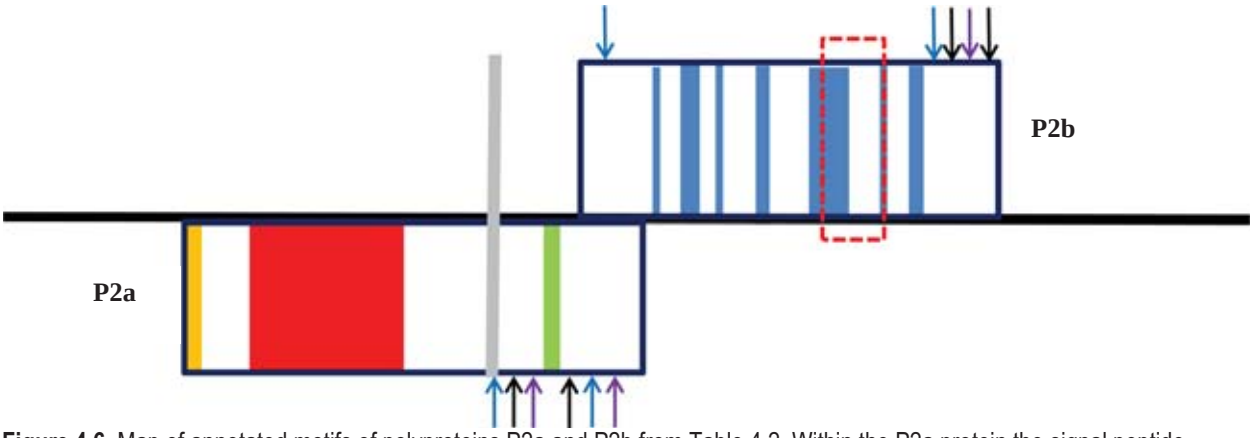


Figure 4.6. Map of annotated motifs of polyproteins P2a and P2b from Table 4.2. Within the P2a protein the signal peptide (gold box), protease domain (red box), ED rich region (green box) and the -1 frameshift (grey line) are shown. In the P2b protein the GDD motif (dashed red region) and the seven RdRp motifs (blue boxes) are shown. Arrows mark potential cleavage sites; EN in blue, ES in purple and ET in black.

P2a contains a signal peptide, but analysis could not determine whether the target was mitochondria, chloroplast, nucleus or a secretory pathway (data not shown). The next motif identified in P2a was the serine protease domain, followed by the VPg domain. The VPg contains the characteristic ED rich region (Tamm & Truve 2000) but there is no upstream [W(A/D)G] (Tamm & Truve 2000) or WNK motifs (McGavin & MacFarlane 2009) which have been observed in other sobemovirus and RuCMV respectively.

P2b is always translated as part of a P2aP2b polyprotein through the -1 frameshift motif (Table 4.1). The P2b translation region comprises the GDD motif (Kamer & Argos 1984) and motifs I-VII (Koonin 1991), which are characteristic of the RdRp.

Potential dipeptide cut sites EN, ES and ET were also identified (arrows in Figure 4.6). These cleavage points cut polyproteins P2a and P2a2b into individual protein components. However, no putative cut sites could be established based on be alignment with other sobemovirus polyproteins (data not shown). Neither could a consensus motif be identified in adjacent sequence to any of the 11 potential cut sites (data not shown).

Coat protein

The VTMoV CP has two large domains; an N-terminal Random (R) domain which is internalised in the capsid of the virion and an outer shell (S) domain. Figure 4.7 shows features of the CP noted in Table 4.2 and indicate positions of the two predicted α-helices and a β-sheet in the R domain.

>*MSKKLTKNQVKQMIQATLPKEQTSARSRRRRRRRSTQQGQSSTVMAPMAGAVIYRKRPMLINGRSG*VTVRHSEV
VLVVQSGTTNFSATSVTLAPNTFTWLAVQGSLYSKWRWISLRATYVPETASTTPGTVAMGFQY**DNTD**VLPTGTAG
MSSLHGFVSGAPWSGFSGSKLLAESPTTPIPAGAIATRLDCQNFGLKWYQ**Y**KSVIPAGDSGNIYIPAQLIVGTLG
TGSTLRYGEVHIQYEIEFIEPLPPSV**NT**LRDIYSRMILSSASDERKGEKSSQTLTDR*

Figure 4.7. Annotation of the VTMoV coat protein amino acid sequence with features described in Table 4.2, including the R domain (italics) with predicted secondary structures highlighted; α-helices are in green, and β-sheets in yellow. The shell domain (plain text) includes the Calcium binding sites in red letters.

The tertiary structure of coat protein was determined by I-TASSER (Figure 4.8), and the 3D model has a confidence score of -2.42, below the -1.5 accuracy point (Roy et al. 2010). All structural analogues identified by I-TASSER analysis were previously described sobemovirus coat proteins.

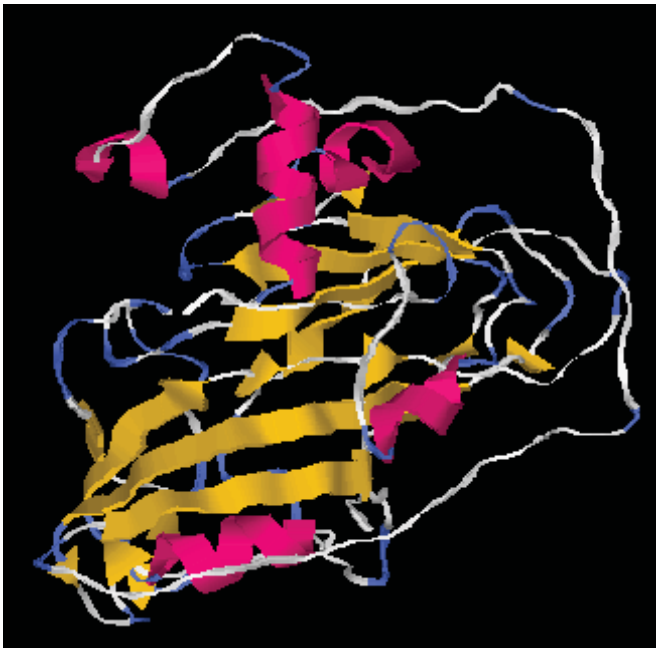


Figure 4.8. Tertiary structure of VTMoV-CP predicted using I-TASSER; α-helices are in pink, β-sheets are in yellow and coiled regions are in blue.

4.3.4 Phylogenetic analysis

Figure 4.9 shows the phylogenetic tree created from a neighbour joining analysis of the complete nucleotide sequences of described sobemovirus species and two outgroup species from the related *Luteovirus* genus (refer to Appendix 1 for multiple sequence alignment).

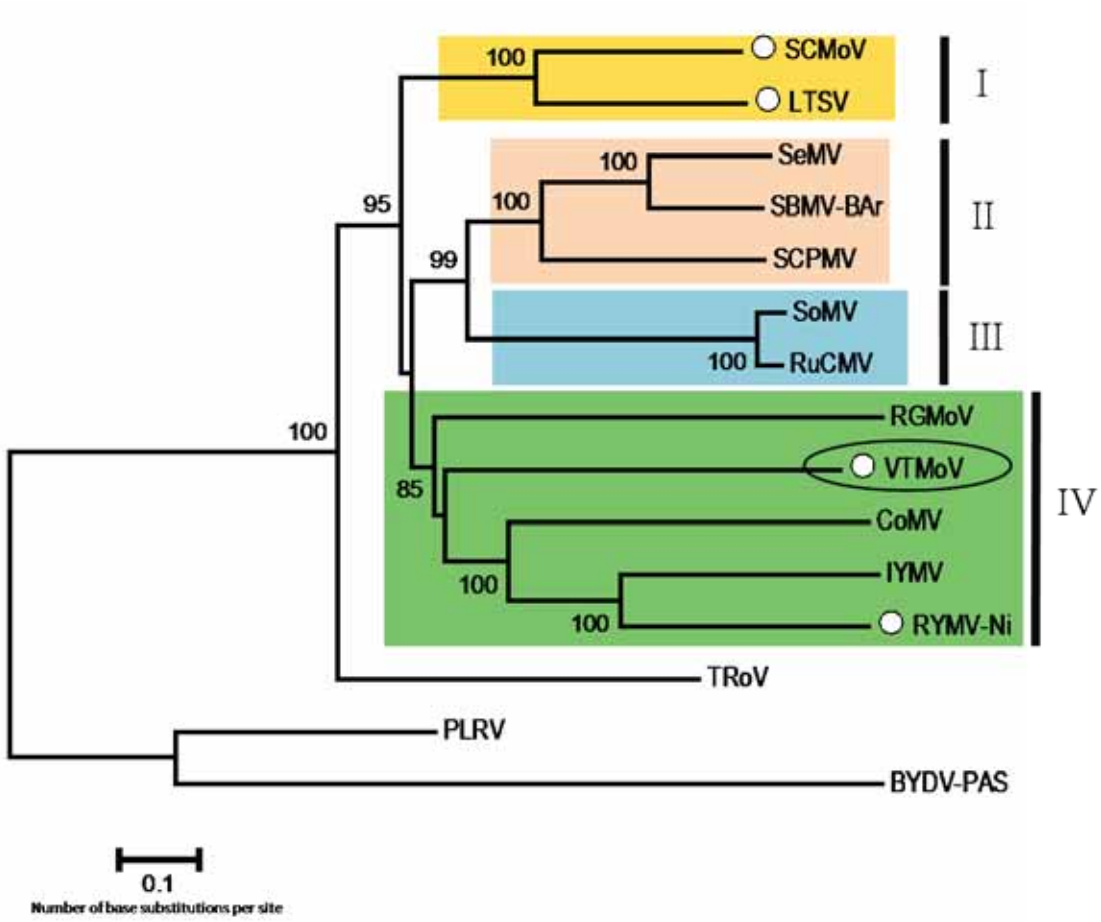


Figure 4.9. Phylogenetic tree from the complete nucleotide sequence of VTMoV and other sobemovirus species, constructed using the Neighbour Joining method. Numbers indicate the percentage of trees in which this grouping occurred after bootstrapping with 1000 replicates (values shown only when >80%. Virus abbreviations and accession numbers are in Chapter 3, Table 3.3. Outgroups are *Barley yellow dwarf virus* (BYDV-PAS, NC002160) and *Potato leaf roll virus* (PLRV, NC001747) luteoviruses. White circles indicate those species reported to contain a satellite/virusoid RNA.

VTMoV grouped with other sobemovirus species and was distinct from the luteovirus outgroups, with high bootstrap support greater than 80 %, which indicates a confidence interval of ~95 % (Xiong 2006). Four distinct groups within the sobemovirus genus were observed with high bootstrap; Group I includes LTSV and SCMoV, Group II includes the type virus SBMV-BAr, SCPMV and SeMV, Group III includes RuCMV and SoMV, and the Group IV includes RYMV-Ni, IYMV, CoMV, RGMoV and VTMoV. Some of these groupings can be characterised by shared features of member species, Group I contains two sobemoviruses isolated from Australia. Group IV except for VTMoV all infect grass host plants (monocotyledons).

Next, phylogenetic trees were created for each sobemovirus ORF, using both nucleotide and amino acid sequences. These trees are presented in Figure 4.10. The phylogenetic analysis of ORF1/P1 was excluded as it had minimal bootstrap support (refer to Appendix 8)

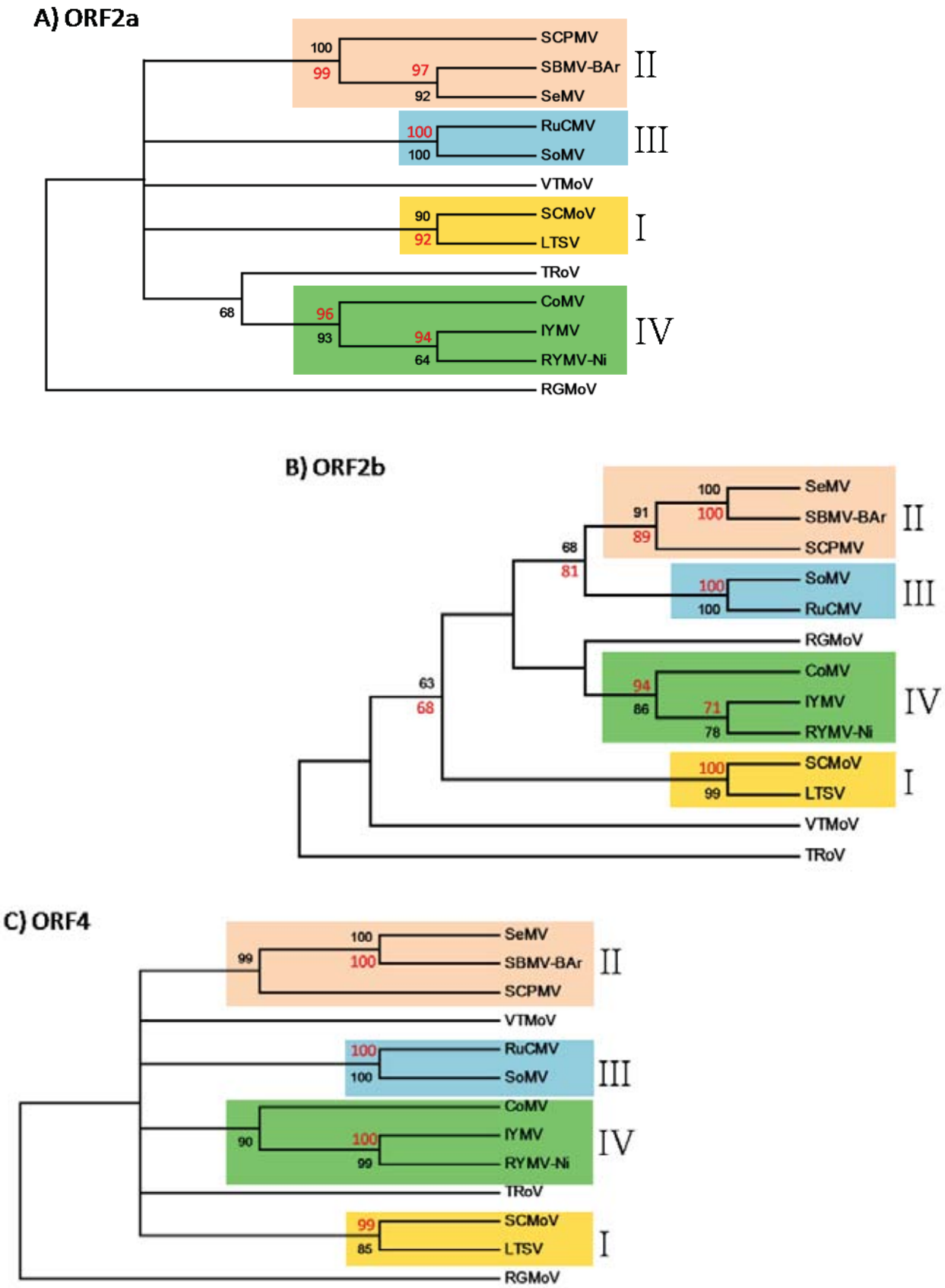


Figure 4.10. Phylogenetic trees of each sobemovirus gene; A) ORF2a; (B) ORF2b (C) ORF4. Trees were created using the Maximum Parsimony method, and bootstrapping was conducted with 1000 replicates. Numbers represent percentage values for bootstrapping support of branches for either amino acid sequence (black) or nucleotide sequence (red). Lineages are colour coded as in Figure 4.9. For details of sequences used refer to Table 3.3.

The three sobemovirus gene trees have the same groups as in Figure 4.9, except that VTMoV and RGMoV do not group together in Group IV as in Figure 4.9. Instead VTMoV, RGMoV and TRoV form 3 monophyletic groups, although there is low support for a grouping between TRoV and VTMoV in the RdRp gene tree (Figure 4.10b).

4.4. Discussion

Annotation of the genome of VTMoV shows that it has most of the features of other sobemoviruses, and in the same genome location. In addition, phylogenetic analyses of VTMoV sequence confirm its grouping with other species within the genus *Sobemovirus*. These analyses and annotations support the current taxonomic classification of VTMoV as a species in this genus (Fauquet et al. 2005).

4.4.1 Annotation of nucleotide sequence

Annotation of the VTMoV nucleotide sequence identified features common to other sobemovirus species. In VTMoV the motif ACAAAA is found in at the start of the gRNA, and upstream of the CP ORF. In SBMV the CP is translated from a sgRNA (Ghosh et al. 1981). The CACAAA motif - where the 5' C may be absent – (Tamm & Truve 2000 ; McGavin & MacFarlane 2009) is found at the 5' end of both the genomic and subgenomic RNAs [except for CoMV (Tamm & Truve 2000)]. The presence of this ACAAA motif at the start of both the gRNA and sgRNA suggests a role in virus replication (Tamm & Truve 2000). This ACAAAA feature is also found in poleroviruses and is thought to function as an enhancer for polymerase binding (Miller et al. 1995), but this is yet to be tested experimentally (Tamm & Truve 2000).

The -1 frameshift heptanucleotide and downstream stem-loop structure are present in all sobemovirus sequences, but the function has only been demonstrated experimentally for CoMV (Tamm et al. 2009). In CoMV, this feature was shown to be important for inducing ribosomal frameshifting, and in particular the length of the stem-loop is important for frameshifting efficiency (Tamm et al. 2009).

The VTMoV genome has other stable putative stem-loop features, upstream of both the first and second ORFs and in the 3' UTR. In SeMV the RdRp can synthesise RNA in the absence of the VPg, but only if the 3' UTR sequence is present (Govind & Savithri 2010). Mutational analysis of the 3' UTR of SeMV demonstrated that the stem loop structure contributed significantly to template specificity of replication. Further analysis showed that the SeMV RdRp was able to interact with different stem-loop structure conformations, suggesting that the RdRp - stem-loop interaction is not sequence specific (Govind &

Savithri 2010). Further experimentation would be required to test whether the stem loop structures play a similar role in replication of VTMoV.

4.4.2 VTMoV proteins

P1

There is no similarity among the P1 protein sequences of the described sobemovirus species. However, the function of P1 may be conserved between sobemoviruses because P1 in both RYMV and CoMV function in cell to cell and systemic movement (Bonneau et al. 1998 ; Meier et al. 2006) and in suppression of gene silencing (Voinnet et al. 1999 ; Sarmiento et al. 2007). To infer the function of the VTMoV P1 protein, the tertiary structure was analysed using I-TASSER (Zhang 2008), which has been ranked as the best tertiary structure prediction server (Zhang 2008). This server predicts the 3D structure of protein sequences and then ascribes potential functions based on structural alignment outputs (Roy et al. 2010). The confidence score of the VTMoV-P1 3D model (-2.08) was low, as generally a score between -1.5 to 2 is considered to be accurate (Roy et al. 2010). This low score could be due to a lack of close protein structure templates (Roy et al. 2010). All structural analogues identified for VTMoV-P1 belonged to the lipocalin protein family [Pfam PF0061 (Finn et al. 2010)]. This protein family is a widely distributed heterogeneous group found in plants, animals and bacteria (Grzyb et al. 2006) with a diverse functionality (Pugalenthi et al. 2010) transporting and binding small hydrophobic molecules (Grzyb et al. 2006). This structural analysis is also supported by the I-TASSER functional analysis, which indicates that P1 is possibly involved in binding and/or transport. Potential ligand homologues for VTMoV-P1 include the same types of small hydrophobic molecules such as heme, retinoic acid and palmitic acid.

In plants, lipocalin proteins are separated on the basis of presence or absence of a chloroplastic targeting peptide (Charron et al. 2005). In VTMoV-P1 this targeting peptide is lacking (data not shown). Plant lipocalins lacking this chloroplastic targeting signal are called Temperature-Induced lipocalins (TILs) and are associated with an abiotic stress response (Charron et al. 2005). If VTMoV-P1 is a lipocalin protein, then it might play a similar role in plant stress responses. If VTMoV-P1 binds ligands such as retinoic acid - which is structurally similar to the plant hormone abscisic acid (Xie et al. 2005), then it may inhibit a plant defence pathway.

Whether P1 affects systemic movement of VTMoV or suppresses gene silencing is not known, although these functions of silencing and movement may explain why there is no homology between P1 proteins within the genus. Sobemoviruses have distinct and narrow host ranges, and different P1 proteins may

be required to perform these functions in different plant hosts. In fact, the distinct interaction of viral MPs with plant host proteins is thought to be a crucial factor in determining if plants can serve as viral hosts (Scholthof 2005). This diversification in the sobemovirus P1 would be expected if it does interact with plant host factors, as both plant and virus factors would be involved in an arms race to overcome each other.

Polyproteins P2a and P2aP2b

VTMoV has the same domain order Protease-VPg-RdRp as other sobemovirus species (Tamm & Truve 2000). The presence of a signal peptide may ensure that VTMoV proteins encoded by these polyproteins are present concurrently and the virus will have all the proteins needed to complete its life cycle efficiently (Satheshkumar et al. 2004). While these three domains can all be identified, the cleavage points between them remain unknown. Serine proteases are extremely specific because primarily they process their respective polyproteins (Gayathri et al. 2006), although in sobemoviruses the serine proteases were suggested to cut at either ET or ES dipeptide motifs (Gorbalenya et al. 1988). Experimentally the specific cleavage motif of SeMV was shown to be [(Q/T)E(T/S)X] (Nair & Savithri 2010), but in CoMV the cleavage motif [VENSxLxPxxSS] was established that cut across an EN dipeptide (Makinen et al. 2000). In RYMV and CoMV the C-terminal processing positions for the VPg were determined using mass spectrometry (Olspert et al. 2011). For the CoMV the VPg was C-terminal was not at an EN dipeptide, but rather at an ET dipeptide (Olspert et al. 2011). Other cleavage motifs in sobemovirus species have been annotated based on the presence of these E/[T/S/N] dipeptide motifs in equivalent locations from protein alignments. However, in VTMoV none of the 11 putative sites could be proposed based on either protein alignment or by the presence of a consensus motif. Therefore, the cleavage points in VTMoV will have to be determined experimentally, perhaps using mass spectrometry to analyse purified VTMoV proteins.

Coat protein

Sobemoviruses have an icosahedral capsid with *T*=3 symmetry with 3 quasi-equivalent subunits (designated A, B and C) (Meier et al. 2008). Coat protein subunits are translated from the sgRNA (Ghosh et al. 1981), but each subunit has minor differences in conformation depending on their axis of symmetry within the capsid (Hebrard et al. 2005). Subunits A and B are disordered in the R domain, but in subunit C the R domain is partially structured (Meier et al. 2008). This plasticity in subunit confirmation may explain the low confidence score of the CP tertiary structure prediction, as sobemovirus CP subunits have three different correct conformations.

In the R domain the formation of protein secondary structure only occurs in the C-subunit and in VTMoV there are three regions of secondary structure, 2 α-helices and a β-sheet in this R domain (Figure 4.7). The β-sheet is closely associated with the β-annulus region, which in SeMV is not crucial for capsid assembly (Satheshkumar et al. 2005).

The two α-helix regions may be involved in RNA binding, because in SCPMV the regions involved in RNA binding are thought to fold into α-helix structures (Lee & Hacker 2001). In addition, the second α-helix region is closely associated with the ARM of VTMoV. In SeMV, the ARM is crucial for RNA encapsidation (Satheshkumar et al. 2005) and in SCPMV, is important but not essential for RNA binding through electrostatic interactions (Lee & Hacker 2001). Recently, in addition to an RNA binding function, the ARM of CoMV has been shown to contain a nuclear localisation signal (NLS) (Olspert et al. 2010). The presence of an NLS in the ARM can also been predicted for VTMoV from the results presented here.

In addition to protein-protein and protein-RNA interactions, there are calcium binding sites in the shell domain that are crucial for subunit-subunit interactions and overall virion stability. For SBMV, removal of calcium and magnesium destabilises the virion shell (Hsu et al. 1976). In VTMoV a similar process may be occurring. When VTMoV virions were treated with 10 mM EDTA before RNA extraction (section 2.3.6), an increased RNA yield compared to untreated virions was observed (data not shown). The removal of calcium by EDTA may have broken the subunit – subunit interactions within the virion as observed in other sobemoviruses (Hull 1977), which may have caused the VTMoV particles to swell. This swelling in the virion may have enhanced the effect of the protease digestion allowing more viral RNA to be extracted.

4.4.3 Phylogenetic analysis

VTMoV is a sobemovirus, as the whole genome sequence alignment clearly grouped VTMoV within the genus. Within this genus, four clear groupings were identified by phylogenetic analysis. These groupings may be linked to either geographical (Group I, Australian viruses) or biological features (Group IV, monocotyledonous plant infecting viruses). VTMoV and RGMoV cluster with this Group IV in one analysis (Figure 4.9). However, analysis of different ORF sequence show VTMoV is a unique monophyletic group, like RGMoV and TRoV. Additional sequences from other sobemovirus species such as SNMoV and completion of the PLYV genome sequence will be needed to be analysed to clarify these groupings.

- Chapter 5 -
VTMoV isolate R17

5.1 Introduction

The sequence obtained from VTMoV-K1 provided primers for determining the sequence of a second isolate of VTMoV. Isolate R17 was selected because it is the type isolate of VTMoV and has a symptom phenotype in *N. clevelandii* different from that of VTMoV-K1 (Francki et al. 1986a). R17 induces necrotic local lesions on inoculated leaves of *N. clevelandii* (not observed in K1) and more severe stunting of growth than for isolate K1. The results described in this chapter include sequencing and annotation of the R17 genome and its associated virusoid and a comparison with sequence of isolate K1.

5.2 Materials and Methods

For recipes of reagents refer to Appendix 3. For details of primers (primer sequence, source and melting temperatures) refer to Appendix 4.

5.2.1 Sequence fragment collection

Total nucleic acid extraction

Dried leaf material (5 –10 mg) from WINC specimen 65783 was ground to a powder with a pestle in a micro centrifuge tube immersed in liquid nitrogen. Total nucleic acids were extracted by adding 50 volumes (w/v) of guanidine thiocyanate extraction buffer, and then continuing as described in section 2.3.5.

RT-PCR

PCR primer combinations used to amplify VTMoV genomic fragments are listed in Table 5.1.

Sequencing fragments

DNA from PCR fragments was isolated as in section 2.3.11, and cloned as outlined in section 2.3.12 before sequencing as described in section 2.3.13. Sequence was determined for multiple clones for each PCR fragment and a consensus sequence was created.

5.2.2 Sequence analysis

Consensus sequences for each fragment were aligned using BioEdit (Hall 1999) and a final consensus sequence was determined. Where sequence mismatches occurred, sequences near the middle of fragments were given precedence over those near the ends of fragments to account for reduction in sequence read quality.

Genome sequence assembly

Overlapping fragments (Table 5.1) were aligned to create a consensus sequence. This was compared with the nucleotide and amino acid sequence of VTMoV isolate K1 using BioEdit (Hall 1999).

5.3 Results

5.3.1 Nucleotide sequence

The sequence fragments from VTMoV isolate R17 (Table 5.1) were assembled to produce a single contiguous sequence (Figure 5.1).

Table 5.1. PCR fragments used to assemble the VTMoV R17 genome sequence and virusoid sequence. Fragments are defined by size, genome region and the primers used to for PCR and first strand cDNA synthesis. Sequences of all fragments are shown in Appendix 9.

Fragment	Primers used		PCR product size (bp)	Map position (nt)
	First strand cDNA synthesis	PCR		
1	VTMoV-2aR2	VTMoV-5UTRF VTMoV-2aR2	1662	1-1662
2	VTMoV-2aR1	VTMoV-2aF1 VTMoV-2aR1	576	1380 -1956
3	VTMoV-CPR2	VTMoV-2bF2 VTMoV-2bR1	1998	1720-3718
4	VTMoV-CPR4	VTMoV-CPF5 VTMoV-CPR4	1041	3067-4108
B	AGSATR	AGSATF AGSATR	366	Virusoid sequence

The assembled sequence from VTMoV isolate R17 comprised 4108 nt and had a base composition of 25 % adenine (1043 nt), 22 % guanine (918 nt), 26 % uracil (1053 nt) and 27 % cytosine (1094 nt), with an overall G+C content of 49 %. This sequence (Figure 5.1) was submitted to GenBank and assigned accession number HQ680399.

>ATATTGAAGGGAAAGTCTTGCACTCAGTACCCATTATATTTTGATATGCCCAGCATTGATGTTGAAGTAGAAAA
GATCCTGCACTTGAGCAACAGAAAGAGACTAAGGTCTGTCTGTGTGGCGAGGAAGAAGACGTTCGTAGACAAGAT
ATCCGAAGGATTTGTCTGCACTCTTTACGGAGAGTACGATCACGACTACGTCACCAGTGTGCATTTGCACATTGT
TTGCTCGTGCGGTAGAGCTTTCTTTGATTTTGTTGAGTTCAAGGACATAAAACTCCGAGATTTCCAATATCAAAG
TCAGTGTATCTGTCCCACCAGGAATTGGTCGACGGTCGCTTACACTGATTTTACTACCGTCAACTGCCAAGTGGA
AGGTTGTGATTACTGTGAGGGGATTGAGTCAGACTCTGATTCGGATTCTGAGGCCATCATAGAGGAATTTCTTCA
AAAGTTTTCCGAGGTTGGCATCTCTGGATCGTCCAGTTCTCCCCAGACTAACTAGGACTCGGTTCCGTTTATTTT
CTTACAAATCATTTAATTGTTTTGTAATTGAGAAAGXATGTTGAGCGAGTTAGTCCAGTTGTGTCTTTCAGCAAC
CATGACTCTAGCGCTGGTAGTGTCAATGATGCTGGACCCGAGCTACGGGTGGCTGATCACCATGTCTGTGATGTT
AGTCTCCTTAGAGCTGACAATCAGGCCGTTCAGGCGCTCTCTAGACTACATGAAAATTGTGGTCCGAGATGAGCC
CAATGAGCCTGAAGTTGCTAGGATAGTAGGCGCGCCAACTTACCACCCAATTCATGGCGTTGTTTGCAAGTGTCT
GTGGAACTCGAAAGAGTTTAATGTGGTTGTGCAACCCAGTTTGTGGCCTCTAGCCTTTCCCCCAAACATATCACA
AGGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCT
CGTTGTCTTATACAACGATGGAGTCAGAATCGGTATGGGCTCTAGAGTTAATTGGCAAGGCGCAGATTACTTGTT
GACAGCCTCCCACGTTTGGTCCCTCGTAACTGGTGATTTTCAAATGGCAAAAGCCAGTAAGATGGTCTCAGTGAA
GGACGCCAAGTGTTATGTTGAGGCTGCTCACGCTAAGCTGGACTTTGCCCTAATTAAAGTTCCGAACAAATATTG
GAGCTCGATAGGGGTTGGTTCGGCTAAATTGTTGTGGCATAAGCCCGGGCAAGTGGTCAAGGTTTACGGAGGTAG
ATCTGATGAATTGGTCTCTTCTGTGGGCCGAGCCGAAAAGGATCCTGATCTGAGCCTACGTCTGACCCATAACGC
TTCAACTGCTCCAGGATGGAGCGGCTCCCCGCTTTACAACTCTGAAAATTTCGTAGTGGGATTGCATACTGGCTT
TTCCGCGGCTGAGCAGAGAAATGAGGCGGTTGATGTAGCGAAGTTGCTGCACCTCGCCCTTAGAACAAAGGAAAC
GACGTTTTCGGAGATCGGTGTCTCTCTTATCGATGAAGATGAGATTGAGTCTCGCGGTTACCAATTCGACGACTT
TGAGCTTAGGGGTGAGGTAAACGTTAAAGGGAAAATGGCGCGGAATGAGATATCTCTCATCGCCTCTAAGAATAA
AGGGAAACCTTGGTATCTTCAAGAGGAAGGTGATGATGAATTCTATGACTCGATCGGTGAGAAAGATTTCCTTGC
CCGCTTCCGTGAACAAACTGGTAAAGAGACAGTGGGGAATTTAAACTGCCAGAGGGCGGCCCAAACATTGGAGCC
GCCCTTCGAGAACTTGAGGCCGTGCGATGGGAAGAACCCAGAGCTATTCAAACCGGCGGGATGGGATTCAACTAT
GTTGGAGTCTCGTCTTGCAAGTTTAGAGAGGGCTCTAAGCACGCTGCTCGCGGAGCAATCAGTTCTGCTGAGCAA
GTTTTCCCAGAACTCAAACAGTATGGTTGGCCAGAAAGAGGCTCTAAAGCCGAGTTCGATTCCCTCCTCCTCCAA
GCCAGCCGTTTCCGGAGAACCTCCTGTCCAGAGCAAACCGAGCGTAAATGCGAAGTCCTCGCTGAGAAGTACCCC
AAAAGTAGAGCCCACCGTTGCTTCAGGAGAGAAAACTTCTGTCAACGCCAGCTCCTCAGGGAACAAATCGAAGCG
ACGACGACGTCGCCGGAAATCAACGACAAAGCCAGCCCCGGCTCCCCCTGGTCAAGACTTGAAAAGACCAATGGT
GAGCTCATCTCCCGTTTCAAAGACCTCTTAATTGAGGCAGTGTTGAGGCGAGTTTTATTGCTTGCCTCCACTCCA
ACCAGTGAGATTTTATGCATGTCTGCTTCAGATCTCGTTAGAGCTAACTTAGTGGATCCTGTGCGATTGTTCGTC
AAGCAGGAACCCCATACGAAGAAGAAGCTGAATGAGCGTCGCTTTAGGCTTATTTCGTCTGTGTCTATAGTTGAT
CAGATAATCGAGAGATTATTGTTTGGTCCTCAGAACAGGCTTGAGATAGCCTTATGGCATCAGATACCTTCCAAA
CCTGGAATGGGACTGAGCGCTCGTACGCAAGCTGATTTGTTGTGGAATGAGTTATTCGCAAAGAGCGAGATTGCC
CCTGCGGCCGAGGCTGACATTTCAGGATTTGACTGGTCTGTGCAGGAATGGGAGTTATGGGCCGATCTGAGCATG
AGGATTTCTTTGTGCGAAGATATGCATGATGGTCTACGGAGGTTGATGGTTAACAGATACCGCTGCTTTATGCTC
TCTTGCTTCCAATTGTCGAATGGGGAATTGTATGAACAAGTTGAGCCTGGACTCATGAAGTCTGGTTCCTACTGC
ACTTCCTCTTCCAACTCCAGAATCAGGTGCCTAATGGGCTATCTAATTGGAGCCCCCTGGATAATAGCCATGGGG
GATGATTCTGTTGAGGGTTATGTGAGAGACGCGAAAGGCAAGTATGAGGAATTAGGACACACTTGTAAGGAATAC
GAGTTGTGTGATGTTGATTCAGACGGCGCGTTGAGATCTGTGAACTTTTGTTCACATTTGATTTCCCGCAACAAG
TTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAA
AGGGAACTTGGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAA
ACATATTGGGAAGAAGATCACCCTGCTGACTATGTCGAAGAAACTCACCAAGAACCAGGTCAAGCAGATGATACA
GGCAACCTTGCCGAAAGAGCAGACTTCTGCTCGGTCCCGACGACGGAGACGCCGAAGGTCGACGCAGCAAGGGCA
GAGTTCTACAGTTATGGCCCCAATGGCTGGAGCTGTGATATACCGGAAGCGACCCATGTTAATCAATGGCCGCTC
TGGGGTTACAGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGT
TACTCTCGCGCCTAACACTTTCACCTGGTTAGCTGTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCT
TAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTTTCCAGTATGACAATAC
TGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTT
TAGTGGTTCTAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTG
TCAAAATTTCGGTCTCAAATGGTATCAGTACAAGTCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCC
AGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGCACATCCAGTATGAGAT
TGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGC
GTCAGATGAAAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCGA

Figure 5.1. Nucleotide sequence from VTMoV isolate R17, showing start codons (ATG) for each ORF(boxed) and single nt insertions, deletions or substitutions compared with VTMoV isolate K1 are shown in red letters, X refers to the site of a deletion mutation

Virusoid sequence

Fragment B (Table 5.1) is a 366 bp PCR product representing the VTMoV virusoid. The R17 virusoid sequence (Figure 5.2) had the same sequence as the virusoid from VTMoV isolate K1 (Figure 3.3). When compared to the sequence from the previously sequenced virusoid (NC003906; isolate unknown) there were 2 substitutions 234T>G and 271A>G. The primer design allowed both linear and circular forms of the virusoid to serve as template for RT-PCR amplification. The virusoid sequence from isolate R17 was submitted to GenBank and assigned accession number HQ680398.

GTGGATGTGTATCCACTCTGATGAGTCCGAAAGGACGAAACGGATGTACCGCTTCTTGTCTCGACCTCGACCTGG
ACTAGTGATCGAGGGAGGCTCAACCTCACGCCCGCTGGGTAGATGTAGTCTCATACTCCAATGACTTGGGGTCAC
TGTGTAAAGGTACTACAGAGCTACGACCATGTGATAGGCGGGGAGCTGGACCCTCTCACCACCTAGGTAGTGTTG
AAGGTCGCGAGGGAGTCAAGGACGCCCGGCATCAGAGGATTGCACGCCACCGGTATCACGGAGGGCAACTGCTT
TCCAGGCTGGCAGGTAACGTTCCTGCCCTTGGGGACTGATTTTTGGTTCGCCTGGTCCGTGTCCGTA

Figure 5. 2. Complete nucleotide sequence of virusoid RNA from VTMoV isolate R17. Nucleotide differences from the previously described sequence (NC003906) are boxed. The putative virusoid self cleavage site is shown in red letters (Keese & Symons 1987). Primer sequences are underlined.

5.3.2 Amino acid sequence

ORF analysis

The VTMoV genome was assessed using Frameplot 2.3.2 for the presence and location of putative ORFs. The output of this analysis is shown in Figure 5.3.

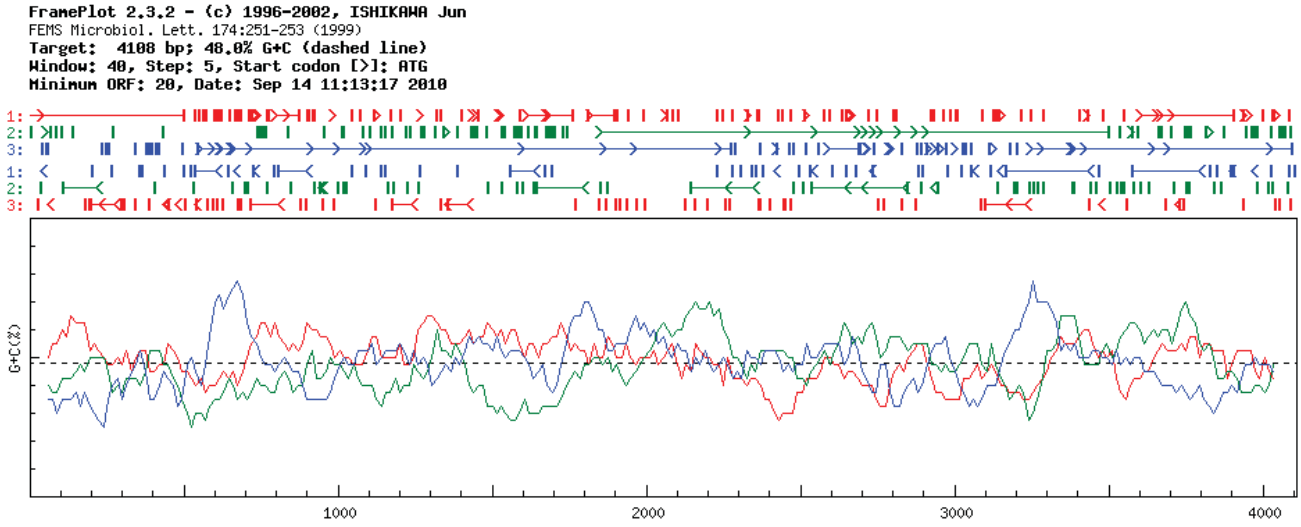


Figure 5.3. Output of Frameplot 2.3.2 analysis of the VTMoV isolate R17 genome sequence. Predicted ORFs are shown as horizontal lines with start codons (ATG) indicated by arrows and stop codons (TAA,TAG, TGA) by vertical lines. Different coloured lines refer to different translation frames on both the sense and anti-sense strands. The G+ C plot refers to the G+ C content across the sequence, and numbers refer to nucleotide positions on the input sequence.

The R17 sequence was near full length, missing 8 nt from the 5' end of genome and 132 nt from the 3' end of the genome when compared to the VTMoV-K1 genome (Chapter 3). However, the R17 sequence does contain the complete ORFs of the coding regions shown in K1 (Chapter 4). VTMoV isolate R17 had the same four ORFs and the same ORF organisation as isolate K1 (Figure 5.4).

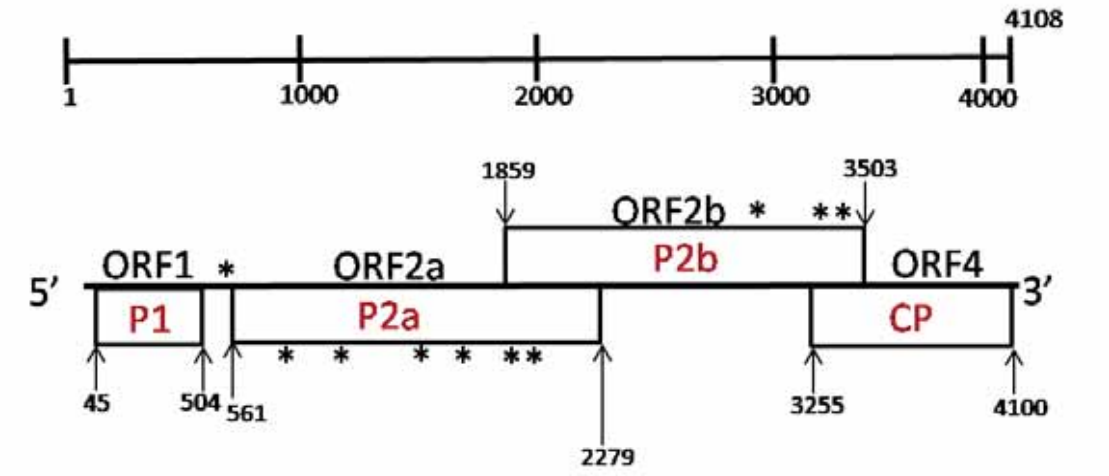


Figure 5.4. Genome map of VTMoV isolate R17. Points of differences with VTMoV-K1 nucleotide sequence are marked with a *. R17 ORFs are labelled in black with the same names as the K1 genome. Annotated sobemovirus protein names are shown in red. The arrows and numbers shown the start and stop codon positions for R17 ORFs on the R17 nucleotide sequence.

In R17 the reading frames for the ORFs differed from K1, in that ORF1 was in a different translation frame from ORF2a and ORF4, whereas in K1, they were both in the same frame as ORF1. The sequence of isolate R17 was identical in the regions surrounding the ATG start codons for both ORF2b and ORF4 and were predicted to be favourable for translation based on the presence of either an A -3 nt upstream of the start codon or a G +4 nt downstream (where the A from ATG start codon is at position +1) (Futterer & Hohn 1996). ORF2a was also in a favourable context for translation despite an adjacent polymorphism (Table 5.2) to the start codon. As in K1, ORF1 was not in a favourable context for translation.

5.3.3. Sequence similarities between isolates of VTMoV

When compared to K1, isolate R17 had the same annotation for the ORFs observed in isolate K1 (see Chapter 4).

5.3.4 Sequence differences between isolates of VTMoV

A comparison of ORFs from R17 and K1 showed there were single nucleotide substitutions, or insertion-deletion differences at 10 sites (Table 5.2).

Table 5.2.Nucleotide sequence differences between VTMoV isolates R17 and K1 are listed sequentially from 5' to 3' based on * in Figure 5.4. The genome positions and nucleotide information from both isolates are presented, including the genome region the substitutions occur in. The types of polymorphisms are labelled; insertion-deletions in the bold outlined box, transitions with ^ and transversions with +. # indicates polymorphisms that cause a change in amino acid sequence.

Polymorphisms	R17 sequence		K1 Sequence		Coding region
	Genome position	nt	Genome position	nt	
1	561	deletion	568	G	UTR
2	842^	T	850	C	ORF2a
3	1070^	C	1078	T	ORF2a
4	1661+	G	1669	C	ORF2a#
5	1704+	G	1712	C	ORF2a#
6	1907^	T	1915	C	ORF2a & ORF2b#
7	1974^	G	1982	A	ORF2a# & ORF2b#
8	2887^	C	2895	T	ORF2b
9	3278^	C	3286	T	ORF2b# & ORF4
10	3494^	C	3502	T	ORF2b# & ORF4

Polymorphism 1 is a deletion in the the UTR between ORF1 and ORF2a. The remaining polymorphisms are point substitutions. The substitutions are either transitions (polymorphisms 2, 3, and 6 - 10) or transversions (polymorphisms 4 and 5). Polymorphisms 4 and 5 are synonymous changes that do not alter the amino acid sequence. The remaining six substitutions are all non-synonymous, because they change an amino acid in the translated protein sequence (Table 5.3).

Table 5.3. Amino acid sequence substitutions between VTMoV isolates R17 and K1. Polymorphism numbers are the same as in Table 5.2. Amino acid substitutions defined as non-conservative are marked in **bold text.**

Polymorphism	Coding region		aa changes	
	Protein	aa position	K1	R17
4	**P2a**	**367**	**C**	**W**
5		**382**	**R**	**G**
7		472	I	V
6	**P2b**	**17**	**P**	**S**
7		**39**	**D**	**G**
9		**474**	**S**	**P**
10		**546**	**W**	**R**

With ORF2a and ORF2b overlapping, polymorphism 7 causes changes in the amino acid sequence of two proteins (P2a and P2b). The change in P2a was regarded as a conservative substitution because both variants encode amino acids with side chains of a similar chemical nature (Mathews et al. 2000), i.e. both have aliphatic side chains (I and V) (Table 5.3). The change caused by polymorphism 7 in P2b was classified as non-conservative because the alternative amino acids were in different amino acid groups, due to the different chemistry of their side chains. The Aspartic acid (D) is acidic while the Glycine (G) is in the aliphatic group. The remaining non-synonymous substitutions were also classified as non-conservative (Table 5.3 bold text) because the different amino acids were in different groups. Refer to Appendix 10 for a list of the amino acid groupings.

Although most of the substitutions are non-synonymous and non-conservative, none are in functional regions annotated in Chapter 4. The only substitution that might be involved in a functional change is substitution 9 (P2b) which is adjacent to a potential dipeptide cut site E/S (see Table 4.2).

5.4 Discussion

The R17 isolate of VTMoV produces symptoms that are more severe than isolate K1. To identify sequence variation that could be contributing to this symptom variation, the genome sequence of isolate R17 was sequenced using RT-PCR with the K1 specific primers. Four overlapping PCR fragments were obtained that covered the complete genomic coding region, but lacked parts of both the 5' and 3' UTRs. The virusoid was also sequenced and was found to

be identical to that of VTMoV-K1, but different from that previously sequenced by Haseloff and Symons (1982).

VTMoV-R17 had the same genome organisation as K1; although isolate R17 has ORF1 in a different translation frame to ORF2a and ORF4 due to the deletion of a nucleotide in the R17 sequence (polymorphism 1). Despite this, the start codons for the R17 ORFs were predicted to be in the same favourable context of translation as described for isolate K1 (Chapter 3).

When the nucleotide sequences of R17 and K1 were compared only 10 polymorphisms were observed. This high similarity between the two isolates could be explained by the origins of the K1 isolate. K1 was discovered after single mirid transmission of isolate R17 in *N. clevelandii*. This mild symptom phenotype was stable even after several rounds of mechanical transmission in *N. clevelandii* plants (Francki et al. 1986a).

There were six polymorphisms (4, 5, 6, 7, 9, and 10, Table 5.3) that altered the VTMoV protein sequences, but none occurred in regions that encode annotated functional protein regions (Chapter 4). It may be that that these amino acid changes are occurring in regions of as yet unknown function, or are in proteins that are associated with symptom variation. In RYMV virulence has been associated with non-synonymous changes in amino acid sequence in a central region of 15 amino acids in the VPg; 80% of these polymorphisms occur at amino acid position 48 of the VPg (Pinel-Galzi et al. 2007). None of these changes in amino acid were found in avirulent RYMV isolates (Pinel-Galzi et al. 2007). In VTMoV polymorphisms 4 and 5 are adjacent to the ED rich region of the VPg motif (Table 4.2).

Alternatively, specific protein polymorphisms may not be directly causing the different symptoms. In RYMV, there has been correlation between symptom expression and virus titre (N'Guessan et al. 2000) and in LTSV virusoid absence has reduced symptom intensity (Jones et al. 1983). The final determination of which, if any VTMoV polymorphism is associated with the symptom phenotypes will require a systematic analysis using reverse genetics of each polymorphism, individually and in combination. Once the polymorphism has been identified the mechanism of how the polymorphism is causing the symptom phenotype will also need to be investigated.

- Chapter 6 -
Sequence variation in the RdRp gene of various isolates of VTMoV

6.1 Introduction

This chapter reports the results of experiments targeted at describing the sequence variation present in the RdRp from VTMoV isolates which had been collected at different times and locations. The RdRp gene was selected as this is the most conserved region of the sobemovirus genome (Fargette et al. 2004). It was considered that analysis of this gene would provide an estimate of the degree of sequence variation present among the VTMoV isolates.

6.2 Materials and Methods

6.2.1 Collection of sequences from isolates

Total nucleic acid extraction

Total nucleic acid was extracted from 5-10 mg of dried infected leaf material as described in section 2.3.5. The VTMoV infected leaf samples are listed in Table 6.1.

Table 6.1 List of Isolates analysed for variation in the RdRp gene. Included are isolate WINC number, details of host species, collection details, collection date and GenBank accession number assigned to each sequence. Additional information on each WINC isolate can be found in Appendix 2. The '--' indicates where specific day is unknown.

WINC number	Host species	Collection details		GenBank accession number
		Specimen name and location	Date	
65754	*N. velutina*	V2 Strzelecki track	--/07/1979	HQ680390
65757		VTMoV (NVV2) # 18 Strzelecki track	--/07/1979	HQ680393
65758		Cunnamulla, QLD	23/08/1982	HQ680395
65771	*N. clevelandii*	VTMoV B1	17/10/1984	HQ680391
65779		Yaningurie water hole	24/11/2000	HQ680386
65780		Cobblers sand hill	24/11/2000	HQ680387
65781		VTMoV-K1	--/02/2004	HM754263
65782		VTMoV-F	--/02/2004	HQ680389
65783		VTMoV-R17	--/02/2004	HQ680399
65795		VTMoV 13 Strz	02/11/1984	HQ680385
65796		VTMoV 14 Strz	02/11/1984	HQ680394
65797		VTMoV 15 Strz	02/11/1984	HQ680396
65798		VTMoV 16 Strz	02/11/1984	HQ680384
65799		VTMoV 17 Strz	02/11/1984	HQ680388
65800		VTMoV 18 Strz	02/11/1984	HQ680392

RT-PCR

First-strand cDNA was synthesised with primer VTMoV- 2bR2 as described in section 2.3.8, and PCR used the primer pair VTMoV-2bF2 and VTMoV-2bR2 to amplify the RdRp gene (1924 nt fragment) from the first-strand cDNA (refer to Appendix 4 for detailed primer information). PCR cycling conditions were denaturation for 3 min at 94 ºC; 30 cycles of 30 s at 94 ºC, 30 s at 48 ºC, and 1 min at 65 ºC, with a final extension step of 10 min at 65 ºC.

Sequencing

Amplicons (1924 nt) were purified as described in section 2.3.11 and cloned as outlined in section 2.3.12. Sequencing reactions contained the universal primers M13F or M13R as in section 2.3.13, or a third primer VTMoV-Re. This third primer enabled sequencing of the middle region of the cloned PCR fragment, which was not often sequenced effectively using the

universal primers. The three primers together enabled complete coverage of the PCR amplicon with good quality sequence.

6.2.2 Analysis of RdRp sequence

Subdivision of RdRp sequence

For some analyses the RdRp sequence was separated into 3 regions. Region 1 included nucleotide positions 1 to 419; where there is overlap between the RdRp gene and ORF2a (genome map positions 1867 to 2287 in Figure 3.6). Region 2 starts at nucleotide position 420 and stops at nucleotide 1396, (genome map positions 2288 to 3262 in Figure 3.6). Region 3, from nucleotides 1397 to the end nucleotide position 1647, includes the end of the RdRp gene and overlaps with ORF4 (genome map positions 3263 to 3513 in Figure 3.6).

Sequence polymorphisms

Sequences were aligned using Clustal X2 (Larkin et al. 2007) and polymorphisms were identified using both BioEdit (Hall 1999) and MEGA (Kumar et al. 2008). A map showing the distribution of these polymorphisms was created using the Hypermut program (Rose & Korber 2000) on the HIV database (www.hiv.lanl.gov). Nucleotide polymorphisms were also sorted according to their polymorphism type, as transitions, transversions, or multi variant sites. Multi variant sites were those sites where more than two different bases were observed.

6.2.3 Estimation of sequence error rate

RT-PCR can introduce sequence errors into individual DNA molecules, which are manifested when sequence is obtained from isolated clones of the amplicon fragment. Multiple clones can be sequenced to detect these errors, which appear as rare sequence variations among the clones (Smith et al. 1997). Due to time and budget constraints, this process was not possible for all of the sequences analysed. To partially compensate, sequence from multiple clones was obtained from a single random VTMoV isolate (65796). These sequences were compared to that of isolate VTMoV-K1 (65781), which had its complete genome sequenced (Chapter 3). This analysis was used to estimate the error rate in the sequence obtained for the other isolates, which was derived by sequencing only a single clone per isolate.

Categorisation of polymorphisms

To identify polymorphisms that may represent sequencing errors the polymorphisms were classified into two types - shared or singleton. When base variants at a polymorphic site were seen in two or more isolates (across all 15 isolates), the polymorphism was defined as shared. These shared mutations were considered more reliable than those that were detected in only one isolate, which were classified as singleton polymorphisms. For the purposes of the analysis, bases unique to K1 (65781), 65796 or R17 (65783) were regarded as reliable and were classified as shared polymorphisms. Singletons were analysed (separately) to gain additional information, although the conclusions gained from the singletons could only be regarded as suggestive, due to the possibility of sequencing errors.

Polymorphism frequency

Polymorphism frequencies were calculated by the ratio of; the number of polymorphisms: total number of nucleotides in the region of analysis, providing a per nucleotide rate value which was comparable between regions. For amino acid polymorphism analysis, amino acid values were substituted and provided a per amino acid rate which was also comparable between different regions.

6.2.4 Detection of Recombination

Sequences were analysed for recombination using the RDP3 analysis suite (Martin et al. 2005b), using default settings and a Bonferroni corrected *P*-value of 0.01. RDP3 identifies putative recombination events, break points, and recombinant parental isolates using the analysis methods; RDP (Martin & Rybicki 2000), GENECOV (Padidam et al. 1999), Chimaera (Posada & Crandall 2001), MaxChi (Smith 1992), BOOTSCAN (Martin et al. 2005a), SISCAN (Gibbs et al. 2000), 3Seq (Boni et al. 2007), Phylpro (Weiller 1998), and LARD (Holmes et al. 1999). Due to the possibility that single analyses might produce false positives, only those recombination events detected by more than 3 methods were reported. Additional recombination analysis using GARD (Kosakovsky Pond et al. 2006b) is described in section 6.2.5.

6.2.5 Assessing evolutionary mechanisms

The MEGA program (Kumar et al. 2008) was used to calculate the mean pairwise genetic distance (d) corrected using the Kimura 2 parameter (Kimura 1980); and the mean pairwise synonymous (d_S) and non-synonymous (d_N) substitution rates using the Pamilo-Bianchi-Li method (Li 1993 ; Pamilo & Bianchi 1993). Standard errors for these parameters were determined using 1000 bootstrap replicates.

The sequences were analysed using Datamonkey[(Delport et al. 2010) http://www.datamonkey.org] which utilises multiple phylogenetic tools to find evidence of selection. The HKY85 model was used for all analyses. Data sets were initially analysed using GARD to detect recombination (Kosakovsky-Pond et al. 2006b). GARD allowed inclusion of any recombinant sequences in subsequent analysis of selection pressures. Selection pressures for each codon was determined using an Integrative selection analyses which the three methods SLAC, FEL and REL to identify what codons under selective pressure (Kosakovsky-Pond & Frost 2005). Codons were considered to be under selection if 2 or 3 methods gave the same results. Sites under differing selection pressures and at the population level (between populations) were determined using IFEL (Kosakovsky-Pond et al. 2006a) and sites under diversifying selection were determined using the PARRIS method (Scheffler et al. 2006). Toggle (Delport et al. 2008) was used to detect positions which may have toggled between multiple amino acid codons. The phenomenon of toggling applies to the situation where an amino acid has changed due to the presence of selective pressure, and then changed again due to change in selective pressure.

6.2.6 Determining relationships between VTMoV isolates

The RdRp sequences (using TRoV RdRp as an outgroup and excluding any identified VTMoV RdRp recombinants) were aligned using Clustal X2 (Larkin et al. 2007). A maximum parsimony analysis was performed using MEGA (Kumar et al. 2008) with 10000 bootstrap replicates. The DnaSP v5 program (Librado & Rozas 2009) was used to calculate nucleotide diversity (π) as outlined by Nei (1987).

6.3 Results

6.3.1 Estimation of error rate

A total of 52 nucleotide positions showed differences across the four sequenced clones of the VTMoV-65796 RdRp amplicon. A base that was unique to a single clone was regarded as a sequencing error. The error rate can be defined as the proportion of differences detected between a single clone and the K1 sequence which are not genuine. For isolate 65796, 8 nucleotide sites were errors, indicating a sequencing error rate of 2 nucleotide positions per clone. Across the whole RdRp gene (1647 nt) this would produce sequencing error with a frequency of 0.001214.

Distribution of sequence polymorphisms

Figure 6.1 shows where the RdRp sequence of the 14 isolates differed from that of VTMoV-K1 (65781).

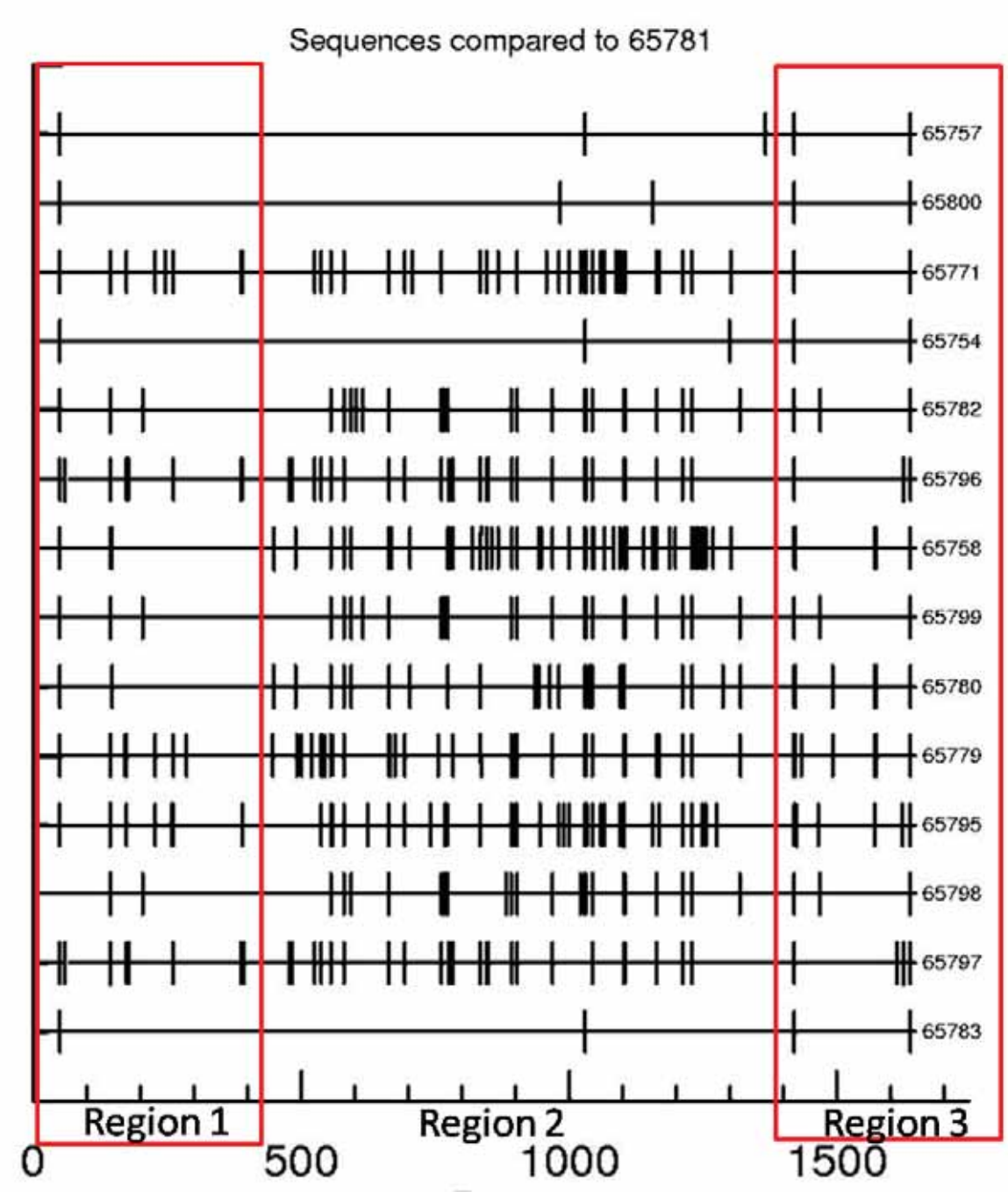


Figure 6.1. Nucleotide polymorphisms in 14 VTMoV isolates (represented by vertical lines) relative to VTMoV-K1 (65781). Numbers on right hand side refer to WINC accessions in Table 6.1. The comparison includes 1647nt of the RdRp gene (ORF2b), and shows regions of overlap with other ORFs, boxed in red. Region 1 marked by the left box shows the 1-419 nt overlap between ORF2b and ORF2a. Region 2 is in the middle of the gene and region 3 is demarcated by the right box showing the overlap of 1397-1647nt between ORF2b and ORF4. The scale bar refers to nucleotide positions of the RdRp starting at genome map position1867 and stopping at position 3513 (Refer to Figure 3.6).

Figure 6.1 shows there was a higher frequency of nucleotide sequence differences relative to K1 (singleton and shared polymorphisms combined) in region 2 (420 to 1396 nt) which is the interval without overlap with other genes.

6.3.2 Polymorphism classification

Types of mutation

Types of mutations observed at polymorphic sites are shown in Table 6.2.

Table 6.2. Occurrence of both transition mutations between purines (A and G) or pyrimidines (C and T) and transversion mutations (between purines and pyrimidines) across the RdRp gene. 'Multiple variant sites' are those where more than 2 different bases were recorded and where both transition and transversion mutations were observed.

Mutation types	Region 1 ORF2a/2b (1-419 nt)			Region 2 ORF2b only (420-1396 nt)			Region 3 ORF2b/CP (1397-1647 nt)		
	Singleton	Shared	Total	Singleton	Shared	Total	Singleton	Shared	Total
Transitions									
A-G	4	8	12	16	13	29	1	2	3
C-T	4	9	13	29	40	69	4	5	9
Transversions									
A-T	0	0	0	5	2	7	0	0	0
G-C	0	0	0	1	0	1	0	0	0
G-T	0	0	0	2	0	2	0	0	0
C-A	0	0	0	0	0	0	0	0	0
Multiple variants	0	0	0	0	0	4	0	0	0

Most polymorphic sites contained transition-type mutations. Where the RdRp gene overlaps with other ORFs (regions 1 and 3), A-G and C-T transitions were in equal frequency, whereas, within region 2, C-T type transitions were more than twice that of the A-G type. Transversions were less common, found only in region 2 and were dominated by the A-T type.

The distributions of singleton and shared polymorphism by the gene region are shown in Table 6.3.

Table 6.3 Distribution of singleton and shared polymorphisms across three defined regions of the RdRp gene. Frequencies of each polymorphism type are presented as rates per nucleotide. The 4 multiple variant sites from Table 6.2 were excluded from this analysis.

Genomic Region	Polymorphism sites					
	Singletons		Shared		Combined	
	#	Frequency	#	Frequency	#	Frequency
Region 1: ORF2a/2b (1-419 nt)	8	0.0191	17	0.0405	25	0.0596
Region 2: ORF2b only (420-1396 nt)	53	0.0542	55	0.0563	108	0.1105
Region 3: ORF2b/CP overlap (1397-1647 nt)	5	0.0199	7	0.0279	12	0.0518
Overall (1647nt)	66	0.0400	79	0.0479	134	0.0814

When polymorphisms were considered, as either combined singleton and shared, or shared only they occurred more frequently in region 2 (0.1105 and 0.0563 respectively). Region 3 had the lowest combined frequency, and region 1 had the lowest frequency of singleton polymorphisms (0.0191).

Most of the 138 polymorphic positions contained synonymous substitutions (i.e. they did not change the amino acid sequence of the RdRp). Only 28.2 % (39/138) were non-synonymous, leading to changes and polymorphisms in the amino acid sequence of the RdRp protein (Table 6.4).

Table 6.4 Distribution and frequency of singleton and shared polymorphisms that result in amino acid changes in the RdRp protein. Frequencies of each polymorphism type are presented as rates per amino acid.

Genomic Region	Polymorphism sites					
	Singletons		Shared		Combined	
	#	Frequency	#	Frequency	#	Frequency
Region 1: ORF2a/2b overlap (1-140 aa)	4	0.0285	8	0.0571	12	0.0857
Region 2: ORF 2b only (141-464 aa)	12	0.0370	4	0.0123	16	0.0494
Region 3: ORF2b/CP overlap (465- 549 aa)	3	0.0350	8	0.0941	11	0.1294
Overall (549 aa)	19	0.0346	20	0.0364	39	0.0710

Over the RdRp gene the frequency of singleton and shared amino acid polymorphisms was 0.0346 and 0.0364 respectively, and both types of polymorphism contributed to changes in amino acid sequence. Singleton polymorphisms in amino acid sequence were distributed across the three regions in similar amount, but shared polymorphisms were highest in region 3

(0.0941) and lowest in region 2 (0.0123). Over the whole gene more polymorphisms of either type, resulted in a higher frequency in regions 1 and 3 compared to region 2. As well as amino acid changes in the RdRp protein, the amino acid changes encoded in the overlapping ORFs are shown in Table 6.5

Table 6.5. Distribution and frequency of singleton and shared polymorphisms that result in amino acid changes in the regions overlapping with the P2a and the CP ORFs. Frequencies of each polymorphism type are presented as rates per amino acid.

Genomic Region	Polymorphism sites					
	Singletons		Shared		Combined	
	#	Frequency	#	Frequency	#	Frequency
Region 1: P2a overlap (432- 572 aa)	5	0.0354	12	0.0851	17	0.0939
Region 3: CP overlap (1- 83 aa)	1	0.0120	0	0	1	0.0120

In region 1, the P2a amino acid sequence had a higher frequency of shared polymorphisms (0.0851) compared to the RdRp amino acid sequence in the same region (0.0571). In region 3, there was a higher frequency of shared polymorphisms in the RdRp amino acid sequence (0.0941, Table 6.4) compared to 0 in the CP part of the amino acid sequence.

6.3.3 Assessing mechanisms of evolution

Detection of Recombination

The RDP3 program detected recombination in the RdRp sequences (Table 6.6).

Table 6.6. Analysis of recombination in the RdRp sequences of VTMoV.

Recombinant break points by nucleotide positions	Parental isolates*			Recombinant detecting programs (P values)
	Recombinant	Major	Minor	
797- 1362	65758	65780	Unknown	MaxChi- 2.89 X10^{-5} Chimaera 2.58 X10^{-6} Siscan 2.89 X 10^{-5} 3Seq 6.51 X10^{-4}
897-1623	65796	Unknown	65782	MaxChi 1.323 X10^{-3} Chimaera 1.99 X10^{-3} SiScan 6.23 X10 $^{-6}$
897-1623	65797	Unknown	65782	MaxChi 1.323 X10^{-3} Chimaera 1.99 X10^{-3} SiScan 6.23 X10 $^{-6}$

*Parental isolates were sequences used in the analysis to identify recombination, and are classified as either as 'major' or 'minor'. The minor parent is the source of the sequence from within the break points (the recombinant break point interval) and the major parent is the source of the remaining sequence. When a WINC isolate number is provided the parent sequence was detected in the dataset. If the parental sequence was not found in the dataset the parent is labelled as unknown and a substitute sequence from the dataset was used for analysis.

Three different recombination events were detected in the VTMoV RdRp data set. Isolate 65758 from Queensland was the product of one event between nucleotide positions 797 and 1362 (interval size: 565 nt). Two different isolates 65796 and 65797 from the Strzelecki track were the result of two other recombination events in the same general region, between nucleotide positions 897 and 1623 (interval size: 726 nt).

Detection of sequence divergence and selective pressures

To determine the nucleotide diversity and to detect selective pressures acting upon the RdRp gene, the 15 VTMoV nucleotide sequences were analysed by region. The mean nucleotide diversity (d) calculation was modified with the Kimura 2 parameter (Kimura 1980) (refer to MEGA software (Kumar et al. 2008) for detailed equations. The mean rates of both synonymous (d_S) and non synonymous (d_N) substitutions per site, and the ratios (ω) between these two values were determined and are shown in Table 6.7. However, because calculation of ω can be sensitive to recombination events (Anisimova et al. 2003), recombinant VTMoV isolates were excluded from the divergence, diversity and ω calculations.

Table 6.7. Nucleotide diversity across the three regions of the RdRp sequences. Recombinant isolates 65758, 65796, 65797 are not included.

Genomic region (nucleotides)	Nucleotide diversity			
	d*	d_N ^	d_S #	ω d_N / d_S
Region 1: ORF2a/2b (1-419)	0.0125 ± 0.0031	0.0047 ± 0.0019	0.0280 ± 0.0090	0.1679
Region 2: ORF2b only (420-1396)	0.2630 ± 0.0031	0.0327 ± 0.0043	0.0097 ± 0.0031	3.37
Region 3:ORF2b/CP (1397-1647)	0.0108 ± 0.0035	0.0153 ± 0.0054	0.0024 ± 0.0020	6.375
Overall (1647)	0.0203 ± 0.0020	0.0067 ± 0.0013	0.0491 ± 0.0060	0.1365

* Units are the number of nucleotide substitutions per nucleotide site between 2 sequences. ^ Units are the number of non-synonymous substitutions per non-synonymous site. # is the number of synonymous substitutions per the total number of synonymous sites.

The d value determines the number of base substitutions per site per pairwise sequence, and was averaged across all comparisons. This value gives an indication of how diverse the analysed sequences were; for regions 1 and 3 the d value was less than one twentieth that of region 2.

ω was highest in region 3 at 6.375 and then region 2 at 3.37. These two values were above 1 indicating that diversifying selection had occurred. Region 1 showed a value of 0.1679 indicating that purifying selection had occurred. The recombinant sequences were not included in these analyses, but inclusion of these isolates in additional analyses did not noticeably alter the results (i.e. values of d and ω were slightly higher when recombinant sequences were included (Refer to Appendix 11).

Specific sites under selection

No sites were detected to be under diversifying selection using the PARRIS program. Neither were any sites determined to have toggled between amino acids (TOGGLE program). A total of 63 sites were identified by integrative selection analysis (SLAC, REL, and FEL) to be under purifying selection. Furthermore, 40 amino acid sites were determined to be under purifying selection at the population level, that is, these sites were under purifying selection by at least some of the sequences within the whole 15 sequence set. Sites under purifying selection are displayed on the RdRp amino acid sequence in Figure 6.2.

Figure 6.2. VTMoV RdRp sequence showing sites under selective pressure. Sites with X are where more than one amino acid is coded for. Sites under purifying selection as assessed from integrative selection (SLAC, FEL and REL analysis) are boxed. Those sites determined by IFEL to be under differential selection at the population level are in bold and a larger text size**.** Functional motifs from Table 4.2 are indicated by colours: blue for Motifs I-VII, red for the GDD motif and yellow for the dipeptide cut sites. Regions of overlap with other ORFs are underlined. Position 300 is shown in green.

Sites under purifying selection are distributed across the whole RdRp sequence and include sites within annotated protein features (Table 4.2). There were 40 sites considered to be under purifying selection at the population level; 39 were included in the 63 sites which were classified as being under purifying selection by integrative selection analysis, except for site 300. This site was not classified as being under purifying selection, but at least some of the sequences within the whole 15 sequence population had site 300 under purifying selection, which is why it was considered to be under purifying selection at the population level.

6.3.4 Relationships between VTMoV isolates

Figure 6.3 shows the relationships among the 12 non-recombinant VTMoV isolates based on the RdRp gene sequences.

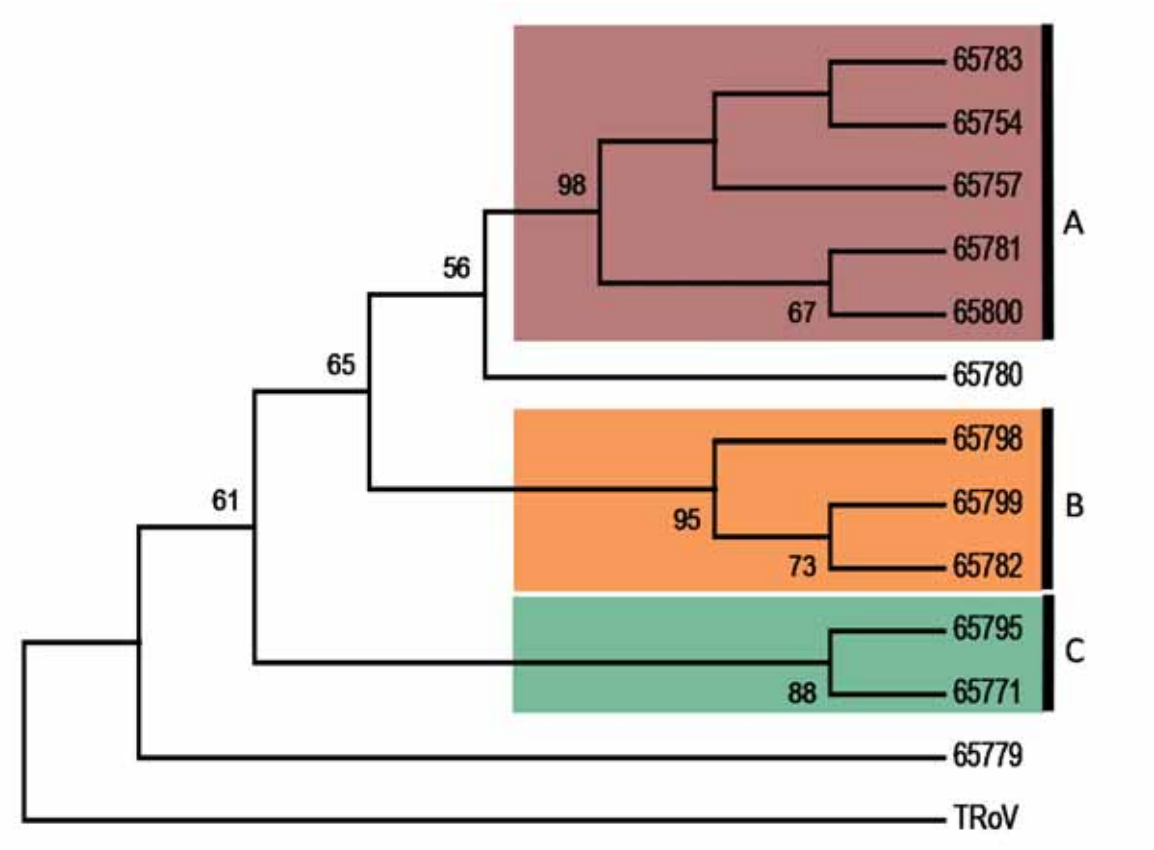


Figure 6.3. Phylogenetic tree based on the RdRp nucleotide sequence of 12 VTMoV isolates, with TRoV as an outgroup. A maximum parsimony analysis was undertaken with bootstrap replication of 10 000. Numbers show % bootstrap support values greater than 50 %.

Groups A, B and C are discriminated by greater than 80% bootstrap support, but these groupings were unrelated to either geographical (location of collection) or temporal factors (date of collection). Isolates which were labelled as originating from the same source were not always grouped together. Isolates 65757 and 65800 are both labelled as VTMoV Strz 18 (the same original source) and both lie in Group A, but are not on the same branches within the group. Isolates 65783 and 65799 (both labelled as VTMoV Strz 17) lie in an entirely different group. Table 6.8 shows values for each of the VTMoV groups for d, ω, and π. π is a measure of polymorphism within a population, calculated by the average number of nucleotide differences per site between any two DNA sequences (Nei 1987).

Table 6.8. Nucleotide diversity within different VTMoV population groups in the RdRp gene.

Group	Nucleotide diversity				
	d*	d_N ^	d_S #	ω (d_N / d_S)	π +
Group A	0.0033 ± 0.0009	0.0030 ± 0.0010	0.0048 ± 0.0023	0.625	0.01840
Group B	0.0024 ± 0.0010	0.0016 ± 0.0009	0.0052 ± 0.0032	0.307	0.00243
Group C	0.0236 ± 0.0043	0.0047 ± 0.0021	0.0627 ± 0.0111	0.075	0.02307
Overall					0.02049

* Units are the number of nucleotide substitutions per nucleotide site between 2 sequences. ^ Units are the number of non-synonymous substitutions per non-synonymous site. # is the number of synonymous substitutions per the total number of synonymous sites. + units are the number of nucleotide differences between any two DNA sequences.

Group C has the highest pairwise genetic distance (d) and nucleotide diversity (π), but has the least number of isolates within it, and has the lowest ω. Group A contains the most isolates (5) and has largest ω, 0.625 [which is 4.5 times greater than overall ω 0.1365 (Table 6.7)]. Group A also has a low d value but a π (nucleotide diversity) value very similar to the π value for all 12 VTMoV sequences.

6.4 Discussion

RdRp gene sequences were determined for 15 isolates of VTMoV. This provided a snapshot of the type of genetic variation present in VTMoV, and where forces of evolution may have generated this variation.

6.4.1 VTMoV polymorphisms

Error rate

The calculated error rate of 0.0012 per nucleotide, or about 2 nucleotides per sequence fragment were regarded as artefacts of the RT-PCR process (Smith et al. 1997). However, of these singleton polymorphisms some may have been genuine polymorphisms. For this reason, the singleton polymorphisms were kept and analysed separately.

Nucleotide polymorphisms

Polymorphisms were distributed across the whole RdRp gene and mostly involved transitions. Transversions were confined to the central region of the gene (region 2). Due to the different sizes of each RdRp region, polymorphism frequencies were calculated to compensate for this

disparity. When polymorphism frequencies were segregated into shared and singletons, there was a clear concentration of shared polymorphisms in region 2. This central region of the gene has no overlap with neighbouring coding sequences (Figure 6.1). In contrast regions 1 and 3 contribute to two protein sequences and could be expected to be restricted in genetic variation to maintain the function of the proteins.

6.4.2 Variation across amino acid sequences

Approximately one quarter of all polymorphisms caused a change in amino acid sequence. Although region 2 had the highest frequency of shared polymorphisms, it had the lowest frequency of polymorphisms that changed amino acid sequence. Shared amino acid polymorphisms were concentrated in regions 1 and 3. This disparity in variation across the RdRp regions was also observed in the calculations of ω. Regions 1 and 2 had ω values greater than one, indicating diversifying selection, but region 3 was constrained and had *a* ω indicative of purifying selection. This difference in polymorphism and selective pressures within the RdRp may be due to a variation in protein domain functional constraints; however, further work on VTMoV RdRp function and protein structure would be required to confirm this.

6.4.3 Evolutionary selective pressures

Overall the VTMoV RdRp had a $\omega = 0.1365$, which indicated that the VTMoV RdRp gene was under purifying selection. These results were supported by the Datamonkey analyses which showed no codons under diversifying selection and a total of 63 codons under purifying selection across the gene.

The ω value for the VTMoV RdRp (0.1365) is higher than that calculated from another sobemovirus (RYMV RdRp, $\omega = 0.090$). The latter was calculated using the non-overlapping region of ORF2b from 16 sequences collected across all Africa, and compared to the values of the ORFs (Fargette et al. 2004). This was taken as evidence that in the RYMV genome, the RdRp is the most constrained ORF (Pinel-Galzi et al. 2009). Whether ORF2b is the most constrained ORF for VTMoV will have to be determined from further experiments with other ORFs of VTMoV. Although partial amino acid sequence from other proteins has been collected, the ω for these regions is unlikely to be accurate as variation in ω was observed in different regions of the RdRp. In order to obtain an accurate calculation of ω for other ORFs, the complete ORF sequences would need to be analysed.

6.4.4 Recombination

Recombination has only been detected in one other sobemovirus species, RYMV (Pinel-Galzi et al. 2009). In this study it was detected in three VTMoV isolates. In RYMV, two recombination events were detected in two parts of the genome; (i) the interval spanning the intergenic region between ORF1 and ORF2a and 5' end of ORF2a, and (ii) the 3' untranslated end of the genome. RYMV recombination events were limited to 5 isolates from Tanzania (Pinel-Galzi et al. 2009). Neither of the recombination intervals reported in the RdRp here match those observed in RYMV and this is the first detection of recombination in the RdRp from a sobemovirus species, but no other sequence diversity studies have analysed sequence variation in a sobemovirus species.

6.4.5 VTMoV isolate relationships

Variation in the RdRp gene was used investigate the relationships between the VTMoV isolates which did not show recombination. Within this small data set, three multiple isolate groups (A, B and C) and two single isolate groups were detected. These groupings did not seem to be related to temporal collection or geographical (collection location) factors (e.g. isolates 65783 and 65799 which are both named as isolate VTMoV-R17). Sequence will need to be collected from additional VTMoV isolates in the WINC (Appendix 2) and from new field collections to determine if any broader patterns or relationships of VTMoV isolates exist. This type of sequence collection could be limited to the 977 nt region 2 of the RdRp (from nucleotides 419 - 1396). This region was found to contain the greatest concentration of genetic variation and would provide a more cost effective means of conducting a large scale sequence diversity study.

- Chapter 7 -
Changes in VTMoV sequence associated with mechanical and mirid transmission

7.1 Introduction

This chapter reports an experiment where VTMoV isolates K1 and R17 were serially passaged by either mechanical inoculation or by mirid transmission for 24 months. The purpose of this experiment was to determine if changes in sequence could be associated with the different modes of transmission. ORFs P1 and CP were selected for comparison because both genes are thought to be involved in movement either within or between plants in other sobemovirus species (Meier et al. 2008). Sequences obtained at 24 months were compared with those of K1 and R17 described in previous chapters, which represent the sequences at the start of the experiment.

7.2 Materials and Methods

7.2.1 Mechanical transmission

Initial virus inoculum was prepared from dried leaf samples of VTMoV WINC isolates K1 (65781) and R17 (65783) crushed in tap water at a concentration of 1 mg ml^{-1}. This was used to inoculate P1 (passage one) plants as described in section 2.3.1. After one month, symptomatic systemically infected leaves were collected from all 6 plants of the first passage, and then dried and stored as outlined in section 2.3.2. Subsequent passages used dried leaf material from the previous passage as inoculum. Isolate K1 inoculum was prepared in water at 1 mg ml^{-1} and R17 at 0.5 mg ml^{-1}. Mechanical passages were maintained on separate benches in the same greenhouse at 21 - 24ºC without supplementary light. Mechanical passages will hereafter be referred to as: Passage Mechanical R17 or K1 and the passage number (#), for example PM_R# or PM_K#.

7.2.2 Mirid transmission

Mirids *(C. nicotianae)* were from a population maintained at the Waite campus, which had been established from insects collected on the Strzelecki track in South Australia and maintained on

N. clevelandii plants. Mirids were isolated by first placing infested plants in a closed box at 4ºC for one hour to reduce their activity. Individual adults and nymphs were collected by aspiration and transferred to a covered petri dish. To ensure that the mirids were not viruliferous at the commencement of the study, they were fed sequentially on 3 sets of detached healthy *N. clevelandii* leaves for 3 days each for a total of 10 days.

Mirid transmission experiment

The mirid transmission experiment started when the sets of 6 mechanically inoculated passage 1 plants were introduced to the separate BugDORM mesh cages (MegaView Science Co, Taiwan). For each passage sets of 6 plants were added to the cages and mirids were allowed to naturally transmit the virus by feeding. These cages were maintained in a glasshouse with no supplementary light, and limited temperature control using an evaporative cooler and shading during the summer months. The mirid transmission experiment with VTMoV K1 and R17 was done in parallel with the mechanical transmission experiment, with inoculations done at the same time every four weeks. Mirid passages will hereafter be referred to as: Passage Insect R17 or K1 and the passage number (#), for example, PI_R# or PI_K#. Figure 7.1 shows a plan for mirid transmission.

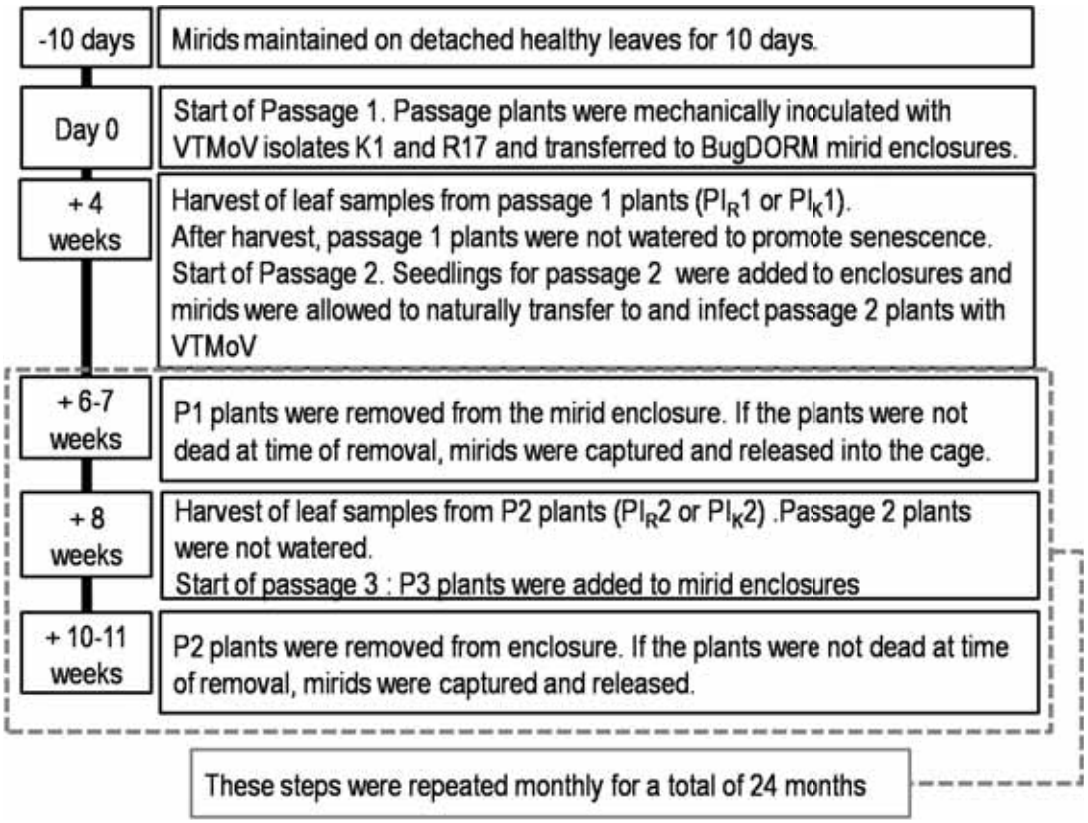


Figure 7.1. Strategy for long term mirid transmission experiment. Four weeks or 1 month was equivalent to one transmission passage.

7.2.3 Sample selection and collection

Each passage consisted of a set of 6 seedlings of *N. clevelandii*, and each seedling was inoculated with either VTMoV isolate K1 or R17, either by mechanical or mirid transmission. Each set was grown for four weeks (one month) before samples were taken and the next passage was started. A total of 96 leaf samples were collected and dried for storage, four samples (2 isolates and 2 modes of transmission) each month for 24 months.

7.2.4 Extraction of nucleic acids

Nucleic acids were extracted from dried leaf material as described in section 5.2.1.

7.2.5 RT-PCR

RT-PCR was used to amplify sequence from all samples collected in month 24. Two regions of the VTMoV genome were targeted, which cover approximately 48 % of the total genome sequence. Region 1 included nucleotides 8 to 990 towards the 5' end of the genome, and region 2 spans from nucleotides 3065 to 4115 as shown in Figure 7.2.

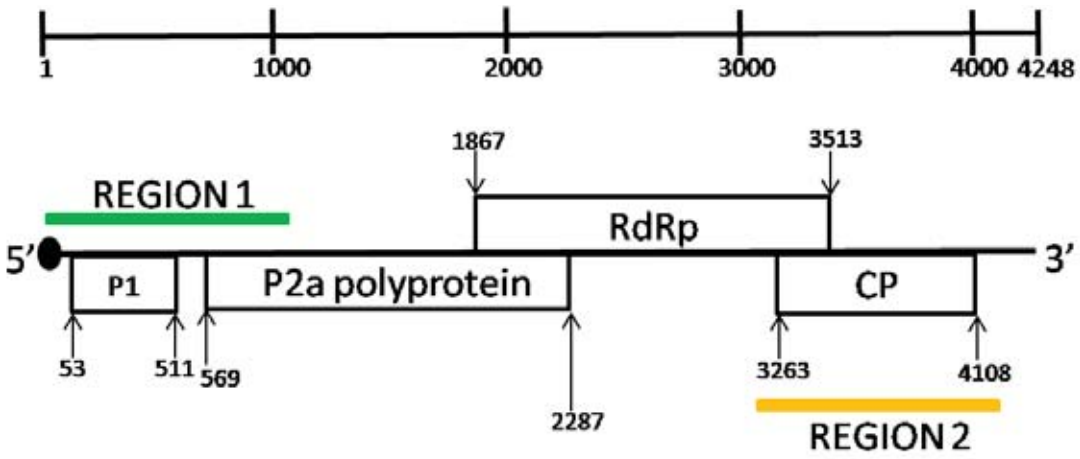


Figure 7. 2. The two regions of the VTMoV genome that were sequenced. Region 1 in green includes nucleotides 8-990. Region 2 in gold covers nucleotides 3560-4115. Each ORF is presented as a box with the genome annotation from Chapter 4, scale bar is in nucleotides.

RT-PCR

First-strand cDNA was synthesised as described in section 2.3.8, using primer VTMoV-2aR6 for region 1 and primer VTMoV-CPR4 for region 2.

PCR conditions

The PCR of region 1 used the primer combination VTMoV-5UTRF and VTMoV-2aR6 to amplify a 982 nt product. Specific cycling conditions for these primers were; denaturation at 94 °C for 3 min, and 30 cycles of 30 s at 94 °C, 30 s at 48 °C, 30 s at 72 °C, followed by an extension step of 7 min at 72 °C.

The PCR of region 2 used the primer pair VTMoV-CPF5 and VTMoV-CPR4 to amplify a 1050 nt product. Specific cycling conditions for these primers were; initial denaturation at 94 °C for 2 min, followed by 30 cycles of 15 s at 94 °C, 15 s at 54 °C, and 30 s at 72 °C, with an extension step of 7 min at 72 °C.

7.2.6 Cloning and sequencing

PCR products were purified as described in section 2.3.11, and cloned as described in section 2.3.12. Three clones were sequenced for each PCR product, except that for PI_R24 region 1 only 2 clones were obtained. For PM_R24 region 2, 6 clones were sequenced as outlined in section 2.3.13.

7.2.7 Sequence analysis

Sequences of the three clones were aligned and consensus sequences derived. All mutations were identified as either singleton or shared. Singleton mutations were those present in only one clone sequence and were considered likely to be an artefactual error (Smith et al. 1997). These singleton mutations are presented in the summary tables and figures to indicate their locations, but were disregarded from any further analysis. Shared mutations were those shared by the majority of clone sequences (2 of the 3) and were considered likely to be genuine mutations and therefore used in all mutation analysis. Clone consensus sequences were compared to genome sequences of K1 and R17 (Chapters 3 and 5 respectively). These genome sequences are referred to as P_R1 or P_K1 as they were identical for both modes of transmission.

Mutation analysis

Mutations were classified as outlined in section 6.2.2. Mutation frequency was calculated from the ratio of shared mutations: total nucleotides in the sequences. The mutation rate was calculated from the ratio of mutation frequency: number passages (or year).

Evolutionary selective pressure analysis

Sequence information from both modes of transmission was analysed as described in section 6.2.5.

7.2.8 Protein analysis

Mutations were analysed to determine if they caused changes in the amino acid sequence of the translation products. Amino acids were grouped according to the chemical nature of their side chains (Mathews et al. 2000) (refer to Appendix 10 for amino acid groupings). Changes in amino acids were then classified as either conservative if the substitution occurred between amino acids from the same group, or non-conservative if the changes substituted amino acids from different groups.

To determine if amino acid changes had an effect on protein structure, each amino acid mutation was analysed using ProtScale (Gasteiger et al. 2005) with default parameters. Hydrophobicity values were plotted according to Kyte and Doolittle (1982).

7.3 Results

7.3.1 Mechanical transmission

Figure 7.3 shows all mutations observed in isolates K1 and R17 after two years of mechanical transmission relative to the P_R1 and P_K1 sequences.

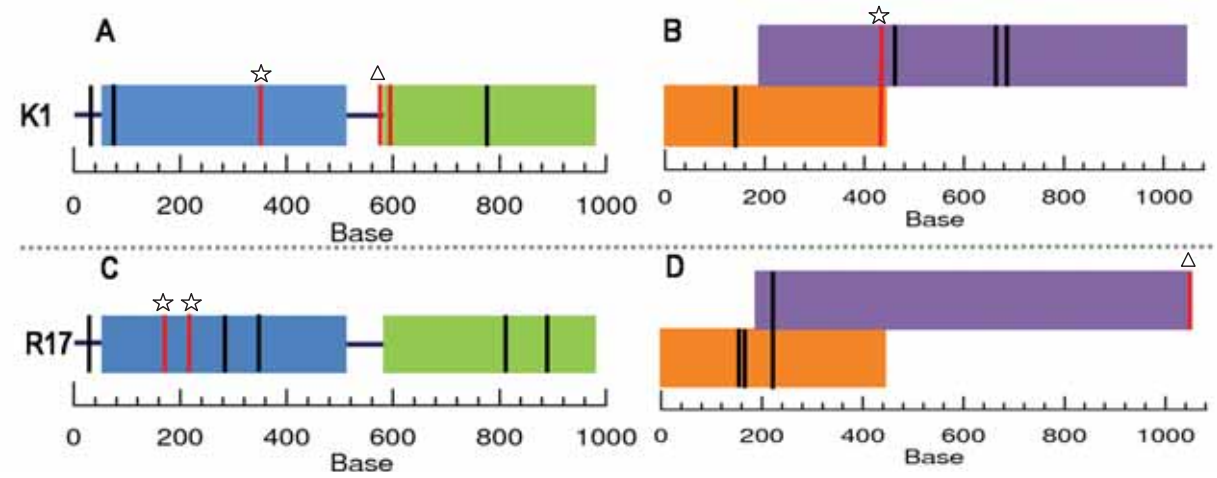


Figure 7.3. Distribution of mutations in isolates K1 (A & B) and R17 (C & D) after 24 serial monthly passages by mechanical transmission. A and C refer to region 1 (8-990 nt), covering the P1 protein represented by the blue box, part of the P2a protein represented by the green box and the untranslated region indicated by the horizontal line. B and D refer to region 2 (3065-4115 nt), the 3' end of the RdRp symbolized by the orange box, and the complete CP shown by the purple box. Mutations are indicated by vertical lines, black for singletons and red for shared mutations. ☆ denotes mutations causing amino acid substitutions, and Δ indicates a deletion mutation. Scale bars show the number of bases for each region.

Table 7.1 summarises the mutations shown in Figure 7.3, classified as either singleton or shared for each region.

Table 7.1. Mutations present in sequence collected from samples PM_R24 and PM_K24.

Transmission passage	Region 1 Mutations			Region 2 Mutations		
	Total	Singleton	Shared	Total	Singleton	Shared
K1 mechanical (PM_K24)	6	3	3	5	4	1
R17 mechanical (PM_R24)	7	5	2	4	3	1

The K1 sequence had 4 shared mutations, while isolate R17 had 3 across both regions. Region 1 had 5 shared mutations, which was more than region 2 which only had 2 shared mutations. Only shared mutations were classified by type in Table 7.2.

Table 7. 2. Types of shared mutations observed in sequence collected from samples PM_R24 and PM_K24. Those mutations marked with a * caused a change in the amino acid sequence.

Passage (Genomic region)	Transitions		Deletion
	A-G	T-C	
PM_K24 (Region 1)	1*	1	1
PM_K24 (Region 2)		1*	
PM_R24 (Region 1)	1*	1*	
PM_R24 (Region 2)			1

Transitions and deletion mutations were observed, but no transversion or multiple variant sites. The deletion mutation in PM_K24 region 1 sequence was identical to polymorphism 1 from Table 5.2 (Chapter 5), mapping to the UTR between ORF1 and ORF2a. The second deletion mutation in PM_R24 region 2 was in the 3'UTR at the end of region 2. Table 7.3 shows the mutations that caused changes in the amino acid sequence.

Table 7.3. Amino acid sequence changes caused by shared mutations in PM_R24 and PM_K24 . Amino acid changes are grouped by protein and indicated as either conservative with plain text or non-conservative by **bold text.**

Transmission passage	Region 1		Region 2	
	P1 (152 aa)	P2a (1-140 aa)	RdRp (401-549 aa)	CP (1-282 aa)
K1 mechanical (PM_K24)	V105I	-	**W546R**	-
R17 mechanical (PM_R24)	**L42F** H58R	-	-	-

Of the four shared mutations in the K1 sequence, two caused a change in amino acid sequence. One of these was a conservative substitution between Valine (V) and Isoleucine (I) in the P1 protein. Both of these amino acids have an aliphatic side chain of a similar chemical nature (Mathews et al. 2000). The second substitution in K1 was a non-conservative change between Tryptophan (W) an aromatic amino acid and Arginine (R) a basic amino acid.

For R17, both shared mutations caused a change in the amino acid sequence of P1. One was a conservative substitution between two basic amino acids Histidine (H) and R. The second was a non-conservative substitution between an aliphatic Leucine (L) and Phenylalanine (F) an amino acid with an aromatic side chain.

Hydrophobicity plots

When the amino acid substitutions were analysed for their effect on protein hydrophobicity the K1 mutation W546R altered hydrophobicity values for adjacent amino acids 542 to 544 (values for 545 to 549 not provided by analysis output). These changes altered the hydrophobicity plot of this protein (Figure 7.4 A and B).

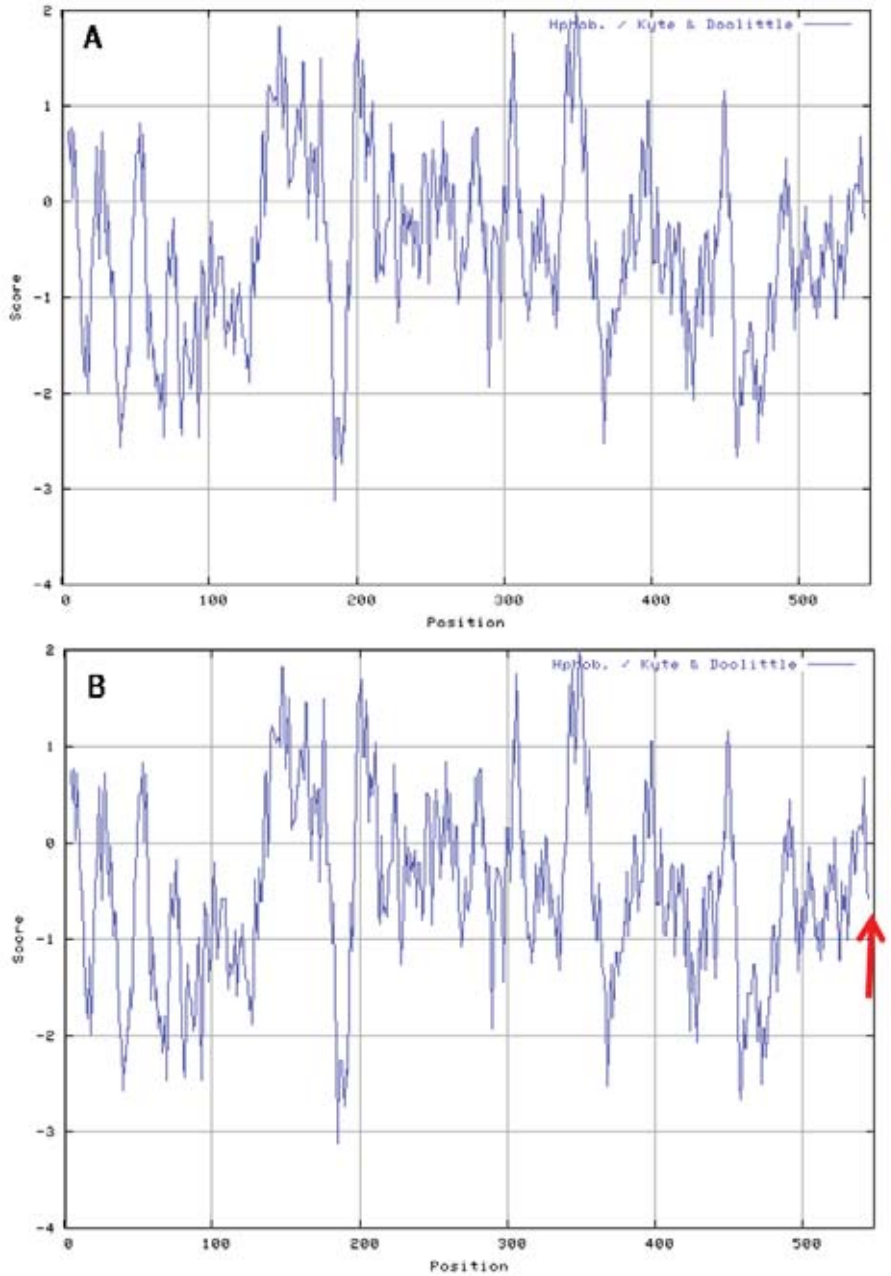


Figure 7.4.Hydrophobicity plots of the RdRp amino acid sequence from isolate K1 before (**A**) and after 24 months of serial mechanical transmission (PM_K24) (**B**). Plots were created using the amino acid hydrophobicity scores of Kyte and Doolittle (1982). Hydrophobicity scores are shown on the Y axis and the amino acid position in the RdRp on the X axis. The red arrows mark the site where the two plots differ.

In R17 the mutations L42F and H58R in the P1 protein were associated with alterations to the hydrophobicity values for adjacent amino acids 39 to 47 and 54 to 62 respectively. These altered values changed the hydrophobicity plot of the P1 protein (Figure 7.5 A and B).

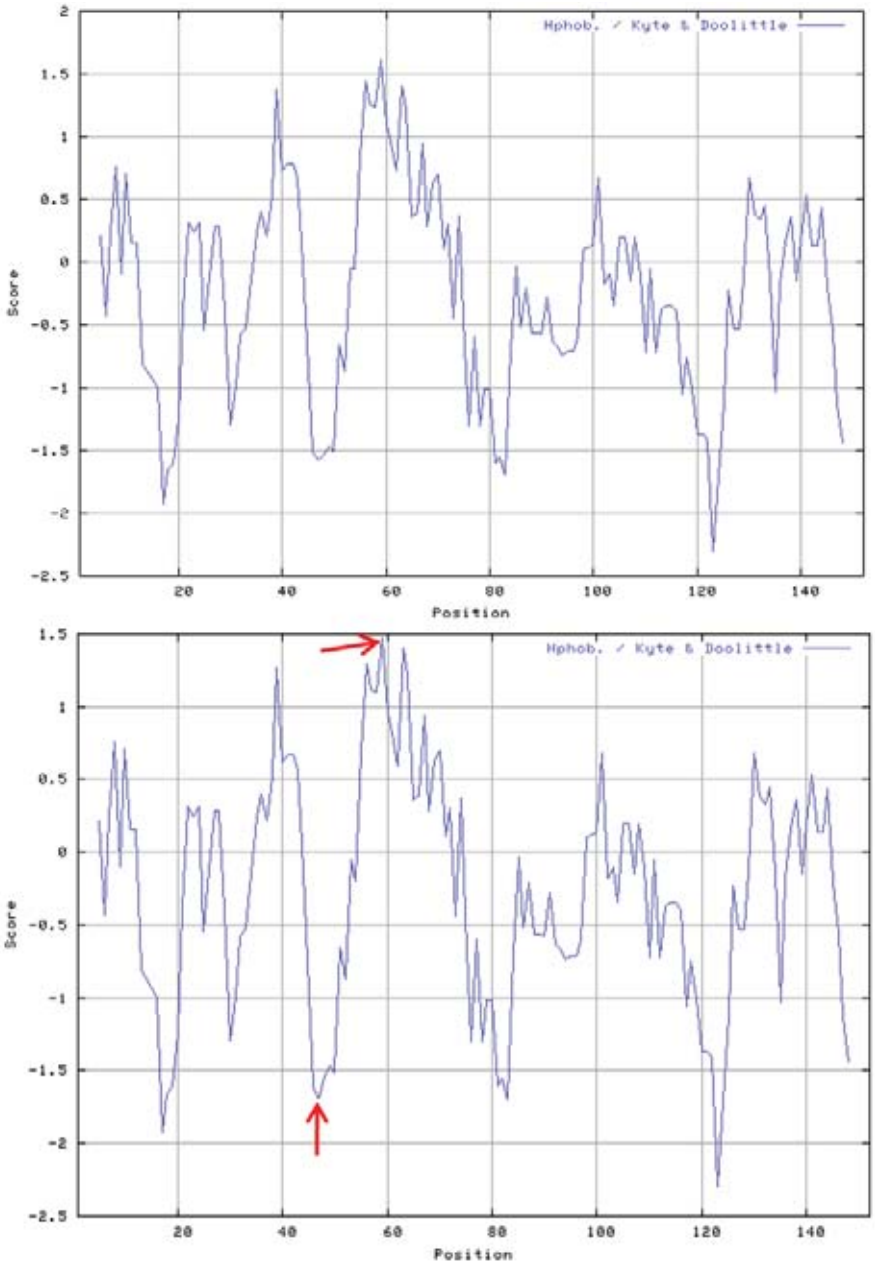


Figure 7.5.Hydrophobicity plots of the P1 amino acid sequence from isolate R17 before (**A**) and after 24 months of serial mechanical transmission (PM_R24) (**B**). Plots were created using the amino acid hydrophobicity scores of Kyte and Doolittle (1982). Hydrophobicity scores are shown on the Y axis and the amino acid position in the P1 on the X axis. The red arrows mark the sites where the plots differ.

Table 7.4 shows the frequency and rate of mutation.

Table 7. 4. Frequency and rate of shared mutations of each isolate following serial mechanical transmission over 24 months.

Isolate and genome region	Fragment size (nt)	Number of clones sequenced	Number of nucleotides sequenced	Shared Mutations	Mutation frequency per nt*	Rate of mutation per nt* /per year
PM_K24 Region 1	982	3	2946	3	10.18×10^{-4}	5.09×10^{-4}
PM_R24 Region 1	982	3	2946	2	6.78×10^{-4}	3.39×10^{-4}
PM_K24 Region 2	1050	3	3150	1	3.17×10^{-4}	1.59×10^{-4}
PM_R24 Region 2#	1050	6	6300	1	1.59×10^{-4}	0.79×10^{-4}

When mutations were observed in 2/6 clones they were classified as shared mutation. No sites in this sample had more than two nucleotides present, which might have indicated variation in the sample sequences. * units are the number of shared nucleotide mutations per number of nucleotides sequenced

In each region the frequency and rate of mutation was higher for K1 than R17. Region 1 had a higher frequency and rate of mutation than region 2.

7.3.2 Mirid transmission

Figure 7.6 shows all mutations observed in isolates K1 and R17 after two years of mirid transmission relative to the P_R1 and P_K1 sequences.

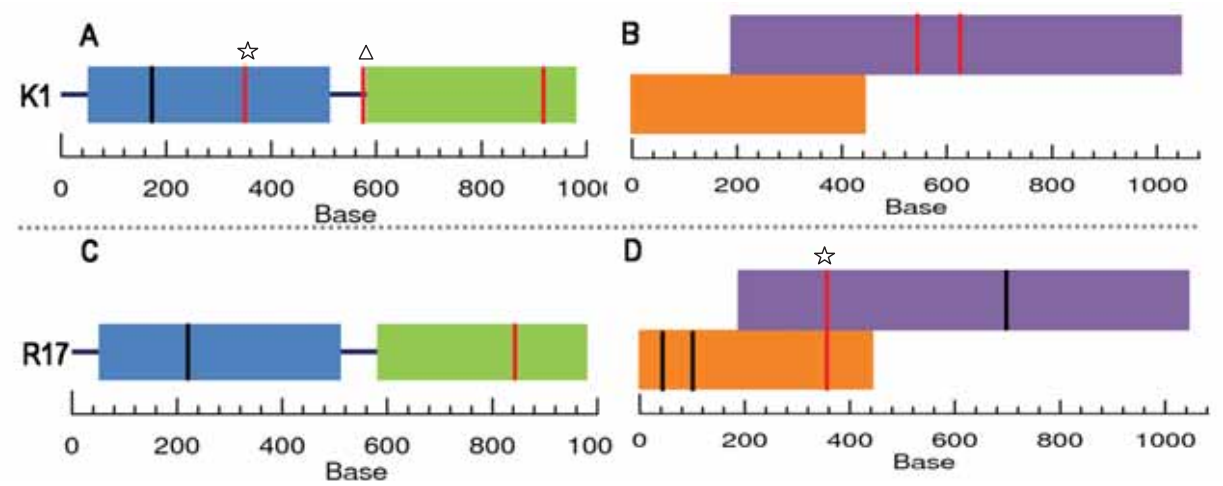


Figure 7. 6. Distribution of mutations in isolates K1 (A & B) and R17 (C & D) after 24 serial monthly passages by mirid transmission. Colours and labelling are the same as described for Figure 7.3.

Table 7.5 summarises the mutations shown in Figure 7.6, classified as either singleton or shared for each region.

Table 7.5 Mutations present in sequence collected from samples PI_R24 and PI_K24 from both region 1 and region 2.

Transmission	Region 1 Mutations			Region 2 Mutations		
	Total	Singleton	Shared	Total	Singleton	Shared
K1 Mirid PI_K24	4	1	3	2	0	2
R17 Mirid PI_R24	2	1	1	4	3	1

Isolate K1 had 5 shared mutations, while isolate R17 had 2. When the two regions were compared, region 1 had 4 shared mutations and region 2 had 3 shared mutations. Singleton mutations were disregarded from any further analysis and only shared mutations were classified by type of mutation (Table 7.6).

Table 7.6. Types of mutations observed in sequence collected from samples PI_R24 and PI_K24 in both region 1 and region 2. Those mutations marked with a * caused a change in the amino acid sequence.

Passage (Genomic region)	Transitions		Deletion
	A-G	T-C	
PI_K24 (Region 1)	1	1*	1
PI_K24 (Region 2)	1	1	
PI_R24 (Region 1)		1	
PI_R24 (Region 2)	1*		

There were an equal number (3) of both types of transition mutations; no transversion or multiple variant sites were observed. The deletion mutation was identical to polymorphism 1 from Table 5.2 (Chapter 5) and one of the deletion mutation observed in sequence from PM_K24 region 1. Table 7.7 shows the amino acid changes associated with the shared substitution mutations.

Table 7.7. Amino acid sequence changes associated with mutations observed in PI_R24 and PI_K24 sequence in region1 and region 2 after comparison with PI_R1 and PI_K1. Amino acid changes are listed by protein and indicated as either conservative with plain text or non-conservative by **bold text.**

Transmission passage	Region 1		Region 2	
	P1 (152 aa)	P2a (1-140 aa)	RdRp (401-549 aa)	CP (1-282 aa)
K1 Mirid PI_k24	**F102L**	-	-	-
R17 Mirid PI_R24	-	-	**I519M**	**Y54C**

Of the four mutations in the K1 sequence, only one caused a change in the amino acid sequence, this was a non-conservative substitution in the P1 protein between an aromatic F and an aliphatic L.

For isolate R17 the single nucleotide mutation in PI_R24 in region 2 caused two non-conservative substitutions in amino acid sequence. The change in the RdRp was from an aliphatic I to a Methionine (M) - which is the hydroxyl and sulphur side change group. The second change was in the CP and was from an aromatic Tyrosine (Y) to Cysteine (C) which is grouped in the hydroxyl and sulphur side chain amino acids.

Hydrophobicity plots

When the amino acid substitutions were analysed for their effect on protein hydrophobicity, all three amino acid substitutions were associated with change in the hydrophobicity plots of their respective proteins. In the K1 mutation F102L was associated with changes in hydrophobicity values of amino acids 98 to 106 in the P1 protein (Figure 7.7 A and B).

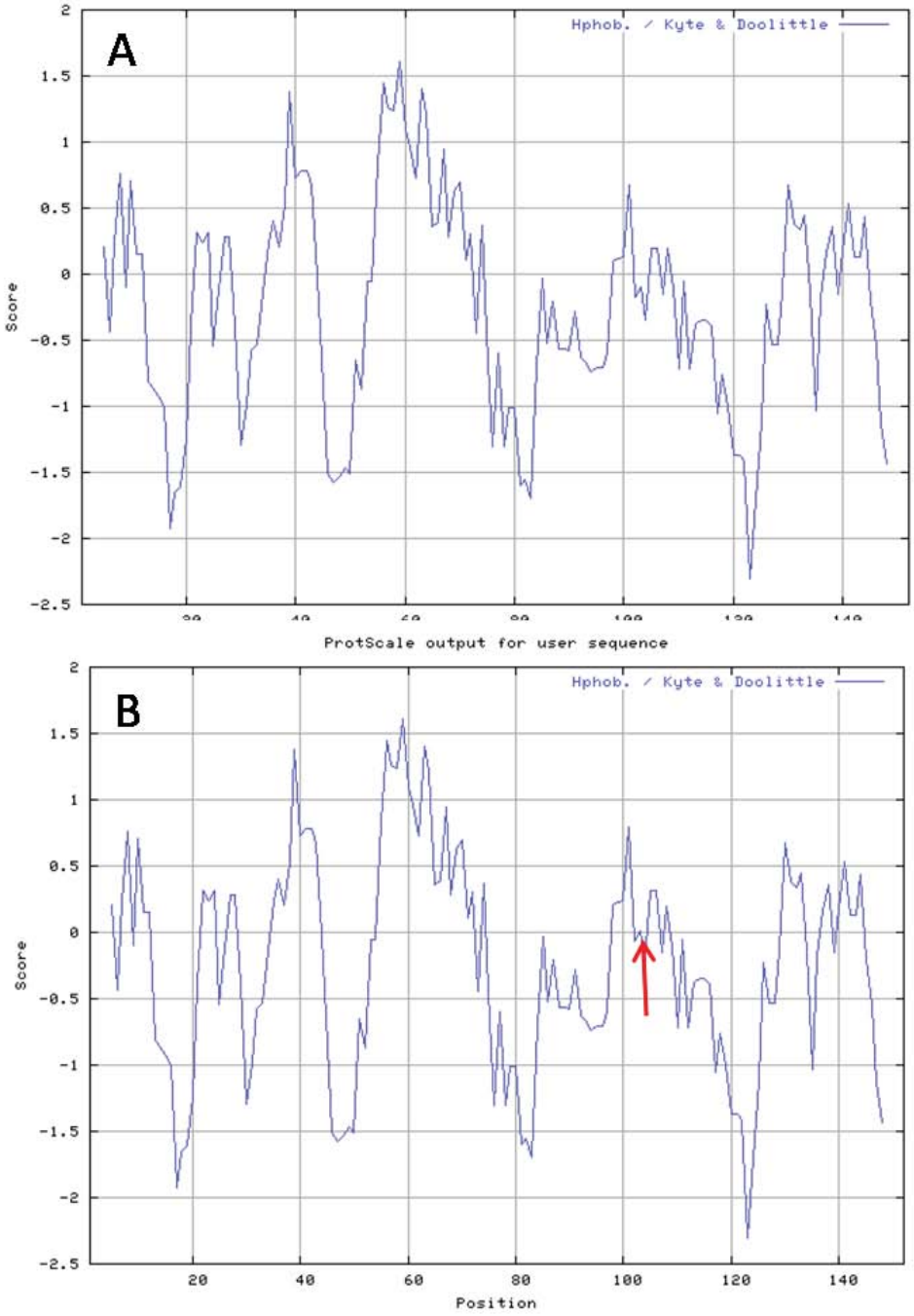


Figure 7.7.Hydrophobicity plots of the P1 amino acid sequence from isolate K1 before (**A**) and after 24 months of serial mirid transmission (PIK24) (**B**). Plots were created using the amino acid hydrophobicity scores of Kyte and Doolittle (1982). Hydrophobicity scores are shown on the Y axis and the amino acid position in the P1 on the X axis. The red arrow marks the site where the plots differ.

The second change was in R17, I519M. This mutation altered the hydrophobicity values from amino acids 515 to 523, which in turn altered the hydrophobicity plot of the RdRp protein (Figure 7.8 A and B).

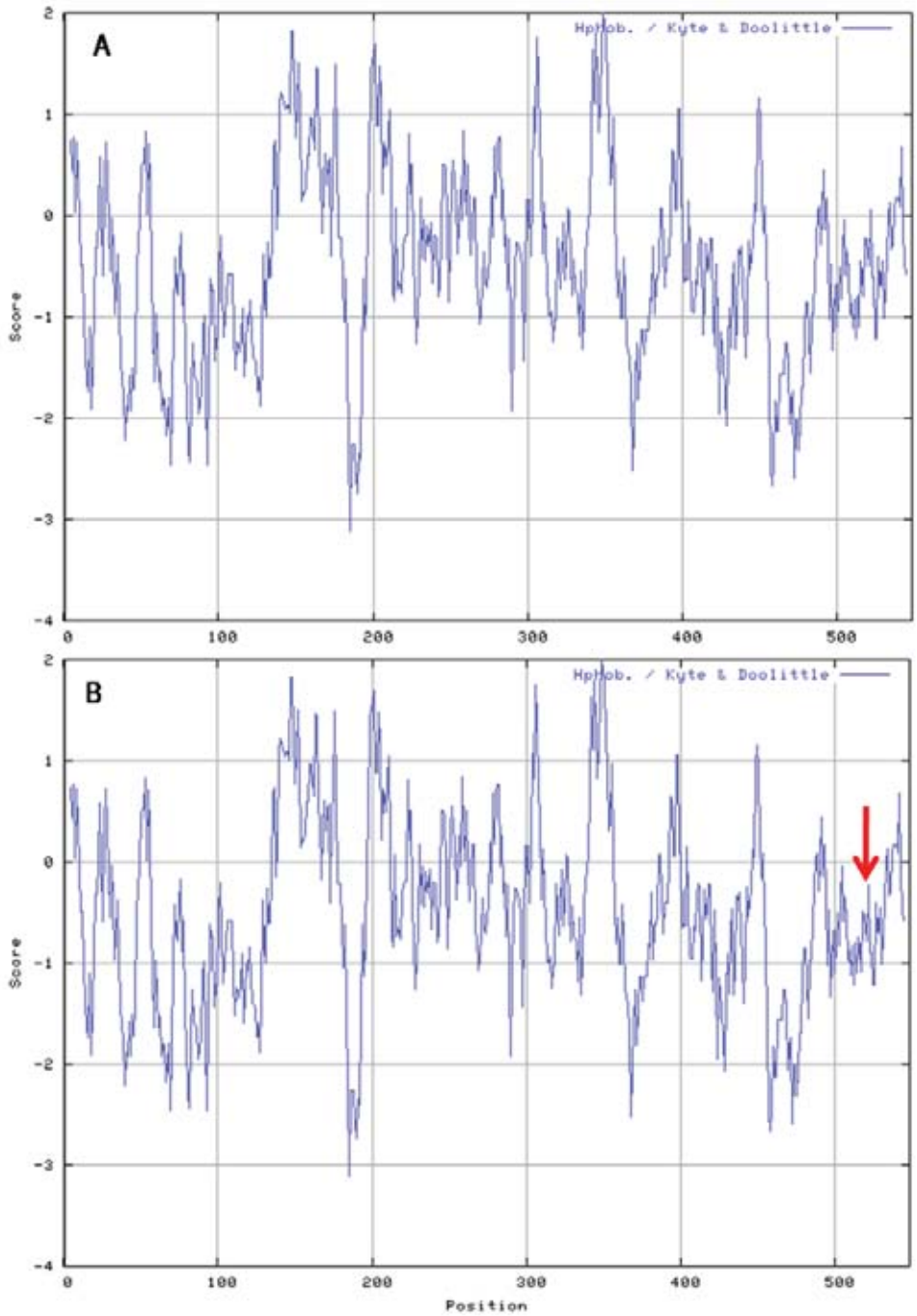


Figure 7.8.Hydrophobicity plots of the RdRp amino acid sequence from isolate R17 before (**A**) and after 24 months of serial mirid transmission (PI_R24) (**B**). Plots were created using the amino acid hydrophobicity scores of Kyte and Doolittle (1982). Hydrophobicity scores are shown on the Y axis and the amino acid position in the P1 on the X axis. The red arrow marks the site where the two plots differ.

The third change was in R17, Y54C. This mutation altered the hydrophobicity values from amino acids 50 to 58, which in turn altered the hydrophobicity plot of the CP protein, as shown in Figure 7.9 A and B.

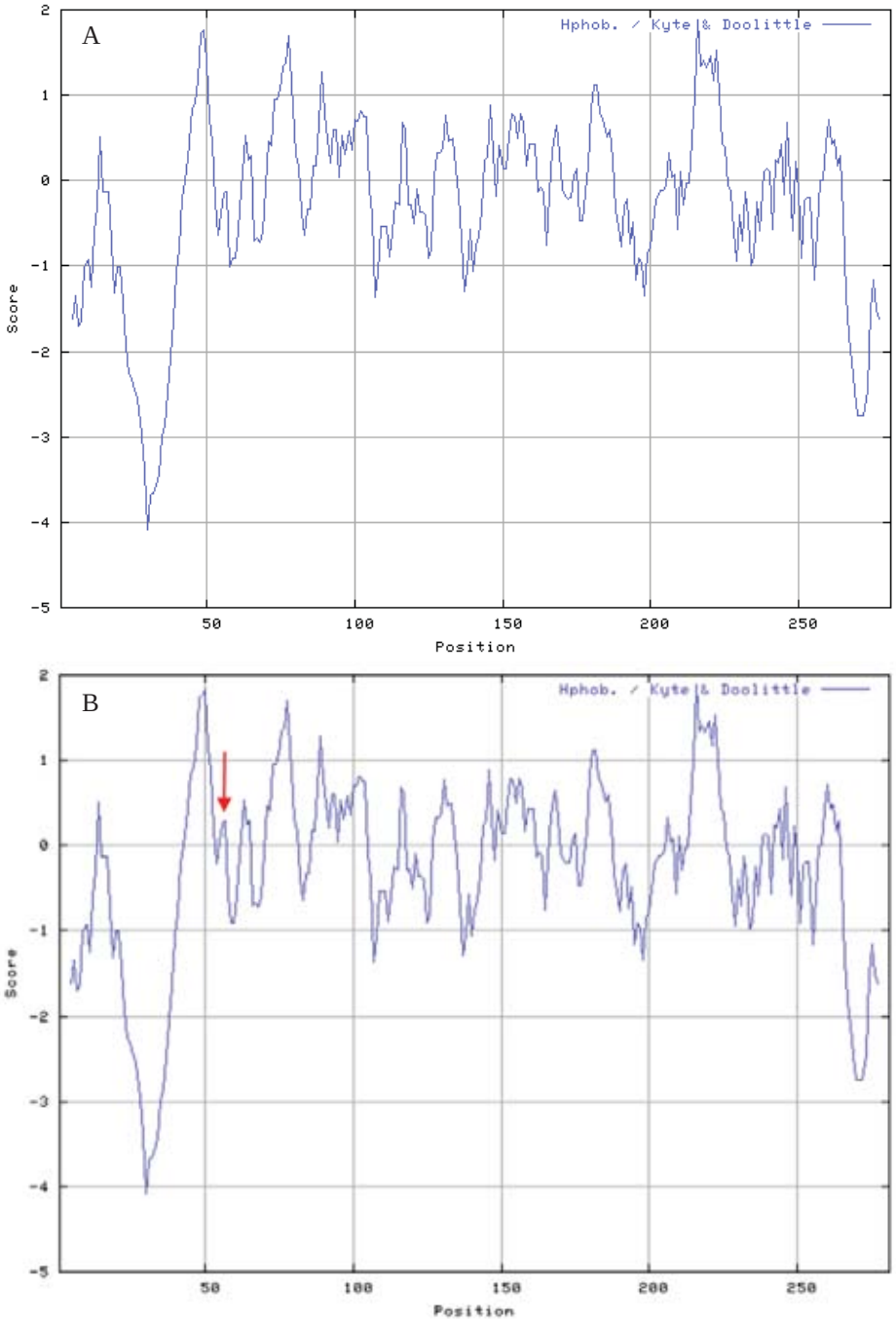


Figure 7.9.Hydrophobicity plots of the CP amino acid sequence from isolate R17 before (**A**) and after 24 months of serial mirid transmission (PI_R24) (**B**). Plots were created using the amino acid hydrophobicity scores of Kyte and Doolittle (1982). Hydrophobicity scores are shown on the Y axis and the amino acid position in the CP on the X axis. The red arrow marks the site where the two plots differ.

The frequency and rate of mutation was calculated (Table 7.8).

Table 7. 8. Frequency and rate of shared mutations of each isolate following serial mirid transmission over 24 months.

Isolate and genome region	Fragment size (nt)	Number of clones sequenced	Number of nucleotides sequenced	Shared Mutations	Mutation frequency per nt[a]	Rate of mutation per nt[a] /per year
PI_K24 Region 1	982	3	2946	3	10.18×10^{-4}	5.09×10^{-4}
PI_R24 Region 1	982	2	1964*	1	5.09×10^{-4}	2.55×10^{-4}
PI_K24 Region 2	1050	3	3134^	2	6.38×10^{-4}	3.19×10^{-4}
PI_R24 Region 2	1050	3	3150	1	3.17×10^{-4}	1.59×10^{-4}

^at the 5' end of one clone sequence nucleotides 9-25 16 are missing.*Sequence could only be collected from two clones, so to be classified as shared, the mutations had to be present on both clone sequences. 'a' units are the number of shared nucleotide mutations per number of nucleotides sequenced

When the two isolates were compared, K1 had a greater mutation frequency and mutation rate than isolate R17. In addition, region 1 had a higher frequency and rate of mutation than region 2. The rate and frequency of mutations in the sequences from mirid transmission samples were greater in region 2 than those values observed for mechanical transmission in Table 7.4.

7.3.3 Analysis of selective pressures

Mechanical and mirid transmission sequence results indicated that over half the non-synonymous mutations observed were located in the P1 protein. This suggests that different selective pressures were acting upon the ORFs. The ORF sequences for both modes of transmission were pooled and assessed for the presence of evolutionary selective pressure. The mean rates of both synonymous (d_S) and non synonymous (d_N) substitutions per site, and the ratio (ω) between these two values were determined and are displayed in Table 7.9

Table 7.9. Rates of non synonymous (d_N) and synonymous (d_S) mutations and the ratio between the two (ω) across the different ORFs analysed.

ORF	Isolate	Selective pressure		
		d_N ^	d_S *	ω d_N / d_S
P1	K1	0.0039 ± 0.0027	0	
	R17	0.0039 ± 0.0028	0	-
	Combined	0.0049 ± 0.0024	0	
Partial RdRp	K1	0	0.0055 ± 0.0030	
	R17	0.0009 ± 0.0008	0	
	Combined	0.0006 ± 0.0005	0.0049 ± 0.0024	0.1224
CP	K1	0	0.0055 ± 0.0030	
	R17	0.0009 ± 0.0008	0	
	Combined	0.0006 ± 0.0006	0.0049 ± 0.0025	0.1224

^ Units are the number of non-synonymous substitutions per non-synonymous site. # is the number of synonymous substitutions per the total number of synonymous sites.

The ratio of d_N / d_S (ω) cannot be calculated when values of zero are observed. However, selective pressures can still be inferred by comparing values of d_N and d_S with each other. When $d_N < d_S$ this is indicative of purifying selection. This was the pattern observed for both the partial RdRp and CP ORFs. When $d_N > d_S$ this indicates diversifying selection, and this pattern was observed for all the P1 sequences. When Datamonkey analysis was used for comparison no specific sites were detected to be under either purifying or diversifying selective pressure.

7.4 Discussion

Only 7 mutations were identified in each of the two modes of transmission so additional work will be needed to verify the observations and hypotheses generated with this data. This work could include repeating the experiment with parallel replications and/or continuation of the experiment for longer.

7.4.1 Mutations in nucleotide sequence

Sequence comparisons from multiple clones permitted classification of mutations as either shared or singleton. Thus singleton mutations, which were likely to be artefactual errors of the RT-PCR process (Smith et al. 1997), could be discarded in subsequent mutation analysis. This analysis showed only transition type mutations. This same bias towards transition mutation has also been observed in other serial passage experiments with CMV, TMV (Schneider & Roossinck 2000), *Wheat streak mosaic virus* (French & Stenger 2005) and *Turnip mosaic virus* (Ohshima et al. 2010). Initially transitional mutation biases were thought to be caused by bias in RdRp error prone replication (Kuge et al. 1989); however,

this explanation seems unlikely as the same transition mutation bias is observed in mammalian DNA genomes as well (Rosenberg et al. 2003). The exact reason for this bias towards transition mutations remains unclear.

The deletion mutation observed in the K1 sequences for both modes of transmission is the same as the deletion mutation observed in Chapter 5 in the R17 genome sequence. It may be that since this site is in a UTR that a nucleotide deletion here can occur without having a deleterious effect on the virus. The deletion occurred in a UTR but was adjacent to the start codon of ORF2a, which means it does cause a frameshift in the ORF2a gene. Despite this the ORF2a start codon is still in a favourable sequence context for translation with an A -3 nt upstream of the start codon (Chapter 5).

Analysis of these two regions from multiple isolates of the WINC would allow the determination of how often transition and deletion mutations occur in the VTMoV population. This may provide an indication of whether the two regions of the genome analysed as part of this study were conserved regions of the genome or were mutation hotspots where mutations occur frequently.

7.4.2 Mutations in amino acid sequence

Many of the mutations in amino acid sequence caused a change in the hydrophobicity plots of the proteins. However, further work would be required to determine if any of these mutations caused functional differences in the proteins. For proteins with assigned functions, mutations can be mapped relative to the annotated features. The R17 mutation in the CP was mapped to the β-annulus region (Chapter 4) in the random domain of the CP. Complete deletion of the β-annulus from mutant SeMV CP has shown this region is not crucial to assembly of empty virions, but its biological activity is still unknown (Satheshkumar et al. 2005). Comparison of the VTMoV CP with CPs from other sobemoviruses for which tertiary protein models exist (Opalka et al. 2000 ; Tars et al. 2003 ; Plevka et al. 2007) may allow for development of a VTMoV CP structure. This VTMoV CP structure would provide a starting point for the generation of specific hypotheses regarding the impact of individual mutations and genomic regions on protein function and virion formation, such as those undertaken for SeMV CP characterisation (Satheshkumar et al. 2005).

The VTMoV-P1 has no similarity to characterised proteins so more steps are required before direct experiments can be used to determine the effect of individual amino acid changes. The first step would be characterisation of the protein function, by creating a protein expression vector with the P1 protein. This should provide sufficient amounts of VTMoV-P1 protein for purification and crystallisation, which would allow for X-ray crystallography to determine the protein structure and identification of amino acids

crucial to this structure. Alternatively, the sequence of P1 from multiple VTMoV isolates could be compared. This comparison would allow identification of those regions of the protein which were conserved and which were more variable. In RYMV, this sequence comparison approach was used to isolate amino acid positions in the VPg that were connected to differences in isolate virulence (Pinel-Galzi et al. 2007).

7.4.3 VTMoV rate of mutation

The highest mutation rate was found in VTMoV- K1, region 1 at 5.09 $x10^{-4}$ nt/site/year for both modes of transmission. This is similar to the evolutionary rate of 5.2 $x10^{-4}$ nt/site/year calculated for RYMV from a field collection spanning 40 years (Fargette et al. 2008b). The evolutionary rate for RYMV after 6 months of serial passage experiments was observed 3 to 6 times higher in the serial passage than estimated from the mutation rate of the sample collection data (Fargette et al. 2008b). The authors argued that this large difference is likely due to the optimal growth conditions of the experimental plants, which allowed for a higher rate of virus replication (Fargette et al. 2008b). This high virus replication rate would allow for the accumulation of more substitutions (Hanada et al. 2004). Whether the experimental rate of mutation for VTMoV data is affected by these same factors will have to be confirmed by additional experiments.

Experiments to determine the replication rate of VTMoV in plants under different growth conditions would require the synthesis of an infectious clone of VTMoV. This clone could be serially passaged through plants under various growth conditions which might restrict plant growth, including limited water and/or limited nutrient availability. Leaf tissue from different infection stages could be collected in parallel from the different growth conditions and compared. These comparisons could include assessment of the substitution rate in VTMoV infectious clone sequences, along with both Real time PCR experiments and enzyme activity assays of the VTMoV-RdRp protein.

7.4.4 Trends in mutations

The only difference in mutations between the two modes of transmission was that sequence in region 2 of mirid transmitted samples had more mutations than region 2 sequence from mechanically transmitted samples. This difference may be caused by the mode of transmission, or it may simply be an artefact due to the low amount of mutations.

A trend that was observed in both modes of transmission was that region 1 sequence had a higher mutation rate than regions 2 sequence. If this difference was associated with transmission it would be

present in only one mode of transmission and not both. Therefore, it is unlikely that this difference in sequences between regions 1 and 2 is connected to the mode of transmission. Similar differences between genomic regions have been reported in studies with TMV. The TMV CP gene had a mutation rate of 5 $x10^{-4}$ nt/site/year, while the Replicase gene had a rate of 3 $x10^{-4}$ nt/site/year (Kearney et al. 1999). These differences in mutation rates suggest that these two regions may be under different selective pressures, but the numbers of mutations observed in this experiment are too small to give an accurate assessment of the evolutionary selective pressures influencing the VTMoV genes. Larger surveys of sequence variation in the CP and P1 genes similar to the RdRp study (Chapter 6) would be needed. These surveys would provide more accurate values for the mean rate of both non-synonymous and synonymous mutations, and ω and for analysis of diversifying or purifying selection.

The other pattern observed was that isolate K1 has a higher mutation rate than isolate R17 across both modes of transmission. It may be that isolate K1 replicates more than R17. This hypothesis could be tested with similar experiments to those mentioned in section 7.4.3. The accumulation of the VTMoV gRNA of these two isolates could be measured using Real Time PCR from a series of leaf samples taken after inoculation with the virus, along with enzymatic activity tests of the VTMoV RdRp.

- Chapter 8 -
General Discussion

This thesis has extended our knowledge of a native Australian virus from a unique ecosystem, describing the viral component of the VTMoV- Mirid - *N. velutina* ecological interaction. The complete genome sequence of VTMoV was reported, confirming the taxonomic position of the virus. Analysis of the sequence variation present in the RdRp gene from multiple VTMoV isolates was also outlined. Finally, the mutation rate of VTMoV after both mirid and mechanical transmission was compared to search for evidence of evolutionary mechanisms that may have influenced VTMoV evolution.

8.1 VTMoV genome sequence

While RNA species of VTMoV were described, no sequence information from the genome was available, except for the classification of VTMoV as a sobemovirus. Conserved regions of a sobemovirus genome alignment were targeted with degenerate primers, and one pair amplified a small fragment of the VTMoV genome. This fragment provided the starting point for a primer walking strategy, and RACE protocol was applied to amplify the 5' and 3' ends of the genome sequence. This approach was preferred to the random cDNA library approach, because it provided immediate information on the VTMoV genome that could be used to undertake the sequence variation studies in parallel, rather than requiring the complete assembly and annotation of the genome first.

Annotation of the genome sequence showed VTMoV had four ORFs, arranged in the same way as the genome model of other sobemovirus species. The identities of ORFs 2a, 2b and 4 were initially based on BLAST results. These were confirmed after annotation of features in both the nucleotide and protein sequences that matched corresponding ORFs in other sobemovirus species (Chapter 4).

The overlapping fragments of VTMoV sequence assembly confirmed that ORF1 was a genuine virus sequence and not of plant origin. However, the identity of the P1 protein from ORF1 could not be confirmed from comparison to characterised sequence, as no matches from BLAST searches were found. This lack of similarity between the P1 of sobemovirus species, meant another approach was required. Information about VTMoV-P1 function was obtained using a bioinformatics approach which predicted the structure of the protein. This approach putatively identified the VTMoV-P1 as a Lipocalin protein. This protein family is a widely distributed heterogeneous group (Grzyb et al. 2006) with a diverse functionality (Pugalenthi et al. 2010). Further experimentation will be required to determine the

specific role of P1 in VTMoV infection, but this is the first record of a gene for any such protein being identified in a plant virus genome.

Phylogenetic analyses of the ORFs and the complete sequence, confirm the taxonomic classification of VTMoV as a sobemovirus. It is a unique virus which does not group with other Australian sobemovirus species (LTSV and SCMoV), but as a single virus lineage not related to any of the extant groupings of sobemoviruses. VTMoV is serologically related to SNMoV (Randles et al. 1981), but no sequence information for this virus is available. The sequencing approach using sobemovirus degenerate primers could be used to sequence the SNMoV genome. cDNA hybridisation studies between VTMoV and SNMoV RNA1 species suggest that these two viruses may have some common regions of sequence (Gould & Hatta 1981), and maybe some VTMoV primers will be useful in sequencing the SNMoV sequence.

Almost all the RNA species of VTMoV have now been sequenced. The genome sequence of VTMoV-K1 was not directly amplified from purified RNA 1 species, but the size of the RNA 1 is very similar to the genome sequence of K1 from this thesis. This could be confirmed by isolation of the RNA 1 species from purified virions and testing the sequence using RT-PCR primers based on the genome sequence of K1.

The remaining RNA species that are not sequenced are RNAs 1a and 1b. These RNA species were initially identified as degradation products of RNA 1 because both RNA1a and 1b hybridised with cDNA probes created from RNA 1 sequence; and because RNA species of similar size to RNA 1a and 1b were observed in analysis of purified RNA 1 after storage (Gould 1981). However, one of these VTMoV RNA species may also be subgenomic RNA. In SBMV, sgRNA is used to translate the CP (Ghosh et al. 1981), and the CACAAA motif is found at the 5' end of both the gRNA and sgRNA of SBMV sequences (Hacker & Sivakumaran 1997). This same motif is found in the VTMoV sequence upstream of the CP gene and at the 5' end of the sequence. This may indicate that VTMoV translates its CP from a sgRNA in the same way as SBMV. Sequencing of RNA 1a and 1b would confirm their identities. If one or both of these RNA species are sgRNA then it would be expected that they contain sequence identical to the 3' end of the sequence downstream of position 3226 where the CACAAA motif was found.

The reason VTMoV-K1 was selected for sequencing was because it was previously shown to have no detectable virusoid (Francki et al. 1986a). K1 was classified as virusoid free based on the lack of RNA 2 and RNA 3 observed after staining of denaturing PAGE (Francki et al. 1986a). The apparent ability of isolate K1 to cause infection without the RNA 2 and RNA 3 species was why the virusoid was classified

as a satellite RNA rather than part of the genome. Satellite RNAs are defined by their reliance on their helper viruses for encapsidation and replication (Roossinck et al. 1992). However, this thesis reported the detection of virusoid RNA in VTMoV-K1 using RT-PCR (Chapter 3). This RT-PCR detection now raises doubts about the satellite nature of the virusoid RNA. It may be that the virusoid are not satellites but may have a function. This RT-PCR result also raises some doubts as to the satellite nature of the virusoids from other sobemovirus species as well. Previous studies with SNMoV and LTSV isolated the RNA 1 from PAGE gels and used this to establish RNA 2 (virusoid) free isolates of both viruses (Jones et al. 1983 ; Jones & Mayo 1984). Confirmation that these isolates were virusoid free needs to be determined again using RT-PCR.

To test the satellite nature of the VTMoV virusoid the RNA 1 of VTMoV must be purified, or an infectious clone of this sequence synthesised. If this RNA 1 sequence is able to replicate and cause infection without virusoid RNA, this would confirm the satellite RNA nature of the virusoid. If the RNA 1 does not cause infection, then it may be that the virusoid plays an as yet unknown function.

8.2 VTMoV sequence variation

The sequence of VTMoV-K1 provided information allowing the sequencing of VTMoV R17. Although this isolate had virusoid RNA detectable by PAGE (Francki et al. 1986a), this did not hinder the collection of sequence as VTMoV-specific primers were able to be used. The R17 sequence was assembled with fewer sequence fragments, but lacked the 3'UTR and part of the 5'UTR. The genome sequence of the R17 genome covered 96.7 % of the complete genome sequence, and showed the same arrangement of ORFs and annotated features as the genome of isolate K1. There were only 10 nucleotide polymorphisms distinguishing the two isolates and the virusoid sequence of the two isolates was identical.

The main differences between R17 and K1 may be explained by a fundamental difference in their RdRp replication rate. The K1 isolate had accumulated more mutations than R17 after two years of transmission (Chapter 7). This variation in mutation may be due to a difference in the replication rate of the isolates. The RdRp of K1 may have the opportunity to accumulate more mutations in the K1 sequence because K1 replicates more frequently than isolate R17 does. The reason for this difference in replication may also be due to the virusoid sequence of R17, which is present at high levels. The R17 virusoid may 'highjack' the R17 RdRp for its own replication providing less RdRp for R17 genome replication. This lower R17 genome replication rate may then result in fewer mutations accumulating in the sequence (Hanada et al. 2004), explaining the low mutation level relative to isolate K1. This

hypothesis could be tested using infectious clones of both VTMoV-K1 and R17. If these clones are infectious without virusoid RNA they could be used to inoculated plants and their replication rates monitored by Real Time PCR experiments. ThePCR targets would need to be designed to avoid the CP in case VTMoV does have a sgRNA component. In addition virusoid RNA could be added in parallel treatments. The replication rates of each of the infectious clones and the virusoid RNAs could be then be compared to see if the patterns agree with the hypothesis or not.

A low amount of sequence diversity was also observed in analysis of the RdRp gene from 15 isolates of VTMoV. The genetic diversity reported in this thesis was 0.02, similar to the genetic diversity observed within West African clades of RYMV based on complete sequence data (Fargette et al. 2004). In comparison the east African groups of RYMV from the same 2004 study had a diversity of approximately 0.07 (Fargette et al. 2004). This low amount of genetic diversity in VTMoV could be due to the limited number of isolates sampled or the evolutionary forces acting upon the gene. In support of the latter idea the majority of nucleotide polymorphisms that were observed did not cause changes in amino acid sequence of the protein. Population genetic analysis also confirmed the virus was under purifying selective pressure to conserve the protein sequence, presumably to maintain protein function.

More sequence from additional isolates from more genomic regions is what is needed to accurately measure the genetic diversity across the VTMoV. A large scale survey of VTMoV could be conducted by testing a range of *N. velutina* plants using the DIBA methodology outlined in section 2.3.4. Sequence could then be collected from the VTMoV positive samples for a small region of the genome such as the P1 or CP genes. This sequence data could then be used to choose which isolates to sequence in full. The complete coding regions of these isolates could be collected using the same approach and methodology as in Chapter 5 to sequence isolate R17.

Furthermore, these phylogenetic data could be used to calculate an evolutionary rate. Either the evolutionary rates can be directly estimated from phylogenetic trees (Gibbs et al. 2010) or from gene sequences of known sampling dates (heterochronous sampling). Heterochronous sampling of RYMV samples collected over ~40 years was used to calculate the evolutionary rate for RYMV (Fargette et al. 2008a). Although, the most direct method of estimating rates of evolution is to compare serially passaged samples (Gibbs et al. 2010), very few mutations were observed in parts of the VTMoV genomes of isolates K1 and R17 after two years. A comparison of the two methods of evolutionary rate calculations can provide different estimates. In the DNA virus, *Banana bunchy top virus* mutation rates varied between 1.4 $x10^{-4}$ from a heterochronous sampling method to 3.9 $x10^{-4}$ from a serial sampling method (Almeida et al. 2009). Obtaining rates of evolution using both methods may provide a more

accurate estimate of the evolutionary rate of VTMoV, and may also help to determine if serial sampling experiment reduces or over estimates the evolutionary rate of VTMoV as discussed in section 7.4.3.

8.3 VTMoV transmission

The mutation observed after the two year experiment provided some evidence of evolutionary mechanisms changing the sequences, but the specific mechanism remains unclear. In studies with mutant CMV, the inoculation phase of non persistent aphid transmission was found to be the source of genetic bottlenecks (Ali et al. 2006). Perhaps, the experimental design did not provide sufficient opportunity for bottlenecks to occur. A lot of mirids were present on the plants for up to four weeks at a time providing an opportunity for a large number of inoculation events to occur. The horizontal non-persistent aphid transmission of *Potato virus Y* was estimated at 0.5 to 3.2 virus particles per aphid (Moury et al. 2007). It may be that a more limited transmission is needed, perhaps using individual mirids to transmit VTMoV between single plants would increase the level of mutation observed. However, this more limiting transmission route may also artificially increase the amount of genetic bottlenecks and may not be reflective of natural transmission.

Alternatively, it may be that there is minimal evolutionary pressure imposed by mirid transmission. For potyvirus-aphid transmission, the maintenance of key sequence features is crucial to maintain aphid transmissibility. The potyvirus virion interacts with the Helper component (Hc) protein, which acts as a 'bridge' between the virus and the vector stylet (Fernandez-Calvino et al. 2010). This interaction is specific and requires the conservation of several key amino acid motifs KITC (Atreya & Pirone 1993) and PTK (Huet et al. 1994) in the HC Pro, and the DAGX motif in the CP (Atreya et al. 1991). Purifying selection would be required to remove any deleterious mutations from arising in these motifs. However, the mirid-VTMoV interaction may not be as specific as this potyvirus-aphid interaction. It may be that VTMoV transmission does not require specific protein interactions between vector and virus. Sobemoviruses are readily transmitted mechanically (Tamm & Truve 2000) and mirids have been shown to transmit other sobemoviruses, including SNMoV, SBMV and SoMV (Gibb & Randles 1988). VTMoV transmission via mirids may simply be a case of 'contamination'. Mirids acquire the virus through feeding on infected sap; the virus contaminates mirid mouthparts, and survives in the mirid gut, to be inoculated in a non-specific mechanical manner when the mirid moves onto new plants. VTMoV may simply 'hitch a ride' on the mirid, and this less specific mode of transmission would not necessarily require the specific interaction of the virion with the insect. Without a specific interaction, natural selection may not be needed to maintain key motifs in the virion, so there would be no detection of

purifying selection in sequences after mirid transmission. This latter hypothesis is supported by experiments that showed mirids have also been shown to transmit other sobemoviruses.

Now that more is known about the viral component of this unique ecological interaction, there is an opportunity to study the interactions between all the elements in the ecological niche of VTMoV. Viruses are widely distributed in the plant kingdom (Cooper & Jones 2006), and most plant viruses are studied as pathogens that cause disease in cultivated plants (Zaitlin & Palukaitis 2000). However, the concept of symbiosis between virus and host plants is gaining acceptance (Roossinck 2011). One specific area in which *N. velutina* may gain advantage from VTMoV infection would be drought tolerance. In different plant species, infection with different viruses improved the drought tolerance of the plants (Xu et al. 2008). This drought tolerance would be useful for *N. velutina* in its natural environment. Alternatively, the stunting symptoms of VTMoV infection may also aid plant survival. Experimental tests comparing VTMoV infected and healthy *N. velutina* plants under drought conditions would help in determining if symbiosis exists between these two species.

- References -

Abubakar, Z., Ali, F., Pinel, A., Traore, O., N'Guessan, P., Notteghem, J.-L., Kimmins, F., Konate, G. and Fargette, D. (2003). "Phylogeography of *Rice yellow mottle virus* in Africa." Journal of General Virology **84**(3): 733-743.

Ali, A., Li, H., Schneider, W. L., Sherman, D. J., Gray, S., Smith, D. and Roossinck, M. J. (2006). "Analysis of genetic bottlenecks during horizontal transmission of *Cucumber mosaic virus*." Journal of Virology **80**(17): 8345-8350.

Ali, A. and Roossinck, M. (2008). Genetic Bottlenecks. Plant virus evolution. M. Roossinck. Heidelberg, Springer- Verlag.

Almeida, R. P. P., Bennett, G. M., Anhalt, M. D., Tsai, C.-W. and O'Grady, P. (2009). "Spread of an introduced vector-borne banana virus in Hawaii." Molecular Ecology **18**(1): 136-146.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhand, Z., Miller, W. and Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**: 3389-3402.

Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schaffer, A. A. and Yu, Y. K. (2005). "Protein database searches using compositionally adjusted substitution matrices." The Federation of European Biochemical Societies Journal **272**(20): 5099-5100.

Amaral, P. P., Resende, R. O. and Souza Júnior, M. T. (2006). "Papaya lethal yellowing virus (PLYV) infects *Vasconcellea cauliflora*." Fitopatologia Brasileira **31**: 517.

Anisimova, M., Nielsen, R. and Yang, Z. (2003). "Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites." Genetics **164**(3): 1229-1236.

Arthur, K., Dogra, S. and Randles, J. W. (2010). "Complete nucleotide sequence of *Velvet tobacco mottle virus* isolate K1." Archives of Virology **155**(11): 1893-1896.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. and Sherlock, G. (2000). "Gene Ontology: tool for the unification of biology." Nature Genetics **25**(1): 25-29.

Atreya, C. D. and Pirone, T. P. (1993). "Mutational analysis of the helper component-proteinase gene of a potyvirus: effects of amino acid substitutions, deletions, and gene replacement on virulence and aphid transmissibility." Proceedings of the National Academy of Sciences of the United States of America **90**(24): 11919-11923.

Atreya, P. L., Atreya, C. D. and Pirone, T. P. (1991). "Amino acid substitutions in the coat protein result in loss of insect transmissibility of a plant virus." Proceedings of the National Academy of Sciences of the United States of America **88**(17): 7887-7891.

Bakker, W. (1975). "Rice yellow mottle virus." AAB Descriptions of Plant Viruses **149**.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Wheeler, D. L. (2008). "GenBank." Nucleic Acids Research **36**(Database Issue): D25-D30.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). "The Protein Data Bank." Nucleic Acids Research **28**(1): 235-242.

Boni, M. F., Posada, D. and Feldman, M. W. (2007). "An exact nonparametric method for inferring mosaic structure in sequence triplets." Genetics **176**: 1035-1047.

Bonneau, C., Brugidou, C., Chen, L., Beachy, R. N. and Fauquet, C. (1998). "Expression of the Rice yellow mottle virus P1 protein *in vitro* and *in vivo* and its involvement in virus spread." Virology **244**(1): 79-86.

Botanic, G. T. (2007). PlantNET-The Plant Information Network System of the Botanic Gardens Trust (2.0).

Catherall, P. L. (1970). "Cocksfoot mottle virus " AAB Descriptions of Plant Viruses **23**.

Charron, J. B., Ouellet, F., Pelletier, M., Danyluk, J., Chauve, C. and Sarhan, F. (2005). "Indentification, expression and evolutionary analyses of plant Lipocalins." Plant Physiology **139**: 2017-2028.

Chomczynski, P. and Sacchi, N. (1987). "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction." Analytical Biochemistry **162**(1): 156-159.

Cole, C., Barber, J. D. and Barton, G. J. (2008). "The Jpred 3 secondary structure prediction server." Nucleic Acids Research **36**(suppl_2): W197-201.

Collins, R. F., Gellatly, D. L., Sehgal, O. P. and Abouhaidar, M. G. (1998). "Self-cleaving circular RNA associated with Rice Yellow Mottle virus is the smallest viroid-like RNA." Virology **241**: 269-275.

Cooper, I. and Jones, R. A. C. (2006). "Wild plants and viruses: Underinvestigated ecosystems." Advances in Virus Research **67**: 1-47.

Davies, C., Haseloff, J. and Symons, R. H. (1990). "Structure, self-cleavage, and replication of two viroid-like satellite RNAs (virusoids) of *Subterranean clover mottle virus*." Virology **177**(1): 216-224.

de la Iglesia, F. and Elena, S. F. (2007). "Fitness declines in *Tobacco etch virus* upon serial bottleneck transfers." Journal of Virology **81**(10): 4941-4947.

Delport, W., Poon, A. F. Y., Frost, S. D. W. and Kosakovsky Pond, S. L. (2010). "Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology." Bioinformatics **26**(19): 2455-2457.

Delport, W., Scheffler, K. and Seoighe, C. (2008). "Frequent toggling between alternative amino acids is driven by selection in HIV-1." PLoS Pathogens **4**(12): e1000242.

Elena, S. F., Dopazo, J., de la Pena, M., Flores, R., Diener, T. O. and Moya, A. (2001). "Phylogenetic analysis of viroid and viroid-like satellite RNA from plants: A reassessment." Journal of Molecular Evolution **53**: 155-159.

Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007). "Locating proteins in the cell using TargetP, SignalP and related tools." Nature Protocols **2**(4): 953-971.

Fargette, D., Pinel-Galzi, A., Sereme, D., Lacombe, S., Hebrard, E., Traore, O. and Konate, G. (2008a). "Diversification of *Rice yellow mottle virus* and related viruses spans the history of agriculture from the Neolithic to the present." PLoS Pathogens **4**(8): e1000125.

Fargette, D., Pinel, A., Abubakar, Z., Traore, O., Brugidou, C., Fatogoma, S., Hebrard, E., Choisy, M., Sere, Y., Fauquet, C. M. and Konate, G. (2004). "Inferring the evolutionary history of *Rice yellow mottle virus* from genomic, phylogenetic, and phylogeographic studies." Journal of Virology **78**(7): 3252-3261.

Fargette, D., Pinel, A., Rakotomalala, M., Sangu, E., Traore, O., Sereme, D., Sorho, F., Issaka, S., Hebrard, E., Sere, Y., Kanyeka, Z. and Konate, G. (2008b). "*Rice yellow mottle virus*, an RNA plant virus, evolves as rapidly as most RNA animal viruses." Journal of Virology **82**(7): 3584-3589.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. and Ball, L. A., Eds. (2005). Virus Taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. London, UK, Elsevier Academic Press.

Fernandez-Calvino, L., Goytia, E., Lopez-Abella, D., Giner, A., Urizarna, M., Vilaplana, L. and Lopez-Moya, J. J. (2010). "The helper-component protease transmission factor of tobacco etch potyvirus binds specifically to an aphid ribosomal protein homologous to the laminin receptor precursor." Journal of General Virology **91**(11): 2862-2873.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L. L., Eddy, S. R. and Bateman, A. (2010). "The Pfam protein families database." Nucleic Acids Research **38**(suppl_1): D211-222.

Flasinski, S. and Cassidy, B. G. (1998). "Potyvirus aphid transmission requires helper component and homologous coat protein for maximal efficiency." Archives of Virology **143**(11): 2159-2172.

Forster, R. L. S. and Jones, A. T. (1980). "Lucerne transient streak virus." AAB Descriptions of Plant Viruses **224**.

Francki, R. I. B., Chu, P. W. G. and Keese, P. (1983). The satellite nature of viroidlike RNA from Lucerne transient streak virus. Plant Infectious Agents. H. D. Robertson, S. H. Howell, M. Zaitlin and R. L. Malmberg. New York, Cold Spring Harbor Laboratory**:** 175-180.

Francki, R. I. B., Grivell, C. J. and Gibb, K. S. (1986a). "Isolation of Velvet tobacco mottle virus capable of replication with and without viroid-like RNA." Virology **148**: 281-384.

Francki, R. I. B., Zaitlin, M. and Palukaitis, P. (1986b). "*In vivo* encapsidation of Potato spindle tuber viroid by Velvet tobacco mottle virus particles." Virology **155**: 469-473.

French, R. and Stenger, D. C. (2005). "Population structure within lineages of *Wheat streak mosaic virus* derived from a common founding event exhibits stochastic variation inconsistent with the deterministic quasi-species concept." Virology **343**: 179-189.

Futterer, J. and Hohn, T. (1996). "Translation in plants- rules and exceptions." Plant Molecular Biology **32**(1): 159-189.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocol Handbook. J. M. Walker. Totowa, NJ, USA, Humana Press**:** 571-607.

Gasteiger, E., Hoogland, C., Ivanyi, I., Appel, R. D. and Bairoch, A. (2003). "ExPASy - the proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Research **31**: 3784-3788.

Gayathri, P., Satheshkumar, P. S., Prasad, K., Nair, S., Savithri, H. S. and Murthy, M. R. N. (2006). "Crystal structure of the serine protease domain of *Sesbania mosaic virus* polyprotein and mutational analysis of residues forming the S1-binding pocket." Virology **346**(2): 440-451.

Ghosh, A., Rutgers, T., Ke-Qiang, M. and Kaesberg, P. (1981). "Characterization of the coat protein mRNA of Southern bean mosaic virus and its relationship to the genomic RNA." Journal of Virology **39**(1): 87-92.

Gibb, K. S. and Randles, J. W. (1988). "Studies on the transmission of Velvet tobacco mottle virus by the mirid, Cyrtopeltis nicotianae." Annals of Applied Biology **112**: 427-437.

Gibb, K. S. and Randles, J. W. (1990). "Distribution of Velvet tobacco mottle virus in its mirid vector and its relationship to transmissibility." Annals of Applied Biology **116**: 513-521.

Gibbs, A. J., Fargette, D., Garcia-Arenal, F. and Gibbs, M. J. (2010). "Time - the emerging dimension of plant virus studies." Journal of General Virology **91**(1): 13-22.

Gibbs, M., Armstrong, J. S. and Gibbs, A. (2000). "Sister-scanning: A Monte carlo procedure for asessing signals in recombinant sequences." Bioinformatics **16**(7): 573-582.

Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. and Koonin, E. V. (1989). "Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases: A distinct protein superfamily with a common structural fold." The Federation of European Biochemical Societies Letters **243**(2): 103-114.

Gorbalenya, A. E., Koonin, E. V., Blinov, V. M. and Donchenko, A. P. (1988). "Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses." The Federation of European Biochemical Societies Letters **236**(2): 287-290.

Gould, A. R. (1981). "Studies on the encapsidated viroid-like RNA: II. Purification and characterisation of a viroid-like RNA associated with Velvet tobacco mottle virus (VTMoV)." Virology **108**: 123-133.

Gould, A. R., Francki, R. I. B. and Randles, J. W. (1981). "Studies on the encapsidated viroid-like RNA: IV. Requirement for infectivity and specificity of two RNA components from Velvet tobacco mottle virus." Virology **110**: 420-426.

Gould, A. R. and Hatta, T. (1981). "Studies on the encapsidated viroid-like RNA: III. Comparitive studies on RNAs isolated from Velvet tobacco mottle virus and *Solanum nodiflorum* mottle virus." Virology **109**: 137-147.

Govind, K. and Savithri, H. S. (2010). "Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase." Virology **401**(2): 280-292.

Greber, R. S. and Randles, J. W. (1986). "Solanum nodiflorum mottle virus." AAB Descriptions of Plant Viruses **318**.

Grzyb, J., Latowski, D. and Strzalka, K. (2006). "Lipocalins - a family portrait." Journal of Plant Physiology **163**(9): 895-915.

Hacker, D. L. and Sivakumaran, K. (1997). "Mapping and expression of Southern bean mosaic virus genomic and subgenomic RNAs." Virology **234**(2): 317-327.

Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symposium Series **41**: 95-98.

Hanada, K. and Francki, R. I. B. (1989). "Kinetics of Velvet tobacco mottle virus satellite RNA synthesis and encapsidation." Virology **170**: 48-54.

Hanada, K., Suzuki, Y. and Gojobori, T. (2004). "A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a riversity of viral infection and transmission modes." Molecular Biology and Evolution **21**(6): 1074-1080.

Harrison, B. D. (1981). "Plant virus ecology: ingredients, interactions and environmental influences." Annals of Applied Biology **99**: 95-209.

Haseloff, J. and Symons, R. H. (1982). "Comparative sequence and structure of viroid-like RNAs of two plant viruses." Nucleic Acids Research **10**(12): 3681-3691.

Hebrard, E., Pinel-Galzi, A., Catherinot, V., Labesse, G., Brugidou, C. and Fargette, D. (2005). "Internal point mutations of the capsid modify the serotype of Rice yellow mottle virus." Journal of Virology **79**(7): 4407-4414.

Hermodson, M. A., Abad-Zapatero, C., Abdel-Meguid, S. S., Pundak, S., Rossmann, M. G. and Tremaine, J. H. (1982). "Amino acid sequence of Southern bean mosaic virus coat protein and its relation to the three-dimensional structure of the virus." Virology **119**(1): 133-149.

Hibi, T. and Saito, Y. (1985). "A Dot immunobinding assay for the detection of Tobacco mosaic virus in infected tissues." Journal of General Virology **66**: 1191-1194.

Hollings, M. and Stone, O. M. (1973). "Turnip rosette virus." AAB Descriptions of Plant Viruses **125**.

Holmes, E. C., Worobey, M. and Rambaut, A. (1999). "Phylogenetic evidence for recombination in dengue virus." Molecular Biology and Evolution **16**: 405-409.

Hsu, C. H., Sehgal, O. P. and Pickett, E. E. (1976). "Stabilizing effect of divalent metal ions on virions of Southern bean mosaic virus." Virology **69**(2): 587-595.

Huet, H., Gal-On, A., Meir, E., Lecoq, H. and Raccah, B. (1994). "Mutations in the helper component protease gene of zucchini yellow mosaic virus affect its ability to mediate aphid transmissibility." Journal of General Virology **75**(6): 1407-1414.

Hull, R. (1977). "The stabilization of the particles of turnip rosette virus and of other members of the southern bean mosaic virus group." Virology **79**(1): 58-66.

Hull, R. (2004). "Southern bean mosaic virus." AAB Descriptions of Plant Viruses **408**.

Ishikawa, J. and Hotta, K. (1999). "FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content." FEMS Microbiology Letters **174**(2): 251-253.

Jones, A. T. and Mayo, M. A. (1984). "Satellite nature of the viroid-like RNA-2 of *Solanum nodiflorum mottle virus* and the ability of other plant viruses to support the replication of viroid-like RNA molecules." Journal of General Virology **65**(10): 1713-1721.

Jones, A. T., Mayo, M. A. and Duncan, G. H. (1983). "Satellite-like properties of small circular RNA molecules inparticles of Lucerne transient streak virus." Journal of General Virology **64**(5): 1167-1173.

Jones, R. A. C., Fosu-Nyarko, J., Jones, M. and Dwyer, G. I. (2001). "Subterranean clover mottle virus." AAB Descriptions of Plant Viruses **387**.

Kado, C. I. (1971). "Sowbane mosaic virus." AAB Descriptions of Plant Viruses **64**.

Kamer, G. and Argos, P. (1984). "Primary structural comparison of RNA-dependent polymerases from plant, animal adn bacterial viruses." Nucleic Acids Research **12**: 7629-7282.

Kearney, C. M., Thomson, M. J. and Roland, K. E. (1999). "Genome evolution of tobacco mosaic virus populations during long-term passaging in a diverse range of hosts." Archives of Virology **144**(8): 1513-1526.

Keese, P. and Symons, R. H. (1987). The structure of viroids and virusoids. Viroids and viroid-like pathogens. J. S. Semancik. Boca Raton, CC Press Inc.

Kimura, M. (1980). "A simple method for estimating evolutionary rate of base substitutions through comparitive studies of nucleotide seqeunces." Journal of Molecular Evolution **16**: 111-120.

Koonin, E. (1991). "The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses." Journal of General Virology **72**: 2197-2206.

Koonin, E. and Dolja, V. V. (1993). "Evolution and taxonomy of positive-strand RNA viruses: Implications of comparative analysis of amino acid sequences." Critical Reviews in Biochemistry and Molecular Biology **28**(5): 375-430.

Koramova, A. V., Haenni, A.-L. and Ramirez, B. C. (2009). "Virus versus host cell translation: Love and hate stories." Advances in Virus Research **73**: 99-170.

Kosakovsky-Pond, S. L. and Frost, S. D. W. (2005). "Not so different after all: A comparison of methods for detecting amino acid sites under selection." Molecular Biology and Evolution **22**(5): 1208-1222.

Kosakovsky-Pond, S. L., Frost, S. D. W., Grossman, Z., Gravenor, M. B., Richman, D. D. and Brown, A. J. L. (2006a). "Adaptation to different human populations by HIV-1 revealed by codon-based analyses." PLoS Computational Biology **2**(6): e62.

Kosakovsky-Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. and Frost, S. D. W. (2006b). "Automated phylogenetic detection of recombination using a genetic algorithm." Molecular Biology and Evolution **23**(10): 1891-1901.

Kuge, S., Kawamura, N. and Nomoto, A. (1989). "Strong inclination toward transition mutation in nucleotide substitutions by Poliovirus replicase." Journal of Molecular Biology **207**: 175-182.

Kumar, S., Dudley, J., Nei, M. and Tamura, K. (2008). "MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences." Briefings in Bioinformatics **9**: 299-306.

Kyte, J. and Doolittle, R. F. (1982). "Amino acid scale: Hydropathicity." Journal of Molecular Biology **157**: 105-132.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lee, L. and Anderson, E. J. (1998). "Nucleotide sequence of a resistance breaking mutant of southern bean mosaic virus." Archives of Virology **143**(11): 2189-2201.

Lee, S.-K., Dabney-Smith, C., Hacker, D. L. and Bruce, B. D. (2001). "Membrane activity of the *Southern cowpea mosaic virus* coat protein: The role of basic amino acids, helix-forming potential, and lipid composition." Virology **291**(2): 299-310.

Lee, S.-K. and Hacker, D. L. (2001). "In vitro analysis of an RNA binding site within the N-terminal 30 amino acids of the *Southern cowpea mosaic virus* Coat Protein." Virology **286**(2): 317-327.

Li, H. and Roossinck, M. (2004). "Genetic bottlenecks reduce population variation in an experimental RNA virus population." Journal of Virology **78**(19): 10582-10587.

Li, W.-H. (1993). "Unbiased estimation of the rates of synonymous and nonsynonymous substitutions." Journal of Molecular Evolution **36**: 96-99.

Librado, P. and Rozas, J. (2009). "DnaSP v5: A software for comprehensive analysis of DNA polymorphism data." Bioinformatics **25**: 1451-1452.

Lokesh, G. L., Gopinath, K., Satheshkumar, P. S. and Savithri, H. S. (2006). "AY004291."

Makinen, K., Makelainen, K., Arshava, N., Tamm, T., Merits, A., Truve, E., Zavriev, S. and Saarma, M. (2000). "Characterization of VPg and the polyprotein processing of Cocksfoot mottle virus (genus Sobemovirus)." Journal of General Virology **81**(11): 2783-2789.

Makinen, K., Naess, V., Tamm, T., Truve, E., Aaspollu, A. and Saarma, M. (1995a). "The putative replicase of the cocksfoot mottle sobemovirus is translated as a part of the polyprotein by -1 ribosomal frameshift." Virology **207**(2): 566-571.

Makinen, K., Tamm, T., Naess, V., Truve, E., Puurand, I., Munthe, T. and Saarma, M. (1995b). "Characterization of cocksfoot mottle sobemovirus genomic RNA and sequence comparison with related viruses." Journal of General Virology **76**(11): 2817-2825.

Martin, D. P., Posada, D., Crandall, K. A. and Williamson, C. (2005a). "A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints." AIDS Research and Human Retroviruses **21**: 98-102.

Martin, D. P. and Rybicki, E. P. (2000). "RDP: detection of recombination amongst aligned sequences." Bioinformatics **16**(6): 562-563.

Martin, D. P., Williamson, C. and Posada, D. (2005b). "RDP2: recombination detection and analysis from sequence alignments." Bioinformatics **21**(2): 260-262.

Mathews, C. K., van Holde, K. E. and Ahern, K. G. (2000). Biochemistry. San Francisco, CA, Addison Wesley Longman.

McGavin, W. J. and MacFarlane, S. A. (2009). "Rubus chlorotic mottle virus, a new sobemovirus infecting raspberry and bramble." Virus Research **139**(1): 10-13.

Meier, M., Olspert, A., Sarmiento, C. and Truve, E. (2008). Sobemovirus. Desk Encyclopedia of plant and fungal virology. B. W. J. Mahy and M. van Regenmortel. Oxford, UK, Academic Press**:** 312-320.

Meier, M., Paves, H., Olspert, A., Tamm, T. and Truve, E. (2006). "P1 Protein of *Cocksfoot mottle virus* is indispensable for the systemic spread of the virus." Virus Genes **32**(3): 321-326.

Meier, M. and Truve, E. (2007). "Sobemoviruses possess a common CfMV-like genomic organisation." Archives of Virology **152**: 635-640.

Miller, W. A., Dinesh-Kumar, S. P. and Paul, C. P. (1995). "Luteovirus gene expression." Critical Reviews in Plant Science **14**(3): 179-211.

Moraes, T. F., Edwards, R. A., McKenna, S., Pastushok, L., Xiao, W., Glover, J. N. M. and Ellison, M. J. (2001). "Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13." Nature Structural & Molecular Biology **8**(8): 669-673.

Moury, B., Fabre, F. and Senoussi, R. (2007). "Estimation of the number of virus particles transmitted by an insect vector." Proceedings of the National Academy of Sciences **104**(45): 17891-17896.

Murant, A. F., Mayo, M. A., Harrison, B. D. and Goold, R. A. (1972). "Properties of virus and RNA components of *Raspberry ringspot virus*." Journal of General Virology **16**(3): 327-338.

N'Guessan, P., Pinel, A., Caruana, M. L., Frutos, R., Sy, A., Ghesquière, A. and Fargette, D. (2000). "Evidence of the presence of two serotypes of rice yellow mottle sobemovirus in Cote d'Ivoire." European Journal of Plant Pathology **106**(2): 167-178.

Nair, R., Carter, P. and Rost, B. (2003). "NLSdb: database of nuclear localization signals." Nucleic Acids Research **31**(1): 397-399.

Nair, S. and Savithri, H. S. (2010). "Processing of SeMV polyproteins revisited." Virology **396**(1): 106-117.

Nascimento, A. K. Q., Lima, J. A. A., Nascimento, A. L. L., Beserra, E. A. and Purcifull, D. E. (2010). "Biological, physical, and molecular properties of a Papaya lethal yellowing virus isolate." Plant Disease **94**(10): 1206-1212.

Nei, M. (1987). Moleuclar Evolutionary Genetics. New York, Columbia University Press.

Ng and Perry (1999). "Stability of the aphid transmission phenotype in *Cucumber mosaic virus*." Plant Pathology **48**(3): 388-394.

Ohshima, K., Akaishi, S., Kajiyama, H., Koga, R. and Gibbs, A. J. (2010). "Evolutionary trajectory of turnip mosaic virus populations adapting to a new host." Journal of General Virology **91**(3): 788-801.

Olspert, A., Paves, H., Toomela, R., Tamm, T. and Truve, E. (2010). "Cocksfoot mottle sobemovirus coat protein contains two nuclear localization signals." Virus Genes **40**(3): 423-431.

Olspert, A., Peil, L., Hebrard, E., Fargette, D. and Truve, E. (2011). "Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs." Journal of General Virology **92**(2): 445-452.

Opalka, N., Tihova, M., Brugidou, C., Kumar, A., Beachy, R. N., Fauquet, C. M. and Yeager, M. (2000). "Structure of native and expanded sobemoviruses by electron cryo-microscopy and image reconstruction." Journal of Molecular Biology **303**(2): 197-211.

Padidam, M., Sawyer, S. and Fauquet, C. (1999). "Possible emergence of new geminiviruses by frequent recombination." Virology **265**: 218-225.

Pamilo, P. and Bianchi, N. O. (1993). "Evolution of the Zfx and Zfy genus: Rates and Interdepedence between the genes." Molecular Biology and Evolution **10**: 271-281.

Pinel-Galzi, A., Mpunami, A., Sangu, E., Rakotomalala, M., Traore, O., Sereme, D., Sorho, F., Sere, Y., Kanyeka, Z., Konate, G. and Fargette, D. (2009). "Recombination, selection and clock-like evolution of *Rice yellow mottle virus*." Virology **394**(1): 164-172.

Pinel-Galzi, A., Rakotomalala, M., Sangu, E., Sorho, F., Kanyeka, Z., Traore, O., Sereme, D., Poulicard, N., Rabenantoandro, Y., Sere, Y., Konate, G., Ghesquiere, A., Hebrard, E. and Fargette, D. (2007). "Theme and variations in the evolutionary pathways to virulence of an RNA plant virus species." PLoS Pathogens **4**(11): 180-1770.

Plevka, P., Tars, K., Zeltins, A., Balke, I., Truve, E. and Liljas, L. (2007). "The three-dimensional structure of *Ryegrass mottle virus* at 2.9A resolution." Virology **369**(2): 364-374.

Posada, D. and Crandall, K. A. (2001). "Evaluation of methods for detecting recombination from DNA sequences: Computer simulations." Proceedings of the National Academy of Sciences of the United States of America **98**(24): 13757-13762.

Pugalenthi, G., Kandaswamy, K., Suganthan, P., Archunan, G. and Sowdhamini, R. (2010). "Identification of functionally diverse lipocalin proteins from sequence information using support vector machine." Amino Acids **39**(3): 777-783.

Qu, C., Liljas, L., Opalka, N., Brugidou, C., Yeager, M., Beachy, R. N., Fauquet, C. M., Johnson, J. E. and Lin, T. (2000). "3D Domain swapping modulates the stability of members of an icosahedral virus group." Structure **8**(10): 1095-1103.

Ramsdell, D. C. (1979). "Blueberry shoestring virus." AAB Descriptions of Plant Viruses **204**.

Randles, J. W., Davies, C., Hatta, T., Gould, A. R. and Francki, R. I. B. (1981). "Studies on encapsidated viroid-like RNA I. Characterization of *Velvet tobacco mottle virus*." Virology **108**: 111-122.

Randles, J. W. and Francki, R. I. B. (1986). "Velvet tobacco mottle virus." AAB Descriptions of Plant Viruses **317**.

Rohozinski, J. (1985). Studies of velvet tobacco mottle virus RNA replication by enzyme-template complexes in extracts from infected leaves. Plant Pathology. Adelaide, University of Adelaide. **PhD:** 141.

Roossinck, M. J. (2011). "The good viruses: viral mutualistic symbioses." Nature Reviews Microbiology **9**(2): 99-108.

Roossinck, M. J., Sleat, D. and Palukaitis, P. (1992). "Satellite RNAs of plant viruses: structures and biological effects." Microbiol. Mol. Biol. Rev. **56**(2): 265-279.

Rose, P. P. and Korber, B. T. (2000). "Detecting hypermutations in viral sequences with an emphasis on G -> A hypermutation." Bioinformatics **16**(4): 400-401.

Rosenberg, M. S., Subramanian, S. and Kumar, S. (2003). "Patterns of transitional mutation biases within and among mammalian genomes." Molecular Biology and Evolution **20**(6): 988-993.

Roy, A., Kucukural, A. and Zhang, Y. (2010). "I-TASSER: a unified platform for automated protein structure ad function prediction." Nature Protocols **5**(4): 725-738.

Sambrook, J. and Russell, D. W. (2001). Molecular cloning: a laboratory guide. Cold Spring Harbor, NY, USA, Cold Spring Harbor Laboratory Press.

Sarmiento, C., Gomez, E., Meier, M., Kavanagh, T. A. and Truve, E. (2007). "*Cocksfoot mottle virus* P1 suppresses RNA silencing in *Nicotiana benthamiana* and *Nicotiana tabacum*." Virus Research **123**(1): 95-99.

Satheshkumar, P. S., Lokesh, G. L., Murthy, M. R. N. and Savithri, H. S. (2005). "The role of Arginine-rich motif and β -annulus in the assembly and stability of *Sesbania mosaic virus* capsids." Journal of Molecular Biology **353**(2): 447-458.

Satheshkumar, P. S., Lokesh, G. L. and Savithri, H. S. (2004). "Polyprotein processing: cis and trans proteolytic activities of Sesbania mosaic virus serine protease." Virology **318**(1): 429-438.

Scheffler, K., Martin, D. P. and Seoighe, C. (2006). "Robust inference of positive selection from recombining coding sequences." Bioinformatics **22**(20): 2493-2499.

Schneider, W. L. and Roossinck, M. J. (2000). "Evolutionarily related sindbis-like plant viruses maintain different levels of population diversity in a common host." Journal of Virology **74**(7): 3130-3134.

Scholthof, H. B. (2005). "Plant virus transport: motions of functional equivalence." Trends in Plant Science **10**(8): 376-382.

Sérémé, D., Lacombe, S., Konaté, M., Pinel-Galzi, A., Traoré, V., Hébrard, E., Traoré, O., Brugidou, C., Fargette, D. and Konaté, G. (2008). "Biological and molecular characterization of a putative new sobemovirus infecting *Imperata cylindrica* and maize in Africa." Archives of Virology **153**(10): 1813-1820.

Smith, D., McAllister, J., Casino, C. and Simmonds, P. (1997). "Virus 'quasispecies': making a mountain out of a molehill?" Journal of General Virology **78**(7): 1511-1519.

Smith, J. M. (1992). "Analyzing the mosaic structure of genes." Journal of Molecular Evolution **34**(2): 126-129.

Subramanya, H. S., Gopinath, K., Nayudu, M. V., Savithri, H. S. and Murthy, M. R. N. (1993). "Structure of *Sesbania* mosaic virus at 4.7A resolution and partial sequence of the coat protein." Journal of Molecular Biology **229**(1): 20-25.

Symons, R. H. and Randles, J. W. (1999). "Encapsidated circular viroid-like satellite RNAs (virusoids) of plants." Current Topics in Microbiology and Immunology **239**: 81-105.

Tamm, T., Suurväli, J., Lucchesi, J., Olspert, A. and Truve, E. (2009). "Stem-loop structure of Cocksfoot mottle virus RNA is indispensable for programmed -1 ribosomal frameshifting." Virus Research **146**(1-2): 73-80.

Tamm, T. and Truve, E. (2000). "Sobemoviruses." Journal of Virology **74**(14): 6231-6241.

Tars, K., Zeltins, A. and Liljas, L. (2003). "The three-dimensional structure of cocksfoot mottle virus at 2.7 å resolution." Virology **310**(2): 287-297.

Toriyama, S., Mikoshiba, Y. and Doi, Y. (1983). "Ryegrass mottle virus, a new virus from *Lolium multiflorum* in Japan." Annals of the Phytopathological society of Japan **49**: 610-618.

Veerisetty, V. and Sehgal, O. P. (1980). "Proteinase-K sensitive factor essential for the infectiviy of Southern bean mosaic virus ribonucleic acid." Phytopathology **70**(4): 282-284.

Voinnet, O., Pinto, Y. M. and Baulcombe, D. C. (1999). "Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants." Proceedings of the National Academy of Sciences of the United States of America **96**(24): 14147-14152.

Weiller, G. F. (1998). "Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences." Molecular Biology and Evolution **15**: 326-335.

White, S. A., Briand, L., Scott , D. J. and Borysik, A. J. (2009). "Structure of rat odorant-binding protein OBP1 at 1.6 A resolution." Acta Crystallographica Section D: Biological Crystallography **65**: 403-410.

Xie, Z., Ruas, P. and Shen, Q. J. (2005). "Regulatory Networks of the Phytohormone Abscisic Acid." Vitamins and Hormones **72**: 235-269.

Xiong, J. (2006). Essential Bioinformatics. New York, NY, USA, Cambridge University Press.

Xu, P., Chen, F., Mannas, J. P., Feldman, T., Sumner, L. W. and Roossinck, M. J. (2008). "Virus infection improves drought tolerance." New Phytologist **180**(4): 911-921.

Zaitlin, M. and Palukaitis, P. (2000). "Advances in understanding plant viruses and virus diseases." Annual Review of Phytopathology **37**: 117-143.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." BMC Bioinformatics **9**(40).

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000). "A greedy algorithm for aligning DNA sequences." Journal of Computational Biology **7**(1-2): 203-214.

Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Research **31**(13): 3406-3415.

APPENDICES

Appendix 1: Published manuscript

Arthur, K., Dogra, S. and Randles, J.W. (2010) Complete nucleotide sequence of Velvet tobacco mottle virus isolate K1
Archives of Virology, v.155 (11), pp. 1893-1896, November 2010

Appendix 2: List of VTMoV isolates in WINC

The WINC isolates of VTMoV are listed in Table A1.1, according to host plant, WINC number, and collection details. The oldest isolates were collected from the natural host plant *N. velutina. N. clevelandii* isolates have been inoculated from older isolates to increase stock, or for use in research.

Table A1.1 VTMoV isolates in the WINC, organised by WINC number and specimen details. Key: Grey boxes indicate isolates where the genome has been sequenced. ɪ Indicates North East South Australia. a- Collector John W. Randles, b- Collector G. Behncken, '-indicates same as previous box. Blank spaces are left where no information is available.

WINC	Nicotiana velutina			Nicotiana clevelandii		Specimens from research		Notes
	Name	Collection location	Date	Name	Date	Name	Date	
65754	VTMoV V2	Strzlecki track,ɪ[a]	--/11/1979					(Randles et al. 1981)
65755	VTMoV (NVV2) #15	"	--/07/1979					
65756	VTMoV (NVV2) # 17	"	--/07/1979					
65757	VTMoV (NVV2) # 18	"	--/07/1979					
65758	VTMoV-QLD	Cunnamulla QLD, b	23/08/1982					
65759	VTMoV V7	Strzlecki track, NE SAɪ,a	--/11/1979					
65760	VTMoV V1	"	--/11/1979					
65761	VTMoV (NVV2) 10A	"	--/07/1979					
65762	VTMoV (NVV2) #19	"	--/07/1979					
65763	VTMoV (NVV2) #14	"	--/07/1979					
65764	VTMoV (NVV2) #13	"	--/07/1979					
65765	VTMoV (NVV2) #16	"	--/07/1979					
65779	VTMoV	Yaningurie water hole ɪ[a]	24/11/2000					
65780	VTMoV	Cobblers sand hill ɪ[a]	24/11/2000					Same collection site as original 1979 samples
65771				VTMoV B1	17/10/1984			
65772				VTMoV B2	"			
65773				VTMoV A	"			
65774				VTMoV C	"			

WINC	Nicotiana velutina			Nicotiana clevelandii		Specimens from research		Notes
	Name	Collection location	Date	Name	Date	Name	Date	
65795				VTMoV 13 Strz [a]	02/11/1984			
65796				VTMoV 14 strz [a]	"			
65797				VTMoV 15 Strz [a]	"			
65798				VTMoV 16 strz [a]	"			
65799				VTMoV 17 strz [a]	"			
65800				VTMoV 18 strz [a]	"			
65766						VTMoV (17)	--/09/1988	
65767						VTMoV	--/05/1983	Jan's mutant (Rohozinski 1985)
65769						VTMoV-F	--/06/1983	RNA1 Only ex. pKAS
65770						VTMoV (K1-isol)	--/02/1986	(Francki et al. 1986b) used for PSTVd encapsidation ex. Cornell
65678						VTMoV stock 17	--/05/1983	
65775						VTMoV	12/09/1989	Repassage of sample Randles et al., 1981)
65776						VTMoV # 17	30/03/1988	
65777						VTMoV	09/11/1988	With sat (p19?)
65778						VTMoV	21/12/1988	
65781						VTMoV-K1	--/02/2004	(Francki et al. 1986a) minus sat
65782						VTMoV-F	--/02/2004	minus sat
65783						VTMoV-R17	--/02/2004	(Randles et al. 1981) with sat

Appendix 3: Solutions, buffers and media

Double distilled water was used to make solutions. All solutions were autoclaved at 121 °C for 20 min. Water used for nucleic acid solutions were treated with 0.1 % DEPC overnight and then autoclaved.

Virion purification solutions

100 mM sodium potassium buffer (pH 7.4)	20 % (10 ml) 1M KH_2PO_4 and 80 % (40 ml) Na_2HPO_4, diluted to 100 mM
Tris-HCl buffers	1 M Tris stock solutions adjusted pH to 7.5, 9.5 and 8 with HCl
Nycodenz	60 %, 50 %, 40 % and 30 % (w/v) nycodenz® solution in 100 mM Tris-HCl (pH 7.5)

Nucleic acid extraction buffers

Guanidine thiocyanate buffer	38 % TE saturated phenol, 1.2 M guanidine thiocyanate, 5 % glycerol, 0.1 M sodium acetate
Phenol: Chloroform	1:1 mix of water saturated phenol and chloroform

Protein digestion solutions

Protease Solution	0.5 mg ml^{-1} protease (from *Streptomyces griseus,* Sigma) in 0.05 M sodium acetate, 0.25 % SDS. Solution incubated for 30 min at 37 ºC before storing at -20 ºC.
Proteinase K solution	40 mg ml^{-1} solution of Proteinase K (31.7 U mg^{-1} isolated from *Triticum album*, Amresco, USA) was prepared in 0.1 M Tris-HCl (pH 7.5). Solution incubated overnight at room temperature before freezing at -20 ºC.

Polyacrylamide gel electrophoresis solutions

Polyacrylamide mixture (19:1)	9.5 g acrylamide, 0.5 g *bis*-acrylamide made up to 50 ml in sterile water.
Polyacrylamide mixture	1X TBE, 3.5 % acrylamide (19:1), 0.007 % ammonium persulphate, 0.035 % Temed
Denaturing polyacrylamide mixture	1X TBE, 3.5 % acrylamide (19:1), 0.007 % ammonium persulphate, 0.035 %Temed, 7 M urea

Agarose gel electrophoresis solutions

1X TE buffer	100 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0).
5X TBE buffer	45 mM Tris, 45 mM Boric acid, 1 mM EDTA (pH 8.0). Diluted to 0.5X for electrophoresis buffer
6X blue loading dye	10 mM Tris-HCl (pH 7.5), 50 mM EDTA, 10 % Ficoll 400, 0.25 % bromophenol blue, 0.25 % xylene cyanol FF 1X working stocks made by dilution with DEPC treated water
DNA size standard ladders	66.6 ng μl^{-1} of either 1Kb ladder (Invitrogen) or 1Kb^{+} ladder (Invitrogen) diluted in 0.8X TE and 1.2X loading dye

Cloning

Luria Agar and Luria Broth (Sigma)	Both liquid and agar media were prepared according to manufacturer's instructions in water and the antibiotic added from 100 µg μl^{-1} stocks after sterilisation
SOC media Provided by Invitrogen Life Technologies (Carlsbad, CA, USA).	2 % tryptone, 0.5 % yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose

DIBA buffers

1X Phosphate buffered saline (PBS)	137 mM NaCl, 1.5 mM KH_2PO_4, 8 mM Na_2HPO_4, 2.7 mM KCl
Wash buffer AP 7.5	0.1 M Tris-HCl (pH7.5), 0.1 mM NaCl, 2 mM $MgCl_2$, 0.05 % Triton X100
Wash buffer AP 9.5	0.1M Tris-HCl (pH 9.5), 0.1 M NaCl, 5 mM $MgCl_2$.
Stop Buffer	10 mM Tris-HCl (pH 7.5), 5 mM EDTA
Blocking buffer A	1X PBS + 2.6 % (w/v) Skim milk powder- final casein concentration 1 %.
Healthy sap sample	Crushed healthy leaf 1:1 in 1X PBS and clarified by centrifugation at 1000 *g* for 5 min
Blocking buffer B	1:10 dilution of healthy sap sample in Blocking buffer A.
75 mg ml^{-1} nitro blue tetrazolium chloride (NBT) stock	Dissolved 75 mg NBT in 700 µl of 100 % dimethylformamide and 300 µl of water
5-Bromo-4-chloro-3-indolyl phosphate (BCIP) 50 mg ml^{-1} stock	Dissolved 50 mg in 1ml of dimethylformamide

Suppliers

Transcriptor Reverse Transcriptase	Roche Diagnostics GmbH
RNase OUT	Invitrogen Life Technologies
RNase H	Invitrogen Life Technologies
1kb / 1kb^{+} ladders	Invitrogen Life Technologies
Taq polymerase	Kapa Taq DNA polymerase (Kapa Biosystems) through (Geneworks, Adelaide)
TOPO cloning kit	Invitrogen Life Technologies, including chemically competent cells
E.coli Poly (A) polymerase	New England Biolabs
Goat anti- rabbit antibody	Sigma Aldrich (MI, USA)

Appendix 4: Primers

All primers were supplied by Geneworks (Adelaide, Australia) in dried form and dissolved in DEPC treated water for 100 µM stock solutions. All primers were designed by the author unless otherwise stated.

Table A4.1 Degenerate primers used

Primer name	Primer sequence (5'-3' orientation)	Genome position[1]	Degeneracy value	Tm[2] (ºC)	Length (nt)
SOBDF2	GAYATHTCNGGHTTYGAYTGGTC	2951- 2974	288	55	23
SOBDR2	CCYTCVACVGARTCRTCDCCCAT	3250-3228	216	59	22
SOBDF4	MTTTAAACTRCYWRSRGGYG	2050-2069	256	50	20
ORF2aDF3	GTGYCKMVSAASDHYTGGTCT	-	1728	55	21

1- Approximate nucleotide positions on the VTMoV-K1 genome sequence, 2- Tm provided by supplier (Geneworks).

Table A4.2 VTMoV sequence specific primers

Primer name	Primer sequence (5'-3' orientation)	Genome position[1]	Tm[2] (ºC)	Length (nt)
VTMoV-Ra	GCATGATGGTCTACGGAGGTTGATGG	2730-2755	61	26
VTMoV-Rf	GTCCAGGCTCAACTTGTTCATAC	2833-2811	55	23
VTMoV-CPF1	CACCAAGAATCAGGTCAAGCAG	3277-3298	55	22
VTMoV-CPR2	TTCAGCGAGGAGTTTAGAACCAC	3782-3759	55	23
VTMoV-2aR2	TTCCCTTTATTCTTAGAGGCGATG	1662-1639	54	24
VTMoV-2aR3	TCTTTCTCACGGATCGAGTCATAG	1722-1699	56	24
VTMoV-CPF3	CAGTTCGTCATTCGGAGGTGG	3465-3485	55	21
VTMoV-CPF4	ATTGAGCCCCTTCCTCCTTCG	3989-4009	56	21
VTMoV-CPF5	GATTTCCCGCAACAAGTTTTGGCTCA	4122-4143	58	26
VTMoV-CPR4	CGCTTGCCTAGCGGTCAGTT	4115-4096	56	20
VTMoV-2aR4	GACACAACTGGACTAACTCGCTC	596-574	57	22
VTMoV-ORF1R1	TTCCACTTGGCAGTTGACGGTA	361-382	22	55
PolyA primer	AAAAAAAAAAAAAAAAAAAAAAAN	-	25	39
VTMoV-5UTRF	ATATTGAAGGGAAAGTCTTGC	8-28	21	48
VTMoV-2aR1	TGCTCAGCAGAACTGATTGC	1955-1936	52	20
VTMoV-Re	AAGAGAGCATAAAGCAGCGG	2785-2766	52	20
VTMoV-2bF2	AGATTTCCTTGCCCGCTTCC	1720-1739	54	20

Primer name	Primer sequence (5'-3' orientation)	Genome position[1]	Tm[2] (ºC)	Length (nt)
VTMoV-2bR1	CACATACGTGGCTCTAAGCG	3622-3603	54	20
VTMoV-2bR2	GAGTCGTAGAAGCGGTTTC	3644-3626	51	19
VTMoV-2aR6	AGACAACGAGAGACTTTGGTTCA	989-967	53	23

1- Nucleotide positions on the VTMoV-K1 genome sequence, 2- Tm provided by supplier (Geneworks).

Table A4.3 Primers from external sources

Primer name	Sequence (5'-3')	Length (nt)	Tm (ºC)*	Reference
Random hexamers		-	-	Invitrogen
Oligo $(dT)_{20}$		-	-	
M13F	GTAAAACGACGGCCAG	16	46	TOPO Cloning kit Invitrogen
M13R	CAGGAAACAGCTATGAC	17	44	
AGSatF	GTGGATGTGTATCCACTCTGATGAGTC	27	37	Designed by Allan Gould (pers. Communication 2009)
AGSatR	TACGGACACGGACCAGGCGAACCAAAAATC	30	37	
PA14	Sequence of primers protected	-	37 or 45	Allan Gould (pers. Communication 2009)
NUP	AAGCAGTGGTATCAACGCAGAGT	23	-	Clontech SMART RACE protocol

*Tm provided by supplier.

Appendix 5: Sequences of fragments 1-12 from VTMoV isolate K1

>Fragment 1
GACATCTCAGGTTTTGATTGGTCTGTGCAGGAATGGGAGTTATGGGCCGATCTGAGCATGAGGATTTCTTTGTGC
GAAGATATGCATGATGGTCTACGGAGGTTGATGGTTAACAGATACCGCTGCTTTATGCTCTCTTGCTTCCAATTG
TCGAATGGGGAATTGTATGAACAAGTTGAGCCTGGACTCATGAAGTCTGGTTCCTACTGCACTTCCTCTTCCAAC
TCCAGAATCAGGTGCCTAATGGGTTATCTAATTGGAGCCCCCTGGATAATAGCCATGGGCGACGACTCTGTTGAG
GG

>Fragment 2
CACCAAGAATCAGGTCAAGCAGGAACCCCATACGAAGAAGAAGCTGAATGAGCGTCGCTTTAGGCTTATTTCGTC
TGTGTCTATAGTTGATCAGATAATCGAGAGATTATTGTTTGGTCCTCAGAACAGGCTTGAGATAGCCTTATGGCA
TCAGATACCTTCCAAACCTGGAATGGGACTGAGCGCTCGTACGCAAGCTGATTTGTTGTGGAATGAGTTATTCGC
AAAGAGCGAGATTGCCCCTGCGGCCGAGGCTGACATTTCAGGATTTGACTGGTCTGTGCAGGAATGGGAGTTATG
GGCCGATCTGAGCATGAGGATTTCTTTGTGCGAAGATATGCATGATGGTCTACGGAGGTTGATGGTTAACAGATA
CCGCTGCTTTATGCTCTCTTGCTTCCAATTGTCGAATGGGGAATTGTATGAACAAGTTGAGCCTGGACTCATGAA
GTCTGGTTCCTACTGCACTTCCTCTTCCAACTCCAGAATCAGGTGCCTAATGGGTTATCTAATTGGAGCCCCCTG
GATAATAGCCATGGGGGATGATTCTGTTGAGGGTTATGTGAGAGACGCGAAAGGCAAGTATGAGGAATTAGGACA
CACTTGTAAGGAATACGAGTTGTGTGATGTTGATTCAGACGGCGCGTTGAGATCTGTGAACTTTTGTTCACATTT
GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAA
TTTTCATGATCTTGAAAGGGAACTTGGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGG
ACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTTGACCTGATTCTTGGTG

>Fragment 3
GGAAAATGGCGCGGAATGAGATATCTCTCATCGCCTCTAAGATTAAAGGGAAACCTTGGTATCTTCAAGAGGAAG
GTGATGATGAATTCTATGACTCGATCGGTGAGAAAGATTTCCTTGCCCGCTTCCGTGAACAAACTGGTAAAGAGA
CAGTGGGGAATTTAAACTGCCAGAGGGCGGCCCAAACATTGGAGCCGCCCTTCGAGAACTTGAGGCCGTGCGATG
GGAAGAACCCAGAGCTATTCAAACCGGCGGGATGGGATTCAACTATGTTGGAGTCTCGTCTTGCAAGTTTAGAGA
GGGCCCTAAGCACGCTGCTCGCGGAGCAATCAGTTCTGCTGAGCAAGTTTTCCCAGAACTCAAACAGTATGATTG
GCCAGAAAGAGGCTCTAAAGCCGAGTTCGATTCCCTCCTCCTCCAAGCCAGCCGTTTCCGGAGAACCTCCTGTCC
AGAGCAAACCGAGCGTAAATGCGAAGTCCTCGCTGAGAAGTACCCCAAAAGTAGAGCCCACCGTTGCTTCAGGAG
AGAAAACTTCTGTCAACGCCAGCTCCTCAGGGAACAAATCGAAGCGACGACGACGTCGCCGGAAATCAACGACAA
AGCCAGCCCCGGCTCCCCCTGGTCAAGACTTGAAAAGACCAATGGTGAGCTCATCTCCCGTTTCAAAGACCTCTT
AATTGAGGCAGTGTTGAGGCGAGTTTTATTGCTTGCCTCCACTCCAACCAGTGAGATTTTATGCATGTCTGCTTC
AGATCTCGTTAGAGCTAACTTAGTGGATCCTGTGCGATTGTTCGTCAAGCAGGAACCCCATACGAAGAAGAAGCT
GAATGAGCGTCGCTTTAGGCTTATTTCGTCTGTGTCTATAGTTGATCAGATAATCGAGAGATTATTGTTTGGTCC
TCAGAACAGGCTTGAGATAGCCTTATGGCATCAGATACCTTCCAAACCTGGAATGGGACTGAGCGCTCGTACGCA
AGCTGATTTGTTGTGGAATGAGTTATTCGCAAAGAGCGAGATTGCCCCTGCGGCCGAGGCTGACATTTCAGGATT
TGACTGGTCTGTGCAGGAATTGGAGTTATGGGCCGATCTGAGCATGAGGATTTCTTTGTGCGAAGAATTGCATGA
TGGTCTACGGAGGTTGATGATTAACGGAATCCGCTGTCTTAGCCTTCCTTGCTTCCA

>Fragment 4
AGATGTATGACAGTTGAGCCTGGACTCATGAAGTCTGGTTCCTACTGCACTTCCTCTTCCAACTCCAGAATCAGG
TGCCTAATGGGTTATCTAATTGGAGCCCCCTGGATAATAGCCATGGGGGATGATTCTGTTGAGGGTTATGTGAGA
GACGCGAAAGGCAAGTATGAGGAATTAGGACACACTTGTAAGGAATACGAGTTGTGTGATGTTGATTCAGACGGC
GCGTTGAGATCTGTGAACTTTTGTTCACATTTGATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACC
TTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTTGGCTCATGTCCCAAGTGG
GCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCT
GACTATGTCGAAGAAACTCACCAAGAATCAGGTCAAGCAGATGATACAGGCAACCTTGCCGAAAGAGCAGACTTC
TGCTCGGTCCCGACGACGGAGACGCCGAAGGTCGACGCAGCAAGGGCAGAGTTCTACAGTTATGGCCCCAATGGC
TGGAGCTGTGATATACCGGAAGCGACCCATGTTAATCAATGGCCGCTCTGGGGTTACAGTTCGTCATTCGGAGGT
GGTTTTGGTAGTCCAATCTGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTG
GTTAGCTGTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGC
TTCTACGACTCCCGGAACTGTGGCAATGGGTTTCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGG
TATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTCTAAACTCCTCGCTGAA

>Fragment 5
TTTATTGCTTGCCTCCACTCCAACCAGTGAGATTTTATGCATGTCTGCTTCAGATCTCGTTAGAGCTAACTTAGT
GGATCCTGTGCGATTGTTCGTCAAGCAGGAACCCCATACGAAGAAGAAGCTGAATGAGCGTCGCTTTAGGCTTAT
TTCGTCTGTGTCTATAGTTGATCAGATAATCGAGAGATTATTGTTTGGTCCTCAGAACAGGCTTGAGATAGCCTT
ATGGCATCAGATACCTTCCAAACCTGGAATGGGACTGAGCGCTCGTACGCAAGCTGATTTGTTGTGGAATGAGTT
ATTCGCAAAGAGCGAGATTGCCCCTGCGGCCGAGGCTGACATTTCAGGATTTGACTGGTCTGTGCAGGAATGGGA
GTTATGGGCCGATCTGAGCATGAGGATTTCTTTGTGCGAAGATATGCATGATGGTCTACGGAGGTTGATGGTTAA
CAGATACCGCTGCTTTATGCTCTCTTGCTTCCAATTGTCGAATGGGGAATTGTATGAACAAGTTGAGCCTGGACT
CATGAAGTCTGGTTCCTACTGCACTTCCTCTTCCAACTCCAGAATCAGGTGCCTAATGGGTTATCTAATTGGAGC

CCCCTGGATAATAGCCATGGGGGATGATTCTGTTGAGGGTTATGTGAGAGACGCGAAAGGCAAGTATGAGGAATT
AGGACACACTTGTAAGGAATACGAGTTGTGTGATGTTGATTCAGACGGCGCGTTGAGATCTGTGAACTTTTGTTC
ACATTTGATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTC
TGAAAATTTTCATGATCTTGAAAGGGAACTTGGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCA
GGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATGTCGAAGAAACTCACCAAGA
ATCAGGTCAAGCAGATGATACAGGCAACCTTGCCGAAAGAGCAGACTTCTGCTCGGTCCCGACGACGGAGACGCC
GAAGGTCGACGCAGCAAGGGCAGAGTTCTACAGTTATGGCCCCAATGGCTGGAGCTGTGATATACCGGAAGCGAC
CCATGTTAATCAATGGCCGCTCTGGGGTTACAGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCTGGGACTA
CTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCTGTACAAGGATCCCTCTATT
CTAAATGGAGGT

>Fragment 6
ACGAACGTTTGGTCTCTCGTAACTGGTGATTTTCAAATGGCAAAAGCCAGTAAGATGGTCTCAGTGAAGGACGCC
AAGTGTTATGTTGAGGCTGCTCACGCTAAGCTGGACTTTGCCCTAATTAAAGTTCCGAACAAATATTGGAGCTCG
ATAGGGGTTGGTTCGGCTAAATTGTTGTGGCATAAGCCCGGGCAAGTGGTCAAGGTTTACGGAGGTAGATCTGAT
GAATTGGTCTCTTCTGTGGGCCGAGCCGAAAAGGATCCTGATCTGAGCCTACGTCTGACCCATAACGCTTCAACT
GCTCCAGGATGGAGCGGCTCCCCGCTTTACAACTCTGAAAATTTCGTAGTGGGATTGCATACTGGCTTTTCCGCG
GCTGAGCAGAGAAATGAGGCGGTTGATGTAGCGAAGTTGCTGCACCTCGCCCTTAGAACAAAGGAAACGACGTTT
TCGGAGATCGGTGTCTCTCTTATCGATGAAGATGAGATTGAGTCTCGCGGTTACCAATTCGACGACTTTGAGCTT
AGGGGTGAGGTAAACGTTAAAGGGAAAATGGCGCGGAATGAGATATCTCTCATCGCCTCTAAGATTAAAGGGAAA
CCCTGGTATCTTCAAGAGGAAGGTGATGATGAATTCTATGACTCGATCCGTGAGAAAGA

>Fragment 7
GGAATTCCGGGGGGGGCTCTAGAGTTAATTGGCAAGGCGCAGATTACTTGTTGACAGCCTCCCACGTTTGGTCCC
TCGTAACTGGTGATTTTCAAATGGCAAAAGCCAGTAAGATGGTCTCAGTGAAGGACGCCAAGTGTTATGTTGAGG
CTGCTCACGCTAAGCTGGACTTTGCCCTAATTAAAGTTCCGAACAAATATTGGAGCTCGATAGGGGTTGGTTCGG
CTAAATTGTTGTGGCATAAGCCCGGGCAAGTGGTCAAGGTTTACGGAGGTAGATCTGATGAATTGGTCTCTTCTG
TGGGCCGAGCCGAAAAGGATCCTGATCTGAGCCTACGTCTGACCCATAACGCTTCAACTGCTCCAGGATGGAGCG
GCTCCCCGCTTTACAACTCTGAAAATTTCGTAGTGGGATTGCATACTGGCTTTTCCGCGGCTGAGCAGAGAAATG
AGGCGGTTGATGTAGCGAAGTTGCTGCACCTCGCCCTTAGAACAAAGGAAACGACGTTTTCGGAGATCGGTGTCT
CTCTTATCGATGAAGATGAGATTGAGTCTCGCGGTTACCAATTCGACGACTTTGAGCTTAGGGGTGAGGTAAACG
TTAAAGGGAAAATGGCGCGGAATGAGATATCTCTCATCGCCTCTAAGATTAAAGGGAAACCTTGGCATCTTCAAG
AGGAAGGTGATGATGAATTCTATGACTCGATCCGTGCGAAAGAGCCCCCCCCCGGACTTCCCA

>Fragment 8
CGGCCGCGAATTCGCCCTTGGAATTCCGGGGGGGGCATCTCTGGATCGTCCAGTTCTCCCCAGACTAACTAGGAC
TCGGTTCCGTTTATTTTCTTACAAATCATTTAATTGTTTTGTAATTGAGAAAGATGTTGAGCGAGTTAGTCCAGT
TGTGTCTTTCAGCAACCATGACTCTAGCGCTGGTAGTGTCAATGATGCTGGACCCGAGCTACGGGTGGCTGATCA
CCATGTCTGTGATGTTAGTCTCCTTAGAGCTGACAATCAGGCCGTTCAGGCGCTCTCTAGACTACATGAAAATTG
TGGTCCGAGATGAGCCCAATGAGCCTGAAGTTGCTAGGATAGTAGGCGCGCCAACTTACCACCCAATTCATGGCG
TTGTTTGCAAGTGTCTGTGGAACTCGAAAGAGTTCAATGTGGTTGTGCAACCCAGTTTGTGGCCTCTAGCCTTTC
CCCCAAACATATCACAAGGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAG
CTGAACCAAAGTCTCTCGTTGTCTTATACAACGATGGAGTCAGAATCGGTATGGGCTCTAGAGTTAATTGGCAAG
GCGCAGATTACTTGTTGACAGCCTCCCACGTTTGGTCCCTCGTAACTGGTGATTTTCAAATGGCAAAAGCCAGTA
AGATGGTCTCAGTGAAGGACGCCAAGTGTTATGTTGAGGCTGCTCACGCTAAGCTGGACTTTGCCCTAATTAAAG
TTCCGAACAAATATTGGAGCTCGATAGGGGTTGGTTCGGCTAAATTGTTGTGGCATAAGCCCGGGCAAGTGGTCA
AGGTTTACGGAGGTAGATCTGATGAATTGGTCTCTTCTGTGGGCCGAGCCGAAAAGGATCCTGATCTGAGCCTAC
GTCTGACCCATAACGCTTCAACTGCTCCAGGATGGAGCGGCTCCCCGCTTTACAACTCTGAAAATTTCGTAGTGG
GATTGCATACTGGCTTTTCCGCGGCTGAGCAGAGAAATGAGGCGGTTGATGTAGCGAAGTTGCTGCACCTCGCCC
TTAGAACAAAGGAAACGACGTTTTCGGAGATCGGGGTCTCTCTTATCGATGAAGATGAGATTGAGTCTCGCGGTT
ACCAATTCGACGACTTTGAGCTTAGGGGTGAGGTAAACGTTAAAGGGAAAATGGCGCGGCATGAGATATCTCTCA
TCGCCTCTAAGATAAAGGGAAA

>Fragment 9
AGCACGATCTCGCGCTATTGTTCTCTGGTTAGCTATATAGGATCCCTCTATTCTATTGGATGTGGATATCGCTTA
GAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTTTCCAGTATGACAATACTG
ATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTA
GTGGTTCTAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTC
AAAATTTCGGTCTCAAATGGTATCAGTACAAGTCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAG
CTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGCACATCCAGTATGAGATTG
AGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGT
CAGA

>Fragment 10
TATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGA
TGAAAGGGAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCGATATAACGCCACCATTGACGTC
TGGAGAGACTTAAATGGGTTTTGCATAACTACTCCGGTTGCTACGTCATAGCACGTTACCAGAAGTTCTCTTGGT
CCCTCTACATGGTTGGGTGAACACCATGTGTTCAT

>Fragment 11
CGAAGGATTTGTCTGCACTCTTTACGGAGAGTACGATCACGACTACGTCACCAGTGTGCATTTGCACATTGTTTG
CTCGTGCGGTAGAGCTTTCTTTGATTTTGTTGAGTTCAAGGACATAAAACTCCGAGATTTCCAATATCAAAGTCA
GTGTATCTGTCCCACCAGGAATTGGTCGACGGTCGCTTACACTGATTTTACTACCGTCAACTGCCAAGTGGAAGG
TTGTGATTACTGTGAGGGGATTGAGTCAGACTCTGATTCGGATTCTGAGGCCATCATAGAGGAATTTCTTCAAAA
GTTTTCCGAGGTTGGCATCTCTGGATCGTCCAGTTCTCCCCAGACTAACTAGGACTCGGTTCCGTTTATTTTCTT
ACAAATCATTTAATTGTTTTGTAATTGAGAAAGATGTTGAGCGAGTTAGTCCAGTTGTGTCA

>Fragment 12
ACAAAATATATTGAAGGGAAAGTCTTGCACTCAGTACCCATTATATTTTGATATGCCCAGCATTGATGTTGAAGT
AGAAAAGATCCTGCACTTGAGCAACAGAAAGAGACTAAGGTCTGTCTGTGTGGCGAGGAAGAAGACGTTCGTAGA
CAAGATTCCGAAGGATTTGTCTGCACTCTTTACGGAGAGTACGATCACGACTACGTCACCAGTGTGCATTTGCAC
ATTGTTTGCTCGTGCGGTAGAGCTTTCTTTGATTTTGTTGAGTTCAAGGACATAAAACTCCGAGATTTCCAATAT
CAAAGTCAGTGTATCTGTCCCACCAGGAATTGGTCGACGGTCGCTTACACTGATTTTACTACCGTCAACTGCCAA
GTGGAAA

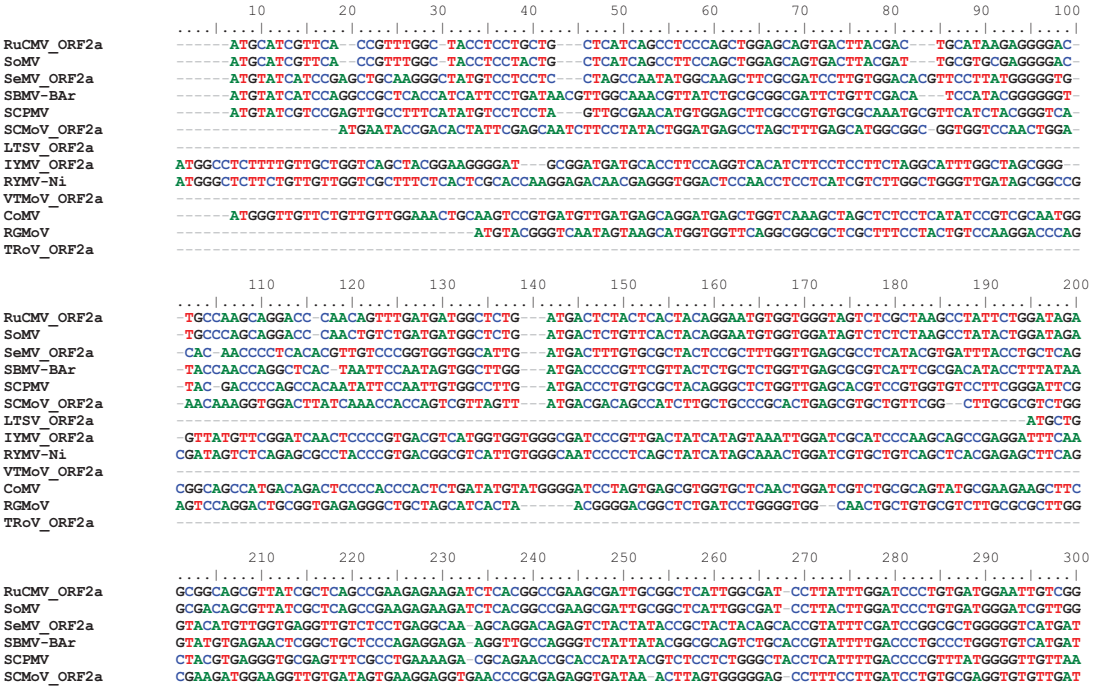
Appendix 6: I-TASSER output of VTMoV-P1 analysis

Table A6.1 Summary of I-TASSER output for P1 protein analysis. Key:*-protein structures from RCSB PDB [www.pdb.org (Berman et al. 2000)], and: ^ are terms from the Gene Ontology Consortium (Ashburner et al. 2000).

3D protein structure	
3D Model assessment	Confidence (c)-score - 2.08 [ranges from -5 to 2 and models with c-scores > -1.5, have a correct fold. Confidence of structure prediction (TM score) TM score 0.47 ± 0.15. With a TM score > 0.5 model of correct topology < 0.17 random score
Threading templates	Only threading normalised z score >1 is: PDB ID:1j74A*(human ubiquitin conjugating enzyme complex, hMms2-hUbc13 (Moraes et al. 2001)) PDB ID: 3figB*[Rat Odorant Binding Protein OBP1 (White et al. 2009)] was top threading template for 2 programs. Little homology between P1 and proteins used as templates in threading alignments
Structural analogues for 3D model	All structural analogues had TM scores between 0.6631 - 0.6226. Detected structural analogues with a TM score > 0.5 can be used for determining the structure class /protein family of predicted query. Structural analogues had a range of molecular classifications including, transporters, allergens, and Lipid and/or retinol/retinoic acid binding proteins. All matches belonged to Pfam (Finn et al. 2010) classification: PF0061.
Protein function prediction	
Enzyme	All Enzyme commission (EC) numbers <1.1. P1 is either not an enzyme or prediction confidence is too low to ascribe function.
Gene Ontology (GO) terms^	All GO scores > 0.5 suggest a prediction with high confidence Molecular function (0005488-binding); The selective, non-covalent, often stoichiometric, interaction of a molecule with one or more specific sites on another molecule (0005215- Transporter activity); Enables the directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within or between cells Biological process (0006810 transport); The directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within or between cells, or within a multicellular organism by means of some external agent such as a transporter or pore. Cellular location (0005576 extracellular region); The space external to the outermost structure of a cell. For cells without external protective or external encapsulating structures this refers to space outside of the plasma membrane. This term covers the host cell environment outside an intracellular parasite. Note that this term is intended to annotate gene products not uniformly attached to cell surface.
Functional analogues for ligand-binding sites	Determines functional analogues with similar ligand binding sites, for matches with BS scores >0.5, and when predicted binding site residues are clustered close together. Ligands bound by functional analogues; Heme, Retinoic acid, Fluorescein, Palmitic acid, Biliverdin IX gamma chromophore, and 3,6 -Bis(Methylene) Decanoic acid.

Appendix 7: Multiple sequence alignments for Figure 4.10A-C

Figure 4.10A




LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
310
320
330
340
350
360
370
380
390
400
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
410
420
430
440
450
460
470
480
490
500
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
510
520
530
540
550
560
570
580
590
600
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a


RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
610 620 630 640 650 660 670 680 690 700
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
710 720 730 740 750 760 770 780 790 800
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
810 820 830 840 850 860 870 880 890 900
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a


CoMV
RGMoV
TRoV_ORF2a
910 920 930 940 950 960 970 980 990 1000
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a


1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500


TRoV_ORF2a
RuCMV_ORF2a
SoMV
SeMV_ORF2a
SBMV-BAr
SCPMV
SCMoV_ORF2a
LTSV_ORF2a
IYMV_ORF2a
RYMV-Ni
VTMoV_ORF2a
CoMV
RGMoV
TRoV_ORF2a
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

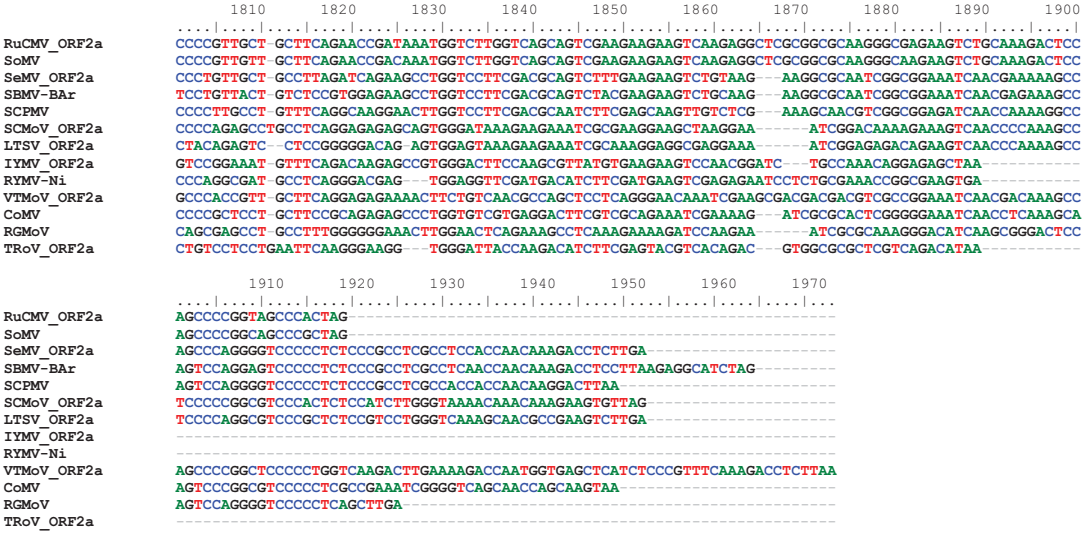
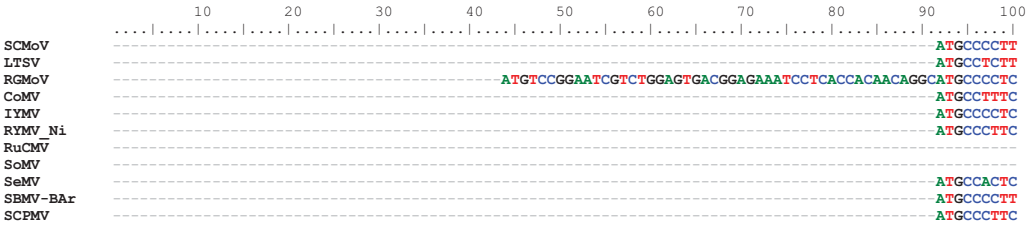
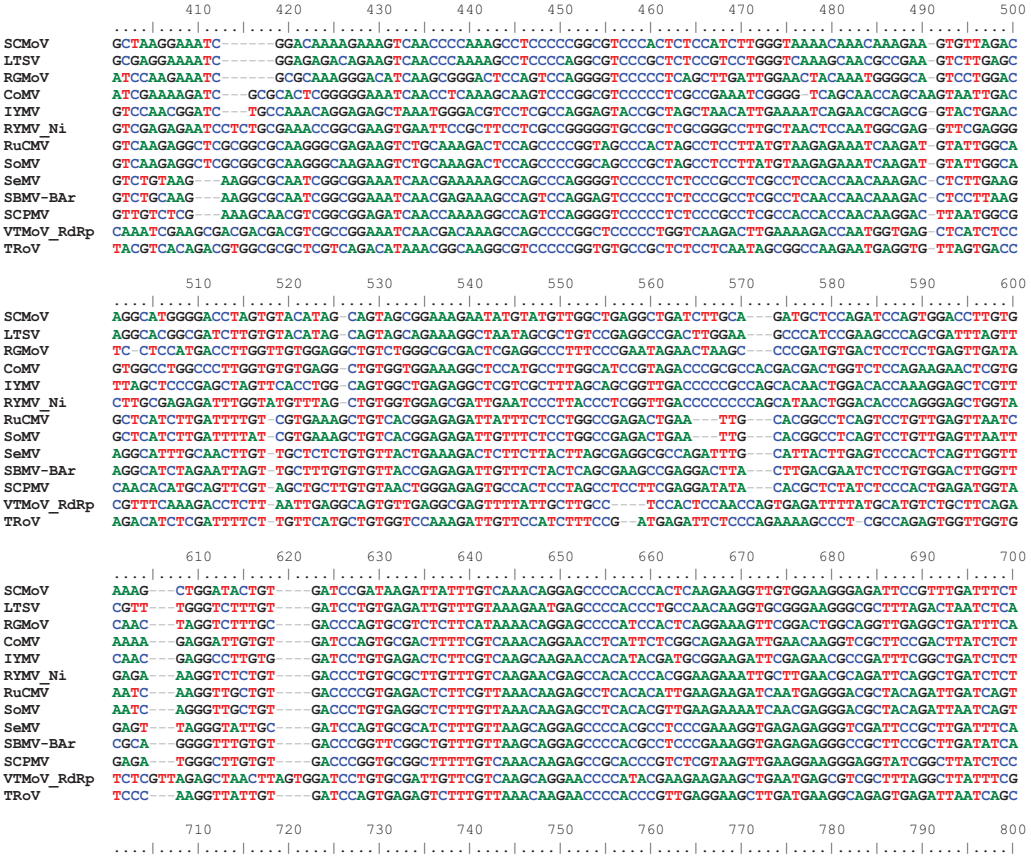

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
RuCMV_ORF2a CCCCGTTGCT-GCTTCAGAACCGATAAATGGTCTTGGTCAGCAGTCGAAGAAGAAGTCAAGAGGCTCGCGGCGCAAGGGCGAGAAGTCTGCAAAGACTCC
SoMV CCCCGTTGTT-GCTTCAGAACCGACAAATGGTCTTGGTCAGCAGTCGAAGAAGAAGTCAAGAGGCTCGCGGCGCAAGGGCAAGAAGTCTGCAAAGACTCC
SeMV_ORF2a CCCTGTTGCT-GCCTTAGATCAGAAGCCTGGTCCTTCGACGCAGTCTTTGAAGAAGTCTGTAAG---AAGGCGCAATCGGCGGAAATCAACGAAAAAGCC
SBMV-BAr TCCTGTTACT-GTCTCCGTGGAGAAGCCTGGTCCTTCGACGCAGTCTACGAAGAAGTCTGCAAG---AAGGCGCAATCGGCGGAAATCAACGAGAAAGCC
SCPMV CCCCTTGCCT-GTTTCAGGCAAGGAACTTGGTCCTTCGACGCAATCTTCGAGCAAGTTGTCTCG---AAAGCAACGTCGGCGGAGATCAACCAAAAGGCC
SCMoV_ORF2a CCCCAGAGCCTGCCTCAGGAGAGAGCAGTGGGATAAAGAAGAAATCGCGAAGGAAGCTAAGGAA-----ATCGGACAAAAGAAAGTCAACCCCAAAGCC
LTSV_ORF2a CTACAGAGTC--CTCCGGGGGACAG-AGTGGAGTAAAGAAGAAATCGCAAAGGAGGCGAGGAAA-----ATCGGAGAGACAGAAGTCAACCCAAAAGCC
IYMV_ORF2a GTCCGGAAAT-GTTTCAGACAAGAGCCGTGGGACTTCCAAGCGTTATGTGAAGAAGTCCAACGGATC---TGCCAAACAGGAGAGCTAA----------
RYMV-Ni CCCAGGCGAT-GCCTCAGGGACGAG---TGGAGGTTCGATGACATCTTCGATGAAGTCGAGAGAATCCTCTGCGAAACCGGCGAAGTGA----------
VTMoV_ORF2a GCCCACCGTT-GCTTCAGGAGAGAAAACTTCTGTCAACGCCAGCTCCTCAGGGAACAAATCGAAGCGACGACGACGTCGCCGGAAATCAACGACAAAGCC
CoMV CCCCGCTCCT-GCTTCCGCAGAGAGCCCTGGTGTCGTGAGGACTTCGTCGCAGAAATCGAAAAG---ATCGCGCACTCGGGGGAAATCAACCTCAAAGCA
RGMoV CAGCGAGCCT-GCCTTTGGGGGGAAACTTGGAACTCAGAAAGCCTCAAAGAAAAGATCCAAGAA-----ATCGCGCAAAGGGACATCAAGCGGGACTCC
TRoV_ORF2a CTGTCCTCCTGAATTCAAGGGAAGG---TGGGATTACCAAGACATCTTCGAGTACGTCACAGAC---GTGGCGCGCTCGTCAGACATAA----------
1910 1920 1930 1940 1950 1960 1970
RuCMV_ORF2a AGCCCCGGTAGCCCACTAG----------
SoMV AGCCCCGGCAGCCCGCTAG----------
SeMV_ORF2a AGCCCAGGGGTCCCCCTCTCCCGCCTCGCCTCCACCAACAAAGACCTCTTGA----------
SBMV-BAr AGTCCAGGAGTCCCCCTCTCCCGCCTCGCCTCAACCAACAAAGACCTCCTTAAGAGGCATCTAG----------
SCPMV AGTCCAGGGGTCCCCCTCTCCCGCCTCGCCACCACCAACAAGGACTTAA----------
SCMoV_ORF2a TCCCCCGGCGTCCCACTCTCCATCTTGGGTAAAACAAACAAAGAAGTGTTAG----------
LTSV_ORF2a TCCCCAGGCGTCCCGCTCTCCGTCCTGGGTCAAAGCAACGCCGAAGTCTTGA----------
IYMV_ORF2a ----------
RYMV-Ni ----------
VTMoV_ORF2a AGCCCCGGCTCCCCCTGGTCAAGACTTGAAAAGACCAATGGTGAGCTCATCTCCCGTTTCAAAGACCTCTTAA
CoMV AGTCCCGGCGTCCCCCTCGCCGAAATCGGGGTCAGCAACCAGCAAGTAA----------
RGMoV AGTCCAGGGGTCCCCCTCAGCTTGA----------
TRoV_ORF2a ----------

Figure 4.10B


10 20 30 40 50 60 70 80 90 100
SCMoV ATGCCCCTT
LTSV ATGCCTCTT
RGMoV ATGTCCGGAATCGTCTGGAGTGACGGAGAAATCCTCACCACAACAGGCATGCCCCTC
CoMV ATGCCTTTC
IYMV ATGCCCCTC
RYMV_Ni ATGCCCTTC
RuCMV
SoMV
SeMV ATGCCACTC
SBMV-BAr ATGCCCCTT
SCPMV ATGCCCTTC


VTMoV_RdRp
TRoV
SCMoV
LTSV
RGMoV
CoMV
IYMV
RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV
SCMoV
LTSV
RGMoV
CoMV
IYMV
RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV
SCMoV
LTSV
RGMoV
CoMV
IYMV
RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV


410 420 430 440 450 460 470 480 490 500
SCMoV GCTAAGGAAATC------GGACAAAAGAAAGTCAACCCCAAAGCCTCCCCCGGCGTCCCACTCTCCATCTTGGGTAAAACAAACAAAGAA-GTGTTAGAC
LTSV GCGAGGAAAATC------GGAGAGACAGAAGTCAACCCAAAAGCCTCCCCAGGCGTCCCGCTCTCCGTCCTGGGTCAAAGCAACGCCGAA-GTCTTGAGC
RGMoV ATCCAAGAAATC------GCGCAAAGGGACATCAAGCGGGACTCCAGTCCAGGGGTCCCCCTCAGCTTGATTGGAACTACAAATGGGGCA-GTCCTGGAC
CoMV ATCGAAAAGATC---GCGCACTCGGGGGAAATCAACCTCAAAGCAAGTCCCGGCGTCCCCCTCGCCGAAATCGGGG-TCAGCAACCAGCAAGTAATTGAC
IYMV GTCCAACGGATC---TGCCAAACAGGAGAGCTAAATGGGACGTCCTCGCCAGGAGTACCGCTAGCTAACATTGAAAATCAGAACGCAGCG-GTACTGAAC
RYMV_Ni GTCGAGAGAATCCTCTGCGAAACCGGCGAAGTGAATTCCGCTTCCTCGCCGGGGGTGCCGCTCGCGGGCCTTGCTAACTCCAATGGCGAG-GTTCGAGGG
RuCMV GTCAAGAGGCTCGCGGCGCAAGGGCGAGAAGTCTGCAAAGACTCCAGCCCCGGTAGCCCACTAGCCTCCTTATGTAAGAGAAATCAAGAT-GTATTGGCA
SoMV GTCAAGAGGCTCGCGGCGCAAGGGCAAGAAGTCTGCAAAGACTCCAGCCCCGGCAGCCCGCTAGCCTCCTTATGTAAGAGAAATCAAGAT-GTATTGGCA
SeMV GTCTGTAAG---AAGGCGCAATCGGCGGAAATCAACGAAAAAGCCAGCCCAGGGGTCCCCCTCTCCCGCCTCGCCTCCACCAACAAAGAC-CTCTTGAAG
SBMV-BAr GTCTGCAAG---AAGGCGCAATCGGCGGAAATCAACGAGAAAGCCAGTCCAGGAGTCCCCCTCTCCCGCCTCGCCTCAACCAACAAAGAC-CTCCTTAAG
SCPMV GTTGTCTCG---AAAGCAACGTCGGCGGAGATCAACCAAAAGGCCAGTCCAGGGGTCCCCCTCTCCCGCCTCGCCACCACCAACAAGGAC-TTAATGGCG
VTMoV_RdRp CAAATCGAAGCGACGACGACGTCGCCGGAAATCAACGACAAAGCCAGCCCCGGCTCCCCCTGGTCAAGACTTGAAAAGACCAATGGTGAG-CTCATCTCC
TRoV TACGTCACAGACGTGGCGCGCTCGTCAGACATAAACGGCAAGGCGTCCCCCGGTGTGCCGCTCTCCTCAATAGCGGCCAAGAATGAGGTG-TTAGTGACC
510 520 530 540 550 560 570 580 590 600
SCMoV AGGCATGGGGACCTAGTGTACATAG-CAGTAGCGGAAAGAATATGTATGTTGGCTGAGGCTGATCTTGCA---GATGCTCCAGATCCAGTGGACCTTGTG
LTSV AGGCACGGCGATCTTGTGTACATAG-CAGTAGCAGAAAGGCTAATAGCGCTGTCCGAGGCCGACTTGGAA---GCCCATCCGAAGCCCAGCGATTTAGTT
RGMoV TC-CTCCATGACCTTGGTTGTGGAGGCTGTCTGGGCGCGACTCGAGGCCCTTTCCCGAATAGAACTAAGC---CCCGATGTGACTCCTCCTGAGTTGATA
CoMV GTGGCCTGGCCCTTGGTGTGTGAGG-CTGTGGTGGAAAGGCTCCATGCCTTGGCATCCGTAGACCCGCGCCCACGACGACTGGTCTCCAGAAGAACTCGTG
IYMV TTAGCTCCCGAGCTAGTTCACCTGG-CAGTGGCTGAGAGGCTCGTCGCTTTAGCAGCGGTTGACCCCCGCCAGCACAACTGGACACCAAAGGAGCTCGTT
RYMV_Ni CTTGCGAGAGATTTGGTATGTTTAG-CTGTGGTGGAGCGATTGAATCCCTTACCCTCGGTTGACCCCCCCCAGCATAACTGGACACCCAGGGAGCTGGTA
RuCMV GCTCATCTTGATTTTGT-CGTGAAAGCTGTCACGGAGAGATTATTTCTCCTGGCCGAGACTGAA---TTG---CACGGCCTCAGTCCTGTTGAGTTAATC
SoMV GCTCATCTTGATTTTAT-CGTGAAAGCTGTCACGGAGAGATTGTTTCTCCTGGCCGAGACTGAA---TTG---CACGGCCTCAGTCCTGTTGAGTTAATT
SeMV AGGCATTTGCAACTTGT-TGCTCTCTGTGTTACTGAAAGACTCTTCTTACTTAGCGAGGCGCCAGATTTG---CATTACTTGAGTCCCACTCAGTTGGTT
SBMV-BAr AGGCATCTAGAATTAGT-TGCTTTGTGTGTTACCGAGAGATTGTTTCTACTCAGCGAAGCCGAGGACTTA---CTTGACGAATCTCCTGTGGACTTGGTT
SCPMV CAACACATGCAGTTCGT-AGCTGCTTGTGTAACTGGGAGAGTGCCACTCCTAGCCTCCTTCGAGGATATA---CACGCTCTATCTCCCACTGAGATGGTA
VTMoV_RdRp CGTTTCAAAGACCTCTT-AATTGAGGCAGTGTTGAGGCGAGTTTTATTGCTTGCC----TCCACTCCAACCAGTGAGATTTTATGCATGTCTGCTTCAGA
TRoV AGACATCTCGATTTTCT-TGTTCATGCTGTGGTCCAAAGATTGTTCCATCTTTCCG---ATGAGATTCTCCCAGAAAAGCCCT-CGCCAGAGTGGTTGGTG
610 620 630 640 650 660 670 680 690 700
SCMoV AAAG---CTGGATACTGT----GATCCGATAAGATTATTTGTCAAACAGGAGCCCCACCCACTCAAGAAGGTTGTGGAAGGGAGATTCCGTTTGATTTCT
LTSV CGTT---TGGGTCTTTGT----GATCCTGTGAGATTGTTTGTAAAGAATGAGCCCCACCCTGCCAACAAGGTGCGGGAAGGGCGCTTTAGACTAATCTCA
RGMoV CAAC---TAGGTCTTTGC----GACCCAGTGCGTCTCTTCATAAAACAGGAGCCCCATCCACTCAGGAAAGTTCGGACTGGCAGGTTGAGGCTGATTTCA
CoMV AAAA---GAGGATTGTGT----GATCCAGTGCGACTTTTCGTCAAACAGGAACCTCATTCTCGGCAGAAGATTGAACAAGGTCGCTTCCGACTTATCTCT
IYMV CAAC---GAGGCCTTGTG----GATCCTGTGAGACTCTTCGTCAAGCAAGAACCACATACGATGCGGAAGATTCGAGAACGCCGATTTCGGCTGATCTCT
RYMV_Ni GAGA---AAGGTCTCTGT----GACCCTGTGCGCTTGTTTGTCAAGAACGAGCCACACCCACGGAAGAAATTGCTTGAACGCAGATTCAGGCTGATCTCT
RuCMV AATC---AAGGTTGCTGT----GACCCCGTGAGACTCTTCGTTAAACAAGAGCCTCACACATTGAAGAAGATCAATGAGGGACGCTACAGATTGATCAGT
SoMV AATC---AGGGTTGCTGT----GACCCTGTGAGGCTCTTTGTTAAACAAGAGCCTCACACGTTGAAGAAAATCAACGAGGGACGCTACAGATTAATCAGT
SeMV GAGT---TAGGGTATTGC----GATCCAGTGCGCATCTTTGTTAAGCAGGAGCCCCACGCCTCCCGAAAGGTGAGAGAGGGTCGATTCCGCTTGATTTCA
SBMV-BAr CGCA---GGGGTTTGTGT----GACCCGGTTCGGCTGTTTGTTAAGCAGGAGCCCCACGCCTCCCGAAAGGTGAGAGAGGGCCGCTTCCGCTTGATATCA
SCPMV GAGA---TGGGCTTGTGT----GACCCGGTGCGGCTTTTTGTCAAACAAGAGCCGCACCCGTCTCGTAAGTTGAAGGAAGGGAGGTATCGGCTTATCTCC
VTMoV_RdRp TCTCGTTAGAGCTAACTTAGTGGATCCTGTGCGATTGTTCGTCAAGCAGGAACCCCATACGAAGAAGAAGCTGAATGAGCGTCGCTTTAGGCTTATTTCG
TRoV TCCC---AAGGTTATTGT----GATCCAGTGAGAGTCTTTGTTAAACAAGAACCCCACCCGTTGAGGAAGCTTGATGAAGGCAGAGTGAGATTAATCAGC
710 720 730 740 750 760 770 780 790 800


SCMoV TCAGTATCACTAGTTGATCAACTAGTTGAGAGACTGCTCTTCGGCCCACAAAACGAAACCGAAATCGATCTCTGGCAATCAGTCCCGTCCAAACCCGGAA
LTSV TCGGTATCGCTGGTTGACCAACTAGTTGAAAGGCTGCTATTCGGGCCCCAAAACAAGGCAGAGATTGCCTTGTGGCAGTCGGTGCCTTCAAAACCTGGTA
RGMoV TCGGTATCCCTGTTGGACCAACTGGTTGAGAGGGTTCTTTTCGGTTTTCAGAATAACCTCGAGATAAGCCGATGGAAGCAGTGTCCATCGAAACCCGGAA
CoMV TCCGTCTCACTCGTAGATCAACTGGTAGAACGCATGCTTTTCGGACCCCAGAATACTACAGAAATTGCTCTGTGGCATTCAAATCCCTCTAAACCCGGGA
IYMV TCTGTGTCCTTGGTGGACCAGCTCGTGGAGCGGATGTTGTTTGGCCCACAGAATGCTACAGAGATATCGATGTGGCACCTTTGTCCGTCGAAACCTGGCA
RYMV_Ni TCTGTTTCTCTAGTCGACCAGCTGGTGGAGCGGATGCTTTTCGGGCCCCAGAACAACACCGAGATTTCGACTTGGTGGCAGTGGCCCTCGAAGCCGGGCA
RuCMV TCAGTAAGTTTGGTAGACCAGATCGTCGAGCGCATGCTCTTCGGGCCTCAGAACCGAGCTGAAATAGCTCTCTGGGATACCATTCCTTCGAAACCCGGCA
SoMV TCAGTGAGTTTGGTGGACCAGATCGTCGAGCGCATGCTCTTCGGGCCTCAGAACCGAGCTGAAATAGCTCTCTGGGACACCATTCCTTCGAAACCCGGCA
SeMV TCAGTCTCGTTGGTTGATCAGCTAGTTGAGCGCATGCTCTTCGGGCCGCAAAATCAGCTTGAGATCGCTGAGTGGGAGAGAATTCCGTCTAAGCCTGGGA
SBMV-BAr TCTGTTTCGCTGGTCGACCAGCTTGTAGAGCGTATGCTCTTTGGGCCTCAAAACCAGCTTGAAATCGCTGAGTGGGAGCATATTCCCTCAAAACCTGGCA
SCPMV TCGGTCTCAATCGTCGACCAATTGGTCGAAAGAATGCTCTTTGGAGCTCAGAACGAGTTAGAGATTGCTGAGTGGCAATCTATCCCTTCAAAACCCGGTA
VTMoV_RdRp TCTGTGTCTATAGTTGATCAGATAATCGAGAGATTATTGTTTGGTCCTCAGAACAGGCTTGAGATAGCCTTATGGCATCAGATACCTTCCAAACCTGGAA
TRoV TCGGTGTCTCTAGTAGACCAGTTAGTTGAAAGGGTCTTGTTCGGCCGCCAAAACCGGAAAGAAATCACACAATGGAAGAGTATCCCGTCGAAACCCGGAA
810 820 830 840 850 860 870 880 890 900
SCMoV TGGGGCTGT-CGCAACCGTGGCAGGTGTCCGCTCTGTGGAATGACTTAGAGCATAAGCATAAGCTCAGTCCTG---CAGCAGAGGCAGACATCTCCGGCT
LTSV TGGGATTGT-CACAGAAGTGGCAGTTCGAGGCGCTGTGGAAGGATTTGCAAATCAAACACGCCAGTGCCCCCG---CCGCTGAGGCGGATATATCGGGTT
RGMoV TGGGTCTCA-CTTCTAAGGAGCAATCCGATGCGTTGTGGGACGAGCTGAAGTTCAAGTCCACCCTAGCCCCCG---CGGCTGAGGCGGACATATCGGGAT
CoMV TGGGTCTGT-CGAAAGCTAGCCAAGTCGCGCTGTTATGGGAGGATTTGGCGCGTACGAACCAGACCCACCCGG---GTGCCATGGCTGACATCTCAGGGT
IYMV TGGGCATGA-GCACTCCAAGTCAAGTAGAGATGCTTTGGAAGGATGTCGCTCACAAGCATTCCCTGCACCAAG---CTGCTGAAGCCGACATTTCGGCGT
RYMV_Ni TGGGTCTTC-TGACCCCTGAACAAATCCGTTTGGTGTGGGACGACGTGTTCCAGAAGCACCAGGCCCACCCTG---CGGCTGAGGCAGACATATCGGGGT
RuCMV TGGGGCTAA-CGCTTAAAAGTCAAGCGAGAAAGATCTTTGGTGATCTACTCGTCAAGCACACCCATTGCCCGG---CGTATGAAGCCGATATATCGGGTT
SoMV TGGGGCTAA-CGCTTAAAAGTCAGGCGAGAAAGATCTTTGGTGATCTACTCGTCAAGCACACCCATTGCCCAG---CATATGAAGCTGATATATCGGGTT
SeMV TGGGCCTGT-CACTTGATAGGCAAGCTCGCAGCTTGTTTGGCGACTTGAGAGTTAAACACTCTCGTTGCCCTG---CTGCTGAGGCTGACATCTCGGGTT
SBMV-BAr TGGGCTTAT-CGCTGCGGCAACAGGCCAAAAGCCTGTTTGACGACTTGAGAGTTAAACACTCCCGTTGTCCTG---CAGCTGAAGCTGACATTTCAGGTT
SCPMV TGGGGCTTT-CCGTCATCCACCAAGCTGACGCGATATTCCGTGACTTGCGAGTCAAACATACCGTGTGTCCTG---CAGCTGAAGCAGACATATCTGGTTT
VTMoV_RdRp TGGGACTGAGCGCTCG-TACGCAAGCTGATTTGTTGTG-GAATGAGTTATTCGCAAAGAGCGAGATTGCCCCT--GCGGCCGAGGCTGACATTTCAGGAT
TRoV TGGGCCTGT-CGCTGACTGAACAAATGAAGTCAGTGTTTGAACAGGTGA---GTAAATTAGCAGCTAGCCGTGAGGCAGCTGAAGCTGATATTTCCGGTT
910 920 930 940 950 960 970 980 990 1000
SCMoV TTGATTGGTCAGTCCAATCATGGGAGATCTTAGCTGATGTGTGCATTAG---GATAGATAGAGGGGGTTTTAAAGGTAACCTCCGTAAAGCCGCCCTG-AA
LTSV TTGATTGGTCTGTGCAGAAATGGGAGTTAGAAGCTGATGTGTGCATGAG---AATAGAAAGGGGAAACTTCCCTGCAAGAATGAGGAAAGCCGCCTTG-AA
RGMoV TCGATTGGTCAGTGCAACATTGGGAGCTAATGGCGGATGTCGAAATGAG---AATTCGACTCGGAGACTTCAGCAACCTTGCTGCAAGGGCTGCTAGG-AA
CoMV TTGACTGGTCCGTTCAGGATTGGGAACTCTGGGCTGACGTATCTATGAG---GATTGAGCTAGGCTCGTTTCCAGCGTTGATGGCGAAGGCAGCTATA-TC
IYMV TCGATTGGTCAGTCCAAGATTGGGAGTTATGGGCTGATCTTGCCATACG---CCTGCAATTGGGATCATTCCCAGATCTCATGAGGAGAGCCGCGATT-TC
RYMV_Ni TTGACTGGTCCGTTCAGGACTGGGAGTTGTGGTCAGACCTCGCGATCCG---TATAAACAGAGGGAATTTCCAGGGAAACCTCAGGAGAGCGGCTATC-AG
RuCMV TCGACTGGACTGTCCAGGACTGGGAACTGTGGGCTGACGTTGAGATTAG---AATCAAGTTGTGCAGTGCTGGAGTTAACCTCGCACAATGTATGCGC-AA
SoMV TCGACTGGACTGTCCAGGACTGGGAACTGTGGGCTGACGTTGAGATTAG---AATCAAGCTGTGCAGTGCTGGAGTTAACCTCGCACAATGCATGCGC-AA
SeMV TTGACTGGTCTGTGCAAGACTGGGAGCTGTGGGCTGATGTTGAGATGAG---AGTAGTTCTAGGCGGCTTTGGCCAGACATTGGCTAGGGCTGCTAGG-AA
SBMV-BAr TTGATTGGTCTGTTCAGGACTGGGAGTTGTGGGCTGATGTAGAGATGAG---AATTGTTCTGGGGGGCTTTGGCCATAAACTGGCTAAAGCCGCCCAG-AA
SCPMV TTGATTGGTCCGTTCAGGACTGGGAATTGTGGGCTGATGTTGAAATGAG---AATCGTTTTGGGCTCTTTCCCACCCATGATGGCTAGAGCTGCTAGG-AA
VTMoV_RdRp TTGACTGGTCTGTGCAGGAATGGGAGTTATGGGCCGATCTGAGCATGAGGATTTCTTTGTGCGAAGATATGCATGATGGTCTACGGAGGTTGATGGTTAA
TRoV TCGATTGGTCTGTGCAAGAGTGGGAATTGGAAATGGACTTGGAGGTTCG---CCTCCGGTTAGGAAACTTTCCCCCTAAGTTGGAGTTAGCCGCTCGA-AA
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
SCMoV TAGGTTCAAATGTTTCTCCAATGCTGTTTTCCAACTCAGCGACGGAACATTGATAAGTCAAGGCTTACCGGGTCTGATGAAATCTGGAAGTTATTGTACC
LTSV CCGCTTCAAGTGTTTCGCCAACGCTGTGTTCCAGCTAAGCAACGGCGAGTTAATAGAGCAAGGCTTACCGGGTCTAATGAAGTCTGGTAGTTACTGTACC


RGMoV CAGGTTTACTTGTTTGTCGAACTCGGTTTTCCAACTGAGTGATGGCACCCTGATTTCTCAGGGACTACCAGGCCTAATGAAGTCCGGGTCTTATTGCACT
CoMV GCGGTTTTACTGCTTGATGAATGCTACCTTCCAATTAACGAATGGAGAGCTCTTGACCCAGGAACTGCCCGGTCTCATGAAGTCCGGGTCGTACTGTACC
IYMV ACGCTTTTACTGTTTCATGAACTCCGTGTTCCAGCTCAGTGATGGGACCTTGATAGAGCAAAACCTGCCAGGTCTCATGAAATCCGGTTCATATTGCACA
RYMV_Ni CCGCTACTACTGTTTTATGAACTCTGTTTTCCAACTCTCAGATGGAACTCTCATCCAGCAGGAATTGCCCGGACTGATGAAGTCTGGCTCTTACTGCACC
RuCMV TCGATTCTATTGCTTCATGAATTCAGTTTTCCAGCTCTCCGATGGAACACTCATCCAGCAGTGTTCGCCGGGTGTTATGAAGTCAGGATCGTACTGTACT
SoMV TCGATTCTATTGCTTCATGAATTCAGTTTTCCAGCTCTCCGACGGAACACTTATTCAACAGTGTTCGCCGGGTGTTATGAAGTCAGGATCGTACTGTACT
SeMV TAGGTTTTCGTGTTTCATGAACTCAGTCTTCCAGTTATCTGACGGTACATTGATTGAGCAACTGCAGCCGGGAATAATGAAGTCTGGATCTTACTGCACC
SBMV-BAr TAGGTTTTCGTGTTTCATGAACTCAGTCTTCCAGCTCTCGGATGGCACACTCATAGAACAGCAACTGCCTGGTATCATGAAGTCTGGTTCATATTGTACC
SCPMV CAGGTTTTCATGCTTCATGAACTCAGTCCTGCAACTCTCAAACGGACAGTTACTCCAGCAAGAGCTGCCGGGGATCATGAAATCTGGATCTTATTGCACC
VTMoV_RdRp CAGATACCGCTGCTTTATGCTCTCTTGCTTCCAATTGTCGAATGGGGAATTGTATGAACAAGTTGAGCCTGGACTCATGAAGTCTGGTTCCTACTGCACT
TRoV TAGATTTAAATGTTTCATGAACTCGGTTTTCCAGCTTAGTAACGGTGAACTGATCTCTCAGGTTAGTCCTGGCCTGATGAAATCTGGATCCTATTGCACA
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
SCMoV TCGAGTACCAACAGCCGGATAAGATGTCTAATGGCTAAAATCATTGGAGCCCCGTGGTGCATTGCCATGGGCGACGACTCGGTAGAAGGATATGTTGAGG
LTSV TCGAGTACTAACAGCCGGATAAGATGCCCTCATGGCTGAAATAATACAAGCCCCGTGGTGCATCGCCATGGGAGACGACTCGGTAGAAGGATATGTTGAGG
RGMoV TCGAGCTCAAATTCACGGATAAGATGCCTAATGGCGAAGATCATCGGCTCACGATGGTGTATAGCCATGGGTGATGATTCCGTGGAGGGTTGGGTTGACA
CoMV TCCAGCTCCAACTCCCGCATTCGCTGCCTTATGGCTGAGCTCATCGGCTCGCCTTGGTGTATCGCAATGGGAGATGATTCTGTTGAGGGTTGGGTTGATG
IYMV TCATCTACCAACTCCAGAATTCGTTGCCTTATGGCTGAGCTTATCGGCTCCCCATGGTGTATTGCTATGGGTGATGATTCTGTTGAGGGATGGGTAGAGG
RYMV_Ni TCATCTACAAACTCGCGCATTCGGTGCCTCATGGGTGAGCTTATCGGCTCGCCGTGGTGCATAGCGATGGGCGACGATTCAGTGGAGGGGTGGATTGAGG
RuCMV TCCTCCACCAACTCCAGAGTCCGCTGCCTTATGGCAGAGATCATTGGGTCCCCGTGGTGTATAGCCATGGGTGATGACTCTGTTGAGGGGTATGTTGAGG
SoMV TCCTCCACCAACTCCAGAGTCCGCTGCCTTATGGCAGAGATCATTGGATCCCCGTGGTGCATAGCCATGGGTGACGACTCTGTTGAGGGGTATGTTGAGG
SeMV TCCTCAACCAATAGCAGAATTCGTTGCCTTATGGCTGAGCTCATTGGCTCCCCATGGTGTATTGCTATGGGTGATGATTCTGTTGAAGGATGGGTTGATG
SBMV-BAr TCCTCGACCAACTCCAGAATTCGTTGCCTTATGGCTGAGCTTATTGGCTCCCCGTGGTGTATCGCCATGGGTGATGATTCTGTTGAAGGTTGGGTTGATG
SCPMV TCCTCCACTAATTCGCGGATACGTTGCCTTATGGCTGAGCTTATTGGTTCCCCATGGTGCATCGCCATGGGTGACGACTCCGTGGAGGGTTTTGTGGAGG
VTMoV_RdRp TCCTCTTCCAACTCCAGAATCAGGTGCCTAATGGGTTATCTAATTGGAGCCCCCTGGATAATAGCCATGGGGGATGATTCTGTTGAGGGTTATGTGAGAG
TRoV TCTTCTACCAACTCAAGAATCAGGGTAGCAATGGCGTATTTAATAGGCTCCCCGTGGTGTATTGCCATGGGTGATGACTCTGTCGAAGGTTACGTTACCA
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
SCMoV GCGCCAGGGAGATGTACGACTCTCTCGGCCACACTTGCAAGGATTATATACCTTGCAA----------AGCTGATCTAGACAAG-----TTAGAGGA
LTSV GGGCCCGAGAGCTCTACGAAGAACTGGGTCACACATGTAAAGACTACATCCCGTGTGC----------TGCTGAGGGAGAGATC-----CTCAAGGA
RGMoV ATGCTGTAGAAAAGTATCATGCACTTGGTCATGAGTGTAAAGAATACTCAGCATGCGCCCGGACATACGTGAGCGGGAGCGGAGAACAAACCCTTCAAAC
CoMV ACGCTCCCAGGAAATATTCAGCACTAGGCCATCTCTGCAAAGAGTATGAAGCGTGTCC---------GGTGCTCCCCGATGGAGATT------TGAAGGA
IYMV GAGCCCAAGAGAAGTACGCCGCCCTAGGCCACACTTGTAAAGAGTACGTGGCTTGCCCTTCTAAACAAGGACGCCTAGGGAAGGAGC------TTCTCAG
RYMV_Ni GGGGCCAGTCCAAGTACGCCGCGCTAGGTCACACCTGCAAGGAGTACTACCCATGT----------AAGAC-CAGAGGGCGCGAGC------TCTTGGA
RuCMV GAGCTAGAGATAAATACGATGCTCTAGGACACAAGTGTAAGGACTACCAAGTCTGTG----------ACTCCGATGGATTGCTT-----CTTCGGTC
SoMV GAGCCAGAGATAAATACGATGCCCTAGGACACAAATGTAAGGACTATCAAGTCTGTG----------ACTCTGATGGATTGCTT-----CTTCGGTC
SeMV GCGCAAAGGACAAGTACATGAGGCTAGGCCATGTTTGCAAGGATTACAAGCCTTGCGC---------CACGACTCCTGAAGGCGACCT-----ATACG-A
SBMV-BAr GCGCAAAGGACAAGTACATGAGACTAGGGCACACGTGCAAGGACTATAAACCCTGTGC---------AACAACCATTTCCGGTCGCTT-----ATACG-A
SCPMV GGGCTAGGGAGAAGTATGCGGGGCTGGGGCACTTGTGCAAGGATTACAAGCCTTGTGC---------AACCACCCCCACAGGCCAGTT-----GTATG-C
VTMoV_RdRp ACGCGAAAGGCAAGTATGAGGAATTAGGACACACTTGTAAGGAATACGAGTTGTGTGA---------TGTTGATTCAGACGGCGCGT-----TGAGATC
TRoV ACGCCAGAGAGAAGTATGAATCCCTCGGACACATCTGTAAGGATTACCTGGTTTGCCA---------AAAGAA---GAAAGGCGAGC-----TAGATGG
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
SCMoV AGTGAATTTTTGCTCCCACA-----CTATAAGGAAAGAC----AGCTACTACCTC-CAATCCTGGGCGAAAACCTTGTTTCGATTCCTC---AGTCAACC
LTSV AGTGAACTTTTGCTCACACA-----GCATAGCTGCGGAT----CGTTGCTATCTC-CAATCCTGGGCTAAAACCTTATTTAGATATCTG---GAACACCC
RGMoV GGTTAATTTCTGCTCGCACG-----AGTTGTCCGAGGGT----AAGTGTCATCTT-ACGACCTGGAGCAAAACGCTTTTCCGATTCTTCCATAGTACGGC
CoMV GGTGAGCTTTTGCTCCCAT-----CTCATCTCTAAGGGT----CGAGCAGAGCTG-GAGACGTGGCCCAAGTGTCTGTTTCGATATCT--CAGTGGACCA


IYMV
RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV
1410
1420
1430
1440
1450
1460
1470
1480
1490
1500
SCMoV
LTSV
RGMoV
CoMV
IYMV
RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV
1510
1520
1530
1540
1550
1560
1570
1580
1590
1600
SCMoV
LTSV
RGMoV
CoMV
IYMV
RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
SCMoV
LTSV
RGMoV
CoMV
IYMV

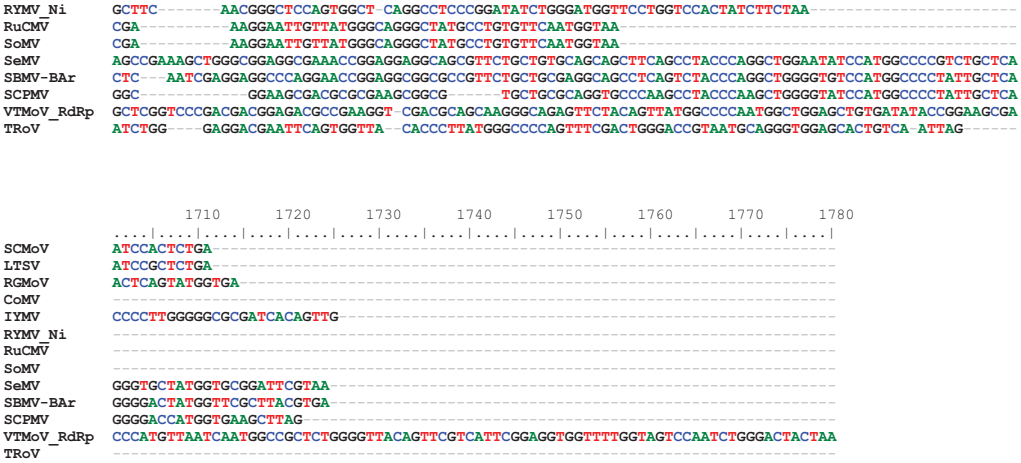
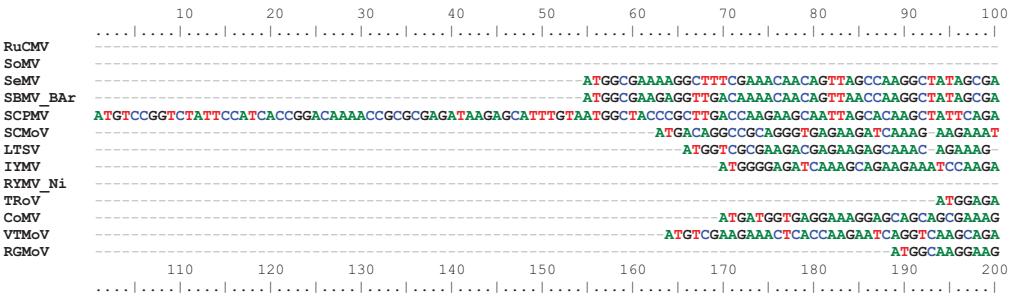
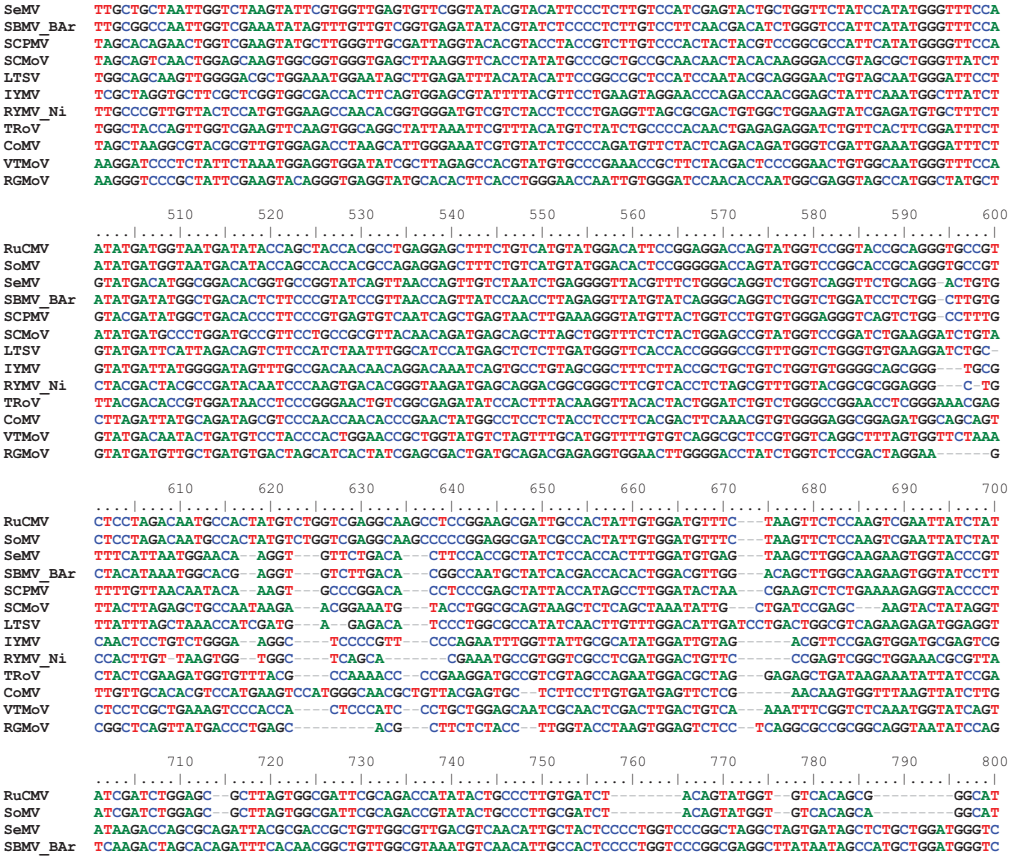
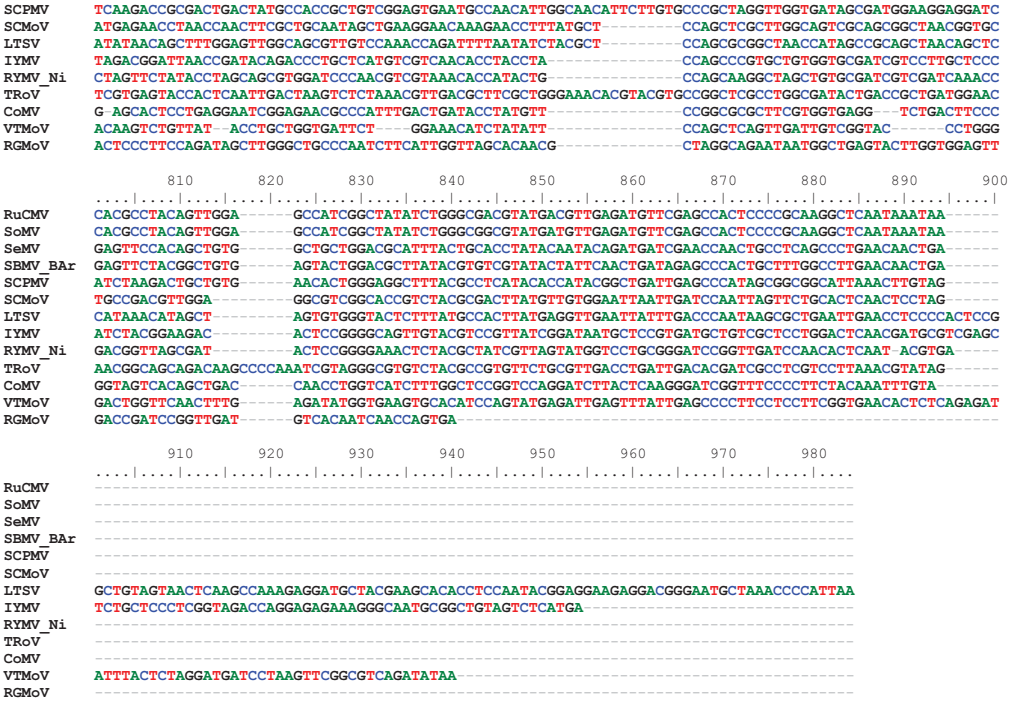

RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV
1710
1720
1730
1740
1750
1760
1770
1780
SCMoV
LTSV
RGMoV
CoMV
IYMV
RYMV_Ni
RuCMV
SoMV
SeMV
SBMV-BAr
SCPMV
VTMoV_RdRp
TRoV

Figure 4.10C


10
20
30
40
50
60
70
80
90
100
RuCMV
SoMV
SeMV
SBMV_BAr
SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV
110
120
130
140
150
160
170
180
190
200


RuCMV
SoMV
SeMV
SBMV_BAr
SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500


SeMV
SBMV_BAr
SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV
510
520
530
540
550
560
570
580
590
600
RuCMV
SoMV
SeMV
SBMV_BAr
SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV
610
620
630
640
650
660
670
680
690
700
RuCMV
SoMV
SeMV
SBMV_BAr
SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV
710
720
730
740
750
760
770
780
790
800
RuCMV
SoMV
SeMV
SBMV_BAr


SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV
810
820
830
840
850
860
870
880
890
900
RuCMV
SoMV
SeMV
SBMV_BAr
SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV
910
920
930
940
950
960
970
980
RuCMV
SoMV
SeMV
SBMV_BAr
SCPMV
SCMoV
LTSV
IYMV
RYMV_Ni
TRoV
CoMV
VTMoV
RGMoV

Appendix 8: Phylogenetic analysis of ORF1 sobemovirus sequences

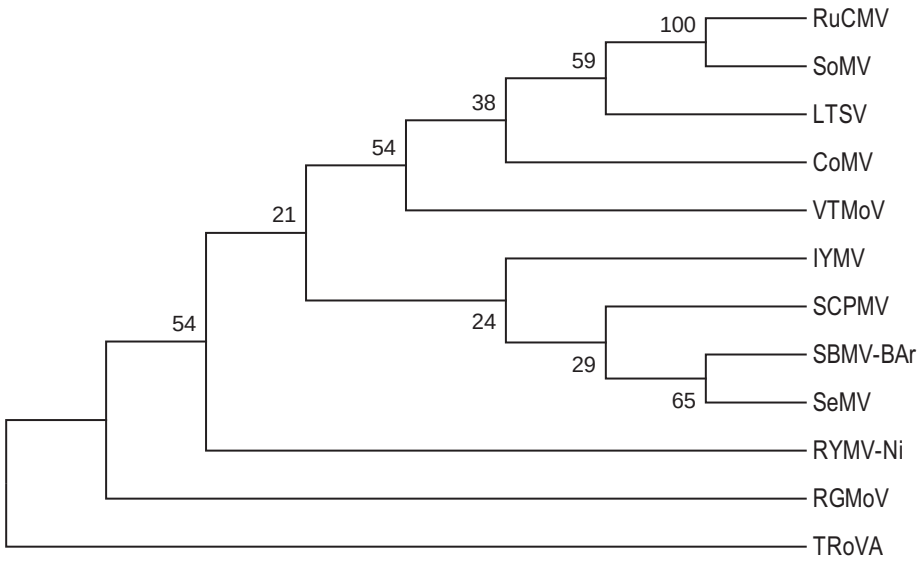


Figure A. Phylogenetic tree of sobemovirus P1 sequence, created using the maximum parsimony method, and bootstrapping was conducted with 1000 replicates. Numbers represent percentage values of bootstrapping support for branching nucleotide sequence (black). For details of sequences used refer to Table 3.3.

Appendix 9: Sequences of fragments 1-4 from VTMoV isolate R17

>Fragment 1
ATATTGAAGGGAAAGTCTTGCACTCAGTACCCATTATATTTTGATATGCCCAGCATTGATGTTGAAGTAGAAAAG
ATCCTGCACTTGAGCAACAGAAAGAGACTAAGGTCTGTCTGTGTGGCGAGGAAGAAGACGTTCGTAGACAAGATA
TCCGAAGGATTTGTCTGCACTCTTTACGGAGAGTACGATCACGACTACGTCACCAGTGTGCATTTGCACATTGTT
TGCTCGTGCGGTAGAGCTTTCTTTGATTTTGTTGAGTTCAAGGACATAAAACTCCGAGATTTCCAATATCAAAGT
CAGTGTATCTGTCCCACCAGGAATTGGTCGACGGTCGCTTACACTGATTTTACTACCGTCAACTGCCAAGTGGAA
GGTTGTGATTACTGTGAGGGGATTGAGTCAGACTCTGATTCGGATTCTGAGGCCATCATAGAGGAATTTCTTCAA
AAGTTTTCCGAGGTTGGCATCTCTGGATCGTCCAGTTCTCCCCAGACTAACTAGGACTCGGTTCCGTTTATTTTC
TTACAAATCATTTAATTGTTTTGTAATTGAGAAAGATGTTGAGCGAGTTAGTCCAGTTGTGTCTTTCAGCAACCA
TGACTCTAGCGCTGGTAGTGTCAATGATGCTGGACCCGAGCTACGGGTGGCTGATCACCATGTCTGTGATGTTAG
TCTCCTTAGAGCTGACAATCAGGCCGTTCAGGCGCTCTCTAGACTACATGAAAATTGTGGTCCGAGATGAGCCCA
ATGAGCCTGAAGTTGCTAGGATAGTAGGCGCGCCAACTTACCACCCAATTCATGGCGTTGTTTGCAAGTGTCTGT
GGAACTCGAAAGAGTTTAATGTGGTTGTGCAACCCAGTTTGTGGCCTCTAGCCTTTCCCCCAAACATATCACAAG
GGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCG
TTGTCTTATACAACGATGGAGTCAGAATCGGTATGGGCTCTAGAGTTAATTGGCAAGGCGCAGATTACTTGTTGA
CAGCCTCCCACGTTTGGTCCCTCGTAACTGGTGATTTTCAAATGGCAAAAGCCAGTAAGATGGTCTCAGTGAAGG
ACGCCAAGTGTTATGTTGAGGCTGCTCACGCTAAGCTGGACTTTGCCCTAATTAAAGTTCCGAACAAATATTGGA
GCTCGATAGGGGTTGGTTCGGCTAAATTGTTGTGGCATAAGCCCGGGCAAGTGGTCAAGGTTTACGGAGGTAGAT
CTGATGAATTGGTCTCTTCTGTGGGCCGAGCCGAAAAGGATCCTGATCTGAGCCTACGTCTGACCCATAACGCTT
CAACTGCTCCAGGATGGAGCGGCTCCCCGCTTTACAACTCTGAAAATTTCGTAGTGGGATTGCATACTGGCTTTT
CCGCGGCTGAGCAGAGAAATGAGGCGGTTGATGTAGCGAAGTTGCTGCACCTCGCCCTTAGAACAAAGGAAACGA
CGTTTTCGGAGATCGGTGTCTCTCTTATCGATGAAGATGAGATTGAGTCTCGCGGTTACCAATTCGACGACTTTG
AGCTTAGGGGTGAGGTAAACGTTAAAGGGAAAATGGCGCGGAATGAGATATCTCTCATCGCCTCTAAGAATAAAG
GGAA

>Fragment 2
TCCCCGCTTTACAACTCTGAAAATTTCGTAGTGGGATTGCATACTGGCTTTTCCGCGGCTGAGCAGAGAAATGAG
GCGGTTGATGTAGCGAAGTTGCTGCACCTCGCCCTTAGAACAAAGGAAACGACGTTTTCGGAGATCGGTGTCTCT
CTTATCGATGAAGATGAGATTGAGTCTCGCGGTTACCAATTCGACGACTTTGAGCTTAGGGGTGAGGTAAACGTT
AAAGGGAAAATGGCGCGGAATGAGATATCTCTCATCGCCTCTAAGATTAAAGGGAAACCTTGGTATCTTCAAGAG
GAAGGTGATGATGAATTCTATGACTCGATCGGTGAGAAAGATTTCCTTGCCCGCTTCCGTGAACAAACTGGTAAA
GAGACAGTGGGGAATTTAAACTGCCAGAGGGCGGCCCAAACATTGGAGCCGCCCTTCGAGAACTTGAGGCCGTGC
GATGGGAAGAACCCAGAGCTATTCAAACCGGCGGGATGGGATTCAACTATGTTGGAGTCTCGTCTTGCAAGTTTA
GAGAGGGCTCTAAGCACGCTGCTCGCGGAGCAATCAGTTCTGCTGAGCAA

>Fragment 3
AGATTTCCTTGCCCGCTTCCGTGAACAAACTGGTAAAGAGACAGTGGGGAATTTAAACTGCCAGAGGGCGGCCCA
AACATTGGAGCCGCCCTTCGAGAACTTGAGGCCGTGCGATGGGAAGAACCCAGAGCTATTCAAACCGGCGGGATG
GGATTCAACTATGTTGGAGTCTCGTCTTGCAAGTTTAGAGAGGGCTCTAAGCACGCTGCTCGCGGAGCAATCAGT
TCTGCTGAGCAAGTTTTCCCAGAACTCAAACAGTATGGTTGGCCAGAAAGAGGCTCTAAAGCCGAGTTCGATTCC
CTCCTCCTCCAAGCCAGCCGTTTCCGGAGAACCTCCTGTCCAGAGCAAACCGAGCGTAAATGCGAAGTCCTCGCT
GAGAAGTACCCCAAAAGTAGAGCCCACCGTTGCTTCAGGAGAGAAAACTTCTGTCAACGCCAGCTCCTCAGGGAA
CAAATCGAAGCGACGACGACGTCGCCGGAAATCAACGACAAAGCCAGCCCCGGCTCCCCCTGGTCAAGACTTGAA
AAGACCAATGGTGAGCTCATCTCCCGTTTCAAAGACCTCTTAATTGAGGCAGTGTTGAGGCGAGTTTTATTGCTT
GCCTCCACTCCAACCAGTGAGATTTTATGCATGTCTGCTTCAGATCTCGTTAGAGCTAACTTAGTGGATCCTGTG
CGATTGTTCGTCAAGCAGGAACCCCATACGAAGAAGAAGCTGAATGAGCGTCGCTTTAGGCTTATTTCGTCTGTG
TCTATAGTTGATCAGATAATCGAGAGATTATTGTTTGGTCCTCAGAACAGGCTTGAGATAGCCTTATGGCATCAG
ATACCTTCCAAACCTGGAATGGGACTGAGCGCTCGTACGCAAGCTGATTTGTTGTGGAATGAGTTATTCGCAAAG
AGCGAGATTGCCCCTGCGGCCGAGGCTGACATTTCAGGATTTGACTGGTCTGTGCAGGAATGGGAGTTATGGGCC
GATCTGAGCATGAGGATTTCTTTGTGCGAAGATATGCATGATGGTCTACGGAGGTTGATGGTTAACAGATACCGC
TGCTTTATGCTCTCTTGCTTCCAATTGTCGAATGGGGAATTGTATGAACAAGTTGAGCCTGGACTCATGAAGTCT
GGTTCCTACTGCACTTCCTCTTCCAACTCCAGAATCAGGTGCCTAATGGGCTATCTAATTGGAGCCCCCTGGATA
ATAGCCATGGGGGATGATTCTGTTGAGGGTTATGTGAGAGACGCGAAAGGCAAGTATGAGGAATTAGGACACACT
TGTAAGGAATACGAGTTGTGTGATGTTGATTCAGACGGCGCGTTGAGATCTGTGAACTTTTGTTCACATTTGATT
TCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTT
CATGATCTTGAAAGGGAACTTGGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTG
GTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATGTCGAAGAAACTCACCAAGAACCAGGTCAA
GCAGATGATACAGGCAACCTTGCCGAAAGAGCAGACTTCTGCTCGGTCCCGACGACGGAGACGCCGAAGGTCGAC
GCAGCAAGGGCAGAGTTCTACAGTTATGGCCCCAATGGCTGGAGCTGTGATATACCGGAAGCGACCCATGTTAAT
CAATGGCCGCTCTGGGGTTACAGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTC
GGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCTGTACAAGGATCCCTCTATTCTAAATGGAG
GTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTTTCCA
GTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGT

>Fragment 4
ATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAAT
TTTCATGATCTTGAAAGGGAACTTGGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGA
CTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATGTCGAAGAAACTCACCAAGAACCAGGT
CAAGCAGATGATACAGGCAACCTTGCCGAAAGAGCAGACTTCTGCTCGGTCCCGACGACGGAGACGCCGAAGGTC
GACGCAGCAAGGGCAGAGTTCTACAGTTATGGCCCCAATGGCTGGAGCTGTGATATACCGGAAGCGACCCATGTT
AATCAATGGCCGCTCTGGGGTTACAGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTT
CTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCTGTACAAGGATCCCTCTATTCTAAATG
GAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTTT
CCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGC
TCCGTGGTCAGGCTTTAGTGGTTCTAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGC
AACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAGTCTGTTATACCTGCTGGTGATTCTGG
AAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGCA
CATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGAT
GATCCTAAGTTCGGCGTCAGATGAAAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCGA

Appendix 10: Amino acid groupings

List of amino acid groupings from Mathews et al (2000), based on the chemical nature of the amino acid side chain.

Amino acid groups	Amino acids
Aliphatic	Glycine (G) Alanine (A) Valine(V) Leucine (L) Isoleucine (I)
Hydroxyl/ sulphur	Serine (S) Cysteine (C) Threonine (T) Methionine (M)
Cyclic	Proline (P)
Aromatic	Phenylalanine (F) Tyrosine (Y) Tryptophan (W)
Basic	Histidine (H) Lysine (K) Arginine (R)
Acidic / Acid amides	Aspartic acid (D) Glutamic acid (E) Aspargine (N) Glutamine (Q)

Appendix 11: Nucleotide diversity figures including recombinant isolates

Table A12: Diversity across the three regions of the RdRp sequences, including recombinant isolates 65758, 65796, 65797

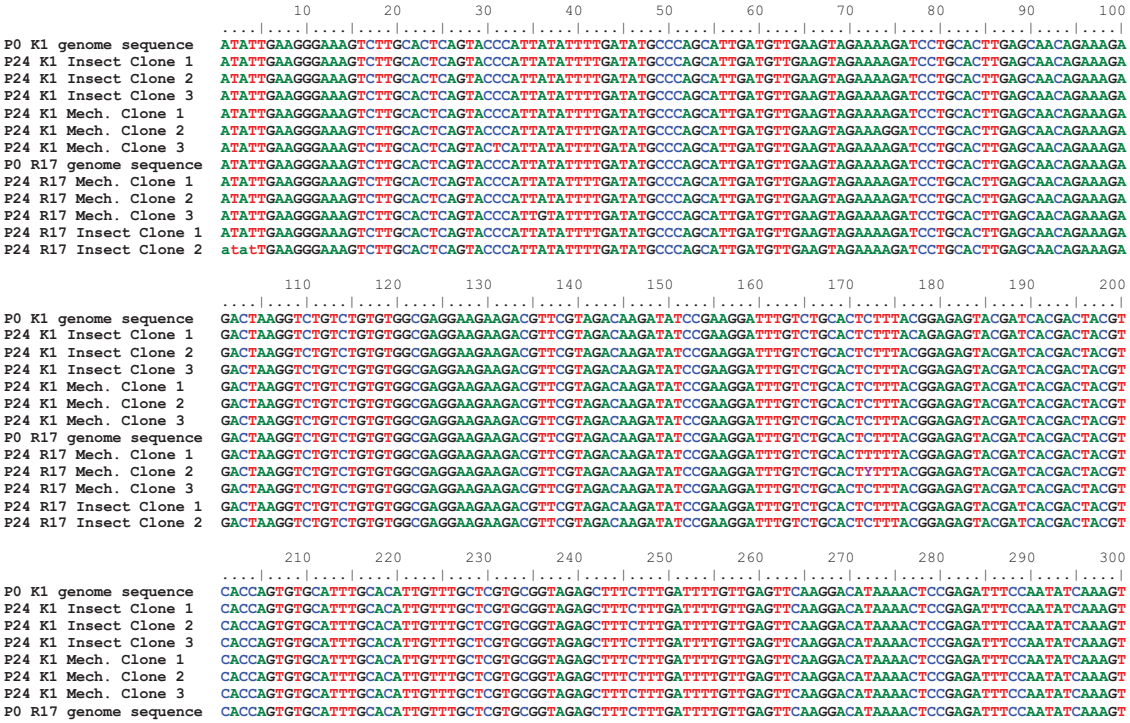
Genomic region (nucleotides)	Nucleotide diversity			
	d*	d_N ^	d_S #	ω d_N / d_S
ORF2a/2b (1-419 nt)	0.0167 ± 0.0035	0.0083 ± 0.0024	0.0332 ± 0.0096	0.25
ORF2b only (420-1396 nt)	0.0302 ± 0.0032	0.0375 ± 0.0044	0.0100 ± 0.0029	3.75
ORF2b/CP overlap (1397-1647nt)	0.0114 ± 0.0033	0.0164 ± 0.0053	0.0022 ± 0.0020	7.454
Overall (1647)	0.0236 ± 0.0021	0.0079 ± 0.0013	0.0575 ± 0.0061	0.13739

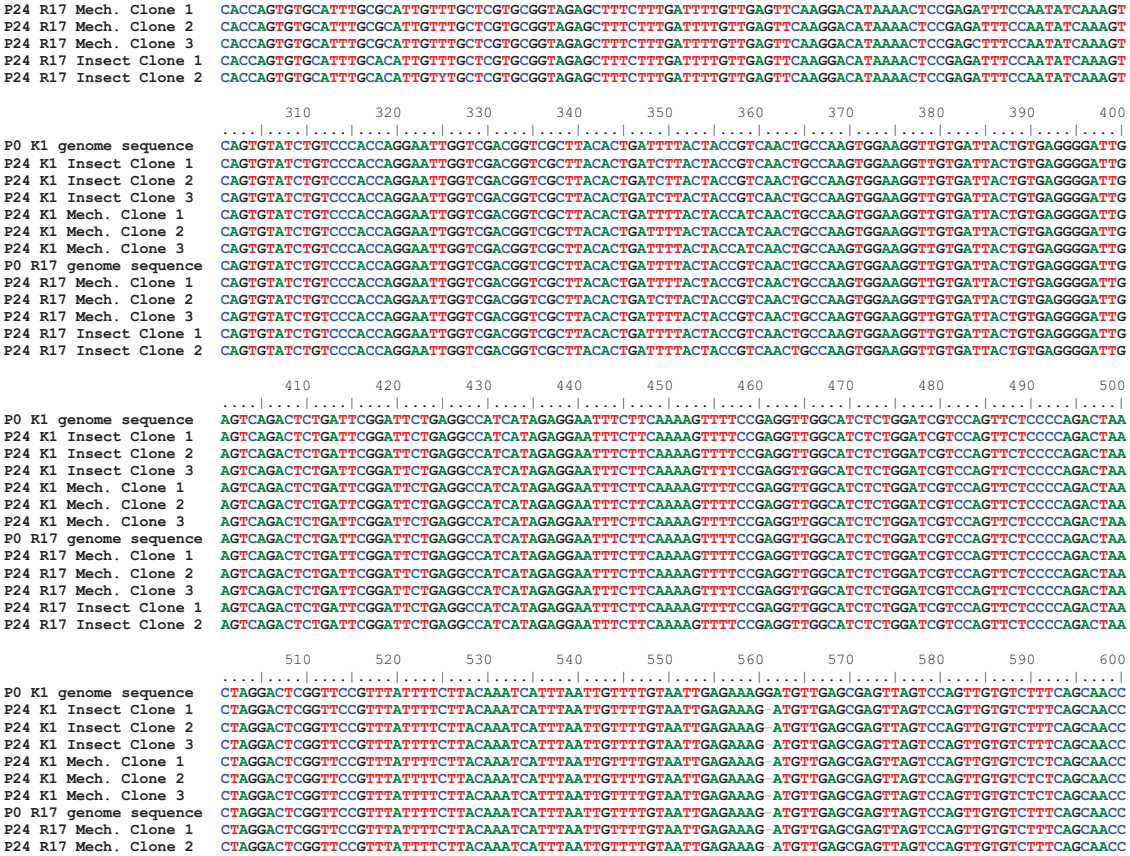
* Units are the number of nucleotide substitutions per nucleotide site between 2 sequences. ^ Units are the number of non-synonymous substitutions per non-synonymous site. # is the number of synonymous substitutions per the total number of synonymous sites

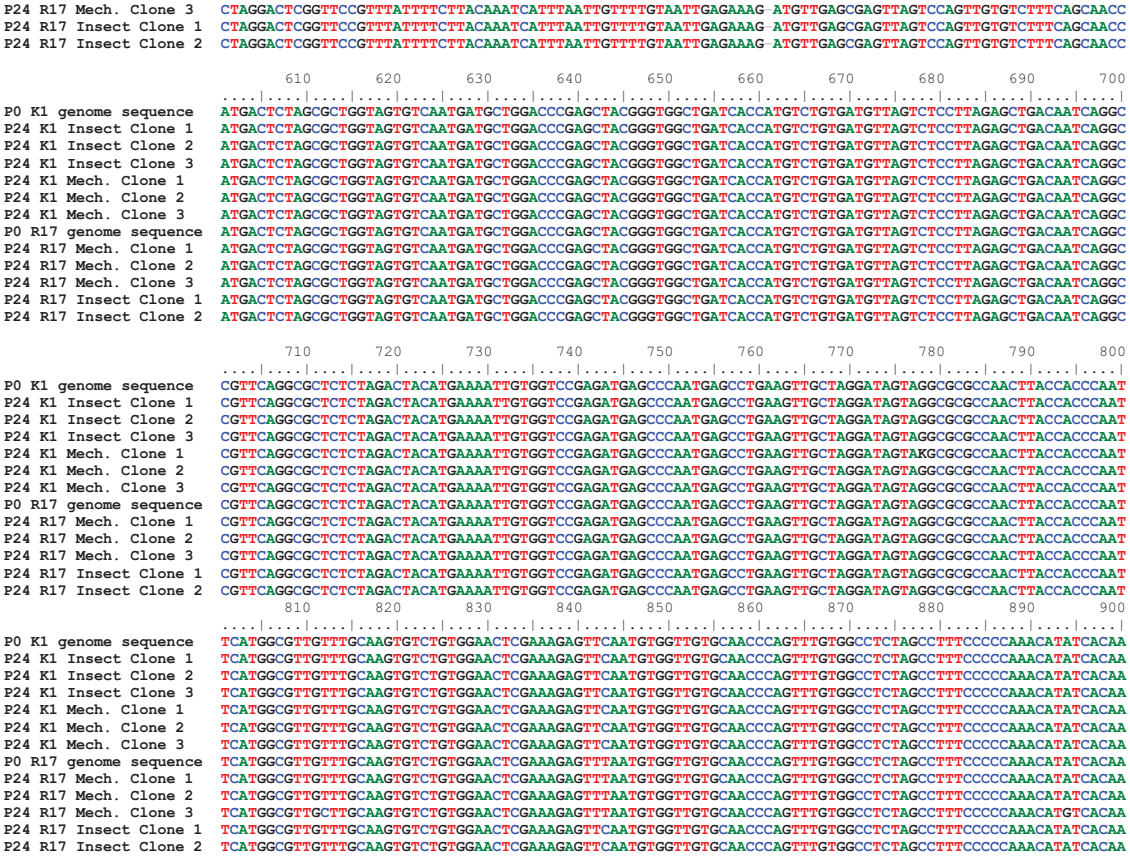
.

Appendix 12: Sequence alignments from the transmission experiment

Region 1 from nucleotides 8-990




P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
310
320
330
340
350
360
370
380
390
400
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
410
420
430
440
450
460
470
480
490
500
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
510
520
530
540
550
560
570
580
590
600
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2

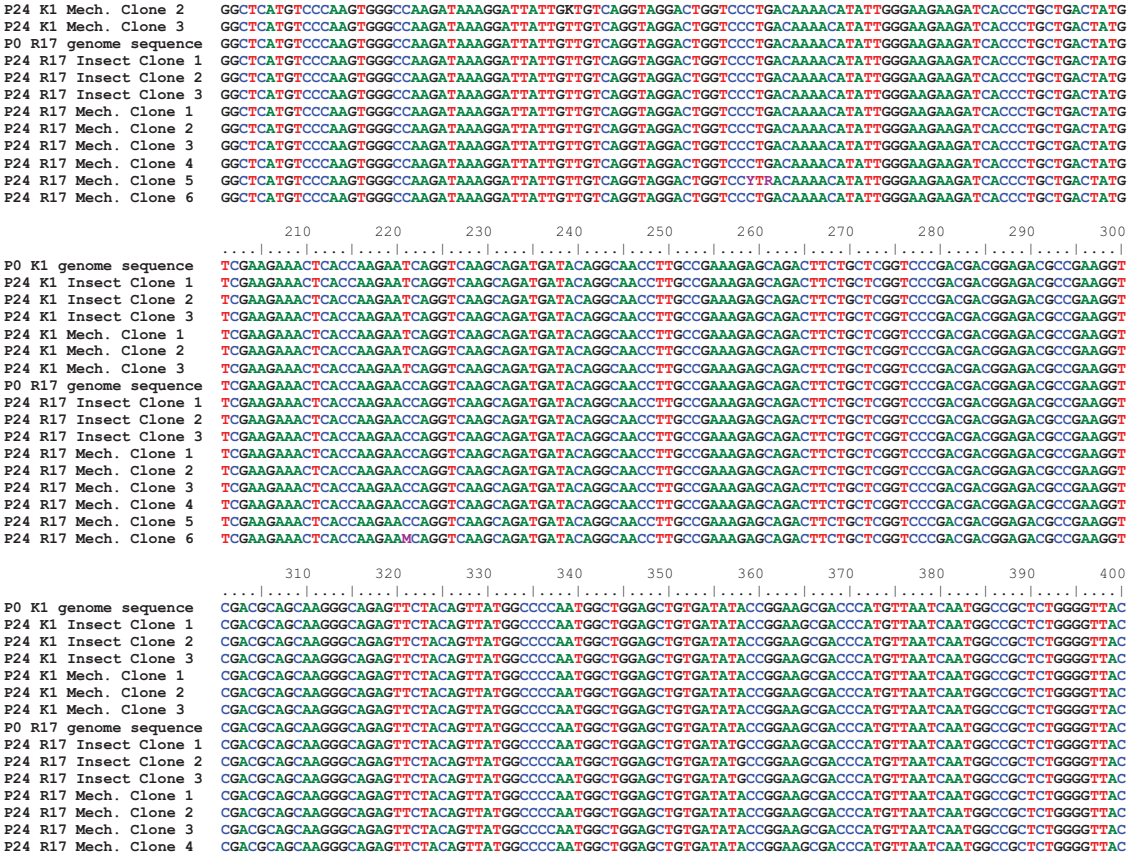

P24 R17 Mech. Clone 3
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
610 620 630 640 650 660 670 680 690 700
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
710 720 730 740 750 760 770 780 790 800
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
810 820 830 840 850 860 870 880 890 900
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2

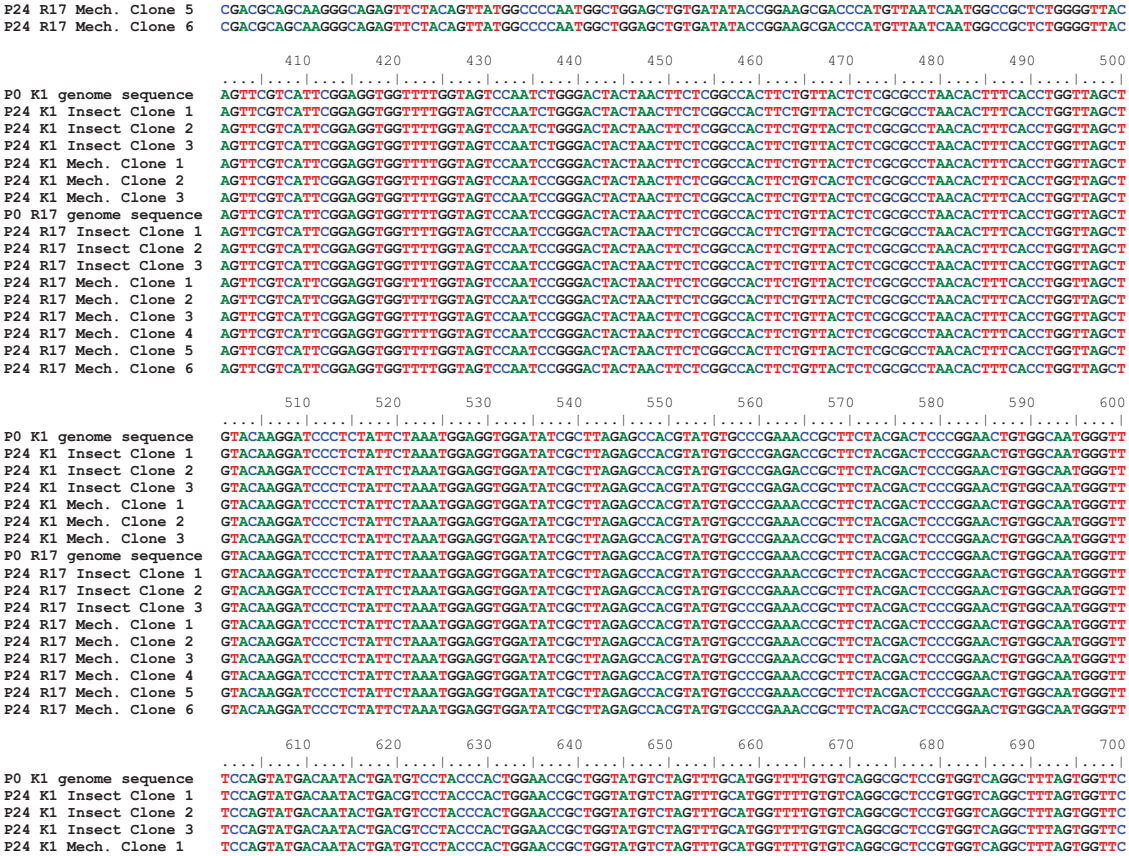
```
                                910       920       930       940       950       960       970       980
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
P0 K1 genome sequence           GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 K1 Insect Clone 1           GGGGTAGTAAAGGAAGGGGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 K1 Insect Clone 2           GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 K1 Insect Clone 3           GGGGTAGTAAAGGAAGGGGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 K1 Mech. Clone 1            GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 K1 Mech. Clone 2            GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 K1 Mech. Clone 3            GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P0 R17 genome sequence          GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 R17 Mech. Clone 1           GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 R17 Mech. Clone 2           GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 R17 Mech. Clone 3           GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 R17 Insect Clone 1          GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
P24 R17 Insect Clone 2          GGGGTAGTAAAGGAAGGAGCTATGATAGGTGTACCATTCACATCCCTACCCTCTGGAGCTGAACCAAAGTCTCTCGTTGTCT
```

Region 2 from nucleotides 3560-4115

```
                                10        20        30        40        50        60        70        80        90        100
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P0 K1 genome sequence           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 K1 Insect Clone 1           ----------------TTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 K1 Insect Clone 2           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 K1 Insect Clone 3           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 K1 Mech. Clone 1            GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 K1 Mech. Clone 2            GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 K1 Mech. Clone 3            GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P0 R17 genome sequence          GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Insect Clone 1          GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Insect Clone 2          GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAGCCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Insect Clone 3          GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGGACTT
P24 R17 Mech. Clone 1           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Mech. Clone 2           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Mech. Clone 3           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Mech. Clone 4           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Mech. Clone 5           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT
P24 R17 Mech. Clone 6           GATTTCCCGCAACAAGTTTTGGCTCACAAGCTGGCCTAAAACCTTGTACAGGTTCCTAGACTCTCCCTCTGAAAATTTTCATGATCTTGAAAGGGAACTT

                                110       120       130       140       150       160       170       180       190       200
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P0 K1 genome sequence           GGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATG
P24 K1 Insect Clone 1           GGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATG
P24 K1 Insect Clone 2           GGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATG
P24 K1 Insect Clone 3           GGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATG
P24 K1 Mech. Clone 1            GGCTCATGTCCCAAGTGGGCCAAGATAAAGGATTATTGTTGTCAGGTAGGACTGGTCCCTGACAAAACATATTGGGAAGAAGATCACCCTGCTGACTATG
```


P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
P24 R17 Insect Clone 3
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Mech. Clone 4
P24 R17 Mech. Clone 5
P24 R17 Mech. Clone 6
210 220 230 240 250 260 270 280 290 300
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
P24 R17 Insect Clone 3
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Mech. Clone 4
P24 R17 Mech. Clone 5
P24 R17 Mech. Clone 6
310 320 330 340 350 360 370 380 390 400
P0 K1 genome sequence
P24 K1 Insect Clone 1
P24 K1 Insect Clone 2
P24 K1 Insect Clone 3
P24 K1 Mech. Clone 1
P24 K1 Mech. Clone 2
P24 K1 Mech. Clone 3
P0 R17 genome sequence
P24 R17 Insect Clone 1
P24 R17 Insect Clone 2
P24 R17 Insect Clone 3
P24 R17 Mech. Clone 1
P24 R17 Mech. Clone 2
P24 R17 Mech. Clone 3
P24 R17 Mech. Clone 4


P24 R17 Mech. Clone 5 CGACGCAGCAAGGGCAGAGTTCTACAGTTATGGCCCCAATGGCTGGAGCTGTGATATACCGGAAGCGACCCATGTTAATCAATGGCCGCTCTGGGGTTAC
P24 R17 Mech. Clone 6 CGACGCAGCAAGGGCAGAGTTCTACAGTTATGGCCCCAATGGCTGGAGCTGTGATATACCGGAAGCGACCCATGTTAATCAATGGCCGCTCTGGGGTTAC
410 420 430 440 450 460 470 480 490 500
P0 K1 genome sequence AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCTGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 K1 Insect Clone 1 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCTGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 K1 Insect Clone 2 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCTGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 K1 Insect Clone 3 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCTGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 K1 Mech. Clone 1 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 K1 Mech. Clone 2 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTCACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 K1 Mech. Clone 3 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P0 R17 genome sequence AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Insect Clone 1 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Insect Clone 2 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Insect Clone 3 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Mech. Clone 1 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Mech. Clone 2 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Mech. Clone 3 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Mech. Clone 4 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Mech. Clone 5 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
P24 R17 Mech. Clone 6 AGTTCGTCATTCGGAGGTGGTTTTGGTAGTCCAATCCGGGACTACTAACTTCTCGGCCACTTCTGTTACTCTCGCGCCTAACACTTTCACCTGGTTAGCT
510 520 530 540 550 560 570 580 590 600
P0 K1 genome sequence GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 K1 Insect Clone 1 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAGACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 K1 Insect Clone 2 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAGACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 K1 Insect Clone 3 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAGACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 K1 Mech. Clone 1 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 K1 Mech. Clone 2 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 K1 Mech. Clone 3 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P0 R17 genome sequence GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Insect Clone 1 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Insect Clone 2 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Insect Clone 3 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Mech. Clone 1 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Mech. Clone 2 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Mech. Clone 3 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Mech. Clone 4 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Mech. Clone 5 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
P24 R17 Mech. Clone 6 GTACAAGGATCCCTCTATTCTAAATGGAGGTGGATATCGCTTAGAGCCACGTATGTGCCCGAAACCGCTTCTACGACTCCCGGAACTGTGGCAATGGGTT
610 620 630 640 650 660 670 680 690 700
P0 K1 genome sequence TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 K1 Insect Clone 1 TCCAGTATGACAATACTGACGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 K1 Insect Clone 2 TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 K1 Insect Clone 3 TCCAGTATGACAATACTGACGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 K1 Mech. Clone 1 TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC

```
P24 K1 Mech. Clone 2    TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 K1 Mech. Clone 3    TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGWTTTGTGTCAGGCGCTCCGTGRTCAGGCTTTAGTGGTTC
P0 R17 genome sequence  TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Insect Clone 1  TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Insect Clone 2  TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Insect Clone 3  TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGCGGTTC
P24 R17 Mech. Clone 1   TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Mech. Clone 2   TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Mech. Clone 3   TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Mech. Clone 4   TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Mech. Clone 5   TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC
P24 R17 Mech. Clone 6   TCCAGTATGACAATACTGATGTCCTACCCACTGGAACCGCTGGTATGTCTAGTTTGCATGGTTTTGTGTCAGGCGCTCCGTGGTCAGGCTTTAGTGGTTC

                                710       720       730       740       750       760       770       780       790       800
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P0 K1 genome sequence   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 K1 Insect Clone 1   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 K1 Insect Clone 2   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 K1 Insect Clone 3   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 K1 Mech. Clone 1    TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 K1 Mech. Clone 2    TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 K1 Mech. Clone 3    TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P0 R17 genome sequence  TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Insect Clone 1  TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Insect Clone 2  TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Insect Clone 3  TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Mech. Clone 1   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Mech. Clone 2   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Mech. Clone 3   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Mech. Clone 4   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Mech. Clone 5   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG
P24 R17 Mech. Clone 6   TAAACTCCTCGCTGAAAGTCCCACCACTCCCATCCCTGCTGGAGCAATCGCAACTCGACTTGACTGTCAAAATTTCGGTCTCAAATGGTATCAGTACAAG

                                810       820       830       840       850       860       870       880       890       900
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P0 K1 genome sequence   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 K1 Insect Clone 1   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 K1 Insect Clone 2   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 K1 Insect Clone 3   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 K1 Mech. Clone 1    TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 K1 Mech. Clone 2    TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 K1 Mech. Clone 3    TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P0 R17 genome sequence  TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Insect Clone 1  TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Insect Clone 2  TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Insect Clone 3  TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGtTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Mech. Clone 1   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Mech. Clone 2   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Mech. Clone 3   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Mech. Clone 4   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
```

```
P24 R17 Mech. Clone 5   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC
P24 R17 Mech. Clone 6   TCTGTTATACCTGCTGGTGATTCTGGAAACATCTATATTCCAGCTCAGTTGATTGTCGGTACCCTGGGGACTGGTTCAACTTTGAGATATGGTGAAGTGC

                                910       920       930       940       950       960       970       980       990       1000
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
P0 K1 genome sequence   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 K1 Insect Clone 1   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 K1 Insect Clone 2   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 K1 Insect Clone 3   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 K1 Mech. Clone 1    ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 K1 Mech. Clone 2    ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 K1 Mech. Clone 3    ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P0 R17 genome sequence  ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCaGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Insect Clone 1  ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Insect Clone 2  ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Insect Clone 3  ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Mech. Clone 1   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Mech. Clone 2   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Mech. Clone 3   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Mech. Clone 4   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Mech. Clone 5   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA
P24 R17 Mech. Clone 6   ACATCCAGTATGAGATTGAGTTTATTGAGCCCCTTCCTCCTTCGGTGAACACTCTCAGAGATATTTACTCTAGGATGATCCTAAGTTCGGCGTCAGATGA

                              1010      1020      1030      1040      1050
                        ....|....|....|....|....|....|....|....|....|....|
P0 K1 genome sequence   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 K1 Insect Clone 1   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 K1 Insect Clone 2   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 K1 Insect Clone 3   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 K1 Mech. Clone 1    AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 K1 Mech. Clone 2    AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 K1 Mech. Clone 3    AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P0 R17 genome sequence  AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 R17 Insect Clone 1  AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 R17 Insect Clone 2  AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 R17 Insect Clone 3  AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGCG
P24 R17 Mech. Clone 1   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGC
P24 R17 Mech. Clone 2   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGC
P24 R17 Mech. Clone 3   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGC
P24 R17 Mech. Clone 4   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGC
P24 R17 Mech. Clone 5   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGC
P24 R17 Mech. Clone 6   AAGGAAAGGTGAGAAGTCTAGTCAGACTTTAACTGACCGCTAGGCAAGC
```