Molecular Analysis
of
Vanilla mosaic virus
from the Cook Islands

Christopher Puli'uvea

A thesis submitted to
Auckland University of Technology
in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

2017

School of Science

Abstract

Vanilla was first introduced to French Polynesia in 1848 and from 1899-1966 was a major export for French Polynesia who then produced an average of 158 tonnes of cured *Vanilla tahitensis* beans annually. In 1967, vanilla production declined rapidly to a low of 0.6 tonnes by 1981, which prompted a nation-wide investigation with the aim of restoring vanilla production to its former levels. As a result, a mosaic-inducing virus was discovered infecting *V. tahitensis* that was distinct from Cymbidium mosaic virus (CyMV) and Odontoglossum ringspot virus (ORSV) but serologically related to dasheen mosaic virus (DsMV). The potyvirus was subsequently named vanilla mosaic virus (VanMV) and was later reported to infect *V. tahitensis* in the Cook Islands and *V. planifolia* in Fiji and Vanuatu. Attempts were made to mechanically inoculate VanMV to a number of plants that are susceptible to DsMV, but with no success. Based on a partial sequence analysis, VanMV-FP (French Polynesian isolate) and VanMV-CI (Cook Islands isolate) were later characterised as strains of DsMV exclusively infecting vanilla. Since its discovery, little information is known about how VanMV-CI acquired the ability to exclusively infect vanilla and lose its ability to infect natural hosts of DsMV or vice versa.

The aims of this research were to characterise the VanMV genome and attempt to determine the molecular basis for host range specificity of VanMV-CI. VanMV-CI has a typical potyvirus genome structure encoding a single polyprotein which can potentially be cleaved into ten proteins, flanked by 5' and 3' untranslated regions (UTR). Sequence comparisons of individual genes and complete genomes from members of the bean common mosaic virus (BCMV) group to which DsMV belongs revealed VanMV-CI is most closely related to DsMV. Sequence analyses revealed VanMV-CI is 75.1-77.5% and 81.8-84.1% identical to DsMV at the nucleotide (nt) and amino acid (aa) levels, respectively in the coat protein (CP) and 75.8-77.1% and 81.9-86.7% identical to DsMV at the nt and aa levels, respectively over the entire genome. These data are consistent with the findings of Farreyrol et al. (2005) who first suggested that VanMV-CI is a strain of DsMV.

The origin of VanMV-CI is unclear. Did it evolve from DsMV or was it the progenitor of DsMV? To try answer this, a recombination and phylogenetic analysis was carried out.

A recombination analysis between DsMV and VanMV-CI genome sequences was carried out against all available potyvirus genomes to determine whether VanMV-CI emerged as a result of a recombination event between DsMV and another virus. This analysis was also to determine whether the N-terminal P1 gene of VanMV-CI arose as a result of a recombination event giving VanMV the ability to exclusively infect vanilla spp. The reasoning for focussing on this genome region is outlined below. This analysis found two recombination events in the P1 gene of DsMV, one in its N-terminal region and another in its C-terminus. Sequence identities between DsMV and its major parent however, were lower than 75% and the recombination event was only detected by one detection method producing insufficient evidence for VanMV arising as a result of one or more recombination events.

A phylogenetic analysis based on the complete genome showed that the emergence of VanMV-CI preceded the publicly available DsMV genomes. However, a maximum likelihood analysis of the entire CP of VanMV-CI and all available DsMV CP sequences showed that VanMV-CI was most closely related to DsMV in the South Pacific and the origin of DsMV in this region preceded VanMV-CI. In addition, this analysis showed that DsMV likely originated in the Asia region (China, Japan and India) where its natural host (*Colocasia esculenta* and *Xanthosoma species*) originated- it is possible that DsMV spread to the Pacific with human migration. Thus, it is most likely that VanMV arose from DsMV, by an as yet unknown mechanism- possibly by random mutation and natural selection or by a recombination event between DsMV and an as yet to be discovered virus.

Upon sequence comparison with DsMV, the P1-HC-Pro region of VanMV-CI became a region of interest for determining a molecular basis for the difference in host range, as this was the most divergent area between the two genomes. This difference was largely due to an insertion/deletion (indel) found in the N-terminal region of the VanMV-CI P1. According to a number of studies, the N-terminal region of the potyviral P1 gene must be compatible with a still unknown host factor before its C-terminus can effectively cleave between itself and HC-Pro. Without successful separation from P1, HC-Pro is unable to counterattack RNA silencing by plants and the virus ceases to infect its host. This suggests that the VanMV-CI P1 may be the region responsible for its ability to infect vanilla but not aroids.

An agroinfiltration experiment was designed to test the functionality of the P1 region as a viral suppressor of RNA interference (VSR) to provide support for its role as a host range determinant. A series of gene constructs were designed to test the function of the VanMV-CI and DsMV P1 genes in combination with the HC-Pro region to determine if the HC-Pro activity of each virus is affected by the origin of the P1. The HC-Pro and P1-HC-Pro regions of VanMV-CI and a New Zealand isolate of DsMV were amplified to generate a series of wild type and hybrid constructs for future use in determining their ability to counterattack RNA silencing by plants (silencing suppressor activity). The wild type constructs (PKP2 and PKP4) and a hybrid PKP6 construct containing DsMV P1/VanMV-CI HC-Pro were generated. Generating the alternative hybrid PKP5 construct (VanMV-CI P1/ DsMV HC-Pro) was unsuccessful, most probably due to a lack of sequence complementarity between the fusion templates (VanMV-CI P1 and DsMV HC-Pro fusions). Attempts were made to transform the constructs PKP1-PKP4 (VanMV HC-Pro, VanMV P1/VanMV HC-Pro, DsMV HC-Pro and DsMV P1/DsMV HC-Pro respectively) into a pHEX2 expression vector for agroinfiltration into *Nicotiana benthamiana* to examine the difference in silencing suppressor activity in light of the difference in P1. Cloning of the PKP6 construct into the entry vector (pCR8) was unsuccessful and transformation of expression vectors containing PKP constructs with *Agrobacterium tumefaciens* GV3101 also encountered a number of problems. This was most likely due to the use of the inefficient freeze-thaw method.

Although the agroinfiltration experiment could not be completed, PKP constructs including HC-Pro and P1/HC-Pro of VanMV and DsMV were amplified and incorporated into the destination expression vector pHEX2 and their sequences confirmed.

Key words

vanilla mosaic virus, dasheen mosaic virus, host range determinant.

Table of contents

List of Figures VIII

List of Tables X

Attestation of Authorship XII

Acknowledgments XIII

Abbreviations XV

Chapter 1: General introduction

1.1 Introduction 2

1.2 Vanilla mosaic virus 2

1.2.1 Symptoms and disease spread 2

1.2.2 Vanilla mosaic virus is a strain of Dasheen mosaic virus 4

1.2.3 Host range 5

1.2.4 Available molecular data 5

1.3 Dasheen mosaic virus 6

1.3.1 Symptoms 6

1.3.2 Host range and disease spread 7

1.3.3 Available molecular data 8

1.4 Potyviridae 8

1.5 Potyvirus 8

1.5.1 Potyvirus genome structure 9

1.5.2 Potyvirus protein functions 10

1.5.2.1 5’ and 3’ untranslated regions 10

1.5.2.2 P1 peptidase 11

1.5.2.3 HC-Pro 12

1.5.2.4 P3, 6K1 and P3N-PIPO 16

1.5.2.5 CI 17

1.5.2.6 NIa and 6K2 18

1.5.2.7 NIb 19

1.5.2.8 CP 19

1.6 Potyvirus evolution 20
1.6.1 Quasispecies 21
1.6.2 Recombination 22
1.7 RNA silencing 23
1.7.1 siRNA pathway 25
1.8 Potyvirus suppressor of RNA silencing 28
1.9 Aims of study 29

Chapter 2: Characterising the VanMV-CI genome

2.1 Introduction 32
2.1.1 RNA virus origins and evolution 32
2.1.2 Picorna-like virus superfamily and potyviridae evolution 34
2.1.3 Potyvirus evolution and the BCMV lineage 36
2.1.4 Dasheen mosaic virus origins and evolution 38
2.1.5 VanMV classification 39
2.2 Aims of study 41
2.3 Methods and materials 42
2.3.1 Plant material 42
2.3.2 Total RNA extraction 42
2.3.3 Sequence assembly 43
2.3.4 VanMV sequence annotation and analysis 43
2.3.5 Phylogenetic analysis of the VanMV genome and genes 44
2.3.6 Recombination analysis of VanMV 47
2.4 Results 48
2.4.1 VanMV genome structure 48
2.4.2 Phylogenetic analysis of VanMV 57
2.4.3 Sequence comparisons of VanMV with other potyviruses 67
2.4.4 Recombination analysis 68
2.5 Discussion 71

Chapter 3: Towards establishing the impact of P1 on HC-Pro in silencing suppression

3.1 Introduction 78
3.1.1 RNA interference 79

3.1.2 Potyvirus silencing suppressor 79
3.1.3 HC-Pro activity in silencing suppression 80
3.1.4 Silencing suppression and host range determination 81
3.2 Aims of study 84
3.3 Methods and materials 84
3.3.1 Total RNA extraction 84
3.3.2 Primer design 84
3.3.3 One-Step reverse transcriptase-polymerase chain reaction (RT-PCR) .. 88
3.3.4 Gel purification of PCR products 88
3.3.5 Overlapping extension PCR 88
3.3.6 Creation of Gateway entry vector 89
3.3.6.1 Ligation reaction 89
3.3.6.2 Transformation of entry vector 89
3.3.7 Colony PCR and gel electrophoresis 90
3.3.8 Plasmid DNA extraction 90
3.3.9 Gateway LR recombination reaction 91
3.3.10 Making competent Agrobacterium tumefaciens cells 92
3.3.11 Transformation of Agrobacterium tumefaciens 92
3.4 Results 93
3.4.1 Total RNA extraction 93
3.4.2 Amplification of P1/HC-Pro 93
3.4.3 Overlapping extension PCR 96
3.4.4 pCR8/GW/TOPO TA cloning (entry vector) 97
3.4.5 Gateway cloning (entry vector to destination vector) 99
3.4.6 Transformation of *Agrobacterium tumefaciens* 101
3.5 Discussion 104

Chapter 4: General discussion 107
References 115

List of Figures

Chapter 1

Figure 1.1: Map locations of VanMV infected vanilla species in French Polynesia……..3

Figure 1.2: VanMV inclusion bodies showing laminated aggregates, pinwheels and cylindrical inclusions ……..4

Figure 1.3: Symptoms of VanMV infections on vanilla plant species……..4

Figure 1.4: VanMV sequences currently available in GenBank……5

Figure 1.5: DsMV symptoms on *Colocasia esculenta* ……..6

Figure 1.6: Potyvirus genome organisation and functions of each protein ……10

Figure 1.7: Functional domains within HC-Pro of potyviruses ……13

Figure 1.8: Summary of major steps in the siRNA pathway ……14

Figure 1.9: Illustration of the quasispecies model ……22

Figure 1.10: Illustration of a recombination event between two viruses ……23

Figure 1.11: Comparison of the (A) miRNA and (B) siRNA pathway ……25

Figure 1.12: Four different ways vsiRNA's can be produced ……27

Figure 1.13: Schematic representation of the (A) siRNA pathway and (B) the methylation step taking place ……28

Chapter 2

Figure 2.1: Phylogenetic tree of RdRp of the picorna-like virus superfamily ……35

Figure 2.2: The Out of Africa model with dates and locations of DsMV discovery around the world ……39

Figure 2.3: Nucleotide sequence alignment of VanMV and DsMV genomes……48

Figure 2.4: Alignment of (A) 5'UTR and P1, (B) aa sequence of the P1 gene of DsMV and VanMV showing indels……50

Figure 2.5: Alignment of the (A) 5' end of the CP coding region and (B) N-terminal region of the CP peptide of DsMV and VanMV showing indels……51

Figure 2.6: Average percent nucleotide identity of VanMV and DsMV vs BCMV group of potyviruses ……56

Figure 2.7: Neighbour-joining phylogenetic tree of potyvirus genomes ……58

Figure 2.8: Maximum likelihood tree of VanMV-CI genome and the BCMV group of potyviruses ……59
Figure 2.9: Maximum likelihood tree of the 5'UTR of VanMV-CI and the BCMV group of potyviruses ……59
Figure 2.10: Maximum likelihood tree of the P1 of VanMV-CI and the BCMV group of potyviruses ……60
Figure 2.11: Maximum likelihood tree of the HC-Pro of VanMV-CI and the BCMV group of potyviruses ……60
Figure 2.12: Maximum likelihood tree of the P3 of VanMV-CI and the BCMV group of potyviruses ……61
Figure 2.13: Maximum likelihood tree of the 6K1 of VanMV-CI and the BCMV group of potyviruses ……61
Figure 2.14: Maximum likelihood tree of the CI of VanMV-CI and the BCMV group of potyviruses ……62
Figure 2.15: Maximum likelihood tree of the 6K2 of VanMV-CI and the BCMV group of potyviruses ……62
Figure 2.16: Maximum likelihood tree of the NIa-VPg of VanMV-CI and the BCMV group ……63
Figure 2.17: Maximum likelihood tree of the NIa-Pro of VanMV-CI and the BCMV group of potyviruses ……63
Figure 2.18: Maximum likelihood tree of the NIb of VanMV-CI and the BCMV group of potyviruses ……64
Figure 2.19: Maximum likelihood tree of the 3'UTR of VanMV-CI and the BCMV group of potyviruses ……64
Figure 2.20: Maximum likelihood tree of the CP of VanMV-CI and the BCMV group of potyviruses ……66
Figure 2.21: Overview of possible sites of recombination over the DsMV and VanMV genome ……68

Chapter 3

Figure 3.1: Overview of strategy used in generating fusion templates and PKP constructs ……87

Figure 3.2: Plasmid pHEX2 Gateway vector 91
Figure 3.3: Agarose gel electrophoresis of total RNA extraction of VanMV and DsMV.....93
Figure 3.4: Agarose gel electrophoresis of expected 1374 bp of HC-Pro of DsMV and 2529 bp of P1-HC-Pro 94
Figure 3.5: Agarose gel electrophoresis of expected 1374 bp of VanMV HC-Pro and 2439 bp of P1-HC-Pro.....94
Figure 3.6: Agarose gel electrophoresis of the expected 1076 bp of VanMV P1 fusion and 1384 bp HC-Pro fusion template 95
Figure 3.7: Agarose gel electrophoresis of the expected 1166 bp of DsMV P1 fusion and 1384 bp HC-Pro fusion template 95
Figure 3.8: Agarose gel electrophoresis of the expected 2529 bp PKP6 construct 97
Figure 3.9: A schematic representation of the ligation reaction taking place between the PCR product and pCR8 entry vector 97
Figure 3.10: Agarose gel electrophoresis of the successful ligation reaction and transformation of PCR products into pCR8 98
Figure 3.11: A schematic representation of the LR recombination reaction taking place between the entry and destination vectors 100
Figure 3.12: Agarose gel electrophoresis of colony PCR of DsMV and VanMV HC-Pro transformed with *E. coli* 100
Figure 3.13: Agarose gel electrophoresis of colony PCR of DsMV and VanMV P1-HC-Pro transformed with *E. coli* 101
Figure 3.14: Agarose gel electrophoresis of (A) colony PCR of *A. tumefaciens* cells transformed with PKP1-4 (B) PCR of purified plasmid from transformed *A. tumefaciens* cells transformed with PKP1-4 constructs 102

List of Tables

Chapter 2

Table 2.1: List of the BCMV group genome sequences obtained from GenBank 45
Table 2.2: Available VanMV sequences obtained from GenBank.....45

Table 2.3: All complete and available DsMV CP sequences obtained from GenBank….46
Table 2.4: Comparison of genome sequence lengths and individual genes of VanMV-CI and DsMV ……49
Table 2.5: Potyvirus cleavage site found in VanMV compared to DsMV……..52
Table 2.6: Potyvirus conserved motifs found in VanMV and DsMV genomes ……54
Table 2.7: Percentage of amino acid and nucleotide identity of VanMV genome and individual genes compared to publically available DsMV sequences ……..67
Table 2.8: Recombination analysis result of DsMV and VanMV using RDP4……..70

Chapter 3

Table 3.1: Primer sequences used to generate constructs and fusion templates for VanMV and DsMV……85
Table 3.2: Summary of progress made throughout the experiment ……103

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Christopher Puli'uvea

Acknowledgments

Ko ‘Atonai Sihova, ‘Ela ‘Eloha ‘a e ‘Afiona, he ‘ikai teu fiu pe fo’i hono tuku ‘a e fakamalo mo e fakafeta’i ki he ‘Afiona, he ‘oku ‘ikai teke fiu pe fo’i he kumi mai ki ho’o sipi, ke ‘oua na’a tuku ha taimi ta’e teu fai ai hano fakahikihiki’i ‘a e ‘Afiona, he ‘oku ‘ikai ha taha ke tatau, ha taha ke kakunga. Talamai, ko ho’o maama mo ho’o huelo, ko e teunga ia ‘a e tokotaha ‘oku koloa ‘aki ‘a e ‘Afiona, ke ne hoko ‘o hange ha mahina, ‘oku ne ho’ata mai e maama, ‘a e huelo mo e mafana ‘a e la’aa, ke ha ai ‘a e kololia mo e langilangi, ke pehe foki Sisu, ho’o Afeitaulalo mai ki he efu kuo ke fakamanava’i, keu koloa ‘aki pe ko e, hei’ilo na’a lava ai ha’o hakeaki’i.

Firstly I would like to express my heartfelt thanks to my supervisor Dr Colleen Higgins. You have been the lighthouse of this project, from its genesis to the very end. Your unwavering guidance and direction has been the driving force behind this achievement. During times of low visibility I have been able to rely on your knowledge and expertise to get back on track and move forward. You create a fun and inspiring atmosphere where weekly update supervision meetings were never dull! Throughout this journey I have grown tremendously, as a scientist, and as a person under your supervision. I am extremely privileged to have had you as a mentor throughout my undergraduate and postgraduate years and am forever indebted to you for your support as a lecturer, supervisor, and now, friend.

I would also like to thank my second supervisor Dr Robin MacDiarmid for all the support and valuable advice you have given me over the years. Your insight and knowledge has been invaluable. Composing this thesis without your input would not have been possible. Thank you for always creating time in your busy schedule to help me and for the advice and words of wisdom. You and Colleen are definitely one mind in two bodies! Your words will not be forgotten. I will miss our PFR meetings and the laughs we have before, during and after!

Thank you to the team at AUT, to Gardett, Priya, Lee and Saeedah for always believing in me even when I felt like all hope was gone! For helping me out with little things around the lab, your support and motivation was much appreciated. A huge thank you to Professor Mike Pearson, Sam, Waqas and the VLO/PFR team for the warm welcome

and support through the experimental and troubleshooting phase of my research. I have learnt a lot during this time and the different perspectives have helped me grow as a scientist.

I would also like to thank my Mum, Dad, Fita, Ana, Fangupo, Vika, Nuni and Funaki for all your love and encouragement throughout this journey. Thank you for believing in me and supporting my goals and dream to be a scientist and continue my education. Thank you Mum for sticking up for me, and Dad, for the 2-hour motivational talks we would have! I will always cherish all you have done for me.

I would also like to acknowledge my Uncles who have helped me along the way but are not here with me to see it to the end. To Paulo ‘Ahosivi Tatuila(‘Alusa), Vikena Tupetaiki Folau and Kulukuku Toatangane Moahengi Latuniupulu. You have all left too soon and wish you were all here to see the final product. Thank you for all the love and support during the early stages of my research, it has been difficult moving forward with out you all and not a day goes by where I don’t think about you all. Ke mou ma’u ha nonga mo ha fiemalie ‘I he fatafata ‘o ‘Atonai.

And last but not least, to my wife, Kristy Anne Rebecca Gibbs Puli’uvea. Completing this thesis would not have been possible without your help and input. For being the shoulder to cry on when experiments weren’t going according to plan, for listening to me troubleshoot out loud and offering your ideas, for putting my needs and wants second to yours to complete this thesis. This thesis is a dedication to you and Israel. You were both the motivation I needed to get through the tough stages and I hope I have made you both proud. Thank you both for everything.

Abbreviations

aa	amino acid
bp	base pairs
BCMNV	bean common mosaic necrosis virus
BCMV	bean common mosaic virus
CaMV	cauliflower mosaic virus
cDNA	complementary DNA
CI	Cook Islands
CI	cylindrical inclusion protein (when referring to potyviral protein)
CMV	cucumber mosaic virus
CP	coat protein
DsMV	dasheen mosaic virus
dsRNA	double stranded ribonucleic acid
eIF4E, eIF (iso) 4E	eukaryotic translation initiation factors 4E, (iso) 4E
E. coli	*Escherichia coli*
ER	endoplasmic reticulum
GFP	green fluorescent protein
HC-Pro	helper component protease
ICTV	International Committee on Taxonomy of Viruses
LB	Luria-Bertani broth
miRNA	micro-RNA
ML	maximum likelihood
N. benthamiana	*Nicotiana benthamiana*
NIa	nuclear inclusion a
NIb	nuclear inclusion b
nt	nucleotide
N. tabacum	*Nicotiana tabacum*
NJ	neighbour joining
OD	optical density
ONMV	oat necrotic mottle virus
ORF	open reading frame
PCR	polymerase chain reaction

PD	plasmodesmata
PPV	plum pox virus
PVA	potato virus A
PVV	potato virus V
PVY	potato virus Y
P1	P1 protein
P3	P3 protein
RdRp	RNA-dependant RNA polymerase
RISC	RNA induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RT-PCR	reverse transcriptase polymerase chain reaction
siRNA	short interfering RNA
ssRNA	single stranded RNA
TBE	tris-borate EDTA
TBSV	tomato bushy stunt virus
TEV	tobacco etch virus
TMV	tobacco mosaic virus
UTR	Untranslated region
VanMV	Vanilla mosaic virus
VPg	viral linked protein
vsiRNA	viral short interfering RNA
VSR	viral suppressor of RNA silencing
V. fragrans	*Vanilla fragrans*
V. planifolia	*Vanilla planifolia*
V. pompona	*Vanilla pompona*
V. tahitensis	*Vanilla tahitensis*
6K1 and 6K2	6kDa proteins 1 and 2

Chapter 1

General Introduction

1.1 Introduction

Plant viruses are infectious agents causing disease that reduce agricultural, pastoral, horticultural and ornamental crop production (Roossinck, 2013; Riechmann et al., 1992). According to the most recent report from the International Committee on Taxonomy of Viruses (ICTV), there are currently about 900 plant virus species identified and officially recognised (King et al., 2012) . Other preliminary studies suggest thousands are yet to be identified (Roossinck, 2013). In addition, numerous viruses have been described in the literature yet have not undergone the official process of recognition as species by ICTV (Roossinck & García-Arenal, 2015). Among those not currently recognised by ICTV is vanilla mosaic virus (VanMV), a vanilla infecting potyvirus (Wisler & Zettler, 1987).

Plant viruses have caused major constraints on exports of a number of countries worldwide, restricting food supply and economic growth (Babu et al., 2011; Ha et al., 2008; Pearson & Grisoni, 2002). These plant pathogens affect crops worldwide resulting in symptoms that may be mild through to large areas of cultivated crops destroyed (Strange & Scott, 2005). In addition to a decline in exports, virus diseases can cause food shortages affecting at least 800 million people worldwide (Strange & Scott, 2005). Countries in the South Pacific such as Tonga, French Polynesia and the Cook Islands have felt the devastating impacts of these pathogens on the cultivation of vanilla spp (Farreyrol, 2005). *Vanilla tahitensis* is one of two main crops cultivated in French Polynesia; introduced in 1848 it produced an annual average of 158 tonnes of cured beans for export from 1899-1966 (Wisler & Zettler, 1987). Between 1967-1981, however, this average decreased dramatically to 0.6 tonnes prompting an investigation of viral infection of the orchid (Wisler & Zettler, 1987). As a result, a number of viruses including VanMV (distinct from watermelon mosaic virus (WMV), formerly known as vanilla necrosis virus) were identified (Wang & Pearson, 1992; Wisler & Zettler, 1987).

1.2 Vanilla mosaic virus

1.2.1 Symptoms and disease spread

VanMV was first discovered in *V. tahitensis* at a number of locations in French Polynesia (underlined in Figure 1.1) as having mosaic-like symptoms distinct from

those caused by Cymbidium mosaic virus (CyMV) and Odontoglossum ringspot virus (ORSV) that may also be present in the same host (Wisler & Zettler, 1987). VanMV causes mosaic like symptoms with mild to severe discolouration of leaves and malformation of both leaves and vines (Figure 1.3) (Farreyrol et al., 2005; Wisler & Zettler, 1987). Symptoms are often well-defined on young leaves and weakened on older leaves, with obvious symptom appearances in cool months compared to warmer months (Odoux & Grisoni, 2010). Following the initial stages of infection there is possible stunting of normal plant growth and in most cases a decline in crop productivity (Odoux & Grisoni, 2010; Wisler & Zettler, 1987). In some instances symptoms may not be detected and may require further molecular diagnostic tests to confirm viral infection (Grisoni et al., 2004). Field observations have revealed an uneven distribution of the potyvirus in its host, suggestive of a non-systematic infection (Odoux & Grisoni, 2010). The virus is primarily transmitted by a number of aphids such as *Cerataphis lataniae, Aphis crassivora* and *A. gossypii* in a non-persistent manner (Wisler & Zettler, 1987). VanMV has an average particle length of 767 nm and width of 11 nm with typical potyviral inclusion bodies in infected cells (Figure 1.2) (Wisler & Zettler, 1987). Since its initial discovery, this newly discovered potyvirus was later reported to infect *V. tahitensis* in the Cook Islands and *V. planifolia* in Fiji and Vanuatu (Farreyrol, Pearson, Grisoni, Cohen, & Beck, 2006; Odoux & Grisoni, 2010).

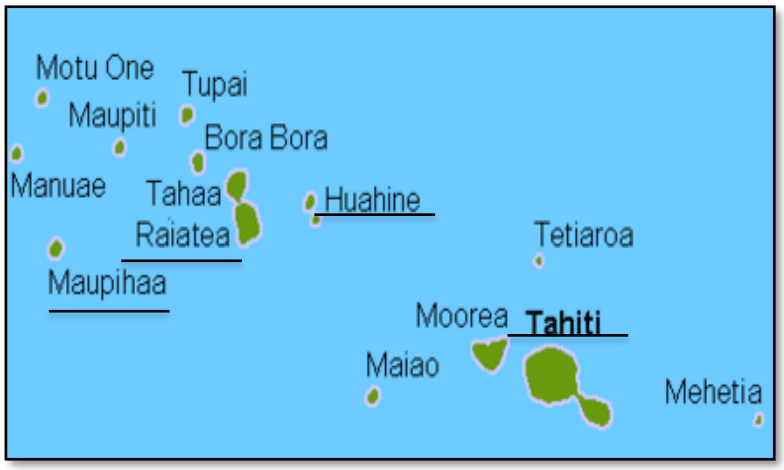


Figure 1.1: Map locations of islands in French Polynesia. The islands underlined were reported to have VanMV infected vanilla species (Wisler & Zettler, 1987).

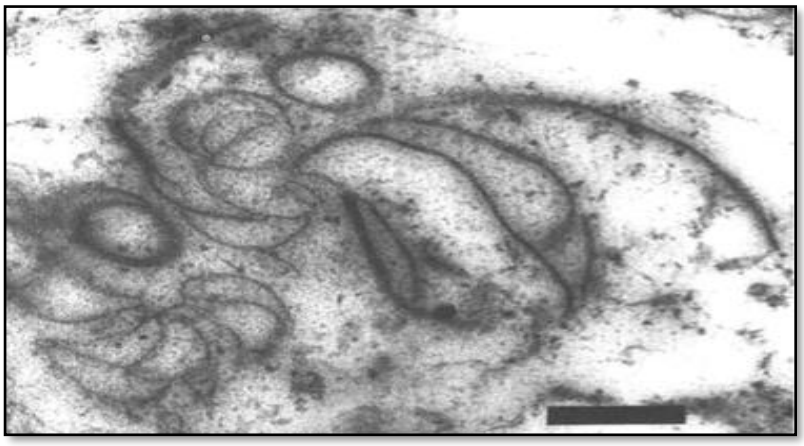


Figure 1.2: VanMV inclusion body typical of a potyvirus showing laminated aggregates, pinwheels and cylindrical inclusions. Bar = 500nm (Wisler & Zettler, 1987).

Figure 1.3: Leaf distortion, discolouration and mosaic like symptoms caused by VanMV on *Vanilla spp.* (Farreyrol, 2005; Wisler & Zettler, 1987).

1.2.2 Vanilla mosaic virus is a strain of *Dasheen mosaic virus*

Since its discovery, VanMV was thought to be similar to dasheen mosaic virus (DsMV), if not identical, based on similarities in serological properties alone (Wisler & Zettler, 1987). A later study identified both VanMV isolates from French Polynesia (FP) and Cook Islands (CI) as strains of DsMV exclusively infecting vanilla species based on the core coat protein (CP) region (Farreyrol et al., 2006). From their data, the authors suggested that VanMV-FP could be a distinct species based on sequence comparison of the entire CP and 3' untranslated region (UTR), a difference in host range and a difference in predicted CP-NIb cleavage site to DsMV (Farreyrol et al., 2006). In contrast, sequences of the Cook Islands isolate were at the boundary of species demarcation for potyviruses. Until the full genome sequence is available, the discriminatory criterion involving the complete genome and comparing polyprotein cleavage sites (Adams et al., 2004) cannot be fulfilled. However, based on current

sequence information available, both VanMV isolates CI and FP can be classified as strains of DsMV exclusively infecting vanilla spp.

1.2.3 Host range

VanMV appears to have a limited host range as it has only been reported to infect species of the Vanilla genus such as *V. tahitensis, V. fragrans, V. pompona* and *V. planifolia* (Farreyrol, 2005). For ease of propagation and further study, attempts have been made to mechanically inoculate a number of different plant species with VanMV. Wisler and Zettler (1987) attempted to inoculate eight different plant species, all of which are not natural hosts of the virus, with a FP isolate of VanMV. A further 20 plant species were mechanically inoculated by Wang and Pearson (1992) with FP and CI isolates. A final attempt was then made by Farreyrol (2005) to mechanically inoculating a further 14 plant species. All attempts were unsuccessful with the hosts failing to develop local or systemic infection.

1.2.4 Available molecular data

The VanMV genome is not publically available. There are however, some partial genome sequences available from GenBank as shown in Figure 1.4. At present, these sequences are available for the 3' end of the virus and one sequence around the 6K2 region.

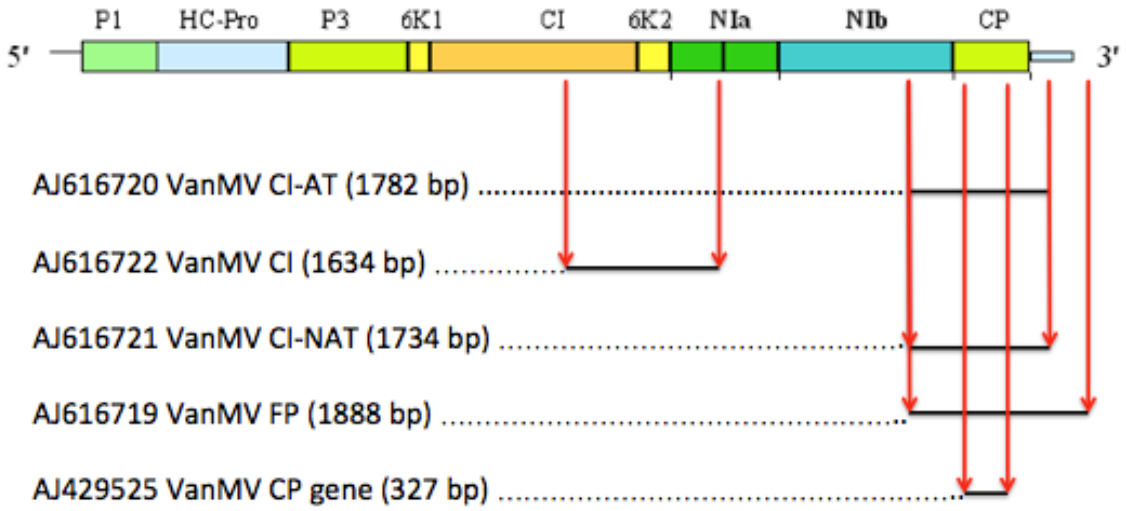


Figure 1.4: VanMV sequences currently available from GenBank. Each solid black horizontal line corresponds to the sequence available in the database. Each vertical line is the approximate region of sequence coverage against a typical potyvirus genome. The description to the left gives the accession numbers for each sequence followed by the isolate name and the size in base pairs (bp) of each (Farreyrol et al., 2006; Grisoni et al., 2006).

1.3 Dasheen mosaic virus (DsMV)

1.3.1 Symptoms

Based on molecular data, morphological and serological properties, DsMV has been classified in the species *Dasheen mosaic potyvirus* of the *Potyvirus* genus and a member of the *Bean common mosaic virus* group (Babu, Hegde, Makeshkumar, & Jeeva, 2011; R. H. Li, Zettler, Purcifull, & Hiebert, 1998; Lima Roberto, Lima, & Aguiar, 2004). DsMV has devastating impacts on aroids worldwide affecting not only indigenous plants, but plants of economic significance (Chen, Chen, Chen, & Adams, 2001; Greber & Shaw, 1986; Ha, 2007; Ha, Revill, Harding, Vu, & Dale, 2008; Kamala, Makeshkumar, Sreekumar, & Chakrabarti, 2015; Ram et al., 2003). DsMV infection of taro induces a range of symptoms from mild mosaic, leaf puckering, yellowing of leaves to severe leaf deformation, low crop yield and stunted plant growth (Figure 1.5) (Kamala et al., 2015).

Figure 1.5: DsMV infected *C. esculenta.* Symptoms consist of discolouration/ yellowing of the leaves and leaf distortion around the edges. Retrieved from http://www.ctahr.hawaii.edu/oc/freepubs/pdf/PD-44.pdf.

1.3.2 Host range and disease spread

DsMV was discovered in 1970 having similar serological properties to tobacco etch virus (TEV) and blackeye cowpea virus (BlCMV) (Abo El-Nil, 1977; Zettler, Foxe, Hartman, Edwardson, & Christie, 1970). From 1971-1987 DsMV was reported to infect *Colocasia esculenta* (taro), *Dieffenbachia picta, D. maculate, Xanthosoma helliborifolium, Chenopodium amaranticolor, C. quinoa, C. ambrosioides* and *Nicotiana benthamiana* in Florida, China, Japan, South Italy, Europe, Venezuela, Puerto-Rico, Netherlands, Egypt, Solomon Islands and Australia (Abo El-Nil, 1977; Chase, 1982; Hill & Wright, 1980; Rana, Vovlas, & Zettler, 1983; Wisler & Zettler, 1987; Zettler, Tsai, Faan, & Ke, 1987). From 1992-1998 DsMV was becoming a common threat to aroid crops worldwide such as *C. esculenta* (taro) from China, Hawaii, Florida and New Zealand (Hu, Meleisea, Wang, Shaarawy, & Zettler, 1995; Pappu et al., 1993; Pearson, Bussell, & Scheffer, 1998). By 2001 and 2003, reports from India showed that DsMV had infected *Amorphophallus paeoniifolius* (elephant foot yam) and ornamental aroid plants (Ram et al., 2003). A review of the geographical locations of potyvirus species indicate a prevalence of DsMV in the Cook Islands, Fiji, French Polynesia and Vanuatu (Pearson & Grisoni, 2002). DsMV has also been reported in Guam, Kiribati, Niue, Papua New Guinea, Solomon Islands and Tonga (Pearson & Grisoni, 2002). DsMV is now known to infect the aroids *Alocasia macrorrhizos, Amorphophallus paeoniifolius, Cyrtosperma chamissonis* and *Xanthosoma sagittifolium* (Pearson & Grisoni, 2002). In addition, DsMV has also been reported to infect the orchid *Spiranthes ceruna* (Jordan et al., 2002). The virus appears to be spread by a number of different aphid species through planting material (Revill et al., 2005). The potyvirus has also been mechanically inoculated into a number of different plant species (Revill et al., 2005) however, is not transmissible to vanilla plant species (Wang & Pearson, 1992; Wisler & Zettler, 1987).

Although closely related to VanMV (Farreyrol et al., 2006; Wisler & Zettler, 1987), there is little information about the molecular basis for the difference in host range and the inability of VanMV to infect DsMV hosts.

1.3.3 Available molecular data

Three DsMV genomes are available from GenBank (KJ86965, AJ298033 KT026108). There are also 89 partial DsMV sequences in GenBank ranging from 327 bp to 1500 bp, primarily from the 3' end of the viral genome.

1.4 Potyviridae

The *Potyviridae* family of plant viruses consists of eight genera, namely *Brambyvirus, Ipomovirus, Macluravirus, Poacevirus, Rymovirus, Tritimovirus, Bymovirus* and *Potyvirus* (Frédéric & Juan, 2015). The family is comprised of 176 species in total, 146 of which belong to the *Potyvirus* genus alone (Frédéric & Juan, 2015). Several members of the family have been associated with serious infections of plants with potyviral species ranked in the top ten plant infecting viruses of host plants including papaya, field-grown vegetables and stone fruits (Ha, 2007; Rybicki, 2015; Scholthof et al., 2011). All members of the *Potyviridae* family presently identified, are disease-causing agents in plants containing a positive-sense single-stranded RNA with flexible and filamentous virus particles (Frédéric & Juan, 2015).

1.5 Potyvirus

The *Potyvirus* genus is the largest of the eight genera within the *Potyviridae* family and includes 83% of all viruses within the family and 30% of all known plant viruses (Babu et al., 2011; Frédéric & Juan, 2015). Potyviruses were first reported in 1920 when potato virus Y (PVY) was discovered and was subsequently designated the species type for the genus in 1971 (Gibbs & K. Ohshima, 2010). Harrison et al. (1971) suggested criteria for plant virus classification by allowing additional attributes of viruses to be considered. As a result, a total of 16 unique virus groups were classified and ordered according to resemblances in molecular characteristics. Among the 16 was the *Potyvirus* group with 25 species including PVY and bean common mosaic virus (BCMV) (Harrison et al., 1971).

All viruses classified in the genus *Potyvirus* have a non-enveloped, flexuous, rod-shaped appearance with diameters ranging from 11-15 nm and lengths between 680-900 nm (Berger & Parrish, 2011). Members of the genus are known to infect a number of ornamental and crop plants predominantly in temperate regions such as the South Pacific and some East Asian countries such as The Philippines, China, Vietnam and

Malaysia (Babu et al., 2011; Chen et al., 2001; Pearson & Grisoni, 2002). Reported modes of transport are clonal propagation of infected crops and/or via aphid transmission by a number of species belonging to the *Aphis* and *Myzus* genera in a non-persistent manner, with each aphid species capable of harbouring more than one virus (Babu et al., 2011; Dombrovsky, Huet, Chejanovsky, & Raccah, 2005; Hampton, Jensen, & Hagel, 2005; Riechmann, Laín, & García, 1992; Rohožková & Navrátil, 2011). Due to the rapid increase in the number of potyviruses, members have been divided into groups such as the BCMV group for ease of identification and classification of new and existing viruses.

1.5.1 Potyvirus genome structure

The potyvirus genome structure, outlined in Figure 1.6, consists of a single open reading frame (ORF) flanked by two highly variable 5' and 3' untranslated regions (UTR) in addition to an overlapping ORF. The genome is approximately 10 kb total in length (Berger & Parrish, 2011). The 5' UTR is covalently bonded to the virus encoded viral genome linked protein (VPg) while the 3' UTR is polyadenylated (Berger & Parrish, 2011). The single ORF is translated into a single polyprotein, subsequently processed by three virus encoded proteases into ten proteins, namely protein 1 (P1), helper component proteinase (HC-Pro), protein 3 (P3), 6K1, cylindrical inclusion protein (CI), 6K2, two nuclear inclusion a proteins (NIa-VPg and NIa-Pro), nuclear inclusion b protein (NIb) and the coat protein (CP) (Berger & Parrish, 2011; Chen et al., 2001; Ha et al., 2008). A more recent publication describes an additional overlapping P3N-PIPO ORF in the P3 region of turnip mosaic virus (TuMV) encoding an approximately 25 kDa protein (Chung et al., 2008). Thus, potyviruses encode 11 proteins. This overlapping ORF region is not only common to potyviruses but can also be found amongst other members of the diverse *Potyviridae* family (Chung, Miller, Atkins, & Firth, 2008).

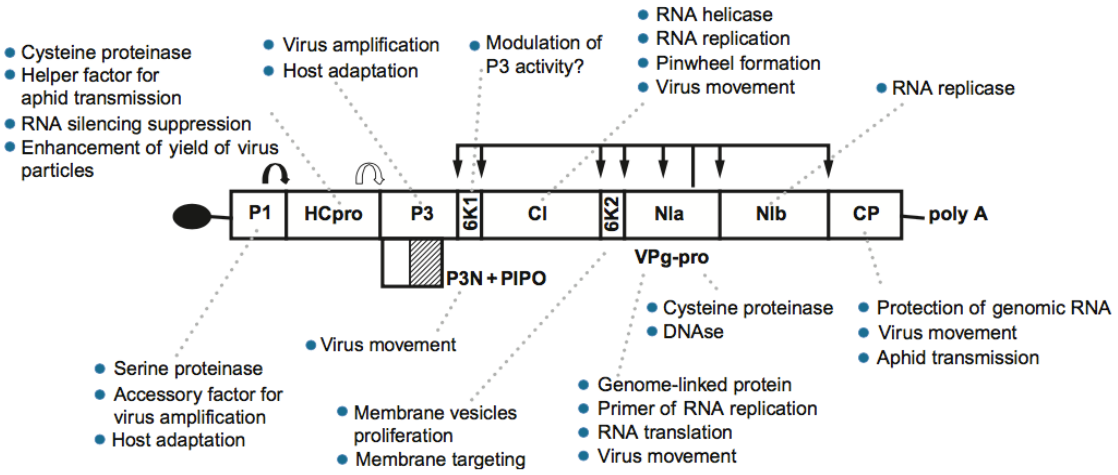


Figure 1.6: The genome organisation of potyviruses consisting of a single ORF, which is subsequently cleaved into ten functional proteins not including P3N-PIPO. General properties and functions of each gene are indicated by the dotted lines. Connected arrows represent cleavage sites processed by the NIa protein while P1 and HC-Pro are self cleaved as indicated by the curved arrows. The genome also consists of a covalently bonded VPg protein at the 5' UTR and a poly (A) tail at the 3' UTR. The recently discovered P3-PIPO conserved region is encoded within the P3 gene of the genome (Frédéric & Juan, 2015).

1.5.2 Potyvirus utr and protein functions

1.5.2.1 5' and 3' untranslated regions

The adenine rich 5'UTR region of the potyvirus genome ranges in length from 144-205 nt with certain regions conserved amongst viruses in certain species within the genus (Riechmann et al., 1992).

Nicolaisen et al., (1992) showed that the 5'UTR of pea seed-borne mosaic virus acts as a translational enhancer while the poly A tail appears to be involved in initiating translation. The 5'UTR enables the entry of ribosomal subunits required for translation. Translation and gene expression appear to be more efficient in viruses containing a 5'UTR than those lacking one (Nicolaisen et al., 1992). This finding was further supported by Zhang et al., (2015) who identified an internal ribosomal entry site (IRES) in the 5'UTR that interacts with the VPg protein to enhance translation in addition to circularisation through the interaction of the 5' and 3'UTRs (Zhang et al., 2015) . Roberts et al. (2017) also identified a 739 nt long IRES-like element in the 5'UTR of triticum mosaic virus that binds to both the eukaryotic initiation factors eIF4G and eIFiso4G subunits, an interaction required for initiating translation.

As well as being highly conserved amongst closely related potyviruses, the 3'UTR may be used in distinguishing viruses at the species level (Nicholson & White, 2011).

1.5.2.2 P1 protease

The P1 gene located at the 5' end of the ORF encodes a protease, which cleaves at a site between itself and the helper component protease (HC-Pro) (Frédéric & Juan, 2015; Verchot, 1995; Verchot & Carrington, 1995a). The P1 protease domain residing at the C-terminal region of the P1 protein is well conserved, however, P1 still remains the most variable gene in the genome owing to the highly evolving N-terminal region (Frédéric & Juan, 2015).

The P1 has been identified as being an accessory in genome amplification, involvement in which has been demonstrated *in trans* (Verchot & Carrington, 1995b). This function has been validated by the insertion of mutations or removal of the N-terminal region and in some cases the entire P1 gene of potato virus A (PVA) and tobacco etch virus (TEV) resulting in a significant decrease in virus accumulation in *Nicotiana benthamiana* by 85-98% (Rajamäki et al., 2005; Verchot & Carrington, 1995a, 1995b). The stimulatory effect of P1 on genome amplification may also be the result of P1 enhancing the silencing suppressor activity of HC-Pro (refer to Chapter 3 Section 3.1.3) (Rajamäki et al., 2005). Studies have shown that P1 may not be required for viral infectivity; however, the cleavage separation of P1 and HC-Pro is essential for virus infection and replication (Shan et al., 2015; Verchot, 1995). A recent study involving constructs of P1a of cucumber vein-yellowing virus (CVYV) (equivalent to potyviral P1), and P1 of plum pox virus (PPV) showed that the potyviral P1 may have host-determining properties (Shan et al., 2015; Valli et al., 2007). These studies suggest that the host-determining factor may be found in either the N-terminus, C-terminus or the entire P1 protein. The P1 of triticum mosaic virus (TriMV), sugarcane streak mosaic virus (SCSMV) and viruses that are taxonomically classified in the *Poacevirus* genus within the *Potyviridae* family, has been found to be primarily responsible for silencing suppression rather than HC-Pro as observed in potyviruses (Tatineni, Qu, Li, Jack Morris, & French, 2012; Young et al., 2012). This evidence further supports the proposition that the P1 protein of potyviruses may be closely associated with silencing-suppression activity.

P1 has also been found to interact with the host protein Rieske Fe/S (Shi, Chen, Hong, Chen, & Adams, 2007), a subunit of the cytochrome b_6f complex enzyme found in the thylakoid membrane in the chloroplasts of plants (Schöttler, Tóth, Boulouis, & Kahlau, 2015). This cytochrome complex is still poorly understood but appears to interact closely with the chloroplast ATP synthase in responding to different environmental stresses and plant development (Schöttler et al., 2015). Implications of this study suggest that the P1 protein maybe involved in symptom development and the variable N-terminus region closely associated with host adaptation (Shi et al., 2007). Other studies however, have shown that the hyper-variable N-terminal region of the P1 protein has negative effects on the proteolytic activity within the C-terminal region (Pasin, Simón-Mateo, & García, 2014).

1.5.2.3 Helper component protease (HC-Pro)

The name of this protein derives from its first discovered function as a helper component in aphid transmission (Govier, Kassanis, & Pirone, 1977). HC-Pro is a cysteine protease that cleaves itself at its C-terminus (Goytia, Fernández-Calvino, Martínez-García, López-Abella, & López-Moya, 2006; Hasiów-Jaroszewska, Fares, & Elena, 2014; Maia, Haenni, & Bernardi, 1996). The self-cleaving protein appears to accumulate in the form of cytoplasmic amorphous inclusions in some viral infections such as pepper mottle virus (PeMV), and in other cases such as PPV, it accumulates as a soluble protein (De Mejia, Hiebert, Purcifull, Thornbury, & Pirone, 1985; Ravelonandro, Peyruchaud, Garrigue, de Marcillac, & Dunez, 1993). Three functional regions have been identified within HC-Pro (Figure 1.7), namely the N-terminal region which is not essential for self-interaction but required for aphid transmission, and the larger C-terminus which is associated with polyprotein processing by its self-cleaving ability (Plisson et al., 2003). The third domain within HC-Pro is the central region comprised of approximately 200-250 amino acids, responsible for most of the HC-Pro functions (Kasschau & Carrington, 2001; Plisson et al., 2003).

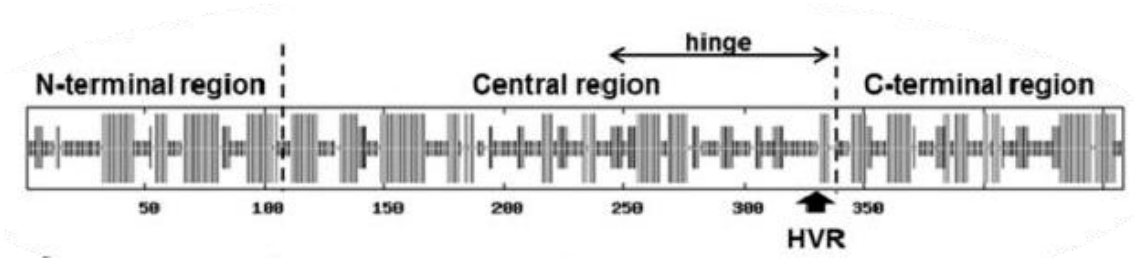


Figure 1.7 – Three domains within HC-Pro of potyviruses consisting of the N-terminal region, central or core region and the C-terminal region. Numbers below indicate the number of amino acids. The hinge region indicated within the central domain has been found to be associated with a number of HC-Pro functions including RNA silencing suppression. The highly variable region (HVR) indicated by the arrow is also associated with the interaction of microtubules (MT)-associated HC-Pro interacting protein 2 (HIP2) in plants. (Tuuli Haikonen, Rajamäki, Tian, & Valkonen, 2013).

One of many functions of HC-Pro is its ability to suppress RNA silencing in plants allowing the virus to enter and replicate in the host cell (Ala-Poikela, Goytia, Haikonen, Rajamäki, & Valkonen, 2011; Frédéric & Juan, 2015). The multifunctional nature of HC-Pro involves its RNA silencing suppression activity, which is not restricted to any particular domain within the protein (Kasschau et al., 2003; Torres-Barceló, Daròs, & Elena, 2010). In addition, the HC-Pro protein has been found to interact with a number of host proteins and other viral proteins such as CP, CI, VPg, NIa and P1 (Ala-Poikela et al., 2011; Frédéric & Juan, 2015).

One study found that TEV HC-Pro binds to a *N. benthamiana* regulator of gene-silencing calmodulin-related protein (rgs-CaM), a protein that is associated with cellular silencing suppression. This interaction may activate rgs-CaM, a calmodulin-like protein in the host often associated with antiviral activity, to suppress silencing in plants (Anandalakshmi, Marathe, Ge, Herr, & et al., 2000). The rgs-CaM–HC-Pro interaction involves a calcium-binding motif in the C-terminal region and 40-50 amino acids in the N-terminal region of rgs-CaM, which may specify the location of interaction with HC-Pro (Anandalakshmi et al., 2000). The same study also revealed that HC-Pro is able to interact with the ethylene-inducible transcription factor RAV2 in mediating the counter-defensive mechanism of viruses against RNA silencing (Anandalakshmi et al., 2000).

The success of HC-Pro in silencing suppression can also be related to the disruption of siRNA and/or miRNA pathways rendering them inactive and, therefore, ineffectively

carrying out RNA silencing (Jamous et al., 2011; Shiboleth et al., 2007). A recent study identified an interaction between the zucchini yellow mosaic virus (ZYMV) HC-Pro and the RNA methyltransferase HEN1 (Hua Enhancer 1) in transgenic *Arabidopsis thaliana.* HEN1 methylates siRNAs at the 3' terminus hydroxide overhang before association with Argonaute. Binding to ZYMV HC-Pro disables the methylation step thereby inhibiting HEN1 activity in RNA silencing (Figure 1.8) (Jamous et al., 2011).

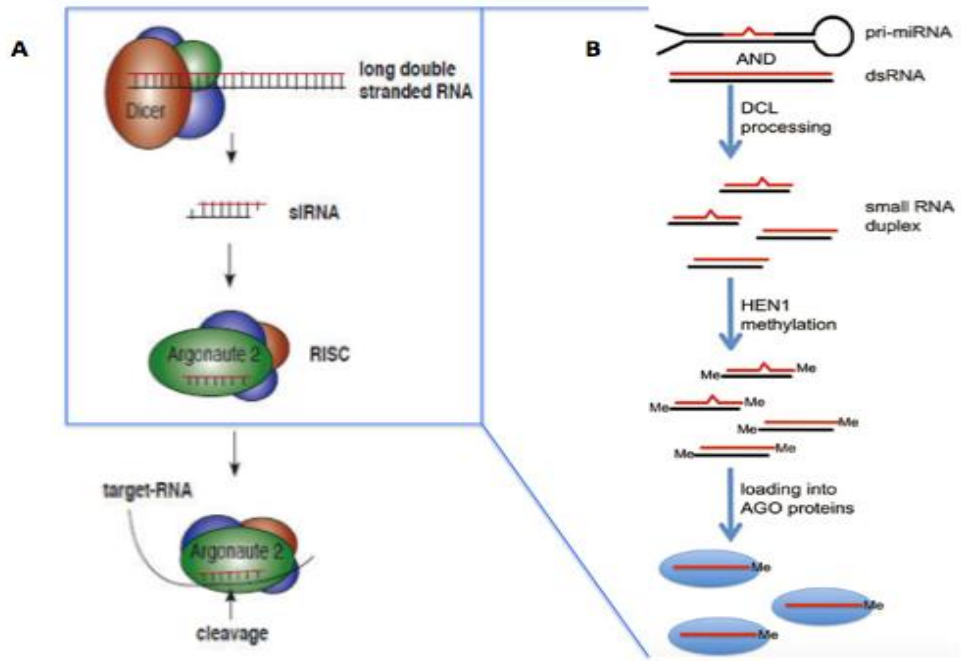


Figure 1.8: A) Summary of the major steps involved in the siRNA pathway. Long double stranded DNA is processed into siRNA fragments approximately 20-25 nt in length by a DICER complex. These siRNA fragments are then loaded on to a RISC complex containing an argonaut protein separating the two strands in which one will act as a guide strand to degrade target-RNA. **B)** A close-up view of the siRNA pathway outlined by the blue square, in which the HEN1 processing step takes place. The methylation of siRNA takes place shortly after it's processing by DICER (or DCL) proteins and just before they are loaded on to argonaut (AGO) protein complex. Retrieved from http://www.biologie.uni-regensburg.de/Biochemie1/Research/index_1.htm

The ubiquitin/26S proteasome system (UPS), primarily associated with protein removal and degradation in the host, is also involved with counter-attacking pathogenic infections such as viruses (Dielen, Badaoui, Candresse, & German-Retana, 2010; Vierstra, 2009). The UPS degradation machinery consists of a core 20S proteasome element with seven α and seven β subunits each capable of interacting with each other where a combined effect can be more efficient (Dielen et al., 2010). In a recent study, a conserved region (KITC) within the HC-Pro N-terminus of papaya ringspot virus (PRSV) interacted with the 20S proteasome α1 subunit in *A. thaliana* and its

homologue in papaya (Sahana et al., 2012). PRSV also interacted with the α5 subunit of *A. thaliana* but did not respond to its homologue in papaya (Sahana et al., 2012). A similar result was observed when the same region in lettuce necrosis virus (LNV) HC-Pro made contact with the α5 subunit rather than the α1 resulting in an inhibited proteasome function (Sahana et al., 2012). These findings indicate that different potyvirus HC-Pro proteins, through the KITC motif, can interact with different 20S proteasome subunits impairing its catalytic functions leading to an increase in viral titre after virus infection.

Tuuli Haikonen et al. (2013) identified a microtubule (MT) associated protein in potato (StHIP2) and tobacco (NtHIP2) interacting with the hinge region within the central domain of PVA HC-Pro. When the HIP2 proteins were depleted, virus accumulation reduced significantly indicating the interaction between the host HIP2 protein and potyviral HC-Pro is necessary for virus infection and accumulation. These findings have narrowed down the region conferring the ability to supress RNA silencing and infection of the host in this particular case to the hinge region within the central domain of HC-Pro as outlined in Figure 1.7.

Another function of the HC-Pro is its involvement in aphid transmission. It has been shown that aphids appear to have access to the HC-Pro either during or before it has contact with the virion particle (Goytia et al., 2006). The ability of the aphid to harbour the virus may involve HC-Pro acting as a bridge between the aphid mouthpart and the virion particle aiding transmission (Frédéric & Juan, 2015; Goytia et al., 2006).

A number of studies have shown that HC-Pro may also be involved in viral amplification and long distance movement where this activity is mediated by a number of conserved regions in the core and N terminal region (Ala-Poikela et al., 2011; Tuuli Haikonen et al., 2013; Kasschau & Carrington, 2001).

Recent studies have identified HC-Pro as enhancing the stability of the coat protein (CP) (Valli et al., 2014). When a chimeric PPV was introduced into *N. benthamiana* with a heterologous silencing suppressor in place of HC-Pro, the virus was able to suppress silencing however; viral progeny had reduced infectivity. In this study, Adrian Valli et al. (2014) demonstrated the requirement of HC-Pro in enhancing the stability

of its CP by site-directed mutagenesis, a function that appears to be highly specific and cannot be complemented by a heterologous potyvirus.

1.5.2.4 P3, 6K1 and P3N-PIPO

P3 is one of the least characterised potyviral proteins studied to date. In contrast to a number of other potyviral proteins such as P1, HC-Pro and CI, P3 along with 6K1 and 6K2 do not bind to viral RNA (Merits, Guo, & Saarma, 1998). A study on the locality of the P3 protein in relation to infected plant tissue indicated a close interaction of the P3 protein with cylindrical inclusions found in the cytoplasm of TVMV-infected *N. benthamiana* (Rodríguez-Cerezo, Ammar, Pirone, & Shaw, 1993). Other findings revealed that TEV and PVA P3 proteins interact with their respective NIb proteins (Merits et al., 1999; Langenburn & Zhang, 1997). In addition, immunocytological studies localised TEV P3 in the nucleolus and nucleus of infected *N. tabacum* cells, suggestive of a role in the early stages of viral replication (Langenburn & Zhang, 1997). More recently, transient expression tests revealed that TEV P3 is directed to the membranes of the endoplasmic reticulum (ER) and creates inclusions closely affiliated with the Golgi apparatus inside the host cell (Cui, Wei, Chowda-Reddy, Sun, & Wang, 2010). This attachment to the ER is largely due to two hydrophobic regions present in the C-terminal region of the P3 protein (Cui et al., 2010). A yeast two-hybrid screen and a co-immunoprecipitation assay revealed that P3 interacted with the host ribose-1, 5-biphosphate, carboxylase/oxygenase (RubisCO) protein, possibly inhibiting its function, and thus contributing to symptom development (Bhat et al., 2013; Lin et al., 2011). The P3 protein, to date, has been found to enhance viral pathogenicity, viral replication and host symptomology; however, the exact mechanisms of how P3 carries out these activities remain obscure (Klein, Klein, Rodriguez-Cerezo, Hunt, & Shaw, 1994).

P3N-PIPO of TuMV has been found to interact with the PCaP1, a calcium binding protein located in the membrane of *A. thaliana,* allowing cell-to-cell movement (Vijayapalani, Maeshima, Nagasaki-Takekuchi, & Miller, 2012). Riechmann et al., (1995) speculated that P3 and 6K1 may be the main functional products in potyvirus infection and their proteolytic processing would have a regulatory function. Supporting this hypothesis, Waltermann and Maiss (2006) identified 6K1 as a distinct

6kDa protein product, which Hong et al., (2007) later identified as occupying the cell periphery independent to that of the P3 involvement with the ER. An undefined role has been proposed for 6K1 in potyvirus multiplication based on these findings (Kekarainen, Savilahti, & Valkonen, 2002).

1.5.2.5 Cylindrical inclusions (CI)

The CI protein forms pinwheel shaped inclusion bodies that are typically found in the cytoplasm of infected cells (Edwardson & Christie, 1996). Like HC-Pro, CI is a multifunctional protein having ATPase and RNA helicase activity required for RNA virus replication (Fernández et al., 1997; Sorel, Garcia, & German-Retana, 2014). Other studies have also revealed that TuMV CI interacts closely with the P3N-PIPO protein to assist the virus in cell-to-cell movement in *N. benthamiana* (Sorel et al., 2014; Wei et al., 2010). In this process, the P3N-PIPO protein localises to the plasmodesmata (PD), microscopic cytoplasmic channels between plant cells, directs the CI protein towards the PD and allows intercellular movement of the virus (Wei et al., 2010). This was further validated when results from an immunogold-labelling electron microscope study revealed a portion of PVA CI obtained from the preparation was localised to one end of the virion. This finding suggests that CI may provide motor function to assist virus translocation through PD and virus disassembly (Gabrenaite-Verkhovskaya et al., 2008). Another study had revealed that the CI protein acts as a virulence factor for a number of different host resistant genes (Sorel et al., 2014). The C-terminal region of the CI protein has been shown to interact with the lettuce mosaic virus (LMV) VPg in overcoming resistance encoded by the lettuce *mo1* gene, a gene which encodes the translation initiation factor 4E (Tavert-Roudet et al., 2012). Other findings have indicated a close interaction between PPV CI and the host photosystem 1 PSI-K protein, a product of the *psaK* gene in *N. benthamiana* (Jiménez, López, Alamillo, Valli, & García, 2006). This study shows that by down-regulating the *psaK* gene in *N. benthamiana*, virus accumulation increased, suggesting a role for the CI - PSI-K interaction in virus infection (Jiménez et al., 2006). LMV and TEV CI have also been shown to interact with the $P58^{IPK}$ (mammalian cellular inhibitor) homologue of *N. benthamiana*. This interaction appears to be required by viruses for viral multiplication and pathogenesis (Bilgin, Liu, Schiff, & Dinesh-Kumar, 2003).

1.5.2.6 Nuclear Inclusion a (NIa) and 6K2

The NIa gene encodes the largest potyviral protein, which forms crystalline inclusions commonly found in the nucleus however, some have also been identified in the cytoplasm of TuMV infected *N. benthamiana* (Cotton et al., 2009; Frédéric & Juan, 2015). The NIa protein is partially processed to form the NIa-VPg and NIa-Pro proteins, two domains with functions that appear to be independent of each other (Dougherty & Dawn-Parks, 1991; Elena & Rodrigo, 2012).

NIa-Pro is responsible for the proteolytic processing of the central and C-terminal regions of the potyviral polyprotein (Adams, Antoniw, & Beaudoin, 2005; Sun, Austin, Tözsér, & Waugh, 2010). Variation in cleavage efficiency of NIa-Pro suggests that the maturation of potyviral proteins is highly controlled and may play an important role in regulating virus infection (Frédéric & Juan, 2015). In addition to its proteinase activity, NIa-Pro also possesses DNase activity (Anindya & Savithri, 2004). It has been speculated that the degradation of host DNA by the NIa protein taking place in the nucleus might play a role in host gene expression relative to viral infection (Anindya & Savithri, 2004).

Jiang and Laliberté (2011) showed that NIa-VPg behaves as a hub protein involved in regulating a number of viral processes such as viral RNA translation, replication and host gene regulation. A study carried out by Hari (1981) identified a viral linked VPg protein attached to the 5' end of tobacco etch virus (TEV). Beauchemin and Laliberté (2007) found in TuMV infections, VPg is found as a 6K2 – VPg - NIa-Pro intermediate product and is confined to the membranous virus factories of the cell. This suggests a likely role of VPg in viral RNA replication and viral infectivity (Beauchemin & Laliberté, 2007). Mathur and Savithri (2012) suggest that VPg is an intrinsically disordered protein that undergoes structural changes when interacting with NIa-Pro to induce ATPase activity of the polyprotein intermediate. A study carried out by Rajamäki, Streng, and Valkonen (2014) indicated a possible role for VPg in suppression of RNA silencing by interfering with the host cell's RNA dependant RNA polymerase 6 (RDR6) that is required to convert single stranded RNA into the prerequisite dsRNA necessary to trigger silencing. VPg can also undergo post-translational modification by multiple phosphorylation events, which can be important in regulating other functions in the

cell (Hafrén & Mäkinen, 2008). A recent review by Frédéric and Juan (2015) reported the interaction of VPg with not only most potyviral proteins, but also with many host factors such as the eukaryotic initiation factor (eIF4E), the nucleolar protein fabrillarin, Poly A binding protein (PABP) and a RNA helicase-like protein identified in peach and Arabidopsis (AtRH8). This functional diversity can be accredited to the structural flexibility of the protein and its intrinsically disordered nature (Rantalainen, Eskelin, Tompa, & Mäkinen, 2011).

1.5.2.7 Nuclear Inclusion b (NIb)

The NIb and NIa proteins together form the nuclear crystalline inclusions described above which can be found in the nucleus and cytoplasm (Frédéric & Juan, 2015; Oruetxebarria et al., 2001). The NIb protein is the RNA-dependant RNA-polymerase (RdRp), which is necessary for potyviral genome replication (Hong & Hunt, 1996). It is thought that RNA replication takes place when NIb interacts with the VPg and NIa-Pro proteins of the 6K2 – VPg – NIa-Pro product (Dufresne et al., 2008). Interactions of NIb with host proteins such as eEF1A, PABP and Hsc70-3 contribute to the formation of efficient replication complexes (Dufresne et al., 2008). A recent study demonstrated how NIb of pepper vein banding virus (PVBV) uridylylates VPg and by its binding to the poly A tail of the viral genome uses the resultant product to prime viral RNA synthesis (Anindya, Chittori, & Savithri, 2005). A form of the Small ubiquitin-like modifier (SUMO) enzyme known as SCE1 found in Arabidopsis has been found to interact with the NIb protein of TuMV in both the nucleus and cytoplasm of the host cell (Xiong & Wang, 2013). This interaction may impact a number of processes such as efficiency of potyviral infection, the SUMOylation regulating NIb activity, or the NIb/SCE1 complex inhibiting the pattern of SUMOylation in cellular proteins. The interaction in each of these cases could in turn create an environment which favours viral multiplication (Xiong & Wang, 2013).

1.5.2.8 Coat protein (CP)

The CP (also known as the capsid protein) functions primarily in encapsulating the genomic RNA of potyviruses (Adams, Antoniw & Fauquet, 2005). About 2000 subunits of CP helically arrange themselves around the genomic RNA of the virus forming flexuous rods of the potyviral virion approximately 11-15 nm in diameter and 680-900

nm in length (Frédéric & Juan, 2015). The CP also functions in vector transmission, regulation of host defence response, cell-to-cell movement and long distant transport (Dolja, Haldeman, Robertson, Dougherty, & Carrington, 1994; Rakitina et al., 2005). A study performed by Ivanov et al. (2003) showed that the CP of PVA is phosphorylated by the host protein kinase CK2, playing an important role in regulating virus infection. It has been hypothesised that the post-translational modifications of CP may behave as control elements regulating genomic RNA allocated for translation, replication and propagation during different steps of the infection cycle (Frédéric & Juan, 2015). Cellular chaperones and the NTPase activity of CP are also thought to be relevant in the described regulatory mechanism (Aparicio et al., 2005; Rakitina et al., 2005). Like P3, CP has also been reported to interact with the host protein Rubisco suggesting a vital role for this host protein in virus infection and/or plant defensive responses (Bhat et al., 2013). The CP gene sequence has also been widely used to discriminate between viral species and to establish taxonomy and evolutionary relationships (Adams, Antoniw & Fauquet, 2005).

1.6 Potyvirus evolution

Initial attempts to decipher the relationships of potyviruses were carried out over 50 years ago, primarily evaluating serological properties using virions as antigens (Gibbs & Oshima, 2012). These attempts encountered a number of difficulties such as the appearance of a continuum of variants and strains linking distinct species. In addition, different relationships have been observed when different antisera and antigens have been used (Gibbs & Oshima, 2010). Further developments, however, allowed the comparison of viruses at the nucleotide and amino acid levels with the former being more informative and consequently providing more evolutionary insight (Gibbs & K. Ohshima, 2010; Moghal & Francki, 1976). It has been proposed that potyviruses may have emerged approximately 6600 years ago during the dawn of agriculture (Gibbs, Ohshima, Phillips, & Gibbs, 2008a; Koonin, Wolf, Nagasaki, & Dolja, 2008). Other studies suggest the lineage leading to potyviruses is 7250 years old (Gibbs & K. Ohshima, 2010) with a significantly high mutation rate making RNA viruses highly evolving and more divergent compared to other pathogens (Sztuba, Urbanowicz, Figlerowicz, & Bujarski, 2011).

1.6.1 Quasi-species

The RdRp of RNA viruses lack proof reading ability leading to high mutation rates and rapid evolution for this type of virus (Sztuba et al., 2011). Such mutation and natural selection have been proposed as the cause of rapid shifts in viral populations, better understood as quasi-species. Quasi-species is a mathematical framework used to understand rapid evolution and population genetics in which a master sequence, through rapid mutations, produces a cloud of all possible variants (Domingo, Sheldona, & Peralesa, 2012 ; Lauring & Andino, 2010). The quasi-species concept was applied to rapidly evolving RNA viruses to explain the high mutation rate (estimated as 1.15×10^{-4} nucleotide substitutions/site/year, Gibbs & Oshima, 2010) and to predict the likely outcomes of virus evolution (Sztuba et al., 2011). Due to the lack of proof reading ability by the replicase, potyviruses are able to produce many point mutations within each generation, which are amplified by large population sizes, a concept known as 'sequence space' (Sztuba et al., 2011). This concept attempts to explain the genetic organisation of a population by understanding that with every cycle of replication, there will be at least one replication that produces a genome that will differ in nucleotide sequence, which also then undergoes many rounds of replication expanding the variation of the population in its sequence space (Figure 1.9) (Domingo et al., 2012 ; Lauring & Andino, 2010). What was the original 'master sequence' can become a different sequence due to genetic drift, positive selection or negative selection acting on the different variant sequences within the sequence (depicted in Figure 1.9). This may explain the different isolates/strains found and an increasing number of potyviruses, which in turn can individually act as 'master sequences' for further mutation and natural selection (Domingo et al., 2012 ; Lauring & Andino, 2010; Sallie, 2005). Using this concept, it can be proposed that VanMV and DsMV are variants of an earlier strain within a sequence space. It is also possible that VanMV was a later or earlier variant of DsMV, or vice versa. In order to fully understand and evaluate the current position of VanMV in potyvirus evolution and sequence space, its complete genome sequence is required as a basis for phylogenetic analysis.

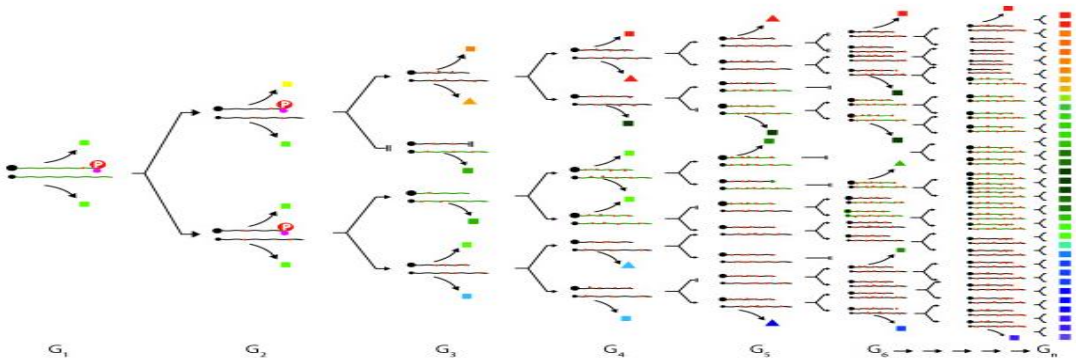


Figure 1.9: Illustration of the quasi-species model. G1 refers to the original master sequence from which a number of random mutations can be created. As generations proceed (G_2-G_n), these mutations, in conjunction with others may be passed on to further generations. A spectrum of all possible variants maybe produced known as the sequence space, which could further act as master sequences for further mutation and natural selection (Sallie, 2005).

1.6.2 Recombination

Recombination is also an important phenomenon when studying variation in genetic populations and evolutionary relationships (Sztuba-Solińska, Urbanowicz, Figlerowicz, & Bujarski, 2011). Recombination is a naturally occurring event in which genetic information is exchanged between two sequences (Boulila, 2010a). This exchange may occur by homologous crossing over of nucleic acids or re-assortment of genomic material as shown in Figure 1.10 (Lauring & Andino, 2010). The exchange of genetic material can also be closely associated with quasi-species where a single viral isolate is comprised of a number of variants within the population. These variants may in turn undergo recombination events with each other producing sequences with distinct virulence (Boulila, 2010a; Domingo, Sheldona, & Peralesa, 2012; Lauring & Andino, 2010). Recombination events can be detected computationally using a number of published detection methods such as GENECONV, BOOTSCAN, MAXIMUM chi(2), CHIMAERA and SISTER SCANNING by making nucleotide sequence alignments and comparisons (Darren P. Martin et al., 2010). A number of recombination events have been identified in the CP genes of plum pox virus (PPV), apple chlorotic leaf spot virus (ACLSV), apple mosaic virus (ApMV), prune dwarf virus (PDV) and prunus necrotic ringspot virus (PNRSV) with virulence properties distinct from each other (Boulila, 2010). A review of RNA-RNA recombination by Sztuba-Solińska et al. (2011) reported a number of cases where RNA viruses were crossing over genomic material at around the CP and 3' UTR regions of the genome. Recombination events have also been

observed in the P1 region of WMV in which the N terminal region showed no relation to soybean mosaic virus (SMV), its closest relative, but revealed high aa sequence similarity to BCMV (Desbiez & Lecoq, 2004a). This finding suggests that WMV arose as a result of a recombination event between SMV and BCMV with virulence properties distinct from either parent. Further, a recombination event was detected in the P1 region of SMV and DsMV enabling the SMV variant to infect *Pinellia ternata*, a natural host for DsMV but not for SMV, and still maintained the ability to infect some soybean cultivars, a natural host for SMV (Valli et al., 2007). This study highlights the significant impact recombination can have on virus evolution and host range determination.

The numerous reports of recombination events documented demonstrate the potential for a rapid shift in RNA virus populations and greater potential to infect plants more efficiently over classical strains (Gibbs & K. Ohshima, 2010; Lecoq et al., 2011). Recombination could be the explanatory mechanism behind the emergence of DsMV and VanMV. It is possible that DsMV may have recombined with another virus to become VanMV, loosing its ability to infect aroids but gaining the ability to infect vanilla. This recombination event could also be in a region of the genome with host specific factors with a preference for orchid plants rather than aroids (Bedhomme, Lafforgue, & Elena, 2012).

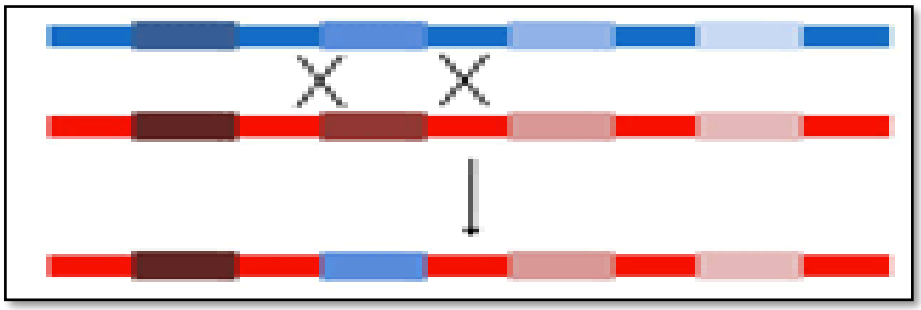


Figure 1.10 A schematic representation of a typical recombination event taking place between two virus strains producing a third strain with some similarities to the original strains.

1.7 RNA silencing

RNA silencing, also known as RNA interference (RNAi), is a natural defence mechanism acquired by eukaryotic species to counter various infections caused by pathogenic organisms (Karthikeyan et al., 2013) . The mechanism not only safeguards the

organism from exogenous foreign genetic material, but also involves post-transcriptional regulation of gene expression (Karthikeyan et al., 2013; Katoch & Thakur, 2013). Its incidental discovery in the early 1990s was a result of an attempt to up-regulate an enzyme responsible for colour pigmentation in petunias. The result was not an intensified pigmentation but rather a variegation and in some cases complete loss of colour from the petals (Napoli, Lemieux & Jorgensen, 1990). This was a clear indication that not only was the exogenous foreign material inactivated, the expression of endogenous loci was also suppressed or silenced, and this silencing occurred at the RNA level (Angaji et al., 2010; Karthikeyan et al., 2013).

RNAi may occur by one of two main pathways (Figure 1.11), the micro RNA (miRNA) (Figure 1.11A) and the small interfering RNA (siRNA) (Figure 1.11B) mediated silencing pathways. In both cases, a longer double stranded RNA (dsRNA) is processed to short RNA molecules (siRNA and miRNA) (Angaji et al., 2010; Kusaba, 2004). In both pathways, the small RNAs serve as specificity elements that direct cellular proteins to target RNA molecules for silencing by degradation or translation inhibition (Carthew & Sontheimer, 2009). The multistep process involves a number of host specific proteins that differ slightly between each pathway; however, the overall principle and mode of action remains the same (Angaji et al., 2010; Kusaba, 2004; Marques et al., 2013). For the miRNA pathway (Figure 1.11A), the source of dsRNA is endogenous RNA transcribed by RNA polymerase II (Lam, Chow, Zhang, & Leung, 2015). The miRNA pathway differs from the siRNA pathway in that the starting dsRNA material (pri-miRNA) has partial complementarity as the pri-mRNA forms a hairpin structure (Figure 1.11A step 2). The pre-miRNA is further processed by a Dicer-like complex (DCL1) to form shorter 20-25 nt structures (step 2-3) followed by the incorporation of a 'guide RNA strand' into an Argonaute protein to form a RNA induced silencing complex RISC (step 4) (Borges & Martienssen, 2015; Lam et al., 2015). While both pathways operate similarly, they each serve a different purpose where the miRNA pathway is primarily concerned with the negative regulation of gene expression, and the siRNA pathway is involved with RNA silencing by degradation or translation inhibition (Angaji et al., 2010)(Lam et al., 2015). In addition, each pathway involves different variants of DCL, double stranded RNA binding (DRB) and Argonaute proteins (Llave, 2010). When RNAi is induced by a viral infection the most likely mode of silencing is via the siRNA

silencing pathway as the foreign genetic material is exogenous and requires silencing rather than down-regulation of gene expression (Carthew & Sontheimer, 2009; Kusaba, 2004).

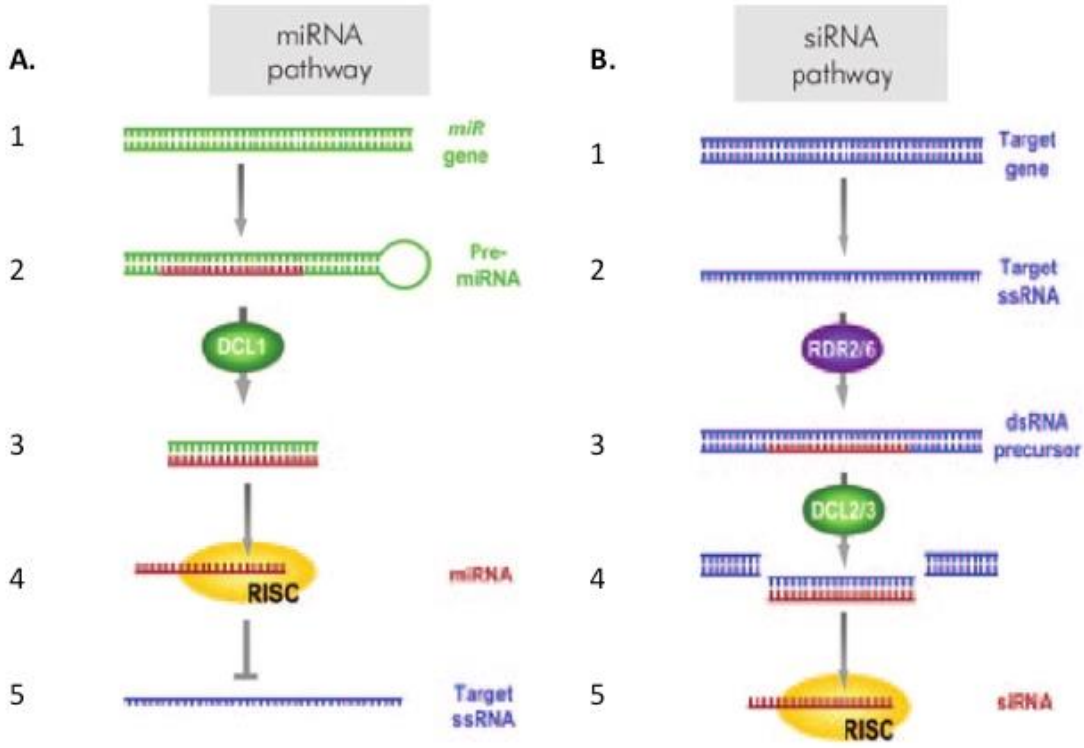


Figure 1.11 Comparison of the **A)** miRNA pathway. Pre-miRNA is transcribed from genomic RNA (step 1) and forms a loop with partial complementarity (step 2). This pre-miRNA undergoes processing by a Dicer-like (DCL1) complex producing 20-25 nt long miRNA (step 3). This processed pre-miRNA is separated into in to two strands one being the guide and the other being a passenger strand. The passenger strand (green) is degraded while the guide strand (red) is incorporated into the RISC complex (step 4). The RISC complex containing the guide strand then targets ssRNA of interest for translation inhibition or negative gene regulation (step 5) **B)** siRNA pathway. A single stranded mRNA transcript is transcribed by polymerase IV (step 2). A complementary strand is synthesised by RDRP2 or RDRP6 to form a long dsRNA precursor (step 3). The dsRNA precursor molecule is processed by a DICER-like (DCL2 or DCL3) complex into shorter 20-25 nt structures (step 4). A guide strand (red) is incorporated into an Argonaute protein to form a RISC complex while a passenger strand (green) is degraded. The RISC complex containing the guide strand then targets foreign genetic material for degradation. Retrieved from http://www.gene-quantification.de/si-rna.html.

1.7.1 siRNA pathway

Unlike the miRNA pathway, the siRNA pathway is induced as a protective mechanism response against the entry and expression of foreign genetic material. Upon entry into the cell, viral ssRNA genomes will replicate rapidly, producing dsRNA intermediates in the process. The plant recognises these structural forms and processes them to form functional viral short interfering RNA (vsiRNA) (Shimura & Pantaleo, 2011). Depending

on the starting genetic material, vsiRNAs can be produced in four distinctive ways as outlined in Figure 1.12. Different DCLs are able to produce vsiRNA which vary in length but are normally between 20-25 nucleotide bases long (Llave, 2010; Martínez de Alba, Elvira-Matelot, & Vaucheret, 2013; Padmanabhan, Zhang, & Jin, 2009; Shimura & Pantaleo, 2011; Xie & Qi, 2008). Some single stranded RNA (ssRNA) viruses whether (+) sense, (-) sense or circular, require the synthesis of a complementary strand by either the host cell's RdRp (Figure 1.12C) or by a viral replicase (Figure 1.12A) (Llave, 2010; Shimura & Pantaleo, 2011; Xie & Qi, 2008). In some cases, rather than synthesising a complementary strand, genomic RNA may form hairpin structures forming a loop with partial complementarity exhibiting the form of dsRNA molecules (Figure 1.12B) (Shimura & Pantaleo, 2011). Circular DNA viruses form dsRNA through the overlaps formed within the more than full-length genomic transcripts or on opposing and overlapping transcripts. These dsRNA molecules, whether synthesised by viral replicase (Figure 1.12A and 1.12C), created by a hairpin loop (Figure 1.12B) or already existing as a dsRNA molecule (Figure 1.12D), are then targeted by DCL, an RNase III nuclease for RNA processing (Shimura & Pantaleo, 2011; Xie & Qi, 2008). These long dsRNA strands are then cut by the PAZ region of the DCL protein into shorter strands known as vsiRNA, consisting of 5' phosphates and 3' overhangs of two nucleotide bases (Llave, 2010; Padmanabhan et al., 2009; Shimura & Pantaleo, 2011). In some cases, hyponastic leaves1 (HY1), a group of DRB proteins, have been found to interact with DCL proteins to assist and enhance the production of vsiRNA (Xie & Qi, 2008). This has been shown for DCL1 and DCL4 proteins interacting with DRB1 and DRB4 respectively to produce shorter *trans-acting* (ta) siRNAs and siRNAs from viral RNA while DCL2 and DCL3 require no assistance from DRB proteins (Curtin et al., 2008)

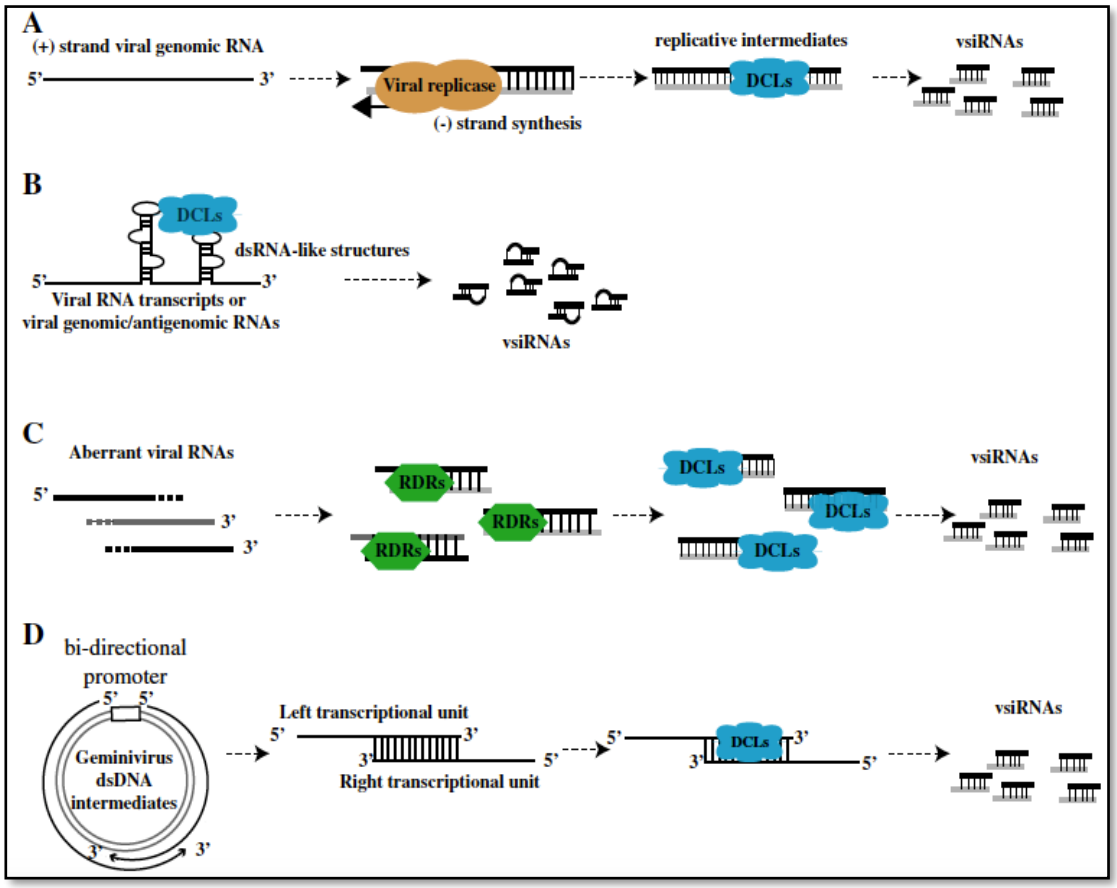


Figure 1.12 A schematic diagram of the four possible ways vsiRNA's can be produced depending on the starting genetic material. **A)** Represents a pathway producing vsiRNA from (+) sense viral RNA. **B)** Single stranded RNA producing overlapping dsRNA-like molecules, which in turn produce vsiRNA. **C)** Starting material involves other ssRNA molecules such as (-) sense RNA genetic material undergoing a number of steps to produce vsiRNA. **D)** dsDNA unwinding to synthesis complementary strands before undergoing dicing by DCL's (Dicer-like proteins) to produce vsiRNA (Shimura & Pantaleo, 2011).

Once the vsiRNAs are synthesised and before they are loaded onto Argonaute, they are further modified by the methylation of the 2'-hydroxyl group of the ribose on the 3' overhang by a dsRNA methyltransferase known as HUA ENHANCER1 (HEN1) (Figure 1.13B) (Vazquez, Legrand, & Windels, 2010; Xie & Qi, 2008). This methylation step protects the vsiRNAs from nucleolytic activity or other types of modifications such as oligouridination (Jamous et al., 2011; Xie & Qi, 2008). Once the methylation step is complete, the vsiRNAs are then loaded on to an Argonaute protein where a 'guide strand' remains intact with Argonaute while the complementary 'passenger strand' is subsequently cleaved. This cleavage is catalysed by the endogenous endonuclease activity of the PIWI domain where the 'passenger strand' is removed and degraded (Danielson & Pezacki, 2013; Xie & Qi, 2008). The RISC complex containing the guide RNA-bound Argonaute identifies complementary sequences in the cytoplasm of the

cell and initiates their cleavage by the Argonaute protein (Figure 1.13A). This cleavage initiates complete degradation of the mRNA by other nucleases thereby preventing its translation (Xie & Qi, 2008).

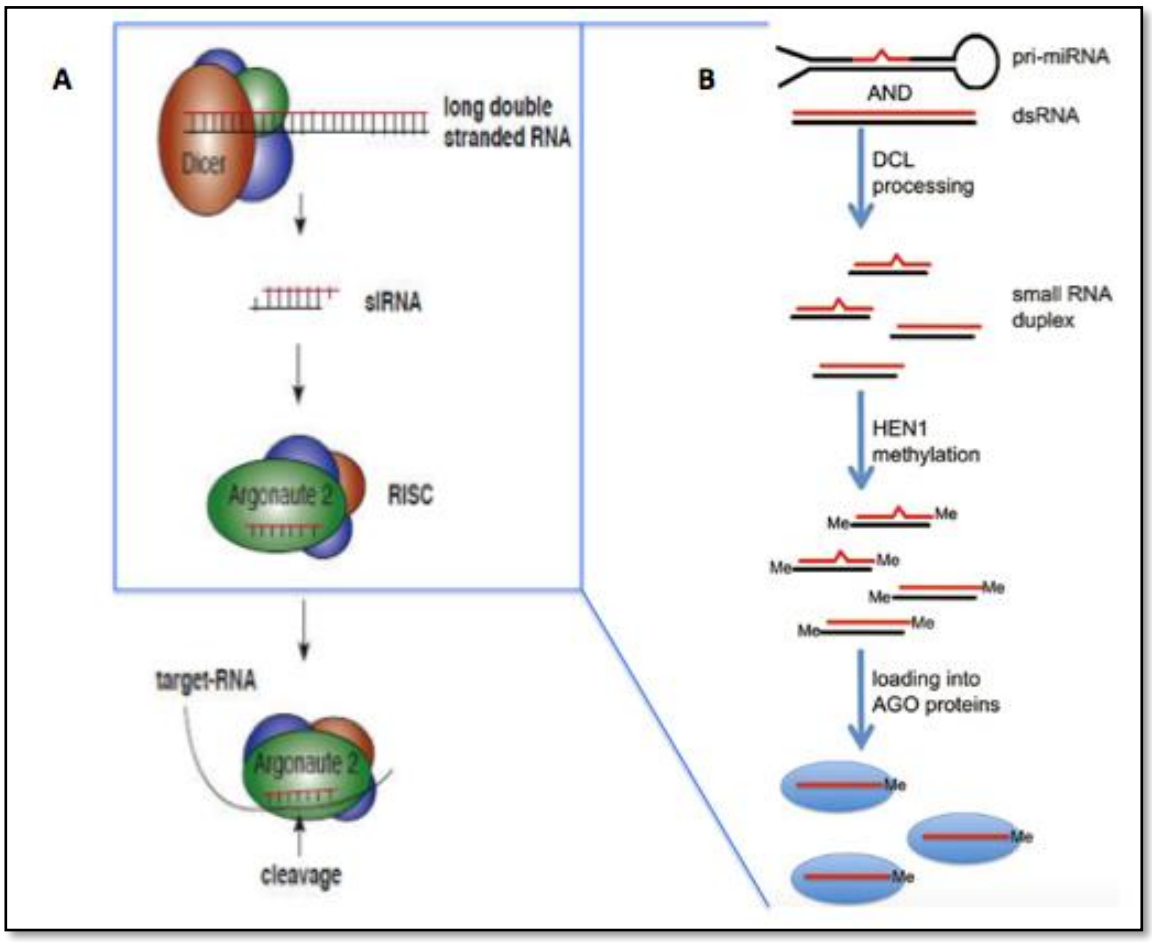


Figure 1.13 A schematic representation of the A) siRNA pathway. Dicer cleaves the long dsRNA molecule into short 20-25nt of siRNA (if silencing was initiated by a virus infection then it would become vsiRNA). Before the siRNA is bound to Argonaute, it undergoes B) a methylation step which stabilises the 'guide strand'. Once the methylation step is complete the methylated strand is loaded onto Argonaute removing the passenger strand. The RISC complex now containing the guide strand then searches for a complementarity match and subsequently cleaves the target RNA (mRNA) and degrades it preventing further translation. Retrieved from http://www.gene-quantification.de/si-rna.html.

1.8 Potyvirus suppressor of RNA silencing

While plants have RNAi as a defensive mechanism against viruses, potyviruses, as well as other viruses, have evolved silencing suppressors to counter-attack this (Gammelgard et al., 2007). The potyvirus HC-Pro, located in the 5' region of the genome upstream of the P1 protein, functions as a viral suppressor of RNA silencing (Haikonen et al., 2013). Different potyvirus HC-Pro proteins counterattack RNAi in

various ways, an interaction that is dependent on the sequence composition of HC-Pro and host (Anandalakshmi et al., 2000; Nakahara et al., 2012; Sahana et al., 2012). These various interactions are discussed further in Chapter 3.

1.9 Aims of study

Vanilla production in the South Pacific has been severely hampered due to a number of orchid infecting viruses, one of which is the potyvirus VanMV (Wisler & Zettler, 1987). A previous study detected and characterised VanMV from FP and the CI as strains of DsMV exclusively infecting the orchid vanilla based on partial genome sequences (Farreyrol et al., 2006). Little information is available on the evolutionary history of the potyvirus and how it had acquired the ability to infect orchids and lose the ability to infect aroid plants, natural hosts of DsMV (Farreyrol, 2005). The overall aim of this research is to characterise the complete genome of the VanMV-CI isolate to better understand its evolutionary history and determine a molecular basis for the difference in host range between VanMV and DsMV. Specific aims of this research are:

1. Determine and analyse the VanMV-CI genome.
2. Compare and contrast the VanMV-CI genome to all available DsMV genome sequences published in GenBank to highlight differences that could account for a difference in host range.
3. Determine the evolutionary relationship of VanMV-CI compared to other members of the BCMV group of potyviruses.
4. Carry out a recombination analysis between DsMV, VanMV and all available potyvirus sequences to determine whether or not VanMV arose as a result of a recombination event.
5. Compare and contrast the silencing suppressor activity of VanMV-CI to that of DsMV with and without the upstream P1 gene.

Chapter 2

Characterising the VanMV-CI genome

2.1 Introduction

2.1.1 RNA virus origins and evolution

The most abundant and genetically diverse biological entities in the biosphere are viral pathogens (Koonin, Dolja, & Krupovic, 2015). Viruses have profound impacts on all three domains of life, namely Archaea, Bacteria and Eukarya, in a parasitic nature capable of causing serious harm to the development and survival of its host (Forterre, 2006). The profusion of each virus class (double stranded (ds) and single stranded (ss) DNA or RNA), in its host may depend largely on the cell type. Double stranded DNA viruses dominate prokaryotes (and are known as phages) in which RNA viruses represent the minority. On the other hand, eukaryotes are predominantly affected by RNA viruses with dsDNA and single-stranded (ss) DNA viruses less common and dsRNA viruses the most rare (Koonin et al., 2015).

Three hypotheses have been developed to explain the origins of viruses, i.e. the virus-first hypothesis (also known as co-evolution), the regressive hypothesis (also known as the degeneracy or reduction hypothesis) and the cellular origin hypothesis (also known as the vagrancy or escape hypothesis) (Durzyńska & Goździcka-Józefiak, 2015; Forterre, 2006). The virus-first hypothesis, as its name proposes, suggests that viruses predated more complex organisms and the cells they now depend on to survive and replicate. This hypothesis suggests viruses were born in a pre-cellular era (Durzyńska & Goździcka-Józefiak, 2015). The majority have rejected this hypothesis due to the fact that all viruses known are dependant on their host for their reproduction (Forterre, 2006). The regressive hypothesis suggests that viruses were once smaller cells that parasitised larger cells. As time went by, certain viral genes were not required and were subsequently lost (Durzyńska & Goździcka-Józefiak, 2015; Forterre, 2006). This hypothesis has also been rejected based on two arguments: firstly, the lack of identified intermediate forms between cells and viruses and secondly, parasites (such as *Microsporidia* in eukaryotes) derived from cells in the three domains of life have retained their own characters (i.e. their own ribosomes and machinery for protein synthesis and ATP production) (Forterre, 2006; Forterre & Krupovic, 2012; Holmes, 2011). The escape hypothesis has become the most prominent hypothesis amongst virologists today with regard to virus origins (Forterre, 2006; Forterre &

Krupovic, 2012). The traditional version of the escape hypothesis states that viruses were originally elements of cell genomes which had then escaped, becoming more autonomous and in turn becoming infectious agents capable of devastating impacts (Forterre & Krupovic, 2012). This hypothesis should be considered in conjunction with the widely accepted RNA-world hypothesis (Sankaran, 2016).

The RNA world hypothesis states that RNA was a forerunner of DNA molecules, a transition brought about by the relative instabilities and poor catalytic abilities of RNA (Bernhardt, 2012; de Farias, Rêgo, & José, 2016; Forterre, 2006; Forterre & Krupovic, 2012; Holmes, 2011). The forerunner RNA, according to a recent study, is thought to have been tRNA, which was also the progenitor of both mRNA and rRNA, a hypothesis otherwise known as the 'tRNA core hypothesis' (de Farias et al., 2016). With regard to RNA virus origins, the RNA-world is considered to have two phases, 'before' and 'after' the development of ribosomes (Forterre & Krupovic, 2012). In light of these two phases, RNA viruses would have had to appear towards the end of the second stage (after ribosomal development) of the RNA-world hypothesis by the widely attested vagrancy process (escape hypothesis). This process would have been followed by a rapid radiation of viral pathogens, giving rise to the different classes of viruses we observe today (Forterre, 2006; Forterre & Krupovic, 2012). These early forms of viruses would have had the capacity to utilise cellular mechanisms in order to replicate, which would have been made possible with RNA dependant RNA polymerase (RdRp) encoded information within the genome. This encoded information acquired by the virion is thought to be one of the main hallmark genes of viruses, which should theoretically correspond to and reflect the evolutionary history of viruses, tracing them back to their likely origins (Koonin et al., 2015). The efficacy of this hallmark gene in sketching the likely evolutionary routes taking place appeared promising when a number of conserved motifs were observed amongst different classes of RNA viruses (-ssRNA, +ssRNA and dsRNA). However, when an overall evaluation was made, comparative RdRp genes presented low similarities, producing a tenuous phylogenetic analysis, which in turn provided little evolutionary information (Koonin et al., 2015; Xiong & Eickbush, 1990). However, when structural analyses were carried out on RdRp homologues of different classes of RNA viruses, structural similarities such as the 'right

handed' fold comprising of palm, fingers and thumb domain confirm their relatedness and similar origins (Koonin et al., 2015; Mönttinen, Ravantti, Stuart, & Poranen, 2014).

2.1.2 Picorna-like virus superfamily and *Potyviridae* evolution

The picorna-like virus super-family is a group of RNA viruses consisting of a positive sense single stranded RNA (+ssRNA) genome. They are major pathogens of plants, animals and insects (Culley, Lang, & Suttle, 2003). There are six divergent families that make up the super-family and they include the *Picornaviridae, Caliciviridae, Comoviridae, Sequiviridae, Dicistroviridae* and *Potyviridae* (Culley et al., 2003). Members of the picorna-like virus super-family are typically associated with high mutation rates due to a lack of proof reading ability, short generation times and large population sizes (Frédéric & Juan, 2015). These characteristics have been crucial to the evolution of RNA viruses since they produce a pool of variants, also referred to as 'sequence space', with individual species showing quasi-species behaviour. This has allowed significant viral dispersal to a number of different hosts (Koonin et al., 2008). These families are also characterised by a number of common features such as a single ORF encoding a single polyprotein and the RdRp (Culley et al., 2003). As mentioned by Koonin et al. (2015), RdRp is of particular importance when attempting to uncover the evolutionary pathways that took place in the recent past. A phylogenetic analysis carried out on the RdRp genes of members of the picorna-like virus family by Koonin et al. (2008) indicated six distinct clades, one of which included the *Potyviridae, Hypoviridae* and *Astroviridae* (Clade 3, Figure 2.1). This clade implies a common ancestor for animal astroviruses, plant potyviruses and double stranded RNA (dsRNA) hypoviruses that existed more recently than the common ancestor shared with other members of the superfamily. The emergence of the different viral lineages within the superfamily would have occurred at the same time as the early stages of eukaryogenesis, before the expansion of major eukaryotic supergroups (Koonin et al., 2008).

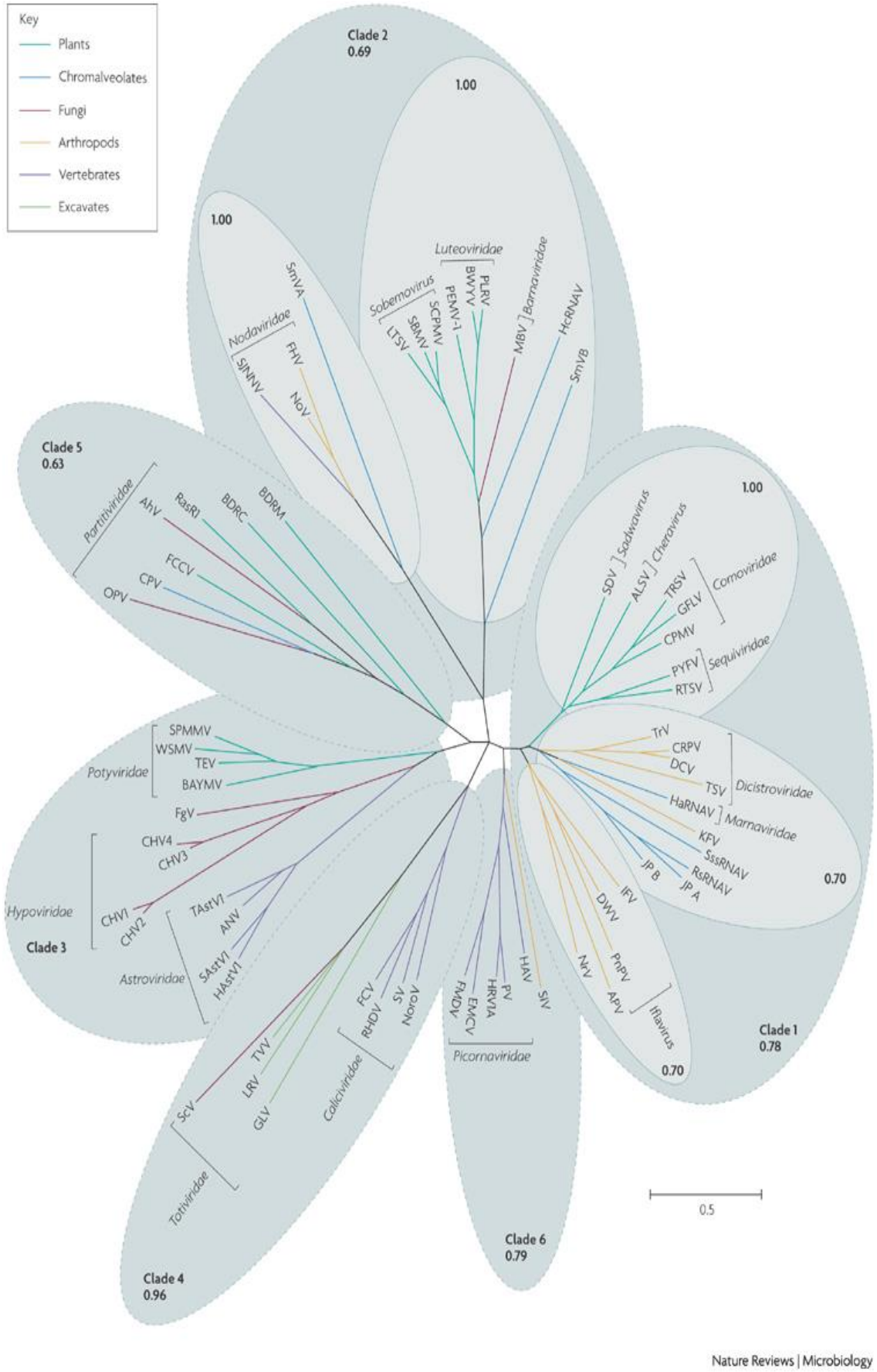


Figure 2.1: The phylogenetic tree of RdRp of the picorna-like virus super family. The *Potyviridae* family is in clade 3 (Koonin et al., 2008).

2.1.3 Potyvirus evolution and the BCMV lineage

Unlike studying RNA virus origins, studying extant virus sequences provides more conclusive findings regarding recent evolutionary histories. Early attempts to evaluate potyvirus evolution and classification were made more than 50 years ago using quantitative serological techniques (Gibbs & Kazusato. Ohshima, 2010). Due to cross reactivity, this became challenging for potyviruses, a short lived problem overcome by later developments in molecular biology by comparing protein and nucleotide sequence alignments (Ward, McKern, Frenkel, & Shukla, 1992). These developments in molecular biology and virus taxonomy allowed the classification of the *Potyviridae* family of plant infecting viruses into seven major genera known as *Brambyvirus, Bymovirus, Ipomovirus, Malcuravirus, Poacevirus, Rymovirus, Tritimovirus* and *Potyvirus* (Valli, García, & López‐Moya, 2015).

A number of different phylogenetic analyses have since been carried out indicating four major groups of potyviruses namely *Potato virus Y*, *Sugarcane mosaic virus*, *Bean yellow mosaic virus* and the *Bean common mosaic virus* group (Barnett, 1992). Potyvirus taxonomy today has encountered a number of new problems that make phylogenetic analyses difficult, such as many of the first deposited potyvirus sequences only providing partial CP sequences that offer limited phylogenetic information. Viral recombination, a phenomenon which seems to be common amongst potyviruses (Gibbs & Kazusato. Ohshima, 2010), also makes conclusive phylogenies difficult to obtain, especially with an incomplete dataset.

However, as more potyvirus sequences become available, the more the evolutionary history, initial radiation and emergence of potyviruses become apparent. Gibbs and Kazusato. Ohshima (2010) suggest potyviruses evolved from a *Rymovirus* by acquiring the ability to be transmitted by aphids, thereby allowing infection of monocotyledon plants in the same area. Assuming a mutation rate of approximately 1.15×10^{-4} nucleotide substitutions per year the emergence of the first potyvirus is estimated to have occurred approximately 7250 years ago during the dawn of agriculture. This appearance is believed to have taken place in the fertile cereal based agricultural region of the Middle East (Gibbs & Kazusato. Ohshima, 2010; Gibbs et al., 2008a; Ohshima, 2013). The genetic structure of potyvirus populations is still poorly understood and it is not certain whether the diverse population structure is primarily

due to rapid mutations, recombination or a combination of both (Gibbs & Kazusato. Ohshima, 2010).

While studying potyvirus sequences available at the time, researchers identified two distinct groups of viruses within the genus *Potyvirus*; those related to BCMV and *Potato virus Y* (Hiebert & Purcifull, 1992; McKern, Ward, & Shukla, 1992). The phenotypic differences observed between the different groups were typical of related viruses following different evolutionary trajectories after descending from a common ancestor. As stated earlier, these distinct potyvirus groups apparently appeared approximately 7250 years ago probably in the same region as their progenitors (Gibbs & Kazusato. Ohshima, 2010). Approximately 3800 years later, progeny from each group diverged independently of each other, one in Southeast Asia (BCMV) and another in North America (PVY), producing viral populations that adapted to infect plants available in each region (Gibbs & Kazusato. Ohshima, 2010). In the past 500 years, migration events between these two locations have allowed virus groups from one region to be exposed to virus groups from another (Gibbs & Kazusato. Ohshima, 2010).

Up until 2008, there were viruses in 30 known potyvirus species classified within the *Bean common mosaic virus* group, two-thirds of which infect dicotyledons and one-third infecting monocots (Gibbs., Trueman, & Gibbs, 2008b). A phylogeographical analysis of these viruses suggests the BCMV lineage originated in South and East Asia, with some migrating towards Australia and Islands of Oceania with a few migrating westward (Gibbs & Kazusato. Ohshima, 2010). Based on BCMV CP and genome sequences, this lineage appears to have emerged between 2940-3580 years ago from plants in the old world and have only been able to infect plants of the new world in the past 500 years (Gibbs & Oshima, 2010; Gibbs et al., 2008a; Gibbs et al., 2008b). It has been suggested that the BCMV lineage could have spread from South East Asia to Australia, coincident with the migration of the Austronesian sea-faring/farming culture from China/Taiwan throughout the southern and eastern Pacific islands (Gibbs. et al., 2008b).

While these studies allow the development of hypotheses regarding emergence and movement of potyviruses, they rely on the assumption that viruses are subject to a

constant rate of evolution, or on certain viruses having the same mutation rates as others. (Duffy, Shackelton, & Holmes, 2008). Further, whether or not a strict or relaxed clock is used determines the mathematical framework, which depicts the evolutionary time scale. When considering these factors, investigating evolutionary emergence can become challenging and results may not mirror the actual events that took place in the recent past (Duffy et al., 2008). The evolutionary rates and dates presented by Gibbs et al. (2008a) however, based on a number of potyvirus sequences from Australia and America, produced similar results using both a relaxed and strict molecular clock. This result suggests a more mathematically reliable method. While tracing the exact origins of the *Potyviridae*, *Potyvirus* and BCMV lineages has proven difficult, the emergence dates and phylogenetic relationships presented are the most reliable estimates currently available. However, a number of genome sequences have been made available since 2008 so these estimates should be revisited.

2.1.4 DsMV origins and evolution

DsMV was initially discovered in 1970 infecting taro (*C. esculenta* (L) Schott) but now has a worldwide distribution with an extensive host range (Abo El-Nil, 1977; Nelson, 2008). A brief review of DsMV is discussed in the general introduction (Section 1.3.2), which includes the date, host and the country it was discovered. This review not only highlights the worldwide distribution and threat it possesses, but also a rough approximate time of appearance in each geographical location. This review should not be interpreted to reflect the emergence of the virus in each geographical location and host but rather to better understand how the virus spreads and possibly evolves. DsMV reports in each case may or may not reflect the exact time of emergence and the period of time it has been present in its host. It is possible that DsMV symptoms have been overlooked and therefore gone unnoticed for some time prior to reporting in certain regions, especially when symptoms have not been obvious visibly in some plants and only apparent during certain seasons of the year (Grisoni et al., 2004; Odoux & Grisoni, 2010).

When investigating the likely origins of DsMV, a factor that should be taken into consideration would be the likely origins of its hosts. It appears that natural and common hosts of the virus are *C. esculenta* (taro) and *Xanthosoma spp* (tannia) as well as other aroid species (Hu et al., 1995; Pappu et al., 1993; Pearson et al., 1998; Pearson

& Grisoni, 2002). The Araceae are monocotyledons with worldwide distribution; certain studies suggest the edible aroids taro and tannia spread with human migration (Chang, 2012).

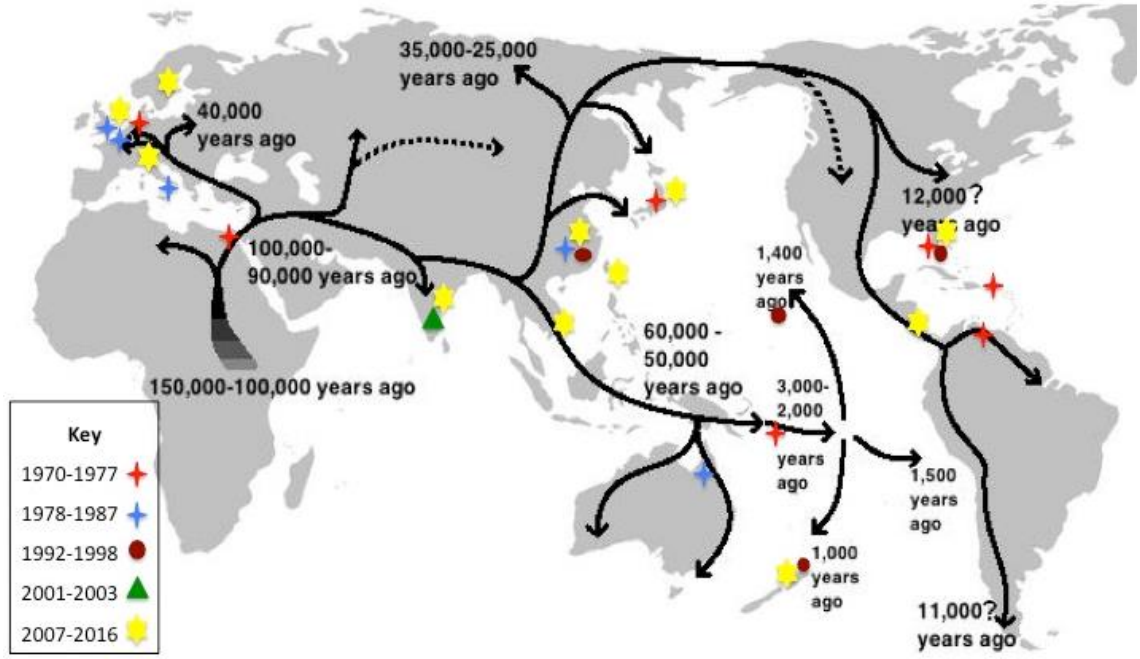


Figure 2.2 The Out of Africa model depicting the migration of *Homo sapiens* (black lines) from Africa throughout different geographical locations around the world. Coloured shapes refer to locations and dates DsMV was reported. Modified from the image retrieved from http://archaeosoup.com/lao-skull-earliest-example-of-modern-human-fossil-in-southeast-asia/.

Given the most likely origins of *Colocasia spp* and *Xanthosoma spp* from South Eastern or Southern Central Asia (Govaerts, 2012), it is plausible to assume that DsMV originated in the same location and travelled to and throughout the South Pacific with human migration as depicted in Figure 2.2. Alternatively, DsMV could have emerged along the migration route from its origins to the South Pacific. The reported DsMV incidences in the literature (Section 1.3.2) show no correlation to human migration suggesting that DsMV had already well-established populations when they were discovered in 1970. These well-established populations appear at various geographical locations worldwide, which would have only been discovered as a result of recent advances in molecular biology.

2.1.5 VanMV classification

Prior to this study, five VanMV partial sequences were available in GenBank representing the 3' region of the genome only (Figure 1.4), with no sequence

information for the 5' region. The available information includes three sequences for the NIb-CP-3'UTR region from CI and FP isolates, one for the CI-6K_2-NIa region from a CI isolate and a partial CP for the FP strain (Farreyrol, 2005; Grisoni et al., 2004). The current species demarcation standards for inferring the same species include CP aa sequence identity greater than 80% and nt sequence identity greater than 75% in either the CP or the whole genome (King et al., 2012). Differences in polyprotein cleavage sites and host range are also properties that are considered. According to previous studies, VanMV satisfies some of these criteria to be considered a strain of DsMV, which is 80% and 76% or greater in aa and nt sequence identity over the CP region, respectively (Adams, Antoniw, & Fauquet, 2005; Farreyrol et al., 2006; King et al., 2012). Other criteria not addressed in previous studies however, include the nt and aa identity over the entire genome, similarities in inclusion body morphologies and polyprotein cleavage sites (Adams, Antoniw, & Fauquet, 2005; King et al., 2012). Based on current standards, it is not clear whether all criteria, at least one or a majority, needs to be satisfied before a virus isolate can be considered a strain of a virus species. Based on these previous studies from the CP alone, VanMV-CI-AT and VanMV-CI-NAT were classified as strains of DsMV however; VanMV-FP, although satisfying some of the criteria (antigenic and serological properties) was suggested to belong to a distinct potyvirus species (Farreyrol et al., 2006). However, the analysis by these authors was based solely on the CP, which may not be a true representative of the overall genome, an important criterion as indicated by Gibbs & Oshima (2010), that could provide improved discriminatory outcomes. From these previous studies of VanMV, the strain and virus demarcation criteria involving the complete genome and polyprotein cleavage sites for both CI and FP isolates have gone unaddressed and there remains a gap in knowledge.

As discussed in Section 1.1, the dramatic decrease in vanilla production between 1967-1981 suggested that VanMV emerged in the South Pacific around 1967 but was only reported in 1987 (Wisler & Zettler, 1987). This appearance is also around the same time DsMV was first reported in 1970 by Zettler et al., (1970) from Florida USA. Due to limited sequence information, especially linked to accurate chronological information, VanMV origins and the events that led to its emergence remain unclear.

2.2 Aims of study

Two independent studies have attempted to evaluate the evolutionary relationship of VanMV to DsMV, both of which have produced conflicting results. Farreyrol (2005) suggested the emergence of the VanMV lineage preceded that of the DsMV lineage, while research by Chang (2012) affirms the contrary. Both studies were carried out based on the CP gene, which may provide limited evolutionary information. Recombination analysis between VanMV and DsMV indicated no significant recombination events taking place in the CP gene (Farreyrol, 2005) however, no other recombination analysis has been carried out on other VanMV genes and between VanMV and other BCMV or other potyviruses. The recombination analysis carried out by Farreyrol (2005) focused mainly on the relationship between VanMV and DsMV rather than viruses of the wider *Potyvirus* genus. The sequence of the VanMV genome had been determined and assembled by Dr Colleen Higgins and students at AUT but had not been analysed. Thus, one aim of this research was to characterise the VanMV-CI assembled genome and compare this genome to available genomes of DsMV and viruses of the BCMV group. The VanMV-CI genome was annotated and potyviral conserved motifs were identified. Further, sequence differences between the VanMV-CI and DsMV genomes were identified in an attempt to pinpoint region(s) of the genome that may be responsible for the host range difference between these viruses. Recombination and phylogenetic analyses was also performed to determine whether or not the emergence of VanMV was the result of a recombination event between DsMV and another virus and if DsMV or VanMV emerged first.

2.3 Methods and Materials

2.3.1 Plant material

A *V. tahitensis* plant showing symptoms of VanMV infection was collected from the Cook Islands in 1988 (Pearson, Jackson, Pone, & Howitt, 1993) and was vegetatively propagated in an insect -proof glasshouse at the University of Auckland, New Zealand since that date. This plant had previously reacted with antiserum to a FP isolate of VanMV (Wang & Pearson, 1992) and VanMV sequences derived from it (Farreyrol et al., 2006).

2.3.2 Total RNA extraction

Total RNA was extracted using the Spectrum Plant Total RNA kit (Sigma Aldrich) according to manufacturer's instructions. Approximately 100 mg of infected leaf tissue was ground to powder form in liquid nitrogen in a sterile mortar and pestle. Lysis solution (500 µl) containing 5 µl of 2-mercaptoethanol (2-ME) was then added to the powder and was vigorously vortexed for 30 seconds before incubating at 56°C for 3-5 minutes. The mixture was then centrifuged (Eppendorf, Hamburg, Germany) at maximum speed (14,000 rpm) for three minutes to pellet the cellular debris. The lysate supernatant was then transferred to a filtration column and centrifuged at maximum speed (14,000 rpm) for one minute to remove residual debris. Binding solution (500 µl) was then added to the clarified lysate and was mixed by pipetting up and down five times. The solution was then transferred to a binding column and was centrifuged for one minute to bind RNA, with the flow-through discarded. Wash solution 1 (300 µl) was then added to the binding column and was centrifuged at maximum speed (14,000 rpm) for one minute discarding the flow-through. On-Column DNase Digestion mixture (10 µl of DNase 1 with 70 µl of DNase digestion buffer, Sigma Aldrich) was added to the binding column and was incubated at room temperature for 15 minutes. Wash solution 1 (500 µl) was then added to the column, which was centrifuged at maximum speed for one minute, discarding the flow-through. A second washing step was carried out by transferring 500 µl of prepared wash solution 2 (containing ethanol) to the column and centrifuged at maximum speed (14,000 rpm) for 30 seconds, discarding the flow-through. This wash step was repeated a third time. The column was allowed to dry by centrifuging the binding

column at maximum speed (14,000rpm) for one minute, discarding any flow-through liquid in the collection tube. The binding column was then transferred into a fresh collection tube and RNA was eluted with 50 µl of elution solution after the column was allowed to sit at room temperature for one minute before centrifuging at maximum speed (14,000 rpm) for one minute. The concentration and purity of the eluted RNA was then measured using a Nanovue ND1000 spectrophotometer (GE Healthcare).

2.3.3 Sequence assembly

The VanMV genome had been assembled prior to the initiation of this project using the following method. Transcriptome sequencing was carried out by Macrogen Inc (South Korea) using HiSeq2000. Ten one-million read files were chosen randomly and assembled separately using the *de novo* assembly tool in GeneiousPro 6.1.8 (http://www.geneious.com) (Kearse et al., 2012; Wylie & Jones, 2011). BLASTn searching against GenBank identified contigs of potyviral sequences; each end of the genome was taken as the most 5' and 3' nucleotides common to all assemblies. Contigs that showed high similarity to available DsMV and VanMV sequences were assembled into one complete genome. The assembled genome was then used as a reference genome to assemble all previously unassembled Illumina reads to confirm the sequence.

2.3.4 VanMV sequence annotation and analysis

A BLASTn search was carried out with this consensus sequence as a query in GenBank (www.ncbi.nlm.nih.gov) using the default parameters to search for any similar sequences. The search revealed three complete genome sequences of DsMV and five partial VanMV sequences showing highest sequence similarity. These sequences and their predicted proteins were used to assist annotation of the VanMV genome. All multiple nucleotide and amino acid sequence alignments were carried out using MUSCLE, and pairwise identities were calculated using the distance option in GeneiousPro 6.1.8.

Sequence motifs were identified by carrying out a sequence search in GeneiousPro 6.1.8 based on readily identified conserved motifs from the literature (Adams, Antoniw, & Beaudoin, 2005; Rohožková & Navrátil, 2011; Xu, Liu, Li, & Li, 2011; Xu, Wang, Kinard, & Li, 2012; Zheng et al., 2008).

2.3.5 Phylogenetic analysis of the VanMV genome and genes

For each potyvirus where a genome sequence was available, at least one sequence was downloaded from GenBank. One genome sequence for each member of the BCMV group of potyviruses was also downloaded from GenBank with the exception of DsMV for which all three genomes were downloaded (Table 2.1). A nucleotide sequence alignment of the VanMV genome was carried out against each potyvirus sequence. Nucleotide sequence alignments of each individual VanMV-CI gene were also carried out against equivalent BCMV group sequences. For the CP gene and 3'UTR, all available CP sequences of DsMV were included in the analysis in addition to the CP and 3'UTR regions of the complete DsMV genomes (Table 2.3). Complete genes from published VanMV sequences were also included in the analysis where appropriate (Table 2.2). All nucleotide sequence alignments were carried out using MUSCLE and pairwise identities calculated using the distance option in GeneiousPro 6.1.8. Phylogenetic analyses for the potyvirus genomes were carried out using the neighbour-joining method in MEGA6 with the default settings (Koichiro, Glen, Daniel, Alan, & Sudhir, 2013). All other phylogenetic analyses were carried out using maximum likelihood in MEGA6. The General Time Reverse (GTR) and Gamma Distributed with Invariable sites (G+I) model was applied to P1, HC-Pro, P3, CI, NIa-VPg, NIa-Pro, Nib and CP while the Tamura 1992 (T92) model and Gamma Distributed with invariable sites (G+I) were used for 6K1 and 6K2. The model used for the 5'UTR and 3'UTR regions was the Hasegawa-Kishino-Yano (HKY) with Gamma distribution model. Statistical support for all phylogenetic analyses was determined by bootstrapping 1000 times. PVY (X12456) was used as an out-group for the phylogenetic analysis of the VanMV genome and individual genes against the BCMV group with the exception of the VanMV CP against DsMV CP for which blue squill virus A (BSVA NC019415), BCMV (AJ312438) and passion fruit woodiness virus (PWV NC014790) was used. Oat necrotic mottle virus (ONMV NC005136) was used as an out-group for the phylogenetic analysis of VanMV against potyvirus sequences. Genome sequences for the BCMV group used in the phylogenetic analyses are outlined in Table 2.1.

Table 2.1: BCMV group genome sequences obtained from GenBank that were used in phylogenetic analyses.

Virus genome	Accession number	Reference
Bean common mosaic necrosis virus (BCMNV)	U19287	Fang, Allison, Zambolim, Maxwell & Gilbertson (1995)
Bean common mosaic virus (BCMV)	AJ312437	Zheng, Chen, Chen, Adams & Hou (2002)
Blackeye cowpea mosaic virus (BICMV)	AY575773	Wang & Fang (2004)
Cowpea aphid borne mosaic virus (CABMV)	AF348210	Mlotshwa, Shithole-Niang, Van Kampen, Van Kammen & Wellink (2001)
Dasheen mosaic virus (DsMV)	KJ786965	Kamala, Makeshkumar, Sreekumar & Chakrabarti (2014)
Dasheen mosaic virus (DsMV)	NC003537	Chen, Chen, Chen & Adams (2001)
Dasheen mosaic virus (DsMV)	KT026108	Liebrecht & Winter (2015) (Unpublished)
East Asian passiflora virus (EAPV)	AB246773	Iwai, Terahara, Yamashita, Ueda & Nakamura (2006)
Fritillary virus Y (FVY)	AM039800	Chen, Zheng, Chi, Adams, Wei, Lin & Chen (2006)
Hardenbergia mosaic virus (HarMV)	HQ161081	Wylie & Jones (2011a)
Passion fruit woodiness virus (PWV)	HQ122652	Wylie & Jones (2011b)
Peanut stripe virus (PStV)	U34972	Flasinski, Gunasinghe, Gonzales & Cassidy (1996)
Potato virus Y (PVY)	X12456	Robaglia, Durand-Tardif, Trochet, Boudazin, Astier-Manifacier & Casse-Delbart (1989)
Soybean mosaic virus (SMV)	D00507	Eggenberger, Stark & Beachy (1989)
Telosma mosaic virus (TeMV)	DQ851493	Ha, Coombs, Revill, Harding, Vu & Dale (2008)
Vanilla mosaic virus (VanMV-CI)	KX505964	This study
Watermelon mosaic virus (WMV)	AY437609	Desbiez & Lecoq (2004)
Wisteria vein mosaic virus (WVMV)	AY656816	Liang, Song, Tian, Li & Fan (2006)
Zantedeschia mild mosaic virus (ZaMMV)	AY626825	Huang, Hu, Yang & Chang (2007)
Zucchini yellow mosaic virus (ZYMV)	AF127929	Lin, Hou & Yeh (2001)

Table 2.2 Available VanMV sequences obtained from GenBank that were used in phylogenetic analyses.

Isolate	Accession number	Reference
Cook Islands	AJ616720	Farreyrol et al. (2006)
Cook Islands	AJ616722	Farreyrol et al. (2006)
Cook Islands	AJ616721	Farreyrol et al. (2006)
French Polynesia	AJ616719	Farreyrol et al. (2006)
French Polynesia	AJ429525	Grisoni et al. (2004)
Cook Islands	KX505964	Puli'uvea et al. (2016)

Table 2.3 DsMV CP sequences obtained from GenBank that were used in phylogenetic analyses.

DsMV accession number	Reference
LC114500	Sasaya et al. (2016) (Unpublished)
LC114507	Sasaya et al. (2016) (Unpublished)
LC114515	Sasaya et al. (2016) (Unpublished)
LC114502	Sasaya et al. (2016) (Unpublished)
LC114499	Sasaya et al. (2016) (Unpublished)
LC114494	Sasaya et al. (2016) (Unpublished)
LC114503	Sasaya et al. (2016) (Unpublished)
LC114508	Sasaya et al. (2016) (Unpublished)
LC114509	Sasaya et al. (2016) (Unpublished)
KT372699	Chung & Chang (2015) (Unpublished)
AF511485	Chen & Chang (2002)
LC114501	Sasaya et al. (2016) (Unpublished)
LC114511	Sasaya et al. (2016) (Unpublished)
LC114496	Sasaya et al. (2016) (Unpublished)
AJ298034	Chen et al. (2001)
NC003537	Chen et al. (2001)
AJ298036	Chen et al. (2001)
AF169832	Chen & Chang (1999) (Unpublished)
LC114512	Sasaya et al. (2016) (Unpublished)
U08124	Pappu et al (1994)
LC114514	Sasaya et al. (2016) (Unpublished)
DQ925465	Ha et al. (2008)
AM910398	Reyes et al. (2009)
AM910400	Reyes et al. (2009)
U00122	Pappu et al (1994)
AM910399	Pappu et al (2009)
AM910401	Pappu et al (2009)
AM910402	Pappu et al (2009)
AJ298035	Chen et al. (2001)
EF199550	Luo et al. (2006) (Unpublished)
JN692173	Shiming & Ni (2011)
AM910404	Reyes et al. (2009)
AM910407	Reyes et al. (2009)
AM910406	Reyes et al. (2009)
AM910403	Reyes et al. (2009)
AM910405	Reyes et al. (2009)
KT026108	Liebrecht & Winter (2015) (Unpublished)
HQ207540	Binoy et al. (2010) (Unpublished)
HQ207531	Binoy et al. (2010) (Unpublished)
HQ207530	Binoy et al. (2010) (Unpublished)
KJ786965	Kamala et al. (2014) (Unpublished)
FJ160764	Binoy et al. (2008) (Unpublished)
HQ207529	Binoy et al. (2010) (Unpublished)
DQ925464	Ha et al. (2008)
DQ925466	Ha et al. (2008)
AY994104	Wei & Pearson (2005) (Unpublished)
AY994105	Wei & Pearson (2005) (Unpublished)
HQ207532	Binoy et al. (2010) (Unpublished)
HQ207535	Binoy et al. (2010) (Unpublished)
HQ207542	Binoy et al. (2010) (Unpublished)
HQ207538	Binoy et al. (2010) (Unpublished)
HQ207539	Binoy et al. (2010) (Unpublished)

DsMV accession number	Reference
HQ207534	Binoy et al. (2010) (Unpublished)
HQ207537	Binoy et al. (2010) (Unpublished)
HQ207536	Binoy et al. (2010) (Unpublished)
LC114504	Sasaya et al. (2016) (Unpublished)
LC114505	Sasaya et al. (2016) (Unpublished)
LC114506	Sasaya et al. (2016) (Unpublished)
HQ207533	Binoy et al. (2010) (Unpublished)
JN692172	Shiming & Ni (2011) (Unpublished)
LC114510	Sasaya et al. (2016) (Unpublished)
LC114495	Sasaya et al. (2016) (Unpublished)
LC114513	Sasaya et al. (2016) (Unpublished)
LC114497	Sasaya et al. (2016) (Unpublished)
LC114498	Sasaya et al. (2016) (Unpublished)
LC114493	Sasaya et al. (2016) (Unpublished)
EU420058	Zhao (2008) (Unpublished)
AJ628756	Shi et al. (2004)

2.3.6 Recombination analysis of VanMV

Recombination analyses were carried out to investigate whether or not VanMV originated as a result of a recombination event between DsMV and another virus. This recombination analysis was carried out on VanMV and all available DsMV genomes against all potyvirus genomes using RDP4 (Martin et al., 2015). Nucleotide sequences were aligned using MUSCLE prior to recombination analysis using GeneiousPro 6.1.8. Neighbour-joining trees of potyvirus genomes and potyvirus P1 genes were calculated using MEGA6 (www.megasoftware.net/download_form) in order to detect any possible shift in branching arrangement before and after predicted recombination points. Oat necrotic mottle virus (ONMV) was used as the out-group for the potyvirus recombination analysis. The analysis was carried out using all default settings with the P-value set at 0.05. The exact WMV, BCMV and SMV sequences used by Desbiez and Lecoq (2004b) in a prior recombination analysis were used as positive controls to determine the reliability of the data.

2.4 Results

2.4.1 VanMV genome structure

The complete genome sequence of VanMV-CI consists of 9,828 nucleotides (nt) excluding the poly (A) tail. It has a typical potyviral genome structure with a main ORF flanked by 5' and 3' untranslated regions (UTR). A start codon is at nucleotide positions 153-155 with a stop codon at nucleotides 9,570-9,572, indicating a putative polyprotein of 3,139 amino acids. The 5'UTR is 152 nt while the 3'UTR is 259 nt. The polyprotein is capable of being cleaved into ten proteins, as is typical of all potyviruses. A nucleotide sequence alignment reveals VanMV has high sequence similarity to DsMV (Figure 2.3). Further, based on an alignment with DsMV KT026108, a putative PIPO protein of 90 aa can be predicted at nucleotides 3063-3331. The PIPO protein has not been experimentally verified so this size may not be accurate.

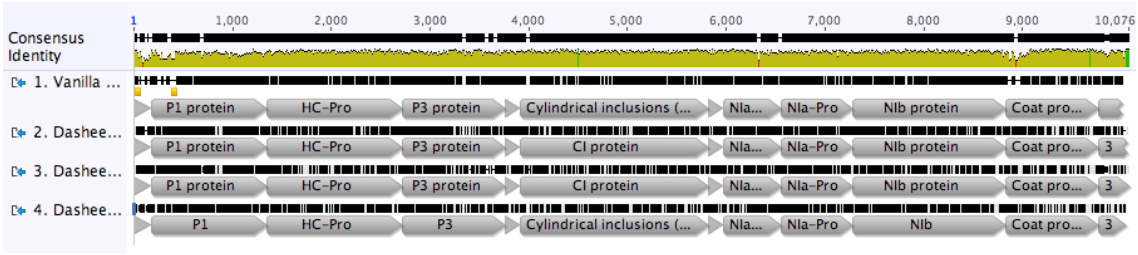


Figure 2.3: Nucleotide sequence alignment of VanMV and DsMV genomes

When comparing sequence lengths (Table 2.4), a majority of VanMV-CI genes show similar sequence lengths to that of DsMV. The regions shown in blue indicate where VanMV-CI differs in length from all DsMV sequences (P1 and CP genes) while red indicates where VanMV-CI differs from at least one DsMV sequence (NIa-VPg). For the P1 gene, VanMV-CI is shorter than all DsMV sequences by at least 90 nt/30 aa while the VanMV-CI CP is shorter than all DsMV sequences by at least 63 nt/21 aa.

Table 2.4: Length of each region of VanMV-CI and publically available DsMV sequences in amino acid (aa) and nucleotides (nt). Numbers indicated in blue is where VaNMV-CI differs from all DsMV sequences while red indicates a difference with at least one DsMV sequence.

Region	VanMV-CI KX505964		DsMV KJ86965		DsMV NC003537		DsMV KT026108	
	nt	aa	nt	aa	nt	aa	nt	aa
5'UTR	152	-	166	-	167	-	167	-
P1	1065	355	1161	387	1155	385	1155	385
HC-Pro	1374	458	1374	458	1374	458	1374	458
P3	1038	346	1038	346	1038	346	1038	346
6K1	156	52	156	52	156	52	156	52
CI	1902	634	1902	634	1902	634	1902	634
6K2	159	53	159	53	159	53	159	53
NIa-VPg	570	190	579	193	570	190	570	190
NIa-Pro	729	243	729	243	729	243	729	243
NIb	1548	516	1548	516	1548	516	1548	516
CP	876	292	939	313	942	314	939	313
3'UTR	239	-	233		233		267	

Sequence comparisons between VanMV-CI and DsMV show the presence of several indels. These sequence differences are within the 5'UTR (Figure 2.4a), the P1 coding sequence (Figure 2.4a and 2.4b) and the variable N-terminal region of the CP gene (Figure 2.5), causing aa sequence changes. The aa repeat region previously identified in the N-terminal region of the DsMV CP is extensive in the French Polynesian isolate of VanMV but absent in the CI isolates (bold in Figure 2.5). The repeat sequence in the French Polynesian isolate of VanMV is similar to that in New Zealand isolates of DsMV. Further, the repeat sequences within the DsMV sequences are varied.

Figure 2.4: Alignment of (a) the 5'UTR and P1 region and (b) the aa sequence of the P1 gene of DsMV and VanMV showing indels. In (a) the start codon is underlined.

(a)

(b)

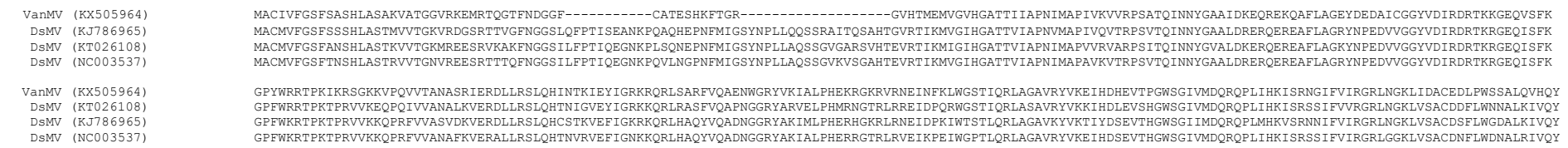
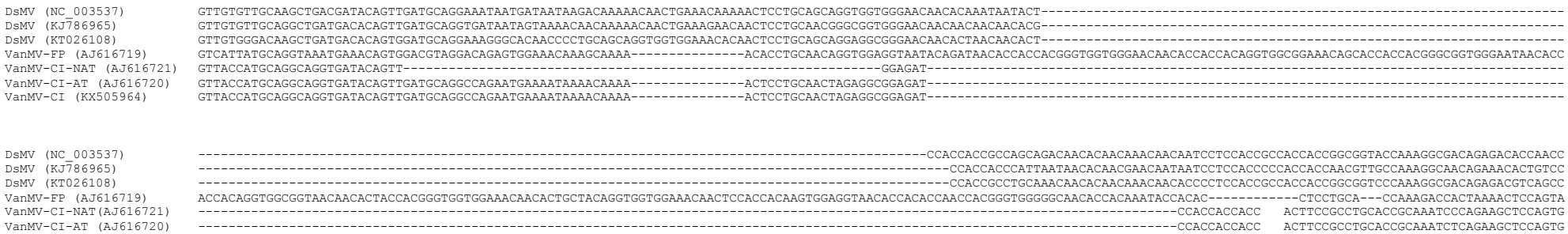
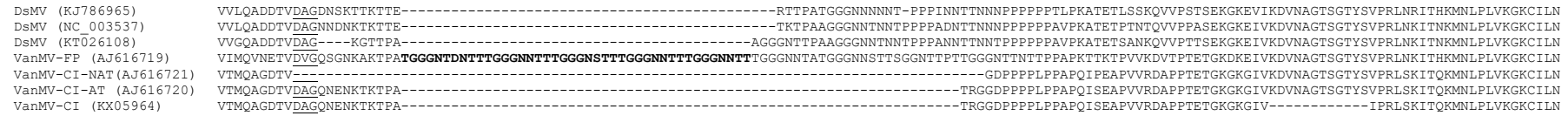


Figure 2.5: Alignment of (a) the 5' end of the CP coding region and (b) the N terminal region of the CP peptide of DsMV and VanMV showing indels. In (b) the aphid transmission motif is underlined. The VanMV-CI sequence recently obtained has the accession KX505964.

(a)



(b)



The predicted cleavage sites of VanMV polyprotein were also compared to those of DsMV (Table 2.5). Although there was very little difference between DsMV isolates, there were some noticeable differences when compared to VanMV-CI. Differences of two amino acids between VanMV and at least one DsMV sequence were observed in the 6K2/NIa-VPg, NIa-VPg/NIa-Pro and NIb/CP junctions. The greatest difference was observed in the P1/HC-Pro junction with three amino acids differing from all DsMV sequences.

Table 2.5: Cleavage sites (predicted) of VanMV-CI compared to those of DsMV. An amino acid indicated in blue is where VanMV-CI differs from all DsMV sequences while red indicates a difference with at least one DsMV sequence.

Protein region	VanMV-CI	DsMV KJ86965	DsMV NC003537	DsMV KT026108
P1/HC-Pro	**VH**QY/**T**	IVQY/A	IVQY/A	IVQYA
HC-Pro/P3	YRVG/G	YRVG/G	YRVG/G	YRVG/G
P3/6K1	VEVQ/A	VEVQ/A	VEVQ/A	VEVQ/A
6K1/CI	V**R**IQ/S	VKIQ/S	VRIQ/S	VRIQ/S
CI/6K2	VQFQ/**S**	VQFQ/N	VQFQ/N	VQFQ/N
6K2/NIa-VPg	V**VT**Q/G	VTAQ/G	VTAQ/G	VTAQ/G
NIa-VPg/NIa-Pro	**VA**ME/G	WQME/G	VAME/G	VAME/G
NIa-Pro/NIb	VQLQ/G	VQLQ/G	VQLQ/G	VQLQ/G
NIb/CP	V**TM**Q/A	VVLQ/A	VVLQ/A	VVGQ/A

Motifs conserved amongst potyviruses were also observed in the VanMV-CI and DsMV genomes (Adams, Antoniw, & Beaudoin, 2005; Rohožková & Navrátil, 2011; Xu et al., 2011; Xu et al., 2012; Zheng et al., 2008). This analysis (Table 2.6) showed a small number of potyvirus conserved motifs absent in both VanMV and DsMV namely FIVRG, $HX_8DX_{32}S$ in the P1 protein and RITC, KITC in the HC-Pro protein. All other known motifs were found in both VanMV-CI and DsMV. The only difference between VanMV-CI and DsMV is the commonly reported IXFG motif at the N-terminus of potyvirus P1

proteins for which VanMV-CI has the IVFG sequence while all DsMV isolates have the MVFG sequence.

The VanMV-CI genome sequence was obtained from the same infected plant material as used by Farreyrol et al. (2006). They reported two deduced amino acid sequences derived from the NIb-CP region, one with the aphid transmission DAG motif and one without. Our more recent sequencing of VanMV-CI from this plant material only identified a sequence with the DAG motif, suggesting this was the dominant sequence present.

Table 2.6: Potyvirus conserved motifs found in the VanMV-CI and DsMV genomes with relevant amino acid positions indicated by numbers before and after the motif. The function of the motif is indicated in brackets in the 2nd column if known. Conserved motifs were analysed using GeneiousPro 6.1.8 (Kearse et al., 2012).

ORF region	Potyvirus Conserved motif and function (*If known*)	Motif in VanMV-CI with aa position	Motif in DsMV and aa positions		
			DsMV KJ786965	**DsMV NC003537**	**DsMV KT026108**
P1	IXFG*	^{4}IVFG^{7}	^{4}MVFG^{7}	^{4}MVFG^{7}	^{4}MVFG^{7}
	FLXG	^{211}FLAG^{214}	^{248}FLAG^{251}	^{246}FLAG^{249}	^{191}FLAG^{194}
	GXSG* (*Serine peptidase*)	^{355}GWSG^{358}	^{392}GWSG^{395}	^{390}GWSG^{393}	^{335}GWSG^{338}
	FIVRG*	-	-	-	-
	$HX_8DX_{32}S$*	-	-	-	-
HC-Pro	$CX_8CX_{18}CX_2C$	$^{433}CX_8CX_{18}CX_2C^{464}$	$^{470}CX_8CX_{18}CX_2C^{501}$	$^{468}CX_8CX_{18}CX_2C^{499}$	$^{413}CX_8CX_{18}CX_2C^{444}$
	LXKA (*Zinc finger*)	^{557}LDKA^{560}	^{594}LDKA^{597}	^{592}LDKA^{595}	^{537}LDKA^{540}
	FRNK	^{586}FRNK^{589}	^{623}FRNK^{626}	^{621}FRNK^{624}	^{566}FRNK^{569}
	CDNQLD	^{602}CDNQLD^{607}	^{639}CDNQLD^{644}	^{637}CDNQLD^{642}	^{582}CDNQLD^{587}
	PTK	^{715}PTK^{717}	^{752}PTK^{754}	^{750}PTK^{752}	^{695}PTK^{697}
	$CX_{72}H$	$^{749}CX_{72}H^{822}$	$^{786}CX_{72}H^{859}$	$^{784}CX_{72}H^{857}$	$^{729}CX_{72}H^{802}$
	RITC*	-	-	-	-
	KITC*	-	-	-	-
P3 (PIPO)	$GSGKSX_3P$	^{1349}GSGKSTGLP^{1357}	^{1386}GSGKSTGLP^{1394}	^{1384}GSGKSTGLP^{1392}	^{1329}GSGKSTGLP^{1337}
CI	DEXH	^{1435}DECH^{1438}	^{1472}DECH^{1475}	^{1470}DECH^{1473}	^{1415}DECH^{1418}
	$G_{1\text{-}2}A_{6\text{-}7}$	$^{3045}G_1A_6{}^{3051}$	$^{3155}G_1A_6{}^{3161}$	$^{3150}G_1A_6{}^{3156}$	$^{3150}G_1A_6{}^{3156}$

Table 2.6: Potyvirus conserved motifs found in VanMV-CI and DsMV genomes with relevant amino acid positions indicated in the numbers before and after the motif. The function of the motif is indicated in brackets if known in the 2nd column. Conserved motifs were analysed using GeneiousPro 6.1.8 (Kearse et al., 2012).

ORF region	Potyvirus Conserved motif and function (*If known*)	Motif in VanMV-CI with nucleotide position	Motif in DsMV and nucleotide positions		
			DsMV KJ786965	**DsMV NC003537**	**DsMV KT026108**
NIa-Pro	$HX_{34}DX_{67}GXCG$	$^{2184}HX_{34}DX_{67}GXCG^{2290}$	$^{2224}HX_{34}DX_{67}GXCG^{2330}$	$^{2219}HX_{34}DX_{67}GXCG^{2325}$	$^{2164}HX_{34}DX_{67}GXCG^{2270}$
NIb	CVDDFN	^{2584}CVDDFN^{2589}	^{2624}CVDDFN^{2629}	^{2619}CVDDFN^{2624}	^{2564}CVDDFN^{2569}
	FTAAPLD	^{2570}FTAAPLD^{2576}	^{2610}FTAAPLD^{2616}	^{2605}FTAAPLD^{2611}	^{2550}FTAAPLD^{2556}
	DGSQFDS	^{2669}DGSQFDS^{2675}	^{2692}DGSQFDS^{2635}	^{2664}DGSQFDS^{2670}	^{2609}DGSQFDS^{2615}
	$SGX_3TX_3NTX_{30}GDD$	$^{2691}SGX_3TX_3NTX_{30}GDD^{2734}$	$^{2732}SGX_3TX_3NTX_{30}GDD^{2774}$	$^{2726}SGX_3TX_3NTX_{30}GDD^{2769}$	$^{2671}SGX_3TX_3NTX_{30}GDD^{2714}$
	[A/S]M[I/V]E[S/A]WG	^{2818}AMIEAWG ^{2824}	^{2858}AMIEAWG^{2864}	^{2853}AMIEAWG ^{2859}	^{2798}AMIEAWG ^{2804}
CP	DAG* (*Aphid transmission*)	^{2903}DAG^{2905}	^{2943}DAG^{2945}	^{2938}DAG^{2940}	^{2883}DAG^{2885}
	MVWCI[E/D]NG	^{3037}MVWCIDNG^{3044}	^{3098}MVWCIDNG^{3105}	^{3094}MVWCIDNG^{3101}	^{3038}MVWCIDNG^{3045}
	W[V/T]MMDG[D/E/N]	^{3053}WVMMDGN^{3059}	^{3114}WVMMDGN^{3120}	^{3110}WVMMDGN^{3116}	^{3054}WVMMDGN^{3060}
	[P/R/A]YMPRYG	^{3100}PYMPRYG^{3106}	^{3161}PYMPRYG^{3167}	^{3157}PYMPRYG^{3163}	^{3101}PYMPRYG^{3107}
	YAFDFYE	^{3119}YAFDFYE^{3125}	^{3180}YAFDFYE^{3186}	^{3176}YAFDFYE^{3182}	^{3120}YAFDFYE^{3126}
	E[N/D]TERH	^{3165} ENTERH ^{3170}	^{3226}ENTERH^{3231}	^{3165} ENTERH ^{3170}	^{3165} ENTERH ^{3170}
	QMKAAA	^{3140}QMKAAA^{3145}	^{3201}QMKAAA^{3206}	^{3197}QMKAAA^{3202}	^{3141}QMKAAA^{3146}

* Reported in some potyviruses

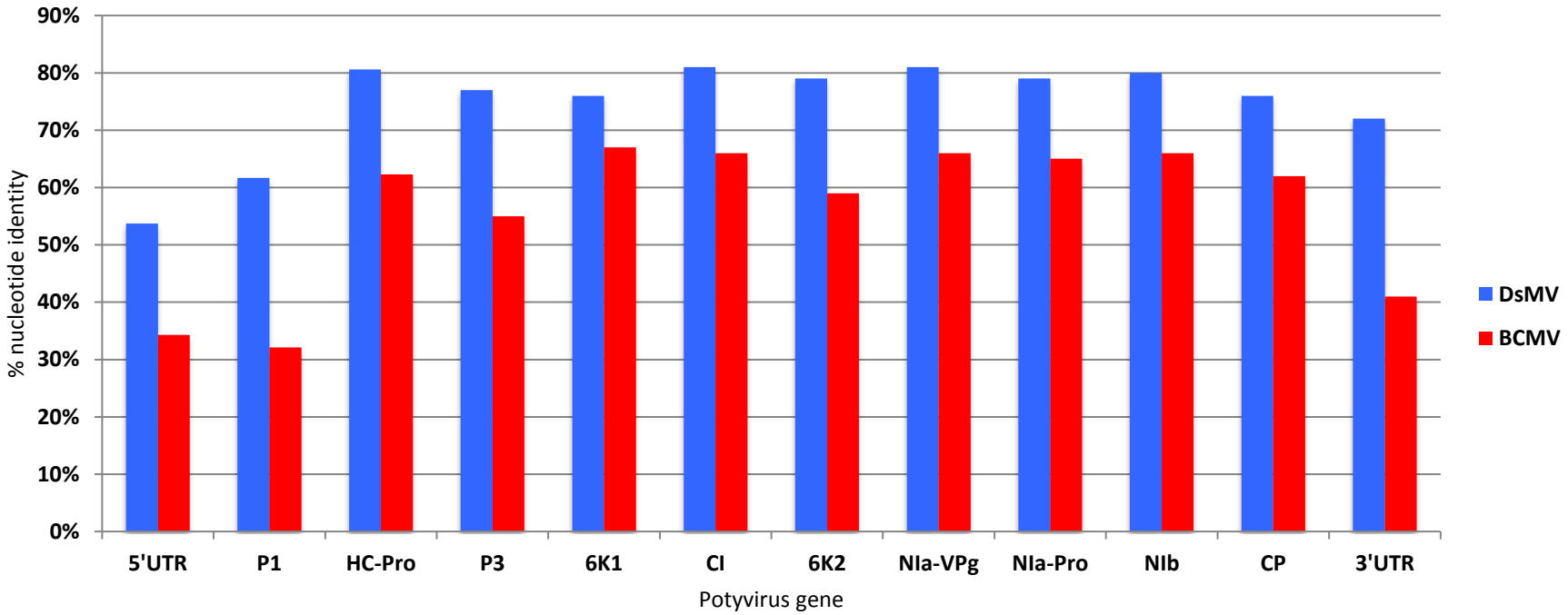


Figure 2.6: Average percentage nucleotide sequence identity of VanMV vs DsMV (blue) and VanMV vs BCMV group (red) for each individual gene within the viral genome. For DsMV, sequences used were DsMV (KJ786965, NC003537, and KT026108) and for BCMV sequences used were ZaMMV (AY626825), WVMV (AY656816) and BCMV (AJ312437).

2.4.2 Phylogenetic analysis of VanMV

The neighbour-joining phylogenetic tree in Figure 2.7 was constructed to determine the relationship of VanMV-CI to other members of the potyvirus genus. This analysis shows that VanMV-CI (indicated in yellow) can be found in the BCMV clade (indicated in green). This was expected since DsMV has been shown to be a member of this group (Reyes et al., 2009). The BCMV clade, distinct from other members of the potyvirus genus, is statistically supported with a bootstrap value of 87. Low bootstrap values for the earliest branch points indicate rapid expansion of potyviral species in the past. This is supported by studies of Gibbs and others looking at the evolutionary history of potyviruses (Gibbs & Oshima, 2010).

Maximum likelihood analysis was also carried out to determine the relationship of VanMV-CI to DsMV and other members of the BCMV group of potyviruses at the genome level (Figure 2.8). This analysis supports previous studies showing a monophyletic relationship where all members of the BCMV group appear to descend from a single common ancestor. This analysis also indicates a clustering of two separate clades A and B that are distinct from one another. Clade A consists of all other BCMV group sequences separate from the VanMV-CI and DsMV sequences that make up clade B. The DsMV KJ786965 and KT026108 sequences (unpublished) were reported from India while the NC003537 sequence was reported from China (Chen et al., 2001). According to this phylogenetic analysis, the lineage leading to VanMV-CI appears to be slightly older than that of the DsMV genomes suggesting that VanMV-CI may have appeared before these DsMV sequences.

Maximum likelihood analysis was carried out on each of the individual viral genes (Figures 2.10-2.18) to determine whether a different evolutionary history has occurred for the different genes. This analysis showed very similar topology to that of the full genome analysis. Apart from the non-coding regions of the genome (Figure 2.9 and Figure 2.19), 6K2 (Figure 2.15) and the CP gene (Figure 2.20), the lineage leading to VanMV-CI appears to be earlier than the lineage of DsMV for all other genes. This branching arrangement is well supported in almost all cases (except for NIa-Pro) with bootstrap values greater than 50. VanMV-CI is highlighted in yellow while DsMV and available published sequences of the BCMV group are indicated in green (Figure 2.7).

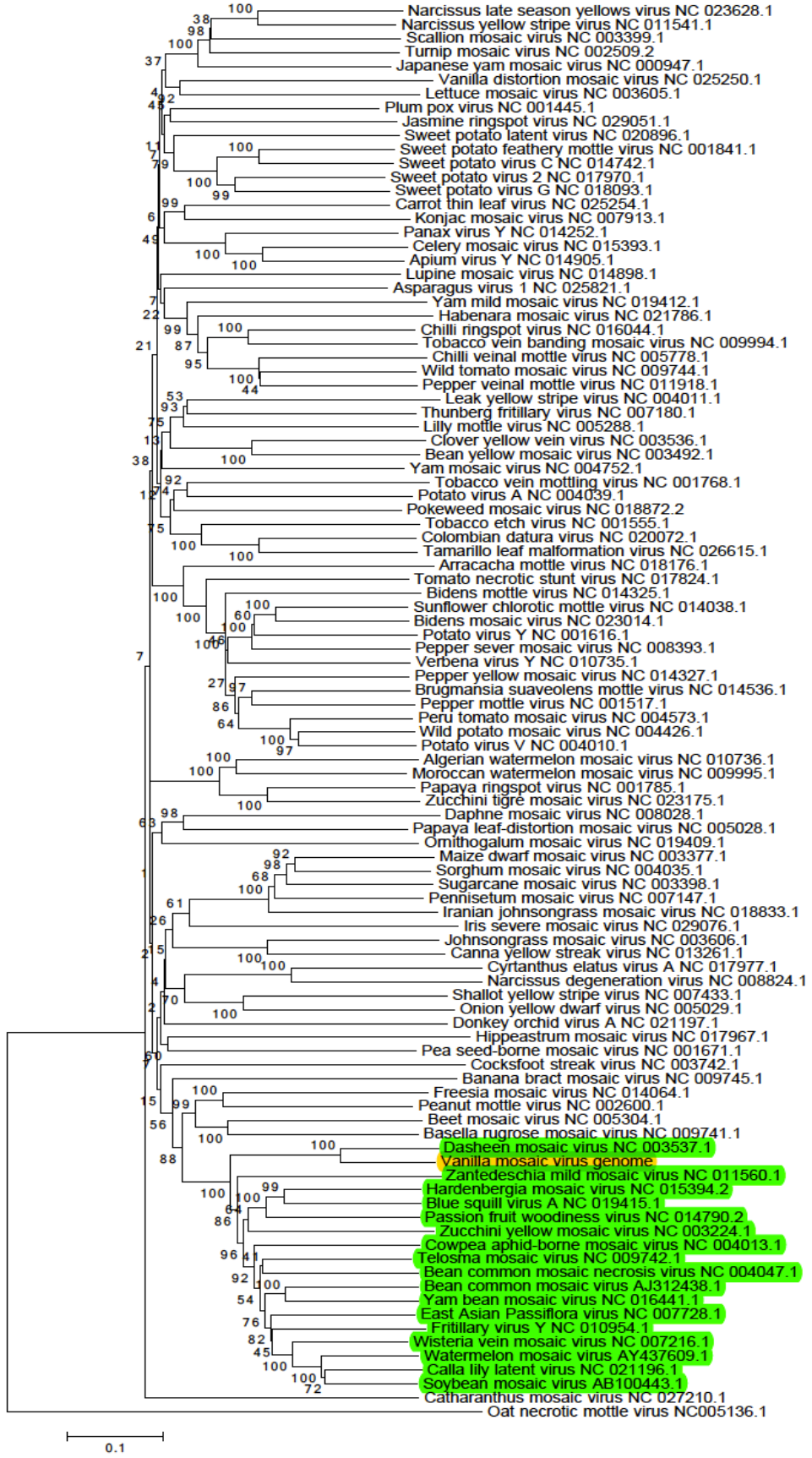


Figure 2.7: A Neighbour joining phylogenetic tree of potyvirus genomes. VanMV-CI (yellow) is classified in the BCMV group (Green). The phylogenetic tree was constructed using 1000 bootstrap values with oat necrotic mottle virus (ONMV) as the out-group. The scale bar represents the number of nucleotide substitutions per site.

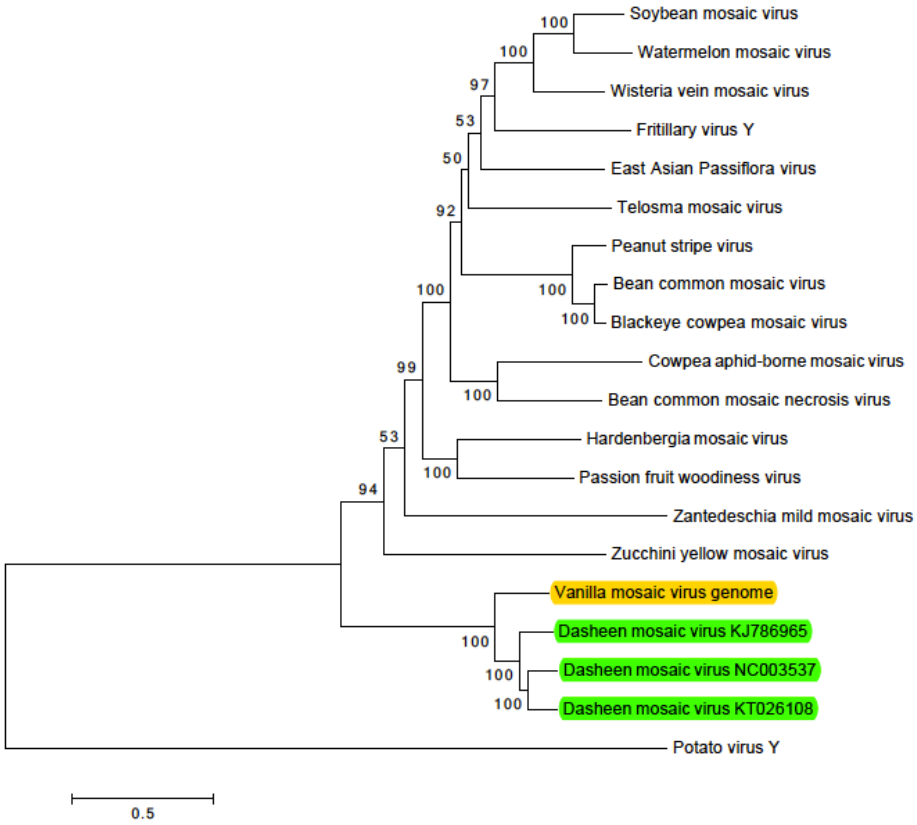


Figure 2.8: Maximum likelihood tree of the VanMV-CI genome against all available genomes of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

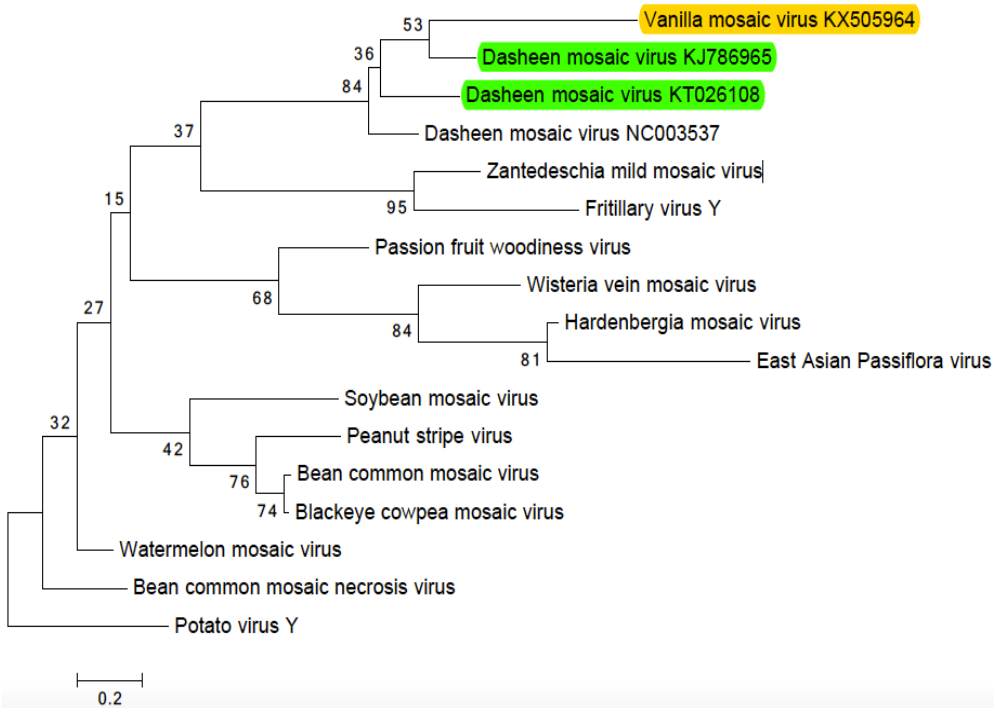


Figure 2.9: Maximum likelihood tree of 5'UTR of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

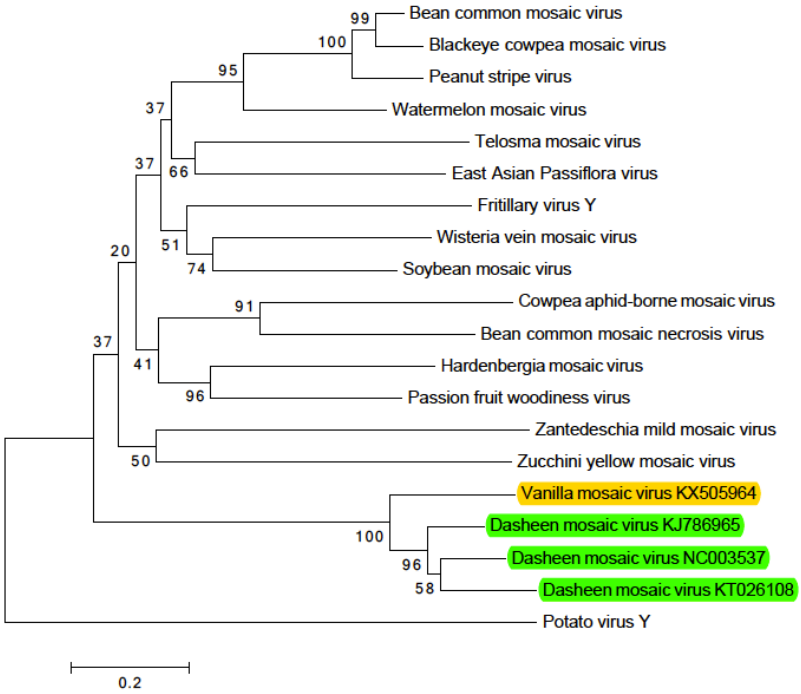


Figure 2.10: Maximum likelihood tree of P1 of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

Figure 2.11: Maximum likelihood tree of HC-Pro of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

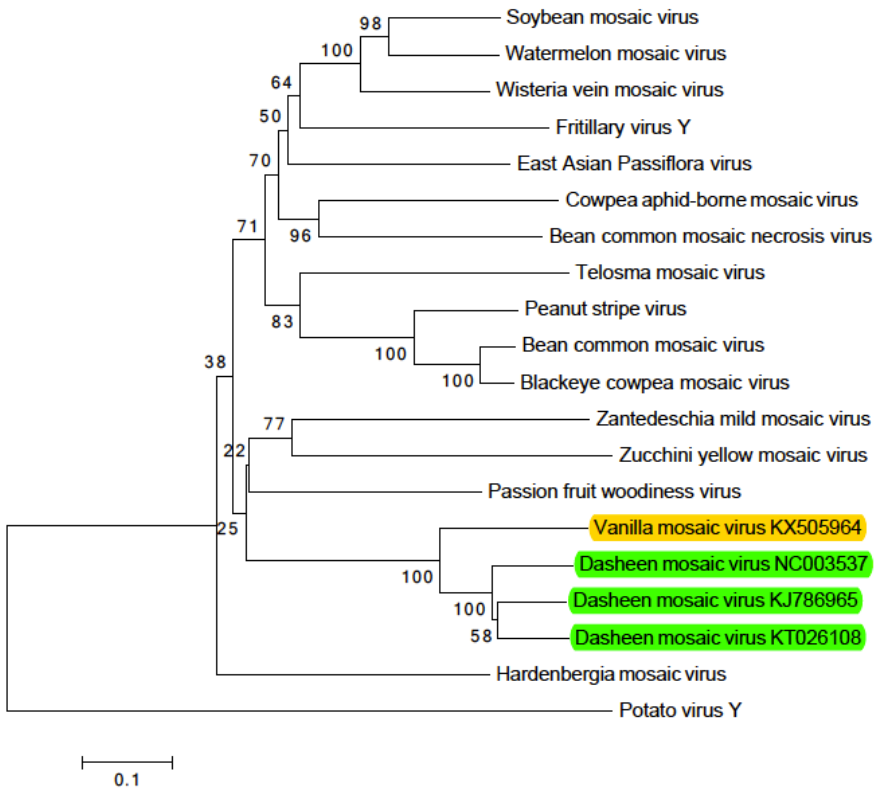


Figure 2.12: Maximum likelihood tree of P3 of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

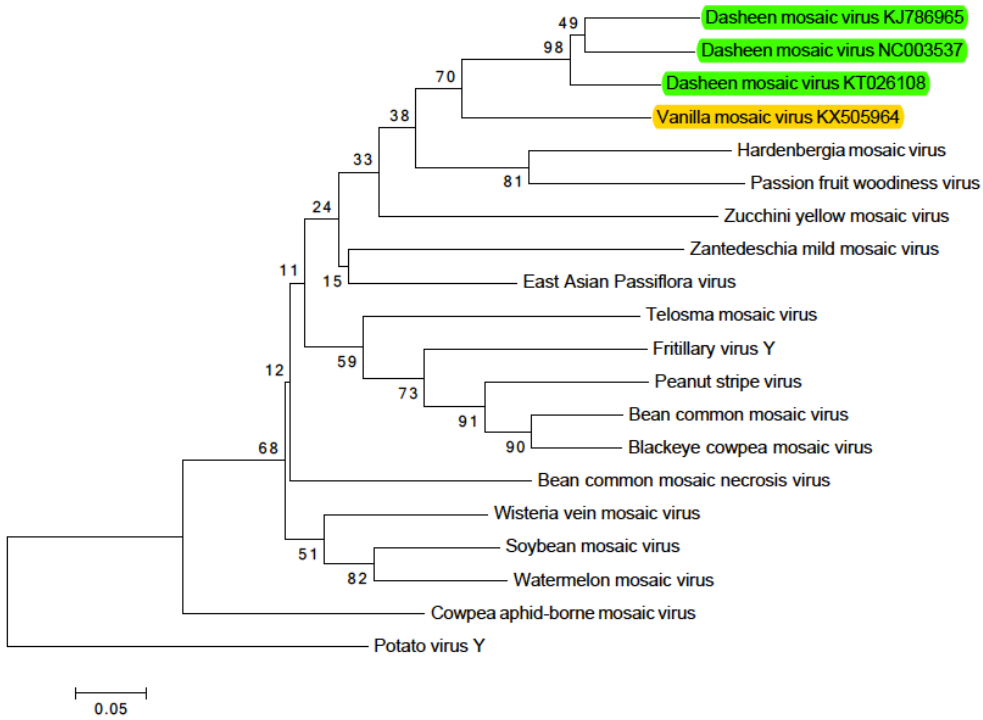


Figure 2.13: Maximum likelihood tree of 6K1 of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

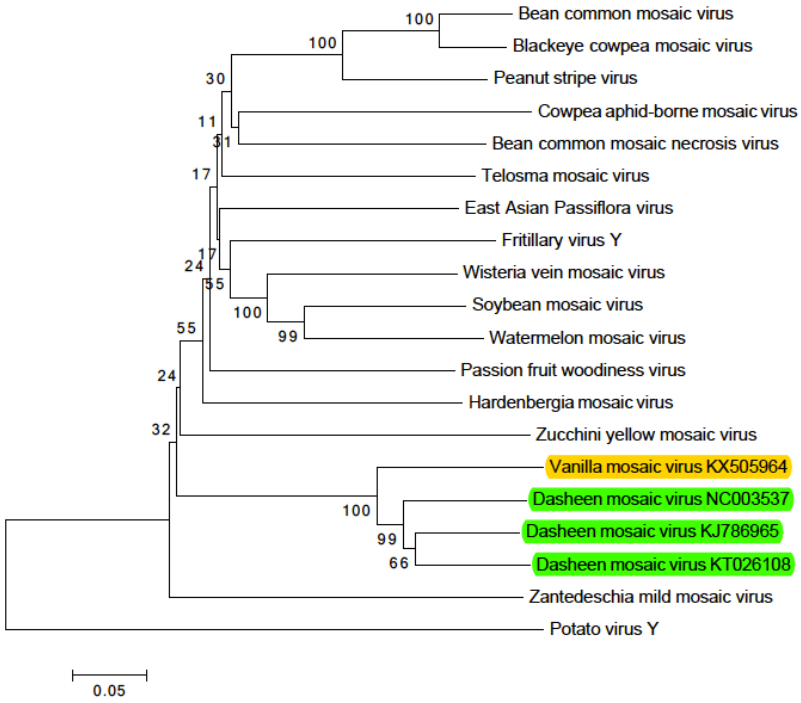


Figure 2.14: Maximum likelihood tree of CI of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

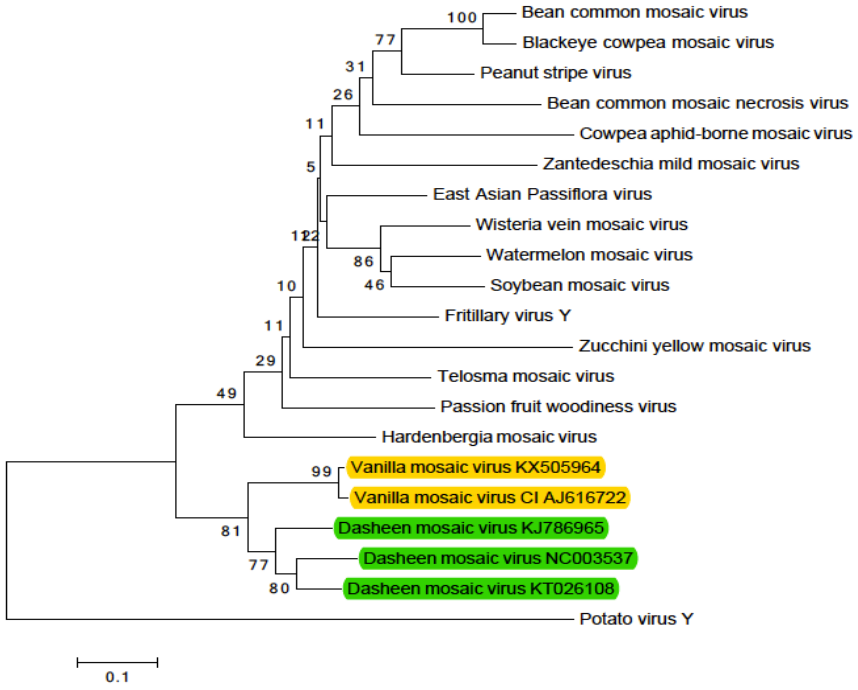


Figure 2.15: Maximum likelihood tree of 6K2 of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

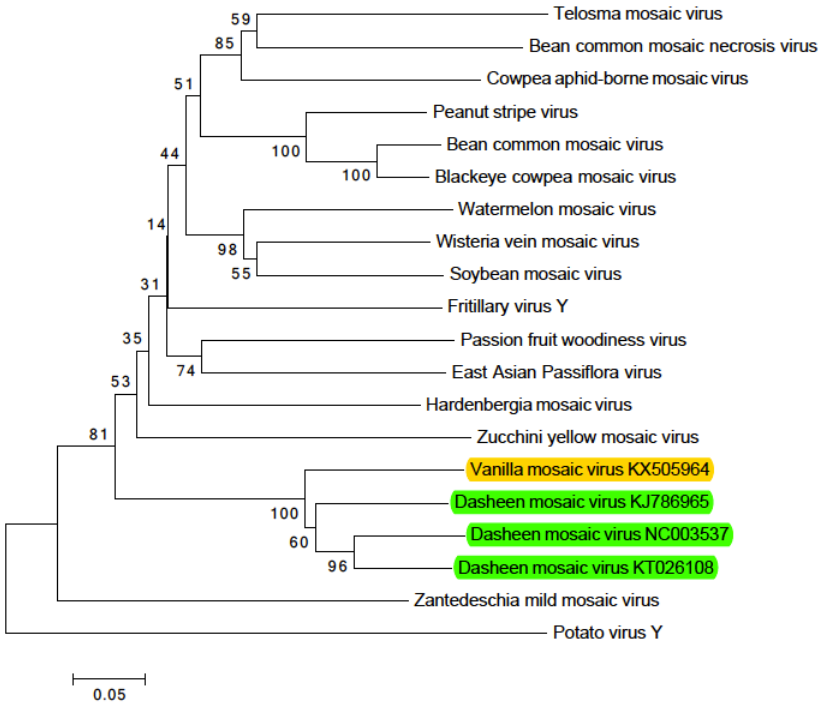


Figure 2.16: Maximum likelihood tree of NIa-VPg of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

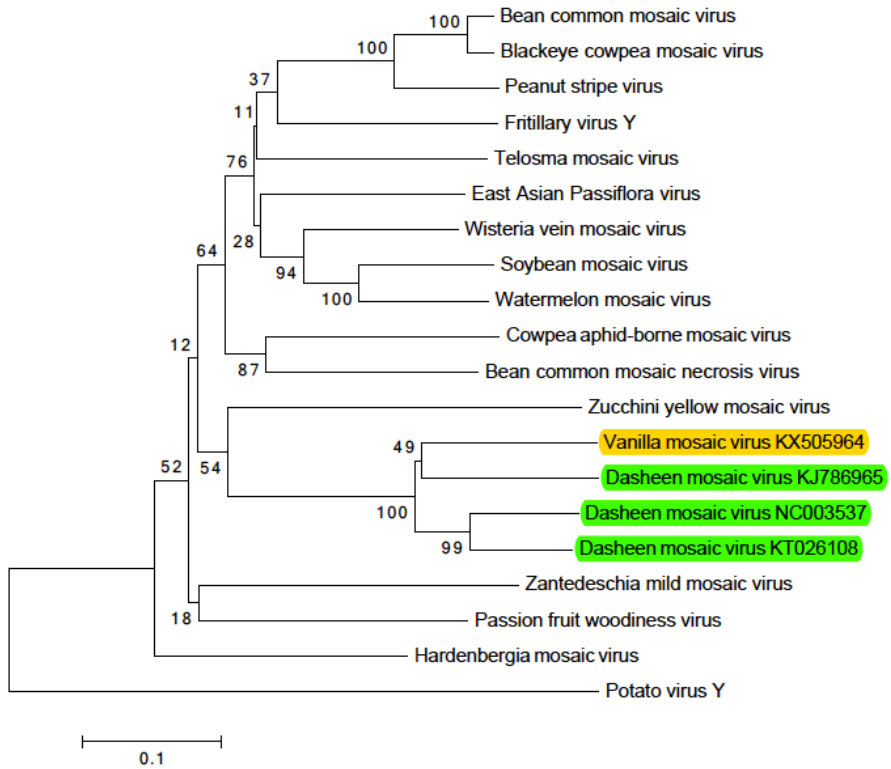


Figure 2.17: Maximum likelihood tree of NIa-Pro of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

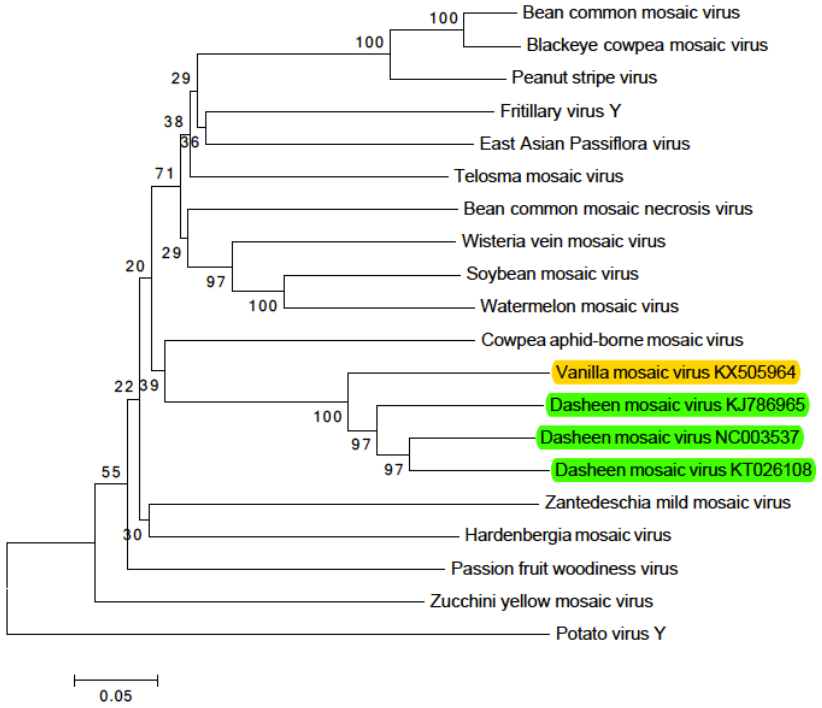


Figure 2.18: Maximum likelihood tree of NIb of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

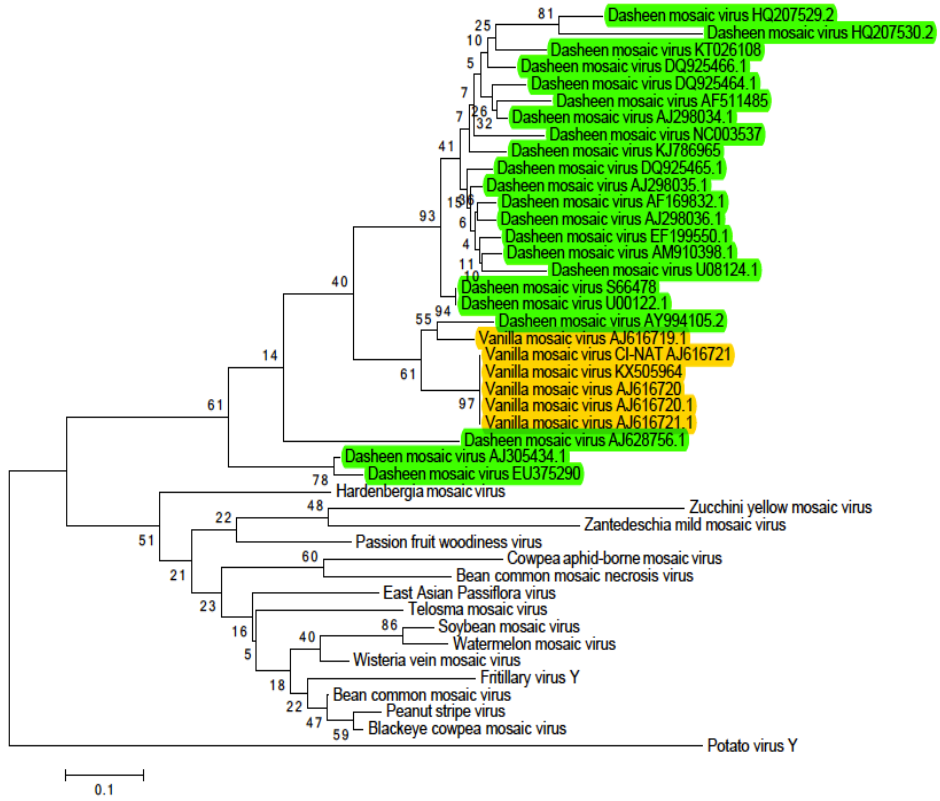


Figure 2.19: Maximum likelihood tree of 3'UTR of VanMV-CI and other members of the BCMV group. The scale bar represents the number of nucleotide substitutions per site. Sequences and accession numbers for BCMV sequences are included in Table 2.1.

In a maximum likelihood analysis of VanMV-CI and all complete CP sequences of DsMV (Figure 2.20), it appears that VanMV-CI is more closely related to DsMV from the South Pacific. This analysis also shows that the VanMV-CI lineage is not older than the DsMV lineage as speculated earlier by the genome analysis (Figure 2.8), but rather is preceded by DsMV sequences from the South Pacific, namely from New Zealand. Further, it also appears that VanMV-CI is preceded by a number of other DsMV sequences from China and Japan, where they are basal to VanMV-CI in Figure 2.20.

Based on current CP sequences available, the phylogenetic analysis in Figure 2.20 suggests a DsMV origin in China or nearby Asian countries. From this analysis, DsMV sequences from China are basal to other CP sequences and the lineage leading to it appears to be the oldest. Based on the CP nucleotide alignment (Figure 2.5), the phylogenetic analysis also shows the lineage leading to the appearance of DsMV in India is slightly older than that leading to VanMV-CI suggesting that DsMV in India preceded VanMV-CI and DsMV in the South Pacific. The phylogenetic analysis presented in this research is consistent with the initial appearance of VanMV-FP prior to the emergence of VanMV-CI in the South Pacific, as reported in the literature (Farreyrol et al., 2006; Wisler & Zettler, 1987).

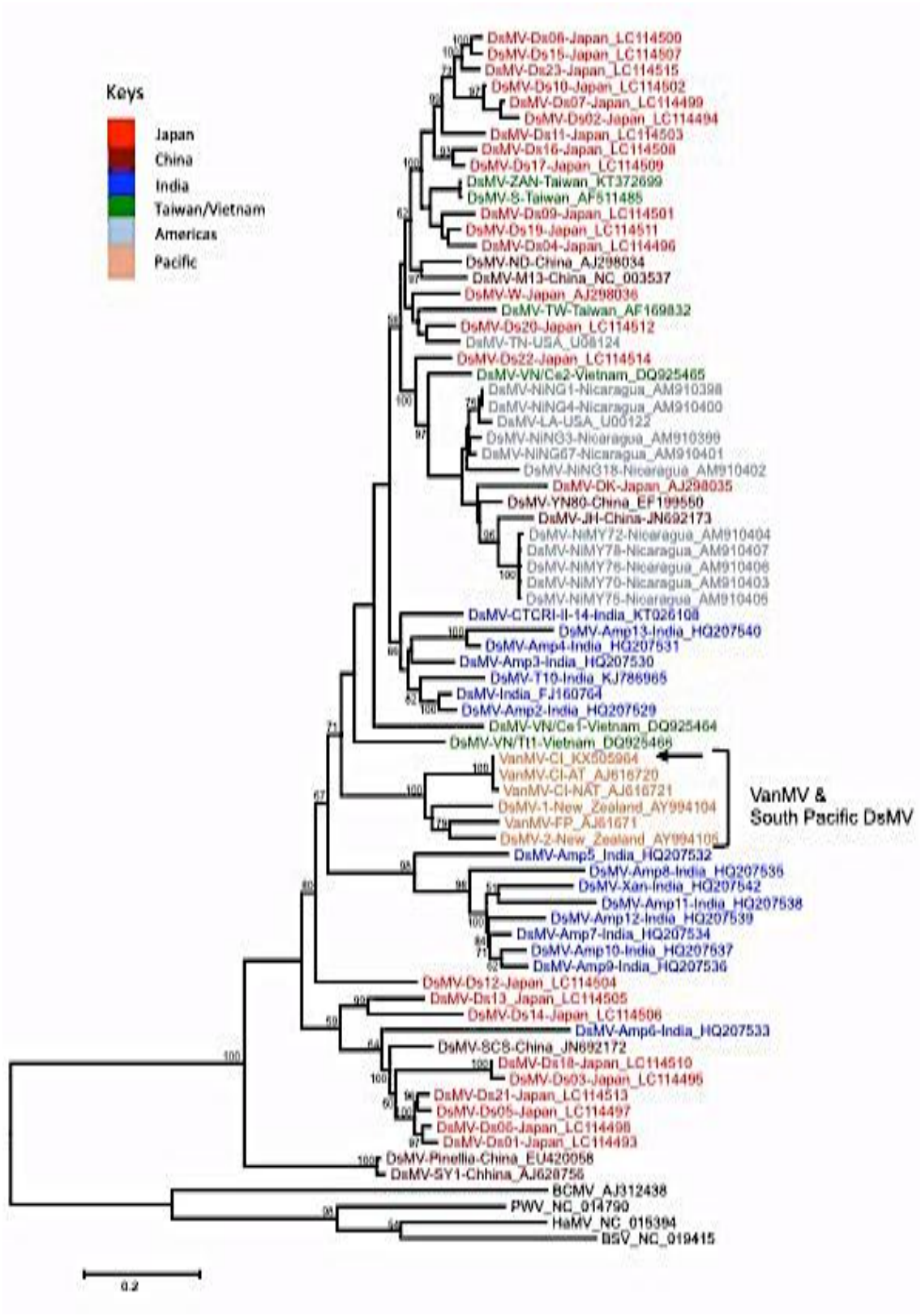


Figure 2.20: Maximum likelihood tree of VanMV-CI CP against all available CP genes for DsMV. The scale bar represents the number of nucleotide substitutions per site. Accession numbers for each sequence is indicated in each taxa. BCMV (AJ312438), PWV (NC014790) and BSV (NC019415) used as the out-group.

2.4.3 Sequence comparisons of VanMV with other potyviruses

Sequence comparisons of VanMV-CI genome with genomes from members of the *Bean common mosaic virus* family showed, as expected, that VanMV has greatest sequence identity to DsMV than to other members of the BCMV family (Figure 2.6). Table 2.7 indicates that at the whole genome level, VanMV-CI shows 75.8-77.1% nt and 81.9-86.7% aa identity to DsMV, compared with 56.1-60.4% nt and 56.4-61.5% aa identity for other members of the BCMV family (Figure 2.4 and Table 2.7). When compared to DsMV, the HC-Pro, CI and NIa-VPg gene shows the highest identity while the 5'UTR and P1 gene show the lowest.

Table 2.7: The percent of amino acid (aa) and nucleotide (nt) sequence identities of VanMV-CI genome and individual genes compared to publically available DsMV sequences (KT86965, NC003537, KT026108).

Region	DsMV KJ86965		DsMV NC003537		DsMV KT026108	
	aa %	nt %	aa %	nt %	aa %	nt %
Genome	81.9	75.8	83.7	76.5	86.7	77.1
5'UTR	-	53.3	-	53.6	-	54.1
P1	61.2	60.7	62.6	62.5	62.3	61.9
HC-Pro	92.4	80.1	93.0	80.9	92.8	80.9
P3	77.7	74.1	72.5	73.2	76.3	82.4
6K1	90.4	71.2	92.3	76.3	92.3	79.5
CI	95.3	80.8	90.4	81.3	95.1	81.3
6K2	86.8	81.8	83.0	76.7	83.0	78
NIa-VPg	65.2	80.0	94.7	82.3	95.8	81.1
NIa-Pro	88.9	78.5	87.7	77.5	89.7	79.8
NIb	88.6	79.1	89.1	79.6	89.7	80.5
CP	81.8	75.1	84.1	77.5	82.4	76.7
3'UTR	-	72.4	-	71.7	-	73.2

2.4.4 Recombination analysis

Comparison of nucleotide sequence trees for different regions of the genome (Figure 2.10-2.18) shows that VanMV-CI is always most closely related to DsMV. Thus, if VanMV arose as a result of a recombination event involving DsMV, or vice versa, this has not been detected by the phylogenetic analysis. Therefore, further analysis of potential recombination events was carried out on available potyvirus genomes using RDP4 software to assess the possibility of recombination at a finer scale. Analysis of the recombination event reported to have occurred between BCMV and SMV leading to WMV specifically in the N-terminal region of the P1 gene (Desbiez & Lecoq, 2004) was used to validate the enquiry. From this analysis, DsMV and VanMV-CI were shown to have had six unique recombination events occurring within their genomes (Figure 2.21). While events 1, 5 and 6 were unique to DsMV, events 2, 3 and 4 were found in both DsMV and VanMV-CI. These events (2, 3 and 4) correlate with the C-terminus of P1, N-terminus of HC-Pro and C-terminus of the P3 gene, respectively. It can be seen that the region identified in event 2 in VanMV-CI is longer than that region of DsMV.

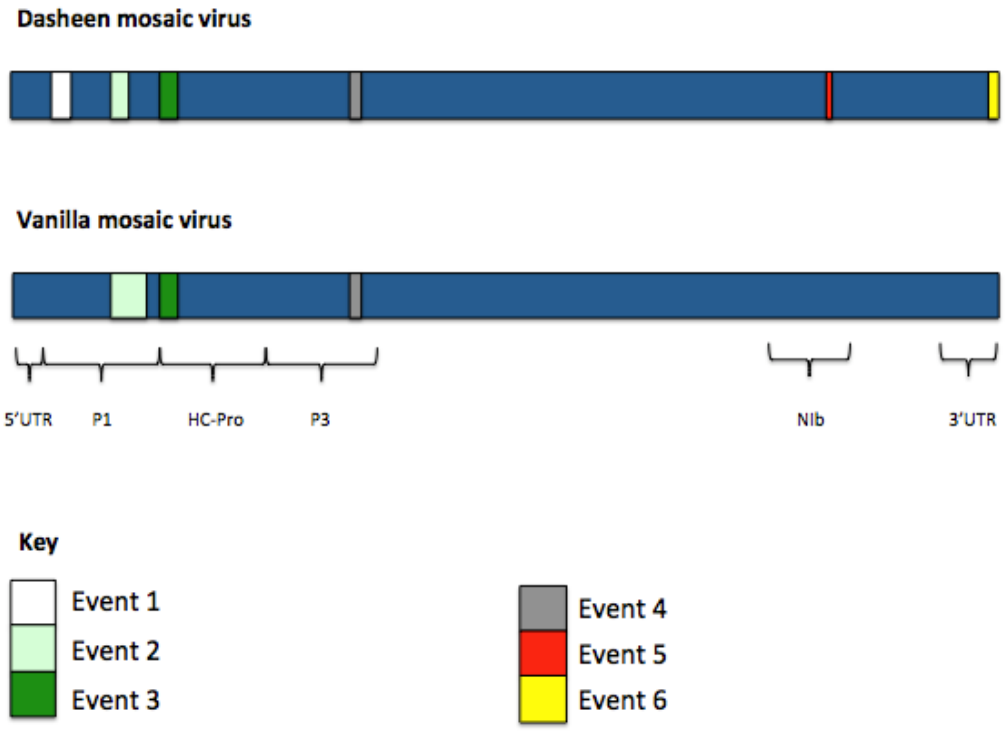


Figure 2.21: Recombination analysis of DsMV and VanMV. Possible recombination events taking place between DsMV (NC003537) and VanMV-CI genome against all available potyvirus genomes. The different colours represent unique recombination events detected in RDP4. Each coloured region covers the approximate area in which the events can be found.

These common predicted recombination events are summarised in Table 2.8. For events 2, 3 and 4, the major parents appear to be BCMV, WMV and telosma mosaic virus (TeMV), respectively. This analysis shows 61.8, 62 and 62.8% nucleotide identity, respectively, in the region of recombination for the major parents. Each of these events has been detected using the GENECOVN method in RDP4 with statistically significant p-values. Minor parents identified in Table 2.8 were sweet potato virus (SPV) (events 2 and 4) and lilly mottle virus (LMoV) (event 3).

Events 1, 5 and 6 were unique to DsMV (Figure 2.21). These correspond to the N-terminus of P1, C-terminus of NIb and the 3'UTR region, respectively. Each of these events were identified by at least one detection method all having statistically significant p-values (Table 2.8). Event 1 was identified by GENECOVN, event 5 by Chimaera and event 6 by both GENECOVN and MaxChi. A major parent for each event appeared to be BSVA (events 1 and 6) and tobacco vein mottling virus (TVMV) (event 5) with nt identity of 61.5% (event 1), 75.5% (event 5) and 63% (event 6). Minor parents identified were LuMV (event 1), BCMV (event 5) and bidens mottle virus (BiMoV) (event 6).

Sequence identities below 75% nt between DsMV/VanMV-CI and its major parent was regarded insufficient evidence for recombination. Further, the overall detection of any event by at least three methods was considered sufficient evidence for recombination. Events 1 (N-terminus of P1) and 5 (C-terminus of NIb) were detected by only one method while recombination events 2 (C-terminus of P1), 3 (N-terminus of HC-Pro), 4 (C-terminus of P3) and 6 (3'UTR) were detected by two methods out of seven used by RDP4 (Table 2.8). Apart from event 5 having 75% nt identity to its major parent, all other recombination events found $\leq$63% nt identity to their major parents. The recombination event for SMV and BCMV as previously identified by Desbiez & Lecoq (2004) was detected by five methods with >75% nt identity to its major parent (data not shown) suggesting a reliable recombination analysis. Although this analysis verified the recombination event between SMV and BCMV, it did not identify sufficient evidence for recombination between VanMV/DsMV with other potyviruses.

Table 2.8: Recombination analysis of DsMV and VanMV. Recombination confirmation analysis based on the events outlined in Figure 2.21. Each break and end break point indicated corresponds to the position in sequence rather than alignment. Only the major parent is shown with percentage sequence similarity to the recombinant. The minor parent is also included as is the detection method with average P-value.

Event	Break point	End break point	Major parent and accession	Nt similarity (%)	Minor parent and accession	Detection method and Average P-Value
1	DsMV 331	DsMV 556	BSVA (NC019415)	61.5	LuMV (NC014898)	GENECOVN $1.919X10^{-4}$
2a	DsMV 773	DsMV 1132	BCMV (AJ312438)	61.7	SPV (NC017970)	GENECOVN $2.245X10^{-7}$ Chimaera $6.066X10^{-11}$
2b	VanMV 646	VanMV 1185	BCMV (AJ312438)	62	SPV (NC017970)	GENECOVN $2.245X10^{-7}$ SiScan $6.066X10^{-11}$
3a	DsMV 1533	DsMV 1658	WMV (AY437609)	62.8	LMoV (NC005288)	MaxChi $2.048X10^{-2}$
3b	VanMV 1427	VanMV 1552	WMV (AY437609)	62.7	LMoV (NC005288)	MaxChi $2.048X10^{-2}$
4a	DsMV 3515	DsMV 3628	TeMV (NC009742)	62.3	SPV (NC017970)	GENECOVN $4.422X10^{-2}$
4b	VanMV 3409	VanMV 3523	TeMV (NC009742)	63	SPV (NC017970)	GENECOVN $4.422X10^{-2}$
5	DsMV 8336	DsMV 8388	TVMV (NC009994)	75.5	BCMV (AJ312438)	Chimaera $1.872X10^{-2}$
6	DsMV 9993	DsMV 10039	BSVA (NC019415)	63	BiMoV (NC014325)	GENECOVN $3.07X10^{-7}$ MaxChi $3.0319X10^{-4}$

2.5 Discussion

Unlike studying larger complex organisms, viruses of various classes are unable to leave behind clues such as fossil records to provide some basis for their identification and evolutionary analysis. For this reason studying the likely origins of RNA viruses has proven challenging, and deciphering the different evolutionary trajectories taking place can become problematic (Durzyńska & Goździcka-Józefiak, 2015; Koonin et al., 2015). It has only been due to the recent developments in molecular biology that it has been possible to sequence and identify virus sequences, distinguishing them more easily from others. Due to the quasi-species nature of RNA viruses, in particular, it is feasible that those virus sequences present today are remnants of ancient random mutation events and natural selection. Viruses also evolve through recombination within and between species (Sztuba et al., 2011). As more and more virus sequences are becoming available, a clearer depiction of the recent past is becoming more possible although not decisive.

Is VanMV-CI a strain of DsMV?

While *Dasheen mosaic virus* is classified as a recognised species within the *Potyvirus* genus, VanMV is not. This study reports the first complete genome of VanMV, which has a typical potyvirus genome structure with high sequence similarity to publically available DsMV genomes. Current criteria for potyvirus species discrimination include greater than 75% and 80% sequence identity at the nt and aa levels, respectively, either across the CP or complete genome. The species demarcation criteria also include virus similarities in inclusion body morphologies, polyprotein cleavage sites and host range (Adams, Antoniw, & Fauquet, 2005). Other demarcation criteria, which have not been satisfied by previous studies, are presented here; namely nt and aa sequence identity of the whole genome and predicted polyprotein cleavage sites. From this study, VanMV-CI is 75.1-77.1% and 81.9-86.7% identical to DsMV at the nt and aa (polyprotein) levels, respectively, with similar results for the complete CP gene. In addition, the VanMV-CI deduced polyprotein shares a number of similar cleavage sites to published DsMV polyproteins with noticeable differences only in the P1/HC-Pro and CI/6K2 junction. Further, VanMV proteins have most of the conserved potyvirus motifs with all predicted motifs being the same as those in DsMV with the exception of the IVFG motif within the P1 protein. Based on these results, this study

provides additional evidence to support the proposition by Farreyrol et al. (2006) that VanMV-CI is a strain of DsMV that exclusively infects the orchid vanilla.

Evolutionary history of VanMV and DsMV

DsMV was first reported in 1970 from Florida, USA while VanMV was later reported in 1986 from French Polynesia infecting *V. tahitensis* (Wisler & Zettler, 1987; Zettler et al., 1970). The discovery of VanMV was a result of a nationwide investigation initiated by French Polynesia to determine the cause of the drop in vanilla production with the aim of restoring vanilla export to its former levels. The decline in vanilla production beginning in 1967 was most likely due to the viruses discovered and reported in 1986 (Wisler & Zettler, 1987). This report suggests that VanMV may have emerged in French Polynesia around 1967. This appearance is approximately the same time DsMV was reported by Zettler et al. (1970) in Florida. VanMV virus was later found in the Cook Islands, Fiji and Vanuatu (Farreyrol et al., 2006). Based on these reports alone it is difficult to discern whether DsMV or VanMV emerged as a pathogen first.

The maximum likelihood analysis of the complete VanMV-CI genome and DsMV genomes published to date and studies carried out by Babu et al. (2011) suggest that the lineage leading to VanMV-CI is older than that of DsMV (Figure 2.8). Apart from the 5'UTR, 3'UTR, 6K2 and the CP gene all other maximum likelihood analyses show a similar topology to the genome analysis proposing the same view. However, this was based only on the limited DsMV sequence information published, with only one VanMV sequence for comparison.

A maximum likelihood analysis of all complete published CP sequences of DsMV (Figure 2.20) does not support the proposition implied by the genome analysis. The CP analysis not only shows that DsMV from the South Pacific (AY994104 and AY994105) precedes VanMV, but also a number of other DsMV variants basal to the South Pacific sequences in Figure 2.20, particularly from China and Japan, show older lineages. Further, the phylogenetic analysis in Figure 2.20 shows that DsMV most likely originated in China, which then later appeared in Japan and India before appearing in the South Pacific. The origin of DsMV proposed by this analysis is further supported by the fact that *Colocasia* and *Xanthosoma spp* (natural hosts of DsMV) most likely originated from China/ South East Asia (Govarts, 2012). It is therefore feasible to

suggest that DsMV originated in the same region as its host that was later distributed to the South Pacific following human migration. The difference in evolutionary trajectories of VanMV and DsMV observed between genome and CP analyses is most probably due to the lack of DsMV and VanMV genome sequences published.

Where is the host range determinant in VanMV?

Since its reporting in 1986, VanMV had only been partially sequenced, specifically towards the 3' region of the genome (Farreyrol et al., 2006). Whilst considered a strain of DsMV, no information has been available regarding the difference in host range where VanMV infects the orchid vanilla while DsMV infects aroid plants and is unable to infect vanilla (Wang & Pearson, 1992; Wisler & Zettler, 1987). Analysis and comparison of DsMV and VanMV-CI provides an opportunity to identify candidate genome regions that may be responsible for host range differences.

The aa repeat region identified in the N-terminal region of DsMV CP is extensive in the French Polynesian isolate of VanMV but absent in the CI isolates (Figure 2.5). In addition, the repeat sequence in the French Polynesian isolate of VanMV is similar to that in New Zealand isolates of DsMV. Further, the repeat sequences within DsMV are varied. This suggests that the host range determinant is not within the highly variable N-terminal region of the CP of VanMV-CI.

Previous studies carried out on potyviruses have suggested a host range determinant in the 5' end of the genome (Sáenz et al., 2002; Salvador et al., 2008; Shan et al., 2015; Verchot, 1995). Sáenz et al. (2002) has suggested that the HC-Pro region, demonstrated in PPV specifically, has host-determining properties. However, the HC-Pro sequence of VanMV, although possibly linked to host range specificity, was more similar to that of DsMV. One of the main differences between VanMV-CI and DsMV can be found in the 5' region in the 5'UTR and the P1 region. This difference owes mainly to the highly variable N-terminal region of P1. Of particular interest is the finding that the VanMV-CI P1 protein is 30 aa shorter than DsMV P1. Previous studies by Salvador et al. (2008) show that a hybrid construct of PPV/TVMV P1 was unable to infect a natural host of PPV. For each hybrid, TVMV either replaced the N-terminus or the entire PPV P1 region. Only the construct containing the natural PPV P1 protein was able to infect its natural host suggesting a host-determining role either in the C-

terminus, N-terminus or both regions of the P1 protein. In addition, a previous study identified a strain of SMV infecting *P. ternate*, a natural host of DsMV, with a P1 region more similar to DsMV than SMV (Valli et al., 2007). These studies support a role for the N-terminus of P1 in host selection. Further, Verchot & Carrington (1995) speculate that the N-terminus of P1 may have regulatory functions in fine-tuning the virus replication cycle. This hyper-variable region may also play an important role in influencing P1 properties such as turnover and structural stability (Verchot & Carrington, 1995). The difference in P1 in this case may influence the activity of this protein in either or both strains, providing host range specificity.

Analysis of the potyviral polyprotein cleavage sites highlighted a difference between DsMV and VanMV-CI in the P1/HC-Pro junction. VanMV-CI differs from all DsMV sequences in that it had a predicted VHQY/T cleavage rather than the IVQY/A site identified in DsMV. Other junctions in the polyprotein also displayed differences without as many aa changes. This junction is a result of the self-cleaving P1 protein in which the core domain encodes a protease responsible for the catalytic activity. In a previous study, Verchot & Carrington (1995) found that accumulation of a number of mutations in the C-terminus of TEV P1 debilitated its proteolytic activity and in turn virus infection. However, when a sequence containing a cleavage site recognised by NIa was inserted, virus infection was restored. This suggests that although the proteolytic domain of P1 is critical for its putative functions, it is not required for virus infection but rather the separation of P1 from HC-Pro. Further, Shan et al. (2015) demonstrate that the separation of P1a and P1b, homologues of potyviral P1 and HC-Pro proteins, is host specific. In that study, constructs were created with P1 and HC-Pro sequences of PPV and P1a and P1b genes of cucumber vein yellowing virus (CVYV). Constructs encoding PPV proteases (PPV-P1/PPV-HC-Pro and PPV-P1/CVYV-P1b) infected *N. benthamiana,* a natural host of PPV, with minor differences, but very little infection in *Cucumber sativus*, a natural host of CVYV. Alternatively, constructs encoding CVYV proteases (CVYV-P1a/CVYV-P1b and CVYV-P1a/PPV-HC-Pro) showed infection in *C. sativus* but very little infection in *N. benthamiana*. The study by Shan and colleagues not only suggests that the separation of P1 and HC-Pro is critical for virus infection, but that the proteolytic activity of P1 is only compatible with certain hosts, and thereby may be involved in determining host range.

Phylogenetic analysis of each gene did not appear sensitive enough to detect recombination events within the VanMV-CI and DsMV genomes. A recombination analysis was therefore performed on the potyvirus sequences using RDP4 software, which incorporated seven detection methods. For this analysis, recombination events detected by more than three methods with more than 75% nt identity to its major parent was considered sufficient evidence of recombination (Edwards, 2014). The recombination event predicted between SMV and BCMV in the N-terminal region of the P1 gene producing WMV (Desbiez & Lecoq, 2004) was detected by five methods all supported by statistically significant P-values (data not shown). This was a good indication of a reliable data set in which any recombination could be investigated for DsMV and VanMV-CI. Interestingly, two unique recombination events were detected in the N-terminus of P1 and 3'UTR of DsMV. These events however, were only detected using one method (GENECOVN) and not identified in six other methods incorporated into the software. In addition, the region of proposed recombination had less than 75% nt identity to its major parent. These results indicate that the recombination event detected, particularly in the N-terminus of P1, is not well supported and therefore is not sufficient evidence for an actual recombination event. It can therefore be inferred that the hyper-variable N-terminus of VanMV-CI P1 is not a result of a recombination event but rather an accumulation of random mutations and natural selection.

Based on the research carried out in this study, it appears that VanMV is a strain of DsMV that exclusively infects vanilla and is most closely related to DsMV sequences from the South Pacific. Based on the CP analysis, the lineage leading to VanMV-CI also appears more recent compared to South Pacific DsMV sequences suggesting that VanMV arose from DsMV in the same region. According to evidence provided in the literature, it appears that the 5' region of the genome, specifically the P1 gene, can determine differences in host range between any two potyviruses. Sequence comparisons of VanMV-CI and DsMV reveal noticeable differences in this region, particularly the N-terminal region of P1, which may be responsible for the shift in host range. A recombination analysis of VanMV-CI and DsMV against all potyvirus sequences available reveals that the hyper-variable N-terminus of VanMV-CI P1 is less likely a result of a recombination event between DsMV and another virus, but rather

an accumulation of random mutations and natural selection. These random mutations would have acquired a P1 protein that was compatible with infecting the orchid vanilla. Sequencing and comparing more vanmv and South Pacific DsMV genomes will potentially create greater clarity about the molecular relationship between these viruses.

Chapter 3

Towards establishing the impact of P1 on HC-Pro in silencing suppression

3.1 Introduction

Positive sense RNA viruses are among the most abundant plant viruses infecting economically and agriculturally important plants worldwide (Hyodo & Okuno, 2016). Their abundance is largely due to the quickly evolving nature of RNA viruses, which enables them to exploit host cell mechanisms for successful infection by manipulating host factors, which evolve only slowly (Hyodo & Okuno, 2016). Virus pathogenesis includes all the processes involved in establishing virus infection by the plant-virus interaction irrespective of disease outcome (Hyodo & Okuno, 2016; Mäkinen & Hafren, 2014). This plant-virus interaction involves the virus entering plant cells by wounds created either mechanically or by vector organisms (Mandadi & Scholthof, 2013). Once in the cell, positive stranded RNA viruses undergo a number of partially overlapping processes in their replication cycle; namely, uncoatingto expose viral RNA, translation of the viral genome by the host cell, replication of the viral RNA genome, short-distance movement between cells, overcoming RNAi by the host cell and encapsidation of the new virus genome (Mäkinen & Hafren, 2014). Because potyviruses lack translation machinery, a critical step in their replication cycle is to have a compatible interaction with the host's translation machinery. If incompatible, the virus will fail to establish infection in its host; however, if compatible, viral factors are able to recruit host ribosomes producing viral replication proteins, which in turn form the virus replication complex (VRC) (Hyodo & Okudo, 2016). Further, an additional ORF known as PIPO is translated in the +2 frame relative to the main ORF (Chung et al., 2008). Following translation, P1, HC-Pro and NIa subsequently cleave potyviral polyproteins into ten mature proteins, which have downstream functions in the VRC, short and long distance movement and counter-attacking host RNAi (Hyodo & Okuno, 2016; Mäkinen & Hafren, 2014). Once the VRC complex is established, most of the potyviral proteins are involved in either genome replication, virion assembly or suppressing RNAi (Mäkinen & Hafren, 2014). In order for infection to occur, the virus must not only be compatible with recruiting the host cell's translation machinery and establish efficient replication; the virus must also be able to overcome RNAi. According to a number of studies (Lacombe, Bangratz, Vignols, & Brugidou, 2010; Maliogka et al., 2012; Merits, Guo, Järvekülg, & Saarma, 1999; Shan et al., 2015; Shiboleth et al., 2007) these processes require the 5' region of the potyvirus genome;

namely the P1 and HC-Pro genes, a region that shows some distinct differences between VanMV-CI and DsMV.

3.1.1 RNA interference

RNAi is a biological process found in eukaryotes in which RNA molecules regulate or silence gene expression (Carthew & Sontheimer, 2009; Karthikeyan et al., 2013; Katoch & Thakur, 2013). This process consists of two pathways, namely, the miRNA and siRNA pathways (Carthew & Sontheimer, 2009; Danielson & Pezacki, 2013). While the miRNA pathway is central to down regulating endogenous gene expression, the siRNA pathway is more concerned with the silencing of foreign pathogenic organisms (Vazquez et al., 2010; Xie & Qi, 2008). This multistep process in plants (Section 1.7.1, Figure 1.13) is able to counter attack the infection by virus pathogens (Danielson & Pezacki, 2013; Xie & Qi, 2008). Sources of exogenous genetic material triggering the siRNA pathway can be one of four types as illustrated in Figure 1.12 (Shimura & Pantaleo, 2011). Once triggered, the siRNA pathway works to protect the plant from viral infection by silencing exogenous viral nucleic acid. This pathway results in the degradation of the foreign viral genetic material, accomplished by a complex set of events beginning with dsRNA processing by DICER and ending in a RISC -containing Argonaute degrading foreign genetic material (Carthew & Sontheimer, 2009; Danielson & Pezacki, 2013; Karthikeyan et al., 2013). However, over the course of evolution, viral pathogens have evolved silencing suppressors that counter-attack the plants defence mechanism, allowing the establishment of virus infection and replication (Karthikeyan et al., 2013). Various plant pathogens have evolved different silencing suppressors, which can either intervene at the transcriptional level by interrupting DNA methylation or at the post-transcriptional level by suppressing the siRNA pathway (Kasschau & Carrington, 1998). These silencing suppressors effectively restrain RNA silencing by interfering with various steps in the silencing pathway. Whether or not a plant can be a host for a virus depends among other things on the virus's ability to suppress the siRNA pathway and enabling the virus to replicate.

3.1.2 Potyvirus silencing suppressors

To overcome RNA silencing, potyviruses have been shown to encode viral suppressors of RNA silencing (Gammelgård, Mohan, & Valkonen, 2007). Different viruses have evolved different mechanisms targeting different steps in the multistep process

(Anandalakshmi et al., 2000; Nakahara et al., 2012; Sahana et al., 2012). In a recent study, Rajamäki et al. (2014) found that the VPg protein of PVA had silencing suppression activity by interacting with the SGS3 protein of *Solanum tuberosum* and *Arbidopsis thaliana* thereby debilitating RNAi. In all other cases, however, the consensus is that the potyvirus HC-Pro is responsible for the suppression of RNAi in plants, allowing virus infection and virus accumulation in its host (Gammelgård et al., 2007; Stenger, Hein, Gildow, Horken, & French, 2005). In addition, numerous studies have also identified the P1 protein, just upstream of HC-Pro, as enhancing the VSR ability of HC-Pro (Gammelgård et al., 2007; Kasschau & Carrington, 1998; Tena Fernández et al., 2013; Adrian Valli et al., 2014).

3.1.3 HC-Pro activity in silencing suppression

As described in Section 1.5.2.3, a number of studies have revealed that HC-Pro may counter-attack RNAi in a number of different ways and may behave differently under different circumstances, an event that is still poorly understood. TEV HC-Pro for example has been shown to interact with a regulator of gene silencing calmodulin protein (rgs-CaM), a protein that appears to suppress the RNAi mechanism in *N. benthamiana* (Anandalakshmi et al., 2000). This study found rgs-CAM has endogenous suppression activity and suggested that when bound to HC-Pro, rgs-CaM-like proteins assist the VSR activity of TEV HC-Pro. In contrast, Nakahara et al. (2012) found that rgs-CaM counter-attacked TuMV and clover yellow vein virus (ClYVV) infection when over expressed in *N. tabacum*. These findings suggest that rgs-CaM may not only support the VSR activity of the potyvirus, but when overexpressed, can also inhibit virus infection by assisting RNAi.

Other VSR activities of HC-Pro reported in two independent studies found that PVY and LMV are able to hijack the ubiquitin/26S proteasome system (UPS) in *A. thaliana*. The UPS system targets misfolded/damaged proteins or functional proteins that carry destruction signals preparing them for degradation (Dielen et al., 2010; Jin et al., 2007). The UPS system also assists RNAi as a defence mechanism against viral pathogens by targeting RNA molecules for destruction by its 20S RNase activity (Dielen et al., 2010). Jin et al. (2007) showed HC-Pro of PVY and LMV were able to interact with the α and β subunits of the UPS system in *A. thaliana* preventing the destruction of viral mRNA/viral siRNA by UPS. Rather than targeting viral mRNA and siRNA, the

HC-Pro-UPS system interaction indirectly manipulated the UPS system to degrade Argonaute proteins instead, thus disturbing the RNAi pathway (Sahana et al., 2012). In addition, papaya ringspot virus (PRSV) was found to interact with the Arabidopsis 20S proteasome homologue of the α1 subunit but not the α5 as previously observed in *A. thaliana* (Sahana et al., 2012). This indicates that the HC-Pro-UPS system interaction may vary depending on the virus, virus strain, or host.

HC-Pro of the potyvirus zucchini yellow mosaic virus (ZYMV) was shown to interfere with the host HEN1 activity (Jamous et al., 2011). HEN1 is a plant methyl transferase that methylates the ribose at the 3' end of the siRNA. By interfering with HEN1 activity, ZYMV HC-Pro exposes siRNAs to nucleolytic activity and oligouridination so that they will not get incorporated into RISC for RNA degradation (Jamous et al., 2011). A similar study of ZYMV indicated the importance of the highly conserved FRNK motif in HC-Pro not only for efficient aphid transmission but also in RNA binding. This RNA binding ability had a significant correlation with high accumulation of siRNA and miRNA duplexes suggesting that the FRNK motif could possibly be a point of contact with the RISC complex (Shiboleth et al., 2007).

In a previous study, researchers found that like TuMV, TEV HC-Pro inhibits accumulation of siRNAs in the siRNA pathway (Mallory et al., 2001). These results demonstrate that silencing suppression in this particular case appears to operate downstream of transgene methylation but at a mobile signal step preceding accumulation of siRNAs (Mallory et al., 2001; Plisson et al., 2003). The different HC-Pro activities identified show a complex network of interactions that vary between virus and host; however, it seems well established that HC-Pro is primarily responsible for VSR activity in potyviruses.

3.1.4 Silencing suppression and host range determination

A number of potyviral proteins have been identified as host determining factors including CP, VPg, P3, 6K1, CI, P1 and HC-Pro (Jayathilake, 2004; Kasschau et al., 2003; Sáenz et al., 2002; Shan et al., 2015). A study carried out by Kasschau and Carrington (1998) highlighted how important a functional HC-Pro is for the initial infection by TEV. When a mutation was introduced into the core domain of HC-Pro, TEV was unable to infect *N. tabacum*. Plants recognized the aberrant viral RNA and subsequently

degraded it with no silencing suppression. When TEV expressed a functional HC-Pro, silencing suppression was enabled and replication occurred after inoculation. Another study carried out by Li et al. (2014) showed that up to three aa differences in the HC-Pro of different SMV isolates corresponded to differences in VSR activity in soybean. Although VanMV-CI and DsMV are very similar in HC-Pro sequence, slight differences may have some impact on their VSR activity.

P1 has been shown to enhance VSR activity of HC-Pro by improving its suppressor activity (Gammelgård et al., 2007; Kasschau & Carrington, 1998; Kasschau et al., 2003). By removing the N-terminal P1 region or the entire P1 of PVA and TEV, infection in *N. benthamiana* was reduced by 85-98% indicating the importance of P1 in genome amplification and virus infectivity (Rajamäki et al., 2005). Successful separation of P1 from HC-Pro is not only required for virus infectivity, but is also required for enhancing VSR activity of HC-Pro (Verchot & Carrington, 1995a). A number of studies have indicated that when a number of protease-debilitating mutations were introduced into TEV P1, separation of P1 from HC-Pro was hampered, VSR activity was disrupted and the virus was rendered less viable resulting in slow infection (Pasin et al., 2014; Verchot & Carrington, 1995a). When a modified cleavage site recognized by NIa protein was introduced in place of the P1/HC-Pro cleavage site, TEV infection of *N. tabacum* was restored (Verchot & Carrington, 1995a). A similar study found that PPV constructs containing cucumber vein yellowing virus (CVYV) P1a (Ipomovirus equivalent of the potyviral P1 protein) and PPV HC-Pro, interfered with PPV VSR activity in *N. benthamiana*, an alternative host (Shan et al., 2015). When a cleavage site recognized by NIa was introduced in place of the CVYV P1a/PPV HC-Pro cleavage site, VSR activity was restored and results were similar to the PPV P1/HC-Pro construct. However, when the CVYV P1a/PPV HC-Pro construct was agro-infiltrated into *Cucumis sativus*, a natural host of CVYV, cleavage of P1a and HC-Pro was successful with VSR activity similar to PPV constructs containing CVYV P1a and P1b (Ipomovirus equivalent to potyvirus P1 and HC-Pro respectively). This study suggests that the separation of P1 from HC-Pro is required for successful silencing suppression and that the proteolytic activity of P1 responsible for the separation is affected by an unknown host factor. This host-determining proteolytic activity of P1 is further supported by results reported by Salvador et al. (2008) in which PPV and tobacco vein mottling virus (TVMV)

constructs containing the N-terminal region of PPV P1 were able to infect plants from the *Prunus* genera, a natural host of PPV, and *Nicotiana* spp, an alternative host for both TVMV and PPV. However, constructs containing the N-terminal region of TVMV P1 were only able to infect *Nicotiana* spp but were unable to establish infection in *Prunus* spp, a natural host for PPV only.

Based on the molecular analysis carried out in Chapter two Section 2.3.4, the most noticeable difference between VanMV and DsMV lies in the P1 gene, in particular its N-terminal region and the cleavage site between itself and HC-Pro. When considering this analysis and various host-determining factors discussed by previous studies, the P1-HC-Pro region has become a region of interest for comparing VanMV-CI and DsMV. The inability of VanMV to infect aroids (hosts of DsMV) could therefore be due to an inefficient silencing suppressor of RNAi, inefficient proteolytic cleavage between P1 and HC-Pro by P1 or an incompatible P1 with an unknown host factor.

The agrobacterium-mediated transient gene expression assay is a rapid and effective system for analysing local VSR activity of viral silencing suppressors without the need to produce transgenic plants (Wydro et al., 2006). The transient gene expression system has a number of advantages over stable gene expression systems namely, its simplicity and easy performance (Wydro et al., 2006). This assay is achieved by infiltrating plant leaves with *Agrobacterium tumefaciens* that is carrying a plasmid containing a gene of interest or a reporter gene such as that for green fluorescent protein (GFP). Where viral silencing suppressions is being studied, the gene of interest is the candidate VSR gene while *gfp* gene acts as a reporter for its suppression activity. Agrobacterium is used to express these co-infiltrated genes in target plant cells. The expression of GFP will be seen by its fluorescence, which will decline over a few days due to gene silencing by RNAi triggered in the plant. If the candidate VSR has silencing suppression activity, then GFP levels do not decrease (Wydro et al., 2006; Zhang et al., 2008; Ma et al., 2009). If silencing suppression occurs, the siRNA pathway will fail to degrade both transcripts (containing the silencing suppressor and GFP) thereby allowing GFP to fluoresce under UV light. However, if silencing suppression does not occur, RNAi may degrade trancripts for both the candidate VSR and GFP. This will be detected by a decrease in mRNA levels for the candidate VSR and GFP and/or by a decrease in fluorescence over a few days (Wydro et al., 2006; Zhang et al., 2008; Ma et

al., 2009). The intensity of GFP fluorescence will correspond to the efficiency in VSR activity of the viral protein (Wydro et al., 2006).

3.2 Aims of study

The aim of this study is to compare and contrast the silencing suppressor activity of VanMV and DsMV HC-Pro. The study described in Chapter 2 identified an indel within the P1 region of VanMV and DsMV. The aim of this part of the study was to examine whether this difference corresponds to an observable difference in local silencing suppression by HC-Pro.

3.3 Methods and Materials

3.3.1 Total RNA extraction

Total RNA was extracted from VanMV-CI and DsMV-NZ leaf samples taken from green house storage at the University of Auckland following the protocol outlined in section 2.2.2.

3.3.2 Primer design

Primers were designed using the VanMV-CI sequence described in section 2.2.3 and a DsMV-NZ isolate identified in an earlier study (Pearson et al., 1998). The primer design targeted the P1 start codon with a modified stop codon at the end of the HC-Pro gene so that a continuous P1-HC-Pro PCR product could be obtained. This was carried out for both VanMV-CI and DsMV-NZ isolates; PKP2 for VanMV and PKP4 for DsMV. Primers were also designed with modified start and stop codons for the HC-Pro gene of both VanMV-CI (PKP1) and DsMV-NZ (PKP3). Fusion primers were also constructed in order to create an overlapping fragment consisting of P1 from one strain and HC-Pro from the other (PKP5 and PKP6). The fusion primers we designed in a way in which the annealing temperature of each overlapping amplicon would be similar to the annealing temperature of the flanking primers used in the overlapping extension step to amplify the overall product. Table 3.1 summarises the constructs and primers designed with the expected product size.

Table 3.1: Primer sequences used to generate fusion templates and constructs for VanMV-CI and DsMV-NZ. Forward and reverse primers were designed to amplify constructs PKP1-6. Expected product sizes are indicated in bp (base pairs).

Construct name	Source of P1	Source of HC-Pro	Forward primer	Reverse primer	Product size (bp)
PKP1	-	VanMV-CI	5'-ACTAAGACCCCAGAGGAACAATTTTTCC-3'	5'-TCAGCCAACTCGATAGAATTTCATTTCGC-3'	1374
PKP2	VanMV-CI	VanMV-CI	5'-ATGGCTTGCATTGTTTTCGGAAGCTTTTC-3'	5'-TCAGCCAACTCGATAGAATTTCATTTCGC-3'	2439
PKP3	-	DsMV-NZ	5'-GCAAAGACTCAAGAAGCGCAATTCTT-3'	5'-CTTCTGGAGTCTTTGTGTACTGGTGC-3'	1374
PKP4	DsMV-NZ	DsMV-NZ	5'-ATGGCTTGTCTCGTTTTCGGTAACTTC-3'	5'-CTTCTGGAGTCTTTGTGTACTGGTGC-3'	2529
PKP5	VanMV-CI	DsMV-NZ	5'-GCAAAGACTCCAGAAGCGCAATTCTT-3'	5'-TCAACCAACCCTATAAAACTTCATCTCACTT-3'	2439
PKP6	DsMV-NZ	VanMV-CI	5'-ATGGCTTGTCTCGTTTTCGGTAACTTC-3'	5'-TCAGCCAACTCGATAGAATTTCATTTCGC-3'	2529
VanMV P1 fusion template	VanMV-CI	-	5'-ATGGCTTGCATTGTTTTCGGAAGCTTTTC-3'	5'-CTTCTGGAGTCTTTGTGTACTGGTGC-3'	1075
VanMV HC-Pro fusion template	-	VanMV-CI	5'-TTGTCAATATGCTAAGACCCCAGAGG -3'	5'-TCAGCCAACTCGATAGAATTTCATTTCGC-3'	1384
DsMV P1 fusion template	DsMV -NZ	-	5'-ATGGCTTGTCTCGTTTTCGGTAACTTC-3'	5'-CCTCTGGGGTCTTAGCATATTGACAA-3'	1166
DsMV HC-Pro fusion template	-	DsMV-NZ	5'-GCACCAGTACACAAAGACTCCAGAAG-3'	5'-CTTCTGGAGTCTTTGTGTACTGGTGC-3'	1384

Figure 3.1 shows the primers used to amplify fusion templates for downstream overlapping extension PCR. The arrow colour corresponds to the region of interest to be amplified.

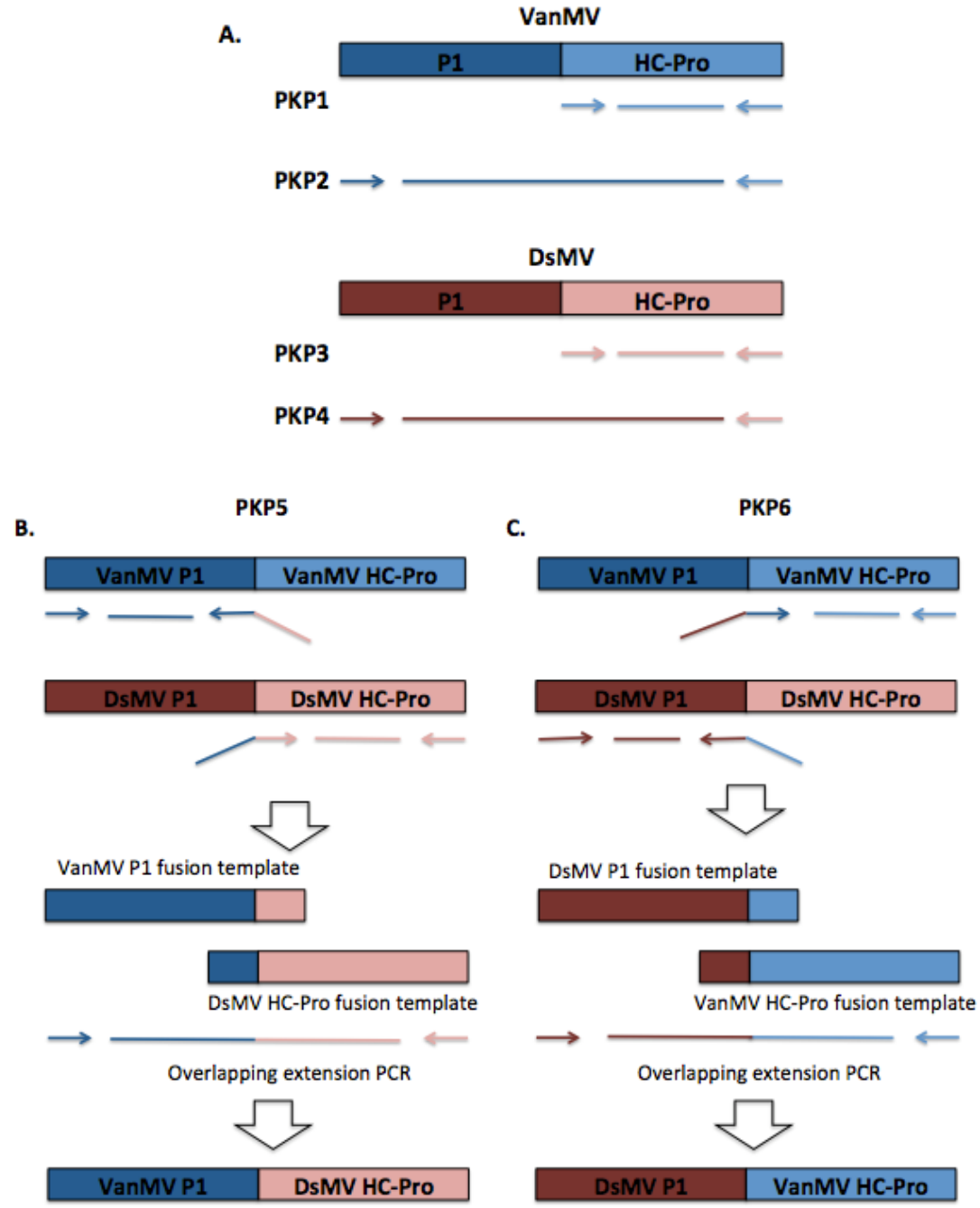


Figure 3.1: Diagrammatic representation of the strategy used to generate hybrid constructs between the P1 and HC-Pro genes of DsMV-NZ and VanMV-CI. Arrow colour corresponds to the region of interest to be amplified. **A.** Constructs PKP1-PKP4. **B.** Constructs PKP5 and **C.** PKP6

3.3.3 One-Step reverse transcriptase-polymerase chain reaction (RT-PCR)

In order to amplify the regions of interest, a One-step RT-PCR was carried out with the gene-specific primers listed in Table 3.1 using SuperScript® III One-Step RT-PCR with Platinum® *Taq* (Invitrogen, Carlsbad, CA, USA). The amount of RNA used in each full reaction (50 µl) varied from 500-700 ng with 2 µl each of 10 µM forward and 10 µM reverse primer. PCR conditions used were as follows: cDNA synthesis at 55°C for 30 minutes, Initial denaturation at 95°C for 2 minutes, 40 cycles of denaturation at 95°C for 15 seconds, annealing at 53-56°C for 30 seconds, 68°C for 1 minute per kilobase and final extension at 68°C for 5 minutes with a final hold at 4°C. Each completed PCR reaction was stored at -20°C. The PCR reactions were carried out using a Techne PCR machine (Straffordshire UK, model FTGRAD2D).

3.3.4 Gel purification of PCR products

PCR products were electrophoresed through 1% agarose in 1 X TBE. All PCR products were gel purified using Zymoclean™ Gel DNA Recovery Kit (Irvine, USA) to obtain products free of primer dimers and non-specific products. The expected PCR product was excised from the agarose gel and purified according to the manufacturer's protocol. The gel slice was dissolved in three volumes of ADB solution by incubating at 55°C for 5-10 minutes. The melted agarose solution was then transferred to a Zymo-spin column, which was then centrifuged at maximum speed (14,000rpm) (Eppendorf, Hamburg, Germany) for 60 seconds, discarding the flow through. DNA washing buffer (200 µl) was then added and further centrifuged at maximum speed for 60 seconds discarding the flow through. DNA was then eluted in 10 µl of elution solution and the concentration/purity measured using a Nanovue ND1000 spectrophotometer (GE Healthcare). A portion of the gel purified PCR product was sent for SANGER sequencing with both forward and reverse primers to confirm the sequences before using for downstream applications. Sequencing was carried out by the Waikato DNA sequencing facility, Waikato University, New Zealand.

3.3.5 Overlapping extension PCR

An overlapping extension PCR was carried out to create hybrid constructs of VanMV-CI and DsMV-NZ using fusion templates as illustrated in the final stages of Figure 3.2.

This PCR reaction was carried out using a Platinum™ *Taq* DNA polymerase kit (Invitrogen, Carlsbad, CA, USA). Equal amounts of gel purified template were used for each hybrid construct (200-500 ng) and PCR was set up following the manufacturer's protocol with 2 µl of each 10 µM forward and 10 µM reverse primer. PCR conditions were as follows: Initial denaturation at 95°C for 2 minutes, 35 cycles of denaturation at 95°C for 30 seconds, annealing at 56°C for 1 minute, 68°C for 1 minute per kilobase and final extension at 68°C for 5 minutes with a final hold at 4°C. Hybrid amplicons were stored at -20°C.

3.3.6 Generation of Gateway Entry Vector

3.3.6.1 Ligation reaction

A ligation reaction was carried out for each PCR product with the pCR8 entry vector using the pCR8™/GW/TOPO TA cloning kit (Life Technologies, Carlsbad, California) following the manufacturer's protocol. For each ligation reaction, the following reagents were added: 1 µl of salt solution, 1 µl of TOPO vector and 10-26 ng of gel purified PCR product and made up to a total volume of 6 µl with nuclease free water. The reaction mix was incubated at room temperature for 30 minutes before proceeding to the transformation step.

3.3.6.2 Transformation of entry vector

The entry constructs were transformed into *E. coli* to obtain multiple copies of the desired PCR-containing plasmid. The transformation was carried out by the addition of 1.25 µl of the ligation reaction to 25 µl of One Shot® Chemically Competent *E. coli* cells (Life Technologies, Carlsbad, California). The transformation mixture was incubated on ice for 10 minutes followed by a heat shock step at 42°C for 30 seconds without shaking. Tubes were then immediately transferred to ice and 250 µl of SOC medium was added and shaken at 200 rpm at 37°C for 1 hour. Once the incubation step was complete, 50 and 100 µl of each transformation reaction was spread on LB agar plates containing 100 mg/ml of spectinomycin and incubated overnight at 37°C. Colonies were taken from each plate and streaked on to a master plate, which was stored at 4°C for downstream protocols.

3.3.7 Colony PCR and gel electrophoresis

A colony PCR was then carried out to determine successful ligation and transformation of the entry vector. One colony was taken from the transformed *E. coli* cells and submerged into 10 µl of nuclease free water in preparation for a colony PCR. Each diluted colony was streaked out onto a master plate of LB agar with 100 mg/ml spectinomycin and incubated overnight at 37°C. A colony PCR was carried out with the gene-specific primers listed in Table 3.1 using GoTaq® green master mix (Promega corporation, USA). The PCR reaction used 1 µl of submerged *E. coli* cells with 1 µl each of 10 µM forward and 10 µM reverse primer. PCR conditions used were as follows; 95°C (2 minutes) and 35 cycles of 95°C (30 seconds), 53-56°C (30 seconds), 72°C (1 minute per kilobase), final extension at 72°C (5 minutes) and final hold at 4°C. The samples were then analysed on a 1% agarose gel at 70 volts for 60 minutes in 1 X TBE buffer.

3.3.8 Plasmid DNA extraction

Colonies grown on the master plate described in Section 3.3.7 were used to inoculate 5 ml of LB liquid medium containing 100 mg/ml of spectinomycin and grown at 37°C with shaking overnight. This was followed by a plasmid DNA extraction using GenElute™ Plasmid Miniprep Kit (Sigma Aldrich) following the manufacturer's protocol. The whole overnight recombinant *E. coli* culture was centrifuged at maximum speed (14,000 rpm) for 1 minute, discarding the supernatant. The pellet was resuspended in 200 µl of resuspension solution (containing RNase A solution) followed by 200 µl of lysis solution with the lysis step not exceeding 5 minutes. A further 350 µl of neutralisation/binding solution was added to precipitate cell debris, which was then centrifuged at maximum speed (14000 rpm) for 10 minutes. GenElute Miniprep Binding Column was prepared by adding 500 µl of column preparation solution then centrifuged at maximum speed (14000 rpm) for 1 minute, discarding the flow-through. Cleared lysate was then transferred to the prepared binding column and centrifuged at maximum speed for 1 minute, discarding the flow-through. The column was washed with 750µl of diluted wash solution by centrifuging at maximum speed for 1 minute, discarding the flow-through. The column was further centrifuged for 2 minutes to remove excess ethanol. DNA was then eluted in 100 µl of elution

solution and centrifuged at maximum speed for 1 minute. The concentration and purity of the plasmid was measured using a Nanovue ND1000 spectrophotometer.

3.3.9 Gateway LR recombination reaction

An LR recombination reaction was set up in order to transfer the gene of interest from the entry vector construct in pCR8 into the destination binary vector pHEX2 (Invitrogen Corp; Fig. 3.2) following the manufacturer's protocol (Life Technologies, CA, USA).

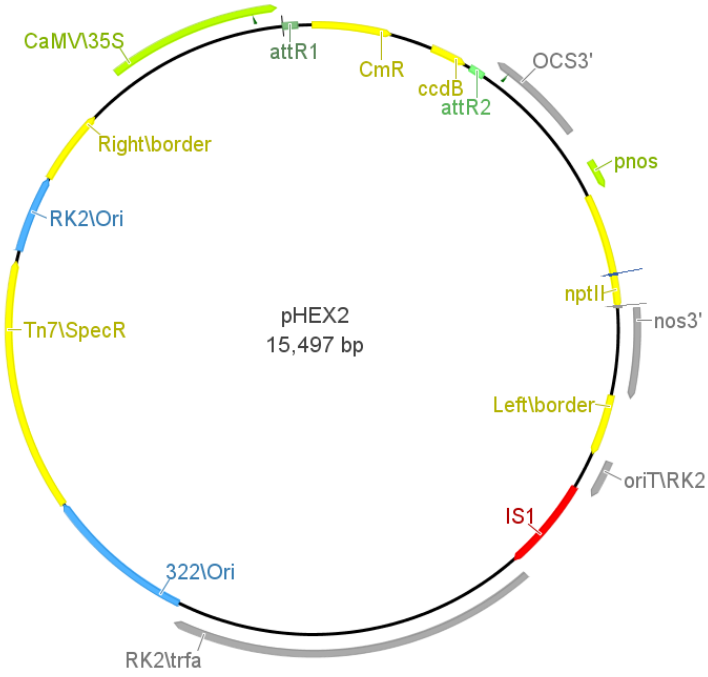


Figure 3.2 Plasmid pHEX2 Gateway vector used to harbour gene of interest for agrobacterium transformation.

The recombination reaction mixture was set up as follows: 100-300 ng of recombinant pCR8, 300 ng of pHEX2 destination vector, 4 µl of 5X LR clonase reaction buffer and the reaction made up to a total volume of 16 µl with TE buffer at pH 8. Each recombination reaction was incubated at 25°C for 1 hour followed by the addition of 2 µl of proteinase K solution and incubation at 37°C for 10 minutes. A transformation reaction was then carried out following the protocol described in Section 3.3.6.2, followed by colony PCR/gel electrophoresis (using gene specific primers described in Table 3.1). Expression clones containing the gene of interest were further grown in LB liquid medium containing 50 mg/ml of spectinomycin with shaking at 37°C overnight. Plasmid purification was then carried out following the protocol outlined in Section

3.3.8. The concentration and purity of the plasmid was measured using a Nano drop Nano vue ND1000 spectrophotometer.

3.3.10 Making competent Agrobacterium tumefaciens cells

LB broth (6 ml) containing 50 mg/ml each of rifampicin and gentamycin were inoculated with 15 µl of *A. tumefaciens* GV3101 from previously prepared glycerol stocks and incubated overnight with shaking at 28°C. A further 50 ml of LB broth containing 50 mg/ml each of rifampicin and gentamycin was inoculated with 5 ml of the overnight culture and incubated at 28°C for 7-8 hours to an OD_{600} of 0.6-0.7. The culture was then chilled on ice for 5 minutes and cells were centrifuged at 4000 rpm for 5 minutes followed by resuspension of the pellet in 1 ml 20 mM $CaCl_2$. Aliquots of 100 µl were made and kept at -80°C.

3.3.11 Transformation of *Agrobacterium tumefaciens*

Empty pHEX2 plasmid and pHEX2 containing gene of interest or GFP was used for the transformation. Purified pART27 plasmid (Gleave, 1992) containing the silencing suppressor P19 was used as a positive control. Purified plasmid DNA (500 ng – 1 pg) was added to 25-50 µl of competent *A. tumefaciens* followed by freezing in liquid nitrogen for 5 minutes. Frozen cells were then incubated at 37°C for 5 minutes followed by the addition of 800 µl of LB broth with no antibiotics. Cells were incubated with shaking at 28°C for 2-3 hours then centrifuged at 5000 rpm for 40 seconds. About 700 µl of LB medium was then poured out and the pellet suspended in the remaining LB medium. The cells were then plated on LB plates containing 50 mg/ml each of rifampicin and gentamycin. Plates were then incubated at 28°C for 2-3 days.

3.4 Results

3.4.1 Total RNA extraction

In order to isolate regions of interest from both VanMV-CI and DsMV-NZ, total RNA was extracted from infected plant material. RNA extraction of VanMV-CI and DsMV-NZ provided sufficient RNA quality and quantity as indicated in Figure 3.3. Bright bands containing ribosomal RNA and 5S RNA in lanes 2 and 3 show the RNA was of sufficient quality and concentration to use. Smears in the background is likely due to overloading of the RNA on the gel

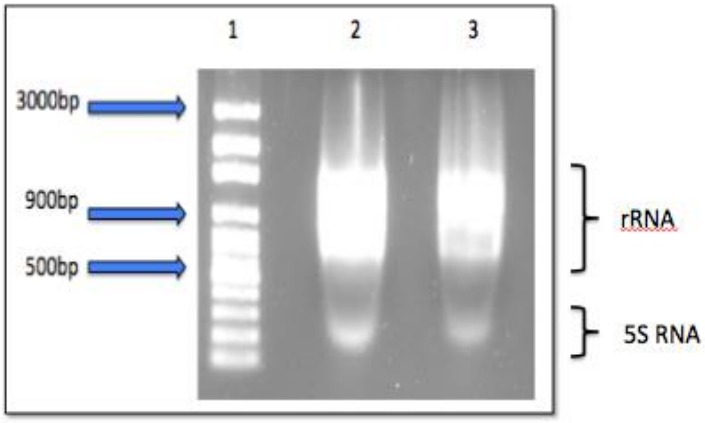


Figure 3.3: Total RNA extracted from DsMV-infected *C. esculenta* in lane 2, VanMV-CI-infected *V. tahitensis* in lane 3 and 100 bp ladder in lane 1.

3.4.2 Amplification of P1/HC-Pro region

Both the HC-Pro (PKP1 and PKP3) and P1-HC-Pro (PKP2 and PKP4) regions from VanMV and DsMV were amplified as the first step to expressing the proteins *in planta*. These sequences were amplified to ensure each coding region began with a start codon and finished in a stop codon. Figure 3.4 shows the expected DsMV-NZ HC-Pro PCR product of 1374 bp (PKP3, lane 2) and DsMV-NZ P1-HC-Pro PCR product of 2529 bp (PKP4, lane 3).

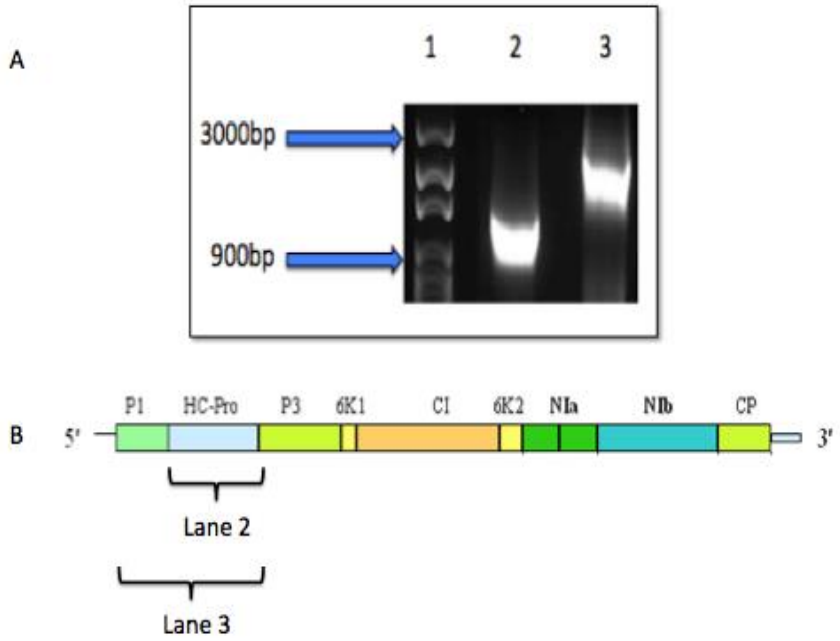


Figure 3.4: A. Agarose gel electrophoresis of the expected 1374 bp from DsMV-NZ HC-Pro (lane 2) and 2529bp from DsMV-NZ P1-HC-Pro (lane 3). Lane 1 shows the 100 bp ladder. **B.** Region of the DsMV-NZ genome amplified.

Figure 3.5 shows the expected VanMV-CI HC-Pro PCR product of 1374 bp (PKP1, lane 2) and VanMV-CI P1-HC-Pro PCR product of 2439 bp (PKP3, lane 3).

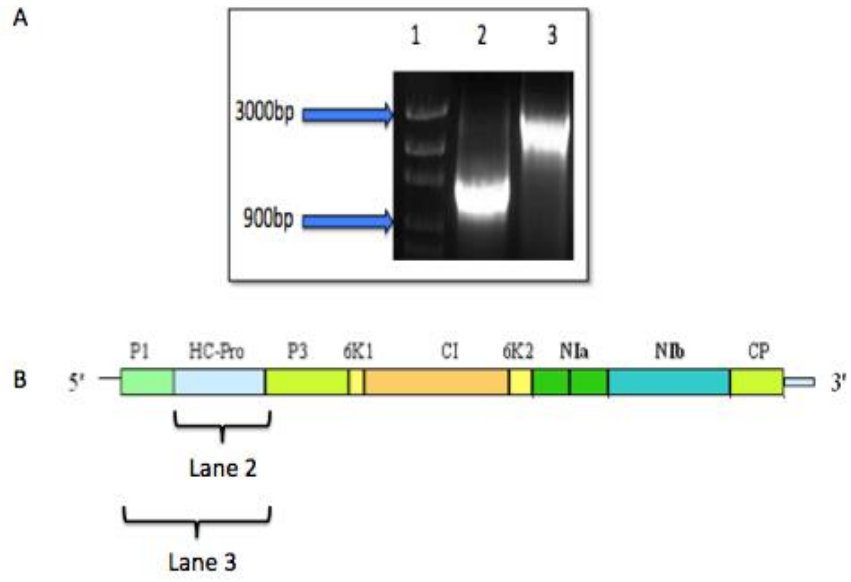


Figure 3.5: A. Agarose gel electrophoresis of the expected 1374 bp from VanMV-CI HC-Pro (PKP1, lane 2) and 2439 bp VanMV-CI P1-HC-Pro (PKP3, lane 3). Lane 1 shows the 100 bp ladder. **B.** Region of the VanMV-CI genome amplified

Figure 3.6 shows the expected PCR product for the VanMV-CI P1 fusion template of 1075 bp (lane 2) and the VanMV-CI HC-Pro fusion template of 1384 bp (lane 3).

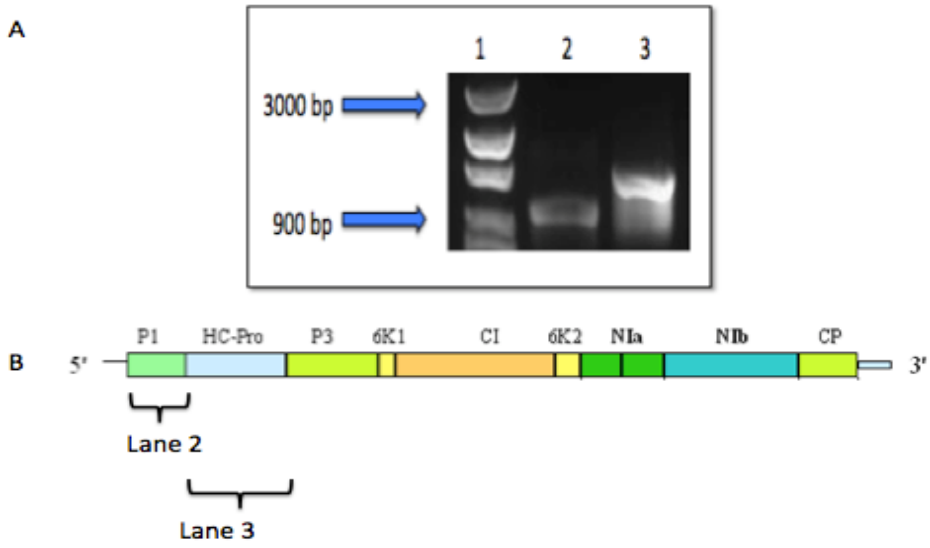


Figure 3.6: A. Agarose gel electrophoresis of the expected 1076 bp of the VanMV-CI P1 fusion template (lane 2) and 1384 bp from VanMV-CI the HC-Pro fusion template (lane 3). Lane 1 shows the 100 bp ladder. **B.** Region of the VanMV-CI genome amplified

Figure 3.7 shows the expected PCR product for DsMV-NZ P1 fusion template of 1166 bp (lane 2) and the DsMV-NZ HC-Pro fusion template of 1384 bp (lane 3).

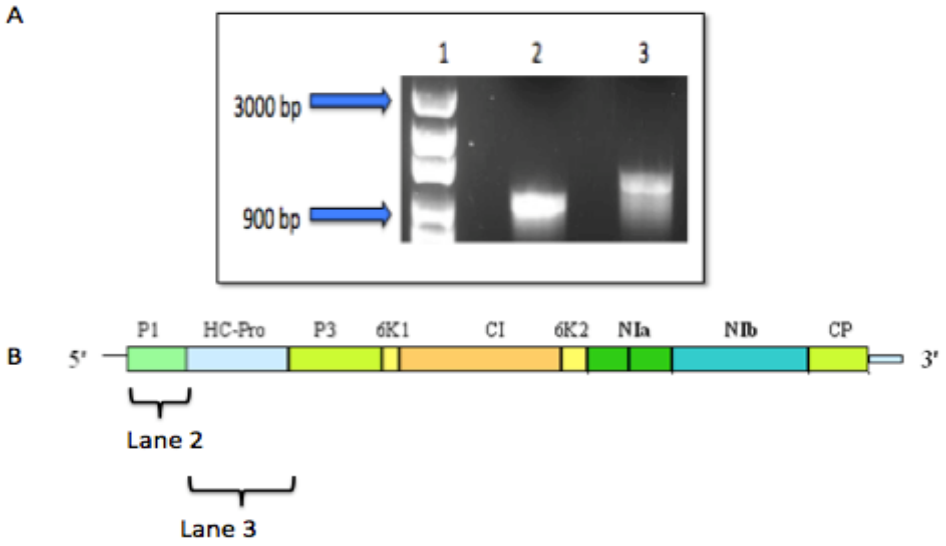


Figure 3.7: A. Agarose gel electrophoresis of the expected 1166 bp of the DsMV-NZ P1 fusion template (lane 2) and 1384 bp from DsMV-NZ HC-Pro fusion template (lane 3). Lane 1 shows the 100 bp ladder. **B.** Region of the DsMV-NZ genome amplified.

3.4.3 Overlap extension PCR

Overlap extension PCR was carried out to create hybrid constructs between the P1 of DsMV-NZ and the HC-Pro of VanMV-CI and vice versa. The overlapping PCR step was carried out using a *Taq* DNA polymerase PCR kit to obtain the PKP5 (VanMV P1/ DsMV HC-Pro) and PKP6 (DsMV P1/ VanMV HC-Pro) constructs using the templates shown in Figure 3.6 and 3.7. The PKP6 construct was successfully amplified creating the expected 2529 bp length (Figure 3.8, lane 3), in contrast, no product was obtained for the PKP5 construct (lane 2).

A number of modifications were made to the overlap extension PCR step in an attempt to generate the PKP5 PCR product including the application of equal amounts of each amplicon (DsMV P1 and VanMV HC-Pro fusion templates) in the overlapping stage as well as varying amounts from 50 ng to 500 ng each. This was to increase the likelihood of each fusion template annealing to its complementary pair. The ratio of $MgCl_2$ to *Taq* DNA polymerase was implemented according to the manufacturer's protocol to optimise PCR conditions. This was to determine the optimal conditions of $MgCl_2$ as a catalyser of *Taq* DNA polymerase but at a suitable ratio to allow for specificity. Further modifications to the procedure were made including a gradient PCR to optimise the annealing temperature for the overlapping region of the fusion templates and for the flanking primers to anneal; however, this also proved unsuccessful. Finally, the overlapping extension PCR was separated into two phases. Phase 1 consisted of all the PCR reagents including templates of equal amounts without the flanking primers. This was to allow for the overlapping regions in the fusion products to anneal to each other in 5-10 cycles, thereby acting as template before the addition of the flanking primers in phase two. This was followed by a phase two PCR reaction with the addition of the flanking primers for a further 25 cycles. This, however, was also unsuccessful.

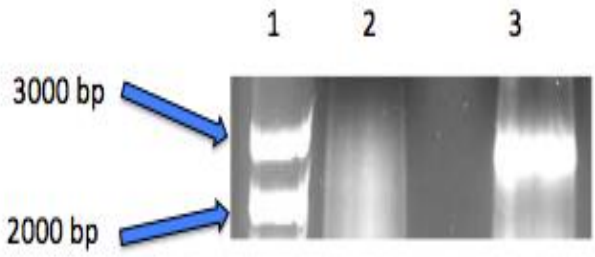


Figure 3.8: Agarose gel electrophoresis of the expected 2529 bp PKP6 construct (lane 3). No product was observed for the PKP5 construct (lane 2). Lane 1 shows the 100 bp ladder.

3.4.4 pCR™8/GW/TOPO® TA cloning (Entry vector)

The pCR™8/GW/TOPO® entry vector contains a single 3'-thymidine (T) overhang (Figure 3.9 A-2) to complement the A overhang (Figure 3.9 A-1) produced by the *Taq* DNA polymerase in the PCR reaction. It also contains Topoisomerase I covalently bound to the linearised vector as indicated in Figure 3.8 B.

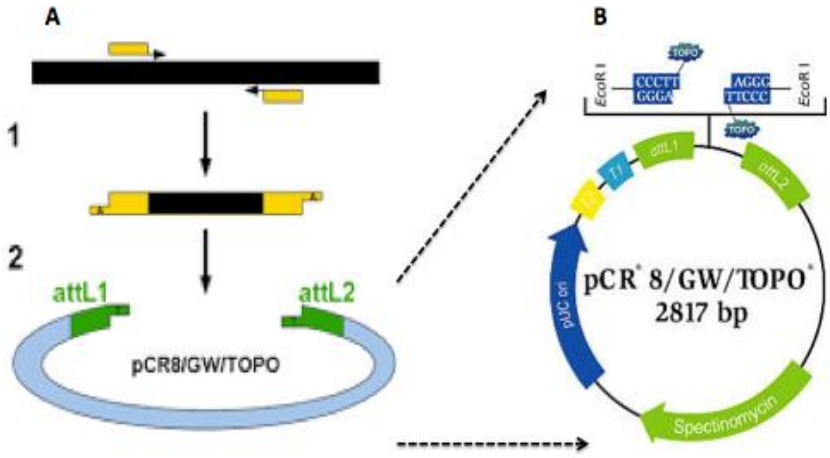


Figure 3.9: **A-1;** Amplification of the region of interest flanked by A overhangs **A-2;** PCR products containing A overhangs are ligated to the T overhangs of the pCR8 entry vector, flanked by attL1 and attL2 sites. **B.** Closer view of the 2817 bp pCR™8/GW/TOPO® entry vector. Obtained from https://www.thermofisher.com/order/catalog/product/K250020

The entry vector supplied was 2817 bp long with each end of the vector containing attL1 and attL2 sites, which enable a recombination reaction to take place between the entry vector and a destination vector of choice. The primers GW1 and GW2 have also been designed to bind within these sites enabling easy confirmation of successful

cloning by PCR. Once the ligation of PCR products into pCR8 was complete (Figure 3.9 A-2), the entry vector containing the PCR product of interest was then transformed into competent One Shot® TOP10 *E. coli* cells. The transformation reaction was then spread on LB agar containing spectinomycin. For the ligation reaction step, an insert to vector ratio was not specified in the manufacturer's protocol so the initial ligation reaction was carried out with an approximately 1:1 ratio of insert (11.25 ng) to vector (10 ng of pCR8). Colonies were observed on agar plates; however, the transformation efficiency was low at 1.2 x 10^4 cfu/μg DNA for PKP2 and PKP4. When 45-90 ng of insert was combined with 10 ng of vector, a higher transformation efficiency of 1.4 x 10^6 cfu/μg DNA was obtained for PKP2 and PKP4 and 1 x 10^7 cfu/μg DNA was obtained for the positive control, PKP1 and PKP3. This indicated the insert to vector ratio was crucial to a successful pCR8 ligation step.

A colony PCR was then carried out using the GW1 and GW2 primers with results presented in Figure 3.10. Results obtained by the colony PCR indicated successful ligation of PKP1, PKP2, PKP3 and PKP4 into the entry vector followed by successful transformation into *E. coli*. Lanes 2 and 4 indicated the expected 1474 bp product (1374 bp size of the HC-Pro region plus 100 bp of upstream nucleotides between the HC-Pro and GW1/GW2 primers) for DsMV-NZ and VanMV-CI, respectively. Lane 3 shows the expected 2629 bp product of DsMV-NZ P1-HC-Pro (2529 bp of P1-HC-Pro plus 100 bp of upstream nucleotides between the P1-HC-Pro and GW1/GW2 primers). Lane 5 shows the expected 2539 bp product of VanMV-CI P1-HC-Pro (2439 bp of P1-HC-Pro plus 100 bp of upstream nucleotides between the P1-HC-Pro and GW1/GW2 primers).

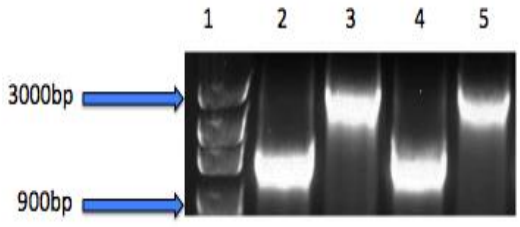


Figure 3.10: Colony PCR verifying successful ligation and transformation of PCR products into pCR8. Lane 1: 100bp marker; lane 2: DsMV-NZ HC-Pro; lane 3: DsMV-NZ P1-HC-Pro; lane 4: VanMV-CI HC-Pro; lane 5: VanMV-CI P1-HC-Pro

Using the same conditions as for the P1-HC-Pro constructs, ligation of PKP6and its subsequent transformation into *E. coli*, was unsuccessful. Changes included the amount of PKP6 PCR product added with the ratio of vector to insert similar to other positive results and increasing the incubation step from 5-30 minutes to allow for the cloning of the large fusion product. Another modification to the procedure was increasing the extension step in the overlapping extension PCR step to ensure the extension of the PCR product was complete with the addition of the 3' adenosine overhang created by *Taq* DNA polymerase, with no success.

3.4.5 Gateway cloning (Entry vector to destination vector)

The Gateway cloning procedure is based on the site-specific recombination system found in bacteriophage lambda phage. The efficiency of the system relies on DNA recombination sites (attL1 and attL2 found in the entry vector and attR1 and attR2 found in the destination vector) and the enzyme LR clonase, which mediates the recombination reaction. The diagram in Figure 3.11 provides an overview of the LR reaction in the Gateway cloning procedure. The recombination reaction takes place via an exchange between the attL1/aatL2 sites of the entry clone (pCR8) and the attR1/attR2 site of the destination binary vector (pHEX2). This recombination reaction in turn exchanges the gene of interest from the entry clone (PKP constructs described in Section 3.3.9) with the ccdB gene (control of cell death B gene) of the destination vector. The outcome of the exchange results in the formation of the expression clone containing the gene of interest (PKP construct) flanked by attB1/attB2 sites. This expression clone contains the spectinomycin resistance gene, allowing selection on LB plates containing spectinomycin. Following this LR reaction, a colony PCR was carried out with results shown in Figures 3.12 and 3.13.

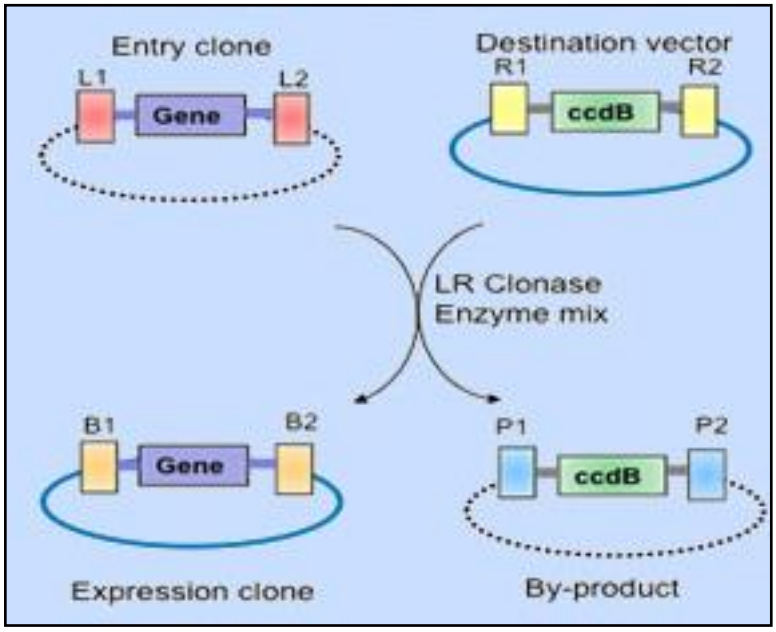


Figure 3.11: LR recombination reaction that takes place between an entry vector for example, pCR8-containing a gene of interest and a destination vector for example, pHEX2. This reaction results in the formation of an expression clone and by-products, which will not be used in downstream applications. L1, L2, R1 and R2 represent the att sites, which will undergo recombination. B1, B2, P1 and P2 are the result of the recombination reaction, being hybrids of the starting att sites. Retrieved from http://nptel.ac.in/courses/102101040/modules/module4/lec25/1.3.html

Figure 3.12 lanes 2 and 3 show the expected 1374 bp HC-Pro product of both DsMV-NZ and VanMV-CI, respectively. Figure 3.13 lanes 2 and 3 show the expected 2439 bp band for VanMV-CI P1-HC-Pro and 2529 bp for DsMV-NZ P1-HC-Pro. These results indicated a successful recombination reaction and subsequent transformation into *E. coli.*

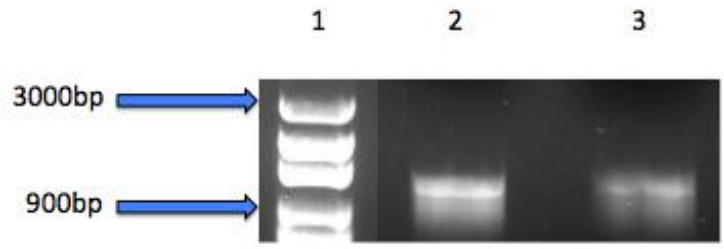


Figure 3.12: Colony PCR verifying LR recombination with pHEX2 and transformation of *E. coli*. Lane 1: 100bp marker; lane 2: DsMV-NZ HC-Pro; lane 3: VanMV-CI HC-Pro.

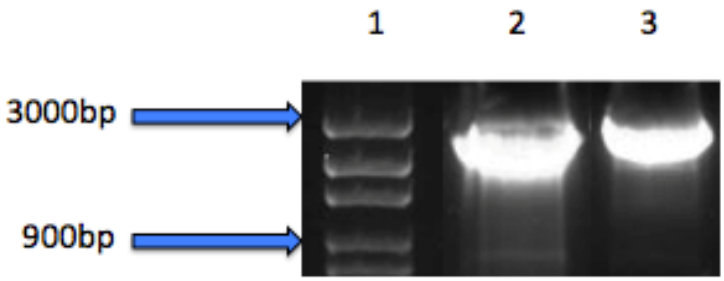


Figure 3.13: Colony PCR verifying LR recombination with pHEX2 and transformation of *E. coli*. Lane 1: 100bp marker; lane 2: VanMV-CI P1-HC-Pro; lane 3: DsMV-NZ P1-HC-Pro.

3.4.6 Transformation of *Agrobacterium tumefaciens*.

The purpose of transforming *A. tumefaciens* with the expression vectors containing the gene of interest was to allow for the expression of the viral proteins in the model plant *N. benthamiana*. Transformation of 25 µl of competent *A. tumefaciens* cells at an OD_{600} of 0.6-0.7 with 200-300 ng of purified pHEX2 containing gene of interest was unsuccessful. Modifications were made to the procedure including the transformation of 1 ml of competent cells with 1 µg of purified plasmid. Although colonies were observed, the transformation efficiency was low with 1.1 x 10^4 cfu/µg DNA. This transformation efficiency is typical of the freeze-thaw method; however, a colony PCR shows the colonies did not contain the plasmid DNA (Figure 3.14A). The colony PCR in Figure 3.14A resulted in non-specific bands similar to that of the negative control (lane 7), which were non-transformed *A. tumefaciens* cells. This suggested that the M13 forward and reverse primers used were binding to the same region as the negative control namely, *A. tumefaciens* DNA or the Helper plasmid. This PCR step was carried out side-by-side with purified pHEX2 plasmids containing PKP1-4 obtained from *E. coli* to ensure the PCR conditions were accurate and that the M13 forward and reverse primers used were specific for the pHEX2 plasmid. This however, does not discount the possibility of isolating enough gDNA that could be amplified by the M13 primers

A plasmid preparation and confirmation PCR step was carried out; however, a number of non-specific products were also observed and the expected band sizes were not obtained (Figure 3.14B). This suggests that the plasmid preparation may have also isolated the Helper plasmid present in Agrobacterium and that the M13 forward and reverse primers were amplifying this rather than the pHEX2 plasmid, if the

transformation was successful. By obtaining the desired PCR product in the positive control (using the M13 primers with previously purified pHEX2 plasmids), this suggests that the transformation step was not successful and that the colonies obtained were false positives.

In addition to constructs PKP1-4 being unable to be transformed into Agrobacterium, pHEX2 empty vector and pHEX2 containing GFP were also unsuccessful; however, the pART27 plasmid containing the silencing suppressor P19 of tomato bushy stunt virus (TBSV) was successfully transformed into Agrobacterium as indicated by the expected product in Figure 3.14A lane 4 and Figure 3.14B lane 7.

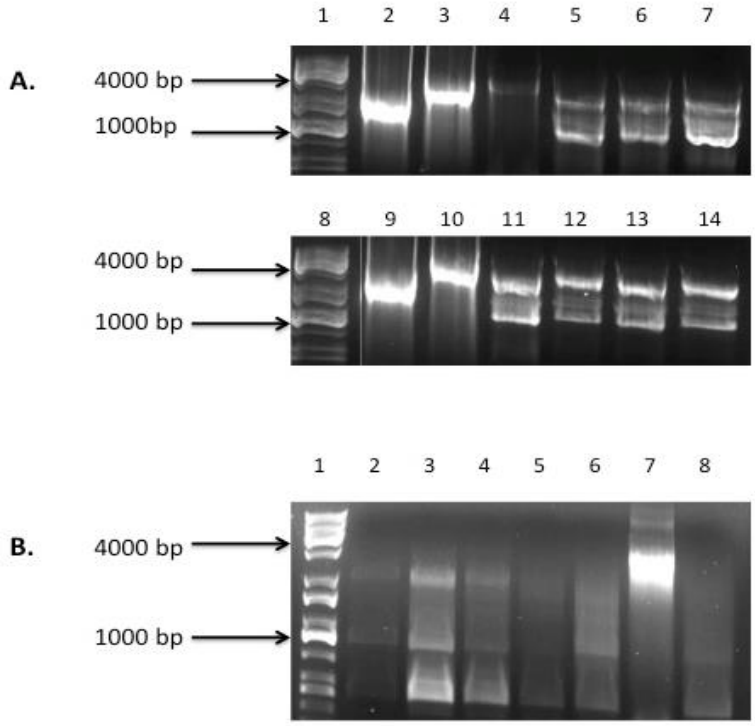


Figure 3.14 A. Colony PCR of transformed A. tumefaciens with pHEX2 plasmids containing PKP constructs. Lane 1: KAPA marker; lane 2: positive control (purified pHEX2 plasmid containing PKP1); lane 3: positive control (purified pHEX2 plasmid containing PKP2); lane 4: P19 from pART27 plasmid; lane 5: pHEX2 empty vector; lane 6: pHEX2 containing GFP; lane 7: untransformed *A. tumefaciens*; lane 8: KAPA marker; lane 9: positive control (purified pHEX2 plasmid containing PKP3); lane 10: positive control (purified pHEX2 plasmid containing PKP4); lane 11: PKP1; lane 12: PKP2; lane 13: PKP3; lane 14: PKP4. **B.** PCR on purified plasmid from Agrobacterium. Lane 1: KAPA marker; lane 2: PKP1, lane 3: PKP2, lane 4: PKP3; lane 5 PKP4; lane 6: pHEX2 empty vector; lane 7: P19 in pART27; lane 8: pHEX2 containing GFP.

Table 3.2 gives a summary of the progress made with the experiments at the time of writing. All constructs apart from PKP5 were successfully amplified. Apart from PKP5, all other constructs could be ligated to the pCR8 entry vector. Constructs PKP1-4 were also successfully transferred to pHEX2 by a recombination reaction and subsequently transformed into One Shot TOP10 *E. coli* cells. These constructs however, were unable to be transformed into *A. tumefaciens* strain GV3103 apart from pART27-P19 as indicated in Figure 3.14.

Table 3.2: Summary of progress made with constructs at each step from amplification of PCR products to transformation with *A. tumefaciens*. pHEX2 empty vector, pART27-P19 and pHEX2-GFP purified plasmids were provided so was only involved in Agrobacterium transformation.

	Amplified	Ligation into pCR8	Recombination transfer to pHEX2	Transformed into *E. coli*	Transformed into Agrobacterium
PKP1	✓	✓	✓	✓	X
PKP2	✓	✓	✓	✓	X
PKP3	✓	✓	✓	✓	X
PKP4	✓	✓	✓	✓	X
PKP5	X	X	X	X	X
PKP6	✓	X	X	X	X
pHEX2 empty vector	-	-	-	-	X
pART27-P19	-	-	-	-	✓
pHEX2-GFP	-	-	-	-	X

3.5 Discussion

The seriousness of virus infection on vanilla plant species was recognised when vanilla production in the South Pacific was severely hindered in 1981 (Wisler & Zettler, 1987). Since its discovery in 1986, little genetic information has been known about VanMV in relation to DsMV. Apart from having similar serological properties, VanMV-CI was classified as a strain of DsMV based on a partial genome sequence (Farreyrol et al., 2006). The inability of VanMV to infect natural hosts of DsMV was apparent when VanMV from FP and the CI failed to infect a number of aroid plants, susceptible to DsMV infection (Farreyrol, 2005; Wang & Pearson, 1992; Wisler & Zettler, 1987). Initially discovered in French Polynesia, VanMV was later identified in the Cook Islands, most probably spread by the propagation of infected plant material (Wisler & Zettler, 1987). Little information was known about the molecular basis for the difference in host range between VanMV and DsMV since the discovery of VanMV.

Studies have been conducted on a number of potyviruses in order to better understand their host range determinants. These studies have suggested a number of different potyviral genes for different potyviruses may be involved (Jayathilake, 2004; Kasschau et al., 2003; Schaad, Lellis, & Carrington, 1997; Shan et al., 2015) suggesting the host determining factor may be dependent on either the virus strain, host or both. In order to provide insight into the host-determining factor of VanMV-CI, the 5' region of the genome was of particular interest. The N-terminus of the P1 gene of PPV, PVA and TEV virus was recently shown to encode a host-determining factor (Pasin et al., 2014; Shan et al., 2015; Verchot & Carrington, 1995a). The P1 of VanMV has an indel within its N-terminus, which may alter its activity to enhance HC-Pro function, thereby affecting its host range. Therefore, the purpose of this study was to determine whether or not there is a significant difference in VSR activity of the VanMV-CI and DsMV HC-Pro proteins because of the difference in P1.

RNA extraction, RT-PCR and overlapping extension PCR

Constructs for the expression of the HC-Pro on its own as well as the P1/HC-Pro region from each virus were created and successfully incorporated into the entry and destination vector and propagated in *E. coli* however, I was unable to transform these constructs into *A. tumefaciens*.

During the overlapping extension PCR step, the PKP6 construct successfully amplified; however, obtaining the PKP5 construct was unsuccessful. A likely problem in creating the PKP6 construct could have been due to the modified A (adenine) attached to the 3' end of each fusion template generated by *Taq* DNA polymerase. Complementarity of each fusion template, particularly in the overlapping region, could thereby be hindered due to the presence of an extra adenine nucleotide. This could have in turn inhibited the production of the PKP6 construct. Poor primer design could have also been a factor however; due to time constraints, new sets of primers were unable to be made. Further improvements would include a longer overlapping region to guarantee higher chances of success. The attached 'A' nucleotide created by *Taq* DNA polymerase could also be taken into consideration when designing primers.

TOPO cloning

Although the PKP6 construct was successfully obtained in the overlapping extension PCR step, its incorporation into pCR8 was not successful. Although highly unlikely, a possible explanation for the failure to clone the PKP6 construct might be due to the absence of the 3' A overhang during the initial PCR step with *Taq* DNA polymerase. Without the 'A' overhang, the PKP5 construct will not have been able to complement the 'T' overhang found in the pCR8 plasmid (Figure 3.9)

The clonase reaction and subsequent transformation of *E. coli* with pHEX2 plasmids containing gene of interest was also successful with high transformation efficiency and positive confirmation by colony PCR (Figure 3.12 and Figure 3.13). Positive and negative controls were in place to ensure no fault in the procedure.

Agrobacterium transformation

A. tumefaciens is used to transfer genes to plant cells, thus the viral sequences had to be transferred to the binary vector pHEX2 before transfer to agrobacterium. Although colonies were observed following the transformation of agrobacterium, the confirmation PCR step showed that these were false positives. Possible problems with the experiment may have been that the agrobacterium Helper plasmid could have spontaneously mutated and become resistant to spectinomycin enabling *A. tumefaciens* cells to grow on medium containing spectinomycin in the absence of the pHEX2 plasmid, which contains resistance to the antibiotic. By already having

resistance to gentamycin and rifampicin, the helper plasmid will only need to develop resistance to spectinomycin.

The PCR conditions used in verifying the successful transformants showed successful and specific amplification of the purified pHEX2 plasmid containing the gene of interest with no non-specific amplification as seen in Figure 3.14A lanes 2, 3, 9 and 10 (PKP1-4 respectively). This shows that the PCR conditions were sufficient enough to detect the plasmid in Agrobacterium and rather than a fault in the PCR conditions, the transformation step itself was unsuccessful.

Although the transformation of PKP1-4, pHEX2 empty vector and pHEX2-GFP was unsuccessful, the transformation of pART27-P19 was successful. This suggests that the fault was not due to incompetency of Agrobacterium itself but rather the inefficient freeze-thaw method used. Further improvements to be made would be carrying out the transformation by the more efficient electroporation method. However, due to equipment and time constraints this method could not be pursued.

Chapter 4

General Discussion

RNA viruses are quickly evolving entities with large and diverse populations created by the lack of proof replicase reading ability (Sztuba et al., 2011). This characteristic has enabled the evolution of a counterattacking silencing suppressor mechanism to overcome the plant's natural defence mechanism, namely RNAi (Gammelgård et al., 2007; T. Haikonen, Rajamäki, & Valkonen, 2013; Kasschau et al., 2003; M.-J. Li et al., 2014; Tena Fernández et al., 2013; Adrian Valli et al., 2014). The HC-Pro gene of potyviruses has been well characterised as having silencing suppressor activity, which varies in efficiency depending on its sequence (Gammelgård et al., 2007; T. Haikonen et al., 2013; Kasschau et al., 2003; M.-J. Li et al., 2014; Tena Fernández et al., 2013; Adrian Valli et al., 2014). Others recognise the importance of the P1 gene and to a lesser extent, the VPg protein, in enhancing the VSR activity of HC-Pro (Lacombe et al., 2010; Maliogka et al., 2012; Martínez & Daròs, 2014; Pasin et al., 2014; Rohožková & Navrátil, 2011; Salvador et al., 2008).

In addition to overcoming RNAi and infecting their natural hosts, virus strains may also develop infection in new hosts as a result of recombination and/or random mutations. The importance of recombination in extending host range was demonstrated when a recombination event took place between SMV and BCMV creating WMV with a much broader host range compared to SMV (Desbiez & Lecoq, 2004). In addition, a previous study identified a strain of SMV infecting *P. ternate*, a natural host of DsMV (Valli et al., 2007). That study identified a recombination event between DsMV and SMV in the P1 region creating a SMV strain that still maintains the ability to infect Soy bean cultivars, but also able to infect *P. ternate*, a natural host of DsMV. A number of other studies have also reported BCMV and CMV infecting peanut, a natural host for *Peanut stripe virus* (PStV) (De Breuil et al., 2015; Park et al., 2015). VanMV, reported in 1987 (Wisler & Zettler, 1987), was later characterised as a strain of DsMV having similar serological properties and significant sequence similarity in the CP coding region yet, differ in host range (Farreyrol, 2005). VanMV exclusively infects vanilla species and has only been reported in the South Pacific while DsMV infects a number of aroid and crop plants with a global distribution (Pearson et al., 1998; Pearson & Grisoni, 2002). Although many attempts have been made to mechanically inoculate a number of plant species susceptible to DsMV infection with VanMV, no attempts have been made to mechanically inoculate vanilla plant species with DsMV. Little is known about the

molecular basis for the difference in host range and distribution and how VanMV developed the ability to infect orchids while losing the ability to infect aroid plants. This has largely been due to the VanMV genome sequence not being available.

The first aim of this research, therefore, was to characterise the VanMV genome sequence in order to determine a possible molecular basis for the difference in host range and thereby establish the likely evolutionary trajectories that took place which led to the emergence of the orchid-infecting VanMV-CI.

Characterisation of VanMV-CI genome

Characterising the complete VanMV-CI genome by making comparisons to all available DsMV genomes in GenBank revealed VanMV-CI is a typical potyvirus with a single ORF flanked by 5' and 3' UTR's. It is predicted to encode a polyprotein of 3139 aa from which ten mature proteins can be predicted. Further, a PIPO protein can be predicted between nt 3063-3331 based on a nt sequence alignment with DsMV KT026108. However, the actual size of the PIPO protein has not been experimentally verified and may differ in size from that of VanMV-FP and other DsMV sequences as other publically available DsMV PIPO sequences have not been annotated. VanMV-CI has all the potyvirus conserved motifs with all predicted motifs found in DsMV with the exception of the IVFG motif in the P1 protein. Further, VanMV-CI is expected to be aphid transmissible since the DAG motif is present in the CP protein. In a previous study, Farreyrol et al (2006) identified CP sequences with and without the DAG motif from the same infected plant material however, only the sequence containing the DAG motif was detected suggesting it is the dominant sequence in the host. Individual genes and the entire genome of VanMV-CI were compared to DsMV indicating high sequence similarity. This analysis further supports VanMV being classified a strain of DsMV according to current demarcation criteria (Adams et. al., 2011).

Comparing VanMV and DsMV

The second aim of this research was to compare and contrast the VanMV-CI genome to all available DsMV genomes. This was done to identify differences between VanMV and DsMV, which could account for differences in host range. Sequence comparisons made between VanMV-CI and available DsMV sequences revealed a number of indels particularly in the 5'UTR, P1 and the CP gene where VanMV-CI differs from all DsMV

sequences included in the analysis. In addition, the putative cleavage site of VanMV-CI between P1/HC-Pro was most divergent when compared to other DsMV sequences (Table 2.4). A number of studies have determined different host range determinants for different viruses including CP, HC-Pro, P1, P3, VPg, 6K1 and CI (Frédéric & Juan, 2015; Jayathilake, 2004; Maliogka et al., 2012; Sáenz et al., 2002; Schaad et al., 1997; Shan et al., 2015). The 5' UTR of potyviruses has not been reported to have host-determining properties suggesting this region of the genome does not contain the host range determinant for VanMV-CI. In addition, the N-terminus of the CP is highly variable between DsMV isolates as well as VanMV isolates and between DsMV and VanMV. Therefore, this region is not likely to contain the host range determinant. A sequence analysis highlighted an indel in the N-terminal region of the DsMV and VanMV P1. This indel of 30 aa was found in the hyper-variable N-terminal region of the P1 protein. In addition, the cleavage site between P1/HC-Pro was the most divergent. The N-terminal region of the P1 protein has been linked to host determining factors in which it determines the compatibility of the virus with host infection and assists in genome amplification and virus infection (Salvador et al., 2008; Shan et al., 2015). Further, P1 can also assist HC-Pro in its VSR activity after successful cleavage from HC-Pro (Gammelgård et al., 2007). For this reason, the P1-HC-Pro was a region of interest for studying the host range difference between VanMV and DsMV.

A limitation to this analysis is the lack of sequence information for VanMV-FP particularly in the 5' region of genome. This information would help distinguish if the difference between VanMV-CI and DsMV in the P1 protein is also apparent in VanMV-FP. In addition, more sequence information for VanMV and DsMV from this region of the genome is necessary to confirm this as a conserved difference between VanMV and DsMV. More sequences will be required to provide a clearer depiction of what the host range determinant might be and also better understand how VanMV lost its possible ability to infect aroids and acquired its ability to infect orchids.

Evolutionary history of VanMV-CI

Vanilla was introduced into French Polynesia in 1848 and from 1899-1866 French Polynesia exported an average of 158 t of cured vanilla beans annually (Wisler & Zettler, 1987). Between 1967-1981 however, vanilla production was reduced

dramatically to an average of 0.6 t annually prompting a nationwide investigation to restore vanilla production (Wisler & Zettler, 1987). VanMV was later discovered in 1986 as a mosaic-inducing potyvirus with similar serological properties to DsMV (Wisler & Zettler, 1987). Interestingly, DsMV was discovered in 1970 and had similar serological properties to TEV and BlCMV (Zettler et al., 1970). These findings suggest VanMV emerged well before its discovery and would have been around the same time DsMV was discovered in Florida (Abo El-Nil, 1977). While DsMV has a worldwide distribution, including a strain infecting the orchid *Spiranthes cernua* from West Virginia USA (Guaragna, Ndum, & Jordan, 2006) VanMV is limited to the South Pacific (Farreyrol et al., 2006). Based on these reports alone, it is difficult to determine whether DsMV or VanMV was the first to emerge. Whole genome analysis suggests that VanMV is an older lineage than DsMV (Babu et al., 2011) while others suggest otherwise (Chang, 2012) leaving the evolutionary trajectory that lead to the emergence of VanMV unclear.

The third aim of this research was therefore to determine the evolutionary relationship of VanMV-CI compared to members of the BCMV group of potyviruses. Phylogenetic analysis of the VanMV-CI genome and individual genes not including the CP gene suggested the lineage ending in VanMV was ancestral as previously proposed by Babu et al. (2011) however, a comprehensive analysis of all available DsMV CP genes suggests DsMV is ancestral. In addition, this phylogenetic analysis suggests that based on the sequences available to date, DsMV originated in China and Japan, which later spread to the South Pacific. This suggests that DsMV would have later evolved to produce the orchid-infecting variant VanMV that exclusively infects vanilla species.

Although an orchid-infecting variant of DsMV has been reported, only a portion of the CP gene of DsMV infecting *S. ceruna* has been documented (unpublished sequence) with no information regarding the 5' region. This indicates that more field samples should be analysed to get a better evaluation of the sequences and geographical distribution of orchid-infecting DsMV variants.

While this thesis has focused on comparisons between VanMV and DsMV in an attempt to answer the question regarding which came first, this study did not attempt

to answer the origin of DsMV/VanMV. To do this, comparisons should be made with other potyviruses, as well as sequences from the genera in the *Potyviridae* family.

Recombination analysis

There are a number of reported instances where recombination events between plant viruses have created new and distinct variants or distinct new viruses. Desbiez & Lecoq (2004) reported a recombination event between SMV and BCMV in the N-terminal region of the P1 protein creating WMV with a much broader host range compared to SMV. In addition, Valli et al. (2007) reported a recombination event between SMV and DsMV in the P1 region enabling a strain of SMV to infect *P. ternate*, a natural host of DsMV. Further, recombination events have been detected in the CP gene of a number of economically important viruses with ancestral decedents (Boulila, 2010b). It is possible that DsMV could have undergone a recombination event in the P1 region with as yet unknown virus thereby creating VanMV.

Therefore, the fourth aim of this research was to determine whether a recombination event may have taken place between DsMV and VanMV or with another virus creating VanMV with the ability to exclusively infect vanilla plants. The *in silico* recombination analysis revealed a putative recombination event between a VanMV-like sequence and a BSVA-like ancestor in the N-terminal region of the P1 gene. The recombination analysis identified the recombinant (BSVA-like ancestor) as sharing 19 out of 30 amino acids (61.5%) in the region corresponding to the P1 indel described in Chapter 2. Another event was detected in the C-terminal P1 region of DsMV (NC003537) with the recombinant (BCMV-like ancestor) sharing 61.7% identity.

A limitation to these findings however, is that the recombination event was detected by only one method and the sequence similarity between the recombinant and ancestral parent was below 75%. Recombination events detected by at least three methods with sequence similarity >75% are considered significant (reference?). Another limitation would be the sequences involved in the analysis that was carried exclusively on all available potyvirid sequences. It is possible that the recombination event was not between potyviruses, but between a potyvirus and a more distant virus of another family. Orchid viruses of other virus families should be included in future analyses to detect more distant recombination events. Further, the recombination

analysis carried out could have been conducted in conjunction with other programs readily available to further validate recombination events detected in this analysis. In addition, the recombination event could have been between DsMV and a yet to be discovered potyvirus so the recombination analysis should be repeated as new potyvirus genomes are reported.

Silencing suppressor activity

A review of host range determining factors identified for members of the *Potyviridae* family of RNA viruses highlighted the P1-HC-Pro region of VanMV and DsMV as regions of interest. This region of the genome is primarily responsible for genome replication, host adaptation, cell-to-cell movement and silencing suppression (Frédéric & Juan, 2015; Pasin et al., 2014; Rajamäki et al., 2005; Rohožková & Navrátil, 2011; Shan et al., 2015). As discussed in Section 3.1.4, the host range of a virus can be determined by the virus's ability to overcome the plants natural defence mechanism. By evaluating the VSR activity of VanMV and DsMV, some inference can be made about their ability to silence RNAi, which is dependant on the proteolytic activity of P1 in separating P1 protein from HC-Pro.

The fifth and last aim of this research was therefore to compare and contrast the VSR activity of VanMV-CI HC-Pro in comparison to DsMV HC-Pro with and without their adjacent P1 proteins. This work was aimed to determine whether the sequence difference between HC-Pro regions and P1 proteins correlated with any detectable difference in VSR activity. Hybrid constructs (PKP5 and PKP6) were also created to compare how the P1 of VanMV-CI and DsMV-NZ would interact with each HC-Pro. However, due to a number of issues encountered with the cloning of these viral sequences, this comparison could not be done. The issues were the result of being unable to construct the PKP5 hybrid construct, unable to ligate the PKP6 construct to PCR8 and the unsuccessful transformation of agrobacterium with constructs PKP1-4. The likely cause of this issue was most probably the inefficient freeze-thaw method used.

Results of this study would have revealed whether or not the indel in the N-terminal region of the P1 protein correlated with a difference in VSR activity. This would have been achieved by Agro-infiltrating VanMV-CI and DsMV constructs into *N.*

benthamiana. It must be noted that *N. benthamiana* is a model plant species and not a natural host for either DsMV or VanMV. Therefore, such constructs should also be tested in aroids and vanilla. Agroinfiltration methods for aroids and vanilla have not been reported, thus these methods need to be established. In order to determine the compatibility of each P1 with subsequent hosts, the following constructs and comparisons would have to be made by Agro-infiltrating into a vanilla plant and a natural host for DsMV:

1. VanMV P1/VanMV HC-Pro
2. DsMV P1/DsMV HC-Pro
3. VanMV P1/DsMV HC-Pro
4. DsMV P1/VanMV HC-Pro

This comparison could determine whether VanMV P1 is compatible with aroid plants and whether DsMV P1 is compatible with vanilla plants. In addition, protein- protein interactions should be evaluated for each construct in each host to determine whether the virus-host interactions are the same for VanMV

In conclusion, this study supports VanMV as being a strain of DsMV. The most likely viral protein to have a role in the host range difference between these viruses is the P1. The hypothesis that the P1 influences the activity of the VanMV/DsMV HC-Pro could not be tested, but some constructs required for this analysis have been generated. Suggestions have been made to improve the chances of creating the other constructs required for analysis and transfer them into agrobacterium for infiltration into plants for expression analysis.

Reference list

Abo El-Nil, M. M. (1977). Purification, Serology, and Some Physical Properties of Dasheen Mosaic Virus. *Phytopathology, 77*(12), 1445. doi:10.1094/Phyto-67-1445

Adams, Antoniw, J. F., & Beaudoin, F. (2005). Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Molecular Plant Pathology, 6*(4), 471-487. doi:10.1111/j.1364-3703.2005.00296.x

Adams, Antoniw, J. F., & Fauquet, C. M. (2005). Molecular criteria for genus and species discrimination within the family Potyviridae. *Archives of Virology, 150*(3), 459-479. doi:10.1007/s00705-004-0440-6

Ala-Poikela, M., Goytia, E., Haikonen, T., Rajamäki, M.-L., & Valkonen, J. P. T. (2011). Helper Component Proteinase of the Genus Potyvirus Is an Interaction Partner of Translation Initiation Factors eIF(iso)4E and eIF4E and Contains a 4E Binding Motif. *Journal of Virology, 85*(13), 6784-6794. doi:10.1128/JVI.00485-11

Anandalakshmi, R., Marathe, R., Ge, X., Herr, J. M., Jr., & et al. (2000). A calmodulin-related protein that suppresses posttranscriptional gene silencing in plants. *Science, 290*(5489), 142-144. doi:10.1126/science.290.5489.142

Angaji, S., Hedayati, S. S., Hosein, R., Samad, S., Shiravi, S., & Madani, S. (2010). Application of RNA Interference in Plants. *Plant Omics, 3*(3), 77. doi:10.1186/1758-907X-3-5

Anindya, R., Chittori, S., & Savithri, H. S. (2005). Tyrosine 66 of Pepper vein banding virus genome-linked protein is uridylylated by RNA-dependent RNA polymerase. *Virology, 336*(2), 154-162. doi:10.1016/j.virol.2005.03.024

Anindya, R., & Savithri, H. S. (2004). Potyviral NIa Proteinase, a Proteinase with Novel Deoxyribonuclease Activity. *Journal of Biological Chemistry, 279*(31), 32159-32169. doi:10.1074/jbc.M404135200

Aparicio, F., Thomas, C. L., Lederer, C., Niu, Y., Wang, D., & Maule, A. J. (2005). Virus Induction of Heat Shock Protein 70 Reflects a General Response to Protein Accumulation in the Plant Cytosol. *Plant Physiology, 138*(1), 529-536. doi:10.1104/pp.104.058958

Babu, B., Hegde, V., Makeshkumar, T., & Jeeva, M. L. (2011). Detection and Identification of Dasheen mosaic virus Infecting Colocasia esculenta in India. *Indian Virological Society, 22*(1), 59-62. doi:10.1007/s13337-011-0030-7

Barnett, O. W. (1992). *Potyvirus taxonomy*. Little Hampton, England: Springer-Verlag. Retrieved from https://books.google.co.nz/books?id=VuZFAAAAYAAJ

Beauchemin, C., & Laliberté, J.-F. (2007). The Poly(A) Binding Protein Is Internalized in Virus-Induced Vesicles or Redistributed to the Nucleolus during Turnip Mosaic

Virus Infection. *Journal of Virology, 81*(20), 10905-10913. doi:10.1128/JVI.01243-07

Bedhomme, S., Lafforgue, G., & Elena, S. F. (2012). Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Mol Biol Evol, 29*(5), 1481-1492. doi:10.1093/molbev/msr314

Berger, P., & Parrish, J. (2011). Potyvirus. In C. Tidona & G. Darai (Eds.), *The Springer Index of Viruses* (pp. 1425-1437): Springer New York. doi:10.1007/978-0-387-95919-1_235

Bernhardt, H. S. (2012). The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biology Direct, 7*, 23-23. doi:10.1186/1745-6150-7-23

Bhat, S., Folimonova, S. Y., Cole, A. B., Ballard, K. D., Lei, Z., Watson, B. S., . . . Nelson, R. S. (2013). Influence of Host Chloroplast Proteins on Tobacco mosaic virus Accumulation and Intercellular Movement. *Plant Physiology, 161*(1), 134-147. doi:10.1104/pp.112.207860

Bilgin, D. D., Liu, Y., Schiff, M., & Dinesh-Kumar, S. P. (2003). P58IPK, a Plant Ortholog of Double-Stranded RNA-Dependent Protein Kinase PKR Inhibitor, Functions in Viral Pathogenesis. *Developmental Cell, 4*(5), 651-661. doi:10.1016/S1534-5807(03)00125-4

Borges, F., & Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology, 16*(12), 727-741. doi:10.1038/nrm4085

Boulila, M. (2010a). Putative recombination events and evolutionary history of five economically important viruses of fruit trees based on coat protein-encoding gene sequence analysis. *Biochem Genet, 48*(5-6), 357-375. doi:10.1007/s10528-009-9317-4

Boulila, M. (2010b). Putative recombination events and evolutionary history of five economically important viruses of fruit trees based on coat protein-encoding gene sequence analysis. *Biochemical Genetics, 48*(5-6), 357-375. doi:10.1007/s10528-009-9317-4

Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell, 136*(4), 642-655. doi:10.1016/j.cell.2009.01.035

Chang, W. L. (2012). *Temporal and Spatial Analysis of Dasheen mosaic virus* (Masters Thesis). Auckland University of Technology, Auckland. Retrieved from http://aut.researchgateway.ac.nz/bitstream/handle/10292/4706/ChangWL.pdf?sequence=3

Chase, A. R. (1982). Dasheen Mosaic Virus Infection of Dieffenbachia Cultivars. *Plant Disease, 66*(1), 891. doi:10.1094/PD-66-891

Chen, J., Chen, J., Chen, J., & Adams, M. J. (2001). Molecular characterisation of an isolate of Dasheen mosaic virus from Zantedeschia aethiopica in China and comparisons in the genus Potyvirus. *Archives of Virology, 146*(9), 1821-1829. doi:10.1007/s007050170068

Chung, B. Y. W., Miller, W. A., Atkins, J. F., & Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences of the United States of America, 105*(15), 5897-5902. doi:10.1073/pnas.0800468105

Cotton, S., Grangeon, R., Thivierge, K., Mathieu, I., Ide, C., Wei, T., . . . Laliberté, J.-F. (2009). Turnip Mosaic Virus RNA Replication Complex Vesicles Are Mobile, Align with Microfilaments, and Are Each Derived from a Single Viral Genome. *Journal of Virology, 83*(20), 10460-10471. doi:10.1128/JVI.00819-09

Cui, X., Wei, T., Chowda-Reddy, R. V., Sun, G., & Wang, A. (2010). The Tobacco etch virus P3 protein forms mobile inclusions via the early secretory pathway and traffics along actin microfilaments. *Virology, 397*(1), 56-63. doi:10.1016/j.virol.2009.11.015

Culley, A. L., Lang, A. S., & Suttle, C. A. (2003). High diversity of unknown picorna-like viruses in the sea. *Nature, 424*(6952), 1054. doi:10.1038/nature01886

Curtin, S., Watson, J., Smith, N., Eamens, A., Blanchard, C., & Waterhouse, P. (2008). The roles of plant dsRNA‐binding proteins in RNAi‐like pathways. *FEBS Letters*, 2753. doi:10.1016/j.febslet.2008.07.004

Danielson, D. C., & Pezacki, J. P. (2013). Studying the RNA silencing pathway with the p19 protein. *FEBS Letters, 587*, 1198-1205. doi:10.1016/j.febslet.2013.01.036

De Breuil, S., Nome, C. F., Bejerman, N., Giolitti, F., Trucco, V., & Lenardon, S. (2015). First report of bean common mosaic virus, peanut strain, infecting peanut in Argentina. *Plant Disease, 99*(5), 735. doi:10.1094/PDIS-11-14-1125-PDN

de Farias, S. T., Rêgo, T. G., & José, M. V. (2016). tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. *Life, 6*(2). doi:10.3390/life6020015

De Mejia, M. V. G., Hiebert, E., Purcifull, D. E., Thornbury, D. W., & Pirone, T. P. (1985). Identification of potyviral amorphous inclusion protein as a nonstructural, virus-specific protein related to helper component. *Virology, 142*(1), 34-43. doi:10.1016/0042-6822(85)90420-9

Desbiez, C., & Lecoq, H. (2004a). The nucleotide sequence of Watermelon mosaic virus (WMV, Potyvirus) reveals interspecific recombination between two related potyviruses in the 5' part of the genome. *Arch Virol, 149*(8), 1619-1632. doi:10.1007/s00705-004-0340-9

Desbiez, C., & Lecoq, H. (2004b). The nucleotide sequence of Watermelon mosaic virus (WMV, Potyvirus) reveals interspecific recombination between two related potyviruses in the 5' part of the genome. *Archives of Virology, 149*(8), 1619-1632. doi:10.1007/s00705-004-0340-9

Dielen, A.-S., Badaoui, S., Candresse, T., & German-Retana, S. (2010). The ubiquitin/26S proteasome system in plant–pathogen interactions: a never-ending hide-and-seek game. *Molecular Plant Pathology, 11*(2), 293-308. doi:10.1111/j.1364-3703.2009.00596.x

Dolja, V. V., Haldeman, R., Robertson, N. L., Dougherty, W. G., & Carrington, J. C. (1994). Distinct functions of capsid protein in assembly and movement of tobacco etch potyvirus in plants. *The EMBO Journal, 13*(6), 1482-1491. doi:10.1111/j.1364-3703.2009.00596.x

Dombrovsky, A., Huet, H., Chejanovsky, N., & Raccah, B. (2005). Aphid transmission of a potyvirus depends on suitability of the helper component and the N terminus of the coat protein. *Archives of Virology, 150*(2), 287-298. doi:10.1007/s00705-004-0407-7

Domingo, E., Sheldona, J., & Peralesa, C. (2012). Viral Quasispecies Evolution. *Mol Biol Evol, 76*(2), 159-216. doi:10.1128/MMBR.05023-11

Domingo, E., Sheldona, J., & Peralesa, C. (2012). Viral Quasispecies Evolution. *Molecular Biology Evolution, 76*(2), 159-216. doi:10.1128/MMBR.05023-11

Dougherty, W. G., & Dawn-Parks, T. (1991). Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: Identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology, 183*(2), 449-456. doi:10.1016/0042-6822(91)90974-G

Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics, 9*(4), 267-276. doi:10.1038/nrg2323

Dufresne, P. J., Thivierge, K., Cotton, S., Beauchemin, C., Ide, C., Ubalijoro, E., . . . Fortin, M. G. (2008). Heat shock 70 protein interaction with Turnip mosaic virus RNA-dependent RNA polymerase within virus-induced membrane vesicles. *Virology, 374*(1), 217-227. doi:10.1016/j.virol.2007.12.014

Durzyńska, J., & Goździcka-Józefiak, A. (2015). Viruses and cells intertwined since the dawn of evolution. *Virology Journal, 12*, 169-169. doi:10.1186/s12985-015-0400-7

Edwards, S. (2014). *Cross Protection and Molecular Characterisation of Three Tamarillo Mosaic Virus Isolates*. Retrieved from https://books.google.co.nz/books?id=Z1VtrgEACAAJ

Edwardson, J. R., & Christie, R. G. (1996). Cylindrical Inclusions. In *Bulletin 894* (Vol. 894). Institute of Food and Agricultural Sciences: Gainsville : Florida Agricultural Experiment Station.

Elena, S. F., & Rodrigo, G. (2012). Towards an integrated molecular model of plant–virus interactions. *Current Opinion in Virology, 2*(6), 719-724. doi:10.1016/j.coviro.2012.09.004

Farreyrol, K. (2005). *Characterisation and detection of viruses (Cucumovirus, Potyvirus) infecting vanilla in Reunion Island and Polynesian Islands* (Doctoral Dissertation). University of Auckland, Auckland, New Zealand Retrieved from https://researchspace.auckland.ac.nz/handle/2292/1289

Farreyrol, K., Pearson, M. N., Grisoni, M., Cohen, D., & Beck, D. (2006). Vanilla mosaic virus isolates from French Polynesia and the Cook Islands are Dasheen mosaic virus strains that exclusively infect vanilla. *Archives of Virology, 151*(5), 905-919. doi:10.1007/s00705-005-0680-0

Fernández, A., Guo, H. S., Sáenz, P., Simón-Buela, L., de Cedrón, M. G., & García, J. A. (1997). The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. *Nucleic Acids Research, 25*(22), 4474-4480. doi:10.1093/nar/25.22.4474

Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research, 117*(1), 5-16. doi:10.1016/j.virusres.2006.01.010

Forterre, P., & Krupovic, M. (2012). The Origin of Virions and Virocells: The Escape Hypothesis Revisited. *Viruses: Essential Agents of Life*, 43. doi:10.1007%2F978-94-007-4899-6_3

Frédéric, R., & Juan, A. G. (2015). Molecular Biology of Potyviruses. *Elsevier Inc, 92*, 101-199. doi:10.1016/bs.aivir.2014.11.006

Gabrenaite-Verkhovskaya, R., Andreev, I. A., Kalinina, N. O., Torrance, L., Taliansky, M. E., & Mäkinen, K. (2008). Cylindrical inclusion protein of potato virus A is associated with a subpopulation of particles isolated from infected plants. *Journal of General Virology, 89*(3), 829-838. doi:10.1099/vir.0.83406-0

Gammelgård, E., Mohan, M., & Valkonen, J. P. T. (2007). Potyvirus-induced gene silencing: the dynamic process of systemic silencing and silencing suppression. *Journal of General Virology, 88*(8), 2337-2346. doi:10.1099/vir.0.82928-0

Gibbs, & Ohshima, K. (2010). Potyviruses and the digital revolution. *Annu Rev Phytopathol, 48*, 205-223. doi:10.1146/annurev-phyto-073009-114404

Gibbs, & Ohshima, K. (2010). Potyviruses and the Digital Revolution. *Annual Review of Phytopathology, 48*(1), 205-223. doi:10.1146/annurev-phyto-073009-114404

Gibbs, Ohshima, K., Phillips, M. J., & Gibbs, M. J. (2008a). The Prehistory of Potyviruses: Their Initial Radiation Was during the Dawn of Agriculture. *Public Library of Science ONE, 3(6)* e2523. doi:10.1371/journal.pone.0002523

Gibbs., Trueman, J. W. H., & Gibbs, M. J. (2008b). The bean common mosaic virus lineage of potyviruses: Where did it arise and when? *Archives of Virology, 153*(12), 2177-2187. doi:10.1007/s00705-008-0256-x

Govaerts, R. (2012). *World Checklist of Araceae*. Retrieved from http://apps.kew.org/wcsp/

Govier, D. A., Kassanis, B., & Pirone, T. P. (1977). Partial purification and characterization of the potato virus Y helper component. *Virology, 78*(1), 306-314. doi:10.1016/0042-6822(77)90101-5

Goytia, E., Fernández-Calvino, L., Martínez-García, B., López-Abella, D., & López-Moya, J. J. (2006). Production of plum pox virus HC-Pro functionally active for aphid transmission in a transient-expression system. *Journal of General Virology, 87*(11), 3413-3423. doi:10.1099/vir.0.82301-0

Greber, R., & Shaw, D. (1986). Dasheen mosaic virus in Queensland. *Australasian Plant Pathology, 15*(2), 29. doi:10.1071/APP9860029

Grisoni, M., Davidson, F., Hyrondelle, C., Farreyrol, K., Caruana, M., & Pearson, M. (2004). Nature, Incidence, and Symptomatology of Viruses Infecting Vanilla tahitensis in French Polynesia. *Plant Disease, 88*(2), 119-124. doi:10.1094/PDIS.2004.88.2.119

Grisoni, M., Moles, M., Farreyrol, K., Rassaby, L., Davis, R., & Pearson, M. (2006). Identification of potyviruses infecting vanilla by direct sequencing of a short RT-PCR amplicon. *Plant Pathology, 55*(4), 523-529. doi:10.1111/j.1365-3059.2006.01397.x

Guaragna, M. A., Ndum, O., & Jordan, R. (2006). Detection and Characterization of Two Previously Undescribed Potyviruses in the Terrestrial Orchid Spiranthes cernua. *Acta horticulturae, 722*, 209-217. doi:10.17660/ActaHortic.2006.722.26

Ha, C. (2007). *Detection and identification of Potyviruses and Geminiviruses in Vietnam* (Doctoral Dissertation). Queensland University of Technology, QUT ePrints. Retrieved from http://www.eprints.qut.edu.au/16540/

Ha, C., Revill, P., Harding, R. M., Vu, M., & Dale, J. L. (2008). Identification and sequence analysis of potyviruses infecting crops in Vietnam. *Archives of Virology, 153*(1), 45-60. doi:10.1007/s00705-007-1067-1

Hafrén, A., & Mäkinen, K. (2008). Purification of viral genome-linked protein VPg from potato virus A-infected plants reveals several post-translationally modified forms of the protein. *Journal of General Virology, 89*(6), 1509-1518. doi:10.1099/vir.0.83649-0

Haikonen, T., Rajamäki, M.-L., Tian, Y.-P., & Valkonen, J. P. T. (2013). Mutation of a Short Variable Region in HCpro Protein of Potato virus A Affects Interactions with a Microtubule-Associated Protein and Induces Necrotic Responses in Tobacco. *Molecular Plant-Microbe Interactions, 26*(7), 721-733. doi:10.1094/MPMI-01-13-0024-R

Haikonen, T., Rajamäki, M. L., & Valkonen, J. P. T. (2013). Improved silencing suppression and enhanced heterologous protein expression are achieved using an engineered viral helper component proteinase. *Journal of Virological methods, 193*(2), 687-692. doi:10.1016/j.jviromet.2013.07.054

Hampton, R. O., Jensen, A., & Hagel, G. T. (2005). Attributes of Bean Yellow Mosaic Potyvirus Transmission from Clover to Snap Beans by Four Species of Aphids (Homoptera: Aphididae). *Journal of Economic Entomology, 98*(6), 1816-1823. doi:10.1093/jee/98.6.1816

Hari, V. (1981). The RNA of Tobacco etch virus: Further characterization and detection of protein linked to RNA. *Virology, 112*(2), 391-399. doi:10.1016/0042-6822(81)90286-5

Harrison, B. D., Finch, J. T., Gibbs, A. J., Hollings, M., Shepherd, R. J., Valenta, V., & Wetter, C. (1971). Sixteen groups of plant viruses. *Virology, 45*(2), 356-363. doi:10.1016/0042-6822(71)90336-9

Hasiów-Jaroszewska, B., Fares, M., & Elena, S. (2014). Molecular Evolution of Viral Multifunctional Proteins: The Case of Potyvirus HC-Pro. *Journal of Molecular Evolution, 78*(1), 75-86. doi:10.1007/s00239-013-9601-0

Hiebert, E., & Purcifull, D. E. (1992). A comparison of pepper mottle virus with potato virus Y and evidence for their distinction [Hiebert1992]. In O. W. Barnett (Ed.), *Potyvirus Taxonomy* (pp. 321-326). Vienna: Springer Vienna. doi:10.1007/978-3-7091-6920-9_34

Hill, S. A., & Wright, D. M. (1980). Identification of dasheen mosaic virus in Dieffenbachia picta and Xanthosoma helliborifolium by immune electron microscopy. *Plant Pathology, 29*(3), 143-144. doi:10.1111/j.1365-3059.1980.tb01200.x/full

Holmes, E. C. (2011). What does virus evolution tell us about virus origins? *Journal of Virology, 85*(11), 5247-5251. doi:10.1128/JVI.02203-10

Hong, Chen, J., Shi, Y. H., & Chen, J. P. (2007). The '6K1' protein of a strain of Soybean mosaic virus localizes to the cell periphery. *Archives of Virology, 152*(8), 1547-1551. doi:10.1007/s00705-007-0972-7

Hong, Y., & Hunt, A. G. (1996). RNA Polymerase Activity Catalyzed by a Potyvirus-Encoded RNA-Dependent RNA Polymerase. *Virology, 226*(1), 146-151. doi:10.1006/viro.1996.0639

Hu, J., Meleisea, S., Wang, M., Shaarawy, M., & Zettler, F. (1995). Dasheen mosaic potyvirus in Hawaiian taro. *Australasian Plant Pathology, 24*(2), 112. doi:10.1071/APP9950112

Hyodo, K., & Okuno, T. (2016). Pathogenesis mediated by proviral host factors involved in translation and replication of plant positive-strand RNA viruses. *Current Opinion in Virology, 17*, 11-18. doi:10.1016/j.coviro.2015.11.004

Ivanov, K. I., Puustinen, P., Gabrenaite, R., Vihinen, H., Rönnstrand, L., Valmu, L., . . . Mäkinen, K. (2003). Phosphorylation of the Potyvirus Capsid Protein by Protein Kinase CK2 and Its Relevance for Virus Infection. *The Plant Cell, 15*(9), 2124-2139. doi:10.1105/tpc.012567

Jamous, R. M., Boonrod, K., Fuellgrabe, M. W., Ali-Shtayeh, M. S., Krczal, G., & Wassenegger, M. (2011). The helper component-proteinase of the Zucchini yellow mosaic virus inhibits the Hua Enhancer 1 methyltransferase activity in vitro. *Journal of General Virology, 92*(9), 2222-2226. doi:10.1099/vir.0.031534-0

Jayathilake, N. (2004). *Defining the molecular basis of host range in Papaya ringspot virus (PRSV) Australia* (Doctoral Dissertation). Queensland University of Technology. Retrieved from http://eprints.qut.edu.au/16102/1/Nishantha_Jayathilake_Thesis.pdf

Jiang, J., & Laliberté, J.-F. (2011). The genome-linked protein VPg of plant viruses — a protein with many partners. *Current Opinion in Virology, 1*(5), 347-354. doi:10.1016/j.coviro.2011.09.010

Jiménez, I., López, L., Alamillo, J. M., Valli, A., & García, J. A. (2006). Identification of a Plum pox virus CI-Interacting Protein from Chloroplast That Has a Negative Effect in Virus Infection. *Molecular Plant-Microbe Interactions, 19*(3), 350-358. doi:10.1094/MPMI-19-0350

Jin, Y., Ma, D., Dong, J., Jin, J., Li, D., Deng, C., & Wang, T. (2007). HC-Pro Protein of Potato Virus Y Can Interact with Three Arabidopsis 20S Proteasome Subunits In Planta. *Journal of Virology, 81*(23), 12881-12888. doi:10.1128/jvi.00913-07

Jordan, R., Guaragna, M. A., Kinard, G., & Lynn, S. (2002). DETECTION AND FIRST REPORT OF DASHEEN MOSAIC VIRUS AND A SECOND POTYVIRUS INFECTING THE TERRESTRIAL ORCHID SPIRANTHES CERNUA. *Acta Horticulture, 568*, 253-260. doi:10.17660/ActaHortic.2002.568.37

Kamala, S., Makeshkumar, T., Sreekumar, J., & Chakrabarti, S. K. (2015). Whole transcriptome sequencing of diseased elephant foot yam reveals complete genome sequence of Dasheen mosaic virus. *Virology Reports, 5*, 1-9. doi:10.1016/j.virep.2014.11.001

Karthikeyan, A., Deivamani, M., Shobhana, V. G., Sudha, M., & Anandhan, T. (2013). RNA interference: evolutions and applications in plant disease management.

Archives of Phytopathology & Plant Protection, 46(12), 1430. doi:10.1080/03235408.2013.769315

Kasschau, K. D., & Carrington, J. C. (1998). A Counterdefensive Strategy of Plant Viruses. *Cell, 95*(4), 461-470. doi:10.1016/S0092-8674(00)81614-1

Kasschau, K. D., & Carrington, J. C. (2001). Long-Distance Movement and Replication Maintenance Functions Correlate with Silencing Suppression Activity of Potyviral HC-Pro. *Virology, 285*(1), 71-81. doi:10.1006/viro.2001.0901

Kasschau, K. D., Xie, Z., Allen, E., Llave, C., Chapman, E. J., Krizan, K. A., & Carrington, J. C. (2003). P1/HC-Pro, a Viral Suppressor of RNA Silencing, Interferes with Arabidopsis Development and miRNA Function. *Developmental Cell, 4*(2), 205-217. doi:10.1016/S1534-5807(03)00025-X

Katoch, R., & Thakur, N. (2013). Advances in RNA interference technology and its impact on nutritional improvement, disease and insect control in plants. *Applied Biochemistry and Biotechnology, 169*(5), 1579-1605. doi:10.1007/s12010-012-0046-5

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data

Kekarainen, T., Savilahti, H., & Valkonen, J. P. T. (2002). Functional Genomics on Potato Virus A: Virus Genome-Wide Map of Sites Essential for Virus Propagation. *Genome Research, 12*(4), 584-594. doi:10.1101/gr.220702

King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. e. (2012). Virus taxonomy; classification and nomenclature of viruses; ninth report of the International Committee on Taxonomy of Viruses. *Reference and Research Book News, 27*(6).

Klein, P. G., Klein, R. R., Rodriguez-Cerezo, E., Hunt, A. G., & Shaw, J. G. (1994). Mutational Analysis of the Tobacco Vein Mottling Virus Genome. *Virology, 204*(2), 759-769. doi:10.1006/viro.1994.1591

Koichiro, T., Glen, S., Daniel, P., Alan, F., & Sudhir, K. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Molecular Biology and Evolution.

Koonin, Dolja, V. V., & Krupovic, M. (2015). Review: Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology, 479-480*, 2-25. doi:10.1016/j.virol.2015.02.039

Koonin, Wolf, Y. I., Nagasaki, K., & Dolja, V. V. (2008). The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nature Reviews. Microbiology, 6*(12), 925-939. doi:10.1038/nrmicro2030

Kusaba, M. (2004). RNA interference in crop plants. *Current Opinion in Biotechnology, 15*, 139-143. doi:10.1016/j.copbio.2004.02.004

Lacombe, S., Bangratz, M., Vignols, F., & Brugidou, C. (2010). The rice yellow mottle virus P1 protein exhibits dual functions to suppress and activate gene silencing. *The Plant Journal, 61*(3), 371-382. doi:10.1111/j.1365-313X.2009.04062.x

Lam, J. K. W., Chow, M. Y. T., Zhang, Y., & Leung, S. W. S. (2015). siRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy Nucleic Acids, 4*, e252. doi:10.1038/mtna.2015.23

Langenberg, W. G., & Zhang, L. (1997). Immunocytology Shows the Presence of Tobacco Etch Virus P3 Protein in Nuclear Inclusions. *Journal of Structural Biology, 118*(3), 243-247. doi:10.1006/jsbi.1997.3856

Lauring, & Andino. (2010). Quasispecies Theory and the Behavior of RNA Viruses *Public Library of Science Pathogens, 6*(7). doi:10.1371/journal.ppat.1001005

Lecoq, H., Fabre, F., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., Schoeny, A., & Desbiez, C. (2011). Search for factors involved in the rapid shift in Watermelon mosaic virus (WMV) populations in South-eastern France. *Virus Res, 159*(2), 115-123. doi:10.1016/j.virusres.2011.05.004

Li, M.-J., Kim, J.-K., Seo, E.-Y., Hong, S. M., Hwang, E.-I., Moon, J.-K., . . . Lim, H.-S. (2014). Sequence variability in the HC-Pro coding regions of Korean soybean mosaic virus isolates is associated with differences in RNA silencing suppression. *Archives of Virology, 159*(6), 1373-1383. doi:10.1007/s00705-013-1964-4

Li, R. H., Zettler, F. W., Purcifull, D. E., & Hiebert, E. (1998). The nucleotide sequence of the 3′ -terminal region of dasheen mosaic virus (Caladium isolate) and expression of its coat protein in Escherichia coli for antiserum production. *Archives of Virology, 143*(12), 2461.

Lima Roberto, C. A., Lima, J. A. A., & Aguiar, J. R. (2004). Serological identification of Dasheen mosaic virus in Anthurium sp. in the State of Ceará. *Fitopatologia Brasileira*(1), 105. doi:10.1590/S0100-41582004000100020

Lin, L., Luo, Z., Yan, F., Lu, Y., Zheng, H., & Chen, J. (2011). Interaction between potyvirus P3 and ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) of host plants. *Virus Genes, 43*(1), 90-92. doi:10.1007/s11262-011-0596-6

Llave, C. (2010). Review: Virus-derived small interfering RNAs at the core of plant–virus interactions. *Trends in Plant Science, 15*, 701-707. doi:10.1016/j.tplants.2010.09.001

Maia, I. G., Haenni, A.-L., & Bernardi, F. (1996). Potyviral HC-Pro: a multifunctional protein. *Journal of General Virology, 77*(7), 1335-1341. doi:10.1099/0022-1317-77-7-1335

Mäkinen, K., & Hafren, A. (2014). Intracellular coordination of potyviral RNA functions in infection. *Frontiers in Plant Science, 5*(110). doi:10.3389/fpls.2014.00110

Maliogka, V. I., Salvador, B., Carbonell, A., SÁEnz, P., LeÓN, D. S., Oliveros, J. C., . . . SimÓN-Mateo, C. (2012). Virus variants with differences in the P1 protein coexist in a Plum pox virus population and display particular host-dependent pathogenicity features. *Molecular Plant Pathology, 13*(8), 877-886. doi:10.1111/j.1364-3703.2012.00796.x

Mallory, A. C., Ely, L., Smith, T. H., Marathe, R., Anandalakshmi, R., Fagard, M., . . . Vance, V. B. (2001). HC-Pro Suppression of Transgene Silencing Eliminates the Small RNAs but Not Transgene Methylation or the Mobile Signal. *The Plant Cell, 13*(3), 571-584.

Mandadi, K. K., & Scholthof, K. B. G. (2013). Plant immune responses against viruses: how does a virus cause disease? *Plant Cell, 25*(5), 1489-1505. doi:10.1105/tpc.113.111658

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics, 26*(19), 2462-2463. doi:10.1093/bioinformatics/btq467

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes: Virus Evolution.

Martínez de Alba, A. E., Elvira-Matelot, E., & Vaucheret, H. (2013). Review: Gene silencing in plants: A diversity of pathways. *BBA - Gene Regulatory Mechanisms, 1829*, 1300-1308. doi:10.1016/j.bbagrm.2013.10.005

Martínez, F., & Daròs, J.-A. (2014). Tobacco etch virus Protein P1 Traffics to the Nucleolus and Associates with the Host 60S Ribosomal Subunits during Infection. *Journal of Virology, 88*(18), 10725-10737. doi:10.1128/JVI.00928-14

Mathur, C., & Savithri, H. S. (2012). Novel ATPase activity of the polyprotein intermediate, Viral Protein genome-linked-Nuclear Inclusion-a protease, of Pepper vein banding potyvirus. *Biochemical and Biophysical Research Communications, 427*(1), 113-118. doi:10.1016/j.bbrc.2012.09.020

McKern, N. M., Ward, C. W., & Shukla, D. D. (1992). Strains of bean common mosaic virus consist of at least two distinct potyviruses. *Potyvirus Taxonomy*, 407-414. doi:10.1007/978-3-7091-6920-9_43

Merits, Guo, D., Järvekülg, L., & Saarma, M. (1999). Biochemical and Genetic Evidence for Interactions between Potato A Potyvirus-Encoded Proteins P1 and P3 and Proteins of the Putative Replication Complex. *Virology, 263*(1), 15-22. doi:10.1006/viro.1999.9926

Merits, Guo, D., & Saarma, M. (1998). VPg, coat protein and five non-structural proteins of potato A potyvirus bind RNA in a sequence-unspecific manner. *Journal of General Virology, 79*(12), 3123-3127. doi:10.1099/0022-1317-79-12-3123

Moghal, S. M., & Francki, R. I. B. (1976). Towards a system for the identification and classification of potyviruses: I. Serology and amino acid composition of six distinct viruses. *Virology, 73*(2), 350-362. doi:10.1016/0042-6822(76)90396-2

Mönttinen, H. A. M., Ravantti, J. J., Stuart, D. I., & Poranen, M. M. (2014). Automated Structural Comparisons Clarify the Phylogeny of the Right-Hand-Shaped Polymerases. *Molecular Biology and Evolution, 31*(10), 2741-2752. doi:10.1093/molbev/msu219

Nakahara, K. S., Masuta, C., Yamada, S., Shimura, H., Kashihara, Y., Wada, T. S., . . . Uyeda, I. (2012). Tobacco calmodulin-like protein provides secondary defense by binding to and directing degradation of virus RNA silencing suppressors. *Proceedings of the National Academy of Sciences of the United States of America, 109*(25), 10113-10118. doi:10.1073/pnas.1201628109

Nelson, S. C. (2008). Dasheen Mosaic of Edible and Ornamental Aroids. *Plant Disease, 44*, 1-9.

Nicholson, B. L., & White, K. A. (2011). 3′ Cap-independent translation enhancers of positive-strand RNA plant viruses. *Current Opinion in Virology, 1*(5), 373-380. doi:10.1016/j.coviro.2011.10.002

Nicolaisen, M., Johansen, E., Poulsen, G. B., & Borkhardt, B. (1992). The 5′ untranslated region from pea seedborne mosaic potyvirus RNA as a translational enhancer in pea and tobacco protoplasts. *FEBS Letters, 303*(2–3), 169-172. doi:10.1016/0014-5793(92)80511-E

Odoux, E., & Grisoni, M. (Eds.). (2010). *Vanilla (Medicinal and Aromatic Plants - Industrial)*. Florida, USA: CRC Press, Taylor & Francis Group.

Ohshima, K. (2013). Studies on the molecular evolution of potyviruses. *Journal of General Plant Pathology, 79*(6), 448-452. doi:10.1007/s10327-013-0488-9

Oruetxebarria, I., Guo, D., Merits, A., Mäkinen, K., Saarma, M., & Valkonen, J. P. T. (2001). Identification of the genome-linked protein in virions of Potato virus A, with comparison to other members in genus Potyvirus. *Virus Research, 73*(2), 103-112. doi:10.1016/S0168-1702(00)00216-1

Padmanabhan, C., Zhang, X., & Jin, H. (2009). Host small RNAs are big contributors to plant innate immunity. *Current Opinion in Plant Biology, 12*, 465-472. doi:10.1016/j.pbi.2009.06.005

Pappu, S. S., Brand, R., Pappu, H. R., Rybicki, E. P., Gough, K. H., Frenkel, M. J., & Niblett, C. L. (1993). A polymerase chain reaction method adapted for selective

amplification and cloning of 3' sequences of potyviral genomes: application to dasheen mosaic virus. *Journal of Virological methods, 41*(1), 9-20. doi:10.1016/0166-0934(93)90158-N

Park, C. Y., Lee, M. A., Park, E. H., Baek, D. S., Lee, S. H., & Kim, J.-S. (2015). First Report of Cucumber mosaic virus Infection of Peanut (Arachis hypogaea) in South Korea. *Plant Disease, 99*(5), 733-733. doi:10.1094/PDIS-10-14-1030-PDN

Pasin, F., Simón-Mateo, C., & García, J. A. (2014). The Hypervariable Amino-Terminus of P1 Protease Modulates Potyviral Replication and Host Defense Responses. *Public Library of Science Pathogens, 10*(3), e1003985. doi:10.1371/journal.ppat.1003985

Pearson, M. N., Bussell, W. T., & Scheffer, J. J. C. (1998). New plant disease record in New Zealand: Dasheen mosaic potyvirus infecting taro (Colocasia esculenta Schott). *New Zealand Journal of Crop and Horticultural Science, 26*(1), 69-70. doi:10.1080/01140671.1998.9514042

Pearson, M. N., & Grisoni, M. (2002). Records of plant viruses for the Pacific Islands. *Australasian Plant Pathology, 31*(1), 15-26. doi:10.1071/AP01049

Pearson, M. N., Jackson, G. V. H., Pone, S. P., & Howitt, R. L. J. (1993). Vanilla viruses in the South Pacific. . *Plant Molecular Biology, 42*, 127-131. doi:10.111/j.1365-3059.1993.tb01479.x

Plisson, C., Drucker, M., Blanc, S., German-Retana, S., Le Gall, O., Thomas, D., & Bron, P. (2003). Structural Characterization of HC-Pro, a Plant Virus Multifunctional Protein. *Journal of Biological Chemistry, 278*(26), 23753-23761. doi:10.1074/jbc.M302512200

Rajamäki, M.-L., Kelloniemi, J., Alminaite, A., Kekarainen, T., Rabenstein, F., & Valkonen, J. P. T. (2005). A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology, 342*(1), 88-101. doi:10.1016/j.virol.2005.07.019

Rajamäki, M.-L., Streng, J., & Valkonen, J. P. T. (2014). Silencing suppressor protein VPg of a potyvirus interacts with the plant silencing-related protein SGS3. *Molecular Plant-Microbe Interactions: MPMI, 27*(11), 1199-1210. doi:10.1094/MPMI-04-14-0109-R

Rakitina, D. V., Kantidze, O. L., Leshchiner, A. D., Solovyev, A. G., Novikov, V. K., Morozov, S. Y., & Kalinina, N. O. (2005). Coat proteins of two filamentous plant viruses display NTPase activity in vitro. *FEBS Letters, 579*(22), 4955-4960. doi:1016/j.febslet.2005.07.083

Ram, R., Joshi, A., Verma, N., Kulshrestha, S., Raikhy, G., Hallan, V., & Zaidi, A. A. (2003). First report of Dasheen mosaic virus infecting four ornamental aroids in India. *Plant Pathology, 52*(3), 411. doi:10.1046/j.1365-3059.2003.00836.x

Rana, G. L., Vovlas, C., & Zettler, F. W. (1983). Manual Transmission of Dasheen Mosaic Virus from Richardia to Nonaraceous Hosts. *Plant Disease, 76*(0191-2917), 891. doi:10.1094/PD-67-1121

Rantalainen, K. I., Eskelin, K., Tompa, P., & Mäkinen, K. (2011). Structural Flexibility Allows the Functional Diversity of Potyvirus Genome-Linked Protein VPg. *Journal of Virology, 85*(5), 2449-2457. doi:10.1128/JVI.02051-10

Ravelonandro, M., Peyruchaud, O., Garrigue, L., de Marcillac, G., & Dunez, J. (1993). Immunodetection of the plum pox virus helper component in infected plants and expression of its gene in transgenic plants. *Archives of Virology, 130*(3), 251-268. doi:10.1007/BF01309658

Revill, P. A., Jackson, G. V. H., Hafner, G. J., Yang, I., Maino, M. K., Dowling, M. L., . . . Harding, R. M. (2005). Incidence and distribution of viruses of Taro (Colocasia esculenta) in Pacific Island countries. *Australasian Plant Pathology, 34*(3), 327. doi:10.1071/AP05032

Reyes, G., Ramsell, J. N. E., Nyman, M., & Kvarnheden, A. (2009). Sequence characterization of Dasheen mosaic virus isolates from cocoyam in Nicaragua. *Archives of Virology, 154*(1), 159. doi:10.1007/s00705-008-0257-9

Riechmann, J. L., Cervera, M. T., & García, J. A. (1995). Processing of the plum pox virus polyprotein at the P3-6K1 junction is not required for virus viability. *Journal of General Virology, 76*(4), 951-956. doi:10.1099/0022-1317-76-4-951

Riechmann, J. L., Laín, S., & García, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology, 73*(1), 1-16. doi:10.1099/0022-1317-73-1-1

Roberts, R., Browning, K. S., Mayberry, L. K., & Rakotondrafara, A. M. (2017). The Triticum mosaic virus 5′ leader binds to both eIF4G and eIFiso4G for translation. *Public Library of Science ONE, 12*(1). doi:10.1371/journal.pone.0169602

Rodríguez-Cerezo, E., Ammar, E. D., Pirone, T. P., & Shaw, J. G. (1993). Association of the non-structural P3 viral protein with cylindrical inclusions in potyvirus-infected cells. *Journal of General Virology, 74*(9), 1945-1949. doi:doi:10.1099/0022-1317-74-9-1945

Rohožková, J., & Navrátil, M. (2011). P1 peptidase – a mysterious protein of family Potyviridae. *Journal of Biosciences, 36*(1), 189-200. doi:10.1007/s12038-011-9020-6

Roossinck, M. J. (2013). Plant Virus Ecology. *Public Library of Science Pathogens, 9*(5). doi:10.1371/journal.ppat.1003304

Roossinck, M. J., & García-Arenal, F. (2015). Ecosystem simplification, biodiversity loss and plant virus emergence. *Current Opinion in Virology, 10*(0), 56-62. doi:10.1016/j.coviro.2015.01.005

Rybicki, E. (2015). A Top Ten list for economically important plant viruses. *Archives of Virology, 160*(1), 17-20. doi:10.1007/s00705-014-2295-9

Sáenz, P., Salvador, B., Simón-Mateo, C., Kasschau, K. D., Carrington, J. C., & García, J. A. (2002). Host-specific involvement of the HC protein in the long-distance movement of potyviruses. *Journal of Virology, 76*(4), 1922-1931. doi:10.1128/JVI.76.4.1922-1931.2002

Sahana, N., Kaur, H., Tena, F., Jain, R. K., Palukaitis, P., Canto, T., & Praveen, S. (2012). Inhibition of the Host Proteasome Facilitates Papaya Ringspot Virus Accumulation and Proteosomal Catalytic Activity Is Modulated by Viral Factor HcPro. *Public Library of Science ONE, 7*(12). doi:10.1371/journal.pone.0052546

Sallie, R. (2005). Replicative homeostasis II: Influence of polymerase fidelity on RNA virus quasispecies biology: Implications for immune recognition, viral autoimmunity and other "virus receptor" diseases. *Virology Journal, 2*. doi:10.1186/1743-422X-2-70

Salvador, B., SaÉNz, P., YangÜEz, E., Quiot, J. B., Quiot, L., Delgadillo, M. O., . . . SimÓN-Mateo, C. (2008). Host-specific effect of P1 exchange between two potyviruses. *Molecular Plant Pathology, 9*(2), 147-155. doi:10.1111/j.1364-3703.2007.00450.x

Sankaran, N. (2016). The RNA World at Thirty: A Look Back with its Author. *Journal of Molecular Evolution, 83*(5-6), 169-175. doi:10.1007/s00239-016-9767-3

Schaad, M. C., Lellis, A. D., & Carrington, J. C. (1997). VPg of tobacco etch potyvirus is a host genotype-specific determinant for long-distance movement. *Journal of Virology, 71*(11), 8624-8631.

Scholthof, K.-B. G., Adkins, S., Czosnek, H., Palukaitis, P., Jacquot, E., Hohn, T., . . . Foster, G. D. (2011). Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology, 12*(9), 938-954. doi:10.1111/j.1364-3703.2011.00752.x

Schöttler, M. A., Tóth, S. Z., Boulouis, A., & Kahlau, S. (2015). Photosynthetic complex stoichiometry dynamics in higher plants: biogenesis, function, and turnover of ATP synthase and the cytochrome b6f complex. *Journal Of Experimental Botany, 66*(9), 2373-2400. doi:10.1093/jxb/eru495

Shan, H., Pasin, F., Valli, A., Castillo, C., Rajulu, C., Carbonell, A., . . . Rodamilans, B. (2015). The Potyviridae P1a leader protease contributes to host range specificity. *Virology, 476*, 264-270. doi:10.1016/j.virol.2014.12.013

Shi, Y., Chen, J., Hong, X., Chen, J., & Adams, M. J. (2007). A potyvirus P1 protein interacts with the Rieske Fe/S protein of its host. *Molecular Plant Pathology, 8*(6), 785-790. doi:10.1111/j.1364-3703.2007.00426.x

Shiboleth, Y. M., Haronsky, E., Leibman, D., Arazi, T., Wassenegger, M., Whitham, S. A., . . . Gal-On, A. (2007). The Conserved FRNK Box in HC-Pro, a Plant Viral Suppressor of Gene Silencing, Is Required for Small RNA Binding and Mediates Symptom Development. *Journal of Virology, 81*(23), 13135-13148. doi:10.1128/JVI.01031-07

Shimura, H., & Pantaleo, V. (2011). Viral induction and suppression of RNA silencing in plants [Article]. *BBA - Gene Regulatory Mechanisms, 1809*, 601-612. doi:10.1016/j.bbagrm.2011.04.005

Sorel, M., Garcia, J. A., & German-Retana, S. (2014). The Potyviridae Cylindrical Inclusion Helicase: A Key Multipartner and Multifunctional Protein. *Molecular Plant-Microbe Interactions, 27*(3), 215-226. doi:10.1094/MPMI-11-13-0333-CR

Stenger, D. C., Hein, G. L., Gildow, F. E., Horken, K. M., & French, R. (2005). Plant Virus HC-Pro Is a Determinant of Eriophyid Mite Transmission. *Journal of Virology, 79*(14), 9054-9061. doi:10.1128/JVI.79.14.9054-9061.2005

Strange, R. N., & Scott, P. R. (2005). Plant Disease: A Threat to Global Food Security. *Annual Review of Phytopathology, 43*(1), 83-116. doi:10.1146/annurev.phyto.43.113004.133839

Sun, P., Austin, B. P., Tözsér, J., & Waugh, D. S. (2010). Structural determinants of tobacco vein mottling virus protease substrate specificity. *Protein Science : A Publication of the Protein Society, 19*(11), 2240-2251. doi:10.1002/pro.506

Sztuba, S. J., Urbanowicz, A., Figlerowicz, M., & Bujarski, J. J. (2011). RNA-RNA Recombination in Plant Virus Replication and Evolution. *Annual Review Phytopathology, 49*(1), 415-443 doi:10.1146/annurev-phyto-072910-095351

Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M., & Bujarski, J. J. (2011). RNA-RNA Recombination in Plant Virus Replication and Evolution. *Annu Rev Phytopathol, 49*(1), 415-443. doi:doi:10.1146/annurev-phyto-072910-095351

Tatineni, S., Qu, F., Li, R., Jack Morris, T., & French, R. (2012). Triticum mosaic poacevirus enlists P1 rather than HC-Pro to suppress RNA silencing-mediated host defense. *Virology, 433*(1), 104-115. doi:10.1016/j.virol.2012.07.016

Tavert-Roudet, G., Abdul-Razzak, A., Doublet, B., Walter, J., Delaunay, T., German-Retana, S., . . . Candresse, T. (2012). The C terminus of lettuce mosaic potyvirus cylindrical inclusion helicase interacts with the viral VPg and with lettuce translation eukaryotic initiation factor 4E. *Journal of General Virology, 93*(1), 184-193. doi:10.1099/vir.0.035881-0

Tena Fernández, F., González, I., Doblas, P., Rodríguez, C., Sahana, N., Kaur, H., . . . Canto, T. (2013). The influence of cis-acting P1 protein and translational elements on the expression of Potato virus Y helper-component proteinase (HCPro) in heterologous systems and its suppression of silencing activity. *Molecular Plant Pathology, 14*(5), 530-541. doi:10.1111/mpp.12025

Torres-Barceló, C., Daròs, J.-A., & Elena, S. F. (2010). Compensatory Molecular Evolution of HC-Pro, an RNA-Silencing Suppressor from a Plant RNA Virus. *Molecular Biology and Evolution, 27*(3), 543-551. doi:10.1093/molbev/msp272

Valli, García, J. A., & López‐Moya, J. J. (2015). Potyviridae. *eLS, John Wiley & Sons Ltd, Chichester*. doi:10.1002/9780470015902.a0000755.pub3

Valli, A., Gallo, A., Calvo, M., Pérez, J. d. J., & García, J. A. (2014). A Novel Role of the Potyviral Helper Component Proteinase Contributes To Enhance the Yield of Viral Particles. *Journal of Virology, 88*(17), 9808-9818. doi:10.1128/JVI.01010-14

Valli, A., López-Moya, J. J., & García, J. A. (2007). Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *Journal of General Virology, 88*(3), 1016-1028. doi:10.1099/vir.0.82402-0

Vazquez, F., Legrand, S., & Windels, D. (2010). Review: The biosynthetic pathways and biological scopes of plant small RNAs. *Trends in Plant Science, 15*, 337-345. doi:10.1016/j.tplants.2010.04.001

Verchot, J. (1995). *Role of the potyvirus P1 protein in polyprotein processing and viral infectivity* (Doctoral Dissertation). Texas A&M University, Ann Arbor. Retrieved fromhttp://ezproxy.aut.ac.nz/login?url=http://search.proquest.com/docview/304286937?accountid=8440. Available from ProQuest Dissertations & Theses Global database.

Verchot, J., & Carrington, J. C. (1995a). Debilitation of plant potyvirus infectivity by P1 proteinase-inactivating mutations and restoration by second-site modifications. *Journal of Virology, 69*(3), 1582-1590.

Verchot, J., & Carrington, J. C. (1995b). Evidence that the potyvirus P1 proteinase functions in trans as an accessory factor for genome amplification. *Journal of Virology, 69*(6), 3668-3674.

Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nature Reviews Molecular Cell Biology, 10*(6), 385-397. doi:10.1038/nrm2688

Vijayapalani, P., Maeshima, M., Nagasaki-Takekuchi, N., & Miller, W. A. (2012). Interaction of the Trans-Frame Potyvirus Protein P3N-PIPO with Host Protein PCaP1 Facilitates Potyvirus Movement. *Public Library of Science Pathogens, 8*(4), e1002639. doi:10.1371/journal.ppat.1002639

Waltermann, A., & Maiss, E. (2006). Detection of 6K1 as a mature protein of 6 kDa in plum pox virus-infected Nicotiana benthamiana. *Journal of General Virology, 87*(8), 2381-2386. doi:doi:10.1099/vir.0.81873-0

Wang, Y. Y., & Pearson, M. N. (1992). Some Characteristics of Potyvirus Isolates from Vanilla tahitensis in French Polynesia and the Cook Islands. *Journal of Phytopathology, 135*(1), 71-76. doi:10.1111/j.1439-0434.1992.tb01252.x

Ward, C. W., McKern, N. M., Frenkel, M. J., & Shukla, D. D. (1992). Sequence data as the major criterion for potyvirus classification [Ward1992]. In O. W. Barnett (Ed.), *Potyvirus Taxonomy* (pp. 283-297). Vienna: Springer Vienna. doi:10.1007/978-3-7091-6920-9_29

Wei, Y., Zhang, C., Hong, J., Xiong, R., Kasschau, D., Zhuo, X., . . . Wang, A. (2010). Formation of complexes at Plasmodesmata for Potyvirus Intracellular Movement Is Mediated by the Viral Protein P3N-PIPO. *Public Library of Science Pathogens, 6*(6), e1000962. doi:10.1371/journalppat1000962

Wisler, G. C., & Zettler, F. W. (1987). Virus Infection of Vanilla and other orchids in French Polynesia. *Plant Molecular Biology, 71*. doi:10.1094/PD-71-1125

Wydro, M., Kozubek, E., & Lehmann, P. (2006). Optimization of transient Agrobacterium-mediated gene expression system in leaves of Nicotiana benthamiana. *Acta Biochimica Polonica, 53*(2), 289-298.

Wylie, S. J., & Jones, M. G. K. (2011). The complete genome sequence of a passion fruit woodiness virus isolate from australia determined using deep sequencing, and its relationship to other potyviruses. *Archives of Virology, 156*(3), 479-482. doi:10.1007/s00705-010-0845-3

Xie, Z., & Qi, X. (2008). Review: Diverse small RNA-directed silencing pathways in plants. *BBA - Gene Regulatory Mechanisms, 1779*, 720-724. doi:10.1016/j.bbagrm.2008.02.009

Xiong, R., & Wang, A. (2013). SCE1, the SUMO-Conjugating Enzyme in Plants That Interacts with NIb, the RNA-Dependent RNA Polymerase of Turnip Mosaic Virus, Is Required for Viral Infection. *Journal of Virology, 87*(8), 4704-4715. doi:10.1128/JVI.02828-12

Xu, D., Liu, H.-Y., Li, F., & Li, R. (2011). Complete genome sequence of Celery mosaic virus and its relationship to other members of the genus Potyvirus. *Archives of Virology, 156*(5), 917-920. doi:10.1007/s00705-011-0951-x

Xu, D., Wang, M., Kinard, G., & Li, R. (2012). Complete genome sequence of two isolates of pokeweed mosaic virus and its relationship to other members of the genus Potyvirus. *Archives of Virology, 157*(10), 2023-2026. doi:10.1007/s00705-012-1329-4

Young, B. A., Stenger, D. C., Qu, F., Morris, T. J., Tatineni, S., & French, R. (2012). Tritimovirus P1 functions as a suppressor of RNA silencing and an enhancer of disease symptoms. *Virus Research, 163*(2), 672-677. doi:10.1016/j.virusres.2011.12.019

Zettler, F. W., Foxe, M. J., Hartman, R. D., Edwardson, J. R., & Christie, R. G. (1970). Filamentous Viruses Infecting Dasheen and other Araceous Plants. *Phytopathology, 60*, 983-987. doi:10.1094/Phyto-60-983

Zettler, F. W., Tsai, J. H., Faan, H. C., & Ke, L. (1987). Dasheen Mosaic Virus Infecting Taro in People's Republic of China. *Plant Disease, 71*, 837. doi:10.1094/PD-71-0837

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

Zheng, L., Wayper, P. J., Gibbs, A. J., Fourment, M., Rodoni, B. C., & Gibbs, M. J. (2008). Accumulating Variation at Conserved Sites in Potyvirus Genomes Is Driven by Species Discovery and Affects Degenerate Primer Design. *Public Library of Science, 3*(2). doi:10.1371/journal.pone.0001586