Temporal and spatial analyses of *Dasheen mosaic virus*

Wee Leong (Joe) Chang

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

2012

School of Applied Science

Table of Contents

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.

Wee Leong (Joe) Chang

Acknowledgement

This thesis arose in part out of years of research that has been done since I started my Masters. During this time, I have worked with a great number of people whose contribution in assorted ways to the research and the making of this thesis deserve special mention. It is a pleasure to convey my gratitude to them all.

In the first place I would like to record my gratitude to Dr. Colleen Higgins for her supervision, advice, and guidance during every stage of this research, as well as giving me extraordinary experiences throughout the work. Above all, and the most needed, she gave me unflinching encouragement and support in various ways. Her truly scientific intuition has made her a constant oasis of ideas and passion in sciences, which are exceptionally inspiring and have enriched my growth as a person, a researcher and a scientist. I am indebted to her more than she knows.

I gratefully acknowledge Dr. Mike Pearson for his advice, supervision, and contribution, which made him a backbone of this research and so to this thesis. His involvement, with his advice, has nourished my intellectual maturity, which I will benefit from for a long time to come. Much gratitude must also be given to Karmun Choi for her support and advice throughout the whole research.

Many thanks to the Molecular Genetics Research Team; Gardette Valmonte, Smriti Nair and Subuhi Khan. I am much indebted for the valuable advice during our research discussions and the fun times we had in the laboratory.

Abbreviations

bp	base pair
BP	Before present day
^{o}C	degrees Celsius
g	gram(s)
Kb	kilobase(s)
L	litre
M	molar
mg	milligram(s)
mL	millilitre
µL	microliter
µg	microgram
nmol	nanomoles
pmol	picomoles
3’ UTR	3’ untranslated region
6K1 and 6K2	6 kDa proteins 1 and 2
aa	Amino acid
ACIAR	Australian Centre for International and Agricultural Research
BCMNV	*Bean common mosaic necrosis virus*
BaMMV	*Barley mild mosaic virus*
cDNA	Complementary DNA
CI	Cylindrical inclusion protein

CIRE	Cap-independent regulatory element
CMV	*Cucumber mosaic virus*
CP	Coat protein
DsMV	*Dasheen mosaic virus*
DTT	Dithiothreitol
eIF4E, eIF (iso) 4E	Eukaryotic translation initiation factors 4E, (iso) 4E
E. coli	*Escherichia coli*
EtoH	Ethanol
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
HC-Pro	Helper component proteinase
HR	Hypersensitive response
HDP	High density probability
IRES	Internal ribosome entry site
kDa	Kilodalton
MP	Movement protein
ML	Maximum Likelihood
NIa	Nuclear inclusion protein a
NIa-Pro	C-terminal proteinase domain of NIa
NIb	Nuclear inclusion protein b, RNA polymerase
nt	Nucleotide
NTR	Non-translated region
ORF	Open reading frame
ONMV	*Oat necrotic mottle virus*
P1	P1 protein

P3	P3 protein
PCR	Polymerase chain reaction
PepMoV	*Pepper mottle virus*
PLDMV	*Papaya leaf distortion mosaic*
PPV	*Plum pox virus*
PSbMV	*Pea seed-borne mosaic virus*
PVA	*Potato virus A*
PVV	*Potato virus V*
PVY	*Potato virus Y*
RdRp	RNA-dependent RNA polymerase
RYGM	*Ryegrass mosaic virus*
RNP	Ribonucleoprotein
RNA	Ribonucleic Acid
RT-PCR	Reverse transcription polymerase chain reaction
SE	Sieve element
TEV	*Tobacco etch virus*
TMV	*Tobacco mosaic virus*
TuMV	*Turnip mosaic virus*
TVMV	*Tobacco vein mottling virus*
TBE	Tris-borate EDTA
WYMV	*Wheat yellow mosaic virus*
YMV	*Yam mosaic virus*
VPg	Viral genome-linked protein
ZYMV	*Zucchini yellow mosaic virus*

List of figures

Chapter 1

Chapter 2

Chapter 3

List of tables

Chapter 1

Chapter 2

Chapter 3

Abstract

Exploring the genetic diversity and evolutionary history of plant viruses is critical to understanding their ecology and epidemiology. Like many other plant RNA viruses, *Dasheen mosaic virus* (DsMV) is an important and conspicuous viral disease of ornamental and edible aroids throughout the South Pacific and worldwide, but its population diversity and variability are poorly understood. To further investigate this virus, phylogenetic and population genetics based methods were used to investigate the temporal and spatial dynamics of the evolutionary mechanism and genetic variability among the DsMV isolates. A selected region of the coat protein (CP) gene was amplified and sequenced to infer genetic relationships between viral isolates at the temporal and spatial scales. This study demonstrated that genetic variation occurs between the DsMV isolates. The population structure of DsMV consisted of the consensus sequence and a pool of mutants that are not identical but are closely related to the consensus sequence, and it coincides with the quasispecies concept described for many RNA viruses. The quasispecies-like nature of the DsMV population suggested that the virus is capable of rapid evolution and adaptation in response to changing ecological factors and agricultural practices. Analysis of DsMV isolates on a temporal scale suggested the role of stochastic and selection-fitness levels are the key determinants in the dynamics of plant virus population genetics and evolution. In contrast, spatial analysis suggested that diversification and spread of DsMV have been concomitant with an extension of human migration and taro/tannia cultivation in the South Pacific islands. The combined actions of genetic drift and selection pressure have continually remoulded this diversity. Thereby, creating a geographic mosaic in the degrees of diversity found within and between geographic regions.

Keywords: *Dasheen mosaic virus*, temporal evolution, spatial evolution, variation, quasispecies

Chapter 1

General Introduction

1.1 Introduction

Over the years, agriculture across the world has been compromised by a succession of devastating epidemics caused by new viruses (Sardanyes & Elena, 2011). These infections either spilled over from reservoir species or were caused by new variants of classic viruses that acquired new virulence factors or changed their epidemiological patterns (Sardanyes & Elena, 2011). The social and economic burden of emerging infectious diseases is both substantial and increasing. However, relatively little is known of the evolutionary mechanism that increases the probability of pathogen emergence. A major goal of evolutionary research is to create a framework for predictability that will enable the use of cost effective measures to prevent the future emergence of potentially devastating pathogens, instead of relying on intervention and control. Viral emergence is often associated with ecological change or with agronomical practices bringing together reservoirs and crop species (Holmes, 2009). The complete picture is, however, much more complex, and results from an evolutionary process in which the main players are ecological factors, viruses' genetic plasticity, and host factors required for virus replication, all mixed with a good measure of stochasticity (Lauring & Andino, 2010).

Using *Dasheen mosaic virus* (DsMV) as the model virus, this thesis examined the population structure and genetic diversity of this plant RNA virus in *Colocasia sp*. and *Xanthosoma sp*. Through rigorous analysis of both the temporal/spatial dynamics and phylogenetic structure of DsMV, the genetic variability among DsMV isolates was studied to understand its evolutionary lineage. Change in genetic diversity over time (temporal analysis) was assessed for glasshouse isolates collected from infected taro (*Colocasia sp.).* Viral sequences isolated from this study were then compared with public DsMV sequences deposited in the National Centre for Biotechnology Information (NCBI) database. Spatial analysis between the genetic distance and geographical distances of the South Pacific DsMV isolates were also carried out from infected *Colocasia sp.* and *Xanthosoma sp.,* to infer patterns of spread of the virus across the landscape over space and time.

The focus of this research work is split into four chapters to describe the relationship between the processes driving the evolution and genetic variation among DsMV

isolates. First, Chapter One provides an overview of the taxonomy and molecular characterization of DsMV, the host crop involved, and the current state of knowledge in the relevant topics of evolutionary virology. Chapter Two focuses on the mechanism behind genetic variation in plant RNA viruses with the major goal being to assess the variation arising from short-term evolution on a temporal scale. In Chapter Three, the same techniques developed in Chapter Two were applied to evaluate the nature of long-term evolutionary processes in plant RNA viruses. Lastly, Chapter Four reviews the findings of this thesis and a concluding remark with a general discussion is highlighted.

1.2 Definition of a Virus

Viruses are sub-microscopic, obligate intracellular organisms that are not functionally active outside their host cells. Like all biological entities, viruses possess genes, reproduce themselves and have the ability to adapt to changing environments. Roger Hall (2002) defined a virus as follows:

> "A set of one or more nucleic acid template molecules normally encased in a protective coat or coats of protein or lipoprotein that is able to organize its own replication only within suitable host cells. It can usually be horizontally transmitted between host protein-synthesizing machinery, organized from pools of the required materials rather than by binary fission, located at sites that are not separated from the host cell contents by a lipoprotein bilayer membrane, and continually gives rise to variants through various kinds of change in the viral nucleic acid''.

In contrast to other microbes and multicellular organisms, the origin and evolution of viruses is mostly unknown. Tracing the origins of viruses is difficult because they don't leave fossils and because of their replication process within the cells they invade (Racaniello, 2001). Some viruses even have the ability to stitch their own genes into those of the cells they infect. This means that studying their ancestry requires untangling it from the history of their hosts and other organisms (Rambaut et al., 2008). What makes the process even more complicated is that viruses don't just infect humans; they infect all forms of cellular life; bacteria, fungi, animals (including insects) and plants.

For example, picornavirus are among the most diverse (more than 200 serotypes) and 'oldest' known viruses (temple record from Egypt ca. 1400 B.C.), which cause disease in humans, animals, insects and plants (Racaniello, 2001). In humans, Picornaviruses are separated into 12 distinct genera. Contained within the *Picornavirus* genus are many organisms of importance as vertebrate and human pathogens, such as foot and mouth disease virus (Semler & Wimmer, 2002). The negative impacts of viruses are not only limited to direct loss of human life, livestock and plants but indirectly, many lives are also lost through hunger and starvation since viruses can cause crop losses worth up to 500 billion dollars (USD) each year to crops (Fermin-Munoz, Meng, Mazumdar-Leighton, & Gubba, 2001).

1.3 Family *Potyviridae*

In 1988, the International Taxonomy of Viruses Committee (ICTV) adopted the classification of phytopathogenic viruses into 34 taxonomic groups (Gnutova & Tolkach, 1998). The greatest number of viruses (more than 180 true and potential representatives or 36% of all known plant viruses) was included in the family Potyviridae. They infect more than 2000 plant species of 550 genera and 81 families (ICTV, 2010).

The family Potyviridae, along with the animal/human infecting Picornaviridae, Sequiviridae, Comoviridae and Caliciviridae families, form the “picorna-like group” of positive-sense, single-stranded RNA viruses. These virus families show similarities in their genome organization and expression strategy (Goldbach, 1986). Within this group, the family Potyviridae contains seven established genera: *Potyvirus*, *Macluravirus*, *Bymovirus*, *Brambyvirus*, *Rymovirus*, *Tritimovirus*, and *Ipomovirus* (ICTV, 2010). potyviruses, brambyvirus and macluraviruses are aphid-transmitted, whereas mites transmit rymoviruses and tritimoviruses. Whiteflies and fungi transmit Ipomoviruses and Bymoviruses, respectively. The genomes are usually monopartite, with the exception of the bipartite genome of Bymoviruses. The virions are flexuous filaments without an envelope and have a diameter of 11 to 15 nm and a length of 650 to 900 nm (250–300 nm for the Bymoviruses and 500–600 nm for the Ipomoviruses) (van Regenmortel, 2000).

1.4 Genus *Potyvirus*

The genus *Potyvirus* is the second largest known virus genus, with 143 species currently assigned to it by the ICTV (2010). Potyviruses infect plants and cause serious diseases in crops worldwide (Fuji, Inoue, Yamamoto, Furuya, & Naito, 2003). Aphids transmit viruses of this genus non-persistently (i.e. does not replicate in the insect vector). Many species of aphids can transmit potyviruses; for example *Myzus persicae, M. ornatus, Aphis rhanni, A. craccivora, A. nasturtii, A. fabae, Macrosiphum euphorbiae, Acyrthosyphon pisum etc*. (Dombrovsky, Huet, Chejanovsky, & Raccah, 2005). For example, the most well studied, widespread and heavily damaging viruses that cause disease are bean yellow mosaic virus (BYMV), bean common mosaic virus (BCMV), potato virus A (PVA), potato virus Y (PVY), soybean mosaic virus (SMV), turnip mosaic virus (TMV), and plum pox virus (PPV) (Gibbs, Ohshima, Philips, & Gibbs, 2008).

1.5 Genome organization

The potyviral genome consists of a single-stranded (ss) positive-sense RNA molecule of about 10 kilobases which is encapsidated by approximately 2000 copies of the coat protein (CP) (Matthews, 1991). The genome contains a single open reading frame (ORF), flanked by an untranslated region (UTR) at both 5' and 3' ends as shown in Figure 1.1. At the 3' end there is a polyadenylated (poly (A) tail and covalently bonded to the 5' end is the viral genome-linked protein, VPg, acting like a cap structure (Jiang & Laliberte, 2011). The ORF encodes a large polyprotein (3000 to 3500 amino acids) that is co- and/or post-translationally cleaved by three viral-encoded proteinases (P1, HC-Pro and NIa-Pro). The P1 and HC-Pro mediate their own cleavage from the polyprotein (Carrington, Freed, & Sanders, 1989) whereas the NIa-Pro is responsible for the cleavage of the C-terminal two-thirds of the polyprotein (Carrington, Cary, Parks, & Dougherty, 1989). As a result of the proteolytical events, ten mature proteins are produced as shown in Figure 1.1 with their relative functions (Huang, Wei, Laliberte, & Wang, 2010). In a recent publication, a small ORF encoding a ~25 kDa protein, P3N-PIPO, was identified embedded in the P3 coding region, thus making a total of eleven mature protein produced by the proteolytical events (Huang et al., 2010).

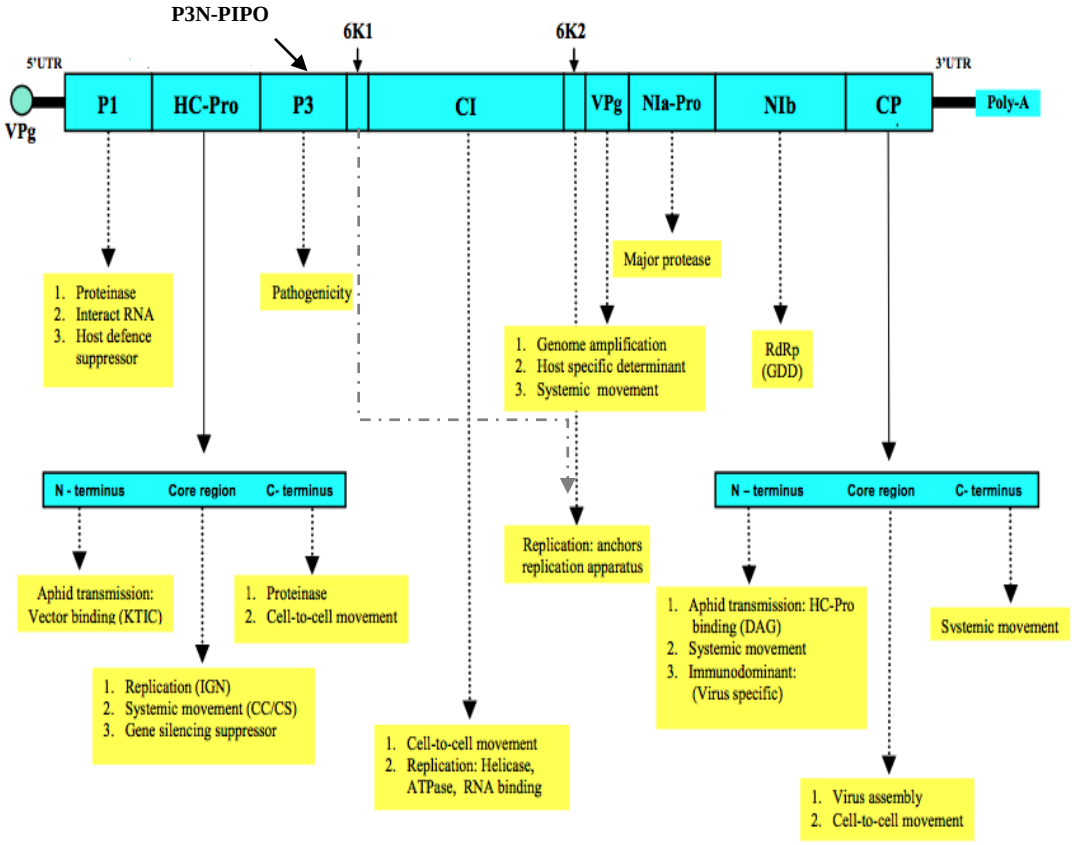


Figure 1.1: Genome organization of the genus Potyvirus with the properties and function of each gene. The genome contains one open reading frame and names of the final protein products are indicated as boxes, separated by lines that indicate the putative cleavage sites of each protein (Ha, 2007).

1.6 Replication and movement of Potyviruses

The potyvirus life cycle consists of entry into cells via wounds (created either by a probing aphid, or through mechanical wounding), uncoating, translation, polyprotein processing, genome replication, virus particle assembly, and movement to other cells, plant parts, and other plants (Urcuqui-Inchima, Haenn & Bernardi, 2001). Similar to other positive-sense ssRNA viruses, potyviruses replicate through a complementary negative strand RNA molecule generated by the viral-encoded RNA-dependent RNA polymerase (RdRp); the nuclear inclusion protein b (NIb). The NIb protein acts as the viral replicase, by generating a negative sense copy of the positive stranded genome. This synthesis is primed by a uridylyated VPg protein that binds to the poly A tail of the viral genome. The RdRp then uses the negative sense strand as a template for synthesis

of a positive sense genome for the new cell. VPg also acts as the protein-primer of the positive strand RNA synthesis allowing host translation factors to translate and replicate its genome (Nagy, 2008).

Viral replication occurs in tight association with cellular membranes and various viral-encoded proteins are also involved (Nagy, 2008). Several potyvirus-encoded proteins are known to, or have been suggested to, participate in the formation of the potyvirus replication complex. The cylindrical inclusion protein (CI) is an RNA helicase and NTPase (Wiesenberg, Missbach, Kahlen, Schreder, & Carlberg, 1995). The viral genome-linked protein (VPg), processed from the N-proximal half of NIa by NIaPro, is covalently bound to the 5′-end of potyvirus RNA (Oruetxebarria et al., 2001). VPg is necessary for infectivity and may function as a primer for minus-strand RNA synthesis. Nuclear inclusion protein b (NIb) is an RNA-dependent RNA-polymerase. Domains of NIa, NIaPro and VPg can interact with NIb and, hence, may participate in the direction of NIb to the replication complex (Merits et al., 2002). The proteins putatively involved in the potyvirus replication complex (CI, VPg, NIaPro and NIb) bind RNA and interact with the non-structural proteins P1 and P3 (Merits et al., 2002). In addition, potyviral coat protein (CP) can interact with NIa and NIb. The P3 protein has no known enzymatic activity but contains a putative transmembrane domain and may be involved in anchoring the replication complex to membranes (Merits et al., 2002). The domains for the small 6K1 and 6K2 peptides in the polyprotein flank the CI domain. It has been hypothesized that both the 6K1 and 6K2 peptide may function as a membrane anchor for CI and VPg during Potyvirus replication (Merits et al., 2002). Subsequently, the replication complex may be released from the membrane by catalysis of cleavage at the polyprotein proteolytic site between 6K2 and VPg (Wei et al., 2010).

After replication and assembling, nascent virus particles will spread to neighbouring cells and further to other parts of the plant. The first process is called cell-to-cell movement the latter long-distance movement. Cell-to-cell movement of viruses occurs through plasmodesmata (PD), a specialized intercellular organelle, unique to the plant kingdom (Shukla, Ward, & Bunt, 1994). This is an active process mediated by virus-encoded protein(s) termed movement protein (MP). Potyviruses do not encode a dedicated MP, and movement functions have been allocated to several proteins. Of the eleven potyvrial proteins, CP, VPg, HC-Pro, CI and P3N- PIPO have been suggested to

have functions in intercellular transport (Dolja, Haldeman, Robertson, Dougherty, & Carrington, 1994). Accumulating evidence indicates that HC-Pro and VPg are essential in other aspects of the infection process such as viral genome replication or suppression of host defense RNA silencing, whereas CP, CI and P3N-PIPO are likely to be MPs of potyviruses (Wen, et al., 2010). The potyviral CP is a three-domain protein with variable N- and C-terminal regions exposed on the particle surface and a conserved core domain that interacts with viral RNA. Mutations in the CP- core domain result in defective cell-to-cell movement and virion assembly (Dolja et al., 1994), suggesting potyviruses likely move as virions. The potyvirus CI has also been suggested to play a role in cell-to-cell movement (Wen et al., 2010).

In the case of the newly found P3N-PIPO protein, it has been shown that mutation of the putative SMV PIPO domain impeded cell-to-cell movement (Wei, Huang, et al., 2010) and that P3N-PIPO direct CI to PD, anchoring the CI proteins therein and/or facilitating the deposition of CI through PD (Wei, Huang, et al., 2010). Based on the model suggested by Wei et al. (2010), cell-to-cell movement may be initiated when the recruitment of nascent virus particles by CI or self-interacting CI structures at membrane-bound sites of replication adjacent to chloroplasts. Then CI-virion complexes associate with the P3N-PIPO followed by trafficking to PD. CI structures accumulate from P3N-PIPO-anchored sites at PD, forming thread-like structures that might recruit additional virus particles for transport. Virus particles fed through the CI structures and PD to enter the adjacent cell may be facilitated by PD- traversing CI complexes.

The molecular details of these proteins involved in Potyvirus replication and movement strategy provides a basis for developing strategies for its control, and to obtain molecular insight into the Potyvirus lifecycle and infection process. A thorough understanding of these protein functions will open a new avenue of research both for understanding the replication mechanisms as well as for devising new diagnostic and control strategies.

1.7 Potyvirus species identification

Traditional criteria to discriminate between viral species and genera have been predominantly based on serology and biological criteria such as host range, cross-protection and symptomatology (Shukla et al., 1994). Although host range assays and symptomatology are valuable tools in the initial description of viruses, conflicting results have been reported, presumably due to differences in the environmental conditions affecting test plant growth or the use of different genotypes of plant species (Bos, 1999).

Serology using antibodies raised against the viral coat protein (CP) has been widely applied for detection of plant viruses. It has also been used to identify new viruses and to estimate how closely related they are as compared to previously described viruses (Gnutova & Tolkach, 1998). When detecting potyviruses, the use of serological methods has sometimes been hampered by the serological relationships between virus species (cross-reactions) or the lack of detection of all strains of a given virus (Spetz et al., 2003). Cross-reactions are probably due to the recognition by antibodies of the highly conserved core region of the CP (Figure 1.1). The carboxy-terminal, two-thirds of this protein is highly conserved, while the N–terminal third is variable between potyviral species (Dombrovsky et al., 2005). Antibodies raised that recognize the variable N-terminus of the CP are less likely to detect some strains of a virus, while antibodies raised against the conserved core may not be species specific.

Extensive studies of potyvirus sequences have shown that potyvirus species can be distinguished by sequence comparisons, and the process of recognizing a species is now in large part determined using this measure (Fuji, Yamamoto, Furuya, & Naito, 2003). Potyvirus isolates with 85% nucleotide sequence identity or more over the whole genome are usually considered to be from the same species (Fauquet, Mayo, Maniloff, Desselberger, & Ball, 2005). This was re-evaluated by Adams et al. in 2005, who suggested that a value of about 76% was most appropriate for the entire polyprotein sequence and that slightly different values might be used for the different regions of the potyvirus genome as shown in Table 1.1.

The most common method of plant virus detection and genetic variation studies are based on the polymerase chain reaction (PCR). PCR-based methods have the advantage

of being both extraordinarily sensitive and requiring reagents that are easily accessible and relatively inexpensive. The downside of PCR-based methods is that, in order to be reliable, the 3′ ends of the oligonucleotides primers employed must be conserved among all viral strains, in order to detect all strains or isolates. Current approaches in plant virus detection focus primarily on using degenerate primers targeted to the coat protein (CP) and the 3' UTR sequences (van Regenmortel, 2000). These regions of the RNA genome are easily obtained by RT-PCR, thus allowing rapid and sensitive detection of viruses.

Degenerate primers used for amplifying the CP 3'UTR regions have been designed from closely related groups of viruses (Chen & Adams, 2001b; Zheng, Rodoni, Gibbs, & Gibbs, 2010) and have been effective for detecting close relatives of known viruses (Zheng et al., 2008). However, they only reliably detect all the target viruses if the degenerate primers contain individual primers that complement and bind to all the virus sequences. As more sequencing of variants is done, the level of known sequences generated in the past may not cover many of the sequence variants in the present time (Zheng et al., 2010). It has been recommended that whole genome analysis should be done for viral identification which allows efficient characterization of the complete viral population, including viruses with circular or linear genomes, and viruses that are too divergent to be detected by PCR assays based on known viral sequences (Ha et al., 2008). Leveraging this technique to study viral sequences, affords sampling of large heterogeneous population of related viral genomes; to quantify genetic variation in viral populations more accurately (Eisen & MacCallum, 2009). As the cost of massively parallel sequencing reduces, this will become more feasible.

Table 1.1: The optimal nucleotide (nt) identity thresholds (% nt identity) for genus and species discrimination in the different genes of the family *Potyviridae*.

Table adapted and modified from Adams, Antoinw and Fauquet, 2005.

Genes	Between genus*	Between species**
	% N.t identity Same genus	% nt identity Same species
P1	41.4	58.0
HC-Pro	46.6	76.0
P3	-	74.0
CI	48.0	78.3
VPg	48.4	76.0
NIa-Pro	46.4	76.5
NIb	54.0	75.0
CP	49.0	76.0
5'-UTR	-	-
3'-UTR	-	76.0
ORF	46.0	76.0

*Below this value, sequences would be expected to be from different genus

**Below this value, sequences would be expected to be from different species

1.8 The host

Taro and tannia are stem tuber crops that are widely cultivated in tropical and subtropical regions of the world (Okonkwo, Ogbuokiri, Ofoegbu, & Klotz, 1993).The two most cultivated species worldwide are *Colocasia esculenta* (L.) Schott and *Xanthosoma sagittifolium* (L.) Schott, both belonging to the monocotyledonous family Araceae. The term taro is used to refer to *Colocasia sp*. It should not be confused with the related aroid *Xanthosoma sp.*, which is called tannia. The family (often referred to as aroids) contains several plants which are cultivated and used for food in various parts of the world (Onwuene, 1999). The Araceae family consists of some 100 genera and more than 1500 species, most of which are tropical and subtropical (Quero-Garcia, Ivancic, & Lebot, 2010). The aroids grow mainly in moist or shady habitats (Onwuene, 1999). Some are terrestrial, while others are vines, creepers or climbers. Many species are also epiphytes (Ekanem & Osuji, 2006).

There are seven species of *Colocasia* (taro) thought to have originated in North Eastern India and *Colocasia esculenta* is the most widely cultivated in Hawaii, South East and Central Asia (Malaysia, Philippines, China, Taiwan and Japan) and the South Pacific Islands (Fiji, Papua New Guineas, Vanuatu, Samoa, Tonga, Palau, Mariana Islands, Cook Islands, French Polynesia and the Solomon Islands) (Quero-Garcia et al., 2010). There are about 40 species of *Xanthosoma* (tannia) which are thought to have originated from the Amazonian region (Cuba, Venezuela, French Guiana, Colombia, Ecuador, Nicaragua, Peru and Brazil) of South America (Quero-Garcia et al., 2010). Of these, *Xanthosoma sagittifolium* is cultivated widely in South/Central America, the Caribbean, West Africa (Cameroon, Ghana and Nigeria) and the South Pacific Islands (Ekanem & Osuji, 2006).

The main economic parts of both species are the corm, and the leaves as shown in Figure 1.2. The corm is usually boiled, baked, roasted or fried and consumed in conjunction with other foods such as fish and coconut preparations. The leaves are usually boiled or prepared in various ways mixed with other condiments (Onwuene, 1999).

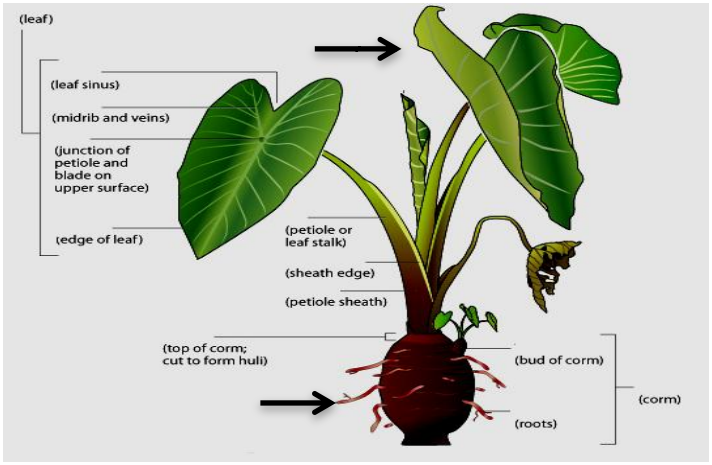


Figure 1.2: Gross anatomy of both *Colocasia sp.* and *Xanthosoma sp*. Main economic parts are indicated by the black arrows.

Image adapted and modified from http://hbs.bishopmuseum.org/botany/taro.

1.9 Origin and distribution of *Colocasia sp.* and *Xanthosoma sp.* and their diseases

Tracing the origins of viruses is difficult as they don't leave fossils like higher organisms; however, the relationship between plants and humans is one of interdependence (Quero-Garcia et al., 2010). It is through this interdependence that it may be possible to decipher the origins of viruses through the movement of crops/plants and human migration (Patrick, 2010). Historic migration of human populations began with the movement of *Homo erectus* out of Africa across Eurasia around a million years ago (Armelagos & Harper, 2005). Following the movement of humans into Central and South Asia (India, China, Pakistan and Bangladesh) 20,000 years ago, the migration of Austronesian people out of Taiwan into South East Asia and the Pacific (6,000- 8,000 years ago) was the last and most far-reaching prehistoric human migration (McCoy & Graves, 2010).

Following the movement of humans, agriculture gradually spread along human migration routes to South East Asia and east into the Pacific (Armelagos & Harper, 2005; McCoy & Graves, 2010). Based on human mitochondrial DNA (mtDNA) mapping (Pierson et al., 2006), the settlement of the many island groups of Near Oceania (Mariana Islands and Palau) and the Bismarck archipelago (Papua New Guinea and Solomon Islands) occurred around 4000-6000 (before present time) BP following the human colonization of South East Asia (Matisoo-Smith & Robins, 2004; McCoy & Graves, 2010; Pierson et al., 2006), then into Remote Oceania.

While it is tempting to say that both taro and tannia arrived in the various locations of the Pacific as man arrived, this is problematic as it appears that taro was introduced on several occasions. In a recent review based on genetic studies on the variation of both modern and ancient mtDNA in pigs, rats and chickens in the South Pacific (Patrick, 2010), a new model for Polynesian origin was proposed as follows (Figure 1.3); (i) the initial Oceania expansion began through Near Oceania (Mariana Islands and Palau) and Remote Oceania (Vanuatu, Cook Islands, Nauru, Fiji, Tonga and Samoa) occurred around 3500 -2700 BP ago, (ii) proposed second migration or arrival of new population (or populations) from Asia into Polynesia occurred approximately 2000-1500 BP and (iii) the last phase indicate settlement of East Polynesia and back migration into

Melanesia around 1200-500 BP (Matisoo-Smith & Robins, 2004; Addison & Matisoo-Smith, 2010).

In contrast, tannia was domesticated and cultivated during pre-Columbian times in the Amazon region of South America (Bradshaw, 2010). According to Garcia-Robledo et. al., (2005), tannia reached West Africa (Cameroon, Nigeria and Ghana) in the early16th century and was spread further by traders and missionaries into the rest of South America. During the late 16th century (500-300 BP), it was introduced into the South Pacific Islands by the early Spanish and Portuguese explorers from South East Asia (east to west movement from South America to West Africa) and South America (west to east movement from Nicaragua) as shown in Figure 1.3.

Understanding the movement of people allows us to understand the movement and spread of crops as people generally carried these with them as they travelled (García-Robledo, Quintero-Marín, & Mora-Kepfer, 2005). It is inevitable that a proportion of these plants were diseased, thus it may be possible to track the spread of disease via the movement of people. As a consequence, diseases of taro and tannia may be tracked by studying the movement of people into and around Remote Oceania.

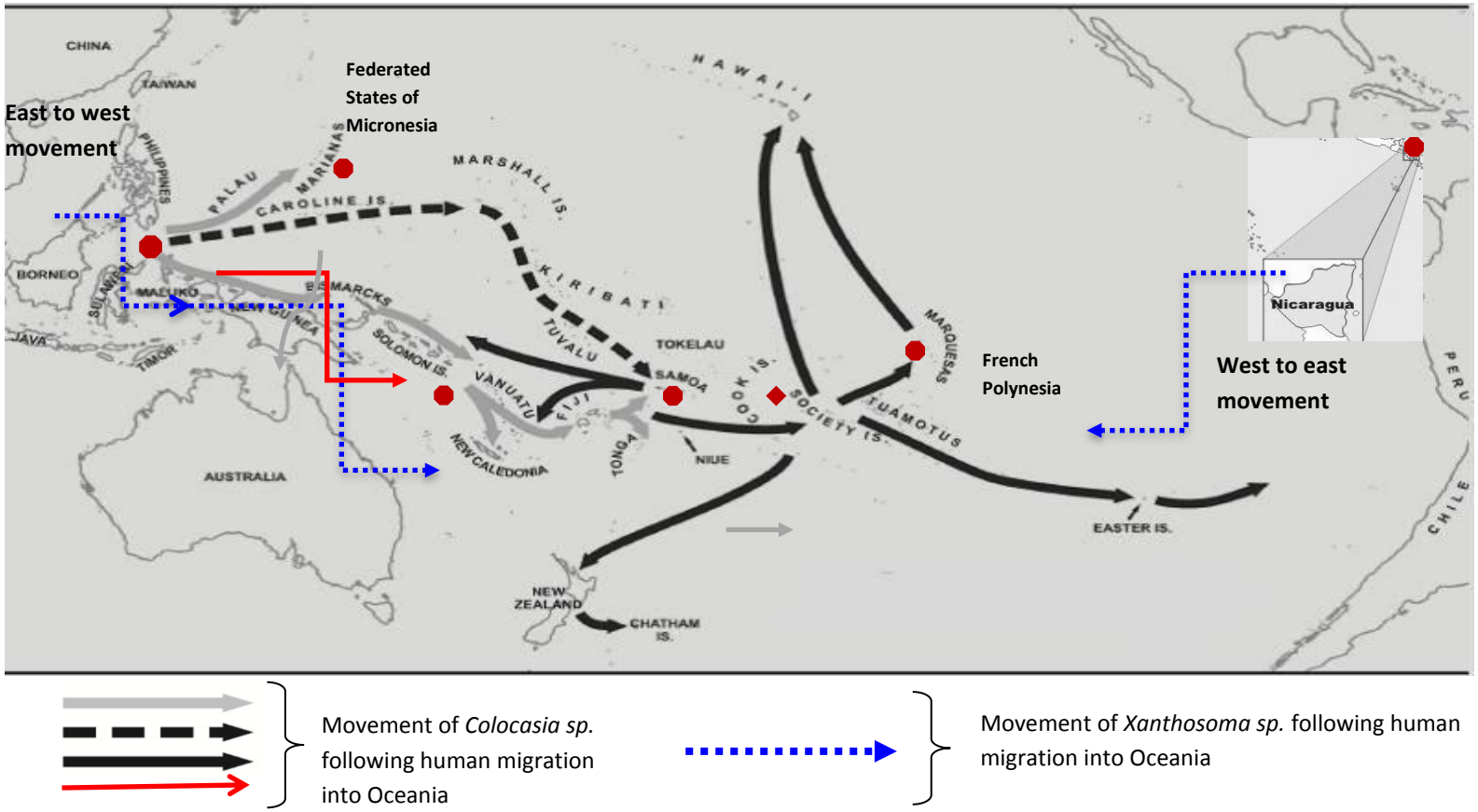


Figure 1.3: Movement of humans into Oceania suggesting possible routes and timing (BP- before present time) for the entry of *Colocasia sp.* and *Xanthosoma sp.* into Oceania. The red arrow indicates the settlement of the many island groups of Near Oceania (Mariana Islands and Palau) and Bismarck Archipelago (Papua New Guinea and Solomon Islands) which occurred relatively early in prehistory (4000-6000 BP). The grey arrows show the initial Austronesian expansion into Oceania (3500-2700 BP). The dotted arrows show the proposed arrival of new population (or populations) from Asia into Polynesia (2000-1500 BP). The black arrows show the settlement of East Polynesia and a back migration into Melanesia (1200-500BP). The dark dotted blue arrow shows introduction of *Xanthosoma sp.* following the early Spanish and Portuguese explorers into Oceania from South East Asia (east movement from South America to West Africa) and Central America (west to east movement from Nicaragua) around the 16th century (500-300 BP).

Image adapted and modified from Addison and Matisoo-Smith (2010).

1.10 Economic and cultural importance

Taro and tannia contribute significantly to food security in producing countries especially in both West Africa and the South Pacific (Quero-Garcia et al., 2010). Both crops rank 14th worldwide in production level, but in the South Pacific Islands, it is the major component of socio-cultural, dietary and economic livelihood (FAO, 2012). According to the Food and Agriculture organization of United Nation (FAO), mere production figures do not convey the full picture of importance of the crop in the producing countries (FAO, 2012). However, combined with Figures on land availability, population and utilization, a clear picture emerges that shows the South Pacific islands having the highest intensity of production, utilization and dependence on taro for food (Table 1.2).

Commonly, production and utilization figures are combined for both taro and tannia. In the South Pacific Islands very little tannia is utilized for food. In West Africa, however, the situation is reversed with more tannia utilized than taro, except in Nigeria (O' Hair, 1990). Most of the crop is produced in Nigeria, Ghana and Cote d'Ivoire (Quero-Garcia et al., 2010). Outside of West Africa, other African producers are Gabon, Egypt, Rwanda, Burundi, Zaire, Central African Republic, Comoros Island, Sao Tome and Principe, Madagascar and Mauritius. Tannia serves as an important food source during the dry season or before yam and cassava harvest in West Africa. Nevertheless, taro is always listed among the staple food crops of coastal West African countries from Nigeria to Guinea (O' Hair, 1990).

Though the bulk of taro is produced in West Africa, South Pacific as a whole has a higher dietary dependence on taro/tannia than any of the other continents of the world (FAO, 2012). No other part of the world can match South Pacific Islands in terms of the intensity of production, utilization and dependence on taro/tannia for food. Even though the figures in Table 1.2 are combined for taro and tannia, in the South Pacific, taro is the predominant crop. Most of the cultures of the South Pacific and South East Asia have evolved on the strength of root crops as the major food source (FAO, 2012). The major producers in South East Asia are China, Japan, Philippines and Thailand; while in the South Pacific, production is dominated by Papua New Guinea, Samoa, Solomon Islands, Tonga and Fiji (FAO, 2012). Both taro and tannia rank among the top two or

three staple food items (FAO, 2012) and play a major role in the socio-cultural aspect, which has for centuries characterized agriculture practice in the South Pacific Islands and South East Asia.

Table 1.2: Production yield and area in hectares (ha) for taro/tannia in 1990 and 2000 for leading producers in Asia, South Pacific and West Africa; Source: (FAO, 2012).

	Area (1000 ha.)		Yield (kg ha^{-1})		Production (1000 tonnes)	
	1990	2000	1990	2000	1900	2000
World	**983**	**1458**	**5314**	**6058**	**5225**	**8835**
Asia	**150**	**133**	**11536**	**13936**	**1727**	**1854**
Africa	**738**	**1274**	**3996**	**5229**	**3130**	**6662**
South Pacific	**47**	**48**	**7142**	**6103**	**337**	**292**

1.11 Production Constraints overview

In spite of intense global coordinated efforts to realize the full potential of taro and tannia as sources of food, animal feed and processed products through reviewed agronomic principles such as classification, growth, structure and soil-plant relationships of field crops, many constraints remain. According to the Australian Centre for International and Agricultural Research (ACIAR), both biophysical and socio-economical constraints are adversely affecting production of both crops in the South Pacific Islands and South America (ACIAR, 2012).

It is clear that there are several constraints that limit the scope of present-day taro/tannia cultivation and production (CIP, 2000a). The major ones, especially as they apply to the Pacific Islands/South America are as follows: i) Susceptibility to pest and pathogens: the most significant viral disease-affecting taro and tannia is *Dasheen mosaic virus* (DsMV) (ii) Laboriousness of the production system due to vegetative propagation process, (iii) Scarcity of planting material: Like most of the other tropical root crops, the planting material for taro/tannia is bulky, making it expensive to transport over long distances. It is also perishable, and cannot be stored over a long time and (iv) Post harvest handling and marketing: At present, the bulk of taro/tannia produced is handled and marketed as the fresh corm. The corm itself has high water content, and cannot be

stored for more than a few days at ambient temperatures and (v) Limited research and extension: The amount of research currently being done on taro/tannia is very little especially in South Pacific Islands and most of the above constraints in production can be effectively tackled and possibly solved through research.

1.12 *Dasheen mosaic virus* (DsMV)

DsMV was first recorded from Dasheen (*Colocasia esculenta* (L) Schott) in 1970 (Nelson, 2008) and affects both taro and tannia throughout the Pacific region and worldwide (Babu, Hegde, Makeshkumar, & Jeeva, 2011). The impact of DsMV on the production of both crops in the South Pacific Islands has been mounting steadily in recent years, reaching devastating proportions in Samoa in 1993, causing export earnings to fall from 49 million to 85,000 NZD (New Zealand currency) in just one year (ACIAR, 2012). This economic loss was coupled with an increase in export costs – staple foods had to be imported to replace both taro and tannia (Nelson, 2008). With many other Pacific countries similarly at risk, several ACIAR projects are providing the scientific resources to help build more effective disease and pest management programs (Braidotti, 2006).

DsMV has been classified in the genus Potyvirus based on the particle morphology, aphid-transmissibility, ability to induce cylindrical ('pinwheel') inclusions in its hosts, and serological relationships to other Potyviruses (Abo El-Nil, Zettler, & Hiebert, 1977). Immunodiffusion tests conducted in agar gels containing sodium dodecyl sulphate indicated that this virus is related to blackeye cowpea mosaic virus and tobacco etch virus (Abo El-Nil et al., 1977). Like other Potyviruses, the genome contains a single ORF, flanked by a UTR at both the 5´-end 3´-ends. From alignments with polyprotein sequences of other Potyviruses the ten mature proteins (except the recently discovered PIPO protein) characteristic of the genus were identified (Chen & Adams, 2001a) and like many Potyviruses, DsMV has a DAG motif (Asp-Ala-Gly) near the N-terminus of the coat protein and this has been shown to be important for virus transmission by aphids (Figure 1.1).

DsMV infection causes serious damage to both crops (taro and tannia) and variability in

symptom expression may depend on the host, the particular strain of DsMV, and the environment. For example, when severely diseased, two or three leaves may show symptoms and then apparently healthy leaves are produced, or leaves may alternate between asymptomatic and symptomatic on the same plant. The pathogen infects leaves, stems, and petioles. The disease results in reduced photosynthesis, reduced corm size or yield, reduction in leaf number or size, and stunted plants. Also, symptoms may be intermittent and vary seasonally, being worse in summer time (Nelson, 2008).

Symptoms produced by DsMV on taro plants as shown in Figure 1.4, display a variety of mosaic patterns (Nelson, 2008) with small, irregular and scattered grey, green, yellow and sometimes white patches along or between the major veins. Symptoms can also display brilliant white or yellow feather-like patterns along the veins and sometimes throughout the leaf blade. Weak symptoms on plants may appear before the virus is detectable using serological methods. This was found recently by Mosahebi, Koohi-Habibi & Okhovvat (2005) testing potato species infected with potato virus M (PVM). Results indicated that the virus concentration increased slowly and that the first symptoms appeared before the concentration was sufficiently high to permit detection with standard serological methods.

The variation in symptoms described above in consecutive vegetative generations may be due to a variety of reasons (Mosahebi, Koohi-Habibi, & Okhovvat, 2005): (1) mixture of strains of the virus in originally infected plants with changing preponderance of the components of the mixture; (2) constant mutation of virus strains from mild to severe and vice versa: (3) variation in the reaction of plants to the presence of the virus, which in turn may be due to various reasons such as variation in growth conditions, variation in defence reaction to the virus invasion, etc.; (4) uneven distribution of virus in the tuber propagants due to chance or causes other than those mentioned in items 1 and 2.

It has also been hypothesized that the coat proteins have profound effect on the symptomology of the host plant. For example, according to a study conducted by Ullah et.al. (2003), the amino terminus (NT) of the Potyvirus coat protein (CP), which is necessary for aphid transmission and systemic infection, is externally located on the virion, strongly antigenic, and highly variable in length and sequence. In this study, chimeras of *Zucchini yellow mosaic virus* (ZYMV) were constructed by substitution of

the native CP-NT with CP-NTs of Potyviruses with overlapping (*Watermelon mosaic virus*, WMV) and non-overlapping (*Tobacco etch virus*, TEV) host range. The chimeric viruses produced strong initial symptoms on ZYMV-susceptible cucurbits. Approximately 6 weeks post-inoculation, zucchini and cucumber plants infected with the TEV chimera, exhibited a distinct recovery characterized by a loss of symptoms on young leaves, reduced virus titer, and virus-specific protection against secondary infection. The chimeric viruses did not overcome naturally occurring ZYMV resistance in cucumber, ZYMV-CP-mediated resistance in transgenic melons, or expand host range to TEV- or WMV-susceptible species. These results demonstrated that despite substantial variability in length and sequence, the CP-NTs from heterologous Potyviruses facilitated systemic infection in ZYMV-susceptible cucurbits, but were not sufficient to cause infection in non-ZYMV hosts. The observed recovery response with the non-cucurbit CP-NT suggests that the CP-NT is important for biological efficiency of the virus and/or adaptation of the virus to its host. Nevertheless, correlations between virus sequence variation and symptomology have been rarely studied (de Assis Filho, Paguio, Sherwood, & Deom, 2002), perhaps this could open up new avenues for future research.

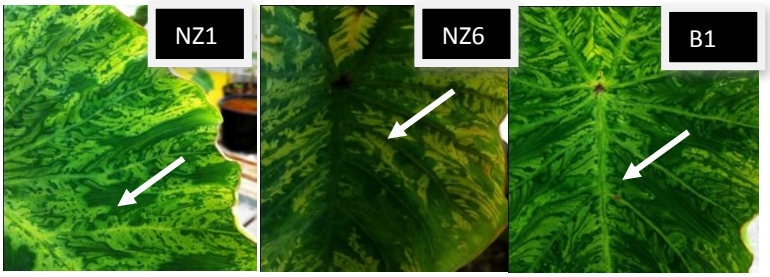


Figure 1.4: Left to right, symptoms of DsMV infection of Colocasia *esculenta* (NZ and B strains): White arrows indicate pale whitish to yellow green patches on the leaves, characteristically as featherlike patterns along the veins, especially near the leaf margins.

Images taken by the author from glasshouse-grown taro located at the School of Biological Sciences, The University of Auckland.

1.13 Sequence variation and quasispecies concept

To understand the concept of variation in an RNA virus, one has to consider the replication strategy of plant RNA viruses. Plant RNA viruses such as Potyviruses have very high mutation rates due to the fact that their viral RNA dependent RNA

polymerase lacks proofreading activity, which leads to frequent errors occurring during viral RNA synthesis (Sztuba-Solińska, Urbanowicz, Figlerowicz, & Bujarski, 2011). Thus, many mutants are created leading to the formation of a population of sequences derived from a founder population. The viral load during RNA virus replication has a strong dynamic component. For example, in persons infected with Human immunodeficiency virus-1 (HIV-1), Hepatitis B virus (HBV), or Hepatitis C virus (HCV), an estimated 10^{10} to 10^{12} new virions are produced each day (Levin & Bull, 1994). Models of HIV-1pathogenesis based on the continuous production of antigenic variants (Nowak et al., 1991) suggested that as the infection progresses the complexity of antigenically distinct mutants that arise due to lack of proof reading may overwhelm the immune system, leading to acquired immunodeficiency syndrome (AIDS) (Nowak et al., 1991). In HIV, and many other infections, new mutant variants are created at astonishingly high rates (Nowak et al., 1991). Hence, the balance between mutation rates and replication rounds is one of the reasons for the great adaptability of RNA viruses (Domingo, 1998).

The quasispecies concept was developed in 1998 by Eigen et.al, to describe the equilibrium process between mutation and natural selection, which generates a population of variable genomes. These genetic variants are organized around one or a set of genotypes of highest fitness known as master sequences (Eigen, McCaskill, & Schuster, 1988). RNA viruses are considered to be quasispecies since they exist as populations of variable genomes. A critical element of quasispecies theory in relation to RNA viruses, is that the frequency of any individual variant virus in the quasispecies is a function of both its own replication rate and the probability that it will arise by the erroneous replication of other members of the population (Miralles, Gerrish, Moya, & Elena, 1999). Consequently, viruses are not independent entities in the quasispecies but are linked by mutational couplings, so that the entire population forms a cooperative structure that evolves as a single unit (Figure 1.5). The consequence of this population structure is that natural selection is no longer directed toward the single fittest variant, as in most population genetic models, but instead acts on the whole mutant distribution—the quasispecies in its entirety—which will then evolve to maximize its average replication rate (Lauring & Andino, 2010).

The genetic organization of a viral population is often depicted using the concept of sequence space, a geometric representation of all possible sequences where physical distance reflects genetic similarity. For example, a given viral genome will undergo replication and generate hundreds of progeny that differ at roughly one position (Lauring & Andino, 2010). Subsequent rounds of replication will generate a more complex mutant distribution with variants lying farther away from each other in sequence space. This ensemble of mutants forms a "cloud" of variants, or quasispecies, in which mutation generates a swarm of candidate genomes that is pruned by natural selection. According to population genetics, the frequency of a given variant in a population is closely approximated by its ability to survive and reproduce—its fitness (Lauring & Andino, 2010). In quasispecies formulations, where mutation rates are elevated, fitness is also subject to the probability that the variant will be generated *de novo* by mutation of its neighbours in sequence space (Domingo et al., 2006). In RNA viruses, the contribution of mutation to genotype frequency is significant, and variants are "coupled" in sequence space (Biebricher & Eigen, 2006). That is, a low fitness variant can be maintained at a higher than expected frequency because it is coupled to a well-represented, higher fitness genotype in sequence space. The phenomenon of mutational coupling is one of the defining characteristics of quasispecies, as it places individual mutants within a functional network of variants (Biebricher & Eigen, 2006).

A viral quasispecies, then, is a cloud of diverse variants that are genetically linked through mutation, interact cooperatively on a functional level, and collectively contribute to the characteristics of the population(Lauring & Andino, 2010). The unit of selection is the population as a whole, and the nature of the functional interactions among genetically distinct variants is therefore of critical importance to pathogenesis in infected hosts (Lauring & Andino, 2010).

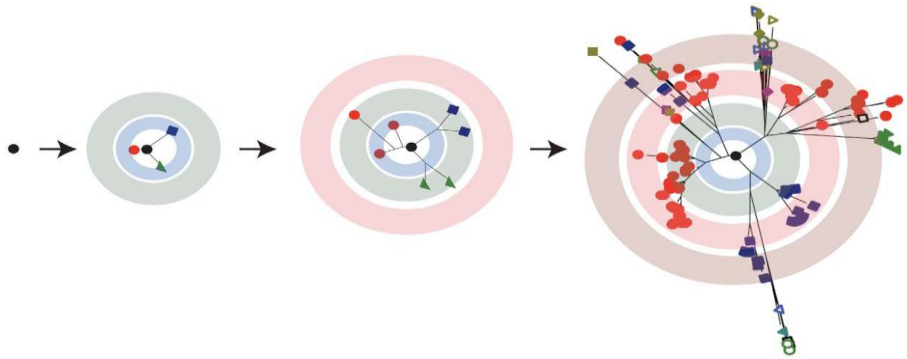


Figure 1.5: Quasispecies concept in a RNA virus population. A virus replicating with a high mutation rate will generate a diverse mutant repertoire over the course of a few generations. In these trees, each branch indicates two variants linked by a point mutation and the concentric circles represent serial replication cycles. The resulting distribution is often represented as a cloud centred on a master sequence.

Image adapted from Lauring and Andino, (2010).

1.14 RNA virus evolution in plants

The genetic evolution of RNA viruses is governed by the interplay of forces such as mutation, genetic exchange (recombination and reassortment), natural selection, and genetic drift. Most of these concepts are described in detail in chapters two and three. Studies of RNA virus evolution primarily consider the dynamics of mutation, selection and genetic drift in relation to the two important RNA virus properties (Roossinck, 2011): (1) explosive replication to produce vast population sizes and (2) high mutation rates acting on a short genome.

First, RNA viruses often have very large population sizes, such that the number of viral particles in a given organism might be as high as 10^{12} (Moya, Holmes, & Gonzalez-Candelas, 2004). Second, such immense population sizes, which are several orders of magnitude larger than those observed for cellular organisms, are a product of explosive replication. For example, in influenza A virus, a single infectious particle can produce an average of 100,000 viral copies in 10 hours (Moya et al., 2004). As natural selection is most efficient with large populations, it is no surprise that experiments using RNA viruses have shown that selection is of fundamental importance in controlling their evolutionary dynamics, such that new mutants with increased fitness, continually appear and out-compete older, inferior variants (Holmes, 2003). Third, owing to the lack of

proofreading activity in their polymerase proteins, RNA viruses exhibit the highest mutation rates of any group of organisms, approximately one mutation per genome, per replication (Holmes, 2003). Finally, the genome sizes of RNA viruses are typically small, ranging from only 3 kb to ~30 kb, with a median size of ~9 kb (Holmes, 2003). These last two properties are intimately related because high-mutation rates are theoretically expected to limit genome size. In particular, a mutation rate that exceeds a notional error threshold (set at approximately the reciprocal of the genome size) generates so many deleterious mutations in each replication cycle that even the fittest viral genomes are unable to reproduce, and population size decreases to extinction (Moya et al., 2004).

In the simple situation outlined above, RNA viruses should evolve in a deterministic manner with the fate of a given mutation being determined by its initial frequency, with the process of natural selection working efficiently on a vast array of mutational variants. Although it is true that RNA virus populations are often highly diverse, this is not sufficient to explain the entirety of RNA virus evolution (Domingo, 1989). In particular, when individual populations may be large the effective population (viral population responsible for initiating infection in the next generation) is often potentially much smaller. The effective population is influenced by the probability of population bottlenecks caused by migration within a host to different replication sites, transmission between populations of hosts, and could also be influenced by variation in replication potential among virus variants (Roossnick, 2011). This would result in effective population numbers much smaller and indicates that random effects due to genetic drift should be considered for understanding the systematic movement of RNA virus within a host (Rouzine, Rodrigo, & Coffin, 2001).

It has been proposed that RNA viruses evolve to maximize their mutation rate such that beneficial mutations can arise at a rate sufficient for the requirement to adapt, yet deleterious mutations are kept at a level at which they can be excised from the population by the forces of selection pressure (Barr & Fearns, 2010). A trade-off between replication efficiency and replication fidelity has been proposed (Barr & Fearns, 2010); however, it has yet to be proved that increases in the mutant spectrum within a population due to low fidelity replication result in an increased rate of adaptation (Keulen, van Wijk, Schuurman, Berkhout, & Boucher, 1999).

1.15 Molecular clock and evolution of RNA virus

Although the ultimate origins of RNA viruses are uncertain, it seems reasonable to assume that these infectious agents have a long evolutionary history, appearing with, or perhaps before, the first cellular life-forms (Strauss & Strauss, 2001). While the RNA viruses we see today may not date back quite this far, evidence that some DNA viruses have evolved with their vertebrate hosts over many millions of years (McGeoch, Dolan, & Ralph, 2000) makes an equally ancient history for RNA viruses a natural expectation. Yet a very different picture of RNA virus origins is painted if their gene sequences are compared. Using the best estimates for the rates of evolutionary change and assuming an approximate molecular clock (Jenkins, Rambaut, Pybus, & Holmes, 2002), it can be inferred that the families of RNA viruses (plants, animals and humans) circulating today could only have appeared very recently, probably not more than about 50,000 years ago. Hence, if evolutionary rates are accurate and relatively constant, present-day RNA viruses may have originated more recently than our own species. The plausible explanation for the recent origin of RNA viruses is that the molecular clock dates only relate to the RNA viruses currently circulating, that is, those that have been identified over the last 100 years or so (Zanotto, Gibbs, Gould, & Holmes, 1996). Thus, RNA virus families like the Flaviviruses may in fact have histories dating back many millions of years, but the early members of these families have gone extinct to be replaced by those we sample today. All that is left following these extinction events are the very long branches relating the different families of RNA viruses to each other.

Despite the growing number of complete genome sequences available for all the major groups of RNA viruses, there is as yet no consensus as to how these viruses came to be (Jenkins et al., 2002). Although RNA viruses are predicted to have been among the first life forms to evolve, molecular clock estimates consistently predict that RNA virus origins are much more recent (Jenkins et al., 2002). The problem stems from their fast rate of evolution, and the key to establishing an accurate evolutionary timescale (Holmes, 2009).

For example, simple molecular clock calculations suggest that the origin of RNA viruses is an extremely recent event (Jenkins et al., 2002). However, in some cases such a recent origin conflicts with other evolutionary data (Holmes, 2003). Perhaps the most

notorious example is that of the primate Lentiviruses, which include the HIV-1 and HIV-2 and a growing list of simian immunodeficiency viruses (SIVs) that infect a wide variety of African monkeys (Hahn, Shaw, De , Cock, & Sharp, 2000). At face value, it would appear that SIVs have been associated with their host species for millions of years. Not only are they asymptomatic in their natural hosts, which when compared to the high virulence of HIV, suggests that they have evolved stable associations over an extended time period. However, the phylogenies of the viruses and the hosts often match, which implies that the viruses and the hosts have undergone cospeciation. Although the divergence times of the primate species in question are often uncertain, it is clear that virus-host cospeciation must mean a viral evolutionary history dating back millions of years (Hahn et al., 2000). In contrast, a study conducted by Sharp et. al., (2001) suggested far more recent timings and maximum sequence divergence observed indicated that the deepest split among the primate Lentiviruses occurred only a few thousand years ago, and clearly far more recently than their host species diverged.

It has been proposed that the molecular clock is not constant and that rates of mutation have changed dramatically both between viruses and along lineages (Sharp et al., 2001). For example, in plant virus evolution, Gibbs et al. (2010), used time-calibrated phylogenetic trees of Nicotiana species to infer the existence of begomoviruses in the Americas around 1.9 million years ago (MYA), and results indicated long-term substitution rates of the begomoviruses CP gene was much lower than (6×10^{-7} substitutions per site per year) the short-term nucleotide substitution rates estimated for the CP genes of begomoviruses. Such low long-term substitution rates are clearly consistent with the hypothesis that some *Begomovirus* species might be co-diverging with either their vector or host species. Specifically, if some RNA viruses evolved more slowly than others or have experienced periods when their rates of mutation were reduced, then divergence times could be greatly extended (Holmes, 2009). The difference observed in the short-term rates are expected to be faster than long-term rates due to their being both upwardly biased by transient substitutions that do not become fixed over the long-term, and blind to the true rates of change at slowly evolving, but apparently invariant, nucleotide sites at which negative selection is acting (Holmes, 2009).

There are a number of factors that influence the accuracy of the molecular clock

analysis (Pavesi, 2001); (1) the probability of different types of base change and variation in base composition, are more likely to be the source of error in generating relative rates of substitution for different sites along the sequence (2) if RNA viruses are biased such that mutation rates vary dramatically between sites, then this could have a major effect on distance estimates and (3) perhaps the most likely way that clock rates could vary between RNA viruses is if there are major differences in rates of replication, that is, in virus generation times. In summary, one cannot expect that short-term rates of nucleotide substitution inferred from sequences sampled over a short-term period to fully capture the complex long-term evolutionary dynamics of geographically sub-divided viral populations.

An important theme of this thesis will be the detection of departure from the strict clock model and the application of the relaxed molecular clock to the analysis of DsMV evolution.

1.16 The genetic variability and evolution of DsMV

As mention previously, DsMV belongs to the *Potyviridae* family and is one the most important viral pathogens attacking wild and cultivated species of the family Araceae worldwide (Zettler, Jackson & Frison, 1989). In recent years, the number of DsMV sequences reported in the literature has increased; however, there are still limited numbers of sequences in the NCBI sequence database. Phylogenetic analyses have shown that DsMV belongs to the bean common mosaic virus (BCMV) group within the genus *Potyvirus* (Adams, Antoniw, & Fauquet, 2005) and that DsMV is most closely related to vanilla mosaic virus (VanMV) (Farreyrol, Pearson, Grisoni, Cohen, & Beck, 2006). While other members of the Potyvirus genus such as bean yellow mosaic *virus* (BYMV), which are, have been analysed in detail very little attention has been paid to the importance of DSMV. As DSMV is the predominant virus infecting both taro and tannia in the South Pacific Islands, there is every justification for wider analysis of this virus.

1.17 Variability of the coat protein among DsMV isolates

As mentioned previously in section 1.6, the potyvirus CP is essential for aphid transmission and may have a role in the movement of the virus both from cell to cell and over longer distances. Based on the structural characterization of the potyvirus CP, there is an accumulating body of knowledge that variation in the CP sequences could clearly differentiate individual potyviruses and their strains (Adams et al., 2005).

Variation between the CP of DsMV isolates have been studied previously, for example, in a study conducted by Li, et.al. (1999) western blot analysis using DsMV polyclonal antiserum demonstrated CP sizes of seven dasheen mosaic virus (DsMV) isolates were estimated within a range 38–47 kilodalton (kDa). The apparent CP sizes of the DsMV isolates extracted from their original hosts were 47 kDa (*Xanthosoma caracu*), 45 kDa (*Colocasia esculenta*, *Zantedeschia aethiopica*), and 38–46 kDa (*Caladium hortulanum*). Results revealed the apparent CP size by Western blots extracts from infected aroids may be used in the characterization and differentiation of DsMV isolates. Result also demonstrated that the variation in CP sizes could be associated with diversity on the N terminal region of the CPs. There is also evidence for the occurrence of a severe strain of the virus in French Polynesia (Zettler & Hartman, 1987). Serological differences were also noticed between a Fiji isolate and Florida isolates (Li, et.al. 1999). Based on these findings, it is obvious that distinct DsMV isolates do exist; however, little is known about their evolutionary relationships occurring in different geographical locations and among various hosts.

1.18 Aims

Although evolution of RNA viruses has been thoroughly studied in animal systems, in recent years plant viruses have gained a more prominent place in the field (Roossinck, 2003; Sardanyes & Elena, 2011). However, experimental evolution and variability studies of plant RNA viruses are generally limited and largely based on descriptive biological experiments or observations. Plant viruses have been used as model systems for virus evolution because of the advantage of using plant as host; no animal systems can provide unlimited numbers of genetically identical hosts for infection. Plant

protoplasts provide a primary cell culture system that avoids some pitfalls of animal tissue culture and there are no ethical concerns about doing whole or vegetatively propagated plants experiments (Roossinck, 2003).

Following the trend of animal and human viruses, plant virologists have been increasingly interested in studying plant virus evolution. For example, for Potyviruses, most evolutionary studies have been generally focused on the genetic variability among isolates of the same species. And yet the ability of Potyviruses to replicate, adapt to survive and infect new hosts remains far from being understood. Analysing the genetic variation of the viral genome can give an understanding of how the virus evolves over time.

From the above discussion, the variability of DsMV and the possible pathological consequences of this raised the need for more detailed analysis of the genetic variations within the populations of the virus. To better understand the nature of evolutionary processes in DsMV, one aim of this study was to determine the genetic variability of DsMV within and between infected plants over a period of time (temporal analysis). DsMV 3' end sequences were isolated and analysed from different propagants inoculated with similar strains of DsMV. Phylogenetic analysis were also carried out to better understand the degree of sequence variation between these samples and to determine the relationships between them.

Another aim of this study was to understand the long-term evolution of DsMV. From the pool of sequence variants that exist within a single plant, movement of infected plants from one location to another as people move, the founding population at each new location will differ. To understand the spatial evolution of DsMV, viral sequences from infected samples from various Pacific Islands were compared to determine the evolutionary relationships between them.

The information generated from such sequence and phylogenetic analysis will contribute to the understanding of Potyvirus sequence variation that occurs in the short and long term; what degree of variation occurs within a plant and then how that variation contributes to the longer term evolution of the virus. This study may provide information to understand the evolution of RNA viruses generally. This information

will also be helpful to the development of better management and control of the Potyvirus'spread.

Chapter 2

Genetic variation and temporal analysis of DsMV

2.1 Introduction

Genetic diversity is the essential component in a viral population that allows a species to evolve in an ever-changing environment with shifting selection pressures. RNA viruses have extreme evolutionary capacities that have allowed them to adapt to parasitize all known groups of organisms and, in many cases, to rapidly adapt to numerous host species within a kingdom (Roossinck, 2003). Considerable effort has been made to demonstrate and understand the population structure and evolutionary capacities of human viruses such as HIV, *Vesicular stomatitis virus* (VSV), and HCV (Hahn et al., 2000) and plant viruses such as WMV and *Papaya ringspot virus* (PRSV) (Moury, Fabre, & Senoussi, 2007). These viruses, with RNA genomes are associated with error-prone replication and short generation times that result in large, highly diverse replicating populations.

In contrast to plant RNA viruses, research in genetic structure and its variability has been relatively scarce compared to abundant research carried out on human and animal RNA viruses (Garcia-Arenal et al., 2001). A limited number of studies with plant RNA viruses have examined population genetic structure using monoclonal antibodies and molecular markers (Acosta-Leal, Duffy, Xiong, Hammond, & Elena, 2011; Elena et al., 2011; Fraile et al., 1997). Over the last 15 – 20 years, molecular analyses using sequence-based methods have been increasing (Lauring & Andino, 2010). An important goal in studying viral evolution is to understand the mechanisms of evolution and how they shape the genetic structure of viral populations (Flynn, 2010). The study of the variability and changes in the genetic structure of plant virus populations is an important aspect of plant pathology and highly relevant to the development of strategies for the control of virus-induced diseases.

2.1.1 Sources of variation

The molecular mechanisms underlying plant RNA virus evolution are mutation, genetic exchange (recombination/reassortment), selection and migration. Genetic diversification is a complex process involving at least two stages: *1)* mutant generation during viral

replication in infected cells, and *2)* stage of the mutants in the cells where they are generated, upon expansion in new individual hosts. The two stages originate and are driven by entirely different mechanisms (Roossinck, 2003). Mutant generation is essentially governed by the stochastic process of mutagenesis, whereas the fate of each individual variant depends on its fitness (or ability to produce infectious propagants) in the ensemble of mutants, and also on random sampling events in a complex network of influences (Barr & Fearns, 2010). This introduction summarizes the basic mechanisms identified as relevant to plant RNA viruses' variation and evolution.

2.1.2 Genome integrity

As described in section 1.5, RNA virus genome replication is performed by the viral replicase complex, which for most viruses is likely to be an assembly of multiple viral and cellular proteins. The catalytic subunit of this complex is referred to as the RNA-dependent RNA polymerase (RdRp). RdRp are notoriously error-prone and so RNA virus genomes are subject to potentially catastrophic alteration from their own error-prone RNA-synthesis machinery (Barr & Fearns, 2010). The difference between positive and negative strand RNA viruses extend beyond the polarity of the RNA assembled into virions (Barr & Fearns, 2010). Positive sense RNA genomes exchange their virion proteins for ribosomes and cellular RNA-binding proteins at the start of infection. Once synthesised and assembled, the virus-specified RdRp and other structural proteins replace ribosomes to accomplish RNA replication. In contrast, negative sense RNA genomes remain associated with their nucleoplasmid proteins, both within virus particle and throughout virus replication cycle. These distinctions mostly reflect the significant differences in the structures of the viral RdRp-template complexes and in the molecular mechanisms of replication of positive and negative RNA virus genomes (Li, et.al. 2009).

Viral genomes may also be susceptible to physical or chemical damage by environmental factors (Sardanyes & Elena, 2011). One such modification is alkylation, the addition of carbon chains to nitrogen and oxygen atoms in RNA bases, in the form of either methylation (addition of a single carbon) or addition of longer carbon chains (Barr & Fearns, 2010). Alkylation could inhibit genome expression and replication,

resulting in an increased mutation rate. For example, bioinformatics analysis performed by van der Born et. al., (2008) revealed a number of different positive-strand RNA plant viruses contain an AlkB domain, and functional studies have shown that some of these domains can repair RNA damage by oxidative demethylation. These findings suggested that some viruses have acquired this protective mechanism and that oxidative demethylation has a role in the virus replication cycle (van den Born et al., 2008).

The terminal sequences of RNA virus genomes are particularly vulnerable to deletion or degradation. RNA viruses have evolved sophisticated mechanisms to avoid truncation of terminal sequences during replication initiation and termination (Poranen, Koivunen, & Bamford, 2008; Tayon, Kim, & Kao, 2001). However, despite these measures, the termini of virus genomes remain susceptible to host cell-mediated RNA degradation, either through specific antiviral host proteins (Silverman, 2007) or due to the components of the normal host-cell RNA-biosynthesis machinery (Houseley & Tollervey, 2009). Virus genomes possess various adaptations helping them to evade these nuclease activities. For example, it has been shown for the Potyvirus *Plum pox virus* (PPV) (Simon-Buela, Osaba, Garcia, & Lopez-Moya, 2000) where full-length *in vitro* transcripts of PPV genomic RNA with mutations altering the number of 5'-terminal adenosine residues were able to infect *Nicotiana clevelandii* plants, whereas a mutant with a substitution of adenosine by guanosine failed to infect. The genomic 5' end was template-independently repaired during *in vivo* RNA synthesis producing wild-type viral progeny. The repair of 5' -end deletions could be explained assuming that synthesis of plus-strand genomic RNA is initiated by primers that have the wild-type sequence. Nagy (1998) have proposed a 3'-end repair mechanism in which small oligonucleotides synthesized by the RNA replicase with a helper viral genome as template serve as primers for the initiation of RNA synthesis, giving rise to the restoration of the wild-type sequence.

Whilst the mechanism(s) for repair in these cases has not been established, the small size of the deletions could potentially allow repair as a consequence of replication initiation. For example, Picornavirus RNA replication begins with uridylylation of a viral protein, VPg, to create a molecule VPg–pU–pU, which acts as a primer for RNA synthesis initiation. Typically, the primer anneals to adenylate residues at the 3′ terminus of the template; however, there are multiple factors that are important for

positioning the replicase complex (Liu, Wimmer, & Paul, 2009). Thus, it is possible that the primer can still function to initiate RNA replication, even if it cannot base pair with the template (Rodrigo, Carrera, Jaramillo, & Elena, 2011). This would restore the missing nucleotides and allow rapid amplification of the repaired genome by using a prime/realign mechanism in which the RdRp initiates internally on the template and then realigns the nascent RNA and utilizes it as a primer. However, this prime/realign mechanism varies from one virus species to another (Barr & Fearns, 2010).

Examples of terminal-repair mechanisms span numerous virus taxonomic lineages indicating that possession of a terminal-repair activity is a fundamental element of RNA virus molecular biology (Codoñer & Elena, 2008).

2.1.3 Mutation rate of RNA viruses

Sequence variation provides the resource for natural selection and the progressive adaptation of the virus population to a changing environment. Viruses face continuous environmental change as they pass from cell to cell within a host and from host to host. From a genetic perspective, viruses can be classified according to whether the genome comprises RNA or DNA. RNA viruses are inherently hypervariable due to the absence of efficient proofreading and post-replicative repair activities associated with RNA replicase and reverse transcriptase (Holmes, 2009; Sardanyes & Elena, 2011).

RNA viruses exploit all known mechanisms of genetic variation to ensure their survival. Distinctive features of RNA virus replication include high mutation rates, high yields, and short replication times (Domingo, 1989). The resulting rate of nucleotide misincorporation in RNA viruses (10^{-3}-10^{-4}) is at least 1000 times that of bacteria or eukaryotes, causing one or more base substitutions each time the viral genome replicates (Holmes, 2009). Some mutations are lethal as they truncate or distort resulting proteins, rendering it non-functional. However, many mutations result in viable genomes that continue to replicate and contribute to the virus population. In this way, RNA viruses continually refine their genetic structure to accommodate the changing environment (Domingo & Holland, 1997). It must be noted that the variation of RNA viruses is extremely dependent on the population size of the virus that is

involved in the infections (Sardanyes & Elena, 2011). An important challenge for any RNA virus variation study is the assignment of phenotypic traits to specific mutations.

Despite the different constellations of mutations, epidemiological as well as functional and structural studies suggest that RNA viruses can tolerate restricted types and numbers of mutations during any specific time point during their evolution (Domingo & Holland, 1994; Roossinck, 2011).

2.1.4 Genetic drift- bottleneck

Genetic drift is the term used in population genetics to refer to the statistical drift over time of gene frequencies in a population due to random sampling effects in the formation of successive generations (Shriner et al., 2004). In the context of viral populations, genetic drift occurs when the viral population size is limited and therefore, by chance, certain variants increase or decrease in frequency. Unlike natural selection, genetic drift is random and rarely produces adaptations to the environment (Holmes, 2009).

Genetic drift has several important effects on plant RNA virus evolution. For example, in a recent study using *Tobacco mosaic virus* (TMV), the number of particles in an infected tobacco leaf has been estimated to be in the range of 10^{11}–10^{12} (Sacristan, Malpica, Fraile, & Garcia-Arenal, 2003). This census population size might differ by a large factor from the effective population size, which is the number of individuals that contribute equally to the following generation. The effective population size, and not the census population size, is what matters for the evolution of the virus population, a point often overlooked in the virological literature. In a population of an RNA virus such as TMV, the effective population size may be much smaller than the actual population size, because a large fraction of the population will consist of mutants that will not multiply, as suggested by their low intrinsic infectivity (in the range 10^{3}–10^{4}). In addition, infection of a new host may be started by a very small number of virus particles, which will reduce even more the effective population size. Based on this study, genetic drift reduces genetic variation in the TMV populations, which potentially decreases the virus population's ability to evolve in response to new selective pressures.

Genetic drift acts faster and has more drastic results in smaller virus population by increasing the rates of genetic fixation and extinction (Garcia-Arenal, Fraile, & Malpica, 2001).

In contrast with genetic drift, genetic bottlenecks occur when a viral population's size is reduced for at least one generation. Because genetic drift acts more quickly to reduce genetic variation in small populations, undergoing a bottleneck can reduce a viral population's genetic variation even more, even if the bottleneck doesn't last for very many generations (Garcia-Arenal et al., 2001). Bottlenecks can also occur at different moments of the life history of the virus, such as (1) each time a virus infects a new host plant, (2) a new plant species becomes a host, or (3) a new geographical area is colonized by the virus (Garcia-Arenal et al., 2001). Using TMV as an example, it was revealed that the number of founder populations in the colonization of a new leaf was small, indicating severe bottlenecks occurred during systemic colonization of tobacco plants, and that effective numbers could be much smaller than census numbers. Thus, genetic drift-bottleneck may be important in the evolution of plant virus populations, as shown for TMV on the changes in population structure during the colonization of new organs or new host plants (Sacristan et al., 2003).

Such bottlenecks result in a type of genetic drift termed the "founder effect", as a new virus population (in a newly infected plant, area, etc.) is started by a small number of genetic variants randomly chosen from the parent population as shown in Figure 2.1 (Flynn, 2010). Thus, founder effects result in a smaller diversity within a viral population.

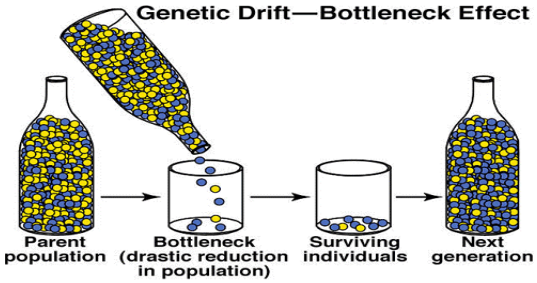


Figure 2.1: Genetic Drift-Bottleneck effect: occurs when a random event causes a drastic reduction in the size of the viral population, which in turn causes genetic drift. Bottlenecking usually reduces the genetic variability within a viral population, since some variant are lost from the parent population.

Image adapted from Flynn, (2010).

As a consequence of bottleneck events, mutations may be fixed at a higher level within the population than before the bottleneck occurred, independently of their selective value. This can lead to an effective population size below the threshold needed to ensure the transmission of the fittest variant. As a result, the viral population becomes progressively dominated by less fit genotypes—a process known as Muller's ratchet—and will succumb by a mutational meltdown (Novella, Domingo, & Holland, 1995).

The process of mutation meltdown refers to the process by which a small population accumulates harmful mutations, which leads to loss of fitness and decline of the population size, which may lead to further accumulation of deleterious mutations, which make them unstable (Zeldovich, Chen, & Shakhnovich, 2007). A population experiencing mutational meltdown is trapped in a downward spiral and will go extinct if the phenomenon lasts for some time. Usually, the deleterious mutations would simply be selected away, but during mutational meltdown, a number of individuals will suffer an early extinction.

2.1.5 Selection- fitness in RNA viruses

Selection is a directional process by which differential reproduction of genetically distinct mutants or variants occurs within a virus population and results in decreased population diversity (Roossinck, 2003). Selection can be positive (increase in the number of the fittest variants) or negative (also known as purifying selection, decrease in the number of less fit variants). Consequently, the fit variants will have more propagants in the next generation than less fit variants (Bhatt, Katzourakis, & Pybus, 2010). Much of the time a newly evolved mutant is of equal fitness with the progenitor, or if, altered, its fitness may be masked in the context of the quasispecies. Such mutation is selectively neutral (neither positive or negative) and its fate is determined not by selection but by chance events (Bhatt et al., 2010).

In spite of more recent theories of evolution, viruses appear to predominantly follow different rules than Darwin's original concepts of survival of the fittest (Saccheri & Hanski, 2006). Traditionally, fitness is defined as the relative reproductive ability in a defined environment. In the context of RNA viruses, the reproductive ability of a given

viral strain in a defined environment includes components such as replication, transcription and encapsidation rates as well as virion stability in the environment (Rambaut et al., 2008). As mentioned earlier, the quasispecies mutant spectrum is at 'selection equilibrium'. This equilibrium is usually metastable and will collapse when the more advantageous mutant appears and be replaced by new selection equilibrium (Lauring & Andino, 2010). Repeated fluctuations in population equilibrium can determine largely the selection-fitness of the virus populations.

Standard Darwinian evolution predicts that the fittest strain of an organism, the one that reproduces fastest in a given environment, will displace the other strains there (Domingo, 1998). However, whether or not the quasispecies nature of RNA virus follows the same rule, as Darwin's prediction is controversial. For example, a virus such as HIV is always mutating, and can mutate back and forth between different strains. For instance, HIV in a patient with advanced infection or AIDS exists as millions of related strains within the same patient (usually only one was transmitted, and then it evolved within that individual into countless slightly different variants). This makes HIV hard to treat, because some members of the quasispecies probably already have resistance mutations to a new drug even by chance alone, and these resistant viruses are ready to be selected and become much more prevalent when the drug treatment is started. It has been suggested that the quasispecies theory challenges the Darwinian dogma that natural selection favours individual variants (Mas, Lopez-Galindez, Cacho, Gomez, & Martinez, 2010). Perhaps the true dogma to be challenged is that all RNA viruses form quasispecies without considering the defining feature of this theory: that natural selection favours a viral populations rather than individual variants (Mas et al., 2010). The true accomplishment of quasispecies theory is that it has introduced useful and important evolutionary ideas into virology. The danger is that over-hyping the RNA virus quasispecies, and giving it properties that are poorly defined, will damage the credibility of virology as an evolutionary science (Saccheri & Hanski, 2006).

The effects of selection and genetic drift are often difficult to separate, because selection also results in a decrease within population diversity and may increase diversity between populations depending on different selection pressures. Muller's ratchet theory predicts that when a viral populations is small and the mutation rate is high, the population will decline in fitness due to the accumulation of deleterious

mutations in a "kind of irreversible ratchet mechanism"(Andersson & Hughes, 1996). However, the opposite appears to occur for large RNA virus population where it has been observed that repeated transmissions of large populations from host to host in a constant environment leads to significant increases in mean population fitness (Sardanyes & Elena, 2011). For example, long-term passage experiment on two large viral populations of the negative strand RNA virus VSIV serotype; BHK-21 cells and wild-type VSIV (Indiana, Mudd-Summers strain) indicated an exponential increase in fitness during large population transmissions. Result also revealed that fitness values reached a high-fitness plateau during which stochastic fitness variations were observed. This effect appears likely to be due to bottleneck effects on very high fitness populations. (Sanjun, Cuevas, Furi, Holmes, & Moya, 2007).

The degree to which a population is affected by drift and selection varies accordingly to circumstances. For a large population, where genetic drift occurs very slowly, even weak selection on a population may push its fitness upwards or downwards depending on whether the mutation is beneficial or deleterious (Sardanyes & Elena, 2011). However, if the population is very small, drift will predominate. In this case, weak selective effects may be overshadowed by drift. This situation conveys the general idea that all plant viruses are confronted with “universal barriers” in plants, imposing repeated transient decreases in their population size, thus making genetic drift a major constant driver of their evolution (Monsion, Froissart, Michalakis, & Blanc, 2008).

2.1.6 Genetic exchange in RNA virus

An increasing number of animal and plant viruses have been shown to undergo RNA-RNA recombination, which is defined as the exchange of genetic information between non-segmented RNAs (White, Enjuanes, & Berkhout, 2011). Only some of these viruses have been shown to undergo recombination in experimental infection of tissue culture, animals, and plants (White et al., 2011). However, a survey of viral RNA structure and sequences suggests that many RNA viruses were derived from homologous or non-homologous recombination between viruses or between viruses and cellular genes during natural viral evolution (White et al., 2011). The high frequency and widespread nature of RNA recombination indicate that this phenomenon plays a

more significant role in the biology of RNA viruses than was previously recognized. For example, recombination in Influenza A virus can cause an antigenic shift in the influenza genome resulting in a new subtype of the virus (Boni, de Jong, van Doorn, & Holmes, 2010). The resulting antigenic shift occurs when a cell becomes simultaneously infected by two different strains of type influenza. Through antigenic shift, the virus is able to largely circumvent the body's immune system, which may not be able to recognize and confer immunity to a new influenza strain even if an individual has already built up immunity to a different strain of the virus. The unusually broad range of hosts susceptible to influenza A appears to increase the likelihood that antigenic shift will occur. In particular, the mixing of strains that can infect birds, pigs, and humans is thought to be responsible for most antigenic shifts (Boni et al., 2010).

Since the discovery of RNA recombination in *Brome mosaic virus* (BMV) twenty years ago, it has become increasingly clear that many plant RNA viruses have the capacity to exchange genetic material with one another, and to acquire genes from their hosts. RNA viruses possess mechanisms for genetic exchange that make their reproduction "just as sexual as those in eukaryotes" (Chao, 1988). Whenever different genetic variants replicate in the same cell, genetic exchange can occur by recombination of genome regions that are switched between nucleotide strands, or by reassortment of complete genome segments in viruses with segmented genomes. These recombination and reassortment events allow the most efficient and environmentally adapted combinations of genes to emerge from the available genetic pool, increasing the potential for viral survival (Flynn, 2010).

By generating novel combinations of pre-existing nucleotide polymorphism, genetic exchange can potentially accelerate evolution by increasing the population genetic diversity upon which selection acts (Sztuba-Solińska et al., 2011). Evidence of recombination has been reported in some members of the genus Potyvirus, including *Plum pox virus* (PPV) (Cervera, Riechmann, Martin, & Garcia, 1993) and *Yam mosaic virus* (YMV) (Bousalem, Douzery, & Fargette, 2000). It has been suggested that recombination events can lead to gene specific virulence where recombination events occurred in one host may have enabled some recombinants to infect new hosts (Ohshima et al., 2002). For example, full- length sequence of WMV analysed in previous study revealed that this virus is very closely related to SMV; however the N-

terminal part of the P1 protein presented striking differences. The P1 of WMV is 135 amino acids longer than the P1 of SMV, and presents a very low percentage of identity (below 30%) with SMV in its N-terminal part. This N-terminal part is more than 80% identical to the equivalent fragment of two strains of BCMV. This study indicate that WMV may be the result of a putative recombination in the P1 between a BCMV and a SMV (Desbiez & Lecoq, 2004).

In contrast with recombination, reassortment in plant RNA viruses with segmented genomes has also been previously documented (Roossinck, 2003). Within the plant viruses studied, it is thought that the broad host range and worldwide distribution of *Cucumber mosaic virus* (CMV, genus *Cucumovirus*) is in part due to reassortment events, contributing to its enormous evolutionary success (Shi, Miller, Symons, & Palukaitis, 2003).

Genetic exchange results in novel genetic combinations that could have important phenotypic effects. It has been documented repeatedly that genetic exchange can result in dramatic changes in the properties of the viruses, and recombinant genotypes have been associated often with host range expansion, with host switches, or with increased pathogenicity (Codoñer & Elena, 2008). In contrast with genetic mutations, frequent recombination can create high fitness variants more rapidly than by mutation alone. Moreover, recombination might also purge deleterious mutations from virus populations, thereby preventing a dramatic decrease in fitness, by avoiding the effects of Muller's ratchets (Codoñer & Elena, 2008). However, recombination rates vary considerably among plant RNA viruses. This is due to the different levels of precision of viral replication proteins (i.e. variations in the error-prone nature of the replicase) during RNA replication and the presence or absence of recombinationally active sequences (recombinant hotspots) (Sztuba-Solińska et al., 2011). For example, phylogenetic analysis of seventy nine gene sequence alignments from thirty five negative-sense RNA viruses (a total of 2154 sequences) were carried out to determine the extent of homologous recombination (Chare, Gould, & Holmes, 2003). This study revealed that rates of homologous recombination in negative-sense RNA viruses were much lower than those of mutation. Results indicated that the possible reason for the low rate of recombination in negative-sense RNA viruses is that this process is hampered by the presence of the ribonucleoprotein complex (RNP), which never

disassembles from the RNA and may, therefore, affect the ability of RNA polymerase to switch templates during replication.

Environmental and host effects are likely to influence the rate of genetic exchange and selection, to ensure the survival of only the fittest variants. Nevertheless, recombination impacts on many aspects of the study of plant RNA viruses, as it might affect the durability of resistance genes, or the ecological risks of virus-resistant transgenic plants for future research especially considering the potential risks of generating new viral strains bearing properties different from the original ones.

2.1.7 Migration

Variation can also occur as a result of one virus population being transferred to another. This is known as gene flow (Coulibaly, Pasquet, Papa, & Gepts, 2002). The concept of gene flow is now widely used in population genetics which predicts patterns of sequence variation within and between closely related virus species to infer the fine-scale genetic structures of virus populations. Information on population structures - particularly that pertaining to stratification and admixture (i.e. gene flow) - is valuable in a variety of situations (Joannon, Lavigne, Lecoq, & Desbiez, 2010). Gene flow is especially important for plant RNA viruses in agro-ecosystems because it introduces new variants into existing virus populations, which can have any appreciable effect on the evolution of virus quasispecies (Shikano, Chiyokubo, & Taniguchi, 2001). More specifically, a detailed knowledge of virus population stratification can provide important insights into how virus genetic diversity generated through mutation and recombination is shaped into discernable taxonomic groupings: a process that involves natural selection and genetic drift in the context of epidemiological fluctuations in virus population sizes and the spatial movement of viruses across land-masses (Shikano et al., 2001).

The number of gene flow and virus population structure studies has greatly increased in recent years. For example, recent phylogenetic analysis of *Moroccan watermelon mosaic virus* (MWMV) from different origins, revealed the presence of three major groups with an interesting geographic structure in the African continent (Yakoubi et al.,

2008). Genetic distances between isolates correlated with geographic distances, suggesting efficient long-distance migration, either through aphid transmission, or more probably, through human migration contributed to the variation of the virus.

The deeper understanding of virus epidemiology and evolutionary history that can potentially be provided by studies of virus population structure is also directly applicable to the formulation of strategies for controlling the dissemination of viral diseases.

2.1.8 Solving the DsMV variation puzzle

Solving the genetic variation of RNA viruses is like having a box filled with thousands of puzzle pieces mixed together from an unknown number of jigsaw puzzles, and you want to solve each puzzle separately. How would you begin to sort this out? This puzzle box is analogous to what many researchers face when trying to discover the underlying genetic and environmental causes linked to complex forms of disease. With the prospect of disentangling the systems that regulate complex diseases, we often do not understand the exact sense in which a complex disease is "complex". Considering gene-environment interactions can improve our understanding of the causes of complex disease, it can also assist in developing diagnostic strategies. Rather than studying genetic and environmental factors separately, we must dump the puzzle pieces out of the box and begin to assemble the jigsaw puzzle one piece at a time in order to look at the whole picture.

Much of this research is centred on two main themes: (1) understanding the mechanisms of sequence variation, through experimental analysis and (2) to understand the nature of evolutionary processes in DsMV.

2.1.9 Aims of the study

Given the variety of genetic strategies, genome complexity, and global ecology of plant RNA viruses, the problem for research into virus variation is that it inevitably digresses

into a web of interlocking questions. How does the sequence population change over time and what are the relationships between variations within the different viral strains? To answer these questions, changes in sequence of different DsMV strains were monitored in taro propagated in the glasshouse over a one-year period.

The research examined the population-genetic structure of these strains by comparing virus sequences with DsMV sequences deposited in the NCBI database. Phylogenetic analysis was also carried out to better understand the evolutionary history between these samples and known DsMV sequences worldwide.

2.2 Materials and Methods

2.2.1 Virus isolates

All DsMV isolates were purified from the host taro, *Colocasia esculenta* shown in Table 2.1. Isolates NZ1, NZ2, NZ4, NZ6, B1, B1A, B1B and B152, were purified from plants grown in the glasshouse at the School of Biological Sciences, The University of Auckland. Isolates B1A, B1B and B152 had been studied previously and all symptomatic plants were identified by visual examination.

Table 2.1: List of glasshouse grown taro samples collected and analysed

Virus Isolate (Parent plant)	Host	Source
NZ1, 2, 4 and 6	*Colocasia sp.*	Taro grown in NZ, at Unitec Campus
B1, B1A, B1B, B152	*Colocasia sp.*	Taro imported into NZ from field trial

2.2.2 Propagation of plants

C. esculenta infected with DsMV isolates NZ1, NZ6 or B1 was propagated to give rise to first and second-generation propagant plants. The upper half-inch of the corm with the shoot, or smaller axillary corms were separated and placed into fresh soil and maintained in the glasshouse at the University of Auckland. First propagation of the parent plant was performed in 2009 into four individual plants (first generation propagants 1-4) shown in Figure 2.2. The second-generation propagants were created in 2011 by splitting each first generation plants into four propagants. The list of propagants is shown in Table 2.2. Plants were allowed to grow for 3 to 4 months to allow young shoots to develop. For the second generation, only two propagants survived. For the original parent plants, samples were collected from mature and young leaves as shown in Figure 2.3. For the first and second-generation propagants, only

young leaves were sampled. All leaf samples were placed immediately at -80°C until required for RNA extraction.

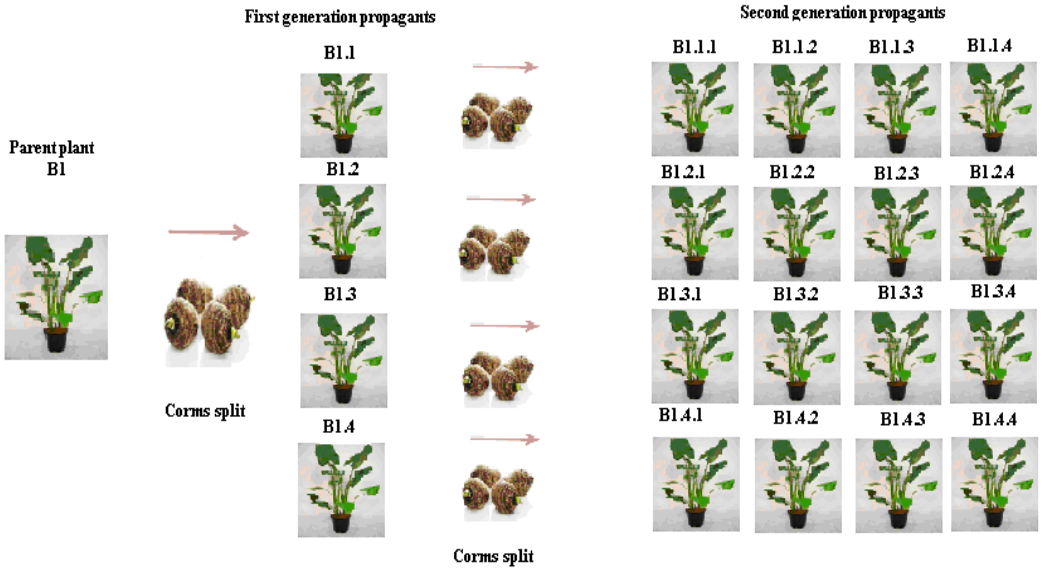


Figure 2.2: Propagation process for taro infected with DsMV- B1 into two generations of propagant plants for temporal analysis. This process was repeated for taro infected with DsMV- NZ1 and – NZ 6.

Image adapted and modified from http://realornamentals.com/plant-store.

Table 2.2: List of propagants collected and analysed as part of the temporal analysis.

Parent	*First Generation propagants	**Second Generation propagants
NZ1	NZ 1.1, 1.2, 1.3, 1.4	NZ 1.1.1, NZ 1.1.2 NZ 1.2.1, NZ 1.2.2 NZ 1.3.1, NZ 1.3.2 NZ 1.4.1, NZ 1.4.2
NZ6	NZ 6.1, 6.2, 6.3, 6.4	NZ 6.1.1, NZ 6.1.2 NZ 6.2.1, NZ 6.2.2 NZ 6.3.1, NZ 6.3.2 NZ 6.4.1, NZ 6.4.2
B1	* B 1.1, 1.2	**B 1.1.1 **B 1.2.1

** Two plants survived from the propagation process

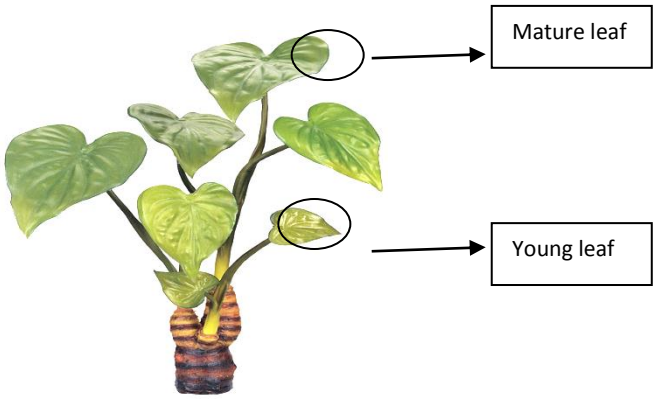


Figure 2.3: Collection site for mature and young infected leaves.

Image adapted and modified from https://miami.miamimart.com.

2.2.3 Preparation of total RNA

Total RNA extractions from infected taro were done using the Spectrum Total RNA kit (Sigma Aldrich). Up to 100 mg of leaf tissue was ground into a fine powder in liquid nitrogen. Lysis buffer supplemented with 2-mercaptoethanol (2-ME) was added to the powder, which was then homogenized by vortexing. The homogenate was centrifuged for 3 minutes at 13,200 rpm in an Eppendorf tabletop centrifuge 5417C (Eppendorf, Hamburg, Germany). The supernatant was transferred to a filtration column and was centrifuged for 1 minute at 13,200 rpm. The clarified flow-through lysate was collected and 500 µL of binding buffer was added to the lysate. The lysate was vortexed briefly and 700 µL of the lysate was transferred to the binding column, which was centrifuged for another minute at 10,000 rpm. The eluate was discarded and 500 µL wash solution 1 was added to the column, which was then centrifuged for 1 minute at 13,200 rpm. A further wash step with 500 µL of diluted wash solution 2 was added to the column, which was then centrifuged again for 1 minute at 13,200 rpm. The wash step was repeated again and the column was transferred to a clean micro-centrifuge tube, and then added 50 µL of RNase-free water was added to the centre of column to elute the total RNA. The column was allowed to sit for one minute and then centrifuged at 13,000 rpm for two minutes. N.P. purity and concentration of total RNA was measured using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington,

and Detroit, USA). The samples were then split into three sets of 10 µL aliquots for short-term (-20 ^{O}C) or long-term (-80 ^{O}C) storage.

2.2.4 First strand cDNA amplification from DsMV RNA

First strand cDNA was synthesized via reverse transcription of mRNA using the PV1/SP6 primer (Table 2.3) with a cDNA Flexi kit (Quanta Biosciences Gaithersburg, USA) as shown in Figure 2.4.The primer was annealed to the RNA as follows: 4 µL (350-450 ng) total RNA was added to a mixture of 2 µL 10pmol reverse primer, 7 µL RNase free water and 2 µL of GSP enhancer, giving a total reaction volume of 15 µL. The reaction was then incubated at 65^{O}C for 5 minutes, and chilled on ice for 1 minute. cDNA synthesis was then carried out by first combining 1 µL qScript reverse transcriptase and 4 µL qScript Flex reaction mixture (5x). This 5 µL solution was mixed thoroughly with first reaction solution, and synthesis was carried out at 42°C for 60 min, 85°C for 5 min and a holding temperature of 4°C using the TECHNE thermocycler (Staffordshire UK, Model FTGRAD2D). This reaction gave a total volume of 20 µL of cDNA. cDNA was used immediately for PCR or stored at -20^{O}C.

Table 2.3: Primers utilised in cDNA and RT-PCR

Primer	Sequence	References
PV1/SP6	5'-GATTTAGGTGACACTATAG (T) 17(A/G/C)-3'	Mackenzie et. al., 1998
MJ1	5'-ATGGTHTGGTGYATHGARAAYGG-3'	Marie-Jeanne, et .al, 2000
MJ2	5'-TGCTGCKGCYTTCATYTG-3'	Marie-Jeanne et.al. 2000
SP6	5'- GATTTAGGTGACACTATAG -3'	Macrogen Inc. South Korea
T7PRO	5'- AATACGACTCACTATAGG-3'	Macrogen Inc. South Korea
Oligo dT	5'-ACTATCTAGAGCGGCCGCTTT16-3'	Quanta Bioscience kitset

Single letter code H=A/C/T, Y=C/T, R=A/G, K=G/T

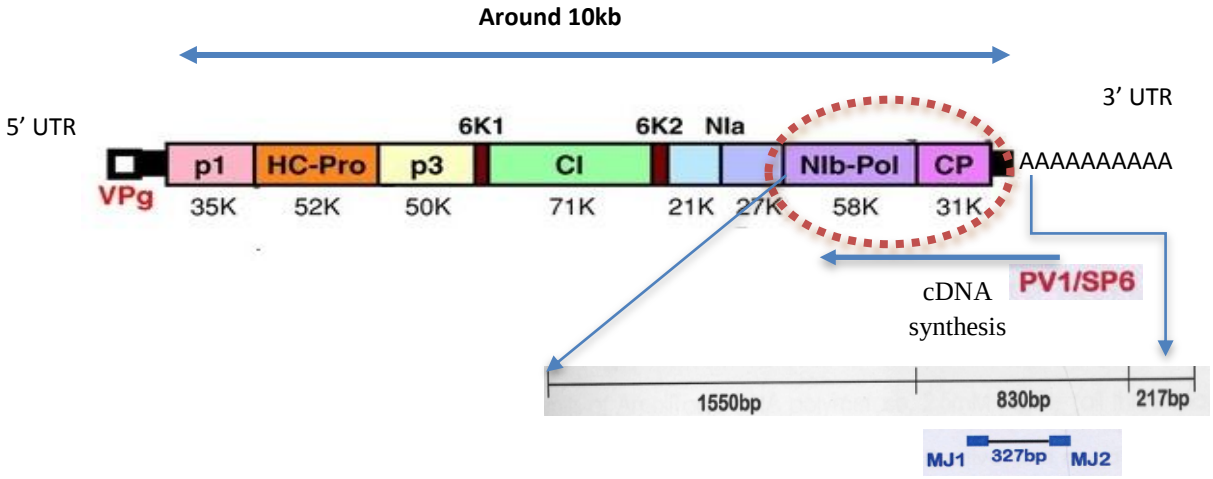


Figure 2.4: Schematic diagram for the priming strategy of the CP region

2.2.5 PCR amplification with primers MJ1 and MJ2

First strand cDNA (5 µL) was mixed with 2 µL of 10µM forward primer (MJ1) and 2 µL of 10 µM reverse primer (MJ2), 12.5 µL of 2X Go®Taq Green master mix (Promega) and RNase free water to make a final total volume of 25 µL. Each PCR had a negative control with 5 µL of sterile water instead of cDNA. Thermocycling was undertaken in a TECHNE thermocycler (Staffordshire UK, Model FTGRAD2D). The PCR programme began with an initial denaturation at 94^{O}C for 3 minutes, followed by 40 cycles of 94^{O}C for 1 minute, 55^{O}C for 45 seconds and 72^{O}C for 2 minutes. After that there was 10 minutes at 72^{O}C, and then PCR was held at 15^{O}C or stored at -20^{O}C for future use.

PCR product (5 µL) was analysed by electrophoresis through 1.5% agarose in 1X TBE buffer for 1.5 hours at 75V and compared with 100 bp marker (Dnature, Gisborne, New Zealand).

2.2.6 PCR product purification

The Invitrogen PureLink PCR Purification kit (Invitrogen, Grand Island, NY, USA) was used to purify PCR products away from any primer dimer formed. PureLink Binding

Buffer with isopropanol (four volumes) was added to one volume of the PCR product. Each sample was transferred to a PureLink Spin Column and centrifuged at room temperature at 10,000 × *g* for 1 minute in an Eppendorf tabletop centrifuge 5417C (Eppendorf, Hamburg, Germany). The flow through was discarded and the spin column was replaced back in the collection tube. Wash Buffer (650 µL) with ethanol was added to the column and centrifuged at room temperature at 10,000 × *g* for 1 minute. The flow through was discarded from the collection tube and the wash step was repeated. The DNA was eluted into a fresh tube using 50 µL of Elution Buffer (10 mM Tris-HCl, pH 8.5) following incubation at room temperature for 1 minute. The column was centrifuged at maximum speed for 2 minutes. The elution tube containing the purified PCR product was stored at -20°C for long term storage. In some instances, PCR products were sent to Macrogen Inc. South Korea for dideoxy sequencing.

2.2.7 Gel purification of PCR product

Some PCR products had to be purified away from non- specific products. To do this, DNA bands were excised from agarose gels and purified using a Gel DNA Recovery Kit (Zymogen, CA, USA), following the manufactures protocol. Each volume of excised DNA was added to three volumes of ADB buffer and incubated at 37-55 °C for 5-10 minutes until the gel slice was completely dissolved. Melted agarose solution was transferred to a spin column and centrifuge at ≥10,000 x *g* for 30-60 seconds in an Eppendorf tabletop centrifuge 5417C (Eppendorf, Hamburg, Germany). The flow through was discarded and 200 µL of Wash Buffer was added to the column and centrifuged at ≥10,000 x *g* for 30 seconds. DNA was eluted with 6-10 µL of water by centrifuging at ≥10,000 x *g* for 30-60 seconds. DNA was stored at -20°C for long term storage.

2.2.8 Single strand conformation polymorphism analysis (SSCP)

Each diluted sample was loaded onto a non-denaturing 12% polyacrylamide gel made up with acrylamide: NN’ bisacrylmide at 29:1 made up in 1x TBE buffer. SSCP

polyacramide gel was run using Bio-Rad Protein II xi Cell vertical gel electrophoresis unit (Bio-Rad laboratories, USA). Electrophoresis was carried out in 1 X TBE buffer, using constant power (180V) for 4 hours at 4ºC. Ladder marker (100 bp) (Dnature, New Zealand, Gisborne) was loaded onto the gel along with the samples, and act as a marker for migration pattern.

2.2.9 Silver staining protocol

Following PAGE, the DNA was visualized following the protocol of Benbouza, et. al. (2006). The gels were fixed in 10% acetic acid/0.5% ethanol solution and gently agitated for 2 min. The gels were placed on a rotary shaker and rinsed quickly in sterile distilled water (30 seconds per wash). Fresh sterile distilled H_2O was used in each wash. The gels were then transferred to a staining solution containing 0.15% (w/v) silver nitrate and 0.15% (v/v) 37% formaldehyde in sterile nanopure water and agitated for 7 minutes. After a minute wash in sterile distilled H_2O, the gels were developed at room temperature in a solution of 0.15% (v/v) 37% formaldehyde, and 0.0015% sodium hydroxide. The developing process was stopped by adding new fixer solution, when bands were clearly resolved with a low background. The image was digitally capture using a MAC iphone4® and contrast was digitally enhanced using MAC Aperture software®. All reagents were made fresh with each run and glassware were soaked in chromic acid and rinsed in sterile distilled water prior to reagent preparation.

2.2.10 Cloning and Sequencing

PCR products of DsMV- NZ1, -NZ6 and -B1 CP fragments were ligated into pGEM-T EASY (Promega, Madison, Wis.) and *Escherichia coli* strain JM109 were transformed by recombinant DNAs through process of (1) heat—shock transformation, (2) culturing of transformants, and (3) finally plasmid extraction following manufacture's protocol as described below.

Ligation

Gel purified DNA (3 µL) was added to a ligation reaction containing 5 µL 2X rapid ligation buffer, 1 µL of 5 ng/ µL pGEM-T Easy vector and 1 µL of 100 units T4 DNA ligase (all these reagents were from Promega Corporation USA). The total volume was 10 µL and components were mixed and incubated overnight at 4^{O}C in a water bath.

Transformation

Five microliters of the ligation mixture was mixed with 30 µL of the *Escherichia coli* JM109 cells (thawed on ice, Promega Corporation USA), and then incubated on ice for 20-30 minutes. The cells were heat shocked in a 42^{O}C water bath for 40 seconds, and immediately returned to ice for 2 minutes. Super optimal broth with catobolite repression (SOC) medium (250 µL) (Invitrogen Life Technologies Carlsbad CA, USA.) was added and incubated at 37^{O}C for 60-90 minutes with constant shaking at 150 rpm.

Culturing of transformants

Cells (100 µL) were spread onto duplicate Luria-Bertani (LB) agar plates (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCL, 15 g/L agar, pH 7.2, in distilled water) which were previously prepared with 2 mg/plate isopropyl β-D-1-thiogalactopyranoside (IPTG) (Invitrogen Life Technologies, Carlsbad CA, USA) 0.8 mg/plate ampicillin (Applichem, Darmstadt, Germany), and 0.8 mg/plate 5-Bromo-4-chloro-3-indolyl-ß-D-galactopyranoside (X-gal) (Invitrogen Life Technologies Carlsbad Ca, USA.). Plates were left to air-dry and then incubated upside down overnight at 37^{O}C.

A minimum of five white colonies per plate was selected and each was sub-cultured into 5 mL of LB broth containing 800 µg ampicillin and an overnight incubation at 37^{O}C with constant shaking. Cultures that grew effectively were then screened through PCR.

Colony screening via PCR protocol

A 2 µL aliquot of *E.coli* overnight culture in LB broth with 800 µg ampicillin was added to PCR reaction containing 2 µL of 10 µM forward primer SP6 primer and 2 µL of 10 µM T7PRO, 12.5 µL of 2X Go® Taq green master mix (Promega) and RNase free water to make a final total volume of 25 µL.

Thermocycling was undertaken in a TECHNE thermocycler. This PCR programme began with initial denaturation at 95^{O}C for 5 minutes, followed by 35 cycles of 95^{O}C 1 minute, 55^{O}C 1 minute and 68^{O}C 1 minutes. After that there was a 10 minutes at 68^{O}C and PCR products were used immediately or stored at -20^{O}C.

Plasmid extraction protocol

Recombinant plasmids were purified utilizing the Zyppy Plasmid Miniprep kit (Zymo Research) following the manufacturer's protocol. An *E.coli* overnight culture in LB (1.5 mL aliquot) was added to a sterile 2 ml microcentrifuge tube, then centrifuged at 16,100 x g for one minute in an Eppendorf 5417C Tabletop centrifuge (Eppendorf, Hamburg, Germany); the supernatant was discarded without disturbing the pelleted cells. An additional 1.5 mL of overnight culture was added to the tube, centrifuged and the supernatant removed as before. The pelleted cells were resuspended in 600 µL of RNase free water. Lysis Buffer (7X) (100 µL) was added and mixed thoroughly by inverting the tube 4-6 times in order to lyse the cells. Neutralization buffer was added subsequently (350 µL) and mixed thoroughly by inverting the tube 4-6 times. Centrifugation was performed for 2 minutes at 13,000 rpm, and the supernatant (containing plasmid DNA) was transferred to a spin column and centrifuged for 1 minute at 13,000 rpm; the flow through was discarded. Endo wash buffer (200 µL containing guanidine hydrochloride) was added and centrifuged for 1 minute at 13,000 rpm to remove trace nuclease activity associated with strains of *E.coli*, and the flow-through was discarded. Wash buffer (400 µL containing ethanol) was added and centrifuged for 1 minute at 13,000 rpm to remove contaminants. Addition centrifugation was performed for 1 minute to remove residual wash buffer. The spin column was placed in a sterile 1.5 mL microcentrifuge tube and plasmid DNA was eluted with 30 µL of water by centrifuging at 13,000 rpm for 1 minute.

The quantity and purity of the extracted plasmid DNA was assessed using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies Wilmington, DE, USA). Plasmid DNA was stored at –20^{O}C.

Restriction enzyme digestion:

To check for the presence of inserts of the correct size (327bp), a restriction digest was carried out. 1 µL extracted plasmid was added to a total of 9 µL of mixture containing 1 µL of reaction buffer (3x) (Fermantas, USA), 1 µL of EcoR1 R1, 7 µL of sterile water.

The reaction was incubated at 37^{O}C for 30-60 minutes. A 5 µL aliquot was electrophoresed through a 1% agarose gel in 1 x TBE at 75V for 30 minutes, and visualized under UV light.

2.2.11 Sequencing recombinant plasmid

Three representative clones for each recombinant plasmid were sent to Macrogen Inc., South Korea and sequenced from one end utilizing the universal primer T7PRO (Table 2.3). Sequences were edited to remove vector and poor quality sequence using GeneiousPro 5.5.4 (Drummond et.al. 2011). The clone sequences were compared with each other and publicly available DsMV sequences. Pairwise percentage identities were calculated using the distance calculation option in GeneiousPro 5.5.4.

2.2.12 Sequence PCR products

Some RT-PCR products were purified and sent to Macrogen Inc., South Korea for direct sequencing. The nucleotide sequences of the partial CP gene of DsMV were generated with the potyvirus primer MJ1 at using a AB3730xI DNA Analyser (Applied Biosystem). The quality of the sequences was examined using the software GeneiousPro 5.5.4 and poor quality sequence removed as described in section 2.2.10.

2.2.13 Phylogenetic analysis

The 327 bp nucleotide sequences of the partial CP gene for each strain were analysed using the application software GeneiousPro version 5.5.4. The DsMV sequences were identified using DsMV- NZ1 as the query sequence in BLASTn search (www.ncbi.nlm.nih.gov) carried out on the NCBI website using the default parameter. Genbank Accession numbers for each sequence from the NCBI database are given in Appendix 1. Pairwise percentage identities between each sequence were calculated using the distance calculation option in the GeneiousPro 5.5.4 software. Multiple

alignments were created using Clustal W (Higgins et. al, 2005) in GeneiousPro 5.5.4 accepting the default parameter.

Phylogeny relationships were determined by Neighbour Joining using default parameters and this was also performed using the Maximum-Likelihood GTR model + invariant sites with four gamma categories implemented in GeneiousPro 5.5.4. Statistical support for tree branching patterns was determined from 1,000 bootstrap replications. *Passiflora foetida virus Y* (PFVY), *Zucchini yellow mosaic virus* (ZYMV) and *Passion woodiness virus* (PWV) were used in the phylogenetic trees as an out-group.

2.2.14 Selection analysis

In order to identify the selection pressure exerted on the partial CP gene, non-synonymous/synonymous substitution ratios (d_N/d_S) were estimated using public free web-interface programme "ka/ks calculation tool" (http://services.cbu.uib.no/tools). The PCR sequences for all isolates as well as the corresponding public domain sequences were entered. Comparisons of the sequences were carried out and average d_N/d_S ratio was identified across the length of the sequences using the default parameter.

2.3 Results

2.3.1 Integrity check of RNA extraction

The quality of the total RNA extracted from infected taro leaves was determined by absorbance ratio and electrophoresis. High quality RNA has a A_{260}:A_{280} value of 1.7-2.0 (Liao et al., 2004) and A_{260}:A_{230} of 2.0-2.3 (Asif, Dhawan, & Nath, 2000). The former ratio indicates the level of protein contamination while the latter indicates solvent contamination. All RNA samples showed the expected ratios (data not shown) indicating they were pure.

Total RNA extracted from taro leaves was also electrophoresed through 1% agarose/1xTBE gels to determine the integrity (Figure 2.5). Ribosomal RNA bands (28S and 18S rRNA) were observed in all samples with little smearing. Furthermore, similar intensity was observed for the 28S band and 18S band which indicates the RNA was intact and suitable for use.

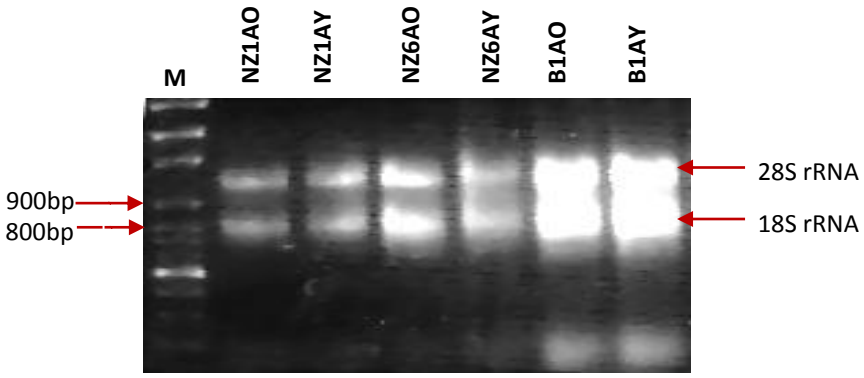


Figure 2.5: Total RNA from *C.esculenta* leaves infected with DsMV- NZ1, NZ6 and B1 were electrophoresed through a 1% agarose/ 1xTBE. Lane: M: 100bp Ladder DNA Marker (100ng); AO denotes mature leaf and AY denotes young leaf.

2.3.2 Optimization of RT-PCR

Amplification of a PCR product from mRNA first requires synthesis of complementary DNA (cDNA) using reverse transcriptase. This synthesis can be primed using a primer that can bind to the poly A tail of the mRNA (e.g. oligo dT or PV1/SP6, Table 2.3) or the reverse primer used in PCR (in this case MJ2, Table 2.3). These three primers were compared for their ability to prime cDNA synthesis from the total RNA prepared using the SPECTRUM total RNA extraction kit. Figure 2.6 shows that PCR amplification with the degenerate primers MJ1/MJ2 was most efficient when cDNA synthesis was primed with PV1/SP6. Surprisingly, priming with oligo dT did not result in high levels of PCR product. This primer does not have a nucleotide to anchor it to the very 3' end of the viral 3'UTR, thus, this primer can bind to anywhere within the poly A tail. The proportion of cDNA synthesized that was long enough, did not appear sufficient for amplification with MJ1 and MJ2 primers (see Figure 2.4 for where these primers bind). The degenerate primer MJ2 also appeared to bind efficiently. However, for this research PV1/SP6 was chosen to prime the cDNA synthesis from the Poly A tail.

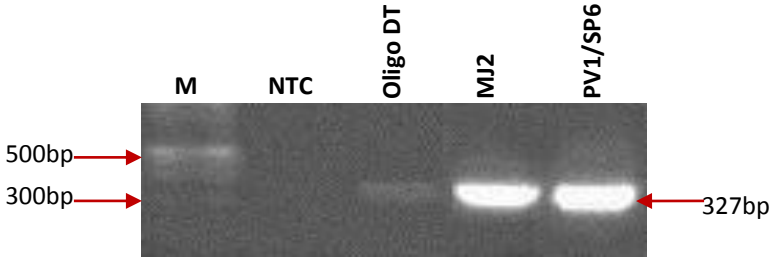


Figure 2.6: Comparison by RT-PCR of different primers (Oligo dT, MJ2 and PV 1/SP6) used for priming synthesis of first strand cDNA. cDNA were then used as template for amplification of DsMV using the primers MJ1/MJ2. Lane M: 100bp Ladder DNA Marker (100ng) and NTC: No template control.

2.3.3 RT-PCR of partial DsMV CP gene using the degenerate primer pair MJ1/MJ2

The degenerate primer pair MJ1/MJ2 (Marie-Jeanne, Ioos, Peyre, Alliot, & Signoret, 2000) have been used by others (Babu et al., 2011; Farreyrol et al., 2006) for the amplification of a 327 bp PCR product from a range of Potyviruses. The primers were shown by Babu et.al, (2011) to amplify this sized fragment from DsMV and were used

to amplify PCR products used in this study. Figure 2.7 shows the amplification of the expected 327 bp products from a range of DsMV strains, with no amplification from healthy leaf tissue showing these primers amplify the virus specifically.

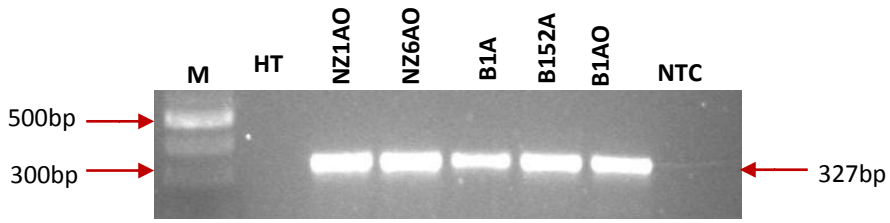


Figure 2.7: Agarose gel electrophoresis (1.5% agarose/ 1x TBE) of the expected 327 bp PCR product amplified with MJ1/MJ2 primers. Lane M: 100bp Ladder DNA Marker (100ng); NTC: Negative template control and HT: Healthy taro.

2.3.4 SSCP analysis of DsMV isolates between old and mature leafs within parent plants

Single strand conformation polymorphism (SSCP) has been used in many variability studies to characterize variation in virus populations (Hall, French, Morris, & Stenger, 2001). This method, applied to PCR products, makes possible rapid identification of any modification of the nucleic acids that affects the 3-dimensional structure of the corresponding single strand (ss) DNA molecules (Delaunay, Rolland, & Jacquot, 2009). Based on the differential migration of ssDNA in a nondenaturant PAGE, the SSCP procedure allows simultaneous processing of samples and the description of their genetic diversity (Delaunay et al., 2009).

Single-strand conformation polymorphism (SSCP) was used in this study to determine if viral population from different sites (old vs. young leaf) within a parent plant were substantially different. The analysis was carried out on taro plants infected with different strains of DsMV present in the glasshouse since 2010. Based on the SSCP analysis (Figure 2.8), no differences were observed between the SSCP patterns for all old and young leaves for all strains, suggesting little or no variation within plant.

Varying SSCP pattern was observed between DsMV strains and isolates. The patterns for DsMV-NZ and B strains were different. Further, patterns varied between NZ isolates (Figure 2.8a) but no variation was seen between B isolates (Figure 2.8b). DsMV-B isolates showed two bands, suggesting one sequence amplified during PCR. The pattern of DsMV NZ2 and NZ4 also indicated one viral sequence was amplified. The migration of the bands for NZ2 and NZ4 was slightly different than the B isolates suggesting a different sequence was amplified. In contrast, DsMV NZ1 and NZ6 showed four bands suggesting two different viral sequences were amplified. The migration pattern of bands for DsMV NZ1 and NZ6 appeared to be the same, suggesting they were comprised of the same sequence. It should be noted that the marker (M) used in the SSCP analysis do not give size estimates as they are linear, while the SSCP patterns are based on different conformation shape of the ssDNA structure. The marker (M) allows comparison of migration patterns.

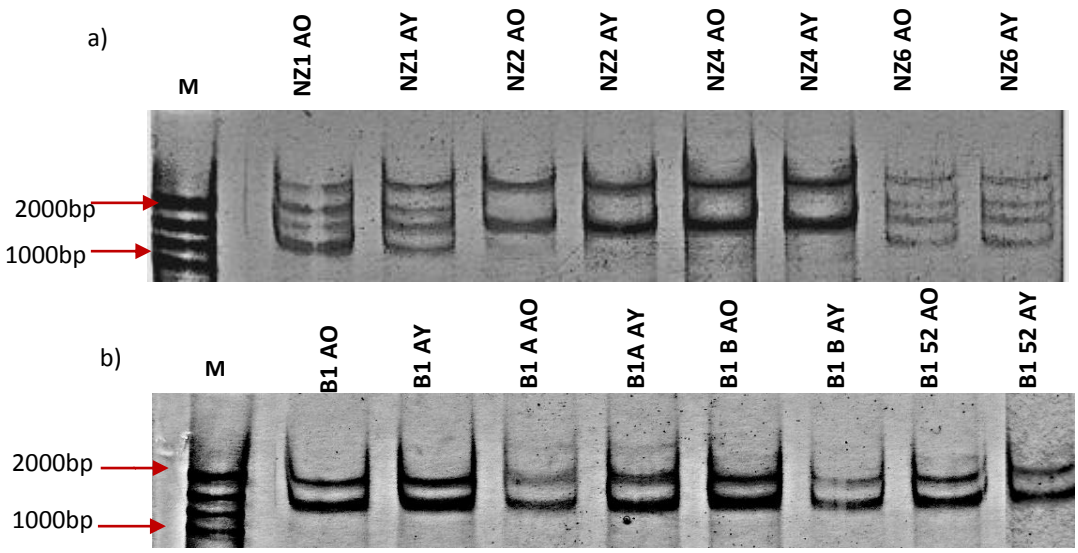


Figure 2.8: SSCP analysis and comparison of fragments amplified from RNA of isolates of DsMV isolates (mature and young leaves) using primers MJ1/MJ2 samples were run at 180 V for 4 hours in 12% PAGE in 1X TBE buffer at 4ºC. Note: (a) parent plants from NZ isolates and (b) parent plants from B isolates. Lane marked M: 100bp Ladder DNA Marker (100ng); AO denotes mature leaf and AY denotes young leaf.

2.3.5 Cloning MJ1/MJ2 RT-PCR products

PCR products of the predominant and variant SSCP patterns for DsMV- NZ1, and -NZ6 (observed in Figure 2.8) were ligated into pGEM-T Easy and transformed into *Escherichia coli* (*E.Coli*) strain JM109 to separate out all possible sequences from the PCR. Additionally, PCR product displaying the predominant SSCP pattern for DsMV-B1 was also cloned into pGEM-T Easy, as there was indication of a possible pair of bands that migrated very closely to the main pair. Inserts of five plasmids derived from each PCR product cloned were identified by colony PCR using SP6/ T7PRO primer pair (Table 2.3) and restriction digestion with *Eco*R1 restriction endonuclease. Figure 2.9a shows amplification of the expected 468bp product containing the 327 bp insert plus the vector sequence from a DsMV-NZ1 clone. Figure 2.9b) shows an *Eco*R1 digest of a clone for the DsMV- NZ1 strain showing the ~ 3 kb vector and 339 bp insert fragment. Similar patterns were observed for clones of each virus strain.

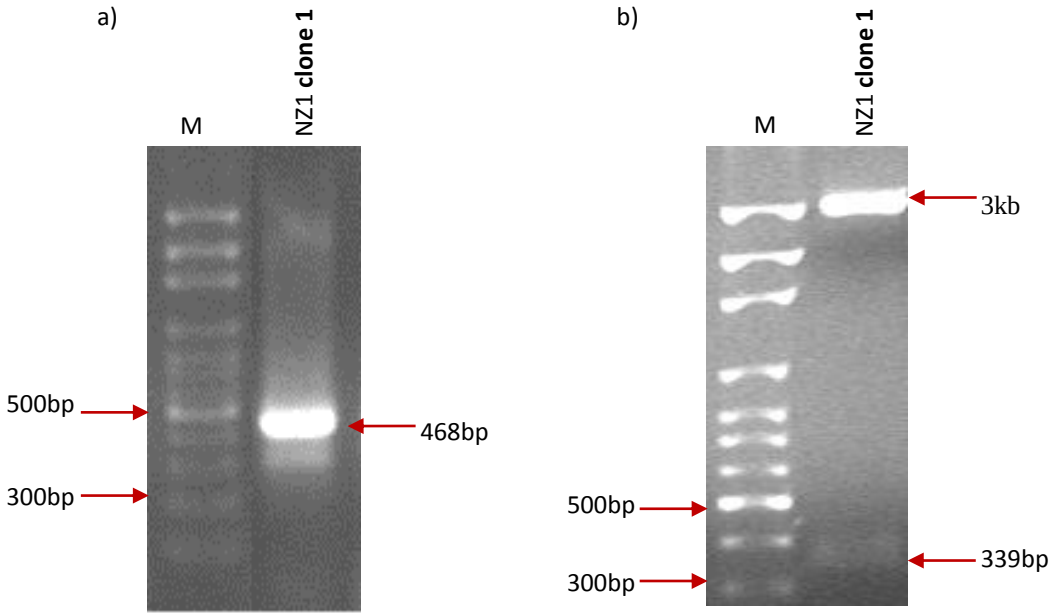


Figure 2.9: Verification of cloned DsMV-NZ1 MJ1/MJ2 PCR product by (a) colony PCR with primer pair SP6/T7 PRO and (b) restriction fragments of NZ1 clone1 digested with EcoR1. Lane: M: 100bp Ladder DNA Marker (100 ng).

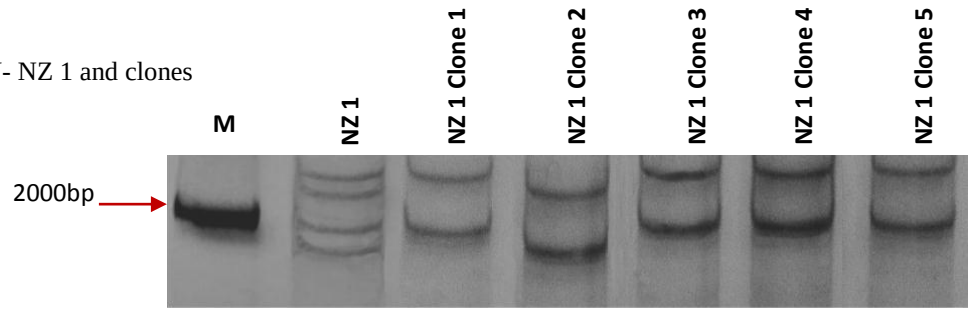
2.3.6 Identification of SSCP bands

Following cloning of the DsMV strains for NZ1, NZ 6 and B1, the insert of several clones were PCR amplified with MJ1/MJ2. SSCP analysis was then carried out on the PCR product on each cloned sequence to determine which SSCP band corresponded to which clone/sequence. SSCP on each clone was compared with the SSCP pattern obtained from infected leaf material. In this way, clones corresponding to each SSCP band could be identified.

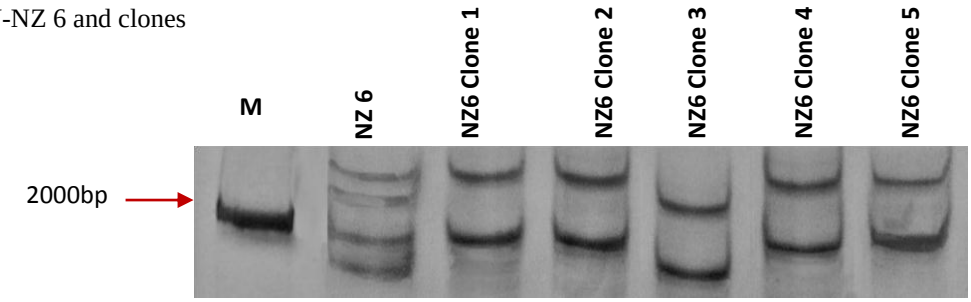
It can be seen in Figure 2.10a) for DsMV- NZ1, clones 1, 3, 4 and 5 had the sequence corresponding to one pair of fragments in the original SSCP pattern, while clone 2 had the other. For DsMV-NZ6 (Figure 2.10b), clones 1, 2, 4 and 5 had the sequence corresponding to one pair of the SSCP bands, while clone 3 had the other.

For DsMV-B1 (Figure 2.10c), there were indications of another pair of bands that migrated closely with the main pair; however, these could not be separated using different concentrations of acrylamide or running the gel for longer times (data not shown). These shadow bands were observed with DsMV -B1A and -B1B (Figure 2.8b). The separation of these bands was achieved; clones 3 and 4 had the main bands, while clones 1 and 2 had the shadow bands.

(a) DsMV- NZ 1 and clones



(b) DsMV-NZ 6 and clones



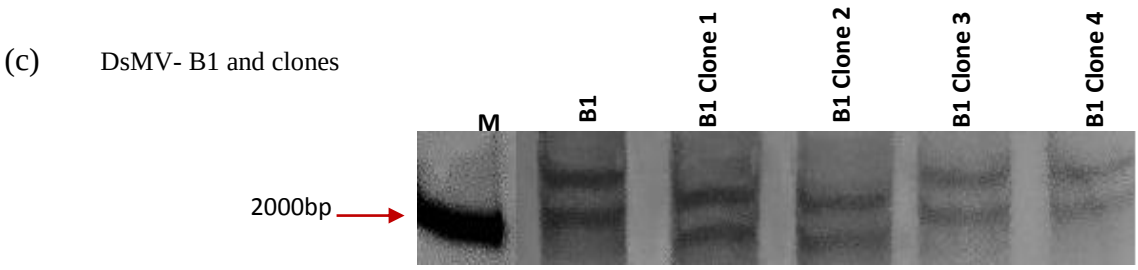


Figure 2.10: PCR–SSCP patterns derived from clones (a) DsMV-NZ1 (b) DsMV-NZ 6 and (c) DsMV-B1 compared with that from taro leaves infected with each virus strain. M: 100bp Ladder DNA Marker (100ng).

2.3.7 Sequence analysis of DsMV – CP of NZ1, NZ 6 and B1 clones

The SSCP profiles seen for each DsMV clone suggested each contained a different nucleotide sequence. To investigate this, three representative clones from each virus strain was sequenced directly with universal primer T7PRO (Table 2.3). Sequence comparisons were then carried out between clones of each strain as well as between strains.

Figure 2.11 shows the sequences for each clone compared to the others for the same strain. For DsMV- NZ1 it can be seen that clone 2 differed from clones 1 and 3 by two nucleotides (Figure 2.11a), which was sufficient to cause a change in the secondary structure thus altering the mobility in the SSCP assay (Figure 2.10a). For DsMV- NZ6, clone 3 differed from clone 1 and 2 by several nucleotides (Figure 2.11b), causing a mobility change Figure 2.10b. DsMV- NZ6 clones 1 and 2 differed by two nucleotides, which did not alter their mobility. For DsMV- B1 clone 3 differed from clones 1 and 2 by three nucleotides (Figure 2.11c), which caused a change in mobility (Figure 2.10c). This mobility difference was difficult to discern without cloning the fragments (Figure 2.8b compared with Figure 2.10c).

As previously described, SSCP profiles for DsMV –NZ1 and –NZ6 (Figure 2.8) showed the same pattern; however a multiple sequence alignment between clones of DsMV –NZ1 and –NZ6 revealed two nucleotide differences (Figure 2.11d), between DsMV –NZ clones and –NZ6 Clone 1. DsMV-NZ6 clone 2 was identical to –NZ1 clones 1 and 3. On the other hand, DsMV- NZ6 clone 3 shows several nucleotide differences

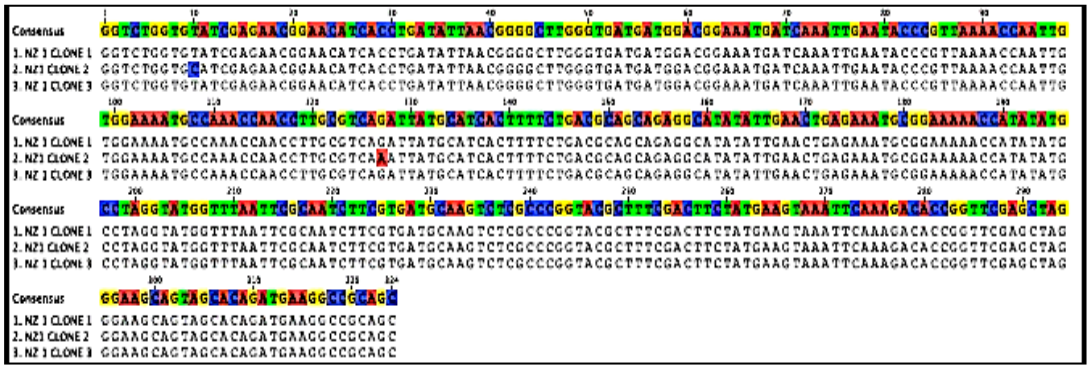
between all clones of –NZ1. This observation suggested that differences of two or several nucleotides sometimes resulted in a changed SSCP pattern, and relatively large numbers of differences did not necessarily do so, therefore there is no simple relationship between the number of nucleotide differences and SSCP patterns.

Pairwise comparisons of nucleotide sequences showed 80%-100% identity between all clones (Table 2.4). DsMV-NZ6 clones shared much higher nucleotide identity with DsMV –B1 clones (89%-100%), than DsMV-NZ1 clones (89%-97%). Overall the results, showed that they all belonged to the same species based on the demarcation value of 76% nucleotide sequence identity for Potyviruses (Adams et al., 2005).

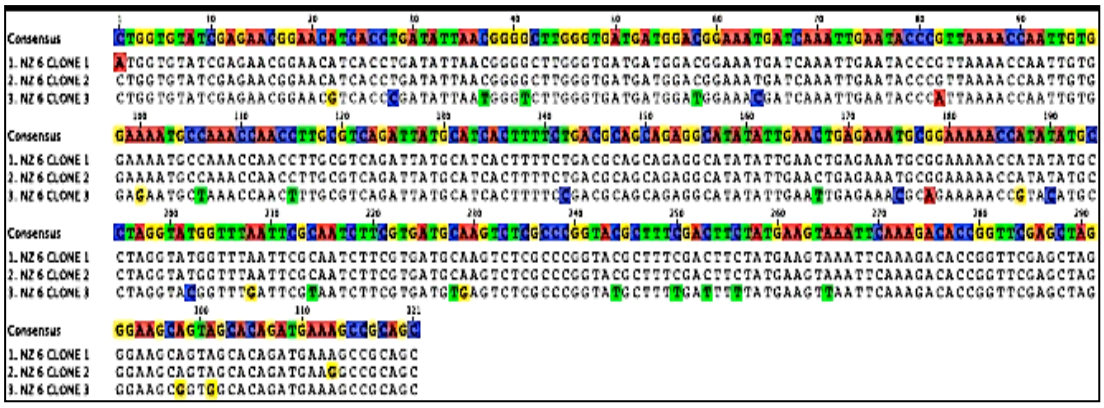
When the deduced amino acid sequences of these isolates were compared, there were no differences between DsMV-NZ1 and –NZ6 apart from –NZ6 clone 3; DsMV –NZ6 Clone 3 had acquired a single base mutation resulting in an amino acid substitution from alanine (ala) to valine (val) (Figure 2.11e), making it the same amino acid sequence as the DsMV- B clones. The amino acid sequences only varied by this one amino acid, resulting in 98%-100% sequence identity between all clones (Table 2.4).

Based on the analysis of sequences from the clones of each viral strain, some of these clones had identical sequences (designated the consensus sequence), and the other clones harboured one to several point mutations compared to the consensus sequence. This result suggested predominant sequences coexisting within the same hosts and viral populations within infected host existing as quasispecies, i.e. as a population of nearly identical sequences.

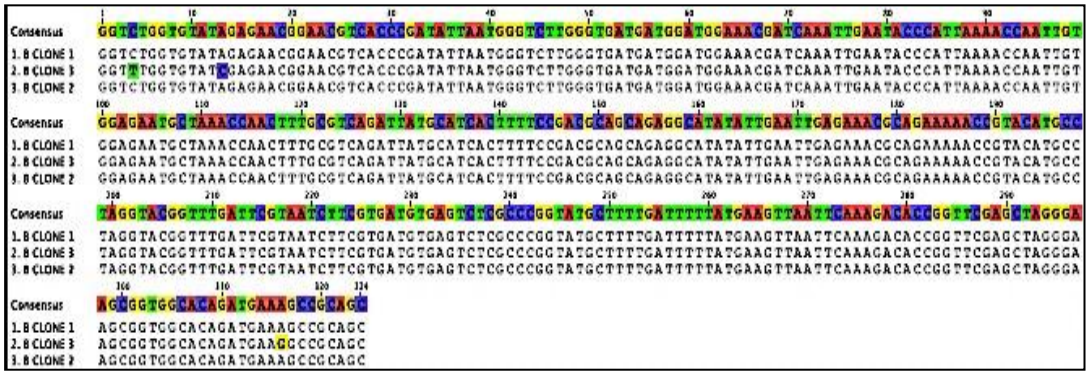
(a) DsMV- NZ1 and clones



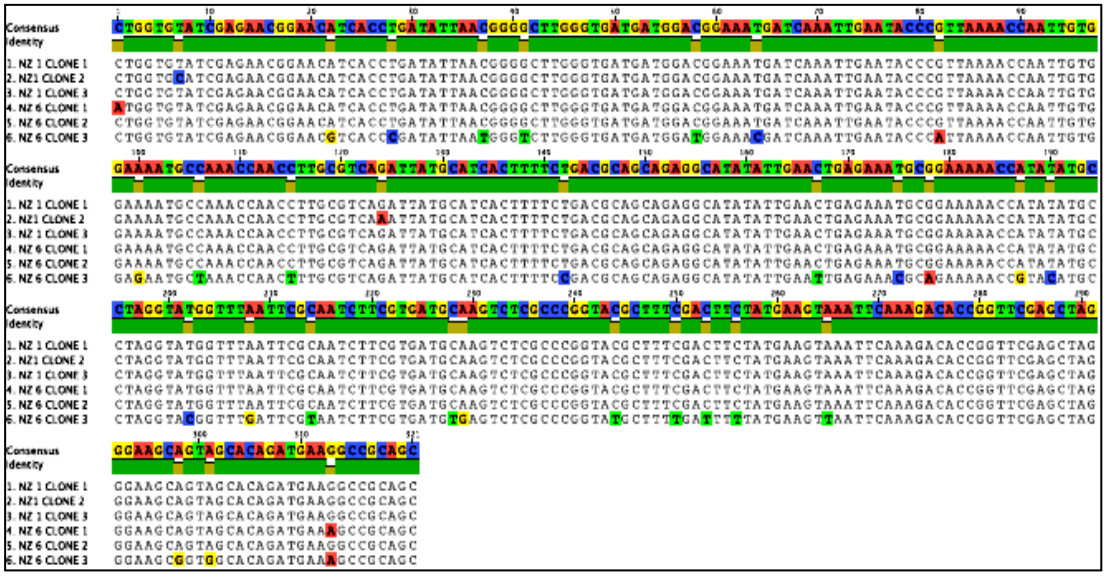
(b) DsMV- NZ6 and clones



(c) DsMV- B1 and clones



(d) DsMV-NZ1 and -NZ6 clones



(e) Amino Acid sequence alignment between clones

Figure 2.11: Multiple sequence alignments of the cloned PCR products for: (a) DsMV-NZ1; (b)-NZ6; (c)-B1; (d) –NZ1 and NZ6 and (e) amino acid sequence alignment between DsMV clones. Consensus sequences with differences between clones are highlighted.

Table 2.4: Nucleotide identity between parent clones (above diagonal in black) and amino acid identity (below diagonal in blue) of clones for DsMV – NZ1, -NZ6 and –B1

	NZ1 clones	**NZ6 Clones**	**B1 Clones**
NZ1 clones		89%-97%	80%-90%
NZ6 Clones	99%-100%		89%-100%
B1 Clones	98%-99%	99%-100%	

2.3.8 SSCP screening and sequence comparison of propagant plants

The parent plants analysed as described above were propagated as shown in Figure 2.2. SSCP analysis was performed on PCR products derived from both first generation and second-generation propagant plants for each isolate (Table 2.2) to determine if the viral populations (i.e. variation) between the parent and propagant plants were substantially different. To further characterize the isolates, multiple sequence alignments were also performed to deduce the nucleotide and amino acid sequence identity. Figure 2.12 shows the comparison of SSCP profiles of DsMV -CP of NZ1, NZ6 and B1 between the parent plants, first generation propagants and second generation propagants.

For DsMV- NZ1 (Figure 2.12a) first generation propagants NZ1.1 and NZ1.2 carried one of the predominant sequences from the parent plant while NZ1.3 and NZ1.4 carried the other. The same profile was observed in the second-generation propagations as in the first generation. For example, NZ1.1.1 and NZ1.1.2 had the same pattern as NZ1.1. Because of this only the first generation propagant isolates (NZ1.1-1.4) were sent for sequencing. Multiple sequence alignments were carried out to compare the parent sequences with the virus sequence amplified in the first generation propagants. These alignments are shown in Figure 2.12a. These analyses revealed up to three nucleotide differences with the parent clones. Sequence comparisons between propagants and parent clones revealed 98%-99% nucleotide identity and 100% at the amino acid identity. This indicated viral isolates are closely related as a result of subdivision from the parental predominant sequences during the propagation process, resulting in stabilization in the first and second generation.

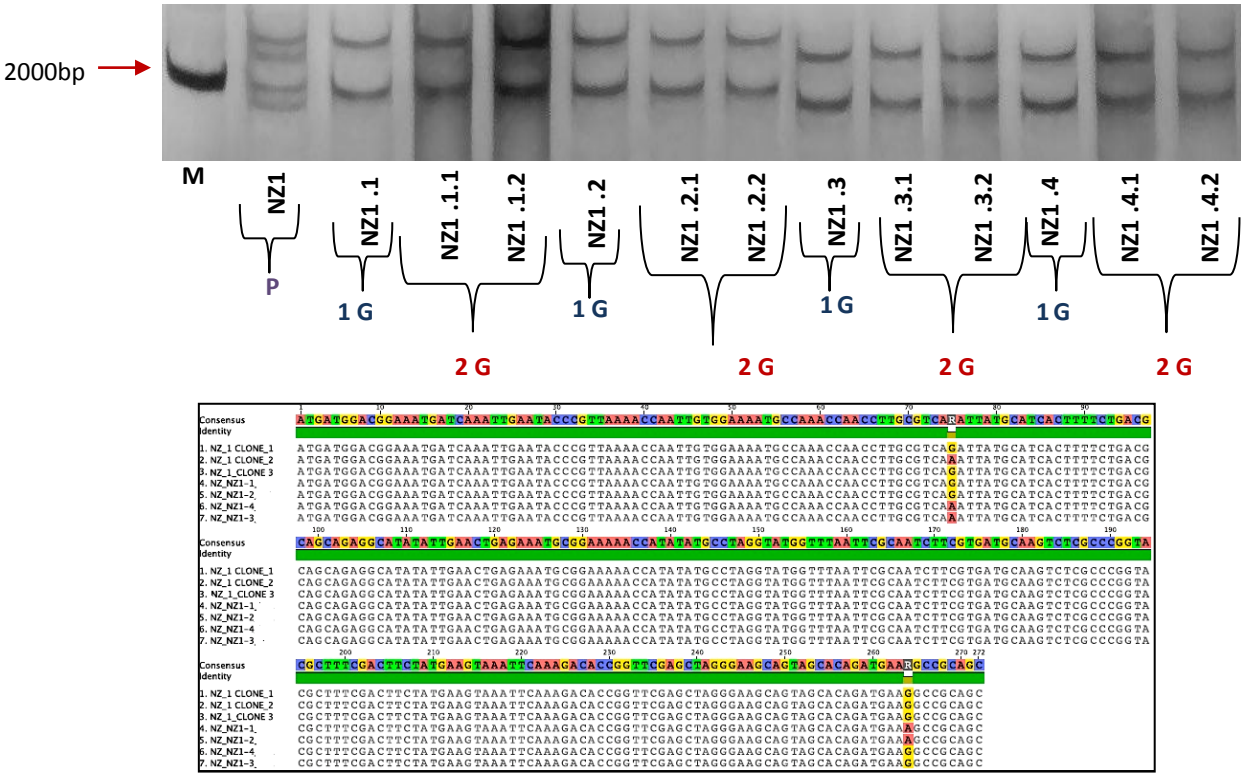
In contrast, the SSCP profile indicated only one of the predominant sequences found in the parent NZ6 plant was propagated to all propagant plants NZ6.1 – NZ6.4 (Figure 2.12b). One of the major NZ6 sequences was lost in the first and second generation again, the pattern was the same between first and second generation DsMV. Thus, only propagant isolates from the first generation (NZ6.1 and 6.2) were sequenced. Multiple sequence alignment between the propagant isolates (Table 2.5b) showed NZ6.1 and 6.2 had up to five nucleotide differences with the parent clones 1 and 3, whereas several nucleotide differences between parent clone 2. Sequence comparisons between

propagants and parent clones revealed 89%-99% nucleotide identity and 98%-99% amino acid identity. The loss of one predominant sequence during the propagation process resulted in stabilization of the other predominant sequence.

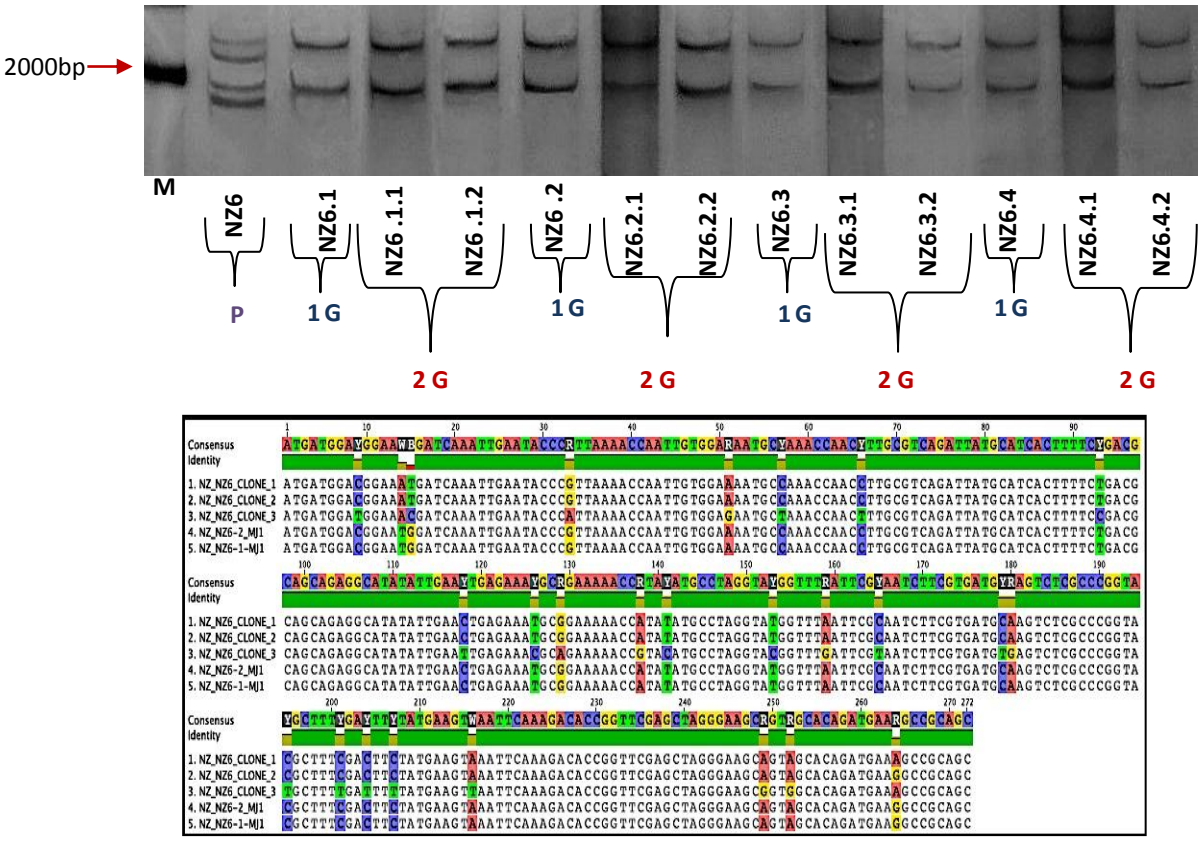
For the B1 strain, propagant plant B1.2 had the same pattern as the B1 parent (Figure 2.12c). However, B1.1 showed a pattern not seen in the original parent SCCP pattern but shows the same SSCP pattern as B1 clone 3 and 4 (see Figure 2.10c). Thus, only propagant isolates from the first generation (B1.1 and 1.2) were sent for sequencing. Multiple sequence alignment between the propagant isolates (Table 2.5c) showed B1.1 and 1.2 had up to seven nucleotide differences with the parent clone sequences. Sequence comparisons between propagants and parent clones revealed 97%-98% nucleotide identity and 97%-98% amino acid identity. The pattern observed for B1 strain indicated viral isolates are closely related as a result of subdivision from the parental predominant sequences during the propagation process. This resulted in stabilization of the predominant sequence in the first and second generation.

Overall, the results showed sequences for the DsMV progeny viral populations represented part of the parental consensus sequences, and predominant sequence in the propagants had been subdivided or lost during greenhouse propagations. The mutant spectra (nucleotide substitution) for each viral population varied between propagant plants, suggesting genetic variation between the isolates of different hosts. For all strains, the second-generation propagants had the same pattern as the first generation, suggesting stabilization of the predominant sequences not changed significantly during greenhouse propagations.

a) DsMV- NZ1 isolates and sequence comparison



b) DsMV- NZ6 isolates and sequence comparison



c) DsMV- B1 isolates and sequence comparison

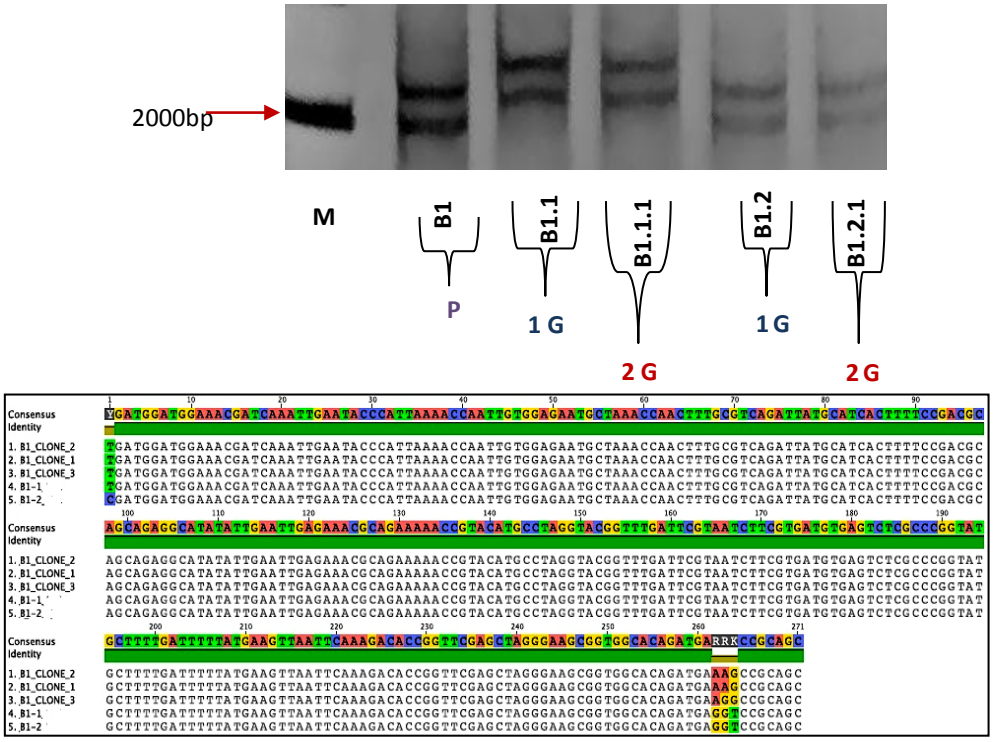


Figure 2.12: SSCP analysis and multiple sequence alignments of DsMV amplified from parent and propagant plants (a) DsMV- NZ1, (b) DsMV- NZ6 and (c) DsMV- B1. Lane marked M: 100bp Ladder DNA Marker (100ng), P refers to parent plants, 1 G refers to 1st generation propagant plant, and 2 G refers to 2nd generation propagant plants. Consensus sequences with differences between clones and propagants are highlighted.

Table 2.5: Nucleotide identity between parent clones (above diagonal in blue) and amino acid identity (below diagonal in black) between (a) NZ1parent clones and propagant isolates, (b) NZ6 parent clones and propagant isolates and (c) B1 parent clones and propagant isolates.

(a)	NZ 1 CLONE 1	NZ 1 CLONE 2	NZ1 CLONE 3	NZ1-1	NZ1-2	NZ1-4	NZ1-3
NZ 1 CLONE 1		99%	100%	98%	98%	98%	98%
NZ 1 CLONE 2			99%	98%	98%	99%	99%
NZ1 CLONE 3				98%	98%	98%	98%
NZ1-1	100%	100%	100%		100%	99%	99%
NZ1-2	100%	100%	100%	100%		99%	99%
NZ1-4	100%	100%	100%	100%	100%		100%
NZ1-3	100%	100%	100%	100%	100%	100%	

(b)	NZ6 CLONE 1	NZ6 CLONE 2	NZ6 CLONE 3	NZ6-1	NZ6-2
NZ6 CLONE 1		99%	90%	99%	99%
NZ6 CLONE 2			89%	99%	99%
NZ6 CLONE 3				89%	89%
NZ6-1	99%	98%	98%		100%
NZ6-2	99%	98%	98%	100%	

(c)	B1 CLONE 1	B1 CLONE 2	B1 CLONE 3	B1-1	B1-2
B1 CLONE 1		100%	99%	97%	97%
B1 CLONE 2			99%	97%	97%
B1 CLONE 3				98%	97%
B1-1	97%	97%	97%		99%
B1-2	98%	98%	98%		97%

2.3.9 Phylogenetic analysis

The sequences from each DsMV strain and isolate were compared with equivalent sequences in the public domain (GenBank). The public DsMV sequences were identified using DsMV- NZ1 as the query sequence in a BLASTn search, implemented in GeneiousPro 5.5.4. Phylogenetic analysis by Neighbour- joining tree and Maximum Likelihood (ML) phylogenetic tree were carried out and calculated to determine the evolutionary history among DsMV isolates (Figure 2.13 and 2.14). Both trees showed the same overall topology, thus the ML tree was used in this study to determine the evolutionary relationships between the DsMV isolates. The ML tree was reasonably well resolved, based on the interpretation of bootstrap support values of greater than or equal to 50%. Low bootstrap values found within the internal nodes of the groups (only values >50% are shown) suggested rapid radiation in the evolutionary process of DsMV.

Phylogenetic analysis indicated two major lineages diverging from a common ancestor among the DsMV isolates, with bootstrap support of 100% (Figure 2.14). Each of these lineages is further split into two major lineages with bootstrap support of 68% and 70%, respectively, making four major clades all together, labelled in Figure 2.14, as Group 1-4. Groups 1 and 2 were comprised of DsMV sequences from India and China isolated from infected aroids other than taro. Group 3 was comprised of sequences from a variety of locations isolated from different aroid hosts including tannia and taro.

Within group 4 the A, B and NZ isolates appeared to form five distinct sub-clades (Figure 2.14), with three sub-clades consisting of sequences from A, B and NZ strains of DsMV. The other two sub-clades included VanMV as well as DsMV from French Polynesia. One clade contained DsMV –NZ sequences from this study as well as previous studies (Cong, 2007; Yuan, 2008), except for clone 3 of DsMV-NZ6. All members of Group 4 have a common ancestor within this group; the NZ and B strains appear to have a common ancestor. The overall observation among the glasshouse isolates appeared to be clustered geographically.

The DsMV-NZ clade was supported with a bootstrap value of 98%. The sequence AY994105 was first identified at the Unitec campus (Pearson, Bussell, & Scheffer, 1998), and propagated in the glasshouse since then. This strain was designated DsMV-NZ and all NZ isolates reported here and in previous studies were derived from this original strain. The sequences 1NZ2, 1NZ3 and 1NZ4 were determined by previous studies (Cong, 2007; Yuan, 2008) and other sequences within the clade were generated by this study. 1NZ2, 1NZ3 and 1NZ4 isolate shared 89%-100% nucleotide identity (data not shown) with the NZ1 isolates in this study. The origin of the NZ samples is unknown; however, given the movement of Pacific Islanders into New Zealand, it is likely that this virus was brought in infected pants from a South Pacific location (M.Pearson, personal communication),

In contrast, B strains from the glasshouse were clustered into two of the five sub-clades within Group 4, suggesting significant diversification from their common ancestor. These strains originally came from taro plants imported from the Solomon Islands and field trialled in New Zealand in 1992. These strains was have been propagated in the glasshouse ever since (M.Pearson, personal communication). Interestingly, NZ6 clone 3 clustered within the DsMV- B1 and A1-5-D sub-group, which indicated this isolate

shared similarity in sequences from the Solomon Islands. This was supported by the nucleotide identities, mentioned earlier. The NZ sequences were clustered together and share an ancestor with the Solomon Islands sequences.

DsMV/VanMV from the Cook Islands forms a separate sub-clade as do VanMV from French Polynesia. The remaining DsMV sequences by and large do not cluster according to where they were sampled, nor by host.

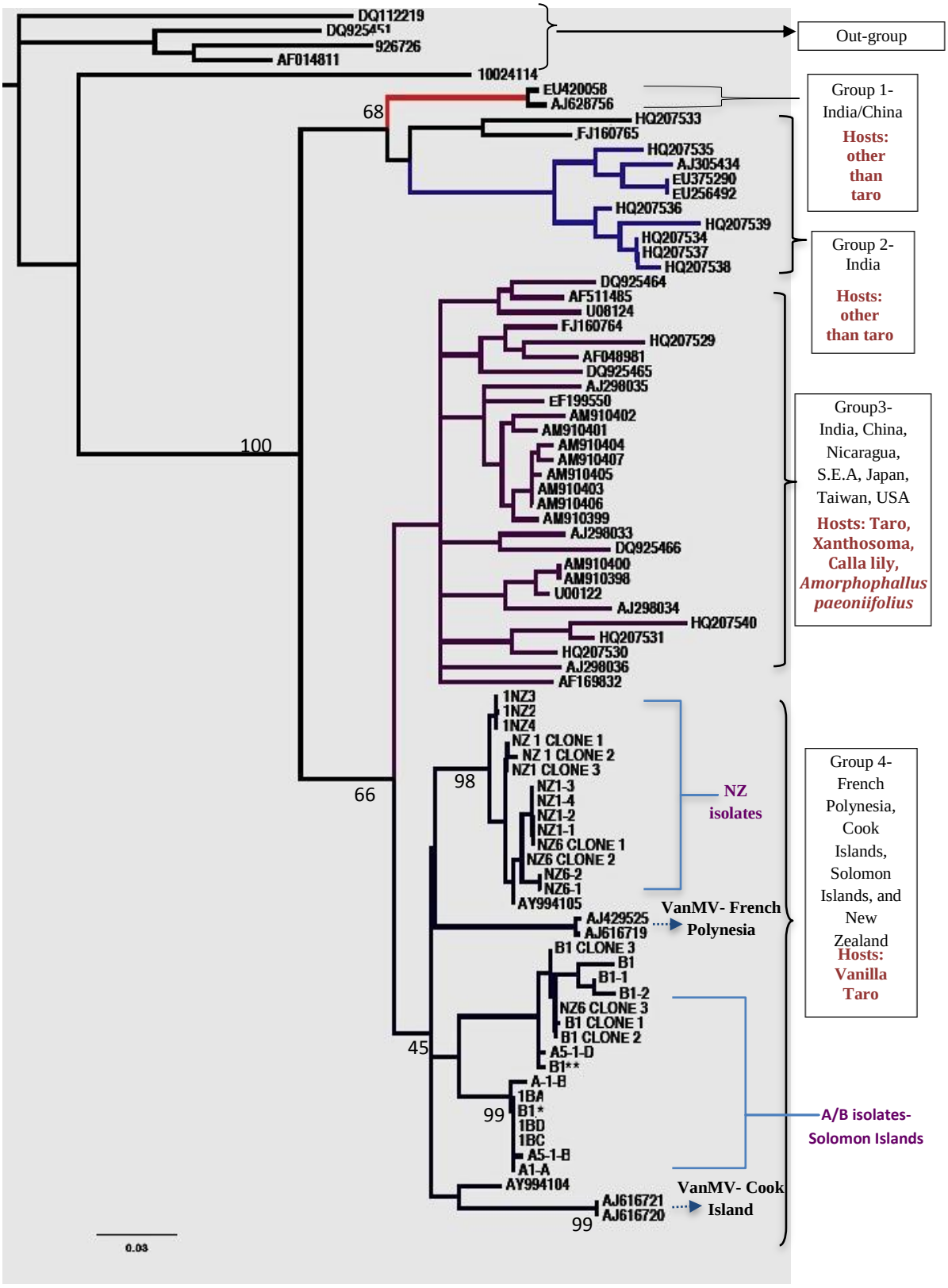


Figure 2.13: Neighbour-Joining tree of the partial CP sequences of DsMV and related Potyviruses. The tree is rooted with PSFV, PWV and ZYMV as the out-group. Bootstrap values above 50 are shown and the scale bar shows the number of substitutions per base. B1* (bottom) and B1 ** (top) indicate sequences from a doublet sequence derive from a previous study (C.Higgins, personal communication). Abbreviations, host range, location, sequences and accession numbers are shown in Appendix 1 and 2. S.E.A denotes South East Asia.

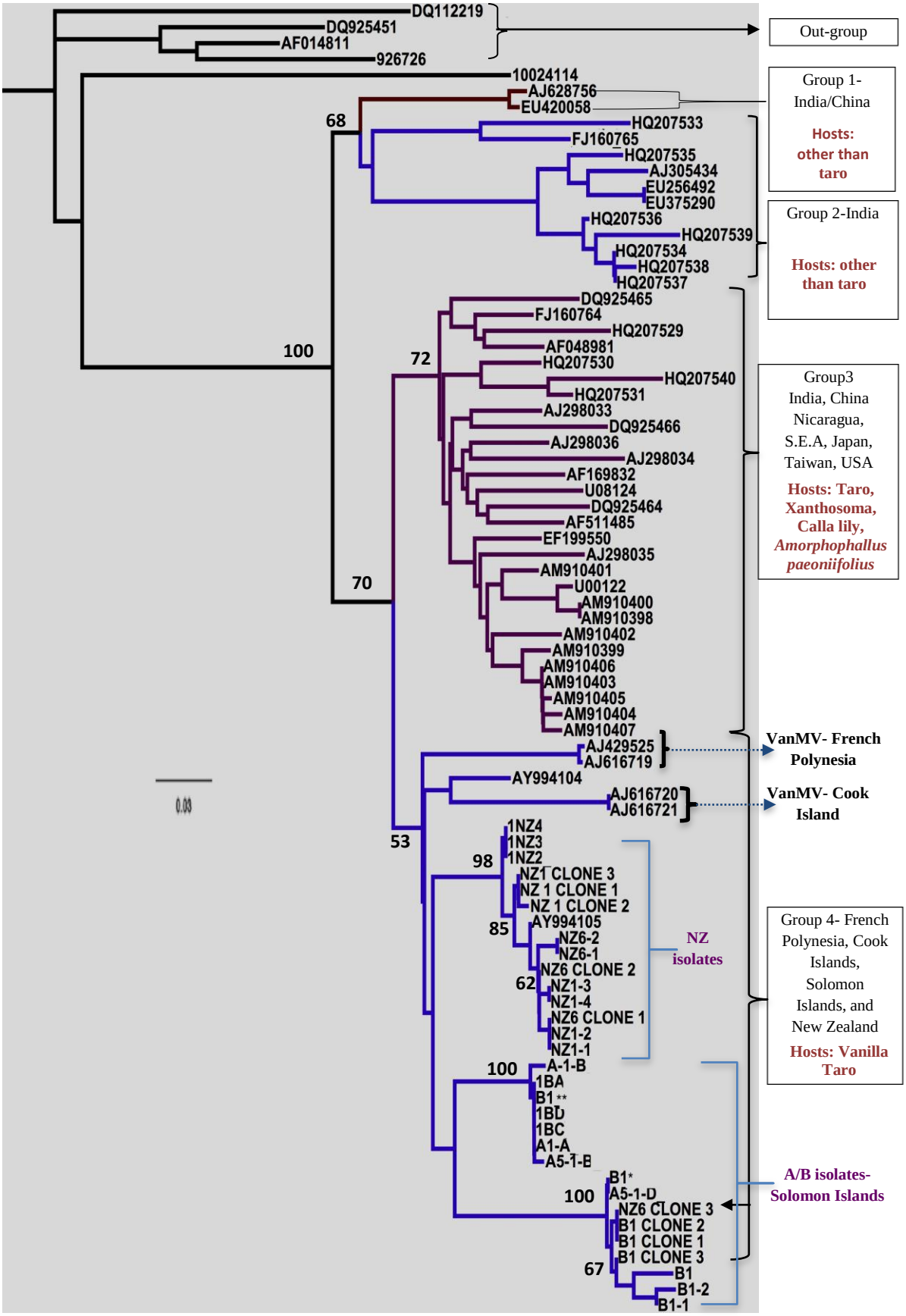


Figure 2.14: Maximum likelihood tree of the partial CP sequences of DsMV and related Potyviruses. The tree is rooted with PSFV, PWV and ZYMV as the out-group. Bootstrap values above 50 are shown and the scale bar shows the number of substitutions per base. B1* (bottom) and B1 ** (top) indicate sequences from a doublet sequence derive from a previous study (C.Higgins, personal communication). Abbreviations, host range, location, sequences and accession numbers are shown in Appendix 1 and 2. S.E.A denotes South East Asia

Nucleotide and amino acid sequences were compared between the glasshouse grown strains of NZ, A and B with published DsMV sequences to determine the nucleotide and amino acid identity (Table 2.6). Overall, the glasshouse strains of DsMV were more similar to each other than to the public sequences (96%-100% compared to 88%-100% amino acid identity and 88%-100% nucleotide identity compared with 80%-92% amino acid identity). While the origin of the NZ strain is unknown, the A and B strains originated from infected taro plants imported from The Solomon Islands (M.Pearson, personal communication). Therefore, this close relationship was not expected. Given the NZ isolates share ancestry with A/B strains (Figure 2.14), it is possible that this strain came from a nearby region.

The nucleotide sequences of DsMV CP between the glasshouse strains and Groups 1, 2 and 3 were equisimilar (80%-92%), with Group 3, closest to the NZ strains (86%-92%). Results indicated greater variation observed between the glasshouse strains with Group 2 (80%-88%) and 1 (83%-90%). At the amino acid (aa) identity level, Group 3 shared higher amino acid identity with the glasshouse isolates (88-100%); however, between NZ, A and B, the amino acid identity were much higher (96-100%).

Despite the fact that the partial CP gene and gene products of all analysed DsMV were very similar, phylogenetic analysis indicated all DsMV had a common ancestor with a bootstrap value of 100%. Two major lineages arose, one leading to groups 1 and 2 which branched off earlier than that leading to Groups 3 and 4 (Figure 2.14). Significant variation in DsMV –NZ, B and A has arisen during propagation of these strains over many years.

This analysis also revealed sequence similarity between DsMV and VanMV. This had been found previously (Farreyrol et al., 2006). Inclusion of VanMV in this analysis showed that it is most similar to Group 4 (82-89%) at the nucleotide level and (89-99%) at the amino acid level. This relationship was supported by the phylogenetic analysis which showed that VanMV shared a common ancestry with DsMV- NZ, A and B in (Figure 2.14).

While partial CP sequences were analysed here, percentage sequence identity and phylogenetic analysis supported these DsMV isolates from the glasshouse belonging to the same species. By these criteria, VanMV should also be considered as belonging to

the same species, even though it has a different host range.

Table 2.6: Percentage of nucleotide identity (nt.) (below diagonal in black) and amino acid identity (a.a) (above diagonal in blue) sequences of the partial CP gene between the groups of DsMV isolates.

		Group 4			Group 1 (India/China)	Group 2 (India)	Group 3 (S.E.A, Japan, Nicaragua. USA)
		NZ isolates	B isolates	A isolates			
Group 4	NZ isolates		96% - 100%	99% - 100%	93% - 99%	94% - 97%	98% - 100%
	B isolates	89% - 100%		98% - 100%	93% - 97%	92% - 99%	88% - 100%
	A isolates	89% - 91%	88% - 100%		93% - 99%	93% - 99%	90% - 99%
Group 1 (India/China)		86% - 88%	83% - 89%	86% - 90%		88% -100%	88% -100%
Group 2 (India)		80% - 88%	81%-85%	82% - 85%	92%-100%		88% -100%
Group 3 (S.E.A, Japan, Nicaragua. USA)		86% - 92%	84% - 90%	84% - 90%	82% - 90%	82% - 88%	
VanMV (Cook Islands and French Polynesia)		82% to 89% (nt.) 89% to 99% (a.a)			76% to 90% (nt.) 76% to 90% (a.a)		

S.E.A: denotes South East Asia

2.3.10 Selection analysis of parent clones and propagants (NZ1, B1 and NZ6)

The ratios of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitution rate ($\omega = d_N/d_S$) were determined to provide an indication of the selective forces at the protein level (Yang, 2007). Values of $\omega < 1$, $\omega = 1$, and $\omega > 1$ indicate purifying (or negative) selection, neutral evolution, and diversifying (or positive) selection, respectively, on a functional protein. The d_N/d_S ratios were calculated for each parent clone and propagant sequences shown in Table 2.7. Most ω were far less than 1 (apart from DsMV -B1 parent plant and DsMV –NZ6 clone 3), indicating of the CP gene was generally under purifying action. However, only a short fragment of the CP genes were analysed in this study, representing approximately one third of the full-length CP gene. Therefore, if different specific regions of the CP have evolved, signatures of selection would have been missed. Further study using full-length sequence of the CP gene should be analysed.

It should be noted that all plants were grown under the same conditions in the glasshouse to minimise environmental variation.

Both DsMV –B1 and DsMV- NZ 6 clone 3 from parent taro plants, on the other hand, appeared to be under diversifying selection; however DsMV –B1indicated much stronger diversifying selection (2.5 compared to 1.2) than DsMV- NZ 6 clone 3. However, if the CP is involved in viral movement, the amino acid substitution could also reduce the systemic movement of virus within the host. Weak diversifying selection acting on amino-acid substitution can result in a large reduction in viral fitness, limiting cell to cell movement and eventually elimination through genetic bottlenecks (Manrubia, Domingo, & Lazaro, 2010). This was observed in the propagant plants of DsMV-NZ 6 where only one predominant sequence was lost throughout the propagated process into two generations.

A similar study has also shown that amino acid substitution interfered with systemic movement of a Potyvirus. In a recent study by Andrejeva et al., (1999) to determine the role of CP gene in the systemic infection of tobacco, partial CP genes of two *Potato virus A* (PVA) variants (PVA-U and B11) were subjected to mutation analysis. They demonstrated different phenotypic effects on tobacco leaf, with amino acid substitution in PVA-U causing no detectable difference in the rate of systemic infection while in PVA-B11 systemic infection in tobacco was reduced, limiting cell-to-cell movement. Although the authors did not analyse the extent of selection forces, results suggested difference in selection forces would have also contributed to the differences in systemic infection between the two PVA variants. It also highlighted the importance of coordinate function of CP in the accumulation and systemic infection of plants, as the CP is classified as one of the movement proteins functioning in intercellular transport (Roossnick, 2011).

Table 2.7: Selection analysis (non-synonymous/synonymous ratio) exerted on the partial CP gene of DsMV -NZ1, NZ6 and B1 parent plants and propagants. * ω ratio greater than one implies positive selection; less than one implies purifying (stabilizing) selection, and a ratio of one implies neutral selection.

Parent	***ω ratio**	**Propagants**	***ω ratio**	**Parent**	***ω ratio**	**Propagants**	***ω ratio**	**Parent**	***ω ratio**	**Propagants**	***ω ratio**
NZ1 Clone 1	0.01	**NZ1-1**	0.01	**NZ6 Clone 1**	0.01	**NZ6-1**	0.04	**B1 Clone 1**	0.01	**B1-1**	0.02
NZ1 Clone 2	0.01	**NZ1-2**	0.04	**NZ6 Clone 2**	0.01	**NZ6-2**	0.05	**B1 Clone 2**	0.04	**B1-2**	0.05
NZ1 Clone 3	0.05	**NZ1-3**	0.01	**NZ6 Clone 3**	**1.2**			**B1 Clone 3**	0.05		
		NZ1-4	0.01					**B1**	**2.5**		

2.4 Discussion

The complex structure of virus populations has been the object of intensive study in bacterial hosts as well as animals and plant hosts for the last 15 -20 years. While it is clear that tremendous genetic diversity is rapidly generated during viral replication, the study of plant RNA virus diversity remains comparatively scarce when compared to human RNA viruses. The purpose of the present work was twofold; firstly, to evaluate the genetic diversity of the DsMV populations within and between host over time. This analysis was done by comparing SSCP patterns and sequence diversities as well as carrying out a phylogenetic analysis. Secondly, changes in sequences of different DsMV strains were assessed in propagated taro plants to determine the genetic variation of the virus population.

RT-PCR and SSCP analysis for polymorphism screening

In this study, RT-PCR was used to amplify a fragment of the CP gene; however amplification errors could have a direct impact on the estimation of genetic variability of the virus isolates. The method of analysis used reverse transcriptase derived from the RNA virus Moloney Murine Leukemia Virus Reverse Transcriptase (MMLV) (Promega, Madison, Wis.) to synthesise cDNA *in vitro* from viral RNA templates. MMLV does not have proof reading activity and has been reported to have an error rate of 1 nucleotide in 30,000 (Promega technical notes, http://www.promega.com). However, any errors created by the reverse transcriptase would be impossible to distinguish between misincorporation that occurs during *in vitro* transcription and mutations already present in the viral RNA population under study (Bracho, Moya, and Barrio, 1998). Another drawback in this study relates to the use of Taq DNA polymerase in the Go®Taq Green master mix, which has an error rate of approximately 1×10^{-5} errors per base (Promega technical notes, http://www.promega.com). This error rate is relatively high due to the enzyme's lack of 3′→5′ exonuclease (proofreading) activity. Considering high nucleotide identity obtained here, the impact of such errors should be minimal. It should be noted that PCR amplifies predominant sequence and minor variant sequences are not detected. If all sequence variants required, transcriptome sequencing with high coverage such as Ilumina based sequencing would be an appropriate technique to use.

SSCP was used in this study to detect single nucleotide changes (Weber, Fukino, Villalona-Calero, & Eng, 2005) and in this regard, could be considered very sensitive. Indeed, in the analysis presented here, two nucleotide differences were detectable. However, not all differences are detected and false negative results can occur. For example, the same SSCP profiles were observed between DsMV-NZ1 and –NZ6; yet alignment of their cloned nucleotide sequences revealed differences in one to several nucleotide positions. Thus, there is no simple relationship between the number of differences and SSCP patterns to predict the extent of nucleotide diversity.

Another limitation of SSCP, is the detection sensitivity decreases when amplicon sizes exceed 200 bp (Weber et al., 2005), therefore, only short PCR products can be analysed with this technique, limiting the amount of genomic information that can be analysed at any time. The reported accuracy for SSCP analysis is highly dependent on several physical factors that need to be optimized for virtually each region of interest.

Therefore, as a polymorphism screen, SSCP may not be a perfect screening tool because a difference in SSCP does not reflect difference in nucleotide sequence and sensitivity decreases with increasing amplicon sizes. Perhaps, the use of PCR-dependent denaturing gel electrophoresis technique (DGGE) would enable the differentiation of nucleotide differences. PCR-generated DNA fragments of the same length but with different base-pair sequences can be fully separated in a fine-tuned gradient gel. Thus, DGGE constitutes a robust procedure by which a single point mutation can be detected. The rationale behind DGGE is that fine separation is based on the melting behavior of double-stranded DNA and that melting behavior in turn depends on the base-pair composition of the target DNA (Carmona, Sepulveda, Cardenas, Nilo, & Marshall, 2012). Molecules with variant DNA sequences may have different melting behavior and will therefore stop migrating at different positions in the gel.

Temporal analysis of DsMV isolates in glasshouse

The first aim of this study was to determine the extent and structure of intra-host and inter-host viral genetic diversity in DsMV, SSCP analysis of the RT-PCR product of the CP gene was performed on mature and young leaves of parent plants. The relationship between the SSCP profiles observed among clones and the variation in their nucleotide sequences suggested that DsMV isolates within each parent plant indicated complex

populations or a “quasispecies like” nature. The partial DsMV- CP between mature and young leaves within each parent plant showed the same SSCP profile, indicating little or no genetic variation. This indicates the predominant sequence within the quasispecies is stable, upon systemic infection. This also suggests that neither genetic drift nor selection pressure of any kind have influenced the viral population. Since genetic drift has had no apparent impact, this suggests the quasispecies population is very large, therefore virus fitness i.e. replication efficiency is very stable.

In contrast, between hosts SSCP patterns differed between DsMV-B1, NZ1 and NZ6. SSCP patterns for DsMV-NZ1 and NZ6 were the same; however sequencing showed that these strains were in fact different viral populations. Therefore, the predominant sequence(s) for each strain differed; hence the quasispecies populations were different. This suggests which sequence becomes predominant is likely to be random; therefore genetic drift is probably an important factor determining the specific profile of each population within different plants.

Selection analysis indicated B1 is under diversifying selection and NZ1 and NZ6 are predominantly under purifying selection. However, NZ 6 clone 3 disappeared from its host, but this sequence was the same as B1 clones 1 and 2 which survived in their host. These results suggest that selection imposed by the host-virus infection cycle is selected for particular sequences specific to the virus-host combination. Given the hosts genetic backgrounds, it is feasible that even though they are the same species, the host plants for B and NZ are not clones and therefore could have different relationship with the virus.

It would be interesting to infect the host genotype of DsMV-B with the NZ strains and vice versa or a co-infection to see if the selection type is consistent. This would demonstrate a role for the host in selecting the quasispecies profile.

The second aim of this research was to compare variation of viral populations between parent plants and propagants. For B1 strain, B1 parent plant is under diversifying pressure and SSCP results indicated one sequence appears predominant with another emerging. These appears to have been unevenly spread throughout the corm of the plant so when it was split for propagation, one sequence was not common in one part of the corm, while the other was more common in the other part. Each acted as the founder sequence as a consequence of genetic bottleneck for a new population within the first

generation plants. These two predominant sequences therefore became separated in the first generation plants. These plants were shown to be under purifying selection, i.e. deleterious sequences were actively being removed so that each of these predominant sequences lead to stabilization of the quasispecies population, Thus stabilisation continued into the second generation suggesting each quasispecies had reached maximum fitness.

For NZ1 family, SSCP results indicated separation of predominant sequences as a consequence of genetic bottleneck. NZ1 showed a similar result to the B1 family, except the parent plant NZ1 was experiencing purifying selection. These suggest that the two predominant sequences had already reached stable replication efficiency and that this was maintained for each separately in the first and second generation.

In contrast, SSCP results for the NZ6 family indicated separation of predominant sequences as a consequence of genetic bottleneck, however only one predominant sequence was carried over to the propagants. The starting point for NZ6 was similar to NZ1 but one predominant sequence was lost to the first and second generations. Muller's ratchet implies the first generation propagant had a small viral population upon propagation, which was more vulnerable to genetic bottleneck, which lead to extinction of one of the predominant sequences (Roossnick, 2011). The remaining sequence was efficiently replicated so that the quasispecies were stable into the second generation.

Phylogenetic analysis of DsMV isolates

Genetic divergence among the DsMV isolates in the ML tree likely reflects time elapsed since separation from a common ancestor. Geographical relationship within Group 4 could have been due to the number of samples analysed, and if an equivalent number of samples had been studied for the other regions a similar finding might have been observed. For example, similar observation was seen in a recent study by Reyes et.al, (2009), where RT-PCR was used to amplify the coat protein (CP) region from ten Nicaraguan DsMV isolates from infected cocoyam in the same geographical region. Phylogenetic analysis showed that the Nicaraguan isolates formed two distinct subgroups correlated with geographic origin, which is also observed in the phylogenetic tree in this study. Results suggested that the distinct origin of the cocoyam genotypes was the possible cause of the genetic differences between virus isolates from the DsMV

subgroups. This supports the geographical clustering of the NZ, A and B strains and that the overlapping populations for NZ 1 and 6 are distinct from the A and B population. The origin of NZ strains was determined to have originated in Samoa, which is discussed in Chapter 3. This suggests the movement of taro and spread of DsMV between Melanesia and Polynesia.

Although a recombination analysis was not performed in this study, the occurrence of recombination events during the evolution of DsMV should not be rule out. The fragment studied was probably too short for recombination to be apparent. Such a study is best done with much larger sequences, preferably whole viral genomes. However, selection is also an important factor. Genetic drift and selection are not mutually exclusive forces and may occur concurrently. Thus, the combined effects of both forces acting simultaneously on the CP gene ultimately determine the extent of divergence between and within the groups. Similar kinds of analysis have been performed with other viruses and most notably in *Wheat streak mosaic virus* (WSMV). For example study conducted by Hall et.al. (2001), genetic variation was determined in every gene except 6K1, 6K2 and VPg between two variants of WSMV using SSCP and serial passage experiments in three different hosts (maize, corn and wheat). Results demonstrated that the stochastic processes of bottlenecking, subdivision, and drift reduced variation among the virus isolates, yet the random nature of these processes facilitated divergence among genetically isolated lineages within the constraints of selection. Although the virus, virus gene sequences, and host are different in this study, the findings presented supported that stochastic processes and negative selection have great impact on the short-term evolution of potyviruses (Hall et. al. 2001).

In conclusion, the aims of this work were to determine the genetic variation of glasshouse DsMV isolates within and between host plants and generations. Indeed, the results indicated lack of variation within host, suggesting large viral population, and fitness stability maintained the genetic stability of the glasshouse DsMV isolates. The differences observed between hosts and propagant generations can be accounted for by the extent of deterministic and stochastic mechanism operating on the effective size of the viral population; between hosts and propagant generations; genetic drift, genetic bottlenecks and selection pressure were the key mechanisms generating diversity among the isolates.

This suggests that the highly diverse DsMV have a better chance of expanding into a new host and thus pose a greater threat of emerging as crop diseases. Although the correlation between viral population diversity and expansion into new hosts is intriguing, the data sets in this study are relatively small, and extrapolation to inter-host analyses with larger numbers of sequences should be made with caution. Further, the sequence analysed was relatively short, analysing larger fragments, if not complete genomes may gain a more complete view of the mechanisms of evolution and history of this virus.

The temporal characterization of DsMV isolates described here offers a view into the short- term evolutionary dynamics of the population structuring of DsMV isolates. However, to more fully understand the epidemiological and the genetic structure of DsMV on a finer geographical scale, spatial analysis of DsMV from the South Pacific Islands was performed to understand the long-term evolution of DsMV in the next chapter. It is possible, if not likely, that the evolutionary and spatial dynamics among this virus are of central importance in understanding the shaping patterns of the viral genetic diversity. Additionally, molecular clock analysis may help in clarifying the genetic diversity and distribution of this virus.

Chapter 3

Spatial Analysis of DsMV

3.1 Introduction

RNA viruses are the most abundant parasites infecting humans, farm animals, and cultivated plants. The past two decades have witnessed an expanding list of infectious diseases characterized as emerging threats to human and/or wildlife health, agricultural production, or public security; for example, pathogens used for bioterrorism (Real et al., 2005). Recent reviews of the mechanisms leading to disease emergence most often point to the simultaneous effects of spatial and evolutionary genetic changes in patterns of first appearance and the spread of viruses (Holmes & Grenfell, 2009). Studies on spatial dynamics of viral populations primarily focus on processes that affect changes in gene frequencies in viral populations over time and space (Tugume, CuÉLlar, Mukasa, & Valkonen, 2010). Epidemic expansion of RNA viruses may prove especially amenable to exploring the linkage between evolutionary and ecological dynamics, because these processes often occur on the same temporal scale (Grenfell et al., 2004).

The importance of spatial analysis of RNA viruses can be accounted for by several reasons; (1) spatial analysis is directly applicable to understanding the spatial variation of disease, and its relationship to environmental factors and the public health care system. For example, spatial analysis of HIV prevalence and incidence among injection drugs users in St Petersburg, Russia were determined to understand the implications for HIV transmission (Heimer, Barbour, Shaboltas, Hoffman, & Kozlov, 2008). Study identified linkages between disease prevalence and risky injection behaviours. The analysis also identified where resources might be allocated geographically for maximum impact in slowing the HIV epidemic among injection drug user; (2) spatial analysis also provides in depth understanding of the epidemiology of diseases which in turn can be used to understand and predict disease prevalence. For example, the recent worldwide distribution of Severe Acute Respiratory Syndrome-coronavirus (SARS-CoV) can be understood only as a combined epidemiology and evolutionary phenomenon, the appearance of a novel genetic type of coronavirus coupled to the movement and dispersal of infected individuals over transportation networks led to an epidemic expansion of this virus (Stavrinides & Guttman, 2004) and (3) understanding the spatial analysis of RNA virus inform future strategies for the surveillance and control of diseases. For instance, the formulation of influenza vaccines before each flu season is essentially a process of evolutionary prediction. An understanding of the

spatial dynamics on how influenza evolution works at host and viral population levels is highly relevant for control strategies.

There has also been considerable interest in using sequence data from plant RNA viruses to infer spatial dynamics and evolutionary history of the associated viruses (Simmonds, 2004). For example, in a recent study by Simmons et. al., (2011) fifty-five CP sequences from ZYMV isolated from different geographical locations, were used to determine their evolutionary relationships. Spatial analysis indicated distinct clusters of viral isolates apparent in the ML tree. These clusters represented significant clustering by country of origins sampled and provide strong evidence for the *in situ* evolution of ZYMV within individual countries. The study highlighted the utility of spatial analysis to reveal key aspects of the epidemiological history of plant RNA viruses.

Based on the examples above, spatial analysis will allow us to tease out the mechanisms underlying the observed evolution of a pathogen. Understanding these is crucial to devising effective control strategies for important human, animal and plant virus diseases, including emerging zoonotic and other diseases.

3.1.1 Evolutionary history of Potyviruses

The earliest attempts to determine the relationships of Potyviruses were made more than 50 years ago using quantitative serological methods with virions as antigens (Bos, 1992). Although these methods worked well with most plant viruses, they did not with Potyviruses (Bos, 1992). Indeed, there seemed to be a continuum of variants or strains linking distinct viruses. In 2009, phylogenetic analyses were carried out with 59 full-length Potyvirus sequences, to compared the ML pairwise patristic distances of the polyprotein with the patristic distances of the coherently evolving region of the complete CP genes (cCP region) (Gibbs & Ohshima, 2010). The phylogeny reveals the initial stages of the Potyvirus starburst. The initial divergence gave rise in quick succession to five basal lineages, namely, the Onion yellow dwarf virus (OYDV) group, Sugarcane mosaic virus (SCMV) group, Pea seedborne mosaic virus, and two super groups, the PVY super group and the BCMV super group. Further analysis of the polyprotein and cCP (complete CP) distances indicated the cCP have evolved more

slowly than the remainder of the open reading frame (ORF). Thus, the cCP ML relationships can be used as a surrogate measure of the relationships of the complete genomes and also used to assign species known only from their cCP sequences to their likely positions in the genomic phylogeny.

The first Potyviruses to provide evidence of sub generic groups were those related to BCMV (Wylie & Jones, 2009) and to PVY (van der Vlugt, 1992). These two groups illustrate clearly the sort of distinctive phenotypic differences that can appear in related viruses that have followed different evolutionary trajectories after diverging from a common ancestor. The most recent common ancestor to the BCMV and PVY groups existed approximately 6,770 years ago, probably in the same agricultural region that the progenitor Potyvirus emerged a few centuries before (Gibbs & Ohshima, 2010). This supported previous study by Gibbs, et.al. (2008), using partial CP sequences of about fifty Potyviruses to infer their evolutionary history of Potyvirus. Results indicated the initial radiation of the Potyviruses occurred only about 6,600 years ago, and hence coincided with the dawn of agriculture.

In a subsequent study by Gibbs et.al, (2008), the BCMV group was further analysed and was found to include at least thirty-seven species. Two-thirds of the primary hosts of the BCMV group are dicotyledons, mostly legumes and passifloras, although two of the most economically damaging infects crops of cucurbits. Of the monocotyledonous plants, one-third is orchids and aroids. Six species of the BCMV group are cosmopolitan and cause damaging diseases in commonly grown crop plants, whereas the other species of the BCMV group have been found only in Southeast or East Asia, Oceania, or Australia in minor crop species, ornamentals, or wild plants. Study also suggested BCMV entered Australia through the northern coasts and nearby islands of Greater Australia on the Austronesia ‘‘speedboat to Polynesia’’. Interestingly, this study also indicated that DsMV isolates infecting taro are closely related to the BCMV subgroup in the genus Potyvirus. This study also indicated the relatively limited range of crops carried by the Austronesia, and/or the small region of China from which they came, selected the lineage of viruses that were carried to Australia (Gibbs, Trueman & Gibbs, 2008). The findings based on the BCMV group by Gibbs et.al (2008) coincide with similar study by McCoy & Graves, 2010, which suggested that the migration of

Austronesian people from Taiwan was dated 6,600 years ago, coincided with the dawn of agriculture.

3.1.2 Spatial analysis of DsMV

In recent years, the number of DsMV sequences reported in the literature has increased (Babu et. al., 2011). Partial sequences from *Colocasia*, *Zantedeschia*, *Pinellia*, and *Caladium* as well as one complete genome sequence (9,991 nt) from *Zantedeschia* have been reported (NCBI, 2010). The study of the genetic structure of DsMV on a geographical scale has previously been attempted for different reasons. For example, PCR-based diagnostic tests were used to survey taro growing in eleven Pacific Island countries and results indicated DsMV was widespread in the South Pacific Islands (Revill et al., 2005). However, no further details were analysed in the extensive occurrence and evolutionary history of DsMV in the South Pacific Islands.

Geographically focused studies of DsMV genetics had also been used to elucidate patterns of spread of the virus across the landscape. For example, RT-PCR was used to amplify the coat protein (CP) region from ten Nicaraguan DsMV isolates from infected cocoyam in the same geographical region. Phylogenetic analysis showed that the Nicaraguan isolates formed two distinct subgroups correlated with geographic origin and genetic differences between virus isolates from the DsMV subgroups were related to different cocoyam genotypes (Reyes, et.al. 2009).

Although the studies mentioned above were carried out to understand the spatial dynamics of DsMV, the host range and geographical location differs from the study carried out in this thesis. To date, no sequences of DsMV isolates from the South Pacific Islands from *Xanthosoma sp.* and *Colocasia sp.* have so far been reported. It is currently unclear whether DsMV genetic variation occurs in the South Pacific Islands.

3.1.3 Aim of the study

Based on the discussion above, the aim of this study was to determine the spatial distribution of the DsMV lineages in the South Pacific and the extent to their evolutionary relationships. The rate of molecular evolution of the lineages was also examined using nucleotide substitution data, as a way to infer whether these lineages may be under different selective forces potentially brought about by, for example, differences in host or geography.

3.2 Materials and Methods

3.2.1 DsMV infected leaf samples

DsMV infected taro leaf material had been collected from several locations in the South Pacific (Table 3.1) and was provided by Associate Professor Michael Pearson of the School of Biological Sciences (SBS), The University of Auckland. All samples apart from those collected from Samoa were freeze dried and stored at -20^{o}C. Samples from Samoa had infected leaf sap sampled onto FTA cards and stored at room temperature. Since the Ministry of Agriculture and Forestry (MAF) classified these samples as quarantine samples, the RNA extractions were carried out within the quarantine laboratory in SBS.

Table 3.1: Sample identification, location, geographical region and host of virus isolates sampled in the South Pacific

Isolates	Location	Geographical region	Host
***VC3451 VX3402 VX3406 VX3415**	**Vanuatu**	Melanesia	*Colocasia* *Xanthosoma*
***CX3626 CX3627 CX3633 CX3664 *CC3688**	**Cook Islands**	Polynesia	*Xanthosoma* *Colocasia*
***FPC3783 *FPC3870**	**French Polynesia**		*Colocasia*
SAMOA1 SAMOA3 SAMOA5 SAMOA7	**Samoa**		*Colocasia*
PC4030 *PC4006 *PC4007	**Palau**	Micronesia	*Colocasia*
NC4549	**Nauru**		*Colocasia*
MAX4442 MAX4481 MAX4492 *MaX4423 *MaC4424	**Mariana Islands**		*Xanthosoma* *Colocasia*
***FSMX4039 *FSMX4041 FSMX4060**	**Federated States of Micronesia**		*Xanthosoma*

Note: *Sequence generated from previous studies (Yuan, 2008)

3.2.2 Total RNA Extraction

Total RNA was extracted from 100 mg freeze dried leaf samples as described in section 2.2.3. For Samoan samples, RNA was extracted from FTA cards, as follows: 2 mm diameter filter disc was extracted into 1.5 mL microcentrifuge with 400 mL RNA Processing Buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 800 U/mL RNase Out™ (Invitrogen) inhibitor, 200 µg/mL glycogen and 2mM DTT). Samples were incubated on ice for 15 minutes with mixing every 5 minutes following which 200 µL of 7.5 M ammonium acetate and 600 µL ice-cold 100% isopropanol were added. The mixture was then incubated at – 20 °C for at least one hour. The disc was removed after incubation and the remaining liquid centrifuged at 12,000 g for 5 minutes and the supernatant was discarded. The pellets were washed with 200 µL of ice-cold 75 % ethanol and centrifuged at 12,000 g for 5 minutes before air drying at room temperature. The pellet was resuspended with 50 µL of 0.1M buffer and either used directly for RT-PCR analysis or stored at -20^{o}C.

3.2.3 RT-PCR amplification using MJ1 and MJ2

First strand cDNA was synthesised as described in section 2.2.5. A 327 bp region of the coat protein gene was amplified by PCR using the potyvirus generic primer MJ1/MJ2 (Table 2.3). PCR products were gel purified as described in section 2.2.6., before sending to Macrogen Inc, South Koprea for directly sequencing using the MJ1 primer.

3.2.4 Phylogenetic analysis

Phylogenetic analysis of South Pacific isolates of DsMV was carried out as described in section 2.2.13. Sequences from PFVY, ZYMV and PWV (see Appendix 1 and 2) were used as the outgroup for this analysis.

3.2.5 Selection analysis

Selection pressure analysis was carried out as described in section 2.2.14

3.2.6 Spatial pattern analysis

Spatial analysis was performed between genetic distances (pairwise genetic distance) and geographic distances among the South Pacific isolates to infer spatial and evolutionary relationship between the isolates. Average pairwise genetic distances were estimated for all sequences between each geographic location using MEGA 5 with the default parameter. Geographic distances based on great circle distance between the various South Pacific Islands were calculated directly from longitudes and latitudes of each geographic location of the island using the free public web interface program (www.csgnetwork.com/marinegrcircalc). Correlations between average pairwise genetic and geographic distances was performed by Pearson statistics on Microsoft ®Excel.

3.2.7 Estimation of the evolutionary rate and application of a molecular clock

Molecular clock analysis was performed to determine the time of divergence for the DsMV isolates in the South Pacific Islands. A phylogenetic tree was constructed using all the partial CP gene sequences listed in Appendix1, under a relaxed-clock model using BEAST v 1.4.8 (Drummond et al., 2006), using the default parameters, to estimate the divergence time of all sequences found within each geographical location. In this Bayesian phylogenetic inference, calibration of the most plausible root was performed using Tree Annotator v1.5.4 software (http://tree.bio.ed.ac.uk/software) to construct the Maximum Likelihood tree. (Stamatakis, Hoover, & Rougemont, 2008). Based on the work of McCoy & Graves (2010), the migration of Austronesian people from Taiwan was dated 6,600 years ago; therefore this time frame was used in the time calibration in the molecular clock in the basal node for Group 2 clade.

The molecular clock was tested by comparing the ML value for the given topology with and without the molecular clock constraints using MEGA 5.

3.3 Results

3.3.1 RT-PCR of partial CP gene using degenerate primers MJ1/MJ2

Degenerate primers MJ1/MJ2 correspond to conserved sequences located in the core of the CP region (Marie-Jeanne et al., 2000) were used to amplify this sequence from South Pacific isolates RT-PCR using these primers showed the amplification of the expected 327 bp products from all South Pacific DsMV isolates (Figure 3.1). This band was successfully amplified from RNA extracted from infected leaves and FTA cards. However, most of the dried leaf samples also gave rise to non- specific products and primer dimers (Figure 3.1a). The PCR products of the correct size therefore had to be gel purified for further characterization by direct sequencing. Prior to sequencing, gel electrophoresis was carried out on the purified PCR products to confirm their purity (Figure 3.1b).

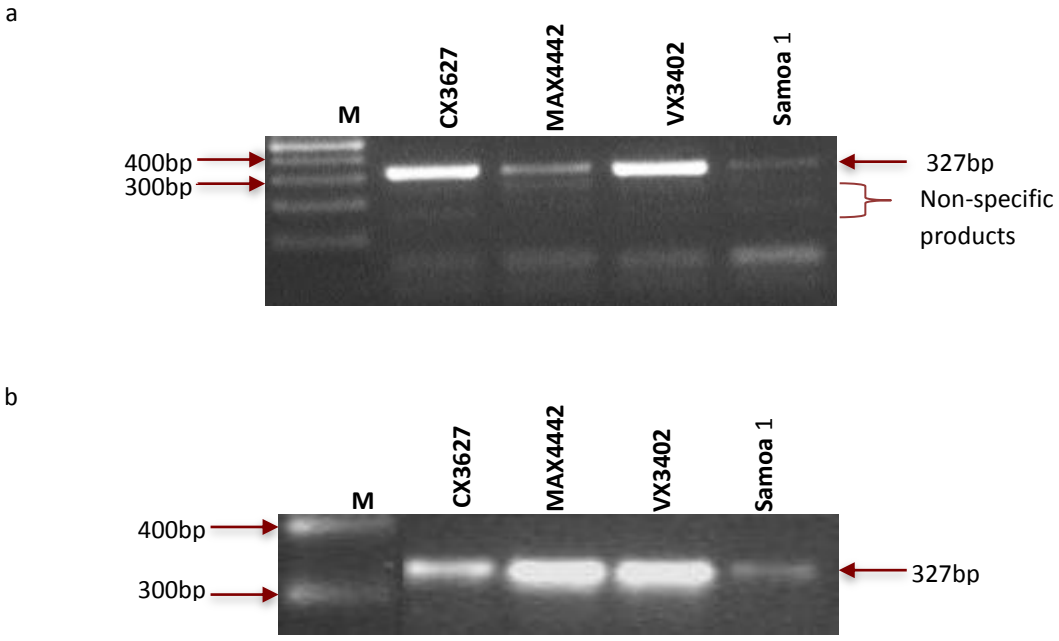


Figure 3.1a): Agarose gel electrophoresis (1.5% agarose/ 1x TBE) of the expected 327 bp PCR product amplified with MJ1/MJ2 primers with the presence of non-specific products, b) Agarose gel electrophoresis (1.5% agarose/ 1x TBE) of the gel purified PCR product (327 bp). Lane M: 100bp Ladder DNA Marker (100ng).

3.3.2 Phylogenetic analysis

PCR products from the South Pacific isolates were sequenced and compared to DsMV sequences held in the NCBI database using GeneiousPro 5.5.4. The DsMV sequences were identified using each sequence from different geographic location as the query sequence in BLASTn searches. Neighbour- joining tree (Figure 3.2) and Maximum Likelihood (ML) (Figures 3.3 and 3.4) phylogenetic trees were generated to determine the evolutionary relationships of each sequence. Both trees showed the same overall topology, the analysis focussed on ML tree since this type of analyse is generally more robust (Gibbs, et.al, 2010). Consistent with the analysis shown in Figure 2.13, two major lineages with bootstrap support of 100% have led to the extant isolates. Each of these lineages is further split into two major lineages with bootstrap support of 73% and 87%, respectively, making four major clades all together, labelled in Figure 3.4, as Groups 1-4. Low bootstrap values found within the internal nodes of the groups (only value >50% are shown) suggest relationships could not be well resolved probably due to rapid radiation.

The Group 1 is made up of DsMV sequences from Indian and Chinese isolates from infected aroids other than taro. It appears that, since this Group occupies a basal position within the tree that these sequences evolved from the most common recent ancestor (TMCRA) earlier than the other DsMV sequences. The second major lineage gave rise to Groups 2, 3 and 4, South Pacific samples are all within this lineage, suggesting common ancestry for all strains in this region (Figure 3.4). Group 2 was also comprised of sequences from a variety of locations isolated from different aroid hosts including tannia and taro. Group 3 was comprised of DsMV sequences from tannia from Cook Islands (e.g. CX3627), Vanuatu (E.g. VX3406) and Mariana Islands (e.g. MX4492) as well as from taro from Palau (PC4030) and Nauru (NC4549). Group 4 was made up of sequences from taro taken from Polynesia, specifically French Polynesia (e.g. FPC3870), Cook Islands (CC3688) and Samoa (e.g. Samoa 1), as well as the glasshouse samples from NZ and the Solomon Islands described in Chapter Two.

Within Group 2, two of the three isolates collected from infected taro in Palau (PC4007 and PC4006) appeared at two distinct positions, while the other isolate (PC4030) was in Group 3, suggesting two to three incursions of DsMV into Palau.

Isolates from Federated States of Micronesia were collected from infected tannia and all within Group 2, however the sequence of FSMX4060 was quite distinct from the others FSMX4039 and 4041, suggesting two incursions into this region. Interestingly, the sequences of FSMX4039 and 4044 are closely related to isolates from Cook Islands (CX3626), Palau (PC4006) and Vanuatu (VC3451), suggesting a dispersal event appears from Palau to the Cook Islands and subsequently into Vanuatu and Federated States of Micronesia.

Isolates from Mariana Islands appeared in both Group 2 and 3, two isolate collected from infected taro and tannia (MaC4424 and MAX4423) were placed in Group 2, and while the other three isolates (MAX4442, 4481, and 4492) collected from infected tannia were placed in Group 3. This suggests DsMV entered this region on at least two occasions. The relationship between each isolate suggests that those sequences in Group 2 entered the Mariana Islands before those in Group 3. The variation within a group was likely to have arisen since arrival.

Similar observation was made for Vanuatu. Isolates VC3451 (from taro in Group2) was a distinct sequence from tannia derived isolate VX3402, VX3406 and VX3415. This suggests two incursions have occurred, with variation in tannia arising since introduction. Glasshouse isolates A and B were placed within Group 4, among the South Pacific isolates. As mentioned in Chapter Two, A and B strains are known to have come from the Solomon Islands, since these viruses where first identified in field trial plants in this region. These isolates were split into two clades, the variability of DsMV in the original plants was not tested at outset, and the variability seen here could have arisen in the glasshouse from two incursions into the islands.

Cook Islands isolates appeared within Group 2, 3 and 4; one isolate from tannia (CX3626) was in Group 2 while two of the isolates (CX3626 and CX3633) collected from infected tannia were placed in Group 3, which are closely related to Vanuatu isolate VX3402. Another isolate collected from infected taro (CC3688) appeared within Group 4, closely related to the French Polynesia isolates FPC3870 and FPC3873 collected from infected taro. This suggested at least three incursions of DsMV into the Cook Islands. Isolates from French Polynesia all clustered in Group 4, suggesting only one incursion into this region. Four DsMV sequences from infected taro sampled in Samoa were included in Group 4, suggesting the virus entered this region also only

once. Isolates from Samoa appeared to be cluster with the NZ glasshouse isolates in Group 4; suggesting the NZ strains appear to have originated from Samoa. Based on this observation, there appear to be three evolutionary groups into NZ, two that originated from the Solomon and one originated from Samoa. VanMV from the Cook Islands and French Polynesia were also clustered together within Group 4; supporting the suggestion that this virus is a strain of DsMV (Farreyrol et al., 2006).

Only one sequence was studied from Nauru (NC4549) which appeared to be closely related to isolates from Vanuatu, the Cook Island, Mariana Islands and the Solomon Islands. This suggests movement of infected plants between locations. However, only one sequence was analysed in this study, therefore relationships may not be well resolved.

Based on these observations, it appears that DsMV originated in outer India or China and subsequently to South East Asia then subsequently throughout the Pacific. Overall, within Group 3, there appeared to have gradual dispersal events from the Micronesia (Palau, Mariana Islands and Nauru) to Polynesia (Cook Islands) and subsequently back incursions into Melanesia (Vanuatu). It is apparent that evolutionary relationships of DsMV in the South Pacific involved the combination of multiple incursions and radiation following several entries into different locations. Addition of South Pacific samples did not change the overall topology of tree when compared to the phylogenetic tree describing the temporal analysis in Chapter Two.

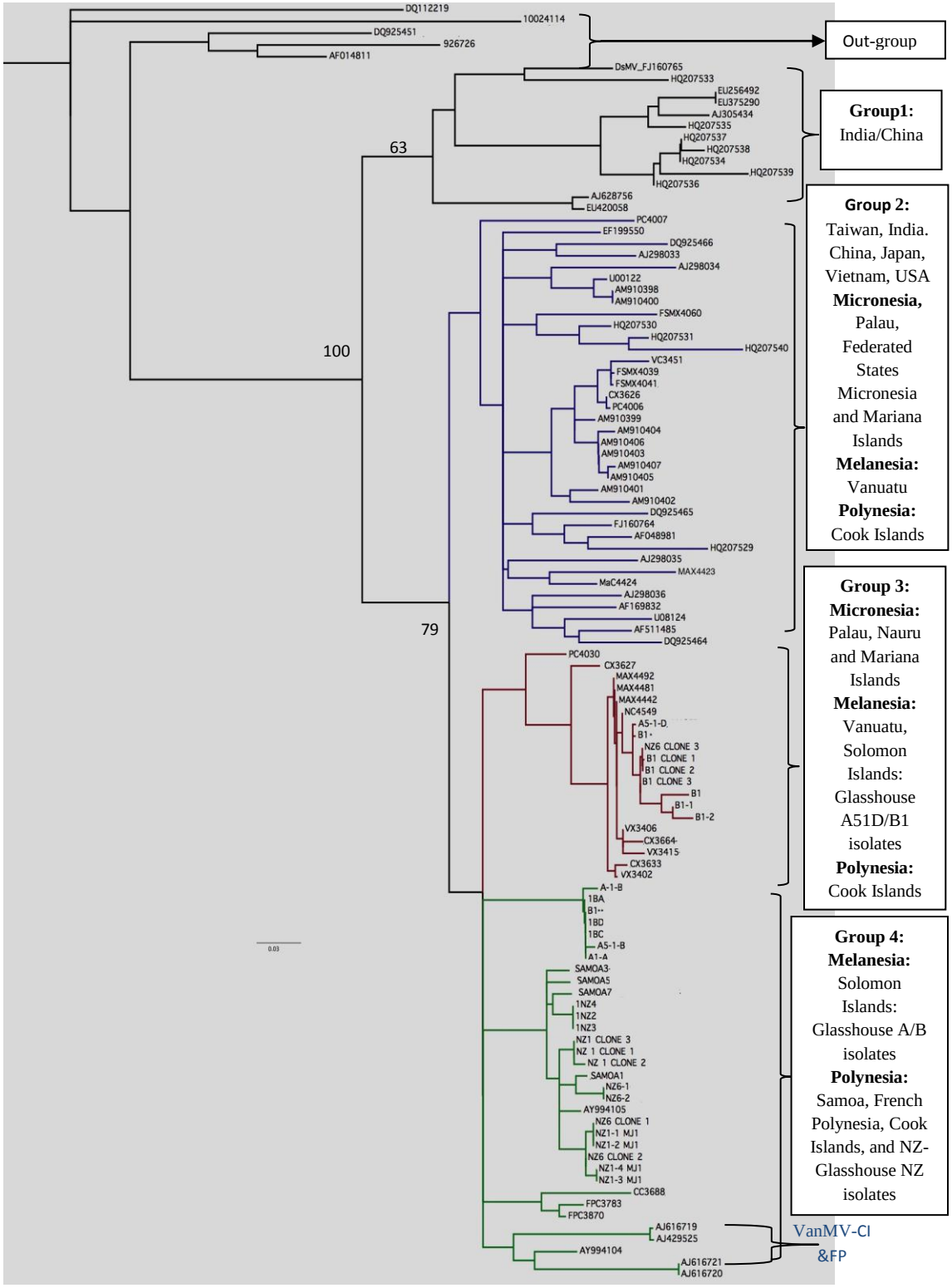


Figure 3.2: Neighbour-joining tree of the partial CP sequences of DsMV and related Potyviruses. The tree is rooted with PSFV, PWV and ZYMV as the out-group. Bootstrap values above 50 are shown and the scale bar shows the number of substitutions per base. B1* (top) and B1 ** (bottom) indicate sequences from a doublet sequence derive from previous study (C.Higgins, personal communication). Abbreviations, host range, location, sequences and accession numbers are shown in Appendix 1 and 2. CI denotes Cook Island and FP denotes French Polynesia

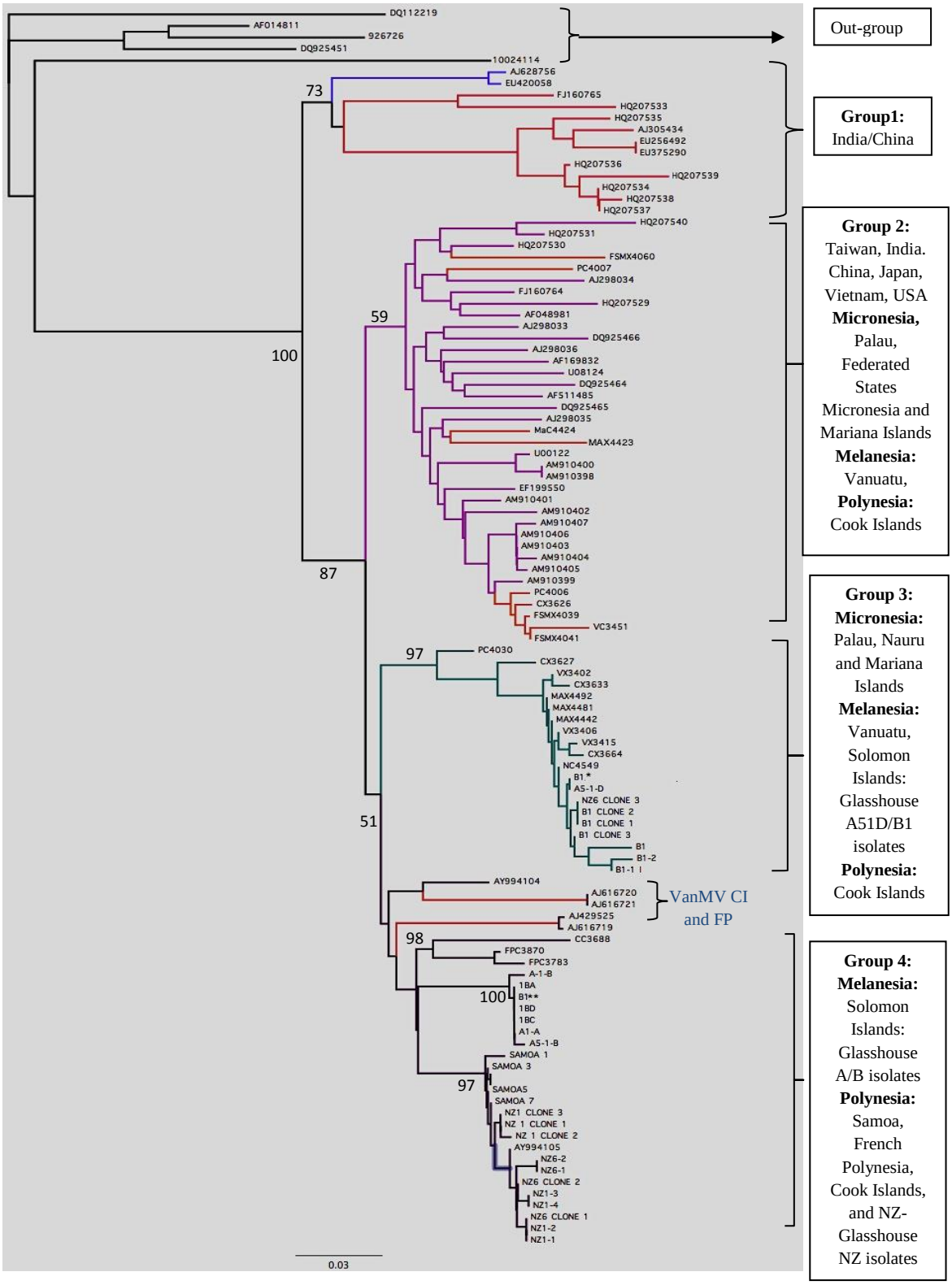


Figure 3.3: Maximum likelihood tree of the partial CP sequences of DsMV and related Potyviruses. The tree is rooted with PSFV, PWV and ZYMV as the out-group. Bootstrap values above 50 are shown and the scale bar shows the number of substitutions per base. B1* (top) and B1 ** (bottom) indicate sequences from a doublet sequence derive from previous study (C.Higgins, personal communication). Abbreviations, host range, location, sequences and accession numbers are shown in Appendix 1 and 2. CI denotes Cook Island and FP denotes French Polynesia

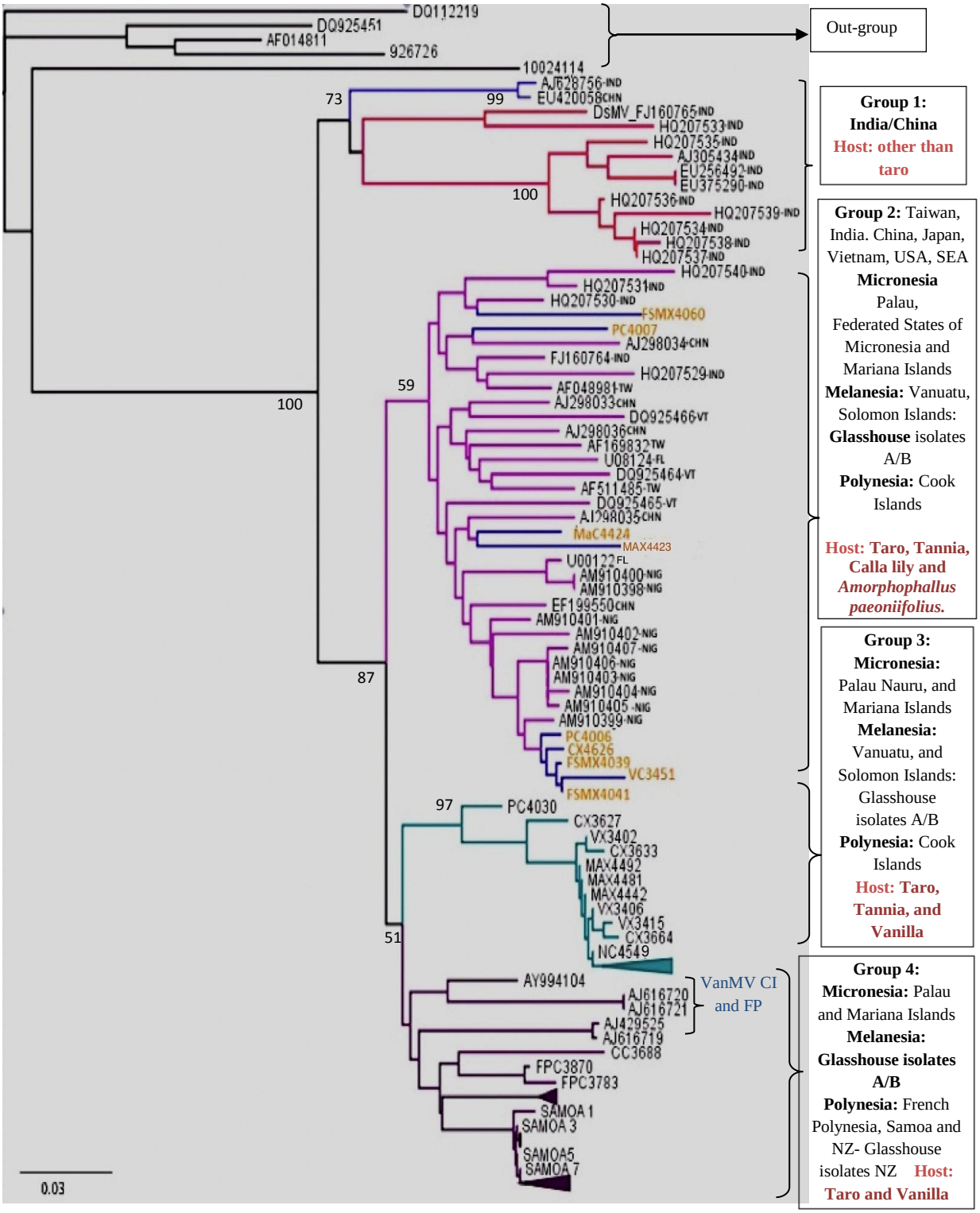


Figure 3.4: Maximum likelihood tree of the partial CP sequences of DsMV and related Potyviruses with glasshouse isolates compressed to make it easier to view the Pacific samples relationships. The tree is rooted with PSFV, PWV and ZYMV as the out-group in GeneiousPro. Bootstrap values above 50 are shown and the scale bar shows the number of substitutions per base. Abbreviations, host range, location, sequences and accession numbers are shown in Appendix 1 and 2. SEA: South East Asia, FP: French Polynesia, CI: Cook Islands, VT: Vietnam, IND: India, CHN: China, FL: Florida and NIG: Nicaragua. The scale bar shows the number of substitutions per base.

Sequences of the South Pacific isolates were compared with published DsMV sequences to determine the nucleotide and amino acid identity (Table 3.2). The nucleotide sequences within Groups indicated Group 3 (92%-100%) and 4 (89%-100%) and showed higher nucleotide identity compared to Group 1 (85%-100%) and 2 (82%-100%).

Between Groups 1, 2, 3 and 4 were equisimilar (79%-100%), with Group 3, closest to Group 4 (86%-100%). Result indicated greater variation observed between Group 1 and 2 (79%-86%) and 3 (79%-88%). At the amino acid (aa) identity level, Group 2 shared higher amino acid identity with the Group 4 (89-100%) compared to Group 3 (81%-100%) and Group 1 (86%-100%).

Comparison among the South Pacific strains of DsMV showed they were more similar to each other than the public sequences (87%-100% nucleotide identity compared to 79%-98%) and 80% -100% amino acid identity compared with 79%-100% (Table 3.3) with Group 2 being closest to the Micronesia at the nucleotide identity. This was also observed in the phylogenetic tree as most of the Micronesian isolates were placed at distinct positions within Group 2. Nucleotide identity within geographical region of the South Pacific isolates was equisimilar (87%-100%) with Melanesian and Polynesian isolates showing slightly lower diversity (88%-100%). Between geographical regions both nucleotide identity (86%-100%) and amino acid identity (80%-100%) were equisimilar.

While partial CP sequences were analysed here, percentage sequence identity and phylogenetic analysis supported these DsMV isolates from the South Pacific belonging to the same species based on the demarcation value of 76% nucleotide sequence identity for Potyviruses (Adams et al., 2005).

VanMV showed highest nucleotide identity on average 83%-91% with French Polynesia isolates compared to other South Pacific isolates (Table 3.4). Amino acid identity between VanMV and South Pacific isolates were 88%-99%, suggesting VanMV should be classified as a strain of DsMV as proposed by Farreyrol, et.al. (2006).

Table 3.2: Percentage of nucleotide identity (above diagonal in black), within group (red) and amino acid identity (below diagonal in blue) of sequences of the partial CP gene between Groups of the phylogenetic tree.

Isolates	Group 1	Group 2	Group 3	Group 4
Group 1	82%-100%	79%-86%	79%-88%	80%-88%
Group 2	86%-100%	85%-100%	81%-94%	82%-93%
Group 3	83%-93%	81%-100%	92%-100%	86%-100%
Group 4	93%-95%	89%-100%	88%-99%	89%-100%

Table 3.3: Percentage of nucleotide identity (above diagonal in black), within group/region (red) and amino acid identity (below diagonal in blue) sequences of the partial CP gene and between the South Pacific isolates and NCBI DsMV sequences.

Isolates	Group 1	Group 2 (Without South Pacific isolates)	MICRONESIA	MELANESIA	POLYNESIA
Group 1	82%-100%	82%-90%	79%-88%	79%-88%	80%-88%
Group 2 (Without South Pacific isolates)	**88%-100%**	88%-100%	83%-98%	81%-95%	82%-91%
MICRONESIA	**86%-95%**	**79%-100%**	87%-100%	87%-100%	86%-100%
MELANESIA	**83%-95%**	**79%-100%**	**80%-100%**	88%-100%	86%-100%
POLYNESIA	**92%-95%**	**90%-100%**	**83%-100%**	**80%-100%**	88%-100%

Table 3.4: Percentage of nucleotide identity (nt) and amino acid identity (aa) sequences of the partial CP gene between the South Pacific isolates and VanMV

Isolates	MICRONESIA	MELANESIA	POLYNESIA
VanMV- CI (Cook Island)	86%-88% (nt) 89%-98% (aa)	87%-88% (nt) 88%-98% (aa)	87%-91% (nt) 89%-99% (aa)
VanMV- FI (French Polynesia)	86%-88% (nt) 89%-98% (aa)	86%-87% (nt) 88%-98% (aa)	83%-90% (nt) 89%-99% (aa)

3.3.3 Selection analysis

The ratio of nonsynonymous (d_N) to synonymous (d_S) nucleotide substitution rate ($\omega = d_N/d_S$) was determined to provide a measure of the selective constraints at the protein level (Yang, 2007). All isolates were subjected to strong purifying selection (Table 3.5). Similar studies have been performed on other Potyviruses. For example in a study conducted by Simmons et. al., (2008) selection analysis on CP sequences of ZYMV indicated predominant evolutionary pressure was that of negative (purifying) selection. A major drawback in detection of positive selection can be attributed to the fact that the test of molecular adaptation is highly conservative; it will fail if positive selection affects only a few amino acid sites along a few lineages on the phylogeny (Twiddy et al., 2002). As only a short fragment of the CP gene was analysed in this study, further studies involving full-length CP sequences with more comprehensive methods should be perform to detect variation of selective forces across the full- length CP sequence. This would allow one to investigate how selective forces affect the patterns of variation change across the full-length CP sequence; especially to the detect signatures of natural selection or identification of changes in mutation or recombination rates.

Table 3.5: Selection analysis (non-synonymous/synonymous ratio) exerted on the partial CP gene between South Pacific isolates. * ω ratio greater than one implies positive selection; less than one implies purifying (stabilizing) selection, and a ratio of one may implies neutral selection.

Isolates	Location	d_N/d_S ratio (*ω) ratio
VC3451	**Vanuatu**	0.03
VX3402		0.02
VX3406		0.04
VX3415		0.01
SAMOA7	**Samoa**	0.01
SAMOA5		0.05
SAMOA3		0.04
SAMOA1		0.03
PC4030	**Palau**	0.01
PC4006		0.05
PC4007		0.07
NC4549	**Nauru**	0.01
MAX4442	**Mariana Islands**	0.05
MAX4481		0.04
MAX4492		0.05
MaC4423		0.04
MaC4424		0.03
FPC3783	**French Polynesia**	0.04
FPC3870		0.04
FSMX4039	**Federated States of Micronesia**	0.02
FSMX4041		0.03
FSMX4060		0.03
CX3626	**Cook Islands**	0.03
CX3627		0.02
CX3633		0.04
CX3664		0.02
CC3688		0.03

3.3.4 Spatial pattern analysis

Comparative analysis of pairwise genetic distances and geographic distances was performed to further elucidate the evolutionary and spatial relationship of South Pacific isolates, and infer the pattern of transfer of viral populations through infected materials among the various South Pacific Islands are involved. Results indicated on average, the greater the geographical distance, the higher the genetic distance suggesting greater

diversity between the isolates. Association between geographic and pairwise genetic distances was calculated between the geographical locations shown in Table 3.6. A plot of the average pairwise genetic distance vs. the great circle (geographical) distance (km) showed a strong correlation (Figure 3.5, R^2 = 0.73). Geographical distances between 4000 and 8000 km were associated with pairwise genetic distances between 0.09 and 0.11 corresponding to significant divergence between two locations. For example Samoan isolates show genetic distance of 0.11 with Palauan isolates in relation to a geographical distance of 6377km. Geographical distances between 1000 and 4000 km were associated with genetic distances ranging from 0.06 to 0.08, corresponding to closely related variants between the geographical locations.

Spatial association between average pairwise genetic and geographic distance was confirmed by the Pearson statistical analysis. The observed statistic in Table 3.7 (r = 0.851, $P < 0.001$) was significantly outside the distribution of statistic values from the randomization test. These results indicate some evidence of isolation of distance, which was also apparent in the phylogenetic analysis. This association also suggested the virus appears to have spread along from Micronesia (Palau, Micronesia and Mariana Islands) and gradually descended towards Melanesia (Vanuatu) and Polynesia (Samoa, Cook Islands, French Polynesia). However, the point of origin is difficult to identify, particularly in some cases distant isolates are genetically similar for example isolates from Cook Islands and Federated States of Micronesia. This indicated several incursions of the virus into each geographical location as inferred in the ML tree in Group 2 and 3 (Figure 3.4). Therefore, it is possible that the spread may have originated from Micronesia to Polynesia (most likely in the Cook Islands) and then incursions back into Melanesia and Micronesia, thus generating genetic homogeneity among some isolates. However, considering only a small fragment of the CP gene and sample sets from the South Pacific Islands were used to analyse the spatial association, the complete spatial pattern of DsMV might not be fully resolved. Further studies using full-length CP sequences and larger sample sets from the South Pacific Islands would help to elucidate the spatial association. This would allow for a better and more complete knowledge of the spatial pattern of DsMV.

Table 3. 6: Average great circle geographic distance (km) (above diagonal) versus average pairwise genetic distance for South Pacific isolates

	MICRONESIA				MELANESIA	POLYNESIA		
Isolates	**Mariana Islands**	**Palau**	**Federated States of Micronesia**	**Nauru**	**Vanuatu**	**Samoa**	**Cook Islands**	**French Polynesia**
Mariana Islands		1536	1989	2982	4504	5745	4509	8025
Palau	0.07		2614	3699	4641	6377	7500	8776
Federated States of Micronesia	0.07	0.06		2614	2958	4037	5548	6344
Nauru	0.07	0.08	0.10		1917	2770	2659	5115
Vanuatu	0.09	0.09	0.07	0.06		2174	3360	4450
Samoa	0.10	0.11	0.11	0.06	0.08		1511	2407
Cook Islands	0.06	0.11	0.10	0.07	0.07	0.06		1143
French Polynesia	0.10	0.11	0.11	0.10	0.10	0.07	0.08	

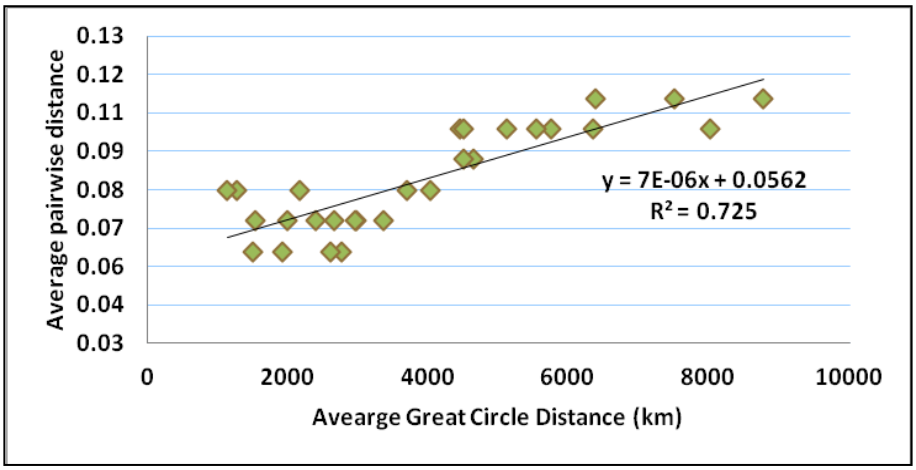


Figure 3.5: Correlation between the average Great circle distance (km) and the average genetic distance between the isolates ($r=^2$ 0.837, $p<0.001$)

Table 3.7: Pearson statistic showing association with geographical distances and genetic distances among South Pacific isolates. The observed Pearson statistic (r = 0.851, P< 0.001) is significantly outside the distribution of statistic values from the randomization test.

Pearson Correlation Statistic	
r =the Pearson product-moment correlation coefficient	0.851
r^2 = the coefficient of determination	0.725
Slope	7.00E-06
Standard error of estimate	0.009
df	26
p-value	<0.001
H_0: no correlation between average genetic and geographical distances between populations. **Reject null hypothesis**	

3.3.5 Molecular clock- Divergence time of DsMV

A molecular clock under the relaxed clock model was used to estimate the divergence time among South Pacific isolates in relation to human migration. The most common branch point of all DsMV isolates and strains included in this analysis placed the origin of this virus around 15,000 years BP (Figure 3.6). This correlated with the overall groupings of DsMV in the phylogenetic tree (Figure 3.4). The MRCA of the primary lineage of the South Pacific isolates including Group 2 isolates was estimated to be 7,000 years BP (95% HDP) shown in Figure 3.6, indicating a spread from Eastern Asia into South East Asia and gradually into the South Pacific Islands. This also correlated with the phylogenetic tree as clades for Group 2, 3 and 4 diverged from a common ancestor. Similar study by Gibbs et.al, (2008), also indicated the initial major radiation of the Potyviruses occurred around 6,600 years BP from South East Asia to Australia and the Pacific Islands.

To further investigate the spread of DsMV into the South Pacific, correlation between human migration into the South Pacific Islands and the divergence time provides plausible explanations for the observations in the molecular clock. According to Patrick et.al, (2010) the initial human settlement in Polynesia originated from the dispersal of humans out of Micronesia between 3,500-2,700 BP (grey arrow) (Figure 3.7). This suggested that DsMV from Mariana Islands, Federated States of Micronesia and Samoa diverged from the MCRA during the first wave of human migration into Oceania (blue

dotted line). Isolates from Cook Islands, New Zealand (NZ) and Vanuatu also appeared to have diverged from the MCRA during the first wave of human migration into Oceania. The dotted black arrow in the map shows the proposed arrival of a new population (or populations) from Micronesia into Polynesia (2,000-1,500 BP). This suggested French Polynesia and Palau isolates diverged from the MCRA (red dotted arrows) during this second wave of human migration into Oceania. The black arrows in the map showed the settlement of East Polynesia and back migration into Melanesia (1200-500 BP). Divergence time for Nauru cannot be calculated accurately as only one sequence was presented.

Overall the molecular clock analysis suggested that for one particular location, the spread of DsMV had to have happened at the same time, which is unlikely based on the phylogenetic tree. For example, according to the phylogenetic tree isolates from Palau were in Group 2 and 3, which according to the clock separated 7,000 years ago. However, the clock estimated the MCRA of Palau isolates appeared 900 years ago. This disparity may be due to the use of a "relaxed-clock" model, which does not assume substitution rate autocorrelation across lineages time scales, particularly at long time scales (Ho, Phillips, Cooper, & Drummond, 2005). When enough time has passed, many sites have undergone more than one change, but it is impossible to detect more than one. This means that the observed number of changes in the Palauan isolates was no longer linear with time, instead flattens out and can affect the divergence time (Ho, Phillips, Cooper, & Drummond, 2005). Further, analysis did not consider recombination of viral genomes. Sequences analysed in this study was too short for a more stringent clock applied. Therefore, for more recent events, longer sequences should be applied.

The same disparity was also observed for the majority NZ samples, which share ancestry with the Samoa DsMV and an earlier ancestry with one sequence from the Cook Islands (CC3688). The molecular clock implies a closer relationship between DsMV from NZ and the Cook Islands diverging 2,126 years BP, with the extant virus arising approximately 800 years ago. Considering the movement of humans, the settlement of the Cook Islands originated from the dispersal of humans out of Samoa around 1,200 years BP, with subsequent arrival of Polynesians from the Cook Islands into NZ and French Polynesia around 500-1,000 years BP (Matisoo-Smith et al., 1998). This could potentially reflect the observation in the molecular clock and phylogenetic

tree for the Cook Islands, NZ and French Polynesia isolates. Regardless of where it originated, the appearance of the virus in NZ coincides with approximation of the movement of people into NZ. It is known that taro was brought by the Maoris from Eastern Polynesia (Matisoo-Smith et al., 1998), hence from this analysis it would seem that some of these plants were infected with DsMV.

It should be noted that glasshouse isolates NZ, A and B were not included in the molecular clock analysis as the timescale (i.e. generation times) was too short for analysis. This can cause differences between samples that do not represent fixation of different sequences in the different viral populations. The inclusion of differences that have not yet become fixed leads to a potentially dramatic inflation of the apparent rate of the molecular clock at very short timescales (Ho, Phillips, Cooper, & Drummond, 2005).

Molecular clock test was performed to determine if there are differences in evolutionary rates among the South Pacific isolates. Statistical analysis was modelled using a discrete Gamma (G) distribution (shape parameter shown). Under this model the mean rate of evolutionary rate for DsMV was 3.51×10^{-5} nucleotide substitutions per site, per year with a standard error of 2.34×10^{-6}, giving a 95% confidence range of 2.78×10^{-5} to 4.22×10^{-5} (Table 3.8). The coefficient variation was estimated to be at 0.95, indicating the evolutionary rate is not constant, opposed to a value of zero which indicates a constant rate of evolution (Drummond, et.al 2005). Hence, the null hypothesis of equal evolutionary rate throughout the tree was rejected at a 5% significance level ($P < 0.05$) confirming different evolutionary rate among the lineages of South Pacific isolates.

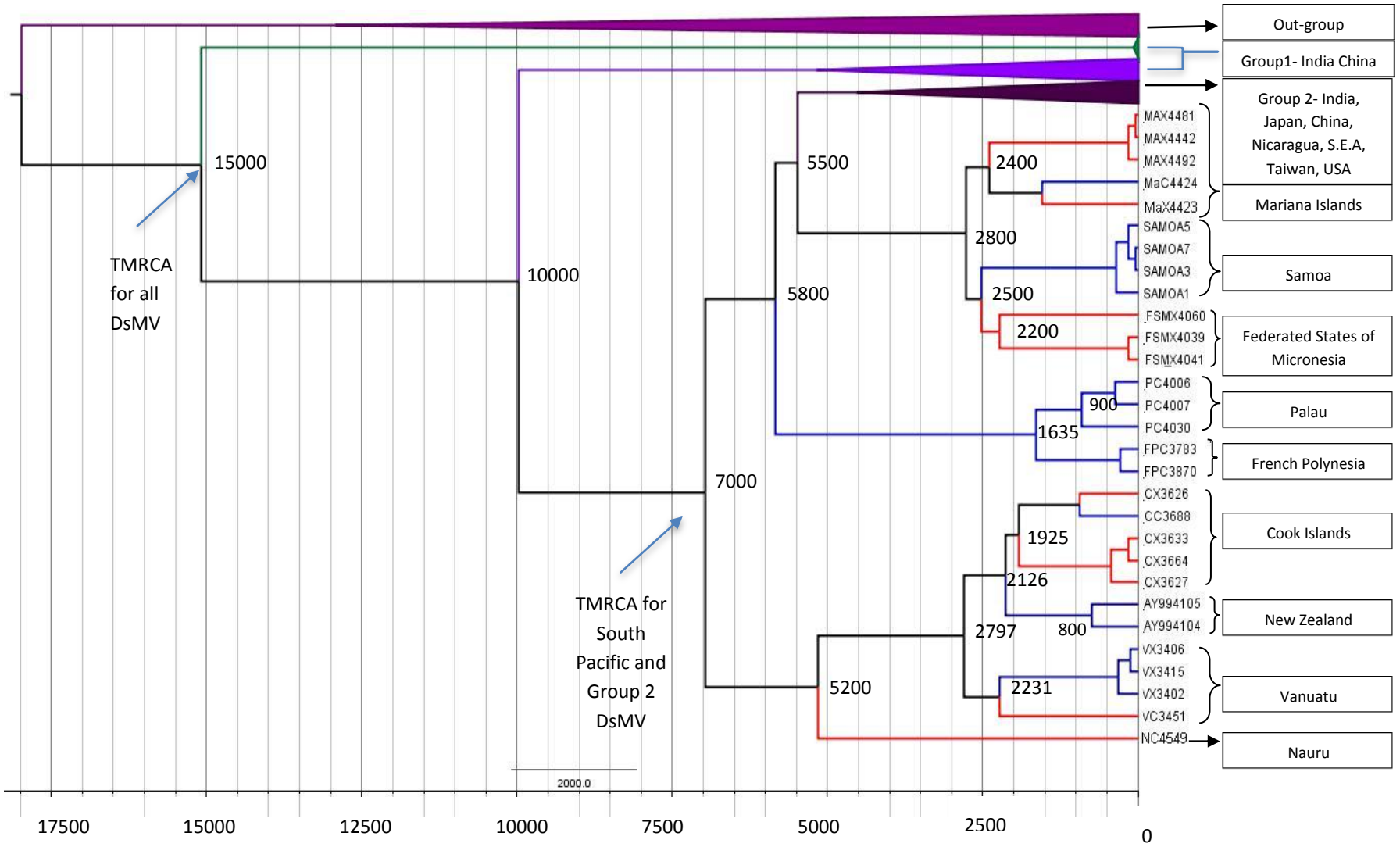


Figure 3.6: Molecular clock of estimated average divergence times of South Pacific DsMV isolates. The tree was reconstructed from the partial CP sequences by Bayesian inference under an uncorrelated lognormal relaxed molecular clock model to calculate the divergence based on their evolutionary rate. Average divergence times based on 95% HDP intervals (BP -years before present) are positioned at the nodes. The species names and the sequence accession numbers are given in Appendix 1.

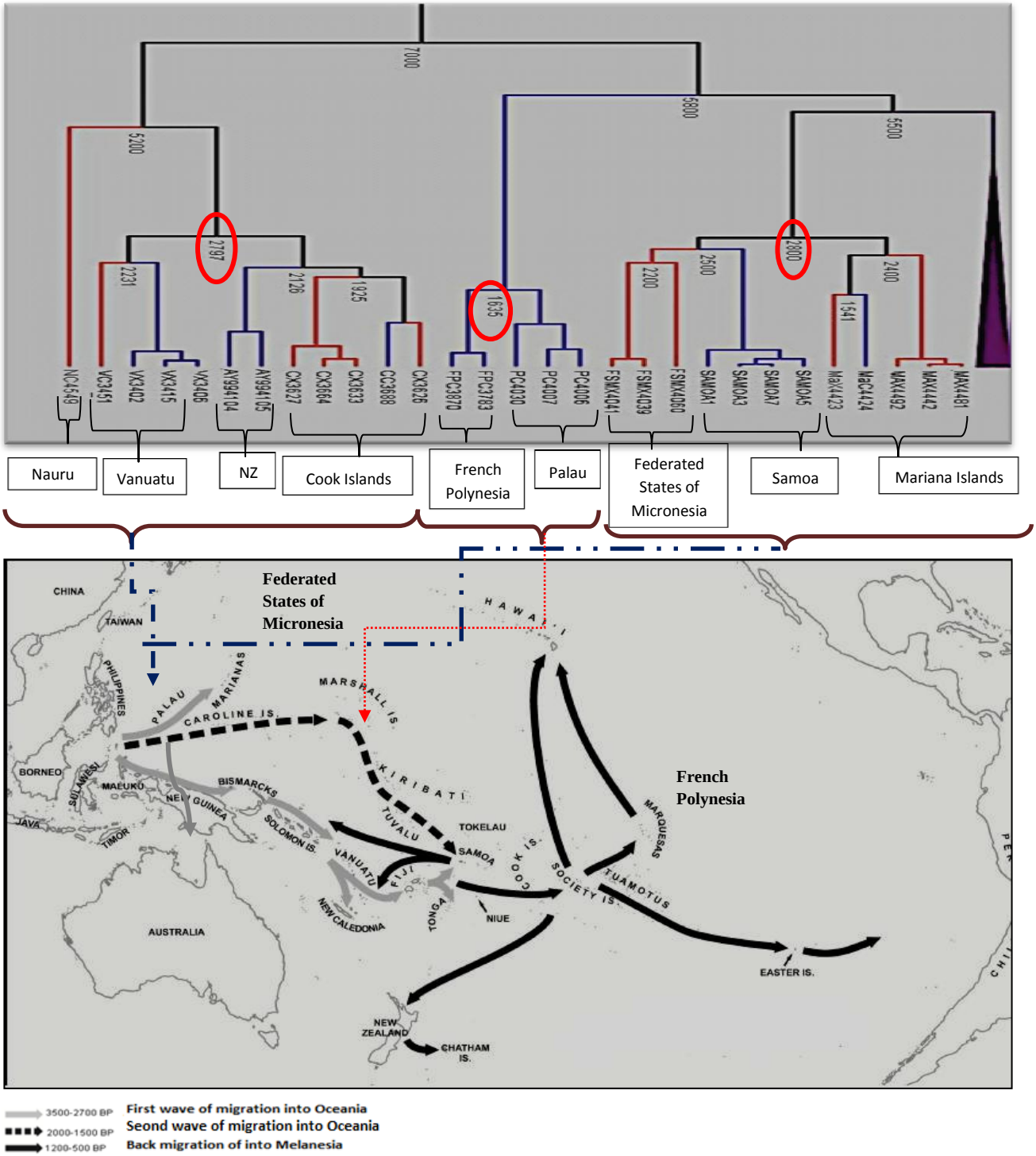


Figure 3.7: Correlation of molecular clock and the movement of humans into Oceania indicating divergence time of South Pacific isolates from the MCRA. Scale for timeline is in BP (before present day). The dotted blue lines represent the dispersal of virus due to the first wave of human migrations. The dotted red lines represent the dispersal of virus due to the second wave of human migrations. The dates in red circles represent the MCRA for the respective South Pacific isolates.

Image adapted and modified from Addison and Matisoo-Smith (2010).

3.3.6 Molecular clock testing

Table 3.8: Statistical test of molecular clocks using the Maximum Likelihood method (lnL) and the tree statistics. HDP denotes 95% highest density probability value.

Tree statistics

Evolutionary rate among the DsMV isolates	**$3.51x10^{-5}$**
Standard error	$2.34x10^{-6}$
Co-efficient of variation	0.905
95% HPD lower	$2.78x10^{-5}$
95% HPD upper	$4.22x10^{-5}$

3.4 Discussion

The aim of this study was to understand the long-term evolution of DsMV. From the pool of sequence variants that exist within a single plant, movement of infected plants from one location to another as humans migrate will result in differing founding population at each new location. To understand the spatial evolution of DsMV, viral sequences from infected samples from various Pacific Islands were compared to determine the evolutionary relationships between them.

Phylogenetic and spatial analyses of South Pacific DsMV isolates

In this study, partial DsMV CP gene sequences from several South Pacific isolates of DsMV were compared to equivalent sequences in the public domain. DsMV infected taro and tannia were sampled from various locations in Vanuatu, Samoa, Polynesia, The Cook Islands, Federated States of Micronesia, Nauru and Palau. This sample set is not exhaustive; several locations such as Tonga, Fiji and Hawaii were not represented. Irrespective of this, the analyses provided here demonstrated that DsMV is variable and had a complex history.

While there have been several variability studies focused on other Potyviruses sequences, for example in PRSV and ZYMV (Olarte Castillo et al., 2011; Simmons, Holmes, & Stephenson, 2011), there is little reported on the variation of DsMV. Other studies have differed from those reported here, as the variability studies focussed on isolates within a region in a country and/or on different gene within the Potyvirus genome. Hence, to some extent the conclusions are not comparable.

Phylogenetic analyses of DsMV revealed distinct clades of viral isolates; however, there was no compelling evidence for a relationship between the host and geographical distribution unless a location had been heavily sampled, such as Nicaragua, Solomon Islands and New Zealand. For example, the large Group 2 clade contained some of the South Pacific isolates and also other isolates as far apart as Central America and India. Further, this clade included DsMV sampled from different hosts such as *Colocasia*, *Pinellia*, *Caladium* and *Zantedeschia*.

South Pacific isolates were also placed in the Group 4 clade in the phylogenetic trees. These analyses revealed significant phylogeographical complexity for DsMV isolates. Nucleotide identity between the South Pacific isolates, indicated isolates was equisimilar and composed of closely related variants, despite the fact they are from different hosts and locations. However, some isolates between geographic regions indicated lower diversity, for example Melanesia (Vanuatu and Solomon Islands) isolates showed lower diversity compared to Micronesia (Mariana Islands, Palau, Nauru and Federated States of Micronesia). This indicates that the maintenance of genetic diversity among the DsMV isolates may be facilitated by three distinct mechanisms; multiple incursions of viral strains into the region, vector transmission bottlenecks, and viral populations that were genetically isolated from one another such that stochastic processes influenced divergence (Elena, et.al, 2011).

Spatial analyses of the South Pacific isolates indicated genetic distance increases with geographic distance. However, in some instances some distant isolates (from sites >4000 km apart) were genetically close. This relationship could potentially result from strains frequently transmitted across geographic distances. This reflects multiple incursions into the geographic location with the transportation of vegetatively propagated infected plant material, which was also inferred by the phylogeny. A major limitation in this study was that only one geographic location represented the outset region of Melanesia; Vanuatu. Hence, for accurate determination of spatial relationship among the South Pacific isolates, further longitudinal sampling of DsMV isolates collected from a larger number of locations within the South Pacific regions would provide more power to the analyses.

Analyses of *dn/ds* ratio demonstrated strong negative selection dominates the partial CP gene sequences. It should be noted that although this research did not specifically test for evidence (or absence) of adaptation here, the results provided no indication that positive selection has played an important role in the long-term evolution of this sequence. A similar kind of analysis was performed which showed similar observation; in a recent study conducted by Bousalem et. al., (2000), the estimation of the *dn/ds* ratio was used to evaluate the hypothesis of an evolutionary response of the YMMV partial CP to accommodate new host species or geographical adaptation. This study showed that

purifying selection dominated the evolution of the YMMV CP gene. However, the overall negative selection can mask punctual positive selection operating on a few positions as evidenced recently with Potyviruses (Bousalem et.al, 2000). Considering that CP gene are highly conserved it is highly likely that adaptation may play a minor role in DsMV evolution. Perhaps, full CP sequences would probably help to elucidate these unclear aspects of DsMV evolution rather than the partial CP sequences.

Molecular clock of DsMV

Molecular clock analysis was performed to estimate the timing of divergence of DsMV within and beyond the South Pacific isolates. The analysis performed here points to India as being the most likely region of origin, around 15,000 years BP which coincide with the notion that taro was thought to have originated in the Indo-Malayan region, in eastern India and Bangladesh around 10,000 – 20,000 years BP (Bradshaw, 2010; McCoy & Graves, 2010). This suggests that DsMV has spread with the movement of taro from India into other geographical regions since that time.

The primary lineage of the South Pacific isolates including the Group 2 isolates was estimated from the molecular clock to have averaged around 7,000 years BP, indicating a spread eastward into Eastern Asia, South East Asia, and the Pacific islands from India. As infected taro has been cultivated and moved into different geographical locations, this allowed the virus to move more effectively and spread among localities. The introduction of tannia to Asia by Spaniards in the mid-16th century and its subsequent establishment in the South Pacific Islands (Bradshaw, 2010) will have represented a new niche for viral infection, where tannia grown nearby would be subsequently be infected and further enhancing the spread of DsMV. Results also demonstrated subsequent incursions of DsMV into many region of the South Pacific, for example the common branch point of all South Pacific DsMV isolates included in this analyses placed the origin of this virus around 2,800 years ago which coincide with the three major migration waves of humans into the South Pacific. This indicates the potential of variant mixing within and among the geographic regions. Radiation following entry into each region would then generate higher genetic diversity among viral populations.

However, given the course of evolution, gene frequencies are similar across large areas because drift and gene flow equilibrate on a time scale longer than that of viral population turnover (Roossnick, 2011). The overall geographic distribution of the South Pacific DsMV isolates suggested a gradual wave-like pattern with expansion of virus with succession of founder effects and subsequent diversification phases. Genetic homogeneity becomes more apparent as a consequence of isolation of distance when agriculture practices become more substantial within a region and human migration is infrequent (Roossnick, 2011). The non-persistent transmission of DsMV by aphids may result in the inoculation of very limited number of particles; hence vectors will transfer a small subset of the virus diversity from one region into contiguous regions.

Comparisons between phylogenetic analysis and molecular clock suggest that the appearance of DsMV in various Pacific locations does not necessary correlate directly with the movement of people throughout Pacific. This could be because a relaxed clock was used, due to short sequence was analysed. Further, molecular clock analysis would not give information about trade being carried out between islands. The introduction of tannia to Asia by Spaniards in the mid-16th century and its subsequent establishment in the South Pacific Islands would have even further complicated the estimation of divergence time. Clearly, the origin of DsMV is more complex that can be analysed by the movement of people into new location.

Under the relaxed molecular clock model, the overall estimated evolutionary rate for DsMV was 3.51 x 10^{-5} nt/site/year. The evolutionary rate estimated here was different to that observed in other studies of plant viruses. For Rice yellow mottle virus (RYMV), the rate was estimated from samples collected over 40 years to be $4–8\times10^{-4}$ ns/s/yr. (Fargette et al., 2008). For Begomovirus viruses, Tomato yellow leaf curl (TYLC) virus was estimated to be evolving at 4.6×10^{-4} ns/s/yr. from samples collected over 18 years (Duffy, Elena & Holmes, 2008; Ho, Phillips, Cooper & Drummond, 2005). These estimates were for different viruses and may be species specific or possibly genus/family specific and it is noticeable that samples analysed over a short time span are several-fold greater than those based on longer time, as has been noted in other studies of the rates of evolution of cellular organisms (Ho, Phillips, Cooper & Drummond, 2005). The discrepancy between short- and

long-term rates can be attributed to the difference in mutation rate, population size and replication fidelity of the virus (Duffy, Elena & Holmes, 2008).

In contrast, evolution rate determined by Gibbs et. al., (2008) with fifty partial coat protein genes of Potyviruses was estimated to be 1.15×10^{-4} nucleotide substitutions/site/year. This suggests the evolutionary rate of DsMV is much lower than expected. The difference in evolutionary rate compared to this study could be due to several reasons (Ho, Phillips, Cooper & Drummond, 2005); (i) samples used in this study were primarily DsMV and this could reflect the difference in mutations accumulate over time when comparing to different species, and (ii) the use of different software (PAUP vs. BEAST) to generate divergence time based on different statistical algorithms could reflect the difference in the evolutionary rate estimation.

It should be noted that although recombination within the South Pacific isolates was not tested for, the divergence times in the molecular clock could be affected by recombination events. (Lefeuvre et al., 2011). For example, in a recent study conducted by Castillo, et.al., (2011), CP and HC-Pro molecular clock trees were analysed in order to compare the resolution of PRSV evolutionary relationships for each sequence separately and together. However, both trees generated different topologies due to recombination events in the isolates. This highlights the limitation of molecular clock analyses where small levels of recombination can invalidate the likelihood ratio test of the molecular clock. It was unlikely that recombination was a major influenced this analysis since the sequence was short.

In conclusion, the spatial analyses of DsMV isolates described here offer a view into the long- term evolutionary dynamic of the DsMV isolates. The overall observation indicates variation among the South Pacific Islands isolates can be summarized as follows; (i) multiple gradual dispersal of DsMV has occurred across the geographical distance, and that spatial pattern analyses observed suggested a shift from multiple gradual to stochastic dispersal (ii) existence of purifying selection driven by the host could indeed lead to fixation of beneficial mutation and (iii) genetic differentiation observed between the South Pacific isolates, suggests isolation of distance over the course of evolution, and illustrating again the stochastic character of these events.

Finally, the spatial analysis of South Pacific DsMV offers a long-term view into the evolution of this virus. Perhaps the most important factors in the long-term evolution have been human migration and the development of agriculture practices, which allowed the rapid dissemination of viruses to diverse geographical areas. It is the relative frequency of migration versus population isolation that is fundamental in shaping the phylogeographical structure of these viral populations.

Relationship between DsMV and VanMV

VanMV was classified as a member of the genus Potyvirus in 1986 when it was reported in French Polynesia causing leaf distortion and mosaic in *Vanilla tahitensis* (Fauquet et al., 2005). Subsequently, it has been detected in *Vanilla sp.* in the Cook Islands, Fiji and Vanuatu. It has been demonstrated that VanMV is serologically related to DsMV (Wisler, Zettler, & Mu, 1987). In a study conducted by Farreyrol et.al, 2006, sequence comparisons between VanMV and DsMV showed them to be within the boundary established by the ICTV between species and strain. Some of the data in the study supported VanMV should be considered distinct from DsMV due to the following reasons: (i) both have separate host ranges, (ii) the CI and NTR nucleotide sequences of VanMV were less than 75% identical to that of DsMV strains, and (iii) VanMV had a rare Q//V CP-NIb cleavage site, different from that of DsMV. However, phylogeny of the CP core region, and the high percentage identities across the CP core and across the CI-6K2-NIa region favoured the hypothesis that VanMV and DsMV are strains of the same virus (Farreyrol et al., 2006).

Similar observation was observed in this study; phylogenetic analyses revealed sequence similarity between DsMV and VanMV, and that VanMV should be classified as a strain of DsMV. Perhaps, a plausible explanation could be attributed to a 'host –shift' that would have occurred for DsMV to acquire ability to replicate in a new host. However, this does not exclude the possibility that VanMV arose from recombination between DsMV and another parent. If both taro and vanilla are planted in the same field, this could increase the chance of aphid transmission of potential variants to cause species jump. Based on a recent publication (Longdon, Hadfield, Webster, Obbard, & Jiggins, 2011), it was suggested that the source of host switching may often be predictable from the host phylogeny, but that the

effect may be more complex than simply causing host shifts between closely related hosts. A host switch may require both fitness cost to the virus and the substantial changes in ecological niches. When ecological conditions allow for frequent host jumping, such as communities rich in biodiversity, allopatric divergence would be expected to be particularly frequent (Pagan & Holmes, 2010).

On the other hand, for host jumping to be successful, viruses have to evolve as combinations of genes whose products interact with cellular components to produce progeny throughout the plants and, in most cases, to interact with a vector to be moved to other plants. A variant of DsMV could have jumped host to infect vanilla, due to the fact that the viral gene involved in cell to cell movement had evolved to allow host range to change. Additionally, plants have host–defense mechanisms including RNA silencing, which must be overcome by the virus for effective movement within the plant. Therefore, it is possible that a particular variant of DsMV may have evolved gene products to suppress these defense mechanisms.

Nevertheless, this observation highlights the key role that viral speciation or host jumping are some of the consequences of the evolution of RNA viruses, yet further emphasizes the need to better study and understand viral biodiversity and host range. Further research should be carried out on the molecular barriers to both cellular and host switching for these RNA viruses to understand viral biodiversity and host range.

Chapter 4

Final Discussion

4.1 Final Discussion and Conclusion

RNA viruses are characterized by a high genetic variability, which has been mainly attributed to the absence of a proofreading activity in RNA replicases (Sztuba-Solińska, Urbanowicz, Figlerowicz, & Bujarski, 2011). DsMV was used as a model to understand the mechanisms of genetic variation of an RNA virus at a temporal and spatial scale. The first aim of this thesis was to study the genetic variation of DsMV within and between host taro plants that had been propagated for three generations. This study was done to determine the mechanisms of short-term evolution among virus populations. To that extent SSCP was used to characterize sequence polymorphism in virus populations within and between host plants. For the second aim of this thesis, spatial variation of DsMV isolates infecting taro and tannia in various South Pacific locations were evaluated in order to understand long-term evolution of the virus. Taro and tannia contribute important staple foods in the South Pacific; hence this study will provide some aspect in the understanding of the epidemiology of DsMV.

In the temporal analyses of glasshouse grown DsMV isolates, SSCP has proven to be a reliable tool for screening polymorphism between virus populations. Through this analysis, it was possible to characterise large sample sets efficiently and to visually identify samples for warranted analysis. However, limitations of this technique meant that cloning and subsequent sequencing were needed to determine the extent of genetic variation. From this study, two contrasting situations were observed; within host analyses revealed little or no genetic variation of the virus, whereas genetic variation was observed between parent and propagants plants. Genetic bottlenecks appeared to limit the number of randomly selected virus genomes, thereby creating a founding population, resulting in genetic drift. This highlighted that genetic bottlenecks may occur frequently during the natural life cycles of RNA viruses. Various events required for virus movement and transmission for example, attachment to an aphid, represent events in the virus life cycle that may impose a bottleneck. Evidence from this study suggested that small population diversity, and genetic drift, were the key determinants for the genetic variation at the temporal scale.

Overall, the differences between intrahost and interhost diversity can be attributed to the systematic movement of an effective size of viral populations which determines whether selection and/or genetic drift is the predominant force shaping their genetic structure and evolution. If the population is large, then selection (influenced by the host) is likely to be the predominant force. If small then genetic drift is more likely to predominate.

In contrast, spatial analysis showed that genetic variation occurs in the South Pacific isolates and this study showed multiple gradual dispersal of DsMV has occurred across the geographical distance, and that spatial pattern analyses suggested a shift from multiple gradual to stochastic dispersal occurs through selection process. As fitness is the product of a genotype in an environment, selection pressure may direct effects on the long term evolutionary process of the South Pacific DsMV isolates by influencing the strength of selection relative to drift (e.g. through changes in population size), with possible consequences for the evolution of genetic robustness and genome complexity.

Nevertheless, fitness trajectory in long-term evolution can be initially high but tend to decelerate over time (Elena, et.al., 2011). Such dynamics indicate that, after being placed in a new environment, populations are evolving from a region of low fitness toward an adaptive peak or plateau. Variants that carry different beneficial mutations compete with one another, thereby interfering with their spread and substitution in the population. Thus, in time many beneficial mutations become transiently common only to be excluded later by interfering mutations, giving rise to a leapfrog event in which the most abundant genotype at a given moment is phylogenetically related to an earlier dominant time than to the immediately preceding one (Sardanyes & Elena, 2011). Host radiation allows the virus to expand its ecological niche by adapting to one or more novel hosts. With fitness trade-offs and ever changing environments, these can promote the long-term evolution of the virus.

Despite the remarkable evolutionary potential of RNA viruses, it is important to stress that factors such as deleterious mutations and purifying selection can create trade-offs, affecting RNA virus evolution. A high mutation rate may or may not have evolved as a strategy for accelerating adaptation and it is possible that the mechanisms of replication fidelity impose a fitness burden on systems, as RNA viruses, that rely critically on their replication rate.

Overall, both studies presented have demonstrated that genetic variation occurred at both temporal and spatial scale. It can be seen that the main steps involved in long-term RNA virus evolution have some overlap with some of the steps that determine short-term evolution in quasispecies dynamics, despite their occurring on very different space-time scales. The triggering event in these evolutionary episodes is replication with genetic variation dependent on infection. Then either positive or negative selection, together with genetic drift takes place within a host, or between host individuals, to shape the genetic composition of the virus. Research into the role of movement proteins and bottlenecks, genetic drift and effective population size would provide further understanding of the speciation and evolution of this virus.

As a final concluding remark, studying virus evolution through the movement of people can provide a door into a closer inspection of human history. By revealing the mechanisms of viral evolution, it might also be possible to shed light on one of the most important topics in disease epidemiology: predicting what new infections, from what reservoir species, and in what locations will emerge and spread in the future. This applies to virus infecting plants, animals and humans. Therefore, epidemiological and virus evolutionary studies must be integrated to obtain a better explanation of viral emergence and establishment of measures for the prediction and control of devastating diseases.

References

Abo El-Nil, M. M., Zettler, F. W., & Hiebert, E. (1977). Purification, serology, and some physical properties of Dasheen Mosaic Virus. *Phytopathology* (67), 1445-1450.

ACIAR. (2012). *Crop Improvement and Management.* Australian Centre for International Agricultural Research. Retrieved from http://www.aciar.gov.au

Acosta-Leal, R., Duffy, S., Xiong, Z., Hammond, R. W., & Elena, S. F. (2011). Advances in plant virus evolution: translating evolutionary insights into better disease management. *Phytopathology, 101*(10), 1136-1148. doi:10.1094/PHYTO-01-11-0017

Adams, M. J., Antoniw, J. F., & Fauquet, C. M. (2005). Molecular criteria for genus and species discrimination within the family Potyviridae. *Archives of Virology, 150*(459-479).

Addison, D.J. and E. Matisoo-Smith.(2010) Rethinking Polynesians origins: a West-Polynesia Triple-I Model. *Archaeology in Oceania* 45:1–12.

Andersson, D. I., & Hughes, D. (1996). Muller's ratchet decreases fitness of a DNA-based microbe. *Proceedings of the National Academy of Sciences, 93*(2), 906-907.

Andrejeva, J., Puurand, U., Merits, A., Rabenstein, F., Jarvekulg, L., & Valkonen, J. P. (1999). Potyvirus helper component-proteinase and coat protein (CP) have coordinated functions in virus-host interactions and the same CP motif affects virus transmission and accumulation. *Journal of General Virology, 80*(5), 1133-1139.

Armelagos, G. J., & Harper, K. N. (2005). Genomics at the origins of agriculture, part one. *Evolutionary Anthropology: Issues, News, and Reviews, 14*(2), 68-77. doi:10.1002/evan.20047

Asif, M., Dhawan, P., & Nath, P. (2000). A simple procedure for the isolation of high quality RNA from ripening banana fruit. *Plant Molecular Biology Reporter, 18*(2), 109-115. doi:10.1007/bf02824018

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

Barr, J. N., & Fearns, R. (2010). How RNA viruses maintain their genome integrity. *Journal of General Virology, 91*(6), 1373-1387. doi:10.1099/vir.0.020818-0

Benbouza, H., Jacquemin, J. L., Baudoin, J. P., & Mergeai, G. (2006). Optimization of a reliable, fast, cheap and sensitive silver staining method to detect ssr markers in polyacrylamide gels. *Biotechnologie Agronomie Societe et Environnement,* 2(10), 77-81.

Bhatt, S., Katzourakis, A., & Pybus, O. G. (2010). Detecting natural selection in RNA virus populations using sequence summary statistics. *Infection, Genetics and Evolution, 10*(3), 421-430. doi:10.1016/j.meegid.2009.06.001

Biebricher, C. K., & Eigen, M. (2006). What is a quasispecies? *Current Topics In Microbiology and Immunology, 299*, 1-31.

Boni, M. F., de Jong, M. D., van Doorn, H. R., & Holmes, E. C. (2010). Guidelines for Identifying Homologous Recombination Events in Influenza A Virus. *PLoS ONE, 5*(5), e10434. doi:10.1371/journal.pone.0010434

Bos, L. (1992). Potyviruses, chaos or order? *Archives of virology. Supplementum, 5*, 31-46.

Bos, L. (1999). *Plant Viruses, Unique and Intriguing Pathogens* (5th ed.). Retrieved from http://books.google.co.nz/books?id=CnzbgZgby60C&pg=PA756&lpg=PA756&dq=Plant+Viruses,+Unique+and+Intriguing+Pathogens+bos&source=bl&ots=FnGqtgZEog&sig=HXKNcY3b2Ov9Xh2Jo5pinmzOEzk&hl=en&ei=7MsfTOiPCcO9ca_JwfkM&sa=X&oi=book_result&ct=result&resnum=4&ved=0CCIQ6AEwAw - v=onepage&q=Plant%20Viruses%2C%20Unique%20and%20Intriguing%20Pathogens%20bos&f=false

Bousalem, M., Douzery, E. J. P., & Fargette, D. (2000). High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of Yam mosaic virus: a contribution to understanding potyvirus evolution. *Journal of General Virology* (81), 243-255.

Bradshaw, J. E. (2010). *Root and Tuber crops*. Retrieved from http://www.springer.com/life+sciences/plant+sciences/book/978-0-387-92764-0?detailsPage=authorsAndEditors

Braidotti, G. (2006). *Fighting back against disease and pests of taro*. Canberra: Australian Centre for Interational Agricultural Research. Retrieved from http://aciar.gov.au/

Carmona, M., Sepulveda, D., Cardenas, C., Nilo, L., & Marshall, S. H. (2012). Denaturing Gradient Gel Electrophoresis (DGGE) as a Powerful Novel Alternative for Differentiation of Epizootic ISA Virus Variants. *PLoS ONE, 7*(5), e37353. doi:10.1371/journal.pone.0037353

Carrington, J. C., Cary, S. M., Parks, T. D., & Dougherty, W. G. (1989). A second proteinase encoded by a plant potyvirus genome. *European Molecular Biology Organization Journal, 8*(2), 365-370.

Carrington, J. C., Freed, D. D., & Sanders, T. C. (1989). Autocatalytic processing of the potyvirus helper component proteinase in Escherichia coli and in vitro. *Journal of General Virology, 63*(10), 4459-4463.

Carrington, J. C., Kasschau, K. D., Mahajan, S. K., & Schaad, M. C. (1996). Cell-to-Cell and Long-Distance Transport of Viruses in Plants. *Plant Cell, 8*(10), 1669-1681. doi:10.1105/tpc.8.10.16698/10/1669 [pii]

Cervera, M. T., Riechmann, J. L., Martin, M. T., & Garcia, J. A. (1993). 3'-terminal sequence of the plum pox virus PS and o6 isolates: Evidence for RNA recombination within the potyvirus group. *Journal of General Virology, 74*, 329-334.

Chao, L. (1988). Evolution of sex in RNA viruses. *Journal of Theorectical Biology, 133*(1), 99-112.

Chare, E. R., Gould, E. A., & Holmes, E. C. (2003). Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *Journal of General Virology, 84*(10), 2691-2703. doi:10.1099/vir.0.19277-0

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

Chen, J., & Adams, M. J. (2001b). A universal PCR primer to detect members of the and its use to examine the taxonomic status of several members of the family. *Archives of Virology, 146*(4), 757-766. doi:10.1007/s007050170144

Christie, R. G., & Edwardson, J. R. (1974). Light and electron microscopy of plant virus inclusions. *Florida Agriculture Experiment Station Monograph* (Series No. 9), 150. Retrieved from http://www.apsnet.org/phyto/PDFS/1983/Phyto73n02_290.pdf

Christie, R. G., & Edwardson, J. R. (1986). Light microscopic techniques for detection of plant virus inclusions. *Plant Disease*(70), 273-279.

CIP. (2000a). *Annual Report, International Potato Centre*. Lima, Peru. Retrieved from http://www.cipotato.org/publications/annual_reports/2002/

Codoñer, F. M., & Elena, S. F. (2008). The promiscuous evolutionary history of the family Bromoviridae. *Journal of General Virology, 89*(7), 1739-1747. doi:10.1099/vir.0.2008/000166-0

Cong, M. (2007). *Mosaic Virus in Its Host, Taro?* AUT unpublished, Undergraduate project.

Coulibaly, S., Pasquet, R. S., Papa, R., & Gepts, P. (2002). AFLP analysis of the phenetic organization and genetic diversity of Vigna unguiculatal. Walp. reveals extensive gene flow between wild and domesticated types. *TAG Theoretical and Applied Genetics, 104*(2), 358-366. doi:10.1007/s001220100740

de Assis Filho, F. M., Paguio, O. R., Sherwood, J. L., & Deom, C. M. (2002). Symptom induction by Cowpea chlorotic mottle virus on Vigna unguiculata is determined by amino acid residue 151 in the coat protein. *Journal of General Virology, 83*(4), 879-883.

Delaunay, A., Rolland, M., & Jacquot, E. (2009). Analysis of Population Structures of Viral Isolates using Single Strand Conformation Polymorphism (SSCP) Method. In R. Burns (Ed.), J. M. Walker, (Vol. 508, pp. 235-247): Humana Press. Retrieved from http://dx.doi.org/10.1007/978-1-59745-062-1_18. doi:10.1007/978-1-59745-062-1_18

Desbiez, C., & Lecoq, H. (2004). 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

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. *European Molecular Biology Organization Journal, 13*(6), 1482-1491.

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. (1989). RNA virus evolution and the control of viral disease. *Progress in Drug Research, 33*, 93-133.

Domingo, E. (1998). Quasispecies and the implications for virus persistence and escape. *Clinical Diagnosis Virology, 10*(2-3), 97-101.

Domingo, E., & Holland, J. J. (1994). Mutation rates and rapid evolution of RNA viruses. *In Evolutionary Biology of Viruses* (Morse, S.S. eds),161-184. Retrieved from http://arjournals.annualreviews.org/doi/pdf/10.1146/annurev.micro.51.1.151

Domingo, E., & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annual Review in Microbiology, 51*, 151-178. doi:10.1146/annurev.micro.51.1.151

Domingo, E., Martín, V., Perales, C., Grande-Pérez, A., García-Arriaza, J., & Arias, A. (2006). Viruses as Quasispecies: Biological Implications Quasispecies: Concept and Implications for Virology. In E. Domingo (Ed.), (Vol. 299, pp. 51-82): Springer Berlin Heidelberg. doi:10.1007/3-540-26397-7_3

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2012). Geneious v5.5.4. Retrieved from http://www.geneious.com/

Eigen, M., McCaskill, J., & Schuster, P. (1988). Molecular quasi-species. *The Journal of Physical Chemistry, 92*(24), 6881-6891. doi:10.1021/j100335a010

Eisen, J. A., & MacCallum, C. J. (2009). Genomics of emerging infectious disease: A PLoS collection. *PLoS biology, 7*(10), e1000224. doi:10.1371/journal.pbio.1000224

Ekanem, A., & Osuji, J. (2006). Mitotic index studies on edible cocoyam. *African Journal of Biotechnology, 5*, 846-849.

Elena, S., Bedhomme, S., Carrasco, P., Cuevas, J. M., de la Iglesia, F., Lafforgue, G., ... Zwart, M. P. (2011). The evolutionary genetics of emerging plant RNA viruses. *Molecular Plant-MicrobeIinteractions, 24*(3), 287-293. doi:10.1094/MPMI-09-10-0214

FAO. (2012). *Food and Agriculture Organization of United Nations*. Retrieved 12/12, 2011, from http://www.fao.org

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

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., & Ball, L. A. (2005). Virus Taxonomy: *VIIIth Report of the International Committee on Taxonomy of Viruses*, 8 Ed. Retrieved from http://www.virologyj.com/content/2/1/64

Fermin-Munoz, G. A., Meng, B., Ko, K., , Mazumdar-Leighton, S., & Gubba, J. E. (2001). *Biotechnology: A new era for plant pathology and plant protection*. Retrieved from http://www.aspnet.org

Flynn, M. (2010). *The evolution of populations*. Retrieved from http://www.google.co.nz/imgres?imgurl=http://4.bp.blogspot.com/_

Fraile, A., Alonso-Prados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M., & Garcia-Arenal, F. (1997). Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *Journal of Virology, 71*(2), 934-940.

Fuji, S., Inoue, M., Yamamoto, H., Furuya, H., & Naito, H. (2003). Nucleotide sequences of the coat protein gene of potyviruses infecting Ornithogalum thyrsoides. *Archives of Virology, 148,* 618-621.

Fuji, S., Yamamoto, H., Furuya, H., & Naito, H. (2003). Characterization of a new potyvirus isolated from Chinese artichoke in Japan. *Archives of Virology, 148,* 2249-2255.

Garcia-Arenal, F., Fraile, A., & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology, 39,* 157-186. doi:10.1146/annurev.phyto.39.1.157

García-Robledo, C., Quintero-Marín, P., & Mora-Kepfer, F. (2005). Geographic Variation and Succession of Arthropod Communities in Inflorescences and Infructescences of Xanthosoma (Araceae)1. *Biotropica, 37*(4), 650-656. doi:10.1111/j.1744-7429.2005.00082.x

Gasteiger, E., Gattiker, A., Hoogland, C., Ivany, i. I., Appel, R. D., & Bairoch, A. (2003). Potyvirus Genome. *Swiss Institute of Bioinformatics*. Retreived from http://viralzone.expasy.org/all_by_species/50.html

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

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

Gibbs, A., Trueman, J., & Gibbs, M. (2008). 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

Gnutova, R. V., & Tolkach, V. F. (1998). Taxanomy of the family Potyviridae. *Archives of Phythopathology, 31,* 543-548.

Goldbach, R. (1986). Molecular evolution of plant RNA viruses. *Annual Review of Phytopathology, 24,* 289-310.

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science, 303*(5656), 327-332. doi:10.1126/science.1090727

Ha, C. (2007). Detection and identification of potyviruses and geminiviruses in Vietnam. PhD by Publication, Queensland University of Technology.

Ha, C., Coombs, S., Revill, P. A., Harding, R. M., Vu, M., & Dale, J. L. (2008). Design and application of two novel degenerate primer pairs for the detection and complete genomic characterization of potyviruses. *Archives of Virology, 153*(1), 25-36. doi:10.1007/s00705-007-1053-7

Hahn, B. H., Shaw, G. M., De , K. M., Cock, & Sharp, P. M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science, 287*(5453), 607-614. doi:10.1126/science.287.5453.607

Hall, J. S., French, R., Morris, T. J., & Stenger, D. C. (2001). Structure and temporal dynamics of populations within Wheat streak mosaic virus isolates. *Journal of Virology, 75*(21), 10231-10243. doi:10.1128/jvi.75.21.10231-10243.2001

Heimer, R., Barbour, R., Shaboltas, A. V., Hoffman, I. F., & Kozlov, A. P. (2008). Spatial distribution of HIV prevalence and incidence among injection drugs users in St Petersburg: implications for HIV transmission. *Acquired Immunodeficeincy Syndrome Journal, 22*(1), 123-130. doi:10.1097/QAD.0b013e3282f244ef

Ho, S. Y., Phillips, M. J., Cooper, A., & Drummond, A. J. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution, 22*(7), 1561-1568. doi:10.1093/molbev/msi145

Holmes, E. C. (2003). Molecular clocks and the puzzle of RNA virus origins. *Journal of Virology, 77*(7), 3893-3897. doi:10.1128/jvi.77.7.3893-3897.2003

Holmes, E. C. (2009). The evolutionary genetics of emerging viruses. *Annual Review of Ecology, Evolution, and Systematics, 40*(1), 353-372. doi:10.1146/annurev.ecolsys.110308.120248

Holmes, E. C., & Grenfell, B. T. (2009). Discovering the Phylodynamics of RNA Viruses. *PLoS Computational Biology, 5*(10), e1000505. doi:10.1371/journal.pcbi.1000505

Houseley, J., & Tollervey, D. (2009). The many pathways of RNA degradation. *Cell, 136*(4), 763-776. doi:S0092-8674(09)00067-1 [pii]10.1016/j.cell.2009.01.019

Huang, T. S., Wei, T., Laliberte, J. F., & Wang, A. (2010). A host RNA helicase-like protein, AtRH8, interacts with the potyviral genome-linked protein, VPg, associates with the virus accumulation complex, and is essential for infection. *Plant Physiology, 152*(1), 255-266. doi:10.1104/pp.109.147983

Hull. R. (2002). *Mathews Plant Virology. Fourth Edition.* Retrieved from http://www.sciencedirect.com/science/book/9780123611604

ICTV. (2010). *International Committe on taxanomy of viruses*. Retrieved 02/06/10, from http://www.csiropedia.csiro.au/display/CSIROpedia/Potyvirus+taxonomy

Jenkins, G. M., Rambaut, A., Pybus, O. G., & Holmes, E. C. (2002). Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. *Journal of Molecular Evolution, 54*(2), 156-165. doi:10.1007/s00239-001-0064-3

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

Joannon, B., Lavigne, C., Lecoq, H., & Desbiez, C. (2010). Barriers to gene flow between emerging populations of Watermelon mosaic virus in Southeastern France. *Phytopathology, 100*(12), 1373-1379. doi:10.1094/PHYTO-04-10-0118

Keulen, W., van Wijk, A., Schuurman, R., Berkhout, B., & Boucher, C. A. (1999). Increased polymerase fidelity of lamivudine-resistant HIV-1 variants does not limit their evolutionary potential. *AIDS, 13*(11), 1343-1349.

Lauring, A. S., & Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *Plos Pathogen, 6*(7), e1001005. doi:10.1371/journal.ppat.1001005

Lefeuvre, P., Harkins, G. W., Lett, J. M., Briddon, R. W., Chase, M. W., Moury, B., & Martin, D. P. (2011). Evolutionary time-scale of the begomoviruses: evidence from integrated sequences in the Nicotiana genome. *PLoS ONE, 6*(5), e19193. doi:10.1371/journal.pone.0019193

Levin, B. R., & Bull, J. J. (1994). Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends in Microbiology, 2*(3), 76-81.

Li, F., & Ding, S.-W. (2006). Virus Counterdefense: Diverse strategies for svading the RNA-silencing immunity. *Annual Review of Microbiology, 60*(1), 503-531. doi:doi:10.1146/annurev.micro.60.080805.142205

Li, M., Lu, Q., Wu, X., Man, N., Chen, X., Jin, W., & Lu, X. (2009). Analysis of RNA-dependent RNA polymerase sequence of infectious Flacherie Virus isolated in China and its expression in BmN cells. *Agricultural Sciences in China, 8*(7), 872-879. doi:10.1016/s1671-2927(08)60290-4

Li, R. H., Zettler, F. W., Hiebert, E., Purcifull, D. E., & Morales, F. J. (1999). Differentiation of Dasheen Mosaic Potyvirus isolates based on variability in the apparent size of the capsid protein in western blots. *Journal of Phytopathology, 147*(6), 359-364. doi:10.1046/j.1439-0434.1999.00095.x

Liao, Z., Chen, M., Guo, L., Gong, Y., Tang, F., Sun, X., & Tang, K. (2004). Rapid isolation of high quality total RNA from Taxus and Ginkgo. *Preparative Biochemistry and Biotechnology, 34*(3), 209-214. doi:10.1081/pb-200026790

Liu, Y., Wimmer, E., & Paul, A. V. (2009). Cis-acting RNA elements in human and animal plus-strand RNA viruses. *Biochimica et Biophysica Acta, 1789*(9-10), 495-517. doi:10.1016/j.bbagrm.2009.09.007

Longdon, B., Hadfield, J. D., Webster, C. L., Obbard, D. J., & Jiggins, F. M. (2011). Host phylogeny determines viral persistence and replication in novel hosts. *Plos Pathogen, 7*(9), e1002260. doi:10.1371/journal.ppat.1002260

Mackenzie, A. M., Nolan, M., Wei, K. J., Clements, M. A., Gowanlock, D., Wallace, B. J., & Gibbs, A. J. (1998). Ceratobium mosaic potyvirus: another virus from orchids. *Archives of Virology, 143*(5), 903-914. doi:10.1007/s007050050341

Manrubia, S. C., Domingo, E., & Lazaro, E. (2010). Pathways to extinction: beyond the error threshold. *Philosophical Transactions of the Royal Society Biological Sciences, 365*(1548), 1943-1952. doi:10.1098/rstb.2010.0076

Marie-Jeanne, V., Ioos, R., Peyre, J., Alliot, B., & Signoret, P. (2000). Differentiation of Poaceae Potyviruses by reverse transcription-polymerase chain reaction and restriction analysis. *Journal of Phytopathology, 148*(3), 141-151. doi:10.1046/j.1439-0434.2000.00473.x

Mas, A., Lopez-Galindez, C., Cacho, I., Gomez, J., & Martinez, M. A. (2010). Unfinished stories on viral quasispecies and Darwinian views of evolution. *Journal of Molecular Biology, 397*(4), 865-877. doi:10.1016/j.jmb.2010.02.005

Matisoo-Smith, E., Roberts, R. M., Irwin, G. J., Allen, J. S., Penny, D., & Lambert, D. M. (1998). Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences, USA, 95*(25), 15145-15150.

Matisoo-Smith, E., & Robins, J. H. (2004). Origins and dispersals of Pacific peoples: Evidence from mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences, USA, 101*(24), 9167-9172. doi:10.1073/pnas.0403120101

Matthews, R. E. F. (1991). *Plant Virology*. 3rd edition. Academic Press, San Diego, CA.

McCoy, M. D., & Graves, M. W. (2010). The role of agricultural innovation on Pacific Islands: a case study from Hawai'i Island. *World Archaeology, 42*(1), 90-107. doi:10.1080/00438240903430340

McGeoch, D. J., Dolan, A., & Ralph, A. C. (2000). Toward a comprehensive phylogeny for mammalian and avian herpesviruses. *Journal of Virology, 74*(22), 10401-10406.

Merits, A., Rajamaki, M.-L., Lindholm, Runeberg-Roos, P., Kekarainen, T., Puustinen, P., ... Saarma, M. (2002). Proteolytic processing of potyviral proteins and polyprotein processing intermediates in insect and plant cells. *Journal of General Virology, 83*(5), 1211-1221.

Miralles, R., Gerrish, P. J., Moya, A., & Elena, S. F. (1999). Clonal interference and the evolution of RNA viruses. *Science, 285*(5434), 1745-1747. doi:10.1126/science.285.5434.1745

Monsion, B., Froissart, R., Michalakis, Y., & Blanc, S. (2008). Large bottleneck size in Cauliflower mosaic virus populations during host plant colonization. *Plos Pathogen, 4*(10), e1000174. doi:10.1371/journal.ppat.1000174

Mosahebi, G., Koohi-Habibi, M., & Okhovvat, S. M. (2005). Study on potato virus M (PVM) occurrence in potato fields in Iran. *Communications in Agriculture and Applied Biological Sciences, 70*(3), 441-443.

Moury, B., Fabre, F., & Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proceedings of the National Academy of Sciences, USA, 104*(45), 17891-17896. doi:10.1073/pnas.0702739104

Moya, A., Holmes, E. C., & Gonzalez-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. *Nature Reviews: Microbiology, 2*(4), 279-288. doi:10.1038/nrmicro863

Nagy, P. D. (2008). Yeast as a model host to explore plant virus-host interactions. *Annual Review of Phytopathology, 46*, 217-242. doi:10.1146/annurev.phyto.121407.093958

Nelson, S. (2008). Dasheen Mosaic of edible and ornamental aroids. *Plant Disease* (44), 1-9.

Novella, I. S., Domingo, E., & Holland, J. J. (1995). Rapid viral quasispecies evolution: implications for vaccine and drug strategies. *Molecular Medicine Today, 1*(5), 248-253.

Nowak, M., Anderson, R., McLean, A., Wolfs, T., Goudsmit, J., & May, R. (1991). Antigenic diversity thresholds and the development of AIDS. *Science, 254*(5034), 963-969. doi:10.1126/science.1683006

O' Hair, S. (1990). Tropical Root and Tuber crops *Timber Press.* Symposium conducted at the meeting of the Advances in new crops, Portland. Retrieved from http://www.actahort.org

Ohshima, K., Yamaguchi, Y., Hirota, R., Hamamoto, T., Tomimura, K., Tan, Z., ... Gibbs, A. (2002). Molecular evolution of Turnip mosaic virus: evidence of host adaptation, genetic recombination and geographical spread. *Journal of General Virology, 83*, 1511-1521.

Okonkwo, P. O., Ogbuokiri, J. E., Ofoegbu, E., & Klotz, U. (1993). Protein binding and ivermectin estimations in patients with onchocerciasis. *Clinical Pharmacology and Therapeutics, 53*(4), 426-430.

Olarte Castillo, X. A., Fermin, G., Tabima, J., Rojas, Y., Tennant, P. F., Fuchs, M., Restrepo, S. (2011). Phylogeography and molecular epidemiology of Papaya ringspot virus. *Virus Research, 159*(2), 132-140. doi:10.1016/j.virusres.2011.04.011

Onwueme, I. C. (1999). *Taro cultivation in the Asia and Pacific region.* Papua New Guinea. Retrieved from http://www.fao.org/

Oruetxebarria, I., Guo, D., Merits, A., Makinen, K., Saarma, M., & Valkonen, J. P. (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.

Pagan, I., & Holmes, E. C. (2010). Long-term evolution of the Luteoviridae: time scale and mode of virus speciation. *Journal of Virology, 84*(12), 6177-6187. doi:10.1128/JVI.02160-09

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

Patrick, V. K. (2010). Peopling of the Pacific: A Holistic Anthropological Perspective. *Annual Review of Anthropology, 39*(1), 131-148. doi:doi:10.1146/annurev.anthro.012809.104936

Pavesi, A. (2001). Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. *Journal of Molecular Evolution, 53*(2), 104-113. doi:10.1007/s002390010198

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

Pierson, M. J., Martinez-Arias, R., Holland, B. R., Gemmell, N. J., Hurles, M. E., & Penny, D. (2006). Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. *Molecular Biology Evolution, 23*(10), 1966-1975. doi:msl063 [pii]10.1093/molbev/msl063

Poranen, M. M., Koivunen, M. R., & Bamford, D. H. (2008). Nontemplated terminal nucleotidyltransferase activity of double-stranded RNA bacteriophage phi6 RNA-dependent RNA polymerase. *Journal of Virology, 82*(18), 9254-9264. doi:JVI.01044-08 [pii]10.1128/JVI.01044-08

Quero-Garcia, J., Ivancic, A., & Lebot, V. (Eds.). (2010). *Root and Tuber Crops* (Vol. 7). New York: Springer.

Racaniello, V. R. (2001). Picornaviridae: The viruses and their replication in Fields Virology (Knipe, D. M., Howley, P. M., Griffin, D. E., Lamb, R. A., Martin, M. A., Roizman, B., and Straus, S. E., eds) 4th Ed. Lippincott-Raven Publishers, Philadelphia, 685-722. Retrieved from http://books.google.co.nz/books?id=5O0somr0w18C&pg=PA795&dq=Picornaviridae:+the+viruses+and+their+replication.+In+Fields+Virology&hl=en&sa=X&ei=KBvMT6-pJsXsmAXLWBDw&ved=0CDcQ6AEwAA#v=onepage&q=Picornaviridae%3A%20the%20viruses%20and%20their%20replication.%20In%20Fields%20Virology&f=false

Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., & Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza A virus [10.1038/nature06945]. *Nature, 453*(7195), 615-619. doi:http://www.nature.com/nature/journal/v453/n7195/suppinfo/nature06945_S1.html

Real, L. A., Henderson, J. C., Biek, R., Snaman, J., Jack, T. L., Childs, J. E., ... Nadin-Davis, S. (2005). Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. *Proceedings of the National Academy of Sciences,USA, 102*(34), 12107-12111. doi:10.1073/pnas.0500057102

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), 327-331. doi:http://dx.doi.org/10.1071/AP05032

Rodoni, B. (1995). *Alomae disease of taro*. Sydney: Australian Centre for International Agricultural Research. Retrieved from http://www.fao.org/docrep/005/ac450e/ac450e0a.htm

Rodoni, B. C., & Moran, J. R. (1998). The detection of dasheen mosaic virus using the enzyme linked immunosorbent assay (ELISA). *Acta Horticulturae* 234, 281-285.

Rodrigo, G., Carrera, J., Jaramillo, A., & Elena, S. F. (2011). Optimal viral strategies for bypassing RNA silencing. *Journal of the Royal Society, Interface of the Royal Society, 8*(55), 257-268. doi:10.1098/rsif.2010.0264

Roossinck, M. J. (2003). Plant RNA virus evolution. *Current Opinion in Microbiology, 6*(4), 406-409. doi:S1369527403000870 [pii]

Roossinck, M. J. (2011). The good viruses: viral mutualistic symbioses. *Nature Review Microbiology, 9*(2), 99-108. doi:nrmicro2491 [pii]10.1038/nrmicro2491

Rouzine, I. M., Rodrigo, A., & Coffin, J. M. (2001). Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiology and Molecular Biology Reviews, 65*(1), 151-185. doi:10.1128/MMBR.65.1.151-185.2001

Saccheri, I., & Hanski, I. (2006). Natural selection and population dynamics. *Trends in Ecology & Evolution, 21*(6), 341-347. doi:10.1016/j.tree.2006.03.018

Sacristan, S., Malpica, J. M., Fraile, A., & Garcia-Arenal, F. (2003). Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco plants. *Journal of Virology, 77*(18), 9906-9911.

Sanjun, R., Cuevas, J., Furi, V., Holmes, E. C., & Moya, A. (2007). Selection for robustness in mutagenized RNA viruses. *PLoS Genetics, 3*, 6(e93). doi:10.1371/journal.pgen.0030093

Sardanyes, J., & Elena, S. F. (2011). Quasispecies spatial models for RNA viruses with different replication modes and infection strategies. *PLoS ONE, 6*(9), e24884. doi:10.1371/journal.pone.0024884

Semler, B. L., & Wimmer, E. (2002). Molecular Biology of Picornaviruses. Retrieved from http://www.im.microbios.org/articles0203/2003/march/15 BR Bosch-Semler.pdf

Sharp, P. M., Bailes, E., Chaudhuri, R. R., Rodenburg, C. M., Santiago, M. O., & Hahn, B. H. (2001). The origins of acquired immune deficiency syndrome viruses: where and when? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 356*(1410), 867-876. doi:10.1098/rstb.2001.0863

Shi, B.-J., Miller, J., Symons, R. H., & Palukaitis, P. (2003). The 2b protein of Cucumoviruses has a role in promoting the cell-to-cell movement of pseudorecombinant viruses. *Molecular Plant-Microbe Interactions, 16*(3), 261-267. doi:10.1094/mpmi.2003.16.3.261

Shikano, T., Chiyokubo, T., & Taniguchi, N. (2001). Temporal changes in allele frequency, genetic variation and inbreeding depression in small populations of the guppy, Poecilia reticulata. *Heredity, 86*(2), 153-160.

Shriner, D., Shankarappa, R., Jensen, M. A., Nickle, D. C., Mittler, J. E., Margolick, J. B., & Mullins, J. I. (2004). Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection. *Genetics, 166*(3), 1155-1164. doi:10.1534/genetics.166.3.1155

Shukla, D. D., Ward, C. W., & Bunt, A. A. (1994). *The Potyviridae.* . Retrieved 12/10/10, from http://www.dpvweb.net/dpv/showdpv.php?dpvno=366

Silverman, R. H. (2007). Viral encounters with 2',5'-oligoadenylate synthetase and RNase L during the interferon antiviral response. *Journal of Virology, 81*(23), 12720-12729. doi:JVI.01471-07 [pii]10.1128/JVI.01471-07

Simmonds, P. (2004). Genetic diversity and evolution of hepatitis C virus – 15 years on. *Journal of General Virology, 85*(11), 3173-3188. doi:10.1099/vir.0.80401-0

Simmons, H. E., Holmes, E. C., & Stephenson, A. G. (2011). Rapid turnover of intra-host genetic diversity in Zucchini yellow mosaic virus. *Virus Research, 155*(2), 389-396. doi:10.1016/j.virusres.2010.11.007

Simon-Buela, L., Osaba, L., Garcia, J. A., & Lopez-Moya, J. J. (2000). Preservation of 5'-end integrity of a potyvirus genomic RNA is not dependent on template specificity. *Virology, 269*(2), 377-382. doi:10.1006/viro.2000.0229

Spetz, C., Taboada, A. M., Darwich, S., Ramsell, J., Salazar, L. F., & Valkonen, J. P. (2003). Molecular resolution of a complex of potyviruses infecting solanaceous crops at the centre of origin in Peru. *Journal of General Virology, 84*(9), 2565-2578.

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology, 57*(5), 758-771. doi:10.1080/10635150802429642

Stavrinides, J., & Guttman, D. S. (2004). Mosaic evolution of the severe acute respiratory syndrome coronavirus. *Journal of Virology, 78*(1), 76-82. doi:10.1128/jvi.78.1.76-82.2004

Strauss, J. H., & Strauss, E. G. (2001). Virus evolution: how does an enveloped virus make a regular structure? *Cell, 105*(1), 5-8. doi:S0092-8674(01)00291-4 [pii]

Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M., & Bujarski, J. J. (2011). RNA-RNA recombination in plant virus replication and evolution. *Annual Review of Phytopathology, 49*(1), 415-443. doi:doi:10.1146/annurev-phyto-072910-095351

Tayon, R., Jr., Kim, M. J., & Kao, C. C. (2001). Completion of RNA synthesis by viral RNA replicases. *Nucleic Acids Research, 29*(17), 3576-3582.

Tugume, A. K., Cuellar, W. J., Mukasa, S. B., & Valkonen, J. P. T. (2010). Molecular genetic analysis of virus isolates from wild and cultivated plants demonstrates that East Africa is a hotspot for the evolution and diversification of Sweet potato feathery mottle virus. *Molecular Ecology, 19*(15), 3139-3156. doi:10.1111/j.1365-294X.2010.04682.x

Twiddy, S. S., Farrar, J. J., Vinh Chau, N., Wills, B., Gould, E. A., Gritsun, T., ... Holmes, E. C. (2002). Phylogenetic Relationships and Differential Selection Pressures among Genotypes of dengue-2 Virus. *Virology, 298*(1), 63-72. doi:10.1006/viro.2002.1447

Ullah, Z., Chai, B., Hammar, S., Raccah, B., Gal-On, A., & Grumet, R. (2003). Effect of substitution of the amino termini of coat proteins of distinct potyvirus species on viral infectivity and host specificity. *Physiological and molecular plant pathology, 63*(3), 129-139.

Urcuqui-Inchima, S., Haenni, A. L., & Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Research, 74*(1-2), 157-175.

van den Born, E., Omelchenko, M. V., Bekkelund, A., Leihne, V., Koonin, E. V., Dolja, V. V., & Falnes, P. l. ò. (2008). Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acids Research, 36*(17), 5451-5461. doi:10.1093/nar/gkn519

van der Vlugt, R. (1992). Is pepper mottle virus a strain of potato virus Y? *Archives of Virology. Supplementum, 5*, 327-335.

van Regenmortel, M. H. V. (2000). Virus taxonomy: The classification and nomenclature of viruses. The Seventh Report of the International Committee on the Taxonomy of Viruses. .Retrieved from http://www.cdc.gov/ncidod/eId/vol10no1/pdfs/03-0279.pdf

Weber, F., Fukino, K., Villalona-Calero, M., & Eng, C. (2005). Limitations of single-strand conformation polymorphism analysis as a high-throughput method for the detection of EGFR mutations in the clinical setting. *Journal of Clinical Oncology, 23*(24), 5847-5848. doi:10.1200/jco.2005.01.5222

Wei, T., Huang, T.-S., McNeil, J., Lalibert, J.F. o., Hong, J., Nelson, R. S., & Wang, A. (2010). Sequential recruitment of the endoplasmic reticulum and chloroplasts for plant Potyvirus peplication. *Journal of Virology, 84*(2), 799-809. doi:10.1128/jvi.01824-09

Wei, T., Zhang, C., Hong, J., Xiong, R., Kasschau, K. D., Zhou, X., ... Wang, A. (2010). Formation of complexes at plasmodesmata for Potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PLoS Pathogen, 6*(6), e1000962. doi:10.1371/journal.ppat.1000962

White, K. A., Enjuanes, L., & Berkhout, B. (2011). RNA virus replication, transcription and recombination. *RNA Biology, 8*(2), 182-183.

Wiesenberg, I., Missbach, M., Kahlen, J.-P., Schreder, M., & Carlberg, C. (1995). Transcriptional activation of the nuclear receptor RZR by the pineal gland hormone melatonin and identification of CGP 52608 as a synthetic ligand. *Nucleic Acids Research, 23*(3), 327-333. doi:10.1093/nar/23.3.327

Wisler, G. C., Zettler, F. W., & Mu, L. (1987). Virus infections of Vanilla and other orchids in French Polynesia. *Plant Disease, 71*(12), 1125-1129. doi:10.1094/pd-71-1125

Wylie, S. J., & Jones, R. A. (2009). Role of recombination in the evolution of host specialization within bean yellow mosaic virus. *Phytopathology, 99*(5), 512-518. doi:10.1094/PHYTO-99-5-0512

Yakoubi, S., Desbiez, C., Fakhfakh, H., Wipf-Scheibel, C., Marrakchi, M., & Lecoq, H. (2008). Biological characterization and complete nucleotide sequence of a Tunisian isolate of Moroccan watermelon mosaic virus. *Archives of Virology, 153*(1), 117-125. doi:10.1007/s00705-007-1074-2

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution, 24*(8), 1586-1591. doi:10.1093/molbev/msm088

Yuan, A. (2008). *Long and short term evolution study of DsMV*. AUT unpublished, Undergraduate project.

Zanotto, P. M., Gibbs, M. J., Gould, E. A., & Holmes, E. C. (1996). A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *Journal of Virology, 70*(9), 6083-6096.

Zeldovich, K. B., Chen, P., & Shakhnovich, E. I. (2007). Protein stability imposes limits on organism complexity and speed of molecular evolution [Research Support, N.I.H., Extramural]. *Proceedings of the National Academy of Sciences, USA*, 104(41), 16152-16157. doi:10.1073/pnas.0705366104

Zettler, F. W., & Hartman, R. D. (1987). Dasheen mosaic virus as a pathogen of cultivated aroids and control of the virus by tissue culture. *Plant Disease* (71), 958-963.

Zettler, F. W., Jackson, G. V. H., & Frison, E. A. F., & R. Rome/International Board for Plant Genetic Resources. (1989). *FAO/IBPGR Technical Guidelines for the Safe Movement of Edible Aroid Germplasm*. Retrieved from http://www.bioversityinternational.org/fileadmin/bioversity/publications/pdfs/400_Edible_aroid.pdf?cache=1336372995

Zheng, L., Rodoni, B. C., Gibbs, M. J., & Gibbs, A. J. (2010). A novel pair of universal primers for the detection of potyviruses. *Plant Pathology, 59*(2), 211-220. doi:10.1111/j.1365-3059.2009.02201.x

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. *PLoS ONE, 3*(2), e1586. doi:10.1371/journal.pone.0001586

Appendix 1.0

Genbank accession	*Virus*	Sequence		Comment
NC014790	*Passion woodiness virus*	AAUGGGGUGUGGGUAAUGAUGGAUGGCGAUGAGCAGG UUGAAUACCCUCUUAAGCCAAUGGUUGAAAAUGCAAA ACCCACUCUCCGACAGAUUAUGCACCAUUUCUCGGACG CAGCUGAAGCAUAUAUCGAGAUGAGAUGUGCAAGUGG GCCAUACAUGCCUAGGUAUGGACUCUUGCGCAAUUUG AGAGACAAGAAUUUGGCACGAUAUGCUUUUGAUUUUU AUGAAGUCAAUGCUAAGACUUCCGAUCGUGCUAGGGA AGCAGUUGCGCAAAUGAAAGCUGCAGC	Outgroup	NCBI sequence Sequence ID: 10024114 (in phylogenetic tree)
NC003224.1	*Zucchini yellow mosaic virus*	AAUGGAGUGUGGUUCAUGAUGGAUGGAGAUGAGCAGG UCGAGUAUCCUUUGAAACCAAUAGUCGAAAAUGCAAA GCCAACGCUGCGACAAAUAAUGCAUCACUUCUCAGAU GCAGCGGAGGCAUACAUAGAAAUGAGAAAUGCAGAGG CACCAUACAUGCCGAGGUAUGGUUUGCUUCGAAAUCU ACGGGAUAGGAGUUUGGCUCGAUACGCUUUCGACUUC UACGAAGUCAACUCUAAAACUCCUGAAAGAGCCCGCG AAGCUGUUGCGCAGAUGAAAGCAGCAGC-		NCBI sequence Sequence ID: 92672 (in phylogenetic tree)
AF014811	*Zucchini yellow mosaic virus*	AAUGGAGUGUGGGUCAUGAUGGAUGGAAAUGAACAAG UUGAAUAUCCUUUAAAACCAAUAGUUGAGAACGCAAA ACCAACGCUGCGACAAAUAAUGCAUCAUUUCUCAGAU GCAGCGGAGGCUUAUAUAGAAAUGAGAAAUGCAGAGG CACCAUACAUGCCGAGGUAUGGUUUGCUUCGAAAUCU ACGGGACAGGAGUUUAGCCCGUUACGCUUUUGACUUC UAUGAAGUUAAUUCAAAGACUCCUGAUAGAGCCCGCG AAGCUGUUGCGCAGAUGAAGGCAGCAGC		NCBI sequence

Genbank accession	Virus	Sequence			Comment
DQ112219	*Passiflora foetida virus Y*	AAUGGAGUUUGGGUCAUGAUGGAUGGAGAUGAACA GAUUGAAUAUCCUUUAAAACCAAUGGUAGAGAAUG CUAAACCAACACUCAGGCAAAUUAUGCAUCACUUUU CAGACGCAGCUGAAGCGUACAUUGAGAUGAGAAAU AGUGAAGCACCAUACAUGCCUAGGUAUGGAUUGCU GAGGAAUUUAAGAGAUCGGAGUCUUGCCAGAUAUG CUUUUGAUUUCUACGAAGUUAACUUUAAAACUUCU GAAAGGGCACGGGAAGCUGUGGCUCAAAUGAAGGC AGCCGC	Outgroup		NCBI sequence
DQ925451	*Zucchini yellow mosaic virus*	AAUGGAGUGUGGUUUAUGAUGGAUGGAGAUGAGCA AGUUGAAUAUCCUUUAAAGCCGAUAGUUGAAAAUG CGAAGCCAACGCUGCGACAAAUAAUGCAUCAUUUUU CCGAUGCAGCAGAGGCUUAUAUUGAAAUGAGAAAU GCAGAAGCACCAUACAUGCCGAGGUAUGGUUUGCU UCGAAACCUGAGGGACAGGAGUUUAGCUCGAUAUG CUUUUGACUUCUAUGAAGUUAAUUCAAAGACUCCU GAAAGAGCUCGCGAGGCCGUGGCUCAAAUGAAAGC AGCAGC			NCBI sequence
AF048981	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCGTTGAAGCCAATTGTTGAAAATGCAA AACCAACCTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAGAGGCATACATTGAACTGAGAAACGCAGA GAAACCGTATATGCCTAGATACGGTCTTATTCGCAAT TTACGTGATGCAAGTCTCGCCCGGTATGCTTTTGACTT TTATGAGGTCAATTCTAAAACACCGGTGCGAGCAAGA GAAGCAGTTGCGCAAATGAAGGCGGCTGC	Location: China	Host: Caladium	NCBI sequence

Genbank accession	Virus	Sequence	Location	Host	Comment
AF169832	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCGTTGAAACCAATTGTGGAGAATGCAA AACCTACCTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAGAGGCATATATTGAACTGAGGAATGCGGA GAAGCCGTACATGCCTAGGTATGGCCTCATTCGCAAC TTACGTGATGCGAGTCTCGCCCGGTATGCATTTGACT TCTATGAGGTCAATTCTAAGACACCGGTTCGAGCAAG GGAAGCAGTCGCGCAGATGAAGGCCGCTGC	Taiwan	Aroid plant	NCBI sequence
AJ298033	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGAACAA ATTGAATACCCGTTAAAGCCTATTGTGGAGAATGCAA AACCCACCTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAGAGGCATACATTGAACTGAGGAATGCGGA GAAACCGTACATGCCTAGGTACGGTCTTATTCGCAAC TTACGTGATGCAAGTCTCGCTCGGTACGCTTTTGACTT CTACGAGGTCAATTCTAAAACACCGGTGCGAGCAAG AGAGGCAGTTGCGCAGATGAAGGCCGCTGC	China	*Zantedeschia aethiopica*	NCBI sequence
AJ298034	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGACCAA ATTGAATACCCGTTGAAGCCAATTGTAGAAAATGCAA AACCCACTTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAGAGGCTTATATAGAATTGAGAAATGCGGA GAAGCCATATATGCCTAGGTATGGTCTCATTCGCAAC TTACGTGATGCGAGTCTCGCTCGGTACGCTTTTGATTT CTATGAGGTTAACTCTAAGACGCCGGTGCGAGCAAG AGAAGCAGTTGCGCAAATGAAGGCCGCTGC--	China	*Zantedeschia aethiopica*	NCBI sequence

Genbank accession	Virus	Sequence	Location	Host	Comment
AJ298035	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGACCAA ATTGAATACCCGTTAAAGCCAATTGTGGAGAATGCTA AACCCACCTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAGAGGCTTATATTGAACTGAGAAACGCAGA AAAACCGTACATGCCTAGGTACGGTCTTATTCGCAAC TTGCGTGATGCTAGTCTTGCCCGGTATGCATTTGATTT CTATGAGGTCAATTCAAAGACACCGGTGCGAGCAAG GGAGGCAGTTGCGCAAATGAAGGCCGCTGC	China	*Zantedeschia aethiopica*	NCBI sequence
AJ298036	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGATCAA ATTGAATACCCGTTGAAACCAATTGTGGAAAACGCAA AACCCACCTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAGAGGCTTATATTGAATTGAGGAATGCGGA GAAACCGTACATGCCTAGGTACGGTCTCATTCGCAAC CTACGTGATGCAAGTCTCGCCCGGTATGCATTTGATT TCTATGAAGTCAATTCTAAAACACCGGTGCGAGCAAG AGAGGCAGTTGCGCAGATGAAGGCCGC-	China	*Zantedeschia aethiopica*	NCBI sequence
AJ305434	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAGCAGTCAA GTTGAATACCC GTTGAGACCCATAGTGGAAAATGCCAAACCAACCTTG CGCCAGATTATGCATCACTTTTCTGACGCAGCAGAGG CCTACATTGAATTGAGAAATGCGGAACGGCCGTACAT GCCTAGGTATGGTCTTATACGCAACTTGCGTGATGCA AGCTTGGCTCGTTATGCGTTTGATTTCTACGAAGTCA ATTCAAAGACGCCCATTCGAGCACGTGAGGCGGTGG CACAAATGACAGCTGCAGC	India	*Zantedeschia aethiopica*	NCBI sequence

Genbank accession	Virus	Sequence	Location	Host	Comment
AJ429525	*Vanilla mosaic virus*	ATGGTATGGTGTATTGAGAATGGAACATCACCCGATA TTAACGGGGCTTGGGTGATGATGGACGGAGATGATC AAATTGAGTACCCGTTAAAACCCATTGTGGAAAATGC TAAACCGACCTTGCGTCAGATCATGCATCACTTCTCT GACGCAGCAGAGGCATATATTGAACTAAGAAATGCG GAGAAACCGTACATGCCTAGGTACGGTTTGATTCGTA ATCTTCGTGATGCAAGTCTCGCTCGGTACGCTTTTGAT TTTTATGAAGTTAATTCAAAAACACCGGTGCGAGCTA GGGAGGCAGTAGCACAGATGAAGGCAGCAGCA	French Polynesia	*Vanilla tahitensis*	NCBI sequence
AJ616719	*Vanilla mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAGATGATCAA ATTGAGTACCCGTTAAAACCCATTGTGGAAAATGCTA AACCGACCTTGCGTCAGATCATGCATCACTTCTCTGA CGCAGCAGAGGCATATATTGAACTAAGAAATGCGGA GAAACCGTACATGCCTAGGTACGGTTTGATTCGTAAT CTTCGTGATGCAAGTCTCGCTCGGTACGCTTTTGATTT TTATGAAGTTAATTCAAAAACACCGGTGCGAGCTAGG GAGGCAGTAGCACAGATGAAGGCTGCAGC	French Polynesia	*Vanilla tahitensis*	NCBI sequence
AJ616720	*Vanilla mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAATGAACAA ATTGAATACCCATTGAAACCAATTGTGGAGAATGCTA AACCAACCTTACGTCAGATTATGCATCACTTTTCTGA CGCAGCAGAGGCATATATCGAACTGAGGAATGCGGA AAAACCGTATATGCCTAGATACGGTTTAATTCGAATC TTCGTGATGCAAGTCTCGCCCGATATGCTTTTGACTTC TACGAGGTCAACTCAAAGACACCAGTTCGGGCTAGG GAGGCAGTAGCACAGATGAAGGCTGCAGC	Cook Islands	*Vanilla tahitensis*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
AM910399	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAATGACCAAA TTGAATACCCGTTAAAACCAATTGTAGAGAATGCAAA ACCCACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAGAA ACCGTACATGCCTAGGTACGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTCGACTTTTAT GAAGTCAACTCAAAGACACCGGTGCGAGCAAGGGAG GCAGTTGCGCAGATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp.*	NCBI sequence
AM910400	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGATCAAA TTGAATACCCGTTAAAGCCGATTGTGGAGAATGCAAA ACCCACTTTGCGTCAGATAATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAAAA ACCATACATGCCTAGGTATGGTCTTATTCGCAACTTAC GTGATGCAAGTCTCGCCCGGTACGCTTTCGACTTCTAT GAAGTTAACTCAAAAACGCCGGTTCGAGCAAGGGAGG CAGTTGCGCAAATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp.*	NCBI sequence
AM910401	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGATCAAA TTGAATACCCGTTAAAGCCGATTGTGGAGAATGCAAA ACCCACCTTGCGTCAGATAATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAAAA ACCGTACATGCCTAGGTACGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTCGACTTTTAT GAGGTCAACTCAAAGACACCGGTGCGAGCAAGGGAG GCAGTTGCGCAGATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp.*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
AJ616721	*Vanilla mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAATGAACA AATTGAATACCCATTGAAACCAATTGTGGAGAATG CTAAACCAACCTTACGTCAGATTATGCATCACTTTT CTGACGCAGCAGAGGCATATATCGAACTGAGGAAT GCGGAAAAACCGTATATGCCTAGATACGGTTTAATT CGCAATCTTCGTGATGCAAGTCTCGCCCGATATGCT TTTGACTTCTACGAGGTCAACTCAAAGACACCAGTT CGGGCTAGGGAGGCAGTAGCACAGATGAAGGCTGC AGC	Cook Islands	*Vanilla tahitensis*	NCBI sequence
AJ628756	*Dasheen mosaic virus*	AACGGGGUUUGGGUGAUGAUGGAUGGAGACACUC AAAUCGAAUACCCGUUGAAACCCAUAGUAGAAAA UGCCAAACCAACCUUGCGUCAGAUUAUGCAUCAC UUUUCUGACGCGGCAGAGGCCUACAUUGAAUUGA GAAACGCAGAAAGACCAUACAUGCCUAGGUAUGG UUUAGUACGCAACUUGCGUGAUGCAAGUCUCGCC CGGUAUGCUUUUGACUUUUAUGAAGUUAACUCAA AAACACCGGUGCGAGCUCGAGAAGCAGUAGCCCA AAUGAAGGCAGCCGC	China	Aroid plant	NCBI sequence
AM910398	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGATCA AATTGAATACCCGTTAAAGCCGATTGTGGAGAATG CAAAACCCACTTTGCGTCAGATAATGCATCACTTTT CTGACGCAGCAGAGGCTTATATTGAACTGAGAAAT GCGGAAAAACCATACATGCCTAGGTATGGTCTTATT CGCAACTTACGTGATGCAAGTCTCGCCCGGTACGCT TTCGACTTCTATGAAGTTAACTCAAAAACGCCGGTT CGAGCAAGGGAGGCAGTTGCGCAAATGAAGGCCGC TGC	Nicaragua	*Xanthosoma sp.*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
AM910402	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGATCAAA TCGAGTACCCGTTAAAACCAATCGTGGAAAATGCAAA ACCCACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCTTACATTGAACTAAGAAATGCGGAGAA ACCGTACATGCCTAGGTACGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTCGACTTTTAT GAGGTCAACTCAAAAACACCGGTGCGAGCAAGGGAG GCAGTTGCGCAGATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp*.	NCBI sequence
AM910403	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGACCAAA TTGAATACCCGTTAAAACCAATTGTAGAGAATGCAAA ACCCACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAGAA ACCATACATGCCTAGGTATGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTTGACTTTTAT GAGGTCAACTCAAAGACACCGGTGCGAGCAAGGGAG GCAGTTGCGCAGATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp*.	NCBI sequence
AM910404	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGACCAAA TTGAATACCCGTTAAAACCAATTGTAGAGAATGCAAA ACCCACCTTGCGTCAGATTATGCATCACTTTTCTGATG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAGAA ACCATACATGCCTAGGTATGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTTGACTTTTAT GAGGTCAACTCAAAGACACCGGTGCGAGCAAGGGAG GCAATTGCGCAGATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp*.	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
AM910405	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGACCAAA TTGAATACCCGTTAAAACCAATTGTAGAGAATGCAAA ACCCACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAGAA ACCATACATGCCTAGGTATGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTTGACTTTTAT GAGGTCAACTCAAAGACACCGGTGCGAGCAAAGGAG GCAGTTGCGCAGATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp.*	NCBI sequence
AM910406	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGACCAAA TTGAATACCCGTTAAAACCAATTGTAGAGAATGCAAA ACCCACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAGAA ACCATACATGCCTAGGTATGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTTGACTTTTAT GAGGTCAACTCAAAGACACCGGTGCGAGCAAGGGAG GCAGTTGCGCAGATGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp.*	NCBI sequence
AM910407	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGACCAAA TTGAATACCCGTTAAAACCAATTGTAGAGAATGCAAA ACCCACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTGAGAAATGCGGAGAA ACCATACATGCCTAGGTATGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTTGACTTTTAT GAGGTCAACTCAAAGACACCGGTGCGAGCAGGGGAG GCAGTTGCGCAGACGAAGGCCGCTGC	Nicaragua	*Xanthosoma sp.*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
AY994104	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAACGATCAAA TTGAATACCCGTTGAAACCAATTGTGGAGAATGCTAA ACCAACTTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCATATATTGAACTGAGAAATGCGGAGAA GCCGTACATGCCTAGGTATGGTCTGATTCGCAATCTTC GTGATGCAAGTCTCGCCCGGTATGCTTTTGACTTCTAT GAGGTCAATTCAAAGACACCGGTTCGAGCTAGGGAAG CAGTAGCACAGATGAAGGCTGCAGC	New Zealand	*Colocasia sp.*	NCBI sequence
AY994105	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGATCAAA TTGAATACCCGTTAAAACCAATTGTGGAAAATGCCAA ACCAACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCATATATTGAACTGAGAAATGCGGAAAA ACCATATATGCCTAGGTATGGTTTAATTCGCAATCTTC GTGATGCAAGTCTCGCCCGGTACGCTTTCGACTTCTAT GAAGTAAATTCAAAGACACCGGTTCGAGCTAGGGAAG CAGTAGCACAGATGAAGGC	New Zealand	*Colocasia sp.*	NCBI sequence
DQ925464	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGGAGCGATCAAA TTGAATACCCGTTAAAACCAATTGTAGAAAATGCAAA ACCCACCTTACGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCTTACATTGAATTGAGGAATGCGGAGAA GCCGTACATGCCTAGGTACGGTCTCATACGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTTGACTTTTAT GAGGTCAATTCTAAAACACCGGTGCGAGCAAGAGAAG CAGTAGCGCAGATGAAGGCCGC	Vietnam	*Colocasia sp.*	NCBI sequence

Genbank accession	Virus	Sequence	Location	Host	Comment
DQ925465	Dasheen mosaic virus	AACGGGGCTTGGGTGATGATGGATGGAAACGACCAAA TTGAATACCCGTTGAAGCCTATTGTGGAAAATGCAAA ACCCACCTTGCGTCAGATAATGCATCACTTTTCTGACG CAGCAGAGGCGTACATTGAACTGAGAAATGCGGAAA AACCGTACATGCCTAGGTACGGTTTAATTCGTAATTTA CGTGATGCAAGTCTCGCCCGGTATGCTTCCGACTTTTA TGAGGTCAATTCAAAAACACCGGTACGAGCAAGGGAG GCAGTTGCGCAGATGAAGGCTGCTGC	Vietnam	*Colocasia sp.*	NCBI sequence
DQ925466	Dasheen mosaic virus	AACGGGGCTTGGGTGATGATGGACGGAAATGAACAA ATTGAATACCCGTTAAAACCAATTGTGGAAAATGCCA AACCAACTTTGCGTCAGATAATGCATCACTTTTCCGAC GCAGCAGAGGCATATATTGAACTGAGGAGCGCGGAG AAACCGTATATGCCTAGGTACGGTCTTATTCGCAACTT ACGTGGTGCAAGCCTCGCCCGCTATGCTTTCGACTTCT GTGAGGTCAATTCGAAAACACCGGTGCGAGCAAGAGA GGCAGTAGCGCAGATGAAGGCCGCTGC	Vietnam	*Typhonium trilobatum*	NCBI sequence
EF199550	Dasheen mosaic virus	AACGGGGCTTGGGTGATGATGGATGGAAATGATCAGA TTGAATACCCGTTAAAACCAATTGTTGAGAATGCAAA ACCAACCTTGCGTCAGATAATGCATCACTTTTCTGACG CAGCAGAGGCTTATATTGAACTAAGAAATGCGGAGAA ACCGTACATGCCTAGGTACGGTCTTATTCGCAACTTAC GTGATGCAAGTCTTGCCCGGTATGCTTTTGACTTCTAT GAGGTTAATTCAAAAACACCGGTGCGAGCAAGGGAG GCAGTCGCGCAGATGAAGGCCGCTGC	China	*Araceae rivieri Durieu*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
EU256492	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAGCAGCCAG GTTGAATACCCGTTGAAGCCCATAGTGGAAAATGCCA AACCAACCTTGCGCCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCCTACATTGAATTGAGAAACGCGGAAC GGCCATACATGCCTAGGTATGGTCTTATACGCAACTTG CGTGATGCAAGCTTGGCTCGTTATGCGTTTGATTTCTA CGAGGTCAATTCAAAGACGCCTATTCGAGCACGTGAG GCAGTGGCACAAATGAAGGCTGCAGC	India	*Amorphophallus paeoniifolius*	NCBI sequence
EU375290	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAGCAGCCAG GTTGAATACCC GTTGAAGCCCATAGTGGAAAATGCCAAACCAACCTTG CGCCAGATTATGCATCACTTTTCTGACGCAGCAGAGG CCTACATTGAATTGAGAAACGCGGAACGGCCATACAT GCCTAGGTATGGTCTTATACGCAACTTGCGTGATGCAA GCTTGGCTCGTTATGCGTTTGATTTCTACGAGGTCAAT TCAAAGACGCCTATTCGAGCACGTGAGGCAGTGGCAC AAATGAAGGCTGCAGC	India	*Amorphophallus paeoniifolius*	NCBI sequence
EU420058	*Dasheen mosaic virus*	AACGGGGUUUGGGUGAUGAUGGAUGGAGACACUCAA AUCGAAUACCCGUUGAAACCCAUAGUAGAAAAUGCA AAACCAACCUUGCGUCAGAUUAUGCAUCACUUUUCU GACGCGGCAGAGGCCUACAUUGAAUUGAGAAACGCA GAGAGACCAUACAUGCCUAGGUAUGGUUUAGUACGC AACUUGCGUGAUGCAAGUCUCGCCCGGUAUGCUUUU GACUUUUAUGAAGUUAACUCAAAAACACCGGUGCGA GCUCGAGAAGCAGUAGCCCAAAUGAAGGCAGCUGC	China	*Pinellia tuberifera*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
FJ160764	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAACGATCAA ACTGAATACCCGTTGAAACCCATTGTGGAAAATGCAA AACCAACCTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAGAGGCATATATTGAACTGAGAAACGCGGA GAAACCGTATATGCCTAGGTACGGTCTTATTCGCAAT TTACGTGATGCAAGTCTCGCACGGTATGCTTTTGACTT CTATGAGGTCAATTCAAAAACACCGGTACGAGCAAG AGAAGCAGTTGCGCAAATGAAGGCCGCTGC	India	*Amorphophallus paeoniifolius*	NCBI sequence
FJ160765	*Dasheen mosaic virus*	AACGGAGCCTGGGTGATGATGGATGGAGAAGAACAA ATTGAATATCCGTTGAAACCAATTGTAGAGAATGCCA AACCAACCTTGCGTCAGATCATGCATCACTTTTCTGA CGCAGCAGAGGCTTACATTGAACTGAGAAACGCGGA ACGTCCATACATGCCTAGGTATGGACTAGTTCGCAAC TTGCGTGATGCGAGCCTTGCACGGTATGCTTTTGATTT CTATGAAGTTAATTCGAAAACACCCGTGCGAGCGCGT GAAGCAGTAGCCCAAATGAAGGCAGCAGC	India	*Colocasia esculenta*	NCBI sequence
HQ207529	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGATGGAACACCAGGGA GTTGAATACCCGTTGAAACCCATAGTGGAGAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGA CGCAGCAGAGGCCTACATTGAATTGAGAAACGCGGA ACGGCCGTACATGCCTAGGTACGGCCTTATACGCAAC TTGCGTGATGCAAGCTTGGCTCGTTATGCGTTTGACTT CTACGAAGTCAATTCAAAGACGCCCATTCGAGCACGT GAGGCAGTGGCACAAATGAAAGCCGC	India	*Amorphophallus paeoniifolius*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
HQ207531	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAACGATCAA ATTGAATACCCGTTAAAACCAATTGCTGAAAACGCAC CAACCTTGCGTCAGATAATGCATCACTTCTCTGACGC GGCAGAGGCATATATTGAACTGAGAAATGCGGAGAA GCCGTACATGCCTAGGTATGGTCTTATTCGCAACTTAC GTGATGCAAGTCTCGCCCGGTATGCTTTTGACTTTTAT GAGGTCAACTCTAAAACACCGGTACGAGCAAGAGAG GCGGTCGCGCAAATGAAGGCCGC-	India	*Amorphophallus paeoniifolius*	NCBI sequence
HQ207533	*Dasheen mosaic virus*	AACGGAGCCUGGGUGAUGAUGGAUGGAGAAGAACA AAUCGAAUAUCCGUUGAAACCAAUAGUGGAAAAUG CCAAACCCACCUUGCGUCAGAUUAUGCAUCACUUUU CUGACGCGGCAGAGGCCUACAUUGAAUUGAGAAACG CGGAACGUCCAUAUAUGCCUAGGUAUGGACUUAUUC GCAACUUGCGUGAUGCGAGCCUCGCACGGAAUGCAU UUGAUUUCUAUGAAGUCAAUUCAAAACCACCUGUGC GAGCGCGUGAAGCAGUUACCCAAAUGAAGGCUGCUG G	India	*Amorphophallus paeoniifolius*	NCBI sequence
HQ207530	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGATCAA ATTGAGTACCCGTTAAAACCAATTGTAGAAAATGCAA AACCAACCTTGCGTCAGATAATGCATCACTTTTCTGA CGCAGCAAAGGCATATATTGAACTGAGAAACGCGGA GAAACCGTACATGCCTAGGTATGGTCTTATCCGCAAC TTACGTGATGCAAGTCTCGCCCGGTATGCTTTTGACTT CTATGAAGTCAACTCTAAAACACCGGTTCGAGCAAGA GAGGCAGTTGCGCAAATGAAGGCCGCTGC	India	*Amorphophallus paeoniifolius*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
HQ207534	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAGCAGTCAA GTTGAATACCCGTTGAAACCCATAGTGGAGAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCCTACATTGAATTGAGAAACGCGGAA CGGCCGTACATGCCTAGGTACGGCCTTATACGCAACT TGCGTGATGCAAGCTTGGCTCGTTATGCGTTTGACTTC TACGAAGTCAATTCAAAGACGCCCATTCGAGCACGTG AGGCAGTGGCACAAATGAAAGCCGC-	India	*Amorphophallus paeoniifolius*	NCBI sequence
HQ207535	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAGCAGTCAA GTTGAATACCCGTTGAAACCCATAGTGGAAAACGCCA AACCAACCTTGCGGCAGATTATGCATCACTTTTCTGA CGCAGCAGAGGCCTACATTGAATTGAGAAACGCGGA ACGGCCGTACATGCCTAGGTATGGTCTTATACGCAAC TTGCGTGATGCAAGCTTGGCTCGCTATGCGTTTGATTT CTATGAAGTCAATTCAAAAACGCCCATTCGAGCACGT GAGGCTGTAGCACAGATGAAAGCTGCAGC	India	*Amorphophallus paeoniifolius*	NCBI sequence
HQ207536	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAGCAGTCAA GTTGAATACCCGTTGAAACCCATAGTGGAGAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCCTACATTGAATTGAGAAACGCGGAA CGGCCGTACATGCCTAGGTACGGCCTTATACGCAACT TGCGTGATGCAAGTCTTGCTCGTTATGCGTTTGACTTC TACGAAGTCAATTCAAAGACGCCCATTCGAGCACGTG AGGCAGTGGCACAAATGAAAGCCGC	India	*Amorphophallus paeoniifolius*	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
HQ207537	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAGCAGTCAA GTTGAATACCCGTTGAAACCCATAGTGGAGAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCCTACATTGAATTGAGAAACGCGGAA CGGCCGTACATGCCTAGGTACGGCCTTATACGCAACT TGCGTGATGCAAGCTTGGCTCGTTATGCGTTTGACTTC TACGAAGTCAATTCAAAGACGCCCATTCGAGCACGTG AGGCAGTGGCACAAATGAAAGCCGC-	India	*Amorphophallus paeoniifolius*	NCBI sequence
HQ207538	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAGCAGTCGA GGTGAATACCCGTTGAAACCCATAGTGGAGAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCCTACATTGAATTGAGAAACGCGGAA CGGCCGTACATGCCTAGGTACGGCCTTATACGCAACT TGCGTGATGCAAGCTTGGCTCGTTATGCGTTTGACTTC TACGAAGTCAATTCAAAGACGCCCATTCGAGCACGTG AGGCAGTGGCACAAATGAAAGCCGC	India	*Amorphophallus paeoniifolius*	NCBI sequence
HQ207539	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGATGGAACACCAGGGA GTTGAATACCCGTTGAAACCCATAGTGGAGAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCCTACATTGAATTGAGAAACGCGGAA CGGCCGTACATGCCTAGGTACGGCCTTATACGCAACT TGCGTGATGCAAGCTTGGCTCGTTATGCGTTTGACTTC TACGAAGTCAATTCAAAGACGCCCATTCGAGCACGTG AGGCAGTGGCACAAATGAAAGCCGC	India	Calla lily	NCBI sequence

Genbank accession	*Virus*	Sequence	Location	Host	Comment
HQ207540	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAACGATCAA AGTGAGTACCCGTTAAAACCAATTGTTGATGATGCAA AACCAACCTTGCGTCAGATAATGCATCACTTCTCTGA CGCGGCAAAGGCATATATTGAATTGAAAAATGCGGA GAAGCCGTACATGCCTAGGTATGGTCTTATTCGCAAC TTACGTGATGCAAGTCTCGCCCGGTATGCTTTTGACTT TTATGAGGTCAACTCTAAAACACCGGTGTACACAAGA GAGGCGGTCGCGCAAATGAAGGCCGC	India	*Amorphophallus paeoniifolius*	NCBI sequence
U00122	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGATGGAAATGATCAA ATTGAATACCCGTTAAAGCCGATTGTGGAGAATGCAA AACCCACCTTGCGTCAGATAATGCATCACTTTTCTGAC GCAGCAGAGGCTTATATTGAACTGAGAAATGCGGAA AAACCATACATGCCTAGGTATGGTCTTATTCGCAACT TACGTGATGCAAGTCTCGCCCGGTACGCTTTCGACTTC TATGAAGTTAACTCTAAAACACCGGTCCGAGCAAGGG AGGCAGTTGCGCAAATGAAGGCCGCTGC	Florida	*Colocasia sp.*	NCBI sequence
U08124	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAACGATCAG ATTGAATACCCGTTAAAACCAATTGTGGAAAATGCAA AACCCACCTTGCGTCAGATAATGCATCACTTTTCTGAC GCAGCAGAGGCTTATATCGAACCGAGGAATGCGGAG AAACCATACATGCCTAGGTATGGTCTCATTCGCAATC TACGTGATGCAAGTCTTGCCCGGTATGCTTTCGACTTC TATGAGGTCAATTCTAAAACACCGGTGCGAGCGAGAG AAGCAGTCGCGCAAATGAAGGCCGC-	Florida	*Colocasia sp.*	NCBI sequence

Genbank accession	Virus	Sequence	Location	Host	Comment
1NZ2	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGATCAA ATTGAATACCCGTTAAAACCAATTGTGGAAAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCATATATTGAACTGAGAAATGCGGAA AAACCATATATGCCTAGGTATGGTTTAATTCGCAATC TTCGTGATGCAAGTCTCGCCCGGTACGCTTTCGACTTC TATGAAGTAAATTCAAAGACACCGGTTCGAGCTAGGG AAGCAGTAGCACAGATGAAGGCTGCTGC	New Zealand	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ
1NZ3	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGATCAA ATTGAATACCCGTTAAAACCAATTGTGGAAAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCATATATTGAACTGAGAAATGCGGAA AAACCATATATGCCTAGGTATGGTTTAATTCGCAATC TTCGTGATGCAAGTCTCGCCCGGTACGCTTTCGACTTC TATGAAGTAAATTCAAAGACACCGGTTCGAGCTAGGG AAGCAGTAGCACAGATGAAGGCTGCTGC	New Zealand	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ
1NZ4	*Dasheen mosaic virus*	AACGGGGCTTGGGTGATGATGGACGGAAATGATCAA ATTGAATACCCGTTAAAACCAATTGTGGAAAATGCCA AACCAACCTTGCGTCAGATTATGCATCACTTTTCTGAC GCAGCAGAGGCATATATTGAACTGAGAAATGCGGAA AAACCATATATGCCTAGGTATGGTTTAATTCGCAATC TTCGTGATGCAAGTCTCGCCCGGTACGCTTTCGACTTC TATGAAGTAAATTCAAAGACACCGGTTCGAGCTAGGG AAGCAGTAGCACAGATGAAGGCTGCTGC	New Zealand	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
NZ1 CLONE 1	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCGTTAAA ACCAATTGTGGAAAATGCCAAACCAACCTTGCGTCAGA TTATGCATCACTTTTCTGACGCAGCAGAGGCATATATTG AACTGAGAAATGCGGAAAAACCATATATGCCTAGGTAT GGTTTAATTCGCAATCTTCGTGATGCAAGTCTCGCCCGG TACGCTTTCGACTTCTATGAAGTAAATTCAAAGACACC GGTTCGAGCTAGGGAAGCAGTAGCACAGATGAAGGCC GCAGC	New Zealand	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ
NZ1 CLONE 2	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCGTTAAA ACCAATTGTGGAAAATGCCAAACCAACCTTGCGTCAAA TTATGCATCACTTTTCTGACGCAGCAGAGGCATATATTG AACTGAGAAATGCGGAAAAACCATATATGCCTAGGTAT GGTTTAATTCGCAATCTTCGTGATGCAAGTCTCGCCCGG TACGCTTTCGACTTCTATGAAGTAAATTCAAAGACACC GGTTCGAGCTAGGGAAGCAGTAGCACAGATGAAGGCC GCAGC	New Zealand	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ
NZ1 CLONE 3	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCGTTAAA ACCAATTGTGGAAAATGCCAAACCAACCTTGCGTCAGA TTATGCATCACTTTTCTGACGCAGCAGAGGCATATATTG AACTGAGAAATGCGGAAAAACCATATATGCCTAGGTAT GGTTTAATTCGCAATCTTCGTGATGCAAGTCTCGCCCGG TACGCTTTCGACTTCTATGAAGTAAATTCAAAGACACC GGTTCGAGCTAGGGAAGCAGTAGCACAGATGAAGGCC GCAGC	New Zealand	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
NZ1-1	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCG TTAAAACCAATTGTGGAAAATGCCAAACCAACC TTGCGTCAGATTATGCATCACTTTTCTGACGCAG CAGAGGCATATATTGAACTGAGAAATGCGGAAA AACCATATATGCCTAGGTATGGTTTAATTCGCAA TCTTCGTGATGCAAGTCTCGCCCGGTACGCTTTC GACTTCTATGAAGTAAATTCAAAGACACCGGTT CGAGCTAGGGAGCAGTAGCACAGATGAAAGCC GCAGC	New Zealand	*Colocasia esculenta*	From first generation propagant plant Sequenced by MJ1/MJ2 Glasshouse grown in NZ
NZ1-2	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCG TTAAAACCAATTGTGGAAAATGCCAAACCAACC TTGCGTCAGATTATGCATCACTTTTCTGACGCAG CAGAGGCATATATTGAACTGAGAAATGCGGAAA AACCATATATGCCTAGGTATGGTTTAATTCGCAA TCTTCGTGATGCAAGTCTCGCCCGGTACGCTTTC GACTTCTATGAAGTAAATTCAAAGACACCGGTT CGAGCTAGGGAAGCAGTAGCACAGATGAAAGC CGCAGC	New Zealand	*Colocasia esculenta*	From first generation propagant plant Sequenced by MJ1/MJ2 Glasshouse grown in NZ
NZ1-3	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCG TTAAAACCAATTGTGGAAAATGCCAAACCAACC TTGCGTCAAATTATGCATCACTTTTCTGACGCAG CAGAGGCATATATTGAACTGAGAAATGCGGAAA AACCATATATGCCTAGGTATGGTTTAATTCGCAA TCTTCGTGATGCAAGTCTCGCCCGGTACGCTTTC GACTTCTATGAAGTAAATTCAAAGACACCGGTT CGAGCTAGGGAAGCAGTAGCACAGATGAAGGC CGCAGC	New Zealand	*Colocasia esculenta*	From first generation propagant plant Sequenced by MJ1/MJ2 Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
NZ1-4	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCGTTA AAACCAATTGTGGAAAATGCCAAACCAACCTTGCGT CAAATTATGCATCACTTTTCTGACGCAGCAGAGGCA TATATTGAACTGAGAAATGCGGAAAAACCATATATG CCTAGGTATGGTTTAATTCGCAATCTTCGTGATGCAA GTCTCGCCCGGTACGCTTTCGACTTCTATGAAGTAA ATTCAAAGACACCGGTTCGAGCTAGGGAAGCAGTA GCACAGATGAAGGCCGCAGC	New Zealand	*Colocasia esculenta*	From first generation propagant plant Sequenced by MJ1/MJ2 Glasshouse grown in NZ
NZ6 CLONE 1	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCGTTA AAACCAATTGTGGAAAATGCCAAACCAACCTTGCGT CAGATTATGCATCACTTTTCTGACGCAGCAGAGGCA TATATTGAACTGAGAAATGCGGAAAAACCATATATG CCTAGGTATGGTTTAATTCGCAATCTTCGTGATGCAA GTCTCGCCCGGTACGCTTTCGACTTCTATGAAGTAA ATTCAAAGACACCGGTTCGAGCTAGGGAAGCAGTA GCACAGATGAAAGCCGCAGC	New Zealand	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ
NZ6 CLONE 2	*Dasheen mosaic virus*	ATGATGGACGGAAATGATCAAATTGAATACCCGTTA AAACCAATTGTGGAAAATGCCAAACCAACCTTGCGT CAGATTATGCATCACTTTTCTGACGCAGCAGAGGCA TATATTGAACTGAGAAATGCGGAAAAACCATATATG CCTAGGTATGGTTTAATTCGCAATCTTCGTGATGCAA GTCTCGCCCGGTACGCTTTCGACTTCTATGAAGTAA ATTCAAAGACACCGGTTCGAGCTAGGGAAGCAGTA GCACAGATGAAGGCCGCAGC	New Zealand	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
NZ6 CLONE 3	*Dasheen mosaic virus*	ATGATGGATGGAAACGATCAAATTGAATACCCATTA AAACCAATTGTGGAGAATGCTAAACCAACTTTGCGT CAGATTATGCATCACTTTTCCGACGCAGCAGAGGCA TATATTGAATTGAGAAACGCAGAAAAACCGTACATG CCTAGGTACGGTTTGATTCGTAATCTTCGTGATGTGA GTCTCGCCCGGTATGCTTTTGATTTTTATGAAGTTAA TTCAAAGACACCGGTTCGAGCTAGGGAAGCGGTGGC ACAGATGAAAGCCGCAGC	New Zealand	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ
NZ6-1	*Dasheen mosaic virus*	ATGATGGACGGAATGGATCAAATTGAATACCCGTTA AAACCAATTGTGGAAAATGCCAAACCAACCTTGCGT CAGATTATGCATCACTTTTCTGACGCAGCAGAGGCAT ATATTGAACTGAGAAATGCGGAAAAACCATATATGC CTAGGTATGGTTTAATTCGCAATCTTCGTGATGCAAG TCTCGCCCGGTACGCTTTCGACTTCTATGAAGTAAAT TCAAAGACACCGGTTCGAGCTAGGGAAGCAGTAGCA CAGATGAAGGCCGCAGC	New Zealand	*Colocasia esculenta*	From first generation propagant plant Sequenced by MJ1/MJ2 Glasshouse grown in NZ
NZ6-2	*Dasheen mosaic virus*	ATGATGGACGGAATGGATCAAATTGAATACCCGTTA AAACCAATTGTGGAAAATGCCAAACCAACCTTGCGT CAGATTATGCATCACTTTTCTGACGCAGCAGAGGCAT ATATTGAACTGAGAAATGCGGAAAAACCATATATGC CTAGGTATGGTTTAATTCGCAATCTTCGTGATGCAAG TCTCGCCCGGTACGCTTTCGACTTCTATGAAGTAAAT TCAAAGACACCGGTTCGAGCTAGGGAAGCAGTAGCA CAGATGAAGGCCGCAGC	New Zealand	*Colocasia esculenta*	From first generation propagant plant Sequenced by MJ1/MJ2 Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
B1*	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCATTGAAACCAATAGTGGAAAACGCAA AACCAACCTTGCGTCAGATTATGCATCACTTCTCTGAC GCAGCAGAGGCATATATTGAACTGAGAAATGCGGAG AAACCATACATGCCTAGGTACGGTTTAATTCGCAATC TTCGTGATGCAAGTCTCGCCCGGTATGCTTTTGACTTT TATGAAGTTAATTCAAAGACACCGGTTCGAGCTAGGG AAGCGGTAGCACAGATGAAGGCTGCTGC	Solomon Islands	*Colocasia esculenta*	PV1/PV2 doublet bottom sequence provided by C. Higgins Glasshouse grown in NZ
B1**	*Dasheen mosaic virus*	AATGGGTCTTGGGTGATGATGGATGGAAACGATCAAA TTGAATACCCATTAAAACCAATTGTGGAGAATGCTAA ACCAACTTTGCGTCAGATTATGCATCACTTTTCCGACG CAGCAGAGGCATATATTGAATTGAGAAACGCAGAAA AACCGTACATGCCTAGGTACGGTTTGATTCGTAATCTT CGTGATGTGAGTCTCGCCCGGTATGCTTTTGATTTTTA TGAAGTTAATTCAAAGACACCGGTTCGAGCTAGGGAA GCGGTGGCACAGATGAAGGCCGCAGC	Solomon Islands	*Colocasia esculenta*	PV1/PV2 doublet bottom sequence provided by C. Higgins Glasshouse grown in NZ
A1B	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCATTGAAACCAATAGTGGAAAACGCAA AACCAACCTTGCGTCAGATTATGCATCACTTCTCTGAC GCAGCAGAGGCATATATTGAACTGAGAAATGCGGAG AAACCATACATGCCTAGGTACGGTTTAATTCGCAATC TTCGTGATGCAAGTCTCGCCCGGTATGCTTTTGACTTT TATGAGGTTAATTCAAAGACACCGGTTCGAGCTAGGG AAGCGGTAGCACAGGTGAAGGCTGCTGC	Solomon Islands	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
A1A	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCATTGAAACCAATAGTGGAAAACGCA AAACCAACCTTGCGTCAGATTATGCATCACTTCTCTG ACGCAGCAGAGGCATATATTGAACTGAGAAATGCGG AGAAACCATACATGCCTAGGTACGGTTTAATTCGCA ATCTTCGTGATGCAAGTCTCGCCCGGTATGCTTTTGA CTTTTATGAAGTTAATTCAAAGACACCGGTTCGAGCT AGGGAAGCGGTAGCACAGATGAAGGCTGCTGC	Solomon Islands	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ
A51B	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCATTGAAACCAATAGTGGAAAACGCA AAACCAACCTTGCGTCAGATTATGCATCACTTCTCTG ACGCAGCAGAGGCATATATTGAACTGAGAAATGCGG AGAAACCATACATGCCTAGGTACGGTTTAATTCGCA ATCTTCGTGATGCAAGTCTCGCCCGGTATGCTTTTGA CTTTTATGAAGTTAATTCAAAGACACCTGTTCGAGCT AGGGAAGCGGTAGCACAGATGAAGGCTGCTGC	Solomon Islands	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ
A51D	*Dasheen mosaic virus*	AATGGGTCTTGGGTGATGATGGATGGAAACGATCAA ATTGAATACCCATTAAAACCAATTGTGGAGAATGCT AAACCAACTTTGCGTCAGATTATGCATCACTTTTCCG ACGCAGCAGAGGCATATATTGAATTGAGAAACGCAG AAAAACCGTACATGCCTAGGTACGGTTTGATTCGTA ATCTTCGTGATGTGAGTCTCGCCCGGTATGCTTTTGA TTTTTATGAAGTTAATTCAAAGACACCGGTTCGAGCT AGGGAAGCGGTGGCACAGATGAAGGCCGCAGC	Solomon Islands	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
1BA	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCATTGAAACCAATAGTGGAAAACGCA AAACCAACCTTGCGTCAGATTATGCATCACTTCTCTG ACGCAGCAGAGGCATATATTGAACTGAGAAATGCGG AGAAACCATACATGCCTAGGTACGGTTTAATTCGAA TCTTCGTGATGCAAGTCTCGCCCGGTATGCTTTTGAC TTTTATGAAGTTAATTCAAAGACACCGGTTCGAGCTA GGGAAGCGGTAGCACAGATGAAGGCTGCTGC	Solomon Islands	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ
1BD	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCATTGAAACCAATAGTGGAAAACGCA AAACCAACCTTGCGTCAGATTATGCATCACTTCTCTG ACGCAGCAGAGGCATATATTGAACTGAGAAATGCGG AGAAACCATACATGCCTAGGTACGGTTTAATTCGCA ATCTTCGTGATGCAAGTCTCGCCCGGTATGCTTTTGA CTTTTATGAAGTTAATTCAAAGACACCGGTTCGAGCT AGGGAAGCGGTAGCACAGATGAAGGCTGCTGC	Solomon Islands	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ
1BC	*Dasheen mosaic virus*	AATGGGGCTTGGGTGATGATGGACGGAAACGATCAA ATTGAATACCCATTGAAACCAATAGTGGAAAACGCA AAACCAACCTTGCGTCAGATTATGCATCACTTCTCTG ACGCAGCAGAGGCATATATTGAACTGAGAAATGCGG AGAAACCATACATGCCTAGGTACGGTTTAATTCGCA ATCTTCGTGATGCAAGTCTCGCCCGGTATGCTTTTGA CTTTTATGAAGTTAATTCAAAGACACCGGTTCGAGCT AGGGAAGCGGTAGCACAGATGAAGGCTGCTGC	Solomon Islands	*Colocasia esculenta*	MJ1/PV1 consensus sequence provided by C.Higgins Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
B1 CLONE 2	*Dasheen mosaic virus*	ATGATGGATGGAAACGATCAAATTGAATACCCATT AAAACCAATTGTGGAGAATGCTAAACCAACTTTGC GTCAGATTATGCATCACTTTTCCGACGCAGCAGAG GCATATATTGAATTGAGAAACGCAGAAAAACCGTA CATGCCTAGGTACGGTTTGATTCGTAATCTTCGTGA TGTGAGTCTCGCCCGGTATGCTTTTGATTTTTATGA AGTTAATTCAAAGACACCGGTTCGAGCTAGGGAAG CGGTGGCACAGATGAAAGCCGCAGC	Solomon Islands	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ
B1 CLONE 3	*Dasheen mosaic virus*	ATGATGGATGGAAACGATCAAATTGAATACCCATT AAAACCAATTGTGGAGAATGCTAAACCAACTTTGC GTCAGATTATGCATCACTTTTCCGACGCAGCAGAG GCATATATTGAATTGAGAAACGCAGAAAAACCGTA CATGCCTAGGTACGGTTTGATTCGTAATCTTCGTGA TGTGAGTCTCGCCCGGTATGCTTTTGATTTTTATGA AGTTAATTCAAAGACACCGGTTCGAGCTAGGGAAG CGGTGGCACAGATGAAGGCCGCAGC	Solomon Islands	*Colocasia esculenta*	Cloned sequence from parent plant Glasshouse grown in NZ
B1-2	*Dasheen mosaic virus*	TATGGGTCTTGAGTGACGATGGATGGAAACGATCA AATTGAATACCCATTAAAACCAATTGTGGAGAATG CTAAACCAACTTTGCGTCAGATTATGCATCACTTTT CCGACGCAGCAGAGGCATATATTGAATTGAGAAAC GCAGAAAAACCGTACATGCCTAGGTACGGTTTGAT TCGTAATCTTCGTGATGTGAGTCTCGCCCGGTATGC TTTTGATTTTTATGAAGTTAATTCAAAGACACCGGT TCGAGCTAGGGAAGCGGTGGCACAGATGAGGTCC GCAGC	Solomon Islands	*Colocasia esculenta*	From first generation propagant plant Sequenced by MJ1/MJ2 Glasshouse grown in NZ

Genbank Accession	Virus	Sequence	Location	Host	Comment
VC3451	*Dasheen mosaic virus*	CTTGGGTGATGATGGACGGAAATGACCAAATTG AATACCCATTTAAAACCAATTGTAGAGAATGCT AAACCCACCTTGCGTCAGATTATGCATCACTTTT CTGACGCAGCAGAGGCTTATATTGAACTGAGAA ATGCGGAGAAACCGTACATGCCTAGGTACGGTT TGATTCGCAACTTACGTGATGCAAGTCTTGCCCG GTATGCTTTCGACTTTTATGAGGTCAACTCAAAG ACACCGGGCCGAGCAAGGGAGGC	Vanuatu	*Colocasia sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
VX3402	*Dasheen mosaic virus*	CTTGGGTGATGATGGACGGAAACGATCAAATTG AATACCCATTTAAAACCAATTGTGGAGAATGCT AAACCAACTTTGCGTCAGATTATGCATCACTTTT CCGACGCAGCAGAGGCATATATTGAATTGAGAA ACGCAGAAAAACCGTACATGCCTAGGTACGGTT TGATTCGTAATCTTCGTGATGTGAGTCTCGCCCG GTATGCTTTTGATTTTTATGAAGTTAATTCAAAG ACACCGGTTCGAGCTAGGGAAGC	Vanuatu	*Xanthosoma sp.*	Sequenced by MJ1/MJ2
VX3406	*Dasheen mosaic virus*	CTTGGGTGATGATGGATGGAAACGATCAAATTG AATACCCATTTAAAACCAATTGTGGAGAATGCT AAACCAACTTTGCGTCAGATTATGCATCACTTTT CCGACGCAGCAGAGGCATATATTGAATTGAGAA ACGCAGAAAAACCGTACATGCCTAGGTACGGTT TGATTCGTAATCTTCGTGATGTGAGTCTCGCCCG GTATGCTTTTGATTTTTATGAAGTTAATTCAAAG ACACCGGTTCGAGCTAGGGAAGC	Vanuatu	*Xanthosoma sp.*	Sequenced by MJ1/MJ2

Genbank Accession	Virus	Sequence	Location	Host	Comment
VX3415	*Dasheen mosaic virus*	CTTGGGTGATGATGGATGGAAACGATCAAATTGAAT ACCCATTTAAAACCAATTGTGGAGAATGCTAAACCA ACTTTGAGTCAGATTATGCATCACTTTTCCGACGCAG CAGACGCATATATTGAATTGAGAAACGCAGAAAAAC CGTACATGCCTAGGTACGGTTTGATTCGTAATCTTCG TGATGTGAGTCTCGCCCGGTATGCTTTTGATTTTTATG AAGTTAATTCAAAGACACCGGTTCGAGCTAGGGAAG C	Vanuatu	*Xanthosoma sp.*	Sequenced by MJ1/MJ2
SAMOA7	*Dasheen mosaic virus*	CTTGGGTGATGATGGACGGAAATGATCAAATTGAAT ACCCGTTAAAACCAATTGTGGAAAATGCCAAACCAA CCTTGCGTCAGATTATGCATCACTTTTCTGACGCAGC AGAGGCATATATTGAACTGAGAAATGCGGAAAAACC ATATATGCCTAGGTATGGTTTAATTCGCAATCTTCGT GATGCAAGTCTCGCCCGGTACGCTTTCGACTTCTATG AAGTAAATTCAAAGACACCGGTTCGAGCTAGGGAAG C	Samoa	*Colocasia esculenta*	Sequenced by MJ1/MJ2
SAMOA5	*Dasheen mosaic virus*	CTTGGGTGATGATGGACGGAAATGATCAAATTGAAT ACCCGTTAAAACCAATTGTGGAAAATGCCAAACCAA CCTTGCGTCAGATTATGCATCACTTTTCTGACGCAGC AGAGGCATATATTGAACTGAGAAATGCGGAAAAACC ATATATGCCTAGGTATGGTTTAATTCGCAATCTTCGT GATGCAAGTCTCGCCCGGTACGCTTTCGACTTCTATG AAGTAAATTCAAAGACACCGGTTCGAGCTAGGGAAG C	Samoa	*Colocasia esculenta*	Sequenced by MJ1/MJ2

Genbank Accession	Virus	Sequence	Location	Host	Comment
SAMOA3	*Dasheen mosaic virus*	CTTGGGTGATGATGGACGGAAATGATCAAATTGAA TACCCGTTAAAACCAATTGTGGAAAATGCCAAACC AACCTTGCGTCAGATTATGCATCACTTTTCTGACGC AGCAGAGGCATATATTGAACTGAGAAATGCGGAA AAACCATATATGCCTAGGTATGGTTTAATTCGCAAT CTTCGTGATGCAAGTCTCGCCCGGTACGCTTTCGAC TTCTATGAAGTAAATTCAAAGACACCGGTTCGAGC TAGGGAAGC	Samoa	*Colocasia esculenta*	Sequenced by MJ1/MJ2
SAMOA1	*Dasheen mosaic virus*	CTTGGGTGATGATGGACGGAATGGATCAAATTGAA TACCCGTTAAAACCAATTGTGGAAAATGCCAAACC AACCTTGCGTCAGATTATGCATCACTTTTCTGACGC AGCAGAGGCATATATTGAACTGAGAAATGCGGAA AAACCATATATGCCTAGGTATGGTTTAATTCGCAAT CTTCGTGATGCAAGTCTCGCCCGGTACGCTTTCGAC TTCTATGAAGTAAATTCAAAGACACCGGTTCGAGC TAGGGAAGC	Samoa	*Colocasia esculenta*	Sequenced by MJ1/MJ2
PC4030	*Dasheen mosaic virus*	CTTGGGTGATGATGGATGGAAACGATCAAATTGAA TACCCAGTTAAAACCAATTGTGGAAAATGCAAAAC CAACCTTGCGTCAGATTATGCATCACTTTTCCGACG CAGCAGAGGCATATATTGAATTGAGAAACGCGGAA AAACCGTACATGCCTAGGTACGGTCTTATTCGCAA CCTACGTGATGCGAGTCTCGCCCGGTATGCTTTTGA TTTTTATGAAGTTAATTCAAAGACACCGGTGCGAG CTAGGGAAGC	Palau	*Colocasia sp.*	Sequenced by MJ1/MJ2

Genbank Accession	Virus	Sequence	Location	Host	Comment
PC4006	*Dasheen mosaic virus*	CTTGGGTGATGATGGATGGAAACGATCAAATCGAA TACCCAGTTAAAGCCAATTGTGGAAAATGCAAAAC CAACCTTGCGTCAGATTATGCATCACTTTTCTGACG CAGCAGAGGCATACATTGAACTGAGAAACGCGGA GAAACCATATATGCCTAGATATGGTCTTATTCGCA ACTTACGTGATGCGAGTCTCGCCCGGTATGCTTTTG ATTTTTATGAGGTTAACTCTAAGACACCGGTGCGA GCAAGAGAAGC	Palau	*Colocasia sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
PC4007	*Dasheen mosaic virus*	TGGGTGATGATGGATGGAAACGATCAAATTGAATA CCCATTAAAACCAATTGTGGAGAATGCTAAACCAA CTTTGCGTCAGATTATGCATCACTTTTCCGACGCAG CAGAGGCATATATTGAATTGAGAAACGCAGAAAA ACCGTACATGCCTAGGTACGGTTTGATTCGTAATCT TCGTGATGTGAGTCTCGCCCGGTATGCTTTTGATTT TTATGAAGTTAATTCAAAGACACCGGTTCGAGCTA GGGAA	Palau	*Colocasia sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
NC4549	*Dasheen mosaic virus*	TGGGTGATGATGGATGGAAACGATCAAATTGAATA CCCATTAAAACCAATTGTGGAGAATGCTAAACCAA CTTTGCGTCAGATTATGCATCACTTTTCCGACGCAG CAGAGGCATATATTGAATTGAGAAACGCAGAAAA ACCGTACATGCCTAGGTACGGTTTGATTCGTAATCT TCGTGATGTGAGTCTCGCCCGGTATGCTTTTGATTT TTATGAAGTTAATTCAAAGACACCGGTTCGAGCTA GGGAA	Nauru	*Colocasia sp.*	Sequenced by MJ1/MJ2

Genbank Accession	Virus	Sequence	Location	Host	Comment
MAX4492	*Dasheen mosaic virus*	TGGGTGATGATGGACGGAAATGATCAAATTGAAT ACCCATTAAAGCCAATTGTGGAGAACGCAAAACC TACTTTGCGTCAGATAATGCATCACTTCTCTGACG CAGCAGAAGCTTACATTGAACTGAGGAACGCGG AAAAACCGTATATGCCTAGGTACGGTCTTATTCG CAACTTACGTGATGCAAGTCTCGCCCGGTATGCT TTCGATTTTTATGAGGTCAATTCTAAGACACCGGT TCGAGCAAGAGAG	Mariana Islands	*Xanthosoma sp*.	Sequenced by MJ1/MJ2
MAX4442	*Dasheen mosaic virus*	TGGGTGATGATGGATGGAAACGATCAAATTGAAT ACCCATTAAAACCAATTGTGGAGAATGCTAAACC AACTTTGCGTCAGATTATGCATCACTTTTCCGACG CAGCAGAGGCATATATTGAATTGAGAAACGCAG AAAAACCGTACATGCCTAGGTACGGTTTGATTCG TAATCTTCGTGATGTGAGTCTCGCCCGGTATGCTT TTGATTTTTATGAAGTTAATTCAAAGACACCGGTT CGAGCTAGGGAA	Mariana Islands	*Xanthosoma sp*.	Sequenced by MJ1/MJ2
MAX4481	*Dasheen mosaic virus*	TGGGTGATGATGGATGGAAACGATCAAATTGAAT ACCCATTAAAACCAATTGTGGAGAATGCTAAACC AACTTTGCGTCAGATTATGCATCACTTTTCCGACG CAGCAGAGGCATATATTGAATTGAGAAACGCAG AAAAACCGTACATGCCTAGGTACGGTTTGATTCG TAATCTTCGTGATGTGAGTCTCGCCCGGTATGCTT TTGATTTTTATGAAGTTAATTCAAAGACACCGGTT CGAGCTAGGGAA	Mariana Islands	*Xanthosoma sp*.	Sequenced by MJ1/MJ2

Genbank Accession	Virus	Sequence	Location	Host	Comment
FPC3783	*Dasheen mosaic virus*	TGGTGCATNGANAACGGAACATCACCTGATATTA ATGGGGATTGGGTGATGATGGACGGAAATGATCA AATTGAGTACCCATTAAAACCAATTGTGGAAAAC GCAAAACCAACCTTGCGTCAGATTATGCATCACT TTTCCGACGCAGCAGAGGCATATATTGAAATGAG AAACGCGGAAAAACCGTACATGCCTAGGTACGGT TTAATACGCAATCTTCGTGATGCAAGTCTCGCCCG GTACGCTTTTGACTTTTATGAAGTTAATTCAAAGA CACCGGTTCGAGCTAGGGAGGCAGTAGCACAGAT GAAGGCCGCAGCAA	French Polynesia	*Colocasia sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
MaC4423	*Dasheen mosaic virus*	TGGGTGATGATGGATGGAAACGATCAAATTGAAT ACCCATTAAAACCAATTGTGGAGAATGCTAAACC AACTTTGCGTCAGATTATGCATCACTTTTCCGACG CAGCAGAGGCATATATTGAATTGAGAAACGCAGA AAAACCGTACATGCCTAGGTACGGTTTGATTCGT AATCTTCGTGATGTGAGTCTCGCCCGGTATGCTTT TGATTTTTATGAAGTTAATTCAAAGACACCGGTTC GAGCTAGGGAA	Mariana Islands	*Xanthosoma sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
MaC4424	*Dasheen mosaic virus*	TGGGTGATGATGGACGGAAATGATCAAATTGAAT ACCCATTAAAGCCAATTGTGGAAAATGCAAAACC CACCTTGCGTCAGATAATGCATCACTTTTCTGACG CGGCAGAGGCTTATATTGAATTGAGAAACGCGGA AAAACCGTACATGCCTAGGTACGGTCTTATTCGC AACTTGCGTGATGCAAGTCTSGCCCGGTATGCTTT CGACTTTTATGAAGTTAATTCAAARACACCGGTG CGAGCAAGGGAG	Mariana Islands	*Colocasia esculenta*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins

Genbank Accession	Virus	Sequence	Location	Host	Comment
FPC3870	*Dasheen mosaic virus*	**T**GGTGCATNGANAACGGAACATCACCTGATATTA ATGGGGATTGGGTGATGATGGACGGAAATGATCA AATTGAATACCCATTAAAACCAATTGTGGAAAAT GCAAAACCAACCTTGCGTCAGATCATGCATCACTT TTCTGACGCAGCAGAGGCATATATTGAAATGAGA AACGCGGAAAAACCGTACATGCCTAGGTACGGTT TAATACGCAATCTTCGTGATGCAAGTCTCGCCCGG TACGCTTTTGACTTTTATGAAGTTAATTCAAAGAC ACCGGTTCGAGCTAGGGAGGCAGTAGCACAGATG AAGGCCGCAGCAA	French Polynesia	*Colocasia sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
FSMX4060	*Dasheen mosaic virus*	TGGTTGATGATGGATGGAAACGATCAAATTGAAT ACCCGTTAGAACCAATTGTAGAAAATGCAAAACC TACCTTGCGTCAAATAATGCATCACTTTTCTGACG CACCAAAGGCTTATATTGAACTGAGAAACGCGGA AAAACCGTACATGCCTAGGTATGGTCTTATTCGCA ATTTACGTGATGCAACTCTCGCCCGGTATGCTTTT GACTTCTACGAAGTTAATTCAAACACACCGGTCCG AGCCAGAGA	Federated States of Micronesia	*Xanthosoma sp.*	Sequenced by MJ1/MJ2
FSMX4039	*Dasheen mosaic virus*	TGGTGCATTGANAACGGAACATCACCCGATATTA ACGGGGCTTGGGTGATGATGGACGGAAATGACCA AATTGAATACCCGTTAAAACCAATTGTAGAGAAT GCAAAACCCACCTTGCGTCAGATTATGCATCACTT TTCTGACGCAGCAGAGGCTTATATTGAACTGAGA AATGCGGAGAAACCGTACATGCCTAGGTACGGTC TTATTCGCAACTTACGTGATGCAAGTCTTGCCCGG TATGCTTTCGACTTTTATGAGGTCAACTCAAAGAC ACCGGTGCGAGCAAGGGAGGCAGTTGCGCAGATG AANGCCGCAGCAA	Federated States of Micronesia	*Xanthosoma sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins

Genbank Accession	Virus	Sequence	Location	Host	Comment
FSMX4041	*Dasheen mosaic virus*	TGGTGCATNGANAACGGAACATCACCCGATATTAA CGGGGCTTGGGTGATGATGGACGGAAATGACCAAA TTGAATACCCATTAAAACCAATTGTAGAGAATGCA AAACCCACCTTGCGTCAGATTATGCATCACTTTTCT GACGCAGCAGAGGCTTATATTGAACTGAGAAATGC GGAGAAACCGTACATGCCTAGGTACGGTCTTATTC GCAACTTACGTGATGCAAGTCTTGCCCGGTATGCTT TCGACTTTTATGAGGTCAACTCAAAGACACCGGTG CGAGCAAGGGAGGCAGTTGCGCAGATGAANGCCG CAGCAA	Federated States of Micronesia	*Xanthosoma sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
CX3664	*Dasheen mosaic virus*	ATGGGACTTGGGTGATGATGGATGGAAACGATCAA ATTGAATACCCATTAAAACCAATTGTGGAGAATGC TAAACCAACTTTGCGTCAGATTATGCATCACTTTTC CGACGCAGCAGAGGCATATATTGAATTGAGAAACG CAGAAAAACCGTACATGCCTAGGTACGGTTTGATT CGTAATCTTCGTGATGTGAGTCTCGCCCGGTATGCT TTTGATTTTTATGAAGTTAATTCAAAGACACCGGTT CGAGCTAGGGAAGCGGTGGCACAGATGAAGCCGC	Cook Islands	*Xanthosoma sp.*	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
CC3688	*Dasheen mosaic virus*	ATGGGGATTGGGTGATGATGGACGGCAAYGATCAA ATTGAATACCCATTGAAACCTATTGTGGAGAATGC CAAACCTACCTTGCGTCAGATAATGCATCACTTTTC TGACGCAGCAGAGGCATATATTGAAATGAGAAACG CGGAGAAACCGTACATGCCTAGGTATGGTTTAATC CGCAATCTTCGTGATGCAAGTCTCGCCCGGTACGC ATTTGATTTCTATGAAGTCAATTCAAAGACRCCGGT TCGAGCCAGGGAAGCAGTAGCACAGATGAAGCCG C	Cook Islands	*Colocasia sp.*	Sequenced by MJ1/MJ2

Genbank Accession	Virus	Sequence	Location	Host	Comment
CX3626	*Dasheen mosaic virus*	ACGGGGCTTGGGTGATGATGGACGGAAATGA TCAAATTGAATACCCGTTAAAACCAATTGTR GAGAATGCAAAACCCACCTTGCGTCAGATTA TGCATCACTTTTCTGACGCAGCAGAGGCTTAT ATTGAACTGAGAAATGCGGAGAAACCGTACA TGCCTAGGTACGGTCTTATTCGCAACTTACGT GATGCAAGTCTTGCCCGGTATGCTTTCGACTT TTATGAGGTCAACTCAAAGACACCGGTGCGA GCAAGGGAGGCAGTTGCGCAGATGAAGCCG	Cook Islands	*Xanthosoma sp*.	Sequenced by MJ1/MJ2 Sequence provided by C.Higgins
CX3627	*Dasheen mosaic virus*	ATGGGACTTGGGTGATGATGGACGGAAACGA TCAAATTGAATACCGATTGAAACCAATTGTG GAGAATGCTAAACCAACTTTGCGTCAGATTA TGCATCACTTTTCTGACGCAGCAGAGGCATAT ATTGAATTGAGAAACGCGGAGAAACCGTACA TGCCTAGGTACGGTTTGATTCGCAATCTTCGT GATGCGAGTCTCGCCCGGTATGCTTTTGATTT TTATGAAGTTAATTCAAAGACACCGGTTCGA GCTAGGGAAGCAGTGGCACAGATGAAGCCG	Cook Islands	*Xanthosoma sp*.	Sequenced by MJ1/MJ2
CX3633	*Dasheen mosaic virus*	ATGGGACTTGGGTGATGATGGACGGAAACGA TCAAATTGAATACCCATTAAAACCAATTGTG GAGAATGCTAAACCAACTTTGCGTCAGATTA TGCATCACTTTTCCGACGCAGCAGAGGCATA TATTGAATTGAGAAACGCAGAAAAACCGTAC ATGCCTAGGTACGGTTTGATTCGTAATCTTCG TGATGCGAGTCTCGCCCGGTATGCTTTTGATT TTTATGAAGTTAATTCAAAGACACCGGTTCG AGCTAGGGAAGCCGTGGCACAGATGAAGCCG C	Cook Islands	*Xanthosoma sp*.	Sequenced by MJ1/MJ2